

# Changes in gene expression following DNA damage in a Non-Transformed Cell Line

Thesis submitted to the University of Leicester  
for the degree of Doctor of Philosophy

Julian Wang MSci (Hons)

University College London

June 2017

Medical Research Council Toxicology Unit

University of Leicester

# **Changes in gene expression following DNA damage in a Non-Transformed Cell Line**

**Julian Wang**, MRC Toxicology Unit, University of Leicester, Leicester, UK, LE1 9HN

Post-transcriptional regulation of gene expression forms an essential part of the DNA damage response (DDR). It has been shown previously that following DNA damage initiated by a number of different agents, selective translational reprogramming occurs, with specific subsets of mRNAs displaying enhanced mRNA translation, despite an overall inhibition of protein synthesis. In the past, these studies were typically conducted using transformed cell lines, which have been shown to differ significantly in their DDR when compared to non-transformed cells.

In this thesis, regulation of gene expression following DNA damage was investigated using the non-transformed MCF10A cell line through the use of the next-generation sequencing based technique, ribosome profiling. To this end, ionising radiation (IR) and ultraviolet radiation (UV-B) were tested as DNA damage sources, with UV-B proving to be more favourable. Using ribosome profiling, over 4000 significant differentially expressed genes were identified following UV-B irradiation, as well as a number of key pathways involved in this regulation.

The use of ribosome profiling however, also presented technical issues. A significant proportion of contaminated reads (corresponding to 5' tRNA halves) were found to be present in the sequencing data, which greatly reduced the effective read depth. Further investigations revealed that size exclusion columns used for the purification step of ribosome profiling selectively purified tRNA fragments (tRF). An alternative purification procedure using sucrose cushions resulted in reduced levels of tRF contamination. Another issue identified was sub-optimal RNase I digestion, a critical step in ribosome profiling. While a number of factors were investigated, the cause of the low level of RNase I digestion remained unresolved.

## Acknowledgements

I would like to express my sincere gratitude to my supervisor Professor Anne Willis for providing the PhD and for all the opportunities to attend conferences and training courses that came with it. Her support and unwavering enthusiasm has been a great source of inspiration throughout the project. I would also like to thank everyone in the Willis lab, past and present, for all the help, advice, friendships and laughs over the last four years. In particular, I would like to thank Dr Lindsay Wilson for teaching me new techniques at the start of my PhD; Dr Mark Stoneley for his guidance and encouragement throughout; Dr Tuija Pöyry and Dr Emilie Horvilleur for all their invaluable suggestions and support regarding my thesis; and Carolyn Jones, Dr Kate Dudek and Dr Ruth Spriggs for all their help with sequencing library preparations and subsequent analysis.

I would also like to acknowledge Dr Thomas Sbarrato, Dr Viktória Vágány, Dr Jana Alexandrova, Dr David Piñeiro Del Rio and Robert Harvey with whom, at one point or another, have all been subjected to my (many) excited ramblings regarding the latest gadgetries or technological marvels, and have always helped to brighten what would have otherwise been a disappointing day in the lab!

A special thanks goes to my grandparents, whom I was fortunate enough to have been raised by, and whom from a young age, inspired me to pursue my academic goals. Last, but not least, to my ever supportive parents who have always encouraged and believed in me, I would like to say a very profound thank you.

## Contents

1. Introduction .....	17
1.1 Eukaryotic translation .....	17
1.1.1 Canonical cap-dependent initiation .....	17
1.1.1.1 43S pre-initiation complex formation .....	17
1.1.1.2 Recruitment of the PIC to mRNA .....	18
1.1.1.3 Ribosome scanning of mRNA .....	18
1.1.1.4 Assembly of the 80s complex.....	19
1.1.2 Cap-independent initiation .....	19
1.1.3 Elongation .....	22
1.1.4 Termination.....	22
1.1.5 Regulation of translation initiation .....	24
1.1.5.1 Regulation by eIF2 $\alpha$ activity and upstream open reading frames .....	24
1.1.5.2 Regulation by eIF4E activity .....	25
1.1.6 Regulation of translation elongation .....	27
1.2 DNA damage .....	27
1.2.1 Ionising radiation induced DNA damage.....	27
1.2.2 Ultraviolet light induced DNA damage .....	28
1.3 DNA damage response.....	30
1.4 DNA damage repair pathways .....	31
1.4.1 Double strand break (DSB) repair .....	31
1.4.2 Base excision repair.....	36
1.4.3 Nucleotide excision repair .....	36
1.5 DNA damage and control of protein synthesis .....	39
1.6 Techniques for studying genome wide changes in translation.....	40
1.7 Project aims.....	40
2. Materials and Methods.....	42

2.1 Cell culture techniques.....	42
2.1.1 Growth and freeze media .....	42
2.1.2 Source and maintenance of cell lines .....	42
2.1.3 Cryopreservation of cell lines.....	43
2.1.4 Thawing frozen cell stocks .....	43
2.1.5 Clonogenic assay of cells in vitro.....	43
2.2 Cell treatment and induction of DNA damage.....	44
2.2.1 Treatment of cells with ionising radiation (IR).....	44
2.2.2 Treatment of cells with ultraviolet radiation .....	44
2.3 Protein techniques.....	44
2.3.1 Buffers and solutions .....	44
2.3.2 Antibodies and dilutions used.....	46
2.3.3 Cell lysate sample preparation.....	46
2.3.4 Protein quantification .....	46
2.3.4.1 Pierce BCA protein assay.....	47
2.3.4.2 Bradford assay.....	47
2.3.5 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) .....	47
2.3.6 Transfer of proteins onto PVDF membrane.....	48
2.3.6.1 Wet transfer method .....	48
2.3.6.2 Semi-dry transfer method.....	48
2.3.7 Immunodetection of proteins.....	49
2.3.8 Stripping Western blot membranes.....	49
2.3.9 <sup>35</sup> S Methionine incorporation for measuring protein synthesis rates.....	49
2.4 RNA techniques.....	50
2.4.1 Buffers and solutions .....	50
2.4.2 Polysome profiling.....	51
2.4.2.1 Preparing sucrose gradients.....	51

2.4.2.2 Harvesting cells .....	51
2.4.2.3 RNA precipitation and purification .....	52
2.4.3 Ribosome profiling .....	52
2.4.3.1 Preparing sucrose gradients.....	52
2.4.3.2 Harvesting cells and digestion.....	52
2.4.3.3 Centrifugation, fractionation and RNA extraction .....	53
2.4.3.4 rRNA depletion.....	54
2.4.3.5 Size selection.....	55
2.4.3.6 Library preparation .....	56
2.4.3.7 Sequencing.....	56
2.4.3.8 Bioinformatics analysis of sequencing data .....	57
2.4.4 Northern blot analysis.....	62
2.4.4.1 Preparation of samples .....	62
2.4.4.2 Gel electrophoresis separation.....	62
2.4.4.3 Transfer and crosslinking .....	62
2.4.4.4 Radioactive probe incorporation and hybridisation .....	63
2.4.4.5 Membrane washing and exposing .....	63
2.4.5 Northern blot analysis for small RNA.....	63
2.4.5.1 Preparation of samples .....	63
2.4.5.2 End-labelling ladders.....	64
2.4.5.3 Gel electrophoresis separation .....	64
2.4.5.4 Transfer and crosslinking .....	64
2.4.5.5 Radioactive probe incorporation and hybridisation .....	65
2.4.5.6 Membrane washing and exposing .....	65
2.4.5.7 Probes used.....	65
2.4.6 RT-qPCR using SYBRGreen .....	65
3. Selection and optimisation of DNA damage model .....	68

3.1 Introduction .....	68
3.2 Establishment of IR treatment conditions .....	69
3.3 Exposure to IR induces rapid induction of DNA damage and repair.....	69
3.4 Global protein synthesis rates were inhibited following IR exposure .....	72
3.5 Changes in subpolysome/polysome distribution following IR exposure .....	72
3.6 Establishment of UV-B treatment conditions .....	77
3.7 Global protein synthesis rates are inhibited following UV-B exposure .....	77
3.8 DNA damage and translational responses to UV-B .....	81
3.9 Substantial shifts in polysome profiles following UV-B exposure.....	81
3.10 Early translational reprogramming requires phosphorylation of eIF2 $\alpha$ .....	85
3.11 Discussion.....	88
4. Investigation of gene expression changes using ribosome profiling .....	91
4.1 Introduction .....	91
4.2 Identifying translational changes using ribosome profiling.....	92
4.2.1 Optimisation of ribosome profiling protocol steps.....	92
4.3 RNA sequencing using illumina technology .....	96
4.3.1 Library preparation .....	96
4.3.2 Cluster generation.....	96
4.3.3 Sequencing .....	97
4.3.4 Alignment and data analysis .....	97
4.4 Analysis of ribosome profiling data.....	98
4.4.1 Pre-processing of sequencing data .....	98
4.4.2 Overview of mapping statistics .....	98
4.4.3 Significant number of contaminating reads.....	103
4.4.4 Transcriptome analysis of ribosome profiling data .....	104
4.4.4.1 Identifying differentially expressed genes using whole transcriptome dataset .....	104

4.4.4.2	Tight inter-replicate correlation in sequencing datasets.....	104
4.4.4.3	Principal component analysis reveals tight clustering of samples based on treatment conditions .....	107
4.4.4.4	UV-B induced upregulation of specific pathways .....	109
4.4.4.5	qPCR validation of DESeq2 analysis outputs.....	116
4.4.5	Translatome analysis.....	118
4.4.5.1	Mapping of sequencing reads reveals distribution of fragments.....	118
4.4.5.2	Low numbers of genes identified to be translationally significantly different between treatment conditions .....	123
4.4.5.3	Insufficient read depth resulting in problematic analysis.....	124
4.5	Discussion.....	128
5.	Identification of the sources of contamination in the ribosome profiling sequencing data	130
5.1	Introduction .....	130
5.2	Sources of contaminating tRNA reads .....	131
5.2.1	Identification of primary contaminating reads .....	131
5.2.2	Basal levels of tRFs in MCF10A cells.....	131
5.2.3	Enrichment of tRNA halves by size exclusion spin columns .....	136
5.2.4	Reduced levels of tRNA halves contamination using sucrose cushions.....	137
5.2.5	Sucrose cushion purified samples reveals problems with RNase I digestion .....	138
5.3	Troubleshooting and optimisation of RNase I digestion.....	142
5.3.1	Repeat run of RNase I titration reveals digestion issues .....	142
5.3.2	Unknown causes for the changes observed in RNase I digestion.....	144
5.3.3	Improved digestion using micrococcal nuclease .....	150
5.4	Discussion.....	152
6.	Summary .....	154
7.	Appendix .....	159
7.1	Adapter sequences.....	159
7.1.1	TruSeq Small RNA library preparation adapter sequences.....	159

7.1.2 TruSeq Stranded mRNA LT library preparation adapter sequences .....	160
7.2 Scripts used for bioinformatics analysis.....	162
7.2.1 Removal of contaminating reads using Bowtie2 .....	162
7.2.2 Alignment to reference genome using TopHat.....	164
7.2.3 Counting reads .....	165
7.2.4 Translatome analysis using R package babel .....	166
7.2.5 Transcriptome analysis using R package DESeq2.....	172
7.2.6 Custom R functions used.....	175
8. References.....	181

## List of Figures

Figure 1.1   Overview of translation initiation .....	21
Figure 1.2   Overview of translation elongation.....	23
Figure 1.3   Translational regulation by uORFs .....	26
Figure 1.4   Overview of DNA Damage Response Signalling .....	33
Figure 1.5   Schematic overview of non-homologous end-joining.....	34
Figure 1.6   Schematic overview of homologous recombination.....	35
Figure 1.7   Schematic overview of nucleotide excision repair.....	38
Figure 2.1   illumina RNA-seq overview .....	60
Figure 3.1   Dose response to ionising radiation exposure .....	70
Figure 3.2   Rapid induction of DNA damage and repair following IR exposure.....	71
Figure 3.3   Changes in protein synthesis rates following IR-induced DNA damage .....	74
Figure 3.4   Separation of mRNAs by sucrose density gradients.....	75
Figure 3.5   Changes in polysome profiles of cells following exposure to IR .....	76
Figure 3.6   Dose response to UV-B irradiation.....	79
Figure 3.7   Global protein synthesis rates inhibited following UV-B irradiation.....	80
Figure 3.8   DNA damage and translational responses to UV-B.....	83
Figure 3.9   Substantial shifts in the polysome profiles of cells following irradiation with UV-B .....	84
Figure 3.10   Early translational regulation following UV-B was mediated by eIF2 $\alpha$ .....	87
Figure 4.1   Schematic diagram of the ribosome profiling protocol .....	94

Figure 4.2   Optimisation of RNase I digestion .....	95
Figure 4.3   Sequence quality post adapter trimming.....	100
Figure 4.4   Overall individual mapping statistics for sequencing runs.....	101
Figure 4.5   Averaged overview of mapping statistics for sequencing runs.....	102
Figure 4.6   Strong correlation between all three biological repeats for both the RPF and whole transcriptome datasets .....	106
Figure 4.7   Principle component analysis of whole transcriptome dataset at 1 and 4 hours.	108
Figure 4.8   MA plot of transcriptome analysis outputs.....	111
Figure 4.9   Differentially expressed genes following UV-B treatment.....	112
Figure 4.10   UVB induced changes in pathways identified using IPA at 1 hour.....	113
Figure 4.11   UVB induced changes in pathways identified using IPA at 4 hours (Part 1) .....	114
Figure 4.12   qPCR validation of DESeq2 analysis outputs .....	117
Figure 4.13   Periodicity mapping of globally averaged ribosome profiling data .....	120
Figure 4.14   Periodicity mapping of ribosome profiling data averaged across conditions.....	121
Figure 4.15   Extended periodicity mapping reveals diminished numbers of reads mapped to the start of CDS.....	122
Figure 4.16   Correlation between RPF and whole transcriptome read counts.....	126
Figure 4.17   Babel translome analysis top differentially translated genes output.....	127
Figure 5.1   Detailed overview of the over-represented sequences in RPF dataset .....	133
Figure 5.2   Distribution of over-represented sequences in RPF dataset .....	134
Figure 5.3   Low levels of 5' tRNA halves present in both control and UV-B treated samples	135
Figure 5.4   Purification of RPFs using size exclusion columns results in the enrichment of tRF5s .....	139
Figure 5.5   Reduced levels of tRF5 enrichment with sucrose cushions .....	140
Figure 5.6   Polysome profiling of sucrose purified samples reveals problems with RNase I digestion.....	141
Figure 5.7   RNase I titrations reveals issues with digestion .....	143
Figure 5.8   Impact of cell seeding density on RNase I digestion .....	146
Figure 5.9   Impact of multiple factors on RNase I digestion (Part 1) .....	147
Figure 5.10   Impact of multiple factors on RNase I digestion (Part 2) .....	148
Figure 5.11   Impact of cell type and RNase inhibitors on RNase I digestion.....	149
Figure 5.12   Improved digestions using micrococcal nuclease .....	151

## List of Tables

Table 2.1   Reagents and volumes used for cell culture growth and freeze media. ....	42
Table 2.2   Antibodies used for Western blotting .....	46
Table 2.3   Gel composition for SDS-PAGE (for 1x 0.75 mm gel). ....	48
Table 2.4   Northern probes used .....	65
Table 2.5   RT 1x reaction mix .....	66
Table 2.6   Probes used for RT-qPCR.....	67

## Abbreviations

4E-BP1	Eukaryotic translation initiation factor 4E-binding protein 1
6-4PPs	Pyrimidine (6-4) pyrimidone (6-4PPs) photoproducts
AMP	Adenosine monophosphate
AP site	apurinic/apyrimidinic site
APS	Ammonium persulphate
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
BCP	1-bromo-3-chloropropane
BER	Base excision repair
BSA	Bovine Serum Albumin
CDS	Coding sequence
Chk1	Checkpoint kinase 1
Chk2	Checkpoint kinase 2
CHX	Cycloheximide
CPD	Cyclobutane pyrimidine dimers
CrPV	Cricket paralysis virus
CS	Cockayne's syndrome
CSA	Cockayne syndrome type A/Excision Repair Cross-Complementation Group 8
CSB	Cockayne syndrome type B/Excision Repair Cross-Complementation Group 6
CT	Chloera toxin

dCTP	deoxycytidine triphosphate
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DSB	Double-strand DNA breaks
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
eIF	eukaryotic initiation factor
eIF2 $\alpha$	Eukaryotic translation initiation factor 2 alpha subunit
eIF4E	Eukaryotic translation initiation factor 4E
eRF	eukaryotic release factor
FBS	Fetal Bovine Serum
GCN2	General control non-derepressible-2 kinase
GDP	Guanosine diphosphate
GEF	Guanine nucleotide exchange factor
GG-NER	Global genome nucleotide excision repair
GTP	Guanosine triphosphate
Gy	Gray (unit of ionising radiation dose)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HR	Homologous recombination

HRI	Heme-regulated inhibitor kinase
IR	Ionising radiation
IRES	Internal Ribosome Entry Sites
MMR	Mismatch repair
mRNA	Messenger RNA
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PARP	Poly (ADP-ribose) polymerase
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PERK	PKR-like endoplasmic reticulum (ER) kinase
PIC	Pre-initiation complex
PKR	Protein kinase R
PNK	Polynucleotide kinase
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPA	Replication factor-C
RPA	Replication Protein A

RPC	Replication factor-A
RPF	Ribosome protected fragment
RPKM	Reads per kilobase of transcript per million mapped reads
rpM	Reads per million
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription polymerase chain reaction
SBS	Sequencing by synthesis
SDS	Sodium dodecyl sulphate
snoRNA	Small nucleolar RNA
SSB	Single-strand DNA breaks
SSC	Saline-sodium citrate buffer
TC	Ternary complex
TCA	Trichloroacetic acid
TC-NER	Transcriptional coupled nucleotide excision repair
TEMED	Tetramethylethylenediamine
TFIIH	Transcription factor-IIH
tRF	tRNA-derived RNA fragments
tRF3	3' tRNA halves
tRF5	5' tRNA halves
tRNA	Transfer RNA
tRNA	Transfer RNA

uORF	upstream open reading frame
UTR	Untranslated region
UV	Ultraviolet radiation
WT	Whole transcriptome
XP	Xeroderma Pigmentosum
XPA	Xeroderma Pigmentosum, Complementation Group A
XPB	Xeroderma Pigmentosum, Complementation Group B

## ***1. Introduction***

Gene expression is highly regulated to allow cells to respond rapidly to maintain homeostasis and in addition following conditions of cell severe stress such as heat shock, or exposure to agents that damage DNA, allow the production of proteins that are necessary for survival. Gene expression is regulated at all levels, including of transcription of genomic DNA to pre-mRNA, splicing of pre-mRNA to form mature mRNA (Weake & Workman, 2010), export and localisation to the cytoplasm (Keene, 2010), and translation (Sonenberg & Hinnebusch, 2009). Of these, regulation at the level of translation allows for a rapid response in the proteins synthesised, without the delays associated with the transcription and processing of a new mRNA (Sonenberg & Hinnebusch, 2009). It is perhaps unsurprising then that translational control in eukaryotic cells plays a key role in the response to cellular stress.

### ***1.1 Eukaryotic translation***

Protein synthesis is an energetically expensive and complex process, consuming the majority of the energy used within cells (Buttgereit, F; Brand, 1995), and as a result is subject to tight regulation. Translation can be divided into three phases: initiation, elongation and termination. Of these, it is the initiation phase that is thought to be the rate limiting step and is subject to the majority of regulation (Sonenberg & Hinnebusch, 2009), although recent data also show that the elongation stage is a major regulatory node (Richter & Coller, 2015).

#### ***1.1.1 Canonical cap-dependent initiation***

Translation initiation is the process that results in the formation of an elongation-competent 80S ribosome at the initiation codon, achieved with the aid of a number of eukaryotic initiation factors (eIF), and methionyl transfer RNA specialised for initiation (Met-tRNA<sub>i</sub>). The major steps involved include the formation of the 48S initiation complex at the initiation codon, followed by the subsequent joining of the 60S ribosomal subunit. The formation of the 48S complex on most mRNA is primarily formed by the actions of a scanning 43S pre-initiation complex (PIC). This process is represented schematically in Figure 1.1, and described in detail below.

##### ***1.1.1.1 43S pre-initiation complex formation***

The 43S PIC is comprised of the ternary complex (TC) and the 40S ribosomal subunit along with a number of eIFs. The TC consists of the heterotrimer eIF2, GTP and methionyl transfer RNA specialised for initiation (Met-tRNA<sub>i</sub>). The 43S pre-initiation complex (PIC) is then subsequently formed by the recruitment of the TC to the small 40S ribosomal subunit with the aid of eIFs 1,

1A, 3 and 5 (Asano et al., 2001; Kolupaeva, Unbehaun, Lomakin, Hellen, & Pestova, 2005; Majumdar, Bandyopadhyay, & Maitra, 2003).

Loading of Met-tRNA<sub>i</sub> onto the small 40S ribosomal subunit occurs through the action of eIF2, which is composed of an  $\alpha$  and  $\beta$  subunits which are bound to a central  $\gamma$  subunit (to which both GTP and tRNA bind to) (Kolitz & Lorsch, 2010).

#### *1.1.1.2 Recruitment of the PIC to mRNA*

The mRNA is prepared for its recruitment to the 43S PIC by PABP and the eIF4F complex, which contains the DEAD-box helicase eIF4A, eIF4B, scaffolding protein eIF4G and the cap-binding factor eIF4E. The binding of eIF4E to eIF4G causes an increase in the affinity of eIF4E to the 7-methyl-guanosine (m<sup>7</sup>G) cap end on the mRNA (Gross et al., 2003), as well as a conformational change which helps to align eIF4A in the orientation required for ATP-dependent helicase activity. The combination of the recognition of the 5' cap structure by eIF4E, and RNA binding sites in eIF4G results in the anchoring of eIF4A to the 5' untranslated region (UTR) of the mRNA. This results in the unwinding of structures in the 5' UTR by eIF4A. The activity of eIF4A is stimulated by accessory proteins, which include eIF4B and eIF4H. Polyadenylate-binding protein (PABP) also binds with the eIF4G scaffolding protein as well as the 3' poly(A) tail of the mRNA, which along with the interaction between eIF4E and eIF4G, causes the formation of a mRNA circular loop (Figure 1.1), preparing the mRNA for recruitment of the PIC (Aitken & Lorsch, 2012). In addition, it has been shown that eIF4B also interacts with PABP further aiding the stability of the complex (Bushell et al., 2001). The 43S PIC is then recruited to the m<sup>7</sup>G cap proximal region on the 5' end of mRNA through the interaction of eIF3 with eIF4G (LeFebvre et al., 2006).

#### *1.1.1.3 Ribosome scanning of mRNA*

Following recruitment to the mRNA, the PIC moves along the mRNA strand in a 5' to 3' direction until the start codon is located using a process termed scanning. In the current model, this process occurs through the binding of the PIC to the mRNA in an "open" conformation, induced by eIF1 and eIF1A and thought to be stabilised by eIF3 (Nanda, Saini, Muñoz, Hinnebusch, & Lorsch, 2013). Scanning is also facilitated by the unwinding of the structured mRNA through eIF4A activity and through non-cognate pairing between the mRNA and the anti-codon on Met-tRNA<sub>i</sub>. Studies have also shown the mammalian DExH-box protein DHX29 to be involved in the scanning of highly structured 5' UTRs, with depletion of the protein causing polysome dissociation and inhibition of translation initiation (Sonenberg & Hinnebusch, 2009).

Once the AUG start codon in the Kozak consensus sequence is encountered in the ribosomal P-site, base pairing with the anti-codon of Met-tRNA<sub>i</sub> occurs (Marilyn Kozak, 1987), triggering a series of events including the hydrolysis of the eIF2-bound GTP via the action of the GTPase-activating protein eIF5, as well as the dissociation of eIF1 from the PIC (Maag, Fekete, Gryczynski, & Lorsch, 2005). More recent studies have also suggested a role for eIF5 in AUG recognition via its enhancement of eIF1 dissociation from the PIC (Nanda et al., 2009, 2013). It is important to note however that both the position of the AUG relative to the 5' end of the mRNA, as well as the context in which it is present in are important factors in determining translation. This optimal sequence of nucleotides surround the AUG, the Kozak consensus sequence, is defined as GCC(A/G)CCAUGG, with a purine at the -3 and a G at the +4 position relative to the A in the AUG codon (which is designated as +1) (Marilyn Kozak, 1987). However translation initiation also occurs from both non-AUG codons, and start codons in weak context. As the PIC scans mRNA sequentially, the position of the AUG is an important factor in initiation, with ribosomes potentially initiating at the first AUG despite a weaker context. The presence of leaky scanning, whereby the PIC scans past an AUG and instead reinitiating further downstream, has been also shown in cases where by the start codon is not in a good context (M. Kozak, 1991).

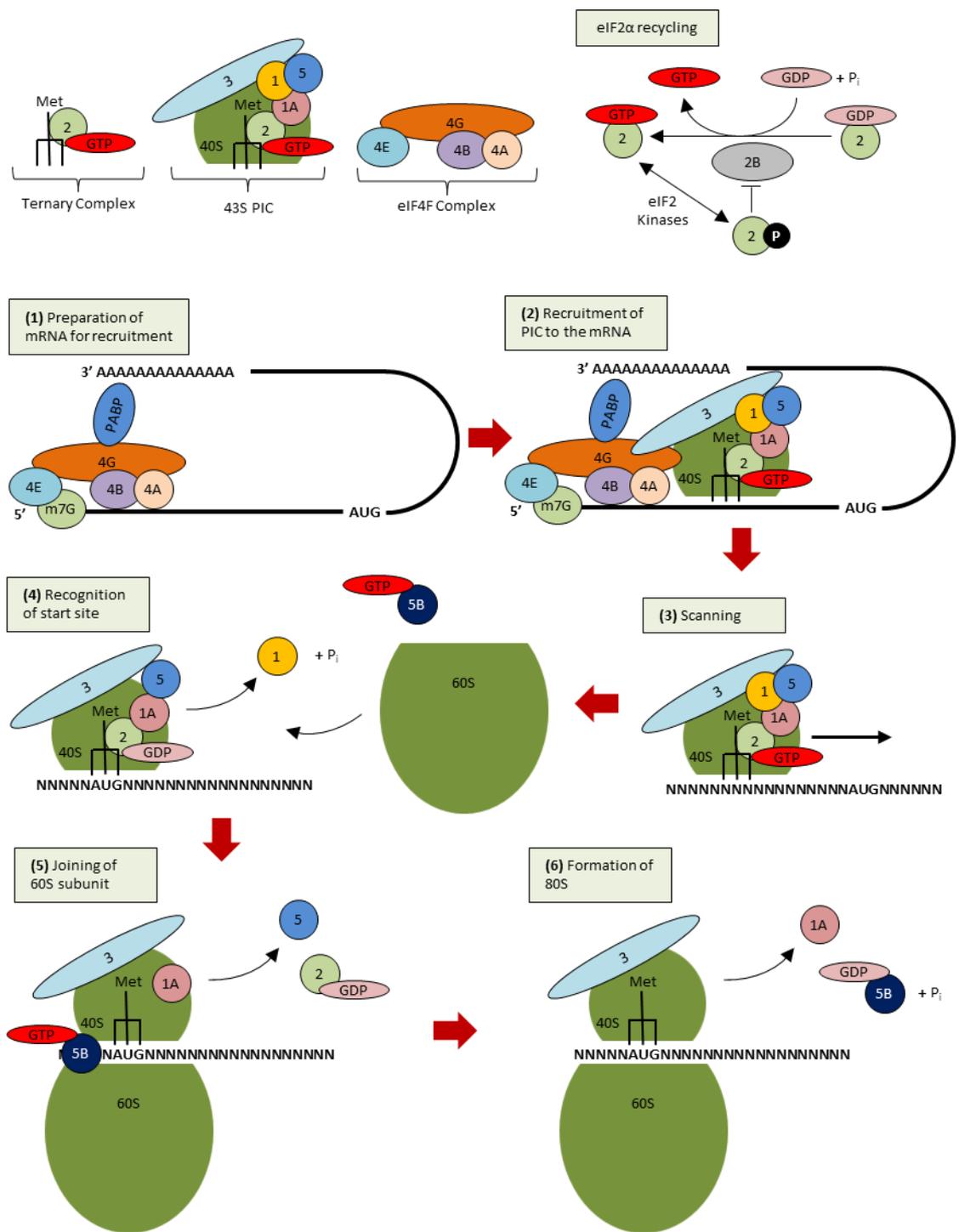
#### *1.1.1.4 Assembly of the 80s complex*

Following the dissociation of eIF1 from the PIC, eIF2-GDP and eIF5 are also released, allowing for the eIF5B mediated 60S ribosomal subunit joining. The interaction between the C-terminal end of both eIF5B and eIF1A is required for the efficient subunit joining, with the C-terminal end of eIF1A only becoming accessible following its displacement from the ribosomal P-site upon start codon recognition (Yu et al., 2009). Upon joining, the GTPase eIF5B hydrolyses GTP, resulting in the formation of the 80S initiation complex and the dissociation of eIF5B and eIF1A (Aitken & Lorsch, 2012).

#### **1.1.2 Cap-independent initiation**

While the majority of translation initiation occurs in the cap-dependent manner described previously, in certain cases, most notably in viruses, the conventional scanning PIC mechanism as well as a number of eIFs can be bypassed by the use of internal ribosome entry sites (IRES). In the case of picornaviruses, the presence of an IRES negates the need for eIF4E for the recruitment of the 40S ribosomal subunit, though in many cases, still requires all other eIFs (Doudna & Sarnow, 2007). Remarkably, in the case of cricket paralysis virus (CrPV), the presence of an IRES eliminates the need for any eIFs, including Met-tRNA<sub>i</sub>, with instead IRES elements performing the functions of the eIFs (Doudna & Sarnow, 2007). IRES have also been identified

in mammalian mRNAs and the data suggest that up to 10% of all messages have the potential to be regulated in this way (Spriggs, Bushell, & Willis, 2010; Weingarten-Gabbay et al., 2016). Cellular IRESs are used to maintain expression of key proteins under conditions of cell stress including apoptosis and heatshock (Spriggs et al., 2010), but more recently they have also been shown that they are required during development (Xue et al., 2014). Indeed, many of the IRESs identified as active during apoptosis (Bushell et al., 2006) may also have a role in development. While there appears to be little commonality between cellular and viral IRESs and in addition even between viral IRES, most are highly structured and display an extraordinary diversity of the mechanisms by which they function (Elroy-Stein & Merrick, 2007; Kieft, 2009; Spriggs et al., 2010).



**Figure 1.1 | Overview of translation initiation**

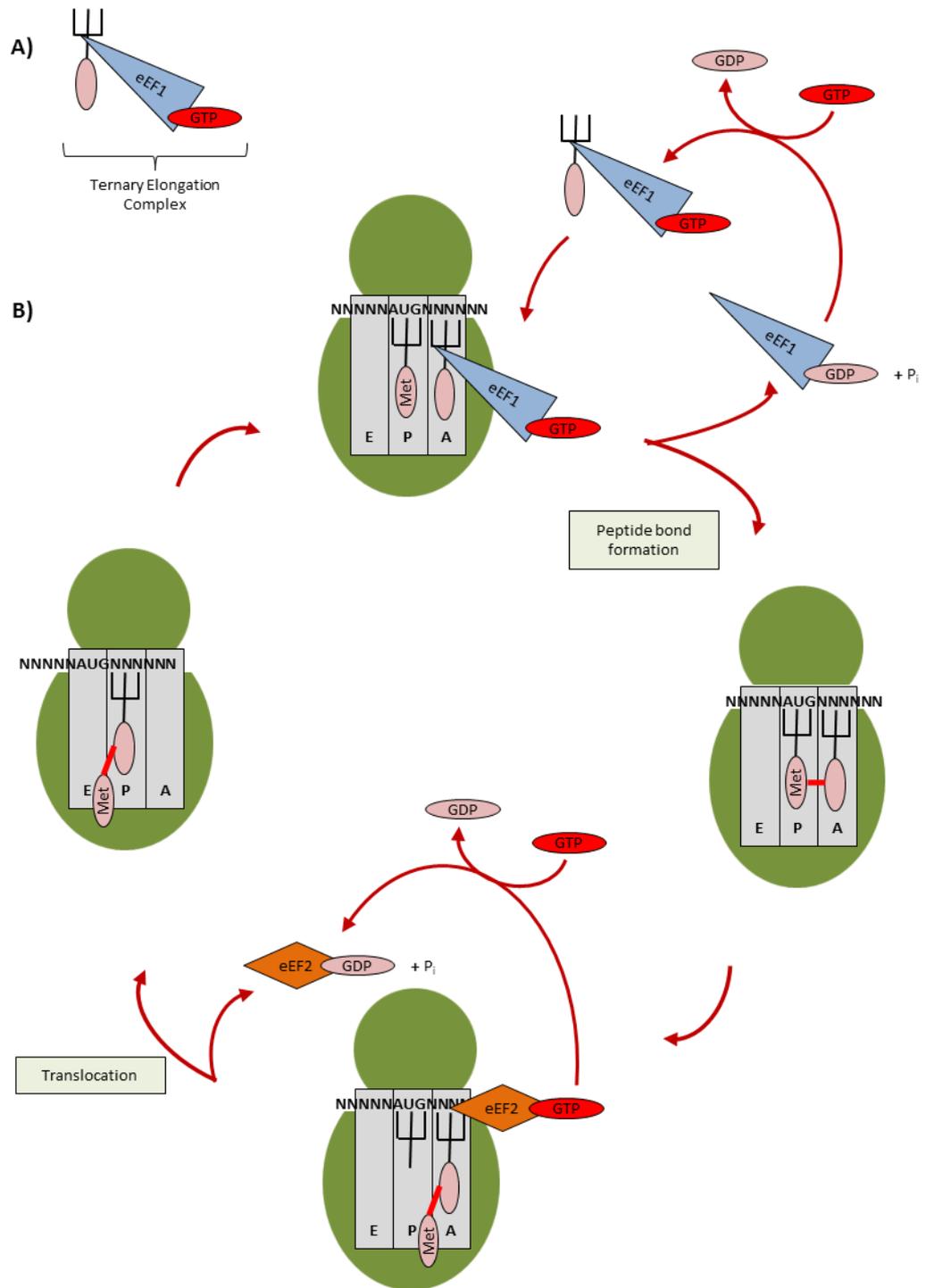
Schematic overview of translation initiation in eukaryotic cells, as described in Section 1.1.1. For clarity, all names have been abbreviated, with objects marked as a number (ie, 1, 1A, 2, 3, 5) referring to the eukaryotic initiation factor (ie eIF1, eIF1A, eIF2). Similarly, 43S PIC refers to the 43S pre-initiation complex.

### ***1.1.3 Elongation***

Elongation is the central process of translation where a polypeptide chain is formed from the mRNA strand, and is composed of three principle steps; the recruitment of a cognate aminoacylated tRNAs (aatRNAs), peptide bond formation, and translocation of the ribosome to the next codon. By the end of the initiation stage of translation, Met-tRNA<sub>i</sub> is located in the P-site, with both the E-site and A-site being unoccupied. With the A-site empty, both cognate and non-cognate aatRNAs can enter this site as a ternary elongation complex, formed with GTP, aatRNAs and the elongation factor 1A (eEF1A) (Figure 1.2A), however it is not until a cognate codon-anticodon reaction occurs that ribosomal conformational changes are induced (Figure 1.2). These stabilise the binding of the tRNA and causes the hydrolysis of GTP, leading to the dissociation of the aminoacylated end of the tRNA by eEF1A. Ribosomal peptidyl transferase then catalyses the formation of a peptide bond, causing deacetylation of the tRNA in the P-site, which results in the transfer of the peptide chain to the tRNA in the A-site. Following this elongation step, the deacetylated tRNA (currently Met-tRNA<sub>i</sub>) is moved to the E-site whilst the peptidyl tRNA is moved to the P-site in a process termed translocation by the action of eEF2, leaving the A-site free to accept the next ternary elongation complex and repeat the elongation cycle.

### ***1.1.4 Termination***

Termination of the elongation cycle occurs when a stop codon, consisting of UAG, UAA or UGA is reached. This is recognised by the eukaryotic release factor 1 (eRF1), which then enters the A-site of the ribosome, leading to the release of the polypeptide chain through the hydrolysis of the ester bond linking the chain to the tRNA in the P-site. The activity of eRF1 in turn, is stimulated by the eukaryotic release factor 3 (eRF3) (Cheng et al., 2009; Dever & Green, 2012).



**Figure 1.2 | Overview of translation elongation**

**A)** The ternary elongation complex is formed of aminoacylated tRNA (aatRNAs), GTP and the elongation factor 1A (eEF1A). These complexes can then enter the ribosomal A-site, however, it is not until a cognate aatRNAs enters the site that conformational changes are induced and elongation proceeds as described in Section 1.1.3. **B)** Schematic overview of translation elongation in eukaryotic cells, showing incorporation of cognate aatRNA into the ribosomal A-site, formation of peptide bond and subsequent translocation, as described in Section 1.1.3.

### ***1.1.5 Regulation of translation initiation***

There are two major pathways involved in the rapid global inhibition of translation at the level of translation, one is the control of active TC availability by the phosphorylation of eIF2 $\alpha$ , and other being regulation of eIF4E availability by the actions of a family of 4E binding proteins (4E-BPs).

#### ***1.1.5.1 Regulation by eIF2 $\alpha$ activity and upstream open reading frames***

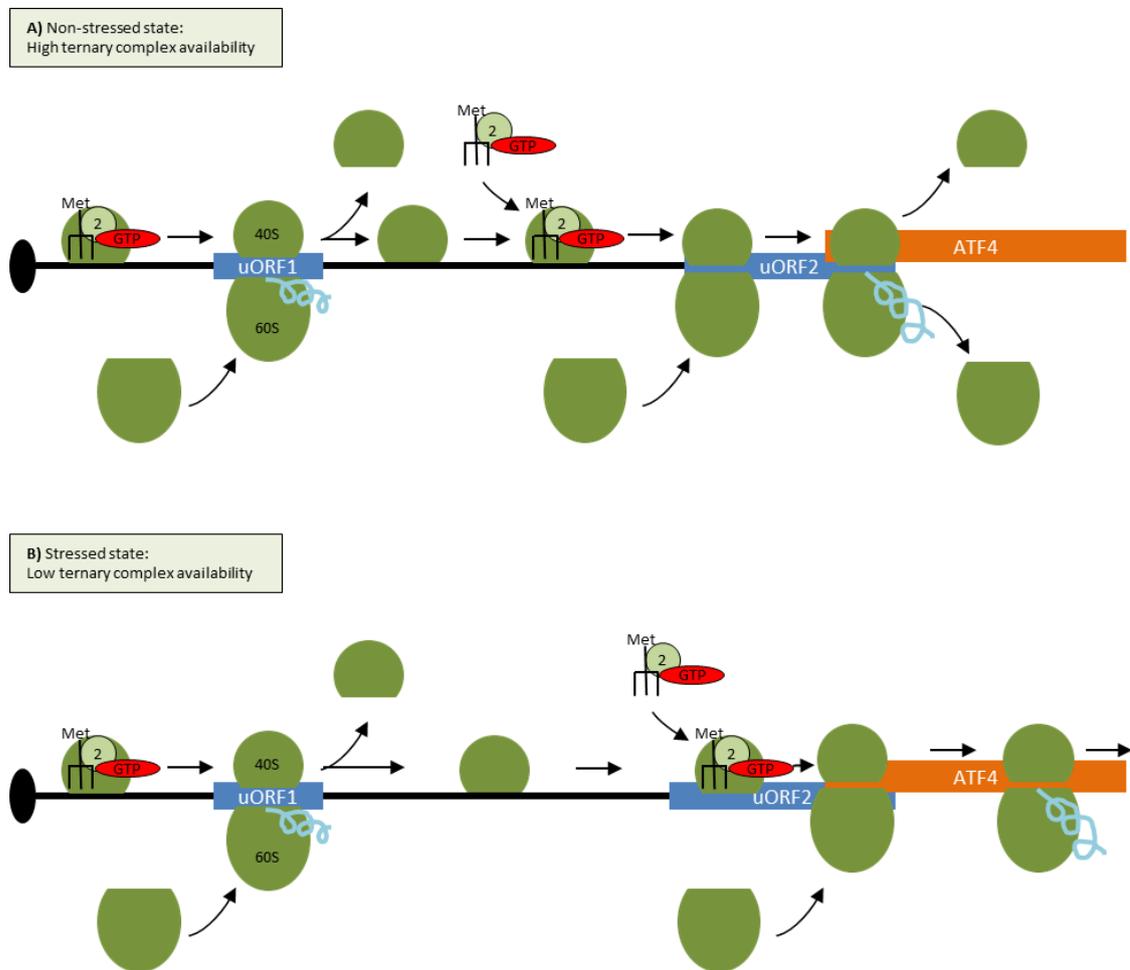
A crucial target of regulation at initiation is the recycling of eIF2-GDP to eIF2-GTP by the GDP-GTP exchange factor eIF2B (Figure 1.1) (Sonenberg & Hinnebusch, 2009), as Met-tRNA<sup>i</sup> has a higher affinity for binding to eIF2-GTP than to eIF2-GDP (Kapp & Lorsch, 2004). However, when eIF2 $\alpha$  is phosphorylated on serine 51, it acts as a competitive inhibitor of eIF2B and so results in a reduction in the efficiency of the GDP-GTP exchange and a reduced rate of TC formation, in turn reducing the rate of PIC formation and thereby overall translation initiation rates (Figure 1.1). The phosphorylation of eIF2 $\alpha$  in mammalian cells is mediated by a diverse family of four kinases which are activated in general following specific stresses although some of their functions are overlapping; these are heme-regulated inhibitor (HRI) activated in response to heme deficiency, protein kinase R (PKR) activated in response to interferon and double-stranded RNA (Williams, 1999), PKR-like endoplasmic reticulum (ER) kinase (PERK) in response to ER stress (Ron, 2002), and general control non-derepressible-2 (GCN2) in response to amino acid starvation or UV irradiation (Deng et al., 2002).

Phosphorylation of eIF2 $\alpha$  results in the reduction of global protein synthesis due to its impact on TC formation, however, paradoxically eIF2 $\alpha$  phosphorylation also causes up-regulation of a specific subset of transcripts, which contain multiple upstream open reading frames (uORFs). Many of these mRNAs encode for stress response genes, resulting in the selective up-regulation of a number of key DNA damage and repair genes as part of the integrated stress response. The classic example of one such mRNA is mammalian activating transcription factor 4 (ATF4), which contains two uORFs, with the second uORF overlapping with the main coding region (Figure 1.3). Under conditions of high eIF2-TC availability, the 40S ribosomal subunits that resume scanning following termination after the uORF1 can reinitiate at the AUG of the uORF2 if the 40S acquires new eIF2-TC before reaching the uORF2 start codon. The resulting translation of uORF2 then prevents initiation at the start of the main coding region (Figure 1.3A). Under stress conditions, however, when the available eIF2-TC is low, there is a reduced probability of reinitiation at the start of uORF2 and instead, it is more likely that the scanning 40S ribosome would continue beyond uORF2 and reinitiate at the AUG of the main coding region, thereby resulting in an

upregulation of ATF4 under these stress conditions (Figure 1.3B) (Vattem & Wek, 2004). Through this mechanism, a reduction in the abundance of available eIF2 can result in the selective up-regulation of those mRNAs which contain the appropriate uORF configuration (Ron & Harding, 2007).

#### *1.1.5.2 Regulation by eIF4E activity*

A second mechanism for the global regulation of protein synthesis involves eIF4E, which forms a critical component of the eIF4F complex, and is responsible for the recognition of the m<sup>7</sup>G cap, thereby playing an essential role in cap-dependent translation initiation. The availability and activity of eIF4E therefore, is a key target for translational regulation. There are a family of 4E binding proteins (4E-BPs) that inhibit the activity of eIF4E through the competitive binding to eIF4E, preventing eIF4E binding to eIF4G and therefore preventing eIF4F complex formation (Marcotrigiano, Gingras, Sonenberg, & Burley, 1999). The activity of 4E-BPs is regulated through phosphorylation of four main residues: threonine 46, threonine 37, threonine 70 and serine 65 (Gingras, Raught, Gygi, et al., 2001; Gingras, Raught, & Sonenberg, 2001). Hypophosphorylated 4E-BPs have a strong affinity to eIF4E, resulting in the inhibition of protein synthesis, while hyperphosphorylated 4E-BPs have weak affinity. The activity of 4E-BPs is primarily mediated through mTOR signalling, which in turn is controlled by the upstream phosphatidylinositol-3-kinase-like (PI3K)/Akt pathway, and was amongst the first discovered mTORC1 targets (Hara et al., 1997). Regulation of the phosphorylation states of 4E-BPs is determined by a number of external stimuli, which are known to activate the mTORC1 pathway such as hormones, growth factors and pH.



**Figure 1.3 | Translational regulation by uORFs**

Schematic overview showing the regulation of translation by uORFs in mRNA such as ATF4. With two uORFs present, one overlapping the primary ORF, differences in the availability of ternary complexes (TC) alters the rate at which ribosomes reinitiate and thereby alters the translation of the primary ORF. In the non-stressed state (A), high TC availability results in rapid reinitiation and thereby translation of uORF2. Whereas in the stressed state (B), the low TC availability results in higher probability of reinitiation further along the mRNA and thereby increased likelihood of translation of the primary ORF.

### ***1.1.6 Regulation of translation elongation***

tRNA- and eEF1A-dependent codon decoding, and eEF2-dependent ribosome translocation, are two processes which together determine the overall rate of translation elongation and a recent data suggest that these are both important to cell survival and tumorigenesis. For example, changes in cellular tRNA composition, particularly increased expression of limiting tRNAs, would affect elongation rates and has been shown that increased synthesis of specific proteins, particularly in a tumour cell environment, is directly related to tRNA content of the cells (Gingold et al., 2014).

As discussed above, eEF2 catalyses ribosome translocation by transferring the deacylated tRNA to the E site, which positions the peptidyl-tRNA in the P site and re-exposes the A site. The ability of eEF2 to modulate this step is controlled by its phosphorylation status. When phosphorylated by its specific kinase, eEF2K (a negative regulator) the interaction of eEF2 with the ribosome decreases and this slows the rate of translation. It has been shown that modulation of elongation is essential following heatshock, proteotoxic shock and nutrient deprivation, supporting a central role for the elongation process in the control of protein synthesis during cell stress (Richter & Coller, 2015). Interestingly, mRNA sequences within coding regions that cause ribosome pausing may contribute to message-specific elongation control (Richter & Coller, 2015). It has been shown that elongation control is important in disease development. For example, in an in vivo colon cancer using APC<sup>-/-</sup> mice there is increased elongation translation of selective mRNAs leading to an upregulation in the synthesis of the corresponding proteins e.g. cyclin D3 (Faller et al., 2014).

## ***1.2 DNA damage***

There is a diverse range of agents that elicit DNA damage, originating both externally and also endogenously. Common sources of DNA damaging agents include ionising radiation (IR), ultraviolet radiation (UV), reactive oxygen species (ROS) and a range chemical compounds which cause both double and single strand breaks, and inter-strand cross links. The investigations described in this thesis are focused on the study into gene expression changes following only IR and UV-B, and therefore, these specific sources are described in more detail below.

### ***1.2.1 Ionising radiation induced DNA damage***

IR encompasses both charged subatomic particles moving at relativistic speeds, such as alpha and beta radiation sources, and also part of the electromagnetic spectrum with high enough energy to cause ions through the release of electrons; this includes gamma rays and x-rays. IR

exposure occurs from multiple sources, such as cosmic radiation, radon gas, medical x-rays, high altitude flights or radiotherapy. Exposure to IR can cause a variety of DNA damage, both from direct effects or via indirect interactions with reactive species generated by the radiation. Amongst these, IR-induced double-strand DNA breaks (DSBs) are the most genotoxic (Mahaney, Meek, & Lees-Miller, 2009), but exposure to IR can also cause single-strand breaks (SSBs) and induce oxidation of DNA bases (Kao, Rosenstein, Peters, Milano, & Kron, 2005; Su, Meador, Geard, & Balajee, 2010).

It has also been reported previously that high doses of IR inhibit global protein synthesis rates both through the disruption of the cap initiation complex, which is required for the recruitment of the ribosomal subunits and subsequent initiation of translation (Paglin et al., 2005), and also causes the disassembly of cap-independent translation complexes such as the DAP5-dependent initiation complex which regulates IRES-dependent translation (Trivigno, Bornes, Huber, & Rudner, 2013). IR disrupts the cap complex by signalling through to 4E-BP. This leads to its dephosphorylation and as a consequence binding to and sequestering of eIF4E, which reduces the levels of cap initiation complex.

### ***1.2.2 Ultraviolet light induced DNA damage***

Ultraviolet light encompasses the spectrum of light with wavelengths ranging from 400nm to 10nm (as defined by ISO 21348), with this region being commonly sub-divided into three main discreet bands designated UV-A (400-315nm), UV-B (315-280nm) and UV-C (280-100nm). Of these, only UV-A and UV-B are generally taken as being of significance, as UV-C, and shorter classes such as extreme UV (121-10nm) and vacuum UV (200-10nm), are blocked by atmospheric oxygen and ozone (Madronich, McKenzie, Björn, & Caldwell, 1998).

Exposure of cells to UVB results in the formation of an array of different photoproducts, of which the most abundant forms are the cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6-4) pyrimidone (6-4PPs) photoproducts (You et al., 2001). Both products have been shown to cause distortions in DNA, with 6-4PPs in particular causing major distortions in the DNA backbone structure (Taylor, Garrett, & Cohrs, 1988), resulting in the loss of base pairing at the site of the lesion. CPDs have also been shown to occur at approximately three times the frequency of 6-4PPs and are responsible for the majority of mutations arising from UV irradiation (You et al., 2001). Formation of CPDs generally occurs between adjacent thymine residues (Gillet & Schärer, 2006b), and causes mutations arising from misincorporation of DNA bases by error prone DNA polymerases. The formation of these lesions is proportional to the UV irradiation dose received.

UV irradiation of cells can also result in further spontaneous oxidative DNA damage arising from increased levels of reactive oxygen species formed (Neeley & Essigmann, 2006; Wang & Kochevar, 2005).

Studies investigating the relative toxicity of these two photoproducts have shown that although both CPDs and 6-4PPs contribute to UV-induced apoptosis in cells lacking nucleotide excision repair (NER) pathways, in NER-proficient cells, only CPDs are responsible for UV-induced deaths (de Lima-Bessa et al., 2008). Studies in NER-deficient cells have also shown that the reduction of apoptosis levels following the removal of either CPD or 6-4PP UV lesions for a given dose was similar (Sage, 1993). This was despite the presence of an approximately 3-5 fold higher abundance of CPDs, suggesting that 6-4PPs may be 3-5 times as toxic in NER-deficient cells. Though less abundant, 6-4PP lesion removal occurs at a much faster rate than CPDs (Costa, Chiganças, Galhardo, Carvalho, & Menck, 2003), with approximately 80-90% of 6-4PP lesions removed within 4 hours following UV exposure in humans and rodents, compared to removal of only 50-60% of CPDs by 24 hours (Riou et al., 2004). This rapid repair of 6-4PPs may explain its inability or inefficiency in causing UV-induced deaths in NER-proficient cells.

### ***1.3 DNA damage response***

The complexity of the DNA damage response in cells has evolved from the need to manage a wide range of disparate damage types, from double strand breaks to single base alterations. Following DNA damage, a number of signalling networks are activated to initiate appropriate actions such as DNA repair, gene expression regulation, cell-cycle arrest, or apoptosis. Collectively, these are referred to as the DNA damage response (DDR). Depending on the type of DNA damage sustained, different signalling pathways are responsible for the activation of key regulators involved. A summary framework of the major components of the DDR signalling pathway, consisting of signal sensors, transducers and effectors, is shown in Figure 1.4. The majority of DNA lesions are processed by one of several pathways, including base excision repair, nucleotide excision repair, mismatch repair, non-homologous end-joining and homologous recombination. Though each response varies, DNA repair is highly coordinated with cell cycle progression, eliciting the activation of an array of signalling pathways termed DNA damage checkpoints (Harper & Elledge, 2007; Lazzaro et al., 2009). Common features of which include the delay, or arrest of cell cycle to allow for repair, alterations in gene expression and in cases whereby the damage sustained exceeds the cells ability for repair, induction of apoptosis (Zhou & Elledge, 2000). These effects are brought on through the complex and coordinated actions of various proteins involved in DNA damage signalling cascades, and can be broadly categorised as being DNA damage sensors, transducers, mediators and effectors. Key components that are central to the DDR signalling pathway include the phosphoinositide-3-kinase (PI3-K)-related protein kinase (PIKK) family, consisting of ataxia telangiectasia mutated (ATM), ataxia telangiectasia and rad-3-related kinase (ATR) and DNA-dependent protein kinase catalytic subunit (DNA-PKcs) (Bartek & Lukas, 2007).

The activation of the DDR requires the sensing of the aberrant DNA structures, examples of such sensors include the MRE11-RAD50-NBS1 (MRN) complex, which is responsible for the detection of double strand breaks (Lavin, 2007), and replication protein A (RPA) which is responsible for sensing single-stranded DNA (ssDNA). Following activation, the MRN complex is also responsible for the recruitment and phosphorylation of ATM, the apical DDR kinase (Cimprich & Cortez, 2008). With the aid of mediator proteins such as BRCA1, p53, MDC1, and effector kinases such as the checkpoint kinases Chk1 and Chk2, DNA damage signalling is spread throughout the nucleus (Polo & Jackson, 2011). Other major proteins classified as effectors include Nbs1 and Cdc25C. Critically, these checkpoint kinases, in addition to co-ordinating a number of effector molecules, are also responsible for the control of cell cycle and other processes (Reinhardt &

Yaffe, 2009). The progression through the distinct cell cycle phases, which are comprised of G<sub>1</sub> (growth phase 1), S (DNA synthesis and replication), G<sub>2</sub> (growth phase 2), and M (mitosis), is tightly regulated by cell cycle checkpoints. Following the activation of the cell cycle checkpoints in response to DNA damage, the progression through cell cycle is arrested, allowing time for DNA repair or in instances whereby the DNA damage sustained exceeds the capabilities of the cell to undergo repair, cell death (Sancar, Lindsey-Boltz, Unsal-Kaçmaz, & Linn, 2004).

#### ***1.4 DNA damage repair pathways***

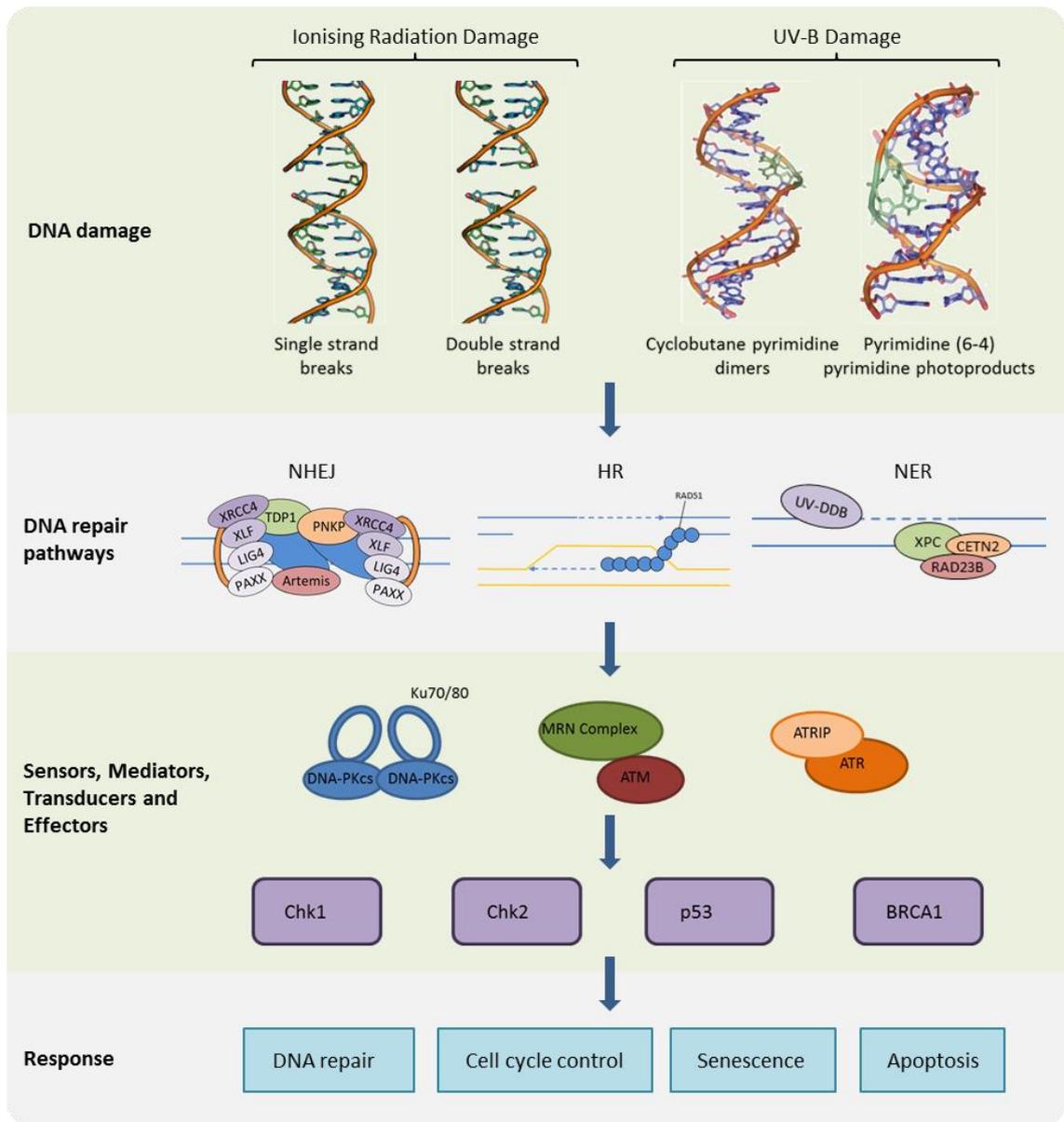
IR has been shown to elicit DNA damage primarily through double strand breaks, which are subsequently repaired primarily by two distinct mechanisms, non-homologous end-joining (NHEJ) and homologous recombination (HR) (Ciccia & Elledge, 2010). Similarly, in mammalian cells, UV-B induced DNA damage is primarily processed by excision repair pathways. These can be distinguished into two major categories, base excision repair (BER) and nucleotide excision repair (NER) (Rastogi, Richa, Kumar, Tyagi, & Sinha, 2010; Sinha & Häder, 2002). As this thesis is focused on the investigation into gene expression changes following only IR and UV-B, these specific repair pathways are described in more detail below.

