

Characterisation of Upf1's Role in Maintaining Genome Stability

Nadia Mohd Fadzli

**Thesis submitted for the
Degree of Doctor of Philosophy at the
Department of Biomedical Science,
University of Sheffield**

March 2018

Contents

Abbreviations	7
Acknowledgement	11
Summary	12
1.0 Introduction	13
1.1 Maintenance of genome stability	13
1.2 Check point regulators	16
1.3 Nonsense mediate mRNA decay (NMD)	17
1.4 Helicase activity of Upf1.....	29
1.5 Nuclear Upf1	30
1.6 Aims and objectives of this work	34
2.0 Materials & Methods	35
2.1 Materials	35
2.2 Antibodies	36
2.3 siRNA	39
2.4 Molecular biology (E.Coli)	40
2.5 Biochemistry assays- Protein Purification & Analysis (E.coli)	45
2.6 Strand displacement & Analysis.....	47
2.7 Cell Culture (mammalian cells)	49
2.8 Protein Extraction and Analysis (mammalian cells).....	50
2.9 Sub-Cellular Fractionation (mammalian cells)	51
2.10 Micrococcal Nuclease (MNase) Digestion (mammalian cells)	52
2.11 Immunoprecipitation	53
2.12 Co-Immunoprecipitation	53

2.13	Proteomics (Mass Spectrometer).....	54
3.0	Nuclear Upf1	57
3.1	Introduction	57
3.2	Results	61
3.3	Conclusion	81
4.0	Introduction	89
4.1	General Introduction.....	89
4.2	Results	91
4.3	Discussion.....	111
5.0	Identifying Upf1 protein interacting partners	116
5.1	Introduction	116
5.2	Results	121
5.3	Conclusion	149
6.0	General discussion and future perspectives	160
6.1	Introduction	160
6.2	The role of known Upf1 phosphorylation sites in maintaining genome stability 162	
6.3	Upf1's role in DNA replication and/ or repair might be orchestrated by its ubiquitin ligase activity	166
6.4	Successful establishment of an immuno-precipitation protocol for chromatin- associated and cytoplasmic Flag-tagged UPF1 for mass-spectrometric analysis.....	167
6.5	Upf1 may regulate DNA replication and/or repair through identified protein interactors	169
6.6	Future perspectives.....	176
	Appendix 1: Sequence coverage of trypsin digestion for Upf1 (Isoform 2)	178

Appendix 2: Comparison of protein interactors from different published lists with own MS data 180

Appendix 3: List of significant proteins identified 202

References..... 212

Figures

Figure 1.1: Mechanism of cell cycle regulation	15
Figure 1.2: Schematic diagram of how Upf1 gets phosphorylated	20
Figure 1.3: Schematic representation of the domain arrangement of human Upf1 full length (fl)	24
<i>Figure 1.4: Schematic illustration of SMG5, SMG6 and SMG7</i>	<i>28</i>
Figure 3.1: The Flp-In expression system.....	59
Figure 3.2: Schematic of Upf1 protein domains and the positions of mutations established using the Flp-In expression system.....	60
Figure 3.3: Optimization of doxycycline concentration for expression of flag-Upf1 protein in Flp-In Hela cells	62
Figure 3.4: Overexpression of ChrmB mutant Upf1, but not phosphomimetic or unphosphorylatable mutants results in the expression of the DNA damage marker H2AX.....	65
Figure 3.5: Optimisation of Upf1 siRNA for Flag-Upf1 cell lines	67
Figure 3.6: Cells expressing phospho-mimetic Flag Upf1 ^{EEE} , unphosphorylatable Flag-Upf1 ^{AAA} or Flag-Upf1 ^{ChrmB} are unable to rescue the effect of Upf1 knockdown.....	70
Figure 3.7: Sub-cellular fractionation of Flag-Upf1 ^{Wt} , Flag-Upf1 ^{EEE} , Flag-Upf1 ^{AAA} and Flag-Upf1 ^{ChrmB} cells.	73
Figure 3.8: Protein purification of full-length Upf1 expressed from E.coli.....	77
Figure 3.9: Mutations does not impair the ability of Upf1 to function as a helicase.	80
Figure 4.1: Micrococcal nuclease digest of FLAG-Upf1 ^{Wt} chromatin fraction.	93
Figure 4.2: Immuno-isolation of FLAG proteins by 3x FLAG peptide.....	96
Figure 4.3: Optimisation of Anti-FLAG M2 beads to maximise recovery of FLAG-Upf1 protein	98

Figure 4.4: Schematic of cell growth treatments for each cell line	101
Figure 4.5: Immuno-precipitation of FLAG-Upf1 for triplicate biological samples.....	103
Figure 4.6: Analysis of FLAG-Upf1 protein by SDS-PAGE prior to in-gel digest	108
Figure 4.7: Volcano plot of detected proteins for FLAG-tagged Upf1 samples.....	109
Figure 5.1: Simplified experimental and data analysis workflow that is used in this study..	119
Figure 5.2: Schematic flow of bottom-up MS	120
Figure 5.3: Purity of FLAG-Upf1 ^{Wt} enriched chromatin fraction.....	123
Figure 5.4: Immunoblot for sample quality validation	124
Figure 5.5: SDS-PAGE of Immuno-isolated samples post reduction and alkylation steps.....	133
Figure 5.6: Biological repeats following imputation	134
Figure 5.7: Biological replicates of FLAG-Upf1 ^{Wt} and FLAG-Upf1 ^{ChmB} correlates within and among groups	135
Figure 5.8: Differential interacting proteins between FLAG-Upf1 ^{Wt} and FLAG-Upf1 ^{ChmB}	137
Figure 5.9: Venn diagram displaying the overlap of proteins identified after filtering based on gene ontology	143
Figure 5.10a: Validation of interacting proteins by Co-IP	147
Figure 5.10b: Validation of interacting proteins by Co-IP.....	148
Figure 6.1: Alternative clamp loaders present in cells.....	175

Abbreviations

ACTL6A	Actin-like protein 6A
AmBic	Ammonium bicarbonate
ACN	Acetonitrile
ATM	Ataxia-telangiectasia mutated
ATP	Adenosine triphosphate
ATR	ATM and Rad3 related
BSA	Bovine serum albumin
BME	β -mercaptoethanol
CH domain	Upf1 N-terminal domain
DECID	Decay inducing complex
DEK	DEK proto-oncogene
DMEM	Dulbecco's minimum essential medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DNA-PK	DNA-dependent protein kinase
Dox	Doxycycline
DTT	Dithiolthreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
eIF4E	Eukaryotic initiation factor 4E
EJC	Exon junction complex

ESI	Electrospray ionisation
FA	Formic acid
FBS	Fetal bovine serum
FDR	False discovery rate
FUBP(1/3)	Far Upstream element binding protein (1/3)
HD	Histone decay
HERC2	HECT and RLD domain containing E3 ubiquitin protein ligase 2
HEPES	4-(2-hydroxyethyl)-1piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
HU	Hydroxyurea
IAA	Indole-3-acetic acid
IgG	Immunoglobulin G
kDa	Dalton
LB	Luria-Bertani
MCM6	Minichromosome maintenance complex component 6
MNase	Micrococcal nuclease
mRNA	Messenger RNA
MS	Mass spectrometry
MQ	MaxQuant analysis software/tool
NASP	Nuclear autoantigenic sperm protein
NIB	Nuclear isolation buffer
NMD	Nonsense-mediated mRNA decay
ORC2	Origin recognition complex subunit 2

PARP1	Poly(ADP-ribose) polymerase 1
PBS	Phosphate-buffered saline
PCNA	Proliferating cell nuclear antigen
PIC	Protease inhibitor cocktail
PIKK	Phosphatidylinositol 3-kinase-related kinases
PMSF	Phenylmethanesulfonylfluoride
PoIδ	DNA polymerase δ
PP2A	Protein phosphatase 2A
PSM	Peptide-spectrum matches
PTC	Premature termination codon
PTM	Post-translational modifications
RFC (1-5)	Replication factor C (1-5)
RNA	Ribonucleic acid
RPA	Replication protein A
S42	Upf1 amino acid- serine 42
S1096	Upf1 amino acid- serine 1096
S1116	Upf1 amino acid- serine 1116
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate poly acrylamide gel electrophoresis
siRNA	Small interfering RNA
SLBP	Stem-loop binding protein
SMD	Staufen mediated mRNA decay

SMARCD (1-2)	SWI/SNF related matrix associated, actin dependent regulator of chromatin subfamily D member (1-2)
SMARCE1	SWI/SNF related matrix associated, actin dependent regulator of chromatin subfamily E member 1
s/n	Immuno-precipitation/ immune-isolation's supernatant
SPIN1	Spindlin 1
SQ domain	Upf1 C-terminal
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
SURF	SMG2-Upf1-eRF1-3 complex
T28	Upf1 amino acid- threonine 28
TBS	Tris-buffered saline
TCEP	Tris (2-carboxyethyl) phosphine
TEMED	1,2-bis (dimethylamino)-ethane
TERRA	Telomeric repeat-containing RNA
Top2A	Topoisomerase 2 α
TRIP13	Thyroid hormone receptor interactor 13
Upf (1-3)	Up-frameshift suppressor (1-3)
UBR5	Ubiquitin protein ligase E3 component N-recognin 5
WDHD1	WD repeat and HMG-box DNA binding protein 1
XRN2	5'-3' exoribonuclease 2
γH2AX	Phosphorylated histone 2AX

Acknowledgement

First of all, I'd like to thank God for presenting me this opportunity to be able and study abroad. Never have I imagined or wanted to do a PhD but this journey has truly taught me invaluable things about life and people.

To my supervisor Professor Carl Smythe, you are truly inspirational. Thank you for all the guidance and help when I truly needed them. And, thank you for not giving up on me when I may seem like a lost cause, your patience and understanding is very much appreciated.

To Dr Mark Collins and Dr Adelina E. Acosta Martin, your commitment in helping me get through all the technical aspects of mass spectrometry is truly amazing and I greatly thank you for that. And Adelina, those hugs and pep talks really helped me get through those depressing times. I can't thank you enough.

For providing me with all the help and support throughout the lab, I would like to thank all past and present Smythe lab members who have always been like family. Special thanks to Dr Dave Turton who has always been there for great discussions and taking time to teach me skills that I have now acquired. I'd also like to thank Dr Cyril Sanders and his group in the medical school for welcoming me and taking time to teach techniques that I never thought I could do. Life would be so dull without the "telekung cakap-cakap" girls. Thank you for keeping me sane, all the pep-talks and making sure my tummy is always filled with food.

To my family, thank you for all your prayers and encouragement throughout this time. Last but not least, I'd like to thank my husband, Azmir for letting me be thousands of miles away from him just to let me fulfil my curiosity. I wouldn't think that I'd be able to have a peace of mind without his blessings and support.

Summary

Genome instability arises when the genome maintenance system fails to recognise or repair any damages that occurs in the cell. It is a prerequisite for the development of cancer. The Upf1 gene has been pointed out to have a role in maintaining genome stability although the mechanism is largely unknown. Upf1 is a DNA/RNA helicase that is responsible for several mRNA quality control mechanism in the cell. Extensive study have been done in regards to its role during nonsense-mediated mRNA decay (NMD) which largely, not exclusively takes place in the cytoplasm.

Here I investigate the role of nuclear Upf1 in regards to maintaining the genome integrity via DNA replication and/or DNA repair, a function which is independent of NMD. Two mutant cell lines which have mutations at key phosphorylation sites (T28, S1096 & S1116) and another mutant which has shown to lose its chromatin binding ability (S42) were used in this study to establish the genomic integrity phenotype. I found that only the chromatin binding mutant portrays as a dominant negative and could cause DNA damage. Further analysis on its helicase activity surprisingly showed that neither had impaired ability to displace dsDNA.

Finally continuing analysis using the chromatin binding mutant, I utilised mass spectrometry to identify novel protein interactors that could be responsible to facilitate Upf1 to maintain genome integrity.

CHAPTER 1

1.0 Introduction

1.1 Maintenance of genome stability

When cellular surveillance systems fail to recognise or eliminate errors caused either by an inherited defect or environmental circumstances (such as reactive chemical species or ionising radiation, then this can lead to genome instability. Such instability have been regarded as a prerequisite for the development of cancer or other disorders associated with loss of genomic integrity (Langie et al. 2015; Negrini et al. 2010; Hanahan & Weinberg 2011).

The term genome instability covers a wide range of circumstances where genome content is altered ranging from microsatellite instability (Ellegren 2004) elevated frequency of mutations to the genome through DNA breaks, chromosomal translocations, inversions, deletions, abnormal numbers of chromosomes (aneuploidy) to the extreme circumstance of chromothripsis (Maher & Wilson 2012). The cell possesses multiple surveillance mechanisms that activate DNA damage responses (DDR) to attempt repair to such challenges (Langie et al. 2015).

In eukaryotes, the cell cycle comprises core machinery driven by cyclin-dependent kinases (Malumbres & Barbacid 2005; Malumbres 2014) that ensure timely progression through the various phases. However, circumstances that give rise to the requirement for a DNA damage response , often require a suspension of the cell cycle schedule. Thus cells also operate checkpoints to ensure that the cell cycle may be delayed for sufficient periods of time until the damage is repaired (Nyberg et al. 2002; Mazouzi et al. 2014)

Figure 1.1.

Checkpoint delay can be divided into three phases; initiation, maintenance of replication structures during repair and termination of repair which would allow cell cycle to proceed (Latif et al. 2004). For example, disruption to the replication process can threaten chromosomal stability by interfering with the progression, stability and

proper resumption of replication fork function after replication arrest (Mazouzi et al. 2014). This occurs due to the DNA damage that is generated by errors during DNA replication which is usually referred to as replication stress, in which could affect the progression of replication forks by either making them slower or stalled (Masai et al. 2010).

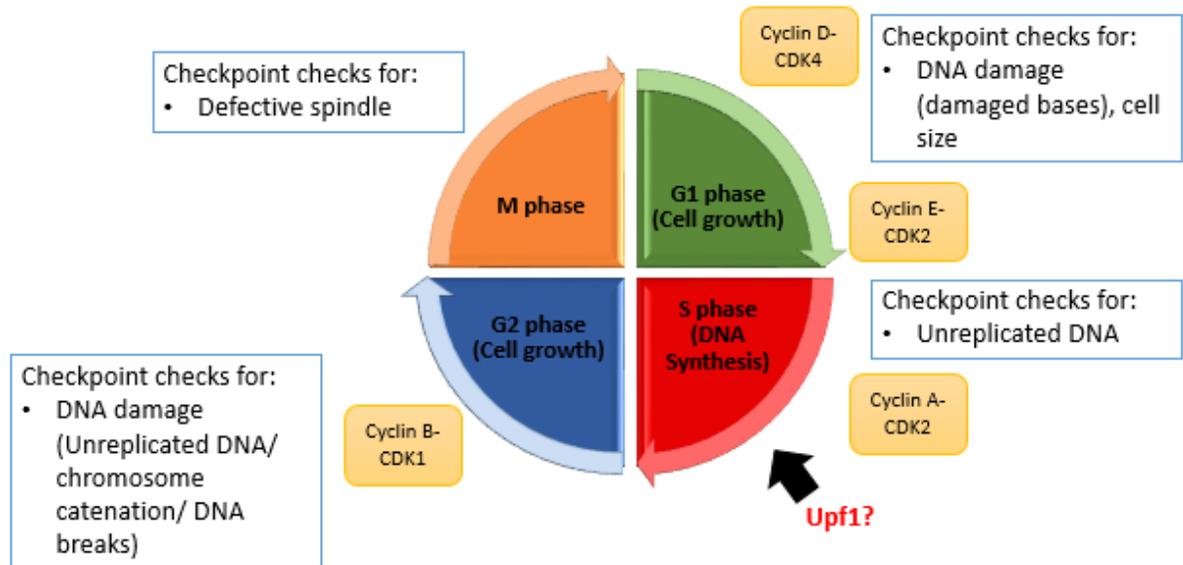


Figure 1.1: Mechanism of cell cycle regulation

The cell cycle is divided into four phases; G1, S (Synthesis), G2 and Mitosis (M). The cell cycle is regulated by cyclins which binds and activate cyclin dependent protein kinases (CDKs) to activate its kinase activity. Cyclin-CDK complexes functions to prevent the progression of the cell through the cell cycle. However should damage is detected, then the cell cycle would either come to a halt or slows down until damage is repaired or cells go into apoptosis. The cyclin-CDK complexes in the respected phases are as indicated in figure.

1.2 Check point regulators

The initial detection of DNA damage in interphase involves members of the phosphatidylinositol-kinase related protein kinases (PIKK) family (Lovejoy & Cortez 2009). In mammals, the PIKK family comprises six proteins which, in addition to the founding member phosphatidylinositol 3-kinase, includes mammalian target of rapamycin (mTOR), ATM (ataxia telangiectasia mutated), ATR (ATM and Rad3-related), Suppressor with Morphogenetic effects on Genitalia (SMG1) and DNA-dependent protein kinase catalytic subunit (DNA-PKcs) (Melero et al. 2014).

Once DNA damage or DNA replication stress is detected, two PIKKs, predominantly responsible for the initiation of a DDR, ATR and ATM, acts as initial signal transducers. In appropriate circumstances, the relevant PIKK is recruited to a damage-containing site, where it is responsible for the phosphorylation of numerous substrates that recruit repair proteins (effectors) and disseminate information more widely throughout the cell via the activation of intermediate signal transduction pathways (transducers) (Hiom 2005; Mazouzi et al. 2014; Gaillard et al. 2015; Maréchal & Zou 2013). Both of these kinases have a strong preference to phosphorylate serine or threonine residues followed by glutamine (the S/T-Q motif) (Stokes et al. 2007; Wagner et al. 2016).

Although both ATR and ATM respond to DNA damage and replication stress, their DNA-damage specificities are distinct and their functions are not redundant (Maréchal & Zou 2013). Interestingly however, there is cross talk between the two signalling pathways, and each can localise to DNA damage sites primarily involving the other, and modulate the DDR should one kinase be compromised (Chanoux et al. 2009; Maréchal & Zou 2013). ATR is activated in response to a wide range of aberrant DNA structures such as UV-induced DNA lesions and replication stress induced by nucleotide depletion which predominantly involves the generation of excess single-stranded DNA (ssDNA) (Lovejoy & Cortez 2009; Cimprich & Cortez 2008; Hall-Jackson et al. 1999). On the other hand, ATM binds to, and is activated by double-stranded breaks (Lambert & Carr 2005;

Lee & Paull 2005). While both are involved in the DDR pathway, ATR is an essential gene in mouse (Cortez 2001; Chanoux et al. 2009). ATR is found to be activated during every cell-cycle, indicating that the aberrant DNA structures to which it responds are generated normally during DNA replication, and when disrupted, leads to either cell cycle arrest or apoptosis (Cortez 2001; Brown & Baltimore 2003).

Double stranded breaks (DSB) also activates DNA-PK, which recruits it to the DNA lesion site. The most dominant DSB repair pathway in human is non-homologous end joining (NHEJ) where it mediates the realignment of broken DNA strands without the need to have a template (Davis & Chen 2013). DNA-PK is recruited to DSBs by interacting to the end binding heterodimer Ku70/80, where its function and activation is dependent on Ku70/80 (Smith & Jackson 1999; Lovejoy & Cortez 2009). The Ku heterodimer is able to localise to DSBs within seconds of its creation and that step initiates NHEJ, where it serves as a scaffold to recruit other NHEJ factors to the damage site which includes DNA-PK (Mari et al. 2006; Davis & Chen 2013).

1.3 Nonsense mediate mRNA decay (NMD)

Mature mRNA formation involves 5' capping, intron removal, 3'-end formation and potentially RNA editing or post-transcriptional nucleotide modifications (Park & Maquat 2013). At a DNA level, premature termination codons (PTCs) can occur as a consequence of mutations due to errors in the gene sequence, or frameshift mutations. In RNA, PTCs can arise from splicing or transcriptional errors (Nicholson & Mühlemann 2010).

An mRNA containing a PTC has a premature translation termination codon upstream of the normal termination codon. PTC containing mRNAs have the potential to give rise to a C-terminal truncated or non-functional proteins (Yamashita et al. 2001; Okada-Katsuhata et al. 2012; Durand et al. 2016). Cells contain a mechanism to detect PTCs in mRNAs and bring about the destruction of such an RNA. This process is referred to as nonsense mediated decay.

Eukaryote mRNA undergoes splicing where introns are removed from exons as part of the formation of a mature, functional mRNA (Lodish et al. 2000). Nonsense-mediated mRNA decay (NMD) is a surveillance mechanism that identifies and degrades mRNAs that harbours PTCs (Baker & Parker 2004; Maquat 2006; Smith & Baker 2015). A mechanism that mainly but not exclusively takes place in the cytoplasm, NMD detects errors that could have been generated from nonsense mutations or errors in cellular processes such as inefficient splicing of pre-mRNAs (Baker & Parker, 2004; Glavan, Behm-Ansmant, Izaurralde, & Conti, 2006; Hug, Longman, & Ceres, 2015; Isken & Maquat, 2008).

The first mRNA identified to be destroyed by NMD was discovered in *Saccharomyces cerevisiae* (*S.cerevisiae*) (Chang & Kan 1979) and later the significance of NMD in humans was established in patients with β^0 -thalassemia (Losson & Lacroute 1979; Kinniburgh et al. 1982; Maquat et al. 1981; Baker & Parker 2004). β^0 -thalassemia is a blood disorder that is characterised by the absence of β -globin which might arise due to mutations that affects the transcription, RNA processing mRNA stability or RNA transport from nucleus to cytoplasm (Maquat et al. 1981). Patients would have severe anaemia and hepatosplenomegaly (Origa n.d.).

Therefore, failure to eliminate PTCs has the potential to give rise to abnormal mRNAs and subsequently truncated proteins that could lead to disease. It has been reported that 30% of all mutations causing human disease generate mRNAs with PTC namely including β -thalassemia, Duchenne Muscular Dystrophy, Cystic fibrosis, Charcot-Marie-Tooth Disease type 2 and Ullrich's disease (Isken & Maquat 2008; Maquat 2006; Brogna & Wen 2009; Bono 2014; Sun & Maquat 2000; Sun et al. 1998; Culbertson 1999).

1.3.1 Assembly of NMD surveillance machinery

In the normal translation process, newly synthesized mRNA is initially capped by the cap binding complex, CBP 80/20, at the 5' end. Newly synthesised mRNA retains residual protein components at the junction of contiguous exons formed as a result of splicing the exon junction complex (EJC). During a pioneer round of translation, a normal mRNA will have the cap replaced with the eukaryotic initiation factor, eIF4E, (Choe et al.

2012; Isken et al. 2008; Hwang et al. 2010; Lejeune et al. 2002) and the components of the EJC will be displaced from the maturing mRNA. However if a pioneering ribosome arrests at a premature termination codon more than ~50–55 nucleotides upstream of a post-splicing exon-exon junction during a pioneer round of translation, then this will result in the initiation of NMD (Sun & Maquat 2000; Le Hir et al. 2000; Maquat 2005).

CBP80, unlike eIF4E, is able to interact with both Upf and SMG NMD factors suggesting a mechanism by which NMD is initiated (Lejeune et al. 2002; Kashima et al. 2006b). In this circumstance, a multi-protein complex termed SURF (Hwang et al. 2010) comprising the proteins Smg1-Upf1-eRF1-eRF3, then interacts with Upf2, which in turn bridges the Upf3-EJC complex (the latter comprising Y14-Magoh, eIF4AIII and MLN51) (**Figure 1.2**). The assembly of the SURF complex in apposition to elements of the EJC initiates structural rearrangements and further protein recruitment resulting in RNA destruction. Details regarding the effect of SURF-Upf2-Upf3-EJC interaction will be discussed in section 1.3.3 in more detail.

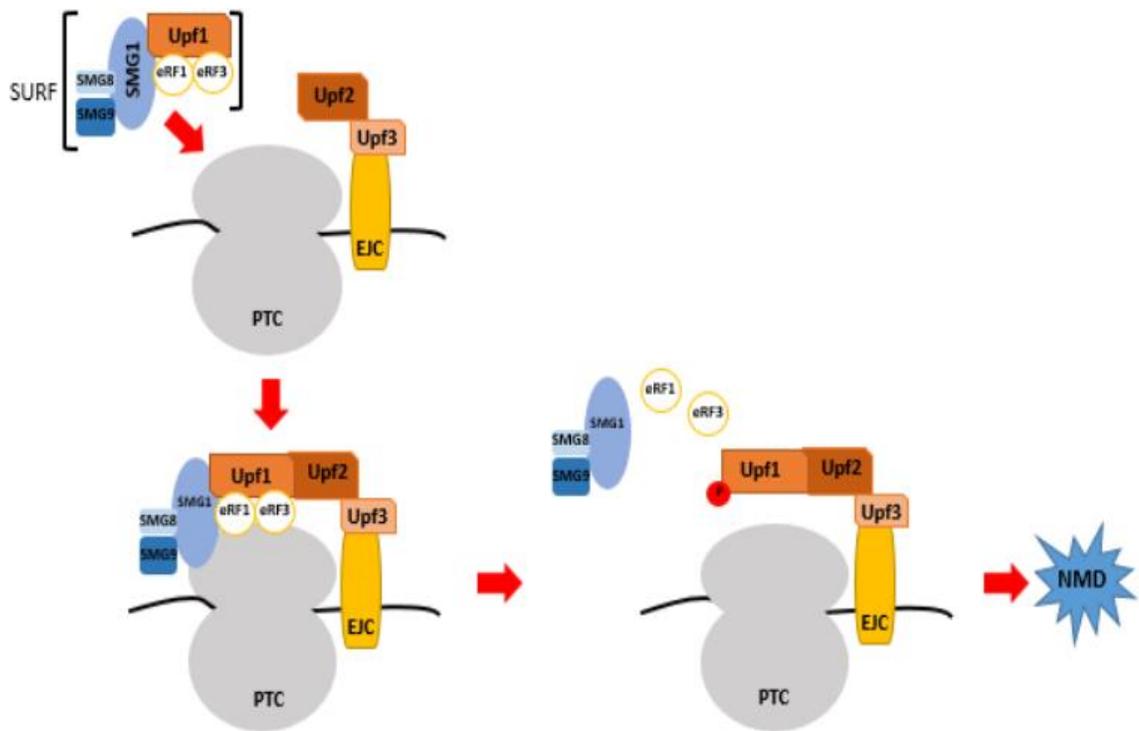


Figure 1.2: Schematic diagram of how Upf1 gets phosphorylated

Once a PTC is detected, the SURF complex, which comprises of Upf1, SMG1, SMG8, SMG9, eRF1 and eRF3 gets assembled at site. Binding of Upf2 at the CH domain of Upf1 bridges SURF complex to Upf2-Upf3-EJC which triggers the phosphorylation of Upf1 and forming the decay inducing complex (DECID). Phosphorylated Upf1 then disengages with the rest of the SURF complex which then triggers a signal transduction pathway that would lead to NMD.

1.3.2 Up-frameshift 1 (Upf1) as a central player of NMD machinery

The human Upf proteins (Upf1, Upf2 and Upf3) works as the core components in NMD where they form a complex at the PTC recognition site. Deletion of any of Upf proteins in yeast results in failure to elicit an NMD response (Matia-González et al. 2013). Additionally, mutations in Upf1 has been shown to stabilise nonsense- containing mRNAs (Leeds et al. 1991; Anders et al. 2003). However of these three Upf proteins, Upf1 has proven to be the most conserved (Maquat, 2006) across species.

Upf1 functions as a principle component of a macromolecular complex that recognises and degrades mRNA that contains PTCs during NMD (Azzalin 2012; Isken & Maquat 2008). Conserved from yeast to man (Kaygun & Marzluff 2005a), Upf1 encodes a DNA/RNA-dependent ATPase helicase (Cheng et al. 2007; Bhattacharya et al. 2000) of the AAA helicase super-family (Maquat, 2006).

hUpf1, which consists of 1,118 amino acids, was first identified based on sequence homology with an orthologue in *S.cerevisiae* (Applequist et al. 1997; Perlick et al. 1996). Studies previously undertaken in yeast had uncovered several characteristics of Upf1. It was shown to be able to bind both RNA and ssDNA, displayed both RNA- and DNA-dependent ATPase activity as well as possessing a 5'-3' helicase activity (Czaplinski et al. 1995; Weng et al. 1996b; Pal et al. 2001).

Structurally, Upf1's helicase domain (HD) is composed of a double RecA-like domain where the ATPase activity is located, and is flanked by two other external domains (**Figure 1.3**). The N-terminal region of the protein is rich in cysteines and histidines, and is thus termed the CH domain, whilst the C-terminal is rich in serine-glutamine clusters (SQ domain). A partial, as well as a complete crystal structure of the protein has been reported. Upf1 is known to be a phospho -protein. High throughput phospho-proteomics have validated at least two of the four C- terminal sites which were suggested to be phosphorylated on the basis of sequence specific phospho-specific antibodies (Ohnishi et al. 2003; Fiorini et al. 2013; Kashima et al. 2006c; Chakrabarti et

al. 2014; Kashima et al. 2006b). Additionally, hUpf1 is known to be phosphorylated on Thr28, based on recognition by phospho-specific antibodies (Imamachi et al. 2012).

The CH domain is reported to act in a cis-inhibitory fashion on the ATPase and unwinding activities of the Upf1 helicase core. In the “inactive” or less active state, the CH domain forces Upf1 to bind to RNA in a clamped conformation which prevents its function. However during NMD, this is relieved once Upf2 binds to the CH domain. This interaction triggers a displacement of the CH domain from its original position which, in turn, causes a conformational change that promotes Upf1 helicase activation (Fiorini et al. 2013; Chakrabarti et al. 2011; Chamieh et al. 2008).

Although the molecular function of the SQ domain is less understood, it also has been shown to have an inhibitory effect on the helicase activity of Upf1 (Fiorini et al. 2013). Deletion at two SQ motifs (S1096 & S1116) was shown to cause an impairment to the binding sites of SMG5-7 which has shown to be important for sequential association and dissociation of NMD factors during specific mRNA decay of Upf1 (Chakrabarti et al. 2014). Thus both N and C-terminal are important Upf1 regulatory domains which act not only to suppress helicase and ATPase activity, but also to modulate Upf1-protein (Chakrabarti et al. 2014).

Upf1 also mediates two other mRNA degradation processes besides the canonical NMD pathway. However, both pathways does not require the presence of Upf2. The first is Staufen mediated mRNA decay (SMD). This mRNA decay pathway actually competes for Upf1 with the NMD pathway to degrade mRNA (Park & Maquat 2013). This pathways degrades mRNAs that harbours a STAU1-binding site at the 3'-untranslated region (Park et al. 2013). STAU1 and its other paralog STAU2 are involved in the microtubule-dependent transport of RNAs to dendrites and plays a crucial role in the formation and maintenance of dendritic hippocampal neurons that are required for memory and learning (Vessey et al. 2008; Park et al. 2013).

Secondly, together with Stem-Loop Binding Protein (SLBP), Upf1 promotes the degradation of histones mRNAs at the end of S-phase or when DNA replication is

inhibited (Isken & Maquat 2008; Azzalin & Lingner 2006a). Histone decay (HD) is a process responsible for maintaining the balance between histone supply and the requirements of newly synthesized DNA (Müller et al , 2007). When DNA replication is blocked, HD results in the destruction of histone mRNAs (Müller et al , 2007) as part of the intra S-phase checkpoint. The efficiency of histone mRNA degradation is reduced following mutation and/ or knockdown of Upf1, suggesting a requirement for Upf1 in the proper functioning of the S-phase checkpoint (Kaygun & Marzluff 2005a). Dysregulation of histone supply during S phase also results in chromosome abnormalities and genome instability (Gunjan & Verreault 2003; Keall et al. 2007).

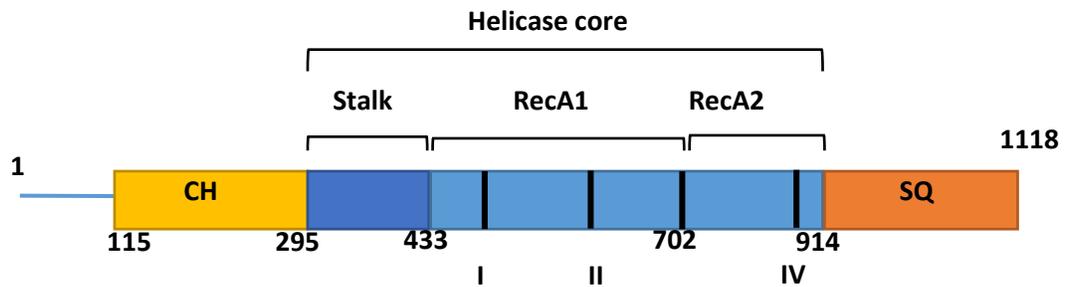


Figure 1.3: Schematic representation of the domain arrangement of human Upf1 full length (fl)

The helicase region (Upf1DCH) features two RecA domains (in light blue) and the “stalk” (in dark blue). Upstream of the helicase region, Upf1 contains a regulatory domain: the CH-domain (in yellow) and another downstream of the helicase region, the SQ-domain (in orange).

1.3.3 SMGs key to Upf1 phosphorylation-dephosphorylation

The PIKK family kinases preferentially phosphorylate at sites containing an S/T-Q motif. Interestingly Upf1 contains 28 SQ/TQ sites on both CH and SQ domains; 14 of which are clustered within the last 88 amino acids (Kim et al. 1999; Page et al. 1999; Chawla & Azzalin 2009). Studies have shown that phosphorylation-dephosphorylation cycles are important for Upf1 NMD function in both *Caenorhabditis elegans* (*C.elegans*) and mammals (Ohnishi et al. 2003; Durand et al. 2016; Anders et al. 2003; Okada-Katsuhata et al. 2012; Hug et al. 2015; Kervestin & Jacobson 2012). In one study, Okada-Katsuhata et al., (2012) had identified the phosphorylation sites that were required for the dephosphorylation of Upf1. They further tested the two sites T28 and S1096 by mutating them to alanine which caused an overexpression of Upf1 and caused the conformation of the protein in an unphosphorylatable state at either the N- or C-terminal. What they found was phosphorylation of these sites was required for the stabilisation of PTC-containing β -globin mRNA.

Melero et al. (2014) reported that SMG1 forms a complex with SMG8 and SMG9, forming SMG1c. This is the complex that is responsible for the phosphorylation of Upf1. Two sites that have been identified to be key phosphorylation sites for NMD on Upf1; T28 and S1096 (Okada-Katsuhata et al. 2012). Interestingly, an additional C-terminal site (S1116) has recently been shown to be phosphorylated in vivo (Turton, Beniston & Smythe, unpublished). This site has also recently been shown to be important for SMG7 binding when it forms a dimer with SMG5 (Chakrabarti et al. 2014). Phosphorylation at these sites then create binding platforms for other SMGs (SMG5, SMG6 & SMG7) that, together with the catalytic PP2A phosphatase subunit (see below), are thought to dephosphorylate Upf1 (Nicholson et al. 2010; Chiu et al. 2003; Cho et al. 2013; Bono 2014; Yamashita et al. 2001; Okada-Katsuhata et al. 2012; Isken et al. 2008).

The current model for NMD is that the SURF complex, which comprises SMG1, SMG8, SMG9, Upf1 and the eukaryotic release factors 1 and 3 (eRF1 & eRF3) assemble at the ribosome when a PTC is encountered. When the PTC and EJC is within 50-55 nucleotides of each other, then an interaction between Upf1 (SURF) complex and Upf2-

Upf3-EJC can occur. This then triggers the phosphorylation of Upf1 and formation of the decay inducing complex (DECID) (Deniaud et al. 2015; Bono 2014; Okada-Katsuhata et al. 2012). Deniaud et al. (2015) proposed that through this interaction ie Upf1-Upf2-Upf3-EJC, the interaction between Upf1 and the other SURF complex components are disrupted and is thought to trigger phosphorylation of Upf1 which in turn starts a cascade of events for mRNA degradation (**Figure 1.1**).

The phosphorylation of Upf1 subsequently promotes the association of other phospho-binding proteins, namely SMG5, SMG6 and SMG7 which sequentially alter the conformation of the substrate complex and induce dephosphorylation of Upf1. That these three SMG proteins are required for the dephosphorylation of Upf1 is demonstrated by observations that mutants in each result in the accumulation of hyperphosphorylated Upf1 in cells and block NMD (Anders et al. 2003; Page et al. 1999). Jonas, Weichenrieder, & Izaurralde (2013) also reported a similar finding where they observed that NMD was strongly inhibited when all three SMGs were absent (Ottens et al. 2017). Although the three SMG proteins share some similar features and are evolutionary conserved, their functions are not redundant (Page et al. 1999; Gatfield et al. 2003).

SMG6's catalytic activity lies on its C-terminal PiIT N-terminal (PIN) domain) (**Figure 1.4**) which is shown to usually be present in proteins with nuclease activity (Matelska et al. 2017; Huntzinger et al. 2008; Clissold & Ponting 2000). SMG6's association with Upf1 at the phospho-T28 site induces endonucleolytic activity, cleaving the 5' of PTC containing mRNA. This nuclease activity is important for NMD as over-expression of a nuclease inactive SMG6 mutant have shown to partially inhibit NMD in a dominant negative manner (Glavan et al. 2006).

Although SMG5 also possesses a PIN domain at its C-terminal, it however lacks key catalytic residues that SMG6 have that would allow it to function as a nuclease (Huntzinger et al. 2008; Glavan et al. 2006). Unlike SMG5 and SMG6, SMG7 does not contain a PIN domain. However it does have a conserved 14-4-3 like domain similar to SMG5 and SMG6 (**Figure 1.4**). SMG7 in its structure also has a low-complexity structure

at its C-terminus that is thought to be important for P-body localisation where NMD may occur, at least in some systems (Chakrabarti, Bonneau, Schüssler, Eppinger, & Conti, 2014; Unterholzner & Izaurralde, 2004; Glavan et al., 2006; Jonas et al., 2013; Nicholson & Mühlemann, 2010).

SMG7-mediated mRNA degradation involves recruitment of the decapping enzyme DCP2 which removes guanosine cap structure leaving it accessible for the 5'-3' exonuclease XRN1 (Nicholson et al. 2010; Jonas et al. 2013). SMG5 has no degradative activity on its own but is required for the nuclease function of SMG7. The SMG7 G100E mutant in the study was unable to bind to SMG5, but could still trigger mRNA degradation (Jonas et al. 2013). They also showed that the activity of SMG7 depended on its interaction with Upf1 because mutations of the 14-3-3- like domain strongly impaired Upf1 binding. Despite these insights, the function of SMG5 is still poorly understood.

Whilst SMG6 is responsible for the nuclease activity associated with the N-terminal end of the Upf1 containing complex, both SMG5 and SMG7 are required for efficient dephosphorylation of residues at the C- terminus of Upf1 in addition to degradation of mRNA occurring at that end of Upf1. Pull downs of the SMG5-7 complex have shown that the complex interacts with both Upf1 and protein phosphatase 2 (PP2A) (Chiu et al. 2003), proposing that this complex might be responsible for the dephosphorylation of Upf1 through the recruitment of PP2A to Upf1's S-1096 site (Anders et al. 2003; Okada-Katsuhata et al. 2012; Chiu et al. 2003; Seshacharyulu et al. 2013). SMG5-7 binds to each other creating a heterodimer through their 14-3-3 like domain and this interaction has been reported to be important for functional mRNA degradation to occur (Unterholzner & Izaurralde 2004; Jonas et al. 2013).

Hence, in addition to triggering mRNA target degradation, the association of the SMG5–SMG7 complex or SMG6 with phosphorylated UPF1 evokes a cascade of events, including the recruitment of PP2A, UPF1 dephosphorylation, and recycling of NMD factors to initiate new rounds of NMD (Nicholson et al. 2010; Jonas et al. 2013).

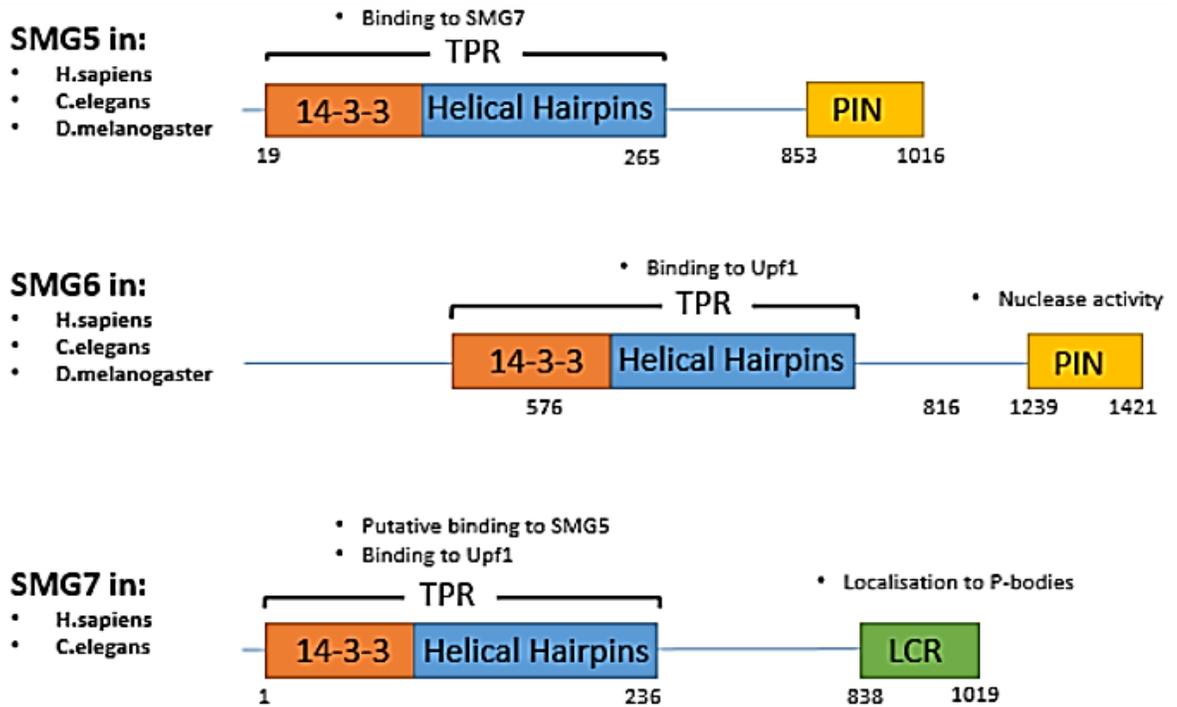


Figure 1.4: Schematic illustration of SMG5, SMG6 and SMG7

The C-terminal of both SMG5 and SMG6 contains the PIN domain; whereas the catalytic centre of SMG5 is mutated, SMG6 has endonuclease activity. LCR, low-complexity region (Nicholson & Mühlemann 2010).

1.4 Helicase activity of Upf1

In order for degrade aberrant mRNA, Upf1 needs to go through the cyclical process of phosphorylation and dephosphorylation. Chakrabarti et al., (2011) stated that in its inactive conformation ie absence of Upf2, the ATPase and helicase activities are low due to the folding of the CH domain towards the helicase domain which contains two RecA domains where RNA and ATP binding occurs . Upf1's helicase domain (residues 295–914) contains seven motifs characteristic of the RNA helicase superfamily 1 (SF1), which use the energy of ATP hydrolysis to rearrange nucleic acid structure or RNA–protein complexes (Chamieh et al. 2008; Cheng et al. 2007).

Notably, the yeast and human UPF1 proteins have been shown to have RNA-dependent ATP hydrolysis and 5'–3' ATP-dependent unwinding activities in vitro, where both activities are important for NMD (Chamieh et al. 2008; Cheng et al. 2007). Studies of strand displacement have shown that, in the absence of ATP, no nucleotide displacement was detected proving that Upf1's helicase activity is dependent on the hydrolysis of ATP (Weng et al. 1996a; Bhattacharya et al. 2000).

Weng et al. (1996b) used a series of mutants harbouring mutations at the CH domain. Some loci demonstrated a 2.5 fold reduction in the helicase activity, which lead to higher levels of nonsense-containing mRNAs compared to wild-type strains. This proves the importance of Upf1 helicase activity for normal, efficient, NMD functionality.

1.5 Nuclear Upf1

The exact sub cellular location where NMD is undertaken is highly debated (Isken & Maquat 2008; Nicholson et al. 2012). However, Upf1 is known to be present in both nucleus and cytoplasm where it has been shown that Upf1 shuttles between nucleus and cytoplasm (Iborra et al. 2004; Mendell et al. 2002). More recently, it has become apparent that ~4% of total cellular Upf1 is associated with chromatin, and that amount of chromatin bound Upf1 increases ~ four-fold when cells are exposed to the replication inhibitor, hydroxyurea (Azzalin & Lingner 2006b).

More recently, it has been shown that siRNA-mediated knock-down of Upf1 causes chromosome and chromatid breakage, the accumulation of telomeric repeat-containing RNAs (TERRA) and also elevated levels of phosphorylated H2AX, a DNA damage marker (Azzalin, Reichenbach, Khoriali, Giulotto & Lingner 2007; Azzalin & Lingner 2006b).

Interestingly, Upf1's nuclear role appears to be separate from its role in NMD as depletion of other NMD factors such as Upf2 does not have the same effect on genome stability. Although depletion of Upf2 does not trigger a DNA damage response, it does impair NMD to the same extent I observed when cells are depleted of Upf1 (Azzalin & Lingner 2006b; Azzalin 2012).

1.5.1 Upf1 and replication

DNA replication is central to cell proliferation which predominantly occur in the S-phase of the cell cycle. It occurs in a semi-conservative manner, involving simultaneous leading and lagging strand synthesis; this results in a chromosome that has two identical sister chromatids in which each one of them contains a newly synthesized strand and also a strand from the parent molecule.

The replication machinery is complex and the replication fork is comprised of numerous components hence tight regulation and monitoring of its mechanism is essential ensuring that new cells generated are without any defects (Masai et al. 2010). DNA replication is initiated by a six-subunit complex, origin recognition complex (ORC)

binding to the replication origin. This complex is able to bind both DNA and binds and hydrolyses ATP which influences its function (Bell 2002). Hydrolysed ORC is then able to bind cell division cycle 6 (Cdc6) protein and later the chromatin licensing and DNA replication factor 1 (Cdt1) protein forming the pre-replication complex (pre-RC). The formation of the pre-RC, occurs during the G₁ -phase is required to load six minichromosome maintenance complex (Mcm2-7) onto the DNA which is removed once MCM is loaded on to DNA (Bell & Dutta 2002). The Mcm2-7 complex is important for both initiation and elongation during DNA synthesis.

Once cells enter the S-phase, the complex now becomes a replisome, responsible to coordinate DNA replication and recruits additional replication factors such as Cdc45p, CDKs, DNA polymerases (α , δ , ϵ) and replication protein A (RPA) (Bell 2002). Other replisome factors such as Claspin, WDHD1, replication factor C (RFC) and fork protection complex are responsible to regulate polymerase functions and coordinate DNA synthesis with the unwinding of the template by Cdc45p-Mcm-GINS complex. The unwinding of the dsDNA can cause replication stress due to the tension that is caused by the introduction of positive supercoils. Topoisomerase is then responsible to remove these supercoils and relax the DNA allowing replication forks to proceed (Mazouzi et al. 2014; Maestroni et al. 2017).

Studies suggested that the presence of Upf1 in the nucleus may be involved in regulating the replication process, potentially by resolving aberrant DNA-DNA, or DNA-RNA hybrids. The basis for the interaction between Upf1 and chromatin is unknown, although previously purification of a DNA helicase activity tightly associated with DNA polymerase δ was previously identified as Upf1 (Carastro et al. 2002), suggesting that Upf1 may be associated with chromatin via a direct interaction with a component of the replication machinery.

Upf1's importance for DNA replication was also highlighted by Azzalin & Lingner (2006b) where they reported that, although Upf1depleted cells start DNA replication, they were unable to complete the process, and had induced an ATR-dependent DNA damage response. More recently, chromatin-bound Upf1 has been shown to be more

directly involved in regulating genome stability (Azzalin & Lingner, 2006; Chawla et al., 2011, Turton & Smythe, unpublished). Upf1-depleted cells show an accumulation of nuclear DNA repair foci, γ H2AX and replication protein A (RPA), a protein that binds preferentially to single stranded DNA (ssDNA) (Azzalin 2012; Azzalin & Lingner 2006b). Damaged DNA results in the phosphorylation of histone H2AX and because phosphorylation of H2AX at ser139 is abundant, occurs quickly and correlates well with double stranded breaks (DSBs), it is a very sensitive marker that may be used to detect DNA damage (Sharma et al. 2012).

Chromatin isolation followed by immunoblotting has established that all Upf (Upf1, Upf2 & Upf3) and SMG (SMG1, SMG5, SMG6 & SMG7) proteins are also associated with chromatin. Interestingly, the amount of chromatin-bound SMG proteins were enriched seven-fold at telomeres compared with ALU-repeat sequences (Azzalin, Reichenbach, Khoraiuli, Giulotto & Lingner 2007). These data imply that aspects of Upf1 helicase function are conserved in its chromatin associated role. Nonetheless an important question is to establish the precise role of this complex on chromatin and whether or not their mechanism of actions resemble those observed in NMD.

1.5.2 UPF1 and telomere maintenance

Maintenance of genomic integrity is achieved by a combination of processes, which includes DNA repair and telomere maintenance. Another role of Upf1 that has recently been highlighted is its involvement during telomere replication and maintenance (Chawla et al., 2011). Upf1 has also been shown to have an interaction with TPP1, a component of the shelterin complex, and localizes at telomeres (Chawla et al. 2011). However, the exact mechanism of how this interaction helps maintain genome integrity at telomeres is unknown.

Telomeres are TTAGGG repeats specific to the ends of human chromosome and contribute to protecting genome stability. The telomere ends with a single-strand overhang. This feature is protected by a complex formed by six proteins specific to telomeres, termed the shelterin complex (De Lange 2005; Chavez et al. 2009). Components of the shelterin complex comprises TRF1, TRF2, POT1, TIN2, TPP1 and Rap1.

One of the roles of the shelterin complex is to protect the ends of telomeres to prevent those ends from triggering a double-stranded DNA damage response, which might result in chromosomal fusion.

The precise role played by Upf1 in ensuring telomere integrity is unclear, however another suggested role for Upf1 is the displacement of TERRA (Chawla et al. 2011; Azzalin, Reichenbach, Khorianti, Giulotto & Azzalin 2007). TERRA is a heterogeneous non-coding RNA that consists of a combination of subtelomeric and telomeric sequences. It has been previously assumed to be transcriptionally silent (Luke & Lingner 2009). TERRA at telomeres has been highlighted to pose as an obstacle for the replication machinery blocking replication fork progression through telomeric DNA.

However, it has also been proposed to help maintain telomere integrity (Feuerhahn et al. 2010; Reig-Viader et al. 2013). A recent study by Chu et al. (2017) also supported this notion whereby they showed in an assay where TERRA was depleted, there was an increase in telomere dysfunction. A FISH assay revealed several telomeric pathologies manifesting loss of integrity after TERRA knockdown. Among them were loss of the (TTAGGG)_n telomeric repeat sequence on both sister chromatids or just on one sister chromatid (heterogeneous), duplications of telomeric repeats at chromosomal ends, insertions of telomeric repeats within internal chromosomal regions, and occasional fusions between sister chromatid telomeres (Chu et al. 2017).

It has also been highlighted that Upf1 and several NMD components ie. SMG1, SMG6 regulates TERRA at telomeres as their depletion caused an increase in telomeres with TERRA foci (Azzalin, Reichenbach, Khorianti, Giulotto & Azzalin 2007).

However, it is known that the ability of Upf1 to ensure telomere integrity requires its ATPase activity as ATP-deficient mutants failed to restore function in transfection experiments (Chawla et al. 2011). The role of Upf1 cyclical phosphorylation in telomere integrity is unknown, and with the exception of the protein TPP1 and telomerase, the identity of Upf1-associated proteins, and the mechanism of action are unknown.

1.6 Aims and objectives of this work

The objective of this research is to investigate the role of nuclear Upf1 in maintaining genome stability, a role which is thought to be independent of its mRNA decay role. Recent discovery regarding the number of nuclear processes that Upf1 might have involved in ranges from DNA replication, DNA damage and repair mechanism and also telomere maintenance.

I have highlighted the canonical role of Upf1, describing the current understanding of how it functions in mRNA surveillance. I have also discussed the components that are involved for that particular process to take place and also linked some of them to the newly discovered function in the nucleus. In addition to that, I have introduced key studies where Upf1 was shown to interact with chromatin and its importance for S-phase progression.

There are several stages of my study in the attempt of uncovering answers regarding Upf1's role in maintaining genome stability. The first is to utilise isogenic HeLa cells, developed in the lab, expressing wild-type and mutant forms of Upf1 where one of which is unable to associate with chromatin (Turton 2014) to observe their effects on genome stability. Secondly, I wished to determine whether such mutants can act as DNA helicases *in vitro*, and thirdly to attempt to establish Upf1's mode of interaction with chromatin by performing a comparative mass spectrometry analysis between a mutant that fails to associate with chromatin and wild-type protein. Using this approach, I hope to gain insight to discovering novel protein interactors that might be of importance for Upf1's role in maintaining genome stability.

CHAPTER 2

2.0 Materials & Methods

2.1 Materials

The chemicals and tissue culture media used were supplied by Sigma Aldrich Limited, unless otherwise indicated. Western blotting detection reagents were obtained from Amersham Biosciences UK Limited. The protein assay reagent (Bradford Assay) and the Mini-Protean II protein gel electrophoresis equipment was purchased from Biorad Laboratories. All molecular reagents such as restriction enzymes, kinases and kits were obtained from New England Biolabs Limited, Promega UK Limited, or Qiagen. Primers required for PCR and antibiotics for tissue culture (Blasticidin, Hygromycin, Doxycycline) were obtained from Invitrogen Limited.

2.2 Antibodies

The below table shows detailed information regarding the primary and secondary used in this project.

