Characterisation of cytoplasmic lncRNA interactions with translation machinery during human neuronal differentiation

Aikaterini Douka

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

Faculty of Biological Sciences

School of Molecular and Cellular Biology

September 2020

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Dedication

My thesis is dedicated:

To Eva Vitsa (MSc), Dr Nikos Babilis my biology teachers and Katerina Skaltsa (MSc), my chemistry teacher, who inspired me to love science and follow my dream to become a Biologist.

To my parents

Acknowledgements

First of all, I would like to thank my supervisor Dr Julie Aspden for giving me the opportunity to join her amazing RNA lab and for her immense support and encouragement throughout this journey. Julie introduced me to the world of RNA and has been the best mentor I could wish for over these years. I will never forget her enthusiasm when I presented even a minor promising result and her most kind words of advice and encouragement, when I burst into tears during more difficult moments of this PhD.

A special thank you to my co-supervisors Prof. Adrian Whitehouse and Prof. James Deuchars, for their guidance and great support during my PhD. I have learned a lot from Ade's constructive criticism on my experiments and I am really grateful to Jim for offering to look at my immunofluorescence slides with such an enthusiasm and providing valuable feedback.

A big thank you to Dr Dapeng Wang for his valuable help and assistance with the analysis of my RNA-Seq data and to Martin Callaghan and the HPC team for their help.

I find it hard to put my feelings into words and I feel like 'thank you' is sometimes not enough to say to the people who have been my second family. Massive thanks to my special friend Michaela Agapiou, with whom we began our journey in the Aspden lab at the same time, for always being there for me to give me a hug when I needed it! A special thank you to Ioannis Tsagakis, whom I consider my 'lab-brother', for his support and valuable help whenever I needed it. I also want to say a big thank you to Dr Eleanor Walton and Tayah Hopes, for the hearty welcome in the lab as well as their emotional support and valuable advice! I will miss Tayah's strong coffee during our coffee/tea breaks! A great thank you to Isabel Birds for our fruitful discussions over lncRNA smORFs and for being such a fantastic collaborator!! I also want to thank Dr Karl Norris, for his advice and his feedback and a special thank you to Dr Mark Handley for his valuable suggestions during writing this thesis. I'm so grateful for every moment I spent as member of the Aspden 'RNA crew', as I call it. I will miss our celebrations, the shuffleboard and pizza nights, and above all I will miss every one of you guys so much!!

I wish to also thank my friend Dr Iosifina Sampson, for her emotional support and our endless conversations during incubation times, at late hours in the lab, as well as for her valuable advice when I needed it.

This thesis wouldn’t have been possible without the support of my partner, George, whom I thank for his kindness and his patience throughout this journey. A special thank you goes to my family, my parents, my brother and my granny for always being there for me to remind me that home is not 3,000km away but... just 30min via skype!!

Abstract

Long non-coding RNAs (lncRNAs) are non-coding transcripts longer than 200nt, many of which are capped, spliced and polyadenylated, like mRNAs. However, lncRNAs exhibit more tissue and developmental-stage specificity than mRNAs and are particularly enriched in the nervous system. Several lncRNAs contribute to neuronal differentiation and lncRNA mis-regulation is implicated in neurological disorders. Whilst nuclear functions of lncRNAs are extensively studied, less is known about their cytoplasmic functions, even though many are localised there. LncRNAs contain small Open Reading Frames (smORFs) a small proportion of which show evidence of active translation, resulting in small peptide synthesis in *S. cerevisiae*, *D. melanogaster*, *M. musculus* and *H. sapiens*. However, these translation events remain poorly understood. This thesis aims to characterise the interactions of cytoplasmic lncRNAs with the translation machinery and investigate their coding potential, using Poly-Ribo-Seq, during the early stages of human neuronal differentiation.

Differentiation of human SH-SY5Y cells with retinoic acid resulted in a significant reduction of active translation, revealed by polysome profiling. To determine the importance of lncRNA-polysome interactions in differentiation, Poly-Ribo-Seq was performed in undifferentiated and differentiated SH-SY5Y. Upon neuronal differentiation 178 lncRNAs are upregulated and 100 downregulated, 71% and 58% of which are associated with polysomes, respectively. Additionally, a dynamic interaction pattern between lncRNAs and polysomes was identified during differentiation, validated by RT-qPCR of specific cytoplasmically enriched lncRNAs across sucrose gradients.

LncRNAs of Control and differentiated cells contain actively translated smORFs. 45 translated lncRNA smORFs were identified by Poly-Ribo-Seq. One of these is LINC01116 smORF, encoding for an 87aa putative peptide. Its translation was validated by FLAG-tagging assays in SH-SY5Y and HEK293 cells. It exhibits cytoplasmic distribution with distinct localisation in cell membrane protrusions. LINC01116 expression is upregulated 4-fold upon differentiation. LINC01116 knockdown results in significant reduction of neurite length and neuronal marker MOXD1 in differentiated cells indicating it contributes to neuronal differentiation.

Table of contents

List of figures

List of tables

Abbreviations

AD	Alzheimer's disease
APS	Ammonium persulfate
BDNF	Brain derived neurotrophic factor
BMP	Bone morphogenetic protein
bp	Basepairs
CDS	Coding sequence
circRNA	circular RNA
CRABP2	Cellular retinoic acid binding protein 2
CREB	c-AMP response element binding protein
DMEM	Dulbecco’s modified eagle medium
DMSO	Dimethylsulfoxide
dNTPs	Deoxyribonucleotide triphosphate
dORF	Downstream open reading frame
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
eIF	Eukaryotic initiation factor
eEF	Eukaryotic elongation factor
eRF	Eukaryotic release factor
ER	Endoplasmic reticulum
ERK	Extracellular signal regulated kinase
eRNA	enhancer RNA
FBS	Fetal bovine serum
FGF	Fibroblast growth factor
FP	Footprint
GABA	Gamma aminobutyric acid
GAPDH	Glyceraldehyde 3-phosphate dehydrogenas
GFP	Green fluorescent protein
GO	Gene ontology
hASCs	Human adipose derived stem cells
HD	Huntington's disease
hESCs	Human Embryonic stem cells
HLA	Human leukocyte antigen

hnRNA	Heterogeneous nuclear RNA
ICC	Immunocytochemistry
iPSCs	Induced pluripotent stem cells
kDa	KiloDaltons
lncRNA	Long non-coding RNA
LTP	Long term potentiation
LTR	Long terminal repeat sequence
mESCs	Mouse embryonic stem cells
miRNA	MicroRNA
mRNA	Messenger RNA
NADH	Nicotinamide adenine dinucleotide (NAD) + hydrogen (H)
NCAM	Neural cell adhesion molecule
ncRNA	non-coding RNA
NPCs	Neural progenitor cells
NSCs	Neural stem cells
ORF	Open Reading Frame
PBS	Phosphate buffer saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PD	Parkinson's disease
piRNA	Piwi interacting RNA
RA	Retinoic acid
RALDH	Retinaldehyde dehydrogenase
RAR	Retinoic acid receptor
RARE	RA response elements
RBPs	RNA binding proteins
RNPs	Ribonucleoprotein
RPF	Ribosome Protected Fragment
rRNA	Ribosomal RNA
RRS	Ribosome release score
RT	Reverse transcription (unless otherwise explained)
RT	Room temperature

RT-qPCR	Real-time quantitative PCR
RXR	Retinoid X receptor
SDS	Sodium dodecyl sulphate
SHH	Sonic hedgehog
siRNA	Small interference RNA
smORF	Small Open reading frame
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
SR	Sarcoplasmic reticulum
TBE	Tris Borate EDTA
TEMED	Tetramethylethylenediamine
tRNA	transfer RNA
TSS	Trancription start site
uORF	Upstream open reading frame
UTR	Untranslated region
WB	Western blot
WNT	Wingless and Int1
WT	Wild type
Δ	Delta (also used for mutant)

Chapter 1

General Introduction

1.1 Non-coding RNAs

The decipher of the genetic code (Nirenberg and Matthaei, 1961) introduced the concept 'messenger RNA' of mRNA functioning as a protein-coding intermediate. Alongside this, the discovery of rRNA and tRNA in the 1950s (Hoagland et al., 1958) introduced the idea of 'non-coding' RNA, as opposed to the protein-coding mRNA (Brenner et al., 1961; Gros et al., 1961). Almost 3 decades later, other important classes of non-coding RNAs, such as snRNAs and snoRNAs, were discovered and characterised (Hodnett and Busch, 1968; Weinberg and Penman, 1968; Lerner and Steitz, 1979). Those types of 'housekeeping' ncRNAs were extensively studied as they play pivotal roles in mRNA splicing and protein synthesis. However, the vast majority of studies during the second half of the 20th century justifiably focused on the characterisation of protein-coding genes, since their functions could be directly related to phenotypes. The technical expertise available at the time allowed the thorough molecular characterisation of a limited number of genes at a time. Consequently, our understanding of non-coding genome functions was very limited until the emergence of high throughput sequencing of the human genome about 20 years ago (Lander, E.S. et al., 2001).

The deep sequencing of the genomes and transcriptomes (ENCODE/NONCODE project) (Liu, 2004; He et al., 2007; Djebali et al., 2012) of various organisms revolutionised our view of gene expression. We now know that ~80% of the human genome can be transcribed (ENCODE) and what was previously referred to as 'junk DNA' currently encompasses some major classes of non-coding RNAs with rapidly emerging structural and regulatory functions (Cech and Steitz, 2014). As of July 2020, the GENCODE project reports that 67% of human genes are non-coding and specifically 13% of human genes encode small non-coding RNAs, whilst 30% encode long non-coding RNAs, a percentage similar to that of protein-coding genes (Figure 1.1). Non-coding RNAs (ncRNAs) are highly diverse but can be broadly classified into 2 categories: housekeeping ncRNAs and regulatory ncRNAs (Zhang et al., 2019). The former comprises rRNAs, tRNAs, snRNAs, snoRNAs and telomeraseRNA (TERC), whilst the latter includes the more recently discovered microRNAs (miRNA) (Lee et al., 1993; Wightman et al., 1993), siRNAs, eRNAs, piRNAs, vaultRNAs, Y RNA, circRNAs and lncRNAs (Table 1.1). Our current understanding of the whole spectrum of functions of ncRNAs still lags behind, compared to that of mRNAs. Nevertheless,

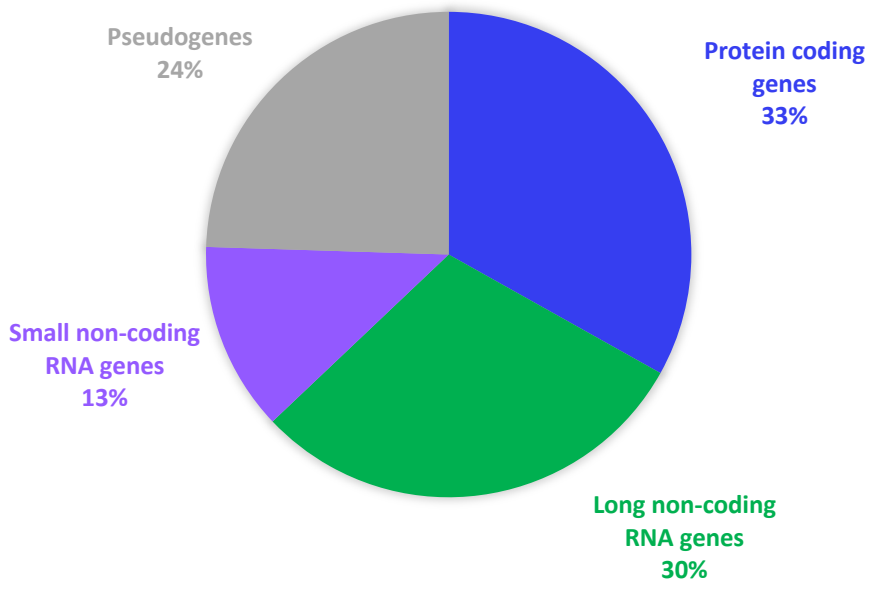


Figure 1.1: Proportion of protein-coding and non-coding genes in the human genome (Gencode release v.34). Data extracted in July 2020.

Table 1.1: Summary of ncRNA functions

Non-coding RNA	Function	Reference
rRNA	RNA component of the small and large ribosomal subunit; ribozyme	(Hoagland et al., 1958)
tRNA	Base pairs to mRNA codon and connects it to its cognate activated amino acid during translation	(Hoagland et al., 1958)
snRNA	Component of spliceosome	(Hodnett and Busch, 1968)
snoRNA	Pre-rRNA processing	(Smith and Steitz, 1997; Tollervey and Kiss, 1997)
hnRNA	Intron containing pre-mRNA	(Edmonds et al., 1971)
miRNA	forms complex with AGO protein and targets mRNAs by base-complementarity, using 22 nt seed sequences near its 5' end, inducing mRNA deadenylation and decay or translational regulation	(Lee et al., 1993; Wightman et al., 1993)
siRNA	Induces the cleavage of a perfectly complementary target RNA	(Tomari, 2005)
piRNA	Forms complexes with PIWI proteins and silences repetitive elements in germline cells	(Siomi et al., 2011)
vaultRNA	Component of 'Vault Complex'; nucleocytoplasmic transport, intracellular detoxification, signalling, apoptosis resistance, innate immune response, DNA damage repair and nuclear pore complex formation	(Chung et al., 2005; Vanzon et al., 2005; Kim et al., 2006; Shimamoto et al., 2006; Kowalski et al., 2007; Ryu et al., 2008; Vollmar et al., 2009)
eRNA	Regulation of transcription	(De Santa et al., 2010; Kim et al., 2010)
circRNA	miRNA inhibition (molecular sponges), formation of RNP complexes to regulate cell cycle, translation regulation	(Abdelmohsen et al., 2017)
lncRNA	Regulation of transcription, translation, formation of RNP complexes	(Cech and Steitz, 2014)
Y RNA	DNA replication, cancer biomarkers	(Christov et al., 2008)
Telomerase RNA (TERC)	Provides template for telomeric DNA synthesis and scaffolds protein assembly	(Blackburn and Gall, 1978)

substantial progress has been made over the last few decades towards the functional characterisation of many classes of ncRNAs.

Housekeeping ncRNAs are primarily involved in the performance of essential cellular metabolic processes, such as mRNA splicing and translation (Figure 1.2). snRNAs, characterised in late 1960s (U1 and U2) and 1970s (U4, U5 and U6) (Hodnett and Busch, 1968; Weinberg and Penman, 1968; Lerner and Steitz, 1979) are components of the spliceosome, which removes the introns from transcribed pre-mRNAs. rRNA and tRNA are in the core of the translation machinery. rRNA was initially thought to be a structural 'scaffold' for the assembly of the ribosome. In the 1970s it was established that rRNA actually has an active role in securing translation initiation site, as shown by the base-pairing of in 16S rRNA with mRNA at a sequence 5' upstream of the start codon in *E. coli* (Steitz and Jakes, 1975). Later studies proved that rRNA is the catalytic moiety of the large ribosomal subunit (Yonath et al., 1982; Ban, 2000; Wimberly et al., 2000) and in fact 23S rRNA is a key participant in peptidyl-transferase function, can still stimulate peptide bond formation even after digestion of ribosomes with proteases (Noller et al., 1992).