##### ***1.4.1 Double strand break (DSB) repair***

DSBs are an extremely genotoxic form of DNA damage as both strands of the helix structure are broken, it leaves the DNA vulnerable to deletions, translocations as well as chromosomal fragmentation. It has also been shown that just one unrepaired DSB is sufficient to illicit permanent growth arrest and cell death (Bennett, Lewis, Baldwin, & Resnick, 1993). To avoid this, two separate DSB repair mechanisms have evolved, NHEJ, and HR. HR is a highly efficient and accurate pathway of DNA DSB repair, achieved through the retrieval of missing genetic information from undamaged DNA strand on a sister chromatid. Repair through NHEJ however re-joins the ends of a DSB without the requirement of a homologous sequence, but at the cost of being an error prone form of repair. The choice of which DSB repair pathway is used is tightly regulated depending on cell cycle phases (Chapman, Taylor, & Boulton, 2012), though recent studies have also suggested a role for 53BP1 in this decision (Panier & Boulton, 2013). As HR requires a homologous sequence located on a sister chromatid, it is restricted to the post-replicative stages of cell cycle, namely S and G<sub>2</sub>. NHEJ on the other hand is active throughout the cell cycle. Studies investigating the proportions of DSB repairs during the different cell cycle phases have found NHEJ to be the dominant repair pathway in G<sub>1</sub> and G<sub>2</sub>, with HR repair at its maximum during mid S phase (Karanam, Kafri, Loewer, & Lahav, 2012).

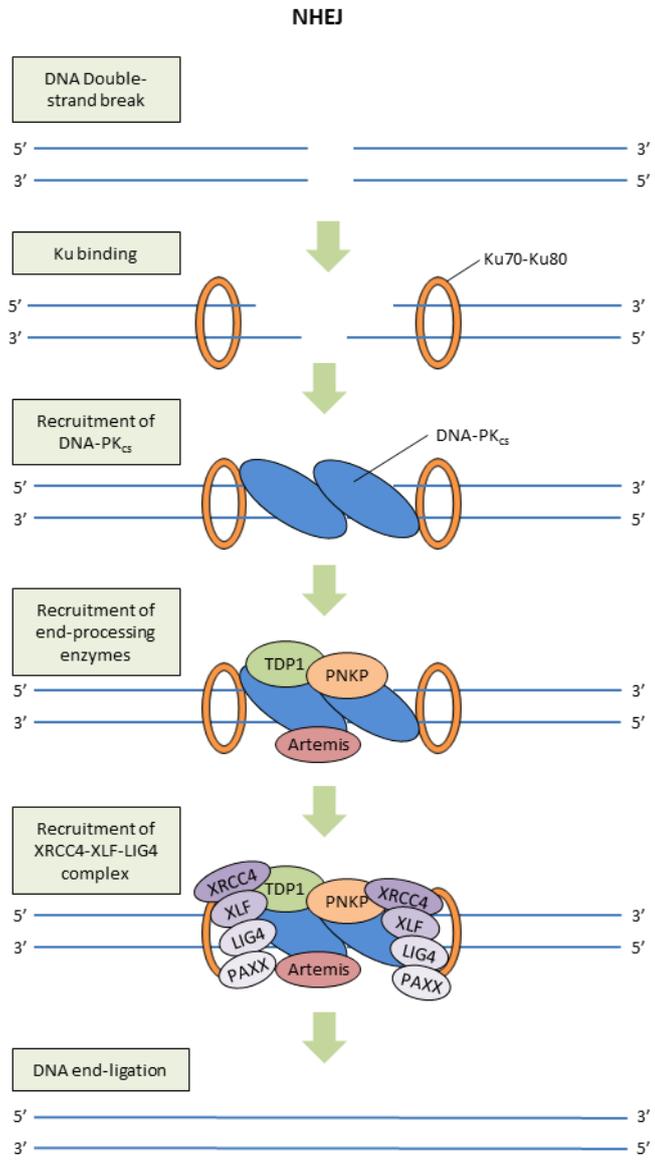
In classic NHEJ, the heterodimeric Ku70 and Ku80 complex rapidly binds to the DNA ends of the damage site (Figure 1.5), resulting in the recruitment and activation of DNA-dependent protein kinase – catalytic subunit (DNA-PK<sub>cs</sub>), priming the complex for the further recruitment of end-processing enzymes AP endonuclease 1 (APE1), tyrosyl-DNA phosphodiesterase 1 (TDP1), as well as Artemis, enabling the access of the core NHEJ proteins to the damage site, such as XRCC4 (X-ray repair cross-complementing protein 4), XLF (XRCC4-like factor), LIG4 (DNA ligase 4), and PAXX (Paralog of XRCC4) (Craxton et al., 2015; Ochi et al., 2015; Xing et al., 2015). Collectively these are responsible for mediating the final re-ligation step (Wyman & Kanaar, 2006).

In HR, DSBs are recognised by the MRE11-Rad50-NBS1 (MRN) complex, resulting in the rapid accumulation of MRN to the damage site (Figure 1.6). The MRN complex is then responsible for the activation of ATM, which in turn causes the phosphorylation of the histone variant H2A.X (Ciccio & Elledge, 2010; Misteli & Soutoglou, 2009). Binding of the mediator protein MDC1 to the chromatin flanking the DSB follows, and subsequently triggers recruitment of chromatin modification complexes. The resulting long stretches of single stranded DNA (ssDNA) then triggers the recruitment of replication protein A (RPA). RAD51 subsequently then displaces RPA with the aid of other proteins including BRCA1 and BRCA2, resulting the formation of a RAD51-ssDNA nucleofilament. This nucleofilament can then invade the homologous sequence, forming a D-loop and a Holliday junction, resulting in the priming of DNA synthesis and repair of the DSB. These junctions are removed through a process termed HJ “dissolution” or “resolution”, by a number of highly specialised endonucleases, termed HJ resolvases, which catalyses the cleavage of these junctions into two separate DNA duplexes. In eukaryotes, the majority of these junctions are removed by “dissolution” (Matos & West, 2014; Wyatt & West, 2014). For canonical HJ resolvases, these result in the formation of a pair of symmetrical nicks across the helices, resulting in nicked DNA duplexes which are then ligated (Figure 1.6). In contrast, non-canonical HJ resolvases result in the formation of asymmetric nicks, resulting in DNA duplexes which require additional processing before ligation.



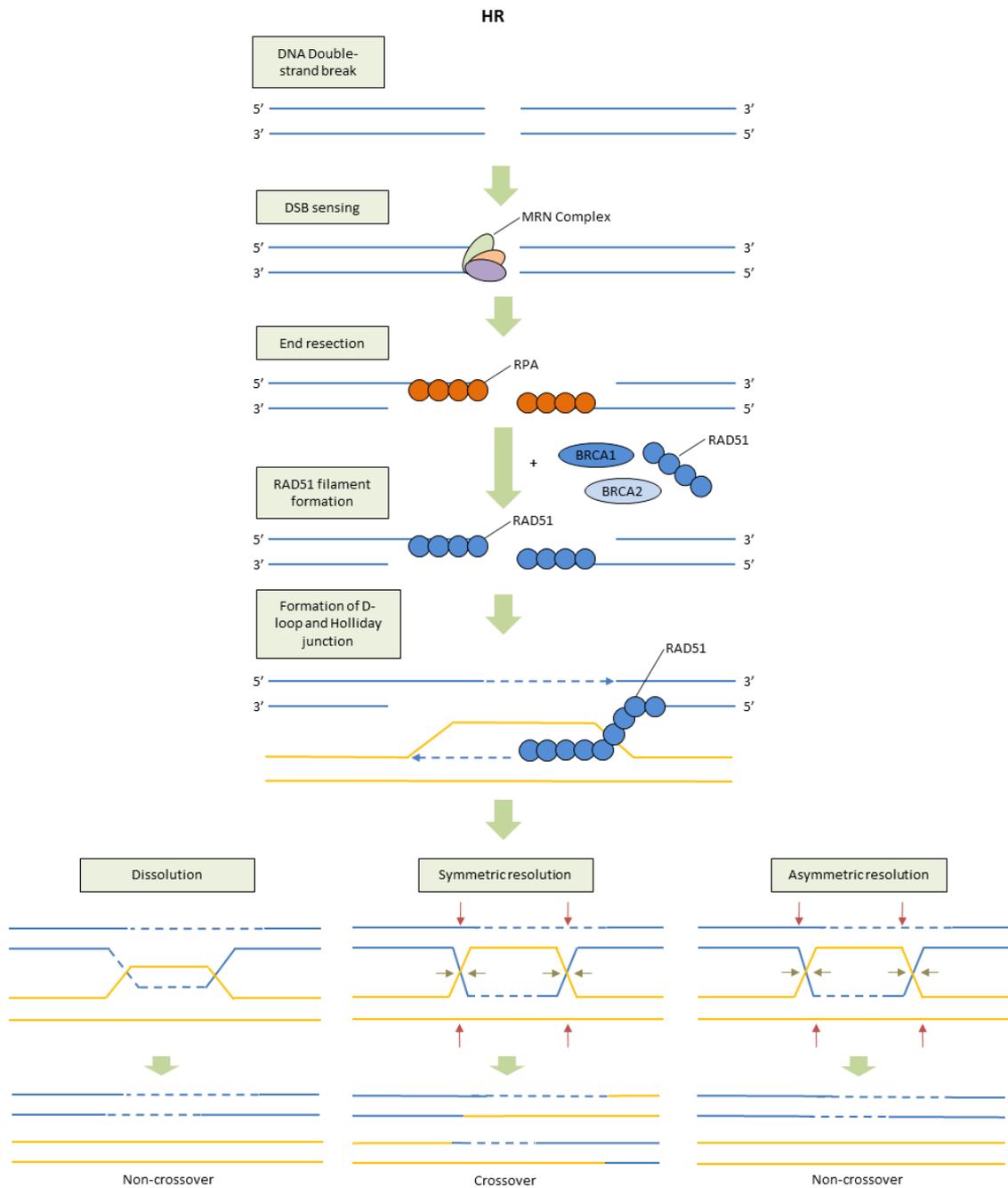
**Figure 1.4 | Overview of DNA Damage Response Signalling**

Overview of the primary types of DNA damage induced by both ionising radiation and UV-B exposure, as well as the main DNA repair pathways involved in their DDR. Various proteins act as the sensors to DNA damage, and help to recruit a number of PI3K kinases such as DNA-PKcs, ATM and ATR, which initiates a cascade of processes and signalling pathways, resulting in a DNA damage response. Such responses include DNA repair, changes in cell cycle, cellular senescence, apoptosis and changes in gene expression.



**Figure 1.5 | Schematic overview of non-homologous end-joining**

Overview of classic NHEJ repair of double-stranded DNA breaks, which acts independently of cell cycle. Ku70-Ku80 heterodimers rapidly binds to the DNA ends, resulting in the recruitment and activation of DNA-dependent protein kinase catalytic subunit (DNA-PK<sub>cs</sub>) which is an ATM related kinase. End-processing enzymes AP endonuclease 1 (APE1), tyrosyl-DNA phosphodiesterase 1 (TDP1), as well as Artemis and the XRCC4-XLF-LIG4-PAXX complex are then recruited, preparing the DNA ends for re-ligation.



**Figure 1.6 | Schematic overview of homologous recombination**

Overview of double-stranded DNA break repair by HR, which is limited to late S and G2 phases of cell cycle due to its requirement of having available sister chromatids as repair templates. In HR, the DSB is recognised by the MRE11-RAD50-NBS1 (MRN) complex, which is rapidly recruited to the damage site. In combination with several other nucleases long stretches of single-stranded DNA are then generated by resection, to which replication protein A (RPA) binds to. RAD51 subsequently then displaces RPA with the aid of proteins such as BRCA1 and BRCA2, forming a RAD51-ssDNA nucleofilament. This nucleofilament then searches for and binds with a homologous sequence, forming a D-loop and a Holliday junction, priming DNA synthesis and restoring lost genetic information from the DSB. These junctions can then be removed by HJ “dissolution” or “resolution”, as described in Section 1.4.1.

### ***1.4.2 Base excision repair***

The base excision repair (BER) pathway is responsible for the protection of cells from a number of endogenous DNA damage sources, such as damage induced by reactive oxygen species, hydrolysis or other metabolites, which may affect DNA base structure. The actions of this repair pathway can be broadly divided into two stages; the detection and excision of the damaged DNA base, and the subsequent repair by nucleotide re-synthesis. Depending on the number of nucleotides requiring re-synthesis, this process is termed short-patch (SP), for single nucleotide replacement, or long-patch (LP) repair, for replacement of more than 2 nucleotides (Frosina et al., 1996).

The detection and removal of modified or damaged bases are carried out by a number of different DNA glycosylases, with a number of glycosylases responsible for different types of damage (Seeberg, Eide, & Bjørås, 1995). Following the removal of the damaged base, the resulting apurinic/apyrimidinic (AP) site is then removed by an AP endonuclease (which incises the DNA strand 5' to the AP site) or an AP lyase (which incises the DNA strand 3' to the AP site). The remaining deoxyribose phosphate residue is then removed by a phosphodiesterase, with the remaining gap subsequently re-synthesised by DNA polymerase, and sealed by DNA ligase (Seeberg et al., 1995).

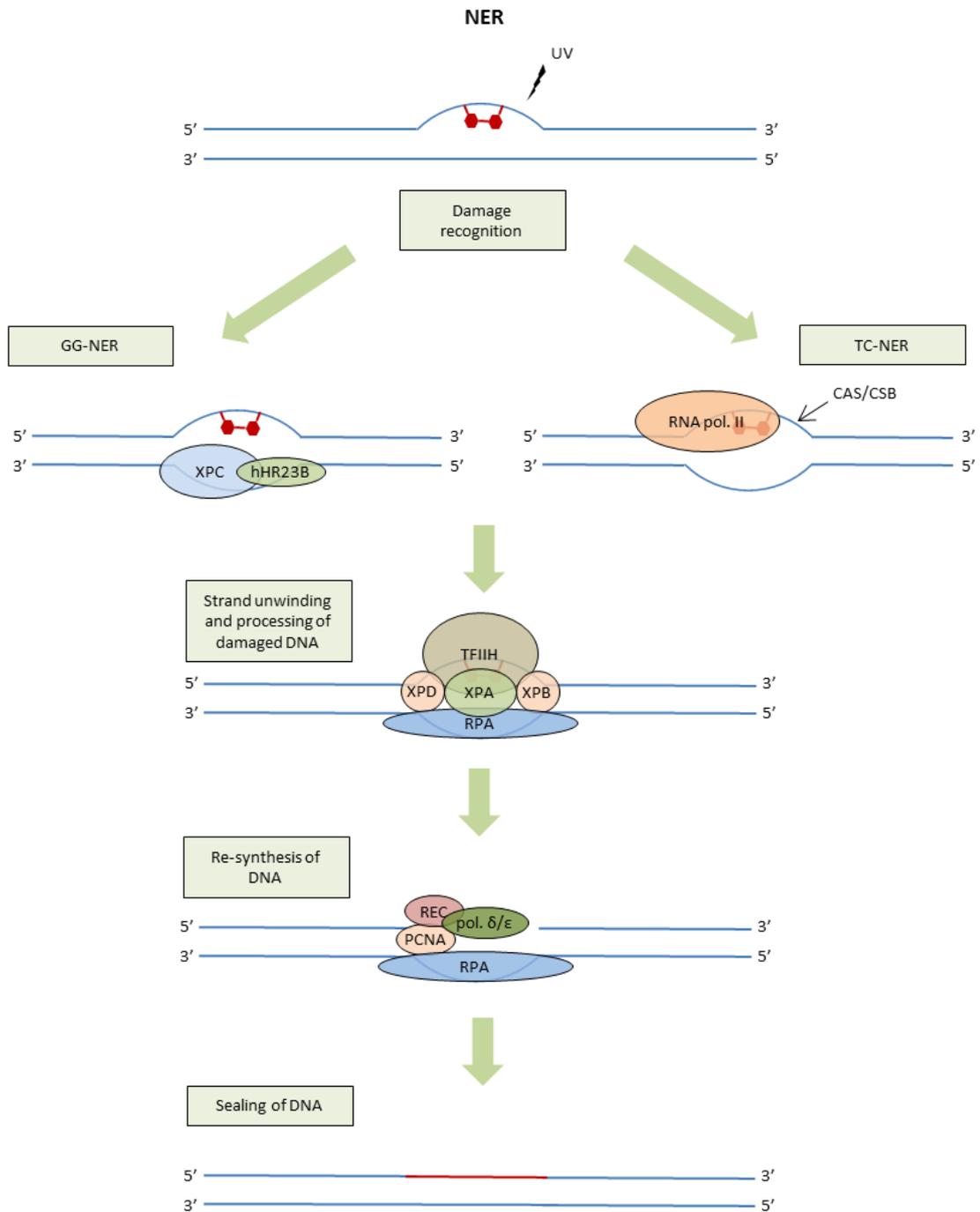
### ***1.4.3 Nucleotide excision repair***

The nucleotide excision repair (NER) pathway is responsible for removing a large number of DNA lesions, including CPDs and 6-4PPs which are generated following UV irradiation, as well as bulky chemical adducts, DNA-intrastrand crosslinks and certain forms of oxidative damage which cause distortions in the DNA double helix (Costa et al., 2003). The NER pathway is found in most organisms, and is highly conserved in eukaryotes, its importance is demonstrated in genetic disorders in humans such as Xeroderma Pigmentosum (XP) and Cockayne's syndrome (CS) whereby NER is defective. For patients with these disorders, this defect in NER results in a 1000 fold increase in the risk of incidence rates for skin cancers, such as melanoma, compared to healthy individuals (Lindahl & Wood, 1999).

NER is composed of two differentially regulated sub-pathways; transcription-coupled NER (TC-NER) which is responsible for the repair of lesions in actively transcribed genes (Hanawalt, 2002), and global genome NER (GG-NER) which is a slower repair process, functioning across the entire genome. For TC-NER, recognition of damage is induced by the stalling of RNA polymerase II upon encountering a lesion, which triggers the recruitment of CSA (ERCC8) and CSB (ERCC6) (Fousteri

& Mullenders, 2008). In GG-NER, lesions are recognised by the hHR23B-XPC protein complex, with the rate of repair highly dependent on lesion type, in particular, with 6-4PPs being removed at a much faster rate compared to CPDs, possibly due to the higher affinity of the hHR23B-XPC complex (Rastogi et al., 2010).

Whilst damage recognition for TC-NER and GG-NER differ, both converge on a common repair pathway (Figure 1.7). Unwinding of the DNA double helix around the lesion occurs through the actions of the multi-subunit transcription factor-IIH (TFIIH), this complex is composed of a core component (with XPB, XPD, p62, p52, p34 and p8), and a cdk activating kinase subunit (Schultz et al., 2000). Following unwinding, RPA, XPA and XPG are recruited to the site. The combined activities of these proteins results in the cleavage of the damaged strand and the remaining gap is subsequently re-synthesised by DNA polymerase  $\delta$  or  $\epsilon$  (along with PCNA, RPA and RFC), and sealed by DNA Ligase I (Figure 1.7) (Gillet & Schärer, 2006a).



**Figure 1.7 | Schematic overview of nucleotide excision repair**

Overview of global genome nucleotide excision repair (GG-NER) and transcriptional coupled nucleotide excision repair (TC-NER), as described in more detail in Section 1.4.3.

### ***1.5 DNA damage and control of protein synthesis***

Although extensive research has been carried out on the DNA damage signalling pathways and the protein complexes that are required to repair damaged DNA there is still much to be learnt. In particular, the post-transcriptional response to DNA damage has not been well documented. This is perhaps surprising given that there is overwhelming evidence to show that post-transcriptional control makes a major contribution to the overall regulation of gene expression pathways (Spriggs et al., 2010). Work from the Willis laboratory has shown that following exposure to UVB light a post-transcriptional regulatory pathway is initiated (Powley et al., 2009; Somers et al., 2015). Mechanistically it was shown that following UVB exposure there was phosphorylation of eIF2 $\alpha$  however, uORFs present in the 5' UTRs of NER enzymes including ERCC5 (XPG), ERCC1 and DDB1 regulated the translation of the downstream cistron and the synthesis of the corresponding proteins following DNA damage (Powley et al., 2009). The data show that following DNA damage these uORFs are bypassed allowing for selective translation of the downstream cistron, although the mechanisms by which this occurs are not understood. There is a large degree of selectivity in the translation of uORF containing transcripts following cell stress. These RNA elements are present in 40-50% the human genome, yet only a very small subset of uORF-containing mRNAs are translated following UV exposure. It is likely that the recruitment of these mRNAs to the translational machinery requires interaction with specific RNA-BP complexes.

Importantly, recent data showed that the pathways that repair bulky adduct DNA damage caused by UV exposure are also employed to repair damage caused by exposure to chemotoxic agents including the commonly used platinum-based compounds such as cisplatin and carboplatin (Somers et al 2015). A common polymorphism was identified in the 5' UTR of ERCC5, a structure specific endonuclease that cleaves 3' of the DNA adduct and is an indispensable core protein of the NER machinery. This polymorphism, rs751402, is present in 35% of the white Caucasian population and generates a uORF (termed uORF1). It was shown that individuals with the uORF had low background level of ERCC5, but following exposure to bulky adduct damage this enzyme was induced allowing for efficient repair of the exogenous DNA damage (Somers et al., 2015). This information was shown to be very important for tumour cell survival and tumours from individuals that contained the uORF were resistant to treatment with widely used platinum-based chemotherapeutic agents.

## ***1.6 Techniques for studying genome wide changes in translation***

Two techniques have been developed to study genome wide changes in mRNA translation; polysome profiling (reviewed in Spriggs et al 2010) and the more recently developed ribosome profiling (Ingolia, Ghaemmaghami, Newman, & Weissman, 2009). Polysome profiling is based on the premise that mRNAs that are associated with more ribosomes are more translated. Therefore, post nuclear extracts are loaded onto sucrose density gradients, which are then fractionated and the corresponding RNAs identified using either cDNA microarray or more recently by RNA-seq. By measuring a shift in position of an mRNA from the subpolymes to the polysome it is possible to assess changes in the translational status of mRNAs in both control and treated cells. In ribosome profiling, polysomes are isolated and digested and the mRNA fragments associated with individual ribosomes are sequenced. Therefore, the major advantage of using ribosome over polysome profiling is that it is possible to obtain changes in ribosome positioning at single-nucleotide resolution. However, there are challenges in the interpretation of ribosome profiling data (King & Gerber, 2016), and it has been argued that when studying genome-wide changes in translational efficiency polysome profiling is able to provide more robust data sets (Gandin et al., 2016). When studying the post-transcriptional response to DNA damage, polysome profiling was carried out to assess changes in the global translation profile following UVB exposure (Powley et al., 2009), and a candidate gene approach was used to study changes in translation following IR exposure. Given that uORFs were identified as the RNA motifs that permitted selective translation following UVB exposure, the use of ribosome profiling would provide further evidence for position dependent changes, and therefore would be the technique of choice in this instance.

## ***1.7 Project aims***

As outlined in this chapter, regulation of gene expression forms an essential part of the DNA damage response, with selective translational reprogramming occurring following DNA damage initiated by a range of different agents. However, previous studies into these responses have typically been conducted using transformed cell lines (e.g. HeLa or MCF7 cells), which have been shown to elicit dramatically different DDRs to non-transformed cells. Moreover, previous work from the Willis laboratory has shown that selective mRNAs, which contain uORFs are selectively translated following UVB exposure. However, it was not possible to probe the use of uORFs in the previous study in detail, as it was carried out using polysome profiling, a technique, which

as discussed above, is not able to provide information about change in position of the ribosome on the mRNA. The overall aims of this project were to investigate selective translational reprogramming following DNA damage using the non-transformed breast epithelial MCF10A cell line and to use the technique ribosome profiling to assess if there were changes in the degree occupancy (e.g. at specific AUGs) of ribosomes on selective mRNAs. Thus chapters 3 and 4 outline the development of the chosen UV-B based DNA damage model, and subsequent sequencing using ribosome profiling, as well as analysis of the resulting data for the identification of differentially expressed genes and their associated pathways. Chapter 5 explores a number of technical issues encountered during ribosome profiling, and attempts to both characterise these problems as well as propose new approaches to mitigate these technical issues.

## 2. Materials and Methods

### 2.1 Cell culture techniques

#### 2.1.1 Growth and freeze media

Cell Line	Growth Media Mix	Freeze Mix
<b>MCF10A</b>	DMEM/F12 (Invitrogen #11330-032) 5% Horse Serum (Invitrogen #16050-122) 10ng/ml EGF 1mg/ml Hydrocortisone 10ng/ml Cholera Toxin 10µg/ml Insulin	MCF10A Growth Media mix without Pen/Strep 25% FBS 7.5% DMSO
<b>HEK293</b>	DMEM (Invitrogen #41966) 10% FBS (Sigma #F9665) 200mM L-glutamine	

Table 2.1 | Reagents and volumes used for cell culture growth and freeze media.

#### 2.1.2 Source and maintenance of cell lines

Cell lines used were sourced directly from Sigma-Aldrich (CLLS1049-1SET), and were grown in the manufacturer recommended media mixes listed in Section 2.1.1. Typically, cells were grown in T175 gamma irradiation sterilised plasticware (Grenier Bio), with the appropriate growth medium (Table 2.1) in a humidified incubator maintained at 37°C containing 5% CO<sub>2</sub>. Cells were typically grown to approximately 80% confluence before they were split; the frequency and ratio of which varied depending on the cell line. For splitting cells, existing media was removed, the cells were washed once with warm PBS and were then trypsinised using 1x Trypsin with 0.5mM EDTA (Sigma) at 37°C. The time required for cells to detach varies with the cell line used, but typically for MCF10A cell line this took approximately 15-25 minutes. Following detachment, a greater than triple volume of fresh media was added to inactivate the trypsin and the cells were centrifuged at 200 xg for 5 minutes. The cell pellet was resuspended in an appropriate volume of fresh media and seeded accordingly.

### **2.1.3 Cryopreservation of cell lines**

Cells were harvested using the method in Section 2.1.2, but were resuspended in the appropriate medium (Table 2.1) at a concentration of approximately  $3 \times 10^6$  cells/ml. Aliquots of 1ml were transferred into autoclaved screw top Eppendorf tubes (Sigma) and were slowly frozen at  $-80^\circ\text{C}$  for a minimum of 24 hours prior to transferring to liquid nitrogen for long term storage.

### **2.1.4 Thawing frozen cell stocks**

Frozen stocks of cells were quickly thawed in a  $37^\circ\text{C}$  water bath, and were then added to conical centrifuge tubes (BD Falcon) containing pre-warmed media before being spun down to remove remaining DMSO. Pelleted cells were suspended in fresh media and seeded accordingly.

### **2.1.5 Clonogenic assay of cells in vitro**

This cell survival assay is based on the ability of a single seeded cell to grow into a colony, defined as being of at least 50 cells. This measurement can be used to determine dose-responses to a range of treatments. Typically, cells were seeded into triplicate wells in a 6-well plate at different densities and incubated for 24 hours before treatment. Following treatment, cells were incubated until colonies of at least 50 cells could be observed (typically around a week depending on starting seeding density). Once colonies were formed, media was removed, rinsed carefully with PBS and were incubated in 6% gluteraldehyde with 0.5% crystal violet at room temperature for a minimum of 30 minutes. The gluteraldehyde crystal violet mixture was then removed and the cells were carefully washed with water and were left to air dry. Stained colonies were counted and the surviving fraction for each condition was calculated using the following formula:

$$SF = \frac{\text{no. of colonies formed after treatment}}{\text{no. of cells seeded} \times PE}$$

Where PE, plating efficiency, is defined as:

$$PE = \frac{\text{no. of colonies formed}}{\text{no. of cells seeded}} \times 100\%$$

## ***2.2 Cell treatment and induction of DNA damage***

### ***2.2.1 Treatment of cells with ionising radiation (IR)***

Cells were typically grown to roughly 70% confluence prior to treatment, and were irradiated using a Pantak X-ray machine at a dose rate of 1 Gy/min at room temperature for the stated doses as shown. Untreated control cells were identically handled except for irradiation. Following irradiation, cells were returned to the incubator for the stated durations shown.

### ***2.2.2 Treatment of cells with ultraviolet radiation***

Cells were grown to approximately 70% confluence prior to treatment. Cells were subjected to a media refresh for 1 hour. Following this, media was removed before being irradiated with UV-B using a UVP CL-1000 cross-linker oven (with a peak output wavelength of 302nm) at the specified doses. After irradiation, the original media was added back to the cells and were incubated at 37°C for the stated durations shown.

## ***2.3 Protein techniques***

### ***2.3.1 Buffers and solutions***

Unless otherwise stated, all chemicals were from Sigma-Aldrich (UK).

**Whole cell lysis buffer:** 10 mM Tris pH 7.5, 50 mM NaCl, 0.5% NP40, 0.5% sodium deoxycholate, 0.5% SDS, 10mM iodoacetamide. Lysis buffer was supplemented with 1X Complete™ Protease Inhibitor Cocktail (Roche Applied Science) and 1X PhoSTOP™ Phosphatase Inhibitor Cocktail

**RIPA buffer:** 150 mM NaCl, 50 mM Tris pH 7.5, 1% Triton X100, 0.5% sodium deoxycholate, 0.1% SDS

**2x SDS-PAGE sample loading buffer:** 100 mM Tris-HCl pH 6.8, 20% glycerol, 8% SDS, 10% β-mercaptoethanol, 2 mM EDTA, 0.02% (w/v) bromophenol blue

**5x SDS-PAGE samples loading buffer:** 200 mM Tris-HCl pH 6.8, 50% glycerol, 20% SDS, 5 mM EDTA, 0.5% bromophenol blue

**Resolving buffer:** 1.5 M Tris-HCl pH 8.8, 1% SDS

**Stacking buffer:** 0.25 M Tris-HCl pH 6.8, 1% SDS

**10x SDS running buffer:** 250 mM Tris-HCl pH 8.8, 1.92 M glycine, 1% SDS

**Tris buffered saline with Tween (TBST):** 10 mM Tris-HCL pH 8.0, 150 mM NaCl, 0.1% Tween-20

**Ponceau stain:** 0.5% w/v Ponceau S in 5% w/v trichloroacetic acid (TCA)

**In-house ECL solution:** 1 ml Luminol solution (0.1 M Tris-HCL pH 8.6, 2.5 mM sodium luminol), 10  $\mu$ l enhancer (11 mg para-coumaric acid in 10 ml DMSO), 3.1  $\mu$ l 3% hydrogen peroxide solution

### 2.3.2 Antibodies and dilutions used

1° Antibody	MW (kDa)	Source	Clonality	Supplier	1° Dilution	2° Dilution
β-actin	45	Mouse	Mono	Cell Signaling #3700	1:5000	1:5000
4E-BP1	15-20	Rabbit	Mono	Cell Signaling #9644	1:1000	1:10000
ATM	370	Mouse	Mono	Cell Signaling #sc23921	1:1000	1:5000
eIF2α	38	Rabbit	Poly	Cell Signalling #9722	1:1000	1:10000
p-4E-BP1 (S65)	15-20	Rabbit	Mono	Cell Signaling #9456	1:1000	1:10000
P53	53	Mouse	Mono	Dako #M7001	1:1000	1:5000
PARP	89,116	Rabbit	Poly	Cell Signaling #9542	1:1000	1:10000
p-ATM (S1981)	350	Mouse	Mono	Cell Signaling #4526	1:1000	1:5000
p-Chk1 (S345)	56	Rabbit	Poly	Cell Signaling #2341	1:1000	1:10000
p-Chk2 (T68)	62	Rabbit	Poly	Cell Signaling #2661	1:1000	1:10000
p-eIF2α (S52)	38	Rabbit	Poly	Life Tech #44728G	1:1000	1:10000
p-eIF4B (S422)	80	Rabbit	Poly	Cell Signaling #3591	1:1000	1:10000
p-H2AX (S139)	15	Rabbit	Mono	Cell Signaling #9718	1:1000	1:10000

Table 2.2 | Antibodies used for Western blotting

### 2.3.3 Cell lysate sample preparation

Typically, cells were grown and treated in 6-well plates or 6cm dishes, after treatment cells were lysed in 100 µl or 200 µl ice-cold whole cell lysis buffer respectively (Section 2.3.1) and were snap frozen and stored at -80°C. Once this process had been completed for all the required samples, the frozen plates/dishes were thawed out on ice and the lysates were transferred to Eppendorf tubes. DNA was sheared by forcing samples through gauge 27 needles 10x (BD Plastipak). Alternatively, samples were incubated with 10U of Benzonase on ice for 30 mins. Lysates were then subsequently placed on ice until protein quantification.

### 2.3.4 Protein quantification

Protein quantification of samples was conducted using Pierce BCA assay or Bradford assay depending on the composition of the lysate. Both assays were conducted using 96-well plates,

with pre-prepared BSA standards ranging from 20 - 2,000 µg/ml, with both samples and standards repeated in triplicate. The concentrations of the protein samples were calculated from the standard curve obtained from the readings of the BSA standards. These concentration readings were then used to normalise all samples within the same set to that of the lowest concentration using additional lysis buffer of the same composition. Appropriate volumes of 2x or 5x SDS-PAGE sample loading buffer (Section 2.3.1) were added, after which lysates were denatured at 75°C for 10 minutes and placed on ice until required or stored at -20°C.

#### *2.3.4.1 Pierce BCA protein assay*

3 µl of each sample and standard were pipetted in 96-well plates, to which 180 µl of BCA reagent (Pierce) was added to each well. The plate was shaken to ensure each well had been mixed thoroughly and was covered and incubated at 37°C for 30 minutes, after which the plate was allowed to cool to room temperature before absorbance readings were measured at 562 nm for each well using a PowerWave XS2 plate reader (Bio-Tek instruments).

#### *2.3.4.2 Bradford assay*

3 µl of each sample and standard were pipetted in 96-well plates, to which 180 µl of Bradford reagent were added to each well. The plate was shaken to ensure each well had been mixed thoroughly, and was incubated at room temperature for 5 minutes before absorbance readings were measured at 595 nm for each well using a PowerWave XS2 plate reader (Bio-Tek instruments).

### **2.3.5 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)**

SDS-PAGE gels were cast with a resolving gel layer, composed of varying acrylamide concentration which was determined by the size of the protein of interest, and a stacking gel layer on top (Table 2.3). Once protein samples and appropriate protein size ladders were loaded, gel electrophoresis was run using a Bio-Rad Mini-PROTEAN® Tetra Cell system in 1X SDS running buffer at a constant voltage of 80 V for the stacking layer and 100 V for the resolving layer until the dye front had moved off the bottom of the gel.

Alternatively, precast Novex (Invitrogen) gels of appropriate compositions were assembled in XCell SureLock™ Mini-Cell (Invitrogen), filled with appropriate running buffers, and ran at 150 V until the dye front moved off the bottom of the gel.

Gel (%)	Stacking Gel	Resolving Gel		
		7	10	15
ddH <sub>2</sub> O (ml)	1.80	2.58	2.03	1.20
30%:0.8% Acrylamide:Bisacrylamide (ml)	0.67	1.17	1.67	2.50
4X Resolving buffer (ml)		1.25	1.25	1.25
2X Stacking buffer (ml)	2.50			
25% Ammonium persulphate* (μl)	50		50	
TEMED* (μl)	3		3	

**Table 2.3 | Gel composition for SDS-PAGE (for 1x 0.75 mm gel).**

\*Polymerisation reagents were added just prior to pouring the gel

### ***2.3.6 Transfer of proteins onto PVDF membrane***

Following separation by SDS-PAGE gel electrophoresis, proteins were transferred and immobilised onto PVDF (Bio-Rad) by wet or semi-dry transfer. Prior to transfer, the PVDF membrane was immersed in methanol for 30 seconds to pre-wet the membrane, followed by immersion in the appropriate transfer buffer for 2 minutes. Reversible Ponceau S staining was conducted following transfer to confirm successful transfer of proteins onto the membrane.

#### *2.3.6.1 Wet transfer method*

Wet transfer was conducted using Bio-Rad Mini-PROTEAN® Tetra Cell with Mini Trans-Blot® module, formed by the gel being placed on the pre-wetted PVDF membrane. The gel and the PVDF membrane were flanked by 3 layers of Whatman® filter paper (GE Healthcare) on either side, pre-soaked in transfer buffer, and finally sandwiched between two sponge pads. Transfers were conducted at 100 V for 1 hour at 4°C, with the presence of an ice block and magnetic stirrer in the tank to maintain a low temperature.

#### *2.3.6.2 Semi-dry transfer method*

Semi-dry transfer was typically used for small proteins (<40 kDa). The gel was placed on top of the pre-wetted PVDF membrane which was sandwiched between 3 layers of Whatman® filter paper (GE Healthcare) on both sides, pre-soaked in transfer buffer. Transfers were conducted at 10 V for 1 hour at room temperature.

### ***2.3.7 Immunodetection of proteins***

Immobilised proteins on the PVDF membrane were detected using antibodies specific to the protein of interest. To prevent non-specific binding to the membrane, it was first blocked in TBST containing 5% dried milk powder (Marvel) w/v for 1 hour at room temperature in Falcon tubes (BD Plastipak) with rotation. Following blocking, the membrane was incubated with a solution of TBST containing the relevant primary antibody (with concentrations used shown in Table 2.2) along with either 5% w/v Marvel or BSA overnight at 4°C with rotation, or for 1 hour at room temperature with rotation. The excess primary antibody was removed by three 10 minute wash steps in TBST, after which the membranes were incubated for a minimum of 1 hour at room temperature with a horseradish peroxidase-conjugated secondary antibody raised against the animal species of the primary antibody that was used. Excess secondary antibody was then removed by another three 10 minute wash steps in TBST. The blots were visualised by incubating with an enhanced chemiluminescence (ECL) kit (GE Healthcare, Amersham) or for abundant proteins, an in-house ECL solution (Section 2.3.1) for 5 minutes, before being exposed to photographic film (Fuji RX X-ray) with exposure times ranging from seconds to hours depending on protein abundance.

### ***2.3.8 Stripping Western blot membranes***

Membranes were stripped of antibodies by washing with PBS for 5 minutes, followed by incubation in Restore Western Blot Stripping Buffer (Thermo Scientific) for 15 minutes. The stripped membranes were washed in PBS for 5 minutes at room temperature with rotation.

### ***2.3.9 <sup>35</sup>S Methionine incorporation for measuring protein synthesis rates***

Cells used for <sup>35</sup>S methionine incorporation were typically grown in 6-well plates to around 60% confluence. Growth media was removed and the cells were gently washed twice with warm PBS. <sup>35</sup>S methionine (MP Biomedicals, California) was added to methionine and cysteine free DMEM media, which was supplemented with 10% dialysed FBS and 2 mM L-glutamine along with any other additives specific to the cell line used (Table 2.1) to reach a final concentration of 30 µCi/ml. 1ml of this media was added to each well and incubated for 20 minutes at 37°C. Following incubation, the plates were placed on ice and the <sup>35</sup>S methionine media was removed and the cells were gently washed twice with ice cold PBS. The cells were then directly lysed by addition of 400 µl 1X Passive Lysis Buffer (Promega) and were snap frozen at -80°C. Once all samples had been processed, they were allowed to thaw on ice, after which lysates were transferred to Eppendorf tubes and were spun at 17,000 xg for 1 minute at 4°C. 50 µl of the supernatant was used for protein quantification while 300µl of the lysate was added to an equal

volume of 25% TCA and incubated for 30 minutes on ice. Following incubation, the precipitated TCA-lysate samples were loaded onto Whatman® GF-A glass fibre filters that were pre-wetted with 25% TCA, and assembled in a vacuum manifold. The filters were washed twice with 70% IMS, twice with acetone and were left to dry. The dried filters were transferred to scintillation vials to which 2 ml of Ecoscint scintillation cocktail (National Diagnostics) was added. The level of <sup>35</sup>S methionine incorporation was determined by scintillation counting using a Wallac LKB RackBeta scintillation counter for 2 minutes per sample, after which the readings were normalised to the protein abundance of each sample using the protein quantification results.

## ***2.4 RNA techniques***

Unless otherwise stated, all chemicals were from Sigma.

### ***2.4.1 Buffers and solutions***

**10x Polysome gradient buffer (high salt):** 3 M NaCl, 150 mM MgCl<sub>2</sub>, 150 mM Tris-HCL pH 7.5, 1 mg/ml cycloheximide

**10x Polysome gradient buffer (standard salt):** 1.5 M NaCl, 50 mM MgCl<sub>2</sub>, 150 mM Tris-HCL pH 7.5, 1 mg/ml cycloheximide

**10X RPF gradient buffer (high salt):** 3 M NaCl, 150 mM MgCl<sub>2</sub>, 150 mM Tris-HCL pH 7.5, 1 mg/ml cycloheximide, 20 mM DTT

**10X RPF gradient buffer (standard salt):** 1.5 M NaCl, 50 mM MgCl<sub>2</sub>, 150 mM Tris-HCL pH 7.5, 1 mg/ml cycloheximide, 20 mM DTT

**PBS-CHX:** 1X PBS, 100 µg/ml cycloheximide

**Polysome lysis buffer (standard salt):** 1% Triton-100, 1X polysome gradient buffer

**Polysome lysis buffer (low salt):** 10 mM NaCl, 15 mM Tris-HCL pH 7.5, 0.1 mg/ml cycloheximide, 0.2 M sucrose, 0.5% NP40 (Igepal), 1 mM DTT, 0.2 U/µl RNasein (Promega)

**RPF lysis buffer:** 1% Triton-100, 1X RPF gradient buffer, RNasin (for ribosome profiling, typically 500 U/ml)

**Blue sucrose:** 65% w/v sucrose solution, 0.01% bromophenol blue

**Sucrose cushion (TruSeq):** 50% (w/v) sucrose in 1x RPF buffer, 20U/ml SUPERase-In (ThermoFisher Scientific)

**Sucrose cushion (Ingolia):** 1M sucrose in 1x RPF buffer, 20U/ml SUPERase-In (ThermoFisher Scientific)

**Methylene blue stain:** 0.02% methylene blue, 0.3M NaOAc pH 5.2

**Church Gilbert solution:** 140mM Na<sub>2</sub>HPO<sub>4</sub>, 70mM NaH<sub>2</sub>PO<sub>4</sub>, 7% SDS

**Stop solution:** 95% deionised formamide, 1mM EDTA, 0.01% bromaphenol blue and xylene cyanol

## ***2.4.2 Polysome profiling***

### *2.4.2.1 Preparing sucrose gradients*

Sucrose gradients were made using 10%, 20%, 30%, 40% and 50% w/v sucrose solutions in the appropriate 1X polysome gradient buffer (Section 2.4.1), 2 ml layers of each concentration were sequentially added to 12 ml polyallomer centrifuge tubes (Sorvall, Newtown, Connecticut, USA) in descending order and were allowed to freeze at -80°C for a minimum of 30 minutes before adding the next layer. This freeze-layer process was repeated until the gradient was fully formed, and was subsequently stored at -20°C. Prior to use, gradients were removed a day in advance and stored at 4°C to equilibrate.

### *2.4.2.2 Harvesting cells*

Cells were typically grown in 15 cm plates to 70% confluence and were supplemented with cycloheximide (final concentration 100 µg/ml) and at 37°C for 5 minutes. Following incubation, plates were kept on ice and the cycloheximide incorporated media was removed, plates were gently washed with 10 ml ice-cold PBS-CHX before they were lysed and scraped in 600 µl of appropriate 1x polysome lysis buffer. The lysates were transferred to Eppendorf tubes and incubated on ice for 3 minutes prior to centrifugation at 1300 xg for 5 minutes at 4°C. The supernatant was loaded onto the pre-prepared standard salt buffer sucrose gradients.

Loaded gradients were placed onto a pre-cooled SW41 Ti rotor (Beckman Coulter) and were spun at 247,000 xg for 2 hours at 4°C in a Sorvall WX Ultra Series centrifuge (Thermo Scientific). Following centrifugation, gradients were loaded onto a gradient density fractionation system (Presearch, Hampshire, UK) in which blue sucrose (Section 2.4.1) was pumped in at the bottom

of the gradient tube using a syringe pump (KD Scientific) at a rate of 1ml/min. This caused the displaced sucrose solution to be pumped through a UA-6 UV/Vis detector (Teledyne Isco) in which the absorbance of the sample was continuously measured at 254 nm. The displaced sample was subsequently collected at 1 minute intervals using a Foxy Jr fraction collector (Teledyne Isco).

#### *2.4.2.3 RNA precipitation and purification*

The RNA was precipitated from each gradient fraction by incubating the sample overnight at -20°C with 3 ml 7.7 M guanidine hydrochloride and 4 ml ethanol. Following precipitation, samples were centrifuged at 4,000 rpm for 50 minutes at 4°C and the pellet was resuspended in 400 µl ddH<sub>2</sub>O. The resuspended RNA was transferred to Eppendorf tubes containing 40 µl 3 M sodium acetate pH 5.2 and 1 ml 100% ethanol and was either stored at -80°C for 20 minutes or -20°C overnight. Following precipitation, samples were centrifuged at 12,000 xg for 30 minutes at 4°C, after which the supernatant was carefully removed and the remaining pellets were washed with 75% ethanol by gentle vortexing. The samples were centrifuged again at 12,000 xg for 10 minutes at 4°C, supernatant was removed, and the pellets were allowed to air dry before resuspended in 40 µl nuclease free water and were stored at -80°C.

### **2.4.3 Ribosome profiling**

#### *2.4.3.1 Preparing sucrose gradients*

Sucrose gradients were prepared as described in Section 2.4.2.1, but with the appropriate 1x RPF gradient buffer (Section 2.4.1), supplemented with 500 U/ml RNaseIn Plus (Promega).

#### *2.4.3.2 Harvesting cells and digestion*

Cells were harvested using the same protocol as described in Section 2.4.2.2, using 1x RPF lysis buffer. For ribosome profiling, generally greater numbers of 15 cm plates were required, typically 2-3 for each condition, to ensure plenty of material for processing. The optical density of each sample was measured on a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Scientific), and appropriate amounts of RNase I (Ambion) was then added to an aliquot of the collected lysates. RNase I digestion was carried out at either room temperature or at 37°C for 30-45 minutes in an Eppendorf ThermoMixer. Following digestion, appropriate volumes of SUPERase-In RNase Inhibitor (ThermoFisher Scientific) was added to the sample to quench the reaction, typically 15 µl for 600 µl of lysate from one 15 cm plate. The digested lysates were then placed on ice until they were loaded onto the prepared sucrose gradients or sucrose cushions.

Total RNA samples were extracted from an aliquot of the undigested lysates using the Zymo Research RNA Clean & Concentrator-25 kit. Typically, 100 µl of undigested lysates was used, to which 10 µl of 10% SDS was added and mixed. This sample was then processed using the illumina TruSeq variant of the manufacturer protocol. Briefly, this involved addition of 220 µl of RNA binding buffer and 220 µl of 100% ethanol to each sample, following which the purification was carried out using the standard manufacturer's instructions, with the final purified samples eluted in 30 µl of ddH<sub>2</sub>O.

#### *2.4.3.3 Centrifugation, fractionation and RNA extraction*

Purification and separation of the 80s monosomes from the digested lysates was achieved in one of three methods, through the use of sucrose gradient fractionation similar to that used for polysome profiling, by the use of illustra MicroSpin S-400 HR size exclusion columns (GE Life Sciences), or alternatively, through the use of sucrose cushions.

##### *2.4.3.3.1 Sucrose gradient fractionation*

Centrifugation and fractionation was carried out using the same protocol as described in Section 2.4.2 **Error! Reference source not found.**, with the exception that the samples were collected in 30s (500 µl) fractions in LoBind Eppendorf tubes and placed immediately on ice until all fractions were collected. Following collection, 1.4 ml Trizol LS (Sigma) was added to all relevant fractions (as determined by the resulting polysome profiling trace), mixed and left to isolate the RNA at room temperature for 10 minutes. 200 µl 1-bromo-2-chloropropane was added to each fraction, mixed thoroughly by shaking for 15 seconds, and left at room temperature for 10 minutes before being centrifuged at 12,000 xg for 15 minutes at 4°C. The upper aqueous layer was transferred into a fresh LoBind Eppendorf tube and mixed with 1µl GlycoBlue and 1ml isopropanol and left at room temperature for 15 minutes. The precipitated RNA was pelleted by centrifugation at 12,000 xg for 30 minutes at 4°C, after which the supernatant was discarded and the pellet was washed carefully with 75% ethanol before another centrifugation step at 12,000 xg for 10 minutes at 4°C. After removal of the supernatant, the pellets were allowed to air dry and were resuspended in 10µl nuclease free water.

##### *2.4.3.3.2 Size exclusion columns*

For purification of monosomes using illustra MicroSpin S-400 HR size exclusion columns (GE Life Sciences), the columns were first vortexed to resuspend the resin, after which they were centrifuged at 2,800 rpm for 1 minute at room temperature on an Eppendorf 5418 benchtop

centrifuge (~670 xg). 400µl of appropriate RPF buffer (Section 2.4.1) was added to each column and centrifuged again using the same conditions, this process was repeated a total of 3 times. On the final wash step, the resin was dried by centrifugation at 2,800 rpm for 4 minutes at room temperature. The flow-through was discarded, and the column was placed in fresh 1.5ml LoBind tubes (Eppendorf). 100µl of the RNase I digested lysates was then carefully loaded per column and were centrifuged at 2,800 rpm for 2 minutes. 10 µl of 10% SDS was added to the flow-through of each column. RNA was extracted and purified with the Zymo Research RNA Clean & Concentrator-25 kit, using the illumina TruSeq variant of the manufacturer protocol. Briefly, the variant involved addition of 220 µl of RNA binding buffer and 495 µl of 100% ethanol to each sample, following which the purification was carried out using the standard manufacturer's instructions, with the final purified samples eluted in 30 µl of ddH<sub>2</sub>O. The number of columns required per sample was dependant on the final yield of RNA that was needed to proceed for subsequent steps. Typically, only 1-2 size exclusion columns were required per condition.

#### *2.4.3.3.3 Sucrose cushion*

For monosome selection using sucrose cushions, 300 µl of digested samples were carefully loaded onto 900 µl of sucrose cushion (for illumina TruSeq protocol, Section 2.4.1) or 1 ml of sucrose cushion (for Ingolia's protocol, Section 2.4.1) in polycarbonate ultracentrifuge tubes (13 mm × 51 mm; Beckman, cat. no. 349622) and centrifuged in a TLA100.3 rotor at 70 krpm for 4 hours at 4°C using an Optima MAX XP ultracentrifuge (Beckham Coulter). Following centrifugation, the supernatant was carefully removed, and the glassy pellet was resuspended in 100 µl ddH<sub>2</sub>O, to which 10 µl of 10% SDS was added and mixed. RNA was then extracted and purified with the Zymo Research RNA Clean & Concentrator-25 kit using the same protocol as in Section 2.4.3.3.2.

#### *2.4.3.4 rRNA depletion*

Following the 80s monosome selection and ribosome protected fragment (RPF) RNA extraction using either sucrose cushions or sucrose gradient fractionation, both total RNA and the RPF samples were subjected to an rRNA depletion step using the Ribo-Zero Magnetic Gold kit (illumina). Again, the illumina TruSeq variant of the manufacturer protocol was used. Briefly, 4 µg of both the total and the RPF RNA were used as input. Preparation of the magnetic beads was carried out according to the manufacturer's protocol in 1.5 ml LoBind microcentrifuge tubes (Eppendorf), following which a separate reaction mix containing the 4 µg of RNA sample, 4 µl of Ribo-Zero reaction buffer, and 10 µl of Ribo-Zero Gold rRNA removal solution was combined in a 0.2 ml LoBind microcentrifuge tube (Eppendorf) and made up to 40 µl total volume with

RNase-free water. This reaction volume was then gently mixed by pipetting and incubated at 68°C for 10 minutes followed by 5 minutes incubation at room temperature. After incubation the reaction mix was transferred to microcentrifuge tubes containing washed magnetic beads and mixed by gentle pipetting at least 10 times. Each sample tube was immediately vortexed at a medium speed for 10 seconds, after which the samples were incubated at room temperature for 5 minutes (omitting the 50°C incubation step as per the illumina TruSeq variant protocol, which improves the removal of small fragmented rRNA).

Ribo-Zero treated total and RPF RNA samples were then purified with the Zymo Research RNA Clean & Concentrator-5 kit, using the illumina TruSeq variant of the manufacturer protocol. Briefly, this variant of the protocol involved bringing the sample volume up to 100 µl with RNase-free water. In total RNA samples, 100 µl of RNA binding buffer and 100 µl of 100% ethanol were then added. For the RPF RNA samples, 200 µl of RNA binding buffer and 450 µl of 100% ethanol were added. Purification was carried out using the standard manufacturer's instructions, with the final purified samples eluted in 21 µl and 11 µl of RNase-free water for the total and the RPF RNA samples respectively, yielding a recovered volume of roughly 20 µl for the total RNA samples and 10 µl for the RPF RNA samples.

#### *2.4.3.5 Size selection*

rRNA depleted RPF samples were size selected using a 15% Novex® TBE-Urea precast gel (Invitrogen). 5µl of samples was mixed with 5 µl of 2x Novex® TBE-Urea Sample Buffer (Invitrogen) and was denatured at 95°C for 5 minutes. 5µl of markers (made up of 28mer and 34mer at 500 nM) was also prepared using the same method. All samples were placed on ice and were loaded within 10 minutes of the denaturing step. Prior to loading, the wells of the precast gels were thoroughly flushed through with 1x TBE buffer to remove precipitated urea. Once samples were loaded, gel electrophoresis was run at 150 V at room temperature until the bromophenol blue marker band had moved off the bottom of the gel. The gel was then incubated in 1x TBE containing 1x SybrGold (Invitrogen) for 5 minutes at room temperature, and visualised using a Safelmager (Invitrogen). Gel regions that lie within the two marker bands for the appropriate samples were excised, the resulting gel slice was shredded by centrifugation through a 0.5 ml LoBind microcentrifuge tube with a hole punched through the bottom using a sterile 20-gauge needle, into a 1.5 ml Lobind tube at 12,000 xg for 2 minutes. The resulting shredded gel slice was eluted in 400 µl 300 mM NaCl, 1 mM EDTA, 0.1% SDS overnight at 4°C on a vertical rotating daisy wheel. The eluate was filtered using Costar Spin-X 0.45 µm centrifuge tubes (Sigma) at 2,000 xg for 3 minutes at 4°C to remove any gel fragments. The filtered solution

was transferred to a fresh 1.5 ml LoBind tube, mixed with 2 µl GlycoBlue, 1 ml ethanol and left to precipitate at -20°C overnight. The precipitated RNA was pelleted by centrifugation, washed with 75% ethanol and resuspended in 10 µl nuclease free water as described in Section 2.4.3.3.1.

Size selected and non-size selected samples were analysed using a small RNA chip on a Bioanalyzer 2100 (Agilent) to check for sample degradation and sample concentration.

Marker name	Sequence
28mer	AGCGUGUACUCCGAAGAGGAUCCAACGU
34mer	GGCAUUAACGCGAACUCGGCCUACAAUAGUGACGU

#### 2.4.3.6 Library preparation

Following sample processing and quality checks using the Bioanalyzer 2100, library preparation was performed by Carolyn Jones and Dr. Kate Dudek (Genomics Facilities, MRC Toxicology Unit). For the processed RPF samples, this was performed using the TruSeq Small RNA Library Prep kit (illumina) using the manufacturer's protocol; for the processed total RNA samples, this was performed using the TruSeq Stranded mRNA LT Library Prep kit (illumina) also using the manufacturer's protocol. A schematic overview of the steps involved in illumina sequencing is shown in Figure 2.1.

#### 2.4.3.7 Sequencing

Sequencing was performed on an illumina NextSeq 500 platform by the Department of Biochemistry at the University of Cambridge. All 12x RPF libraries were multiplexed, pooled and run on 1x sequencing lane. The 12x total RNA libraries were multiplexed, pooled and run on 2x sequencing lanes. Sequencing was conducted using 75 bp single-end cycle on the high output mode, following which all reads were de-multiplexed and stored on the illumina Base Space platform before being transferred to local storage. A schematic overview of the steps involved in illumina sequencing is shown in Figure 2.1.