Primary Antibody	Species	Raised against	Supplier	Dilution
Anti-Upf1	Sheep affinity purified	Human Upf1	Scottish National Blood Transfusion Service ref:C5B9	WB= 1:5000
Anti-Rent 1	Goat affinity purified	Human Upf1	Bethyl Prod code: (A301-038)	WB=1:1000
Anti-Flag	Mouse affinity purified	FLAG Tag	Sigma (Prod. code: F1804)	WB= 1:5000
Anti-Actin	Mouse affinity purified	Modified β -actin peptide	Sigma (Prod. code: A1978)	WB= 1:10000
Anti-alpha Tubulin	Mouse affinity purified	Rat brain tubulin	Sigma (Prod. code: T8203)	WB= 1:10000
Anti-phospho histone 2AX	Mouse affinity purified	Human H2AX (phosphoser139)	Milipore (Prod. code: 05-636)	WB= 1:1000

Primary Antibody	Species	Raised against	Supplier	Dilution
Anti-nucleolin	Mouse affinity purified	Human nucleolin	Santa Cruz (Prod. code: sc-17826)	WB= 1:10000
Anti-ORC2	Goat affinity purified	Human ORC2	BD Pharmingen (Prod. code: 559266)	WB= 1:1000
Anti-WDHD1	Rabbit affinity purified	Human WDHD1	Abcam (Prod code: ab72436)	WB= 1:2000-10000
Anti-Top2A	Rabbit affinity purified	Human Top2A	Abcam (Prod code: ab12318)	WB= 1:10000-30000
Anti-RFC4	Rabbit affinity purified	Human RFC4	Abcam (Prod code: ab96852)	WB= 1:500-3000
Anti-SMARCE1	Rabbit affinity purified	Human SMARCE1	Abcam (Prod code: ab131328)	WB= 1:1000-10000
Anti-RFC1	Goat affinity purified	Human RFC1	Abcam (Prod code: ab3566)	WB= 1:1000-10000

Secondary Antibody	Species	Raised against	Supplier	Dilution
Anti-sheep HRP	Sheep affinity purified	Sheep IgG	Santa Cruz (Prod code: sc-2473)	WB: 1:5000
Anti-goat HRP	Goat affinity purified	Goat IgG	Santa Cruz (Prod code: sc-2020)	WB: 1:5000
Anti-mouse HRP	Mouse affinity purified	Mouse IgG	Santa Cruz (Prod code: sc-2060)	WB: 1:5000
Anti-rabbit HRP	Rabbit affinity purifies	Rabbit IgG	Santa Cruz (Prod code: sc-2004)	WB: 1:5000

2.3 siRNA

ON-Target plus siRNAs was obtained from Thermo Scientific

Upf1 ON-Target plus siRNAs siRNA sense sequence:

CAGCGGAUCGUGUGAAGAAUU

A non-targeting siRNA by Thermo Scientific was used as a negative control.

2.4 Molecular biology (E.Coli)

2.4.1 Construction of pET11c-Upf1 (AAA & EEE)

A (15-2862) BsiW1/Fse1 Upf1^{res} fragment was digested out from PCDNA5/FRT/TO/CAT-UPF1^{T28A/E:S1096A/E:S1116A/E} (given by Dr Dave Turton) and ligated into pET11c-UPF1Wt (given by Dr Cyril Sanders) vector containing GST tag and TEV cleavage site generating pET11c-Upf1^{T28A/E}. Mutagenesis on residues S1096 and S1116 was performed via PCR (section 2.2.7) and was digested with Fse1/Stu1. Fragment DNA was then ligated to pET11c-Upf1^{T28A/E} to generate pET11c-Upf1^{T28A/E:S1096A/E:S1116A/E}.

2.4.2 Transformation of Upf1Wt, Upf1EEE, Upf1AAA & Upf1S42A

100 µl of competent E.coli cells (DH5α) were added 1µl of sample DNA and allowed to incubate on ice for 20 mins. The mix was then heat shocked at 42°C for 30 seconds and placed on ice for 2 mins before adding outgrowth media (1x luria-Bertani (LB), 0.01M MgCl₂, 25mM KCl, 0.4% glucose) and incubated at 37°C for an hour in an orbital shaker. Transformed cells were plated on M9ZB agar plates and incubated overnight at 37°C. A control plate and different volumes of transformed cells were also plated in parallel.

2.4.3 M9ZB Agar plate

Plates were prepared according to Sanders Lab protocol. Agar plates contained 1x M9 salt (10mM ammonium chloride, 22mM potassium dihydrogen phosphate, 42.2mM disodium hydrogen phosphate), 1x ZB (10g/L tryptone, 5g/L yeast extract, 5g/L sodium chloride), 1x agar 15g/L and 100µg/ml ampicillin.

2.4.4 Isolation of plasmid DNA (QIAprep Spin Miniprep Kit. Prod. code: 27104)

A single bacterial colony was picked and inoculated into 4ml 1x ZB media containing 100mg/L ampicillin. Incubated overnight at 37°C in an orbital shaker. A miniprep was done on the culture according to manufacturer's instructions. 3ml of the culture was transferred to a clean Eppendorf tube and pelleted by centrifugation at 13,000rpm for 1

min (the remaining 1ml was used for glycerol stock). Pellet was then resuspended with 250µl P1 buffer. 250µl P2 was added and mixed thoroughly. Then, N3 buffer was added and mixed thoroughly making the sample cloudy. Sample was centrifuged at 13,000rpm for 10 mins. Supernatant then transferred to a spin column and centrifuged at 13,000rpm for 1 min. Flow-through was discarded, 750µl of PE buffer was added to column and centrifuged at 13,000rpm for 1 min. Flow-through discarded and recentrifuged allowing the excess buffer to wash out. Finally, spin column was transferred to a new Eppendorf tube and 50µl of EB buffer was added, stand for 1 min before centrifuging at 13,000rpm for 1 min. DNA obtained was used for transfection, other molecular biology assays, and sequencing. DNA stored in -20°C.

2.4.5 DNA Sequencing

10µl of 100ng/µl plasmid DNA and 10µl of 1µM primers were sent to the University of Sheffield Medical School Core Genomic Facility for sequencing. Sequencing was done using Applied Biosystems 3730 DNA Analyser according to manufacturing instructions.

2.4.6 Glycerol stocks of transformed cells

A single bacterial colony was picked, inoculated into 4ml 1x ZB media containing selective antibiotics and incubated overnight at 37°C in an orbital shaker. 700µl of the culture was mixed with 300µl of sterile 50% glycerol and stored in -80°C.

2.4.7 Restriction Digest

Digests were done on PCR products and miniprep DNA products. A 50µl reaction volume was prepared using 5U of enzyme per microgram of DNA, enzyme buffers (according to manufacturer's recommendation) and distilled water. 10% of total digest then used for analysis using gel electrophoresis and gel extraction (section 2.4.11).

2.4.8 Ligation

0.1µg of digested plasmid was transferred to new Eppendorf tubes and a range of DNA fragment (insert) in 1:1 to 1:2 ratio was added. To each tube, 1µl of T4 DNA ligase, 2µl of T4 ligase buffer were added and made up with MilliQ H₂O to a final volume of 20µl. Reaction tubes were left overnight at room temperature. Control reactions including plasmid-only and ligase-free were also prepared.

2.4.9 Gel Agarose & Electrophoresis

A 1% agarose gel was used for all molecular work. Mass of agarose was weight and dissolved in TAE while heated until dissolved. 0.1µl/ml ethidium bromide was added while agarose mix is slightly cooled and the mixture was poured into a Biorad cast. To run samples, 5x DNA loading buffer was added to each sample prior to loading into wells. Gel was run at 120V until bands have moved to the center of gel. New England BioLabs (Protein ladder 10-250kDa; Prod. code: P7706) ladder was use as a marker.

2.4.10 DNA Purification (QIAquick PCR purification kit. Prod. code: 28104)

Steps were done according to manufacturer's instructions on DNA fragments that were produced by PCR or following other enzymatic steps. 5x volumes of PB buffer was added to 1 volume of the sample reaction. It is then added to a spin column with a collection tube attached allowing flow-through to be collected. The sample was spun at 13,000rpm for 1 min to allow DNA to bind. Flow-through was discarded and 750µl PE buffer was added to the column. The sample was spun again at 13,000rpm for 1 min, flow-through discarded and re-spun to remove excess wash buffer. The column was transferred to a clean Eppendorf tube and 50µl of EB buffer was added, allowing it to stand for 1 min before eluting the DNA by spinning down at 13,000rpm for 1 min. DNA is stored at -20°C.

2.4.11 DNA extraction & purification (QIAquick gel extraction kit. Prod. code: 28704)

Instead of ethidium bromide, SYBR Safe (Invitrogen) was used to stain DNA allowing visualizing of DNA without UV light. The DNA band of interest was excised from agarose

gel and purified according to manufacturer's instructions. 3x volumes of QG buffer was added to 1 volume of gel weight and incubated for 10 mins at 50°C until the gel is completely dissolved. 1 volume of isopropanol was added to the sample and mixed. Sample was then added to a spin column with a collection tube attached before centrifuging at 13,000rpm for 1 min. Flow-through discarded and 750µl PE buffer was added to the column. The sample was spun again at 13,000rpm for 1 min, flow-through discarded and respun to remove excess wash buffer. Column was transferred to a clean Eppendorf tube and 50µl (plasmids) or 30µl (inserts) of EB buffer was added, allowing it to stand for 1 min before eluting the DNA by spinning down at 13,000rpm for 1 min. DNA is stored at -20°C.

2.4.12 PCR

Primers used for

Forward Primer: Upf2833 (gift from dr Cyril Sanders)

TCCGTCTATGATCGGAGCAGC

Reverse Primer:

UPF1_R_Stu (EEE)

GTTCTCGAGGCCTGATGCATACTGCTCCAGCCCCTCACCCC

StuS1116A (AAA)

ACGTGAGGCCTGATGCATACTGG

	Cycle(s)	Temperature (°C)	Time
1	1	94	30 secs
2	25	94	30 secs
		50	60 secs
		72	15 secs
3	1	72	150 secs

2.5 Biochemistry assays- Protein Purification & Analysis (E.coli)

2.5.1 Bacterial Expression System

A starter culture was prepared by inoculating a colony into 10ml of expression media supplemented with 1x M9 salt, 0.5% glucose, 1mM MgSO₄ and 100µg/ml ampicillin. Bacteria culture was incubated at 37°C in a New Brunswick Innova 4200 incubator shaker for 7-8 hours. 300µl of the starting culture was used to inoculate 1 litre of expression media. 30 litres of culture was prepared per run and 60-90 litres of culture was prepared for each cell line. Culture was incubated overnight at 22°C in a New Brunswick Excell E25 incubator shaker at 190rpm. Cultures were grown until OD₆₀₀ reached an optimal of 0.8-1.0 and was induced with sterile 1M IPTG. Cells were harvest after 6 hours of incubation at 16°C by centrifugation at 4200rpm for 15 mins using a Beckman Coulter J6 MI centrifuge. Supernatant was removed and pellets were resuspended with bacteria wash buffer (20mM Tris pH8.0, 200mM NaCl) before transferring to a clean 1ml centrifuge tube. The suspension was centrifuged at 5000rpm, 4°C for 10 mins. Pellet was snapped freeze with liquid nitrogen and stored at -80°C.

2.5.2 Protein Purification

All procedures were done at 4°C. Pellet obtained from 2.5.1 was weight and calculated. Afterwards resuspended in 1.5ml per gram of low-salt lysis buffer (1x lysis buffer (100mM Tris pH 7.5, 20mM EDTA, 20% glycerol, 20mM DTT); 0.1M NaCl, 1mM PMSF). 1/50 volume of 50mg/ml lysozyme was added to suspension and incubated for 30 mins. An equal volume high-salt lysis buffer (1x lysis buffer, 1.9M NaCl, 1mM PMSF) was added and sonicated for 3 rounds using a Sanyo Soniprep 150 at 40 secs pulses. Suspension was then centrifuged at 11,500rpm for 30 mins. Supernatant was measured and 1/20 volume of polymin P was added dropwise with mixing to precipitate nucleic acids, left 5 mins at 4°C to equilibrate. Then centrifuged for 10 mins at 4°C, 11,500 rpm using Beckman Coulter (Avanti J-26 XPI- rotor F14BA 6x250y). Supernatant measured and 0.291g/ml ammonium sulphate which allows protein precipitation was added. It was incubated

whilst mixing for 10 mins at 4°C and centrifuged for 20 mins at 11,500rpm. The supernatant was discarded and pellet resuspended with 1ml per gram cell paste lysis buffer (1x lysis buffer, 0.25M, 1mM PMSF).

Glutathione sephorase resin was washed 3 times with at least 10x bead volume lysis buffer (1x lysis buffer, 0.25M, 1mM PMSF) and spun down at 1000 xg for 90 seconds. Beads added to the total protein solution and left overnight to incubate at 4°C.

2.5.3 GST-tag Elution

All steps were done at 4°C. Beads from 2.5.2 were collected by spinning down sample at 1000 xg for 90 seconds. Afterwards, beads were washed 3 times 10x beads volume with lysis buffer (1x lysis buffer, 1M NaCl, 1mM PMSF), resuspend with the same buffer and applied to drip column. Buffer was drained and beads were washed with 40x volume lysis buffer (1x lysis buffer, 0.2M NaCl). Buffer was allowed to flow-through. Tagged proteins were then eluted with the GST elution buffer (5mM Tris pH8.8, 300mM NaCl, 1mM EDTA, 5mM DTT, 10% glycerol, 20mM reduced glutathione, pH8). The eluate was collected and checked for the presence of protein using Bio-rad protein assay dye.

2.5.4 Size Exclusion Chromatography (SEC)

A superdex 200 column (16/100) was used to separate monomeric and multimeric proteins and other contaminants. First, column was equilibrated with SEC buffer (25mM Tris-Cl, pH7.5, 200mM NaCl, 5mM DTT, 10% glycerol, 1mM PMSF). Concentrated protein sample was run through column and fractions corresponding to the peaks on chromatograph were analysed on SDS PAGE. After the final run on the SEC, protein fractions corresponding to peak was concentrated to ~1 mg/ml and stored at -80°C.

2.5.5 His-Trap (IMAC) column

The column was equilibrated with both high and low imidazole buffer (2x His Buffer (100mM tris-Cl pH7.5, 1M NaCl, 4mM DTT, 10% glycerol), 20mM or 500mM imidazole). Protein sample was injected into column and eluted by imidazole gradient. Fractions

corresponding to the peaks on chromatograph were collected and analysed on SDS PAGE. Fractions containing Upf1 were pooled.

2.5.6 N- and C-terminal Tag Digest

Digestion of the N-terminal tag was performed after 2.5.3 and before running protein sample through a His-Trap column. N-terminal was digested by adding 4 units/mg of thrombin to protein sample and incubate overnight at 4°C. For C-terminal tag digest, pooled fractions obtained from SEC column were mixed with TEV protease in a dialysis tube and placed in 1L of dialyzing buffer (25mM Tris-Cl pH7.5, 200mM NaCl, 5mM DTT, 10% glycerol, 1mM PMSF) overnight at 4°.

2.5.7 Coomassie Blue Staining

A polyacrylamide gel electrophoresis was run like in 2.8.3. To detect for protein, the gel was submerged in staining solution (1.2% w/v Coomassie brilliant blue, 40% methanol, 10% w/v acetic acid) and destained in destaining solution (7.5% methanol, 7.5% acetic acid).

2.6 Strand displacement & Analysis

Helicase assay and radiolabeling of oligonucleotides were performed as described (George et al. 2009). All reactions were done on ice and using a siliconized tube.

2.6.1 Radio-labelling DNA substrates

A DNA substrate was created using generic oligonucleotides (Sigma). PST55 consists of a 20 bp duplex (PSB) with a 55 base 5' tail. PSB was first end-labeled with ³²P using 1 U/μl T4 polynucleotide kinase (pnk), 5μM PSB oligonucleotide, (γ-³²P) ATP and 1x PNK buffer. Reaction was performed at 37°C for 90 mins and was stopped by heating the reaction to 95°C before immediately placing tubes on ice. Once PSB has been radiolabeled, a reaction containing annealing buffer (1M sodium chloride, 10mM Tris-Cl pH8, 1mM EDTA) and an equimolar of radiolabeled substrate and complementary oligonucleotide

(PST55) was set up to enable the two strands to anneal. Reaction was mixed, spun down and boiled for 5mins before cooled down to 25°C.

2.6.2 Gel purification & substrate quantification

Substrate was ran on a 8% (19:1) polyacrylamide gel (1x TBE running buffer) at 180V. Gels were exposed to film and the band representing substrate was cut out. Gel piece was incubated with probe elution buffer (150mM NaCl or KCl, 1mM tris-Cl pH8, 0.1mM EDTA) at 4°C overnight. Substrate was then filtered using a Spin X filter at 3000g for 3 mins. 1 µl of control sample and purified substrate were spotted on a DEAD81 anion-exchanger chromatography paper. The paper was first washed with 0.5mM sodium phosphate, then 70% ethanol and finally 100% ethanol. Paper was dried and exposed to a phosphor-imaging screen and imaged using Fujifilm FLA-3000. Based on the image obtained, the substrate's concentration was calculated.

2.6.3 Helicase Assay

A dilution series was prepared from 0 to 600nM. Proteins were first diluted with protein dilution buffer (25mM Tris pH7.5, 200mM NaCl₂, 5mM DTT, 10% Glycerol, 1mM PMSF). Subsequent dilutions were done with an addition of 0.1mg/ml BSA. For each protein sample, reaction set up was 1x helicase buffer (0.05mM HEPES pH7.0, 2mM DTT), 0.2nM radiolabeled substrate, 1mM ATP, 10mM magnesium chloride, 100ng/µl BSA, and 75mM NaCl. Tubes were incubated at 37°C for 30 mins. The reaction was stopped by the addition of STOP buffer (100mM EDTA, 50% glycerol, 0.5% (w/v) SDS, 1% (w/v) bromophenol blue), 0.025mg/ml plasmid and 2.5µM T55 oligonucleotide. Samples were run on an 8% polyacrylamide gel at 200V, visualised and quantified following exposure of dried gel to a phosphoimager. A known helicase "dead" (K498A) mutant was used as a control.

2.7 Cell Culture (mammalian cells)

2.7.1 Growth conditions (FLP-IN cell lines)

Stable FLP-IN cell lines containing a FLAG tagged Upf1 which contained a siRNA site and mutated phosphosites were used in this study (a generous gift by Dr D Turton). Cells were maintained in DMEM (Sigma) supplemented with 10% tetracycline free foetal bovine serum (FBS), 1x glutamine and supplemented with or without antibiotics (blasticidin & hygromycin). Cells kept in T25 flasks were incubated at 37°C with 5% CO₂ and split 1:10- 1:20 every three-four days. This step were repeated for all cell FLAG-Upf1^{Wt}, (FLAG-Upf1^{EEE}, FLAG-Upf1^{AAA} and FLAG-Upf1^{ChrmB}).

2.7.2 Doxycycline treatment

The expression of Flag-Upf1 was induced with the addition of 100mg/ml doxycycline to a final concentration of 0.7ug/ml in media for the required length of time.

2.7.3 Overexpression Time Course Assay

In 6-well plates, 2x10⁵ cells (FLAG-Upf1^{Wt}, FLAG-Upf1^{EEE}, FLAG-Upf1^{AAA} and FLAG-Upf1^{ChrmB}) were seeded independently in 2 ml of growth media mentioned in 2.3.1. Two hours after seeding cells, doxycycline was added to the designated wells and left to incubate for the required length of time i.e. 24, 48 or 72 hours. A control well was established on the 72 hours plate with an addition of 2mM hydroxyurea at 48 hours.

2.7.4 Knockdown (KD) Time Course Assay- Electroporation

Protocol used was based on the manufacturer's instructions of the Neon kit (Prod. code: MPK10025) by invitrogen. Cells for each cell line grown in T75 flasks were trypsinised and counted using a haemocytometer. The seeding density for each well of the 6-well plate should be 2.2x10⁵ cells. The calculated volume required was then spun down and washed twice with PBS. Supernatant was discarded and pellet was resuspend with the calculated volume of R-buffer. 10nM of both non-targeting siRNA and Upf1-siRNA were

pipetted into individual Eppendorf tubes. Using the 100µl Neon tip, cells were mixed with the siRNA before electroporating using the Neon transfection system (1400V, 1 pulse, 20ms). Electroporated cells were then added to wells containing 2ml of growth media stated in 2.3.1 before adding doxycycline to (+) doxycycline wells 2 hours after seeding. Plates were incubated for the 24 hours, media was changed to new and allow to incubate for 48, or 72 hours before cell lysis. A control well was established on the 72 hours plate with an addition of 2mM hydroxyurea at 48 hours.

2.8 Protein Extraction and Analysis (mammalian cells)

2.8.1 Whole cell extract

HeLa cells grown in either dishes or plates were placed on ice, media was aspirated and washed twice with ice-cold PBS. Cells were lysed with IPLB (20mM Tris Acetate pH7.5, 270mM sucrose, 1mM EGTA, 1mM sodium orthovanadate, 10mM sodium β-glycerophosphate, 1% Triton X-100, 50mM sodium fluoride, 0.1% β-mercaptoethanol, 0.2mM PMSF, 1x Protease Inhibitor Cocktail, 1µM microcystin) and transferred to new Eppendorf tubes. Lysates were then freeze-thawed 3x before centrifuged at 13,000rpm for 5 mins (Eppendorf 5417R) to pellet insoluble material. Supernatant then transferred to new Eppendorf tube and protein concentration determined. Lysates then stored at -20°C.

2.8.2 Bradford assay

Protein concentration was determined using Bradford dye. The 5x Bio-Rad Protein Assay reagent was diluted to 1x with MiliQ H₂O. Diluted (1:10) protein samples were mixed with 1x Bradford dye and measured at OD₅₉₅. Readings were compared to standard curve established using known BSA standards to calculate protein concentration.

2.8.3 SDS-polyacrylamide gel electrophoresis

Protocol was done according to the method of Laemmli (Laemmli 1970) and performed under denaturing conditions using mini gels. Samples prepared were boiled for 5 mins in 1x SDS loading buffer (50mM Tris-Cl, 100mM DTT, 2% SDS, 10% glycerol, 0.01% (w/v) bromophenol blue) before loading onto gel. Bio-Rad Mini-PROTEAN II apparatus were used and both inner and outer reservoirs were filled with electrophoresis running buffer. Conditions for gel running were 140V, for 60-90 mins.

2.8.4 Immunoblotting

Immunoblotting was performed as described (Burnette 1981). After electrophoresis, gels were removed and overlaid with nitrocellulose paper (Amersham Protran). Gel-nitrocellulose was sandwiched between 2 sheets of Whatmann paper and proteins transferred at 100V, 400mA for 75 mins using the Bio-Rad Mini-PROTEAN II apparatus. The nitrocellulose membrane was blocked in TBS (Tris-CL pH7.5, 150mM NaCl) with 5% of either milk powder (Tesco) or BSA depending on the primary antibody used for 1 hour at room temperature. After blocking, the nitrocellulose membrane was incubated with a primary antibody in either TBS with 5% milk or BSA for 1 hour at room temperature or 4°C overnight. Membranes were washed with 1x TBS 6x for 5 mins and incubated with 5% milk in TBS containing a secondary antibody for 1 hour at room temperature. Membranes were again washed 6x 5 mins with 1x TBS before incubating with ECL reagents according to manufacturer's instructions and chemiluminescence detected using film or Bio-Rad imager.

2.9 Sub-Cellular Fractionation (mammalian cells)

Protocol was done according to Méndez & Stillman (2000). Cells were grown for 3 days after splitting where doxycycline and hydroxyurea (for S-phase enrichment) was added at 24 hours and 48 hours respectively. At 72 hours, cells were washed with PBS, new media was added and was incubated for another 3 hours before harvesting. Cells were then trypsinised, suspended in PBS and total cell number were counted using a

haemocytometer. Afterwards, the suspension was centrifuged at 1000rpm for 5 mins in a Biofuge Primo Heraeus bench top centrifuge (Heraeus #7591 rotor). Pellet was resuspended in Buffer A (10mM HEPES, 10mM KCl, 1.5mM MgCl₂, 0.34M sucrose, 10% glycerol, 1mM DTT, 1x Protease inhibitor cocktail, 0.1mM PMSF) to give final cell concentration 4x10⁴cells/μl. 0.1% TritonX-100 was added to cells and was incubated on ice for 5 mins. Cells were then centrifuged at 1,300g for 5 mins 4°C and supernatant was marked as S1 and pellet as P1. Clearing of S1 was done by centrifugation at 20,000g for 15 mins and supernatant collected was marked as S2 which contains the cytoplasm. The nuclear pellet (P1) was washed with 2x initial volume of Buffer A and incubated on ice for 5 mins before spun down at 1,300g for 5 mins at 4°C. Supernatant was discarded and pellet was resuspended in the same volume of Buffer B (3mM EDTA, 0.2mM EGTA, 1mM DTT, 1x Protease inhibitor cocktail) similar to the initial volume of buffer A used. Nuclear suspension was left to incubate on ice for 30 mins to lyse before being spun down at 1,700g for 4 mins at 4°C. Supernatant was removed and marked as S3, contains the soluble nuclear fraction. Pellet was resuspended with 2x initial volume of Buffer B and incubated on ice for 5 mins before a final spin at 1,700g for 5 mins at 4°C. Supernatant was discarded and the pellet is marked as P3. The volume of all fractions were recorded and all samples were stored in -20°C. However, prior to running on SDS-PAGE, an equal amount of 5x SDS loading buffer was added to P3. Each sample was calculated and normalized to the number of cells. A representative of an equal number of cells was then loaded into each lane.

2.10 Micrococcal Nuclease (MNase) Digestion (mammalian cells)

P3 fraction which is the chromatin fraction obtained during sub-cellular fractionation was used. The pellet was resuspended using Nuclear Isolation buffer (NIB) from (Torrente et al. 2011) (15mM Tris, pH7.5, 15mM NaCl₂, 60mM KCl, 5mM MgCl₂, 1mM CaCl₂ 250mM sucrose, 1x PIC). 10% of the total volume was transferred to a new tube and would be used as a control. The amount of micrococcal nuclease (Sigma- N3755) used was according to manufacturers' recommendations. Once nuclease was added,

sample was incubated at 28°C for 2 mins in the presence of 1mM CaCl₂, vortexed and incubation repeated. 2mM EGTA was added to stop reaction and sample was incubated on ice before spun down at 1,700g for 5 mins at 4°C. Supernatant was transferred to a new tube. Stored both supernatant and pellet at -20°C.

2.11 Immunoprecipitation

Protocol performed as per manufacturer's recommendation. Anti-Flag M2 (M2) beads (Sigma-Aldrich) were washed in TBS (50mM Tris HCl with 150mM NaCl, pH7.4) overnight before use. Chromatin fraction was mixed with of 50% slurry prewashed M2 beads or Protein G beads (incubated with 1µg of antibody of same species). Volume was normalised to highest volume with lysis buffer (50mM Tris Cl pH7.4, 50mM NaCl, 50mM NaF, 1mM EGTA, 1% Triton x-100, 1x Protease Inhibitor Cocktail) and incubated overnight at 4°C with rotation. Immunoprecipitated complexes were collected by centrifugation at 5000g for 30 secs. Beads were washed 3x in 500µl TBS. 30µl of sample loading buffer was added to beads and boiled for 5 mins, centrifuged and step repeated with adding another 20µl of sample loading buffer. Supernatant collected after each spin and used for SDS-PAGE.

2.12 Co-Immunoprecipitation

For both whole cell lysate and chromatin fraction, 2mg of protein lysate were used. The lysates used were the same samples as the ones that were sent for MS analysis. The triplicate protein lysates of both (nuclear fraction) FLAG-Upf1^{Wt} and (WCL) FLAG-Upf1^{ChrmB} were combined and protein concentration measured. 30µl of 50% slurry of prewashed beads were incubated together with lysate overnight. Volume was normalised to highest volume with lysis buffer (50mM Tris Cl pH7.4, 50mM NaCl, 50mM NaF, 1mM EGTA, 1% Triton x-100, 1x Protease Inhibitor Cocktail) and incubated overnight at 4°C with rotation. Immunoprecipitated complexes were collected by centrifugation at 5000g for 30 secs. Beads were washed 3x in 500µl TBS. 50µl of sample

loading buffer was added to beads and boiled for 5 mins, centrifuged and supernatant used for SDS-PAGE.

2.13 Proteomics (Mass Spectrometer)

2.13.1 Reduction and alkylation

1 μ l TCEP was added to lysates as a reduction step and left on heating block (70°C) for 10 mins. Lysates were left to cool down before 2mM IAA was added. Samples were incubated at room temperature for 30 mins in the dark. Reduced and alkylated samples were then ran on SDS-PAGE (precast gel- NuPAGE 4-12% (Prod. Code: NP0321BOX), running buffer: 1x NuPAGE MOPS SDS running buffer (Prod. code: NP0001), protein ladder: SeeBlue Plus2 Prestained Protein Standard) to separate proteins and visualised with coomassie. Gel was scanned to determine the number of fractions that will be cut out. Individual gel fractions were transferred to clean 2.0ml Eppendorf tubes and destained with 1ml 50% ACN.

2.13.2 In-gel digestion peptide extraction

Tubes from 2.9.1 were retrieved and spun down. Supernatant was discarded and 600 μ l of 100mM Ambic (MS grade water need to be used) was added. Tubes were placed on shaker for 10mins at room temperature. Supernatant was discarded and step was repeated before adding 600 μ l of 50% ACN/50% Ambic. Tubes were placed on shaker for 10 mins at room temperature and step was again repeated. 600 μ l of 100% ACN was added to tubes and incubated on shaker at room temperature for 10 mins. Supernatant was discarded and tubes were left uncapped to dry. 1ng/ μ l of Trypsin was added to samples and placed on ice allowing gel to be hydrated before incubating overnight in 37°C incubator.

100 μ l of 100% ACN was added to samples and incubated for 15mins at 37°C. Supernatant transferred to new collecting tubes and 50 μ l of 0.5% formic acid was added. Tubes were incubated at room temperature for 15mins on shaker. 100 μ l of 100% ACN was added to

samples and incubate for 15mins at room temperature. Supernatant transferred to collecting tubes. Repeat washes with 0.5% formic acid and 100% ACN, supernatants transferred to collecting tubes. Final wash with 100µl of 100% ACN and supernatant transferred. Collecting tubes containing peptides were then dried out using a SpeedVac concentrator. 40µl of formic acid was added to dried pellet and sonicate using VWR Ultrasonic bath for 2mins allowing pellet to dissolve completely. Samples were then centrifuged for 5mins at 13,000g. 20µl of sample was then transferred to a clean vial for LC-MS/MS analysis.

2.13.3 Mass Spectrometry

Resuspended peptides prepared in 2.13.2 were injected onto a PepMap100 C18 2 cm x 75µm I.D. trap column (ThermoFisher Scientific) at 5µl/min in 0.1% formic acid, 2% ACN with 45°C in column oven and 6°C in autosampler. Samples were separated over a 125-min gradient of increasing ACN (2.4%-72%), in 0.1% formic acid using a 50cm PepMap100 C18 analytical column (2µm particle size, 100Å pore size 75µm I.D.) at 250nl/min, 45°C.

Using the ThermoFisher-Scientific Orbitrap Elite (with Nanospray Flex Ion ESI), ionisation was carried out at 2.0kV, with the ion transfer capillary at 250°C and S-lens set at 60%. The MS1 spectra was acquired at a resolving power of 60,000 with an AGC (automatic gain control) target value of 1×10^6 ions by the Orbitrap detector, range 350-1850 m/z.

Once the MS1 analysis is done, the top 20 most abundant precursors were selected for data dependent activation (MS2 analysis) using CID (collision induced dissociation) with a 10 ms activation time and an AGC setting of 10,000 ions in the dual cell linear ion trap set at normal scan rate resolution. The precursor ions of single charge were rejected and a 30 sec dynamic exclusion window setting was used after a single occurrence of an ion.

2.13.4 Data analysis

MS data was analysed data using MaxQuant version 1.5.5.1. Data was searched against a human UniProt sequence databases using following search parameters: trypsin with a maximum of 2 missed cleavages, 7 ppm for MS mass tolerance, 0.5 Da for MS/MS mass tolerance, with Acetyl (Protein N-term), Phospho (STY) and Oxidation (M) set as variable modifications and carbamidomethyl (C) as a fixed modification. A protein FDR of 0.01 and a peptide FDR of 0.01 were used for identification level cut offs. Once proteins were identified, quantitative comparisons between samples with statistical analysis were done using Perseus version 1.5.6.0.

CHAPTER 3

3.0 Nuclear Upf1

3.1 Introduction

Upf1 has an essential role in maintaining the stability of the genome which is independent of its cytoplasmic role in NMD (Azzalin & Lingner 2006; Varsally & Brogna 2012). Within the nucleus, a distinct fraction of Upf1 - about 4% of all cellular Upf1 (Azzalin & Lingner 2006b) has been identified. Upf1's role in genome stability has been shown to be separate to its' NMD function as evidenced by the fact that knockdown of the essential NMD factor UPF2 has no effect on genome stability. Nonetheless, a number of components involved in NMD have also been shown to be present in the nucleus (Azzalin, Reichenbach, Khoraiuli, Giulotto & Lingner 2007; Chawla & Azzalin 2009). These includes SMG1, SMG 5/7 and Upf2 which are known to be involved in regulating UPF1 when participating in NMD. In order to begin to delineate aspects of Upf1 function unique to its presence on chromatin, therefore, I wished to find out whether Upf1's nuclear role is disrupted by mutations of specific amino acids within the primary sequence of UPF1 which have been shown to be required for regulation of Upf1 in NMD.

This study utilises the Flp-In system (Life Technologies Corporation 2012) whereby the inducible expression of the gene of interest may be undertaken with the addition of doxycycline to the growth media in isogenic cell lines where the alleles of interest are integrated at the same site within the genome. This is possible using a cell line containing a single FLP recombinase site integrated in the genome, and is based on the binding of tetracycline to the Tet repressor and the derepression of the promoter controlling expression of gene of interest (**Figure 3.1**) Here, doxycycline was used to control the expression of Upf1 for both wild-type and mutants.

The cell lines (gift from Dr Turton) used in this study contain integrated forms of Flag-tagged Upf1, containing silent mutations to ensure siRNA resistance against

sequences targeted at wild type-type Upf1 (**Figure 3.2**). Two out of the three mutants used have mutations at known key phosphorylation sites essential for NMD function (Schweingruber et al. 2013; Nicholson et al. 2014; Durand et al. 2016). The amino acid residues for those sites (T28, S1096 & S1116) using the literature sequence notation for Upf1 isoform 2 were substituted to either an alanine or glutamic acid. Flag-Upf1^{EEE} is a phosphomimetic mutant whereas Flag-Upf1^{AAA} is an unphosphorylatable mutant. The third cell line used has a mutation at the N-terminal of S42 site and was mutated to an alanine residue (herein referred to Flag-Upf1^{ChrmB}). This mutant was previously shown (Turton, D. and Smythe, C., unpublished) not to associate with chromatin.

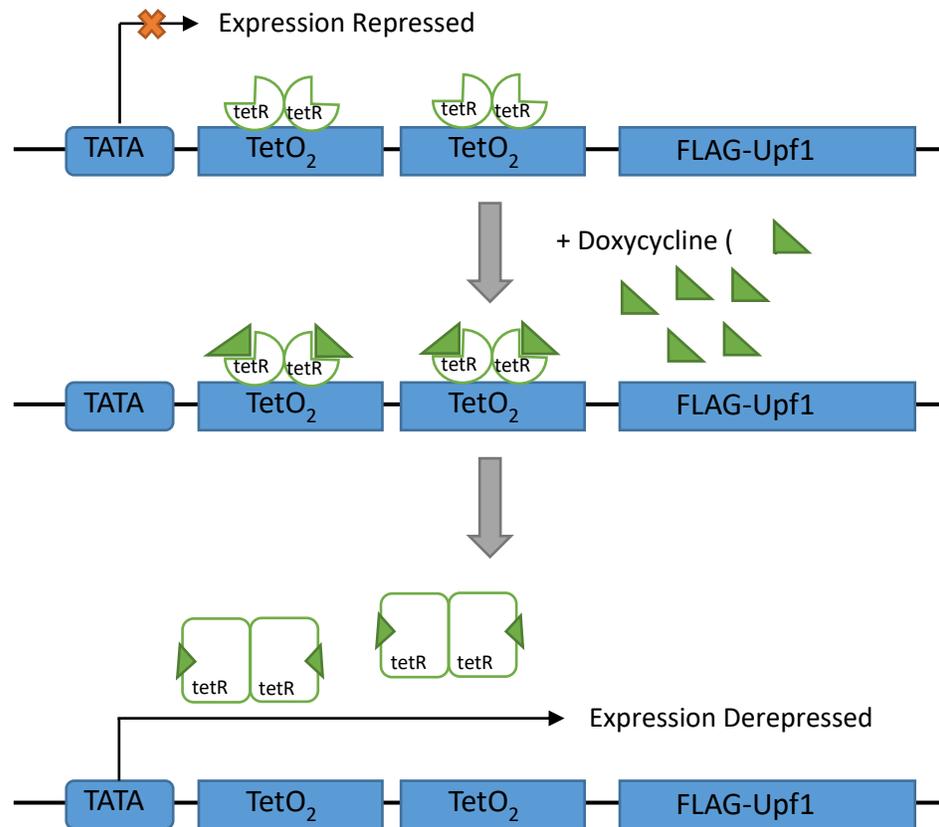


Figure 3.1: The Flp-In expression system.

This system allows the expression of Flag-Upf1 proteins with the addition of doxycycline. Without doxycycline, the tet repressor (tetR) blocks transcription but addition of doxycycline changes its conformation and allows transcription to proceed.

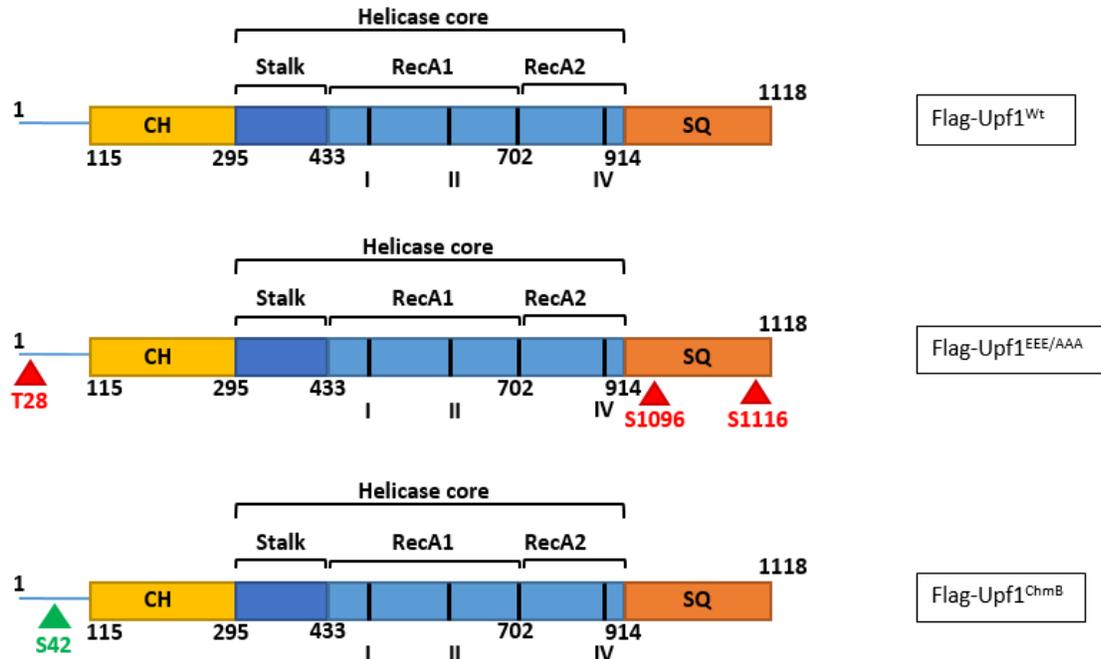


Figure 3.2: Schematic of Upf1 protein domains and the positions of mutations established using the Flp-In expression system

Upf1 is 1118 amino acids in length. It is made up by the Cysteine/Histidine rich domain (CH domain in yellow) at the N-terminal, a Serine-glutamine rich domain (SQ domain in orange) at the C-terminal end and a helicase core that encompasses two recA-like domains typical of the Super family 1 helicases and a “stalk” domain that is essential for NMD and regulates RNA binding affinity in response to ATP (Dehghani-tafti & Sanders 2017). The Ch-domain is structurally similar to the RING domains found in E3 ubiquitin ligase, hence Upf1 is also thought to have ubiquitin ligase activity (Feng et al. 2017). Red triangles represents mutated phosphorylation sites essential during NMD for Upf1 (T28, S1096 & S1116), where changes were made to either alanine (Flag-Upf1^{AAA}) or glutamic acid (Flag-Upf1^{EEE}). The green triangle represents mutated site of the chromatin-binding mutant (S42) where serine was mutated to alanine.

3.2 Results

3.2.1 Quantifying Upf1's expression level in FLP-IN system

Previous work (Turton and Smythe, unpublished) characterized conditions for expression of the unphosphorylatable Upf1 mutant (Flag-Upf1^{AAA}) in Flp-In HeLa cells. To first establish an optimal doxycycline-induced expression level and expression conditions for Flag-Upf1^{EEE} expression in HeLa cells, cell lines containing Flag-Upf1^{EEE} or Flag-Upf1^{Wt} were plated in a 6-well plate and incubated as described in section 2.7. Cells were exposed to a range of doxycycline concentrations for 24 or 48 hours (**Figure 3.3**).

Immunoblotting, using an anti-Flag Antibody, indicated the absence of ectopic expression of mutant Upf1 protein in the absence of any added doxycycline. In contrast, the presence of doxycycline resulted in robust expression of Flag-Upf1^{EEE}, similar to those obtained with the wild type ectopically expressed protein. There was little variation in the levels of expression over the range of doxycycline concentrations used in this experiment. Both Flag-Upf1^{Wt} and Flag-Upf1^{EEE} were expressed at levels similar to those of the endogenous protein, as judged by the limited increase in intensity of the appropriate molecular weight band when probed with anti-Upf1 antibody. (**Figure 3.3**). As higher concentrations above those recommended by the manufacturer (1.0 µg/µl) did not significantly increase the levels of protein expression, the concentration of doxycycline chosen for future experiments was 0.7 µg/µl.

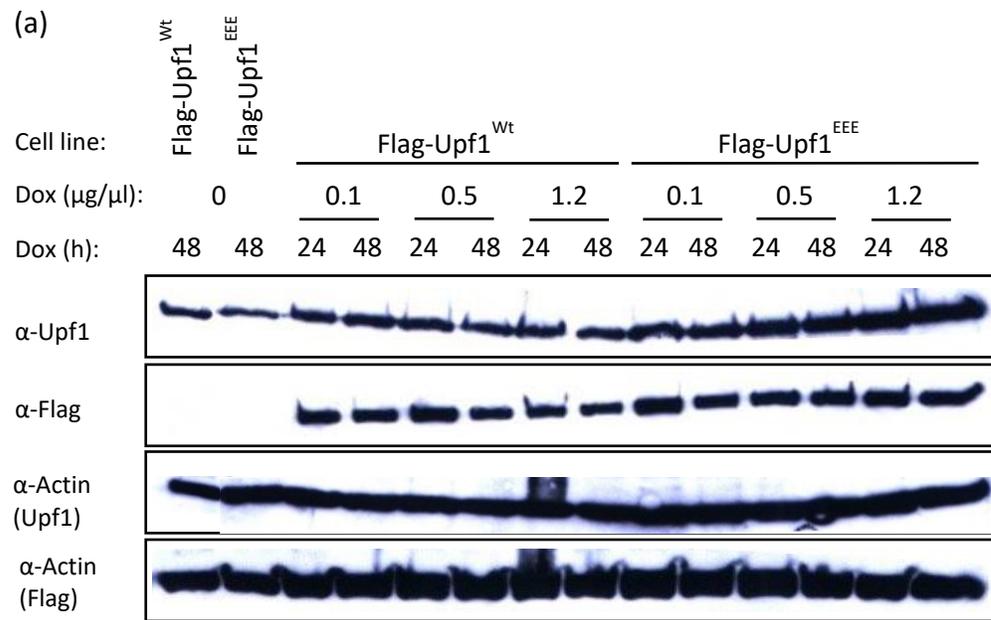


Figure 3.3: Optimization of doxycycline concentration for expression of FLAG-Upf1 protein in Flp-In HeLa cells

Cells were incubated in media containing the indicated range of doxycycline concentrations, 0-1.2 μg/ml for 24 and 48 hours. Lysates were prepared as described in section 2.8.1 and western blot analysis was performed with both α-Flag and α-Upf1 antibodies was done to determine the expression levels. Detection of β-actin was used as a loading control.

3.2.2 Upf1 and genome integrity

Mammalian cell Upf1 hypomorphs have been shown previously to result in loss of genome integrity, as evidenced by increased expression of a DNA damage marker and loss of telomeric DNA (Chawla et al. 2011; Azzalin & Lingner 2006b). In NMD, Upf1 is known to undergo cycles of phosphorylation and dephosphorylation to ensure destruction of NMD targets in which recruitment of key components of the NMD machinery is dependent on the phosphorylation status of Upf1 (Conti & Izaurralde 2005; Page et al. 1999; Ohnishi et al. 2003). In a first attempt to determine whether Upf1 might act via a similar mechanism in execution of its genomic integrity role, I wished to determine whether co-expression of phosphorylation site mutants might have a dominant negative effect on genome integrity in the HeLa model cell line. To do this, I assessed the levels of γ H2AX expression, a DNA damage marker. Damage to DNA results in the phosphorylation of histone H2AX and its localization to sites of damage. Because phosphorylation of H2AX at ser139 is abundant, fast and correlates well with double stranded breaks (DSBs), it is a sensitive marker that may be used to detect DNA damage (Sharma et al. 2012). In addition, I assessed the ability of the Flag-Upf1^{ChrmB} mutant to act in a dominant-negative fashion.

A time-course experiment was undertaken to evaluate the impact on H2AX expression, following overexpression of the mutants- Flag-Upf1^{EEE}, Flag-Upf1^{AAA} and Flag-Upf1^{ChrmB} cells using Flag-Upf1^{Wt} as a control comparison. All cells 2×10^5 were grown in 6-well plates and incubated for 24, 48 or 72 hours in the presence of 0.7 μ g/ml doxycycline before whole cell lysates were generated and the presence of DNA damage determined by immunoblotting. As additional controls, cells were also grown either in the absence of doxycycline, or in the presence of both doxycycline and the replication inhibitor hydroxyurea (2mM). The latter inhibits DNA synthesis, and as a consequence of replication fork collapse results in a generation of significant numbers of double-strand breaks and thus expression of γ H2AX is expected.

The results shown in **Figure 3.4a** indicates that, at the levels of overexpression achieved, none of the phosphorylation site mutants appear to have any effect on γ H2AX expression. This was not due to a failure to detect γ H2AX, as treatment of mutant cell lines with the replication inhibitor hydroxyurea resulted in a robust signal as expected. This is consistent with previous data involving isogenic cell lines expressing forms of Upf1 (Turton & Smythe, unpublished). Somewhat surprisingly, overexpression of the ChrmB mutant Upf1 (last 4 lanes), which fails to bind to chromatin, did induce an increase in γ H2AX, suggesting that this mutant can act as a dominant negative, interfering with genomic integrity.

The results shown in **Figure 3.4a** indicates that, at the levels of overexpression achieved, none of the phosphorylation site mutants appear to have any effect on γ H2AX expression. This was not due to a failure to detect γ H2AX, as treatment of mutant cell lines with the replication inhibitor hydroxyurea resulted in a robust signal as expected. This is consistent with previous data involving isogenic cell lines expressing forms of Upf1 (Turton & Smythe, unpublished). Somewhat surprisingly, overexpression of the ChrmB mutant Upf1 (last 4 lanes), which fails to bind to chromatin, did induce an increase in γ H2AX, suggesting that this mutant can act as a dominant negative, interfering with genomic integrity.

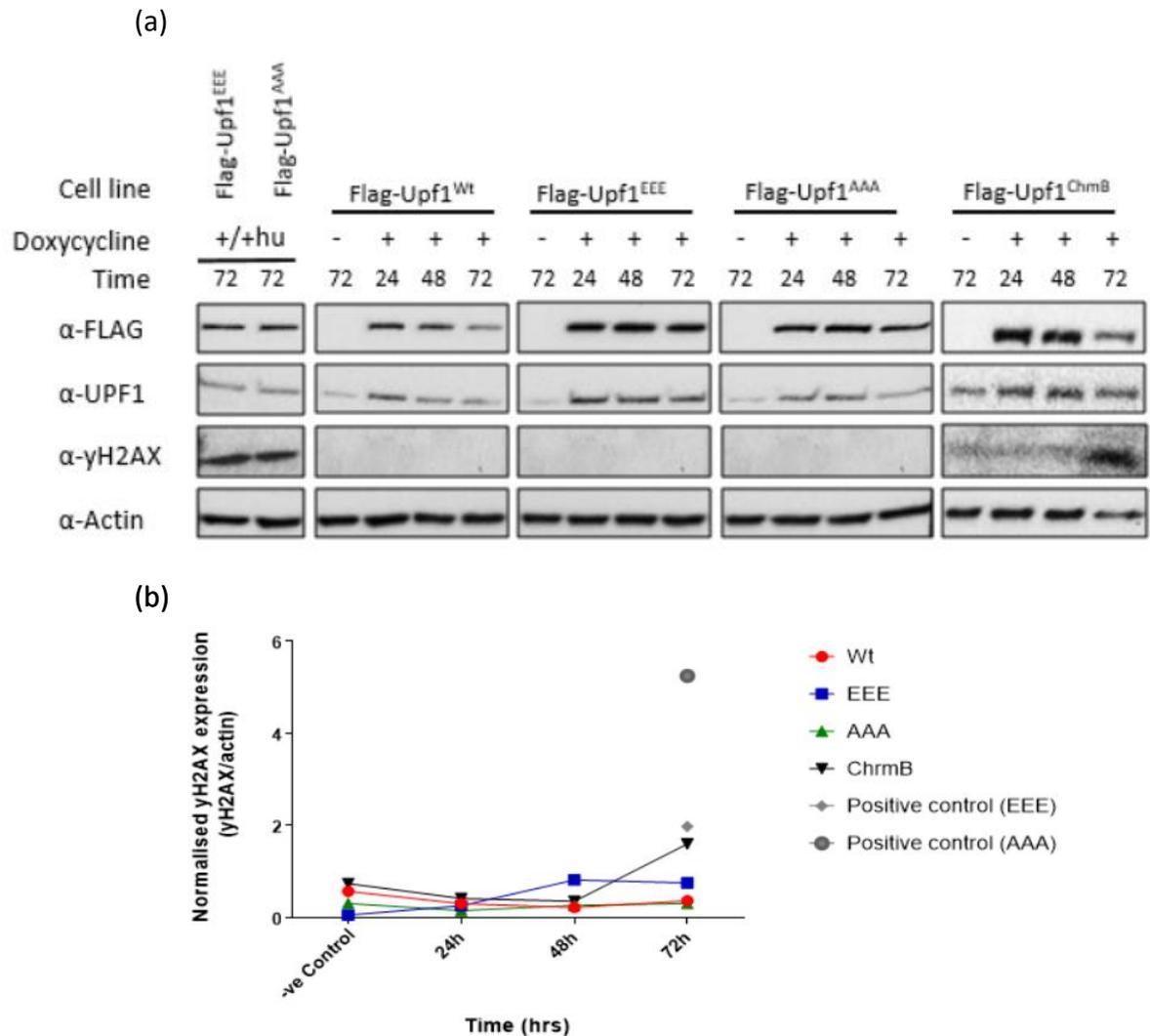


Figure 3.4: Overexpression of ChrmB mutant Upf1, but not phosphomimetic or unphosphorylatable mutants results in the expression of the DNA damage marker H2AX

Cells were treated with 0.7µg/ml doxycycline to induce Flag-tagged protein expression for the indicated lengths of time to induce Flag-tagged protein expression prior to cell lysis. (a) All proteins were analysed by immunoblotting and detected using antibodies listed in section 2.2. (b) Quantification of (a) using ImageJ and graph GraphPad Prism 7 software.

3.2.3 Optimisation of siRNA concentration for time course assay

It is conceivable that N- and C- terminal phosphorylation of Upf1 is not required for its chromatin-associated role. In this circumstance, it might be expected that either non-phosphorylatable or phosphor-mimetic mutants might both be capable of rescuing known genomic phenotypes associated with Upf1 deficiency. In order to investigate this, it was necessary to establish conditions where significant siRNA-mediated knockdown of Upf1 was obtained. Using electroporation (section 2.3.4), two concentrations of siRNA targeting wild-type Upf1 were used, 50nM and 200nM. A non-targeting siRNA was used as a control. Cell lysates were prepared at the indicated times without the addition of doxycycline and immunoblotting was undertaken using anti-Upf1 antibodies.

As observed from the immunoblot and its quantification, siRNA-mediated knockdown of Upf1 occurs efficiently over 24 or 48 period with residual levels being approximately 10-20%, irrespectively of the concentration of siRNA used (**Figure 3.5a**). Noticeably, Upf1 knockdown was transient and endogenous Upf1 levels start to increase after 48 hours of treatment for both concentrations, recovering to levels of approximately 20% after 72 hours. Based on **Figure 3.5b**, the concentration of 200mM siRNA seems to have a significant affect the knockdown even at 72 hours when recovery is observed.

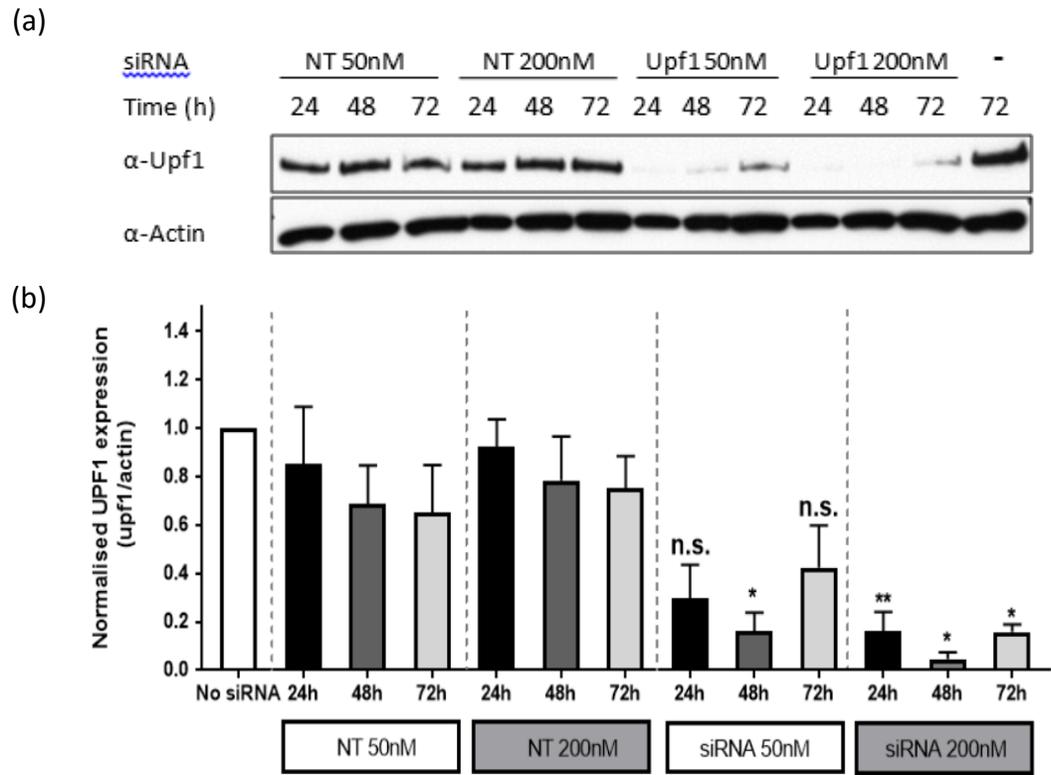


Figure 3.5: Optimisation of Upf1 siRNA for Flag-Upf1 cell lines

Flag-Upf1 cells (Wt and mutants) were transfected with either 50nM or 200nM of Upf1 or non-targeting siRNA using electroporation. Lysates were prepared from cells that have been grown for 24, 48 or 72 hours. No doxycycline was added to the cell culture and expression of endogenous was observed by immunoblotting. (b) Quantification to analyse efficiency of knockdown based on the blot using GraphPad Prism 7 software. (n=3)

3.2.4 Time-course assay for Upf1 knockdown

Results in section 3.2.2 indicate that overexpression of the phosphorylation site mutants described previously have no effect on integrity of DNA as judged by the DNA damage marker γ H2AX, while overexpression of the mutant which fails to bind to chromatin does results in DNA damage. Previous work in this laboratory (Turton 2014) and elsewhere (Azzalin & Lingner 2006b; Chawla et al. 2011) have shown that reduced expression of Upf1 results in the expression of γ H2AX. I therefore wished to establish whether any of the mutants described above could rescue the DNA damage phenotype following siRNA-mediated knockdown of Upf1.