Regulatory ncRNAs, which are often involved in the regulation of gene expression, have been identified in various organisms (Figure 1.2). One of the first types of regulatory RNAs to be discovered, was microRNAs (miRNAs), which were identified in *C. elegans*, almost 30 years ago (Lee et al., 1993; Wightman et al., 1993). miRNAs are 21-23nt long and derive from transcribed hairpin structures, through processing by the endonucleases Drosha and Dicer (Denli et al., 2004; Okada et al., 2009; Alarcón et al., 2015). They are particularly important in the regulation of gene expression as in most cases they interact with complementary regions of mRNAs and lead to mRNA degradation and translation repression (O'Brien et al., 2018). Endogenous short-interfering RNAs (siRNAs) were firstly discovered in plants (Tomari, 2005) but we know for more than a decade now that they are also expressed in *D. melanogaster* and *Mus musculus* (Okamura and Lai, 2008). siRNAs are thought to regulate gene expression in a similar way as miRNAs do. Another class of small regulatory non-coding RNAs are piRNAs (25-33nt long). They derive from piRNA

clusters related to transposon sequences, they form complexes with piwi proteins and are essential in silencing repetitive elements in germline cells (Siomi et al., 2011).

VaultRNAs (vtRNAs) form a different class of ncRNAs of 'intermediate length' (88-100nt) and although they have been described almost 35 years ago (Kedersha and Rome, 1986) very little was known about their function until recently. vtRNAs are integral components of the gigantic 'Vault' RNP complex (13MDa) found in the cytoplasm of most eukaryotes (Stadler et al., 2009) (Figure 1.2). Vault complex, and thus vtRNAs, are involved in a multitude of cellular processes such as nucleocytoplasmic transport, intracellular detoxification, signalling, apoptosis resistance, innate immune response, DNA damage repair and nuclear pore complex formation (Chung et al., 2005; Vanzon et al., 2005; Kim et al., 2006; Shimamoto et al., 2006; Kowalski et al., 2007; Ryu et al., 2008; Vollmar et al., 2009).

Another class of ncRNAs that was discovered early on but remained largely unexplored until recently, is circRNAs (Hsu and Coca-Prados, 1979). circRNAs are generated by back-splicing in which 5' terminus of a pre-mRNA upstream exon is non-colinearly spliced with the 3' terminus of a downstream exon (Yu and Kuo, 2019). The number of studies on circRNAs is increasing and although most of them derive from mRNAs and can be translated, they are classified as long non-coding RNAs (lncRNAs). The best characterised function of circRNAs is their interaction with miRNAs, wherein they act as 'molecular sponges' by base-pairing with the miRNA and inhibit its activity, a mechanism also employed by lncRNAs (Figure 1.2). circRNAs have some more functions in common with lncRNAs, such as their interaction with proteins and the formation of complexes that regulate the cell cycle and translation (Abdelmohsen et al., 2017).

EnhancerRNAs (eRNAs) are a more recently discovered class of ncRNAs often intersecting with lncRNAs, due to their length and structural features (Figure 1.2). Discovered in 2010 (De Santa et al., 2010; Kim et al., 2010) eRNAs are transcribed by RNA polymerase II from enhancer regions and regulate the transcription of the target gene.

Long non-coding RNAs (lncRNAs) are typically longer than 200nt and share the main structural features of mRNAs. They are transcribed by RNA polymerase II and the majority of them are 5'capped. ~50% of lncRNAs contain a poly-A tail and 98% of

human lncRNAs are spliced (Djebali et al., 2012). A common feature of non-coding RNAs is that they can act as part of a network of RNA interactions (P. Zhang et al., 2019).

Functional interactions between ncRNAs and proteins are well established since the characterisation of RNP complexes, such as ribosomes and the spliceosome (P. Zhang et al., 2019). ncRNAs interact with mRNA as well as with other lncRNAs. LncRNAs often interfere with miRNA induced RNA degradation, by base-pairing with the miRNA binding sites on the mRNA, therefore stabilising the mRNA and protecting it from degradation (Faghihi et al., 2008; Faghihi et al., 2010; Modarresi et al., 2012).

Interaction of ncRNAs with DNA plays a decisive role in the regulation of gene expression. Three well characterised examples of lncRNAs that interact with DNA are XIST (Brockdorff et al., 1992; Brown et al., 1992), MALAT1 (Ji et al., 2003) and NEAT1 (Hutchinson et al., 2007). One of the first lncRNAs to be discovered, XIST is a 17kb lncRNA, transcribed only from one of the 2 X chromosomes in human developing embryo after the stage of pluripotency and mediates the gene dosage compensation between male and female placental mammals. XIST transcript coats the chromosome from which it is transcribed and then recruits Polycomb repressive complex 2 (PRC2) histone methyltransferase, which deposits H3K27me3 on the chromosome, triggering its transcriptional silencing (Brockdorff et al., 1992; Brown et al., 1992).

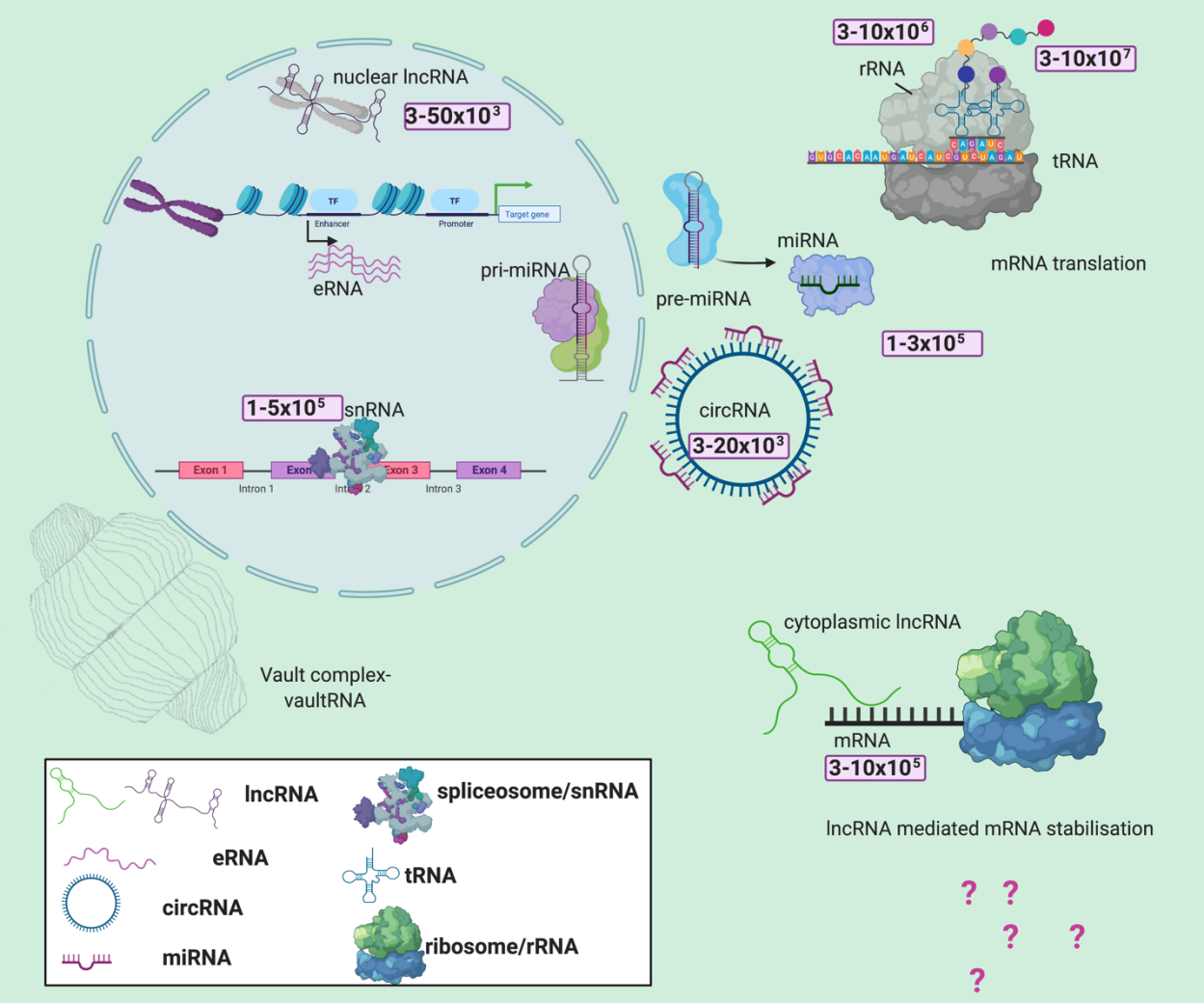


Figure 1.2: Schematic representation of the main types of non-coding RNAs in the cell. Schematic representation of the main types of non-coding RNAs withing the cell and their average number of molecules per cell (from (Palazzo and Lee, 2015)). Question marks represent the types of non-coding RNAs, yet to be discovered. Image created with BIORENDER.

1.2 LncRNAs: Molecular nature and origins

The molecular nature of lncRNAs is in many ways similar to that of mRNAs. However, they are annotated as non-coding because their RNAs do not contain large open reading frames (>100 codons) that encode protein products (Tsagakis et al., 2020). In contrast, they can contain small open reading frames, termed smORFs (Basrai et al., 1997). Our understanding of lncRNAs vastly improved over the last decade, following the advent of high throughput sequencing methods, such as RNA-Seq, CLIPseq and Ribosome profiling (Guttman et al., 2013; Chew et al., 2013). It is now possible to accurately detect and characterise lncRNAs and their interactions with other RNAs and proteins. Since the first reports of lncRNAs, almost 30 years ago (Bartolomei et al., 1991; Brockdorff et al., 1992; Brown et al., 1992), several lncRNAs have been extensively studied. Nonetheless, only a handful of them have had their functions characterised in detail. The lack of coding potential, as well as the lack of obvious phenotypes resulting from their mutation, has led to a debate as to whether lncRNAs are functional or not. Accumulating evidence of differential expression of lncRNAs during disease (Nagalakshmi et al., 2008; Cloonan et al., 2008; Ingolia et al., 2009; Thomson et al., 2011), as well as interaction of lncRNAs with key metabolic pathway components (Yoon et al., 2012) suggests that they do have functional roles within the cell.

1.2.1 Categories of lncRNAs

There are several ways to classify lncRNAs, the most common of which is by their genomic location. Based on their chromosomal position, with respect to the position of protein-coding genes, lncRNAs are broadly classified into five, non-mutually exclusive, categories. Sense lncRNAs (Figure 1.3: A), are located between or within other genes, and are transcribed from the same strand as the nearest protein-coding gene. Anti-sense lncRNAs (or Natural Antisense Transcripts, NATs) (Figure 1.3: B) are transcribed from the opposite strand of the nearest protein-coding gene, therefore can potentially base-pair to the mRNA transcribed in the ‘sense’ orientation. LncRNA genes that are located on the opposite strand to a protein-coding gene whose transcription is initiated less than 1000 bp away are considered as bi-directional (Figure 1.3: C) as they share the same promoter with the protein-coding gene. Intronic

lncRNAs are located inside the introns of protein-coding genes (Figure 1.3: D), whilst intergenic lncRNAs (Figure 1.3: E) are located in intervening regions between two protein-coding genes (Ponting et al., 2009). The majority of human lncRNA genes are either ‘intergenic’ (52%) or ‘anti-sense’ (42%) (Figure 1.4).

Classification of lncRNAs based on their genomic location is plausible and was definitely a means to initially characterise lncRNAs, when very little was known about their functions. However, the category into which a lncRNA falls does not necessarily dictate its function. For example, an anti-sense lncRNA is traditionally thought to regulate its ‘sense’ mRNA by base-pairing. However, this is not always the case and anti-sense lncRNAs,such as NOP14-AS may also have additional functions irrespective of their ‘sense’ mRNA (Goyal et al., 2017).

Although lncRNAs share some basic molecular characteristics, their length and structural features vary significantly. These features have the potential to influence the localisation and the function of a lncRNA within the cell. For example, the 5’ -m^7G cap, polyA tail and splicing contribute to a lncRNA’s ability to exit the nucleus and become part of a regulatory pathway in the cytoplasm (Cabili et al., 2011; Derrien et al., 2012).

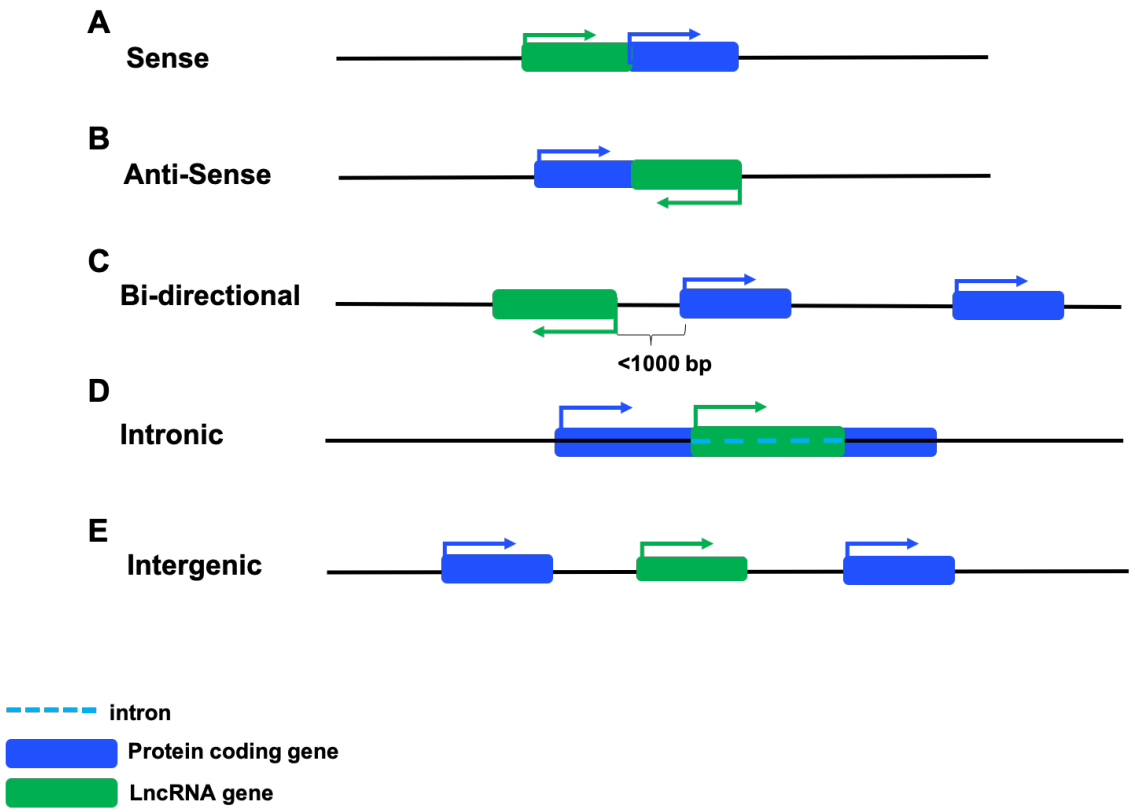


Figure 1.3: Categories of lncRNA. Types of lncRNA based on their genomic location, orientation, and relative position to nearby protein-coding genes. (A) sense, (B) anti-sense, (C) bi-directional, (D) intronic, and (E) intergenic. LncRNA genes are colour-coded in green and protein-coding genes in blue. Intronic sequences are depicted as dotted line.