#### *2.4.3.8 Bioinformatics analysis of sequencing data*

##### *2.4.3.8.1 Sequencing reads quality checks*

Output from the de-multiplexed sequencing data resulted in reads from all samples being stored across four separate zip compressed FASTQ files. These were first concatenated to form a single compressed FASTQ file per sample.

Following concatenation, reads for all samples were passed through FastQC (Andrews, 2016) to generate comprehensive quality control reports for each condition. Particular attention was paid to the outputs for “per base sequence quality” and “overrepresented sequences”, with the former being measured by a Phred quality score. This logarithmic Phred score (measured as Q values) represents the quality of the base call and the probability of an incorrect call, and thus is used to provide a quantitative measure of the accuracy of the sequencing, with a cut-off value of Q20 (99% accurate) being the accepted norm for most applications. For the sequencing data that was generated for this project, almost all of the reads had a Phred score of Q30+ (99.9%+ accurate), with only a small fraction dropping below that level, and none below a score of Q20.

##### *2.4.3.8.2 Adapter trimming*

Due to the length of the ribosome protected fragments that were being sequenced (~28 - 30 nt), a large portion of the 75 bp read cycle would have been composed of the illumina barcode and adapter sequences (more detailed description in Section 7.1). To a much lesser extent, this also applied to the total RNA sequence reads. To enable the use of bioinformatics analysis packages, these adapters and barcode sequences must first be removed from the sequencing reads data. This was achieved with the program cutadapt (Martin, 2011), using the appropriate reference sequences from the illumina Adapter Sequences Nov 2015 guide (details listed in the appendix, Section 7.1), and using a default error tolerance rate of 10%. Following adapter trimming, all samples were re-run through FastQC (Andrews, 2016) to verify the resulting reads quality.

##### *2.4.3.8.3 Removing rRNA and tRNA contaminating sequencing reads*

Removal of contaminating rRNA and tRNA sequencing reads was done in a two-step process, this was to allow for the collection of detailed statistics on the individual removal rates for both rRNAs and tRNAs. This process was performed using the program bowtie2 version 2.2.5 (Langmead & Salzberg, 2012) and a custom bash script to process each sample to first remove rRNA contaminating reads, followed by tRNA reads. The required bowtie2 indices were built using the reference dataset for abundant human ribosomal sequences from the illumina iGenome *homo sapiens* hg38 UCSC dataset, and from the UCSC tRNA hg38 database

(<http://gtrnadb.ucsc.edu/>) for the rRNA and tRNA indices respectively. Removed rRNA and tRNA reads were written to separate zip compressed FASTQ files, with only the uncontaminated reads taken forward for further processing.

#### *2.4.3.8.4 Alignment to reference genome*

Sequence alignment to the reference genome was carried out using TopHat version 2.1.0 (Trapnell, Pachter, & Salzberg, 2009) and custom bash code (listed in Appendix, Section 7.2.2), with both the reference gene annotation (GTF) file and the Bowtie2 indices for the genome, being provided by the illumina iGenome *homo sapiens* hg38 UCSC dataset. Sequence alignment was performed on an Intel Core i7 desktop with access to 16GB of RAM running Ubuntu 14.04 LTS, taking approximately 1 - 2 hours to process per sample.

#### *2.4.3.8.5 Counting reads and differential gene expression analysis*

Aligned sequencing reads mapping to known genomic features were counted with HTSeq-count (Anders, Pyl, & Huber, 2015), using the gene annotation file from the illumina iGenome *homo sapiens* hg38 UCSC dataset, and with the HTSeq counting mode set to the default of “union”. This results in the generation of a tab delimited text file containing a list of genes and their respective counts for each sample. The raw count matrices were imported into R and transformed to form a single data frame for all the total RNA counts, and another for the RPF counts.

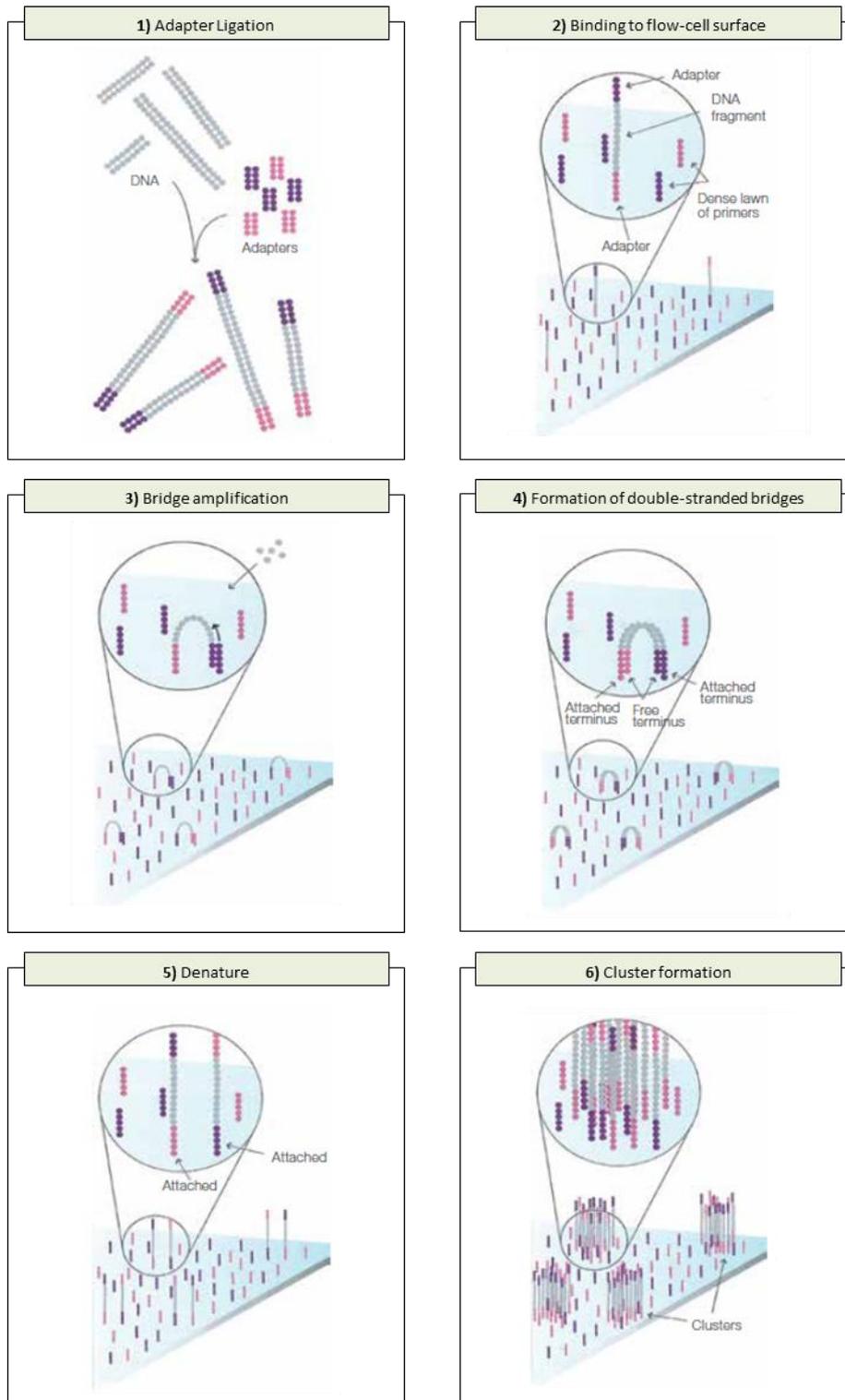
Using the raw counts from the total RNA samples, differential gene expression analysis was conducted using the R package DESeq2 (Love, Huber, & Anders, 2014). Scripts used for both procedures above are listed in the Appendix, Section 0.

#### *2.4.3.8.6 Translatome analysis*

Translatome analysis was performed using the R package Babel (Olshen et al., 2013b). Briefly, this analysis pipeline involved using both the total RNA and the RPF raw count matrices as input, which were then normalised and modelled using a negative binomial distribution. The package then primarily outputs three things, a list of identified genes that have an unusual RPF/total RNA count ratio within conditions; a combined p-value across repeats within the same condition; and finally, a list of identified genes where the RPF/total RNA count ratios were significantly different between conditions by using combined p-values. Unless stated otherwise, an FDR cut off of 10% was used for the analysis.

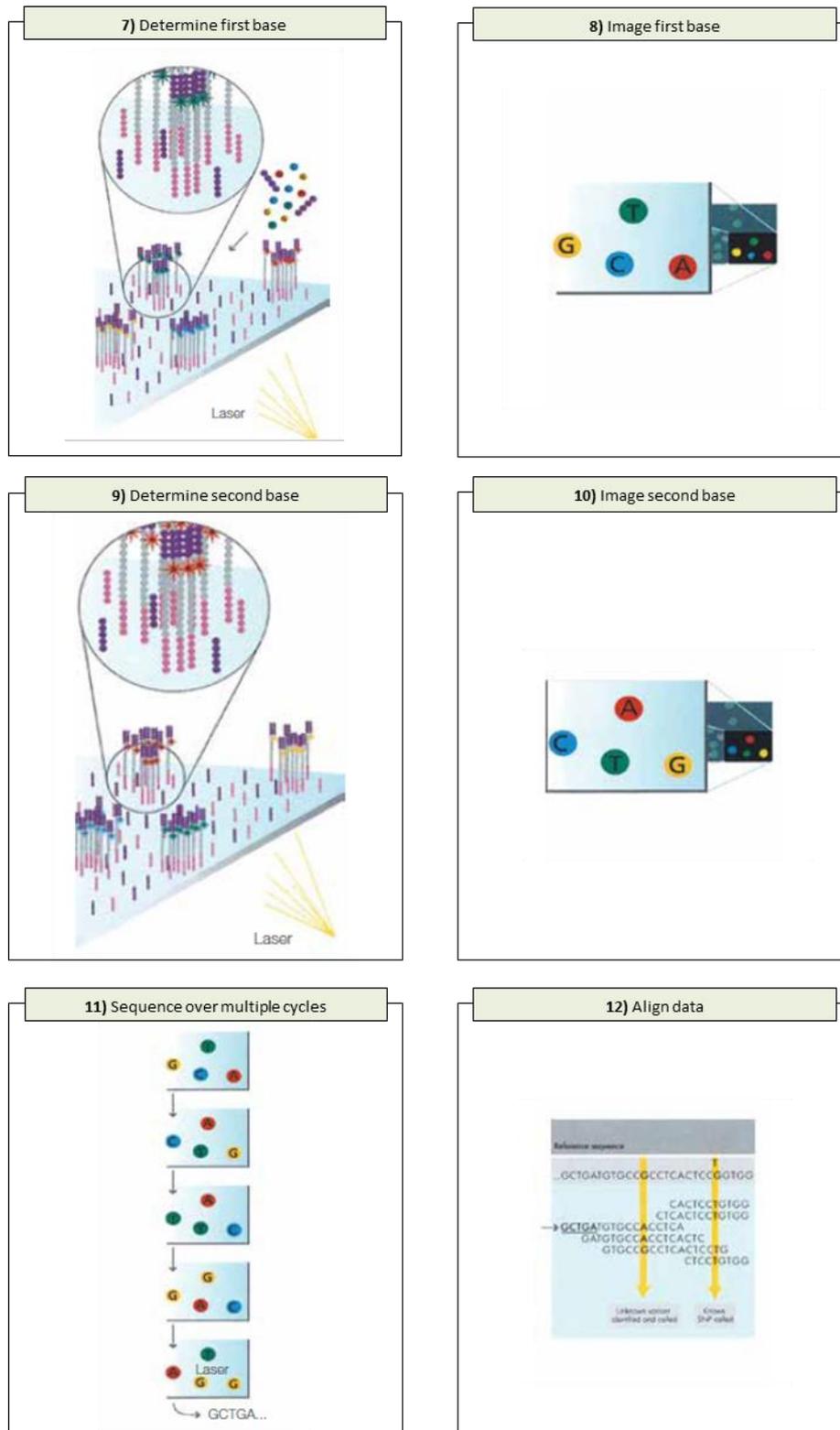
#### *2.4.3.8.7 Sequence periodicity mapping*

Periodicity plots and analysis were performed by Dr Ruth V. Spriggs, using custom Perl scripts on the processed and aligned BAM files generated from Section 2.4.3.8.4. Briefly, the number of reads mapped at each position relative to the start or the end of the coding sequence (CDS) were summed, and plotted in graphs displaying the overall normalised reads (reads per million) between positions -60 and +60 nucleotides relative to the start and end of the CDS. This was done for individual conditions, averaged across biological repeats and also averaged across all repeats to check for abnormal sequencing results.



**Figure 2.1 | illumina RNA-seq overview**

Schematic overview of the steps involved in illumina-based DNA sequencing. First, adapters are ligated onto the genomic fragments (1), the single-stranded fragments are then bound to the surface of the flow cell (2). Unlabelled nucleotides and enzymes are then added to initiate the solid-phase bridge amplification process, following which they are denatured and the process repeated until dense clusters of identical sequences are generated (3-6). Figures were adapted from [www.illumina.com](http://www.illumina.com) and TruSeq RNA and DNA Library Preparation Kits v2 Data Sheets (illumina).



**Figure 2.1 | illumina DNA-seq overview (continued)**

Schematic overview of the steps involved in illumina-based RNA sequencing. Following cluster generation, the sequencing cycle is initiated by the addition of labelled reversible terminators, primers and DNA polymerase (7), the emitted fluorescence (and thereby the first base) of each individual cluster following laser excitation is then recorded and the fluorophores are cleaved (8). This process is repeated until the sequencing cycle is complete (9-11). The individual reads are then aligned to a reference genome (12) and can be used for further analysis as described in detail in Section 2.4.3.8. Figures were adapted from [www.illumina.com](http://www.illumina.com) and TruSeq RNA and DNA Library Preparation Kits v2 Data Sheets (illumina).

## **2.4.4 Northern blot analysis**

### *2.4.4.1 Preparation of samples*

RNA samples for northern blot analysis were obtained by either Trizol RNA extraction, using 1 ml of Trizol Reagent (Sigma) per 10 cm plate using the manufacturer's protocol, or by processing the collected fractionated polysome profiling samples as per the protocol in Section 2.4.2.3.

For each processed RNA sample, 5 - 10  $\mu$ l were taken and added to a fresh 0.2 ml LoBind microcentrifuge tube, along with 2  $\mu$ l of 10x MOPS solution, 3.5  $\mu$ l formaldehyde, 10  $\mu$ l formamide and 2  $\mu$ l formaldehyde gel-loading buffer. These samples were incubated at 65°C for 15 minutes and were placed immediately on ice. Prior to loading onto gels, all tubes were briefly centrifuged to deposit all the sample to the bottom of the tubes.

### *2.4.4.2 Gel electrophoresis separation*

Samples were typically ran on 1% agarose gels, formed by dissolving 1 g of RNase free agarose in 72.5 ml of RNase free water, to which 17.5 ml of formaldehyde and 10 ml of 10x MOPS solution were added and allowed to set in gel trays. Once set, the gels were transferred to appropriate tanks and submerged in 1x MOPS solution. The denatured RNA samples were then carefully loaded into the submerged wells of the gel and were run at 100 V at room temperature for 1 to 2 hours depending on the level of separation required.

### *2.4.4.3 Transfer and crosslinking*

RNA from the gel was passively transferred onto Zeta-Probe membranes (Bio-Rad) through capillary action overnight for a minimum of 14 hours. Briefly, the capillary action was setup using 20x SSC transfer buffer soaked 3MM blotting paper as a wicking material, above which the gel was placed. The Zeta-Probe membrane was first washed in RNase-free water, then soaked in 20x SSC before placed on top of the gel. Three more sheets of 3MM blotting paper soaked with RNase-free water was placed on the Zeta-Probe membrane. An impermeable border along the edges of the gel slice was made using parafilm (Sigma) to force capillary action only through the Zeta-Probe membrane. The setup was completed with the addition of absorbent tissues on top with the entire stack compressed using an evenly distributed 1kg weight on top.

Following an overnight transfer, RNA was UV crosslinked to the Zeta-Probe membrane using a Stratalinker (Stratagene) oven set at 1200  $\mu$ J/cm<sup>2</sup>. The crosslinked RNA was stained using methylene blue stain (Section 2.4.1), and visualised following repeated washing in RNase-free water. Positions of the ribosomal RNA bands as well as the wells were marked using a pencil.

The stain was subsequently removed by several washes in 1x SSC containing 1% SDS for no longer than 15 minutes. Membranes were washed twice in RNase-free water and stored dry.

#### *2.4.4.4 Radioactive probe incorporation and hybridisation*

Radioactive probes were made using 30 ng of template DNA in 15 µl of nuclease-free water, which was denatured at 95°C for 3 minutes then placed on ice. Following the denaturing step, 5 µl of 5x labelling buffer (Promega), 1µl bovine serum albumin (BSA), 0.5 µl of 25 mM dNTP mix (containing just dATP, dTTP and dGTP), 1 µl of Klenow fragment polymerase (Promega) and 2.5 µl [ $\alpha$ -<sup>32</sup>P] dCTP (10 mCi/ml; Hartmann Analytic) were added and incubated at 37°C for 1 hour. Following incubation, the mixture was filtered through a Sephadex G-50 microspin size exclusion column (GE Life Sciences) to exclude unincorporated radiolabelled dCTP. The flow-through was denatured at 95°C for 3 minutes and placed on ice until required or stored at -20°C for future use.

For probe hybridisation, the northern blot membranes were first pre-hybridised in 10 ml of Church Gilbert solution (Section 2.4.1) in a rotating incubator oven for a minimum of 30 minutes at 65°C. The cooled purified denatured radiolabelled probe was added to the Church Gilbert solution and the membranes were allowed to hybridise overnight at 65°C in the rotating incubator oven.

#### *2.4.4.5 Membrane washing and exposing*

Following hybridisation, the blots were washed twice each in three separate wash solutions at gradually decreasing temperatures by rotation. The first two washes were done in 2x SSC, 0.1% SDS, followed by two more in 0.5x SSC, 0.1% SDS and finally twice in 0.1x SSC, 0.1% SDS solutions. Membranes were air dried, wrapped in cling film and exposed to phosphoimaging plates (Fujifilm) for hours or days depending on the signal strength. These phosphoimaging plates were visualised using a Typhoon FLA 9500 phosphorimager (GE Life Sciences).

### **2.4.5 Northern blot analysis for small RNA**

#### *2.4.5.1 Preparation of samples*

RNA samples for small RNA northern blot analysis were obtained by using a miRNeasy RNA extraction kit (Qiagen) from unprocessed 10 cm plates of PBS washed cells, RNase I digested lysates, size exclusion column purified RNase I digested lysates (as per the protocol in Section 2.4.3.3.2) or sucrose cushion purified RNase I digested lysates (as per the protocol in Section 0) using the manufacturer's protocol.

RNA samples were normalised to the same quantity using nuclease free water. Typically, between 3 - 10 µg of RNA was used per sample. Following normalisation, equal volumes of stop solution was added (Section 2.4.1) and samples were denatured at 80°C for 10 minutes and cooled on ice until required for loading.

#### *2.4.5.2 End-labelling ladders*

10 bp ladders (Invitrogen) were used for size reference, these were end-labelled with [ $\gamma$ -<sup>32</sup>P] ATP (10 mCi/ml). 2 µl of 10 bp ladders was combined with 25 µCi [ $\gamma$ -<sup>32</sup>P] ATP, 1 µl T4 PNK buffer and 1 µl T4 PNK (NEB) in 10 µl total volume with nuclease free water and incubated at 37°C for 30 minutes. Following incubation, the total volume was made up to 50 µl with nuclease free water and the mixture was filtered through a Sephadex G-50 microspin size exclusion column (GE Life Sciences). 50 µl of stop solution (Section 2.4.1) was added to the flow-through. The end-labelled ladders were kept on ice until required or stored at -20°C.

#### *2.4.5.3 Gel electrophoresis separation*

All samples were run on 10-well 15% TBE-Urea Novex pre-cast gels (Invitrogen) in 1x TBE running buffer along with the end-labelled ladders. All wells were thoroughly flushed through immediately prior to loading to remove precipitated urea. Once the samples were loaded, the gel was run at a constant 150 V for around 1.5-2 hours at room temperature until the dye front reached the bottom of the gel.

#### *2.4.5.4 Transfer and crosslinking*

RNA was transferred onto Hybond NX membranes (Amersham) using semi-dry transfer apparatus and ran at 12 V for 1 hour at 4°C. The Hybond NX membrane was first soaked in 0.5x TBE buffer and placed on top of the gel with the stack being sandwiched between 3 layers of 0.5x TBE soaked 3MM Whatmann blotting paper on either side.

Following the transfer, RNA was fixed to the membrane by using chemical crosslinking (Pall & Hamilton, 2008), which results in improved detection for small RNAs compared to traditional UV crosslinking. Briefly, the protocol first involved preparation of fresh chemical EDC crosslinking solution (0.16 M 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) in 0.13 M 1-methylimidazole adjusted to pH 8 with HCl), in which a single sheet of 3MM Whatmann filter paper was soaked. The Hybond NX membrane was placed on top of the soaked filter paper, with the side containing the transferred RNA not touching the filter paper. This assembly was then wrapped in cling film and incubated at 60°C for 1 hour. Following crosslinking,

the Hybond NX membrane was rinsed in nuclease free water to remove excess EDC crosslinking solution, air dried and stored, wrapped in cling film, until required for hybridisation.

#### 2.4.5.5 Radioactive probe incorporation and hybridisation

Northern blot probes were made using 50 pmol of desired oligos, which were end-labelled with [ $\gamma$ -<sup>32</sup>P] ATP (10 mCi/ml). The 50 pmol of oligos were combined with 25  $\mu$ Ci [ $\gamma$ -<sup>32</sup>P] ATP, 1  $\mu$ l T4 PNK buffer and 1  $\mu$ l T4 PNK (NEB) in 10  $\mu$ l total volume with nuclease free water and incubated at 37°C for 30 minutes. After incubation, the probes were purified following the same protocol used for the end-labelled ladders (Section 2.4.5.2). 1 $\mu$ l of the purified probe was read on a Wallac LKB RackBeta scintillation counter for 1 minute to determine the quantity of the probe required to reach a final concentration of 1 million counts per minute of probe per ml of hybridisation buffer.

For probe hybridisation, northern blot membranes were first incubated in 10 mls of ULTRAhyb Ultrasensitive Hybridisation Buffer (Applied Biosystems) for 30 minutes at 42°C in a rotating oven. The purified end-labelled probe was added directly to the buffer at a concentration of 1 million cpm/ml and left to hybridise overnight at 42°C in a rotating oven.

#### 2.4.5.6 Membrane washing and exposing

Following hybridisation, the hot ULTRAhyb buffer was discarded and northern blot membranes were washed twice for 30 minutes each at 42°C with 2xSSC, 0.1% SDS in a rotating oven. Membranes were air dried, wrapped in cling film and visualised as described in Section 2.4.4.5.

#### 2.4.5.7 Probes used

All probes were synthesised by Sigma-Aldrich (UK).

Probe	Sequence	Length
tRF5-Gly-GCC	GGCGAGAATTCTACCACTGAACCACCAA	28
tRF5-Val-AAC	GGCGAACGTGATGACCACTACACTACGGA	29
tRF3-Val-AAC	TTCCGCCCGTTTCGAACCGGGGACCTTTCGCG	33

Table 2.4 | Northern probes used

### 2.4.6 RT-qPCR using SYBRGreen

The required cDNA for qPCR was generated through reverse transcription reactions using 1  $\mu$ g of total RNA, along with 100 ng of random hexamers (Invitrogen, Cat #48190-011),

SuperScript III reverse transcriptase (RT) kit (Life Technologies, Cat #18080044), and a pre-prepared mix of dNTPs (Life Technologies, Cat #10216018, #10218014, #10217016, #10219012). RNase inhibitor RNasin Plus (Promega) was also added to minimise RNA degradation during the reverse transcription reaction.

Required quantities for a 1x reaction mix is listed below.

Component	Volume
RNA (+ddH <sub>2</sub> O for normalisation)	5 µl
Nuclease free water	6 µl
Random hexamers	0.5 µl
dNTPs mix (10 mM)	0.5 µl
5x First strand buffer	4 µl
RNasin Plus	1 µl
DTT	1 µl
SuperScript III	1 µl

**Table 2.5 | RT 1x reaction mix**

Following the RT reaction, 5 µl of the generated cDNA was used to form the template for qPCR, along with gene specific primers added at a concentration of 250 nM, in a final volume of 20 µl per well in a MicroAmp® Fast 96-well Reaction Plate (Life Tech) with 10 µl SYBR Green Master Mix (Life Tech). Fold-changes in expression between treated and untreated samples were then calculated using the  $2^{\Delta\Delta CT}$  method of normalisation. The list of the gene specific primers used is shown below. All probes were ordered from Sigma-Aldrich.

Probe	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
CHAC1	GTGGTGACGCTCCTTGAAGATC	GAAGGTGACCTCCTTGGTATCG
JUN	TTCTATGACGATGCCCTCAACGC	GCTCTGTTTCAGGATCTGGGGTTAC
ATF3	TGATGCTTCAACACCCAGGC	GGATGGCAAACCTCAGCTCT

RHOB	CAGTAAGGACGAGTTCCCG	GTCCACCGAGAAGCACATGA
FOSB	TCTGTCTTCGGTGGACTCCTTC	GTTGCACAAGCCACTGGAGGTC
BTG2	GGCTCCATCTGCGTCTTGTA	AGCACTTGGTTCTTGCAAGT
EBLN2	ACTGAGTGTGGTGTGAGCA	CACTTGATCCAGCCGCAATG
HIST1H1D	TGCTCCTACCATTCTGCAC	GGATGCTTTGCGTTTCCCAG
HIST1H3J	CTGTGCTATTCACGCCAAGC	ACCCAAGATAGAGCAGGGGA
HIST1H2AB	CGGCAAACAAGGCGGTAAAG	GCCACAGGAAACTGCAAAC
PARD6B	GGCTCCAGTGCAGGGTAAC	TCCTGGGACAAGCCTGGATA
DDIT4	GCTTACCTGGATGGGGTGTC	GCATCAGGTTGGCACACAAG

**Table 2.6 | Probes used for RT-qPCR**

### ***3. Selection and optimisation of DNA damage model***

#### ***3.1 Introduction***

As outlined in the introduction, a large number of pathways are activated following DNA damage. In general, following cell stress there is a global down-regulation of protein synthesis, coupled with a simultaneous up-regulation of a small subset of certain genes (Forrester, Li, Hovan, Ivashkevich, & Sprung, 2012; Jen & Cheung, 2003). This has been shown to be the case in response to a range of DNA damaging agents such as ionising radiation (IR), ultra-violet (UV) radiation and alkylating agents (Piccirillo, Bjur, Topisirovic, Sonenberg, & Larsson, 2014; Powley et al., 2009; Young, 2011).

Two of the most common sources of DNA damage encountered on a daily basis are IR and UV radiation. Previous investigations into DNA damage responses, which have used IR and UV, have primarily utilised tumour derived cell lines and have typically involved the use of relatively high doses of radiation (Braunstein, Badura, Xi, Formenti, & Schneider, 2009; Minafra et al., 2015; Trivigno et al., 2013; Uehara et al., 2014; Young, 2011). In the case of IR, the supra-lethal doses used are often much greater than those used in a clinical radiotherapy environment and therefore the ability to translate the data obtained into a clinical setting is limited. Nevertheless, these studies have shown that gene expression changes following exposure to IR are greater at the level of mRNA translation compared to that of transcription (Lü, De La Peña, Barker, Camphausen, & Tofilon, 2006). This possibly reflects the fact that a significant proportion of the total energy expenditure of a cell is consumed by protein synthesis (Buttgereit, F; Brand, 1995), and is therefore, perhaps not unexpectedly, this process is highly regulated and responsive to the surrounding stimuli.

Surprisingly, there are relatively few studies investigating gene expression changes following DNA damage using non-tumour derived cell lines. However, the few studies which have used non-tumour derived cell lines, have observed very different DNA damage responses between transformed and non-transformed cells (Braunstein et al., 2009; J.-H. Lee, Choy, Ngo, Foster, & Marks, 2010). In a study by the Schneider lab, it was discovered that while in non-transformed cells IR rapidly induced translational regulation, enhancing the translation of specific mRNAs, which were involved in DNA repair, as well as the activation of a number of key regulatory protein such as p53 and 4E-BP1, these responses were absent in transformed cells. The aim of this project was to further investigate the DNA damage response using the immortalised, non-transformed breast epithelial MCF10A cell line and polysome profiling and ribosome profiling.

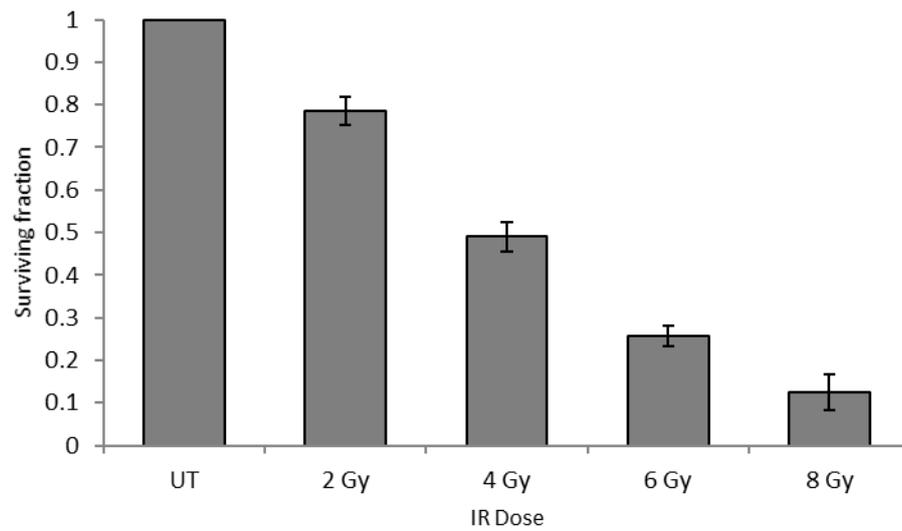
Previous studies in the Willis lab have demonstrated the viability of using both IR and UV-B irradiation as effective source of DNA damage. As such, this chapter will outline the development and optimisation of DNA damage models using both sources in succession, and investigate their feasibility for conducting further investigations using ribosome profiling.

### ***3.2 Establishment of IR treatment conditions***

To determine the effect of ionising radiation on the non-transformed MCF10A breast epithelial cells, and to optimise the dose of IR required, clonogenic assays were performed (as described in Section 2.1.5). Based on previous work which investigated the translational response in transformed HeLa and FaDu cells lines (Young, 2011), an IR dose range of 2 - 8 Gy was chosen. As expected, the long term survival rates for the MCF10A cells were directly proportional to the dose of IR that the cells were exposed to (Figure 3.1), with only approximately 10% of the cells surviving the highest IR dose of 8 Gy compared to approximately 80% at the lowest dose.

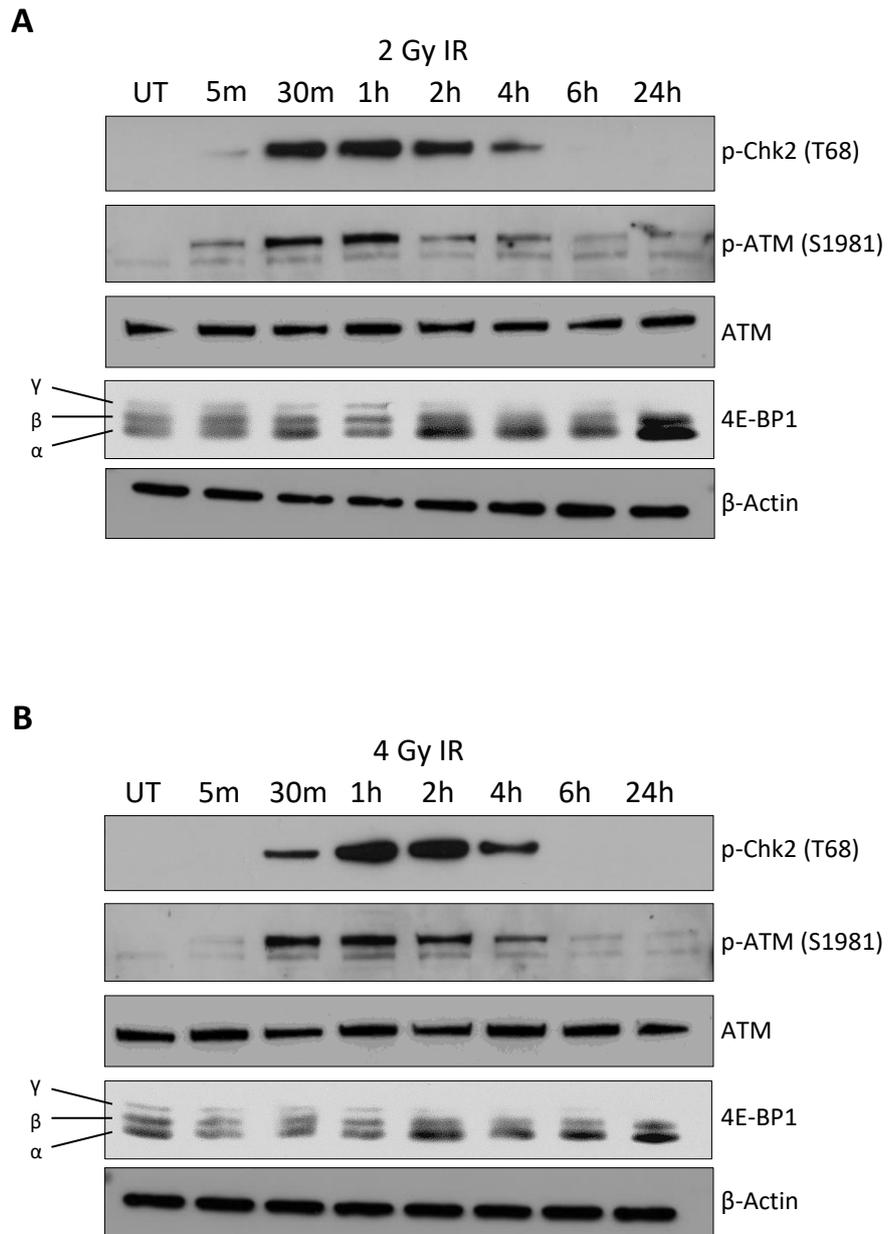
### ***3.3 Exposure to IR induces rapid induction of DNA damage and repair***

To investigate responses to IR, the sublethal doses of 2 Gy and 4 Gy were selected to study the DNA damage and repair pathways involved following irradiation. Cellular responses to IR-induced DNA damage, primarily double-strand breaks, act primarily via the ATM-Chk2 kinase signalling pathway (Sancar et al., 2004). Phosphorylation of both of these kinases occurred in a time dependent manner, with an onset from around 30 minutes and lasting until at least 4 hours post IR exposure (Figure 3.2). Western blot analysis of the differently phosphorylated forms of 4E-BP1 showed a major increase in the hypophosphorylated ( $\alpha$ ) form from 2 hours onwards for both doses of IR exposure (Figure 3.2). This suggested inhibition of global protein synthesis at these time points.



**Figure 3.1 | Dose response to ionising radiation exposure**

MCF10A cells were irradiated with an IR dose range of 2-8 Gy following the protocol in Section 2.2.1 (with UT referring to the untreated control sample), after which clonogenic assays were conducted. The numbers of surviving colonies were counted, and the summary of which is shown (n=3, error bars representing standard error of the mean).



**Figure 3.2 | Rapid induction of DNA damage and repair following IR exposure**

Identical plates of MCF10A cells were treated and harvested at the doses and time points shown, and were analysed by western blotting, with UT representing the untreated control samples, and m and h representing minutes and hours following IR exposure respectively.

### ***3.4 Global protein synthesis rates were inhibited following IR exposure***

To further assess the effect of IR exposure on translation, MCF10A cells were treated with 2 and 4Gy IR, and global protein synthesis was measured by pulse radiolabelling using <sup>35</sup>S methionine incorporation (incubated for 20 minutes prior to harvest) (Figure 3.3). Incorporation of the radiolabel was determined by liquid scintillation counting as described in Section 2.3.9. Exposure of MCF10A cells to an IR dose of 4 Gy caused a consistent and reproducible time-dependant effect on the rate of protein synthesis (Figure 3.3B).

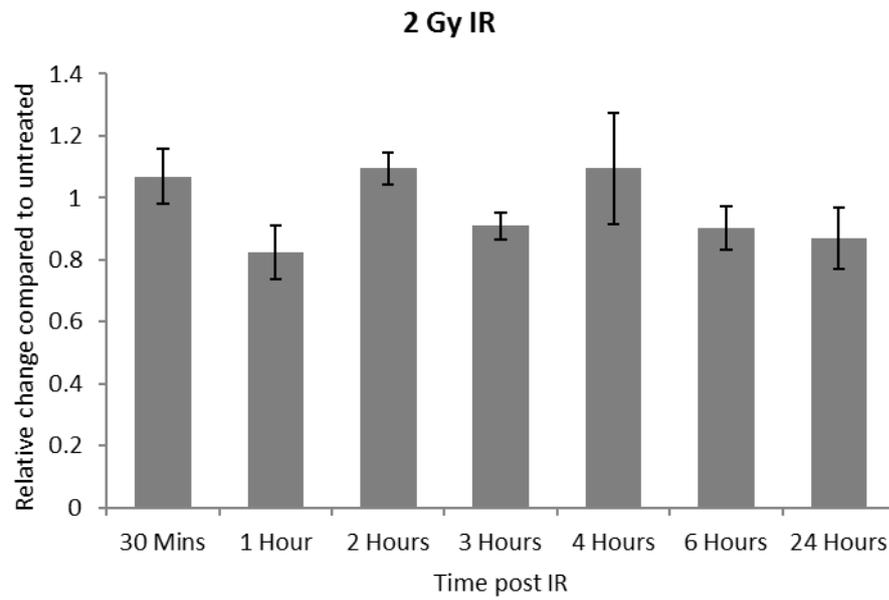
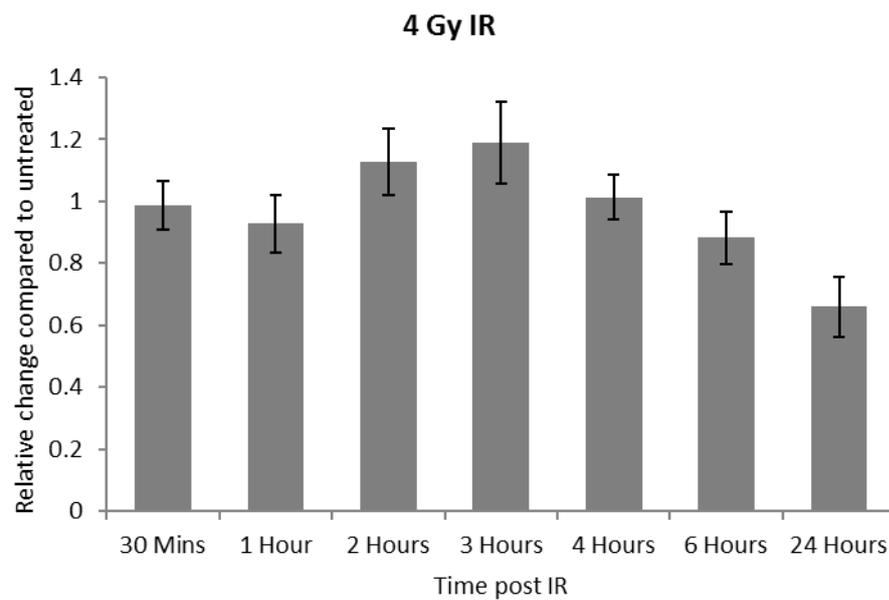
Following a 1 hour incubation post IR treatment, there was a consistent minor decrease in protein synthesis rates, which then increased to approximately 120% of that value in untreated cells by 3 hours, which was followed by inhibition to approximately 65% by 24 hours. This was consistent with the immunoblotting data, which showed an increase in the hypophosphorylated ( $\alpha$ ) form of 4E-BP1 at the later time points (Figure 3.2). Cells treated with an IR dose of 2 Gy also showed a similar, but smaller response on the individual repeats; however there was a large variation between individual experiments (Figure 3.3A).

### ***3.5 Changes in subpolysome/polysome distribution following IR exposure***

Given the response in protein synthesis rates observed following 4 Gy IR exposure, polysome profiling was conducted at time points showing maximal changes, namely at 1, 3 and 24 hours post IR treatment. Polysome profiling is a technique that relies on the separation of mRNAs based on the density of their associated ribosomes. An actively translating mRNA is like to have many associated ribosomes, which are termed polysomes. However for mRNA that is not being actively translated, or inefficiently translated, there are fewer associated ribosomes. The different number of associated ribosomes allow for the effective separation of these mRNAs by mass, using ultracentrifugation through a sucrose density gradient followed by analysis using a UV spectrometer at 254nm. These readings form a visual representation of the different ribosomal populations of the cells at the point of harvest and so can be used to compare changes in translational effects across conditions on a global scale, as demonstrated in Figure 3.4.

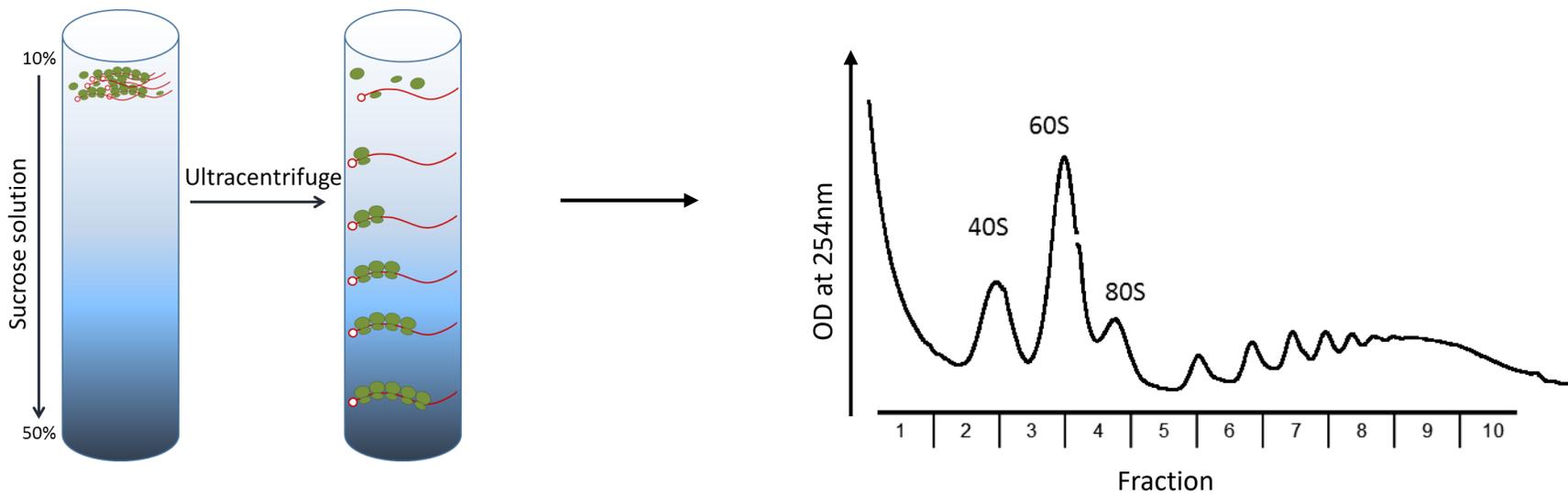
Cells were exposed to 4Gy IR and harvested at 1, 3 and 24 hours, and processed as described in Section 2.4.2. Exposure of MCF10A cells to IR resulted in a slight decrease in the amount of mRNAs associated with polysomes at the 1 hour post irradiation time point (Figure 3.5), which was consistent with earlier <sup>35</sup>S methionine incorporation results showing a minor decrease in protein synthesis rates at that time point (Figure 3.3). The difference was very minor, but it was observed consistently across three biological replicates.

By 3 hours post treatment however, there was a slightly larger change in the profiles, with a general shift from the subpolysomal fractions (the 60s peak in particular) to the polysomal regions when compared with that of untreated cells. Though the changes in the subpolysomal fractions were relatively small compared to those observed in the polysomal fractions, this could be potentially explained by slight differences in the sensitivity of the UV spectrometer across the subpolysomal and polysomal regions. The changes in the polysome profiling traces at 3 hours were consistent with the increased rates of protein synthesis observed at the same time point by <sup>35</sup>S-methionine incorporation assay (Figure 3.3). By 24 hours however, whereas <sup>35</sup>S methionine incorporation showed a dramatic decrease in protein synthesis rates to approximately 65% of untreated samples, the polysome profiles for treated and untreated cells displayed negligible differences. Whilst a marginally greater abundance of 60s ribosomal complexes was observed in the irradiated cells, no such changes were observed for the distribution of ribosomes in the polysomal fractions as might have been expected given the sharp decrease in protein synthesis.

**A****B**

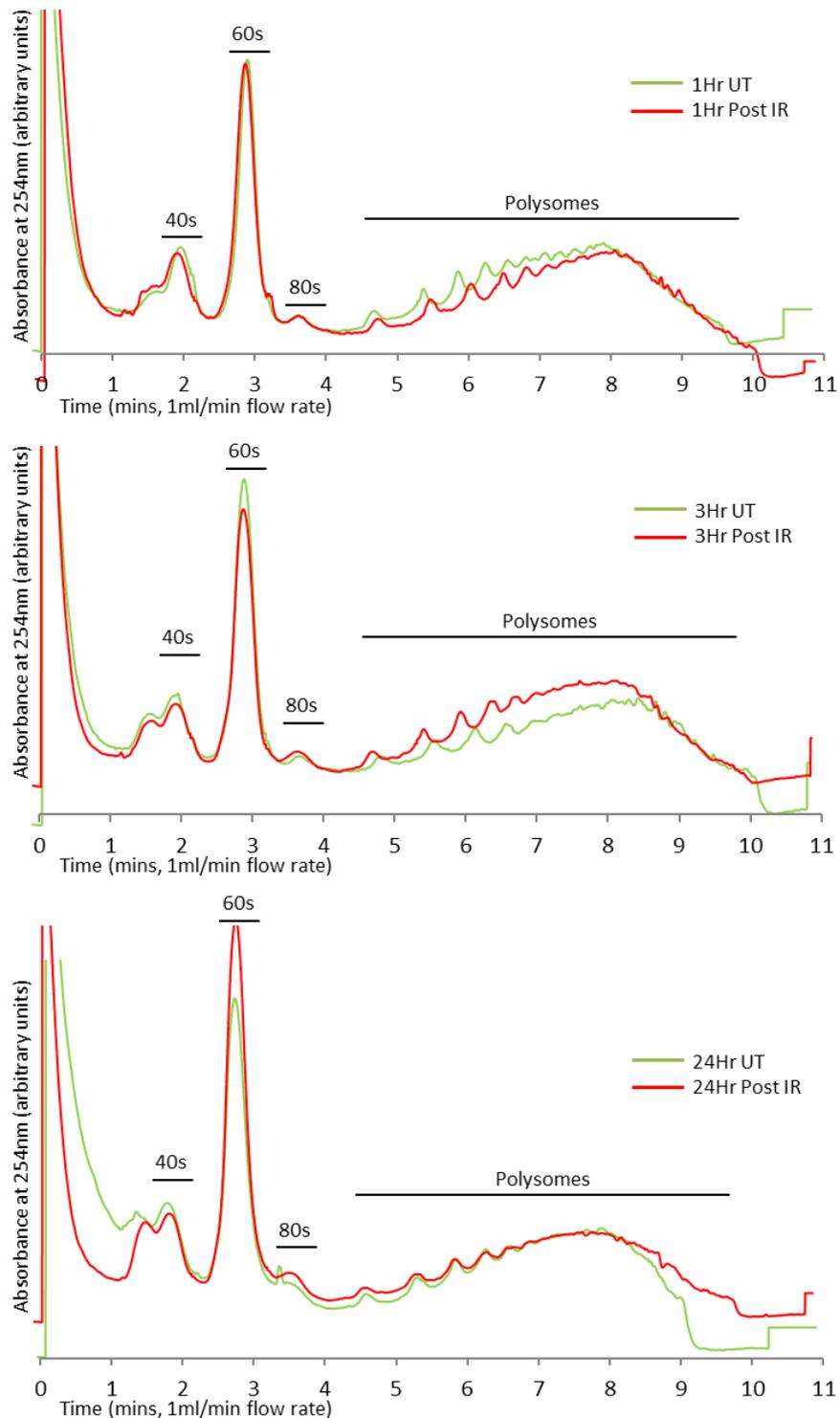
**Figure 3.3 | Changes in protein synthesis rates following IR-induced DNA damage**

MCF10A cells were treated with IR at the doses stated, protein synthesis rates were then determined by scintillation counting of incorporated <sup>35</sup>S methionine at the time points shown (for method see Section 0). Measurements shown were all represented as relative change compared to matched untreated samples, summarised across 3 biological repeats (error bars representing standard error of the mean).



**Figure 3.4 | Separation of mRNAs by sucrose density gradients**

Schematic diagram illustrating the main steps in polysome profiling. Lysates collected from cells that were pre-treated with cycloheximide were loaded onto sucrose density gradients (typically 10-50% w/v), and were separated by mass through ultracentrifugation. The content of this separated mix was then pumped through a UV/Vis detector, through which the translational status of the sample was then determined by the absorbance profile read at 254nm, generating the trace shown. The contents of the sucrose density gradient could then be fractionated, and collected for further analysis. This process is detailed in Section 2.4.2.



**Figure 3.5 | Changes in polysome profiles of cells following exposure to IR**

Identical plates of MCF10A cells were exposed to 4 Gy IR, and treated with 100 $\mu$ g/ml cycloheximide for 5 minutes prior to harvest at the time points shown (as described in Section 2.4.2.2), with UT referring to the untreated sample. Data shown were from a single repeat, but was representative of a total of 3 biological repeats. Slight variances in the alignment of traces were due to minor pipetting errors when preparing the sucrose gradient.

### ***3.6 Establishment of UV-B treatment conditions***

Similar to the IR work presented above, clonogenic assays were performed to determine the survival rates following irradiation with UV-B (with a peak wavelength of 302nm). Based on previous work conducted in the Willis lab, a dose range of between 50 and 250 Jm<sup>-2</sup> of UV-B radiation was used to investigate the effect of exposure on the long term survival of MCF10A cells.

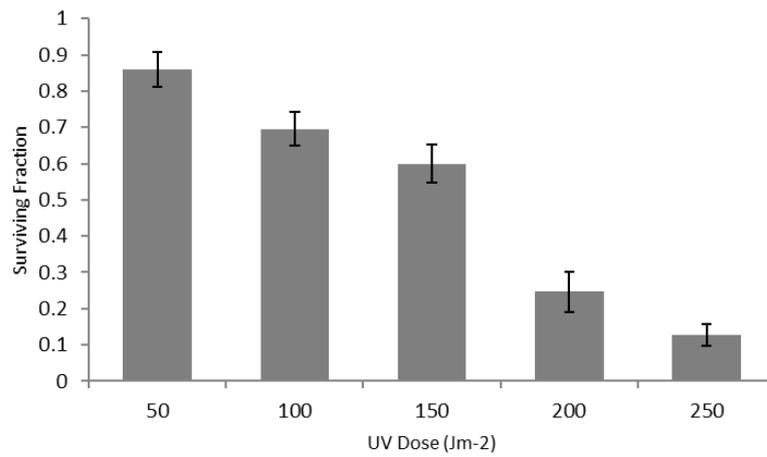
At the lower end of the dose range used, UV-B resulted in the survival of around 85% of the initial colonies (Figure 3.6). Irradiation with the highest dose tested, UV-B lead to the survival of only just over 10%, with intermediate doses following an approximately linear correlation to the survival rates.

### ***3.7 Global protein synthesis rates are inhibited following UV-B exposure***

The effect of UV-B irradiation on the global levels of protein synthesis was measured by pulse radiolabelling using <sup>35</sup>S methionine incorporation as described in Section 2.3.9. This was conducted using a slightly expanded UV-B dose range of 50 to 300 Jm<sup>-2</sup> to examine the response in protein synthesis levels at 4, 8 and 24 hour time points following treatment.

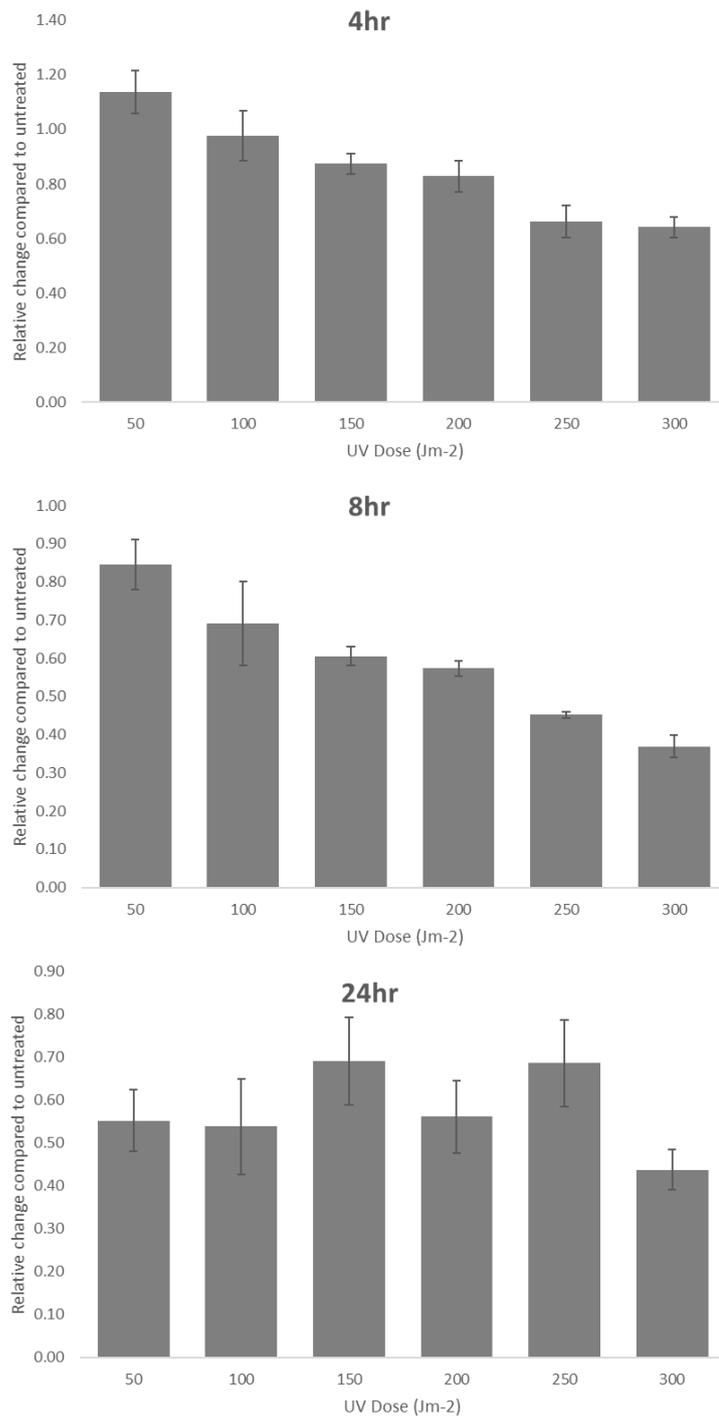
Irradiation with UV-B elicited a dose and time dependant reduction of global protein synthesis rates with the exception of the 24 hour time point (Figure 3.7). The apparent minor increase in protein synthesis rates observed following exposure to 50 Jm<sup>-2</sup> UV-B at the relatively early time point of 4 hours was most likely due to the effects of the media refresh conducted 1 hour prior to treatment, as measurements were all set relative to t=0. At 4 and 8 hours, the reduction in protein synthesis rates were dose dependent with the highest dose of 300 Jm<sup>-2</sup> resulting in a decrease in protein synthesis rates to approximately 65% by 4 hours and approximately 35% by 8 hours. For the low doses of UV-B irradiation used such as 50 and 100 Jm<sup>-2</sup>, the protein synthesis decreased to approximately 85% and 70% by 8 hours respectively and also showed greater degree of fluctuation. Similarly, by 24 hours, all doses of UV-B irradiation resulted in much higher variation in the protein synthesis rates between the biological replicates, and with no discernible overall trend. This was most likely due to the use of unsynchronised cells, with the impact of the UV-B exposure being the primary cause of the response at the earlier time points (4 and 8 hours) and the effect of cell cycle only becoming apparent by the later time point of 24 hours.

Based on the clonogenic and the  $^{35}\text{S}$  methionine incorporation assays performed, a UV-B dose of  $200 \text{ Jm}^{-2}$  was selected as the treatment condition for this project. At this dose, protein synthesis rates decreased to 83% and 57% by 4 and 8 hours respectively relative to  $t=0$ .



**Figure 3.6 | Dose response to UV-B irradiation**

MCF10A cells were irradiated with UV-B at a dose of between 50 and 250 Jm<sup>-2</sup>, following the protocol in Section 2.2.1, after which clonogenic assays were conducted. The numbers of surviving colonies were counted, and the summary of which is shown (n=3, error bars representing standard error of the mean).