Cells were exposed to either non-targeting or Upf1 targeting siRNA (200nM) for a range of time period (24, 48 & 72 hours). Doxycycline was added to all of the cells 2 hours after electroporation and grown to the indicated time. As positive controls for γ H2AX expression, both FLAG-Upf1^{EEE} and FLAG-Upf1^{AAA} cells were harvested after 72 hours, grown in the presence of doxycycline and hydroxyurea (HU) which was added for the final 24 hrs. The addition of HU into media was to arrest cell growth at S-phase.

Consistent with previous results, significant depletion of endogenous Upf1 was achieved. In this experiment, knockdown effects observed were more persistent than before probably due to a higher concentration of siRNA used. Both Flag-Upf1^{EEE} and Flag-Upf1^{AAA} lines showed a progressive increase in expression of γ H2AX. Induction, by doxycycline addition, of an siRNA-resistant form of Flag-Upf1^{EEE} did not have any significant impact on either the rate or the extent of γ H2AX expression, while expression of Flag-Upf1^{AAA} did reduce the extent of γ H2AX expression slightly. Interestingly, γ H2AX levels were also not reduced, when siRNA-resistant Flag-Upf1^{ChrmB} was expressed in cells lacking endogenous wild-type Upf1. These data show that mutations in both N- and C-terminal phospho-sites sites render Upf1 non-functional with respect to its role in genome integrity. That neither phospho-mimetic nor non-phosphorylatable mutants demonstrated ability to rescue the defect induced by knockdown of the endogenous protein suggests that, as with Upf1 function in NMD, cyclical phosphorylation and

dephosphorylation may also be involved in the mechanism of Upf1-mediated maintenance of genome integrity. Interestingly, Upf1^{ChrmB} also failed to rescue wild-type function, strongly suggesting that chromatin association is important for Upf1 role in genome integrity. Taken together, these data indicate that this system is useful for the functional analysis of the role of Upf1 in genome integrity.

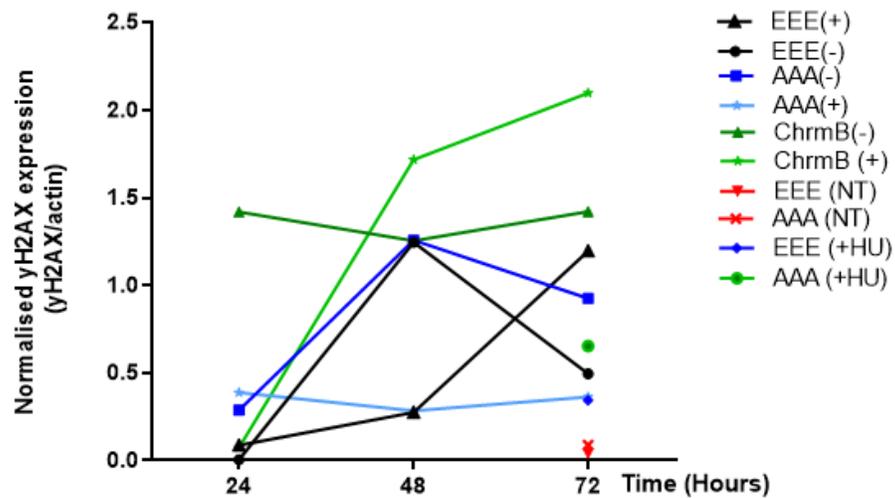
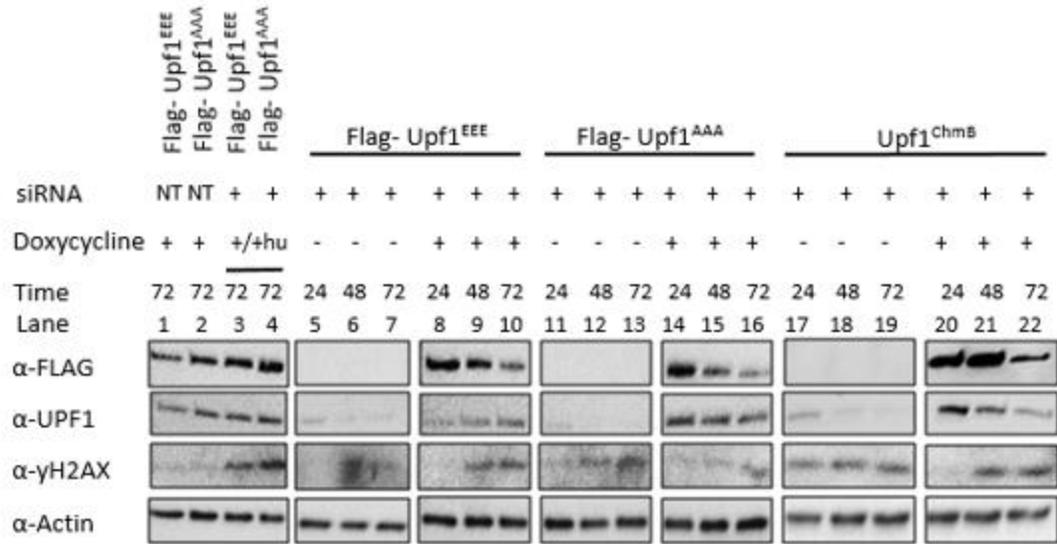


Figure 3.6: Cells expressing phospho-mimetic Flag Upf1^{EEE}, unphosphorylatable Flag-Upf1^{AAA} or Flag-Upf1^{ChrmB} are unable to rescue the effect of Upf1 knockdown. Flag-Upf1^{Wt} and mutants were transfected with either Upf1 (200nM) or non-targeting siRNA. Cells were grown for indicated times and lysed. (a) Lysates were analysed by immunoblotting with the indicated antibodies. Cell treated with both doxycycline and hydroxyurea (HU, added for only 24hrs) were used as a positive control for the expression of γH2AX. Actin was used as a loading control with upper actin band representing the same blot as α-Flag and α-γH2AX. (b) Quantification of (a). **Legend:** (-) uninduced/ dox not added, (+) induced/ dox added, (NT) non-targeting siRNA, (-HU) no Hydroxyurea added, (+HU) Hydroxyurea added.

3.2.5 Chromatin-bound Upf1

The association of a subpopulation of total cellular Upf1 with chromatin has been described a number of times (Azzalin & Lingner 2006b; Chawla et al. 2011; Turton), although the mechanistic basis for this association is not understood. The recruitment of UPF1 to chromatin is dependent on the presence of the PIKK family member and checkpoint kinase ATR (Azzalin & Lingner 2006b; Chawla et al. 2011). However, it is unclear whether ATR phosphorylates UPF1 directly, and if so, whether ATR acts in a manner analogous to that of SMG1 in NMD, phosphorylating PIKK consensus sites in the N- and C-terminal regions of Upf1 - which may be necessary for chromatin association, or, whether ATR acts directly to phosphorylate Upf1 at a distinct site, enabling its recruitment to chromatin. Previously, work in this laboratory has shown that the Upf1^{ChrmB} mutation (Ser42-Ala) fails to bind chromatin, resulting in telomere loss and γ H2AX expression (Turton 2014). In order to further understand the relationship between Upf1 phosphorylation status, chromatin binding, and consequent DNA damage, I wished to establish whether the Upf1 phosphorylation site mutants described above were associated with chromatin.

In order to do this, sub-cellular fractionation, based on the method of Mendez and Stillman (2000), was undertaken as described in Section 2.5 to generate chromatin, soluble nuclear, and cytosolic fractions (**Fig. 3.7a**). Only a small fraction of total cellular Upf1 is chromatin-associated and maximum levels of chromatin-bound Upf1 occurs in S-phase (Azzalin & Lingner 2006b). Thus in experiments described below, cells were enriched in S-phase by the addition of 2mM hydroxyurea for 24 hours prior to fractionation. Flag-Upf1^{Wt}, Flag-Upf1^{AAA}, Flag-Upf1^{EEE} and Flag-Upf1^{ChrmB} cells were treated with 0.7 μ g/ml doxycycline for 48 hours and 2mM HU was added for the final 24 hrs. HU was washed off and incubated in new media for 3 hours before fractionation. This would allow cells to pass into S-phase and producing synchronised cells. Sub cellular fractions were subject to immunoblotting using anti-Flag antibodies to detect forms of Upf1, anti-tubulin antibodies for tubulin as a marker of the cytoplasmic fraction, and

anti-ORC2 antibodies to detect ORC2, a component of the Origin Recognition Complex (Méndez & Stillman 2000; Cuadrado et al. 2006) as a marker for chromatin.

As revealed by immunoblotting (**Figure 3.7**) and as expected (Azzalin & Lingner 2006b; Turton 2014), Flag-Upf1^{Wt} was distributed between cytoplasm and the nuclear soluble fraction, with a small proportion in the chromatin fraction, which was not reduced by repeated washing of the chromatin pellet (Azzalin & Lingner 2006b). That Upf1 is associated with chromatin, and is not simply a consequence of cytoplasmic contamination is supported by the absence of detectable levels of tubulin in the chromatin fraction. Both Flag-Upf1^{EEE} and Flag-Upf1^{AAA} were found in all fractions, with a distribution similar to that of the wild-type protein. These data indicate that mutations in both N- and C-terminal PIKK phosphorylation sites, that either generates phosphomimetic or non phosphorylatable forms, do not interfere with Upf1's ability to associate with chromatin. This however was not observed for the Flag-Upf1^{ChrmB} mutant, confirming previous data (Turton 2014). Taken together with results shown in figure 3.6, these data suggest that distinct mechanisms underpin the inability of either Flag-Upf1^{EEE} or AAA to suppress γ H2AX expression compared with Flag-Upf1^{ChrmB}.

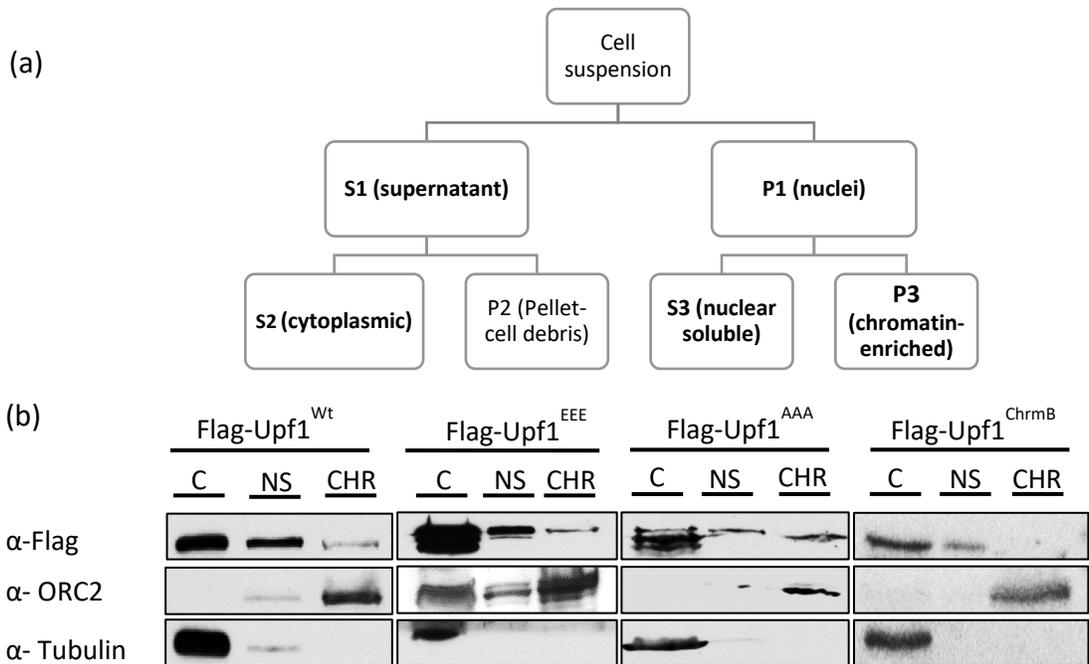


Figure 3.7: Sub-cellular fractionation of Flag-Upf1^{Wt}, Flag-Upf1^{EEE}, Flag-Upf1^{AAA} and Flag-Upf1^{ChrmB} cells.

(a) Scheme of sub-cellular fractionation process. FLAG-Upf1^{Wt} cells were grown in 0.7µg/ml doxycycline for 48hr, then 2mM of hydroxyurea (HU) was added for 24hr to arrest cells in S-phase. HU was washed out and cells released into S-phase for 3hr. Cell suspension was spun down and pellet was resuspended in Buffer A (10mM HEPES, 10mM KCl, 1.5mM MgCl₂, 0.34M sucrose, 10% glycerol, 1mM DTT, 1x Protease inhibitor cocktail, 0.1mM PMSF) before centrifuged to separate the S1 and P1 fraction. S1 was further clarified by spinning down at 20,000g for 15 mins to obtain the S2 fraction. Cell pellet (P1) was resuspended in Buffer B (3mM EDTA, 0.2mM EGTA, 1mM DTT, 1x Protease inhibitor cocktail) allowing the nuclei to be lysed before centrifuge to separate the S3 and P3 cell fractions. (b) Samples representative of equal no of cells were loaded onto an SDS-PAGE and visualised by western blotting. Purity of samples were determined by using antibodies against α-ORC2 nuclear marker and α-Tubulin as cytoplasmic marker. S2-cytoplasmic (C), S3- nuclear soluble (NS) and P3- chromatin enriched fractions (CHR). The FLAG-Upf1^{EEE} was not a clean prep hence there were bands present for ORC2 in both cytoplasm and nuclear soluble fractions.

3.2.6 Expression and purification of full-length wild-type and mutant forms of Upf1

Upf1 is a member of the SF1 helicase superfamily and uses the energy of ATP to unwind both DNA and RNA substrates (Czaplinski et al. 1995; Perlick et al. 1996; Carastro et al. 2002). In vivo, all of the known functions of Upf1 require ATPase activity, as mutations in the ATPase domain inhibit function (Weng et al. 1998; Weng et al. 1996b). Structurally, Upf1 is composed of a characteristic double RecA-like helicase domain (HD) flanked by two external regulatory domains - the so-called cysteine and histidine-rich (CH) domain at the N-terminus, and the C-terminal SQ domain, thus named as it contains a large number of SQ motifs. Biochemical studies have been restricted due to difficulties in producing intact Upf1 and have focused mainly on the analysis of truncated species (Chamieh et al. 2008; Chakrabarti et al. 2011). However, in vitro, characterization of recombinant, truncated forms of Upf1 indicate that both the N- and C-terminal domains can partially inhibit the intrinsic activity of the core helicase domain (Chakrabarti et al. 2014; Fiorini et al. 2013). For NMD in vivo, phosphorylation of T28, S1096 & S1116 are believed to be involved in the sequential recruitment of auxiliary factors that relieve helicase inhibition, in turn enabling processivity, possibly by altering the conformational arrangement of N and C termini with respect to the helicase domain (Okada-Katsuhata et al. 2012; Yamashita et al. 2006).

The inability of either of the phospho-site mutations, or the ChrmB mutant, analysed in this work, to rescue DNA damage induced by knockdown of endogenous wild-type Upf1, led me to address the question whether these mutants retained affinity for DNA, and helicase activity, compared to the wild-type protein.

I proceeded to develop an expression system in *Escherichia coli*, in collaboration with Dr Cyril Sanders (University of Sheffield) for full-length Upf1. This involved a strategy involving the cloning of full-length Upf1 into a bacterial expression vector containing both N-terminal and C-terminal purification tags, with distinct protease cleavage sites to enable removal of tags prior to biochemical analysis. Expression of a protein of interest in *E.coli* is advantageous, in that relatively large quantities may be

obtained relatively economically. During the course of this work, the Sanders' lab established a protocol for purifying full-length Upf1 (Dehghani-tafti & Sanders 2017).

Upf1^{EEE}, Upf1^{AAA} and Upf1^{ChromB} sequences were inserted into the pET11c plasmid backbone described in (Dehghani-tafti & Sanders 2017), containing a 5' glutathione S-transferase tag (GST) and a 3' poly histidine affinity tag (**Figure 3.8a**). Using the protocol described in section 2.5, I expressed and purified 90 liters each of full-length Upf1^{Wt}, Upf1^{EEE}, Upf1^{AAA} and Upf1^{ChromB} proteins, each to homogeneity (**Figure 3.8b**). All steps for purification was done at 4°C. Cells were resuspended in lysis buffer (1x lysis buffer (100mM Tris pH 7.5, 20mM EDTA, 20% glycerol, 20mM DTT); 0.1M NaCl, 1mM PMSF) and 1/50 volume of 50mg/ml lysozyme was added to the cell suspension, incubating for 30mins. Cell suspension was sonicated and lysate cleared by centrifugation (11,500rpm, 30mins). Nucleic acid acids were removed by precipitating with Polymin P (0.5% w/v) and proteins precipitated with ammonium sulphate (50% saturation) before GST affinity chromatography. Eluted protein was then digested with thrombin, to get rid of N-terminal tag. Protein was then concentrated by binding and step elution from a 1ml His-Trap column (GE Healthcare) before applying to a Superdex 200 column. Protein was applied again to a His-Trap column and eluted in a gradient 20 to 500mM imidazole. Fractions containing Upf1 was then pooled and digested with TEV removing the His-Tag. For each cell line, an approximately 1mg of purified protein was obtained from an average of 360g wet-weight of E.coli cells.

There are however several caveats to this method. Firstly, because this experiment was done *in vitro*, in a more artificial and controlled condition, not all components that are usually available in an organism is present. Belasco (2010) pointed out that there are no ribonuclease or ancillary protein with a specialized role in NMD that has been identified in bacteria, which means that it lack homologs of the UPF and SMG proteins. Hence, the results obtained might not be as accurate as if the same experiment was done *in vivo*. On top of that, PTMs are important for eukaryotes as it can significantly change the integral characteristics of a protein that might affect things

like stability and solubility, hydrophobicity and also solvent accessibility (Tokmakov et al. 2012). E.coli however lacks or has limited eukaryotic PTM machinery function which is a disadvantage since it could not produce eukaryotic phosphoproteins such as serine/threonine kinases (Khow & Suntrarachun 2012).

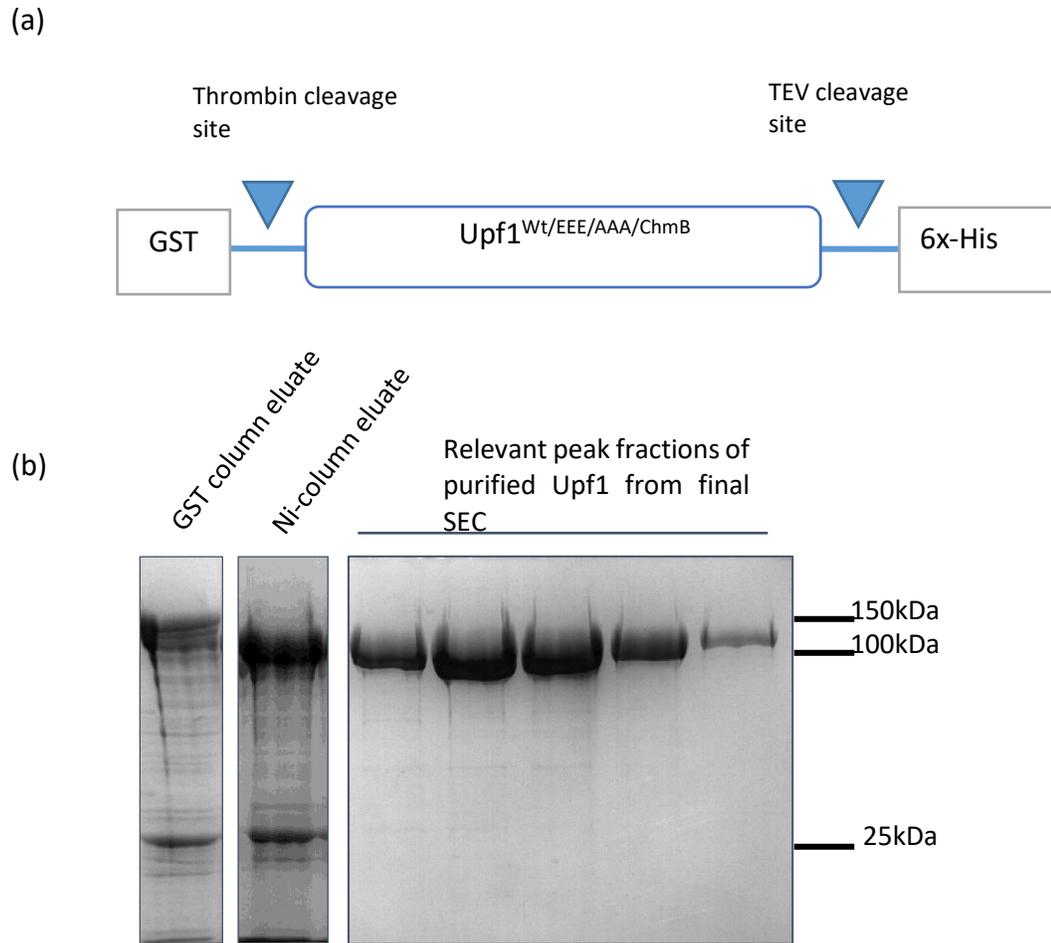


Figure 3.8: Protein purification of full-length Upf1 expressed from *E.coli*

(a) Schematic of Upf1 construct that was made and expressed in *E.coli*. The plasmid was produced in-house by Dr Cyril Sanders 9 (pET11cGST) where the N-terminal has a GST (P08515) tag followed by a thrombin cleavage site. On the C-terminal, the construct consists of a TEV cleavage sequence (ENLYFQS) followed by six histidine residues. (b) Tagged Upf1 was expressed in *E.coli* BL21 (DE3) and purification performed. Protein fractions representing Upf1 were run through twice each of nickel column (Ni-column) and a size exclusion column (SEC) before a purified full length Upf1 was obtained. Tags were cleaved off in the process.

3.2.7 Strand displacement assay

Once purified protein was obtained, I wished to establish whether any or all of the mutants retained DNA binding and helicase function, by performing both electrophoretic mobility shift (EMSA) and strand displacement assays (Dehghani-tafti & Sanders 2017). Initially, a strand displacement assay as described in section 2.6 was undertaken with both full-length Upf1^{Wt} and the equivalent mutants (Upf1^{EEE/AAA/ChrmB}).

The assay utilised the partially single-stranded, double-stranded substrate, described in section 2.6.1, consisting of a 55 base 5' tail and 20bp dsDNA (ss/dsDNA substrate) and helicase activity of the protein obtained in **Figure 3.8**. The displacement activity was determined by measuring the fraction of annealed substrate undergoing separation in a 30 minute, 37°C assay by analysis of radiolabelled reaction products on a polyacrylamide gel following autoradiography, and subsequent quantitation of radiolabelled bands. To determine the maximum degree of strand separation possible, the substrate was heated to 100°C for 5 minutes and slowly left to cool to 4°C to ensure total separation of DNA strands (Dehghani-tafti & Sanders 2017). An ATPase-dead mutant, Upf1^{K498A} (kind gift of C.Sanders, Sheffield) was used as a negative control. This Upf1 variant protein contains an amino acid substitution in the ATPase catalytic domain, which has been shown previously to abolish both ATPase activity and Upf1 helicase activity (Kaygun & Marzluff 2005a).

As expected, no unwinding of the substrate was detected for Upf1^{K498A} **Figure 3.9a** (lanes 25-29). In the case of each of the three mutants Upf1^{EEE} (**Figure 3.9**, lanes 19-24), Upf1^{AAA}, (**Figure 3.9a**, lanes 13-18) Upf1^{ChrmB} (**Figure 3.9a**, lanes 7-12) displayed unwinding activity in a concentration-dependent manner, very similar to Upf1^{Wt}. (**Figure 3.9a**, lanes 1-6). A graph quantifying the reaction rate as a function of added enzyme concentration was plotted, where all data presented are derived from three independent experiments (**Figure 3.9b**). Based on the graph plot, it is shown that there were no significant differences between the helicase activity of Upf1^{Wt} and the different

forms of mutants *in vitro*. However, all three mutants (EEE/AAA/ChrmB) seems to be a bit more active compared to wild-type.

It was anticipated that one (particularly FLAG-Upf1^{ChrmB}) or more of these mutants would not exhibit allocates activity. Because enzyme activity requires DNA binding, electro-mobility shift assays were not undertaken. However that was not observed in this study.

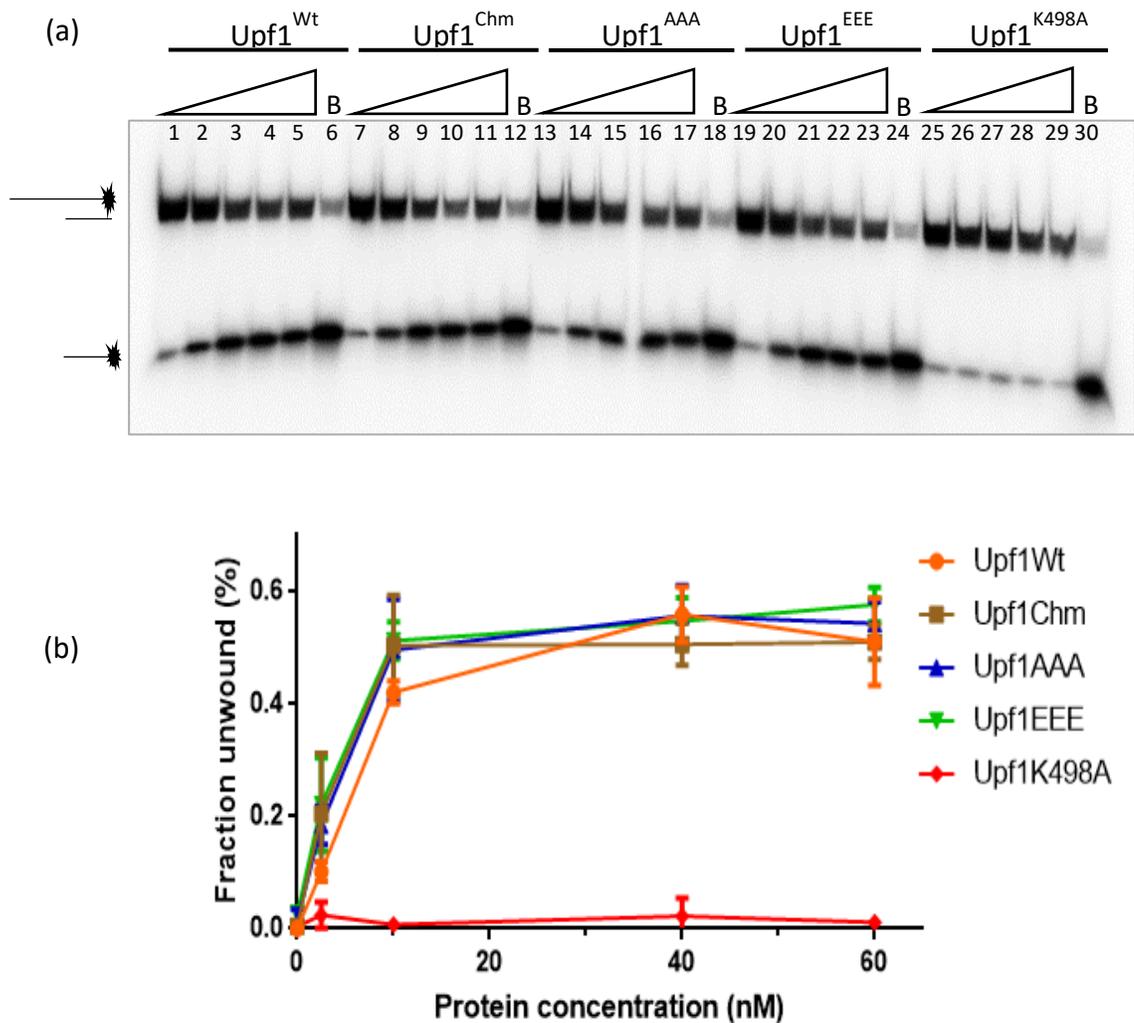


Figure 3.9: Mutations does not impair the ability of Upf1 to function as a helicase.

(a) Strand displacement assay (as detailed in section 2.6) was used to determine the helicase activity of Upf1^{Wt} and mutant forms of Upf1 (Upf1^{AAA/EEE/ChmB}). The upper band represents double stranded DNA with the shorter strand having ³²P-labelled and the lower band represents ³²P-labelled single stranded DNA. Reactions were performed with increase Upf1 proteins concentrations (0, 2.5, 10, 40 and 60 nM). A boil control was included as a positive control. (b) is the quantification of (a).

3.3 Conclusion

In addition to its canonical role as a central component of the nonsense mediated decay pathway (Nagy & Maquat 1998; Pal et al. 2001) and its role in histone mRNA decay (Müller et al. 2007; Kaygun & Marzluff 2005a; Kaygun & Marzluff 2005b; Panamwan 2017), the presence of Upf1 in the cell has been shown to be essential for maintaining genome stability in mammalian cells, via a poorly characterized role in DNA replication and/or DNA repair (Varsally & Brogna 2012; Azzalin & Lingner 2006a; Imamachi et al. 2012; Chawla et al. 2011; Carastro et al. 2002). Azzalin & Lingner (2006a) was first to identify that nuclear Upf1 has a separate role from NMD, which occurs predominantly in the cytoplasm, by showing that the DNA damage effects observed when Upf1 was depleted in cells were not seen in cells depleted of another core NMD component, Upf2. In addition, cells depleted of Upf2 was able to proceed normally through the cell-cycle unlike cells that were depleted of Upf1. Upf1 is recruited to chromatin during S-phase, and levels are increased in response to application of exogenous replication stress, which suggest that Upf1 is involved in DNA replication or repair in higher eukaryotes. siRNA-mediated depletion of Upf1 results in telomere dysfunction which includes accumulation of telomere free ends and fragile telomeres, and accumulation of cells containing nuclear γ H2AX foci. Upf1 has a central role in nonsense mediated decay, and this role involves ATP-dependent helicase activity (Chakrabarti et al. 2011; Czaplinski et al. 1995; Weng et al. 1996b), the cyclical phosphorylation and dephosphorylation of key amino acid residues in both its C- (S1096 & S1116) and N-terminus (T28) that are required for specific recruitment of other NMD components (SMG6, SMG5 & SMG7). More recently, Upf1 has also been shown to act as a ubiquitin ligase (Feng et al. 2017). With the exception of its ATPase activity, it is not at all clear which, if any, of these molecular functions are required for its genomic integrity function (Chawla et al. 2011).

Although DNA replication is a tightly regulated process, a range of replication associated aberrant DNA structures are known to occur, and a range of DNA repair systems operate to resolve them. Replication stress is defined as slow or stalling of the

replication fork which can cause from several factors such as DNA lesions, secondary DNA structures or dormant replication origins (Mazouzi et al. 2014). Efficient replication fork progression requires that replisome gains access to the DNA through remodelling of the chromatin structure (Jones & Petermann 2012). The presence of secondary structures like G-quadruplex especially at telomeres require resolution since these structure could stall fork progression and be critical to genome integrity.

In summary, published data are consistent with the model in which UPF1 recruitment to the replication fork is required to complete telomere replication, in addition to other replication sites throughout the genome.

I wished to investigate functional similarities and differences in the molecular mechanism of Upf1 activity in vivo to begin to characterize molecular mechanisms underpinning its role in genome stability. Hence the aim of the experiments described in this chapter was to investigate the relevance of phosphorylation sites which have been shown previously to be important during NMD (Okada-Katsuhata et al. 2012; Durand et al. 2016; Yamashita et al. 2001). Previous work using ChIP analysis of telomere DNA has indicated that a number of regulatory components of the NMD pathway, specifically the protein kinase SMG1 (Yamashita et al. 2001), and the phosphatase regulatory components SMG5/7, are enriched at telomeres (Azzalin & Lingner 2006b; Azzalin, Reichenbach, Khoriantuli, Giulotto & Lingner 2007).), suggesting that a phosphoregulatory cycle is likely to be important in UPF1-mediated genome stability. Importantly, SMG1 is a member of the PIKK family of protein kinases, which also includes the checkpoint kinase ATR (Lovejoy & Cortez 2009). Recruitment of Upf1 to chromatin is dependent on functional ATR (Chawla et al. 2011), however the precise role of ATR is unclear. For example, it is unclear whether ATR acts directly or indirectly in the recruitment of Upf1, or functions redundantly with SMG1 in phosphorylation of Upf1 on previously identified sites to maintain chromatin-associated function, or whether the sites of phosphorylation on chromatin-associated Upf1 are distinct from those known to play a role in NMD.

In addition, previous work from this laboratory identified Ser42 as an important residue required for the association of Upf1 with chromatin, as mutation to alanine at this position results in loss of chromatin binding, while a phospho-mimetic substitution to glutamic acid results in chromatin association. It is not clear, however, whether Ser42 is phosphorylated on chromatin associated Upf1 (Turton 2014). I therefore wished to examine further the significance of these observations.

3.3.1 Upf1 and genome stability

Once an optimal expression level of FLAG-Upf1 was determined, I wanted to establish whether over-expression of either phospho-site or chromatin-binding mutants might have a dominant negative effect, and chose to monitor effects by measuring the expression of γ H2AX. Of the 3 mutant cell lines; FLAG-Upf1^{EEE}, FLAG-Upf1^{AAA} and FLAG-Upf1^{ChrmB}, only FLAG-Upf1^{ChrmB} appeared to give rise to increased expression of γ H2AX. Phosphorylated H2AX (γ H2AX) is a sensitive proxy marker for DNA double strand breaks, that, among other genomic insults such as ionising radiation (Sharma et al. 2012) may be caused by the collapse of a stalled replication fork (Chanoux et al. 2009; Sharma et al. 2012).

Previous studies (Ohnishi et al. 2003; Kashima et al. 2006a; Cheng et al. 2007) have identified the importance of sequential phosphorylation-dephosphorylation of Upf1 for NMD as this process is believed to regulate the remodeling of the SURF surveillance complex, during the process of identification of premature stop codons. I wished to establish whether phosphosite mutants had a dominant negative effect on genome stability, as this might indicate that a specific post-translationally modified form of Upf1 directly interfered with some component of the replication/repair machinery.

While overexpression of either FLAG-Upf1^{EEE} or FLAG-Upf1^{AAA} showed no significant effect on γ H2AX expression (**Figure 3.4**), interestingly FLAG-Upf1^{ChrmB} did result in elevated levels of the DNA damage marker following 72 hours of overexpression. This suggests with the presence of endogenous chromatin associated Upf1, the

presence of another form of Upf1 that is not capable of binding to chromatin (FLAG-Upf1^{ChrmB}) does interfere in some ways with DNA homeostasis. The phosphorylation sites which controls the phospho state of Upf1 however, has never been focused on in relations to Upf1's nuclear role of maintaining genome stability.

The absence of any DNA damage effect following over-expression of phospho-site mutants did not provide any insight as to whether phosphorylation of residues T28, S1096 and S1116 is necessary for Upf1 chromatin associated function. To test this, I proceeded to knockdown endogenous Upf1 using electroporation. The aim was to determine whether expression of FLAG-Upf1 phospho-site mutants could rescue the DNA damage phenotype. When endogenous Upf1 levels were reduced by siRNA-mediated knockdown and in the absence of doxycycline, both cell lines had elevated levels of γ H2AX expression compared to the equivalent Flag-Upf1^{Wt} cell line. After knockdown of the endogenous wild type protein and in the presence of doxycycline inducing the expression FLAG-Upf1^{AAA}, γ H2AX levels were elevated after 72 hours. A similar effect was observed following knockdown and expression of FLAG-Upf1^{EEE}. Noticeably, the ability of the FLAG-Upf1^{EEE} mutant to suppress the appearance of γ H2AX expression was reduced compared to FLAG-Upf1^{AAA}, suggesting that Upf1^{AAA} may retain some chromatin-associated function compared to the phospho-mimetic form.

Thus, in the presence of endogenous Upf1, normal cell surveillance functions proceed even in the presence of FLAG-Upf1 phospho-site mutants. However, these mutants were unable to rescue the wild-type phenotype when endogenous Upf1 expression levels were significantly reduced (**Figure 3.6**). The results of these experiments are consistent with those of Azzalin and Lingner (2006b) and together these data highlight the important of a functional Upf1 in maintaining genome stability, as depletion of Upf1 causes a DNA damage response. Additionally, the data shown here imply that cycles of phosphorylation at both N- and C-terminal sites are required for the chromatin-associated function of Upf1.

As expected (Turton 2014), cells expressing FLAG-Upf1^{ChrmB} expressing cells were unable to rescue the emergence of DNA damage, resulting from knockdown of endogenous Upf1, as judged by the expression of γ H2AX which was observed after 48 hours. The inability of FLAG-Upf1^{ChrmB} to rescue the emergence of γ H2AX levels in both experiments (overexpression and also knockdown) shows that it portrays itself as a dominant negative.

3.3.2 Chromatin-bound Upf1

Upf1^{WT}-depleted cells that express FLAG-Upf1^{ChrmB} (ie S42A) are capable of undertaking NMD (Turton 2014) but express γ H2AX, presumably as a consequence of a lack of chromatin-associated function (Turton 2014). One possible explanation for the observed expression of γ H2AX in Flag-Upf1^{EEE} and FLAG-Upf1^{AAA} expressing cells lacking endogenous wild-type Upf1, was that these mutants might also lack the capability to bind chromatin.

Biochemical fractionation was undertaken to prepare cytoplasmic, soluble nucleoplasmic and chromatin-enriched fractions from all FLAG-Upf1 expressing cell lines. To maximize the amount of chromatin-associated Upf1, all lines were treated with hydroxyurea for 24 hours, to accumulate cells in S phase, and because hydroxyurea inhibits DNA synthesis and also induces replication fork stalling (Masai et al. 2010; Shechter et al. 2004; Ho et al. 2006) which may provoke Upf1 recruitment (Azzalin and Lingner, 2006b).

I found (**Figure 3.7**) that, unlike FLAG-Upf1^{ChrmB}, FLAG-Upf1^{EEE} and FLAG-Upf1^{AAA} were associated with chromatin. Taken together these data indicate that the presence of wild-type N and C- terminal phospho-sites are not required for chromatin recruitment, but that nonetheless both of these mutants are unable to prevent the expression of γ H2AX, suggesting that chromatin associated Upf1 is likely to require phosphorylation at one or more of these sites in order to carry out its genome integrity function.

3.3.3 Helicase activity of Upf1

Sequential phosphorylation and dephosphorylation of Upf1 (Chamieh et al. 2008; Behm-Ansmant & Izaurralde 2006; Imamachi et al. 2012; Kervestin & Jacobson 2012; Schweingruber et al. 2013; Chiu et al. 2003) is essential for its role in mRNA surveillance. The various phospho-forms of the protein are believed to change its conformation and affinity for key proteins during the process of NMD (Chakrabarti et al. 2011; Conti & Izaurralde 2005). Phosphorylation of Upf1 at key phosphorylation sites are predicted to activate its ATPase and helicase activities that leads to mRNA degradation through the interaction of SMG5-7 complex and SMG6 (Kervestin & Jacobson 2012; Schweingruber et al. 2013).

Previously, the non-availability of purified full-length Upf1 has precluded careful biochemical analysis of Upf1 enzymatic function. I wished to compare the DNA binding and associated helicase activity of EEE, AAA and ChromB mutants with the wild-type protein, and establish a role, if any for phosphorylation in the activity of the protein, as well as to establish whether the failure of the ChromB mutant to bind chromatin was a consequence of an inability to directly bind its DNA substrate.

Thus, I established and undertook strand displacement assays to determine how these mutations affect the helicase activity of Upf1. During the course of this work, a collaborative approach (Dehghani-tafti & Sanders 2017) led to the development of a purification protocol for full-length Upf1 (Dehghani-tafti & Sanders 2017). Using this protocol, I purified Upf1^{AAA}, Upf1^{EEE}, Upf1^{ChromB}, and Upf1^{K498A}. **Figure 3.9** Upf1^{K498A} is mutated at the ATPase motif 1 which abolishes ATPase activity of the Upf1 helicase core, and is inactive in strand displacement assays (Lee et al. 2015; Kashima et al. 2006b).

As expected Upf1^{K498A} was inactive at all concentrations assayed. Somewhat surprisingly, there was little but no significant difference in unwinding ability was observed in any of the other mutants when compared to the wild-type Upf1. This result is surprising for FLAG-Upf1^{EEE} and Flag-Upf1^{AAA}, as it has been suggested that such

mutants are locked in a specific conformational state precluding efficient processive helicase activity (Chakrabarti et al. 2014; Chakrabarti et al. 2011). One possible explanation might be that efficient processive helicase function in vivo involves the cyclical phospho-dependent, displacement of N and C-terminal domains that each can reduce helicase activity and that the use of triple A and triple E mutants gives rise to a structurally constrained form of Upf1 that cannot exhibit domain specific inhibition of helicase activity. This could be resolved by determining the helicase activity of combinations of mutants (ie N-terminal A, C-terminal EE, N-terminal E, C-terminal AA) mutants in comparison with those produced here. As the overall goal of this work was to identify mechanism specific to Upf1 chromatin-associated function specifically, this issue was not pursued further.

However, Chakrabarti et al. (2011) did observe a higher unwinding activity with a Upf1 Δ CH than that of wild-type. In the absence of Upf2, the CH domain is positioned on the helicase domain keeping the helicase activity low. Binding of Upf2 would on the other hand move the CH domain from the helicase domain which displaces it to the opposite site of the helicase allowing an increase of helicase activity. Therefore, they stipulated that this conformation change could be mimicked with the removal of the CH domain altogether.

Results from previous experiments indicate that the FLAG-Upf1^{ChrmB} is unable to associate with chromatin, however, interestingly, Upf1^{ChrmB} displayed strand displacement activity comparable to the wild-type protein (**Figure 3.9**). This data indicate that the ChrmB mutant exhibits normal DNA binding and helicase activity. It follows that the reason that the ChrmB mutant is not associated with chromatin in vivo is not as a consequence of a failure in DNA-binding activity.

In conclusion, this chapter aimed to characterise the genome integrity phenotype of different FLAG-Upf1 mutants; Flag-Upf1^{EEE}, Flag-Upf1^{AAA} and Flag-Upf1^{ChrmB} through the usage of Flp-In HeLa cell lines. I have demonstrated that wild-type, but not non-phosphorylatable or phospho-mimetic Upf1 mutants, are able to rescue γ H2AX

expression after endogenous Upf1 knockdown, that expression of FLAG-Upf1^{ChrmB} acts as a dominant-negative, and that failure of the ChrmB to associate with chromatin is not the consequence of a loss in DNA binding or helicase activity associated with the S42A mutation. These latter observations suggest that recruitment of Upf1 to a chromatin-bound form must involve some change in Upf1's molecular composition or environment. With that in mind, I chose this mutant for further analysis using mass spectrometry to determine Upf1 interacting proteins.

CHAPTER 4

4.0 Introduction

4.1 General Introduction

Following the initial analysis (Chapter 3) of isogenic cell lines that inducibly express wild-type, phospho-mutant, and a chromatin non-binding form of Upf1, I wished to investigate potential mechanistic differences in Upf1 function in the nucleus, via an approach involving the identification of its protein interactors. With the recent advancement in peptide based mass spectrometry (MS) (Marcotte 2007) coupled with the availability of human genome sequence information coupled with bio informatics-based tools for matching and quantifying specific peptides (Cox & Mann 2008), MS has become an immensely powerful tool for identifying protein-protein interactions.

The aim of the work outlined here was to optimise conditions for immuno-isolation (IP) of Flag-Upf1 to allow a one-step enrichment of Upf1 and its interacting proteins followed up by MS-based proteomics analysis (Flury et al. 2014). Hence, the first step was to optimise conditions for optimal immuno-isolation. Monoclonal Anti-FLAG antibodies specifically recognise target proteins within lysates with relatively low cross-reactivity with other cellular proteins (Gerace & Moazed 2015). Anti-Flag antibodies are commercially available (<https://www.sigmaaldrich.com/catalog/search?term=anti+flag&interface=All&N=0&mode=match%20partialmax&lang=en®ion=GB&focus=product>) and are also available already coupled to an inert matrix such as agarose (<https://www.sigmaaldrich.com/catalog/substance/antiflagm2affinitygel1234598765>) Furthermore, epitopes bound to these antibodies may be specifically recovered by incubating antibodies in the presence of excess epitope peptide (Lykke-Andersen et al. 2000; Zemp & Lingner 2014). Consequently, I chose to use an ANTI-FLAG® M2 affinity resin for the immuno-isolation step.

Based on the results presented in Chapter 3, I chose to compare the interacting partners of chromatin-associated wild-type Upf1 compared to those of cellular Upf1^{ChmB} mutant. This decision was based on the following considerations. While all three mutants, when expressed in wt-Upf1 knockdown cells, fail to rescue the loss of genome integrity phenotype as evidenced by γ H2AX expression, the ChrmB mutant acts in a dominant-negative fashion, and does not bind chromatin, suggesting that proteomic comparison with the wild-type protein might be expected to reveal differences in binding partners. Secondly, that the S42A mutation in the ChrmB mutant has no effect on DNA binding, as inferred from measurements of helicase activity (Chapter 3), increased the possibility that the inability of this mutant to bind to chromatin is a consequence of the loss of a protein-protein interaction function.

Several approaches are available for the quantitation of interacting proteins by mass spectrometry, including both SILAC (Dunham et al. 2012; Ong & Mann 2006; Mann 2006) as well as label-free quantitation (Griffin et al. 2010; Choi et al. 2015; Cox et al. 2014). SILAC is often used where identical isolation regimes are utilised for two or more distinct molecular entities, and a ratiometric comparison is sufficient (Ong et al. 2002; Mann 2006). As the proposed experiment involves a comparison of chromatin-bound Upf1 with total cellular Upf1, a label free, mass spectrometry approach was adopted. For subsequent experiments, endogenous Upf1 was not depleted and experiments were undertaken with the additional of doxycycline to induce FLAG protein expression.

4.2 Results

4.2.1 Optimisation of Micrococcal Nuclease (MNase) conditions

Both whole cell lysate and isolated chromatin fractions as described in section 2.8.1. and 2.9 were used for subsequent experiments. However, due to the viscous nature of the chromatin fraction, it was necessary to subject this fraction to nuclease digestion prior to subsequent manipulation. Micrococcal nuclease (MNase) was selected to treat chromatin fraction samples. This enzyme preferentially cleaves DNA in internucleosomal regions, as well as RNA, and it does not disrupt specific protein-DNA complexes (Horz & Altenburger 1981).

The aim of this experiment was to establish conditions for chromatin digestion. Two conditions were tested, one based on literature use (Torrente et al. 2011; Jeong & Stein 1994) and this was compared with the manufacturer's recommended conditions (https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Product_Information_Sheet/2/n3755pis.pdf), Flag-Upf1^{Wt} cells were harvested from a sub-confluent growing population and sub-cellular fractionation was performed to obtain the chromatin fraction, as previously described in Section 3.2.5. Immunoblotting of the chromatin fraction was undertaken to ensure that contamination from other soluble fractions was minimised (**Figure 5.3**).

In each case, isolated chromatin fraction was resuspended in an equal volume of nuclear isolation buffer (15mM Tris, pH7.5, 15mM NaCl₂, 60mM KCl, 5mM MgCl₂, 1mM CaCl₂ 250mM sucrose, 1x PIC) MNase (3 units, 1 units= 5 µl) together with 1mM CaCl₂, or buffer control, was added to start the indicated reaction (**Figure 4.1**). At the end of the reaction, 2mM of EGTA was added to samples before centrifuged to pellet any residual with chromatin, and the supernatant and pellet fractions separated and subjected to immunoblotting.

As revealed by immunoblotting using an anti-Upf1 antibody (**Figure 4.1**) incubation of the chromatin fraction for 2 mins at 28°C, resulted in the release of approximately 50% of chromatin-associated Upf1. Elevated incubation temperature

(which is expected to increase micrococcal nuclease activity), together with a fivefold increase in incubation time has little or no effect on the proportion of Upf1 released by this treatment. The reason for this is unknown, but may reflect the possibility that Upf1 may bind to other insoluble nuclear components such as the nuclear matrix. As elevated temperature and extended incubation time is likely to be detrimental to the preservation of intact protein-protein interactions, it was decided to use the minimum time and temperature required to release 50% of the isolated material for future experiments.

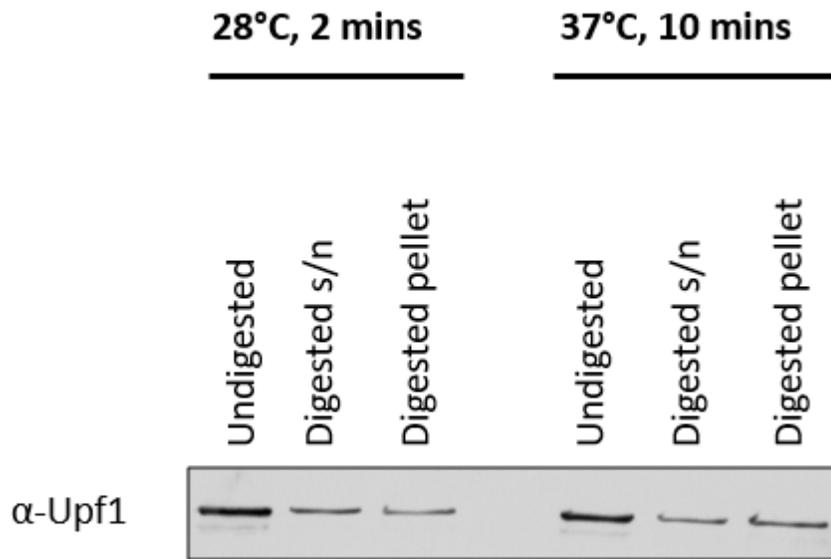


Figure 4.1: Micrococcal nuclease digest of FLAG-Upf1^{Wt} chromatin fraction.

Chromatin fraction of Flag-Upf1^{Wt} were digested at two different conditions. The first chromatin treatment was incubation at 28°C for 2 mins and the second treatment was an incubation at 37°C for 10 mins. One protein sample was first diluted with NIB (15mM Tris, pH7.5, 15mM NaCl₂, 60mM KCl, 5mM MgCl₂, 1mM CaCl₂ 250mM sucrose, 1x PIC), before 2units of MNase and 1mM of CaCl₂ was added. The protein mix was then divided into 2 tubes for individual incubation treatments. At the end of each incubation, 2mM of EGTA was added to stop the reaction and placing on ice. All samples were centrifuged at 1700g for 5 mins, supernatant transferred to a clean tube, and pellet resuspended in an equivalent amount of NIB. All fractions of digestion was kept at -20°C for subsequent analysis. Immunoblotting was performed as described in section 2.8.4 and probed with α- Upf1 antibody (Scottish National Blood Transfusion Service).

4.2.2 Optimisation of Immuno-isolation (IP) conditions for FLAG- tagged Upf1 protein samples

Having established conditions for solubilising chromatin-associated Upf1, it was necessary to identify conditions in which all of the expressed Upf1 could be depleted from a cell lysate using a defined ratio of affinity resin to cell lysate, and to identify conditions required for its elution.

For these steps, I utilised the ANTI-FLAG® M2 affinity resin (Sigma Aldrich) which has a FLAG monoclonal antibody covalently attached to agarose resin. This is expected to allow the specific purification and immune-precipitation of Flag-Upf1 protein that has been expressed in HeLa cells following the addition of doxycycline.

Cell lysate (200µg protein in 1X NIB buffer), produced from 2x 145mm plates of synchronously growing cells (and confluency of approximately 80%) was used in this experiment. The lysate was incubated with 20µl of prewashed M2 affinity resin overnight at 4°C on a rotator. Bound and unbound material was separated by micro-centrifugation (5000g for 30 sec), the supernatant (s/n) removed, and following a wash with 500ul of TBS, the affinity resin was eluted first using 3µL of 5µg/µl 3x Flag peptide dissolved in 100µl TBS (incubated at 4°C on rotator for 30min), with eluted material again separated by micro-centrifugation as above, followed by final elution using 50µl of 10% (w/v) SDS in sample loading buffer (section 2.11) at 100°C. Subsequently all samples were subjected to SDS-PAGE, and immunoblotted using anti-Flag antibodies.

Under these conditions, it was found that this volume of M2 affinity resin was sufficient to specifically deplete all of the doxycycline -induced Upf1 (compare lane 4 with lanes 1, 2, and 3, **Figure 4.2a**). Surprisingly, these data also revealed that FLAG-Upf1 was not eluted using 3x Flag peptide under the conditions used, as no band was observed in the FLAG-Upf1 eluate lane (**Figure 4.2a**, lane 6). However, the Flag-Upf1 was recovered by elution using 10% SDS at elevated temperature (**Figure 4.2a**, lane 8), and intensities of the eluted band compared to that of the input material indicated that the recovery approached 100%. This meant that the

FLAG-Upf1 protein was still bound to the beads and could not be eluted using the competitive elution method.

In order to establish whether the apparent inability to elute Flag-Upf1 was a consequence of the methodology adopted, or specific to the tagged protein under investigation, I compared the efficiency of elution using the same protocol as described above for a distinct Flag-tagged gene product (stem loop binding protein (SLBP), (Panamwan 2017)). Flag-SLBP containing 2×10^6 HeLa cells were seeded in 100 cm dishes. After 24 h, 0.7 $\mu\text{g}/\text{ml}$ doxycycline was added to induce protein expression, and after a further 24 hours, 2mM HU was added to arrest cells in S phase (Maquat 2006). Subsequently cells were harvested, lysates prepared according to section 2.8.1 and subjected to the immune-isolation protocol described above.

As observed in **Figure 4.2b**, incubation of the M2 affinity resin with a working concentration of 300ng/ μl 3x Flag peptide resulted in the specific elution of a protein (compare Lane 5 with lanes 1-4) at the expected molecular-weight size for FLAG-SLBP (approximately 38kD, data not shown) while no further material was observed with subsequent elution using 10% SDS.

The data confirmed that competitive peptide elution using the batch of antibody matrix worked efficiently in my hands with at least one other protein, and that the inability to elute Upf1 under similar conditions must reflect some aspect of the structural configuration of the epitope tag at the N-terminus of the protein, and its interaction with anti-Flag antibodies.