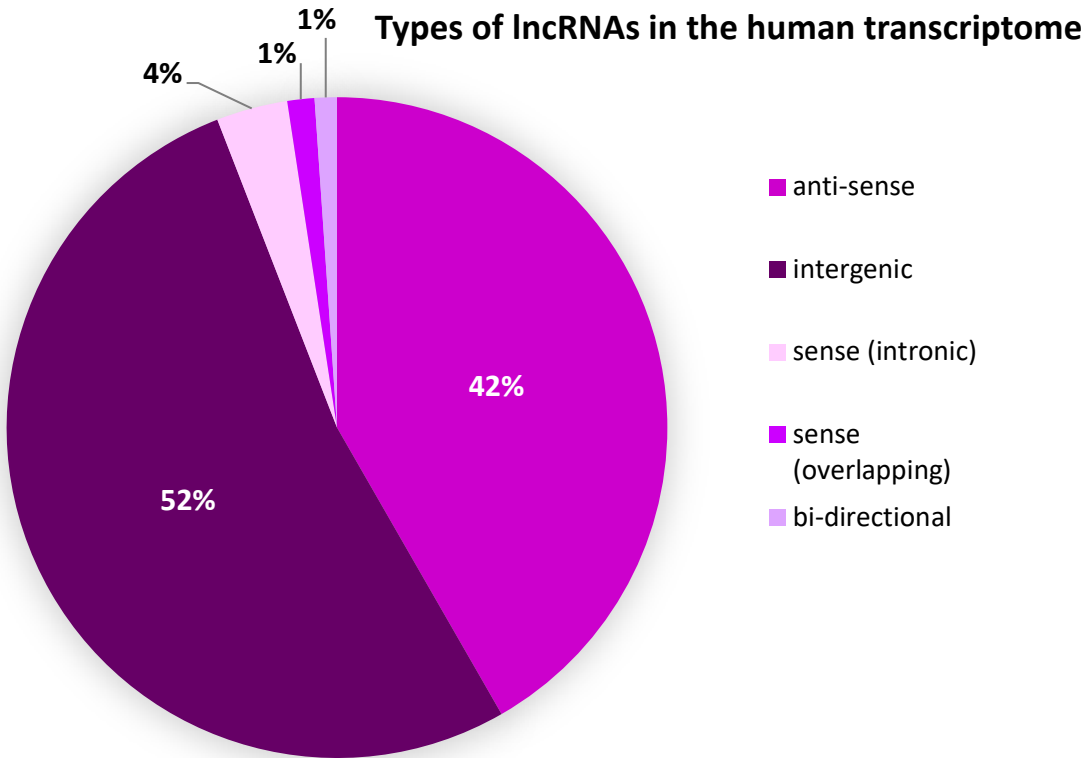


Figure 1.4: Proportion of lncRNAs present in the human transcriptome by genomic location. Annotation from Gencode January 2019 (release 29, GRch38).

1.2.2 The origins, evolution and conservation of lncRNAs

Despite the many similarities between lncRNAs and mRNAs, there is one key difference: lncRNAs are far less conserved across species than mRNAs. This initially raised some concerns regarding the functionality of lncRNAs. The lack of orthologs in other species as well as the faster rate of evolution of lncRNA exons, compared to protein-coding exons and UTRs suggest that lncRNA genes could be 'junk DNA' or 'transcriptional noise' (Marques and Ponting, 2009; Young et al., 2012). However, lncRNAs are in fact under stronger selective pressure than neutrally evolving sequences such as introns and transposable elements (Ponting et al., 2009; Haerty and Ponting, 2013). Moreover, lncRNAs exhibit high tissue and developmental stage specific expression (Ponting et al., 2009), a fact that is not consistent with neutral evolution and lack of function.

Conservation at the sequence level may not be very common, however several lncRNAs, such as XIST (Brockdorff et al., 1992; Brown et al., 1992), H19 (Bartolomei et al., 1991), HOTAIR (Rinn et al., 2007), CHASERR (Rom et al., 2019), Cerox1 (Sirey et al., 2019) and LINC00261 (Hezroni et al., 2015) show evidence of sequence conservation across eutherian mammals. LINC00261 even shows sequence homology, in its first exon, between mammals and fish and as well as syntenic homology from human to sea urchin (Hezroni et al., 2015). Interestingly, certain lncRNAs have been reported to exhibit conserved synteny as well as high sequence identity at promoter regions and exon-intron boundaries, across amniotes (from *Mus musculus* to *Gallus gallus*) (Chodroff et al., 2010; Hezroni et al., 2017). The expression of 3 murine lncRNAs conserved across amniotes, namely AK082072, AK082467, and AK043754, is restricted in the brain and appears to be developmentally regulated. Interestingly, the expression pattern of the conserved regions of each of these lncRNAs in opossum and chicken brains of early and late developmental stages is similar to that observed in the mouse brain. Furthermore, they are expressed in brain regions associated with the evolution of cerebral cortex (Molnár et al., 2006; Cheung et al., 2010) suggesting a functional role. Although more and more lncRNAs are being discovered and characterised, the identification of conserved lncRNAs is still limited by the fact that their evolution remains poorly understood.

The origin of lncRNAs is largely unknown and not all lncRNAs have emerged in the same way. It has been estimated that 5% of conserved mammalian lncRNAs derive from lost protein-coding genes (Hezroni et al., 2017). So far, there are five working hypotheses as to the origins and evolution of lncRNAs. According to the first (Figure 1.5: A), a protein-coding gene acquires frame disruptions and is transformed into a functional non-coding RNA that incorporates some previous coding sequence. This was the case with the XIST lncRNA that is critical for the inactivation of the X chromosome in eutherian mammals. It has been revealed that several XIST exons and its promoter are derived from a protein-coding gene, called Lnx3, that acquired frame- disrupting mutations, although it is not known whether this was a “one step” process or included more than one steps during evolution (Elisaphenko et al., 2008).

During vertebrate evolution, protein-coding capacity could be abolished by the insertion of a transposable element in the coding sequence (Figure 1.5: B), leading to ORF disruption and generation of a lncRNA. Moreover, it has been shown that 34% of the lncRNAs in eutherian mammals, originating from ancestral protein-coding genes, have isoforms with transcription start sites overlapping with a transposable element. Therefore, it is possible that the current promoter of a lncRNA was adopted during or after the loss of protein-coding potential (Hezroni et al., 2017). Interestingly, some rodent and human brain specific lncRNAs, namely the BC1 and BC200, as well as several anti-sense lncRNAs, have emerged following insertions of transposable elements, reviewed in (Conley et al., 2008; Ponting et al., 2009).

A third hypothesis (Figure 1.5: C) suggests that following a chromosomal rearrangement, two un-transcribed, previously well-separated sequence regions are juxtaposed and give rise to a multi-exon non-coding RNA. A dog testis-derived lncRNA (supported by ESTs BM537447, C0597044, and DN744681) appears to have arisen in that way (Ponting et al., 2009). Another possibility is that the duplication of a non-coding gene by retro-transposition can generate either a functional non-coding retro-gene or a non-functional non-coding retro-pseudo-gene (Ponting et al., 2009) (Figure 1.5: D).

The lack of sequence conservation in lncRNAs does not support lncRNA evolution from gene duplications, as is the case with protein-coding genes. However, local tandem duplications in a previously protein-coding gene may explain the neighbouring

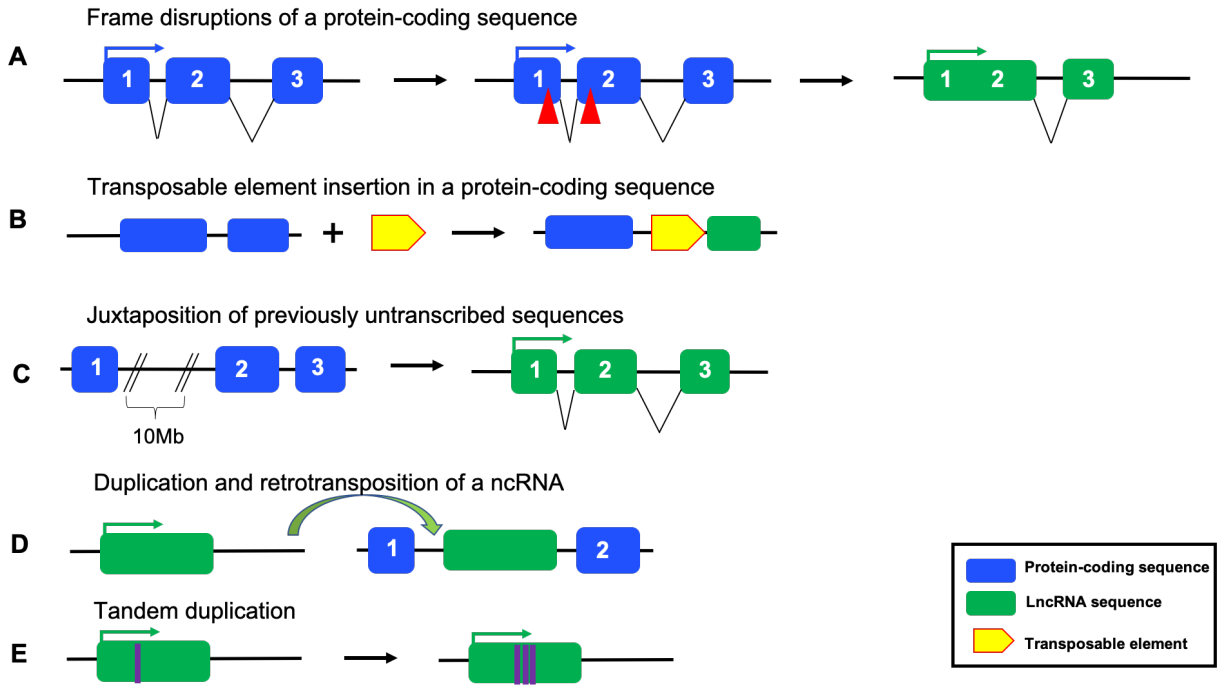


Figure 1.5: Five hypotheses explaining the origin of lncRNA genes. (A) A protein-coding gene acquires frame disruptions and becomes a noncoding RNA that incorporates some previous coding sequence. (B) Insertion of a transposable element in a previously protein-coding sequence gives rise to a lncRNA because the TE disrupts the protein-coding gene. (C) After a chromosomal rearrangement, two untranscribed and previously well-separated sequence regions are juxtaposed and give rise to a multi-exon noncoding RNA. (D) A noncoding gene duplicates by retrotransposition and generates a functional non-coding retrogene or a non-functional non-coding retropseudo- gene. (E) Neighbouring repeats within a noncoding RNA have their origins in local tandem duplication events. Based on Ponting et al., 2009 and Hezroni et al., 2017.

repeats that have been observed in some lncRNAs (Figure 1.5: E) (Ponting et al., 2009).

Conservation of lncRNAs is not limited to the nucleotide sequence level but it also exists at the structural level. Genome wide studies of RNA secondary structure over the last decade, reviewed in (Sanbonmatsu, 2016) have demonstrated that lncRNAs tend to be more structured than mRNAs (Spitale et al., 2015) and contain evolutionary conserved modular secondary structure subdomains (Wilusz et al., 2012). A study of XIST lncRNA has shown that the silencing of X chromosome is mediated by 2 conserved stem loops at its 5'-end region. These stem loop repeats are separated by AU-rich spacers whose sequence is not conserved. Further investigation showed that the length of these spacers is not important for the function of the stem loops (Wutz et al., 2002).

MALAT1 and MENβ (multiple endocrine neoplasia β) lncRNAs both possess a highly conserved triple helical structure at their 3' end that protects them from degradation. Moreover, when these structures were placed downstream from an ORF, the ORF was efficiently translated *in vivo* despite the lack of a poly(A) tail, thus proving the function of these structures as translation enhancers (Wilusz et al., 2012). Structural characterisation of the human steroid receptor RNA activator lncRNA (SRA-1), using Selective-2'-hydroxyl acylation analysed by primer extension (SHAPE), In-line, RNAse-V1 and Dimethyl-Sulfate (DMS) probing methods, has revealed that it contains four modular secondary structure subdomains, each of which containing multiple secondary structure motifs (Novikova et al., 2012). Also, an isoform of SRA has been reported to encode SRA protein (SRAP) (Kawashima et al., 2003). Structural information for one of these domains improved the ability to perform phylogenetic alignment of SRA-1 sequences across mammals and showed that the structure of the lncRNA, rather than its translational product (Novikova et al., 2012). Taken together, the above evidence suggests that secondary structure significantly contributes to the function and potentially the conservation of lncRNAs. During the last 7 years, the structures of several lncRNAs have been solved, reviewed in (Jones and Sattler, 2019) providing new insight into their evolution. Characterisation of lncRNA functions based on their structure is a novel and rapidly evolving field and it is particularly important that secondary structure is considered when assessing the evolution and conservation

of lncRNAs (Sanbonmatsu, 2016). Moreover, lncRNAs tertiary structures can be utilised as a means of lncRNA classification, in addition to their classification based on chromosomal location.

1.3 Sub-cellular localization and functions of lncRNAs

The function of a lncRNA is highly dependent on its sub-cellular localisation. Until recently, the majority of studies have focused on the characterisation of nuclear lncRNAs, as early high- throughput studies indicated that lncRNAs were predominantly enriched in the nucleus (Derrien et al., 2012). However, an increasing number of lncRNAs have now been detected in the cytoplasm (Rashid et al., 2016; Mas-Ponte et al., 2017). Enrichment of a lncRNA in either the nucleus or the cytoplasm does not preclude it from being functional in both compartments (Tsagakis et al., 2020). In fact, it is possible for lncRNAs to shuttle between the nucleus and the cytoplasm in response to signals (e.g. UCHL1-AS; (Carrieri et al., 2012)).

1.3.1 Nuclear lncRNAs

LncRNAs are transcribed and processed in the nucleus. Within the nucleus, specific lncRNAs have key roles in the regulation of transcription and splicing, and in epigenetic modifications of chromatin. The specific localisation of lncRNAs in nuclear sub-compartments e.g. NEAT1 in nuclear paraspeckles and PAPAS in the nucleolus (Tsagakis et al., 2020) suggests that their function is carried out at these locations. Therefore, nuclear retention could be mediated by certain signals. MALAT1 and BORG lncRNAs are shown to be targeted to specific locations in the nucleus as a result of nuclear localisation signals within their sequence (Miyagawa et al., 2012; Zhang et al., 2014). Particular RNA motifs have been identified that correlate with RNA retention in the nucleus (Shukla et al., 2018; Kirk et al., 2018). Inefficient splicing of lncRNAs has also been linked to nuclear retention e.g. A-ROD lncRNA (Ntini et al., 2018).

Functions of nuclear lncRNAs can be broadly divided into three categories, based on the relative position of their target gene. **“Transcription-only”** lncRNAs have no known function other than their transcription. This means that the act of transcription can affect regulatory elements overlapping the lncRNA locus.

Specifically, the transcription of some lncRNAs triggers chromatin remodelling that regulates the transcription of neighbouring protein-coding genes by facilitating access to enhancers and promoters for transcriptional machinery. Other lncRNAs, when transcribed, act to repress transcription of protein-coding RNAs by blocking RNA polymerase II elongation or their splicing and polyadenylation (Kornienko et al., 2013; Perry and Ulitsky 2016). One such example, described in yeast, is SRG1 lncRNA (Figure 1.6: A) whose transcription represses the transcription of SER3 gene by causing dense nucleosome packing over the SER3 gene promoter (Figure 1.6: A) (Martens et al., 2004).