**Figure 3.7 | Global protein synthesis rates inhibited following UV-B irradiation**

MCF10A cells were irradiated with UV-B at the doses stated, protein synthesis rates were then determined by scintillation counting of incorporated <sup>35</sup>S methionine at the time points shown (for method see Section 0). Measurements shown are all represented as relative change compared to matched untreated samples, summarised across 3 biological repeats (error bars representing standard error of the mean).

### ***3.8 DNA damage and translational responses to UV-B***

To investigate the time frame of the response to UV-B, western blot analyses of a number of key proteins involved in translation and DNA damage response were conducted. Initially, time points of 4 and 7 hours were selected due to the sustained dose-dependent reduction in protein synthesis following UV-B up to 8 hours (Figure 3.7). As expected, treatment with 200 Jm<sup>-2</sup> UV-B resulted in the activation of p53, at both time points, and the phosphorylation of eIF2 $\alpha$ , Chk1 and H2AX (Figure 3.8B). However, by 7 hours, cleavage of Poly (ADP-ribose) polymerase (PARP), an indicator of cell death (Kaufmann et al., 1993), was detected (Figure 3.8A). To prevent the added complication of studying cell death response as well as the response to UVB irradiation, an expanded earlier time range of 1 to 4 hours was investigated. Activation of p53 was observed from 1 hour onwards, along with the phosphorylation of Chk1 (S345) and H2AX (S139), suggesting activation of the ATR-Chk1 pathway and confirming the presence of DNA damage following UV-B (Figure 3.8C). In contrast to the responses seen following IR (Figure 3.2), levels of the activated form of 4E-BP1 remained unchanged following UV-B, but phosphorylation of eIF2 $\alpha$  (S51) was observed from 1 hour onwards in the treated samples (Figure 3.8D).

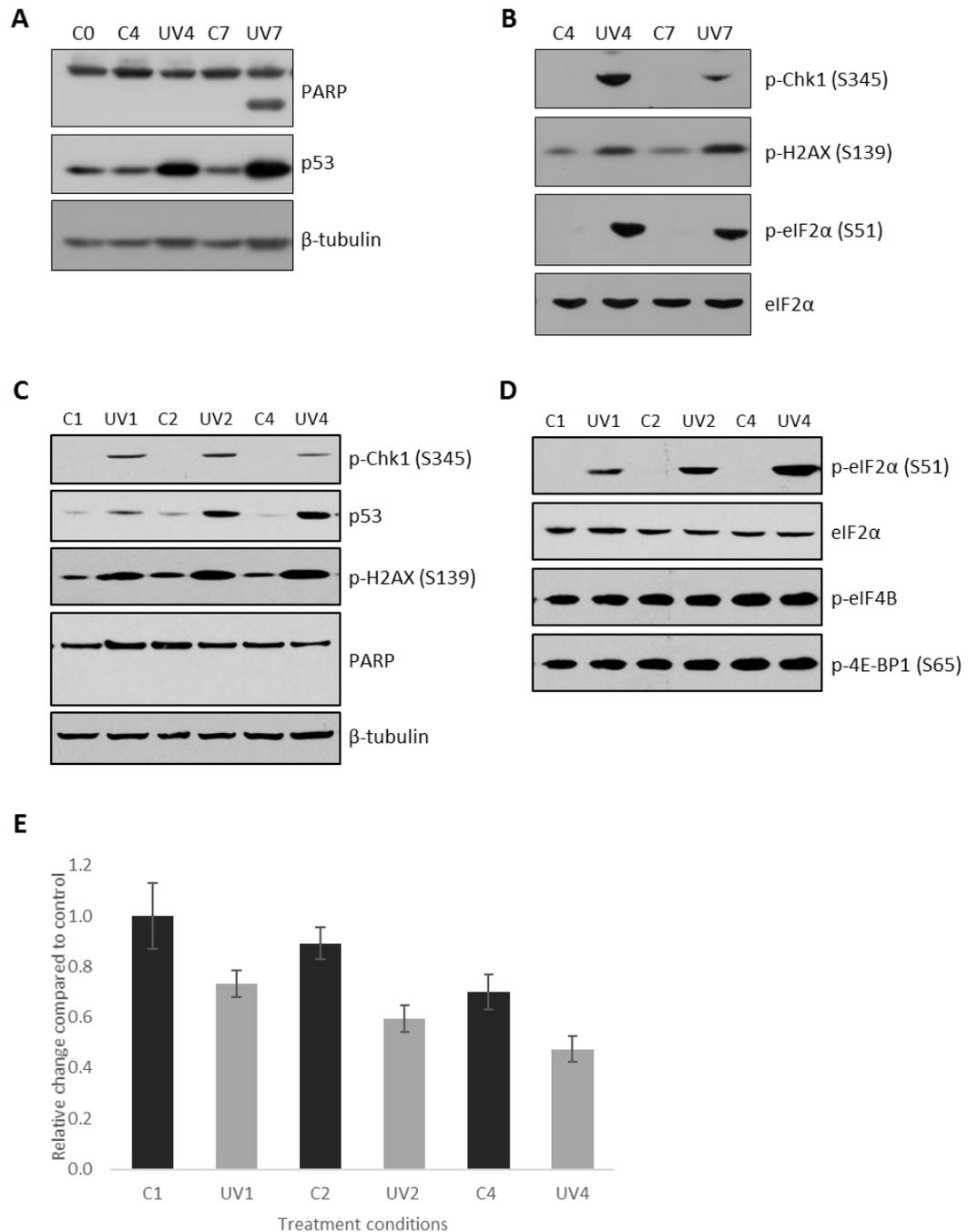
<sup>35</sup>S methionine incorporation assays performed at these earlier time points following UV-B exposure revealed a rapid inhibition of protein synthesis rates, with the levels decreasing to 73% by 1 hour following treatment when compared with the untreated control at the same time point (Figure 3.8E). The rate of protein synthesis decreased further to 60% and 48% by 2 and 4 hours post treatment respectively when compared to the untreated control at the 1 hour time point.

Due to the presence of a substantial decrease in protein synthesis rates, absence of PARP cleavage, as well as the increased phosphorylation of eIF2 $\alpha$ , an early time point of 1 hour post treatment was selected for further analysis using the polysome profiling method described previously (Figure 3.4). In addition, a later time point of 4 hours following treatment was also chosen.

### ***3.9 Substantial shifts in polysome profiles following UV-B exposure***

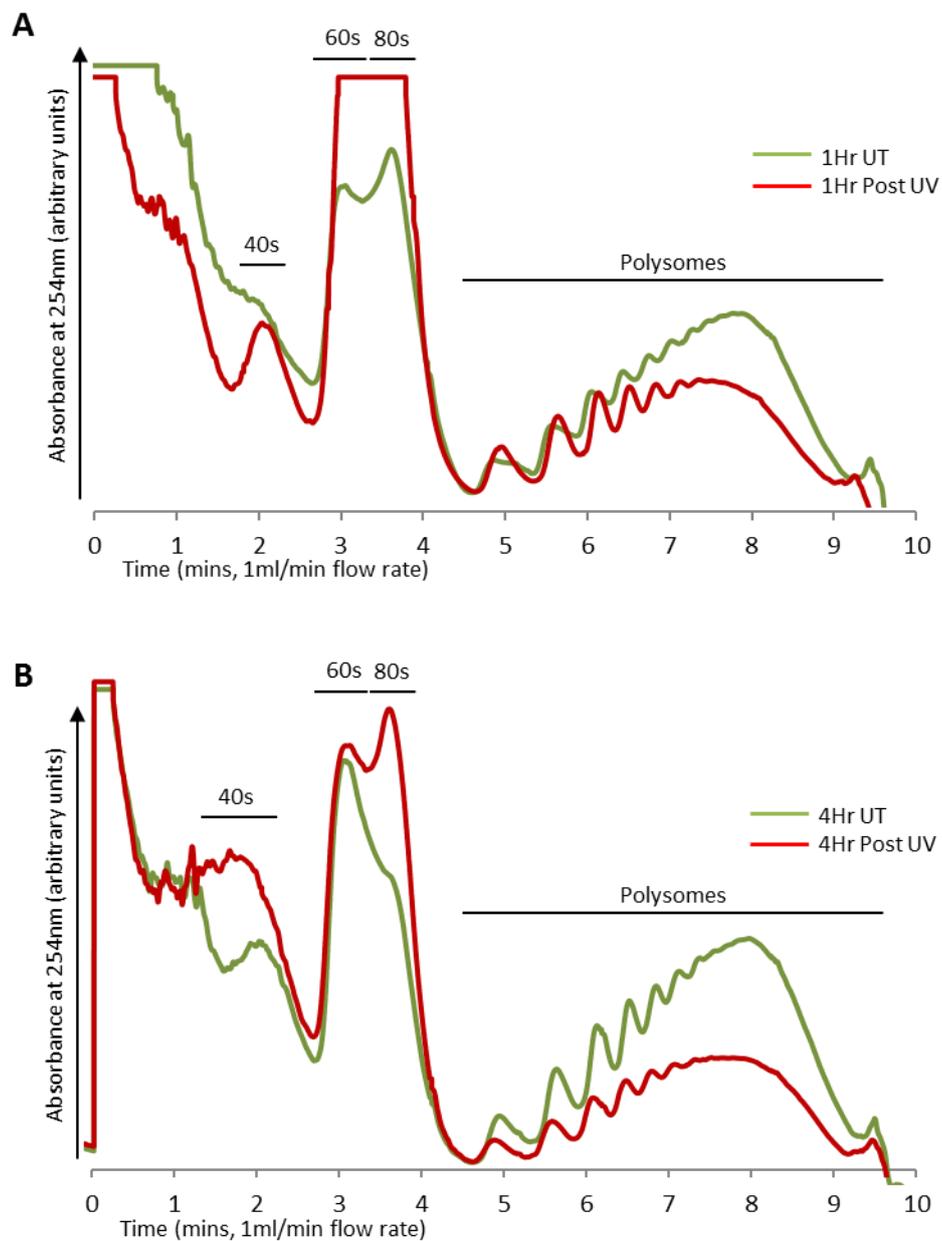
MCF10A cells were irradiated with 200 Jm<sup>-2</sup> UV-B, after which lysates were harvested at the 1 and 4 hour time points using the protocol described in Section 2.4.2. The resulting traces from polysome profiling at both time points showed a significant decrease in the amount of mRNAs associated with polysomes, and a corresponding increase in the amount of mRNAs associated with the subpolysomal fractions was observed at the 1 hour time point (Figure 3.9A). This

observation was consistent with the decrease to 73% of that of the matched control samples in protein synthesis rates 1 hour following UV-B treatment. By 4 hours post treatment, the decrease in the numbers of associated polysomes was even more substantial, again consistent with the methionine incorporation results (Figure 3.7, Figure 3.8D). Interestingly, the significant decrease in the polysomal fractions at 4 hours, did not appear to be associated with a large increase in the subpolysomal fractions following treatment. While this could have perhaps arisen due to non-linear sensitivity of the UV/Vis detector across the whole gradient sample, this seems unlikely given the presence of the large increase in subpolysomal fractions that were detected at 1 hour following treatment. A possible cause of this result could be due to potential disruption of the 80s complex at this later time point resulting in a loss of intact 80s monosomes. Indeed, an increase in the quantities of 40s ribosomal complexes was observed by 4 hours, but not at 1 hour (Figure 3.9).



**Figure 3.8 | DNA damage and translational responses to UV-B**

**A-D)** Identical plates of MCF10A cells were treated and harvested at the time points shown, and were analysed by western blotting, with C and UV representing the untreated control and UV-B samples respectively, followed by the time point. For example, C1 refers to the untreated control sample at the 1 hour time point. **E)** Protein synthesis rates were determined by scintillation counting of incorporated <sup>35</sup>S methionine at the time points shown (for method see Section 2.3.9), using the same nomenclature as above. Measurements shown were all represented as relative change compared to the untreated samples at 1 hour, and summarised across 3 biological repeats (error bars representing standard error of the mean). Western blotting in panels B, C and D were kindly performed by Dr. Emilie Horvilleur and Dr. Tuija Poyry.



**Figure 3.9 | Substantial shifts in the polysome profiles of cells following irradiation with UV-B**

Identical plates of MCF10A cells were irradiated with  $200\text{Jm}^{-2}$  UV-B, and treated with  $100\mu\text{g/ml}$  cycloheximide for 5 minutes prior to harvest at **A)** 1 hour and **B)** 4 hour time points (as described in Section 2.4.2.2). Data shown were from a single repeat, but were representative of a total of 3 biological repeats. Slight variances in the alignment of traces were due to minor pipetting errors when preparing the sucrose gradient.

### ***3.10 Early translational reprogramming requires phosphorylation of eIF2 $\alpha$***

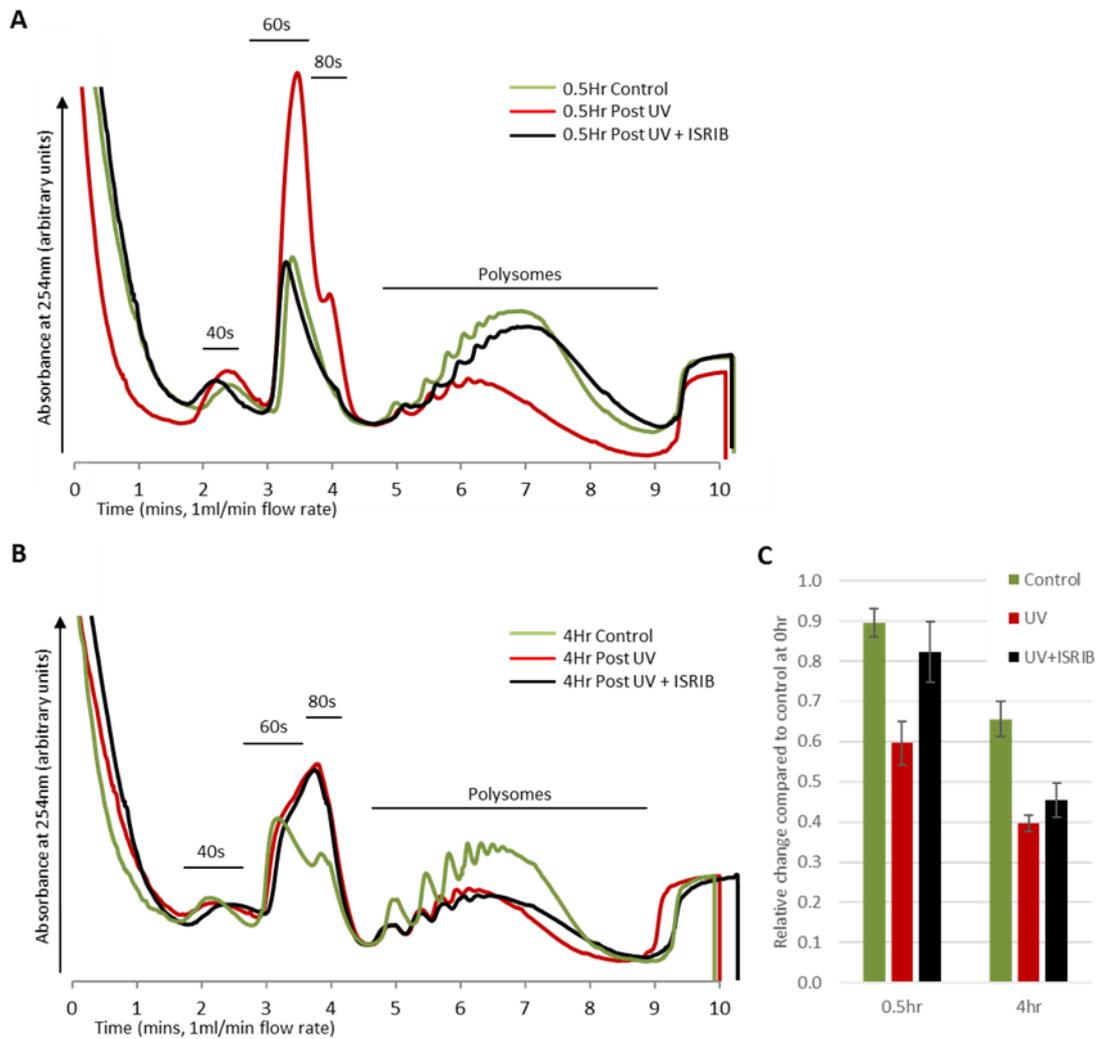
UV-B elicited translational reprogramming at both the 1 and 4 hour time points (Figure 3.9), as such, further investigations were conducted to the key regulators involved. As mentioned previously, a number of treatment conditions, including UV-B, activate an integrated stress response. A key regulator of this stress response is eIF2, which as discussed in the introduction, forms part of the ternary complex that is required to bring the initiator methionine to the start codon. Phosphorylation of eIF2 on the alpha subunit decreases ternary complex formation. Phosphorylation of eIF2 $\alpha$  was observed from 1 hour to 7 hours following UV-B (Figure 3.8).

To investigate the role of eIF2 $\alpha$  in the translational responses observed following UV-B, a relatively recently identified potent inhibitor of the integrated stress response (termed ISRIB) was used (Sidrauski et al., 2013; Sidrauski, McGeachy, Ingolia, & Walter, 2015). This drug was identified using a cell based screen for inhibitors of PERK signalling, and was discovered to potently reverse the effects of eIF2 $\alpha$  phosphorylation ( $IC_{50} = 5$  nM), as well as reducing the viability of cells which were subjected to PERK activation through chronic ER stress (Sidrauski et al., 2013). It was also found to readily cross the blood-brain barrier as well as exhibiting no obvious toxicity in mice. Experiments conducted using ISRIB treated, ER-stressed cells showed that ISRIB itself had no effect on the phosphorylation levels of eIF2 $\alpha$ , instead, it is thought to act downstream of eIF2 $\alpha$ , resulting in the insensitivity of cells to the effects of eIF2 $\alpha$  phosphorylation. Although not fully understood, it is thought the actions of the molecule ISRIB are brought about by weakening the interaction of phospho-eIF2 $\alpha$  with eIF2B, or through the increase in activity of eIF2B itself, ultimately resulting in an increase in the available levels of ternary complex (Sidrauski et al., 2013).

Polysome profiles were conducted at two time points following treatment with ISRIB, kindly provided by the Mallucci lab. Time points selected were 4 hours, and also the earliest time point in which a UV-B response was detected, which corresponded to 30 minutes following treatment (Figure 3.10). The resulting polysome profiling traces obtained at the 30 minute time point were consistent with the previous traces taken at 1 hour, showing large reductions in the amount of mRNAs associated with polysomes, and a corresponding increase in the amount of mRNAs associated with the subpolysomal fractions (Figure 3.9). Interestingly, the addition of ISRIB at this earlier 30 minute time point resulted in the near complete reversal of the observed UV-B responses, resulting in a polysome profiling trace that was similar to the matched control sample

(Figure 3.10A). The reversal of the UV-B responses in the ISRIB treated samples at 30 minutes was also reflected in the almost complete restoration of protein synthesis to approximately untreated levels as measured by <sup>35</sup>S methionine incorporation, inducing an increase from 60% in the UV-B only samples to over 80% in samples treated with UV-B and ISRIB compared to approximately 90% in the matched untreated samples (Figure 3.10C). By 4 hours however, the impact of ISRIB was reduced, causing only a minor increase in the numbers of mRNA associated with the heavy polysomal regions compared to UV-B samples without ISRIB, with both substantially lower than those for untreated samples (Figure 3.10B). Similarly, protein synthesis levels only increased by approximately 5% in ISRIB treated UV-B samples compared to UV-B only samples in the <sup>35</sup>S-methionine assays (Figure 3.10C).

The observation that ISRIB was only able to almost completely restore protein synthesis at the 30 minute time point suggests that the early translational reprogramming was mediated by the phosphorylation of eIF2 $\alpha$ , whereas by 4 hours, other factors were involved. Though it is important to note that higher concentrations of ISRIB were not tested, and therefore it is unknown whether a higher concentration of ISRIB would result in improved recovery of protein synthesis by 4 hours.



**Figure 3.10 | Early translational regulation following UV-B was mediated by eIF2 $\alpha$**

Identical plates of MCF10A cells were treated using UV-B or using UV-B in combination with 200nM ISRIB. Samples were then treated with 100 $\mu$ g/ml cycloheximide for 5 minutes prior to harvest at **A**) 0.5 hours and **B**) 4 hour time points (as described in Section 2.4.2.2). **C**) Protein synthesis rates were then determined by scintillation counting of incorporated <sup>35</sup>S methionine at the time points shown (for method see Section 2.3.9). Measurements shown are all represented as relative change compared to untreated samples at 0 hour, summarised across 3 biological repeats.

### ***3.11 Discussion***

The data shown in this chapter represents the development of two model systems using sub-lethal doses of IR and UV for investigating the DNA damage responses in the non-transformed MCF10A cell line. The purpose of this investigation was to both optimise each DNA damage model, and also to test their feasibility for use in ribosome profiling. To achieve this, low, but clinically relevant IR doses of 2 Gy and 4 Gy, and UV-B dose of 200 Jm<sup>-2</sup> were used.

Exposure to 4 Gy of IR on MCF10A cells was shown to have a subtle effect on the regulation of protein synthesis, with an initial increase at 3 hours post treatment, followed by a dramatic reduction in global protein synthesis rates to approximately 65% that of mock treated cells by 24 hours. This was consistent with the observed increase in the hypophosphorylated ( $\alpha$ ) form of 4E-BP1 by 24 hours. Exposure to 2 Gy of IR exhibited similar responses as measured by western blot analysis, albeit to a slightly lesser degree, though due to the smaller changes in observed protein synthesis rates. The 2 Gy of IR failed to produce consistent results from <sup>35</sup>S methionine incorporation, and so no general trends could be detected when combined across the repeats. Indeed, most previous studies investigating the regulation of gene expression following IR have used significantly higher (and not clinically relevant) doses (Braunstein et al., 2009; Minafra et al., 2015; Trivigno et al., 2013).

Similarly, exposure to all UV-B doses (with the exception of the lowest dose tested of 50 Jm<sup>-2</sup> at 4 hours) resulted in sustained reduction in global protein synthesis until at least 24 hours following treatment. While earlier time points of 4 and 8 hours showed a dose dependent response in the reduction of global protein synthesis, by 24 hours this was no longer the case, indicating the importance of other potential factors such as cell cycle on the DNA damage response (Figure 3.7). At the chosen dose of 200 Jm<sup>-2</sup> UV-B, protein synthesis rates decreased to 73%, 60% and 48% by 1, 2 and 4 hours post treatment respectively when compared to the untreated control at the 1 hour time point (Figure 3.8). Western blotting at these time points showed increase in eIF2 $\alpha$  phosphorylation was time dependent, while 4E-BP1 phosphorylation did not change, suggesting that p-eIF2 $\alpha$  was the main cause of the inhibition of the protein synthesis. Phospho-Chk1, p-H2AX levels as well as an increase in total p53 level in the UV-B samples, all indicating the presence of DNA damage. While PARP cleavage was detected in UV-B samples by 7 hours, it was not present at the earlier 4-hour time point (Figure 3.8).

Polysome profiling analysis on the IR and UV-B treated samples revealed major differences in the responses observed. For 4 Gy IR treated samples, polysome profiling conducted at 1, 3 and

24 hours (time points showing maximal changes in global protein synthesis) displayed only minor differences between the treated and untreated conditions (Figure 3.5). While the changes observed were consistent with the <sup>35</sup>S methionine incorporation data, they were nevertheless very small (Figure 3.3). Interestingly, although global protein synthesis was reduced to approximately 65% by 24 hours in the IR treated condition, this reduction was not reflected in the polysome profiling trace, with the IR sample showing no reduction in the levels of polysomal fractions, and only minor increases in the 60s peak when compared to the untreated sample. In comparison, UV-B samples revealed large changes in their subpolysomal/polysomal distributions between the treated and untreated conditions, with both 1 and 4 hour time points showing significantly larger changes compared to IR (Figure 3.9). These changes observed in the polysome profiling traces at both time points were consistent with the <sup>35</sup>S methionine incorporation results, with the 4 hour trace showing a larger reduction in the number of polysomally associated mRNAs. This was expected given the larger decrease in global protein synthesis to 48%, compared to 73% at 1 hour when compared to untreated control at 1 hour.

Investigations using the drug ISRIB, a potent inhibitor of the integrated stress response, also revealed that the significant translational reprogramming occurring following UV-B was mediated by the phosphorylation of eIF2 $\alpha$ , though only at the early time point of 30 minutes (Figure 3.10A). By 4 hours, the addition of ISRIB to the UV-B samples elicited only minor changes to the polysome profiling traces, and was only able to increase protein synthesis by approximately 5%, in contrast to the near complete restoration of protein synthesis at the 30 minute time point (Figure 3.10B, C). The results obtained suggest the observed translational reprogramming responses following UV-B by 4 hours was mediated in a p-eIF2 $\alpha$  independent manner. Although as higher concentrations of ISRIB were not tested, it is not known whether increasing the dose of ISRIB would have also counteracted the translational responses to UV-B at this later time point.

The overall aim of this project was to characterise the translational responses to DNA damage via ribosome profiling. As such, having a model system, such as using IR as the DNA damage source, which only induced minor changes in the polysome profiles of the cells and conditions in question is not ideal. It is still entirely plausible that significant changes in gene expression were occurring in response to IR on an individual gene level, and therefore detectable by the ribosome profiling technique. However, the lack of large changes in polysome formation in response to IR damage is potentially problematic as previous ribosome profiling investigations have shown that the technique itself results in the generation of large inter-repeat variances

(Jackson, 2013), meaning that small changes between conditions might not be observed. For these reasons, the use of UV-B, with its significantly greater responses as measured by polysome profiling, was deemed more suitable for further investigation using ribosome profiling.

## ***4. Investigation of gene expression changes using ribosome profiling***

### ***4.1 Introduction***

Previous methods of investigating gene expression changes, such as using polysome profiling or microarrays, have a number of limitations. Whilst the polysome profiling method used for identifying translational changes between the treatment conditions has enabled the characterisation of responses to UV-B in the MCF10A cells, further investigation using this method presents a number of limitations. First, polysome profiling is only able to characterise the translational responses on a global level, and whilst it is possible to process and extract RNA from the fractionated samples and probe for individual genes of interest, that too has certain limitations.

Typically, previous procedures employed in the Willis lab and elsewhere for investigating gene expression relied on the use of microarray analysis using RNA samples extracted from the pooled subpolysomal and polysomal fractions of the sucrose density gradient (Powley et al., 2009). cDNAs generated by reverse transcription using different fluorescent dyes for the subpolysomal and polysomal fractions were then hybridised to microarrays containing known probes for a range of mRNAs. The resulting ratio of the detected intensities of the different dyes corresponding to the subpolysomal and the polysomal fractions can be used as a measure of the translational status and efficiency. This approach however is limited in a number of ways. Perhaps most significantly, due to the necessity of comparing the pooled fractions from the subpolysomal and polysomal regions, the resulting outcome is only able to identify very large translational changes. Shifts within the polysomal region, for example a change from an mRNA with many associated ribosomes to only a few following treatment, will remain undetectable using this method, despite the potentially considerable difference in translational efficiency. Another major limitation of this approach is the number of probes that are present on the microarrays. Due to technical limitations, only a subset of probes are included, meaning that inevitably certain genes would not be detectable, despite any potential changes in its translational status.

To mitigate the limitations and problems listed above, it was decided that the relatively recently developed technique of ribosome profiling would be used for this project. This technique enables the identification of differentially translated mRNAs on a genome-wide scale (Brar & Weissman, 2015; Ingolia, Brar, Rouskin, McGeachy, & Weissman, 2012; Ingolia et al., 2009). The

ribosome profiling approach involves the simultaneous deep sequencing of ribosome protected mRNA fragments (the translome), as well as the extracted total RNA (the transcriptome). Together, these provide data on both the abundance of an mRNA, as well as the number of associated ribosomes on the same mRNA that can be used to estimate the translational status of this mRNA. The ability to map the precise positions of ribosome protected fragments (RPF) enables additional information to be potentially gained, such as the ribosomal density across specific regions of a transcript, the presence of translation initiation stalls and detection of possible upstream open reading frames.

This chapter will focus on the use of this technique to investigate the UV-B induced responses in MCF10A cells using the treatment conditions optimised in the previous chapter. Both transcriptional and translational analyses will be conducted on the resulting ribosome profiling sequencing datasets. Using these techniques, I aim to further investigate the selective up-regulation of specific subsets of genes, as well as the identification of the specific pathways in response to UV-B.

## ***4.2 Identifying translational changes using ribosome profiling***

The process involved in ribosome profiling is shown schematically in Figure 4.1, and described in detail in Section 2.4.3. Briefly, preparation for ribosome profiling involved the collection of post nuclear lysate, from which an aliquot was used to extract total RNA, and another aliquot was used for isolating RPFs. The RPFs were generated by digestion with RNase I, a ribonuclease with minimal sequence specificity, making it ideal for this sequencing technique. The digestion results in fragments of RNA of approximately 30 nucleotides in length which were protected from digestion by the enclosing ribosome structure (Wolin & Walter, 1988). For this chapter, these RPFs were isolated using illustra MicroSpin S-400HR size exclusion columns (GE Life Sciences). The resulting RNA samples were then treated using a Ribo-Zero Gold rRNA depletion kit (illumina), following which they were size selected through gel purification using markers of 28 and 34 nucleotides. RNA was then extracted from the excised gel slices, reverse transcribed and used to produce cDNA libraries, which were sequenced on an illumina NextSeq 500 platform.

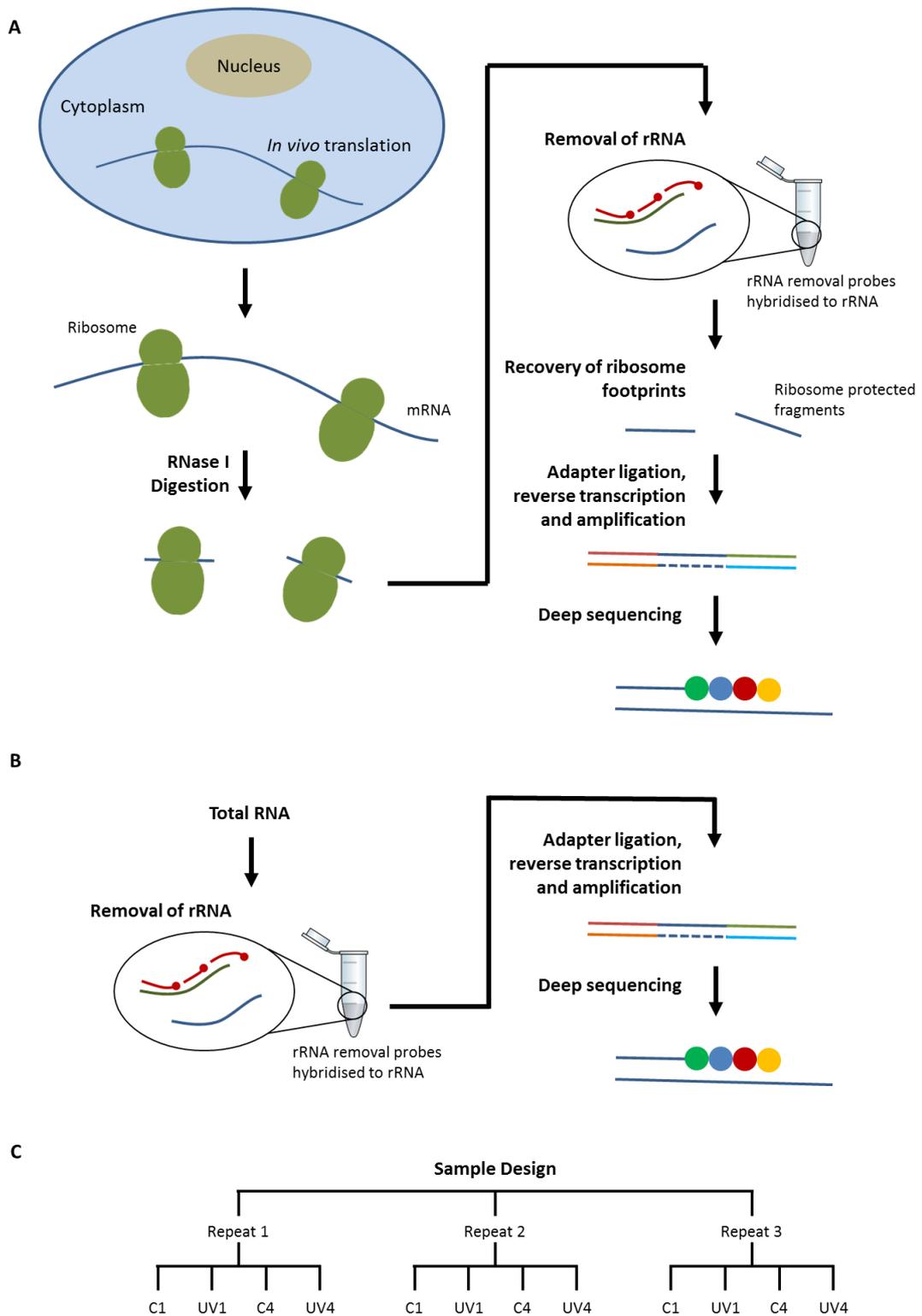
### ***4.2.1 Optimisation of ribosome profiling protocol steps***

A crucial step in this technique is the digestion of mRNA strands using RNase I, a non-specific ribonuclease with high activity (Spahr & Hollingworth, 1961), resulting in the formation of the RPFs. Insufficient digestion would lead to suboptimal yields of monosomes with protected RNA fragments as well as an excess of mRNA fragments with multiple associated ribosomes (eg.

disomes, trisomes etc). These disomes, trisomes and heavier polysomes would result in exclusion of those particular RPFs from sequencing during the subsequent fractionation and size selection steps. Similarly, over digestion would result in excessive degradation causing digestion of the ribosomes themselves and exposing the protected mRNA fragment for degradation. This again would result in the loss of protected fragments and thereby would also lead to loss of sequencing reads from those fragments, highlighting the need for optimisation of the digestion step. To this end, untreated MCF10A cells were harvested, lysed in 1x RPF lysis buffer as described in Section 2.4.3.2, and incubated with varying concentrations of RNase I at room temperature for 30 minutes with gentle rotation. Digested lysates were then separated on a sucrose gradient by ultracentrifugation.

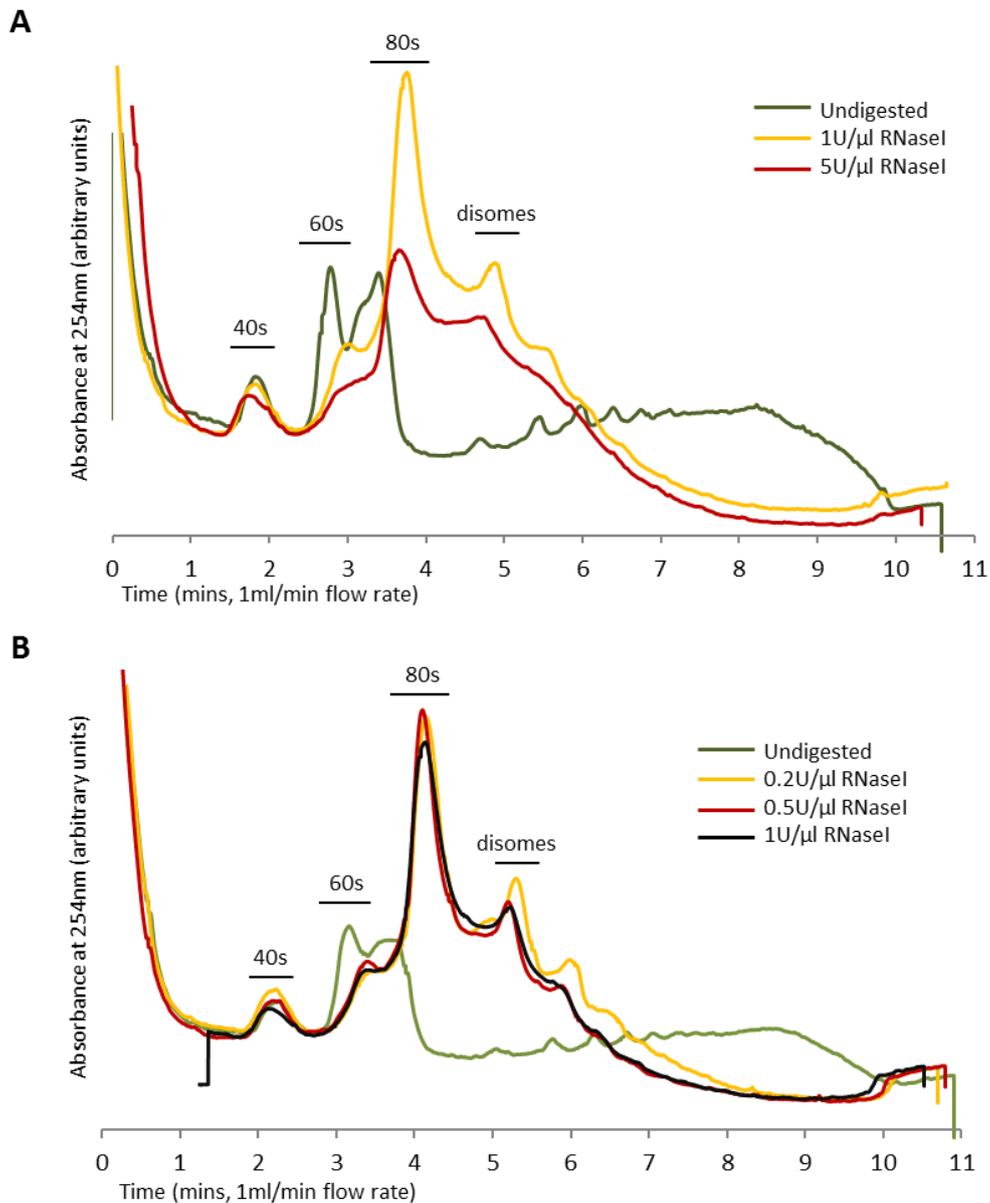
Following incubation with RNase I, the majority of the samples were digested to single 80s monosomes as shown by the loss of polysomes. Using RNase I at a concentration of 5U/ $\mu$ l showed a significant reduction in the yield of monosomes with protected RNA fragments (Figure 4.2A), indicating that this concentration leads to over digestion of the ribosomes and may be too high. The lower concentration of RNase I (1U/ $\mu$ l) however, showed a significantly higher 80s peak and was therefore more suitable for the ribosome profiling work. At this low concentration however, there was also a greater disome peak (present) compared to that of the 5U/ $\mu$ l RNase I digested samples.

To further optimise the RNase I digestion step, lower concentrations of 0.2 and 0.5U/ $\mu$ l were tested. At lower concentrations, the effects of RNase I digestion were somewhat more subtle. The samples treated with 0.5U/ $\mu$ l RNase I showed the best compromise, being a high enough concentration to digest the polysomes into single monosomes, while not excessively degrading the samples and reducing the yield (Figure 4.2B). It is important to note that a complete digestion of all polysomes into single 80s monosomes is in practice unachievable due to the large variances in the possible numbers of associated polysomes present within cells, and therefore disomes and trisomes will always be present to an extent, regardless of the concentrations of RNase I used.



**Figure 4.1 | Schematic diagram of the ribosome profiling protocol**

Overview of the ribosome profiling protocol, with the major steps involved for RPF samples shown in **A**), and those for total RNA samples shown in **B**). More detailed descriptions of the individual steps involved are described in Section 0. **C**) Overview of the samples taken for sequencing for both RPF and total RNA samples, with C1 and C4 referring to untreated controls at 1 and 4 hours respectively and UV1, UV4 referring to UV-B irradiated samples at 1 and 4 hours.



**Figure 4.2 | Optimisation of RNase I digestion**

Identical plates of MCF10A cells, lysed in RPF lysis buffer, treated using RNase I digestion at stated concentrations using the protocol in Section 2.4.3.2, and run on RPF sucrose gradients. **A)** RNase I over-digestion dramatically reduces yield of both monosomal 80s peak as well as disome peak. **B)** Optimisation of RNase I concentration, 0.5U/ $\mu$ l shows best compromise for obtaining greatest 80s monosome peak whilst achieving lowest disome peak. Data shown were from a single repeat, but were representative of a total of 3 biological repeats. Slight variances in the alignment of traces are due to pipetting error when preparing the sucrose gradient.

### ***4.3 RNA sequencing using illumina technology***

Following the preparation of both the RPF and total RNA samples, as detailed in Section 2.4.3.2, samples were processed as shown schematically in Figure 4.1. Briefly, following extraction of the RPFs and the total RNA samples, both were subjected to rRNA depletion using the Ribo-Zero Magnetic Gold kit (illumina), after which adapter sequences were ligated to the RNA. Reverse transcription followed by PCR was then used to generate cDNA constructs, following which sequencing was conducted on an illumina NextSeq 500 platform with a 75 bp single-end read cycle on the machine's high-output mode, using illumina's proprietary sequencing by synthesis (SBS) technology (described briefly below). The aim of this section is to outline the sequencing procedure as a whole, and to highlight the possible points during the protocol where sources of variance could be introduced.

#### ***4.3.1 Library preparation***

For the purposes of this project, two library preparation kits from illumina were required. For the RPF RNA samples, the illumina TruSeq Small RNA kit was used, and for the total RNA samples, the illumina TruSeq stranded mRNA LT kit was used. While the exact procedure for each kit varies slightly, the fundamental principle is the same. 5' and 3' adapters, containing unique barcoding to enable multiplexing, were ligated onto the RNA fragments. Reverse transcription followed by PCR was then conducted to create cDNA constructs. These constructs were then gel purified and used for sequencing. Depending on the read depth requirements, several samples can be multiplexed together and sequenced in one run, though care must be taken to only multiplex specific combinations of uniquely barcoded samples to prevent codon bias during sequencing. For this project, all 12 of the RPF samples were multiplexed and sequenced on one lane, whereas the 12 total RNA samples were multiplexed and sequenced across two lanes. Overview of the sample layout is shown in Figure 4.1C.

#### ***4.3.2 Cluster generation***

Following library preparation, the multiplexed or individual libraries were immobilised randomly onto the flow cell, which contains a pre-distributed lawn of surface bound oligos that are complementary to the adapters used for the library prep. Unlabelled nucleotides and enzymes were added to initiate the solid-phase bridge amplification, resulting in the formation of double-stranded bridges on the solid-phase substrate. These are then denatured, leaving single-stranded templates that are anchored to the substrate. This solid-phase amplification process

creates up to 1000 identical copies of each of the starting template constructs, resulting in the formation of densely packed clusters, each composed of identical strands of DNA.

### **4.3.3 Sequencing**

Illumina's SBS technology utilises four reversible terminator-bound fluorescently labelled nucleotides, which are used to sequence the clusters present on the surface of the flow cell. At each sequencing cycle, a single terminator-bound fluorescently labelled deoxynucleoside triphosphate (dNTP) is incorporated onto the nucleic acid chain anchored on the flow cell. These nucleotide labels also serve to inhibit polymerisation, thus preventing further incorporation after the initial, single nucleotide. The fluorescent dyes are then used to image the flow cell, and to identify the newly incorporated base, after which the labels are enzymatically cleaved to enable the process to be repeated, and for the next labelled nucleotide to be incorporated. As all four labelled dNTPs are presented as single, individual molecules, incorporation biases are minimised due to natural competition. This procedure is repeated until the read cycle is complete.

### **4.3.4 Alignment and data analysis**

The resulting raw sequencing reads generated from the NextSeq 500 were then processed using a combination of custom bash, R and Perl scripts, before being analysed using a number of R packages. Each of these processes is described in more detail in Section 2.4.3.8, along with the custom scripts used in the appendix in Section 7.2. Briefly however, the raw sequencing reads were first concatenated to form single files per sample, after which reads corresponding to adapter sequences were trimmed. Contaminating rRNA and tRNA reads were removed in a two-step process using the reference sequences from the illumina iGenome *homo sapiens* hg38 UCSC dataset and from the UCSC tRNA hg38 database (<http://gtrnadb.ucsc.edu/>) respectively. The processed reads were then aligned to the UCSC hg38 reference genome and their genomic features counted using HTSeq-count (Anders et al., 2015).

The level of contaminating reads present in the obtained data directly impacts the effective read depth of each sample. The impact of this, as well as possible sources of contamination will be discussed in a later section.

## ***4.4 Analysis of ribosome profiling data***

### ***4.4.1 Pre-processing of sequencing data***

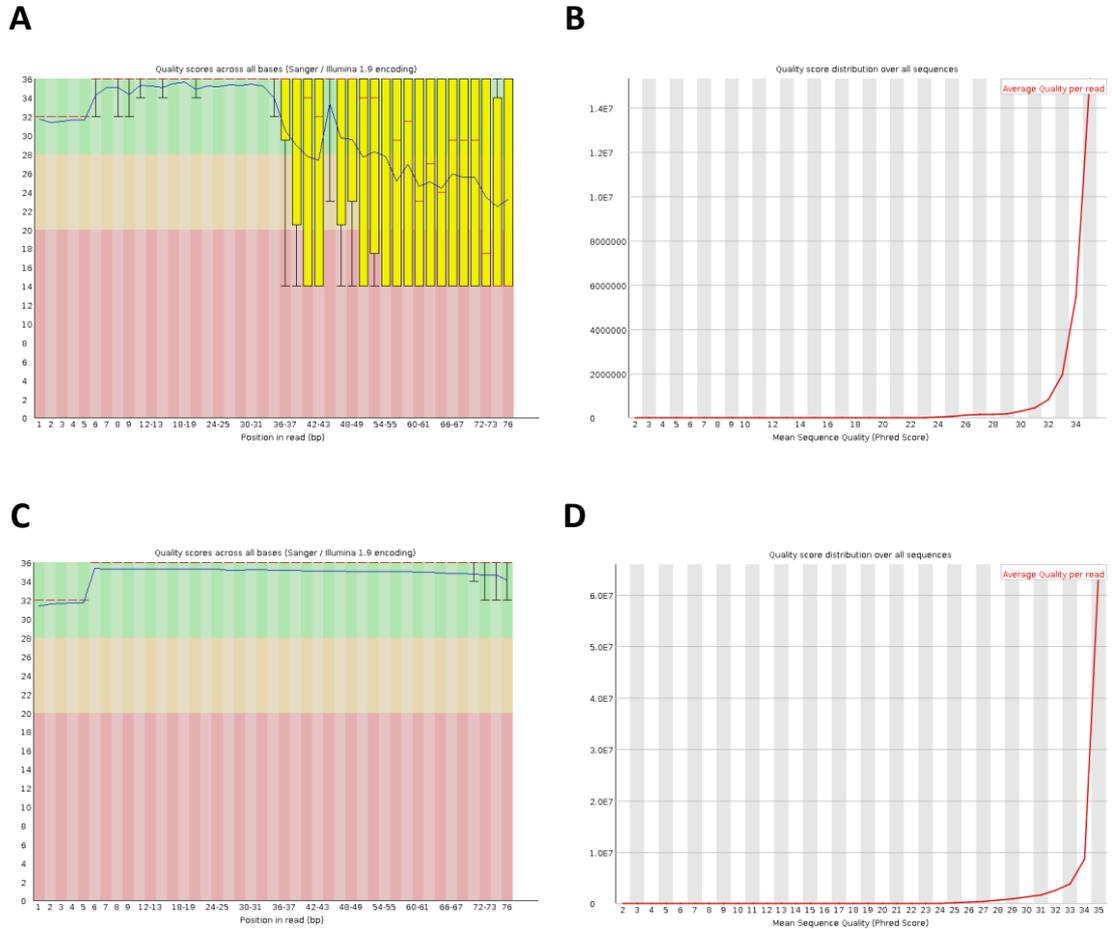
Sequencing output from ribosome profiling was processed as described in the previous section, and in detail in Section 2.4.3.8. Briefly, raw sequencing reads were processed using custom scripts, following which adapter sequences were trimmed using the program cutadapt (Martin, 2011), with the parameters listed previously in Section 2.4.3.8. The resulting quality of the trimmed reads was then verified using the program FastQC (Andrews, 2016). This measure of quality uses the logarithmic Phred score (measured as Q values), which represents the probability of an incorrect call, and thus is used to provide a quantitative measure of the accuracy of the sequencing, with a cut-off value of Q20 (99% accurate) being the accepted norm for most applications. The FastQC-generated output for the base sequence quality of one of the RPF samples and the corresponding whole transcriptome sample is shown in Figure 4.3. As mentioned in the methods (Section 2.4.3.8), because the RPFs are only approximately 30 nucleotides in length, a large proportion of the sequencing reads were composed of the barcode and adapter sequences. Reads trimming results in significantly shorter reads, containing only RPF sequences. The extremely low Phred scores observed from approximately 35 nucleotides onwards are caused by these now discarded and so non-existent reads. Whilst all samples varied slightly in terms of the read quality achieved, almost all samples possessed a Phred score of above Q30 (corresponding to 99.9%+ certainty of an accurate base call).

### ***4.4.2 Overview of mapping statistics***

The resulting numbers of reads for each individual sample following each step of the pre-processing stages were tabulated and summarised in Figure 4.4. Reads averaged across all three biological repeats for each condition, as well as globally, are shown in Figure 4.5. These raw read counts, whilst lower than those obtained from previous ribosome profiling experiments conducted in the Willis lab, resulted in approximately the same numbers of uniquely mapped reads for both the RPF and whole transcriptome datasets (Jackson, 2013). This was due to much more efficient rRNA removal from using the Ribo-Zero Gold (illumina) kit compared to methods used previously.

Comparing the two sequencing datasets, multiplexing all 12 of the RPF samples onto the same sequencing lane has resulted in an average read count of around 19 million reads per sample, compared to that of just under 63 million reads per sample for the total RNA (or whole transcriptome) data, which were multiplexed across two sequencing lanes (Figure 4.5).

Variation of the total read counts within each dataset also exists, with the RPF dataset varying from a low of around 11.5 million to a high of around 26.6 million reads per sample. The whole transcriptome dataset was, on the whole, reasonably uniform in terms of total read counts when comparing across conditions (Figure 4.5), but did however contain individual samples with a low of around 31.5 million and a high of around 111 million reads (Figure 4.4).



**Figure 4.3 | Sequence quality post adapter trimming**

Sequence quality as measured using the logarithmic Phred score (a measure of the probability of an incorrect base call, with values of Q20 corresponding to 99% accuracy, and Q30 corresponding to 99.9% accuracy) using the program FastQC. **A)** Per base quality and **B)** distribution of the sequence quality of the RPF dataset following adapter trimming. **C)** Per base quality and **D)** distribution of the sequence quality of the whole transcriptome dataset following adapter trimming.

RPF												
	Repeat 1				Repeat 2				Repeat 3			
	C1	C4	UV1	UV4	C1	C4	UV1	UV4	C1	C4	UV1	UV4
Total reads	25,182,365	12,747,827	19,929,994	25,050,023	15,037,139	26,586,927	15,930,515	18,647,153	19,553,288	14,164,850	22,554,768	11,558,086
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Reads remaining following rRNA removal	24,079,215	11,842,003	19,012,726	22,794,897	13,164,077	25,264,617	14,743,537	18,005,615	18,243,871	10,119,895	21,456,299	10,467,089
	95.6%	92.9%	95.4%	91.0%	87.5%	95.0%	92.5%	96.6%	93.3%	71.4%	95.1%	90.6%
Reads remaining following rRNA & tRNA removal	6,939,386	3,164,525	5,035,078	4,901,677	5,131,881	5,310,366	2,979,089	3,728,624	3,036,678	2,344,624	4,458,429	1,544,912
	27.6%	24.8%	25.3%	19.6%	34.1%	20.0%	18.7%	20.0%	15.5%	16.6%	19.8%	13.4%
Reads uniquely mapped	5,283,969	2,361,022	3,202,839	3,224,232	3,255,513	3,406,769	1,934,175	2,258,863	2,052,483	1,551,822	3,014,866	951,215
	21.0%	18.5%	16.1%	12.9%	21.6%	12.8%	12.1%	12.1%	10.5%	11.0%	13.4%	8.2%
Whole transcriptome												
	Repeat 1				Repeat 2				Repeat 3			
	C1	C4	UV1	UV4	C1	C4	UV1	UV4	C1	C4	UV1	UV4
Total reads	86,867,985	31,533,001	61,891,336	50,429,937	42,744,019	56,730,254	111,180,546	56,874,598	57,109,139	71,594,776	62,120,042	66,690,239
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Reads remaining following rRNA & tRNA removal	76,990,384	12,215,473	40,779,633	42,715,854	27,120,010	47,195,265	28,236,500	51,978,486	42,498,423	47,840,500	40,192,699	37,423,422
	88.6%	38.7%	65.9%	84.7%	63.4%	83.2%	25.4%	91.4%	74.4%	66.8%	64.7%	56.1%
Reads uniquely mapped	74,859,764	11,692,697	39,521,869	41,433,115	26,285,703	45,974,212	26,923,893	50,663,366	41,314,008	46,522,613	38,970,040	36,290,728
	86.2%	37.1%	63.9%	82.2%	61.5%	81.0%	24.2%	89.1%	72.3%	65.0%	62.7%	54.4%

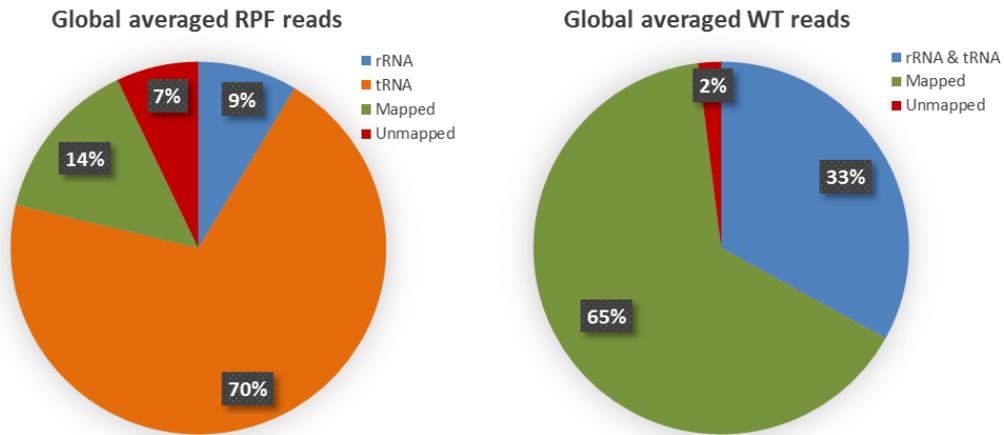
**Figure 4.4 | Overall individual mapping statistics for sequencing runs**

Overview of the mapping statistics for all samples in the sequencing runs. Reads were filtered using reference sets for tRNA and rRNA sequences. Filtered reads were then mapped to the reference *homo sapiens* hg38 UCSC genome. The processing of these reads are described in more detail in Section 2.4.3.8. For each sample, the total numbers of reads remaining after each stage of processing are displayed, along with that figure as a percentage relative to the starting raw reads.

**A**

RPF										
	C1		C4		UV1		UV4		Global	
Total reads	19,924,264	100.0%	17,833,201	100.0%	19,471,759	100.0%	18,418,421	100.0%	18,911,911	100.0%
Reads remaining following rRNA removal	18,495,721	92.2%	15,742,172	86.5%	18,404,187	94.4%	17,089,200	92.7%	17,432,820	91.4%
Reads remaining following rRNA & tRNA removal	5,035,982	25.7%	3,606,505	20.5%	4,157,532	21.2%	3,391,738	17.6%	4,047,939	21.3%
Reads uniquely mapped	3,530,655	17.7%	2,439,871	14.1%	2,717,293	13.9%	2,144,770	11.1%	2,708,147	14.2%
Whole transcriptome										
	C1		C4		UV1		UV4		Global	
Total reads	62,240,381	100.0%	53,286,010	100.0%	78,397,308	100.0%	57,998,258	100.0%	62,980,489	100.0%
Reads remaining following rRNA & tRNA removal	48,869,606	75.5%	35,750,413	62.9%	36,402,944	52.0%	44,039,254	77.4%	41,265,554	67.0%
Reads uniquely mapped	47,486,492	73.3%	34,729,841	61.0%	35,138,601	50.3%	42,795,736	75.2%	40,037,667	65.0%

**B**



**Figure 4.5 | Averaged overview of mapping statistics for sequencing runs**

**A)** Overview of the mapping statistics for each condition, averaged across all three repeats, as well as globally. Reads are displayed identically to Figure 4.4, with the total number of reads remaining following rRNA and tRNA filtering, and unique mapping being displayed. Remaining reads at each stage are also displayed as a percentage relative to the starting raw reads. Detailed descriptions of each stage of the filtering and aligning processes are described in Section 2.4.3.8. **B)** Breakdown of the proportions of reads for the global averaged data corresponding to tRNAs, rRNAs, unmapped and mapped reads for both the RPF dataset and the whole transcriptome (WT) dataset.

#### ***4.4.3 Significant number of contaminating reads***

Prior to the library preparation, samples were rRNA depleted using the Ribo-Zero Gold Magnetic kit (illumina), as described in Section 2.4.3.4. Previous ribosome profiling attempts within the Willis lab using a comparatively much less efficient rRNA depletion method resulted in a significant proportion of total reads corresponding to rRNA sequences, with on average close to 50% of reads being filtered due to contamination (Jackson, 2013). The Ribo-Zero Gold based depletion process was conducted in an attempt to alleviate this problem, and thereby to also lower the required total read depth per sample, which enabled the multiplexing of all 12 RPF samples onto the one sequencing lane.