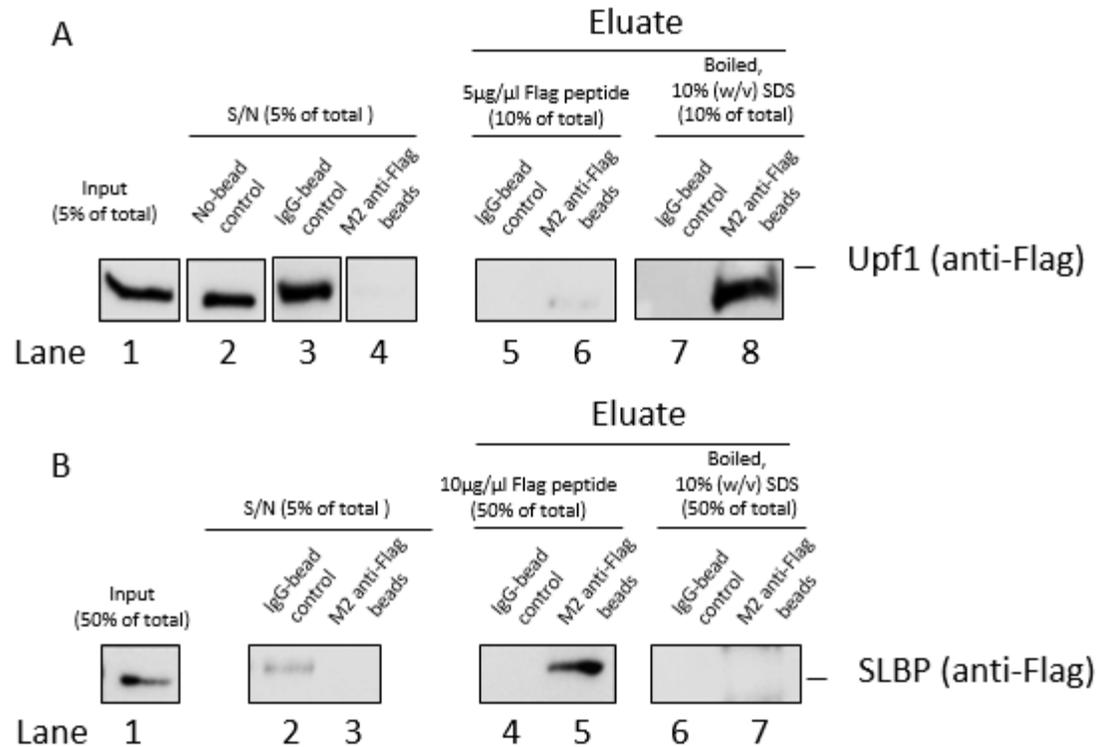


Figure 4.2: Immuno-precipitation of FLAG proteins by 3x FLAG peptide

(a) FLAG-Upf1^{Wt} cells were grown and whole cell lysate was prepared as described in section 2.8.1. 1mg of protein lysate in 38 μ l (Lane 1, 5% of total) was used for IP using 20 μ l of either anti-FLAG M2 beads (lane 4) or as control, IgG beads (lanes 3), and in each case, bound material was eluted first by application of 100 μ l of (150ng/ μ l) 3x Flag peptide (lanes 5 and 6), followed by boiling of beads in 50 μ l of sample loading buffer containing 10% SDS (lanes 7 and 8). (b) FLAG-SLBP cells were grown and whole cell lysate was prepared as above. 1mg of protein lysate in 97 μ l was used for IP exactly as described in (a) above, except that the proportion of material subjected to electrophoresis and immunoblotting was 50% to enable detection of SLBP.

4.2.3 Optimisation of bead volume for Immuno-precipitation (IP)

Having established conditions for solubilising chromatin-associated Upf1, it was necessary to identify conditions in which all of the expressed Upf1 could be depleted from a cell lysate using a defined ratio of affinity resin to cell lysate, and to identify conditions required for its elution.

For these steps, I utilised the ANTI-FLAG® M2 affinity resin (Sigma Aldrich) which has a FLAG monoclonal antibody covalently attached to agarose resin. This is expected to allow the specific purification and immune-precipitation of Flag-Upf1 protein that has been expressed in HeLa cells following the addition of doxycycline.

Cell lysate produced from 5 T175 flasks of synchronously growing cells (and confluency of approximately 80%) was used in this experiment. Different amounts of protein lysate was used (0.5mg, 1.0mg & 1.2mg) however the volume of ANTI-FLAG M2 beads remained constant at 25µl. Lysates were incubated overnight at 4°C on a rotator. Bound and unbound material was separated by micro-centrifugation (5000g for 30 sec), the supernatant (s/n) removed, and following a wash with 500ul of TBS buffer, the affinity resin was eluted by boiling beads in 50µl of 10% (w/v) SDS in sample loading buffer. Subsequently all samples were subjected to SDS-PAGE, and immunoblotted using anti-Flag antibodies.

Under these conditions, it was found that this volume of M2 affinity resin was sufficient to specifically deplete all of the doxycycline -induced Upf1 (lanes 3-5), with recovery of ~80-90% for all samples. There was no residual Upf1 band present in the supernatant, indicating that all tagged Upf1 could be depleted from a cell lysate, and therefore that lysates do not contain a fraction of this protein which is refractory to immunoprecipitation as a consequence, for example, of being part of a larger multi-protein complex (**Figure 4.3**- lanes 6-8). In addition, however, the data suggested that beads were still not fully saturated under these conditions. Saturating the beads in principle would ensure that I have manage to capture all FLAG-tagged Upf1 that is present in the sample and are bound to the antibody present on beads.

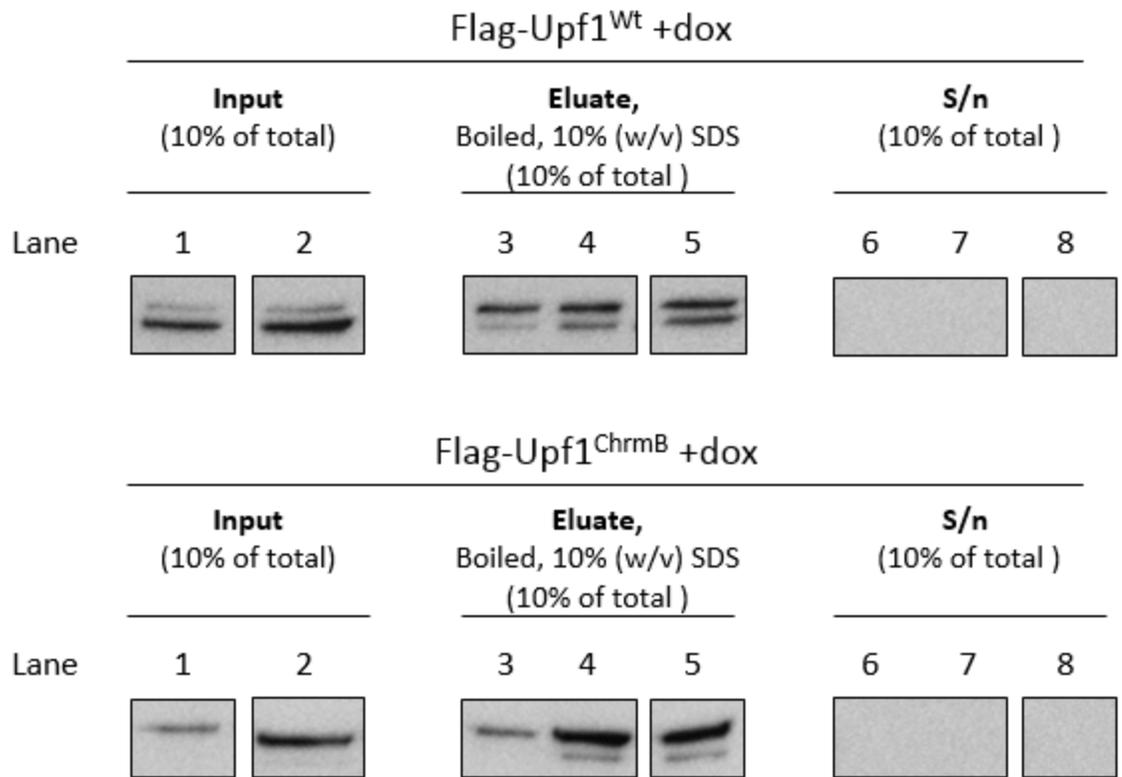


Figure 4.3: Optimisation of Anti-FLAG M2 beads to maximise recovery of FLAG-Upf1 protein

FLAG-Upf1^{Wt} whole cell lysate grown in the presence of doxycycline (induced FLAG expression) was used per pull down experiment. The amount of ANTI-FLAG M2 beads used was 25µl and the amount of protein used for each pull down reaction varied from (**lanes 1 & 3**- 0.5mg, **lanes 2 & 5**- 1.2mg and **lane 4**- 1.0mg). Boiled elution was performed for protein retrieval. 10% of input, eluate and supernatant (s/n) was used for visualisation using western blot.

4.2.4 Optimisation of protein concentration for Mass spectrometry (MS) purpose

Azzalin et al. (2007) have reported that some components of the Upf1-mediated NMD pathway that are known to directly interact with Upf1 (Upf2, SMG1 and SMG5) are also associated with chromatin-associated Upf1. This suggests that the mechanism by which Upf1 executes its chromatin-associated function has some similarities with its mechanism in NMD. The results presented in the previous chapter indicated that, Upf1^{ChromB} associates with, and acts to unwind, duplex DNA. It follows that the inability of this mutant to bind chromatin cannot be explained by loss of DNA binding, and an alternative hypothesis is that the association of Upf1 with chromatin is mediated by another protein, distinct from those Upf1-interacting proteins that are engaged in NMD.

I wished therefore to compare the interactome of Upf1-isolated from a cytoplasmic fraction (involved in NMD) with that of the chromatin-associated form of the protein, to test the hypothesis that Upf1 associates with chromatin via a distinct chromatin-specific protein or proteins, and that a quantitative comparison of the composition of both Upf1 interactomes would identify molecular differences.

In order to do this, I wished to carry out a shotgun proteomics approach to identify proteins specifically associated with populations of Upf1. This approach enables the identification of proteins in a complex mixture using proteolytic digestion to generate peptides followed by a combination of high-performance liquid chromatography to fractionate the peptide mixture, together with the use of tandem mass spectrometry to identify individual peptides (Wolters et al. 2001). The use of shotgun proteomics to characterise that set of polypeptides that interact with a protein of interest is well established (Ito et al. 2001). Software has been developed for the automated analysis of mass spectra providing high-resolution, accurate quantitative data. This coupled with an availability of genomic databases enables highly accurate automated identification of specific peptides (Cox & Mann 2008). More recently, the development of statistical packages that enable determination of the extent of *enrichment* of a protein that appears to associate with the protein of interest (sometimes termed the *bait*), based on

mass spectrometric quantitation of peptides derived from that protein, have facilitated the use of label-free methods for the identification of specific interacting proteins (Tyanova, Temu, Sinitcyn, Carlson, Marco Y. Hein, et al. 2016). The Perseus platform allows for the calculation of the degree of enrichment, compared to an appropriate control sample, together with a statistical analysis of the significance of the data for every identified protein.

In order to determine the scale of biological material required to undertake quantitative proteomics of Upf1 interacting proteins from both chromatin and cytoplasmic sources, an initial experiment was undertaken using cell numbers which have been used previously in similar experiments (Flury et al. 2014; Méndez & Stillman 2000). To this end, the following experiment was undertaken in triplicate. FLAG-Upf1^{Wt} HeLa cells (7×10^7) or FLAG-Upf1^{ChromB} cells (8×10^7) were harvested from sub-confluent culture, after exposure to 0.7 µg/ml doxycycline, or mock treatment, for 48 hr prior to harvest (**Figure 4.4**). The chromatin-associated protein fraction was prepared from FLAG-Upf1^{Wt} HeLa cells, as described in Section 2.9, followed by digestion with MNase, as discussed in Section 4.2.1. Additionally, whole cell lysate was prepared from FLAG-Upf1^{ChromB} cells as described in section 2.8.1, which was clarified to remove the nuclear fraction. All subsequent steps were performed using each of the relevant cellular fractions indicated in **Figure 4.4** in either of the indicated conditions.

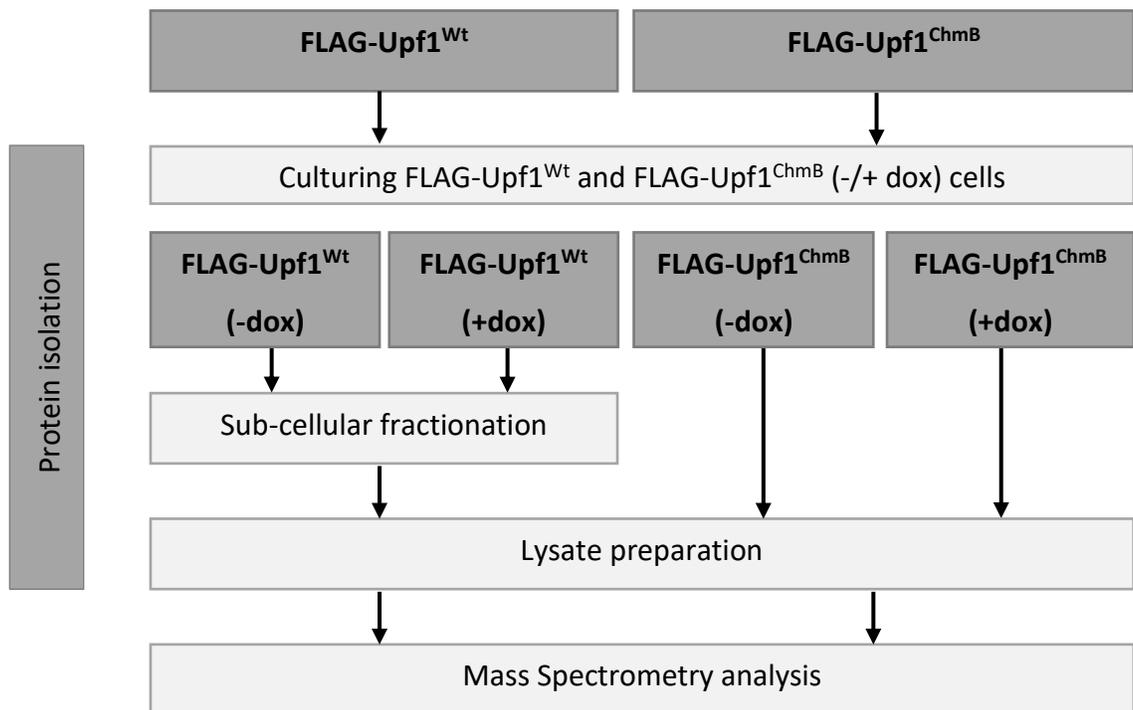


Figure 4.4: Schematic workflow from cell growth until analysis for each cell line

FLAG-Upf1^{Wt} HeLa cells (7×10^7 cells) or FLAG-Upf1^{ChmB} cells (8×10^7 cells) were harvested from sub-confluent culture, after exposure to $0.7\mu\text{g/ml}$ doxycycline, or mock treatment, for 48 hr prior to harvest. 2mM of HU was added to cells for 24 hours to arrest cells at S-phase and was washed out before releasing cells into S-phase for 3 hrs in new media. For FLAG-Upf1^{Wt} HeLa cells, a subcellular fractionation was performed according to section 2.9, while whole cell lysates from FLAG-Upf1^{ChmB} cells undertaken using procedure described in section 2.8.1. Once lysates have been purified using immunoprecipitation, samples were then ready to be subjected for mass spectrometry analysis.

Following the preparation of cell lysates, protein concentrations of each sample were determined using the Bradford protein assay (Section 2.8.2) and the immunoprecipitation protocol established in section 4.2.2 was independently undertaken for three biological replicates of the four conditions shown in **Figure 4.3**.

Aliquots of M2 affinity resin treated with cell fractions either mock-treated or containing doxycycline-induced Flag-Upf1 were eluted with 10% (w/v) SDS as described above. 10% of each Flag-Upf1 -containing eluate was assessed by western blot (**Figure 4.5 a & b**). All replicates, showed approximately 80-90% recovery, with the exception of replicate 3 for FLAG-Upf1^{ChrmB} where recovery was less than 30% (**Figure 4.5b**). Separately another immunoprecipitation experiment using the conditions required for FLAG-Upf1^{ChrmB} replicate 3 was undertaken and processed together for MS analysis (**Figure 4.5c**).

In two out of the six Upf1 isolations undertaken, immunoblotting with an anti-Flag antibody revealed a Upf1 doublet. The reason for the appearance of a doublet in a subset of samples both from the cytoplasmic fraction and the chromatin fraction is not understood. Upf1 exists as two isoforms (<http://www.uniprot.org/uniprot/Q92900#sequences>) although the immunoblotting used here would only be expected to identify isoform 2 which has been used exclusively in analyses of Upf1 function, is the form used to generate doxycycline inducible cell lines, and was the isoform used previously in functional analyses of cytoplasmic and nuclear Upf1 function (Turton 2014). However, should the additional bands in lanes in **Figure 4.5 (a)**-1 & 5 and **(b)** 1 & 4 be a consequence of the presence of a phospho- form of Upf1, no obvious difference in preparation of samples can explain the apparent variation.

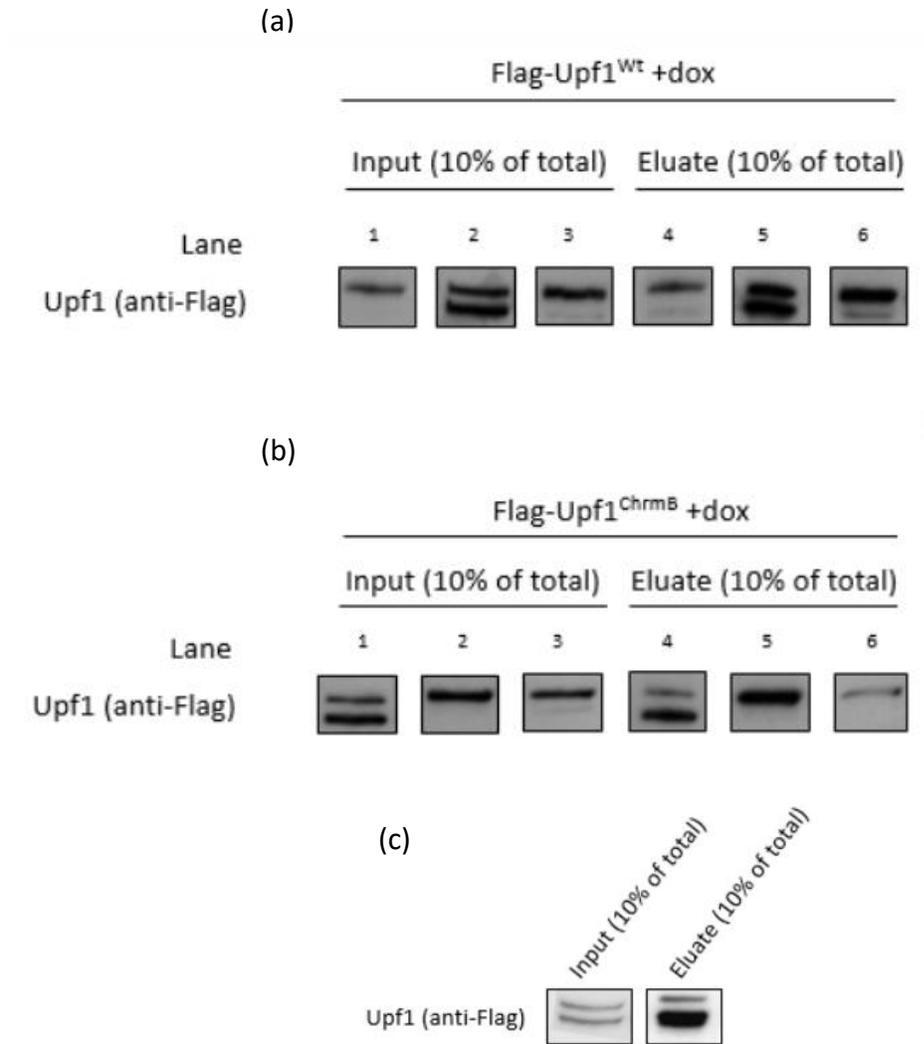


Figure 4.5: Immuno-precipitation of FLAG-Upf1 for triplicate biological samples

IP using chromatin fraction of (a) FLAG-Upf1^{Wt} and whole cell lysate of (b) FLAG-Upf1^{ChmB} cell lines for all three biological replicates. 400µg of protein lysate was used for IP and proteins were eluted by boiling in sample buffer. Protein recovered was observed by immunoblotting using anti Flag antibody. (c) IP of new replicate 3 for FLAG-Upf1^{ChmB} cell line to replace the one in lane 6 of (b). Doublets observed in some of the lanes could be due to the different isoforms present for Upf1. However further possibilities are given in later chapter.

4.2.5 Pilot mass spectrometry (MS) analysis of FLAG-Upf1^{Wt} and FLAG-Upf1^{ChrmB} and associated proteins

When undertaking mass spectrometric analysis, there are two main proteomic strategies: top-down and bottom-up analyses. The top-down method analyses whole proteins, whereas bottom-up investigates peptides from proteolytically digested proteins (Feist & Hummon 2015).

A bottom-up, label-free MS method was used in this study, as it facilitates greater coverage of proteins present in a complex mixture, while being very cost-effective. In a typical protocol (**Figure 5.2**) proteins are first digested into peptides using a sequence-specific endoprotease. Here trypsin was used in combination with digestion of protein fractions after denaturation following SDS-PAGE. Peptide mixtures are resolved using reverse phase HPLC and then are analysed using MS. Spectral data are stored and peptides assigned to their cognate proteins *in silico* by comparing identified peptides to a library of peptide MS/MS spectra (Eriksson & Fenyö 2008; Hein 2014).

The bottom-up method however does have a few disadvantages as digestion could cause problems such as lost of labile PTMs or sequence variants due to inability to detect regions of proteins and also modifications or sequence variations may occur on disparate peptides, causing their relation to one another to be lost (Catherman et al. 2014). Top-down approach could actually eliminate these problem as no digestion is required prior to MS analysis. In this approach, intact protein is introduced to the mass spectrometer which allows both intact and fragment ion masses to be measured (Catherman et al. 2014). However, due to the complexity of the sample used in this experiment the top-down method would not be able to provide data as comprehensive as the bottom-up method in terms of proteome coverage, sensitivity, and throughput.

During sample preparation, all of the 12 separate immunoprecipitates described above were subjected to reduction using tris(2-carboxyethyl) phosphine hydrochloride (TCEP), incubating for 10 mins at 70°C and alkylation using 2mM iodoacetamide (IAA) (30

mins at room temperature, in dark) (Gundry et al. 2009) to reduce any disulphide bonds present and prevent free cysteines from forming or reforming disulphide linkages. This step increases the accessibility of protease during digestion (Gundry et al. 2009), as well as ensuring that resultant peptides may be identified by mass spectrometry software, which relies on the expectation that peptides generated during digestion are linear fragments. Reduced and alkylated samples were subjected to SDS-PAGE as described in Section 2.13.1 for protein separation, and polypeptides were visualised by staining with Coomassie blue (**Figure 4.6**). Eluates containing expressed FLAG-Upf1 (**Figure 4.6** lanes 4-6, 10-12; +doxycycline) showed a very faint band between 100-150kDa mark (Arrow Upf1= 124kDa), which was not observed in eluates derived from cell fractions where Flag-Upf1 expression was not induced (**Figure 4.6** lanes 1-3, 7-9; - doxycycline). All lanes showed significant levels of a band at 50 kDa and 25 kDa, which correspond to the sizes expected of the heavy and light chains of IgG. While in principle, the M2 affinity resin is prepared using an anti-Flag monoclonal antibody which is chemically cross-linked to the agarose matrix, the cross-linking efficiency appears to be less than 100%, and elution with detergent results in some co-elution of antibody polypeptides (P. Panomwan, personal communication).

Gels were stained with Coomassie blue for visualisation of protein bands. Noticeably, neither Upf1 nor any putative interacting proteins were the major bands observed. Most of the polypeptides observed appeared to be present at similar levels in both eluates derived from doxycycline-treated and mock-treated cells. This is consistent with the notion that the majority of polypeptide bands observed in **Figure 4.6** represent non-specific low affinity association of proteins with the agarose matrix, and that the numbers of cells used to prepare lysates for this experiment may not contain sufficient Upf1 to saturate the anti-FLAG antibody binding sites on the matrix.

The polyacrylamide gels shown in **Figure 4.6** were sliced into individual lanes, and each slice further sub-divided into segments A, B and C, as indicated by the lines drawn across each gel in **Figure 4.6**. Each segment was cut into smaller gel fragments as described in Section 2.13.1. Gel slices were destained with 50% acetonitrile (ACN) and

digested with trypsin (1ng/ μ l) overnight in 37°C. Trypsin diffuses into the gel pores and cleaves the protein into peptides which is then extracted from the gel (Gundry et al. 2009). After several washes with 0.5% formic acid and 100% ACN, supernatant was removed by centrifugation and prepared for LC/MS/MS analysis using a Dionex 3000 HPLC upstream of an Orbitrap Elite mass spectrometer.

Peptide fragment spectra were assigned using MaxQuant software (Cox & Mann 2008) and the resulting fragments and searched against human Uniprot complete proteome set. All data was then analysed using the Perseus statistical package to establish the extent of enrichment (summed ratio of peptides detected for any given protein from sample prepared in the presence versus absence of doxycycline) plotted with associated statistical determination of significance for each protein identified. (Tyanova, Temu, Sinitcyn, Carlson, Marco Y Hein, et al. 2016). The data for the samples shown in **Figure 4.6** is displayed as a semi-log volcano plot in **Figure 4.7a, b & c**. In each case, the X axis shows the fold enrichment (where +2 indicated a two-fold enrichment of a protein in samples to which doxycycline was added, and -2 indicates two-fold enrichment of a protein in samples where no doxycycline was present, prior to sample processing). Thus proteins that appear at X axis values less than 0, were more abundant in the uninduced sample compared to the corresponding expressed FLAG-Upf1 sample. The Y axis is the negative log of p value obtained following a t-test, using the abundance data for each protein from each of the three biological replicates for each experimental condition. The solid black line indicates a weighted threshold of significance, which varies as a function of the absolute difference of mean values between experimental and control samples (Goss Tusher et al. 2001). At $S(0) = 0$, the threshold of significance is entirely dictated by p value, while non zero values of $S(0)$ allow a weighting to account for absolute differences in mean values of abundance.

As shown in **Figure 4.7** (a & b) each point represents an identified protein identified. Q92900 (indicated in red) represents Upf1. As indicated, Upf1 is located well below the significant threshold in both FLAG-Upf1^{Wt} and FLAG-Upf1^{ChmB} samples. This was not expected, as Flag-tag affinity isolation following addition of doxycycline should

have facilitated significant enrichment of FLAG-Upf1 protein. In these circumstances, it is not possible to identify additional proteins that are specifically and significantly enriched in addition to Upf1 in each case. **Figure 4.7c** shows the volcano plot comparing the comparative abundance data for FLAG-Upf1^{Wt} associated proteins compared with FLAG-Upf1^{ChmB} associated proteins. Q92900 is positioned corresponding to an X axis value of zero, indicating that the abundance of Upf1 in each sample was equal, confirming that the semi-quantitative estimates of material obtained by immunoblotting above are correct. That there are proteins in this plot that are significantly enriched in each sample reflects the fact that because of the nature of this specific experiment, the Flag-Upf1 in each case is derived from a distinct cellular source, and thus the identity of non-specifically bound proteins derived from each source are likely to be differentially enriched to some degree.

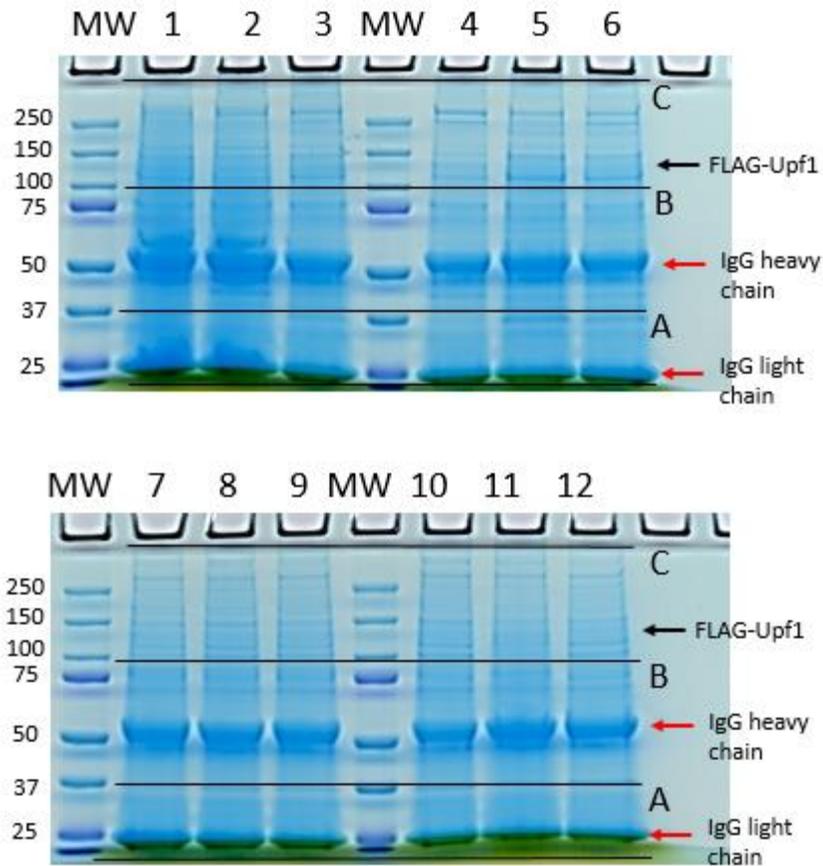


Figure 4.6: Analysis of FLAG-Upf1 protein by SDS-PAGE prior to in-gel digest

Equal volumes (45 μ l) of each of the 10% SDS eluates obtained from the indicated cells using the IP protocol described in section 4.2.2 were subjected to SDS-PAGE. MW: SeeBlue Plus2 Prestained Protein Standard, sizes are indicated in kDa. **Lanes 1-3:** replicates of Flag-Upf1^{Wt} (-dox), **lanes 4-6:** replicates of Flag-Upf1^{Wt} (+dox), **lanes 7-9:** replicates of Flag-Upf1^{ChmB} (-dox) and **lanes 10-12:** replicates of Flag-Upf1^{ChmB} (+dox)

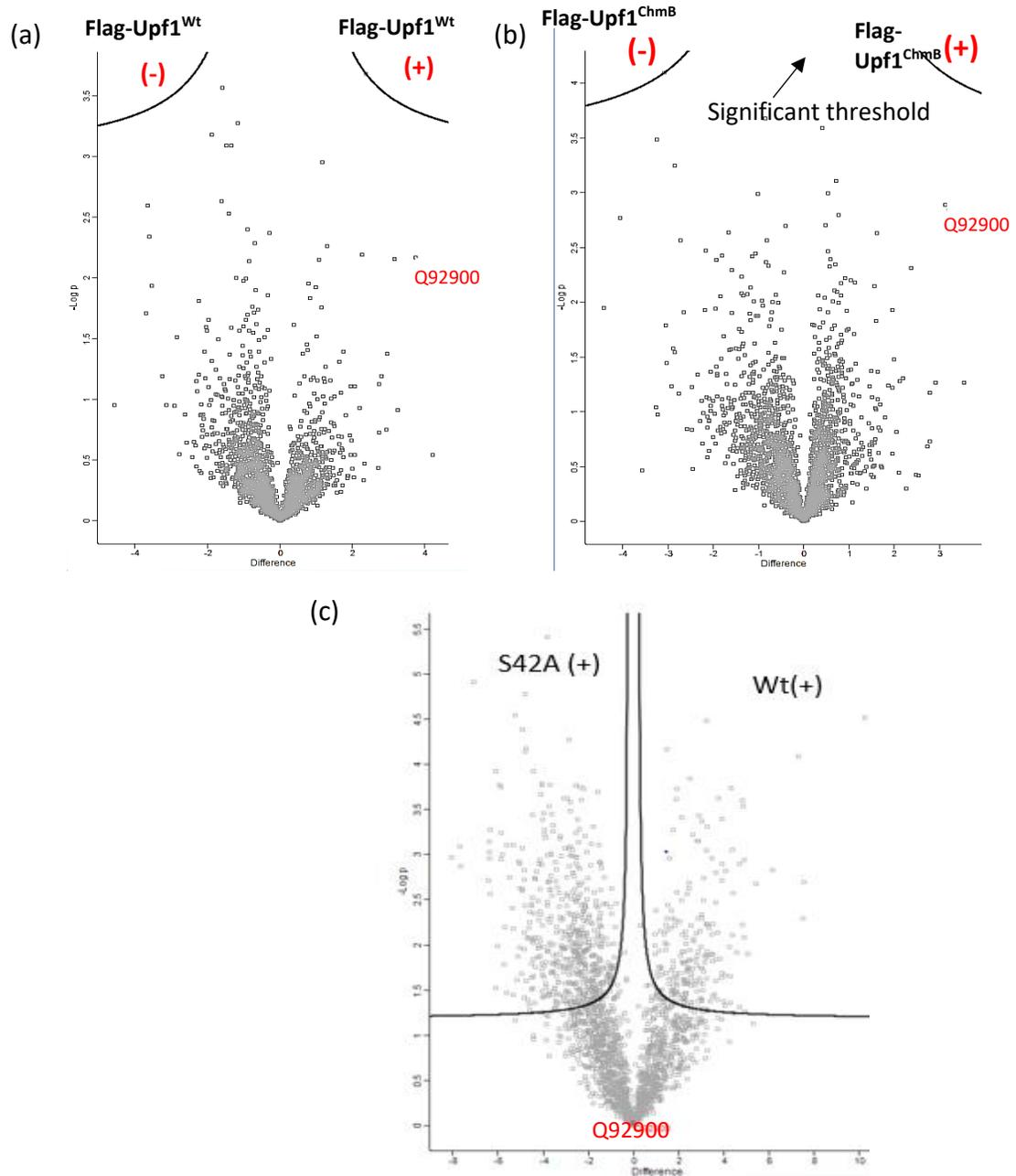


Figure 4.7: Volcano plot of detected proteins for FLAG-tagged Upf1 samples.

Once peptide is put through the LC-MS/MS, identified peptides corresponding to proteins are listed after data processing using MaxQuant software. To further identify and visualise the interactors according to significance, a volcano plot is derived from the analysis software Perseus. (a) Analysis for protein interaction in Flag-Upf1^{Wt} sample in

the absence or presence of doxycycline. (b) Analysis of the protein interaction in Flag-Upf1^{ChmB} sample between uninduced and induced sample (-/+ dox). (c) Analysis of the interacting proteins identified between induced Flag-Upf1^{Wt} and Flag-Upf1^{ChmB}. Y axis represents the $-\text{Log } p$ T test and X axis represents the fold change (s0) difference. Black curves (arrow) represents the significant threshold and Q92900 represents Upf1. The setting used for false discovery rate (FDR): 0.05 and s0: 0.1.

4.3 Discussion

The main aim of this section of work was to establish conditions and methodology for the immuno-precipitation of FLAG-Upf1 protein from both cytosolic and chromatin compartments of the cell, in a manner likely to preserve protein-protein interactions in order to determine differences in Upf1 interactors depending on its cellular location.

The genomes of the eukaryotic organism are packaged into chromatin where histones and DNA are packed into chromatin fibres (Chung et al. 2010; Kornberg 1974). The insoluble and viscous nature of chromatin creates challenges for sample handling and processing, in addition to removal of tightly bound proteins from DNA. Therefore it needs to be digested making it more accessible. In this study I used MNase based digestion to prepare a soluble fraction derived from each chromatin sample. MNase cleaves both single-stranded and double-stranded DNA and RNA, yielding mono and oligonucleotides with 3'-phosphates. Its mode of digest is also site specific which preferentially cuts at A-T rich regions (Dingwall et al. 1981; Horz & Altenburger 1981).

In section 4.2.1, the chromatin fraction which contained bound FLAG-Upf1^{Wt} was digested using two different conditions that have been used widely to solubilise chromatin-associated proteins (Torrente et al. 2011; Jeong & Stein 1994). Based on the results obtained here, I concluded that both conditions works equally well. All subsequent digestion steps used for chromatin fractions utilised a 10 min incubation at 28°C to minimise exposure of biological samples to higher temperatures.

Once soluble sample materials were available for further analysis, it was necessary to establish optimal conditions for immune-isolation of Upf1 from both chromatin and cytoplasmic fractions. IP is a well-established technique (Kaboord & Perr 2008) that has been used widely for protein purification.

Published work (Melero et al. 2014) together with previous observations in the Smythe laboratory (Panamwan 2017) suggested that competitive elution using 150 ng/ μ l 3x Flag peptide would facilitate specific elution, yielding a purer Upf1 preparation,

while at the same time allowing for the reuse of M2 affinity matrix. However based on **Figure 4.2a** it appeared that FLAG-Upf1 was unable to be eluted using this method. Before I proceeded to elute protein by boiling in sample loading buffer, I attempted several alternative methods including those recommended by the manufacturer.

The first strategy was to increase the final concentration of 3x Flag peptide that was recommended from 150ng/ μ l to 300ng/ μ l. However, this increment of peptide concentration also failed to elute Upf1. The effect of adding β -mercaptoethanol (BME) together with the 3x Flag peptide during the elution step was also attempted. Theoretically, reducing agents such as DTT and BME break disulphide bonds between the heavy and light chains of an antibody, and that might be expected to facilitate protein elution. This too failed to elute FLAG-Upf1.

The reason that Flag-peptide-mediated elution failed to yield any Upf1 is unclear, although conceivably, the steric arrangement of the Flag-tag with respect to the antibody might be such that binding of the protein precluded access of the peptide to the antibody binding site. Due to time limitations, I chose not to consider the construction of an alternative tagging affinity purification system such as utilising a HA-tag. Like the FLAG-tag, HA-tag is also a small-size tag where it could be beneficial as it would minimise any effects on structure, activity and characteristics of the recombinant protein and usually does not have to be removed for in vivo functionality (Zhao et al. 2013).

Despite the use of recommended concentrations of salt and detergent in lysozyme washing buffer is to minimise non-specific material bound directly to the matrix, nonetheless there is a significant remaining non-specific binding as evidenced by the complexity of protein bands observed in Figure 4.5. Elution by protein denaturation using 10% SDS, which generated a high yield of FLAG-Upf1 protein, does not enable differential elution of specific versus non-specifically bound proteins and consequently yields less pure material. An obvious disadvantage with this approach is the extent to which peptides derived from irrelevant non-specific proteins might mask signals from

low abundance specific Upf1 interactors. This point will be discussed further in the next chapter.

The next step was to establish an appropriate amount of chromatin derived protein material fraction for IP that would generate significant data after MS analysis. a range of publications have used between 60µg-600µg for chromatin sample per immunoprecipitation sample. Therefore, 400µg of protein was chosen for the initial runs of IP.

In this experiment, western blots of IP performed revealed that recover almost all of the FLAG-Upf1 protein was recovered in both samples. Surprisingly, one of each sample replicates contains a doublet when probed with anti-FLAG antibody. Upf1 is a phospho-protein, and phosphorylation is known to bring about electrophoretic band shifts (Yamashita et al. 2001). Similarly, Upf1 contains a ubiquitin ligase RING domain and recently has been shown to contain ligase activity (Feng et al. 2017). It is possible that Upf1 may be a substrate for itself or be a target for another modification via this or a similar mechanism. Perhaps the most likely explanation is that Upf1 may be susceptible to C-terminal targeted proteolysis, and that slight differences in sample preparation conditions resulted in partial proteolysis of the protein. As all of these explanations were potentially resolvable by analysis of the protein by mass spectrometry, the samples were processed for MS analysis as described below.

MS analysis using the prepared samples had generated data that was analysed statistically using Perseus, a plug-in of the MaxQuant software. This software performs bioinformatics and statistical analysis, while also allowing visualisation of analysed data. A comprehensive analysis of MS data is discussed further in the next chapter. However, here I only extracted the results that was obtained from the Perseus generated volcano plot.

The main objective in utilising a volcano plot is to identify the interactors of a specific “bait” protein (FLAG-Upf1) in the presence of background proteins. It uses a

student's t-test and compares it to the label free quantitation (ie ion intensities) of peptides derived from all proteins identified in replicates of both bait and control samples (Singh et al. 2016). Singh et al. (2016) explains that non-specific background binding proteins are expected to centre around a value equal to 0, when the differences between logarithmic mean protein intensities between bait and control are plotted against the negative p-values derived from the statistical test. Enriched interactors would appear on the right section of the plot, with highest confidence for true interaction found in the upper right quadrant.

Figure 4.7 (a & b) are volcano plots that were derived from replicates of each protein sample (FLAG-Upf1^{Wt} and FLAG-Upf1^{ChmB}). Ideally, Q92900 which is the protein ID for Upf1 should be enriched and appears to be above the significant threshold line. However, this was not observed for both samples in each case. Choi et al. (2012) has noted that a low bait abundance in some replicates has an overall negative effect which includes “dilution” of signal, and loss of *bona fide* interaction, and recommends that in such circumstances the sample replicate be replaced.

Low expression of FLAG-Upf1 protein could be one of the reasons why the enrichment of Upf1 protein observed in samples derived from doxycycline treated cells was only about 10-fold compared to samples from uninduced cells. Although I have optimised the concentration levels of doxycycline used to express FLAG-Upf1 protein, it seems that the expression level was not significant enough. In all subsequent protein expression preparations, I increased the doxycycline concentration from 0.7 µg/ml to 1.0 µg/ml.

A second possibility was the enrichment during IP was insufficient. It is conceivable that the agarose bead volume used was in excess for the amount of protein lysate used, with the consequence of a relatively large degree of non-specific protein binding compared to the specific association of UPF1 with the antibody present on the beads. Although Western blotting data indicated efficient recovery of bound Upf1 (**Figure 4.2**), this data alone does not provide quantitative indication of enrichment. The presence of

excess beads during IP might be expected to result in higher background protein identification in mass spectrometry that could mask true protein interactors, especially if the stoichiometry of interaction is low.

One obvious explanation for a low level of enrichment might be that the volume of agarose beads used in the experiment above was disproportionately large for the amount of UPF1-containing protein material used. Based on literature values and manufacturer recommendations, I chose 400µg of protein lysate to be used for IP. of Upf1^{ChrmB} (and of an equivalent amount of chromatin-derived protein containing the same amount of Upf1^{Wt}). However, based on the results obtained from MS analysis described above, it is clear that a larger protein amount is required in order to obtain statistically significant enriched quantities of UPF1 together with its associated interacting proteins. Hence, subsequently the amounts of protein lysate projected for use was increased from 400µg to 4mg.

CHAPTER 5

5.0 Identifying Upf1 protein interacting partners

5.1 Introduction

Polypeptides may form stable complexes to generate quaternary structures that are essential for overall molecular function (such as phosphorylase kinase, or the ribosome (Berggård et al. 2007; Rao et al. 2014) or more frequently, transiently with other proteins in the execution of complex cellular function. Examples of transient interactions include signalling pathways (Rao et al. 2014; Smaczniak et al. 2012). Over 80% of polypeptides do not function in isolation but instead operate within a complex (Berggård et al. 2007). For example, NMD is triggered by a weak transient interaction of Upf1 with the messenger RNA 5' cap protein CBP80 which in turn triggers the phosphorylation of Upf1 by the phosphatidylinositol 3 kinase (PIKK)-related kinase Smg1 and the interaction of these polypeptides with release factor components of the ribosome. When Upf2 associated with components of the exon junction complex interacts with Upf1 this results in the formation of the DECID complex (Deniaud et al. 2015; Bono 2014; Okada-Katsuhata et al. 2012) which results in specific messenger RNA decay. By uncovering interaction networks of a given polypeptide protein, it creates the opportunity to acquire mechanistic information regarding molecular function of an unknown component. In principle, the functionality of unknown proteins may be inferred, based on evidence of their interaction with other proteins of known function.

As discussed previously, Upf1 is not exclusively involved in NMD. It also has roles in Staufen1-mediated mRNA decay (SMD) (Isken & Maquat 2008; Park et al. 2013; Park & Maquat 2013), histone mRNA (Maquat 2006; Müller et al. 2007) and cell cycle progression and genome stability (Azzalin & Lingner 2006b; Varsally & Brogna 2012). However, with the exception of its role in nonsense-mediated decay, limited biochemical information is available for these additional roles of Upf1, especially in regards to its role in genome stability. Hence it is the aim of the work described in this

chapter to identify a protein interaction network for that population of Upf1, which associates with chromatin and regulates genome stability.

Several approaches have been developed to look at protein-protein interactions either by *in vitro* (protein microarray, tandem affinity purification-mass spectrometry), *in vivo* (Yeast 2 hybrid) or, *in silico* (phylogenetic tree, gene neighbour) approaches (Rao et al. 2014). This study utilises an affinity purification-mass spectrometry approach to identify and investigate Upf1 interactomes. This method couples immune-isolation (IP) which allows the enrichment of FLAG-tagged proteins and later analysing the sample using mass spectrometry (MS).

In previous studies (Flury et al. 2014; Schweingruber et al. 2016), MS has been used as the method of identifying proteins associated with Upf1 and with other known components of the NMD pathway. Flury et al. utilised the stable isotope labelling (SILAC) method in cell culture experiments to determine Upf1 interactors by quantitative proteomics. Schweingruber's group on the other hand utilised BioID to detect protein-protein interactions. This strategy uses a biotin-protein ligase BirA that is fused to bait proteins which biotinylate proteins that are in close proximity (~50 nm). Both studies used whole cell lysate for analysis. However, the aim of this study was to focus specifically on nuclear proteins expected to be involved in the chromosome-associated functions of Upf1, such as DNA replication and repair (Chawla & Azzalin 2009; Chawla et al. 2011; Azzalin, Reichenbach, Khoraiuli, Giulotto & Lingner 2007; Dehghani-tafti & Sanders 2017). Hence, chromatin fractions were used to isolate chromosome associated Upf1, for comparison with a mutant form of Upf1 (Upf1^{ChmB}) that is incapable of binding chromatin.

In the analysis discussed below, 4 biological samples derived from 2 specific isogenic cell lines (HeLa FLAG-Upf1^{Wt} and HeLa FLAG-Upf1^{ChmB}) were analysed as described previously in Chapter 4. In the case of each cell line, cells were either mock-treated or exposed to doxycycline, the latter treatment to enable the specific expression of the relevant Flag-Upf1 protein form (ie either wild-type or non-chromatin binding).

As before, the protein content derived from micrococcal nuclease digestion of the chromatin fraction from FLAG-Upf1^{Wt} cells, and protein lysates from whole cell extracts of FLAG-Upf1^{ChmB} cells were used in the following experiments. To facilitate the detection of common contaminants uninduced samples were included (Mellacheruvu et al. 2013). Protein samples were prepared according to the immuno-precipitation protocol optimised in chapter 4 with the workflow undertaken according to **Figure 5.1**.

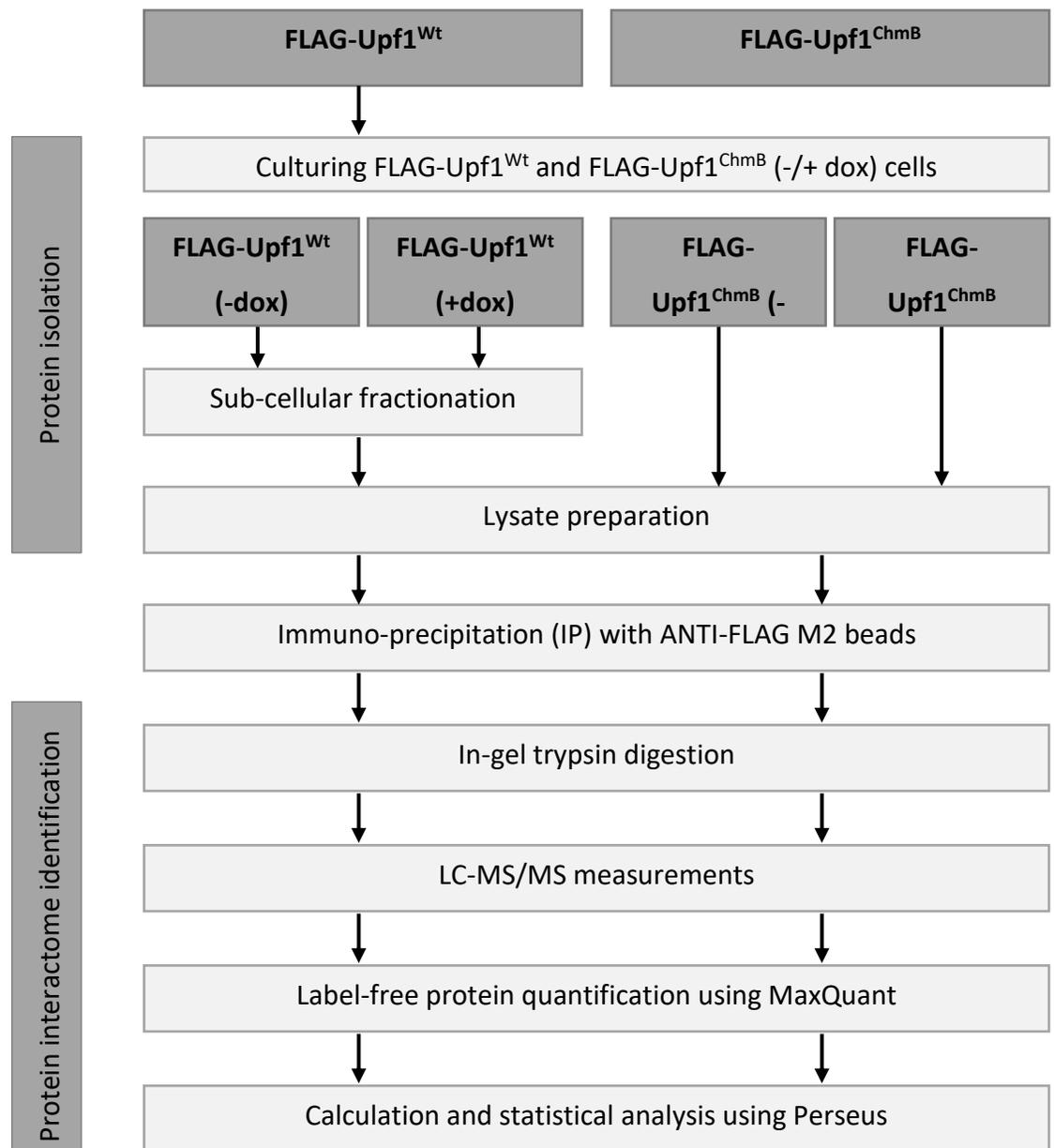


Figure 5.1: Simplified experimental and data analysis workflow that is used in this study

Flag-Upf1^{Wt} and FLAG-Upf1^{ChmB} cells were grown in the presence or absence of doxycycline as described in section 2.7 to induce FLAG protein expression. The chromatin fraction from FLAG-Upf1^{Wt} cells and whole cell lysate (WCL) from FLAG-Upf1^{ChmB} cells was prepared. Protein lysates were then clarified and enriched by immuno-precipitation before gel electrophoresis and LC-MS/MS for analysis.

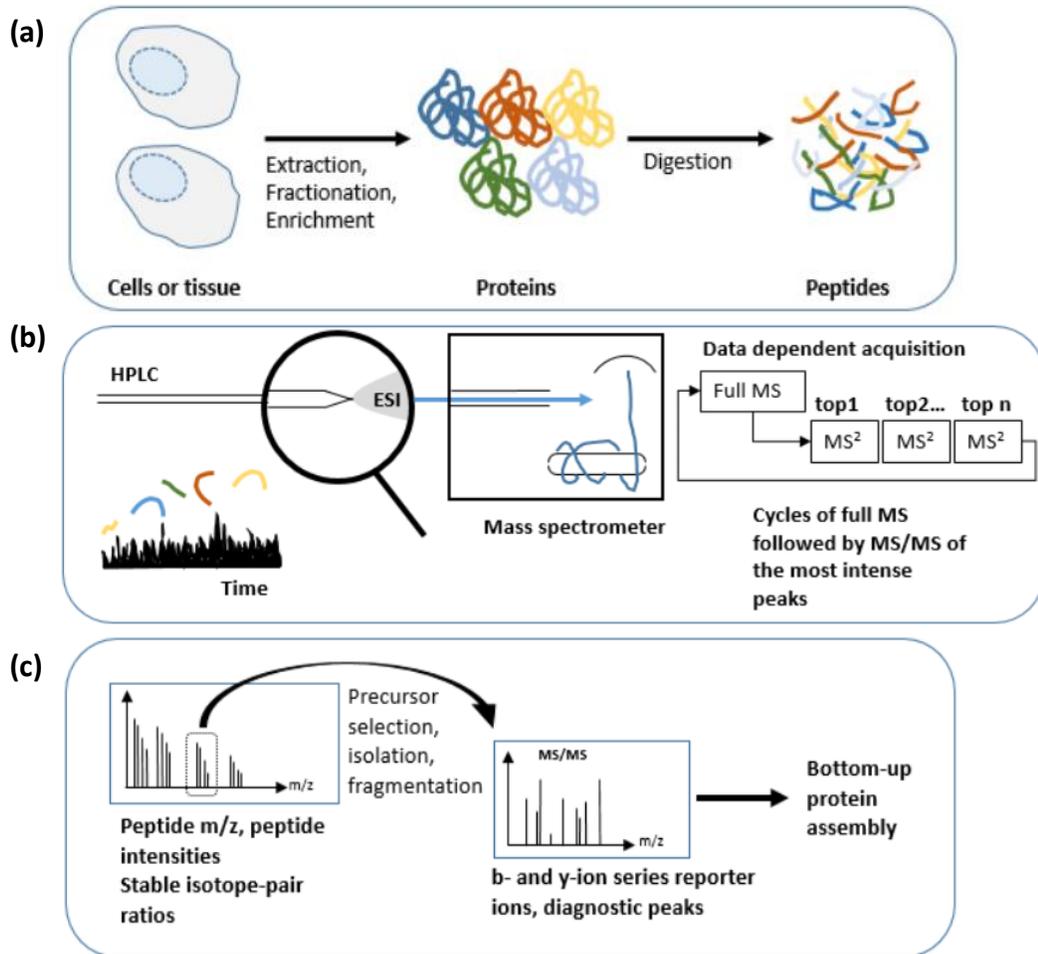


Figure 5.2: Schematic flow of bottom-up MS

Figure as depicted by Hein et al. (2013). (a) **Sample preparation:** Protein lysates are prepared from cells/tissue and are either used as a whole or fractionated. Lysates undergoes enrichment and subjected to digestion (in-gel or in-solution) to produce peptides. (b) **Liquid-chromatography-MS:** Peptides are separated using high-performance liquid chromatography (HPLC) and electro sprayed directly into the mass spectrometer. Peptide ions are measured at high resolution in a data-dependent mode with the most intense peptide ions fragmented to generate an MS/MS spectra. (c) **Spectra interpretation:** A full MS spectra would give information regarding the peptide mass, intensity, presence of post-translational modifications (PTM) and stable isotope pairs. The mass of each fragmented peptide together with its fragment ion pattern is searched against databases for peptide identification and bottom-up protein assembly.