"cis-acting" lncRNAs, which constitute a substantial fraction of lncRNAs with an attributed function, act in the vicinity of their site of their transcription, at various genomic distances (Rom et al., 2019). It seems that 'cis-acting' lncRNAs do not have a particular orientation relative to their target genes, therefore can belong to any of the categories discussed above (intergenic, anti-sense, bidirectional, sense-overlapping and sense-intronic) (Rinn and Chang, 2012; Ma et al., 2013). In some cases, the transcription start site (TSS) of the lncRNA overlaps with enhancer regions. 'Cis-acting' act either as activators of gene expression, similarly to enhancers, or as repressors. The exact mechanisms by which 'cis-acting' lncRNAs interact with transcription factors to regulate transcription are case-dependent and vary considerably, reviewed in (Gil and Ulitsky, 2020). One well characterised example of an enhancer transcribed lncRNA that regulates the expression of its target gene 'in-cis' by promoting chromatin-loops formation is CCAT1-L, which is upregulated in colorectal cancer. CCAT1-L is transcribed from a super-enhancer region ~500 kb upstream of *MYC* gene. CCAT1-L transcript localises to its site of transcription, where it directly interacts with CTCF transcription factor (Figure 1.6: B). This interaction results in the formation of chromatin loops between the *CCAT1-L* and *MYC* loci, leading to increased *MYC* transcription and enhanced tumorigenicity (Xiang et al., 2014).

Repression of gene transcription can be mediated in-cis by lncRNAs in multiple ways. Firstly, by recruitment of proteins that repress gene expression on the site of transcription. Secondly, by competition over available enhancers in the area. Thirdly, via transcriptional interference, in which the transcription of a lncRNA near the target

gene represses the target gene transcription through nucleosome remodelling or by deposition of epigenetic modifications (Gil and Ulitsky, 2020). One such example of in-cis repression is the yeast lncRNA SRG1, which upon its transcription, represses the expression of SER3 gene through nucleosome repositioning that block SER3 transcription start site (Thebault et al., 2011).

LncRNAs that act away from their site of transcription are classified as **"trans-acting"** lncRNAs (Perry and Ulitsky, 2016). Such lncRNAs act either in the nucleus or anywhere else in the cell. The most widespread trans-acting functions in the nucleus are; the recruitment of chromatin-altering complexes to specific loci in trans (Koziol and Rinn, 2010) and the binding of lncRNAs, such as MALAT1, to splicing factors, to regulate mRNA splicing events (Tripathi et al., 2010). Several 'trans-acting' lncRNAs have been found to regulate stability and translation of target mRNAs in the cytoplasm, by base-pairing with them (Wang et al., 2002; Faghihi et al., 2008; Carrieri et al., 2012; Dimartino et al., 2018) (Figure 1.6: C).

1.3.2 Cytoplasmic lncRNAs

The export of lncRNAs to the cytoplasm usually indicates a specific function, given that transport from the nucleus requires some level of activity (Tsagakis et al., 2020). Given that the majority of lncRNAs are 5'-capped and spliced, and ~50% are polyadenylated (Derrien et al., 2012), it is not surprising that they are exported to the cytoplasm. Methylation of adenosine residues within lncRNAs (N^6-methyladenosine-m6A), mediated by YTHDC1, (Roundtree et al., 2017) is also thought to promote the export of lncRNAs in the cytoplasm. LncRNA localization in the cytoplasm can also be dictated by the presence of specific transposable elements in their sequence (e.g. the endogenous retrovirus class ERVL-MaLR) (Carlevaro-Fita et al., 2016) or by external cues. For example, UCHL1-AS moves from the nucleus to the cytoplasm upon stress, induced with rapamycin in mouse dopaminergic MN9D cells (Carrieri et al., 2012). Cytoplasmic lncRNAs take part in the regulation of several different processes within the cytoplasm. Their known actions to date can be broadly classified into 5 distinct types, although certain lncRNAs may act in more than one way, depending on the context.

LncRNAs in the cytoplasm can interact with RNA binding proteins (RBPs) and/or mRNAs. In some cases, they compete with mRNAs for the binding of a specific protein (Figure 1.6: D). Murine lnc-31, which is required for myoblast proliferation, was recently reported to interact with Y-box 1 (YB1) protein as well as Rock1 mRNA and to regulate its translation. During myoblast proliferation, lnc-31 binds to YB-1 and inhibits its proteasome-mediated degradation. Lnc-31 also base-pairs to Rock1 mRNA. As a result, Rock1 translation is promoted via its interaction with lnc-31 and YB-1 protein. In differentiated myotubes, lnc-31 is not expressed in sufficient levels to perform the same function, therefore leading to the downregulation of Rock1 expression (Dimartino et al., 2018).

Some lncRNAs harbour miRNA-binding sites. Therefore, they can act as 'molecular sponges', sequestering miRNAs. This leads to the protection of specific mRNAs from miRNA-mediated degradation. (Figure 1.6: E). A recently described 'miRNA sponge' is MACC1-AS1, which contains binding sites for some tumour suppressor miRNAs (miR-145-3p and miR-384). The binding of these miRNAs prevents them from repressing the translation of their target mRNAs, thus leading to abnormal cell growth. Further investigation revealed that MACC1-AS1 is stabilised in the cytoplasm through its interaction with PTBP1 protein (X. Zhang et al., 2019). According to the above study, MACC1-AS1 is an interesting example of an anti-sense lncRNA whose apparent main function is in the cytoplasm, rather than the nucleus.

Some cytoplasmic lncRNAs regulate the translation of mRNAs and in some cases, they also undergo translation (Figure 1.6: F, G). The association of cytoplasmic lncRNAs with components of the translation machinery, such as the 43S preinitiation complex (e.g. BC200) and polysomes (e.g. UCHL1-AS), acts to either impede or facilitate mRNA translation (Wang et al., 2002; Carrieri et al., 2012; Carlevaro-Fita et al., 2016). Ribosome profiling in human, mouse, fly, and yeast, has detected lncRNAs in ribosome-bound complexes (Ingolia et al., 2009; Duncan and Mata, 2014; Aspden et al., 2014)(Douka et al., unpublished data). This RNA-Seq–based method, which accurately detects translation events by revealing the exact sites of ribosome decoding, has revealed that smORFs <100 codons present in some lncRNAs can be translated. This should not be surprising, given that many cytoplasmic lncRNAs possess the same molecular characteristics as mRNAs that facilitate translation. The

mechanistic details of the interaction of lncRNAs with the translation machinery will be discussed later in the chapter (1.5.1 and 1.5.2).

1.3.3 Localisation of lncRNAs in cytoplasmic compartments and extracellular vesicles

Cytoplasmic lncRNAs can be specifically localised to organelles or other cytoplasmic compartments. Of interest, some cytoplasmic lncRNAs are encoded by mitochondrial DNA, and are therefore found and operate in the mitochondria (Mercer et al., 2011; Sirey et al., 2019). One such example is Cerox1 (Cytoplasmic endogenous regulator of oxidative phosphorylation 1) lncRNA, which is abundantly expressed in the mitochondria and conserved across placental mammals. Cerox1 regulates NADH:ubiquinone oxidoreductase (complex I) activity by co-ordinately regulating the abundance of at least 12 complex I transcripts, via an miRNA (miR-488-3p) dependent mechanism (Sirey et al., 2019).

LncRNAs can also participate in the formation of P- bodies and can be present in extracellular vesicles (EVs). The latter act as ‘cellular messengers’, since they are secreted from one cell and received by another (Figure 1.6: H). It has been proposed that lncRNA molecules with relatively low expression levels (e.g. lincRNA-p21, HOTAIR) are highly enriched in exosomes. The level of secretion of certain lncRNAs can be correlated with the maintenance of cell homeostasis, as it has been proved that lincRNA-p21 and ncRNACCND1 exosome levels reflect the cellular response to DNA damage (Gezer et al., 2014). Therefore, secreted lncRNAs can easily be measured in bodily fluids and ustilised as potential biomarkers for disease diagnosis. One example is lincRNA-p21 transcript, whose levels in exosomes isolated from urine samples of prostate cancer patients, appear to be significantly elevated compared to those of patients with benign prostatic hyperplasia (Isin et al., 2015).

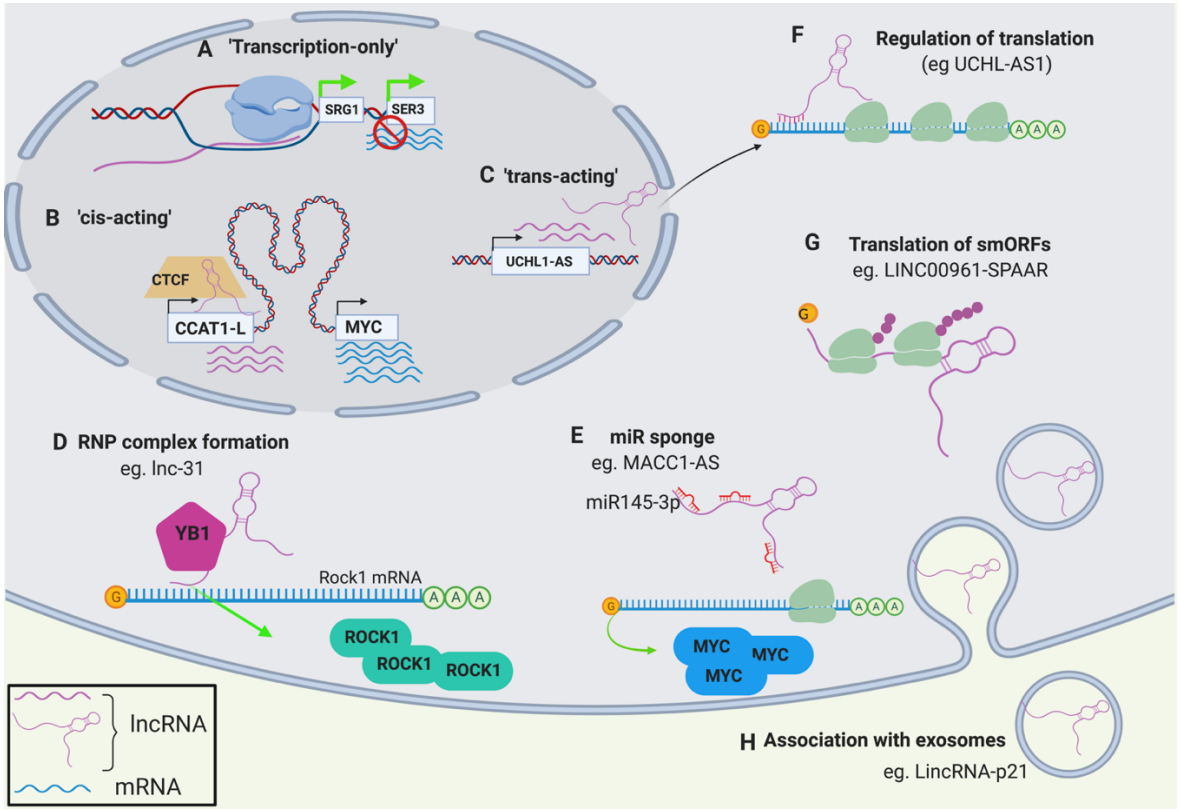


Figure 1.6: Representative functions of lncRNAs in the nucleus and cytoplasm: Generally, lncRNAs either function in either the nucleus (A-C) or in the cytoplasm (D-H). (A) 'Transcription-only' lncRNAs regulate the transcription of other genes by altering the chromatin during their transcription. (B) 'Cis-acting' lncRNA CCATL-1 occupies its transcription site and enhances MYC transcription by interacting with CTCF transcription factor and forming chromatin loops. (C) 'Trans-acting' lncRNAs exert their functions away from their site of transcription. (D) Lnc-31 interacts with YB1 protein and stabilises it and also binds Rock1 mRNA. Lnc-31 mediated YB1 stabilisation facilitates the translation of Rock1. (E) MACC1-AS acts as a miRNA sponge for miR145-3p, preventing it from targeting MYC mRNA, leading to increased MYC translation. (F) UCHL1-AS interacts with polysomes and facilitates the translation of UCHL1 mRNA. (G) LINC00961 is translated to a peptide called SPAAR. (H) LincRNA-p21 is secreted from the cells via exosomes.

1.4 Eukaryotic translation

Translation in eukaryotes, as well as in prokaryotes, is comprised by four main stages, initiation, elongation, termination and recycling (Kapp and Lorsch, 2004). The basic goals of translation are highly conserved between eukaryotes, bacteria and archaea (Schuller and Green, 2018). However, the steps and the machinery required to accomplish these goals are much more complicated in eukaryotic cells (Kapp and Lorsch, 2004). During translation initiation the assembly of an 80S ribosome, with a methionyl-tRNA bound in its P site, at the AUG start codon is guided by several eukaryotic translation initiation factors (eIFs) (Figure 1.7) (Jackson et al., 2010; Schuller and Green, 2018). Subsequently, an aminoacyl-tRNA enters the acceptor site (A site) of the ribosome, where decoding takes place and elongation begins. The ribosome catalyses the formation of a peptide bond between the first and the second amino-acid, the peptide chain starts to form and the first methionyl-tRNA is released from the exit site (E site). Elongation proceeds as the ribosome moves progressively across the mRNA three nucleotides per step, in a 5' to 3' direction, synthesising the encoded peptide. During this process a new aminoacyl-tRNA, carrying the next amino-acid, binds to the A site (decoding), a peptide bond forms between the last and the new amino-acid and the peptide chain is loaded to the new tRNA, which translocates to the P site (Figure 1.7) (Kapp and Lorsch, 2004; Jackson et al., 2010; Schuller and Green, 2018).

Elongation is co-ordinately regulated by aminoacyl-tRNAs and eukaryotic elongation factors (eEFs) (Kapp and Lorsch, 2004; Schuller and Green, 2018). Therefore, as the ribosome moves to the next codon, the P site is always occupied by the tRNA carrying the peptide chain and the A site accepts the next aminoacyl-tRNA. When the ribosome reaches the end of the ORF, the stop codon is recognised by the eukaryotic peptide chain release factors (eRFs) which promote the release of the newly synthesised peptide chain from the tRNA and the ribosome. Following that the 80S ribosome complex is recycled by the ATP-binding cassette (ABCE1) and eIF3, eIF1 and eIF1A into free 60S and 40S subunits, ready to begin a new round of translation (Jackson et al., 2010; Schuller and Green, 2018).

Each step of the translation process needs to be tightly regulated to ensure the faithful decoding of the mRNA in the appropriate time. Initiation is a crucial step as it dictates which mRNAs will be translated at a given time and cell type. There are two main types of initiation in eukaryotes: the canonical 5' cap-dependent initiation and the non-canonical IRES-mediated (cap-independent) translation initiation (Jackson et al., 2010). Canonical translation initiation entails six distinct steps. Step 1 is the formation of the 43S pre-initiation complex when the ternary complex (eIF2-GTP-$tRNA_i^{Met}$) binds the 40S subunit-eIF3 complex. Step 2 comprises the attachment of the 43S pre-initiation complex to the 5'-cap proximal region of the mRNA, facilitated by the cooperative action of eIF4F and eIF4B. In step 3 the 43S complex scans the mRNA sequence, downstream of the 5' cap to find the initiation codon. Step 4 is the initiation codon recognition. In step 5 the ribosome is committed to the start codon. During step 6 the 60S ribosomal subunit joins the 40S-mRNA complex and initiation proceeds (Jackson et al., 2010). About 45-50% of mammalian mRNAs have at least one short uORF (Resch et al., 2009). In these cases, ~50% of the ribosomes translate the uORF and then resume scanning and reinitiate at downstream sites (Jackson et al., 2010). Several factors (summarised in Figure 1.7: A) can affect the progression of translation initiation and the mechanisms of initiation regulation broadly fall into two main categories: the mechanisms that impact on eIFs or ribosomes (e.g. phosphorylation of eIF2) and those that impact on the mRNA (e.g. miRNAs) (Jackson et al., 2010).