Whilst there are differences in both the variation and read depths for each of the samples in the two datasets, the greatest difference between them was the proportions of contaminating reads, corresponding to tRNA and rRNA sequences. For the RPF dataset, the filtering for contaminating tRNA and rRNA reads was carried out in a two stage process, enabling the calculation of the proportions of both species independently that were present in each sample. For the whole transcriptome dataset however, the removal of tRNA and rRNA reads was performed in a single step. Initial preliminary work on filtering the whole transcriptome dataset using the same two stage filtering process as that used for the RPFs showed a negligible amount of reads present that corresponded to tRNA sequences, as was expected, with on average less than 0.001% of the reads being filtered. This minute quantity of reads, coupled with the significantly higher read counts on average per sample, did not justify the massively increased computational time that would have been required to use the same two stage filtering process.

The effectiveness of the Ribo-Zero Gold based rRNA depletion process was shown by the low percentage of reads corresponding to rRNA sequences present in the RPF dataset. On average, only around 9% of the total reads corresponded to rRNA. The reduction in the amount of contaminating reads in the whole transcriptome dataset, however, was somewhat less pronounced, with roughly 33% of all reads mapping to rRNA and tRNA sequences on average. Despite the greater percentage of contaminating reads in the whole transcriptome dataset, as the initial starting read counts for all of the samples were very high, this resulted in approximately 40 million uniquely mapped reads per sample (Figure 4.5).

Whilst the rRNA depletion step during sample preparation seemed to have alleviated the high amount of rRNA contaminating reads previously observed, within the RPF dataset, a significant portion of the reads were found to correspond to tRNA sequences. On average, around 70% of

all reads within the RPF dataset were marked as tRNA sequences and removed during the bowtie2 filtering stage (Figure 4.5). This was significantly higher than what was expected. The source of the contamination of tRNA reads and its consequences will be discussed in detail in a later chapter, but for the purposes of conducting analysis on the remaining reads, this contamination effectively dramatically reduced the read depths of each RPF sample to a very small fraction of the starting count. The low read count presented its own problems when performing transcriptome analysis, which will be discussed in the later section.

Aside from contaminating reads, another major difference between the two datasets was the number of reads that uniquely map to the genome, with those in the whole transcriptome dataset having a significantly higher successful mapping percentage whilst for the RPF dataset, only roughly 2/3rds of the remaining filtered reads were able to be uniquely mapped (Figure 4.5). This difference was most likely due to the lengths of reads involved as longer reads in general are much easier to map to a unique location. The reads in the whole transcriptome dataset were formed of much longer 75 bp reads, as compared to those in the RPF dataset, which contained much shorter ribosome protected fragments of approximately 30 nucleotides. This fact, coupled with the significant amounts of rRNA and tRNA contaminating reads meant that on average, only about 2.7 million reads were able to be uniquely mapped per RPF sample. The amount of reads mapped did not appear to vary with treatment condition, though did appear to vary slightly between repeats, with the third replicate showing the lowest overall uniquely mapped read counts (Figure 4.4).

#### ***4.4.4 Transcriptome analysis of ribosome profiling data***

##### *4.4.4.1 Identifying differentially expressed genes using whole transcriptome dataset*

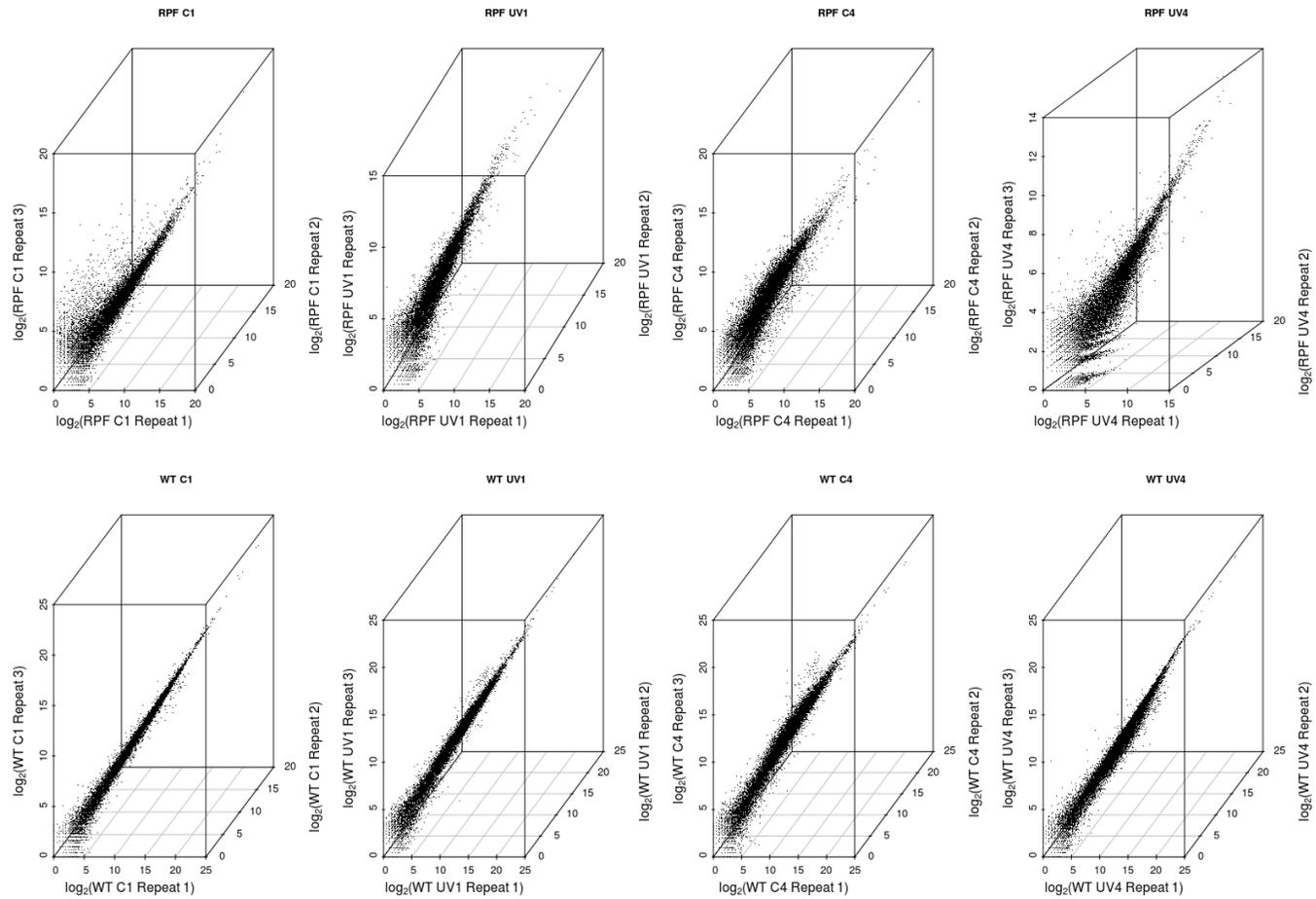
The whole transcriptome dataset, with its approximately 40 million uniquely mapped reads per sample (Figure 4.5), was used to perform transcriptome analysis to identify differentially expressed genes following UVB treatment. The analysis was performed with the R package, DESeq2 (Love et al., 2014), using read counts obtained from HTSeq-count (Anders et al., 2015). Detailed description of the method used is described in Section 2.4.3.8.5.

##### *4.4.4.2 Tight inter-replicate correlation in sequencing datasets*

As a crude method of visualising the variances present between the three replicates of sequencing runs, the log transformed values of the normalised read counts for each individual sample across both the RPF and the whole transcriptome datasets were plotted. For each treatment condition, the log normalised counts corresponding to each repeat were represented

on the x, y and z axes of a 3D scatter plot, with the resulting spread being an indicator of the level of variance present between the replicates.

Plots for both datasets displays an overall tight correlation between the replicates, though with the RPF dataset showing more of a disparity between each replicate compared to that of the whole transcriptome dataset (Figure 4.6). This was not overly unexpected, and was most likely due to the increased complexity, and therefore increased sources of potential variances, involved in the RPF sample and library preparation protocol as compared to that for the whole transcriptome samples.

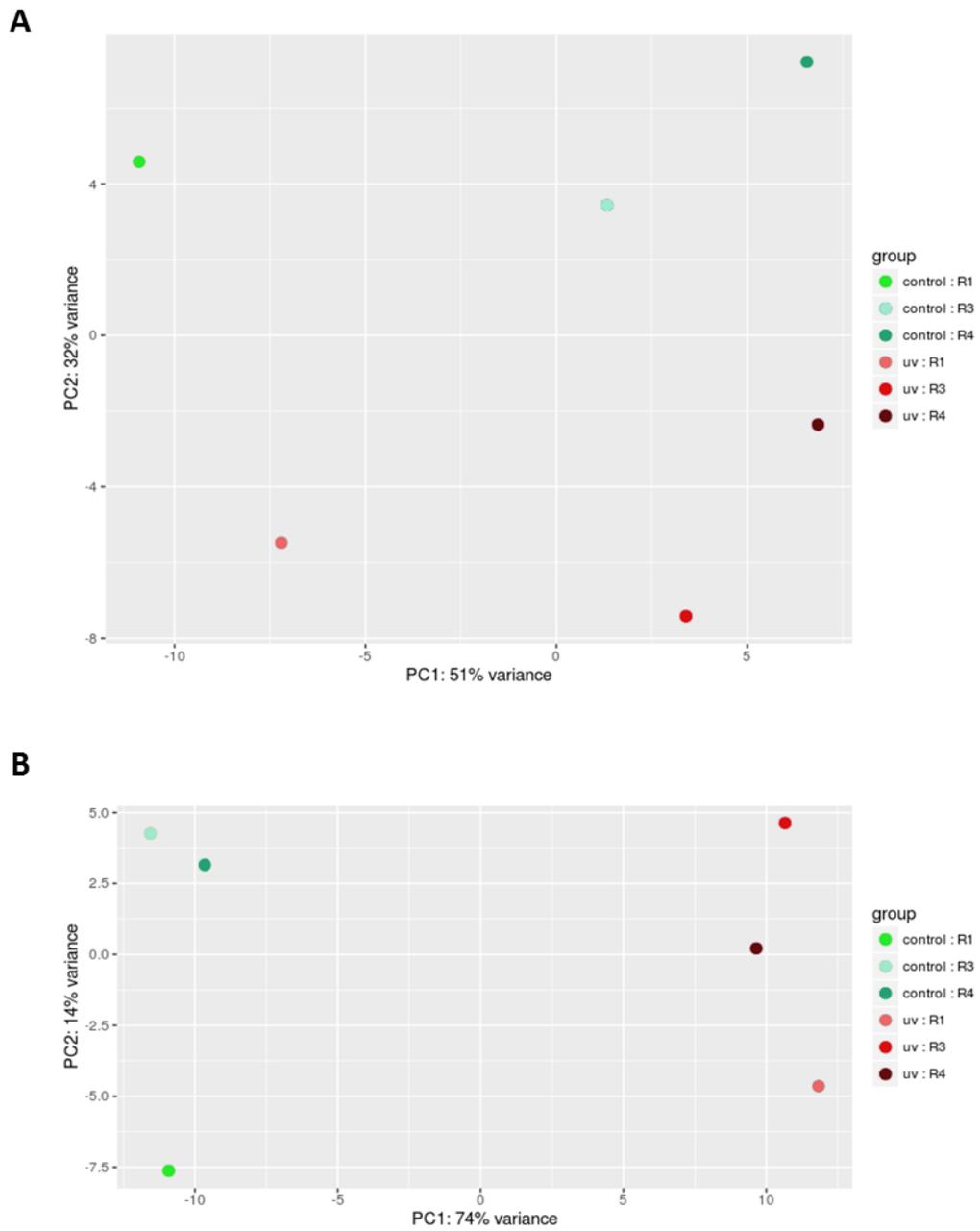


**Figure 4.6 | Strong correlation between all three biological repeats for both the RPF and whole transcriptome datasets**

Log values of the normalised read counts for the three biological replicates were plotted using 3D scatter plots, corresponding to the x,y and z axes. These plots were generated within R using the package “scatterplot3d”.

#### *4.4.4.3 Principal component analysis reveals tight clustering of samples based on treatment conditions*

Raw read counts were processed and normalised by DESeq2. The resulting dataset was then used to generate principal component analysis plots for all whole transcriptome samples at the 1 and 4 hour time points. At the later time point of 4 hours, the primary source of variance was caused by the treatment condition, as was expected (Figure 4.7). Two distinct clusters were present, one containing all biological replicates of the mock treated samples, and the other containing all UVB treated samples. The variance between the biological replicates at this time point was comparatively small. The PCA plot generated for the 1-hour time point however revealed that the primary source of variance in that case, was between the biological repeats.



**Figure 4.7 | Principle component analysis of whole transcriptome dataset at 1 and 4 hours**

PCA plots of the whole transcriptome dataset at **A)** 1 hour time point and **B)** 4 hour time point. Plots were generated using the R module DESeq2, more detailed description of its use is listed in Section 2.4.3.8.5, and scripts used for analysis and plotting are included in the appendix (Section 7.2).

#### 4.4.4.4 UV-B induced upregulation of specific pathways

MA plots for the data at both time points were also generated using the DESeq2 package, and are shown in Figure 4.8. As expected, the level of variance present in the 1 hour time point was relatively small, with the majority of genes unchanged following UVB treatment. The genes which were identified as significant are highlighted in red. For clarity, the 0 mark on the y-axis (log fold change) has also been highlighted in red. The MA plot generated for the 4 hour time point dataset showed both considerably larger variance in the log fold change of genes, as well as the numbers of genes identified to be significantly changed following treatment. The top 30 most significantly up or down-regulated genes for both time points are shown in Figure 4.9.

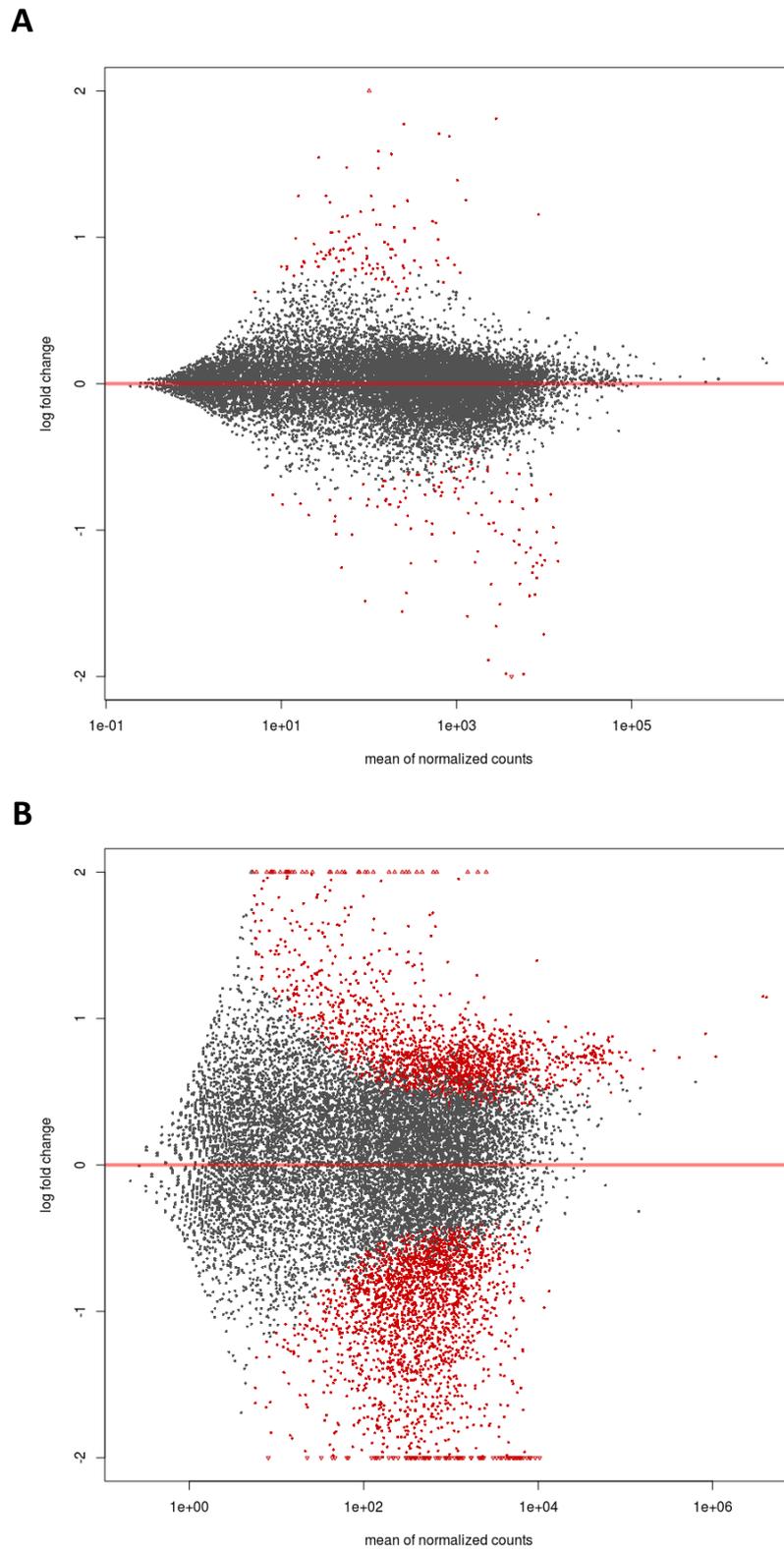
Transcriptome analysis using DESeq2 was performed on the whole transcriptome dataset, with the 1 and 4 hour samples being processed separately. A false detection rate cut-off threshold of 5% was set for the analysis. Details of the parameters and code used for this analysis is listed in the appendix, Section 7.2.

At the 1 hour time point, DESeq2 analysis resulted in the identification of 234 differentially expressed genes with the previously mentioned parameters. Of these, 117 genes were up-regulated following UVB treatment, and 117 genes were down-regulated. Unsurprisingly, given the significantly greater difference between the treatment conditions as shown by the PCA plots (Figure 4.7), at the 4 hour time point, substantially more genes were identified to be differentially expressed. A total of 4083 genes were detected by DESeq2, of which 1797 and 2286 were found to be, respectively, up- and down-regulated following UVB treatment.

The DESeq2-generated significant genes output lists for the 1 and 4 hour datasets were then processed using the Ingenuity Pathway Analysis program (IPA, <http://www.ingenuity.com>). This was carried out to enable the identification and analysis of the pathways present within the differentially expressed gene lists. Analysis using the IPA program for the earlier time point of 1 hour identified a handful of pathways, all possessing  $-\log(p\text{-values})$  of greater than 1.5 (equal to  $p\text{-values}$  smaller than roughly 0.03) (Figure 4.10). Encouragingly, pathways associated with DNA damage and repair such as p53 signalling and ATM signalling were both present in the analysis.

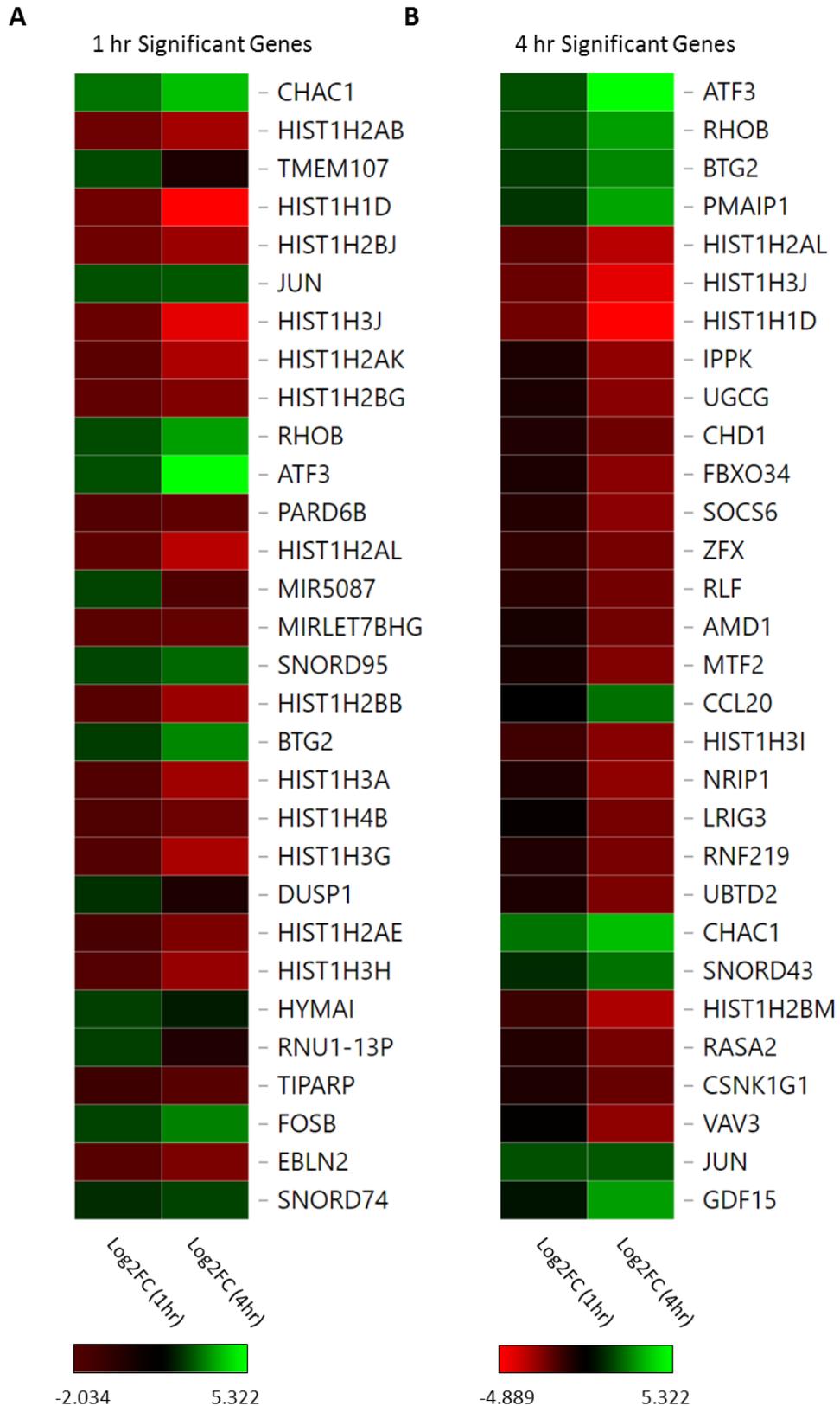
Unsurprisingly, IPA output for the much larger 4 hour DESeq2 significant gene list displayed considerably greater numbers of pathways that were identified (Figure 4.11). Of these, both p53 signalling and ATM signalling pathways which were present in the 1-hour analysis were also identified within the 4 hour dataset. Other major pathways identified included eIF2 signalling, and cell cycle pathways. The eIF2 pathway, and in particular the phosphorylation of eIF2 $\alpha$  is

known to be a key regulator in overall protein synthesis following a wide range of stresses by its interaction with eIF2B (as described in the Introduction). The identification of this pathway was consistent with the western blot data from the previous chapter showing the presence of phospho-eIF2 $\alpha$  from 1 hour to 7 hours following exposure. Indeed, the role of eIF2 $\alpha$  following UV-B was also demonstrated by the use of ISRIB, a drug which potently reverses the effects of eIF2 $\alpha$  phosphorylation, where the UV-B induced responses observed in polysome profiling traces were almost entirely negated (Figure 3.10). Several of the other pathways identified to be changing most significantly, such as oxidative phosphorylation (second only to eIF2 signalling in terms of p-value), have been previously shown to be persistently activated following UVB irradiation (Tsai et al., 2009). This observation was consistent with the 4 hour whole transcriptome dataset, in which 60% of the known genes associated with the oxidative phosphorylation pathway were identified to be significantly changed by DESeq2, of which all exhibited up-regulation following UV-B. Similarly, out of the approximately 50% of genes associated with mitochondrial dysfunction, the vast majority up-regulated following UV-B, which is consistent with previous studies (Torregrosa-Munumer, Goffart, Haikonen, & Pohjoismaki, 2015).



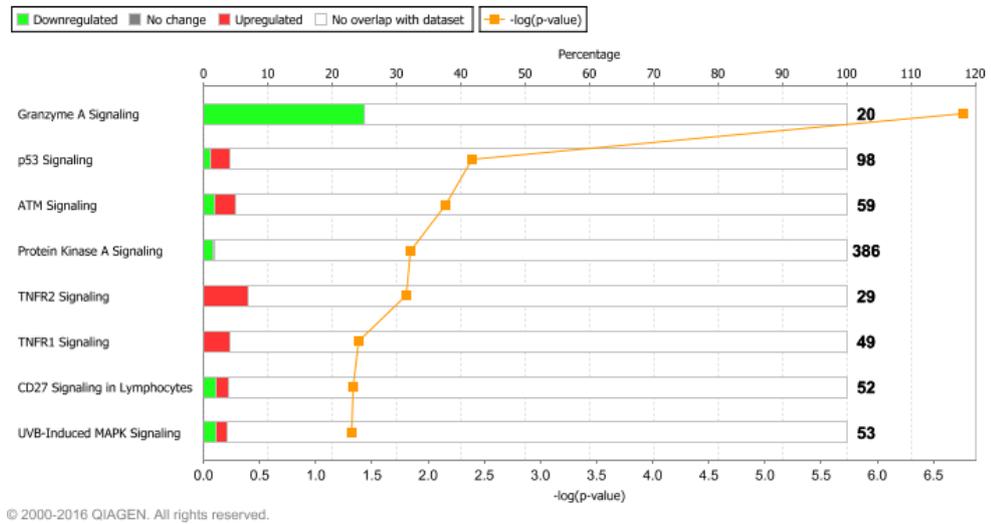
**Figure 4.8 | MA plot of transcriptome analysis outputs**

MA plots for the **A)** 1 hour and **B)** 4 hour transcriptome datasets, generated using R module DESeq2, more detailed description of its use is listed in Section 2.4.3.8.5, and scripts used for analysis and plotting are included in the appendix (Section 7.2).

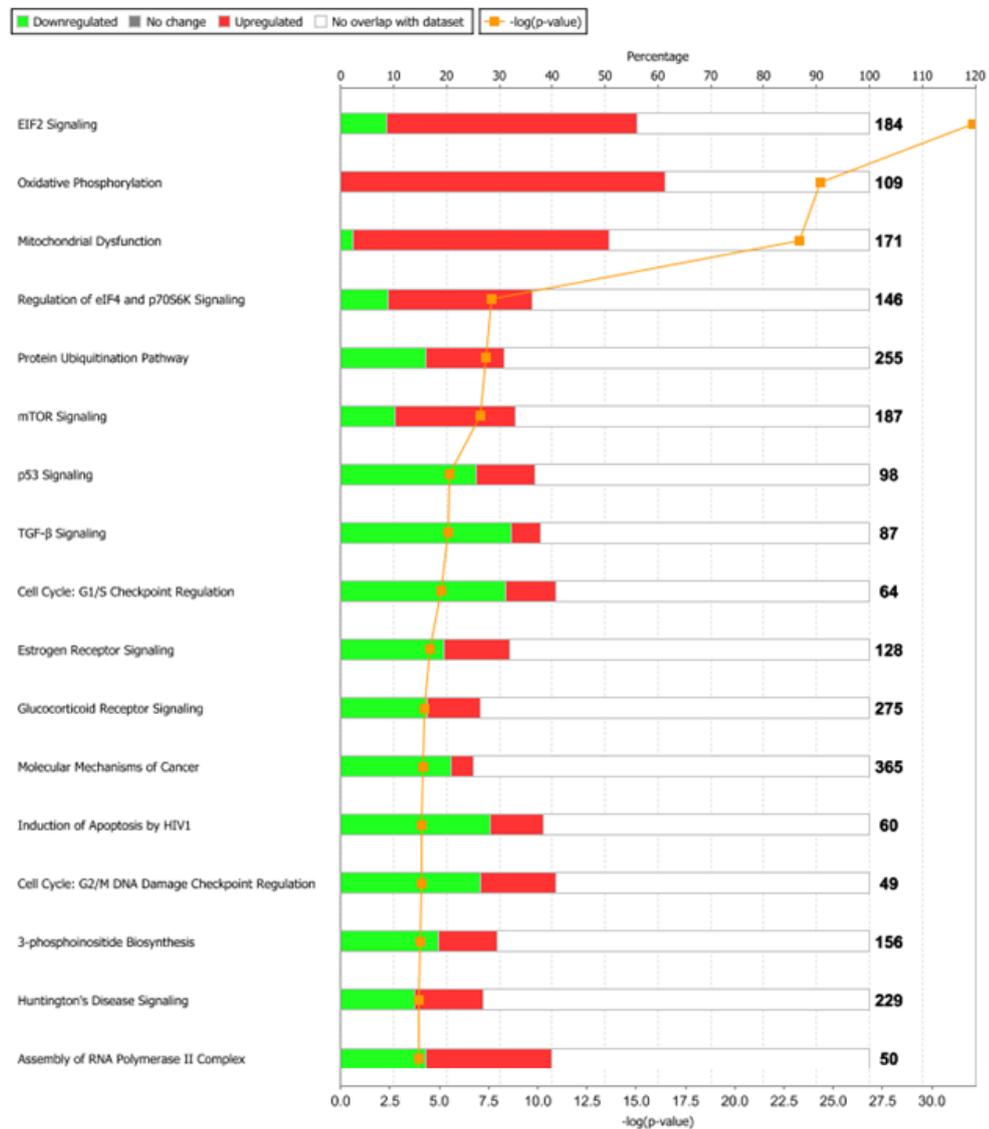


**Figure 4.9 | Differentially expressed genes following UV-B treatment**

The top 30 most significant differentially expressed genes following UV-B as identified by DESeq2, sorted by adjusted p-value at the **A)** 1 hour and **B)** 4 hour time points, with the corresponding log fold change values for those genes at both time points represented on a colour scale, with green representing an up-regulation and red representing down-regulation following UV-B, and the range of log fold change shown in the colour bar below the heatmaps. The cut off adjusted p-value corresponding to the top 30 genes was 3.5e-08 and 6.3e-25 for the 1 and 4 hour time points respectively.



**Figure 4.10 | UVB induced changes in pathways identified using IPA at 1 hour**  
 Pathways found to be regulated following UV-B in the 1 hour whole transcriptome dataset were identified by IPA, and ranked by their p-values.



© 2000-2016 QIAGEN. All rights reserved.

**Figure 4.11 | UVB induced changes in pathways identified using IPA at 4 hours (Part 1)**

Pathways found to be regulated following UV-B in the 4 hour whole transcriptome dataset were identified by IPA, and ranked by their p-values.



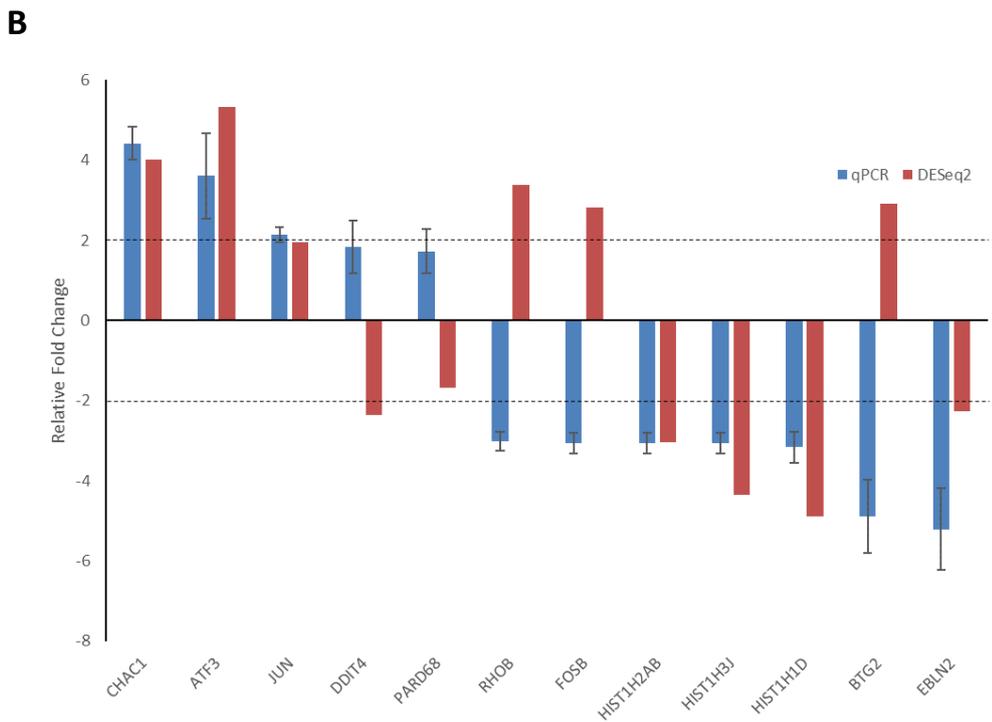
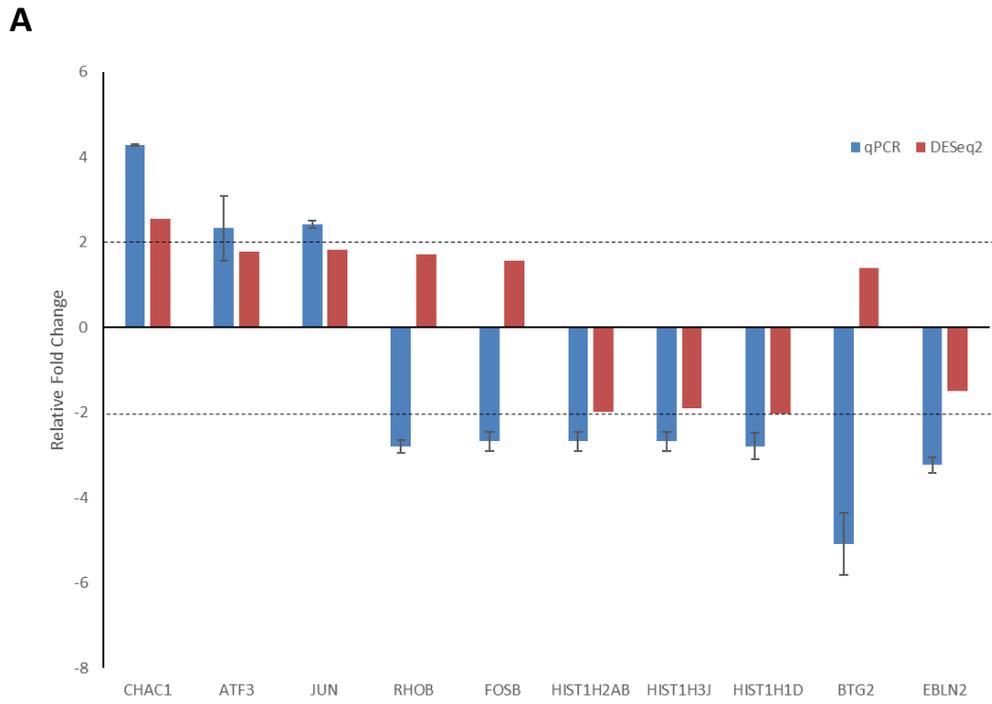
**Figure 4.11 | UVB induced changes in pathways identified using IPA at 4 hours (Part 2)**

Pathways found to be regulated following UV-B in the 4 hour whole transcriptome dataset were identified by IPA, and ranked by their p-values.

#### *4.4.4.5 qPCR validation of DESeq2 analysis outputs*

Having identified the top most differentially expressed genes following UV-B treatment at both time points using DESeq2, further validation of these results was conducted using quantitative PCR. RNA was extracted from samples treated per the conditions described in Section 3.6, following which cDNA was generated through reverse transcription reactions with random hexamers using SuperScript III kits (Life Technologies). Probes for a select number of the most differentially expressed genes were generated using Primer-BLAST (Ye et al., 2012), with the exact sequences for all primer pairs used described in detail in Section 2.4.6. qPCR analysis using SYBR Green was then conducted on MicroAmp® Fast 96-well reaction plates (Life Technologies), and read on a QuantStudio 6 Flex Real-Time PCR system, with the reaction mix used described in Section 2.4.6. The resulting fold-changes in expression between treated and untreated samples were then calculated using the  $2^{\Delta\Delta CT}$  method of normalisation, with the results for both 1 and 4 hour time points shown in Figure 4.12.

Whilst the majority of genes tested using qPCR were concordant with the outputs from the DESeq2 analysis, at both time points, a number of the probed genes did exhibit differing outcomes from the two methodologies (Figure 4.12). A possible cause of these differences in the observed results could be due to the large number of potential (and different) biases inherent in both techniques. For qPCR, these include the choice of using random hexamers for the initial cDNA generation, biases in nucleotide incorporation, and the effect of the chosen sequences of the primer pairs used. Similarly, due to the multitude of steps involved in illumina sequencing (detailed in Section 2.4.3.7), biases could potentially arise during the initial RNA fragmentation step, rRNA depletion, cDNA generation, library amplification, bridge amplification, as well as during the final sequencing by synthesis step itself. Combined, all these factors, in addition to the differing normalisation methods required by DESeq2 and qPCR data, could explain the observed non-concordant results for a few of the genes tested.



**Figure 4.12 | qPCR validation of DESeq2 analysis outputs**

Measurements shown represent the relative fold change of 200 Jm<sup>-2</sup> UV-B treated samples compared to untreated controls at **A**) 1 hour and **B**) 4 hours post treatment. qPCR measurements shown represent the average of 3 biological repeats, with error bars representing the standard error of the mean. Probes and method used are detailed in Section 2.4.6.

#### ***4.4.5 Translatome analysis***

##### *4.4.5.1 Mapping of sequencing reads reveals distribution of fragments*

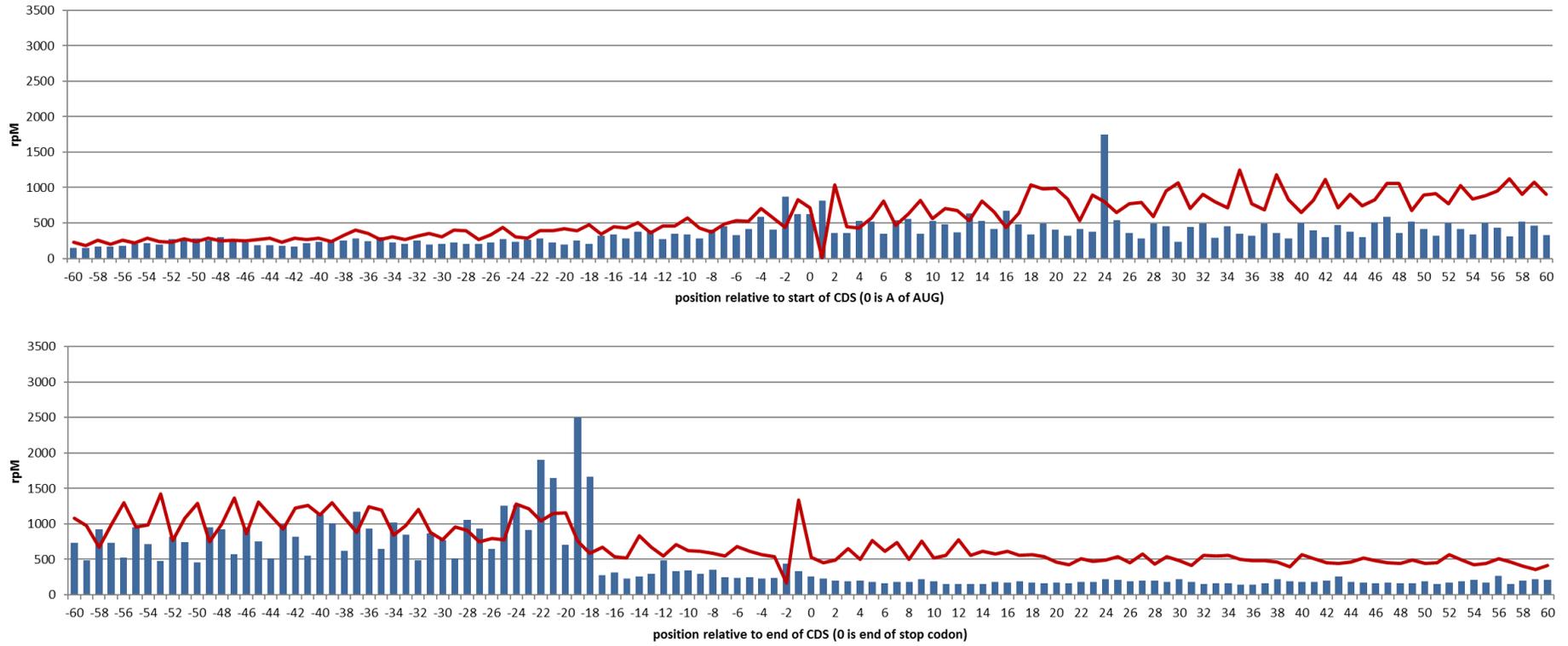
Despite the high proportions of contaminating reads and the resulting low average read counts for the RPF dataset, further analysis was undertaken to investigate the results of the sequencing runs. Periodicity and read density mappings for all samples were conducted by Dr Ruth V. Spriggs using custom scripts (see Section 2.4.3.8). The 5' end of each of the mapped reads were aligned relative to the start or end of the coding sequence (CDS) of the reference sequences they were mapped to. The read density count at each relative position across all the reference sequences was then calculated and normalised to the total number of aligned reads, forming the read density metric, reads per million (rpM).

The periodicity mapping was conducted for all individual samples and the global averaged read density counts are shown in Figure 4.13. Due to the size and structure of the actively translating ribosome, during translation initiation when the AUG start codon is in the P-site of the ribosome, the resulting ribosome protected fragment will be mapped at position -13 (when mapped relative to the position of the A nucleotide of the AUG). Similarly, for terminating ribosomes, when the stop codon is present in the A-site the resulting ribosome protected fragment will be mapped to position -18.

As was expected, the RPF read densities for regions within the CDS were both higher than regions outside of the CDS, and also displayed distinct 3-nucleotide periodicity. Somewhat surprisingly though, both of those factors were more noticeable towards the end of the CDS compared to that at the start (Figure 4.13). Another surprising, but possibly linked observation was the lack of the distinct peak in read density at the start of the CDS that is characteristic of ribosome profiling sequencing runs which utilise cycloheximide pre-treatment. The addition of cycloheximide prevents the translocation of the ribosome by blocking the E-site (Pestova, 2003; Schneider-Poetsch et al., 2010), though it does not prevent further translation initiation, causing the stacking of initiating ribosomes upstream of the initiation site. This in turn translates to a peak in the numbers of resulting ribosome protected fragments, which map to the initiation site, as well as the region immediately upstream due to the said stacking effect. However, no stacking effect was observed in either the global averaged data (Figure 4.13), nor in any individual treatment conditions (Figure 4.14). One possible cause of this could be a failure in capturing the RPFs corresponding to the mentioned stacked ribosomes, the deficiency in the capture of these fragments could be explained by insufficient RNase I digestion. The short distances between the

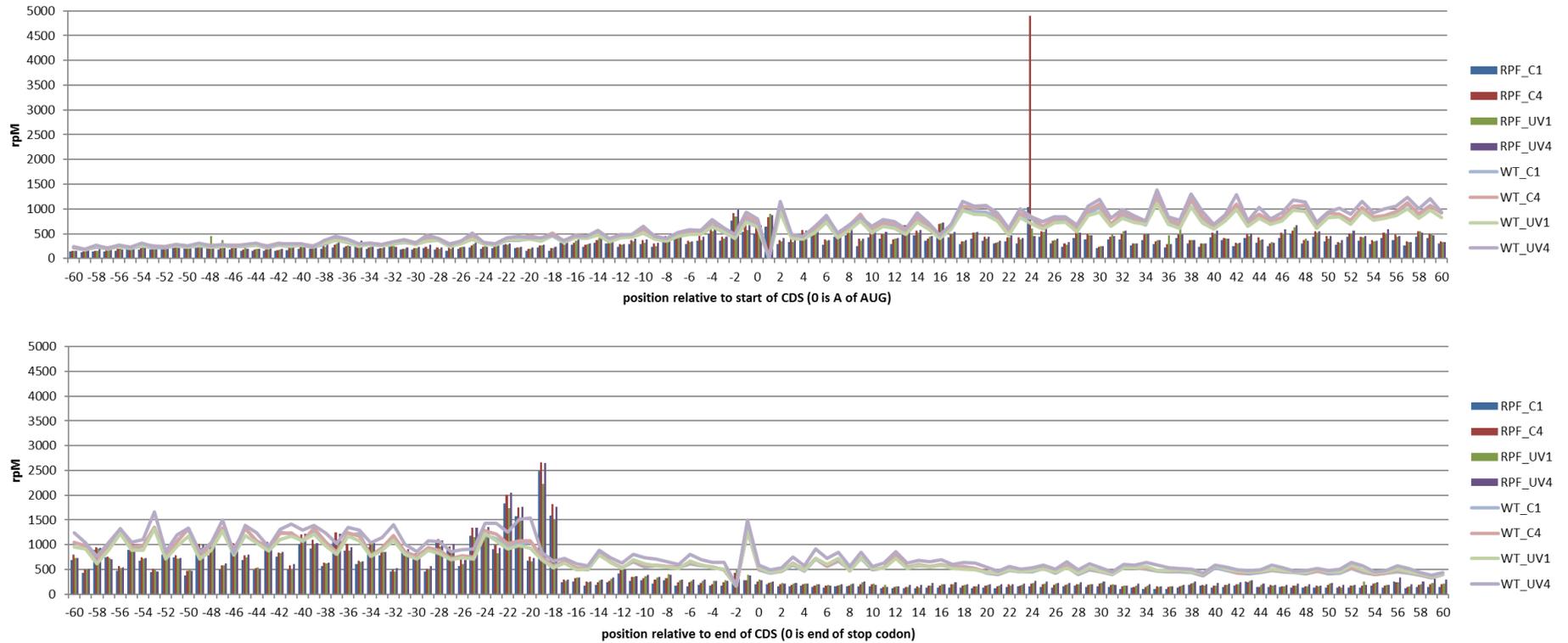
ribosomes may act as a hindrance to the RNase I for access and thereby digestion, resulting in incomplete digestion of the polysomes to 80s monosomes. The polysome containing regions would have been subsequently excluded during the size selection steps of the RPF sample preparation protocol. Regions containing less densely packed ribosomes would remain largely unaffected by this phenomenon, and thereby potentially explaining the observed data in which characteristic hallmarks of RPFs such as 3-nucleotide periodicity and higher read counts were prevalent towards the end of the CDS compared to the start. Indeed, when the read densities were plotted across an extended range, there appeared to be diminished numbers of reads mapping to the very start of the CDS as compared to regions slightly further into the CDS (Figure 4.15). Of course, one cannot fail to acknowledge the sheer impact of the overall low read counts for the RPF dataset, due to the tRNA contamination. From the data shown here, it is not possible to predict whether the same effects would have been observed had there been greater overall RPF read counts as input.

In spite of overall low mapped read counts in the global averaged data for regions at the start of the CDS, there was a distinct spike in reads at position 24 relative to the start of the CDS (Figure 4.13). When investigated further however, this increase was found to be due to non-reproducible variations, and as this effect was only observed in one of the three repeats, for each of the two control condition samples, it is highly likely that this is simply the result of bias or an artefact during the library preparation or sequencing stages rather than a biological cause.



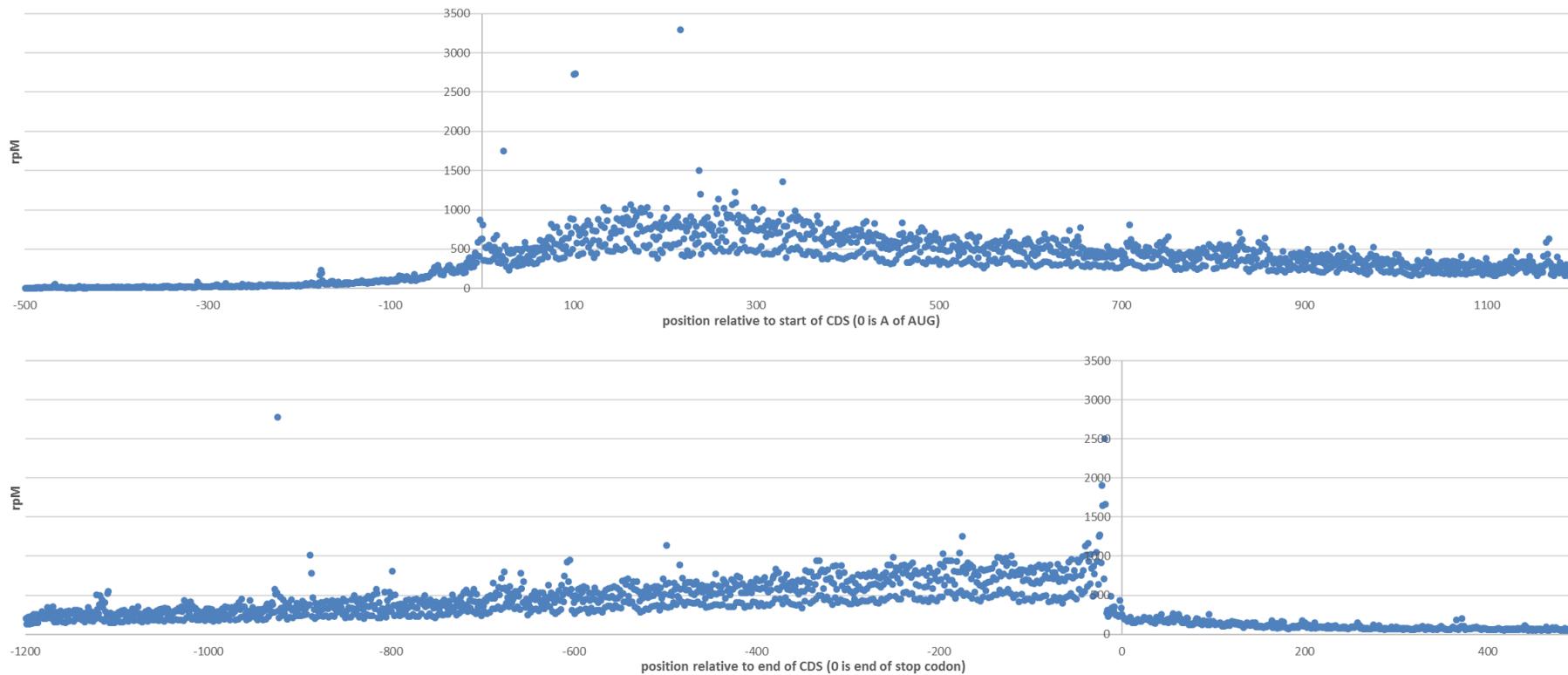
**Figure 4.13 | Periodicity mapping of globally averaged ribosome profiling data**

Global averaged read density mapping of both RPF and whole transcriptome datasets, with plots shown representing the normalised read counts (reads per million) at positions -60 to +60 relative to the start (top) and end (bottom) of the CDS. Normalised read counts for the RPF dataset are represented by the blue bar chart, whole transcriptome dataset counts are represented by the overlaid red line graph.



**Figure 4.14 | Periodicity mapping of ribosome profiling data averaged across conditions**

Averaged read density mapping across all three biological repeats for each treatment condition for both the RPF and whole transcriptome datasets, with the same layout as in Figure 4.13. Normalised read counts for the RPF dataset are represented by the blue bar charts, whole transcriptome dataset counts are represented by the overlaid red line graphs.



**Figure 4.15 | Extended periodicity mapping reveals diminished numbers of reads mapped to the start of CDS**

Read density mapping using the global averaged data from the RPF dataset across an extended range, showing a diminished number of reads mapped relative to the start of the CDS as compared to the region immediately following the start. This extended mapping, along with the previous mapping figures (Figure 4.13, Figure 4.14), lacks the characteristic spike in reads mapped relative to the start of the CDS region as would have been expected in a cycloheximide pre-treated sequencing sample set.

#### *4.4.5.2 Low numbers of genes identified to be translationally significantly different between treatment conditions*

Translatome analysis of the ribosome profiling data was performed using the R package Babel (Olshen et al., 2013b), a brief explanation of its use is described in Section 2.4.3.8.6, with a much more detailed description available in the Babel vignette on the CRAN server hosting the package (<https://cran.r-project.org/web/packages/babel/index.html>). This R package was created specifically for the analysis of ribosome profiling data, and fundamentally operates to identify genes which contain unusual RPF counts given their whole transcriptome read counts. This is then used to calculate p-values across each replicate for each treatment condition. These p-values are then used to identify genes where the relationship between the RPF and whole transcriptome read counts has changed between the treatment conditions.

As expected, when plotting the two datasets together, an increasing function between the RPF and whole transcriptome counts was observed (Figure 4.16). Meaning that on average, highly abundant transcripts (which results in high read counts) were also highly translated, this was largely consistent between all the biological replicates.

The output list from the Babel translatome analysis however, was more surprising. Only 91 and 116 genes were identified as significantly differently translated at the 1 and 4 hour time points respectively, between the treatment conditions. A situation that was not improved even when the false detection rate cut-off was increased to 25%. Of the 116 genes that were significant at the 4-hour time point, the vast majority (105 genes in total) were found to be up-regulated in response to the UV-B, with only 11 genes found to be down-regulated following UV-B. Amongst the identified genes, a significant proportion corresponded to small nucleolar RNA (snoRNA) and zinc finger proteins (Figure 4.17). snoRNAs are a group of non-protein coding RNA molecules, and are associated with ribosomes and rRNA maturation (Lafontaine, 2015). In humans, over 90% of snoRNA genes are within introns (Dieci, Preti, & Montanini, 2009). Interestingly though, of the intronic snoRNA host genes, most belong to the TOP (terminal oligopyrimidine) gene family, a class of genes characterised by their highly regulated translation in response to stress (Yamashita et al., 2008). The presence of these reads corresponding to snoRNAs in the RPF dataset however was intriguing, as these are not translated it seem unlikely that they represent true RPF sequences. Translating 80s ribosomes however are not the only structure capable of protecting short RNA fragments from digestion, protein-RNA complexes may also shield against RNase I digestion, and would therefore result in the co-purification of these sequences along with RPFs.

#### *4.4.5.3 Insufficient read depth resulting in problematic analysis*

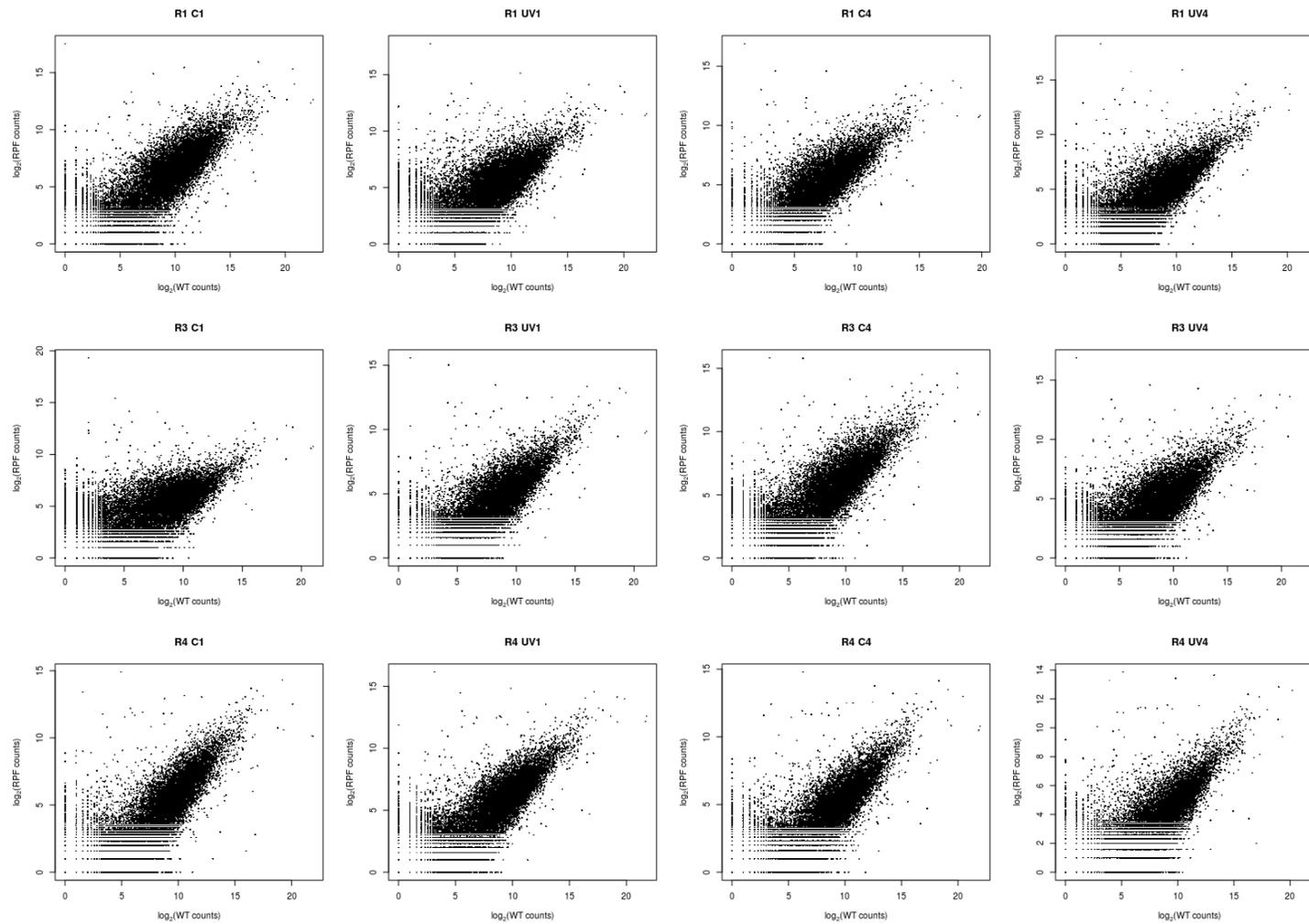
The low numbers and types of genes identified during the attempted translome analysis were perhaps not unexpected given the low numbers of reads associated with the RPF dataset. Insufficient RPF reads counts in these cases may have resulted in the inability for the Babel R package to accurately identify significant genes.