5.2 Results

5.2.1 Validation of protein samples for subsequent usage

For each replicate protein samples, about 15×10^7 cells of FLAG-Upf1^{Wt} were cultured to obtain sufficient amounts of chromatin fraction and around 5×10^7 of FLAG-Upf1^{ChmB} cells cultured to produce whole-cell lysate (WCL) in the presence or absence of 0.7 µg/ml doxycycline. Hydroxyurea (HU) were added to media for 24hr for S-phase enrichment for all cell lines used in this study. Prior to IP, subcellular fractionation was performed on FLAG-Upf1^{Wt} cells to isolate fractions from the cytoplasmic and the nuclear fractions whilst a whole cell extraction was performed on FLAG-Upf1^{ChmB} cells to obtain WCL.

Immunoblotting was undertaken to ensure that FLAG-Upf1^{Wt} containing material isolated from the chromatin fraction derived from the nuclear compartment was free of contaminants from the cytoplasmic compartment. In each of three replicate experiments, protein lysates were fractionated as described in Section 2.9, and S2 (cytoplasmic fraction) and P3 (enriched chromatin fraction) was separated by SDS-PAGE and analysed by western blotting **Figure 5.3**. Upf1 was detected as a doublet of the expected molecular weight when probed with α-FLAG antibody.

All three P3 fractions were free of cytoplasmic contamination as judged by absence of alpha tubulin in the isolated chromatin-associated fraction. As expected, all three samples had a band when probed with α-ORC2 antibody which is a chromatin marker. This supports the notion that protein samples used for future experiments were pure and free of cytoplasmic contamination.

Once the relative purity of samples had been determined, 4mg of protein lysates were incubated overnight with 50 µl of 50% slurry ANTI-FLAG M2 beads and utilised for IP using the protocol in section 2.11. IP was performed for each sample derived from lysates originating from induced and uninduced FLAG-Upf1^{Wt} and FLAG-Upf1^{ChmB} HeLa cells.

An immunoblot was performed to determine the degree of recovery. 5% of total protein used for IP and total eluate were used for western blot **Figure 5.4**. From observation, all induced samples showed high levels of recovery and bands of expected size were observed when probed with α -Upf1 antibody. As controls, uninduced samples were also analysed and no Upf1 cross-reacting protein was observed when these blots were probed with the α -Upf1, as expected. Observed were doublet bands for all samples, similar to that observed in **Figure 5.3**.

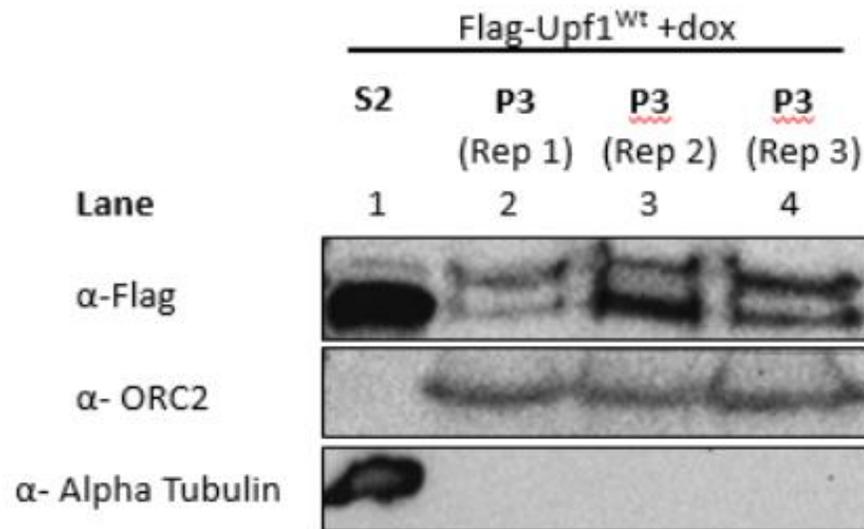


Figure 5.3: Purity of FLAG-Upf1^{wt} enriched chromatin fraction

FLAG-Upf1^{wt} cells were grown in 0.7µg/ml doxycycline for 48hr, then 2mM of hydroxyurea (HU) was added for 24hr to arrest cells in S-phase. HU was washed out and cells released into S-phase for 3hr. Cells were then subjected to a sub-cellular fractionation giving S2-cytoplasmic and P3- chromatin enriched fractions. Samples representative of equal volumes (~60ul) were loaded onto an SDS-PAGE and visualised by western blotting. Purity of samples were determined by using antibodies against α-ORC2 nuclear marker and α-Tubulin as cytoplasmic marker. Lane 1- S2 fraction and lane 2-4 are the replicates for P3 fraction. **Legend:** Rep 1, replicate 1; Rep 2, replicate 2 & Rep 3, replicate 3 samples.

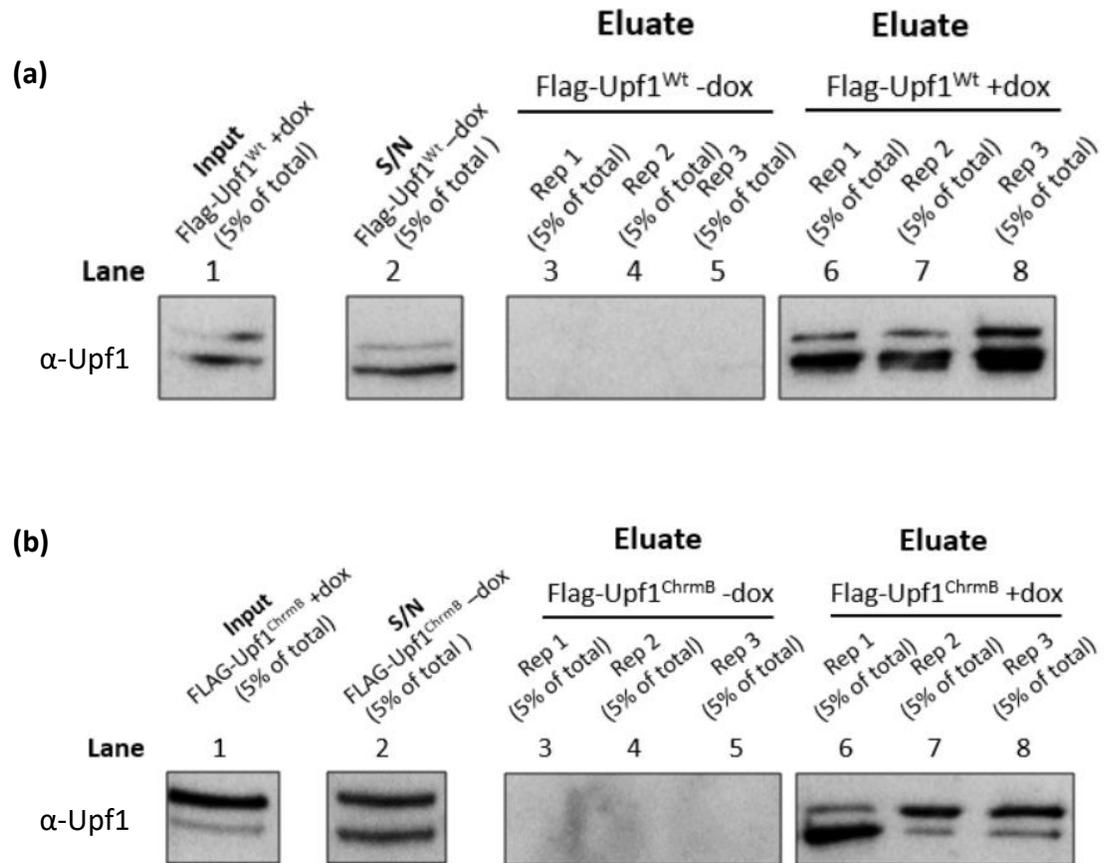


Figure 5.4: Immunoblot for sample quality validation

(a) FLAG-Upf1^{Wt} cells were grown in either the absence or presence of doxycycline (0.7μg/ml) and a sub-cellular fractionation to obtain chromatin fraction was performed it. 4mg of protein lysate for FLAG-Upf1^{Wt} uninduced and induced samples were immunoprecipitated overnight with 50μl 50% slurry ANTI-FLAG M2 beads. The following day, samples were washed three times and eluted by boiling beads in sample loading buffer. Input are representative of 5% total protein (lane 1), 5% total supernatant (s/n) (lane 2) and total eluate in each IP (lanes 3-8) were analysed by SDS-PAGE and western blotting. Eluates were assessed by probing with α-Upf1 antibody. (b) FLAG-Upf1^{ChmB} under the same conditions as in (a) except no sub-cellular fractionation was performed and whole cell lysate was used. Doublets observed could be due to the presence of the different isoforms in the sample or it could be the presence of both phosphorylated and endogenous Upf1.

5.2.2 Mass spectrometry- identification of protein interactors

Mass spectrometry-based proteomics has the potential to make a major contribution to the understanding of complex biological systems and make a significant contribution to the quest of improvement in diagnosis, and amelioration or cure for a variety of human diseases. It can do this by facilitating quantitative comparisons of protein levels from normal and pathological disease states and also under different experimental conditions (Berg et al. 2006). Its capability to detect and quantify large numbers of proteins, together with their post-translational status, in a given biological sample, coupled with ever advancing analytical software, makes this technology a profoundly important analytical tool (Baldwin 2004; Aebersold & Mann 2003).

In order to analyse IP protein samples obtained in section 5.2.1, eluted samples were reduced and alkylated as described previously in section 2.13.1. Equal loadings of samples were separated by SDS-PAGE and visualised by Coomassie staining (**Figure 5.5**). Gel fractions, prepared as described in chapter 4 previously, were excised into smaller gel pieces (See **Figure 5.5**) as an increased surface area to volume ratio is predicted to facilitate more efficient proteolytic digestion of proteins. Each sample replicate (A, B & C) gave rise to 4 fractions (1, 2, 3 & 4) generating 48 samples for individuals LC-MS/MS analysis. Sample preparation and tryptic digestion were performed as stated in section 2.13.

Samples were analysed using liquid chromatography tandem-mass spectrometry (LC-MS/MS) and quantified using a label-free MS approach. Peptides eluting from the LC column are usually ionised by electro-spray (Fallis 2013) and then introduced into the mass-spectrometer where mass and intensities are determined (Berg et al. 2006). The mass spectra obtained were analysed using Mascot Software and databases to identify the corresponding peptides and proteins (Fenyö 2000; Aebersold & Mann 2003).

MaxQuant analysis was performed under the direction of Dr Mark Collins (the University of Sheffield biOMICS facility) with settings as described in section 2.13.4. In total, 3,383 proteins were identified in all 48 fractions. MaxQuant analyses raw MS data and generates a tabulated output file which contains profiles of label-free quantification

(LFQ) intensities per replicate for each protein identified, together with other meta information about the quantity and quality of the recovered proteins with calculated statistics such as sequence coverage, no of unique peptide identified and score (for the best associated MS/MS spectrum) (Singh et al. 2016).

A table, “Protein Groups” contains information regarding the identified proteins which includes information such as sequence coverage, unique peptides, intensity based absolute quantification (iBAQ) value which indicates the abundance of peptide and also an identification score (**Table 1**).

The identification of 3383 proteins includes multiple alternative isoforms where specific identification criteria for a unique identification is not possible, either because tryptic digestion alone cannot provide such a distinction, or because the set of observed peptides during a spectrometry analysis for a specific protein isoform may be significantly fewer than that which might theoretically be generated from a given protein. Thus MaxQuant does not necessarily provide a single protein identification, but rather provides a list of all proteins and isoforms which could be explained by a given set of identified peptides (Cappadona et al. 2012).

The MaxQuant output file may then be used for further analysis using the Perseus statistical software package (Tyanova, Temu, Sinitcyn, Carlson, Marco Y. Hein, et al. 2016). This software enables a more detailed statistical analysis to be accomplished. Perseus integrates data cleansing and normalisation and multiple methods for exploratory analysis such as histogram charts, intensity curves and scatterplots (Tyanova, Temu, Sinitcyn, Carlson, Marco Y. Hein, et al. 2016). In addition to providing a statistical test of significance for quantitative change, this software allows the application of threshold values for a number of parameters to exclude unnecessary or incorrect protein identification to be removed from the data frame of the Perseus software (Laboratory of Mass Spectrometry, LNBio 2014).

By using the default parameters (“only identified by site, Reverse and Contaminant), any incorrect protein identification based on those databases were

eliminated which brought down the number of proteins to 2,825. Elimination of “reverse proteins” is done based on a search against a target-decoy database (Hein 2014). Protein groups which match the proteins from reverse database or contaminants will be discarded (Elias & Gygi 2007; Schwanhüusser et al. 2011b).

Following that, LFQ intensities were logarithmised as a normalisation step and filtered again for proteins having a minimum of 3 valid values in a replicate group ie. there should be at least three valid LFQ values in a replicate, across all 4 samples (FLAG-Upf1^{Wt} +/-dox & FLAG-Upf1^{ChrmB} +/-dox. The LFQ intensities are normally logarithmised because this simplifies the statistical analysis that follows as log intensities are approximately normally distributed (Karpievitch et al. 2012; Callister et al. 2006).

Finally, missing values were imputed with values representing a normal distribution around the detected limit of MS allowing a new distribution with a downshift of 1.8 standard deviation and a width of 0.3 standard deviation to be created (Keilhauer et al. 2015; Hein 2014). This further reduced the identified number of proteins to 1,248. Missing values could arise due to any of these three reasons; i- the peptide is truly present but is not detected or is incorrectly identified, ii- the peptide is truly present but at an abundance below the instrument’s detection limit, and iii- the peptide is not present (Karpievitch et al. 2012).

Histograms were plotted using individual LFQ intensities to verify that datasets obtained for all experimental samples have a normal distribution curve (**Figure 5.6**). Normally distributed data is usually needed because it indicates the sample data has the same characteristics as the population which allows parametric statistical methods like t-test, ANOVA to be undertaken, allowing a generalisation in analysing the data (Mukaka 2012).

The reproducibility of technical and biological replicates can be assessed by displaying scatter plots where the ranked intensities of identified proteins in any given replicate (biological or technical) is compared with every other replicate. Significant variation between replicates is predicted to generate highly disperse data with a low

correlation coefficient. Conversely, consistent data between replicates is expected to produce scatter plots with an approximate slope equal to unity, and high correlation coefficient. The correlation coefficients for pairwise comparison between replicates is determined utilising the LFQ intensities calculated using Perseus software (Luczak et al. 2016). Most of the Pearson correlation values obtained and shown in **Figure 5.7** were > 0.8 which indicates an excellent correlation of LFQ intensities in triplicates between induced and uninduced samples. Additionally, **Table 2** displays the correlation coefficients for all combination of protein samples analysed in triplicate.

To identify proteins that specifically interact either with Upf1^{ChrmB} or with Upf1^{Wt}, Perseus undertakes a Student t-test for significant difference in expression level, for all proteins, identified in samples derived from cells exposed to doxycycline (induced “bait”), and thus expressing the Flag-tagged version of UPF1, compared with the level of the same protein in the unexpressed control sample. When the calculated differences between log mean protein intensities of induced “bait” vs control (X axis) is plotted against the negative log p values derived from the t-test (Y axis), proteins that are not specifically enriched in either sample would centre around a value of zero on the X axis, and indicate a statistically insignificant p value on the Y axis (Singh et al. 2016). Conversely, proteins that are significantly enriched in induced “bait” samples will deviate significantly to the right-hand side of such a plot, with the extent of significance, derived from an analysis of the variation between biological replicates, resulting in a higher y-axis value.

From the “Protein Group” file that was tabulated in MaxQuant, Upf1 was identified to be one of the most enriched proteins with a high identification score of 323.31 (**Table 3**) and a sequence coverage of 50% (**Appendix 1**). The score is a probabilistic scoring that is based on the peptide search engine Mascot, hence a higher number indicates that the probability the presence of a protein is high. The enrichment of FLAG-Upf1^{Wt} and FLAG-Upf1^{ChrmB} in their respective replicate immunoprecipitates, was 20 and 80-fold respectively. **Figure 5.8** (a & b) shows the enrichment of Upf1 (Q92900) for both sets of samples which is above the significant threshold. A default

setting of FDR 0.05, and S0 at 1 was used. These data show significant enrichment of both wild-type and Upf1^{ChrmB}. This was substantially better than what was observed in the experiment reported in **Figure 4.6**, increasing the confidence that statistically significant enrichment of specific interactors could be detected in this approach and not be masked by the presence of non-specifically bound protein.

Figure 5.8c is a volcano plot of FLAG-Upf1^{Wt} versus FLAG-Upf1^{ChrmB} induced samples. As expected, Q92900 appears at 0 which indicates that the ratio of Upf1 present in both samples were 1:1. Hence, other proteins enriched in either set of samples (ie showing a significant enrichment difference, together with a statistically significant $-\log P$ value) identified at this point could be considered as true interactors without bias. The higher the difference between the group means (ie enrichment) and the p value (ie reproducibility), the more the interactors move to the top right or left corner of the plot which is the area of highest confidence (ie true interaction) (Keilhauer et al. 2015).

Name	Description
Protein IDs	Identifier(s) of protein(s) contained in the protein group. They are sorted by number of identified peptides in descending order
Majority protein IDs	These are the IDs of those proteins that have at least half of the peptides that the leading protein has.
Protein counts	Number of peptides associated with each protein in protein group, occurring in the order as the protein IDs occur in the 'Protein IDs' column. Here distinct peptide sequences are counted. Modified forms or different charges are counted as one peptide.
Peptide counts (razor + unique)	Number of peptides associated with each protein in protein group, occurring in the order as the protein IDs occur in the 'Protein IDs' column. Here distinct peptide sequences are counted. Modified forms or different charges are counted as one peptide.
Peptide counts (unique)	Number of peptides associated with each protein in protein group, occurring in the order as the protein IDs occur in the 'Protein IDs' column. Here distinct peptide sequences are counted. Modified forms or different charges are counted as one peptide.
Fasta headers	Fasta headers(s) of protein(s) contained within the group.
Proteins	Number of proteins contained within the group. This corresponds to the number of entries in the column 'Protein IDs'.
Peptides	The total number of peptide sequences associated with the protein group (i.e. for all the proteins in the group)
Sequence coverage	Percentage of the sequence that is covered by the identified peptides of the best protein sequence contained in the group.
Unique + razor sequence coverage (%)	Percentage of the sequence that is covered by the identified unique and razor peptides of the best protein sequence contained in the group
Unique sequence coverage (%)	Percentage of the sequence that is covered by the identified unique peptides of the best protein sequence contained in the group
Mol weight (kDa)	Molecular weight of the leading protein sequence contained in the protein group.

Name	Description
Sequence length	The length of the leading protein sequence contained in the group.
Sequence lengths	The length of all sequences of the proteins contained in the group.
PEP	Posterior Error Probability of the identification. This value essentially operates as a p-value, where smaller is more significant.
Intensity	Summed up eXtracted Ion Current (XIC) of all isotopic clusters associated with the identified AA sequence. In case of a labeled experiment this is the total intensity of all the isotopic patterns in the label cluster.
iBAQ	Σ intensity/#theoretical peptides (intensity-based absolute quantification of proteins)
Only identified by site	When marked with '+', this particular protein group was identified only by a modification site.
Reverse	When marked with '+', this particular protein group contains no protein, made up of at least 50% of the peptides of the leading protein, with a peptide derived from the reversed part of the decoy database. These should be removed for further data analysis. The 50% rule is in place to prevent spurious protein hits to erroneously flag the protein group as reverse.
Contaminant	When marked with '+', this particular protein group was found to be a commonly occurring contaminant. These should be removed for further data analysis.
Id	A unique (consecutive) identifier for each row in the proteinGroups table, which is used to cross-link the information in this file with the information stored in the other files.
Peptide IDs	Identifier(s) of the associated peptide sequence(s) summary, which can be found in the file 'peptides.txt'.
Best MS/MS	The identifier of the best (in terms of quality) MS/MS scans identifying the peptides of this protein, referenced against the MS/MS table.
Phospho (STY) site IDs	Positions of the sites in the leading protein of this group.

Name	Description
Oxidation (M) site positions	Positions of the sites in the leading protein of this group.
Phospho (STY) site positions	Positions of the sites in the leading protein of this group.

Table 1: Information summary for the headers that are present in the “Protein Group” file from MaxQuant

Table of summarised information obtained from the MassIVE website (<ftp://massive.ucsd.edu/MSV000080315/result/MaxquantResults/Trypsin/combined/xt/tables.pdf>) regarding “Protein Group” file that is obtained from MaxQuant. It consists of information regarding the identified proteins in the processes raw-files. Each single row contains the group of proteins that could be reconstructed from a set of peptides (Schwanhüusser et al. 2011a; MassIVE 2011)

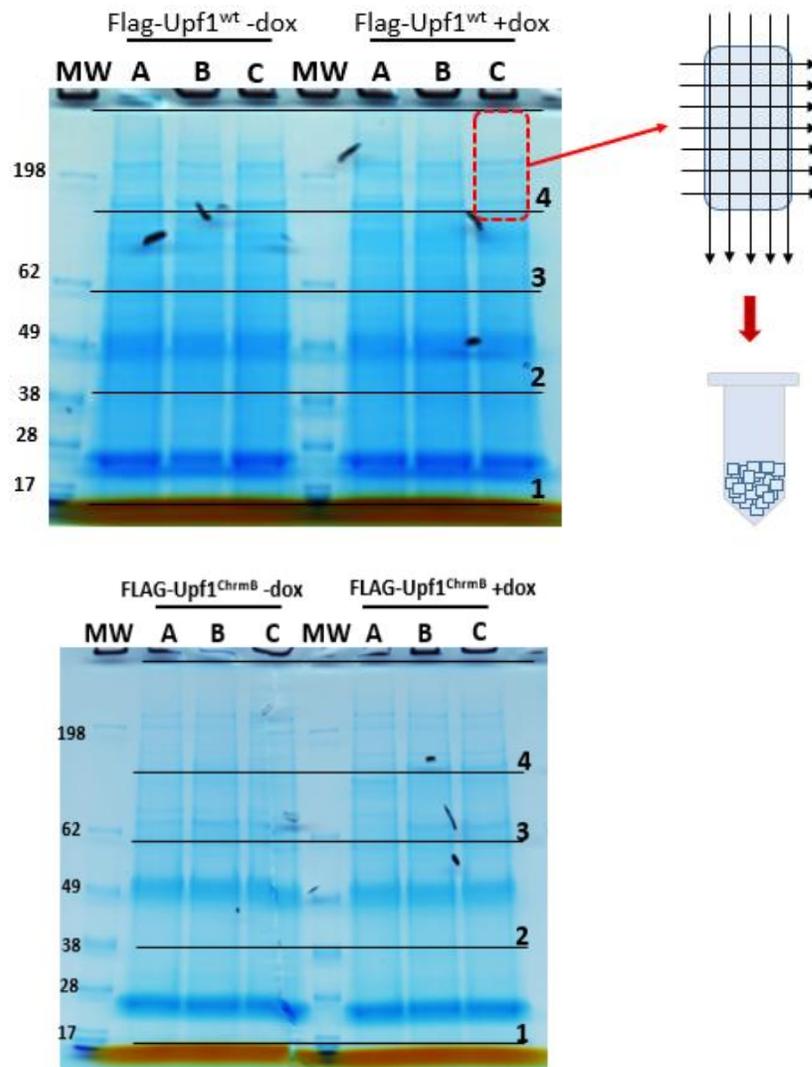


Figure 5.5: SDS-PAGE of IP samples post reduction and alkylation steps.

Triplicate samples of FLAG-Upf1^{wt} induced (+dox) and uninduced (-dox); plus triplicates of FLAG-Upf1^{ChmB} induced (+dox) and uninduced (-dox) were loaded onto a pre-cast NuPAGE 4-12% gel at same volumes. Gels were run for 30 mins before staining with coomassie. Fractions were determined accordingly and gel was cut into smaller pieces to enable better chemical reaction in subsequent steps. A-C represents the replicates whereas 1-4 represents the fraction number for ease of sample labelling purposes.

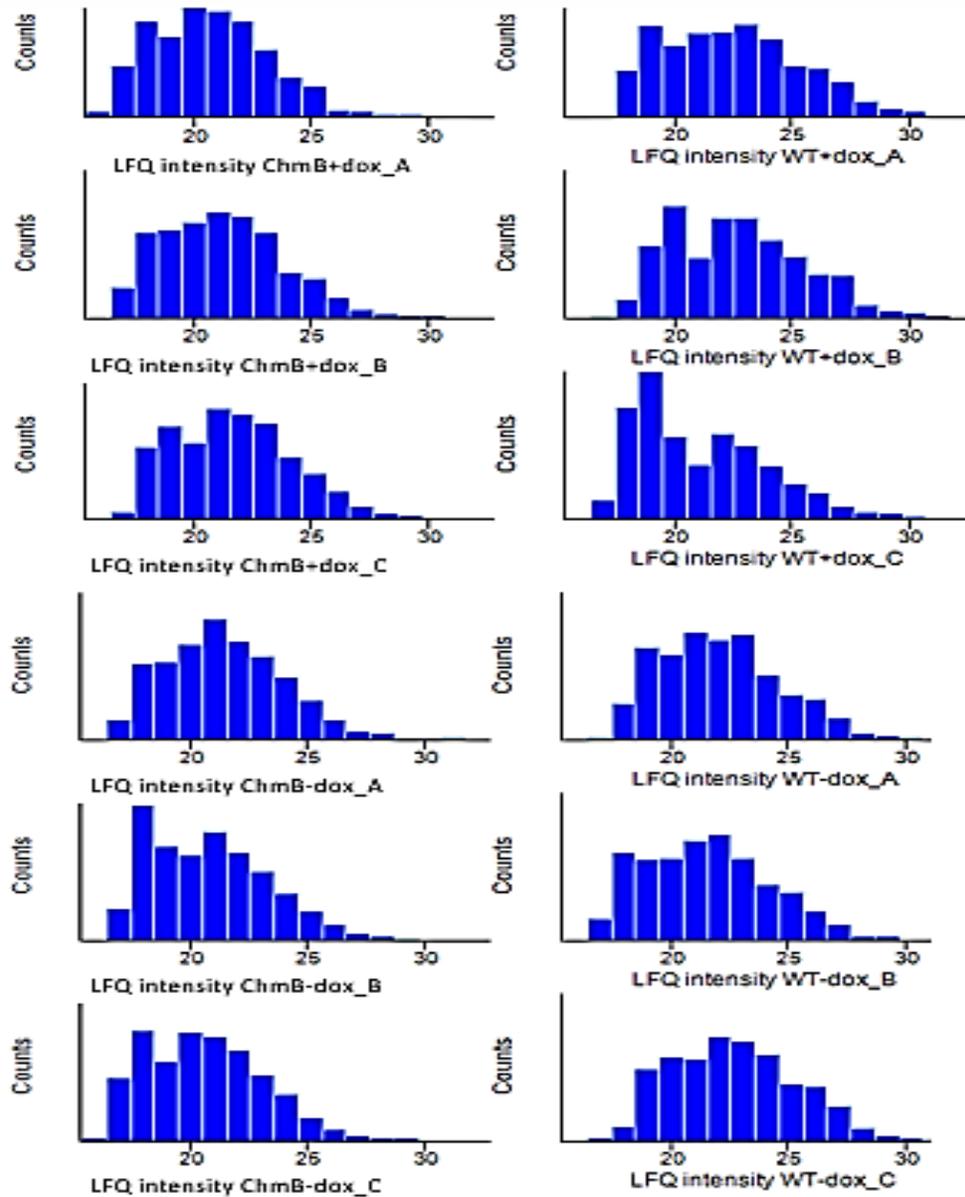


Figure 5.6: Biological repeats following imputation

LFQ intensity values were plotted into histograms to show the pattern based on normal distribution for all replicate conditions of FLAG-Upf1^{Wt} and FLAG-Upf1^{ChmB}. Initial filtering to eliminate reverse proteins and contaminants were done using the options under the “Categorical annotation” tab in Perseus. The LFQ values were then logarithmised before imputing missing LFQ values with those representing a normal distribution around the detection limit of the MS (Keilhauer et al. 2015).

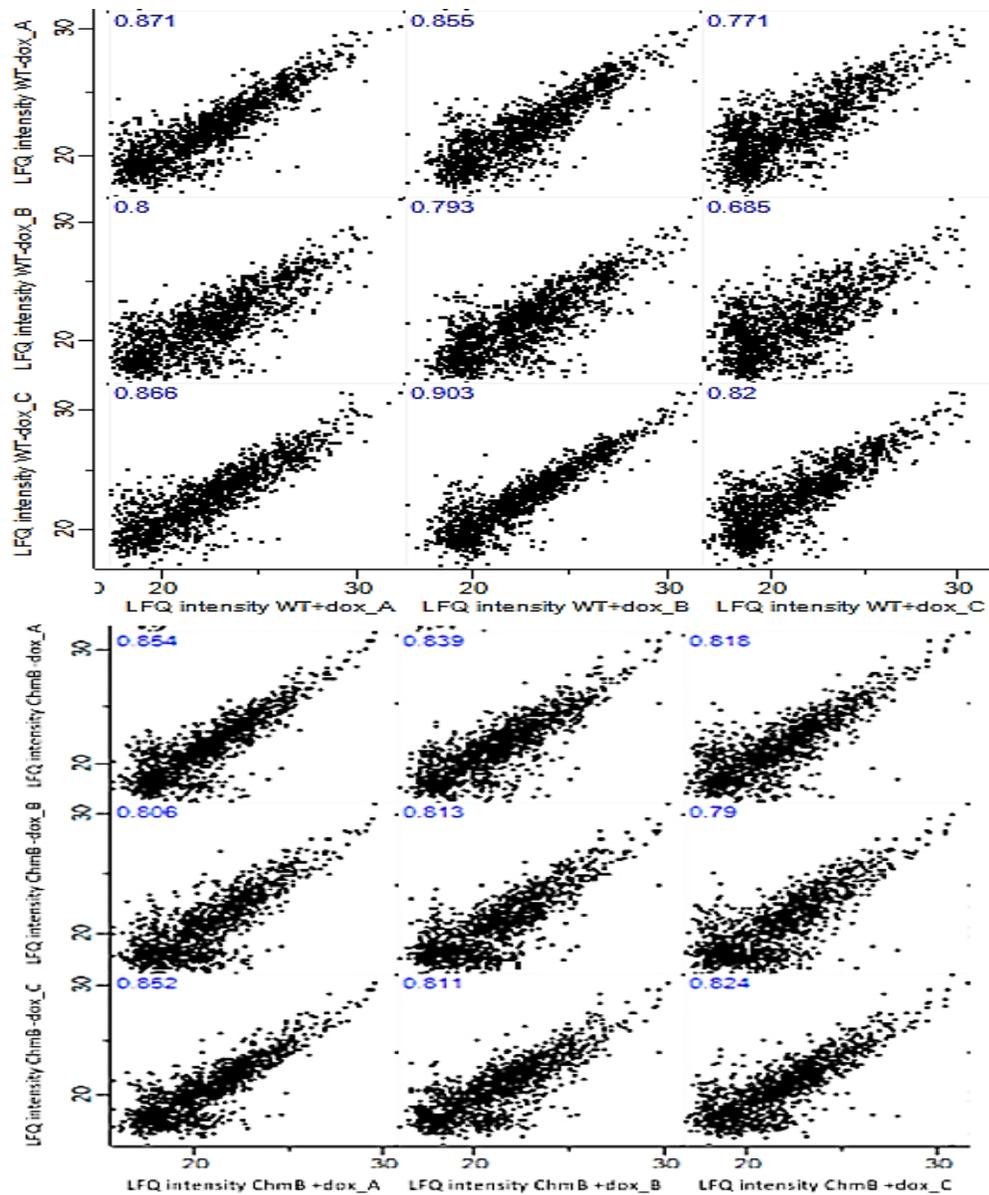


Figure 5.7: Biological replicates of FLAG-Upf1^{Wt} and FLAG-Upf1^{ChmB} correlates within and among groups

Scatter plot showing the correlation coefficients between the LFQ intensities between induced and uninduced FLAG-Upf1^{Wt} and FLAG-Upf1^{ChmB} sample groups. Calculations were derived from Perseus software. The values at the top left corner of each quadrant is the Pearson correlation value. An upward pattern displays a positive correlation between samples and the strength of correlation is determined by the numerical value of the correlation and how closely the dots clustered around a line (Mukaka 2012).

LFQ intensity ChrmB+dox_A	LFQ intensity ChrmB+dox_B	LFQ intensity ChrmB+dox_C	LFQ intensity ChrmB-dox_A	LFQ intensity ChrmB-dox_B	LFQ intensity ChrmB-dox_C	LFQ intensity WT+dox_A	LFQ intensity WT+dox_B	LFQ intensity WT+dox_C	LFQ intensity WT-dox_A	LFQ intensity WT-dox_B	LFQ intensity WT-dox_C	Group1	Name
ChrmB+dox	ChrmB+dox	ChrmB+dox	ChrmB-dox	ChrmB-dox	ChrmB-dox	WT+dox	WT+dox	WT+dox	WT-dox	WT-dox	WT-dox		
NaN	0.84	0.83	0.85	0.81	0.85	0.59	0.57	0.58	0.56	0.54	0.55	ChrmB+dox	LFQ intensity ChrmB+dox_A
0.84	NaN	0.87	0.84	0.81	0.81	0.58	0.59	0.64	0.53	0.48	0.57	ChrmB+dox	LFQ intensity ChrmB+dox_B
0.83	0.87	NaN	0.82	0.79	0.82	0.53	0.56	0.59	0.50	0.48	0.52	ChrmB+dox	LFQ intensity ChrmB+dox_C
0.85	0.84	0.82	NaN	0.82	0.87	0.56	0.56	0.59	0.55	0.53	0.56	ChrmB-dox	LFQ intensity ChrmB-dox_A
0.81	0.81	0.79	0.82	NaN	0.85	0.55	0.53	0.60	0.57	0.50	0.54	ChrmB-dox	LFQ intensity ChrmB-dox_B
0.85	0.81	0.82	0.87	0.85	NaN	0.51	0.50	0.54	0.52	0.49	0.50	ChrmB-dox	LFQ intensity ChrmB-dox_C
0.59	0.58	0.53	0.56	0.55	0.51	NaN	0.89	0.78	0.87	0.80	0.87	WT+dox	LFQ intensity WT+dox_A
0.57	0.59	0.56	0.56	0.53	0.50	0.89	NaN	0.83	0.85	0.79	0.90	WT+dox	LFQ intensity WT+dox_B
0.58	0.64	0.59	0.59	0.60	0.54	0.78	0.83	NaN	0.77	0.68	0.82	WT+dox	LFQ intensity WT+dox_C
0.56	0.53	0.50	0.55	0.57	0.52	0.87	0.85	0.77	NaN	0.87	0.87	WT-dox	LFQ intensity WT-dox_A
0.54	0.48	0.48	0.53	0.50	0.49	0.80	0.79	0.68	0.87	NaN	0.80	WT-dox	LFQ intensity WT-dox_B
0.55	0.57	0.52	0.56	0.54	0.50	0.87	0.90	0.82	0.87	0.80	NaN	WT-dox	LFQ intensity WT-dox_C

Table 2: Correlation coefficient of all protein samples and their replicates

As mentioned above, the strength of correlation is determined by the numerical value that is displayed. The closer the value to 1, the stronger the correlation is between samples. Values in green implicates that there is good correlations between sample combinations within a cell line. Whilst the values in yellow to red indicates that correlations between samples are weak which is expected when comparison is done between different cell lines. There are however, two values within the FLAG-Upf1^{WT} samples that has a correlation value of less than 0.7.

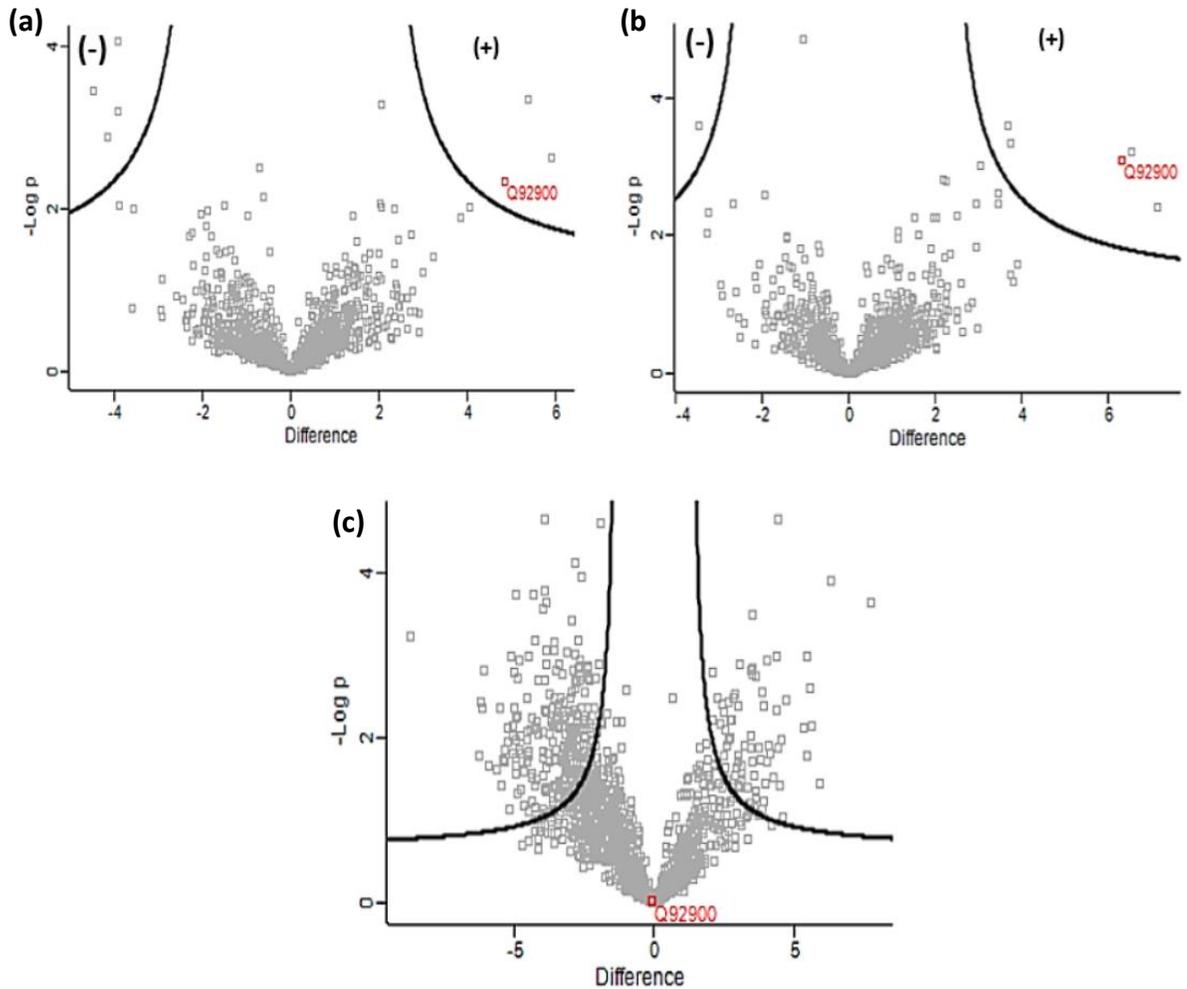


Figure 5.8: Differential interacting proteins between FLAG-Upf1^{Wt} and FLAG-Upf1^{ChmB}.

Volcano plot showing quantitative enrichment of interacting proteins associated with (a) FLAG-Upf1^{Wt} (+) and (b) FLAG-Upf1^{ChmB} (+), compared with control immunoprecipitates from the relevant cell line not exposed to doxycycline (-). (c) Volcano plot showing quantitative enrichment of interacting proteins between FLAG-Upf1^{Wt} and FLAG-Upf1^{ChmB}. Upf1-interacting proteins were identified from statistical analysis, using t-testing, of triplicate data sets, discussed in the text. Solid lines indicate significant threshold using the default setting for false discovery rate (FDR) of 0.05 and an artificial within groups of variance $S_0 = 1$.

5.2.3 Comparison of known and novel interactors identified

Using the list of filtered proteins (2825) identified, it was important to establish the validity of the dataset obtained in this study. The following NMD protein components; Upf2, Upf3B, SMG1, SMG5, SMG7, SMG8 and SMG9 were identified here with Upf2 and Upf3B found to be significantly enriched in induced FLAG-Upf1^{ChrmB} compared to uninduced (Okada-Katsuhata et al. 2012; Bono 2014; Gatfield et al. 2003; Unterholzner & Izaurralde 2004; Nicholson et al. 2010). However, another NMD component SMG6, which acts as a nuclease to degrade nonsense codon-containing mRNAs was not identified in the work presented here. In addition, another well-established mRNA destruction mechanism, Staufen-mediated decay (SMD) is known to involve Upf1 and associated proteins (Isken & Maquat 2008; Park et al. 2013; Park & Maquat 2013). Both STAU1 and STAU2 proteins of SMD are also identified in this study.

A comparative analysis of interacting “hits” identified in this work was performed against previously reported Upf1 interactomes (Flury et al. (2014); Schweingruber et al. (2016)) (**Appendix 2**). Flury et al. (2014) generated data using a similar immunoprecipitation approach to that described here as stated in section 2.11 while Schweingruber et al. (2016), used bioID. In this latter approach, Upf1 and two other NMD factors (Upf2 & SMG5) were tagged with a mutant form of the bacterial BirA protein, which catalyses the biotinylation of proteins into which it comes in contact, and these are subsequently isolated by avidin affinity chromatography for subsequent identification by mass spectrometry. In addition, the BioGRID (<https://thebiogrid.org>) database acts as a general repository for protein-protein interactions identified using a variety of methodologies (Stark 2006; Chatr-Aryamontri et al. 2015). Ten proteins were identified in both this study and Flury et al. (2014) and eleven proteins identified here were also reported by Schweingruber et al. (2016). Whilst only 4 proteins (Upf2, Upf3B, STAU2 and EIF4A3) were identified by all three approaches.

5.2.4 Significant interactors

As indicated previously, volcano plots generated using Perseus facilitate the identification of significant differences in protein interactors between two data sets. The statistical test performed in Perseus produced a new list of proteins and 313 were found to be significantly altered when FLAG-Upf1^{Wt}+dox was compared against FLAG-Upf1^{ChrmB}+dox samples (**Appendix 3**). Additionally, 296 proteins are identified as novel potential interactors.

The aim of this study was to identify proteins that were enriched in a Upf1 population associated with chromatin. Volcano plots are displayed where relevant enrichment in this population, compared to the cytosol fraction is designated negative enrichment, while material more abundantly found in Flag-Upf1^{ChrmB} associated material is deemed to be positively enriched. Based on the Volcano plot and looking at the “Difference” column in **Appendix 3**, proteins were found to be enriched either in FLAG-Upf1^{Wt} or FLAG-Upf1^{ChrmB} immunoprecipitates as determined by their t-test difference (**Appendix 3**). Out of 313 proteins that were identified as significant protein interactors, 238 were found to be more abundant in FLAG-Upf1^{Wt} containing fractions while 75 proteins were more abundant in Flag-Upf1^{ChrmB}.

As an example, RFC4 is one component of a five-subunit complex that acts as a clamp loader to install the replisome at the point of DNA replication initiation has an enrichment value of -3.211, indicating that RFC4 is more abundant in FLAG-Upf1^{Wt} immunoprecipitates. In contrast, DNA topoisomerase 2A (TOP2A), which plays a role in DNA topological resolution, which is more abundant in FLAG-Upf1^{ChrmB} immunoprecipitates. Also abundant in FLAG-Upf1^{Wt} containing fractions were proteins involved in chromatin remodelling such as SWItch/Sucrose Non-Fermentable (SWI/SNF) family proteins (SMARCE1, SMARCA4 & ACTL6A). Actin and vimentin were also found in this fraction, which may reflect their abundance compared to chromatin proteins (Torrente et al. 2011). However, it has recently been suggested that both vimentin and

actin might also be involved in chromatin remodelling (Olave et al. 2002; Ivaska et al. 2007).

From the list in Appendix 3, I applied a bioinformatics filter to select for proteins based on localization using information from the GeneCards database (www.genecards.org) and subsequently according to gene ontology using Uniprot's ID mapping. Key words used in the latter filter were DNA binding, DNA damage, DNA replication, telomere maintenance, telomeric DNA binding and chromatin binding/remodeling. This allowed for the identification of the following proteins which were differentially enriched in Flag-Upf^{wt} versus the FLAG-Upf1^{chromB} fraction: UBR5, WDHD1, SMARCD2, RFC1, RFC2, RFC4, HIST1H2BN, HERC2, ACTL6A, PARP1, TOP2A, DEK, NASP, MCM6, TRIP13, SMARCB1, SMARCE1, SMARCD1, XRN2 and SMARCA4.

A Venn diagram was constructed based on the 20 proteins mentioned above (**Figure 5.9**) categorising them according to their gene ontology: biological processes (DNA damage, DNA replication, telomere maintenance and chromatin binding/remodeling). Unsurprisingly, there are overlaps where some proteins were found to be involved in more than one biological process. For example, RFC2 and Upf1 are reported to be involved in DNA damage response and telomere maintenance in addition to an involvement in DNA replication (Tomida et al. 2008; Ellison & Stillman 2003; Azzalin & Lingner 2006a; Higa et al. 2017).

WDHD1 is also recognised as being involved in both DNA damage response and during DNA replication. WDHD1 is a component of the replisome that regulates DNA replication and interacts with human primase-DNA polymerase/MCM10 (Zhou et al. 2016; Yoshizawa-Sugata & Masai 2009). It is involved in augmenting Claspin function as well as checkpoint activation and appears to be part of the stalled replication fork complex (Leman & Noguchi 2013; Hao et al. 2015; Jones & Petermann 2012). Additionally, Yoshizawa-Sugata & Masai (2009) reported that during S-phase, WDHD1 is phosphorylated in a response to replication arrest by ATR and ATM and its depletion, results in a delay of S-phase progression and an increase in DNA damage.

5.2.5 Comparative analysis of protein interactors in Upf1^{Wt} and Upf1^{ChmB} enriched fractions by immunoblotting

Co-Immunoprecipitation (Co-IP), a technique that determines whether two proteins interact in a physiological condition in the cell (Rao et al. 2014), is a common technique used for verifying protein-protein interactions. The “bait” protein, FLAG-Upf1 is used to capture the proteins with which it is proposed to interact - often referred to as the “prey” protein (Choi et al. 2012). Thus, the next step was to validate the interaction between Upf1 and identified proteins. Time constraints limited the range of proteins for which antibodies, immunoblot conditions, and source material could be generated, and so not all 20 proteins could be investigated independently using this approach.

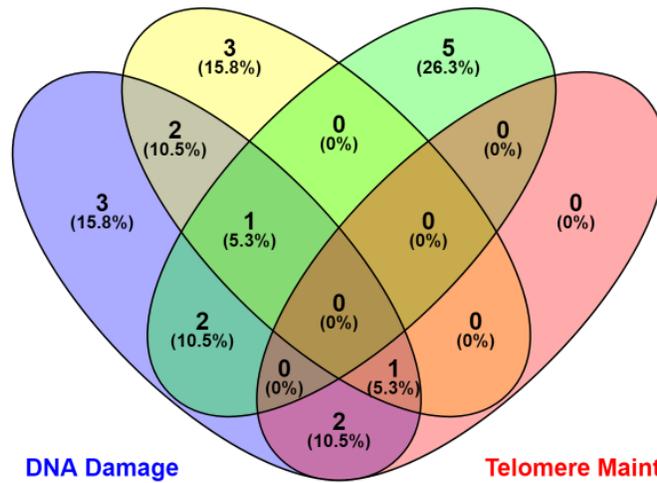
For the purposes of this experiment, the biological replicates of each FLAG-Upf1^{Wt} and FLAG-Upf1^{ChmB} IP used for MS analysis were pooled and, in each case, 2mg of protein lysate were mixed with 30µl of anti-FLAG M2 beads for IP, and proteins were eluted by boiling in sample loading buffer for direct comparison with each other.

I chose one protein from each category in **Figure 5.9** for immunoblot validation and subjected both FLAG-Upf1^{Wt} and FLAG-Upf1^{ChmB} immune-precipitates to immunoblotting with commercially sources antibodies. The five proteins chosen were RFC1, RFC4, WDHD1, SMARCE1 and TOP2A, all of which are known to play significant chromatin associated roles (Ellison & Stillman 1998; Gaillard et al. 2015; Leman & Noguchi 2013; Bermejo et al. 2007; Sethuraman et al. 2016). Based on the significance data analysis of the mass spectrometry data, proteins that were found to be more abundant in FLAG- Upf1^{Wt} IP samples were RFC4, WDHD1 and SMARCE1. On the other hand, RFC1 and TOP2A were more abundant in the FLAG-Upf1^{ChmB}containing samples. Because RFC1 and RFC4 are known to form a multimeric complex to facilitate clamp loading (Gary Schmidt et al. 2001; Corrette-Bennett et al. 2004; Overmeer et al. 2010), it was important to establish whether these two forms of Upf1 differentially bound components of this complex. Most of the proteins found to be abundant in FLAG-

Upf1^{Wt}+dox samples were involved in DNA binding or chromatin binding (Redondo-Muñoz et al. 2013; Majka & Burgers 2004). Whereas proteins that were identified as abundant in Upf1^{ChmB}+dox samples mostly have been implicated in DNA damage responses and repair (Yu et al. 2017; Broderick et al. 2015).

Immunoblotting using an α -Upf1 antibody and antibodies of chosen proteins supported some of the interactions with Upf1 that had been identified by MS (**Figure 5.10a**). As expected, a band for RFC4 was visible in the FLAG- Upf1^{Wt}+dox lane when probed with the α -RFC4 antibody. Also expected were bands for RFC1 and TOP2A appearing in the Upf1^{ChmB}+dox lane when probed with their respective antibodies. The data shown in **Figure 5.10a** were quantified, normalised to Upf1 levels and the resultant comparisons shown in **Figure 5.10b**. These data show that interestingly, chromatin-associated wild-type Upf1 is associated with RFC4, but not RFC1 or TOP2a, while the reverse is true for Upf1 that is unable to bind to chromatin. Unfortunately, when probed with α -WDHD1 and α -SMARCE1, neither fraction showed detection of both WDHD1 and α -SMARCE1. This is likely due to low levels of expression of these proteins in the sample or that immunoblotting conditions (as recommended by the manufacturer) were not optimal.

(a) **Replication** **Chrm Remodel.** **Figure 5.9**



(b)

DNA DAMAGE	REPLICATION	CHRM REMODELING	TELOMERE MAINT.
UBR5	WDHD1	SMARCD2	PARP1
WDHD1	RFC4	SMARCB1	RFC2
HERC2	RFC2	ACTL6A	RFC1
ACTL6A	NASP	DEK	
PARP1	MCM6	SMARCE1	
RFC2	HERC2	SMARCD1	
RFC1	TOP2A	SMARCA4	
DEK		TOP2A	
TRIP13			
XRN2			

Figure 5.9: Venn diagram displaying the overlap of proteins identified after filtering based on gene ontology

Proteins were categorised according to their gene ontology using BioGRID and UniProt focusing on nuclear protein which were involved in DNA damage, DNA replication, chromatin remodelling and telomere maintenance. Purple oval represents DNA damage, yellow oval; replication, green oval; chromatin remodelling and red oval; telomere maintenance. Five proteins (in red); WDHD1, TOP2A, RFC4, SMARCE1 and RFC1 were chosen for validation using Co-IP. Venn diagram was created using Venny 2.1 (<http://bioinfogp.cnb.csic.es/tools/venny>).

No	Gene names	Score	Average ratio plus/minus dox WT	Average ratio plus/minus dox ChrmB	Average ratio WT/ChrmB plus dox
1	UPF1	323.31	22.2718811	85.33202033	0.931014478
2	FUBP3	323.31	2.514339674	3.939182688	5.06617812
3	UBR5	323.31	0.97572616	0.586438665	14.78861919
4	AKAP8	160.76	0.838286122	6.741417038	10.53299606
5	WDHD1	147.13	1.433415754	0	∞
6	SMARCD1	121.36	1.280673756	7.189027173	6.888393256
7	TOP2A	111.47	0.685573594	1.264624375	0.073775014
8	FUBP1	80.41	2.340177519	1.104510428	6.733404594
9	HIST1H2BN	78.697	0.698132669	1.053760535	7.691494273
10	SPIN1	60.732	1.217589231	0.539415826	3.930077341
11	RFC2	58.93	0.989634187	0	∞
12	RFC1	56.401	0	2.07193525	0
13	SMARCE1	53.236	1.739380042	1.227118343	4.713629292
14	MCM6	47.172	0.909307635	0.452575089	7.187394675
15	PARP1	42.958	0	1.598579271	0
16	TRIP13	37.719	0	0.970896164	0
17	XRN2	32.652	N/A	17.19778809	0
18	NASP	24.058	0.322192898	0.859497012	0.051770799
19	RFC4	22.719	0.982296195	0.571524271	7.130526253
20	SMARCD2	17.585	2.985509664	0.467493797	18.9974219
21	ACTL6A	17.343	1.261710655	1.136859895	4.670437652
22	DEK	16.026	1.652253432	N/A	N/A
23	SMARCB1	15.972	0.991839189	0.345858072	14.83177821
24	HERC2	10.755	#DIV/0!	1.678381723	0

Table 3: Filtered “proteingroups.txt” file produced by MaxQuant showing most significant protein interactors

This table shows the raw data from the “proteingroups.txt” produced by MaxQuant and has been filtered to show significant proteins only. The “Score” column represents the identification score value. A high value indicates good protein identification in sample. The “Average ratio plus/minus” are calculations of the average LFQ intensities for induced samples over uninduced for all replicates of both FLAG-Upf1^{Wt} and FLAG-Upf1^{ChrmB}. The final column is the calculated value of the average LFQ intensity of FLAG-Upf1^{Wt} replicates (induced) over the average LFQ intensity of FLAG-Upf1^{ChrmB} replicates (induced).