Successful initiation does not guarantee that translation will be seamlessly completed. Elongating ribosomes can encounter various problems at each step of the elongation process, which force them to pause (Schuller and Green, 2018) (Figure 1.7: B). For example, the presence of secondary structures such as stem-loops or pseudo-knots on mRNA impedes the movement of the ribosome, therefore causing ribosome stalling or proceeding to programmed ribosome frameshifting (Dinman, 2012). Codon optimality is also important for elongation. An 'optimal' codon is a codon which is preferentially used, relative to others to encode an amino-acid and therefore has a pool of tRNAs readily available for elongation (Sabi and Tuller, 2014). The encounter of rare or sub-optimal codons prevents the first step of elongation (tRNA selection) and causes

ribosome stalling. Certain amino-acid structures, such as that of proline, are particularly challenging as substrates for A or P sites (Pavlov et al., 2009) and hinder the peptide bond formation as well as the translocation steps. In eukaryotes the challenge of Pro-Pro bond formation is dealt with by eIF5A (Gutierrez et al., 2013) by stabilizing the conformation of the peptidyl-tRNA for nucleophilic attack by the aminoacyl-tRNA in the A site (Schmidt et al., 2016).

Translation termination is mainly catalysed by eRF1, which has a similar structure to that of tRNA, in complex with the GTPase eRF3. eRF1 recognises the stop codons (UGA, UAG, UAA) at the end of the ORF and promotes the release of the nascent peptide (Schuller and Green, 2018) The efficiency of translation termination is also affected by the sequence context. More specifically, 18S rRNA interacts with the first nucleotide of the 3'-UTR (Brown et al., 2015; Matheisl et al., 2015). Reporter assays in yeast and mammalian cells have shown that specific stop codons and the identity of the first nucleotide of the 3'-UTR are more or less likely to promote stop codon read through (Floquet et al., 2012; Beznosková et al., 2015). The interaction of 18S rRNA with the first base of the 3'-UTR is in line with the observation that the size of ribosome protected fragments (RPFs) at stop codons is one nucleotide longer than RPFs on sense codons (Ingolia et al., 2011).

Translation is a highly regulated and resource-heavy procedure. The mechanistic details of translation regulation elucidated so far focus on the interactions of the ribosomes and initiation/elongation/release factors with the mRNA but far less is known about the potential role of lncRNAs in the regulation of translation. Over the last 20 years, some studies have shown that lncRNAs can play key roles in the regulation of translation initiation either by interacting with translation initiation factors, such as eIF4A (Wang et al., 2002), or by interacting with specific sequences of the mRNA (Carrieri et al., 2012).

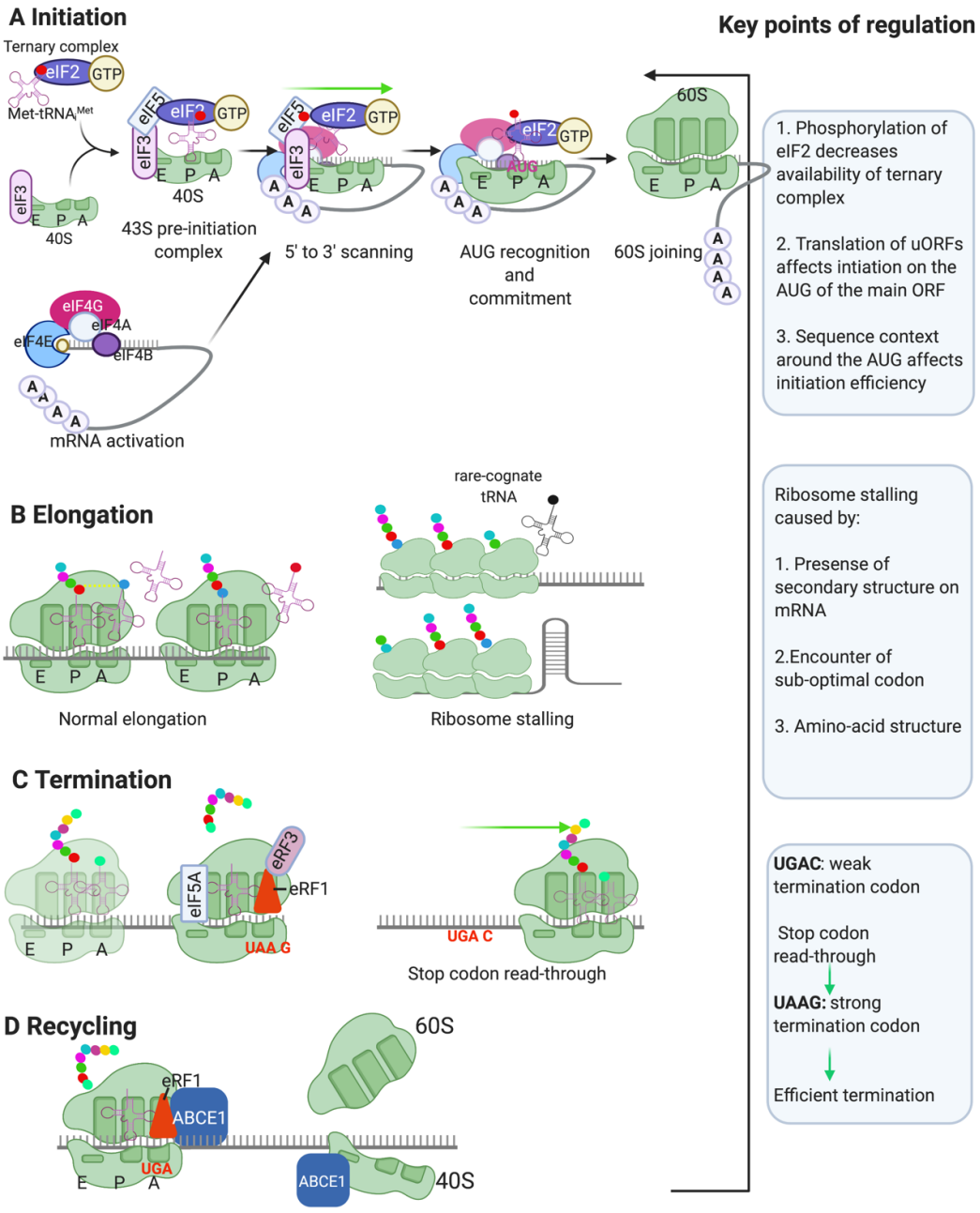


Figure 1.7: Brief schematic of the 4 fundamental stages of translation in eukaryotic cells. A. Initiation requires the formation of the 48S pre-initiation complex, consisting of the 40S subunit bound by eIF3 and the ternary complex. The pre-initiation complex scans across the 5'-UTR of the mRNA to find the AUG start codon B. Elongation follow successful initiation and is comprised by 3 main steps, tRNA selection, peptide bond formation and translocation of tRNA/mRNA. Ribosome stalling can be caused by the presence of secondary structures on the mRNA, the

encounter of a sub-optimal codon or the complex structure of an amino-acid. C. Termination is catalysed by eRF1, which recognises the stop codon and releases the peptide chain. D. Ribosomal subunits recycling is catalysed by ATP-binding cassette subfamily 1 (ABCE1) which remains bound to the 40S to stimulate a new translation initiation. Created based on Jackson et al., 2010 and Schuller and Green 2018. Image created using BIORENDER.

1.5 LncRNAs and translation

The lack of large open reading frames within lncRNAs has led researchers to annotate them as such and to overlook the possibility that they are translated. However, the detection of lncRNAs in ribosome profiling datasets (Ingolia et al., 2011) gave rise to one of the biggest debates in RNA research: can lncRNAs encode small functional peptides? About a decade ago, a comparative analysis of mass spectrometry and RNA-Seq data from 2 human cell lines (K562 and GM12878), produced by the ENCODE project, revealed that ~8% of the GENCODEv7 lncRNAs can encode peptides. Specifically, the study reports 85 unique peptides encoded by 69 lncRNAs (Banfai et al., 2012).

Around the same time, a ribosome profiling study, coupled with an algorithm that defined the exact sites of translation initiation, identified highly translated smORFs within lincRNAs, termed short, polycistronic, ribosome-associated coding RNAs (sprcRNAs) (Ingolia et al., 2011). The study reported that the ribosome occupancy of most lincRNAs resembles that of protein-coding genes. However, a follow-up study that re-assessed these data, this time comparing the ribosome occupancy of lncRNAs to that of 5'-UTRs as well as other classical ncRNAs, demonstrated that lncRNAs show similar ribosome occupancy to ncRNAs and 5'-UTRs. Therefore, the authors argue that ribosome occupancy alone is not sufficient to classify transcripts as coding or noncoding and other (Ingolia et al., 2013).

Since then, several ribosome profiling studies (Duncan and Mata, 2014; Aspden et al., 2014; Chen et al., 2020), have provided evidence that some lncRNAs encode small peptides. Some of these small peptides have been found to possess functions (Magny et al., 2013; Anderson et al., 2015; Szafron et al., 2015; Lewandowski et al., 2019; van Heesch et al., 2019; Spencer et al., 2020; L. Wang et al., 2020; Chong et al., 2020). Other lncRNAs have been found to play key roles in the regulation of translation, by functionally interacting with components of the translation machinery (Wang et al., 2002; Carrieri et al., 2012; Carlevaro-Fita et al., 2016).

1.5.1 Functional interactions of lncRNAs with the translation machinery

Evidence of translational regulation by lncRNAs emerged almost 20 years ago. The rodent lncRNA BC1 as well as its primate counterpart BC200, both expressed in dendrites, were shown to repress local translation. They interact with eIF4A and inhibit the formation of the 48S initiation complex by preventing the binding of eiF4F complex to the mRNA (Figure 1.8: A) (Wang et al., 2002). LncRNAs can also regulate the association of mRNAs with polysomes and therefore, their translation. UCHL1-AS lncRNA, which exhibits sequence conservation between mouse to human, promotes the translation of its anti-sense neuronal mRNA transcript UCHL1 (ubiquitin carboxy-terminal hydrolase L1) (Carrieri et al., 2012). This enhancement of translation is mediated by the interaction of UCHL1-AS 5'-end sequence with its complementary sequence on UCHL1 mRNA as well as a SINEB2 repeat element within the UCHL1-AS lncRNA sequence. Interestingly, UCHL1-AS is enriched in the nucleus, but is exported to the cytoplasm upon stress induced by rapamycin treatment. Although cap-dependent translation is generally inhibited by rapamycin, UCHL1 mRNA translation is enhanced via its interaction with UCHL1-AS (Figure 1.8: B) The authors identified 31 other mRNA/lncRNA sense-antisense pairs with the same structural features and proposed lncRNA mediated translation as a mechanism to maintain synthesis of pro-survival proteins, such as UCHL1, involved in cellular response to stress (Carrieri et al., 2012).

In addition to these 'case-specific' studies, over the last few years several transcriptome-wide studies have identified dynamic interactions of lncRNAs with the translation machinery, during cancer (Carlevaro-Fita et al., 2016), as well as during cell differentiation (Dallagiovanna et al., 2017; Zeng and Hamada, 2018; Pereira et al., 2020). In the K562 human myelogenous leukaemia cell line, ~70% of cytoplasmic lncRNAs have more than 50% of their cytoplasmic copies associated with polysomes. Surprisingly those polysome-associated lncRNAs have significantly higher expression levels and exhibit a more homogeneous expression pattern across different cell lines, compared to free cytoplasmic and nuclear lncRNAs (Carlevaro-Fita et al., 2016). This suggests that the interaction of lncRNAs with polysomes is not a random event but rather a means to potential function. It has been previously suggested that lncRNAs associated with

ribosomes become targets of the nonsense-mediated decay pathway (Tani et al., 2013). This is supported by the stabilisation of polysome-associated lncRNAs, following treatment with drugs that induce ribosome-stalling, such as cycloheximide and emetine (Carlevaro-Fita et al., 2016).

An analysis of ribosome profiling data from human and mouse has revealed that on average, 40% and 48% of expressed lncRNAs interact with ribosomes in human and mouse, respectively. Futhermore, the lncRNAs that associate with ribosomes tend to be more enriched in the cytoplasm (Zeng and Hamada, 2018). The interaction of lncRNAs with polysomes is specifically important for adipocyte differentiation (Dallagiovanna et al., 2017). LncRNAs in human adipose derived stem cells (hASCs) associate with polysomes in undifferentiated cells as well as during early differentiation to adipocytes. The level of polysome association of lncRNA is comparable to that of mRNAs. Moreover, it is dynamic during the early stages of adipogenesis (Dallagiovanna et al., 2017). A recent study demonstrates that interaction of lncRNAs with the translation machinery is important for cardiomyogenesis (Pereira et al., 2020). Polysome-associated lncRNAs exhibit differential expression across the 5 different stages of cardiomyocyte differentiation, suggesting that the differential association with polysomes is important in this process. Interestingly, the association with polysomes is independent of the transcript abundance. The majority of the upregulated lncRNAs in early differentiation were polysome-associated (Pereira et al., 2020), suggesting that the association of lncRNAs is implicated in the process of differentiation.

The mechanistic details of how lncRNAs associate with ribosomes and polysomes still remains elusive. However, some lncRNA sequence features, such as the high GC content, transcript length and the presence of long terminal repeat sequences (LTRs) have been positively correlated with ribosome association (Zeng and Hamada, 2018). Overall, it is evident that lncRNAs interact with the translation machinery and can function as translational regulators, regardless of their own coding potential. However, it is important to be able to distinguish between the lncRNAs that simply associate with polysome complexes, as part of a regulatory mechanism and those that are being actively decoded. This will enable us to better understand the different populations of lncRNAs in

the cytoplasm and uncover transcripts mis-annotated as non-coding or transcripts that have both coding and non-coding functions (e.g SRA; (Kawashima et al., 2003)).

1.5.2 Active translation of lncRNAs

Essentially, any sequence with a start and a stop codon in frame is an Open Reading Frame (ORF) which can potentially be translated. LncRNAs contain several small ORFs (smORFs) Therefore, it should not be surprising for a smORF within a lncRNA to be translated. Of course, one should consider some more parameters before characterising a lncRNA as translated. In eukaryotes, the Kozak consensus GCC(A/G)CC**ATG**G sequence surrounding the start codon provides the necessary context for the scanning 43S pre-initiation complex to recognise the start codon and is required for proper translation initiation (Kozak, 1987). Hence, the calculation of a context score for a given start codon is a good proxy for predicting the translation of an ORF (Zeng and Hamada, 2018). The presence of a poly-A tail in a lncRNA, as well as the 5'-cap is also key for its translation (Jackson et al., 2010).

It has been argued that ribosomes can scan or engage with lncRNA transcripts and initiate translation at multiple sites, without necessarily producing a peptide (Guttman et al., 2013). This would result in the co-sedimentation of a lncRNA with 80S ribosomes (unproductive formation of 80S complexes) or small polysome complexes, even though the lncRNA is not translated. During translation termination of protein-coding transcripts, the ribosome is actively disassembled and recycled, upon the encounter of a stop codon (Jackson et al., 2010). Based on that, a metric called Ribosome Release Score (RRS), which detects the termination of translation at the end of an open reading frame was developed to distinguish genuine translation events from ribosome occupancy that does not result in translation (Guttman et al., 2013).