As a crude method of judging the effective coverage given by the RPF dataset, one can estimate the average number of reads mapped per transcript. The number of transcripts present within cells will vary depending on a number of conditions such as cell type, but an upper estimate gauges this value at about 300,000 transcripts per cell (Marinov et al., 2014). Using this number, and the global average number of reads for the RPF dataset (2.7 million reads), results in approximately 9 mapped reads per transcript. This was clearly a very rough method of estimating the coverage as it assumed reads were evenly distributed across all transcripts, which is not the case. However, it serves to illustrate the problem that this RPF dataset simply did not contain enough read depth. This was especially apparent in the RPF sample with the lowest read count (of only around 950,000 reads), resulting in only roughly 3 mapped reads per transcript using this crude measure.

An alternative, perhaps somewhat less crude, method of calculating the approximate coverage obtained in the RPF dataset would be to divide the length of all the mapped reads by the total exon length. The former can be approximated by using 30 nucleotides as the average RPF length, multiplied by the average total number of reads mapped for all genes for the RPF dataset, as determined by HTSeq-count, which came to 1,750,935 reads. Calculation of the total length of all transcripts was performed by parsing the UCSC hg38 gene annotation file and summing the lengths of all exons present, resulting in a total length of 165,159,533 nucleotides. It is important to note, however, that this value was an overestimation of the real total length, as it did not take into account overlapping exons, which were instead summed as if they were non-overlapping. Using these figures, the approximate coverage was calculated to be only 0.32x. By comparison, using the same method, coverage in the whole transcriptome dataset was estimated to be 15.4x (using an average read length of 75 nucleotides, and a total of 33,967,841 reads mapped for all genes by HTSeq-count).

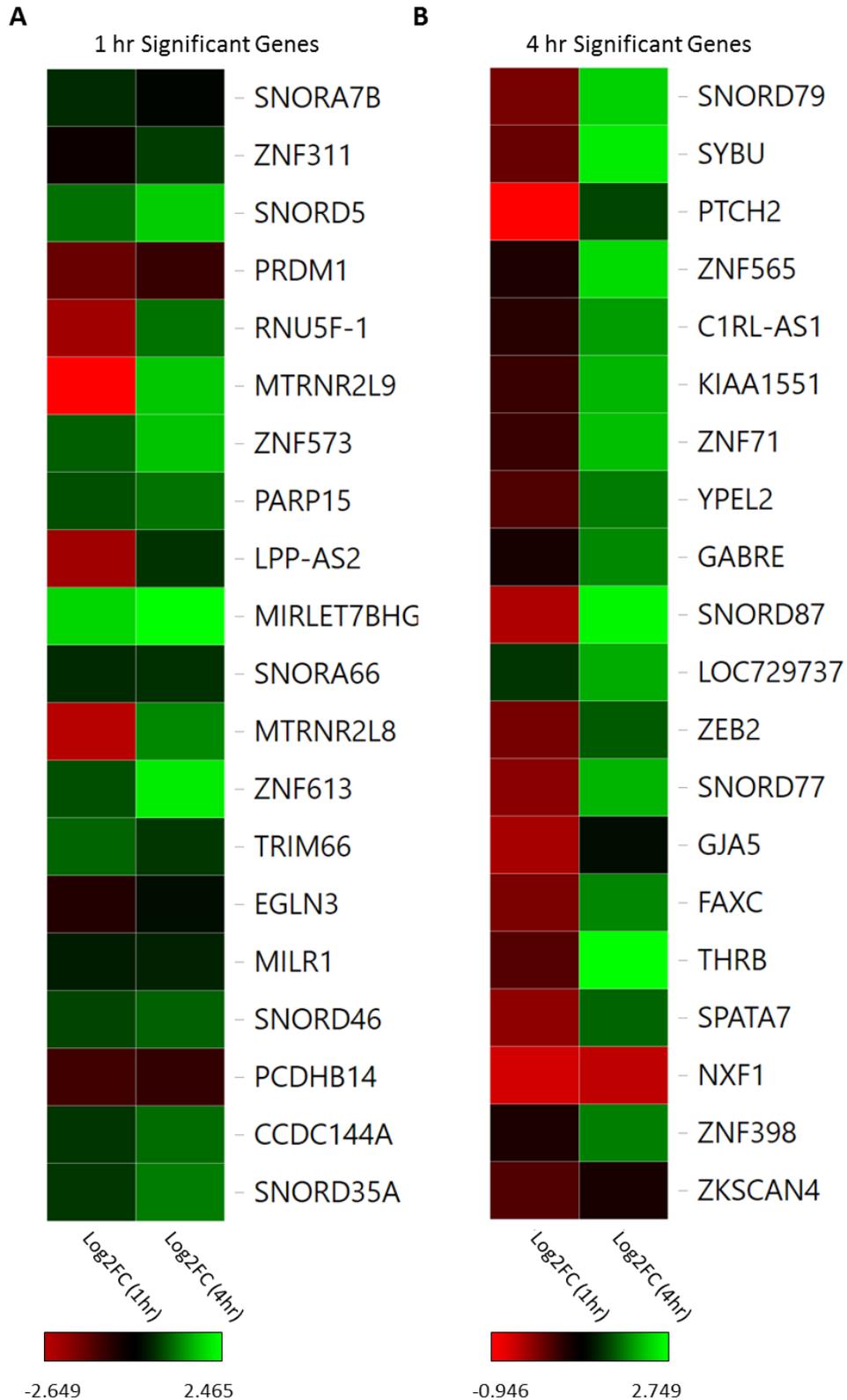
Whilst these values represent fairly crude estimations, they do serve to highlight the significantly lower coverage of reads achieved in the RPF dataset for this sequencing project.

The causes and potential solutions to this insufficient read depth issue are discussed in more detail in the following chapter.



**Figure 4.16 | Correlation between RPF and whole transcriptome read counts**

Scatter plots using log values of the RPF and whole transcriptome read counts were generated for each individual sequencing run, covering all four treatment conditions, and the three biological repeats (R1, R3 and R4). Plots were generated in R using the built in plotting module.



**Figure 4.17 | Babel translome analysis top differentially translated genes output**

The top 20 most significant differentially translated genes following UV-B as identified by Babel at the **A)** 1 hour and **B)** 4 hour time points, with the corresponding log fold change values for those genes at both time points represented on a colour scale, with green representing an up-regulation and red representing down-regulation following UV-B, and the range of log fold change shown in the colour bar below the heatmaps. For the 1 hour time point, a total of 51 genes were identified with an FDR value of 0.05 or lower, and at 4 hours, 72 genes were identified.

## ***4.5 Discussion***

In this chapter, ribosome profiling data is presented for UV-B treated MCF10A cells using the conditions optimised in Chapter 3. The proportions of the different species of RNA present within the RPF sequencing datasets were surprising. While within the whole transcriptome dataset, the majority of the reads mapped to the genome, with, on average, 65% mapped, and 33% corresponding to rRNA contaminating reads. The RPF dataset was more complicated, with, on average, approximately 80% of the reads corresponding to contaminating sequences. Detailed breakdown of the distribution of reads revealed that on average 70% of the reads corresponded to tRNA sequences, with 9% matching rRNA sequences, and only 14% of the reads were mapped to the genome (Figure 4.5). Though the presence of contaminating reads is seen in all ribosome profiling sequencing data, the levels at which they occur in this dataset is abnormally high. The potential sources responsible for the high level of contamination, as well as possible methods to mitigate them are discussed in detail in the following chapter.

The low resulting numbers of reads within the RPF datasets following the removal of contaminating sequences ultimately proved to be problematic for the transcriptome analysis that was conducted. With an average of only approximately 2.7 million reads per sample, the dataset simply did not achieve enough coverage for successful transcriptome analysis. Periodicity mapping conducted on the RPF dataset revealed only minor differences in the numbers of reads mapped to regions within and outside of the CDS, though a characteristic three nucleotide periodicity was present, verifying the mapped reads as RPFs. Interestingly, the peak in the numbers of reads typically found mapped to the start of the CDS in cycloheximide treated experiments was absent in this dataset (Figure 4.13). This may potentially be indicative of issues with the RNase I digestion step, possibly resulting in the insufficient digestion of unprotected mRNA between closely stacked ribosomes, as would be the case in cycloheximide treated samples. This incomplete digestion would then ultimately result in the loss of those RPFs during the subsequent size selection stages. These issues are investigated in more detail in the following chapter.

Whilst the RPF sequencing dataset was marred with a number of issues, the whole transcriptome dataset contained no such problems. Preliminary transcriptome analysis conducted on this dataset identified 234 and 4083 significantly differentially expressed genes at 1 and 4 hours following UV-B irradiation, respectively. Results from qPCR analysis on a select number of the top most significantly differentially expressed genes were largely consistent with these transcriptome analysis results generated by DESeq2 (Figure 4.12). Further analyses

conducted on these lists using the Ingenuity Pathway Analysis (IPA) program revealed the regulation of a large number of pathways such as oxidative phosphorylation and mitochondrial dysfunction, which were consistent with previous studies using UV-B (Torregrosa-Munumer et al., 2015; Tsai et al., 2009). Indeed, a number of key DNA damage and repair pathways such as p53 signalling and ATM signalling were present in the IPA analysis at both time points, with other major pathways such as eIF2 signalling identified to be the most significantly changing following UV-B by the 4 hour time point. The role of the eIF2 pathway in translational control is well known, which is activated by a family of protein kinases, which are able to phosphorylate eIF2 in response to a number of environmental stresses. These data are consistent with the western blot analysis, which showed phospho-eIF2 $\alpha$ , from 1 hour to 7 hours following UV-B. The critical role of eIF2 $\alpha$  was also demonstrated using ISRIB, which potently reverses the effects of eIF2 $\alpha$  phosphorylation, showing that at 1 hour post treatment, the observed UV-B induced responses were almost entirely reversed when investigated using polysome profiling. However, further investigation is required to study the impact of the regulation of these individual pathways on the translational responses observed following UV-B exposure.

## ***5. Identification of the sources of contamination in the ribosome profiling sequencing data***

### ***5.1 Introduction***

As highlighted in the previous chapter, significant proportions of the RPF sequencing dataset contained contaminating reads, with, on average just below 80% of all reads filtered out. Of all the reads, on average, approximately 9% corresponded to rRNA sequences, an encouraging indicator of the efficacy of the Ribo-Zero Gold rRNA depletion kit. Surprisingly, over 70% of RPF reads corresponded to tRNA sequences. The significant proportion of tRNA reads was both unexpected, and also ultimately resulted in too few reads remaining for translome analysis to be successfully conducted. Specifically, the sequences of these identified fragments, which were approximately 31-34 nucleotides in length, all corresponded exactly to the 5' terminal of the respective full length tRNA.

Though historically largely ignored, with the recent development of high throughput sequencing technologies, an increasing number of small RNA types have been identified and associated with regulatory functions (Reon & Dutta, 2016). Amongst these were tRNA-derived RNA fragments (Y. S. Lee et al., 2009). These fragments generally vary between 17 to 35 nucleotides in length depending on their exact formation. Of the different classes of tRNA-derived RNA fragments (tRF), the most common classes include tRNA halves (sometimes referred to as tiRNAs), and Dicer cleaved tRNA fragments originating from both the 5' and 3' termini of the mature tRNA structure (resulting in 5' tRF and 3' CCA tRF sequences respectively). tRNA halves are typically composed of fragments of 30-35 nucleotides (Raina & Ibba, 2014), and are derived from either the 5' or 3' ends of the full length mature tRNA (termed tRF5 or tRF3 respectively). These tRNA halves are produced in response to a number of different stresses including, but not limited to, nutrition deficiency, arsenite treatment, UV irradiation and hypoxia (Fu et al., 2009; Yamasaki, Ivanov, Hu, & Anderson, 2009). In mammalian cells, these tRNA halves are formed by the cleavage within the anticodon loop of the mature tRNA through the actions of angiogenin, a member of the RNase A family of nucleases (Fu et al., 2009).

The impact of these detected contaminating reads in the RPF dataset detailed in the previous chapter, as well as the potential source and possible methods for reduction of these reads are discussed in this chapter.

## ***5.2 Sources of contaminating tRNA reads***

### ***5.2.1 Identification of primary contaminating reads***

The sequencing data for all RPF samples were processed through FastQC (Andrews, 2016), following which the generated list of over-represented sequences was run through the Clustal Omega web service (McWilliam et al., 2013) for multiple sequence alignment and thereby clusters of analogous reads were identified. The resulting output is summarised in Figure 5.1. The top three overall most abundant sequences present in the RPF dataset were processed using NCBI BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and were identified as tRNA<sup>Val(AAC)</sup>, tRNA<sup>Gly(GCC)</sup> and tRNA<sup>Lys(CTT)</sup>, which are highlighted in red, green and yellow respectively in Figure 5.1. These three sequences were present in all 12 RPF samples, and combined, accounted for 49% of all reads on average, with the sequence tRNA<sup>Val(AAC)</sup> alone accounting for just under 28% of all reads on average (Figure 5.2). The sequences of these fragments (including fragments which were less abundant), which were approximately 31-34 nucleotides in length, all corresponded exactly to the 5' terminal half of the respective full length tRNA. The length, and specificity of these abundant 5' tRNA halves (tRF5) in our dataset matched observations in several previous studies which had also discovered their presence in a number of different cell types and conditions (Dhahbi et al., 2013; Fu et al., 2009; Honda et al., 2015).

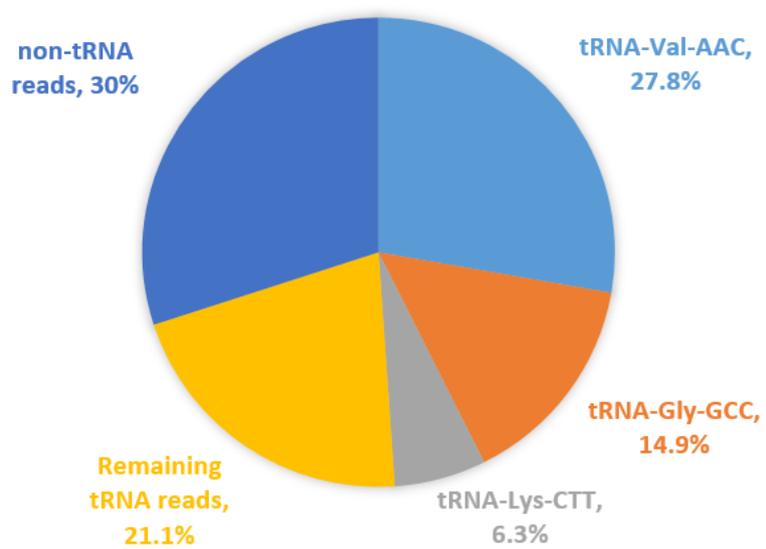
### ***5.2.2 Basal levels of tRFs in MCF10A cells***

The tRF5s identified and summarised in Figure 5.2 showed comparable levels of the top three most abundant tRF sequences, with their combined abundances displaying no obvious changes after UV-B treatment, or between biological replicates. As mentioned previously, the abundance of tRFs has been found to vary significantly depending on the cell or tissue type (Fu et al., 2009). In an attempt to investigate the basal levels of tRFs present in MCF10A cells in both control and UV-B treated conditions, northern blotting was conducted. This procedure was performed according to the methods described in Section 2.4.5, with the exception of the preparation of RNA samples. For this experiment, total RNA samples were obtained directly from 10 cm plates of PBS washed cells using the miRNeasy RNA extraction kit (Qiagen) according to the manufacturer's protocol. As a positive control, HeLa cells were treated with 500  $\mu$ M sodium arsenite (a known inducer for the production of tRNA halves) (Yamasaki et al., 2009) and collected at 2 and 4 hours. These samples were then probed for the two most abundantly over-represented tRF species, namely tRF5-Gly(GCC) and tRF5-Val(AAC). The resulting exposed blots are shown in Figure 5.3.

Despite these two species of tRF5s accounting for over 40% of all reads in the ribosome profiling RPF dataset, they were largely undetectable in both the control and the UV-B treated MCF10A cells. Distinct bands of around 30 nt in size were detected using both probes for the sodium arsenite treated HeLa cells at both 2 and 4 hour time points. As expected the full length tRNA pool remained unchanged between the different treatment conditions (Figure 5.3).

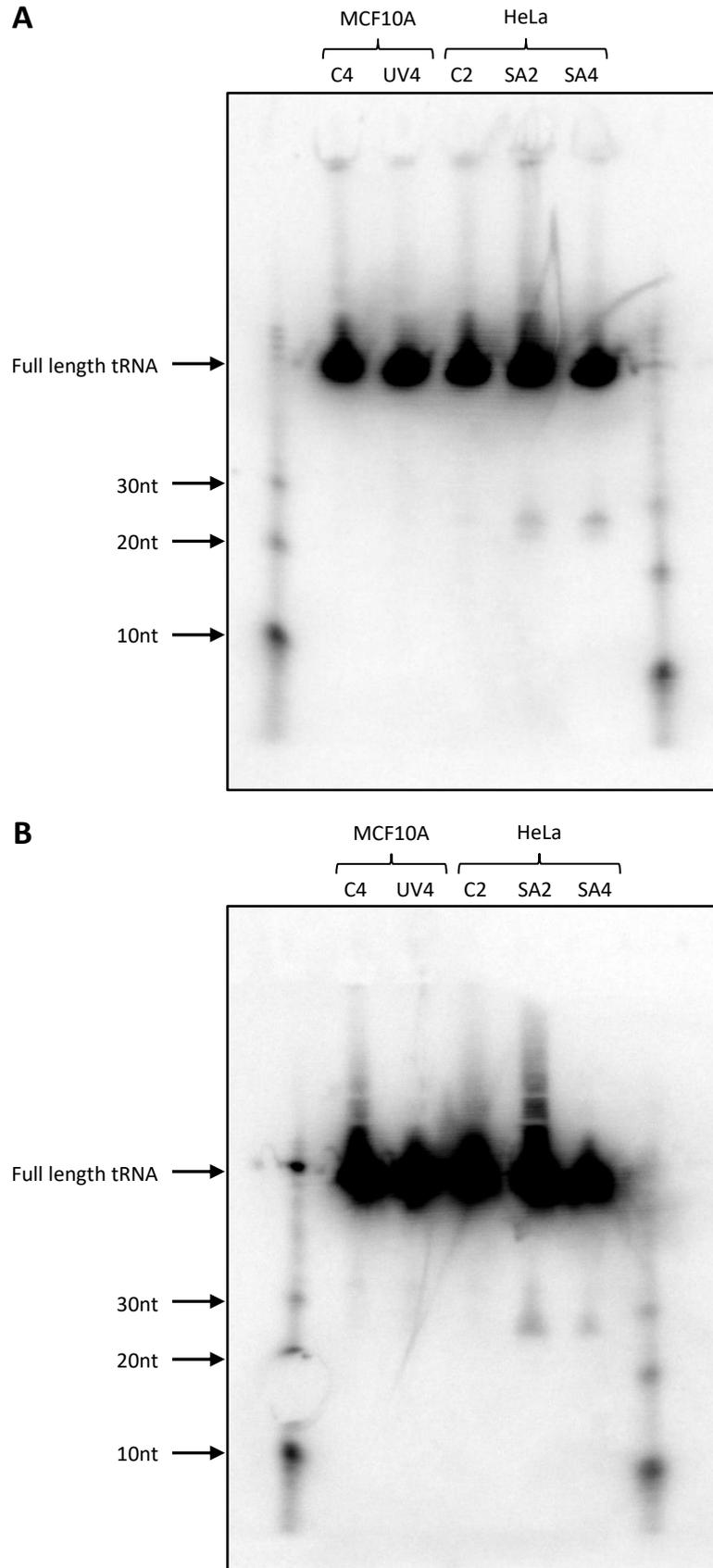
The absence of noticeable levels of these tRF5s in the total RNA extracts from both the control and UV-B treated MCF10A samples suggested that the high levels of contaminating tRF reads in the RPF dataset were possibly a result of either the sample preparation or other factors following it, rather than the cells containing innately high levels of tRF5s.





**Figure 5.2 | Distribution of over-represented sequences in RPF dataset**

Summary of the output generated in Figure 5.1, highlighting the distribution of the top 3 most abundant sequences which were present in all samples, as well as the proportions of the remaining reads.



**Figure 5.3 | Low levels of 5' tRNA halves present in both control and UV-B treated samples**  
 10µg of RNA were loaded per lane and run as per the method described in Section 2.4.5. MCF10A cells were either mock treated or exposed to 200 Jm<sup>-2</sup> UV-B. HeLa samples were collected following either mock, or 500µM sodium arsenite treatment after 2 and 4 hours. Blots were probed for **A)** tRF5-Gly-GCC and **B)** tRF5-Val-AAC.

### **5.2.3 Enrichment of tRNA halves by size exclusion spin columns**

As the endogenous levels of tRF5s in MCF10A cells did not appear inherently high, additional northern blots were conducted at different stages of the RPF sample preparation process. As mentioned previously and speculated in the previous chapter, it was thought that the RNase I digestion step during the sample preparation process could have had a large impact on the final outcome, and as such, three different RNase I concentrations were tested (0.1, 0.5 and 5 U/ $\mu$ l). For reference, the optimal RNase I concentration found from titrations conducted in the previous chapter, and the final concentration used for the ribosome profiling samples was 0.5 U/ $\mu$ l. Samples for each of the three concentrations were collected, and aliquots were processed through the illustra MicroSpin S-400 HR size exclusion columns (GE Life Sciences) as described in Section 2.4.3.3.2. An additional sample, digested with 0.5U/ $\mu$ l RNase I and run through the same size exclusion columns, but centrifuged at 600xg for 2 minutes according to the illumina TruSeq protocol was also included, instead of 670xg for 1 minute as described in the methods. The samples were then run on 15% TBE-Urea Novex pre-cast gels (Invitrogen), and transferred onto Hybond NX membranes as described in the methods, after which, they were incubated using radiolabelled probes corresponding to both the 5' and 3' tRNA halves of the most abundant tRF species, tRNA<sup>Val(AAC)</sup> (Figure 5.4).

There was a clear correlation between the concentration of RNase I used, and the abundance of tRFs that were detected. For the total RNA sample digested using the highest concentration of RNase I (5U/ $\mu$ l, lane 7 in Figure 5.4A, B), the levels of both the full length tRNA as well as the 5' and 3' tRNA halves were dramatically reduced when compared to those for 0.5 U and 0.1 U/ $\mu$ l RNase I digested samples. While both the 5' and 3' tRNA halves were detected following RNase I digestion, significantly higher levels of the 5' species were present in all samples.

Interestingly though, simply processing the samples through the size exclusion columns also seemed to have a large effect on the abundance of the resulting tRFs. Whilst the use of these columns resulted in a large decrease in the abundance of the full length tRNA for all concentrations of RNase I used, the levels of the tRNA halves remained very high. This was especially true for the 5' tRNA halves (Figure 5.4A), to a degree, using the size exclusion columns effectively resulted in purification of these tRNA halves. This was also observed in the sample purified using the illumina TruSeq variant of the protocol, which differs minutely in the preparation and speed used for the size exclusion columns, indicating that this result was caused by the columns themselves rather than the spin conditions used (compare lane 1 to 3 in Figure 5.4A, B).

A seemingly obvious, but naïve conclusion to draw from these results would be that to reduce the levels of contaminating tRNA reads in future ribosome profiling experiments, one could simply increase the concentrations of RNase I used for digestion. This would certainly achieve a reduction in the abundance of tRFs present (and therefore the numbers of reads corresponding to those sequences), as shown in the previous chapter, digestion conditions also drastically changes the levels of RPFs which are able to be recovered (Figure 4.2). This approach would therefore ultimately result in either low proportions of real RPFs being captured, or introduce significant biases into the final RPF pool due to effectively only capturing a subset of all ribosomes.

#### ***5.2.4 Reduced levels of tRNA halves contamination using sucrose cushions***

The purification of RPFs is a critical step in the ribosome profiling technique, and although the use of size exclusion columns resulted in significant levels of tRNA halves contamination, it is not the only method for extracting the RPFs. Alternative approaches include the use of sucrose gradient fractionation and subsequent RNA extraction of monosome containing fractions, and the use of sucrose cushions. Only the latter approach is shown here, however the use of sucrose gradient fractionation was investigated during the preliminary setup and optimisation stages of the ribosome profiling project; this approach resulted in very low extracted RNA yields and was not pursued further.

Samples were digested using 0.1, 0.5 and 5 U/ $\mu$ l of RNase I and each of these samples were then purified using two different sucrose cushion purification protocols, as described in Section 2.4.3.3.3. A reference sample consisting of 0.5 U/ $\mu$ l RNase I digested lysates purified using the size exclusion columns, as well as a sodium arsenite treated HeLa control sample were also included. The Northern blot was run using the protocol described in Section 0, and the resulting blots were incubated with radiolabelled probes corresponding to the 5' and 3' tRNA halves of tRNA<sup>Val(AAC)</sup>. The intensities of the observed bands corresponding to the 5' and 3' tRNA halves were quantified using ImageJ (Schneider, Rasband, & Eliceiri, 2012), and the resulting values were plotted relative to the reference sample in both panels A and B (Figure 5.5, lane 7).

With one exception, all of the digested samples purified using sucrose cushions displayed drastically reduced abundances of tRNA halves when compared to the reference sample purified using size exclusion columns (Figure 5.5). The only increase relative to the reference sample observed was for the level of tRF3-Val-AAC in the 0.5 U/ $\mu$ l digested sample purified using Ingolia's sucrose cushion method (lane 2 in Figure 5.5B), although it should be noted that the

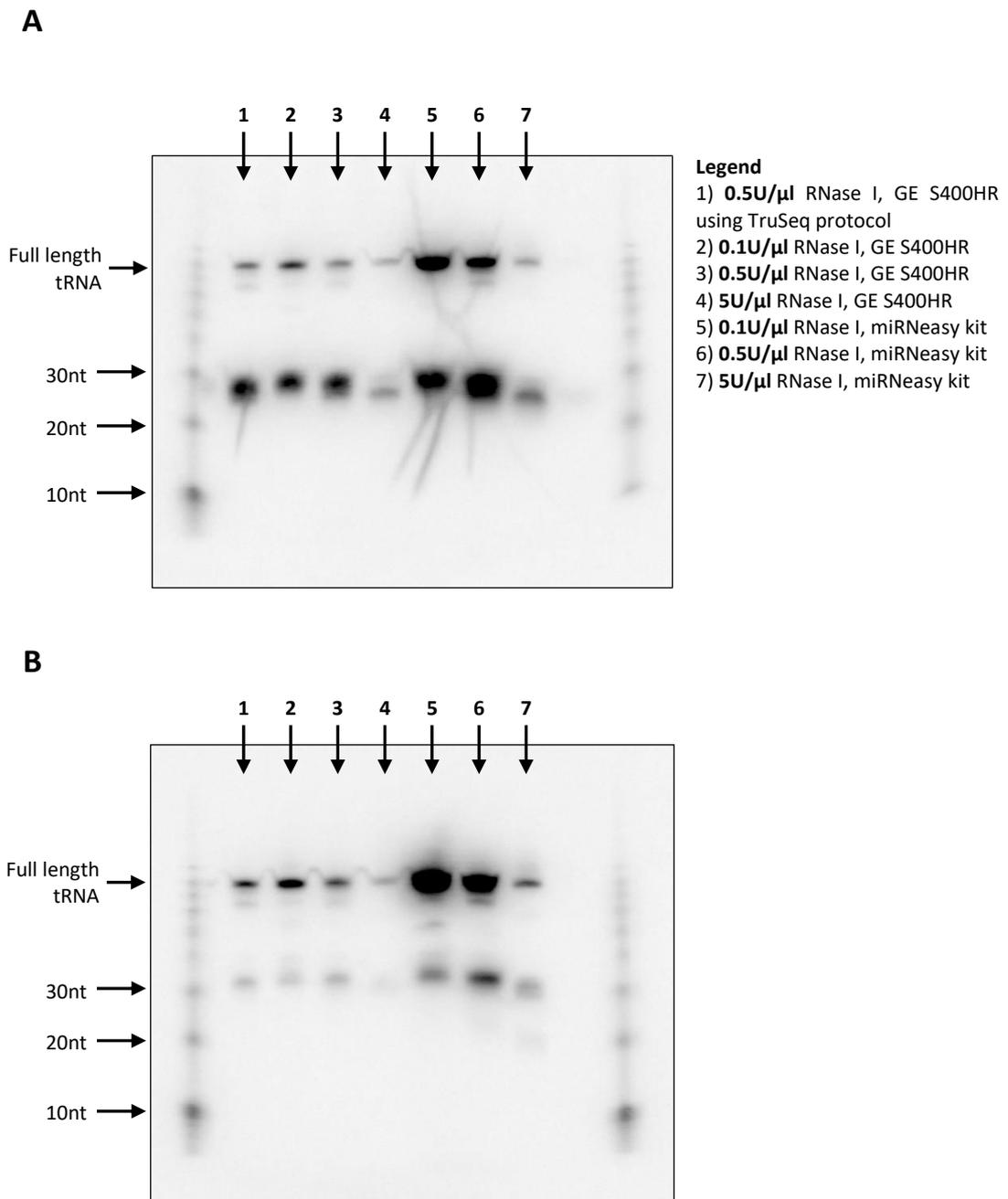
levels of tRF3-Val-AAC were relatively low in all samples (Figure 5.5B). This was especially noticeable when comparing the ratio of the full-length tRNA band signal to that of the tRF3 band signal between these samples in samples processed using the size exclusion columns vs sucrose cushions (Figure 5.4B vs Figure 5.5B).

Between the two sucrose cushion protocols used, those samples that were processed using the illumina TruSeq variant method displayed consistently lower abundances of tRFs across all the different digestion conditions. This was most likely due to the higher concentration of sucrose used for the formation of the cushion, and therefore resulted in a more conservative purification of the RPFs.

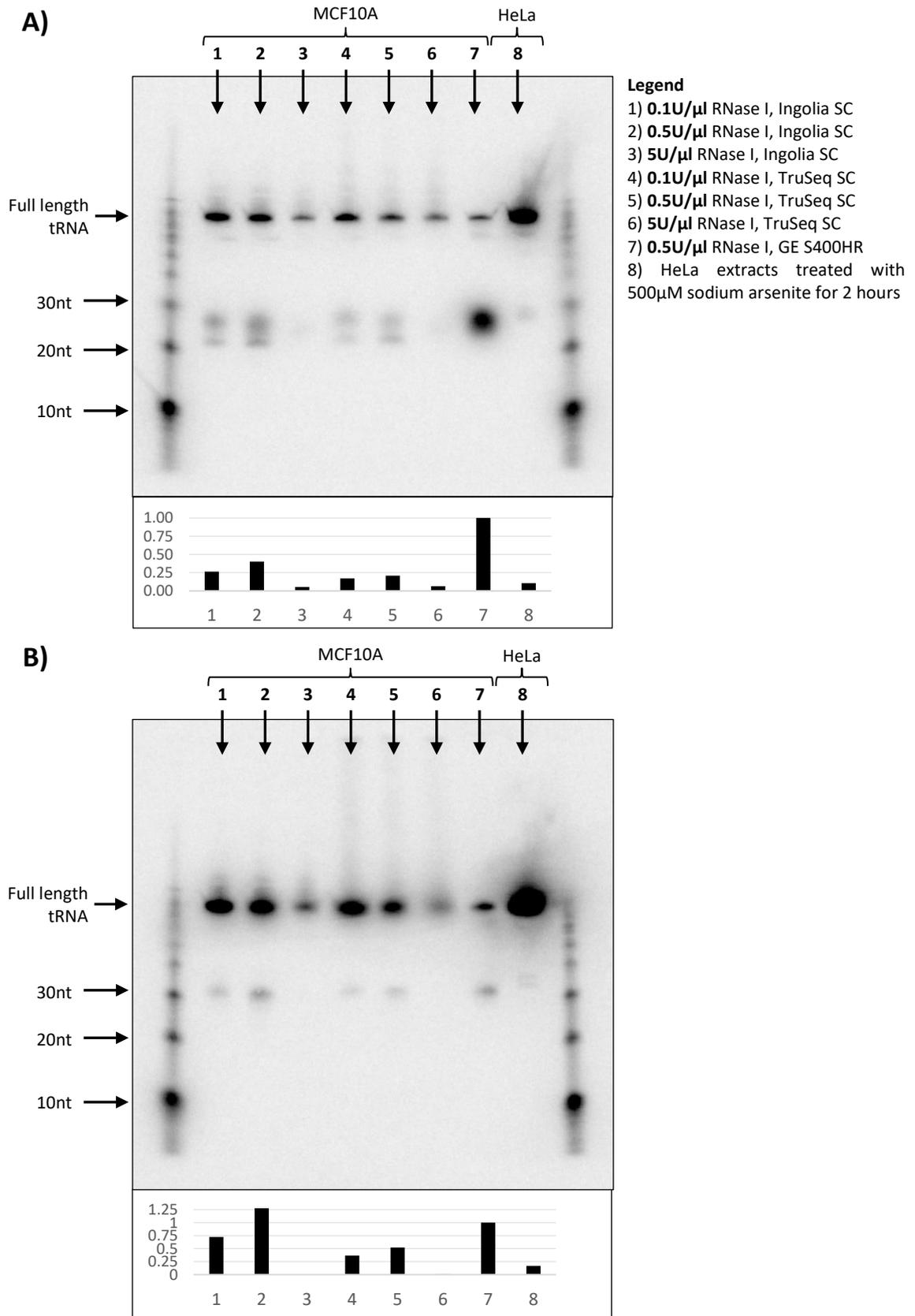
### ***5.2.5 Sucrose cushion purified samples reveals problems with RNase I digestion***

Given the apparent benefits of using sucrose cushion purification of RPFs instead of size exclusion columns, further investigations were performed. Polysome profiling was performed on samples which were digested with 0.1, 0.5 and 1 U/ $\mu$ l of RNase I, as well as samples extracted following sucrose cushion purification using both the Ingolia and the illumina TruSeq methods. In contrast to the previous sections, an upper concentration of 1 U/ $\mu$ l of RNase I was selected instead of 5 U/ $\mu$ l, based on the extremely strong over-digestion found for the 5U/ $\mu$ l concentration in Section 5.2.4. The polysome profiling traces of all samples are shown in Figure 5.6. It should be noted, however, that the traces for the unpurified digested samples are shown at a lower sensitivity scale compared to the samples purified by the sucrose cushion.

Interestingly, the numbers of monosomes obtained following the illumina TruSeq variant of the sucrose cushion purification procedure seemed to be substantially greater across all three digestion conditions, which also had fewer tRFs being purified. Worryingly however, the traces for the non-purified digested samples seemed to indicate problems with the RNase I digestion itself, with clear signs of over-digestion at both the 0.5 and 1 U/ $\mu$ l RNase I concentrations (Figure 5.6C, E). Even at the lowest tested concentration of 0.1 U/ $\mu$ l, the size of the peak corresponding to monosomes was substantially lower than that observed in the previous RNase I titrations (Figure 4.2).

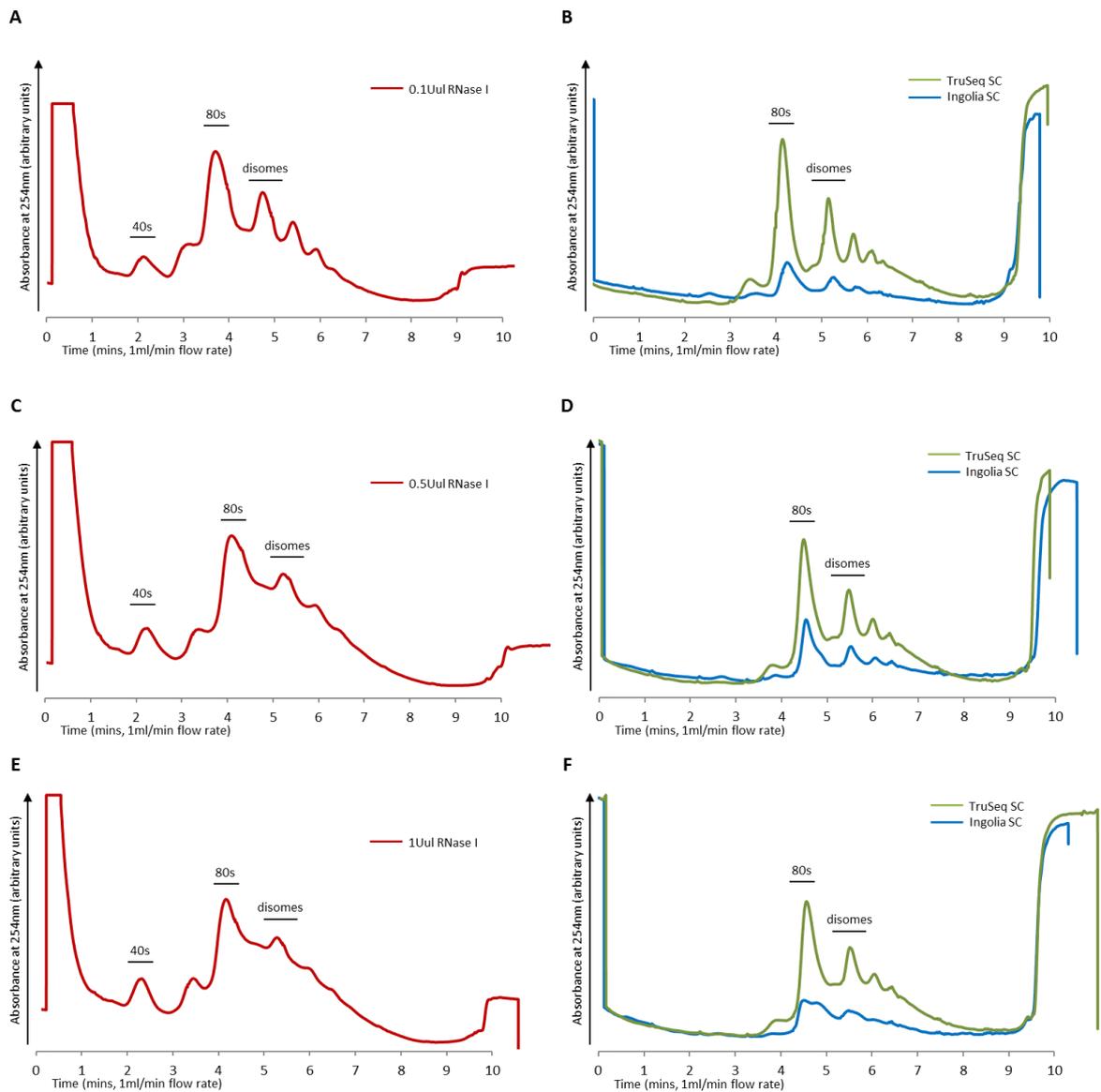


**Figure 5.4 | Purification of RPFs using size exclusion columns results in the enrichment of tRF5s**  
 6μg of RNA were loaded per lane and run as per the method described in Section 2.4.5. Untreated MCF10A cells were RNase I digested and purified using the GE S400HR size exclusion columns or extracted using the miRNeasy kit as per the conditions in the figure legend. Blots were then probed for **A)** tRF5-Val-AAC and **B)** tRF3-Val-AAC.



**Figure 5.5 | Reduced levels of tRF5 enrichment with sucrose cushions**

3μg of RNA were loaded per lane and run as per the method described in Section 2.4.5. Untreated MCF10A cells were RNase I digested and purified using either sucrose cushions (SC) using the TruSeq or Ingolia protocol (Section 0), or purified using size exclusion columns as listed in the figure legend. Blots were then probed for **A)** tRF5-Val-AAC and **B)** tRF3-Val-AAC.



**Figure 5.6 | Polysome profiling of sucrose purified samples reveals problems with RNase I digestion**  
 Identical plates of untreated MCF10A cells were harvested and digested using RNase I at concentrations of **A)** 0.1U/ $\mu$ l of lysate **B)** 0.5U/ $\mu$ l and **C)** 1U/ $\mu$ l, following which aliquots were purified with sucrose cushions (SC) using either the illumina TruSeq protocol or Ingolia's protocol (as described in Section 2.4.3), with the resulting traces displayed in **E, D** and **F** respectively. Detection sensitivity was altered for the SC purified samples, and so the traces from A, B and C cannot be directly compared to E,D and F.

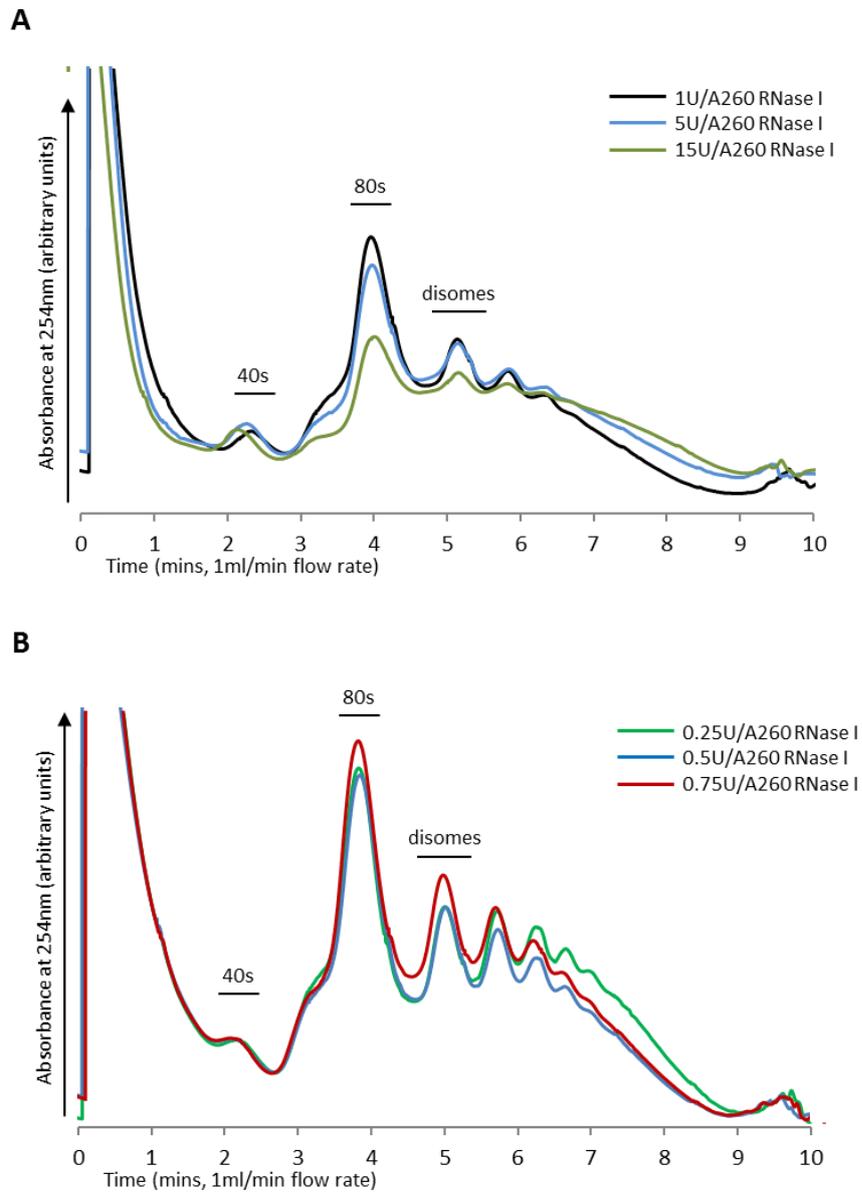
## ***5.3 Troubleshooting and optimisation of RNase I digestion***

### ***5.3.1 Repeat run of RNase I titration reveals digestion issues***

As mentioned previously, the RNase I digestion of the post nuclear lysates is a crucial step in the ribosome profiling technique which has a significant impact on the outcome of the final RPF RNA samples and therefore eventual sequencing dataset. The results in the previous section have highlighted concerns over the use of size exclusion columns for the purification of the RPFs as these appear to disproportionately select for the tRF species which were highly abundant in the ribosome profiling project detailed in Chapter 4. However, the data shown in Figure 5.6 also highlights issues with the RNase I digestion prior to purification. While these results did not discount the impact of the use of size exclusion columns on the levels of tRF contamination observed, they exposed the possibility that sub-optimal RNase I digestion could also have had a role in the issues encountered in the RPF dataset.

To better explore the impact of RNase I digestion, and to investigate the reasons why the previously determined optimal concentration of RNase I no longer seemed to result in ideal digestion, additional titrations were performed covering a range of 0.25U to 15U/A<sub>260</sub> of RNase I. In an attempt to provide more consistent results, the method in which the quantity of RNase I was determined was changed from using units per microliter of lysate (U/μl) to measuring using units per absorbance at 260 nm (U/A<sub>260</sub>). For reference, the previously used RNase I concentrations of 0.1, 0.5 and 1 U/μl approximately corresponds to 2, 10 and 20 U/A<sub>260</sub> respectively.

The polysome profiling traces in Figure 5.7 showed a clear over-digestion of the lysates in all but the lowest of concentrations of RNase I tested. Even at 2 U/A<sub>260</sub> of RNase I (a concentration equivalent to 5x lower than what was used during the ribosome profiling work for the previous chapter), resulted in sub-optimal digestion. A lower RNase I titration concentration range revealed that 0.75 U/A<sub>260</sub> resulted in the best digestion, as determined by both the area of the 80s monosome peak, as well as the ratio between the 80s peak and the subsequent disome, trisome and heavier polysomal fractions (Figure 5.7B). Lowering the RNase I concentration to 0.25 U/A<sub>260</sub> showed slight under-digestion with the clear presence of undigested heavy polysomal fractions, whilst RNase I concentrations above approximately 1 U/A<sub>260</sub> showed increasingly diminishing sizes of the 80s monosome peaks (Figure 5.7A).



**Figure 5.7 | RNase I titrations reveals issues with digestion**

Identical plates of untreated MCF10A cells were harvested and digested using RNase I at concentrations shown. Different detection sensitivities were used for traces from panel **A** and **B**, and so cannot be directly compared.

### ***5.3.2 Unknown causes for the changes observed in RNase I digestion***

The RNase I titration experiments in the previous section revealed a new, substantially lower optimal concentration of RNase I for ideal digestion compared to those used in the previous experiments (Section 4.2.1 and Figure 4.2). This approximately 13-fold decrease in the ideal RNase I concentration (going from the equivalent of 10 U/A<sub>260</sub> to just 0.75 U/A<sub>260</sub>), was unexpected, and a number of possible causes for this change were explored.

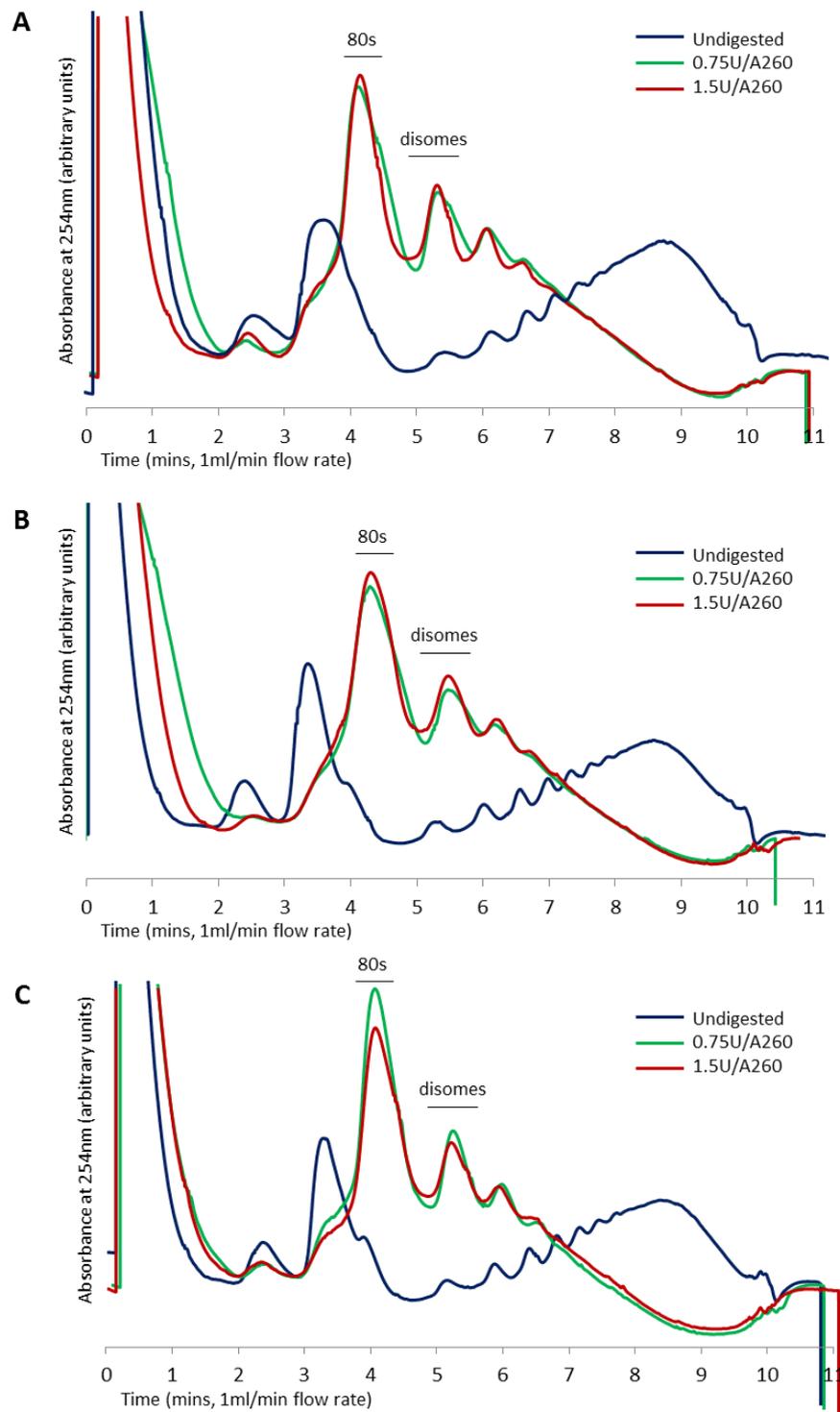
As the efficacy of the digestion is greatly dependant on the quantity of RNA present in the post nuclear lysates, the effect of the cell seeding density was investigated. Whilst the method for determining the quantity of RNase I that was used in this chapter takes into account potential variations in RNA quantities between different samples (U/ A<sub>260</sub>), the method previously used in Chapter 4 did not. Previously, RNase I quantity was determined based on units per µl of lysate. This relied on the assumption that the quantity of RNA extracted per sample remained approximately the same. While this assumption was broadly accurate, potential deviations in either the cell seeding or RNA extraction process could have resulted in differences in the quantities of RNA extracted between the samples from the previous chapter compared to those used for this chapter. To investigate this parameter, 15cm plates seeded with 10 million (as used previously), as well as 7.5 and 5 million MCF10A cells were prepared identically to before, and as described in Section 2.4.2 and 2.4.3.2. The extracted post nuclear lysates from these were then digested using both the optimal RNase I concentration of 0.75 U/A<sub>260</sub>, and a higher concentration of 1.5 U/A<sub>260</sub>. As is clear in Figure 5.8, whilst the traces were obtained using different detection sensitivities on the UV/Vis detector due to the differing RNA quantities present, there appeared to be minimal changes in the digested traces for all three seeding densities. It should be noted however that the trace obtained for the samples seeded with 5 million cells (half of what was used in Chapter 4) appeared to contain the best digested trace out of the three seeding conditions (Figure 5.8C), though this still in no way accounts for the drastic change observed in the optimal RNase I concentrations between the previous chapter (Figure 4.2) and the current chapter.

Other considerations for potential causes of the observed change in the efficacy of the RNase I digestion include the use of cycloheximide pre-treatment during the sample collection process. As described previously, cycloheximide blocks the translocation step, and thereby effectively blocks elongation. This has a useful effect of “freezing” the ribosome in place during the sample collection process and thus allowing for the accurate capture of RPFs at the time of collection. However, its use has been shown to cause the stacking of ribosomes near the initiation site

(Ingolia et al., 2012). It is possible that this tight stacking of ribosomes could have restricted the access of RNase I to the mRNA regions between them, and therefore resulted in these polysomal regions being sub-optimally digested. While cycloheximide pre-treatment was used in both Chapter 4 and the current chapter, differences in its potency may have resulted in variations in the degree of the stacking of ribosomes present around the initiation sites, which could have thereby affected digestion. To investigate this possibility, samples were prepared with and without cycloheximide pre-treatment before RNase I digestion, and examined by polysome profiling method as described above. The obtained traces were overlaid and are shown in Figure 5.9A. Evidently, given the near identical traces obtained with or without cycloheximide pre-treatment, this premise that its use could have impacted RNase I digestion was incorrect.

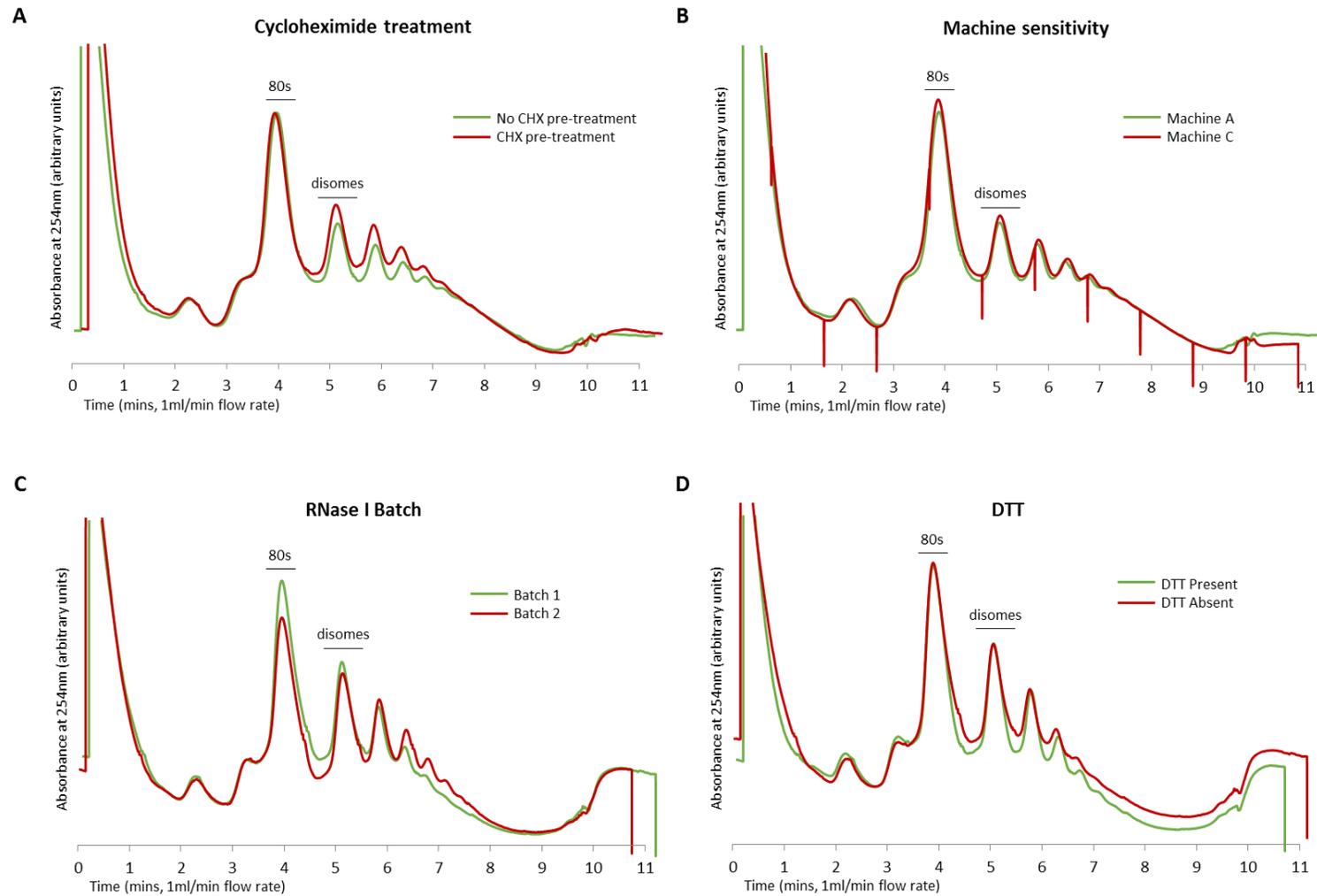
A number of other factors that could possibly have had an impact on the large change in RNase I digestion were tested, and are summarised in Figure 5.9 and Figure 5.10. These included factors such as possible differences in sensitivity between the UV/Vis detectors used (Figure 5.9B), differences in the efficacy of the RNase I itself between batches (Figure 5.9C), the presence of DTT in the buffers used (as oxidised DTT exhibits strong absorbance at 260 nm and could therefore interfere with the UV/Vis detector) (Figure 5.9D), as well as digestion incubation time (Figure 5.10A) and temperature (Figure 5.10B, C). Unfortunately, none of these factors impacted on the resulting digested traces in a meaningful way, and did not lead to an explanation of the differences observed between the results obtained in this chapter, and those obtained during the ribosome profiling project in the previous chapter.