No	Student's T-test Significant CHRMB+dox_WT+dox	Student's T-test significant__	Gene names	Student's T-test Difference CHRMB+dox_WT+dox
1	+	ChrmB+dox_WT+dox	WDHD1	-4.916542053
2	+	ChrmB +dox_WT+dox	DEK	-3.909390132
3	+	CHRMB+dox_WT+dox	SMARCD2	-3.868571599
4	+	CHRMB+dox_WT+dox	RFC2	-3.807206472
5	+	CHRMB+dox_WT+dox	UBR5	-3.743015289
6	+	CHRMB+dox_WT+dox	SMARCB1	-3.37991333
7	+	CHRMB+dox_WT+dox	RFC4	-3.211526235
8	+	CHRMB+dox_WT+dox	MCM6	-3.160982132
9	+	CHRMB+dox_WT+dox	SMARCD1	-3.040451686
10	+	CHRMB+dox_WT+dox	HIST1H2BN	-2.696095785
11	+	CHRMB+dox_WT+dox	SMARCE1	-2.283793132
12	+	CHRMB+dox_WT+dox	ACTL6A	-2.159737269
13	+	CHRMB+dox_WT+dox	SPIN1	-1.944556554
14	+	CHRMB+dox_WT+dox	NASP	2.744192759
15	+	CHRMB+dox_WT+dox	RFC1	2.788565954
16	+	CHRMB+dox_WT+dox	PARP1	2.970174789
17	+	CHRMB+dox_WT+dox	XRN2	3.050047557
18	+	CHRMB+dox_WT+dox	TRIP13	3.473620097
19	+	CHRMB+dox_WT+dox	HERC2	3.521046956
20	+	CHRMB+dox_WT+dox	TOP2A	3.927849452

Table 4: Student t-test for protein interactors extracted from Perseus

The list above is the tabulated proteins identified using t-test and after filtering according to gene ontology: biological function. Proteins are arranged from the smallest to biggest enrichment value based on the statistical test performed. A negative value indicates that the protein is found to be more enriched in FLAG-Upf1^{Wt} sample whereas a more positive value indicates enrichment in FLAG-Upf1^{ChrmB} sample. The (+) sign in first column indicates that the proteins listed are found to be statistically significant in induced ChrmB and WT samples when compared. The last column is the enrichment value obtained from the statistical test performed between induced ChrmB and WT samples.

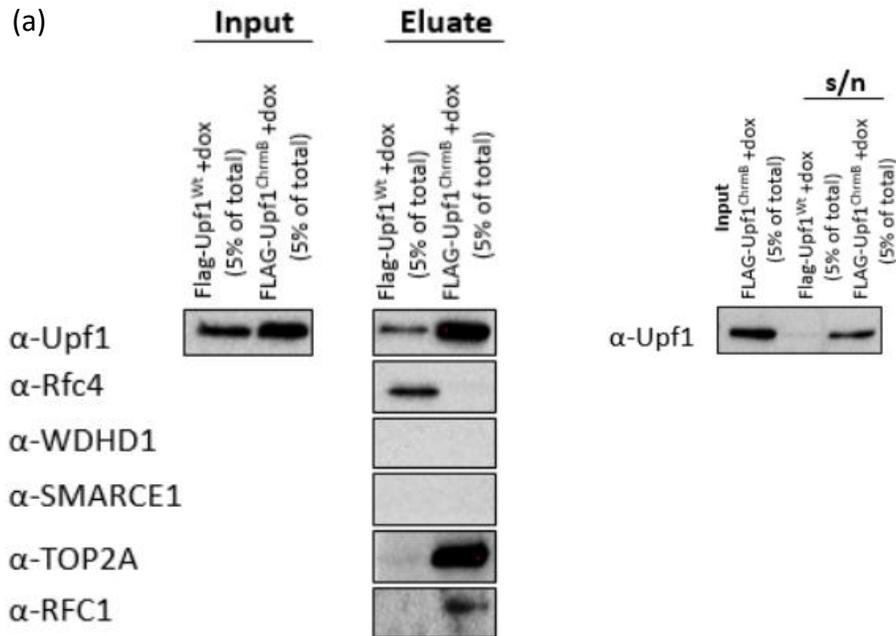


Figure 5.10a: Validation of interacting proteins by Co-IP

Immuno-precipitation was performed using 2mg of protein lysate, 30 μ l of a 50% slurry of M2 anti-Flag-agarose derived from cells in which either FLAG-Upf1^{Wt} or FLAG-Upf1^{ChmB} had been induced as described in section 2.11. Elution was by boiling in sample loading buffer. (a) Immunoblotting using 5% of input samples and eluate was performed to determine the relative amount of indicated proteins interacting with Upf1 using antibodies stated above and listed in detail in Chapter 2. The absence of a band in for both WHDH1 and SMARCE1 could be due to very low interaction that made it harder to detect using western blotting or that the conditions of blotting were not as optimal as it should be.

(b)

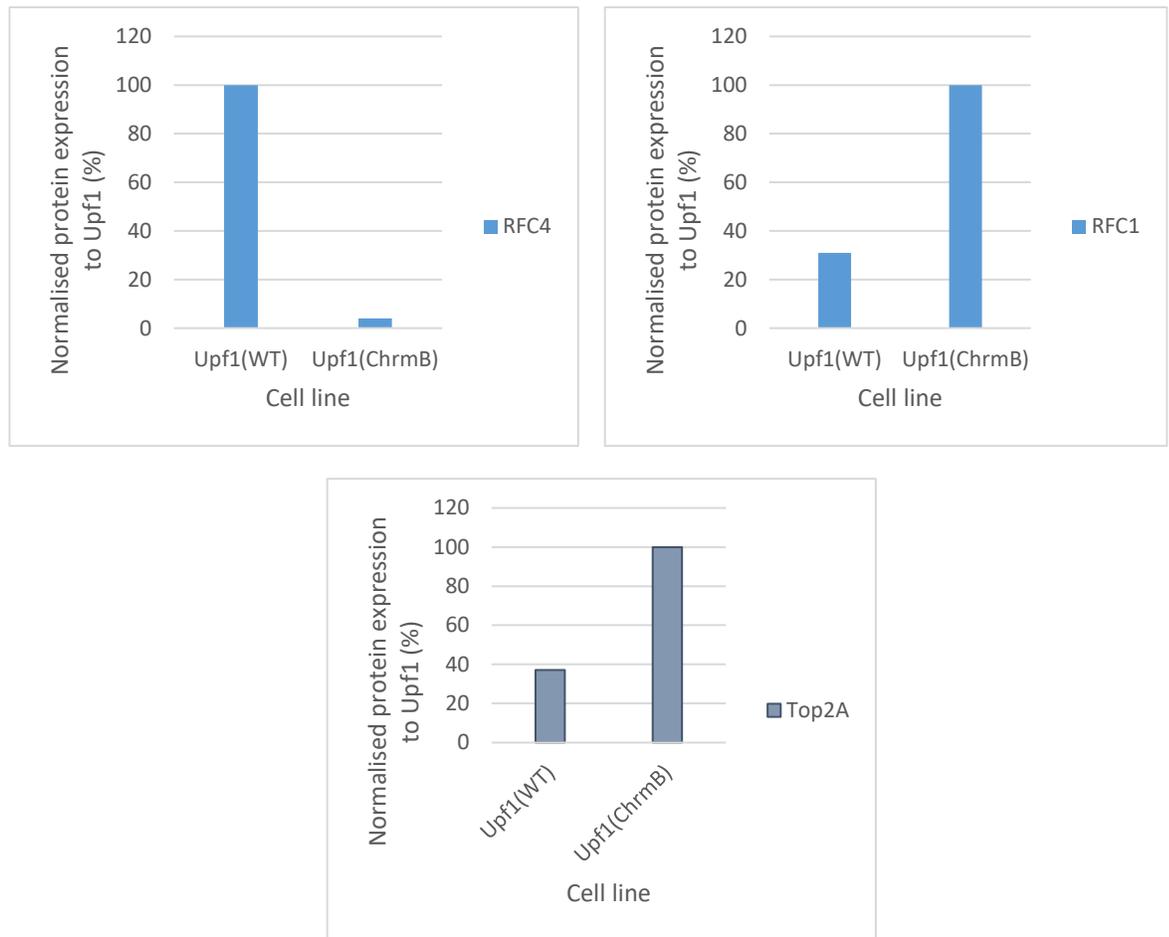


Figure 5.10b: Validation of interacting proteins by Co-IP

Immuno-precipitation was performed using 2mg of protein lysate, 30 μ l of a 50% slurry of M2 anti-Flag-agarose derived from cells in which either FLAG-Upf1^{Wt} or FLAG-Upf1^{ChrmB} had been induced as described in section 2.11. Elution was by boiling in sample loading buffer. (a) Immunoblotting using 5% of input samples and eluate was performed to determine the relative amount of indicated proteins interacting with Upf1 using antibodies stated above and listed in detail in Chapter 2. (b) is the quantification of (a), where proteins RFC4, TOP2 α and RFC1 was normalised to either Upf1^{Wt} or Upf1^{ChrmB}.

5.3 Conclusion

Mass spectrometry (MS) has an exceptional ability to identify very small amounts of protein without prior knowledge of its nature or function, and in principle is capable of determining the protein composition of any molecular assembly (Hein et al. 2013). Its sensitivity routinely allows identification of peptides that are present at femtomole level making MS the method of choice (Gingras et al. 2007) for the identification of proteins that are present at very low levels in biological samples. Coupling MS with an affinity purification step makes it a powerful tool for interaction proteomics. In this study, the “bait”; FLAG-Upf1 is used to isolate “prey” proteins that interact with it, which then can be identified by MS.

This approach was applied to investigate the differences in the nature of chromatin associated and cytoplasmic Upf1 interacting proteins using FLAG-Upf1^{Wt} and FLAG-Upf1^{ChmB} expressing cell lines by comparing each IP samples in three separate biological replicates. This method may not be optimal for the identification of weak or transient interactions, because, in the first case, the dissociation of interactors during sample processing (which involves multiple washing steps to reduce the extent of non-specific protein binding) can result in severe reduction of representation in the sample to the point where no material may remain (Yeung et al. 2008; Hein et al. 2013). Transient interactions may reflect a mechanism involving significant change in the polypeptide composition of a protein complex, and depending on the precise mechanism may result in a failure to identify some components. In the latter scenario, the use of the bioID approach (Roux et al. 2013) may be preferable. Washing steps are unavoidable and thereafter, non-specific binding proteins will still be present. Hence, quantitative proteomic analysis was utilised to differentiate enrichment of true interactors from non-specific background proteins by comparing specific affinity-purified samples to controls using an inducible expression system (Hein et al. 2013).

5.3.1 Advantages and disadvantages of affinity purification-mass spectrometry (AP-MS)

Immuno-precipitation (IP) is the term used in this study to describe the process of isolating FLAG-Upf1 and its interacting partners from cell lysates using the ANTI-FLAG M2 agarose. It is a subset of the broader concept of AP-MS (affinity purification -mass spectrometry) where the same concept of purification and enrichment of proteins and associated complexes is performed by exploiting available affinities or specific antibodies (Hein 2014).

Compared to other methods that could be used for detecting protein-protein interactions such as the yeast-two hybrid system (Gingras et al. 2007; Aebersold & Mann 2003; Schweingruber et al. 2016), AP-MS has several advantages. AP-MS offers a method that allows protein complexes to be purified in their native state from any cell or tissue lysate, and allows multiple isoforms to be interrogated simultaneously (Dunham et al. 2012). Additionally, this method allows posttranslational modifications (PTM) to be identified without further PTM-focussed enrichment. This is important as it can enable an in-depth assessment of the role that these modifications play in regulating interactions. AP-MS also enables the characterisation of protein-protein interactions in a near physiological context (Nesvizhskii & Aebersold 2005; Dunham et al. 2012). This is an important aspect, especially in the context described here, as the aim was to preserve the protein interactions in as native a state as possible.

However, there are drawbacks to this method. Firstly, it requires substantial amount of input material especially in this study, as the requirement to isolate the chromatin fraction meant that the number of cells required to yield 4mg of lysate and its chromatin equivalent was substantial. Secondly, it is only applicable to protein-containing complexes for which antibodies are available, or which can be tagged without affecting normal function. In this case, Upf1 was tagged at the N-terminus with the Flag epitope, and cells expressing Flag-Upf1 alone have been shown previously to be competent for both nonsense-mediated decay as well as telomere maintenance (Turton 2014). An additional drawback is that MS is prone to identifying “contaminant”

proteins, generating many false positives (Hein 2014; Dunham et al. 2012; Nesvizhskii & Aebersold 2005). However the approach used here, in which relative quantitative enrichment, coupled with statistical analysis of biological replicates helps to address this problem, as only significantly enriched proteins, associated with doxycycline expression in each case, are identified (Figure 5.8a and b) and subsequently compared for differential enrichment depending on the precise subcellular source of the protein (in this case cytoplasm versus chromatin-associated).

5.3.2 Identification of protein interactors

In this study, protein extracted from chromatin fraction of cells expressing FLAG-Upf1^{Wt}, and cytoplasmic lysates from cells expressing FLAG-Upf1^{ChmB} were used to compare against control fractions from non-expressing cells to enable the distinction between statistically significant interactors and non-specific proteins binding either to the antibodies or to the matrix. With the help of statistical and analytical software packages such as MaxQuant and Perseus, quantitative analysis was performed to eliminate background contaminants. MaxQuant settings used were as described in section 2.13.4.

A total of 91 known protein interactors were identified in this MS analyses. These results and the protein identifications are similar to what has been reported previously in the literature (Flury et al. 2014; Schweingruber et al. 2016) together with the ones reported in the BioGRID database . All proteins involved in NMD except SMG6 were identified, consistent with the finding of Azzalin et al. (2007) that all seven SMG proteins are present on chromatin in HeLa cells.

However, some proteins that were previously identified by the three sources were not found in this MS analysis. There are several possibilities as to why previously identified interactors were not detected here. Firstly, it is conceivable that the proteins did not interact under the tested conditions (Cox et al. 2014; Karpievitch et al. 2012). As many biological interactions are of low affinity, transient and dependent on specific

cellular environment, this could be a potential reason why some proteins were not detected in this analysis (Aebersold & Mann 2003).

Secondly, the location of the affinity tag at the N-terminus of Upf1 could have disrupted possible interactions or that the conditions used for isolation during IP were too harsh to preserve the interaction. Even though a previous study from this laboratory (Turton 2014) has shown that the presence of an N-terminal affinity tag does not interfere with normal NMD function, it did not exclude the possibility that the tag might interfere with some molecular interactions. Thirdly, the nature of the protein samples used in this study compared to the ones used in Flury et al. (2014); Schweingruber et al. (2016) is different ie chromatin-enriched fraction was used here and although whole cell lysate was also used, the focus of protein identification was also concentrated to nuclear/ chromatin binding proteins.

Finally, if the bait protein is overexpressed or that the prey protein is in much lower abundance than other components in the sample then these circumstances may explain why known interactors were absent in this analysis (Gingras et al. 2007). Also, weak interactions by nature are more prone to variable dissociation, making the detection of such an interaction less reproducible (Hein 2014), as biological variation has a negative impact on the statistical significance of a low level of enrichment. Since the amount of chromatin-bound Upf1 is known to be relatively low even after overexpression, detection of low abundant interactors of Upf1 could pose a challenge.

As seen in **Figure 5.7**, the correlation values calculated using Pearson correlation was between 0.6-0.9 which indicates that there was a reasonably high degree of reproducibility between the replicates.

Using Perseus-based statistical analysis, enriched interacting proteins were determined and can be visualised via a volcano plot (**Figure 5.8**). I was able to identify 313 proteins that were considered as significant protein interactors and using a series of bio-informatic filtering tools, 24 were identified as potential proteins of interest. 20

proteins categorised according to ontological functionality (DNA damage, DNA replication, telomere maintenance, DNA binding and chromatin binding/remodelling) were plotted in a Venn diagram for visualisation (**Figure 5.9**). The remaining four proteins were excluded based on their ontology which did not fall into the categories stated above. A total of 5 proteins were selected, one from each category (WDHD1- DNA damage, TOP2A- replication, RFC1- telomere maintenance, RFC4- replication & SMARCE1- chromosome remodelling) to undertake validation of the mass spectrometric analysis. Importantly, of five proteins tested, 3 showed a direct interaction with Upf1 when co-immunoprecipitation was performed (**Figure 5.10**)

5.3.3 Role of Upf1 in DNA replication and damage response

Molecular functions such as DNA repair, replication, mitosis and transcription when disrupted lead to a number of diseases (Torrente et al. 2011) and all contribute to homeostasis of chromatin. Azzalin and Lingner (2006b) were the first to identify Upf1 as a chromatin-associated protein, with maximal chromatin association observed in synchronised cells progressing through S phase. In addition they showed that DNA damage resulted in increased association of Upf1 with chromatin and that this association was dependent on the action of the checkpoint PIKK family protein kinase ATR (Azzalin & Lingner 2006b). They also showed that Upf1 co-immuno-precipitated with p125 (Azzalin & Lingner 2006b; Azzalin & Lingner 2006a), a component of the replisome, although it was not clear whether this interaction is direct or indirect. Depletion of Upf1 results in an increase in γ H2AX expression, significant increase in telomere-free ends, and loss of Upf1-associated telomerase activity (Chawla et al. 2011), suggesting that chromatin-associated Upf1 has a role in replication of telomeric DNA.

Evidence that the function of Upf1 in its chromatin-associated role is independent of its NMD function is based on the observation that depletion of Upf2, a key component of NMD does not show any of the phenotypes described above (Azzalin & Lingner 2006b).

Although evidence has been gained for the involvement of Upf1 in telomere maintenance, there is little to no information regarding the interacting partners of Upf1 that might be responsible for its nuclear function. Azzalin & Lingner (2006b) reported several indicators that Upf1 might play a role in DNA replication or repair. In their study, they managed to show that Upf1 interacted with a subunit of the polymerase δ , p125 but Upf2 did not. This interaction was enhanced during S-phase. It can thus be assumed that Upf1 may act as a replicative helicase in regulating the replication fork progression or during DNA repair. Another nuclear protein that has been shown to associate with Upf1 is TPP1. TPP1 is a component of shelterin, a safe guard complex that is present at the ends of every chromosome, both to protect it, and to prevent it from activating the DDR cellular machinery that detects and attempts to repair double strand breaks, with which telomeres are indistinguishable (Palm & de Lange 2008). Immunoprecipitates obtained using two independent antibodies against TPP1 were enriched for endogenous Upf1 protein (Chawla et al. 2011) while Upf2 was not detected in the same analysis.

DNA replication and DNA damage responses share many critical proteins (Izawa et al. 2011). Supporting this, several proteins found in this study have been implicated in DNA replication and repair mechanisms.

Based on the filtering that was performed on the 313 protein interactors, 24 were of potential interest; UBR5, WDHD1, SMARCD2, RFC1, RFC2, RFC4, HIST1H2BN, HERC2, ACTL6A, PARP1, TOP2A, DEK, NASP, MCM6, TRIP13, SMARCB1, SMARCE1, FUBP1, SMARCD1, SMARCA4, FUBP3, XRN2, AKAP8 and SPIN. I performed another round of filtering based on their gene ontology (DNA binding, DNA damage, DNA replication, telomere maintenance and chromatin binding/remodelling) reducing the number to 20. 14 proteins out of the 20 identified are involved in either DNA replication or DNA damage response or both (UBR5, WHDH1, HERC2, ACTL6A, PARP1, TOP2A, RFC1, RFC2, RFC4, DEK, TRIP13, XRN2, NASP & MCM6).

5.3.4 Validation of protein interactors

To confirm the data that was been obtained through MS analysis, a co-immunoprecipitation experiment followed by immunoblotting was performed for five proteins; WDHD1, RFC1, RFC4, TOP2A & SMARCE4. Detection was undertaken using commercially available antibodies (Section 2.2).

As seen in **Table 4**, WDHD1 was the most enriched protein found in FLAG-Upf1^{Wt} enriched samples (-4.91654). This protein is associated with DNA replication where its presence is essential for replication initiation (Zhu et al. 2007). In addition, a recent study by Hao et al. (2015) has shown that WDHD1 protein is recruited to DNA damage sites in response to replication stress and its phosphorylation is required for both for its interaction with Claspin and for efficient Chk1 activation. The interaction between Upf1 and WDHD1 protein, would be consistent with a role for Upf1 in resolving aberrant DNA or DNA-RNA hybrid structures formed as a result of the emergence of stalled replication forks (Yoshizawa-Sugata & Masai 2009; Jones & Petermann 2012). Consistent with this notion, in addition to evidence placing Upf1 at telomeres (Chawla et al. 2011; Azzalin, Reichenbach, Khoriantuli, Giulotto & Lingner 2007; Azzalin & Lingner 2006b), indirect immunofluorescence microscopy has shown that Upf1 colocalises with sites of DNA damage (H2AX positive) that do not contain the telomere component Rap1 (Chawla et al. 2011) indicating that in addition to ensuring telomere stability, Upf1 must play a role in resolving DNA damage at sites distinct from telomeric regions. However, WDHD1's interaction with Upf1 could not be validated in the time available, by western blotting. Further optimisation of antibody based detection methods and increased loadings on SDS-PAGE gels could be used to establish the validity of this observation.

In this study, a number of RFC (Replication Factor C) protein subunits were identified as significant protein interactors. RFC is a five subunit protein complex that is required for coordinating the leading and lagging strand DNA synthesis during S-phase and DNA repair in eukaryotic cells (Ellison & Stillman 1998). The RFC protein family functions as a clamp loader that loads PCNA onto DNA in an ATP-dependent process

during DNA synthesis (Xiang et al. 2014). RFC1 appears to have additional functionality, as this gene encodes the large subunit of replication factor C, a five subunit DNA polymerase accessory protein, which is a DNA-dependent ATPase required for eukaryotic DNA replication and repair. The large subunit acts as an activator of DNA polymerases, binds to the 3' end of primers, and promotes coordinated synthesis of both strands. It may also have a role in telomere stability (Shore 2001; Hedglin et al. 2013).

Out of the five RFC subunits, three were identified as enriched in Upf1 containing fractions; RFC1, RFC2 & RFC4. RFC2 and RFC4 were found to be more enriched in the FLAG-Upf1^{Wt} containing sample while RFC1 was more enriched in FLAG-Upf1^{ChrmB}. A stable interaction between Upf1 and RFC 1 and 4 in the relevant Upf1 fractions was detected by western blotting confirming their differential interactions.

RFC1 is a component of the 5 subunit clamp loading complex (Ellison & Stillman 2003; Yao & Donnell 2012; Overmeer et al. 2010). However, it has recently emerged that the RFC2-5 core complex may associate with distinct large subunits other than RFC1. Three other proteins, Elg1, Ctf18 and Rad17 have been shown to associate with the RFC2-5 core complex, each taking the place of RFC1 and conferring distinct functions on the resultant RFC isoform (Majka & Burgers 2004; Yao & Donnell 2012).

Perhaps the best understood alternative clamp loader is the Rad17-RFC, in which RFC1 is replaced by Rad17 (Rad24 in *S. cerevisiae*) (Yao & Donnell 2012; Bermudez et al. 2003). Unlike the other two RFC1-like proteins, Rad17-RFC does not load PCNA onto DNA but loads the Rad9/Rad1/Hus1 complex at DNA damage sites during a replication checkpoint response (Leman & Noguchi 2013) as discussed further in Chapter 6.

It is intriguing that the chromatin-associated form of Upf1 was found associated with 2 of the 5 components involved in loading proteins onto replication forks, and additionally that the mutant which does not bind chromatin is associated with another distinct component of the same complex. Although there are insufficient details to

make specific conclusions about these observations, it seems reasonable to propose that Upf1 interacts with RFC clamp loading function in some way to associate with chromatin, and perform a function at the replication fork. The distinct functions of the RFC components, and an incomplete understanding of the significance of the potential structural alteration of the Chrm B mutant makes further interpretation difficult. However, these data are potential reasons why RFC1 associates with a form of Upf1 that is incapable of binding chromatin, while RFC2 and 4 do. A more detailed discussion regarding these proteins is elaborated in chapter 6.

The structural nature and supercoiling of DNA is essential for nuclear packaging, however processes such as replication and transcription requires the separation of strands (Harkin et al. 2016). Type II topoisomerase makes transient double-stranded breaks into one segment of DNA and pass an intact duplex through the broken DNA before resealing (Watt & Hickson 1994; Nitiss 2009). TOP2A is the isoform that was found in this analysis and is validated for its interaction with Upf1 by immunoblotting. The immunoblot, (**Figure 5.10**) clearly shows that TOP2A was present in FLAG-Upf1^{ChrmB} IP proving an interaction with Upf1.

It is known that TOP2A are expressed in regions where cell division is abundant and is important for replication (Harkin et al. 2016; Nitiss 2009). Based on the functions that have been indicated, it might have been expected that TOP2A would be more abundant in FLAG-Upf1^{Wt} sample. There could be a possibility that although FLAG-Upf1^{ChrmB} mutant cell line is unable to bind to chromatin, it is still able to initiate the recruitment of proteins that are involved in resolving DNA damage events, possible acting in a dominant negative fashion.

SMARCE1 was also significantly enriched in FLAG-Upf1^{Wt} enriched samples (**Table 2**). SMARCE1 is a component of the SWI/SNF chromatin remodelling complexes that carries out enzymatic activities that changes chromatin structure (Euskirchen et al. 2012). Also enriched in FLAG-Upf1^{Wt} were three other SWI/SNF family of chromatin-remodelling complexes; SMARCD1, SMARCD2 AND SMARCB1. Importantly SWI/SNF

chromatin remodelling complexes regulate DNA topoisomerase 2A function (<http://cancerres.aacrjournals.org/content/75/19/4176>). The interaction of forms of Upf1, itself a DNA/RNA helicase with both of these chromatin remodelling complexes provides for the possibility that Upf1 has an additional function in chromatin remodelling, which may or may not be independent of its role in a DNA damage response. Similar to WHDH1 protein, I was unable to detect SMARCE1 by immunoblotting.

The inability to detect a band for WDHD1 and SMARCE1 by immunoblotting does not necessarily mean that interaction is not present. It is possible that the interaction between Upf1 with each of the proteins is weak or occurs at a low stoichiometry. MS has a very high sensitivity which can make it incompatible to compare with antibody detection. Therefore even if a protein is detected during a MS analysis, it might not be enough to be visualised through the means of western blotting.

Another possibility is that conditions of processing the blot might not have been optimal enough for bands to appear. Due to the amount of sample that was left for validation, I could only afford to use 5% of total IP eluate. Therefore, if the amount present in the sample is already at low levels, hence there would be harder to visualise on western blot. I have also tried optimising the conditions for blocking and incubation of both primary and secondary antibody but bands were still invisible. Should I had more time, it would be interesting to do another round of western blotting using a larger amounts of sample for validation.

Another technique to validate protein interaction is Proximity Ligation Assay (PLA) also could be used if there was sufficient time. PLA is a technique that allows transient protein interactions to be detected. This technique allows exceptional specificity and sensitivity of protein detection and quantification for immunocytochemistry (ICC) and immunohistochemistry (IHC) applications in unmodified cells (<http://www.abnova.com>). PLA utilises a pair of oligonucleotide labelled antibodies (PLA probes) that is able to generate an amplified signal only when the probes are in

close proximity (Sigma-Aldrich 2017). This signal could be detected using traditional immunofluorescence (IF) or IHC protocols.

CHAPTER 6

6.0 General discussion and future perspectives

6.1 Introduction

Upf1 is a highly conserved, ubiquitously expressed protein present in a number of subcellular locations. Its cytoplasmic role is most widely studied. The presence of Upf1 in the nucleus at low levels raises questions to what processes operating there might require the involvement of Upf1. Currently, nuclear Upf1 has been implicated in both telomere maintenance as well as global DNA replication/repair, although the link between these remains unclear.

Previously, it has been reported that nuclear Upf1's role was independent of its NMD function, which predominantly takes place in the cytoplasm (Azzalin & Lingner 2006b; Azzalin & Lingner 2006a). Levels of chromatin-bound Upf1 have been reported to be low at M and early G₁ phase but starts to increase mid G₁ and reaching a peak in S-phase before reducing again at end of the S-phase (Azzalin & Lingner 2006b; Varsally & Brogna 2012). This is the first indicator that Upf1 is involved in the replication process.

Secondly, Upf1-depleted HeLa cells were reported to be unable to proceed through S phase suggesting that loss of Upf1 results in S-phase arrest (Azzalin & Lingner 2006b). Cell cycle arrest was suggested to have been triggered by DNA damage arising from the inability of Upf1 to repair or rectify some aberrant nucleic acid structures (Azzalin & Lingner 2006b). However, other investigators had not observed S phase arrest as a consequence of Upf1 depletion (Turton 2014) and thus the contribution of Upf1 to S phase progression remains controversial.

Depletion of Upf1 also results on the accumulation of markers of DNA damage such as expression of γ H2AX (Chawla et al. 2011) and leads to an ATR-mediated DNA damage response (Azzalin & Lingner 2006b; Varsally & Brogna 2012). ATR, a member of the PIKK family is thought to be one of the regulators for Upf1. In a study where cells were depleted of ATR, chromatin association and levels of phosphorylated Upf1 were

reduced (Azzalin & Lingner 2006b; Imamachi et al. 2012). Although it is known that ATR can directly phosphorylate Upf1 in vivo, it is not clear whether or not it participates in the cyclical phosphorylation/dephosphorylation cycle is mediated by SMG1 during NMD (Chawla et al. 2011). These observations together with those of Carastro et al. (2002), who identified Upf1 as the helicase activity that co-purifies with DNA polymerase delta (pol δ) from bovine spleen, contributes to the notion that Upf1 is involved in DNA replication and may participate in the resolution of DNA damage that arises during the normal progress of S phase.

Upf1 has also been implicated in the maintenance of telomeres. Telomeres are repeating DNA sequence at the ends of the chromosomes and during each round of replication, telomeres become shortened, because of the inherent problem associated with lagging strand synthesis at chromosome ends (Axelrad et al. 2013). Chawla et al. (2011) highlighted the importance of ATR for Upf1's ability to associate with telomerase, which regulates telomere length directly. They showed that depletion of Upf1 results in an accumulation of the DNA damage marker, γ H2AX, which partially co-localised with a component of the shelterin protein complex, Rap1. However, in addition, it is clear from the studies that expressed γ H2AX observed is not only localised at telomeres, but it also co-localises with RPA32, a subunit of the single-stranded DNA binding protein, replication protein A which coats unwound DNA at the replication fork and is essential for DNA replication and repair (Chawla et al. 2011).

The main aim of this study was to characterise the role of nuclear Upf1 in regards to maintaining the genome integrity via DNA replication and/or DNA repair. There are three major lines of this study 1) to establish the genomic integrity phenotype of different Upf1 mutants by investigating the importance of known phosphorylations sites, 2) to determine their functionality as helicases to function and to uncover novel protein interactors of Upf1 that might be responsible for assisting Upf1 in maintaining genome stability.

6.2 The role of known Upf1 phosphorylation sites in maintaining genome stability

Although it has been established that nuclear Upf1 is independent of its NMD role, I wanted to investigate whether there are similarities and differences of Upf1's molecular mechanism that enables it to maintain genome stability. With that in mind, I proceed to investigate the importance of phosphorylation sites that have been recognized as essential for NMD (Okada-Katsuhata et al. 2012; Schweingruber et al. 2013; Ohnishi et al. 2003). The cell lines used in this study comprised one capable of expression of a form of Upf1 (Upf1^{AAA}) which is unable to be phosphorylated at any of the previously established canonical phosphorylation sites (T28A, S1096A and S1116A), together with one (Upf1^{EEE}) in which all three canonical sites contain phospho-mimetic residues at all three sites (T28E, S1096E and S1116E). The third cell line is capable of expression of a form of Upf1 containing an S42A mutation previously shown to have lost the ability to associate with chromatin (Turton 2014).

Neither FLAG-Upf1^{AAA} or FLAG-Upf1^{EEE}, unlike FLAG-Upf1^{Wt}, were able to rescue the DNA damage phenotype - as measured by expression of γ H2AX - that results when endogenous Upf1 was depleted **Figure 3.6**. One explanation for this observation might be that both mutant forms lacked sufficient helicase activity, as siRNA-resistant Upf1 helicase mutants have been shown previously to be unable to restore function in Upf1-depleted cells (Kaygun & Marzluff 2005a; Chamieh et al. 2008). However, as discussed below, expressed forms of these mutants showed similar levels of helicase activity compared to the wild-type protein. Taken together, these data suggest that phosphorylation /dephosphorylation at one or more of these sites is likely to be important for chromatin-associated Upf1 function. Such a conclusion is consistent with the observation (Azzalin, Reichenbach, Khoriantuli, Giulotto & Azzalin 2007) that Smg1, Smg5/7 were shown to be associated, at least with telomeric DNA, if not, to date, with replication fork components generally. Models for NMD suggest that phosphorylation of the N-terminal site of Upf1 (T28) may not occur simultaneously with phosphorylation

of either of the C-terminal sites discussed here (Okada-Katsuhata et al. 2012; Nicholson et al. 2014). It is conceivable that only N- or C-terminal sites undergo regulatory phosphorylation on chromatin-associated Upf1. Should opportunity have permitted, it would have been useful to examine the phenotypes associated with individual phosphorylation sites. In addition, this latter approach would have been informed by mass spectrometric analysis of immunoprecipitated Upf1 and the determination of stoichiometry of phosphorylation at each site. Due to limitations associated with the relatively low levels of Upf1 associated with chromatin, it was not possible to devise a reliable strategy in the available time to determine the phosphorylation status at each of the sites discussed here although two of the C-terminal sites were found to be phosphorylated (**Appendix 1**).

Experiments were undertaken in which endogenous Upf1 remained present, and either Upf1^{EEE} or Upf1^{AAA} were overexpressed. The rationale behind these experiments was that, given that Upf1 participating in NMD requires cyclical phosphorylation/dephosphorylation, it might be expected that one or other mutant form might act in a dominant-negative fashion, trapping a molecular complex in such a way that further resolution was not possible. However this was not the case and in the absence of further phosphorylation data, additional experiments were not pursued.

In contrast, Upf1^{ChrmB} did act as a dominant-negative (**Figure 3.6**). Thus in the presence of normal, wild-type Upf1, γ H2AX is expressed indicating that genome integrity is compromised in these circumstances. This indicates that, despite the fact that Upf1^{ChrmB} cannot bind to chromatin, the ChrmB mutant is likely to associate with a key factor required to prevent the emergence of DNA damage. One explanation for this result might be that Upf1^{ChrmB} acts directly or indirectly to deplete chromatin of Upf1^{Wt}. However, the structure of Upf1 has been reported (Schell et al. 2003; Karousis et al. 2016; Fiorini et al. 2015) and it is a monomer, suggesting that Upf1^{ChrmB} does not form inactive multimers with wild-type Upf1.

That Upf1^{ChrmB} expression has a dominant negative phenotype gave rise to the consideration that, irrespective of the precise mechanism, it might be possible to identify a selective binding partner for this mutant by mass spectrometry that might provide insight into the mechanism of action. In parallel, it was clear it might also be possible to identify differential binding partners for both chromatin associated Upf1^{Wt} and Upf1^{ChrmB}, if it were the case that the lack of chromatin association of Upf1^{ChrmB} was not a consequence of a failure to bind or unwind DNA via a conformational change in the N-terminus of the protein (Fiorini et al. 2013; Okada-Katsuhata et al. 2012). It followed therefore, that an analysis of the helicase functionality of Upf1 - wild type and the mutants described was called for. During the course of the work undertaken for this thesis, a collaboration between the C.Sanders (University of Sheffield medical school) and the Smythe laboratory was established to develop a purification procedure for bacterially expressed, wild-type Upf1 (Dehghani-tafti & Sanders 2017). I utilised this new procedure to purify to homogeneity, preparations of full-length Upf1AAA, Upf1EEE, and Upf1ChrmB, involving expression in *E. coli*, and subsequent purification by nickel agarose chromatography.

Protein corresponding to wild-type Upf1 and mutants (EEE, AAA & ChrmB) were used in strand displacement assays (Dehghani-tafti & Sanders 2017). These experiments established that all forms of Upf1 investigated displayed equivalent levels of displacement activity, and, by inference, DNA binding ability comparable to the wild-type. These results are intriguing, as they are not simply reconciled with the data of others (Chamieh et al. 2008; Chakrabarti et al. 2011) suggesting that both N- and C-terminal regions of Upf1 are auto inhibitory unless de-inhibited by phosphorylation, as this circumstance would be expected to give rise to results in which the UPF1EEE mutant would be more active than other forms of the protein. However, it is conceivable that subtle variations in strand displacement activity may require some of the additional components known to bind either the N-terminus or the C-terminus of Upf1. Importantly, the data showed that Upf1ChrmB both binds to DNA and undertakes strand displacement activity equivalent to that observed with the wild-type protein.

This results suggested that the failure of the ChrmB mutant to bind to chromatin is not caused by a failure in DNA binding or helicase activity mediated association, but might conceivably be due to the inability of this form of the protein to interact with other chromatin-associated proteins in vivo to enable it to tether to DNA.

6.3 Upf1's role in DNA replication and/ or repair might be orchestrated by its ubiquitin ligase activity

Different PTM influences different enzymatic activity, protein turnover and localisation, protein-protein interactions and also DNA repair pathways (Karve & Cheema 2011). It is known that NMD absolutely requires the helicase activity of Upf1 (Chamieh et al. 2008; Chakrabarti et al. 2011). The essential cyclical phosphorylation and dephosphorylation (Ohnishi et al. 2003) is required to ensure Upf1-mediated degradation of aberrant mRNA. However it has been highlighted recently that Upf1 has ubiquitin ligase (E3 ligase) activity (associated with the N-terminal CH domain) and this activity is necessary to regulate human skeletal muscle differentiation (Feng et al. 2017).

Ubiquitination, an irreversible modification has been known to regulate many processes in the cell such as membrane trafficking, DNA repair, histone regulation and transcription (Miranda & Sorkin 2007; Gumeni et al. 2017). This PTM results in the covalent attachment of ubiquitin to a lysine residue on target proteins (Panier & Durocher 2009). The type of ubiquitination might be expected to affect the fate or function of target proteins. Ubiquitination has been highlighted to have an important regulatory role in DNA damage response (DDR) where all major DNA-repair pathways, damage-avoidance mechanisms and checkpoint responses are in some way regulated by ubiquitination (Bergink & Jentsch 2009). Unfortunately a determination of the consequences of expressing ubiquitin ligase deficient forms of Upf1 in cells deficient in wild-type protein was not feasible in this work, although it will be an important future avenue of research. Ubiquitination of Upf1 was not examined in the mass spectrometry analysis undertaken here. Should it turn out that the ubiquitin ligase function of Upf1 is required for genome integrity and/or telomere maintenance, then it would be very important to identify the molecular targets of this activity. One additional possibility is that Upf1 may well be a substrate for itself. In several experiments, tagged versions of Upf1 was observed to migrate as a doublet (**Figures 4.3 & 4.5**). The basis for this is unknown, and it might involve some form of additional post-translational modification.

The possibility that it also reflects a degree of proteolysis during sample workout cannot be ignored.

6.4 Successful establishment of an immuno-precipitation protocol for chromatin-associated and cytoplasmic Flag-tagged UPF1 for mass-spectrometric analysis

In this work, the method developed for the identification of Upf1 peptides with around 50% sequence coverage provided insights into the strategy and scale of samples needed to undertake IP, sufficient for LC-MS/MS analysis. Previous studies have used whole cell lysate for MS analysis, whereas in this study, both cytoplasmic and chromatin fractions were used.

During the IP step, the inability to successfully elute Upf1 specifically using the Flag peptide had the consequence that a significant number and amount of non-specific proteins were also precipitated together with Upf1. The reason for the inability of the peptide to compete with protein specifically bound to anti-Flag antibodies is unknown.

Some proteins that were previously identified by the three sources were not identified in this study. Several reasons could be highlighted. The most likely explanation for some of these differences relates to one of the common limitations of shotgun proteomics analysing proteins derived from complex samples, and is related to the ability of the spectrometer itself to correctly identify specific peptides in complex mixtures. The variation in peptide and thus protein identification arises because MS peptide identification relies both on the abundance and number of peptides present at any one time in the mass spectrometer and the time duration (effectively the elution peak width from the upstream HPLC) that they are available for analysis. MS analysis uses data dependent analysis. Thus in any sample, the most abundant 10 peptide(s) are analysed first and will be subjected to MS/MS, with acquisition of parental ion spectra and subsequently fragment ion spectra that enable sequence identification of each peptide. Only then, are less abundant peptides analysed, and depending on the cycle

time of the instrument, the time taken to perform the first analysis may mean that low abundance peptides will have passed through the device without being analysed. It follows that the complexity (ie number of different peptides present) within a sample also has the effect of reducing the overall sequence coverage of any protein. This in turn has the effect of decreasing the calculated probability associated with any given protein, at a particular abundance.

Secondly, as mentioned before in Chapter 5, it is conceivable that the proteins did not interact under the tested conditions (Cox et al. 2014; Karpievitch et al. 2012). Many biological interactions are transient or of low affinity hence that is why some proteins that were detected before was not detected here (Aebersold & Mann 2003). Also, weak interactions by nature are more prone to variable dissociation, making the detection of such interaction less reproducible (Hein 2014). Thirdly, is the presence of affinity tags. Although FLAG-tag used in this study is considered small, the notion of how it could disrupt possible interactions should not be dismissed since it has not been tested. And finally, both protein sources and experimental conditions used in this study are comparably different to the other three sources, for example chromatin enriched lysate vs whole cell lysate. Since it has been reported that chromatin bound Upf1 are of low levels even after overexpression, therefore detection of low abundant proteins interactors of Upf1 could pose a challenge. However, despite these issues, I was able to identify a number of novel interacting proteins of Upf1 using MaxQuant and Perseus analytical software.

6.5 Upf1 may regulate DNA replication and/or repair through identified protein interactors

Azzalin & Lingner (2006b) and Chawla et al. (2011) have proposed that Upf1 might be involved in DNA replication and also in resolving DNA damage. A number of proteins involved in NMD, which predominantly takes place in the cytoplasm (Reichenbach et al. 2003; Azzalin et al. 2007), have also been found in the nucleus which suggests that there is significant overlap in molecular mechanism undertaken by Upf1 in NMD and in its genome integrity role.

The discussion below relates to the set of novel proteins associated with the gene ontology topics of relevance that were identified by mass spectrometry as being distinctly associated with one or other form of immunoprecipitated Upf1 and validated, using immunoblotting to interact with the relevant form of Upf1. It is unknown, at this stage, whether the interactions detected occur directly or require tethering via other proteins.

6.5.1 Upf1 might bind to RFC 2-5 complex forming RFC-like complex (RLC)

The canonical DNA clamp loading complex, RFC that operates during normal DNA replication is made up of five RFC subunits; a large RFC1 subunit and four smaller core RFC subunits termed RFC2-5 (Yao & Donnell 2012; Majka & Burgers 2004). All subunits of RFC (with the exception of Ctf8) are members of the AAA+ ATPase family and contain ATP binding Walker A and B motifs (Kubota et al. 2013). In humans, there are at least 3 alternative clamp loading complexes (Lee et al. 2010; Majka & Burgers 2004; Overmeer et al. 2010; Leman & Noguchi 2013) with distinct functions (**Figure 6.1**). Each clamp loader function is conferred by the identity of the large subunit associated with RFC2-5. Thus CTF, RAD24 and ELG1/ATAD have each been shown to bind to the RFC2-5 complex, and all of which still function as clamp loaders, but with distinct consequences

(**Figure 6.1** ; Lee et al. 2010; Majka & Burgers 2004; Overmeer et al. 2010; Leman & Noguchi 2013).

RFC (ie. RFC1 plus RFC2-5) acts to load onto DNA the clamp proliferating cell nuclear antigen (PCNA), which then acts as a processivity factor for eukaryotic DNA polymerase δ (pol δ) and ϵ (pol ϵ) both of which are essential for DNA replication (Uhlmann et al. 1997). In contrast, the complex containing Elg1/ATAD appears to perform the opposite function, the proper removal of PCNA and disassembly of replication factories in human cell lines (Lee et al. 2013) and in yeast. Reduced expression of Elg1/ATAD5 in human cells causes increased chromosomal rearrangements, increased rate of sister chromatid exchange, as well as sensitivity to DNA damaging agents (Sikdar et al. 2009). DNA damage responses by human ELG1 in S phase are important to maintain genomic integrity (Shkedy et al. 2015). Mutations in Elg1 in yeast result in elongated telomeres, suggesting that compromised unloading or recycling of PCNA delays lagging strand synthesis, in turn resulting in the exposure of single stranded DNA capable of being elongated by telomerase. The Rad24 containing RFC complex recruits RAD1, RAD9, and HUS1 (sometimes referred to as the 9-1-1 complex), which form a heterotrimeric complex that resembles PCNA, and acts as a checkpoint component to facilitate ATR- mediated phosphorylation and activation of Chk1 (Parrilla-Castellar et al. 2004). CTF18-RFC contains an additional subunit Dcc1 at least in yeast (Grabarczyk et al. 2018) and is required for chromosome cohesion (Mayer et al. 2001), although recent data has suggested that at least in yeast, this complex acts specifically to load PCNA onto the leading strand.

All of the observations above suggest a global model for a central role of RFC in DNA replication and genome integrity in which a central core of RFC components form specific complexes required for distinct circumstances, both normal and aberrant, that arise during DNA replication.

The data obtained in this study found that both RFC2 and 4 were significantly enriched in preparations of UPF1 isolated from chromatin, and this enrichment is

validated by immunoblotting in the case of RFC4. As discussed in Chapter 5 these preparations did not show enrichment for the RFC1 component, which somewhat surprisingly was found associated with Upf1^{ChrmB} derived from cytoplasmic extracts. A working hypothesis would be that RFC 2-5 is involved in loading Upf1, and potentially additional associated proteins, onto chromatin, in order to resolve aberrant nucleic acids structures that may be generated at either stalled or ongoing replication forks, as well as the special circumstances associated with telomere maintenance. The apparent absence of significant enrichment of the other components of the RFC2-5 complex may simply be a consequence of the nature of proteomic analysis, although it is also conceivable that either RFC2 and 4 act uniquely in concert with Upf1, or that the Upf1-RFC2-RFC4 represents an abundant trimeric intermediate in the assembly of a mature Upf1-RFC complex. Further work will obviously be required to test these ideas, and whether UPF1 interacts directly with the core RFC2-5 complex or requires additional polypeptides to do so. Given the existence of the ChrmB mutation, a reasonable prediction is that the N-terminus of UPF1 is important for any specific interaction, and transfection experiments using fragments of the RFC components are likely to provide useful information about the nature of that interaction initially in vitro, and subsequently in vivo.

Additional work will also be required in order to understand the significance of the interaction of Upf1^{ChrmB} with RFC1. In the model described above, UPF1 is proposed to be conceptually analogous to ELG1, Rad24 or Ctf18. However an alternative is that the canonical RFC1-5 complex directly recruits Upf1 to chromatin, and that the structural alterations induced by the S42A mutation in Upf1^{ChrmB} results in incomplete assembly of the hexameric complex.

The amino acid residue S42 is followed by a glutamine residue, thus forming a motif preferred by members of the PIKK family of protein kinases. It is currently unclear whether this residue undergoes phosphorylation in vivo, although previous work with another mutant, Upf1^{S42E}, indicated that this form of the protein was capable of binding

directly to chromatin. Future work will be directed towards establishing whether phosphorylation occurs at this site, and whether it might be a target for ATR, which has been shown previously to be required for efficient recruitment of Upf1 to chromatin. Should S42 turn out to be a target phosphorylation sites for ATR, a clear prediction would be that the extent of chromatin association of Upf1^{S42E} would be independent of the functionality of ATR. Additionally a proteomic comparison between Upf1^{S42E} and Upf1^{Wt} is expected to provide important insights into the differential binding of Upf1 interacting partners.

The data presented in Chapter 3 showed that Upf1^{ChrmB}, but not either of the other mutants investigated, acted in a dominant-negative fashion, prompting the subsequent investigation into the identification of components which might be sequestered by this form of the protein. A number of the proteins (**Chapter 5 , Table 4**) including RFC1, which were enriched in Upf1^{ChrmB} containing fractions might conceivably, if sequestered un-productively, give rise to circumstances where replication forks are compromised due to lack of available protein, resulting in an increase in levels of double strand breaks and thus elevated levels of γ H2AX.

As indicated in several studies, after performing clamp loading functions, RFC may remain engaged with DNA by interacting with replication protein A via some of its core subunits (Yuzhakov et al. 1999; Waga & Stillman 1998). If this is indeed true then RFC may also be a component that facilitates the interaction of Upf1 with the replication fork machinery. Consistent with this idea, previous studies by Carastro et al. (2002) and Azzalin & Lingner (2006b) reported that ectopically expressed human Upf1 co-immunoprecipitated with the endogenous pol δ subunit, p66. Turton (2014) also detected a weak association between Upf1 with p66 in immunoprecipitates from asynchronous HeLa cells nuclear extract by immunoblotting.

6.5.2 The significance of TOP2A (TOP2 α) and Upf1 interaction: regulation of telomere replication?

DNA topoisomerase has been classified into two categories. Type I topoisomerases mediate transient breakage of one DNA strand at a time, whilst Type II topoisomerases generate cuts in both DNA strands (Bermejo et al. 2007; Higgins 2012). There are two isoforms of DNA topoisomerase II (TOP2), α and β . Although their biochemical properties show no apparent significant difference (Akimitsu et al. 2003), Top2 α has mainly been implicated in DNA relaxation/decatenation and segregation during DNA replication, and protein expression of Top2a increases as cells progress from S-phase to M. In contrast, Top2 β is mostly associated with facilitating transcription, and its expression remains consistent throughout the cell cycle (de Campos-Nebel et al. 2010; Higgins 2012). In this study, type II α topoisomerase was found to be significantly enriched in Upf1^{ChrmB} preparations.

Bermejo et al. (2007) reported that during a study which carried out a parallel ChIP analysis on a subunit of the DNA polymerase ϵ together with Top2 in cells experiencing prolonged HU treatment following a G1 block indicated that Top2 binds in close proximity to the replication fork. A recent study had discovered that TOP2 α binds telomeres in a telomere repeats binding factor 1 (TRF1)-mediated manner to help regulate DNA replication at telomeres in vivo, relieving topological stress created by the advancement of the replication fork (Stagno et al. 2014). Additionally, Morere et al. (2008) had previously shown that TRF2 but not TRF1 preferentially binds to supercoiled DNA and is enriched at telomeres upon the loss of TOP2 α activity, most probably acting as a topological stress sensor. TRF1 and TRF2 are components of the 6-subunit shelterin complex (TRF1, TRF2, TIN2, RAP1, TPP1 and POT1) which helps protect chromosome ends, prevent the initiation of DNA damage responses arising from the presence and detection of a double-stranded DNA end, and regulate telomere length maintenance by telomerase (Palm & de Lange 2008; Takai et al. 2010). Upf1 has also been reported to associate with a component of the shelterin complex, TPP1 (Chawla et al. 2011). These findings, together with the work reported here suggest that both Upf1 and Top2 α

may associate either directly or indirectly, and thus may play a co-operative role in regulating replication at the telomere region.

However in this study, the interaction between Upf1 and Top2a was observed in the ChrmB mutant cells, a mutant which is unable to associate with chromatin, in the presence of the replication inhibitor hydroxyurea. Again it is conceivable that this interaction reflects a trapped intermediate, or an aberrant cellular response to the lack of Upf1 bound to chromatin. Interestingly, de Campos-Nebel et al. (2010) have reported that in the absence of Top2a a significant increase in γ H2AX expression was observed. Hence as Top2 α is believed to be involved in the regulation of telomere length, γ H2AX expression observed in Upf1^{ChrmB} expressing cells may conceivably be the consequence of Top2 α sequestration by Upf1^{ChrmB}, reducing the ability of the former to associate appropriately with chromatin

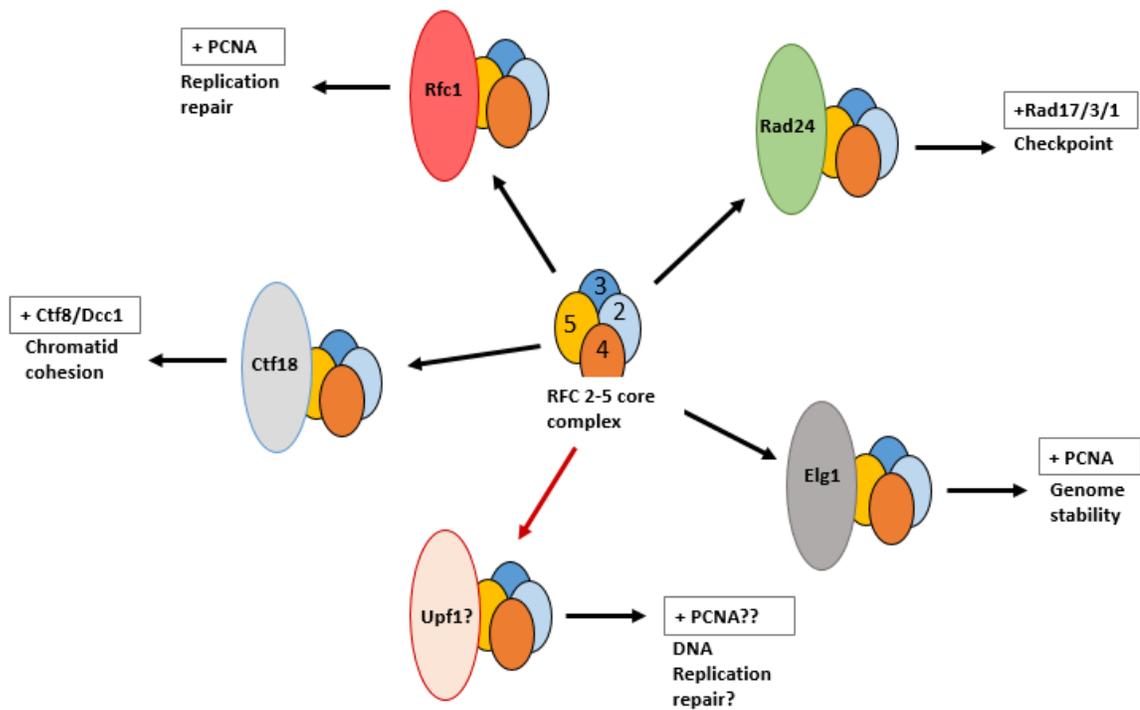


Figure 6.1: Alternative clamp loaders present in cells

Core RFC2-5 complex is able to form complexes with other proteins that functions in different pathways. Canonical clamp loader RFC, is made up by large subunit RFC1 and smaller RFC2-5 subunits. The larger subunit is interchangeable with RAD24, Ctf18 and Elg1 proteins and even though they go down different pathways, still act as clamp loaders. RFC2-5

6.6 Future perspectives

In this study I undertook an initial examination of the impact of phosphorylation site mutants of Upf1 on genome integrity as determined by measuring expression of γ H2AX following overexpression of the relevant mutants. Durand et al. (2016) showed that no single phosphorylation site is essential for Upf1's function but multiple sites cumulatively contribute to its activity with individual sites contributing to different effects. In future work, an experiment examining phosphorylation site mutants on genome stability could be performed using cells depleted of wild-type Upf1, and using multiple combinations of the phosphorylation states; for example a mutation of alanine at the N-terminal combined with a double glutamic acid mutation at the C-terminal (T28A, S1096E & S1116E) or any other combinations.