Classic ribosome profiling (Ribo-Seq) (Figure 1.8), first developed about a decade ago (Ingolia et al., 2009) is an extremely powerful method that has revealed unanticipated mechanistic details of translation, such as the initiation of translation from an unannotated upstream start codons, stop-codon read-through, use of non-canonical start codons and the translation of short open reading frames (smORFs, uORFs) (Jackson and Standart,

2015). However, the original method does not distinguish between RNAs engaged by an 80S ribosome (monosome) and those engaged by multiple ribosomes (polysomes). As previously discussed, an RNA can be sporadically bound by ribosomal subunits or 80S (monosomes) without necessarily undergoing elongation (Guttman et al., 2013; Aspden et al., 2014). Previously, a study in yeast has proved that translation can be inhibited after the formation of an 80S complex (Balagopal and Parker, 2011). Moreover, the monosome fraction in HEK293 cells has been found to contain a large number of inactive 80S ribosomes that are not bound to mRNAs to direct translation (Liu and Qian, 2016). Therefore, in order to determine with confidence whether a lncRNA is translated, a way to distinguish between 80S-bound and polysome-bound mRNAs and lncRNAs is required.

Poly-Ribo-Seq (Figure 1.8) was developed as an improvement of ribosome profiling, aiming to more accurately measure active translation (Aspden et al., 2014). This is achieved by separating 80S ribosomes from polysomes by fractionation, prior to RNaseI digestion. Moreover, Poly-Ribo-Seq distinguishes the actively translated RNAs amongst those that are associated with the polysomes but not necessarily translated (Aspden et al., 2014). Therefore, it has the potential to determine if lncRNAs interact with the translation machinery and identify those that are translated. Sometimes translation could be important for regulation of a downstream ORF (Geballe and Morris, 1994). Another factor limiting the ability to detect the translation of small ORFs in lncRNAs is the detection of small peptides. Mass spectrometry (MS) for small peptides is limited by the low capacity for accurate detection of peptides smaller than 20aa (Banfai et al., 2012), the fact that lncRNAs are expressed in low levels, compared to protein-coding genes, as well as the standard protein sequence databases that only contain annotated proteins (Chong et al., 2020). Nevertheless, recent studies report protein sequences derived from the translation of ORFs identified from ribosome profiling, in MS-based searches (Aspden et al., 2014; Ingolia et al., 2014; Pearson et al., 2016; Erhard et al., 2018; Laumont et al., 2018). More and more lncRNA-encoded small peptides are being discovered and some of them are suggested to be of vital importance for organismal function (Magny et al., 2013; Anderson et al., 2015).

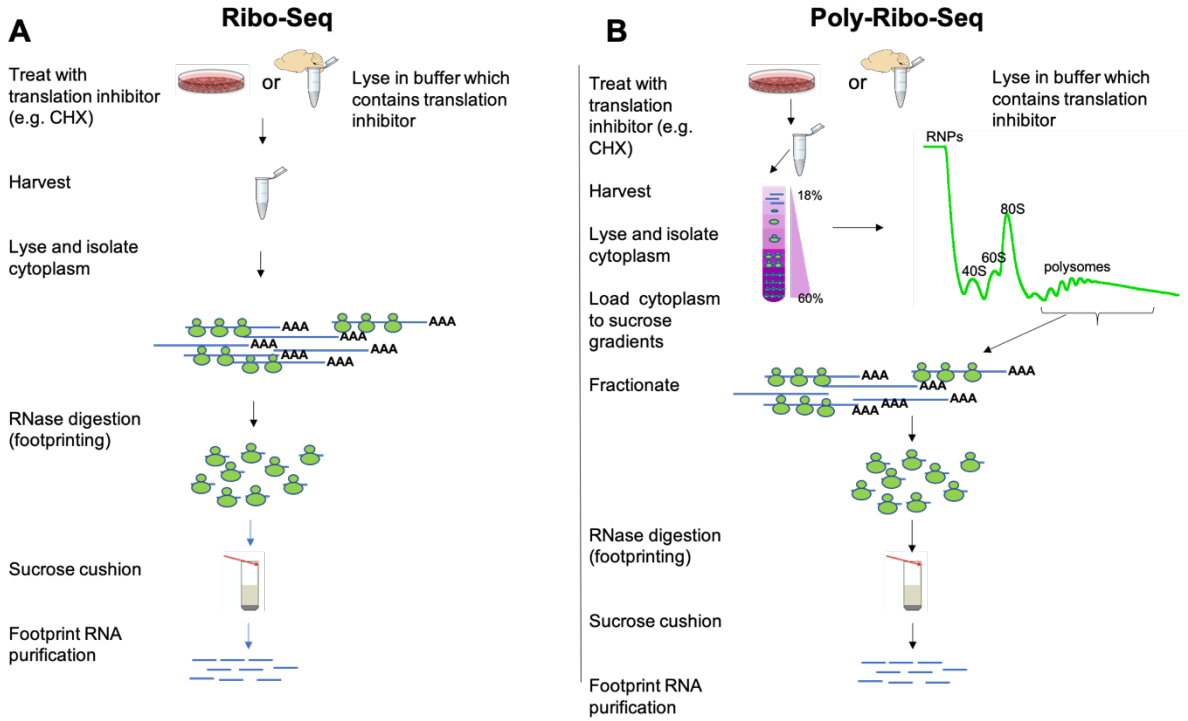


Figure 1.8: Poly-Ribo-Seq is the Ribo-Seq of polysomes. Schematic depicting the conceptual difference between (A) Ribosome profiling (Ribo-Seq) and (B) Poly-Ribo-Seq. Cells or tissue are lysed in the presence or not of a translation inhibitor (e.g. cycloheximide, CHX) and cytoplasm is isolated. In Ribo-Seq (A) the whole cytoplasmic lysate containing the ribosome bound is subjected to RNaseI digestion followed by ultracentrifugation on a sucrose cushion and is subsequently processed for sequencing. In Poly-Ribo-Seq (B), cytoplasm is subjected to polysome fractionation prior to RNaseI treatment and only the polysome fractions (corresponding to actively translating ribosomes) are treated with RNaseI. Subsequently, ribosome protected fragments are sedimented through a sucrose cushion and are processed for sequencing.

1.5.3 The production of functional peptides from lncRNA smORFs

The translation of small ORF from annotated lncRNAs was first discovered in *D. melanogaster* over a decade ago. Pri/tarsaless(tal) and sarcolamban small peptides have been extensively studied in *D. melanogaster* and are shown to be important regulators of developmental processes and cardiac muscle contraction respectively (Galindo et al., 2007; Kondo et al., 2007; Pueyo and Couso, 2008; Magny et al., 2013). 'Polished rice' (pri) or tarsal-less (tal), previously annotated as non-coding RNA in *D. melanogaster melanogaster*, is transcribed into a polycistronic mRNA that contains smORFs encoding for small peptides of 11 and 32 amino acids, highly conserved in other *D. melanogaster* species (Galindo et al., 2007; Kondo et al., 2007). Pri is expressed in all epithelial tissues during *D. melanogaster* embryogenesis and pri mutant flies exhibit abnormalities in F-actin organization during denticle formation, development of the tarsal region, as well as in the development of the trachea. All those abnormalities can be rescued by the ectopic expression of the pri-smORFs in a non-cell autonomous manner (Kondo et al., 2007). Pri/tal peptides (Figure 1.9: C) also regulate leg development in *D. melanogaster* in a non-cell autonomous manner, by triggering a cell signal that results in the formation of second, third and fourth tarsal segments (Pueyo and Couso, 2008).

The lncRNA Putative noncoding RNA 003 (pncr003), expressed in somatic muscles and the post-embryonic heart of *D. melanogaster*, has been discovered to encode two functional peptides of 28 and 29 aa. Pncr003 peptides localise near the sarco-endoplasmic reticulum membrane in muscle cells and is thought to play a pivotal role in Ca^{2+} trafficking at the sarcoplasmic reticulum (SR). Pncr003 peptides exhibit structural homology to human *sarcolipin* (*sln*) and its longer paralogue *phospholamban* (*pln*) and conservation analysis revealed a common origin. Therefore, pncr peptide is termed *sarcolamban* (*scl*) (Magny et al., 2013). LncRNA encoded micropeptides are involved in Ca^{2+} metabolism in mammalian muscle cells as well. Myoregulin (MLN) is encoded by the muscle specific lncRNA LINC00948 in humans and AK009351 in mice and is conserved across mammals. MLN forms a type II transmembrane a-helix (Anderson et al., 2015) and the transmembrane regions are conserved between human *phospholamban* (PLN) and *sarcolipin* (SLN) as well as the invertebrate ortholog

sarcolamban (scl) (Magny et al., 2013). MLN regulates Ca^{2+} transport in a similar way to PLN and SLN, by lowering the affinity of the SERCA pump for Ca^{2+}, and therefore inhibiting its reuptake by the SR (Anderson et al., 2015).

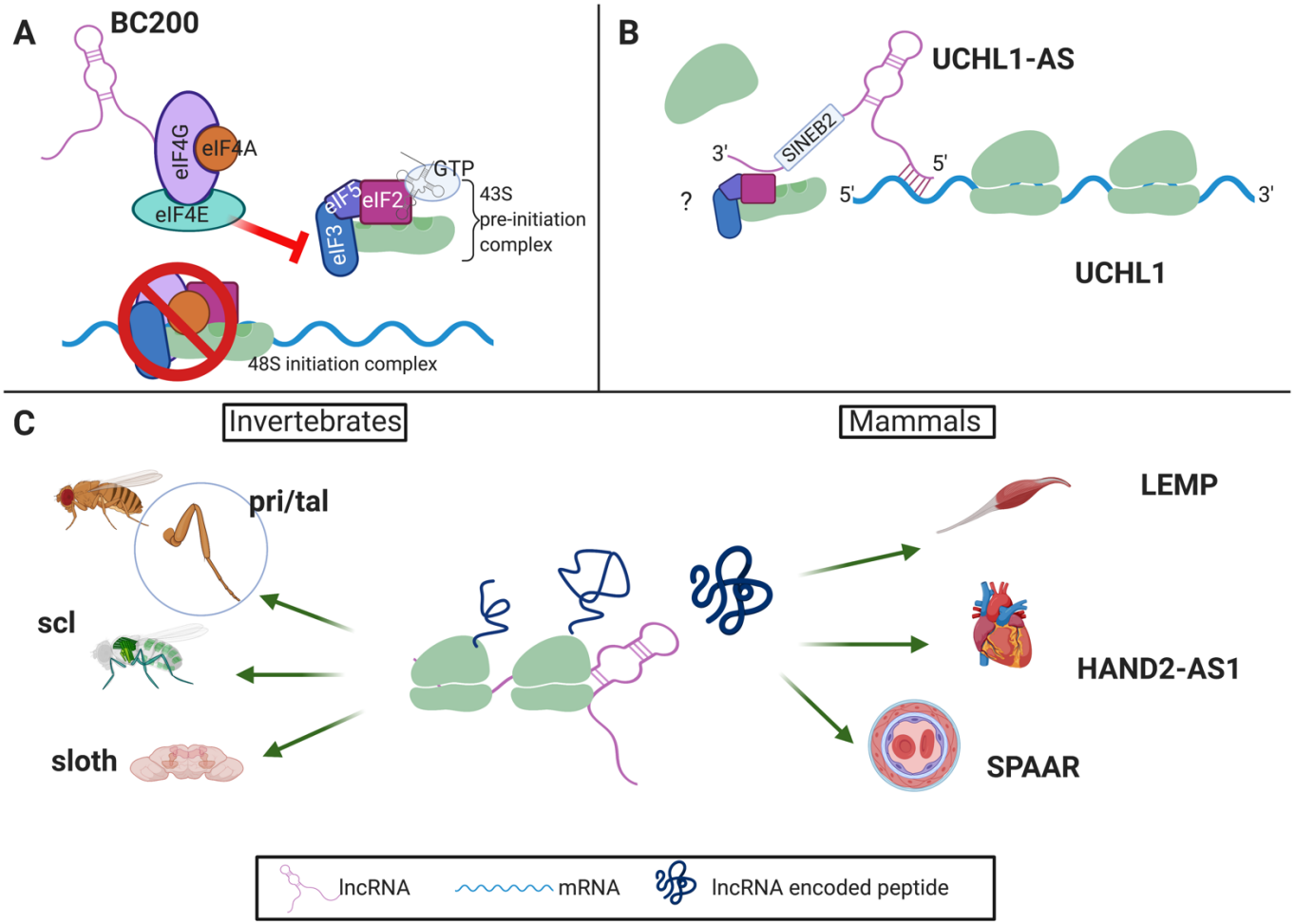


Figure 1.9: LncRNAs interact with translation machinery in different ways. (A) LncRNAs (e.g. BC200) can inhibit translation initiation by preventing 48S complex formation. (B) LncRNAs (e.g. UCHL1-AS) can enhance the translation of an mRNA by base-pairing with it and recruiting ribosomes to it. (C) LncRNAs can be translated and produce small peptides that have regulatory functions in different tissues both in invertebrates and mammals.

Recently, two neuronal small peptides, sloth1 and sloth2, (Figure 1.9: C) were discovered to be translated from the same bicistronic transcript in *D. melanogaster* (Bosch et al., 2020). Sloth1 and sloth2 are located 5 nucleotides apart from each other and are thought to be paralogs because they exhibit sequence similarity and both contain a predicted single transmembrane domain. Notably, both peptides are 50% conserved between *D. melanogaster* and human. Interestingly, the translation of sloth1 inhibits translation of sloth2, but due to the fact that sloth1 is translated sub optimally, balanced translation of both smORFs occurs and the 2 peptides are functionally non-redundant. Loss of both Sloth1 and 2 results in defects in synaptic communication between photoreceptor and post-synaptic neurons as well as neurodegeneration. Sloth1 and 2 peptides localise in mitochondria and, potentially acting as a complex, are involved in ATP production (Bosch et al., 2020).

A recent study of the human and mouse heart translatomes (van Heesch et al., 2019) identified several small peptides encoded by 27 human and 5 mouse lncRNAs, amongst them NEAT1 (Clemson et al., 2009), DANCR (Kretz et al., 2012), BANCR (Flockhart et al., 2012), GATA6-AS1 (Zhu et al., 2018), HAND2-AS1 (Anderson et al., 2015) and CRNDE (Graham et al., 2011). CRNDE was reported to produce a small nuclear peptide expressed in high proliferative cancer tissues (Szafron et al., 2015). NEAT1, GATA6-AS1 and HAND2-AS1 sequences are syntenic between human and rodents and the smORFs are translated in both human and rodent hearts (Figure 1.9: C) (van Heesch et al., 2019).

Small peptides encoded by lncRNAs have also been found to be involved in the regulation of cell differentiation in mouse and human (Spencer et al., 2020; L. Wang et al., 2020). For example, the LEMP 56aa peptide, encoded by MyolncR4 lncRNA in mouse, that localizes in mitochondria and is essential for skeletal muscle differentiation (Figure 1.9: C). Notably, LEMP function is conserved between mouse and zebrafish (*Danio rerio*) (L. Wang et al., 2020). The SPAAR micropeptide, encoded by human LINC00961 was previously reported to be involved in muscle regeneration (Matsumoto

et al., 2017). A recent study demonstrates that it is also involved in the regulation of endothelial cell differentiation and it is shown to interact with SYNE1 protein (Spencer et al., 2020), a regulator of endothelial cell shape and migration (Chancellor et al., 2010).

Extensive characterisation of translation events from human non-canonical ORFs by means of ribosome profiling, HLA-peptidomics and CRISPR-based screening revealed hundreds of peptides expressed in healthy and cancer tissues (Chen et al., 2020; Chong et al., 2020). Further investigation of the subcellular localisation of 6 micro-peptides encoded by lncRNAs, revealed that they locate in mitochondria, ER, Golgi or plasma membrane and they interactions with known protein complexes, such as cytochrome c oxidase complex (COX) (Chen et al., 2020). The production of functional small peptides from the translation of lncRNAs suggests that actually these RNAs were incorrectly annotated.