Similarly, the possible impact of the RNase inhibitor used was also investigated. As increasing the concentrations of RNase I did not appear to result in more complete digestion, but rather primarily the degradation of the monosome peak, it was thought that perhaps the RNase inhibitor that was used in the experiments was interfering with the activity of RNase I. However, samples digested in the presence of an alternative RNase inhibitor (Cambridge Bioscience NxGen RNasin), or processed without any RNase inhibitor, all showed the same response (Figure 5.11). Similar issues were encountered when RNase I digestion was tested using HEK293 cells (Figure 5.11C, D), which have been previously successfully used for ribosome profiling (Ingolia et al., 2012). Despite the use increasing concentrations of RNase I for both MCF10A and HEK293 cells, there appeared to be an inability to digest the heavier polysomes down into single monosomes as was observed in the RNase I titration shown in the previous chapter (Figure 4.2, Figure 5.11).



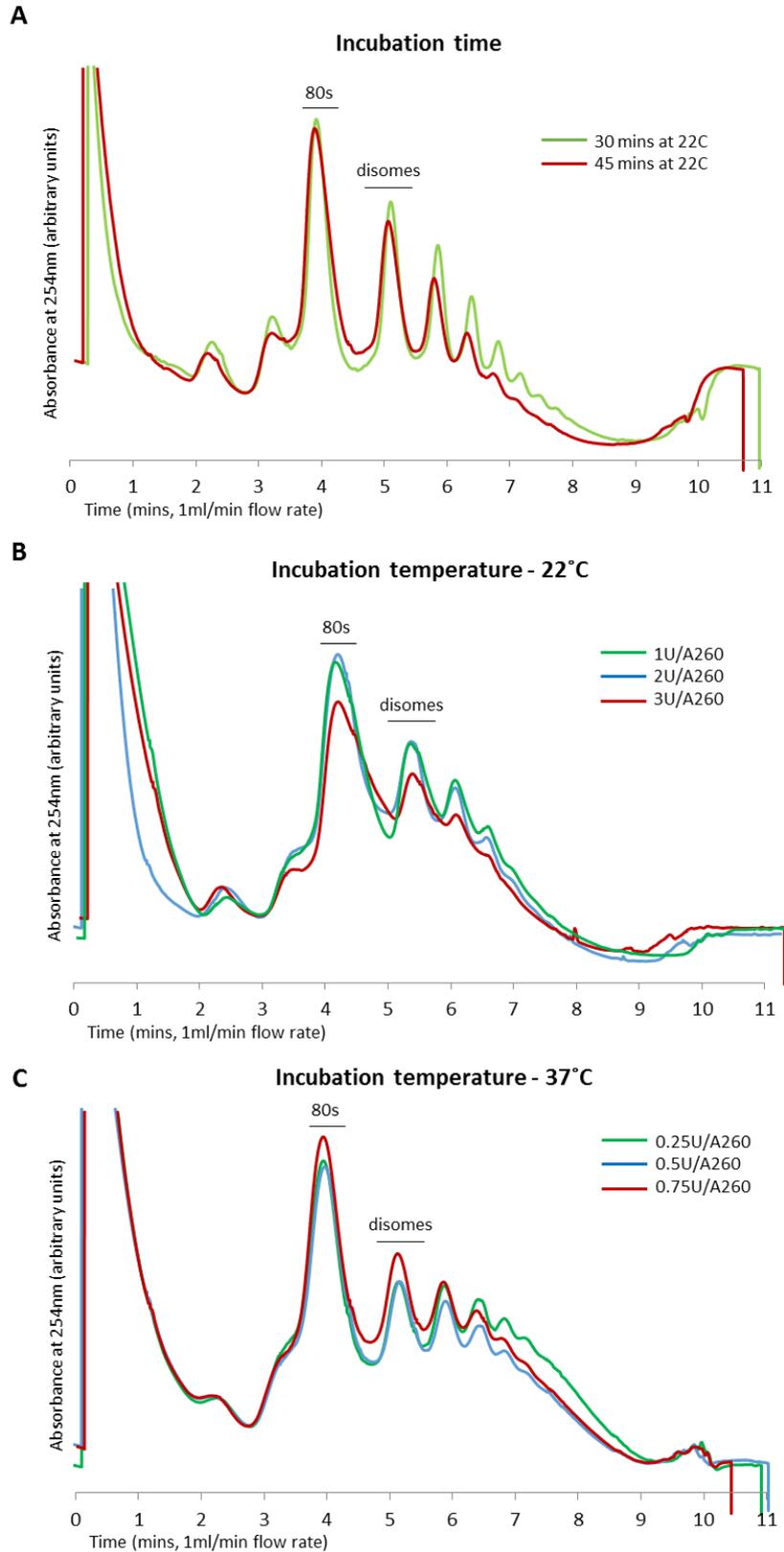
**Figure 5.8 | Impact of cell seeding density on RNase I digestion**

15cm plates seeded using **A)** 10 million cells **B)** 7.5 million cells or **C)** 5 million cells of MCF10A were harvested and digested using RNase I at concentrations shown. Different detection sensitivities were used for traces from panel **A**, **B** and **C**, and so cannot be directly compared.



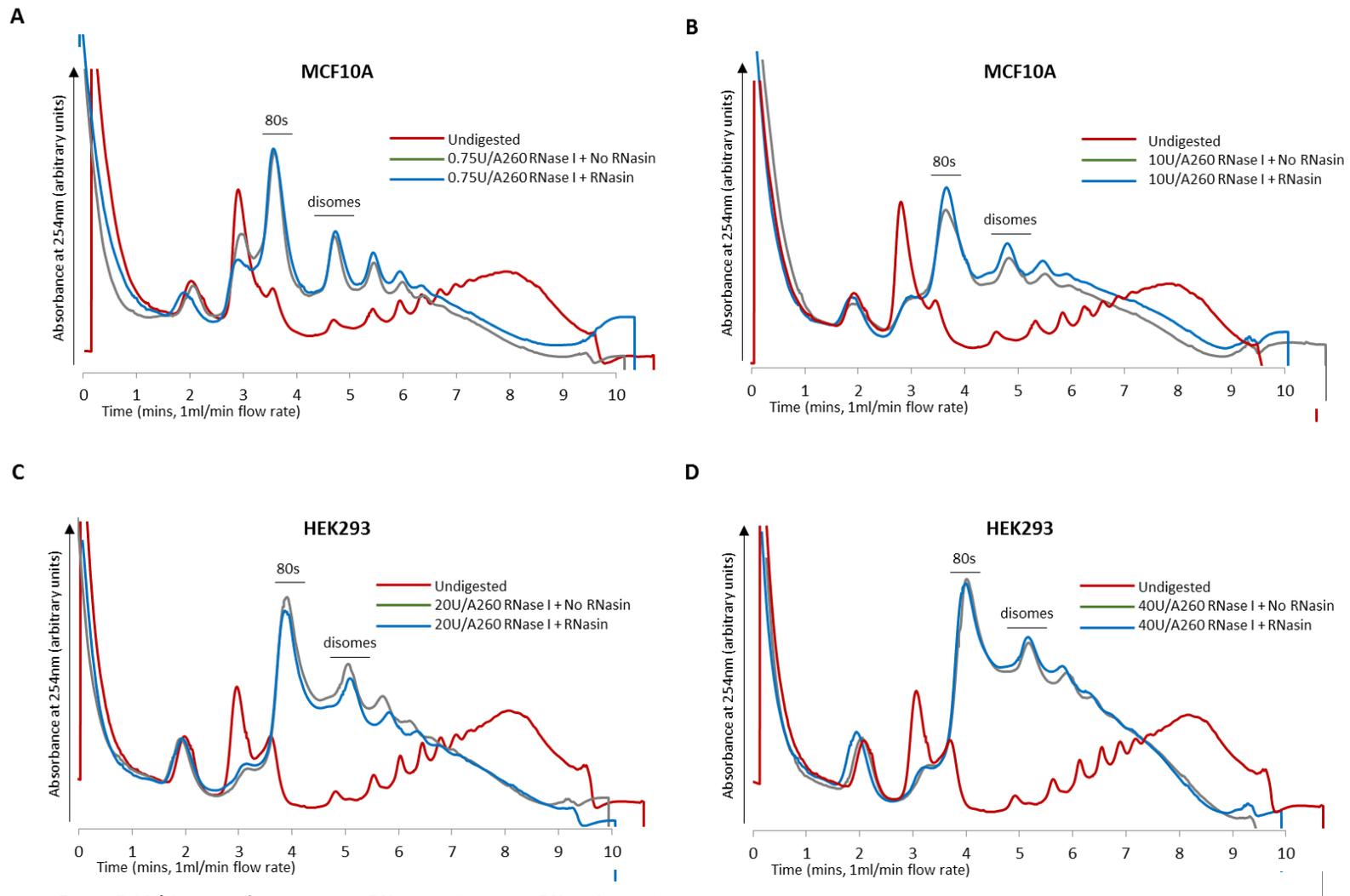
**Figure 5.9 | Impact of multiple factors on RNase I digestion (Part 1)**

Identical plates of untreated MCF10A cells were harvested and digested using 0.75U/A260 RNase I, and used to test **A)** impact of cycloheximide during sample harvest **B)** detection machine sensitivity variation **C)** RNase I batch variation **D)** impact of DTT in lysis buffers used. Samples used for the traces were collected as described in Section 2.4.3.2.



**Figure 5.10 | Impact of multiple factors on RNase I digestion (Part 2)**

Identical plates of untreated MCF10A cells were harvested and digested using concentrations of RNase I shown, and used to test **A)** impact of digestion incubation time, and **B-C)** impact of digestion temperature. Samples used for the traces were collected as described in Section 2.4.3.2.



**Figure 5.11 | Impact of cell type and RNase inhibitors on RNase I digestion**

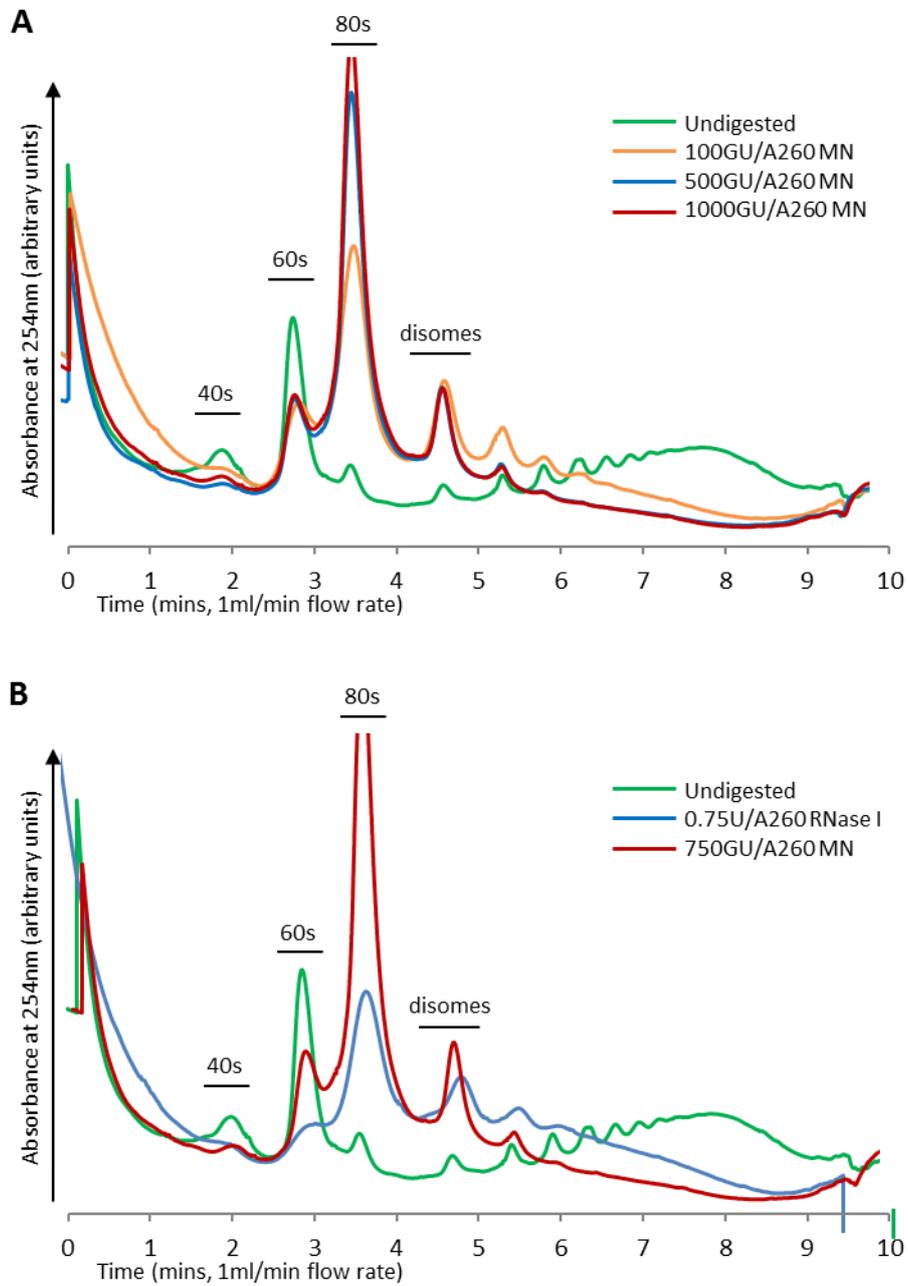
Identical plates of untreated cells were harvested and digested using concentrations of RNase I shown, and used to test the efficacy of digestion with no RNasin present, or with 200U/ml NxGen RNasin (Cambridge Bioscience) in **A-B**) MCF10A cells and **C-D**) Hek293 cells. Samples used for the traces were collected as described in Section 2.4.3.2.

### ***5.3.3 Improved digestion using micrococcal nuclease***

The results of Sections 5.3.1, and 5.3.2 highlighted issues surrounding both the digestion of RNA with RNase I, and also the as yet unresolved cause of these sub-optimal digestion issues. As previously mentioned, this digestion of the post nuclear lysate sample is a crucial step in ribosome profiling, and so in this section, the use of micrococcal nuclease as an alternative to RNase I for this step was investigated. Though many recent ribosome profiling studies in eukaryotic systems have employed the use of RNase I due to its lack of strong sequence specificity and robust footprinting (Ingolia et al., 2012), profiling of bacterial systems have historically been largely reliant on micrococcal nuclease, with some of the early identified limitations such as reduced footprint resolution due to stronger nucleotide preferences been mitigated through recent analysis methodologies (Oh et al., 2011; Woolstenhulme, Guydosh, Green, & Buskirk, 2015). Indeed, recent studies comparing the use of a number of nucleases have shown that the resulting profiling data displays largely comparable estimations of gene expression (Gerashchenko & Gladyshev, 2017). The use of micrococcal nuclease also appears to spare ribosomal RNA more so than RNase I (Dunn, Foo, Belletier, Gavis, & Weissman, 2013), which may result in fewer contaminating reads in the final library and thus potentially increasing sequencing read depth.

Similar to Section 5.3.1, post nuclear lysates were extracted from untreated 15 cm plates seeded with 10 million MCF10A cells were prepared as detailed in Section 2.4.2. These lysates were then digested using micrococcal nuclease at a concentration ranging from 100 GU/A260 to 1000 GU/A260. For comparison, a sample digested using RNase I at the previously determined optimal concentration of 0.75 U/A260 was also included. As is clear from the polysome profiling traces in Figure 5.12A, all concentrations of micrococcal nuclease digested samples resulted in the efficient digestion of polysome to mostly single 80s ribosomes. This is especially apparent in Figure 5.12B when directly comparing with the sample digested using RNase I where the size of the 80s monosome peak obtained was approximately 3 times as large.

Interestingly, the micrococcal nuclease titrations in Figure 5.12A also appear to show much greater tolerance for the exact concentration of the nuclease required, with there being only minor differences in the polysome profiling traces for samples digested using 500 and 1000 GU/A260 of micrococcal nuclease. This sharply contrasts with the RNase I titration results where there was a steep decline in the size of the resulting 80s monosome peak with increasingly sub-optimal concentrations (Figure 5.7).



**Figure 5.12 | Improved digestions using micrococcal nuclease**

Identical plates of untreated cells were harvested and digested using concentrations of RNase I and Micrococcal Nuclease (MN) shown, with **A** showing the subset of the range of MN concentrations tested, and **B**) used to compare the efficacy of digestion between the two nucleases using the determined optimal concentrations for both.

## ***5.4 Discussion***

The work described in this chapter focused on investigating the causes of, and possible solutions for the significant quantities of contaminating reads present within the RPF dataset from the ribosome profiling project detailed in Chapter 4. Analysis of the contaminating reads revealed that the majority of the over-represented sequences were formed by just three species of 5' tRNA halves, specifically tRF5-Val(AAC), tRF5-Gly(GCC) and tRF5-Lys(CTT). Combined, these were responsible for almost half of all reads from the RPF dataset (Figure 5.2).

It was also discovered that while the endogenous levels of these tRF5 species appeared to be relatively low in undigested MCF10A cells for both mock and UV-B treated samples (Figure 5.3), the use of the illustra MicroSpin S-400 HR size exclusion columns (GE Life Sciences) for the purification of the RPFs using RNase I digested samples seemed to lead to significant increases in the levels of tRFs (Figure 5.4). A recent study has also observed the abundance of the same specific species of tRF5s described above, and crucially, have also demonstrated that these 5' tRNA halves are present in serum and in a number of tissues as abundant complexes of between 100 - 300 kDa (Dhahbi et al., 2013). This size range is well within the limits of the size exclusion column, and may explain the large increase in tRF levels following their use.

Interestingly, the purification of RPFs using sucrose cushions instead of size exclusion columns resulted in a much lower abundance of tRFs (Figure 5.5). The reduction in the amount of these tRFs should also translate to fewer contaminated reads in future ribosome profiling experiments, and result in greater amounts of usable data. Another possible RPF purification method, which was not tested in this project, would be to use sucrose gradient density fractionation, and then subsequent extraction of RNA from the appropriate collected fractions. This approach has the benefit of the ability to precisely extract only the fractions containing the 80s monosomes, and should effectively filter out most if not all of these significantly shorter (and therefore lighter) tRF species. While this procedure would in theory produce the purest RPF containing sample, when trialled during the preliminary stages of the ribosome profiling work, very low yields of RNA were obtained compared to the other two methods described above. The low yield presented issues when performing rRNA depletion using the Ribo-Zero Gold kits (illumina), which involves a PAGE size selection step following the rRNA removal, and resulted in quantities of RNA, which were too low for library preparation to be carried out. One possible work-around for this problem, which was not tested, would be to carry out the PAGE size selection step prior to treatment using the rRNA depletion kit. This would, in effect, dramatically increase the quantity of RNA that is able to be loaded onto the Ribo-Zero Gold kits. However, reversing these

steps would most likely also impact the efficacy of the rRNA depletion step, though the extent to which this would occur and whether it would be a worthwhile trade-off, is unknown without further experimentation.

Another issue discovered during the troubleshooting was the apparent problems with RNase I digestion. Compared to the previous concentration of RNase I that was required for optimal digestion as determined by the titrations conducted in the previous chapter (Figure 4.2), the optimal concentration determined when these titrations were re-run was around 13 times lower. In addition, there appeared to be less complete digestion, with greater proportions of heavier polysomal fractions remaining, even when concentration of RNase I was increased. A number of factors that could have potentially been responsible for the difference observed, such as the use of cycloheximide pre-treatment, cell seeding density, batch variation, digestion incubation temperature and digestion time were tested. However, none of these factors had a significant impact, and it was not possible to resolve the differences observed between the RNase I digestion results in the previous chapter and in this chapter (Figure 5.8, Figure 5.9, Figure 5.10, Figure 5.11). Though, as there were no aliquots remaining from the batch of RNase I used in Chapter 4, it is impossible to know fully the extent to which batch variation had on the digestions. However, investigations into using micrococcal nuclease as an alternative to RNase I resulted in significantly improved digestion, with an approximately 3 fold increase in the size of the 80s monosome peak obtained from polysome profiling (Figure 5.12B). This drastic increase in the quantity of resulting monosome following digestion, in addition to the increased range in concentration in which optimal digestion occurs (Figure 5.12A) strongly positions micrococcal nuclease as a viable replacement for RNase I for future ribosome profiling projects.

## ***6. Summary***

The overall aims of this project were to investigate the selective translational reprogramming which occurs following DNA damage, as well as to assess and develop a feasible protocol for the use of ribosome profiling to measure these changes. To this end, the data shown in this thesis has outlined this investigation using the non-transformed breast epithelial MCF10A cell line. The feasibility of using two DNA damage sources were initially investigated in Chapter 3. Due to previous experience in the Willis lab, ionising radiation (IR) and UV-B were chosen as the DNA damage sources for investigation (Powley et al., 2009; Young, 2011). Though both IR and UV-B elicited dose and time dependant DNA damage responses (Figure 3.2, Figure 3.8), as well as reductions in protein synthesis (Figure 3.3, Figure 3.7), only UV-B resulted in large translational changes as measured by polysome profiling (Figure 3.5, Figure 3.9). These larger observed translational changes were deemed more favourable, as previous experiments conducted in the Willis lab using ribosome profiling have shown that the technique itself results in the generation of large inter-repeat variances (Jackson, 2013), meaning that minor changes between conditions might not be observed. Therefore, the use of UV-B as the DNA damage source was selected for further investigations using ribosome profiling.

Chapter 4 outlined the exploration of the ribosome profiling data obtained from UV-B treated MCF10A cells using the conditions optimised in Chapter 3. Analysis of the resulting whole transcriptome sequencing dataset using the R package DESeq2 (Love et al., 2014) identified 234 and 4083 significant differentially expressed genes at 1 and 4 hours following UV-B irradiation respectively, with the top most differentially expressed genes shown in Figure 4.9. qPCR analysis run on a select number of the top most significantly differentially expressed genes were largely consistent with the results generated by DESeq2 (Figure 4.12). Whilst the majority of the genes tested were concordant with the DESeq2 outputs, a few of the probed genes were not. Principle component analysis conducted on this dataset also revealed whilst the largest source of variance at the later 4 hour time point was the treatment itself, at the earlier time point of 1 hour, the different biological repeats seemed to be the larger source (Figure 4.7), which may in part explain some of the non-concordant qPCR analysis results. Further analysis on these lists of identified significant genes using Ingenuity Pathway Analysis (IPA) revealed the regulation of a large number of pathways, including key DNA damage and repair pathways such as p53 and ATM signalling (present at both 1 and 4 hour time points), as well as pathways involved in translation regulation such as the eIF2 pathway (Figure 4.10, Figure 4.11). Indeed, the crucial role of the eIF2 pathway is well known, with changes to the phosphorylation status of eIF2 $\alpha$  in

particular, having a major role in translational regulation following stress. Interestingly, experiments conducted using the drug ISRIB, which potently reverses the effects of eIF2 $\alpha$  phosphorylation, were able to almost fully reverse the translational responses observed following UV-B, though only by the 30 minutes and not at the later time point of 4 hours (Figure 3.10). In the future it would be of interest to examine this pathway in more detail and in particular the cross-talk that occurs between signalling through mTOR and eIF2 alpha. It has been shown that eIF2 $\alpha$  phosphorylation has different effects on cell fate which are dependent on the cell type and the extent of DNA-damage induced such that it can initiate either an apoptotic programme or pro-survival signals. This information is clearly important as it has been suggested that modulators of the eIF2 $\alpha$  response such as ISRIB could be used therapeutically to enhance the efficacy of chemotoxic drugs.

Translatome analysis using the RPF sequencing dataset presented a number of challenges. Due to the heavy contamination present in the RPF dataset, an insufficient number of remaining reads were available for useful analysis with the R package Babel (Olshen et al., 2013a). On average, approximately 80% of all reads from the RPF dataset were identified as contaminated, around 70% in total corresponded to tRNA sequences, with the remainder corresponding to rRNA sequences (Figure 4.5B). Detailed analysis of the contaminated tRNA sequences revealed that the overwhelming majority of the sequences corresponded to 5' tRNA halves (tRF5), with approximately 50% of all reads in the RPF dataset corresponded to just three species of tRF5's (tRNA<sup>Val(AAC)</sup>, tRNA<sup>Gly(GCC)</sup> and tRNA<sup>Lys(CTT)</sup>) (Figure 5.2). The 5' tRNA half corresponding to Val(AAC) specifically accounting for approximately 28% of all reads from the RPF dataset alone. In comparison, the whole transcriptome dataset only contained 33% contaminated reads, of which less than 0.1% corresponded to tRNAs, with the rest being rRNA reads. Whilst these contaminated reads were an issue for analysis, the relatively low quantities of rRNA reads (especially in the RPF dataset) serves to highlight the effectiveness of the Ribo-Zero Gold (illumina) kit. The use of this ribosomal RNA depletion kit has resulted in a marked increase in the proportion of usable reads, excluding tRNA, when compared to previous methods of rRNA depletion which was used in the Willis lab (Jackson, 2013).

Interestingly, previous studies have shown that the production of the tRNA halves correlates with the degree of damage sustained (Mishima et al., 2014), with tRF5's having also been shown to be involved in translational control (Ivanov et al., 2014; Kedersha, Ivanov, & Anderson, 2013). The formation of the tRF5's during stress having been linked to an ability to trigger the inhibition of protein synthesis, as well as the formation of phospho-eIF2 $\alpha$  independent assembly of stress

granules, which were primarily composed of stalled pre-initiation complexes (Ivanov, Emara, Villen, Gygi, & Anderson, 2011; Yamasaki et al., 2009). This may in part explain the disparity observed in the polysome profiles of UV-B treated samples at 4 hours showing significant decreases in the numbers of associated polysomes, but only apparently minor increases in the numbers of subpolysomal fractions (Figure 3.9). Indeed, the formation of stress granules in a phospho-eIF2 $\alpha$  independent manner may also explain the inability of ISRIB in restoring protein synthesis at 4 hours following UV-B despite returning protein synthesis back to near matched untreated sample levels at an earlier time point of 30 minutes (Figure 3.10).

Intriguingly, it has also been shown that transfection of tRF5's, but not tRF3's, also resulted in a slight reduction of protein synthesis (Yamasaki et al., 2009). Though the formation of these tRNA cleavage fragments did not appear to alter the pool of the full length tRNA significantly, and only represent a small proportion of the overall tRNA pool (Ivanov et al., 2011; Saikia et al., 2012). Crucially, the formation of these tRF products did not appear to be simply due to tRNA degradation as the abundance of the different types of tRFs generated did not correlate with the abundance of the parent full length tRNA (Sobala & Hutvagner, 2011). Indeed, it has been previously shown in *Haloferax volcanii* that stress induced tRFs generated from the 5' terminal end of tRNAs (specifically tRNA<sup>Val</sup>, which was the most abundant tRF5 in the RPF dataset) were able to bind directly to the small ribosomal subunit and inhibit translation by the interference of the peptidyl transferase activity (Gebetsberger, Zywicki, Künzi, & Polacek, 2012). Select 5' tRFs, such as 5' tRNA<sup>Ala</sup> and 5' tRNA<sup>Cys</sup>, have also been shown to interact with the translational repressor Y-box binding protein 1 (YB-1) resulting in the displacement of the eIF4F cap-binding complex from capped mRNA, thereby inhibiting translational initiation (Ivanov et al., 2014; Kedersha et al., 2013). Similar mechanisms by 5' tRFs have also been observed in *Saccharomyces cerevisiae* and human cells (Bąkowska-Żywicka, Kasprzyk, & Twardowski, 2016; Sobala & Hutvagner, 2013).

Interestingly, a previous study had also observed the presence of the same specific species of tRF5s found in the RPF dataset (tRNA<sup>Val(AAC)</sup>, tRNA<sup>Gly(GCC)</sup> and tRNA<sup>Lys(CTT)</sup>) (Figure 5.2), and crucially, have also demonstrated that these 5' tRNA halves are present in serum and in a number of tissues as abundant complexes of between 100 - 300 kDa (Dhahbi et al., 2013). Whilst the exact source and the biological relevance of the presence of these specific species of tRNA halves within the RPF dataset remain unknown without further investigation, basal levels of tRFs were found to be near undetectable, suggesting that the abundance of tRFs in the sequencing data was not due to the cell type used (Figure 5.3), although, previous genome wide analyses have

revealed that tRNA levels were enriched in a number of different human cells (Kawaji et al., 2008; Pavon-Eternod et al., 2009). The discovery that the use of the illustra MicroSpin S-400 HR size exclusion columns (GE Life Sciences) resulted in the purification of tRFs was unexpected, in part explained the source of the contaminated tRNA reads in the RPF dataset (Figure 5.4). A potential explanation for the presence of these tRFs following purification using the size exclusion columns could be the formation of these 100 - 300 kDa abundant complexes, which falls well within the working range of the columns. Reduced levels of contamination were found following the use of sucrose cushions for the purification of RPFs (Figure 5.5), for this reason, its use over size exclusion columns in future ribosome profiling projects is recommended. An alternative RPF purification method would be to use sucrose gradient density fractionation, extracting RNA only from the appropriate collection fractions corresponding to the 80s monosomes. This technique in theory should produce the purest RPF containing sample, and was investigated during the preliminary stages of this project. However, due to the low yields of RNA obtained following the purification and extraction, the method was not pursued beyond this RPF purification stage.

During investigations into the tRFs, it was observed that RNase I digestion using previously determined optimal concentration of 0.5U/ $\mu$ l of lysate had resulted in comparatively severe over-digestion (Figure 5.6). Repeat RNase I titrations revealed a significantly lower optimal required concentration, and substantially larger presence of disomes, trisomes and heavier polysomes regardless of RNase I concentration used (Figure 5.7). Despite the testing of a sizable number of possible causal factors for this change, the root cause was not discovered (Figure 5.8, Figure 5.9, Figure 5.10, Figure 5.11). Further investigations into the cause of, and solutions for this observed digestion issue are needed. As the purification of RPFs is a critical step in the ribosome profiling procedure, it is essential that this digestion issue is resolved prior to any future ribosome profiling sequencing projects. One variable that could not be investigated due to insufficient stock was the possibility of significant batch to batch variation of the RNase I between what was used for the original titration, and what was used for the subsequent ribosome profiling and troubleshooting.

In lieu of an explanation for the poor digestion observed, the use of micrococcal nuclease (MN) was investigated as an alternative to RNase I. The use of MN resulted in significantly improved digestion as measured by polysome profiling, with a 80s monosome peak roughly 3 times larger than that previously seen using RNase I (Figure 5.12A). An additional benefit of using MN seemed to be a far greater working range in which it optimally digests the polysomes into 80s

monosomes without either significant over-digestion, or the presence of high levels of disomes, trisomes and higher polysomes (Figure 5.12A). Historically, the profiling of bacterial systems have been largely reliant on the use of micrococcal nuclease, though many ribosome profiling studies in eukaryotes favour RNase I due to its lack of strong sequence specificity and robust footprinting (Ingolia et al., 2012). However, recent studies using cultured *Drosophila* and human cells have investigated and demonstrated the viability of the use of a number of ribonucleases, including micrococcal nuclease, for ribosome profiling (Gerashchenko & Gladyshev, 2017; Miettinen & Bjorklund, 2015).

In conclusion, the work carried out thesis has identified a number of key pathways shown to be involved in the regulation of gene expression following UV-B induced DNA damage, and outlined the development of the ribosome profiling method, in particular demonstrating the efficacy of rRNA depletion using the Ribo-Zero Gold kit (illumina) as well as highlighting several potential pitfalls. Although a number of technical issues remain left to be solved before future ribosome profiling projects can be conducted, the identification of specific 5' tRNA halves in the RPF dataset accounting for the majority of contaminating reads, and their link with roles in translational control as well as their interactions with ribosomal subunits are fascinating and warrant further investigation in their own right (Bąkowska-Żywicka et al., 2016; Gebetsberger et al., 2012; Sobala & Hutvagner, 2013).

## 7. Appendix

### 7.1 Adapter sequences

#### 7.1.1 TruSeq Small RNA library preparation adapter sequences

RNA 5' Adapter (RA5)

5' GUUCAGAGUUCUACAGUCCGACGAUC

RNA 3' Adapter (RA3)

5' TGGAAATCTCGGGTGCCAAGG

Stop Oligo (STP)

5' GAAUCCACCACGUUCCCGUGG

RNA RT Primer (RTP)

5' GCCTTGGCACCCGAGAATTCCA

RNA PCR Primer (RP1)

5' AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCGA

Index sequence are in bold and underlined.

RNA PCR Primer, Index 1 (RPI1)

5'  
CAAGCAGAAGACGGCATAACGAGAT**CGTGAT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 2 (RPI2)

5'  
CAAGCAGAAGACGGCATAACGAGAT**ACATCG**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 3 (RPI3)

5'  
CAAGCAGAAGACGGCATAACGAGAT**GCCTAA**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 4 (RPI4)

5'  
CAAGCAGAAGACGGCATAACGAGAT**TGGTCA**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 5 (RPI5)

5'  
CAAGCAGAAGACGGCATAACGAGAT**CACTGT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 6 (RPI6)

5'  
CAAGCAGAAGACGGCATAACGAGAT**ATTGGC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 7 (RPI7)

5'  
CAAGCAGAAGACGGCATAACGAGAT**GATCTG**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 8 (RPI8)

5'  
CAAGCAGAAGACGGCATAACGAGAT**TCAAGT**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 9 (RPI9)

5'  
CAAGCAGAAGACGGCATAACGAGAT**CTGATC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 10 (RPI10)

5'  
CAAGCAGAAGACGGCATAACGAGAT**AAGCTA**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 11 (RPI11)

5'  
CAAGCAGAAGACGGCATAACGAGAT**GTAGCC**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 12 (RPI12)

5'  
CAAGCAGAAGACGGCATAACGAGAT**TACAAG**GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

### ***7.1.2 TruSeq Stranded mRNA LT library preparation adapter sequences***

TruSeq Universal Adapter

5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

Index sequence are in bold and underlined.

TruSeq Adapter, Index 2

5'  
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 4

5'  
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTGACCAATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 5

5'  
GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 6

5'  
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 7

5'  
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCAGATCATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 12

5'  
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTTGTAATCTCGTATGCCGTCTTCTGCTTG

TruSeq Adapter, Index 13

5'  
GATCGGAAGAGCACACGTCTGAACTCCAGTCACAGTCAACAATCTCGTATGCCGTCTTCTGCTT  
G

TruSeq Adapter, Index 14

5'  
GATCGGAAGAGCACACGTCTGAACTCCAGTCACAGTTCCGTATCTCGTATGCCGTCTTCTGCTT  
G

TruSeq Adapter, Index 15

5'  
GATCGGAAGAGCACACGTCTGAACTCCAGTCACATGTCAGAATCTCGTATGCCGTCTTCTGCTT  
G

TruSeq Adapter, Index 16

5'  
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCCGTCCCGATCTCGTATGCCGTCTTCTGCTT  
G

TruSeq Adapter, Index 18

5'  
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTCCGCACATCTCGTATGCCGTCTTCTGCTT  
G

TruSeq Adapter, Index 19

5'  
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTGAAACGATCTCGTATGCCGTCTTCTGCTT  
G

## 7.2 Scripts used for bioinformatics analysis

### 7.2.1 Removal of contaminating reads using Bowtie2

Removal of contaminating reads was performed using the custom shell script shown below. The example script was for the processing of RPF samples, code for the processing of the whole transcriptome samples differed only in the setting of the sequencing reads location, and the building of the bowtie2 index for tRNA and rRNA which was done in one command instead of two (as described in the methods).

```
#!/bin/bash

#####
# Remove Contaminants for RPFs

#Input files from step 2b) Cut using complete index primers
#Filter out reads smaller than 2nt (bowtie limit, any reads smaller
#than that are ignored)
#Create bowtie indices using the abundant contaminants fa files from
iGenome and the tRNA database (http://gtrnadb.ucsc.edu)
#Align and remove all contaminants using bowtie in 2 steps, first
#removing tRNA reads then rRNA reads.
#
#####

# Add bowtie2 to PATH
export PATH=~/.Downloads/bowtie2-2.2.6:$PATH
export PATH=~/.Downloads/ngsutils-0.5.7/bin:$PATH

# Set sequencing reads location
cd "/mnt/databank_doc/MRC_Tox_Unit/2014Q4 Ribosome Profiling
Data/Combined/RPF/3 Remove Contaminants"

# Name/address of the reference sequences used to build the bowtie
index
tRNAs_fa="hg38-tRNAs.fa"
hg38humrib="/mnt/databank_doc/MRC_Tox_Unit/Homo_sapiens/UCSC/hg38/Sequ
ence/AbundantSequences/humRibosomal.fa"
hg385s="/mnt/databank_doc/MRC_Tox_Unit/Homo_sapiens/UCSC/hg38/Sequence
/AbundantSequences/hum5SrDNA.fa"

# Filter short reads
```

```

mkdir cut_filtered
for d in *.fastq.gz; do
    echo "Processing $d"

    # Regex to shorten sample name
    out=$(echo $d | sed -E "s/([A-Z]+.*)_cut_long.fastq.gz/\1/")

    # Generate output file names
    out_final="cut_filtered/"$out"_cut_filtered.fastq.gz"

    # Filter out reads smaller than 2nt (bowtie limit)
    fastqutils filter -illumina -size 2 $d | gzip -c > $out_final

    echo "Processing of $out complete"

done

cd cut_filtered

# Create bowtie indicies
mkdir tRNAs_bt2
# Build tRNA bowtie index
bowtie2-build $tRNAs_fa tRNAs_bt2/tRNAs
# Build rRNA bowtie index
mkdir rRNAs_bt2
bowtie2-build $hg38humrib,$hg385s rRNAs_bt2/humRib

mkdir tRNA_rRNA_removed
# Align and remove contaminants
for d in *.fastq.gz; do
    echo "Processing $d"

    # Regex to shorten sample name
    out=$(echo $d | sed -E "s/([A-Z]+.*)_cut_filtered.fastq.gz/\1/")

    # Generate output file names
    no_rRNA="tRNA_rRNA_removed/1_"$out"_no_rRNA.fastq.gz"
    rRNA="tRNA_rRNA_removed/1_"$out"_rRNA.sam.gz"
    no_rRNA_tRNA="tRNA_rRNA_removed/2_"$out"_no_rRNA_tRNA.fastq.gz"

no_rRNA_yes_tRNA="tRNA_rRNA_removed/2_"$out"_no_rRNA_yes_tRNA.sam.gz"

    # Filter for rRNA
    bowtie2 -p 8 --un-gz $no_rRNA -x rRNAs_bt2/humRib -U <(zcat $d) |
gzip -c > $rRNA
    echo "Processing for $out rRNA complete"

    # Filter for tRNA
    bowtie2 -p 8 --un-gz $no_rRNA_tRNA -x tRNAs_bt2/tRNAs -U <(zcat
$no_rRNA) | gzip -c > $no_rRNA_yes_tRNA
    echo "Processing for $out tRNA complete"

    echo "Processing of $out complete"

done

```

## 7.2.2 Alignment to reference genome using TopHat

Reads alignment to the UCSC hg38 reference genome was performed using TopHat and a custom shell script shown below. Code used for the RPF and whole transcriptome samples different only in the setting of their respective storage locations.

```
#!/bin/bash

#####
# Tophat Alignment
#
# Alignment and with tophat, using input from tRNA and rRNA removed
# fastq files.
#
#####

# Add required programs to PATH
export PATH=~/.Downloads/bowtie2-2.2.6:$PATH"
export PATH=~/.Downloads/tophat:$PATH"
export PATH=~/.Downloads/samtools-1.3:$PATH"

# Set sequencing reads location
cd "/mnt/databank_doc/MRC_Tox_Unit/2014Q4 Ribosome Profiling
Data/Combined/RPF/4 TopHat"

# Name/address of reference files
genes="/home/julian/Documents/Reference_files/hg38/Genes/genes.gtf"
bt2idx="/home/julian/Documents/Reference_files/hg38/Bowtie2Index/genom
e"

for d in *.fastq.gz; do
    echo "Processing $d"

    # Regex to shorten sample name
    ident=$(echo $d | sed -E "s/^2_([A-
Z]+.*)_no_tRNA_rRNA.fastq.gz/\1/")

    # Generate output file names
    th_out=$ident"_tophat_out"
    mkdir $th_out

    #----- Sequence alignment with TopHat
    tophat --GTF $genes -p 8 -g 1 -o $th_out $bt2idx $d

    echo "Processing of $ident complete"
done
```

### 7.2.3 Counting reads

Aligned sequencing reads mapping to known genomic features were counted with HTSeq-count using the gene annotation file from the UCSC hg38 dataset and a custom shell script shown below. Code used for the RPF and whole transcriptome samples different only in the setting of their respective storage locations.

```
#!/bin/bash

#####
# HTSeq_count bash script
#
# Counting reads using HTSeq_count
#
#####

# Set sequencing reads location
cd "/mnt/databank_doc/MRC_Tox_Unit/2015Q4 Ribosome Profiling
Data/Combined/RPF/4 Tophat"

# Name/address of reference files
genes="/mnt/databank_doc/MRC_Tox_Unit/Homo_sapiens/UCSC/hg38/Annotation/Genes/genes.gtf"

# Rep 1
python -m HTSeq.scripts.count -s no -f bam
R1_C1_tophat_out/accepted_hits.bam $genes > R1_C1_gc.txt
python -m HTSeq.scripts.count -s no -f bam
R1_UV1_tophat_out/accepted_hits.bam $genes > R1_UV1_gc.txt
python -m HTSeq.scripts.count -s no -f bam
R1_C4_tophat_out/accepted_hits.bam $genes > R1_C4_gc.txt
python -m HTSeq.scripts.count -s no -f bam
R1_UV4_tophat_out/accepted_hits.bam $genes > R1_UV4_gc.txt
echo "HTSeq-count for Rep 1 complete"

# Rep 3
python -m HTSeq.scripts.count -s no -f bam
R3_C1_tophat_out/accepted_hits.bam $genes > R3_C1_gc.txt
python -m HTSeq.scripts.count -s no -f bam
R3_UV1_tophat_out/accepted_hits.bam $genes > R3_UV1_gc.txt
python -m HTSeq.scripts.count -s no -f bam
R3_C4_tophat_out/accepted_hits.bam $genes > R3_C4_gc.txt
python -m HTSeq.scripts.count -s no -f bam
R3_UV4_tophat_out/accepted_hits.bam $genes > R3_UV4_gc.txt
echo "HTSeq-count for Rep 3 complete"

# Rep 4
python -m HTSeq.scripts.count -s no -f bam
R4_C1_tophat_out/accepted_hits.bam $genes > R4_C1_gc.txt
python -m HTSeq.scripts.count -s no -f bam
R4_UV1_tophat_out/accepted_hits.bam $genes > R4_UV1_gc.txt
python -m HTSeq.scripts.count -s no -f bam
R4_C4_tophat_out/accepted_hits.bam $genes > R4_C4_gc.txt
python -m HTSeq.scripts.count -s no -f bam
R4_UV4_tophat_out/accepted_hits.bam $genes > R4_UV4_gc.txt
echo "HTSeq-count for Rep 4 complete"
```

## 7.2.4 Translatome analysis using R package babel

Translatome analysis was conducted using the R package babel, as described in the methods.

Scripts used are shown below, with the subroutine script shown in Section 7.2.6.

```
#####
# Translatome Analysis using babel
# using gene counts calculated by HTSeq-count
#
#####

library(babel)

# Set project ID and working directory
proj_id <- "Babel"
setwd("/mnt/databank_doc/MRC_Tox_Unit/2015Q4 Ribosome Profiling
Data/Combined/R_Analysis/Babel/")

# Import subroutines
source("/mnt/databank_doc/MRC_Tox_Unit/2015Q4 Ribosome Profiling
Data/Combined/R_Analysis/JW_R_subroutine_functions.R")

# Import in data
rpfdircloc <- "/mnt/databank_doc/MRC_Tox_Unit/2015Q4 Ribosome Profiling
Data/Combined/RPF/4 Tophat/"
wtDirloc <- "/mnt/databank_doc/MRC_Tox_Unit/2015Q4 Ribosome Profiling
Data/Combined/WT/4c Tophat (fr-firststrand)/"
importReads(rpfdircloc, wtDirloc)
rm(rpfdircloc,wtDirloc)

#####
# Run babel
#####
options(mc.cores = 8)
set.seed(12345)
test.babel <- babel(wt, rpf,
                    group = test.group,
                    nreps = 1e+07,
                    min.rna = 10)

# use biomaRt to obtain gene name descriptions
library(biomaRt)
ensembl = useMart("ensembl",dataset="hsapiens_gene_ensembl")
# configure mRNA_FDR and FDR settings
fdr <- 0.1
mrna_fdr <- 0.1
mrna_logfc <- 1.5

# configure xlsx write parameters
library(xlsx)
xlsx_output <- paste(Sys.Date(), proj_id, "summary.xlsx")

#----- Counting numbers of unusual counts between conditions
between.babel <- test.babel$between

###----- 1hr time point
clvsuv1 <- between.babel[[2]]
```

```

# lhr Transcriptome - Obtain gene descriptions
print(paste("lhr Transcriptome:",
nrow(clvsuv1[complete.cases(clvsuv1) & clvsuv1$mRNA_FDR<mrna_fdr,]),
"DEGs detected in total"))
transcriptomelhr <- clvsuv1[complete.cases(clvsuv1) &
clvsuv1$mRNA_FDR<mrna_fdr,]
Gene_description <- getBM(attributes=
c('hgnc_symbol','description'), filters = 'hgnc_symbol', values =
transcriptomelhr$Gene, mart = ensembl)
matched <- match(transcriptomelhr$Gene,
Gene_description$hgnc_symbol)
transcriptomelhr <- cbind(transcriptomelhr,
Gene_description[matched,2])
rm(matched, Gene_description)

# lhr Transcriptome - Sort into up/down lists, sort,
and write to file
# Down list
transcriptomelhr_down <-
transcriptomelhr[transcriptomelhr$mRNA_logFC > 0,]
print(paste("lhr Transcriptome:",
nrow(transcriptomelhr_down), "DEGs down-regulated"))
transcriptomelhr_down <-
transcriptomelhr_down[order(transcriptomelhr_down$mRNA_FDR),]
write.xlsx2(transcriptomelhr_down, xlsx_output,
sheetName="transcriptomelhr_down", col.names=TRUE, row.names=FALSE,
append=FALSE)

# Up list
transcriptomelhr_up <-
transcriptomelhr[transcriptomelhr$mRNA_logFC < 0,]
print(paste("lhr Transcriptome:",
nrow(transcriptomelhr_up), "DEGs up-regulated"))
transcriptomelhr_up <-
transcriptomelhr_up[order(transcriptomelhr_up$mRNA_FDR),]
write.xlsx2(transcriptomelhr_up, xlsx_output,
sheetName="transcriptomelhr_up", col.names=TRUE, row.names=FALSE,
append=TRUE)

# lhr Translatome - Obtain gene descriptions
print(paste("lhr Translatome:",
nrow(clvsuv1[complete.cases(clvsuv1) & clvsuv1$FDR<fdr,]), "DEGs
detected in total"))
translatomelhr <- clvsuv1[complete.cases(clvsuv1) &
clvsuv1$FDR<fdr,]
Gene_description <- getBM(attributes=
c('hgnc_symbol','description'), filters = 'hgnc_symbol', values =
translatomelhr$Gene, mart = ensembl)
matched <- match(translatomelhr$Gene,
Gene_description$hgnc_symbol)
translatomelhr <- cbind(translatomelhr,
Gene_description[matched,2])
rm(matched, Gene_description)

# lhr Translatome - Sort into up/down lists, sort, and
write to file
# Down list

```

```

        translatomelhr_down <-
translatomelhr[translatomelhr$Direction == "1",]
        print(paste("1hr Translatome:",
nrow(translatomelhr_down), "DEGs down-regulated"))
        translatomelhr_down <-
translatomelhr_down[order(translatomelhr_down$FDR),]
        write.xlsx2(translatomelhr_down, xlsx_output,
sheetName="translatomelhr_down", col.names=TRUE, row.names=FALSE,
append=TRUE)

        # Up list
        translatomelhr_up <-
translatomelhr[translatomelhr$Direction == "-1",]
        print(paste("1hr Translatome:",
nrow(translatomelhr_up), "DEGs up-regulated"))
        translatomelhr_up <-
translatomelhr_up[order(translatomelhr_up$FDR),]
        write.xlsx2(translatomelhr_up, xlsx_output,
sheetName="translatomelhr_up", col.names=TRUE, row.names=FALSE,
append=TRUE)

###----- 4hr time point
c4vsuv4 <- between.babel[[5]]

        # 4hr Transcriptome - Obtain gene descriptions
        print(paste("4hr Transcriptome:",
nrow(c4vsuv4[complete.cases(c4vsuv4) & c4vsuv4$mRNA_FDR<mrna_fdr &
abs(c4vsuv4$mRNA_logFC)>mrna_logfc,]), "DEGs detected in total"))
        transcriptome4hr <- c4vsuv4[complete.cases(c4vsuv4) &
c4vsuv4$mRNA_FDR<mrna_fdr & abs(c4vsuv4$mRNA_logFC)>mrna_logfc,]
        Gene_description <- getBM(attributes=
c('hgnc_symbol','description'), filters = 'hgnc_symbol', values =
transcriptome4hr$Gene, mart = ensembl)
        matched <- match(transcriptome4hr$Gene,
Gene_description$hgnc_symbol)
        transcriptome4hr <- cbind(transcriptome4hr,
Gene_description[matched,2])
        rm(matched, Gene_description)

        # 4hr Transcriptome - Sort into up/down lists, sort,
and write to file
        # Down list
        transcriptome4hr_down <-
transcriptome4hr[transcriptome4hr$mRNA_logFC > 0,]
        print(paste("4hr Transcriptome:",
nrow(transcriptome4hr_down), "DEGs down-regulated"))
        transcriptome4hr_down <-
transcriptome4hr_down[order(transcriptome4hr_down$mRNA_FDR),]
        write.xlsx2(transcriptome4hr_down, xlsx_output,
sheetName="transcriptome4hr_down", col.names=TRUE, row.names=FALSE,
append=TRUE)

        # Up list
        transcriptome4hr_up <-
transcriptome4hr[transcriptome4hr$mRNA_logFC < 0,]
        print(paste("4hr Transcriptome:",
nrow(transcriptome4hr_up), "DEGs up-regulated"))
        transcriptome4hr_up <-
transcriptome4hr_up[order(transcriptome4hr_up$mRNA_FDR),]

```

```

write.xlsx2(transcriptome4hr_up, xlsx_output,
sheetName="transcriptome4hr_up", col.names=TRUE, row.names=FALSE,
append=TRUE)

# 4hr Translatome - Obtain gene descriptions
print(paste("4hr Translatome:",
nrow(c4vsuv4[complete.cases(c4vsuv4) & c4vsuv4$FDR<fdr,]), "DEGs
detected in total"))
translatome4hr <- c4vsuv4[complete.cases(c4vsuv4) &
c4vsuv4$FDR<fdr,]
Gene_description <- getBM(attributes=
c('hgnc_symbol','description'), filters = 'hgnc_symbol', values =
translatome4hr$Gene, mart = ensembl)
matched <- match(translatome4hr$Gene,
Gene_description$hgnc_symbol)
translatome4hr <- cbind(translatome4hr,
Gene_description[matched,2])
rm(matched, Gene_description)

# 4hr Translatome - Sort into up/down lists, sort, and
write to file
# Down list
translatome4hr_down <-
translatome4hr[translatome4hr$Direction == "1",]
print(paste("4hr Translatome:",
nrow(translatome4hr_down), "DEGs down-regulated"))
translatome4hr_down <-
translatome4hr_down[order(translatome4hr_down$FDR),]
write.xlsx2(translatome4hr_down, xlsx_output,
sheetName="translatome4hr_down", col.names=TRUE, row.names=FALSE,
append=TRUE)
# Up list
translatome4hr_up <-
translatome4hr[translatome4hr$Direction == "-1",]
print(paste("4hr Translatome:",
nrow(translatome4hr_up), "DEGs up-regulated"))
translatome4hr_up <-
translatome4hr_up[order(translatome4hr_up$FDR),]
write.xlsx2(translatome4hr_up, xlsx_output,
sheetName="translatome4hr_up", col.names=TRUE, row.names=FALSE,
append=TRUE)
rm(xlsx_output)

# Annotate generated up/down gene lists
# Transcriptome
annotateExisting(transcriptome1hr_up, output.name="Annotated
DEGs", sheet.name="Transcriptome1U")
annotateExisting(transcriptome1hr_down, output.name="Annotated
DEGs", sheet.name="Transcriptome1D", append=TRUE)
annotateExisting(transcriptome4hr_up, output.name="Annotated
DEGs", sheet.name="Transcriptome4U", append=TRUE)
annotateExisting(transcriptome4hr_down, output.name="Annotated
DEGs", sheet.name="Transcriptome4D", append=TRUE)

# Translatome
annotateExisting(translatome1hr_up, output.name="Annotated
DEGs", sheet.name="Translatome1U", append=TRUE)

```

```

        annotateExisting(translatome1hr_down, output.name="Annotated
DEGs", sheet.name="Translatome1D", append=TRUE)
        annotateExisting(translatome4hr_up, output.name="Annotated
DEGs", sheet.name="Translatome4U", append=TRUE)
        annotateExisting(translatome4hr_down, output.name="Annotated
DEGs", sheet.name="Translatome4D", append=TRUE)

```

```

runstats <- paste("Run on", Sys.Date(), "using RPF dataset. Ran using
babel with 1e+07 permutations
Gene counting done using HTSeq-count
RPF: used --stranded=no
WT: used --stranded=reverse")

```

```

#####
# Plotting heatmaps of significant genes in translatome
#####

```

```

library(d3heatmap)
library(gplots)
library(htmlwidgets)

```

```

###----- 4hr time point
trans4sorted <- translatome4hr[order(translatome4hr$`P-value`),]
trans4SH <- as.data.frame(trans4sorted[,2])
rownames(trans4SH) <- trans4sorted$Gene
colnames(trans4SH) <- "4hr_log2FC"

```

```

# Retrieve matched genes from 1hr time point
trans4sig_in_res1 <- match(row.names(trans4SH), clvsuv1$Gene)
trans4SH[,2] <- clvsuv1[trans4sig_in_res1,2]
colnames(trans4SH) <- c("4hr_log2FC", "1hr_log2FC")

```

```

# Reorder the columns so 1hr time point is first
trans4SH <- trans4SH[c(2,1)]

```

```

# Select top 20 genes
trans4SH <- trans4SH[1:20,]

```

```

d3heatmap(data.matrix(trans4SH), scale = "none",
dendrogram='none', col=redgreen(75))

```

```

# Saving to HTML file
hm <- d3heatmap(data.matrix(trans4SH), scale = "none",
dendrogram='none', col=redgreen(75))
saveWidget(hm, paste0(Sys.Date(), " Heatmap Output (v2) - 4hr",
".html"))

```

```

###----- 1hr time point
trans1sorted <- translatome1hr[order(translatome1hr$`P-value`),]
trans1SH <- as.data.frame(trans1sorted[,2])
rownames(trans1SH) <- trans1sorted$Gene
colnames(trans1SH) <- "1hr_log2FC"

```

```

# Retrieve matched genes from 1hr time point
trans1sig_in_res4 <- match(row.names(trans1SH), c4vsuv4$Gene)
trans1SH[,2] <- c4vsuv4[trans1sig_in_res4,2]

```

```
colnames(trans1SH) <- c("1hr_log2FC", "4hr_log2FC")

# Select top 30
trans1SH <- trans1SH[1:20,]

d3heatmap(data.matrix(trans1SH), scale = "none",
dendrogram='none', col=redgreen(75))

# Saving to HTML file
hm <- d3heatmap(data.matrix(trans1SH), scale = "none",
dendrogram='none', col=redgreen(75))
saveWidget(hm, paste0(Sys.Date(), " Heatmap Output (v2) - 1hr",
".html"))

# Save workspace
save.image(file = paste(Sys.Date(), proj_id, "analysis output (RPF -s
no, WT -s reverse).RData"))
```

## 7.2.5 Transcriptome analysis using R package DESeq2

Transcriptome analysis was conducted using the R package DESeq2, as described in the methods.