Although novel interactions between Upf1 with Top2 α and RFC subunits are demonstrated here, more information regarding their interactions and role in maintaining genomic integrity are needed. In addition, high-resolution co-localisation microscopy, CHIP and fluorescent in situ hybridization (FISH) could help to determine the co-localisation of identified interactors with Upf1 especially at telomeres to uncover more information regarding their role and significance in the maintenance of genome integrity. One of the significant challenges associated with studying an important function of a protein, the vast majority of which localises to a distinct cellular compartment is an inability to undertake definitive microscopy for minority population of molecules. In the future, this problem might be overcome by judicious use of proximity ligation assay (PLA). In this technique, antibodies recognising discrete proteins that are within 50 nm of each other, are each modified with one of a pair of unique oligonucleotides. These are capable of enabling rolling circle DNA amplification using fluorescent nucleotides, on microscopy slides, for subsequent high-resolution imaging by microscopy. This technique is being increasingly used for *in situ* detection of low-frequency and transient protein interactions, PTMs and spatial relationships of antigens in cells and tissues (Zatloukal et al. 2014). Several proteins that were identified

during MS analysis in this study, but were not sufficiently abundant to be detected by immunoblotting -such as WDHD1 - could be validated using this approach, since failure to detect by immunoblotting may have been due to overall low levels of protein interactor, or because the In vivo interaction is relatively transient.

The outcome of this study and that of Turton (2014) have helped to establish the significance of the ChrmB motif for UPF1 function. Although further research needs to be done to highlight the full picture of Upf1's role in maintaining genome stability, results from this study have given new insights of potential interactors that might pave new research direction.

Appendix 1: Sequence coverage of trypsin digestion for Upf1
(Isoform 2)

	10	20	30	40	50
MSVEAYGPSS	QTLTFLDTEE	AELLGADTQG	SEFEFTDFTL	PSQTQTTPGG	
60	70	80	90	<u>100</u>	
FGGPGGGGAG	GPGGAGAGAA	AGQLDAQVGP	EGILQNGAVD	DSVAKTSQLL	
<u>110</u>	<u>120</u>	<u>130</u>	<u>140</u>	<u>150</u>	
AELNFEDEE	DTYYTKDLPI	HACSYCGIHD	PACVVYCNTS	KRWFCNGRGN	
<u>160</u>	<u>170</u>	<u>180</u>	<u>190</u>	<u>200</u>	
TSGSHIVNHL	VRAKCKEVTL	HKDGPLGETV	LECYNCGCRN	VFLLGFIKAK	
210	220	230	240	250	
ADSVVLLCR	QPCASQSSLK	DINWDSSQWQ	PLIQDRCFLS	WLVKIPSEQE	
260	270	280	290	300	
<hr/>					
QLRARQITAQ	QINKLEELWK	ENFSATLEDL	EKPGVDEEPQ	HVLLRYEDAY	
310	320	330	340	350	
<hr/>					
QYQNIIFGLV	KLEADYDKKL	KESQTQDNIT	VRWDLGLNKK	RIAYFTLPKT	
360	370	380	390	400	
<hr/>					
DSGNEDLVII	WLRDMRLMQG	DEICLRYKGD	LAPLWKGIGH	VIKVPDNYGD	
410	420	430	440	450	
<hr/>					
EIAIELRSSV	GAPVEVTHNF	QVDFVWKSTS	FDRMQSALKT	FAVDETSVSG	
460	470	480	490	500	
<hr/>					
YIYHLLGHE	VEDVIIKQQL	PKRFTAQGLP	DLNHSQVYAV	KTVLQRPLSL	
510	520	530	540	550	
<hr/>					
IQGPPGTGKT	VTSATIVYHL	ARQGNPVLV	CAPSNIAVDQ	LTEKIHQTGL	
560	570	580	590	600	
<hr/>					
KVVRLCAKSR	EAIDSFVSFL	ALHNQIRNMD	SMPELQKLQQ	LKDETGELSS	
610	620	630	640	650	
<hr/>					
ADEKRYRALK	RTAERELLMN	ADVICTCTCVG	AGDPRLAKMQ	FRSILIDEST	
660	670	680	690	700	
<hr/>					
QATEPECMVP	VVLGAKQLIL	VGDHCQLGFP	VMCKKAAGAG	LSQSLFERLV	
710	720	730	740	750	
<hr/>					
VLGIRPIRLQ	VQYRMHPALS	AFPSNIFYEG	SLQNGVTAAD	RVKKGFDLQW	
760	770	780	790	800	
PQPDKPMFFY	VTQGQEEIAS	SGTSYLNRTS	AANVEKITTK	LLKAGAKPDQ	
810	820	830	840	850	
<hr/>					
IGIITPYEGO	RSYLVOYMOF	SGSLHTKLYQ	EVEIASVDAF	QGREKDFIIL	

<u>860</u>	<u>870</u>	<u>880</u>	<u>890</u>	<u>900</u>
SCVRANEHQG	IGFLNDPRRL	NVALTRARYG	VIIIVGNPKAL	SKQPLWNHLL
<u>910</u>	<u>920</u>	<u>930</u>	<u>940</u>	<u>950</u>
<u>NYYKEQKVLV</u>	<u>EGPLNNLRES</u>	<u>IMQFSKPRKL</u>	<u>VNTINPGARF</u>	<u>MTTAMYDARE</u>
<u>960</u>	<u>970</u>	<u>980</u>	<u>990</u>	<u>1000</u>
AIIPGSVYDR	SSQGRPSSEMY	FQTHDQIGMI	SAGPSHVAAM	NIPIPFNLVM
<u>1010</u>	<u>1020</u>	<u>1030</u>	<u>1040</u>	<u>1050</u>
PFMPPPGYFG	QANGPAAGRG	TFKGKTGRGG	RQKNRFG LPG	PSQTNLFNSQ
<u>1060</u>	<u>1070</u>	<u>1080</u>	<u>1090</u>	<u>1100</u>
ASQDVASQPF	SQGALTQGYI	SMSQPSQMSQ	PGLSQPELSQ	DSYLGDEFKS
<u>1110</u>	<u>1120</u>			
QIDVALSQDS	TYQGERAYQH	GGVTGLSQY		

Appendix 2: Comparison of protein interactors from different published lists with own MS data

NO	PROTEINS IDENTIFIED	UNIPROT Accession code	FLURY PAPER (SILAC)	SCHWEINGRUBER PAPER (BIOID)	UNIPROT BIOGRID (INTERACTORS)	MS DATA
1	Regulator of nonsense transcripts 2 (Upf2)	Q9HAU5	√	√	√	√
2	Regulator of nonsense transcripts 3B (Upf3B)	Q9BZI7	√	√	√	√
3	Protein CASC3 (MLN51)	O15234	√	√		
4	Eukaryotic initiation factor 4A-III (EIF4A3)	P38919	√	√	√	√
5	Eukaryotic translation initiation factor 3 subunit A (EIF3A)	Q14152	√		√	√
6	Eukaryotic translation initiation factor 3 subunit B (EIF3B)	P55884	√		√	√
7	Eukaryotic translation initiation factor 3 subunit C (EIF3C)	Q99613	√			
8	Eukaryotic translation initiation factor 3 subunit D (EIF3D)	O15371	√			
9	Eukaryotic translation initiation factor 3 subunit E (EIF3E)	P60228	√			
10	Eukaryotic translation initiation factor 3 subunit I (EIF3I)	Q13347	√			

11	Eukaryotic translation initiation factor 3 subunit L (EIF3L)	Q9Y262	√			
12	40S ribosomal protein S3 (RPS3)	P23396	√		√	√
13	40S ribosomal protein S3a (RPS3a)	P61247	√			
14	40S ribosomal protein S6 (RPS6)	P62753	√		√	√
15	40S ribosomal protein S2 (RPS2)	P15880	√	√		
16	60S ribosomal protein L3 (RPL3)	P39023	√			
17	60S ribosomal protein L4 (RPL4)	P36578	√			
18	60S ribosomal protein L5 (RPL5)	P46777	√			
19	60S ribosomal protein L8 (RPL8)	P62917	√			
20	Ataxin-2 (ATXN2)	Q99700	√	√		
21	ATP-dependent RNA helicase A (DHX9)	Q08211	√	√		
22	ATP-dependent RNA helicase DDX1	Q92499	√	√		
23	ATP-dependent RNA helicase DDX3X	O00571	√	√		
24	ATP-dependent RNA helicase eIF4A-1 (EIF4A1)	P60842	√			
25	Centrosomal protein 55 kDa (CEP55)	Q53EZ4	√	√		
26	DnaJ homolog subfamily A member 3,	Q96EY1	√			

	mitochondrial (DNAJA3)					
27	EH domain-containing protein 4 (BCOR)	Q9H233	√			
28	Heat shock 70 kDa protein 1A/1B	P08107	√			
29	Heat shock cognate 71 kDa protein (HSPA8)	P11142	√			
30	60 kDa heat shock protein, mitochondrial (HSPD1)	P10809	√			
31	78 kDa glucose-related protein (HSPA5)	P11021	√			
32	Interferon-inducible double stranded RNA- dependent protein kinase activator A (PRKRA)	O75569	√			
33	La-related protein 4B (LARP4B)	Q92615	√		√	√
34	Nuclear cap-binding protein subunit 1 (NCBP1)	Q09161	√		√	√
35	Nuclear fragile X mental retardation-interacting protein 2 (NUFIP2)	Q7Z417	√			
36	Protein FAM98A	Q8NCA5	√	√		
37	Putative RNA-binding protein 15 (RBM15)	Q96T37	√			
38	Retrotransposon- derived protein PEG10	Q86TG7	√			
39	Rho GTPase-activating protein 29 (ARHGAP29)	Q52LW3	√			

40	RNA-binding protein 14 (RBM14)	Q96PK6	√	√		
41	Serine/threonine- protein kinase D3 (PRKD3)	O94806	√			
42	Serine/threonine- protein kinase receptor- associated protein (UNR-interacting protein) (STRAP)	Q9Y3F4	√	√		
43	Solute carrier family 12 member 2 (SLC12A2)	P55011	√			
44	STAGA complex 65 subunit gamma (SUPT7L)	O94864	√			
45	Double-stranded RNA – binding protein Staufen 2 (STAU2)	Q9NUL3	√	√	√	√
46	T-complex protein 1 subunit gamma (CCT3)	P49368	√			
47	Thioredoxin reductase 1, cytoplasmic (TXNRD1)	Q16881	√			
48	tRNA-splicing ligase RtcB homolog (RTCB)	Q9Y3I0	√	√		
49	Tudor domain- containing protein 3 (TDRD3)	Q9H7E2	√			
50	Ubiquitin-associated protein 2 (UBAP2)	Q5T6F2	√			
51	Protein PRRC2C	Q9Y520		√		
52	Crk-like protein (CRKL)	P46109		√		

53	Eukaryotic translation initiation factor 4 gamma 2 (EIF4G2)	P78344		√		
54	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial (ACADM)	P11310		√		
55	Centrosomal protein of 131 kDa (CEP131)	Q9UPN4		√		
56	Calponin-3 (CNN3)	Q15417		√		
57	Adapter molecule crk	P46108		√		
58	DnaJ homolog subfamily B member 1 (DNAJB1)	P25685		√		
59	Leucine zipper protein 1 (LUZP1)	Q86V48		√		
60	Ribonucleoside-diphosphate reductase subunit M2 (RRM2)	P31350		√		
61	YTH domain-containing family protein 2 (YTHDF2)	Q9Y5A9		√		
62	ATP synthase subunit alpha, mitochondrial (ATP5A1)	P25705		√		
63	Staphylococcal nuclease domain-containing protein 1 (SND1)	Q7KZF4		√		
64	Coatomer subunit alpha (COPA)	P53621		√		
65	Serine/threonine-protein kinase SMG1	Q96Q15		√	√	√

66	Double-stranded RNA-binding protein Staufen homolog 1 (STAU1)	Q95793		√	√	√
67	Apoptosis-stimulating of p53 protein 2 (TP53BP2)	Q13625		√		
68	mRNA-decapping enzyme 1A (DCP1A)	Q9NPI6		√	√	
68	Segment polarity protein dishevelled homolog DVL-2	O14641		√		
70	Eukaryotic translation factor 4B (EIF4B)	P23588		√		
71	Protein PRRC2A	P48634		√		
72	Cytosolic carboxypeptidase (AGTPBP1)	Q9UPW5		√		
73	Ankyrin repeat and SAM (ANKS1A)	Q92625		√		
74	AP-3 complex subunit delta-1 (AP3D1)	O14617		√		
75	Ubiquitin-like-conjugating enzyme ATG3	Q9NT62		√		
76	Probable ATP-dependent RNA helicase DDX20	Q9UHI6		√		
77	Protein diaphanous homolog 3 (DIAPH3)	Q9NSV4		√		
78	GRIP and coiled-coil domain-containing protein 2 (GCC2)	Q8IWJ2		√		
79	Histone deacetylase (HDAC5)	Q9UQL6		√	√	

80	Kinesin-like protein KIF14 (KIF14)	Q15058		√		
81	Nuclear migration protein nudC (NUDC)	Q9Y266		√		
82	Perilipin-3 (PLIN3)	O60664		√		
83	Tyrosine-protein phosphatase non- receptor type 23 (PTPN23)	Q9H3S7		√		
84	Sperm-associated antigen 5 (SPAG5)	Q96R06		√		
85	AP2- associated protein kinase 1 (AAK1)	Q2M2I8		√		
86	Annexin A2 (ANXA2)	P07355		√		
87	T-complex protein 1 subunit gamma (CCT3)	P49368		√		
88	Elongation factor 1- gamma (EEF1G)	P26641		√		
89	Isoleucine—tRNA ligase, cytoplasmic (IARS)	P41252		√		
90	Kinesin-like protein KIF1B	O60333		√		
91	PDZ and LIM domain protein 5 (PDLIM5)	Q96HC4		√		
92	Pre-mRNA splicing factor RBM22	Q9NW64		√		
93	RNA polymerase II- associated protein 3 (RPAP3)	Q9H6T3		√		
94	Paired amphipathic helix protein Sin3b	O75182		√		

95	TATA element modulatory factor (TMF1)	P82094		√		
96	E3 ubiquitin-protein ligase TRAF7	Q6Q0C0		√		
97	Serum albumin (ALB)	P02768		√		
98	E3 ubiquitin-protein ligase CBL	P22681		√		
99	Dynactin subunit 1 (DCTN1)	Q14203		√	√	√
100	Segment polarity protein dishevelled homolog DVL-3	Q92997		√		
101	Uncharacterised protein FLJ45252	Q6ZSR9		√		
102	Melanoma-associated antigen D1 (MAGED1)	Q9Y5V3		√		
103	Methionine- tRNA ligase, cytoplasmic (MARS)	P56192		√		
104	MLLT protein	Q96C95		√		
105	Myomegalin (PDE4DIP)	Q5VU43		√		
106	ATP-dependent 6-phosphofructokinase, liver type (PFKL)	P17858		√		
107	ATP-dependent 6-phosphofructokinase, muscle type (PFKM)	P08237		√		
108	ATP-dependent 6-phosphofructokinase, platelet type (PFKP)	Q01813		√		
109	Pleckstrin homology domain-containing	Q9HAU0		√	√	√

	family A member 5 (PLEKHA5)					
110	E3 ubiquitin/ISG15 ligase TRIM25	Q14258		√		
111	Hamartin (TSC1)	Q92574		√		
112	60S ribosomal protein L23 (RPL23)	P62829		√		
113	Histone H1.4 (HIST1H1E)	P10412		√	√	√
114	40S ribosomal protein S11 (RPS11)	P62280		√		
115	40S ribosomal protein S16 (RPS16)	P62249		√		
116	60S ribosomal protein L17 (RPL17)	P18621		√		
117	Polyadenylate-binding protein 4 (PABPC4)	Q13310		√		
118	Polyadenylate-binding protein 1 (PABPC1)	P11940		√	√	√
119	ATP-dependent RNA helicase DDX39A	O00148		√		
120	Ras GTPase- activating protein-binding protein 2 (G3BP2)	Q9UN86		√		
121	Bcl-2-associated transcription factor 1 (BCLAF1)	Q9NYF8		√		
122	Ataxin-2-like protein (ATXN2L)	Q8WWM7		√		
123	Ras GTPase-activating protein-binding protein 1 (G3BP1)	Q13283		√		
124	60S ribosomal protein L10 (RPL10)	P27635		√		

125	Thyroid hormone receptor-associated protein 3 (THRAP3)	Q9Y2W1		√		
126	40S ribosomal protein S17 (RPS17)	P08708		√		
127	60S ribosomal protein L24 (RPL24)	P83731		√		
128	40S ribosomal protein S15a (RPS15A)	P62244		√		
129	Protein LSM12 homolog	Q3MHD2		√		
130	Protein mago nashi homolog (MAGOH)	P61326		√	√	
131	Serine/arginine-rich splicing factor 1 (SRSF1)	Q07955		√		
132	40S ribosomal protein S30 (FAU)	P62861		√		
133	60S ribosomal protein L27a (RPL27A)	P46776		√		
134	Cleavage and polyadenylation specificity factor subunit 1 (CPSF1)	Q10570		√		
135	Pre-mRNA processing factor 19 (PRPF19)	Q9UMS4		√		
136	60S ribosomal protein L35 (RPL35)	P42766		√		
137	Heterogeneous nuclear ribonucleoprotein H (HNRNPH1)	P31943		√		
138	Heterogeneous nuclear ribonucleoprotein K (HNRNPK)	P61978		√		

139	Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1)	Q9NZI8		√		
140	Insulin-like growth factor 2 mRNA-binding protein 3 (IGF2BP3)	O00425		√		
141	Tubulin beta-4B chain (TUBB4B)	P68371		√		
142	40S ribosomal protein S8 (RPS8)	P62241		√		
143	Guanine nucleotide-binding protein-like 3 (GNL3)	Q9BVP2		√	√	√
144	40S ribosomal protein S13 (RPS13)	P62277		√		
145	U5 small nuclear ribonucleoprotein 200 kDa helicase (SNRNP200)	O75643		√		
146	Histone H3.3 (H3F3A)	P84243		√		
147	40S ribosomal protein S9 (RPS9)	P46781		√		
148	Probable ATP-dependent RNA helicase DDX6	P26196		√		
149	Splicing factor 3B subunit 3 (SF3B3)	Q15393		√		
150	Interleukin enhancer-binding factor 3 (ILF3)	Q12906		√		
151	Eukaryotic peptide chain release factor GTP-binding subunit ERF3B (GSPT2)	Q8IYD1			√	

152	Histone RNA hair-pin binding protein (SLBP)	Q14493			√	
153	Eukaryotic peptide chain release factor GTP-binding subunit ERF3A (GSPT1)	P15170			√	√
154	mRNA-decapping enzyme subunit 2 (DCP2)	Q8GW31			√	
155	RNA binding protein 8A (RBM8A)	Q9Y5S9			√	√
156	Regulator of nonsense transcripts 3A (Upf3A)	Q9H1J1			√	
157	5'-3' exoribonuclease 1 (XRN1)	Q8IZH2			√	√
158	COP9 signalosome complex subunit 5 (COPS5)	Q92905			√	
159	Telomerase-binding protein EST1A (SMG6)	Q86US8			√	
160	Elongation factor 2 (EEF2)	P13639			√	√
161	Protein SMG7	Q92540			√	√
162	Protein SMG5	Q9UPR3			√	√
163	Double-stranded RNA-specific adenosine deaminase (ADAR)	P55265			√	√
164	Vacuolar protein sorting-associated protein 35 (VPS35)	Q96QK1			√	√
165	Hsp90 co-chaperone cdc37	Q16543			√	√
166	Protein MEMO1	Q9Y316			√	

167	Zinc finger CCCH domain-containing protein 3	Q8IXZ2			√	
168	Exosome complex component RRP4 (EXOSC2)	Q13868			√	√
169	Interleukin enhancer-binding factor 2 (ILF2)	Q12905			√	√
170	Histone H1.1 (HIST1H1A)	Q02539			√	
171	Proline-rich protein 11 (PRR11)	Q96HE9			√	
172	Transformer-2 protein homolog alpha (TRA2A)	Q13595			√	√
173	60S ribosomal protein L6 (RPL6)	Q02878			√	√
174	MKI67 FHA domain-interacting nucleolar phosphoprotein (NIFK)	Q9BYG3			√	
175	Splicing factor U2AF 65 kDa subunit (U2AF2)	P26368			√	√
176	Heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1)	P09651			√	√
177	CCR4-NOT transcription complex subunit 2 (CNOT2)	Q9NZN8			√	√
178	Serine/threonine-protein kinase TAO2 (TAOK2)	Q9UL54			√	
179	Ornithine decarboxylase antizyme 1 (OAZ1)	P54368			√	

180	Sortilin (SORT1)	Q99523			√	√
181	Atherin (SAMD1)	Q6SPF0			√	
182	ELAV-like protein (ELAVL1)	Q15717			√	√
183	Exportin-1 (XPO1)	Q14980			√	√
184	Small ubiquitin-related modifier 3 (SUMO3)	P55854			√	
185	Protein HIRA	P54198			√	
186	Centrosomal protein of 170 kDa (CEP170)	Q5SW79			√	√
187	SCL-interrupting locus protein (STIL)	Q15468			√	√
188	Cell division cycle 5-like protein (CDC5L)	Q99459			√	√
189	Ubiquitin-like protein 7 (UBL7)	Q96S82			√	
190	Probable ATP- dependent RNA helicase DDX5	P17844			√	√
191	High affinity nerve growth factor receptor (NTRK1)	P04629			√	
192	Gamma-interferon- inducible protein 16 (IFI16)	Q16666			√	
193	F-box/WD repeat- containing protein 11 (FBXW11)	Q9UKB1			√	
194	RNA-binding protein EWS (EWSR1)	Q01844			√	√
195	Centrosomal protein of 104 kDa (CEP104)	O60308			√	
196	Centrosomal protein of 19 kDa (CEP19)	Q96LK0			√	

197	Centrosomal protein of 162 kDa (CEP162)	Q5TB80			√	
198	Centriolin (CNTRL)	Q7Z7A1			√	
199	Ninein (NIN)	Q8N4C6			√	√
200	Protein fantom (RPGRIP1L)	Q68CZ1			√	
201	Sodium channel and clathrin linker 1 (SCLT1)	Q96NL6			√	
202	SNW domain-containing protein 1 (SNW1)	Q13573			√	√
203	DNA dC-dU-editing enzyme APOBEC-3D (APOBEC3D)	Q96AK3			√	
204	Polycomb group RING finger protein 1 (PCGF1)	Q9BSM1			√	
205	Polypyrimidine tract-binding protein 3 (PTBP3)	Q95758			√	
206	Zinc finger C3H1 domain-containing protein (ZFC3H1)	Q60293			√	√
207	Fibroblast growth factor 8 (FGF8)	P55075			√	
208	Zinc finger protein 576 (ZNF576)	Q9H609			√	
209	Ataxin-7 like protein 1 (ATXN7L1)	Q9ULK2			√	
210	Ensconsin (MAP7)	Q14244			√	√
211	RING finger protein 151 (RNF151)	Q2KHN1			√	
212	RNA-binding protein 47 (RBM47)	A0AV96			√	√

213	Ribosome production factor 1 (RPF1)	Q9H9Y2			√	
214	Epithelial splicing regulatory protein 1 (ESRP1)	Q6NXG1			√	
215	60S ribosomal protein L23a (RPL23A)	P62750			√	√
216	Heterogeneous nuclear ribonucleoprotein C-like 1 (HNRNPCL1)	O60812			√	
217	RNA –binding protein Raly (RALY)	Q9UKM9			√	√
218	RNA-binding motif protein, X-linked 2 (RBMX2)	Q9Y388			√	
219	Fibroblast growth factor 17 (FGF17)	O60258			√	
220	Ribosome biogenesis protein NSA2 homolog	O95478			√	
221	Suppressor of WWI4 1 homolog (PPAN)	Q9NQ55			√	
222	Developmental pluripotency-associated protein 4 (DPPA4)	Q7L190			√	
223	U1 small nuclear ribonucleoprotein 70 kDa (SNRNP70)	P08621			√	√
224	POU domain, class 5, transcription factor 1 (POU5F1)	Q01860			√	
225	WD repeat-containing protein 46 (WDR46)	O15213			√	
226	ELAV-like protein 2	Q12926			√	
227	GLTSCR2 protein	Q96CS0			√	√

228	Microprocessor complex subunit DGCR8	Q8WYQ5			√	
229	Heterogeneous nuclear ribonucleoprotein Q (SYNCRIP)	Q60506			√	√
230	Homeobox protein CDX-1	P47902			√	
231	Histone H1t (HIST1H1T)	P22492			√	
232	Zinc finger CCCH-type antiviral protein 1 (ZC3HAV1)	Q7Z2W4			√	√
233	PR domain zinc finger protein 5 (PRDM5)	Q9NQX1			√	
234	Zinc finger CCHC-type and RNA binding motif-containing protein 1 (ZCRB1)	Q8TBF4			√	
235	Spermatid perinuclear RNA-binding protein (STRBP)	Q96S19			√	√
236	Zinc finger CCH domain-containing protein 18 (ZC3H18)	Q86VM9			√	√
237	Transcription factor E4F1	Q66K89			√	
238	Histone H2AX (H2AFX)	P16104			√	
239	Nuclear RNA export factor 2 (NXF2)	Q9GZY0			√	
240	Tetratricopeptide repeat 4 (TTC4)	Q95801			√	
241	Influenza virus NS1A-binding protein (IVNS1ABP)	Q9Y6Y0			√	√

242	Transcription activator BRG1 (SMARCA4)	P51532			√	√
243	Vigilin (HDLBP)	Q00341			√	√
244	Enhancer of mRNA- decapping protein 4 (EDC4)	Q6P2E9			√	√
245	Eukaryotic peptide chain release factor subunit 1 (ETF1)	P62495			√	√
246	Hemoglobin subunit beta (HBB)	P68871			√	
247	Estrogen receptor (ESR1)	P03372			√	
248	Nuclear cap-binding protein subunit 2 (NCBP2)	P52298			√	√
249	60S ribosomal protein L7a (RPL7A)	P62424			√	√
250	60S ribosomal protein L11 (RPL11)	P62913			√	√
251	Protein SMG 9	Q9H0W8			√	√
252	Protein SMG 8	Q8ND04			√	√
253	Thioredoxin-related transmembrane protein 1 (TMX1)	Q9H3N1			√	√
254	Heterogeneous nuclear ribonucleoprotein U (HNRNPU)	Q00839			√	√
255	Small nuclear ribonucleoprotein- associated protein N (SNRPN)	P63162			√	
256	Casein kinase II subunit beta (CSNK2B)	P67870			√	√

257	Exosome complex component RRP41 (EXOSC4)	Q9NPD3			√	√
258	Poly(A)-specific ribonuclease PARN	Q95453			√	√
259	Exosome component 10 (EXOSC10)	Q01780			√	√
260	Serine/threonine-protein kinase ATR	Q13535			√	
261	DNA polymerase delta catalytic subunit (POLD1)	P28340			√	√
262	Protein NDRG1	Q92597			√	
263	RNA-binding protein with serine-rich domain 1 (RNPS1)	Q15287			√	√
264	NAD-dependent protein deacetylase sirtuin-7 (SIRT7)	Q9NRC8			√	
265	Tumor susceptibility gene 101 protein (TSG101)	Q99816			√	
266	Cullin-3 (CUL3)	Q13618			√	√
267	Cullin-5 (CUL5)	Q93034			√	
268	Cyclin-dependent kinase 2 (CDK2)	P24941			√	
269	Cullin-associated NEDD8-dissociated protein 1 (CAND1)	Q86VP6			√	√
270	Enhancer of mRNA-decapping protein 3 (EDC3)	Q96F86			√	√
271	Small nuclear ribonucleoprotein-	P14678			√	√

	associated proteins B (SNRPB)					
272	Heterogeneous nuclear ribonucleoprotein R (HNRNPR)	O43390			√	√
273	U6 snRNA-associated Sm-like protein LSm8	O95777			√	
274	TAR DNA-binding protein 43 (TARDBP)	Q13148			√	√
275	Protein lin-28 homolog A (LIN28A)	Q9H9Z2			√	
276	Putative helicase MOV- 10 (MOV10)	Q9HCE1			√	√
277	Ribonucleases P/MRP protein subunit POP1	Q99575			√	√
278	DNA replication ATP- dependent helicase/nuclease DNA2	P51530			√	
279	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 10 (NDUFB10)	O96000			√	√
280	Histone-lysine N- methyltransferase EZH2 (EZH2)	Q15910			√	
281	Polycomb protein SUZ12	Q15022			√	√
282	E3 ubiquitin-protein ligase RING2 (RNF2)	Q99496			√	
283	Polycomb complex protein BMI-1 (BMI1)	P35226			√	
284	ATp-binding cassette sub-family F member 3 (ABCF3)	Q9NUQ8			√	√

285	Protein diaphanous homolog 1 (DIAPH1)	O60610			√	√
286	Decapping and exoribonuclease protein (DXO)	O77932			√	
287	N-acetylglucosamine-1-phosphotransferase subunit gamma (GNPTG)	Q9UJJ9			√	
288	Vascular cell adhesion protein 1 (VCAM1)	P19320			√	
289	Integrin alpha-4 (ITGA4)	P13612			√	
290	Fibronectin (FN1)	P02751			√	
291	14-3-3 protein theta (YWHAQ)	P27348			√	√
292	Eukaryotic translation initiation factor 2 subunit 1 (EIF2S1)	P05198			√	√
293	Eukaryotic translation initiation factor 2 subunit 2 (EIF2S2)	P20042			√	√
294	RuvB-like 1 (RUVBL1)	Q9Y265			√	√
295	Regulation of nuclear pre-mRNA domain-containing protein 2 (RPRD2)	Q5VT52			√	√
296	RuvB-like 2 (RUVBL2)	Q9Y230			√	√
297	Pleckstrin homology domain-containing family B member 2 (PLEKHB2)	Q96CS7			√	
298	Glutamine-dependent NAD(+) synthetase (NADSYN1)	Q6IA69			√	

299	Acetyl-coenzyme A synthetase, cytoplasmic (ACSS2)	Q9NR19			√	
300	Protein ABHD16A	O95870			√	
301	Rhox homeobox family member 2 (RHOXF2)	Q9BQY4			√	
302	RNA-binding protein 3 (RBM3)	P98179			√	√
303	RNA-binding protein FUS	P35637			√	√

Appendix 3: List of significant proteins identified

	Signific.↑	-Log(P-value)	Difference	Protein IDs	Majority protein	Protein names	Gene names
1	+	1.877035504...	3.755585...	Q9Y6A5...	Q9Y6A5	Transfor...	TACC3
2	+	1.607583375...	-3.08957...	Q9Y697...	Q9Y697	Cysteine...	NFS1
3	+	2.892938503...	-1.94455...	Q9Y657	Q9Y657	Spindlin-1	SPIN1
4	+	1.664169017...	3.104119...	Q9Y4W6...	Q9Y4W6	AFG3-lik...	AFG3L2
5	+	1.755468161...	-2.63313...	Q9Y3F4...	Q9Y3F4	Serine-th...	STRAP
6	+	2.488037486...	-4.53901...	Q9Y3D9...	Q9Y3D9...	28S ribo...	MRPS23
7	+	1.759569164...	-3.93033...	Q9Y3B7...	Q9Y3B7	39S ribo...	MRPL11
8	+	2.162756951...	-3.39642...	Q9Y399...	Q9Y399	28S ribo...	MRPS2
9	+	1.961365131...	-5.18429...	Q9Y383...	Q9Y383...	Putative R...	LUC7L2...
10	+	1.866269726...	-3.46869...	Q9Y305...	Q9Y305...	Acyl-coe...	ACOT9
11	+	2.282772268...	-3.22183...	Q9Y2Q9...	Q9Y2Q9...	28S ribo...	MRPS28
12	+	1.709471855...	-4.69219...	Q9UNM6...	Q9UNM6...	26S prot...	PSMD13
13	+	1.540334109...	2.900797...	Q9ULW0	Q9ULW0	Targetin...	TPX2
14	+	1.627811061...	-2.65019...	Q9ULR0...	Q9ULR0...	Pre-mRN...	ISY1
15	+	3.567060118...	-3.96168...	Q9UKM9...	Q9UKM9...	RNA-bin...	RALY
16	+	1.031907845...	-4.23826...	Q9UHX1...	Q9UHX1...	Poly(U)-...	PUF60
17	+	2.356806802...	4.373818...	Q9UH99...	Q9UH99	SUN dom...	SUN2
18	+	2.736706069...	-4.83466...	Q9UG63...	Q9UG63	ATP-bind...	ABCF2
19	+	2.123910209...	-5.20380...	Q9UER7	Q9UER7	Death do...	DAXX
20	+	3.067156085...	-3.81033...	Q9P2R3...	Q9P2R3	Rabanky...	ANKFY1
21	+	2.420283598...	2.443056...	Q9NZN4	Q9NZN4	EH doma...	EHD2
22	+	2.124786056...	-4.20680...	Q9NZB2...	Q9NZB2...	Constitu...	FAM120A
23	+	1.610223425...	-4.63632...	Q9NYY8	Q9NYY8	FAST kin...	FASTKD2
24	+	2.834056452...	-6.03736...	Q9NX05...	Q9NX05...	Constitu...	FAM120C
25	+	1.418257751...	4.502027...	Q9NVI7	Q9NVI7	ATPase f...	ATAD3A
26	+	2.325402526...	-3.04385...	Q9NUL7	Q9NUL7	Probable...	DDX28
27	+	2.720551469...	-2.04574...	Q9NU22...	Q9NU22	Midasin	MDN1
28	+	1.837606617...	-4.47041...	Q9NSE4	Q9NSE4	Isoleucin...	IARS2
29	+	2.214651755...	-3.18486...	Q9NSC5...	Q9NSC5...	Homer p...	HOMER3
30	+	2.244266193...	-4.34465...	Q9NRG7...	Q9NRG7...	Epimera...	SDR39U1
31	+	2.225926170...	-3.51844...	Q9HCS7	Q9HCS7	Pre-mRN...	XAB2
32	+	1.554000076...	-2.76267...	Q9HCE1...	Q9HCE1...	Putative h...	MOV10

	Signific.↑	-Log(P-value)	Difference	Protein IDs	Majority protein	Protein names	Gene names
33	+	1.243284791...	-4.36652...	Q9HC36...	Q9HC36...	rRNA me...	RNMTL1
34	+	1.920313405...	-2.63401...	Q9H9Y6...	Q9H9Y6...	DNA-dire...	POLR1B
35	+	1.599242143...	-2.98520...	Q9H7Z7...	Q9H7Z7...	Prostagl...	PTGES2
36	+	1.925146541...	-2.52747...	Q9H2U1...	Q9H2U1...	ATP-dep...	DHX36
37	+	2.451872235...	2.665886...	Q9H2M9...	Q9H2M9	Rab3 GT...	RAB3GA...
38	+	1.409540884...	3.050047...	Q9H0D6	Q9H0D6	5-3 exor...	XRN2
39	+	1.061120071...	-3.94670...	Q9H078...	Q9H078...	Caseino...	CLPB
40	+	2.328279180...	-3.81956...	Q9BYK8	Q9BYK8	Helicase...	HELZ2
41	+	2.552794470...	-2.93951...	Q9BXP5...	Q9BXP5...	Serrate R...	SRRT
42	+	2.864869409...	-2.63714...	Q9BWF3...	Q9BWF3...	RNA-bin...	RBM4;R...
43	+	2.990629650...	4.360050...	Q9BUF5...	Q9BUF5...	Tubulin b...	TUBB6
44	+	1.747410459...	-3.56129...	Q9BTZ2...	Q9BTZ2	Dehydro...	DHRS4
45	+	1.441884116...	3.988514...	Q9BRS2...	Q9BRS2	Serine/th...	RIOK1
46	+	1.773584840...	-3.15651...	Q9BPW8...	Q9BPW8...	Protein N...	NIPSNA...
47	+	2.452430878...	-6.16952...	Q99714;...	Q99714;...	3-hydrox...	HSD17B...
48	+	1.960508202...	-3.60944...	Q99613;...	Q99613;...	Eukaryo...	EIF3C;E...
49	+	2.362536228...	-2.36426...	Q96ST3...	Q96ST3	Paired a...	SIN3A
50	+	2.047738470...	-3.17581...	Q96SN8...	Q96SN8...	CDK5 re...	CDK5RA...
51	+	2.571728497...	-2.48427...	Q96JM3...	Q96JM3	Chromos...	CHAMP1
52	+	1.802799721...	-2.33420...	Q96I24	Q96I24	Far upst...	FUBP3
53	+	1.944482688...	-3.04045...	Q96GM5...	Q96GM5...	SWI/SNF...	SMARCD...
54	+	2.035978389...	-5.04770...	Q96C36...	Q96C36...	Pyrroline...	PYCR2
55	+	1.480452842...	-2.57548...	Q96AE4...	Q96AE4...	Far upst...	FUBP1
56	+	1.817331224...	-5.05673...	Q969Z0;...	Q969Z0;...	Protein T...	TBRG4
57	+	1.973412489...	2.663689...	Q969V3...	Q969V3...	Nicalin	NCLN
58	+	2.699868707...	-2.28379...	Q969G3...	Q969G3...	SWI/SNF...	SMARCE...
59	+	1.391830420...	-4.72338...	Q92947;...	Q92947	Glutaryl-...	GCDH
60	+	1.701685470...	-3.17033...	Q92805;...	Q92805	Golgin s...	GOLGA1
61	+	2.227977780...	2.066427...	Q92598;...	Q92598;...	Heat sho...	HSPH1
62	+	1.797553865...	2.479708...	Q8WXI9	Q8WXI9	Transcrip...	GATAD2B
63	+	2.193320206...	-2.74867...	Q8WWY...	Q8WWY...	U4/U6 sm...	PRPF31
64	+	2.366821244...	-3.63242...	Q8TEQ6	Q8TEQ6	Gem-ass...	GEMIN5
65	+	2.376073045...	-4.98390...	Q8NBN7...	Q8NBN7	Retinol d...	RDH13

	Signific.↑	-Log(P-value)	Difference	Protein IDs	Majority protein	Protein names	Gene names
66	+	3.159867448...	-3.52238...	Q8IX12;...	Q8IX12;...	Cell divis...	CCAR1
67	+	1.836725342...	-2.94115...	Q8IWI9;...	Q8IWI9;...	MAX gen...	MGA
68	+	4.615982171...	-1.91097...	Q7Z2T5	Q7Z2T5	TRMT1-...	TRMT1L
69	+	1.295557369...	-3.05165...	Q7LBC6	Q7LBC6	Lysine-s...	KDM3B
70	+	1.898896627...	-3.07176...	Q7L2J0	Q7L2J0	7SK snR...	MEPCE
71	+	2.362028288...	-5.47844...	Q7L2H7...	Q7L2H7...	Eukaryo...	EIF3M
72	+	1.524353853...	-2.72873...	Q7L0Y3...	Q7L0Y3	Mitochon...	TRMT10C
73	+	1.474489530...	-2.90499...	Q6UVJ0	Q6UVJ0	Spindle a...	SASS6
74	+	1.636496943...	2.942221...	Q6P996...	Q6P996...	Pyridoxa...	PDXDC1
75	+	3.637100304...	-3.86259...	Q6P2E9...	Q6P2E9	Enhance...	EDC4
76	+	1.520877277...	-2.73067...	Q68CP9...	Q68CP9...	AT-rich i...	ARID2
77	+	3.787985945...	-3.90408...	Q5W0B1	Q5W0B1	RING fin...	RNF219
78	+	2.497905729...	2.252964...	R4GNB2...	R4GNB2...	DENN d...	DENND4...
79	+	1.393830759...	-3.59009...	Q5TDF0...	Q5TDF0...	Cancer-r...	NTPCR
80	+	1.409811353...	2.810204...	Q5T6F2...	Q5T6F2...	Ubiquitin...	UBAP2
81	+	2.939761862...	4.048751...	Q5T4S7...	Q5T4S7	E3 ubiqu...	UBR4
82	+	1.761247118...	-3.16095...	Q5LJA9;...	Q5LJA9;...	Ubiquitin...	UHL5
83	+	1.320736663...	-3.44339...	Q5JSZ5...	Q5JSZ5	Protein P...	PRRC2B
84	+	3.968383636...	-2.58634...	Q5BKZ1...	Q5BKZ1	DBIRD c...	ZNF326
85	+	2.483599379...	-3.56812...	Q4G0J3...	Q4G0J3...	La-relate...	LARP7
86	+	2.685576038...	-2.75556...	Q1KMD3...	Q1KMD3...	Heteroge...	HNRNPU...
87	+	1.788684139...	5.465847...	Q15750	Q15750	TGF-bet...	TAB1
88	+	3.423336738...	-2.92847...	Q15717;...	Q15717;...	ELAV-lik...	ELAVL1
89	+	2.819479079...	3.473620...	Q15645;...	Q15645	Pachyten...	TRIP13
90	+	3.016767049...	-2.80103...	Q15459;...	Q15459	Splicing f...	SF3A1
91	+	1.507244776...	-4.17878...	Q15393;...	Q15393	Splicing f...	SF3B3
92	+	1.780601700...	-6.21213...	Q15365	Q15365	Poly(rC)...	PCBP1
93	+	1.896824844...	4.437393...	Q14697;...	Q14697;...	Neutral a...	GANAB
94	+	2.176226511...	-2.47672...	Q14692	Q14692	Ribosom...	BMS1
95	+	1.616784714...	-3.16098...	Q14566	Q14566	DNA rep...	MCM6
96	+	2.096669515...	2.568626...	Q14432	Q14432	cGMP-in...	PDE3A
97	+	2.617659145...	5.534034...	Q14247;...	Q14247	Src subs...	CTTN
98	+	1.789972297...	-2.83361...	Q14152	Q14152	Eukaryo...	EIF3A

	Signific.↑	-Log(P-value)	Difference	Protein IDs	Majority protein	Protein names	Gene names
99	+	1.574278274...	-4.32859...	Q13895;...	Q13895;...	Bystin	BYSL
100	+	1.374183841...	-3.91853...	Q13838;...	Q13838;...	Spliceos...	DDX39B
101	+	1.918054439...	-3.20578...	Q13813	Q13813	Spectrin...	SPTAN1
102	+	2.310299357...	2.469503...	Q13464;...	Q13464	Rho-ass...	ROCK1
103	+	2.181059404...	-3.53839...	Q13409	Q13409	Cytoplas...	DYNC112
104	+	2.495319274...	-3.34523...	Q13243;...	Q13243	Serine/a...	SRSF5
105	+	3.193606874...	-2.71623...	Q13151	Q13151	Heteroge...	HNRNPA...
106	+	1.918717119...	2.605533...	Q13085;...	Q13085;...	Acetyl-C...	ACACA
107	+	1.639971036...	2.531583...	Q12965;...	Q12965	Unconve...	MYO1E
108	+	1.100673998...	-3.92089...	Q12931;...	Q12931;...	Heat sho...	TRAP1
109	+	2.127257292...	-2.66506...	Q12874	Q12874	Splicing f...	SF3A3
110	+	1.417010148...	-3.03478...	Q09161;...	Q09161	Nuclear c...	NCBP1
111	+	1.609346254...	-2.57586...	Q07666;...	Q07666	KH doma...	KHDRBS1
112	+	2.891574339...	3.080054...	Q04917;...	Q04917	14-3-3 p...	YWHAH
113	+	2.060998605...	-2.91751...	Q02878;...	Q02878	60S ribo...	RPL6
114	+	2.467950975...	4.698867...	Q01813;...	Q01813	ATP-dep...	PFKP
115	+	1.395755396...	-3.81977...	Q01085;...	Q01085;...	Nucleoly...	TIAL1
116	+	2.194081061...	-3.63523...	Q01082;...	Q01082;...	Spectrin...	SPTBN1
117	+	2.040315472...	3.469167...	P84090;...	P84090;...	Enhance...	ERH
118	+	2.277312811...	-3.14647...	P82933	P82933	28S ribo...	MRPS9
119	+	1.774866691...	-2.45468...	P82912;...	P82912	28S ribo...	MRPS11
120	+	2.374776257...	-3.43173...	P78347;...	P78347	General t...	GTF2I
121	+	1.446220489...	5.901783...	P68371;...	P68371	Tubulin b...	TUBB4B
122	+	1.281245323...	3.625636...	P68366;...	P68366	Tubulin a...	TUBA4A
123	+	1.544576955...	3.586419...	P68363;...	P68363;...	Tubulin a...	TUBA1B...
124	+	1.322410556...	-3.08024...	P68032;...	P68032;...	Actin, alp...	ACTC1;A...
125	+	1.715530363...	3.514684...	P67936	P67936	Tropomy...	TPM4
126	+	2.178949200...	2.384538...	P63151;...	P63151	Serine/th...	PPP2R2A
127	+	4.666108732...	4.414793...	P63104;...	P63104;...	14-3-3 p...	YWHAZ
128	+	2.755907708...	3.651889...	P62258;...	P62258;...	14-3-3 p...	YWHAE
129	+	2.138089932...	-3.05957...	P62249;...	P62249;...	40S ribo...	RPS16;Z...
130	+	2.093946318...	-2.51635...	P61978;...	P61978;...	Heteroge...	HNRNPK
131	+	2.422479777...	-3.38087...	P63261;...	P63261;...	Actin, cy...	ACTG1;A...

	Signific.↑	-Log(P-value)	Difference	Protein IDs	Majority protein	Protein names	Gene names
132	+	2.071430413...	-4.51102...	P60228;...	P60228;...	Eukaryo...	EIF3E
133	+	3.912196207...	6.296787...	P58107	P58107	Epiplakin	EPPK1
134	+	2.951363260...	-2.63378...	P55795	P55795	Heteroge...	HNRNPH...
135	+	2.360452329...	-6.12775...	P55084;...	P55084;...	Trifunctio...	HADHB
136	+	1.198000905...	-3.59737...	P54886	P54886	Delta-1-p...	ALDH18...
137	+	2.387838277...	3.005594...	P54105;...	P54105;...	Methylos...	CLNS1A
138	+	1.217797973...	3.369752...	P53621	P53621	Coatome...	COPA
139	+	2.039681873...	-2.86156...	P51991;...	P51991	Heteroge...	HNRNPA...
140	+	1.681335996...	-4.92500...	P51665;...	P51665;...	26S prot...	PSMD7
141	+	1.126640335...	-3.68657...	P51659;...	P51659;...	Peroxiso...	HSD17B4
142	+	1.593411237...	-4.64575...	P51553;...	P51553;...	Isocitrate...	IDH3G
143	+	2.864578311...	-2.36752...	Q9HBD4...	Q9HBD4...	Transcrip...	SMARCA...
144	+	3.744465570...	-4.90391...	P51398;...	P51398;...	28S ribo...	DAP3
145	+	1.420873877...	-3.54349...	P51149;...	P51149;...	Ras-rela...	RAB7A
146	+	1.138436502...	-5.16132...	P50213;...	P50213;...	Isocitrate...	IDH3A
147	+	2.133179092...	-2.18805...	P49792;...	P49792	E3 SUM...	RANBP2
148	+	3.229956982...	-8.65314...	P49748;...	P49748	Very lon...	ACADVL
149	+	1.913615655...	3.222403...	P49588;...	P49588;...	Alanine-...	AARS
150	+	3.503221397...	3.528795...	P49327	P49327	Fatty aci...	FASN
151	+	2.231092541...	2.744192...	P49321;...	P49321;...	Nuclear a...	NASP
152	+	3.180141838...	-4.25680...	P46940;...	P46940;...	Ras GTP...	IQGAP1
153	+	1.480746803...	-2.61829...	P46939;...	P46939	Utrophin	UTRN
154	+	2.610240981...	-2.80426...	P46777;...	P46777	60S ribo...	RPL5
155	+	2.268094719...	-4.86723...	P46379;...	P46379	Large pr...	BAG6
156	+	2.110002911...	-2.98668...	P42704;...	P42704	Leucine-...	LRPPRC
157	+	2.216043888...	2.694225...	P41250;...	P41250	Glycine-...	GARS
158	+	1.717693835...	-3.28879...	P40939;...	P40939;...	Trifunctio...	HADHA
159	+	2.723850654...	-2.32664...	P39019;...	P39019;...	40S ribo...	RPS19
160	+	1.147968790...	-3.95136...	P38117;...	P38117	Electron...	ETFB
161	+	1.448029537...	-2.75823...	P35998;...	P35998	26S prot...	PSMC2
162	+	4.659568869...	-3.90939...	P35659;...	P35659;...	Protein D...	DEK
163	+	2.273474149...	-2.17978...	P35658;...	P35658;...	Nuclear p...	NUP214
164	+	2.149932747...	5.606121...	P35579;...	P35579	Myosin-9	MYH9

	Signific.↑	-Log(P-value)	Difference	Protein IDs	Majority protein	Protein names	Gene names
165	+	1.402936369...	2.788565...	P35251;...	P35251	Replicati...	RFC1
166	+	1.651895045...	-3.80720...	P35250;...	P35250	Replicati...	RFC2
167	+	2.187603287...	-2.27513...	P32969;...	P32969;...	60S ribo...	RPL9
168	+	1.560997439...	-3.93141...	P32322;...	P32322;...	Pyrroline...	PYCR1
169	+	3.733644048...	-4.31161...	P31942	P31942	Heteroge...	HNRNPH...
170	+	1.455237840...	-3.35803...	P31930	P31930	Cytochro...	UQCRC1
171	+	1.915326626...	4.061540...	P30041	P30041	Peroxi...	PRDX6
172	+	1.917743394...	-2.95325...	P26599;...	P26599;...	Polypyrim...	PTBP1
173	+	1.116907063...	-4.46019...	P26196;...	P26196	Probable...	DDX6
174	+	1.712193508...	-4.49624...	P26038;...	P26038	Moesin	MSN
175	+	1.609504957...	-2.85815...	P25685;...	P25685;...	DnaJ ho...	DNAJB1
176	+	1.957313500...	-2.97529...	P23396;...	P23396;...	40S ribo...	RPS3
177	+	1.516290119...	4.275276...	P23381;...	P23381	Tryptoph...	WARS
178	+	2.366968426...	-2.59423...	P22626;...	P22626;...	Heteroge...	HNRNPA...
179	+	1.981012875...	-2.21317...	P22087;...	P22087;...	rRNA 2-O...	FBL
180	+	1.949895516...	-2.94608...	P20340;...	P20340;...	Ras-rela...	RAB6A
181	+	2.135712374...	5.337679...	P19474;...	P19474	E3 ubiqu...	TRIM21
182	+	1.563837990...	-2.71716...	P18615;...	P18615;...	Negative...	NELFE
183	+	1.572415591...	-3.06219...	P15880;...	P15880;...	40S ribo...	RPS2
184	+	1.819507648...	-2.27668...	P14866;...	P14866;...	Heteroge...	HNRNPL
185	+	1.990751155...	4.547108...	P14324;...	P14324;...	Farnesyl...	FDPS
186	+	1.395093213...	-3.23751...	P13861;...	P13861;...	cAMP-de...	PRKAR2A
187	+	1.387872981...	-3.84322...	P13804;...	P13804;...	Electron...	ETFA
188	+	2.703169192...	-2.82641...	P12270;...	P12270	Nucleop...	TPR
189	+	2.257641150...	-3.48276...	P11498;...	P11498	Pyruvate...	PC
190	+	1.225520834...	3.927849...	P11388;...	P11388	DNA top...	TOP2A
191	+	2.497141233...	-2.95399...	P10809;...	P10809	60 kDa h...	HSPD1
192	+	1.787973869...	2.970174...	P09874	P09874	Poly [AD...	PARP1
193	+	1.532246761...	-2.97542...	P09661;...	P09661;...	U2 smal...	SNRPA1
194	+	2.272478581...	-2.67496...	P09012;...	P09012;...	U1 smal...	SNRPA
195	+	1.617256632...	-2.94917...	P0CW22...	P0CW22...	40S ribo...	RPS17L...
196	+	2.146084890...	-5.03764...	P08670;...	P08670;...	Vimentin	VIM

	Signific.↑	-Log(P-value)	Difference	Protein IDs	Majority protein	Protein names	Gene names
197	+	1.734364191...	-5.39341...	P08621;...	P08621	U1 smal...	SNRNP70
198	+	1.040868436...	4.601729...	P08238;...	P08238	Heat sho...	HSP90A...
199	+	1.365025918...	3.621362...	P07900;...	P07900	Heat sho...	HSP90A...
200	+	1.735687870...	3.822292...	Q5JP53;...	Q5JP53;...	Tubulin b...	TUBB
201	+	1.429049627...	3.024677...	P06737;...	P06737;...	Glycoge...	PYGL
202	+	2.385556464...	3.893431...	P06733;...	P06733	Alpha-en...	ENO1
203	+	3.077762838...	-3.57014...	P05186;...	P05186	Alkaline p...	ALPL
204	+	1.143808877...	3.878394...	P04792;...	P04792;...	Heat sho...	HSPB1
205	+	2.106779520...	-2.78102...	P02545;...	P02545;...	Prelamin...	LMNA
206	+	1.809699476...	-3.76392...	P00387;...	P00387;...	NADH-c...	CYB5R3
207	+	1.716362330...	4.008316...	P00338;...	P00338;...	L-lactate...	LDHA
208	+	1.965732442...	-2.15973...	O96019;...	O96019	Actin-like...	ACTL6A
209	+	1.799255094...	-2.77564...	O95782;...	O95782	AP-2 com...	AP2A1
210	+	2.852352292...	3.521046...	O95714	O95714	E3 ubiqu...	HERC2
211	+	2.361729068...	-2.10661...	O95602;...	O95602;...	DNA-dire...	POLR1A
212	+	1.403628673...	3.362051...	O95373;...	O95373	Importin-7	IPO7
213	+	2.044944607...	-3.74301...	O95071;...	O95071;...	E3 ubiqu...	UBR5
214	+	1.713062011...	-3.73895...	O94992;...	O94992	Protein H...	HEXIM1
215	+	1.161265354...	-4.24612...	O94925;...	O94925	Glutamin...	GLS
216	+	2.813368263...	-4.95609...	O94915;...	O94915	Protein f...	FRYL
217	+	1.884104285...	-3.20802...	O76031;...	O76031	ATP-dep...	CLPX
218	+	1.421611844...	-4.21836...	O75822;...	O75822	Eukaryo...	EIF3J
219	+	2.260535564...	-4.91654...	O75717;...	O75717	WD repe...	WDHD1
220	+	3.656395119...	7.702748...	O75688;...	O75688;...	Protein p...	PPM1B
221	+	1.325388933...	-3.22781...	O75569;...	O75569;...	Interfero...	PRKRA
222	+	2.014120387...	2.709432...	O75477	O75477	Erlin-1	ERLIN1
223	+	2.088273437...	-2.03860...	O75146;...	O75146	Huntingt...	HIP1R
224	+	1.593141134...	-2.69609...	U3KQK0...	U3KQK0...	Histone H...	HIST1H2...
225	+	1.337658666...	-4.86882...	O60701;...	O60701	UDP-glu...	UGDH
226	+	2.304212973...	-2.78026...	O60506;...	O60506	Heteroge...	SYNCRIP
227	+	2.529576440...	2.859638...	O60502;...	O60502;...	Protein O...	MGEA5
228	+	2.539990531...	-3.72584...	O60341;...	O60341	Lysine-s...	KDM1A
229	+	2.028484023...	-2.46151...	O60306;...	O60306	Intron-bi...	AQR