1.6 LncRNAs are important regulators of the nervous system development and function

The development and the preservation of a functional nervous system is an intricate and sophisticated process, which requires a tight control of gene expression. Notably, ~40% of the lncRNAs expressed in the mammalian genome are specifically expressed in the brain, where they exhibit precise spatiotemporal expression profiles as well as cellular localisation variability (Ponting et al., 2009; Derrien et al., 2012). Neurogenesis comprises of 8 stages: 1) induction and patterning of a neurogenic region, 2) birth and migration of neurons and glia, 3) specification of cell fates, 4) guidance of axonal growth cones to specific targets, 5) formation of synaptic connections, 6) binding of trophic factors for survival and differentiation, 7) competitive rearrangement of functional synapses and 8) continued synaptic plasticity during the organism's lifetime (Goodman and Doe, 1993). LncRNAs have emerged over the last few years as key regulators in all of these stages (Wu et al., 2013). Mis-regulation of lncRNAs is highly correlated with the occurrence of neurodevelopmental and neurodegenerative diseases, such as Angelman Syndrome (UBE3A-ATS) (Meng et al., 2012), Alzheimer's disease (BACE1-AS, BC200) (Mus et al., 2007; Faghihi et al., 2008), Parkinson's disease (UCHL1-AS) (Carrieri et al., 2015), Huntington's disease (HTT-AS) (Chung et al., 2011) and Amyotrophic Lateral Sclerosis (ALS) (NEAT1) (An et al., 2019).

1.6.1 Human neuronal differentiation

The nervous system is one of the first systems to begin developing during embryogenesis, at the gastrula stage (week 3 in gestation), and the last to be completed, after birth (SF and Sunderland, 2000). Neural development begins by the formation of neural plate, which then folds to form a neural groove and subsequently the neural tube (Figure 1.10). This process is governed by specific morphogen gradients that dictate the anteroposterior (WNT, FGF, RA) and dorsoventral (WNT, BMP, SHH) patterning of the neural tube. Those gradients of morphogens dictate the transcriptional identity, hence the specific type of neural progenitors in a particular area across the anteroposterior and dorsoventral axes (Tao and Zhang, 2016).

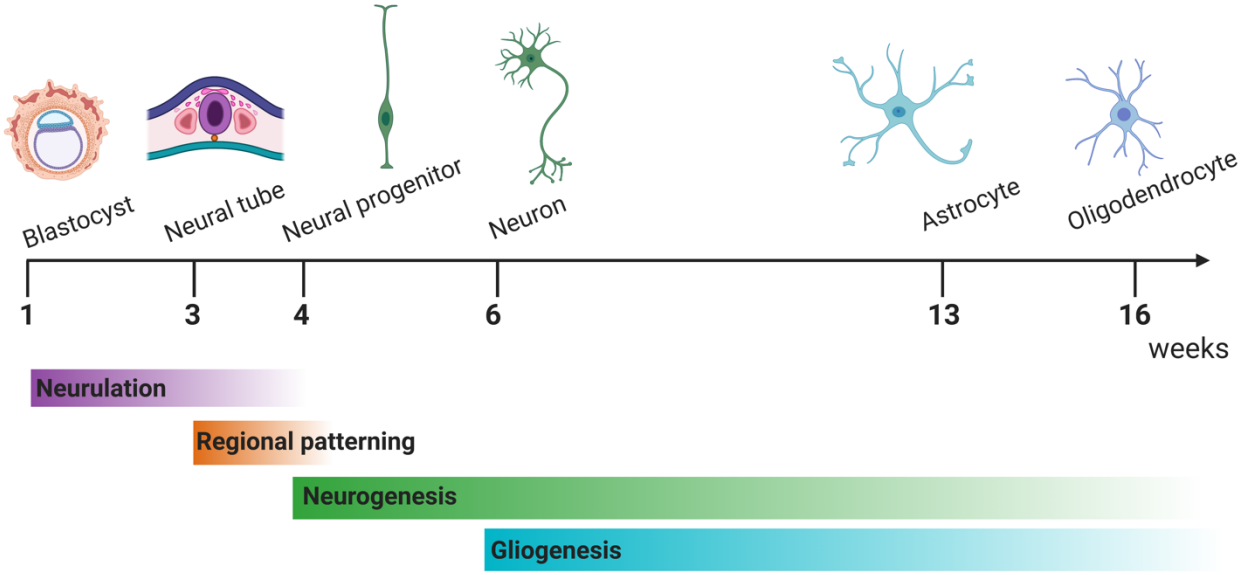
Neurons and glial cells of the central and peripheral nervous system emerge from the neural tube and neural crest respectively, during development (SF and Sunderland, 2000). The neural tube is composed of the germinal neuroepithelium, a layer of rapidly dividing neural stem cells. Cell division occurs vertically and one of the daughter cells remains a stem cell, adjacent to the ventricular surface of the neural tube, whilst the other daughter cell migrates away. Hence, each vertical cell division marks the last time that one of the daughter cells will divide and is called that neuron's "birthday". This process generates the neuronal and glial cells, which migrate and differentiate outside of the neural tube (Figure 1.10). Different pro-neural signals will drive further differentiation of the neural progenitors to specific neural subtypes (e.g. midbrain dopaminergic neurons). Different types of neuronal and glial cells have their birthdays at different times (SF and Sunderland, 2000), depending on their fate and stage of migration.

The primary structural features of the brain (forebrain, midbrain, hindbrain) emerge following the patterning of the neural tube at week 4 in gestation. Cell division and migration out of the neural tube gradually gives rise to the mantle (or intermediate) zone and the ventricular zone. The cells originating from the mantle zone will differentiate into neurons and glia (astrocytes, oligodendrocytes, microglia). Neurons will subsequently form connections among themselves and their axons will form the marginal zone. Therefore, by the 5^{th} week in gestation, the developing neural tube consists of 3 distinct layers, namely the ventricular, the intermediate (mantle) and the marginal zone which become the sources of neurons and glial cells for all brain structures (SF and Sunderland, 2000).

The differentiation of neuronal stem cells (NSCs) into distinct subtypes is influenced by their location in the tissue and by developmental stage and the specific region of tissue. For example, dl4 domain of the developing spinal cord will give ride to dorsal inhibitory neurons. Evidence from some studies suggests that neural progenitors have a certain degree of flexibility and cells can switch identity with cells of adjacent progenitor domains (Dessaud et al., 2007; Dessaud et al., 2010), however they must be stable enough to avoid uncontrollable blending of neuronal identities (Sagner and Briscoe, 2019). Neuronal differentiation timeline is precisely coordinated and any deviation from that timeline results in premature or delayed differentiation. Excessive proliferation of neuronal stem cells can

lead to malformations of the cortex (eg cortical dysplasia type II) or even more extended defects such as megalencephaly. Several mutations have been liked to overproliferation of NSCs, most of which converge on the hyperactivity of mTOR pathway (Guarnieri et al., 2018). On the contrary, reduced proliferation and premature differentiation of NSCs, results in a smaller number of mature neurons, hence microcephaly. This is the result of mutation of genes involved in centrosome maturation and mitotic spindle formation and chromosome condensation during mitosis, such as MCPH1 (Guarnieri et al., 2018).

An essential part of development is the correct wiring of neurons. During development, neuronal cells migrate to find their target cells in order to form functional synapses. To this end, they extend their axonal processes following specific extrinsic biochemical cues which influence the orientation of the neuronal migration. Growth cones, at the tip of the axons, are the receptors of attractant or repellent cues and navigate the neuron accordingly (Tamariz and Varela-EchavarrÃ-a, 2015). Specific protein families such as Netrins, Slits, Ephrins and Semaphorins can act as axon guidance cues, but axons are also guided by morphogens such as BMP, FGF, WNT, SHH or cell adhesion molecules (eg NCAMs).



Adapted from Tao and Zhang 2016

Figure 1.10: Human neural development timeline. Schematic illustrating the major developmental milestones of *in vivo* neural development. Neurulation commences during the first 3 weeks in gestation, with the formation of the neural plate which folds to generate the neural tube. Concomitantly, regional patterning takes place between 3 and 4 weeks in gestation, followed by the development of neural progenitors (NPCs). NPCs divisions give rise to neurons (week 4 in gestation, onwards) and at a later stage to oligodendrocytes. Image adapted from Tao and Zhang, 2016 and created using BIORENDER.

1.6.2 LncRNAs participate in fine-tuning the development of the mammalian CNS

The decision of a neural stem cell to differentiate and migrate outside of the neural tube, is governed by its intrinsic transcription program, as well as by external stimuli. LncRNAs are involved in the regulation of neural stem cell differentiation. The lncRNAs discovered so far to be involved in the fine-tuning of neuronal differentiation are mostly nuclear-enriched trans-acting and function through interaction with key transcription factors and RNA binding proteins.

An RNAi screen for lncRNAs involved in early murine neuronal differentiation, revealed the nuclear lncRNA lncR492 as a lineage-specific inhibitor of neuroectodermal differentiation of murine embryonic stem cells (Figure 1.11: A). LncR492, whose levels are highest in pluripotent stem cells, interacts with the RNA binding protein HuR and together enhance the Wnt signalling, which in turn has an inhibitory effect on neuronal differentiation (Winzi et al., 2018). LncBRN1a and lncBRN1b were recently reported to also interact with HuR to maintain the stemness of mouse neural stem cells (NSCs). PANTR1 is the human homolog of lincBRN1 and is reported to be highly expressed in human induced pluripotent stem cells (iPSCs) (Carelli et al., 2019), therefore it may function in the same way in humans.

Rhabdomyosarcoma 2-associated transcript (RMST) is a nuclear lncRNA conserved from frog to human (Chodroff et al., 2010) and essential for neurogenesis in human (Figure 1.10: B). Its expression is restricted to the brain and increases during neuronal differentiation of human embryonic stem cells (hESCs) (Ng et al., 2013). RMST expression is regulated by the transcription factor REST, which triggers the activation of neuronal genes (Ballas et al., 2005). RMST interacts with SOX2 transcription factor and is required for SOX2 binding to promoter regions of specific neurogenic transcription factors, such as ASCL1 and NEUROG2 (Ng et al., 2013).

Another CNS-specific lncRNA regulator of neuronal differentiation is Tcl1 Upstream Neuron-Associated lincRNA (TUNA) (Figure 1.11: C). TUNA is conserved across vertebrates and essential for the ESC self-renewal and maintenance of pluripotency, as well as for neuronal differentiation in zebrafish, mouse and human (Lin et al., 2014). RNA pulldowns demonstrated that the conserved region of TUNA binds to the pluripotency-

associated RBPs: PTBP1, hnRNP-K and Nucleolin (NCL) to form a multiprotein complex that mediates the transcriptional activation of pluripotency and neuronal differentiation genes, such as NANOG, SOX2 and FGF4.

PNKY is a nuclear neural-specific lncRNA, conserved from mouse to human, that regulates neurogenesis from neural stem cells (NSCs). Via its interaction with the splicing regulator PTBP1, Pnky regulates the expression and alternative splicing of transcripts, such as *Ntsr2, Igfbp5*, *Scrg1*, and *Ppp1r3* that dictate cellular phenotype in mice. Knockdown of Pnky increases murine neuronal differentiation *in vitro* and *in vivo* in the developing cortex. Hence, Pnky acts to restrain neuronal commitment and regulates the production of young neurons during embryonic development and postnatally (Ramos et al., 2015).

The lncRNAs that have so far been reported to regulate neuronal differentiation are enriched and function in the nucleus. Far less is known about the roles of cytoplasmic lncRNAs during neuronal differentiation, let alone their coding potential during this process. However, given that small peptides encoded by lncRNAs play pivotal roles in the regulation of the development of major tissues and organs such as the muscles and the heart (Magny et al., 2013; Anderson et al., 2015; L. Wang et al., 2020) both in invertebrates and in mammals, it is likely that lncRNA-encoded peptides are important in the nervous system development.

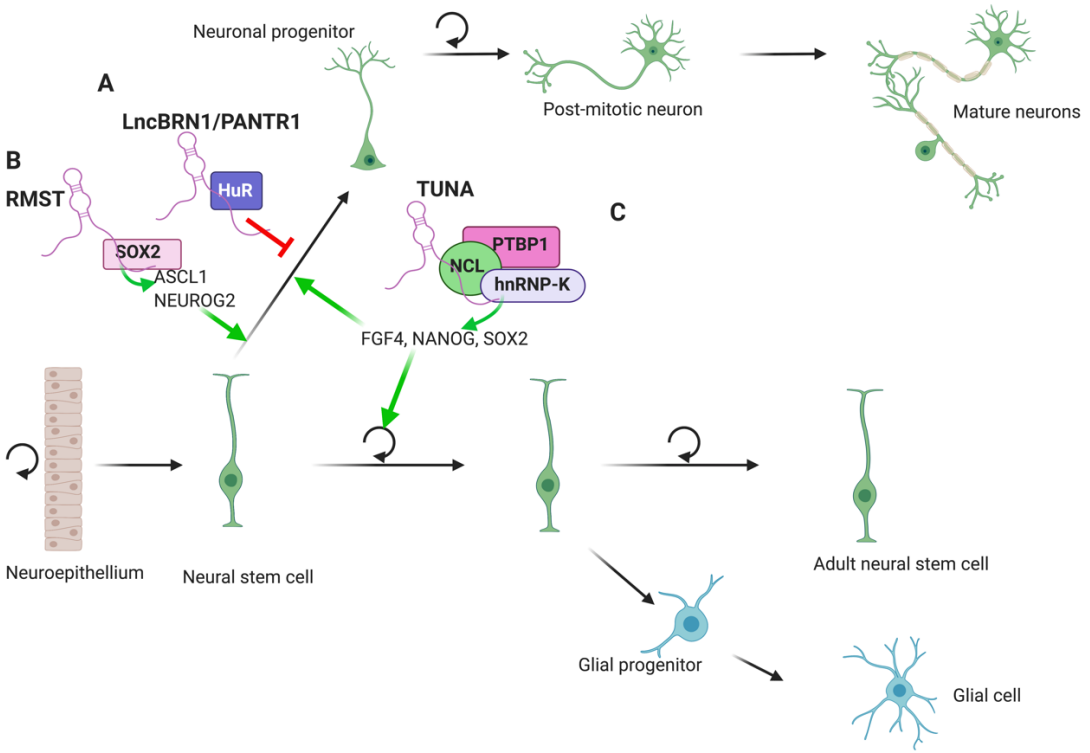


Figure 1.11: Brief overview of neural differentiation and specification in the developing CNS and the roles of lncRNAs in key points of the process. Neural stem cells originate from the neuroepithelium and have the potential to self-renew or differentiate, firstly into neuronal progenitors that will become mature neurons and secondly to glial progenitors that will give rise to glial cells. (A) LncBRN1/PANTR1 lncRNA interacts with the RBP HuR to inhibit neural stem cell differentiation. (B) RMST lncRNA binds to SOX2 transcription factor to promote expression of pro-neural genes and drive neuronal differentiation. (C) TUNA lncRNA interacts with the multiprotein complex, consisting of PTBP1, hnRNP-K and NCL to drive expression of pluripotency and pro-neural genes and regulates both self-renewal and differentiation. Image created with BIORENDER.