Scripts used are shown below, with the subroutine script shown in Section 7.2.6.

```
#####
# Transcriptome analysis using DESeq2
#
#####

# Set project ID and working directory
proj_id <- "DESeq2 Transcriptome Analysis"
setwd("/mnt/databank_doc/MRC_ToX_Unit/2015Q4 Ribosome Profiling
Data/Combined/R_Analysis/DESeq2/")

# Import subroutines
source("/mnt/databank_doc/MRC_ToX_Unit/2015Q4 Ribosome Profiling
Data/Combined/R_Analysis/JW_R_subroutine_functions.R")

library("DESeq2")
library("BiocParallel")
register(MulticoreParam(8))

# Import in data & split into 1 & 4hr data frames
rpfdirloc <- "/mnt/databank_doc/MRC_ToX_Unit/2015Q4 Ribosome Profiling
Data/Combined/RPF/4b Tophat (tRNA not removed)/"
wtmdirloc <- "/mnt/databank_doc/MRC_ToX_Unit/2015Q4 Ribosome Profiling
Data/Combined/WT/4c Tophat (fr-firststrand)/"
importReads(rpfdirloc, wtmdirloc)
rm(rpfdirloc,wtmdirloc)
  rpf1 <- rpf[,c(1,3,5,7,9,11)]
  rpf4 <- rpf[,c(2,4,6,8,10,12)]
  wt1 <- wt[,c(1,3,5,7,9,11)]
  wt4 <- wt[,c(2,4,6,8,10,12)]

# Set colData for wt's
colData1 <-
matrix(c(rep(c("control","uv"),3),c(rep("R1",2),rep("R3",2),rep("R4",2)
)), nrow=6, ncol=2)
row.names(colData1) <- colnames(wt1)
colnames(colData1) <- c("condition","repeat")

colData4 <- colData1
row.names(colData4) <- colnames(wt4)

dds1 <- DESeqDataSetFromMatrix(countData = wt1,
                              colData = as.data.frame(colData1),
                              design = ~ condition)

dds4 <- DESeqDataSetFromMatrix(countData = wt4,
                              colData = as.data.frame(colData4),
                              design = ~ condition)

# Pre-filtering reads
dds1 <- dds1[ rowSums(counts(dds1)) > 1, ]
dds4 <- dds4[ rowSums(counts(dds4)) > 1, ]

# Run & order by p-adj
```

```

dds1 <- DESeq(dds1)
res1 <- results(dds1, alpha=0.05)
res1Ordered <- res1[order(res1$padj),]
summary(res1)

dds4 <- DESeq(dds4)
res4 <- results(dds4, alpha=0.05)
res4Ordered <- res4[order(res4$padj),]
summary(res4)

# Check numbers of DEGs given p-value
padj <- 0.05
sum(res1$padj < padj, na.rm=TRUE)
sum(res4$padj < padj, na.rm=TRUE)

# MA Plot
plotMA(res1, alpha=0.05, main="DESeq2", ylim=c(-2,2))
plotMA(res4, alpha=0.05, main="DESeq2", ylim=c(-2,2))

# PCA plot
rld1 <- rlog(dds1)
rld4 <- rlog(dds4)
plotPCA(rld1, intgroup=c("condition","repeat."))
plotPCA(rld4, intgroup=c("condition","repeat."))

# Dispersion plots
plotDispEsts(dds1, main="Dispersion of 1hr counts")
plotDispEsts(dds4, main="Dispersion of 4hr counts")

#####
# Write DESeq2 output to Excel
#####

library(xlsx)

res1sig <- na.omit(as.data.frame(res1)[res1$padj < 0.05,])
res1sig$gene <- row.names(res1sig)
res1up <- res1sig[res1sig$log2FoldChange > 0,]
res1down <- res1sig[res1sig$log2FoldChange < 0,]

res4sig <- na.omit(as.data.frame(res4)[res4$padj < 0.05,])
res4sig$gene <- row.names(res4sig)
res4up <- res4sig[res4sig$log2FoldChange > 0,]
res4down <- res4sig[res4sig$log2FoldChange < 0,]

# Write lists to file
write.xlsx(res1up[order(res1up$log2FoldChange, decreasing = TRUE),],
file=paste(Sys.Date(), "DESeq2 Sig Output List.xlsx"), sheetName =
paste0("1U (", nrow(res1up), " total)"), append=FALSE)
write.xlsx(res1down[order(res1down$log2FoldChange, decreasing =
TRUE),], file=paste(Sys.Date(), "DESeq2 Sig Output List.xlsx"),
sheetName = paste0("1D (", nrow(res1down), " total)"), append=TRUE)
write.xlsx(res4up[order(res4up$log2FoldChange, decreasing = TRUE),],
file=paste(Sys.Date(), "DESeq2 Sig Output List.xlsx"), sheetName =
paste0("4U (", nrow(res4up), " total)"), append=TRUE)
write.xlsx(res4down[order(res4down$log2FoldChange, decreasing =
TRUE),], file=paste(Sys.Date(), "DESeq2 Sig Output List.xlsx"),
sheetName = paste0("4D (", nrow(res4down), " total)"), append=TRUE)

```

```

#####
# Plot heatmaps of the most significant DEGs
#####

library(d3heatmap)
library(gplots)
library(htmlwidgets)

#----- 1hr time point
# Sort by significance (padj)
res1sigsorted <- res1sig[order(res1sig[,6]),]
res1sigSH <- as.data.frame(res1sigsorted[,2])
rownames(res1sigSH) <- rownames(res1sigsorted)
colnames(res1sigSH) <- "1hr_log2FC"

# Retrieve matched genes from 4hr time point
res1sig_in_res4 <- match(row.names(res1sigSH), row.names(res4))
res1sigSH[,2] <- res4[res1sig_in_res4,2]
colnames(res1sigSH) <- c("1hr_log2FC", "4hr_log2FC")

# Select top 30
res1sigSH <- res1sigSH[1:30,]

d3heatmap(data.matrix(res1sigSH), scale = "none",
dendrogram='none', col=redgreen(75))

# Saving to HTML file
hm <- d3heatmap(data.matrix(res1sigSH), scale = "none",
dendrogram='none', col=redgreen(75))
saveWidget(hm, paste0(Sys.Date(), " Heatmap Output (v2) - 1hr",
".html"))

#----- 4hr time point
# Sort by significance (padj)
res4sigsorted <- res4sig[order(res4sig[,6]),]
res4sigSH <- as.data.frame(res4sigsorted[,2])
rownames(res4sigSH) <- rownames(res4sigsorted)
colnames(res4sigSH) <- "4hr_log2FC"

# Retrieve matched genes from 1hr time point
res4sig_in_res1 <- match(row.names(res4sigSH), row.names(res1))
res4sigSH[,2] <- res1[res4sig_in_res1,2]
colnames(res4sigSH) <- c("4hr_log2FC", "1hr_log2FC")

# Select top 30
res4sigSH <- res4sigSH[1:30,]

# Re-order columns to place 1hr TP first
res4sigSH <- res4sigSH[c(2,1)]

d3heatmap(data.matrix(res4sigSH), scale = "none",
dendrogram='none', col=redgreen(75))

# Saving to HTML file
hm <- d3heatmap(data.matrix(res4sigSH), scale = "none",
dendrogram='none', col=redgreen(75))
saveWidget(hm, paste0(Sys.Date(), " Heatmap Output (v2) - 4hr",
".html"))
# Save workspace
save.image(file = paste(Sys.Date(), proj_id, "output.RData"))

```

## 7.2.6 Custom R functions used

The custom subroutine functions used in R based analyses are shown below.

```
#####
# Subroutine functions for R based analysis
#
#####

importReads <- function(rpfdirloc, wtdirloc, suffix="_gc.txt"){
  #----- Read in RPFs
  # Read in first sample as expression("log"[2]*"(RPF counts)")
  and add subsequent samples in as additional columns to the data frame
  rpf <- read.table(paste0(rpfdirloc,"R1_C1",suffix),
row.names=1, col.names=c("gene","R1_C1"))
  rpf_r1_c4 <- read.table(paste0(rpfdirloc,"R1_C4",suffix))
  rpf_r1_uv1 <- read.table(paste0(rpfdirloc,"R1_UV1",suffix))
  rpf_r1_uv4 <- read.table(paste0(rpfdirloc,"R1_UV4",suffix))

  rpf_r3_c1 <- read.table(paste0(rpfdirloc,"R3_C1",suffix))
  rpf_r3_c4 <- read.table(paste0(rpfdirloc,"R3_C4",suffix))
  rpf_r3_uv1 <- read.table(paste0(rpfdirloc,"R3_UV1",suffix))
  rpf_r3_uv4 <- read.table(paste0(rpfdirloc,"R3_UV4",suffix))

  rpf_r4_c1 <- read.table(paste0(rpfdirloc,"R4_C1",suffix))
  rpf_r4_c4 <- read.table(paste0(rpfdirloc,"R4_C4",suffix))
  rpf_r4_uv1 <- read.table(paste0(rpfdirloc,"R4_UV1",suffix))
  rpf_r4_uv4 <- read.table(paste0(rpfdirloc,"R4_UV4",suffix))

  # Merge all data frames into one
  rpf$R1_C4 <- rpf_r1_c4[,2]
  rpf$R1_UV1 <- rpf_r1_uv1[,2]
  rpf$R1_UV4 <- rpf_r1_uv4[,2]

  rpf$R3_C1 <- rpf_r3_c1[,2]
  rpf$R3_C4 <- rpf_r3_c4[,2]
  rpf$R3_UV1 <- rpf_r3_uv1[,2]
  rpf$R3_UV4 <- rpf_r3_uv4[,2]

  rpf$R4_C1 <- rpf_r4_c1[,2]
  rpf$R4_C4 <- rpf_r4_c4[,2]
  rpf$R4_UV1 <- rpf_r4_uv1[,2]
  rpf$R4_UV4 <- rpf_r4_uv4[,2]

  # Remove last 5 rows (these contain HTSeq-count stats)
  rpf <<- head(rpf, -5)

  # Remove intermediate variables
  rm(list=ls(pattern="rpf_"))

  #----- Read in WTs
  # Read in first sample as expression("log"[2]*"(WT counts)")
  and add subsequent samples in as additional columns to the data frame
  wt <- read.table(paste0(wtdirloc,"WT_R1_C1",suffix),
row.names=1, col.names=c("gene","R1_C1"))
  wt_r1_c4 <- read.table(paste0(wtdirloc,"WT_R1_C4",suffix))
  wt_r1_uv1 <- read.table(paste0(wtdirloc,"WT_R1_UV1",suffix))
  wt_r1_uv4 <- read.table(paste0(wtdirloc,"WT_R1_UV4",suffix))
}
```

```

wt_r3_c1 <- read.table(paste0(wtdirloc,"WT_R3_C1",suffix))
wt_r3_c4 <- read.table(paste0(wtdirloc,"WT_R3_C4",suffix))
wt_r3_uv1 <- read.table(paste0(wtdirloc,"WT_R3_UV1",suffix))
wt_r3_uv4 <- read.table(paste0(wtdirloc,"WT_R3_UV4",suffix))

wt_r4_c1 <- read.table(paste0(wtdirloc,"WT_R4_C1",suffix))
wt_r4_c4 <- read.table(paste0(wtdirloc,"WT_R4_C4",suffix))
wt_r4_uv1 <- read.table(paste0(wtdirloc,"WT_R4_UV1",suffix))
wt_r4_uv4 <- read.table(paste0(wtdirloc,"WT_R4_UV4",suffix))

# Merge all data frames into one
wt$R1_C4 <- wt_r1_c4[,2]
wt$R1_UV1 <- wt_r1_uv1[,2]
wt$R1_UV4 <- wt_r1_uv4[,2]

wt$R3_C1 <- wt_r3_c1[,2]
wt$R3_C4 <- wt_r3_c4[,2]
wt$R3_UV1 <- wt_r3_uv1[,2]
wt$R3_UV4 <- wt_r3_uv4[,2]

wt$R4_C1 <- wt_r4_c1[,2]
wt$R4_C4 <- wt_r4_c4[,2]
wt$R4_UV1 <- wt_r4_uv1[,2]
wt$R4_UV4 <- wt_r4_uv4[,2]

# Remove last 5 rows (these contain HTSeq-count stats)
wt <<- head(wt, -5)

# Remove intermediate variables
rm(list=ls(pattern="wt_"))
rm(rpfdirloc,wtdirloc,suffix)

# Group by condition
test.group <<-
c("C1","C4","UV1","UV4","C1","C4","UV1","UV4","C1","C4","UV1","UV4")
}

outputGeneList <- function(genes, data, output.name="Output",
sheet.name="Sheet", col.names=TRUE, row.names=FALSE, append=FALSE){
  library(xlsx)
  library(biomaRt)

  if (output.name=="Output" | sheet.name=="Sheet"){
    cat("WARNING: output.name and/or sheet.name not
specified.\nUsing default value(s) instead.\n\n")
  }

  xlsx_output <- paste0(Sys.Date(), " ",output.name, ".xlsx")
  ensembl = useMart("ensembl",dataset="hsapiens_gene_ensembl")

  genelist <- na.omit(data[genes,])
  Gene_description <- getBM(attributes=
c('hgnc_symbol','description'), filters = 'hgnc_symbol', values =
genelist$Gene, mart = ensembl)
  matched <- match(genelist$Gene, Gene_description$hgnc_symbol)
  genelist <- cbind(genelist, Gene_description[matched,2])
  rm(matched, Gene_description)
  write.xlsx2(genelist, xlsx_output, sheetName=sheet.name,
col.names=col.names, row.names=row.names, append=append)

```

```

        cat("Matched", nrow(genelist), "Genes out
of", length(genes), "from input list.\n",
        "Results successfully written to", xlsx_output)
}

plotloglog <- function(wt, rpf, pch=16, cex=0.4){
  # Plot log counts of all samples
  plot.new()
  par(mfrow=c(3, 4))
  plot(log2(wt[, 1]), log2(rpf[, 1]),
xlab=expression("log"[2]*"(WT counts)"),
ylab=expression("log"[2]*"(RPF counts)"), main="R1 C1",
pch=pch, cex=cex)
  plot(log2(wt[, 3]), log2(rpf[, 3]),
xlab=expression("log"[2]*"(WT counts)"),
ylab=expression("log"[2]*"(RPF counts)"), main="R1 UV1",
pch=pch, cex=cex)
  plot(log2(wt[, 2]), log2(rpf[, 2]),
xlab=expression("log"[2]*"(WT counts)"),
ylab=expression("log"[2]*"(RPF counts)"), main="R1 C4",
pch=pch, cex=cex)
  plot(log2(wt[, 4]), log2(rpf[, 4]),
xlab=expression("log"[2]*"(WT counts)"),
ylab=expression("log"[2]*"(RPF counts)"), main="R1 UV4",
pch=pch, cex=cex)

  plot(log2(wt[, 5]), log2(rpf[, 5]),
xlab=expression("log"[2]*"(WT counts)"),
ylab=expression("log"[2]*"(RPF counts)"), main="R3 C1",
pch=pch, cex=cex)
  plot(log2(wt[, 7]), log2(rpf[, 7]),
xlab=expression("log"[2]*"(WT counts)"),
ylab=expression("log"[2]*"(RPF counts)"), main="R3 UV1",
pch=pch, cex=cex)
  plot(log2(wt[, 6]), log2(rpf[, 6]),
xlab=expression("log"[2]*"(WT counts)"),
ylab=expression("log"[2]*"(RPF counts)"), main="R3 C4",
pch=pch, cex=cex)
  plot(log2(wt[, 8]), log2(rpf[, 8]),
xlab=expression("log"[2]*"(WT counts)"),
ylab=expression("log"[2]*"(RPF counts)"), main="R3 UV4",
pch=pch, cex=cex)

  plot(log2(wt[, 9]), log2(rpf[, 9]),
xlab=expression("log"[2]*"(WT counts)"),
ylab=expression("log"[2]*"(RPF counts)"), main="R4 C1",
pch=pch, cex=cex)
  plot(log2(wt[, 11]), log2(rpf[, 11]),
xlab=expression("log"[2]*"(WT counts)"),
ylab=expression("log"[2]*"(RPF counts)"), main="R4 UV1",
pch=pch, cex=cex)
  plot(log2(wt[, 10]), log2(rpf[, 10]),
xlab=expression("log"[2]*"(WT counts)"),
ylab=expression("log"[2]*"(RPF counts)"), main="R4 C4",
pch=pch, cex=cex)
  plot(log2(wt[, 12]), log2(rpf[, 12]),
xlab=expression("log"[2]*"(WT counts)"),
ylab=expression("log"[2]*"(RPF counts)"), main="R4 UV4",
pch=pch, cex=cex)
}

```

```

}

plot3dvar <- function(wt, rpf){
  #----- Checking inter-repeat variability
  library(scatterplot3d)

  plot.new()
  par(mfrow=c(2,4))
  # RPF C1
  scatterplot3d(log2(rpf[,1]), # x axis
                log2(rpf[,2]), # y axis
                log2(rpf[,3]), # z axis
                pch=".",
                main="RPF C1",
                xlab=expression("log"[2]*"(RPF C1 Repeat 1)"),
                ylab=expression("log"[2]*"(RPF C1 Repeat 2)"),
                zlab=expression("log"[2]*"(RPF C1 Repeat 3)"))
  )
  # RPF UV1
  scatterplot3d(log2(rpf[,7]), # x axis
                log2(rpf[,8]), # y axis
                log2(rpf[,9]), # z axis
                pch=".",
                main="RPF UV1",
                xlab=expression("log"[2]*"(RPF UV1 Repeat 1)"),
                ylab=expression("log"[2]*"(RPF UV1 Repeat 2)"),
                zlab=expression("log"[2]*"(RPF UV1 Repeat 3)"))
  )

  # RPF C4
  scatterplot3d(log2(rpf[,4]), # x axis
                log2(rpf[,5]), # y axis
                log2(rpf[,6]), # z axis
                pch=".",
                main="RPF C4",
                xlab=expression("log"[2]*"(RPF C4 Repeat 1)"),
                ylab=expression("log"[2]*"(RPF C4 Repeat 2)"),
                zlab=expression("log"[2]*"(RPF C4 Repeat 3)"))
  )

  # RPF UV4
  scatterplot3d(log2(rpf[,10]), # x axis
                log2(rpf[,11]), # y axis
                log2(rpf[,12]), # z axis
                pch=".",
                main="RPF UV4",
                xlab=expression("log"[2]*"(RPF UV4 Repeat 1)"),
                ylab=expression("log"[2]*"(RPF UV4 Repeat 2)"),
                zlab=expression("log"[2]*"(RPF UV4 Repeat 3)"))
  )

  ##### WT
  # WT C1
  scatterplot3d(log2(wt[,1]), # x axis
                log2(wt[,2]), # y axis
                log2(wt[,3]), # z axis
                pch=".",
                main="WT C1",
                xlab=expression("log"[2]*"(WT C1 Repeat 1)"),

```

```

        ylab=expression("log"[2]*"(WT C1 Repeat 2)"),
        zlab=expression("log"[2]*"(WT C1 Repeat 3)")
    )
    # WT UV1
    scatterplot3d(log2(wt[,7]), # x axis
                 log2(wt[,8]), # y axis
                 log2(wt[,9]), # z axis
                 pch=".",
                 main="WT UV1",
                 xlab=expression("log"[2]*"(WT UV1 Repeat 1)"),
                 ylab=expression("log"[2]*"(WT UV1 Repeat 2)"),
                 zlab=expression("log"[2]*"(WT UV1 Repeat 3)")
    )

    # WT C4
    scatterplot3d(log2(wt[,4]), # x axis
                 log2(wt[,5]), # y axis
                 log2(wt[,6]), # z axis
                 pch=".",
                 main="WT C4",
                 xlab=expression("log"[2]*"(WT C4 Repeat 1)"),
                 ylab=expression("log"[2]*"(WT C4 Repeat 2)"),
                 zlab=expression("log"[2]*"(WT C4 Repeat 3)")
    )

    # WT UV4
    scatterplot3d(log2(wt[,10]), # x axis
                 log2(wt[,11]), # y axis
                 log2(wt[,12]), # z axis
                 pch=".",
                 main="WT UV4",
                 xlab=expression("log"[2]*"(WT UV4 Repeat 1)"),
                 ylab=expression("log"[2]*"(WT UV4 Repeat 2)"),
                 zlab=expression("log"[2]*"(WT UV4 Repeat 3)")
    )

}

```

```

annotateExisting <- function(genelist, output.name="Output",
                             sheet.name="Sheet", namespace="molecular_function", col.names=TRUE,
                             row.names=FALSE, append=FALSE) {
    # #for testing
    #
    #     data <- transcriptome1hr_up
    #     output.name = "GO grouping test DELETE ME"
    #     col.names=TRUE
    #     row.names=FALSE
    #     append=FALSE
    #     namespace="molecular_function"

    library(xlsx)
    library(biomaRt)
    library(dplyr)

    if (output.name=="Output" | sheet.name=="Sheet"){
        cat("WARNING: output.name and/or sheet.name not
        specified.\nUsing default value(s) instead.\n\n")
    }
}

```

```

        xlsx_output <- paste0(Sys.Date(), " ", output.name, "
GO_annotation.xlsx")
        ensembl = useMart("ensembl", dataset="hsapiens_gene_ensembl")

        summarised_go_names <- getBM(attributes=
c('hgnc_symbol', 'go_id', 'name_1006', 'description'), filters =
c('hgnc_symbol', 'go_parent_name'), values = list(genelist$Gene,
namespace), mart = ensembl)

        summarised_go_names <- summarised_go_names %>%
            group_by(hgnc_symbol) %>%
            summarize(go_names = do.call(paste,
c(as.list(name_1006), sep=",")))
            options(dplyr.width = Inf)

        matched <- match(genelist$Gene,
summarised_go_names$hgnc_symbol)
        genelist <- cbind(genelist, summarised_go_names[matched,2])

        write.xlsx2(genelist, xlsx_output, sheetName=sheet.name,
col.names=col.names, row.names=row.names, append=append)

        cat("Annotated", nrow(summarised_go_names), "out of",
nrow(genelist), "Genes\n",
            "Results successfully written to", xlsx_output)
    }

getGeneName <- function(genelist, output){

    library(biomaRt)
    library(dplyr)
    ensembl = useMart("ensembl", dataset="hsapiens_gene_ensembl")

    Gene_description <- getBM(attributes=
c('hgnc_symbol', 'description'), filters = 'hgnc_symbol', values =
genelist$Gene, mart = ensembl)

    matched <- match(genelist$Gene, Gene_description$hgnc_symbol)
    output <<- cbind(as.data.frame(genelist),
Gene_description[matched,2])
    rm(matched, Gene_description)

    cat("Obtained", nrow(Gene_description), "gene descriptions out
of", nrow(genelist), "from input list.\n",
        "Results outputted to the global environment")
}

```

## 8. References

- Aitken, C. E., & Lorsch, J. R. (2012). A mechanistic overview of translation initiation in eukaryotes. *Nature Structural & Molecular Biology*, *19*(6), 568–76. doi:10.1038/nsmb.2303
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. doi:10.1093/bioinformatics/btu638
- Andrews, S. (2016). FastQC: a quality control tool for high throughput sequence data. Retrieved from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Asano, K., Shalev, a, Phan, L., Nielsen, K., Clayton, J., Valásek, L., ... Hinnebusch, a G. (2001). Multiple roles for the C-terminal domain of eIF5 in translation initiation complex assembly and GTPase activation. *The EMBO Journal*, *20*(9), 2326–37. doi:10.1093/emboj/20.9.2326
- Bąkowska-Żywicka, K., Kasprzyk, M., & Twardowski, T. (2016). tRNA-derived short RNAs bind to *Saccharomyces cerevisiae* ribosomes in a stress-dependent manner and inhibit protein synthesis in vitro. *FEMS Yeast Research*, fow077. doi:10.1093/femsyr/fow077
- Bartek, J., & Lukas, J. (2007). DNA damage checkpoints: from initiation to recovery or adaptation. *Current Opinion in Cell Biology*, *19*(2), 238–245. doi:10.1016/j.ceb.2007.02.009
- Bennett, C. B., Lewis, a L., Baldwin, K. K., & Resnick, M. a. (1993). Lethality induced by a single site-specific double-strand break in a dispensable yeast plasmid. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(12), 5613–7. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=46771&tool=pmcentrez&rendertype=abstract>
- Brar, G. A., & Weissman, J. S. (2015). Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol*, *16*(11), 651–664. Retrieved from <http://dx.doi.org/10.1038/nrm4069>
- Braunstein, S., Badura, M. L., Xi, Q., Formenti, S. C., & Schneider, R. J. (2009). Regulation of protein synthesis by ionizing radiation. *Molecular and Cellular Biology*, *29*(21), 5645–56. doi:10.1128/MCB.00711-09
- Bushell, M., Stoneley, M., Kong, Y. W., Hamilton, T. L., Spriggs, K. A., Dobbyn, H. C., ... Willis, A. E. (2006). Polypyrimidine Tract Binding Protein Regulates IRES-Mediated Gene Expression during Apoptosis. *Molecular Cell*, *23*(3), 401–412. doi:10.1016/j.molcel.2006.06.012
- Bushell, M., Wood, W., Carpenter, G., Pain, V. M., Morley, S. J., & Clemens, M. J. (2001). Disruption of the Interaction of Mammalian Protein Synthesis Eukaryotic Initiation Factor 4B with the Poly(A)-binding Protein by Caspase- and Viral Protease-mediated Cleavages. *Journal of Biological Chemistry*, *276*(26), 23922–23928. doi:10.1074/jbc.M100384200
- Buttgereit, F; Brand, M. D. (1995). A hierarchy of ATP-consuming processes in mammalian cells. *The Biochemical Journal*, *312* ( Pt 1, 163–167. doi:10.1210/er.2008-0019
- Chapman, J. R., Taylor, M. R. G., & Boulton, S. J. (2012). Playing the end game: DNA double-strand break repair pathway choice. *Molecular Cell*, *47*(4), 497–510. doi:10.1016/j.molcel.2012.07.029
- Cheng, Z., Saito, K., Pisarev, A. V, Wada, M., Pisareva, V. P., Pestova, T. V, ... Song, H. (2009). Structural insights into eRF3 and stop codon recognition by eRF1. *Genes & Development*, *23*(9), 1106–18. doi:10.1101/gad.1770109
- Ciccia, A., & Elledge, S. J. (2010). The DNA damage response: making it safe to play with knives.

- Molecular Cell*, 40(2), 179–204. doi:10.1016/j.molcel.2010.09.019
- Cimprich, K. a, & Cortez, D. (2008). ATR: an essential regulator of genome integrity. *Nature Reviews. Molecular Cell Biology*, 9(8), 616–27. doi:10.1038/nrm2450
- Costa, R. M. A., Chiganças, V., Galhardo, R. da S., Carvalho, H., & Menck, C. F. M. (2003). The eukaryotic nucleotide excision repair pathway. *Biochimie*, 85(11), 1083–1099. doi:10.1016/j.biochi.2003.10.017
- Craxton, A., Somers, J., Munnur, D., Jukes-Jones, R., Cain, K., & Malewicz, M. (2015). XLS (c9orf142) is a new component of mammalian DNA double-stranded break repair. *Cell Death and Differentiation*, 22(6), 890–897. doi:10.1038/cdd.2015.22
- de Lima-Bessa, K. M., Armelini, M. G., Chiganças, V., Jacysyn, J. F., Amarante-Mendes, G. P., Sarasin, A., & Menck, C. F. M. (2008). CPDs and 6-4PPs play different roles in UV-induced cell death in normal and NER-deficient human cells. *DNA Repair*, 7(2), 303–12. doi:10.1016/j.dnarep.2007.11.003
- Deng, J., Harding, H. P., Raught, B., Gingras, A. C., Berlanga, J. J., Scheuner, D., ... Sonenberg, N. (2002). Activation of gcn2 in uv-irradiated cells inhibits translation. *Current Biology*, 12(15), 1279–1286. doi:10.1016/S0960-9822(02)01037-0
- Dever, T. E., & Green, R. (2012). The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 4(7), a013706. doi:10.1101/cshperspect.a013706
- Dhahbi, J. M., Spindler, S. R., Atamna, H., Yamakawa, A., Boffelli, D., Mote, P., & Martin, D. I. (2013). 5' tRNA halves are present as abundant complexes in serum, concentrated in blood cells, and modulated by aging and calorie restriction. *BMC Genomics*, 14, 298. doi:10.1186/1471-2164-14-298
- Dieci, G., Preti, M., & Montanini, B. (2009). Eukaryotic snoRNAs: A paradigm for gene expression flexibility. *Genomics*, 94(2), 83–88. doi:10.1016/j.ygeno.2009.05.002
- Doudna, J. A., & Sarnow, P. (2007). Translation Initiation by Viral Internal Ribosome Entry Sites. *Cold Spring Harbor Monograph Archive*, 48, 129–153. doi:10.1101/087969767.48.129
- Dunn, J. G., Foo, C. K., Belletier, N. G., Gavis, E. R., & Weissman, J. S. (2013). Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *eLife*, 2013(2), 1–32. doi:10.7554/eLife.01179
- Elroy-Stein, O., & Merrick, W. C. (2007). Translation Initiation Via Cellular Internal Ribosome Entry Sites. *Translational Control in Biology and Medicine*, (monograph 48), 155, 934 . doi:10.1101/087969767.48.155
- Faller, W. J., Jackson, T. J., Knight, J. R. P., Ridgway, R. A., Jamieson, T., Karim, S. A., ... Sansom, O. J. (2014). mTORC1-mediated translational elongation limits intestinal tumour initiation and growth. *Nature*, 517(7535), 497–500. doi:10.1038/nature13896
- Forrester, H. B., Li, J., Hovan, D., Ivashkevich, A. N., & Sprung, C. N. (2012). DNA repair genes: alternative transcription and gene expression at the exon level in response to the DNA damaging agent, ionizing radiation. *PloS One*, 7(12), e53358. doi:10.1371/journal.pone.0053358
- Foster, M., & Mullenders, L. H. (2008). Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Research*, 18(1), 73–84. doi:10.1038/cr.2008.6
- Frosina, G., Fortini, P., Rossi, O., Carrozzino, F., Raspaglio, G., Cox, L. S., ... Dogliotti, E. (1996). Two Pathways for Base Excision Repair in Mammalian Cells \*, 271(16), 9573–9578.

- Fu, H., Feng, J., Liu, Q., Sun, F., Tie, Y., Zhu, J., ... Zheng, X. (2009). Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Letters*, *583*(2), 437–442. doi:10.1016/j.febslet.2008.12.043
- Gandin, V., Masvidal, L., Hulea, L., Gravel, S.-P., Cargnello, M., McLaughlan, S., ... Topisirovic, I. (2016). nanoCAGE reveals 5' UTR features that define specific modes of translation of functionally related MTOR-sensitive mRNAs. *Genome Res*, *26*(5), 636–648. doi:10.1101/gr.197566.115
- Gebetsberger, J., Zywicki, M., Künzi, A., & Polacek, N. (2012). tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in *Haloflex volcanii*. *Archaea*, *2012*, 10–12. doi:10.1155/2012/260909
- Gerashchenko, M. V., & Gladyshev, V. N. (2017). Ribonuclease selection for ribosome profiling. *Nucleic Acids Research*, *45*(2), e6. doi:10.1093/nar/gkw822
- Gillet, L. C. J., & Schärer, O. D. (2006a). Molecular Mechanisms of Mammalian Global Genome Nucleotide Excision Repair. *Chemical Reviews*, *106*(2), 253–276. doi:10.1021/cr040483f
- Gillet, L. C. J., & Schärer, O. D. (2006b). Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chemical Reviews*, *106*(2), 253–76. doi:10.1021/cr040483f
- Gingold, H., Tehler, D., Christoffersen, N. R., Nielsen, M. M., Asmar, F., Kooistra, S. M., ... Pilpel, Y. (2014). A dual program for translation regulation in cellular proliferation and differentiation. *Cell*, *158*(6), 1281–92. doi:10.1016/j.cell.2014.08.011
- Gingras, A., Raught, B., Gygi, S. P., Niedzwiecka, A., Miron, M., Burley, S. K., ... Sonenberg, N. (2001). Hierarchical phosphorylation of the translation inhibitor 4E-BP1, 2852–2864. doi:10.1101/gad.912401.7
- Gingras, A., Raught, B., & Sonenberg, N. (2001). Regulation of translation initiation by FRAP / mTOR. *Genes & Development*, *15*, 807–826. doi:10.1101/gad.887201
- Gross, J. D., Moerke, N. J., Haar, T. Von Der, Lugovskoy, A. A., Sachs, A. B., McCarthy, J. E. G., & Wagner, G. (2003). Ribosome Loading onto the mRNA Cap Is Driven by Conformational Coupling between eIF4G and eIF4E. *Cell*, *115*, 739–750.
- Hanawalt, P. C. (2002). Subpathways of nucleotide excision repair and their regulation. *Oncogene*, *21*(58), 8949–8956. doi:10.1038/sj.onc.1206096
- Hara, K., Yonezawa, K., Kozlowski, M. T., Sugimoto, T., Andrabi, K., Weng, Q., ... Avruch, J. (1997). Regulation of eIF-4E BP1 Phosphorylation by mTOR \*, *272*(42), 26457–26463.
- Harper, J. W., & Elledge, S. J. (2007). The DNA Damage Response: Ten Years After. *Molecular Cell*, *28*(5), 739–745. doi:10.1016/j.molcel.2007.11.015
- Honda, S., Loher, P., Shigematsu, M., Palazzo, J. P., Suzuki, R., Imoto, I., ... Kirino, Y. (2015). Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(29), E3816-25. doi:10.1073/pnas.1510077112
- Ingolia, N. T., Brar, G. a, Rouskin, S., McGeachy, A. M., & Weissman, J. S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols*, *7*(8), 1534–50. doi:10.1038/nprot.2012.086
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (New York, N.Y.)*, *324*(5924), 218–23. doi:10.1126/science.1168978
- Ivanov, P., Emara, M. M., Villen, J., Gygi, S. P., & Anderson, P. (2011). Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation. *Molecular Cell*, *43*(4), 613–623.

doi:10.1016/j.molcel.2011.06.022

- Ivanov, P., O'Day, E., Emara, M. M., Wagner, G., Lieberman, J., & Anderson, P. (2014). G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments. *Proceedings of the National Academy of Sciences*, *111*(51), 18201–18206. doi:10.1073/pnas.1407361111
- Jackson, T. J. (2013). *Translational control: mTOR signalling and the use of next-generation sequencing methods*. University of Nottingham. Retrieved from <http://etheses.nottingham.ac.uk/3511/>
- Jen, K.-Y., & Cheung, V. G. (2003). Transcriptional response of lymphoblastoid cells to ionizing radiation. *Genome Research*, *13*(9), 2092–100. doi:10.1101/gr.1240103
- Kao, J., Rosenstein, B. S., Peters, S., Milano, M. T., & Kron, S. J. (2005). Cellular response to DNA damage. *Annals of the New York Academy of Sciences*, *1066*, 243–58. doi:10.1196/annals.1363.012
- Kapp, L. D., & Lorsch, J. R. (2004). GTP-dependent Recognition of the Methionine Moiety on Initiator tRNA by Translation Factor eIF2. *Journal of Molecular Biology*, *335*(4), 923–936. doi:10.1016/j.jmb.2003.11.025
- Karanam, K., Kafri, R., Loewer, A., & Lahav, G. (2012). Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. *Molecular Cell*, *47*(2), 320–9. doi:10.1016/j.molcel.2012.05.052
- Kaufmann, S. H., Desnoyers, S., Ottaviano, Y., Apoptosis, C., Davidson, N. E., & Poirier, G. G. (1993). Specific Proteolytic Cleavage of Poly ( ADP-ribose ) Polymerase : An Early Marker of Chemotherapy-induced Apoptosis Specific Proteolytic Cleavage of Poly ( ADP-ribose ) Polymerase : An Early Marker of. *Cancer Res*, *53*, 3976–3985.
- Kawaji, H., Nakamura, M., Takahashi, Y., Sandelin, A., Katayama, S., Fukuda, S., ... Hayashizaki, Y. (2008). Hidden layers of human small RNAs. *BMC Genomics*, *9*, 157. doi:10.1186/1471-2164-9-157
- Kedersha, N., Ivanov, P., & Anderson, P. (2013). Stress granules and cell signaling: More than just a passing phase? *Trends in Biochemical Sciences*, *38*(10), 494–506. doi:10.1016/j.tibs.2013.07.004
- Keene, J. D. (2010). Minireview: Global Regulation and Dynamics of Ribonucleic Acid. *Endocrinology*, *151*(4), 1391–1397. doi:10.1210/en.2009-1250
- Kieft, J. S. (2009). Comparing the three-dimensional structures of Dicistroviridae IGR IRES RNAs with other viral RNA structures. *Virus Research*, *139*(2), 148–156. doi:10.1016/j.virusres.2008.07.007
- King, H. A., & Gerber, A. P. (2016). Translatome profiling: Methods for genome-scale analysis of mRNA translation. *Briefings in Functional Genomics*, *15*(1), 22–31. doi:10.1093/bfgp/elu045
- Kolitz, S. E., & Lorsch, J. R. (2010). Eukaryotic initiator tRNA: finely tuned and ready for action. *FEBS Letters*, *584*(2), 396–404. doi:10.1016/j.febslet.2009.11.047
- Kolupaeva, V. G., Unbehaun, A., Lomakin, I. B., Hellen, C. U. T., & Pestova, T. V. (2005). Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. *RNA (New York, N.Y.)*, *11*(4), 470–86. doi:10.1261/rna.7215305
- Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research*, *15*(20), 8125–48. Retrieved from

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=306349&tool=pmcentrez&endertype=abstract>

- Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. *Journal of Biological Chemistry*, 266(30), 19867–19870.
- Lafontaine, D. L. J. (2015). Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nature Structural & Molecular Biology*, 22(1), 11–19. doi:10.1038/nsmb.2939
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. doi:10.1038/nmeth.1923
- Lavin, M. F. (2007). ATM and the Mre11 complex combine to recognize and signal DNA double-strand breaks. *Oncogene*, 26(56), 7749–58. doi:10.1038/sj.onc.1210880
- Lazzaro, F., Giannattasio, M., Puddu, F., Granata, M., Pelliccioli, A., Plevani, P., & Muzi-Falconi, M. (2009). Checkpoint mechanisms at the intersection between DNA damage and repair. *DNA Repair*, 8(9), 1055–1067. doi:10.1016/j.dnarep.2009.04.022
- Lee, J.-H., Choy, M. L., Ngo, L., Foster, S. S., & Marks, P. A. (2010). Histone deacetylase inhibitor induces DNA damage, which normal but not transformed cells can repair. *Proceedings of the National Academy of Sciences of the United States of America*, 107(33), 14639–44. doi:10.1073/pnas.1008522107
- Lee, Y. S., Shibata, Y., Malhotra, A., Lee, Y. S., Shibata, Y., Malhotra, A., & Dutta, A. (2009). A novel class of small RNAs : tRNA-derived RNA fragments ( tRFs ) A novel class of small RNAs : tRNA-derived RNA fragments ( tRFs ), 2639–2649. doi:10.1101/gad.1837609
- LeFebvre, A. K., Korneeva, N. L., Trutschl, M., Cvek, U., Duzan, R. D., Bradley, C. a, ... Rhoads, R. E. (2006). Translation initiation factor eIF4G-1 binds to eIF3 through the eIF3e subunit. *The Journal of Biological Chemistry*, 281(32), 22917–32. doi:10.1074/jbc.M605418200
- Lindahl, T., & Wood, R. D. (1999). Quality Control by DNA Repair. *Science*, 286(5446), 1897–1905. doi:10.1126/science.286.5446.1897
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. doi:10.1186/s13059-014-0550-8
- Lü, X., De La Peña, L., Barker, C., Camphausen, K., & Tofilon, P. J. (2006). Radiation-induced changes in gene expression involve recruitment of existing messenger RNAs to and away from polysomes. *Cancer Research*, 66(2), 1052–1061. doi:10.1158/0008-5472.CAN-05-3459
- Maag, D., Fekete, C. a, Gryczynski, Z., & Lorsch, J. R. (2005). A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Molecular Cell*, 17(2), 265–75. doi:10.1016/j.molcel.2004.11.051
- Madronich, S., McKenzie, R. L., Björn, L. O., & Caldwell, M. M. (1998). Changes in biologically active ultraviolet radiation reaching the Earth's surface. *Journal of Photochemistry and Photobiology. B, Biology*, 46(1–3), 5–19. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9894350>
- Mahaney, B. L., Meek, K., & Lees-Miller, S. P. (2009). Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *The Biochemical Journal*, 417(3), 639–50. doi:10.1042/BJ20080413
- Majumdar, R., Bandyopadhyay, A., & Maitra, U. (2003). Mammalian translation initiation factor eIF1 functions with eIF1A and eIF3 in the formation of a stable 40 S preinitiation complex. *The Journal of Biological Chemistry*, 278(8), 6580–7. doi:10.1074/jbc.M210357200

- Marcotrigiano, J., Gingras, A., Sonenberg, N., & Burley, S. K. (1999). Cap-Dependent Translation Initiation in Eukaryotes Is Regulated by a Molecular Mimic of eIF4G, *3*, 707–716.
- Marinov, G. K., Williams, B. A., McCue, K., Schroth, G. P., Gertz, J., Myers, R. M., & Wold, B. J. (2014). From single-cell to cell-pool transcriptomes: Stochasticity in gene expression and RNA splicing. *Genome Research*, *24*(3), 496–510. doi:10.1101/gr.161034.113
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10. doi:10.14806/ej.17.1.200
- Matos, J., & West, S. C. (2014). Holliday junction resolution: Regulation in space and time. *DNA Repair*, *19*, 176–181. doi:10.1016/j.dnarep.2014.03.013
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., ... Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*, *41*(Web Server issue), 597–600. doi:10.1093/nar/gkt376
- Miettinen, T. P., & Bjorklund, M. (2015). Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3' untranslated regions. *Nucleic Acids Research*, *43*(2), 1019–1034. doi:10.1093/nar/gku1310
- Minafra, L., Bravat??, V., Russo, G., Forte, G. I., Cammarata, F. P., Ripamonti, M., ... Gilardi, M. C. (2015). Gene expression profiling of MCF10A breast epithelial cells exposed to IOERT. *Anticancer Research*, *35*(6), 3223–3234.
- Mishima, E., Inoue, C., Saigusa, D., Inoue, R., Ito, K., Suzuki, Y., ... Abe, T. (2014). Conformational change in transfer RNA is an early indicator of acute cellular damage. *Journal of the American Society of Nephrology : JASN*, *25*(10), 2316–26. doi:10.1681/ASN.2013091001
- Misteli, T., & Soutoglou, E. (2009). The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nature Reviews. Molecular Cell Biology*, *10*(4), 243–54. doi:10.1038/nrm2651
- Nanda, J. S., Cheung, Y.-N., Takacs, J. E., Martin-Marcos, P., Saini, A. K., Hinnebusch, A. G., & Lorsch, J. R. (2009). eIF1 controls multiple steps in start codon recognition during eukaryotic translation initiation. *Journal of Molecular Biology*, *394*(2), 268–85. doi:10.1016/j.jmb.2009.09.017
- Nanda, J. S., Saini, A. K., Muñoz, A. M., Hinnebusch, A. G., & Lorsch, J. R. (2013). Coordinated movements of eukaryotic translation initiation factors eIF1, eIF1A, and eIF5 trigger phosphate release from eIF2 in response to start codon recognition by the ribosomal preinitiation complex. *The Journal of Biological Chemistry*, *288*(8), 5316–29. doi:10.1074/jbc.M112.440693
- Neeley, W. L., & Essigmann, J. M. (2006). Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chemical Research in Toxicology*, *19*(4), 491–505. doi:10.1021/tx0600043
- Ochi, T., Blackford, A. N., Coates, J., Jhujh, S., Mehmood, S., Tamura, N., ... Jackson, S. P. (2015). PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science*, *347*(6218), 185–188. doi:10.1126/science.1261971
- Oh, E., Becker, A. H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., ... Bukau, B. (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell*, *147*(6), 1295–1308. doi:10.1016/j.cell.2011.10.044
- Olshen, A. B., Hsieh, A. C., Stumpf, C. R., Olshen, R. A., Ruggero, D., & Taylor, B. S. (2013a). Assessing gene-level translational control from ribosome profiling. *Bioinformatics*, *29*(23), 2995–3002. doi:10.1093/bioinformatics/btt533

- Olshen, A. B., Hsieh, A. C., Stumpf, C. R., Olshen, R. A., Ruggero, D., & Taylor, B. S. (2013b). Genome analysis Assessing gene-level translational control from ribosome profiling, 1–8. doi:10.1093/bioinformatics/btt533
- Paglin, S., Lee, N.-Y., Nakar, C., Fitzgerald, M., Plotkin, J., Deuel, B., ... Yahalom, J. (2005). Rapamycin-sensitive pathway regulates mitochondrial membrane potential, autophagy, and survival in irradiated MCF-7 cells. *Cancer Research*, 65(23), 11061–70. doi:10.1158/0008-5472.CAN-05-1083
- Pall, G. S., & Hamilton, A. J. (2008). Improved northern blot method for enhanced detection of small RNA. *Nature Protocols*, 3(6), 1077–1084. doi:10.1038/nprot.2008.67
- Panier, S., & Boulton, S. J. (2013). Double-strand break repair: 53BP1 comes into focus. *Nature Reviews. Molecular Cell Biology*, 1(December). doi:10.1038/nrm3719
- Pavon-Eternod, M., Gomes, S., Geslain, R., Dai, Q., Rosner, M. R., & Pan, T. (2009). tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Research*, 37(21), 7268–7280. doi:10.1093/nar/gkp787
- Pestova, T. V. (2003). Translation elongation after assembly of ribosomes on the Cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. *Genes & Development*, 17(2), 181–186. doi:10.1101/gad.1040803
- Piccirillo, C. a, Bjur, E., Topisirovic, I., Sonenberg, N., & Larsson, O. (2014). Translational control of immune responses: from transcripts to translatoemes. *Nature Immunology*, 15(6), 503–511. doi:10.1038/ni.2891
- Polo, S. E., & Jackson, S. P. (2011). Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes & Development*, 25(5), 409–33. doi:10.1101/gad.2021311
- Powley, I. R., Kondrashov, A., Young, L. a, Dobbyn, H. C., Hill, K., Cannell, I. G., ... Willis, A. E. (2009). Translational reprogramming following UVB irradiation is mediated by DNA-PKcs and allows selective recruitment to the polysomes of mRNAs encoding DNA repair enzymes. *Genes & Development*, 23(10), 1207–20. doi:10.1101/gad.516509
- Raina, M., & Ibba, M. (2014). TRNAs as regulators of biological processes. *Frontiers in Genetics*, 5(JUN), 1–14. doi:10.3389/fgene.2014.00171
- Rastogi, R. P., Richa, Kumar, A., Tyagi, M. B., & Sinha, R. P. (2010). Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *Journal of Nucleic Acids*, 2010, 592980. doi:10.4061/2010/592980
- Reinhardt, H. C., & Yaffe, M. B. (2009). Kinases that control the cell cycle in response to DNA damage: Chk1, Chk2, and MK2. *Current Opinion in Cell Biology*, 21(2), 245–255. doi:10.1016/j.ceb.2009.01.018
- Reon, B. J., & Dutta, A. (2016). Biological processes discovered by high-throughput sequencing. *American Journal of Pathology*, 186(4), 722–732. doi:10.1016/j.ajpath.2015.10.033
- Richter, J. D., & Collier, J. (2015). Pausing on Polyribosomes: Make Way for Elongation in Translational Control. *Cell*, 163(2), 292–300. doi:10.1016/j.cell.2015.09.041
- Riou, L., Eveno, E., Hoffen, A. Van, Zeeland, A. A. van, Sarasin, A., & Mullenders, L. H. F. (2004). Differential Repair of the Two Major UV-Induced Photolesions in Trichothiodystrophy Fibroblasts Differential Repair of the Two Major UV-Induced Photolesions in Trichothiodystrophy Fibroblasts. *Cancer Research*, 64, 889–894.
- Ron, D. (2002). Translational control in the endoplasmic reticulum stress response, 110(10), 1383–1388. doi:10.1172/JCI200216784.From

- Ron, D., & Harding, H. (2007). eIF2 Phosphorylation in Cellular Stress Responses and Disease. *Translational Control in Biology and Medicine*, 349–372. doi:10.1101/087969767.48.345
- Sage, E. (1993). DISTRIBUTION AND REPAIR OF PHOTOLESIONS IN DNA: GENETIC CONSEQUENCES AND THE ROLE OF SEQUENCE CONTEXT. *Photochemistry and Photobiology*, 57(1), 163–174. doi:10.1111/j.1751-1097.1993.tb02273.x
- Saikia, M., Krokowski, D., Guan, B. J., Ivanov, P., Parisien, M., Hu, G. F., ... Hatzoglou, M. (2012). Genome-wide identification and quantitative analysis of cleaved tRNA fragments induced by cellular stress. *Journal of Biological Chemistry*, 287(51), 42708–42725. doi:10.1074/jbc.M112.371799
- Sancar, A., Lindsey-Boltz, L. a, Unsal-Kaçmaz, K., & Linn, S. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual Review of Biochemistry*, 73, 39–85. doi:10.1146/annurev.biochem.73.011303.073723
- Schneider-Poetsch, T., Ju, J., Eyler, D. E., Dang, Y., Bhat, S., Merrick, W. C., ... Liu, J. O. (2010). Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nature Chemical Biology*, 6(3), 209–217. doi:10.1038/nchembio.304
- Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7), 671–675. doi:10.1038/nmeth.2089
- Schultz, P., Poterszman, A., Moras, D., Egly, J. M., Fries, L., & Cedex, F.-I. (2000). Molecular Structure of Human TFIIF, 102, 599–607.
- Seeberg, E., Eide, L., & Bjørås, M. (1995). The base excision repair pathway. *Trends in Biochemical Sciences*, 20(10), 391–397. doi:10.1016/S0968-0004(00)89086-6
- Sidrauski, C., Acosta-Alvear, D., Khoutorsky, A., Vedantham, P., Hearn, B. R., Li, H., ... Walter, P. (2013). Pharmacological brake-release of mRNA translation enhances cognitive memory. *eLife*, 2013(2), 1–22. doi:10.7554/eLife.00498
- Sidrauski, C., McGeachy, A. M., Ingolia, N. T., & Walter, P. (2015). The small molecule ISRIB reverses the effects of eIF2 $\alpha$  phosphorylation on translation and stress granule assembly. *eLife*, 4, 1–16. doi:10.7554/eLife.05033
- Sinha, R. P., & Häder, D.-P. (2002). UV-induced DNA damage and repair: a review. *Photochemical & Photobiological Sciences*, 1(4), 225–236. doi:10.1039/b201230h
- Sobala, A., & Hutvagner, G. (2011). Transfer RNA-derived fragments: Origins, processing, and functions. *Wiley Interdisciplinary Reviews: RNA*, 2(6), 853–862. doi:10.1002/wrna.96
- Sobala, A., & Hutvagner, G. (2013). Small RNAs derived from the 5' end of tRNA can inhibit protein translation in human cells. *RNA Biology*, 10(4), 553–63. doi:10.4161/rna.24285
- Somers, J., Wilson, L. A., Kilday, J., Horvilleur, E., Cannell, I. G., Pöyry, T. A. A., ... Willis, A. E. (2015). A common polymorphism in the 5' UTR of ERCC5 creates an upstream ORF that confers resistance to platinum-based chemotherapy. *Genes & Development*, 29(18), 1891–1896. doi:10.1101/gad.261867.115
- Sonenberg, N., & Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, 136(4), 731–45. doi:10.1016/j.cell.2009.01.042
- Spahr, P. F., & Hollingworth, B. R. (1961). Purification and Mechanism of Action of Ribonuclease from Escherichia coli Purification and Mechanism from Escherichia of Action of Ribonuclease coli Ribosomes. *J. Biol. Chem*, 236, 823–831.
- Spriggs, K. a, Bushell, M., & Willis, A. E. (2010). Translational regulation of gene expression during conditions of cell stress. *Molecular Cell*, 40(2), 228–37. doi:10.1016/j.molcel.2010.09.028

- Su, Y., Meador, J. a, Geard, C. R., & Balajee, A. S. (2010). Analysis of ionizing radiation-induced DNA damage and repair in three-dimensional human skin model system. *Experimental Dermatology*, *19*(8), e16-22. doi:10.1111/j.1600-0625.2009.00945.x
- Taylor, J., Garrett, D. S., & Cohrs, M. P. (1988). Solution-State Structure of the Dewar Pyrimidinone Photoproduct of Thymidyl-(3'->5')-thymidine. *Biochemistry*, *27*, 7206–7215.
- Torregrosa-Munumer, R., Goffart, S., Haikonen, J. A., & Pohjoismaki, J. L. O. (2015). Low doses of UV and oxidative damage induce dramatic accumulation of mitochondrial DNA replication intermediates, fork regression and replication initiation shift. *Molecular Biology of the Cell*, *26*, 4197–4208. doi:10.1091/mbc.E15-06-0390
- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105–1111. doi:10.1093/bioinformatics/btp120
- Trivigno, D., Bornes, L., Huber, S. M., & Rudner, J. (2013). Regulation of protein translation initiation in response to ionizing radiation. *Radiat. Oncol.*, *8*(1), 35. doi:10.1186/1748-717X-8-35
- Tsai, M.-L., Chang, K.-Y., Chiang, C.-S., Shu, W.-Y., Weng, T.-C., Chen, C. R., ... Hsu, I. C. (2009). UVB radiation induces persistent activation of ribosome and oxidative phosphorylation pathways. *Radiation Research*, *171*(6), 716–24. doi:10.1667/RR1625.1
- Uehara, F., Miwa, S., Tome, Y., Hiroshima, Y., Yano, S., Yamamoto, M., ... Hoffman, R. M. (2014). Comparison of UVB and UVC effects on the DNA damage-response protein 53BP1 in human pancreatic cancer. *Journal of Cellular Biochemistry*, *115*(10), 1724–1728. doi:10.1002/jcb.24837
- Vattem, K. M., & Wek, R. C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proceedings of the National Academy of Sciences*, *101*(31), 11269–11274. doi:10.1073/pnas.0400541101
- Wang, H., & Kochevar, I. E. (2005). Involvement of UVB-induced reactive oxygen species in TGF-beta biosynthesis and activation in keratinocytes. *Free Radical Biology & Medicine*, *38*(7), 890–7. doi:10.1016/j.freeradbiomed.2004.12.005
- Weake, V. M., & Workman, J. L. (2010). Inducible gene expression: diverse regulatory mechanisms. *Nature Reviews Genetics*, *11*(6), 426–437. doi:10.1038/nrg2781
- Weingarten-Gabbay, S., Elias-Kirma, S., Nir, R., Gritsenko, A. A., Stern-Ginossar, N., Yakhini, Z., ... Segal, E. (2016). Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science*, *351*(6270), aad4939. doi:10.1126/science.aad4939
- Williams, B. R. (1999). PKR; a sentinel kinase for cellular stress. *Oncogene*, *18*(45), 6112–6120. doi:10.1038/sj.onc.1203127
- Wolin, S. L., & Walter, P. (1988). Ribosome pausing and stacking during translation of a eukaryotic mRNA. *The EMBO Journal*, *7*(11), 3559–3569.
- Woolstenhulme, C. J., Guydosh, N. R., Green, R., & Buskirk, A. R. (2015). High-Precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Reports*, *11*(1), 13–21. doi:10.1016/j.celrep.2015.03.014
- Wyatt, H. D. M., & West, S. C. (2014). Holliday junction resolvases. *Cold Spring Harbor Perspectives in Biology*, *6*(9), 1–29. doi:10.1101/cshperspect.a023192
- Wyman, C., & Kanaar, R. (2006). DNA double-strand break repair: all's well that ends well. *Annual Review of Genetics*, *40*, 363–83. doi:10.1146/annurev.genet.40.110405.090451
- Xing, M., Yang, M., Huo, W., Feng, F., Wei, L., Jiang, W., ... Xu, D. (2015). Interactome analysis

- identifies a new paralogue of XRCC4 in non-homologous end joining DNA repair pathway. *Nature Communications*, 6, 6233. doi:10.1038/ncomms7233
- Xue, S., Tian, S., Fujii, K., Kladwang, W., Das, R., & Barna, M. (2014). RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. *Nature*, 517(7532), 33–38. doi:10.1038/nature14010
- Yamasaki, S., Ivanov, P., Hu, G. F., & Anderson, P. (2009). Angiogenin cleaves tRNA and promotes stress-induced translational repression. *Journal of Cell Biology*, 185(1), 35–42. doi:10.1083/jcb.200811106
- Yamashita, R., Suzuki, Y., Takeuchi, N., Wakaguri, H., Ueda, T., Sugano, S., & Nakai, K. (2008). Comprehensive detection of human terminal oligo-pyrimidine (TOP) genes and analysis of their characteristics. *Nucleic Acids Research*, 36(11), 3707–3715. doi:10.1093/nar/gkn248
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13, 134. doi:10.1186/1471-2105-13-134
- You, Y. H., Lee, D. H., Yoon, J. H., Nakajima, S., Yasui, a, & Pfeifer, G. P. (2001). Cyclobutane pyrimidine dimers are responsible for the vast majority of mutations induced by UVB irradiation in mammalian cells. *The Journal of Biological Chemistry*, 276(48), 44688–94. doi:10.1074/jbc.M107696200
- Young, L. A. (2011). *Translational control of gene expression following exposure to DNA-damaging agents*. University of Nottingham. Retrieved from <http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.555405>
- Yu, Y., Marintchev, A., Kolupaeva, V. G., Unbehauen, A., Veryasova, T., Lai, S.-C., ... Pestova, T. V. (2009). Position of eukaryotic translation initiation factor eIF1A on the 40S ribosomal subunit mapped by directed hydroxyl radical probing. *Nucleic Acids Research*, 37(15), 5167–5182. doi:10.1093/nar/gkp519
- Zhou, B. S., & Elledge, S. J. (2000). The DNA damage response: putting checkpoints in perspective. *Nature*, 408(November), 433–439.