	Signific.↑	-Log(P-value)	Difference	Protein IDs	Majority protein	Protein names	Gene names
230	+	2.851515732...	-2.67349...	O43896;...	O43896	Kinesin-I...	KIF1C
231	+	2.117434152...	-3.39911...	O43823;...	O43823	A-kinase...	AKAP8
232	+	2.800446417...	2.088748...	O43707;...	O43707;...	Alpha-ac...	ACTN4
233	+	2.163914970...	-3.58093...	O43615;...	O43615;...	Mitochon...	TIMM44
234	+	2.437476293...	-3.18358...	O43390;...	O43390;...	Heteroge...	HNRNPR
235	+	1.169425384...	3.595551...	O43318;...	O43318;...	Mitogen-...	MAP3K7...
236	+	2.994832045...	-4.45276...	O15020;...	O15020	Spectrin...	SPTBN2
237	+	2.101670098...	-4.37668...	O00487;...	O00487	26S prot...	PSMD14
238	+	1.741714718...	-3.24312...	O00425;...	O00425	Insulin-li...	IGF2BP3
239	+	2.060814600...	-4.09202...	O00303;...	O00303	Eukaryo...	EIF3F
240	+	1.822660609...	-4.72071...	O00232;...	O00232	26S prot...	PSMD12
241	+	1.667466011...	-5.90442...	O00231;...	O00231;...	26S prot...	PSMD11
242	+	2.119008866...	-2.82451...	M0R2Z9...	M0R2Z9...	SURP an...	SUGP2
243	+	1.277807008...	-3.35727...	M0R208...	M0R208...	ATP-dep...	CLPP
244	+	1.623926802...	-2.98923...	M0QZN2...	M0QZN2...	40S ribo...	RPS5
245	+	2.941059802...	-4.81374...	K7ERF1...	K7ERF1...	Eukaryo...	EIF3K
246	+	1.632963466...	-5.60363...	K7EKE6...	K7EKE6...	Lon prot...	LONP1
247	+	2.716609364...	-4.09441...	K7EJ78;...	K7EJ78;...	40S ribo...	RPS15
248	+	2.334078832...	-2.85789...	J3QLS3;...	J3QLS3;...	28S ribo...	MRPS7
249	+	1.845643690...	-2.26996...	J3KQ37;...	J3KQ37;...	RANBP2...	RGPD8
250	+	2.060786135...	-3.00748...	J3QL05;...	J3QL05;...	Serine/a...	SRSF2
251	+	1.440957873...	-3.48301...	I3L2C7;P...	I3L2C7;P...	Gem-ass...	GEMIN4
252	+	1.741306163...	-3.19624...	I3L0N3;P...	I3L0N3;P...	Vesicle-f...	NSF
253	+	1.604677623...	-2.91061...	H7BXY3...	H7BXY3...	Putative A...	DHX30
254	+	2.848204584...	-2.33927...	H0YA96...	H0YA96...	Heteroge...	HNRNPD
255	+	2.094231251...	-2.85861...	G8JLB6;...	G8JLB6;...	Heteroge...	HNRNPH...
256	+	1.784076857...	-4.44305...	G5EA30...	G5EA30...	CUGBP E...	CELF1
257	+	4.121315384...	-2.78860...	G5E9W7...	G5E9W7...	28S ribo...	MRPS22
258	+	2.888968282...	-3.37991...	G5E975...	G5E975...	SWI/SNF...	SMARCB...
259	+	2.137278695...	-2.29617...	G3XAC6...	G3XAC6...	RNA-bin...	RBM39
260	+	2.008149272...	-2.45834...	G3V203...	G3V203...	60S ribo...	RPL18
261	+	1.730743813...	-5.31576...	G3V1C3...	G3V1C3...	Apoptos...	API5
262	+	1.497836193...	2.578734...	G3V1D1...	G3V1D1...	Ferritin;F...	FTH1

	Signific.↑	-Log(P-value)	Difference	Protein IDs	Majority protein	Protein names	Gene names
263	+	1.982044947...	-4.93451...	F8WJN3...	F8WJN3...	Cleavag...	CPSF6
264	+	1.579825634...	3.086867...	F8W9S7...	F8W9S7...	GTPase...	GAPVD1
265	+	1.875196186...	-4.77440...	F8VQE1...	F8VQE1...	LIM dom...	LIMA1
266	+	2.771245122...	-3.25907...	F5H2X7...	F5H2X7...	Cip1-inte...	CIZ1
267	+	1.225998559...	-3.61520...	E9PLD0...	E9PLD0...	Ras-rela...	RAB1B;R...
268	+	2.479744537...	2.825022...	E9PG15...	E9PG15...	14-3-3 p...	YWHAQ
269	+	1.127810167...	-3.55328...	F8VZJ2;...	F8VZJ2;...	Nascent...	NACA
270	+	1.896723698...	3.432910...	E7EVA0...	E7EVA0...	Microtub...	MAP4
271	+	1.703507367...	-2.73821...	E5RIZ4;...	E5RIZ4;...	39S ribo...	MRPL15
272	+	2.309653134...	-3.24565...	D6RH20...	D6RH20...	28S ribo...	MRPS27
273	+	1.361439621...	-3.21152...	C9JZI1;P...	C9JZI1;P...	Replicati...	RFC4
274	+	2.195651157...	-4.19390...	C9JVN9...	C9JVN9...	L-2-hydr...	L2HGDH
275	+	2.282898212...	-2.46676...	C9JXB8...	C9JXB8...	60S ribo...	RPL24
276	+	1.822933891...	-2.69619...	C9JG87...	C9JG87...	39S ribo...	MRPL39
277	+	2.341779376...	-3.86857...	J3KMX2...	J3KMX2...	SWI/SNF...	SMARCD...
278	+	1.196679808...	-4.62595...	B8ZC5...	B8ZC5...		GLS
279	+	2.099685681...	-4.29782...	B7Z9I1;Q...	B7Z9I1;Q...	Medium-...	ACADM
280	+	3.001585679...	5.433712...	F8W1R7...	F8W1R7...	Myosin I...	MYL6
281	+	2.994291916...	-5.11124...	B7WPG3...	B7WPG3...	Heteroge...	HNRNPL...
282	+	1.683566484...	-2.83947...	B4DY09...	B4DY09...	Interleuk...	ILF2
283	+	2.078104631...	-2.92055...	J3QS39;...	J3QS39;...	Ubiquitin...	UBB;RP...
284	+	1.776792160...	-4.93892...	B3KS98...	B3KS98...	Eukaryo...	EIF3H
285	+	2.551329585...	3.876916...	B3KNJ4...	B3KNJ4...	SUMO-a...	SAE1
286	+	3.054194400...	-3.11134...	B2R5W2...	B2R5W2...		HNRNPC
287	+	2.083844472...	-3.93297...	B0QY89...	B0QY89...	Eukaryo...	EIF3L
288	+	1.848994244...	-3.93948...	A3KN83	A3KN83	Protein s...	SBNO1
289	+	2.148826680...	-3.23433...	A0A0D9...	A0A0D9...	Ran-bind...	RANBP10
290	+	2.700832327...	-3.87938...	A0A0D9...	A0A0D9...	Serine/a...	SRSF4
291	+	1.991555650...	-2.75714...	A0A0C4...	A0A0C4...	Enoyl-Co...	ECI2
292	+	1.572601541...	-4.94254...	A0A0C4...	A0A0C4...	Probable...	DDX46
293	+	2.027200185...	-4.72523...	C9J9K3;...	C9J9K3;...	40S ribo...	RPSA;R...
294	+	2.779714446...	3.496110...	A0A0B4...	A0A0B4...	Tubulin b...	TUBB3
295	+	2.218159123...	-2.51885...	A0A0B4...	A0A0B4...	Serine/a...	SRSF7

	Signific.↑	-Log(P-value)	Difference	Protein IDs	Majority protein	Protein names	Gene names
296	+	1.493722106...	-3.19459...	E9PH62...	E9PH62...	Double-s...	STAU2
297	+	2.715498826...	-2.66225...	A0A0A0...	A0A0A0...	Eukaryo...	EIF4G3
298	+	1.358482525...	-3.84293...	A0A0A0...	A0A0A0...	A-kinase...	AKAP9
299	+	1.436776360...	-3.13016...	A0A087X...	A0A087X...	26S prot...	PSMC6
300	+	1.520357695...	-4.36467...	A0A087X...	A0A087X...	AP-2 com...	AP2B1
301	+	1.345705162...	-2.91081...	A0A087X...	A0A087X...	Double-s...	STAU1
302	+	2.899286091...	-3.80767...	A0A087X...	A0A087X...	Kanadap...	SLC4A1...
303	+	2.831003552...	-2.85699...	D6R9P3...	D6R9P3...	Heteroge...	HNRNPA...
304	+	1.858305209...	-5.30947...	A0A087W...	A0A087W...	Isocitrate...	IDH3B
305	+	1.210975085...	3.445697...	A0A087W...	A0A087W...	Ensconsin	MAP7
306	+	1.424375333...	-5.40314...	A0A087W...	A0A087W...	ATP-dep...	DHX29
307	+	1.979265517...	-4.51611...	A0A087W...	A0A087W...	TAR DN...	TARDBP
308	+	1.821567960...	4.111637...	A0A087W...	A0A087W...		TPM3;D...
309	+	1.490772336...	3.100472...	A0A087W...	A0A087W...	Clathrin h...	CLTC
310	+	1.722580213...	2.991138...	A0A087W...	A0A087W...	Antigen K...	MKI67
311	+	2.753237809...	-2.62967...	A0A087W...	A0A087W...	Retrotran...	PEG10
312	+	2.826484117...	-3.51336...	A0A087W...	A0A087W...	Heteroge...	HNRNPD...
313	+	1.482426432...	-3.35459...	A0A024R...	A0A024R...	Vigilin	HDLBP

References

- Aebersold, R. & Mann, M., 2003. Mass spectrometry-based proteomics. *Nature*, 422(6928), pp.198–207.
- Akimitsu, N. et al., 2003. Enforced cytokinesis without complete nuclear division in embryonic cells depleting the activity of DNA topoisomerase II α . *Genes to Cells*, 8(4), pp.393–402.
- Anders, K.R., Grimson, A. & Anderson, P., 2003. SMG-5, required for *C. elegans* nonsense-mediated mRNA decay, associates with SMG-2 and protein phosphatase 2A. *The EMBO journal*, 22(3), pp.641–650.
- Applequist, S.E., Selg, M. & Ramanl, C., 1997. Cloning and characterization of HUPF1, a human homolog of the *Saccharomyces cerevisiae* nonsense mRNA-reducing UPF1 protein. *Nucleic acids research*, 25(4), pp.814–821.
- Axelrad, M.D., Budagov, T. & Atzmon, G., 2013. Telomere length and telomerase activity; a yin and yang of cell senescence. *Journal of visualized experiments : JoVE*, (75), pp.1–8.
- Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Azzalin, M., 2007. Telomeric and RNA Surveillance Factors Chromosome RNA Mammalian. , 318(5851), pp.798–801.
- Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J., 2007. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science (New York, N.Y.)*, 318(5851), pp.798–801.
- Azzalin, C.M., 2012. UPF1: A leader at the End of Chromosomes. *Nucleus*, 3(1), pp.16–21.
- Azzalin, C.M. & Lingner, J., 2006a. The Double Life of UPF1 in RNA and DNA Stability Pathways. *Cell Cycle*, 5(July), pp.1496–1498.
- Azzalin, C.M. & Lingner, J., 2006b. The Human RNA Surveillance Factor UPF1 Is Required for S Phase Progression and Genome Stability. *Current biology : CB*, 16(4), pp.1–5.
- Baker, K.E. & Parker, R., 2004. Nonsense-mediated mRNA decay: Terminating erroneous gene expression. *Current Opinion in Cell Biology*, 16(3), pp.293–299.

- Baldwin, M.A., 2004. Protein Identification by Mass Spectrometry. *Molecular & Cellular Proteomics*, 3(1), pp.1–9.
- Behm-Ansmant, I. & Izaurralde, E., 2006. Quality control of gene expression: A stepwise assembly pathway for the surveillance complex that triggers nonsense-mediated mRNA decay. *Genes and Development*, 20(4), pp.391–398.
- Belasco, J.G., 2010. All things must pass: Contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nature Reviews Molecular Cell Biology*, 11(7), pp.467–478.
- Bell, S.P., 2002. The origin recognition complex: From simple origins to complex functions. *Genes and Development*, 16(6), pp.659–672.
- Bell, S.P. & Dutta, A., 2002. DNA Replication in Eukaryotic Cells. *Annual Review of Biochemistry*, 71(1), pp.333–374.
- Berg, M. et al., 2006. Reproducibility of LC-MS-based protein identification. *Journal of Experimental Botany*, 57(7), pp.1509–1514.
- Berggård, T., Linse, S. & James, P., 2007. Methods for the detection and analysis of protein-protein interactions. *Proteomics*, 7, pp.2833–2842.
- Bergink, S. & Jentsch, S., 2009. Principles of ubiquitin and SUMO modifications in DNA repair. *Nature*, 458(March), pp.461–467.
- Bermejo, R. et al., 2007. Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation. *Genes and Development*, 21(15), pp.1921–1936.
- Bermudez, V.P. et al., 2003. Loading of the human 9-1-1 checkpoint complex onto DNA by the checkpoint clamp loader hRad17-replication factor C complex in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 100(4), pp.1633–1638.
- Bhattacharya, A. et al., 2000. Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. *RNA*, 6, pp.1226–1235.

- Bono, F., 2014. Juggling Key Players in NMD Initiation. *Structure (London, England : 1993)*, 22(8), pp.1074–5.
- Broderick, R. et al., 2015. TOPBP1 recruits TOP2A to ultra-fine anaphase bridges to aid in their resolution. *Nature Communications*, 6, pp.1–27.
- Brogna, S. & Wen, J., 2009. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature structural & molecular biology*, 16(2), pp.107–13.
- Brown, E.J. & Baltimore, D., 2003. Essential and dispensable roles of ATR in cell cycle arrest and genome maintenance. *Genes & development*, 17(5), pp.615–28.
- Burnette, W.N., 1981. Western Blotting: Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Analytical Biochemistry*, 112(2), pp.195–203.
- Callister, S.J. et al., 2006. Normalization Approaches for Removing Systematic Biases Associated with Mass Spectrometry and Label-Free Proteomics. *Journal of Proteome Research*, 5(2), pp.277–286.
- de Campos-Nebel, M., Larripa, I. & González-Cid, M., 2010. Topoisomerase ii-mediated DNA damage is differently repaired during the cell cycle by non-homologous end joining and homologous recombination. *PLoS ONE*, 5(9), pp.1–13.
- Cappadona, S. et al., 2012. Current challenges in software solutions for mass spectrometry-based quantitative proteomics. *Amino*, 43, pp.1087–1108.
- Carastro, L.M. et al., 2002. Identification of delta helicase as the bovine homolog of HUPF1: demonstration of an interaction with the third subunit of DNA polymerase delta. *Nucleic acids research*, 30(10), pp.2232–43.
- Catherman, A.D., Skinner, O.S. & Kelleher, N.L., 2014. Top Down proteomics: Facts and perspectives. *Biochemical and Biophysical Research Communications*, 445(4), pp.683–693.
- Chakrabarti, S. et al., 2011. Molecular mechanisms for the RNA-dependent ATPase activity

- of Upf1 and its regulation by Upf2. *Molecular cell*, 41(6), pp.693–703.
- Chakrabarti, S. et al., 2014. Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5-SMG7 and SMG6. *Nucleic acids research*, 1(23), pp.1–14.
- Chamieh, H. et al., 2008. NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nature structural & molecular biology*, 15(1), pp.85–93.
- Chang, J.C. & Kan, Y.W., 1979. Beta 0 Thalassemia, a Nonsense Mutation in Man. *Proceedings of the National Academy of Sciences of the United States of America*, 76(6), pp.2886–2889.
- Chanoux, R.A. et al., 2009. ATR and H2AX cooperate in maintaining genome stability under replication stress. *Journal of Biological Chemistry*, 284(9), pp.5994–6003.
- Chatr-Aryamontri, A. et al., 2015. The BioGRID interaction database: 2015 update. *Nucleic Acids Research*, 43(D1), pp.D470–D478.
- Chavez, A., Tsou, A.M. & Johnson, F.B., 2009. Telomeres do the (un)twist: helicase actions at chromosome termini. *Biochimica et biophysica acta*, 1792(4), pp.329–40.
- Chawla, R. et al., 2011. Human UPF1 interacts with TPP1 and telomerase and sustains telomere leading-strand replication. *The EMBO journal*, 30(19), pp.4047–58.
- Chawla, R. & Azzalin, C.M., 2009. The telomeric transcriptome and SMG proteins at the crossroads. *Cytogenetic and Genome Research*, 122(3–4), pp.194–201.
- Cheng, Z. et al., 2007. Structural and functional insights into the human Upf1 helicase core. *The EMBO journal*, 26(1), pp.253–64.
- Chiu, S.-Y. et al., 2003. Characterization of human Smg5/7a: A protein with similarities to *Caenorhabditis elegans* SMG5 and SMG7 that functions in the dephosphorylation of Upf1. *Rna*, 9(1), pp.77–87.
- Cho, H. et al., 2013. SMG5-PNRC2 is functionally dominant compared with SMG5-SMG7 in

- mammalian nonsense-mediated mRNA decay. *Nucleic Acids Research*, 41(2), pp.1319–1328.
- Choe, J. et al., 2012. Translation initiation on mRNAs bound by nuclear cap-binding protein complex CBP80/20 requires interaction between CBP80/20-dependent translation initiation factor and eukaryotic translation initiation factor 3g. *The Journal of biological chemistry*, 287(22), pp.18500–9.
- Choi, H. et al., 2012. Analyzing Protein-Protein Interactions from Affinity Purification-Mass Spectrometry Data with SAINT. In *Current Protocols in Bioinformatics*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Choi, H. et al., 2015. QPROT: statistical method for testing differential expression using protein-level intensity data in label-free quantitative proteomics. *J Proteomics*, 129, pp.121–126.
- Chu, H.P. et al., 2017. TERRA RNA Antagonizes ATRX and Protects Telomeres. *Cell*.
- Chung, H.R. et al., 2010. The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS ONE*, 5(12).
- Cimprich, K. & Cortez, D., 2008. ATR: An Essential Regulator of Genome Integrity. *Nature Reviews Molecular Cell Biology*, 9(8), pp.616–627.
- Clissold, P.M. & Ponting, C.P., 2000. PIN domains in nonsense-mediated mRNA decay and RNAi. *Current Biology*, 10(24), pp.R888–R890.
- Conti, E. & Izaurralde, E., 2005. Nonsense-mediated mRNA decay: Molecular insights and mechanistic variations across species. *Current Opinion in Cell Biology*, 17(3), pp.316–325.
- Corrette-Bennett, S.E. et al., 2004. DNA polymerase delta, RFC and PCNA are required for repair synthesis of large looped heteroduplexes in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 32(21), pp.6268–6275.
- Cortez, D., 2001. ATR and ATRIP: Partners in Checkpoint Signaling. *Science*, 294, pp.1713–1716.

- Cox, J. et al., 2014. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular & Cellular Proteomics*, 13(9), pp.2513–2526.
- Cox, J. & Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26(12), pp.1367–1372.
- Cuadrado, M. et al., 2006. ATM regulates ATR chromatin loading in response to DNA double-strand breaks. *The Journal of experimental medicine*, 203(2), pp.297–303.
- Culbertson, M.R., 1999. RNA surveillance unforeseen consequences for gene expression, inherited genetic disorders and cancer. *Trends in Genetics*, 15(2), pp.254–259.
- Czaplinski, K. et al., 1995. Purification and characterization of the UPF1 protein: A factor involved in translation and mRNA degradation. *RNA (New York, N.Y.)*, 1, pp.610–623.
- Davis, A. & Chen, D., 2013. DNA double strand break repair via non-homologous end-joining. *Translational Cancer Research*, 2(3), pp.130–43.
- Dehghani-tafti, S. & Sanders, C.M., 2017. DNA substrate recognition and processing by the full-length human UPF1 helicase. *Nucleic acids research*, 45(12), pp.7354–7366.
- Deniaud, A. et al., 2015. A network of SMG-8, SMG-9 and SMG-1 C-terminal insertion domain regulates UPF1 substrate recruitment and phosphorylation. *Nucleic acids research*, 43(15), pp.7600–11.
- Dingwall, C., Lomonosoff, G.P. & Laskey, R.A., 1981. High sequence specificity of micrococcal nuclease. *Nucleic Acids Research*, 9(12), pp.2659–2674.
- Dunham, W.H., Mullin, M. & Gingras, A.-C., 2012. Affinity-purification coupled to mass spectrometry: Basic principles and strategies. *PROTEOMICS*, 12(10), pp.1576–1590.
- Durand, S., Franks, T.M. & Lykke-Andersen, J., 2016. Hyperphosphorylation amplifies UPF1 activity to resolve stalls in nonsense-mediated mRNA decay. *Nature communications*, 7, p.12434.

- Elias, J.E. & Gygi, S.P., 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods*, 4(3), pp.207–214.
- Ellegren, H., 2004. Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics*, 5(6), pp.435–445.
- Ellison, V. & Stillman, B., 2003. Biochemical characterization of DNA damage checkpoint complexes: Clamp loader and clamp complexes with specificity for 5' recessed DNA. *PLoS Biology*, 1(2).
- Ellison, V. & Stillman, B., 1998. Reconstitution of recombinant human replication factor C (RFC) and identification of an RFC subcomplex possessing DNA-dependent ATPase activity. *Journal of Biological Chemistry*, 273(10), pp.5979–5987.
- Eriksson, J. & Fenyö, D., 2008. Optimizing Sensitivity and Specificity in Mass Spectrometric Proteome Analysis. In *Mass Spectrometry*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 211–221.
- Euskirchen, G., Auerbach, R.K. & Snyder, M., 2012. SWI/SNF Chromatin-remodeling Factors: Multiscale Analyses and Diverse Functions. *Journal of Biological Chemistry*, 287(37), pp.30897–30905.
- Fallis, A., 2013. Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway. *Journal of Chemical Information and Modeling*, 53(9), pp.1689–1699.
- Feist, P. & Hummon, A.B., 2015. Proteomic challenges: Sample preparation techniques for Microgram-Quantity protein analysis from biological samples. *International Journal of Molecular Sciences*, 16, pp.3537–3563.
- Feng, Q., Jagannathan, S. & Bradley, R.K., 2017. The RNA Surveillance Factor UPF1 Represses Myogenesis via Its E3 Ubiquitin Ligase Activity. *Molecular Cell*, 67(2), p.239–251.e6.
- Fenyö, D., 2000. Identifying the proteome: Software tools. *Current Opinion in Biotechnology*, 11(4), pp.391–395.
- Feuerhahn, S. et al., 2010. TERRA biogenesis, turnover and implications for function. *FEBS*

- letters*, 584(17), pp.3812–8.
- Fiorini, F., Boudvillain, M. & Le Hir, H., 2013. Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. *Nucleic acids research*, 41(4), pp.2404–15.
- Flury, V. et al., 2014. Characterization of phosphorylation- and RNA-dependent UPF1 interactors by quantitative proteomics. *Journal of Proteome Research*, 13(6), pp.3038–3053.
- Gaillard, H., García-Muse, T. & Aguilera, A., 2015. Replication stress and cancer. *Nature Reviews Cancer*, 15, pp.276–289.
- Gary Schmidt, S.L., Gomes, X. V & Burgers, P.M.J., 2001. ATP utilization by yeast replication factor C: III. The ATP-binding domains of Rfc2, Rfc3, and Rfc4 are essential for DNA recognition and clamp loading. *Journal of Biological Chemistry*, 276(37), pp.34784–34791.
- Gatfield, D. et al., 2003. Nonsense-mediated mRNA decay in *Drosophila*: At the intersection of the yeast and mammalian pathways. *EMBO Journal*, 22(15), pp.3960–3970.
- George, T. et al., 2009. Human Pif1 helicase unwinds synthetic DNA structures resembling stalled DNA replication forks. *Nucleic Acids Research*, 37(19), pp.6491–6502.
- Gerace, E. & Moazed, D., 2015. Affinity Pull-Down of Proteins Using Anti-FLAG M2 Agarose Beads. In *Methods in Enzymology*. pp. 99–110.
- Gingras, A.-C. et al., 2007. Analysis of protein complexes using mass spectrometry. *Nature Reviews Molecular Cell Biology*, 8(8), pp.645–654.
- Glavan, F. et al., 2006. Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *The EMBO journal*, 25, pp.5117–25.
- Goss Tusher, V. et al., 2001. Significance Analysis of Microarrays Applied to the Ionizing Radiation Response Significance analysis of micre ionizing radiation response. *Source: Proceedings of the National Academy of Sciences of the United States of America*, 98(9), pp.5116–5121.

- Grabarczyk, D.B., Silkenat, S. & Kisker, C., 2018. Structural Basis for the Recruitment of Ctf18-RFC to the Replisome. *Structure*, 26(1), p.137–144.e3.
- Griffin, N.M. et al., 2010. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nature Biotechnology*, 28(1), pp.83–89.
- Gumeni, S. et al., 2017. Proteome Stability as a Key Factor of Genome Integrity. *International Journal of Molecular Sciences*, 18(10), p.2036.
- Gundry, R.L. et al., 2009. Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. In *Current Protocols in Molecular Biology*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 342–355.
- Gunjan, A. & Verreault, A., 2003. A Rad53 Kinase-Dependent Surveillance Mechanism that Regulates Histone Protein Levels in *S. cerevisiae*. *Cell*, 115, pp.537–549.
- Hall-Jackson, C. a et al., 1999. ATR is a caffeine-sensitive, DNA-activated protein kinase with a substrate specificity distinct from DNA-PK. *Oncogene*, 18(November), pp.6707–6713.
- Hanahan, D. & Weinberg, R.A., 2011. Review Hallmarks of Cancer : The Next Generation. *Cell*, 144(5), pp.646–674.
- Hao, J. et al., 2015. And-1 coordinates with Claspin for efficient Chk1 activation in response to replication stress. *The EMBO Journal*, 34(15), pp.1845–1985.
- Harkin, L.F. et al., 2016. Distinct expression patterns for type II topoisomerases IIA and IIB in the early foetal human telencephalon. *Journal of Anatomy*, 228(3), pp.452–463.
- Hedglin, M., Kumar, R. & Benkovic, S.J., 2013. Replication clamps and clamp loaders. *Cold Spring Harbor Perspectives in Biology*, 5(4), pp.1–19.
- Hein, M.Y., 2014. *A Human Interactome*. Ludwig-Maximilian University of Munich.
- Hein, M.Y. et al., 2013. Proteomic Analysis of Cellular Systems. In M. A. . Walhout, M. Vidal, & J. Dekker, eds. *Handbook of Systems Biology*. Elsevier, pp. 3–25.
- Higa, M., Fujita, M. & Yoshida, K., 2017. DNA Replication Origins and Fork Progression at Mammalian Telomeres. *Genes*, 8(112), pp.1–24.

- Higgins, N.P., 2012. A human TOP2A core DNA binding X-ray structure reveals topoisomerase subunit dynamics and a potential mechanism for SUMO modulation of decatenation. *Journal of Molecular Biology*, 424(3–4), pp.105–108.
- Hiom, K., 2005. DNA repair: How to PIKK a partner. *Current Biology*, 15(12), pp.473–475.
- Le Hir, H. et al., 2000. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *EMBO Journal*, 19(24), pp.6860–6869.
- Ho, C.C. et al., 2006. Stalled replication induces p53 accumulation through distinct mechanisms from DNA damage checkpoint pathways. *Cancer Research*, 66(4), pp.2233–2241.
- Horz, W. & Altenburger, W., 1981. Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic Acids Research*, 9(12), pp.2643–2658.
- Hug, N., Longman, D. & Ceres, J.F., 2015. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research*, 44(4), pp.1483–1495.
- Huntzinger, E. et al., 2008. SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA (New York, N.Y.)*, 14, pp.2609–2617.
- Hwang, J. et al., 2010. UPF1 association with the cap-binding protein, CBP80, promotes nonsense-mediated mRNA decay at two distinct steps. *Molecular Cell*, 39(3), pp.396–409.
- Iborra, F.J. et al., 2004. Molecular cross-talk between the transcription, translation, and nonsense-mediated decay machineries. *Journal of Cell Science*, 117(6), pp.899–906.
- Imamachi, N., Tani, H. & Akimitsu, N., 2012. Up-frameshift protein 1 (UPF1): Multitalented entertainer in RNA decay. *Drug Discoveries & Therapeutics*, 6(2), pp.55–61.
- Isken, O. et al., 2008. Upf1 Phosphorylation Triggers Translational Repression during Nonsense-Mediated mRNA Decay. *Cell*, 133, pp.314–327.
- Isken, O. & Maquat, L.E., 2008. The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nature reviews. Genetics*, 9(9), pp.699–712.

- Ito, T. et al., 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences*, 98(8), pp.4569–4574.
- Ivaska, J. et al., 2007. Novel functions of vimentin in cell adhesion, migration, and signaling. *Experimental Cell Research*, 313(10), pp.2050–2062.
- Izawa, N. et al., 2011. HERC2 interacts with claspin and regulates DNA origin firing and replication fork progression. *Cancer Research*, 71(17), pp.5621–5625.
- Jeong, S. & Stein, A., 1994. Micrococcal nuclease digestion of nuclei reveals extended nucleosome ladders having anomalous DNA lengths for chromatin assembled on non-replicating plasmids in transfected cells. *Nucleic Acids Research*, 22(3), pp.370–375.
- Jonas, S., Weichenrieder, O. & Izaurralde, E., 2013. An unusual arrangement of two 14-3-3-like domains in the SMG5-SMG7 heterodimer is required for efficient nonsense-mediated mRNA decay. *Genes and Development*, 27, pp.211–225.
- Jones, R.M. & Petermann, E., 2012. Replication fork dynamics and the DNA damage response. *Biochemical Journal*, 443(1), pp.13–26.
- Kaboord, B. & Perr, M., 2008. Isolation of Proteins and Protein Complexes by Immunoprecipitation. In *2D PAGE: Sample Preparation and Fractionation*. Humana Press, pp. 349–364.
- Karousis, E.D., Nasif, S. & Mühlemann, O., 2016. Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdisciplinary Reviews: RNA*, 7, pp.661–682.
- Karpievitch, Y. V, Dabney, A.R. & Smith, R.D., 2012. Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics*, 13(Suppl 16), p.S5.
- Karve, T.M. & Cheema, A.K., 2011. Small Changes Huge Impact : The Role of Protein Posttranslational Modifications in Cellular Homeostasis and Disease. *Journal of Amino Acids*, 2011.
- Kashima, I. et al., 2006a. Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA

- decay. *Genes & development*, 20(3), pp.355–67.
- Kashima, I. et al., 2006b. 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 and Development*, 20(3), pp.355–367.
- Kashima, I. et al., 2006c. 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 and Development*, 20, pp.355–367.
- Kaygun, H. & Marzluff, W.F., 2005a. Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1. *Nature structural & molecular biology*, 12(9), pp.794–800.
- Kaygun, H. & Marzluff, W.F., 2005b. Translation Termination Is Involved in Histone mRNA Degradation when DNA Replication Is Inhibited. *Molecular and cellular biochemistry*, 25(16), pp.6879–6888.
- Keall, R. et al., 2007. Histone gene expression and histone mRNA 3' end structure in *Caenorhabditis elegans*. *BMC molecular biology*, 8, p.51.
- Keilhauer, E.C., Hein, M.Y. & Mann, M., 2015. Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) Rather than Affinity Purification Mass Spectrometry (AP-MS). *Molecular & Cellular Proteomics*, 14(1), pp.120–135.
- Kervestin, S. & Jacobson, A., 2012. NMD: a multifaceted response to premature translational termination. *Nature reviews. Molecular cell biology*, 13(11), pp.700–712.
- Khoo, O. & Suntrarachun, S., 2012. Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pacific Journal of Tropical Biomedicine*, 2(2), pp.159–162.
- Kim, S. et al., 1999. Substrate Specificities and Identification of Putative Substrates of ATM Kinase Family Members Substrate Specificities and Identification of Putative Substrates of ATM Kinase Family Members *. *The Journal of Biological Chemistry*, 274(53), pp.37538–37543.

- Kinniburgh, A. et al., 1982. mRNA-deficient β 0-thalassemia results from a single nucleotide deletion. *Nucleic Acids Research*, 10(18), pp.5421–5427.
- Kornberg, R., 1974. Chromatin Structure : A Repeating Unit of Histones and DNA Chromatin structure is based on a repeating unit of eight. *Science*, 184, pp.868–871.
- Kubota, T., Myung, K. & Donaldson, A.D., 2013. Is PCNA unloading the central function of the Elg1/ ATAD5 replication factor C-like complex? *Cell Cycle*, 12(16), pp.2570–2579.
- Laboratory of Mass Spectrometry, LNBio, C., 2014. *Tutorial for proteome data analysis using the Perseus software platform*,
- Laemmli, U.K. (1970):, 1970. Cleavage of Structural Proteins during Assembly of Head of Bacteriophage T4. *Nature*, 227, pp.680–685.
- Lambert, S. & Carr, A.M., 2005. Checkpoint responses to replication fork barriers. *Biochimie*, 87(7), pp.591–602.
- De Lange, T., 2005. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & development*, 19(18), pp.2100–10.
- Langie, S.A.S. et al., 2015. Causes of genome instability : the effect of low dose chemical exposures in modern society. *Carcinogenesis*, 36, pp.61–88.
- Latif, C., den Elzen, N.R. & O’Connell, M.J., 2004. DNA damage checkpoint maintenance through sustained Chk1 activity. *Journal of cell science*, 117(16), pp.3489–3498.
- Lee, J.H. & Paull, T.T., 2005. ATM Activation by DNA Double-Strand Breaks Through the Mre11-Rad50-Nbs1 Complex. *Science*, 308, pp.551–554.
- Lee, K.Y. et al., 2013. ATAD5 regulates the lifespan of DNA replication factories by modulating PCNA level on the chromatin. *Journal of Cell Biology*, 200(1), pp.31–44.
- Lee, K.Y. et al., 2010. Human ELG1 regulates the level of ubiquitinated proliferating cell nuclear antigen (PCNA) through its interactions with PCNA and USP1. *Journal of Biological Chemistry*, 285(14), pp.10362–10369.
- Lee, S.R. et al., 2015. Target Discrimination in Nonsense-Mediated mRNA Decay Requires

- Upf1 ATPase Activity Article Target Discrimination in Nonsense-Mediated mRNA Decay Requires Upf1 ATPase Activity. *Molecular Cell*, 59(3), pp.413–425.
- Leeds, P. et al., 1991. The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes and Development*, 5, pp.2303–2314.
- Lejeune, F. et al., 2002. The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: Dynamics of mRNP remodeling. *EMBO Journal*, 21(13), pp.3536–3545.
- Leman, A.R. & Noguchi, E., 2013. The replication fork: Understanding the eukaryotic replication machinery and the challenges to genome duplication. *Genes*, 4, pp.1–32.
- Life Technologies Corporation, 2012. *Flp-In™ T-REx™ Core Kit For Generating Stable, Inducible Mammalian Expression Cell Lines by Flp Recombinase-Mediated Integration*.
- Lodish, H. et al., 2000. RNA Processing, Nuclear Transport, and Post-Transcriptional Control. In *Molecular Cell Biology*.
- Losson, R. & Lacroute, F., 1979. Interference of nonsense mutations with eukaryotic messenger RNA stability. *Proceedings of the National Academy of Sciences of the United States of America*, 76(10), pp.5134–5137.
- Lovejoy, C.A. & Cortez, D., 2009. Common mechanisms of PIKK regulation. *DNA Repair*, 8(9), pp.1004–1008.
- Luczak, M. et al., 2016. Label-free quantitative proteomics reveals differences in molecular mechanism of atherosclerosis related and non-related to chronic kidney disease. *International Journal of Molecular Sciences*, 17(5), pp.1–18.
- Luke, B. & Lingner, J., 2009. TERRA: telomeric repeat-containing RNA. *The EMBO journal*, 28(17), pp.2503–10.
- Lykke-Andersen, J., Shu, M.D. & Steitz, J.A., 2000. Human Upf proteins target an mRNA for nonsense-mediated decay when downstream of a termination codon. *Cell*, 103, pp.1121–1131.

- Maestroni, L., Matmati, S. & Coulon, S., 2017. Solving the Telomere Replication Problem. *Genes*, 8(55), pp.1–16.
- Maher, C.A. & Wilson, R.K., 2012. Chromothripsis and human disease: Piecing together the shattering process. *Cell*, 148(1–2), pp.29–32.
- Majka, J. & Burgers, P.M.J., 2004. The PCNA-RFC Families of DNA Clamps and Clamp Loaders. *Progress in Nucleic Acid Research and Molecular Biology*, 78, pp.227–260.
- Malumbres, M., 2014. Cyclin-dependent kinases. *Genome Biology*, 15(6), pp.1–10.
- Malumbres, M. & Barbacid, M., 2005. Mammalian cyclin-dependent kinases. *Trends in Biochemical Sciences*, 30(11), pp.630–641.
- Mann, M., 2006. Functional and quantitative proteomics using SILAC. *Nature Reviews Molecular Cell Biology*, 7(12), pp.952–958.
- Maquat, L.E., 2006. *Nonsense-mediated mRNA decay*, Landes Bioscience.
- Maquat, L.E., 2005. Nonsense-mediated mRNA decay in mammals. *Journal of Cell Science*, 118(9), pp.1773–1776.
- Maquat, L.E. et al., 1981. Unstable beta-globin mRNA in mRNA-deficient beta o thalassemia. *Cell*, 27, pp.543–553.
- Marcotte, E.M., 2007. How do shotgun proteomics algorithms identify proteins? *Nature Biotechnology*, 25(7), pp.755–757.
- Maréchal, A. & Zou, L., 2013. DNA Damage Sensing by the ATM and ATR Kinases. *Cold Spring Harbor Perspectives in Biology*, 5, pp.1–18.
- Mari, P.-O. et al., 2006. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proceedings of the National Academy of Sciences*, 103(49), pp.18597–18602.
- Masai, H. et al., 2010. Eukaryotic chromosome DNA replication: where, when, and how? *Annual review of biochemistry*, 79, pp.89–130.
- MassIVE, 2011. Summary.

- Matelska, D., Steczkiewicz, K. & Ginalski, K., 2017. Comprehensive classification of the PIN domain-like superfamily. *Nucleic Acids Research*, 45(12), pp.6995–7020.
- Matia-González, A.M. et al., 2013. Functional characterization of Upf1 targets in *Schizosaccharomyces pombe*. *RNA biology*, 10(6), pp.1057–65.
- Mayer, M.L. et al., 2001. Identification of RFC(Ctf18p, Ctf8p, Dcc1p): An alternative RFC complex required for sister chromatid cohesion in *S. cerevisiae*. *Molecular Cell*, 7(5), pp.959–970.
- Mazouzi, A., Velimezi, G. & Loizou, J.I., 2014. DNA replication stress: Causes, resolution and disease. *Experimental Cell Research*, 329(1), pp.85–93.
- Melero, R. et al., 2014. Structures of SMG1-UPFs complexes: SMG1 contributes to regulate UPF2-dependent activation of UPF1 in NMD. *Structure*, 22, pp.1105–1119.
- Mellacheruvu, D. et al., 2013. The CRAPome: a contaminant repository for affinity purification–mass spectrometry data. *Nature Methods*, 10(8), pp.730–736.
- Mendell, J.T., Collete, M. & Dietz, H.C., 2002. Separable Roles for rent1/hUpf1 in Altered Splicing and Decay of Nonsense Transcripts. *Science*, 298(5592), pp.419–422.
- Méndez, J. & Stillman, B., 2000. Chromatin Association of Human Origin Recognition Complex , Cdc6 , and Minichromosome Maintenance Proteins during the Cell Cycle : Assembly of Prereplication Complexes in Late Mitosis Chromatin Association of Human Origin Recognition Complex , Cdc6 , and. *Molecular and cellular biology*, 20(22).
- Miranda, M. & Sorkin, A., 2007. Regulation of Receptors and Transporters by Ubiquitination: New Insights into Surprisingly Similar Mechanisms. *Molecular Interventions*, 7(3), pp.157–167.
- Morere, J. et al., 2008. TRF2 and Apollo Cooperate with Topoisomerase 2 a to Protect Human Telomeres from Replicative Damage. *Cell*, 142, pp.230–242.
- Mukaka, M.M., 2012. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), pp.69–71.

- Müller, B. et al., 2007. DNA-activated protein kinase functions in a newly observed S phase checkpoint that links histone mRNA abundance with DNA replication. *The Journal of cell biology*, 179(7), pp.1385–98.
- Nagy, E. & Maquat, L.E., 1998. A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends in Biochemical Sciences*, 23(6), pp.198–199.
- Negrini, S., Gorgoulis, V.G. & Halazonetis, T.D., 2010. Genomic instability — an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology*, 11, pp.220–228.
- Nesvizhskii, A.I. & Aebersold, R., 2005. Interpretation of Shotgun Proteomic Data: The protein inference problem. *Molecular & Cellular Proteomics*, 4(10), pp.1419–1440.
- Nicholson, P. et al., 2014. A novel phosphorylation-independent interaction between SMG6 and UPF1 is essential for human NMD. *Nucleic acids research*, 42(14), pp.9217–9235.
- Nicholson, P. et al., 2010. Nonsense-mediated mRNA decay in human cells: Mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cellular and Molecular Life Sciences*, 67(5), pp.677–700.
- Nicholson, P., Joncourt, R. & Muhlemann, O., 2012. Analysis of Nonsense-Mediated mRNA Decay in Mammalian Cells. *Current Protocols in Cell Biology*.
- Nicholson, P. & Mühlemann, O., 2010. Cutting the nonsense: the degradation of PTC-containing mRNAs. *Biochemical Society transactions*, 38, pp.1615–1620.
- Nitiss, J.L., 2009. DNA topoisomerase II and its growing repertoire of biological functions. *Nature Reviews Cancer*, 9(5), pp.327–337.
- Nyberg, K. a et al., 2002. Toward maintaining the genome: DNA damage and replication checkpoints. *Annual review of genetics*, 36, pp.617–56.
- Ohnishi, T. et al., 2003. Phosphorylation of hUPF1 Induces Formation of mRNA Surveillance Complexes Containing hSMG-5 and hSMG-7. *Molecular Cell*, 12(5), pp.1187–1200.
- Okada-Katsuhata, Y. et al., 2012. N- and C-terminal Upf1 phosphorylations create binding

- platforms for SMG-6 and SMG-5:SMG-7 during NMD. *Nucleic acids research*, 40(3), pp.1251–66.
- Olave, I.A., Reck-Peterson, S.L. & Crabtree, G.R., 2002. Nuclear Actin and Actin-Related Proteins in Chromatin Remodeling. *Annual Review of Biochemistry*, 71(1), pp.755–781.
- Ong, S.-E. et al., 2002. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics*, 1(5), pp.376–386.
- Ong, S.-E. & Mann, M., 2006. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature protocols*, 1(6), pp.2650–60.
- Origa, R., Beta-Thalassemia. In M. P. Adam et al., eds. *GeneReview*. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK1426/>.
- Ottens, F. et al., 2017. Transcript-specific characteristics determine the contribution of endo- and exonucleolytic decay pathways during the degradation of nonsense-mediated decay substrates. *RNA*, 23(8), pp.1224–1236.
- Overmeer, R.M. et al., 2010. Replication Factor C Recruits DNA Polymerase to Sites of Nucleotide Excision Repair but Is Not Required for PCNA Recruitment. *Molecular and Cellular Biology*, 30(20), pp.4828–4839.
- Page, M.F. et al., 1999. SMG-2 is a phosphorylated protein required for mRNA surveillance in *Caenorhabditis elegans* and related to Upf1p of yeast. *Molecular and cellular biology*, 19(9), pp.5943–5951.
- Pal, M. et al., 2001. Evidence that phosphorylation of human Upf1 protein varies with intracellular location and is mediated by a wortmannin-sensitive and rapamycin-sensitive PI 3-kinase-related kinase signaling pathway. *RNA (New York, N.Y.)*, 5(3), pp.2808–2839.
- Palm, W. & de Lange, T., 2008. How shelterin protects mammalian telomeres. *Annual review of genetics*, 42, pp.301–34.
- Panamwan, P., 2017. *Analysis of the Mechanism of DNA Damage and Replication Arrest-*

- induced Histone mRNA Decay*. University of Sheffield.
- Panier, S. & Durocher, D., 2009. Regulatory ubiquitylation in response to DNA double-strand breaks. , 8, pp.436–443.
- Park, E., Gleghorn, M.L. & Maquat, L.E., 2013. Stauf2 functions in Stauf1-mediated mRNA decay by binding to itself and its paralog and promoting UPF1 helicase but not ATPase activity. *Proceedings of the National Academy of Sciences of the United States of America*, 110(2), pp.405–12.
- Park, E. & Maquat, L.E., 2013. Stauf-mediated mRNA decay. *Wiley Interdisciplinary Reviews: RNA*, 4(4), pp.423–435.
- Parrilla-Castellar, E.R., Arlander, S.J.H. & Karnitz, L., 2004. Dial 9-1-1 for DNA damage: The Rad9-Hus1-Rad1 (9-1-1) clamp complex. *DNA Repair*, 3(8–9), pp.1009–1014.
- Perlick, H.A. et al., 1996. Mammalian orthologues of a yeast regulator of nonsense transcript stability. *Proceedings of the National Academy of Sciences of the United States of America*, 93(20), pp.10928–32.
- Rao, V.S. et al., 2014. Protein-Protein Interaction Detection: Methods and Analysis. *International Journal of Proteomics*, 2014, pp.1–12.
- Redondo-Muñoz, J. et al., 2013. Phosphoinositide 3-kinase beta controls replication factor C assembly and function. *Nucleic Acids Research*, 41(2), pp.855–868.
- Reig-Viader, R. et al., 2013. Telomeric repeat-containing RNA and telomerase in human fetal oocytes. *Human reproduction (Oxford, England)*, 28(2), pp.414–22.
- Roux, K.J., Kim, D.I. & Burke, B., 2013. BioID: A screen for protein-protein interactions. *Current Protocols in Protein Science*, (Unit 19.23), pp.1–14.
- Schell, T. et al., 2003. Complexes between the nonsense-mediated mRNA decay pathway factor human upf1 (up-frameshift protein 1) and essential nonsense-mediated mRNA decay factors in HeLa cells. *The Biochemical journal*, 373, pp.775–783.
- Schwanhüsser, B. et al., 2011a. Global quantification of mammalian gene expression

- control. *Nature*, 473(7347), pp.337–342.
- Schwanhüusser, B. et al., 2011b. Supplement: Global quantification of mammalian gene expression control. *Nature*, 473(7347), pp.337–342.
- Schweingruber, C. et al., 2016. Identification of Interactions in the NMD Complex Using Proximity-Dependent Biotinylation (BioID). *Plos One*, 11(3), pp.1–27.
- Schweingruber, C. et al., 2013. Nonsense-mediated mRNA decay - Mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(6–7), pp.612–623.
- Seshacharyulu, P. et al., 2013. Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer. *Cancer Letters*, 335(1), pp.9–18.
- Sethuraman, A. et al., 2016. SMARCE1 regulates metastatic potential of breast cancer cells through the HIF1A/PTK2 pathway. *Breast Cancer Research*, 18(1), pp.1–15.
- Sharma, A., Singh, K. & Almasan, A., 2012. *Histone H2AX phosphorylation: a marker for DNA damage*. 3rd ed. L. Bjergbaek, ed., Humana Press.
- Shechter, D., Costanzo, V. & Gautier, J., 2004. Regulation of DNA replication by ATR: signaling in response to DNA intermediates. *DNA repair*, 3(8–9), pp.901–8.
- Shkedy, D. et al., 2015. Regulation of Elg1 activity by phosphorylation. *Cell Cycle*, 14(23), pp.3689–3697.
- Shore, D., 2001. Telomeric chromatin: replicating and wrapping up chromosome ends. *Current opinion in genetics & development*, 11(2), pp.189–98.
- Sigma-Aldrich, 2017. Amplified Detection Duolink[®] Proximity Ligation Assay (PLA).
- Sikdar, N. et al., 2009. DNA damage responses by human ELG1 in S phase are important to maintain genomic integrity. *Cell Cycle*, 8(19), pp.3199–3207.
- Singh, S., Hein, M.Y. & Stewart, A.F., 2016. msVolcano: A flexible web application for visualizing quantitative proteomics data. *PROTEOMICS*, 16(18), pp.2491–2494.
- Smaczniak, C. et al., 2012. Proteomics-based identification of low-abundance signaling and

- regulatory protein complexes in native plant tissues. *Nature Protocols*, 7(12), pp.2144–2158.
- Smith, G. & Jackson, S., 1999. The DNA-dependent protein kinase. *Genes & Development*, 13(8), pp.916–934.
- Smith, J. & Baker, K., 2015. Nonsense-mediated RNA decay- a switch for regulating gene expression. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 37(6), pp.612–623.
- Stagno, M. et al., 2014. Topoll α prevents telomere fragility and formation of ultra thin DNA bridges during mitosis through TRF1-dependent binding to telomeres Topoll α prevents telomere fragility and formation of ultra thin DNA bridges during mitosis through TRF1-dependent bind. *Cell Cycle*, 13(9), pp.1463–1481.
- Stark, C., 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, 34(90001), pp.D535–D539.
- Stokes, M.P. et al., 2007. Profiling of UV-induced ATM/ATR signaling pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50), pp.19855–60.
- Sun, X. et al., 1998. A mutated human homologue to yeast Upf1 protein has a dominant-negative effect on the decay of nonsense-containing mRNAs in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, 95(August), pp.10009–10014.
- Sun, X. & Maquat, L.E., 2000. mRNA surveillance in mammalian cells: the relationship between introns and translation termination. *Rna*, 6, p.1.
- Takai, K.K. et al., 2010. In vivo stoichiometry of shelterin components. *The Journal of biological chemistry*, 285(2), pp.1457–67.
- Tokmakov, A.A. et al., 2012. Multiple post-translational modifications affect heterologous protein synthesis. *Journal of Biological Chemistry*, 287(32), pp.27106–27116.
- Tomida, J. et al., 2008. DNA damage-induced ubiquitylation of RFC2 subunit of replication

- factor C complex. *Journal of Biological Chemistry*, 283(14), pp.9071–9079.
- Torrente, M.P. et al., 2011. Proteomic interrogation of human chromatin. *PLoS ONE*, 6(9).
- Turton, D., 2014. *The functional analysis of Upf1 in S-phase progression and genome stability*. University of Sheffield.
- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., et al., 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, 13(9), pp.731–740.
- Uhlmann, F. et al., 1997. Identification of Regions within the Four Small Subunits of Human Replication Factor C Required for Complex Formation and DNA Replication. *Journal of Biological Chemistry*, 272(15), pp.10058–10064.
- Unterholzner, L. & Izaurralde, E., 2004. SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. *Molecular Cell*, 16, pp.587–596.
- Varsally, W. & Brogna, S., 2012. UPF1 involvement in nuclear functions. *Biochemical Society transactions*, 40(4), pp.778–83.
- Vessey, J.P. et al., 2008. A loss of function allele for murine Staufen1 leads to impairment of dendritic Staufen1-RNP delivery and dendritic spine morphogenesis. *Proceedings of the National Academy of Sciences*, 105(42), pp.16374–16379.
- Waga, S. & Stillman, B., 1998. Cyclin-dependent kinase inhibitor p21 modulates the DNA primer-template recognition complex. *Molecular and cellular biology*, 18(7), pp.4177–87.
- Wagner, S.A. et al., 2016. ATR inhibition rewires cellular signaling networks induced by replication stress. *Proteomics*, 16(3), pp.402–416.
- Watt, P.M. & Hickson, I.D., 1994. Structure and function of type II DNA topoisomerases. *Biochemical Journal*, 303(Pt 3), pp.681–695.
- Weng, Y., Czaplinski, K. & Peltz, S.W., 1998. ATP is a cofactor of the Upf1 protein that modulates its translation termination and RNA binding activities. *RNA (New York, N.Y.)*,

4(2), pp.205–214.

Weng, Y., Czapinski, K. & Peltz, S.W., 1996a. Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Molecular and cellular biology*, 16(10), pp.5477–90.

Weng, Y., Czapinski, K. & Peltz, S.W., 1996b. Identification and characterization of mutations in the UPF1 gene that affect nonsense suppression and the formation of the Upf protein complex but not mRNA turnover. *Molecular and cellular biology*, 16(10), pp.5491–5506.

Wolters, D.A., Washburn, M.P. & Yates, J.R., 2001. An automated multidimensional protein identification technology for shotgun proteomics. *Analytical Chemistry*, 73(23), pp.5683–5690.

Xiang, J. et al., 2014. Levels of human replication factor C4, a clamp loader, correlate with tumor progression and predict the prognosis for colorectal cancer. *Journal of translational medicine*, 12, p.320.

Yamashita, A. et al., 2001. Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes and Development*, 15(17), pp.2215–2228.

Yamashita, A., Kashima, I. & Ohno, S., 2006. *Role of SMG-1-Mediated Phosphorylation of Upf1 in NMD.*

Yao, N.Y. & Donnell, M.O., 2012. The RFC clamp loader: Structure and Function. *Subcellular Biochemistry*, 62, pp.259–279.

Yeung, Y. et al., 2008. Removal of detergents from protein digests for mass spectrometry analysis. *Analytical Biochemistry*, 382(2), pp.135–137.

Yoshizawa-Sugata, N. & Masai, H., 2009. Roles of human AND-1 in chromosome transactions in S phase. *Journal of Biological Chemistry*, 284(31), pp.20718–20728.

Yu, X. et al., 2017. Genome-wide TOP2A DNA cleavage is biased toward translocated and

- highly transcribed loci. *Genome Research*, 27(7), pp.1238–1249.
- Yuzhakov, A. et al., 1999. Multiple competition reactions for RPA order the assembly of the DNA polymerase ?? holoenzyme. *EMBO Journal*, 18(21), pp.6189–6199.
- Zatloukal, B. et al., 2014. Sensitivity and Specificity of In situ Proximity Ligation for Protein Interaction Analysis in a Model of Steatohepatitis with Mallory-Denk Bodies. *PLoS ONE*, 9(5).
- Zemp, I. & Lingner, J., 2014. The shelterin component TPP1 is a binding partner and substrate for the deubiquitinating enzyme USP7. *The Journal of biological chemistry*, 289(41), pp.28595–606.
- Zhao, X., Li, G. & Liang, S., 2013. Several affinity tags commonly used in chromatographic purification. *Journal of Analytical Methods in Chemistry*, 2013(Table 1).
- Zhou, Y. et al., 2016. Role of WDHD1 in Human Papillomavirus-Mediated Oncogenesis Identified by Transcriptional Profiling of E7-Expressing Cells. *Journal of Virology*, 90(13), pp.6071–6084.
- Zhu, W. et al., 2007. Mcm10 and And-1 / CTF4 recruit DNA polymerase α to chromatin for initiation of DNA replication. *Genes & Development*, pp.2288–2299.