1.6.3 Neuronal translation regulation during development and in adulthood

The nervous system consists of a plethora of different parts, which need to function cooperatively. Therefore, intercellular communication between neurons and between neurons and glial cells as well as intracellular communication between the cell soma and the axon, is crucial during development and in adult life. Given that a typical excitatory neuron receives signal from 1–10,000 other neurons and transmits information to 50–100,000 neurons, through synapses (Holt et al., 2019), tight control of protein synthesis is required, in order to maintain the nervous system's correct wiring during development and synaptic plasticity in adulthood.

More than 2,500 proteins are present in synapses (Pielot et al., 2012) and as a group, synaptic proteins and proteins present at growth cones, exhibit a shorter half-life, compared to the average population of cellular proteins (Cohen et al., 2013; Deglincerti et al., 2015; Dörrbaum et al., 2018; Fornasiero et al., 2018) suggesting a quicker turnover occurs, to accommodate synaptic transmission requirements. To reach their targets in the central or peripheral nervous system, axons can extend for millimeters or even up to meters in large vertebrates (Sahoo et al., 2018). The fastest transport across an axon has been reported to be ~1μm/s, for organelles and <0.1μm/s for proteins, suggesting that it would take days to transport a protein from the cell soma across the axon of a human sciatic nerve, which is ~1m long (Holt et al., 2019). However, distal axons should be able to instantly respond to external stimuli, therefore local translation in the axon is an elegant solution to this distance 'restriction'.

Translation in the synapse has been underestimated until recently (Sahoo et al., 2018) but it is now established that mRNAs, bound to RBPs are transported from the soma to the axon and vice-versa (Maday et al., 2014). RNA-Seq of rat striatal synaptic neuropil, which consists of dendrites, axons, glial cells and interneurons but lacks cell somata, has revealed ~2,500 mRNAs that localize to the dendrites and axons of hippocampal pyramidal neurons (Cajigas et al., 2012). mRNAs selective transport to synapses is thought to be mediated by certain motifs in their 5' and 3'-UTRs, although no

single ‘consensus’ motif has been identified to drive axonal localisation (Holt et al., 2019). Secondary structures of mRNAs also play a role in axonal localisation, as they guide the mRNA’s interaction with RBPs, which stabilise their target mRNAs (Gomes et al., 2014). It has been suggested that an RBP can bind to several different mRNAs, since the number of RBPs detected in axons so far is vastly smaller than that of mRNA transcripts (Kar et al., 2018).

Notably, individual mRNAs interact with different RBPs, in order to be transported to axons of different neuronal types. For example, the interaction of β-actin mRNA with ZBP1 and hnRNP R is required for its localisation in motor axons, whilst interaction with ZBP1 alone or together with HuD is sufficient for β-actin mRNA transport to axons of other neurons (Glinka et al., 2010; Kim et al., 2015). mRNAs are translationally repressed during their transport to synapse (Wells, 2006) and as soon as they reach their destination, their affinity to RBPs is decreased. For example, ZBP1 is subjected to phosphorylation, regulated by SHH signal which increases translation in distal axons, and releases β-actin mRNA for local protein synthesis (Hüttelmaier et al., 2005). mRNA transport to synapses needs to be precisely regulated and any alterations or abnormalities of this process can affect axon growth, function and survival. Mis-regulation of several axonal RBPs, such as TDP-43 and FUS/TLS and SMN is linked to the onset of ALS and SMA neurodegenerative diseases (Sahoo et al., 2018).

The first evidence of the presence of ribosomes and polysomes in dendrites and axons of mammalian hippocampal and spinal cord neurons emerged more than 4 decades ago (Bodian, 1965; Tennyson, 1970; Steward and Levy, 1982). A recent electron microscopy study demonstrates that single ribosomes (monosomes) are sparsely scattered in axons and axon terminals and very rarely group into polysomes, in the nodes of Ranvier (Shigeoka et al., 2016). Although several studies argue that monosome-bound mRNAs do not necessarily undergo active translation (Balagopal and Parker, 2011; Guttman et al., 2013; Liu and Qian, 2016), this does not appear to be the case in the synapse. A recent ribosome profiling study of mouse hippocampal synaptic neuropil revealed that synaptic mRNAs are preferentially translated by monosomes (Biever et al., 2020).

Axonal translation is under tight regulation, both during axon growth and in mature synapses (Holt et al., 2019). When axon regeneration is required, following an injury, translation is enhanced, triggered by the increase of axoplasmic Ca^{2+} (Sahoo et al., 2018). Axon elongation and migration during development requires active protein synthesis and is governed by attractant and repellent guidance cues. Notably, a single cue is capable of inducing the rapid and specific regulation of ~100 proteins and different guidance cues trigger distinct proteomic cascades (Cagnetta et al., 2018). Netrin-1, as an attractant, prompts the local translation of cytoskeleton proteins, such as β-actin (Leung et al., 2006). On the contrary, Sema3A and Slit2, as repellents trigger the translation of cytoskeletal-disassembly factors (Wu et al., 2005; Piper et al., 2006). Interestingly, the number of ribosomes in the axon decreases during synapse formation, as the neurons transition from an elongation phase to a more mature phase that does not require the same level of active translation. It is in fact suggested that high levels of active translation might prevent the axon terminal from forming a physically stable connection (Costa et al., 2019).

Synaptic plasticity is the ability of the synapse to strengthen or weaken, according to the specific stimuli. For example, memory storage for long term requires the strengthening of synapses through a process called Long term potentiation (LTP), which requires local protein synthesis (Gkogkas et al., 2010). Translation of mRNAs required for synaptic plasticity is regulated by several mechanisms, such as mRNA modification, RBP phosphorylation or the phosphorylation of eukaryotic translation initiation factors. One of the most well-established examples is the regulation of the CREB-dependent expression of synaptic plasticity genes by the phosphorylation of eIF2α (Gkogkas et al., 2010). ATF4 is one of the repressors of CREB-mediated gene expression and contains two regulatory uORFs. Under normal conditions, when eIF2-GTP is abundant, ribosomes scan downstream of the first uORF, reinitiating at the second uORF, which is an inhibitory element that blocks translation of ATF4. As a result, CREB-mediated synaptic gene expression is no longer repressed, and the formation of strong synapses (LTP) is allowed. Under cellular stress conditions, eIF2α is phosphorylated and eIF2-GTP levels are reduced. Consequently, the amount of ternary complex is reduced and therefore the time required for the scanning ribosomes to reinitiate is increased. This delay results in

ribosomes scanning through the second inhibitory uORF and enhancing translation of ATF4, which represses CREB-mediated synaptic plasticity gene expression, therefore LTP is inhibited (Harding et al., 2000; Vattem and Wek, 2004; Costa-Mattioli et al., 2007).

1.6.4 LncRNAs role in synaptic plasticity and neuronal translation regulation: a balance between health and disease

As in numerous other intricate cellular processes, unsurprisingly, several lncRNAs have been implicated in neuronal translation regulation, regeneration and synaptic plasticity. Recently, more than 15 lncRNAs have been reported to be involved in the regulation of some process within the nervous system (Table 1.2) (Wei et al., 2018). Notably, in some cases, a lncRNA with a well-established function, irrelevant to the nervous system, is shown to also have a neuronal-specific function. Such an example is XIST, which has been reported to be involved in the regulation of apoptosis after spinal cord injury (Gu et al., 2017). The importance of the lncRNAs functions in the nervous system is evident from the fact that their deregulation is highly correlated or even triggers the occurrence of neurodegenerative diseases.

MALAT1, a nuclear lncRNA with well characterised functions in cancer (Ji et al., 2003), is an important regulator of synaptic plasticity. MALAT1 is highly expressed in neurons and enriched in nuclear speckles. MALAT1 has been shown to recruit SR-family splicing proteins to transcription sites to control the expression of synaptogenesis-related genes, in cultured hippocampal neurons. Moreover, knockdown of MALAT1 induces a decrease in synaptic density, whilst its overexpression counteracts this effect (Bernard et al., 2010).

Evf2 lncRNA is essential for the normal development of GABAergic neurons. Evf2 is transcribed from the highly conserved locus Dlx5/6 and, together with Dlx-2 and MECP2 transcription factors, increases the transcriptional activity of Dlx5/6 and Gad1 (enzyme required for the conversion of glutamate to GABA) (Colasante et al., 2008) and regulates the gene expression of GABAergic interneurons in the developing mouse brain. Notably,

silencing of Evf2 results in abnormal formation of GABAergic circuits in the hippocampus and dentate gyrus (Bond et al., 2009).

Regulation of protein synthesis in pre- and post- synaptic terminals is crucial for the maintenance of long-term synaptic plasticity. LncRNA BC200 has been shown to regulate translation in the synapse, by interacting with eIF4A and preventing the formation of the 48S initiation complex (Wang et al., 2002). Synaptic plasticity failure is linked with multiple neurodegenerative disorders, one of which is Alzheimer's disease (AD). Interestingly, in normal aging BC200 RNA levels in cortical areas were reduced by >60% between 49 and 86-year-old individuals (Figure 1.12). On the contrary, in AD patients' brain, BC200 was significantly upregulated in specific areas of the brain, associated with the disease, such as Brodmann's area 9 (Mus et al., 2007).

In some cases, the function of a lncRNA is only triggered by an external stimulus. For example, KCNA2-AS is an antisense lncRNA to voltage-dependent potassium channel KCNA2 mRNA, lowly expressed in a few dorsal root ganglionic (DRG) neurons of the spinal cord but highly upregulated upon nerve injury. Its expression is enhanced by the binding of the transcriptional activator Myeloid zinc finger protein 1 (MZF1) in the promoter region of KCNA2-AS (Zhao et al., 2013). KCNA2-AS selectively targets KCNA2 mRNA and reduces total voltage-gated potassium current, hence increasing excitability of DRG neurons and provoking neuropathic pain. Furthermore, blocking the KCNA2-AS by exogenously overexpressing KCNA2 sense transcript, led to a significant attenuation of neuropathic pain in rats, suggesting that KCNA2-AS can be a therapeutic target for the treatment of neuropathic pain, following spinal cord injury (Zhao et al., 2013).

Mis-regulation of the intronic lncRNA 17A, is associated with neurodegeneration caused by impaired GABA-B signalling. 17A is located in intron 3 of the G protein-coupled receptor- 51 (*GRP51*) gene and induces the production of an alternative spliced transcript (GRP1) that abolishes GABA-B2 intracellular signalling. Notably, transcription of 17A is activated by the inflammatory reaction in the brain of AD patients and therefore is highly correlated with Alzheimer's disease (Figure 1.12) (Massone et al., 2011).

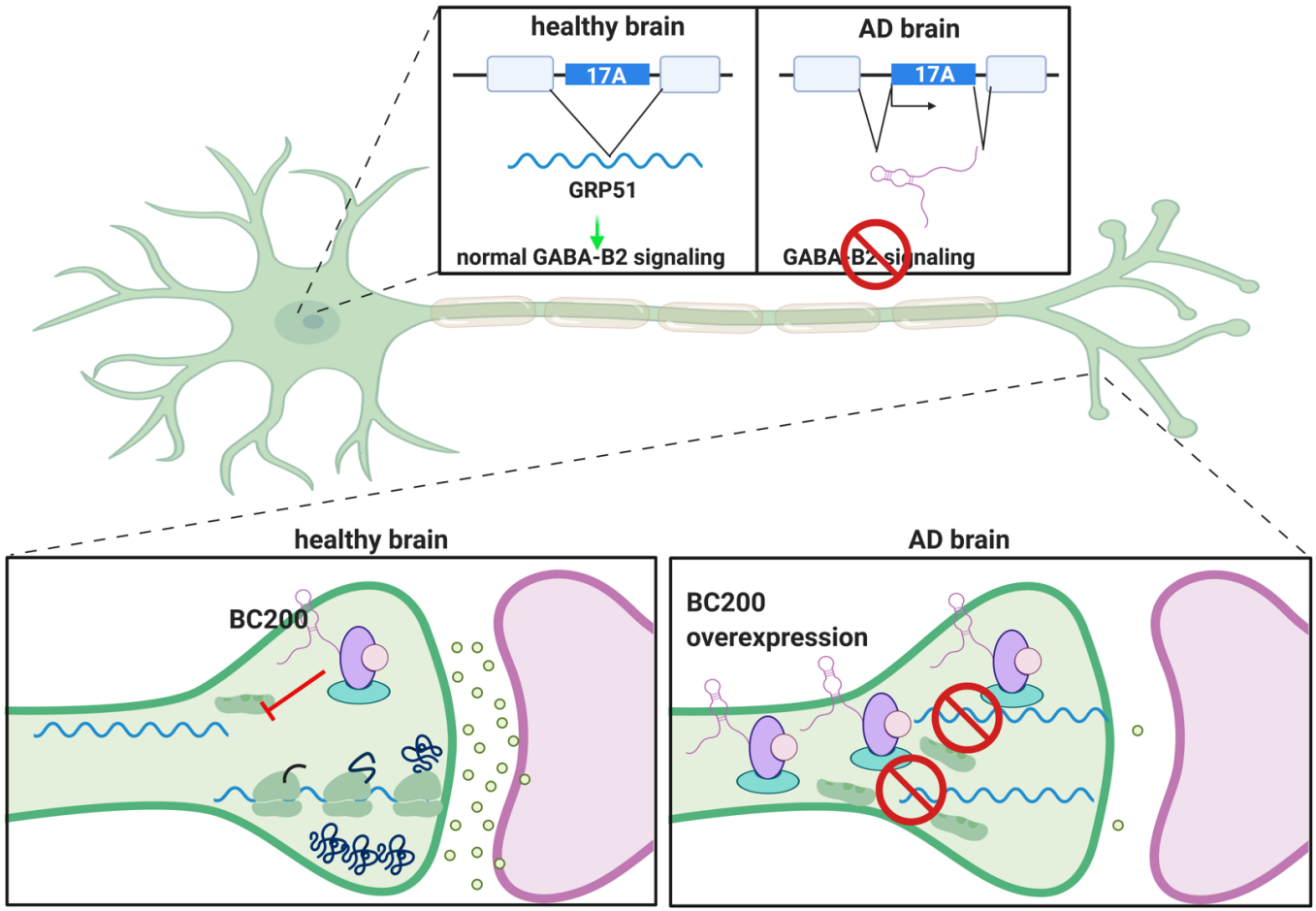


Figure 1.12: Mis regulation of lncRNAs disrupts neuronal translation and synaptic plasticity. Intronic lncRNA 17A is included in the alternative GRP51 transcript, triggered by AD induced inflammation and inhibits the canonical transcription of GABA(B2) receptors, thereby, significantly affecting the GABA-B signaling pathway. BC200 prevents inhibits translation initiation in the synapse and is reduced in normal aging, but in AD BC200 is highly expressed and therefore translation is disrupted, and synaptic plasticity is decreased. Image created with BIORENDER.

Table 1.2: Summary of the lncRNAs associated with the regulation nervous system development and function

LncRNA	Conservation	Disease associated	Mechanism	Reference
UBE3A-AS	Mouse (ube3a-ats)	Angelman's syndrome	Silencing of the paternal allele of UBE3A gene	(Meng et al., 2012)
BACE1-AS		Alzheimer's disease	Stabilisation of BACE1 mRNA by covering the miR485-5p binding site	(Faghihi et al., 2008; Faghihi et al., 2010)
BC200	mouse (BC1)	Alzheimer's disease	Regulation of translation in the synapse	(Wang et al., 2002; Mus et al., 2007)
17A		Alzheimer's disease	Production of alternative transcript GRP1 inhibits canonical transcription of GABA-B2 receptors	(Massone et al., 2011)
PINK1-AS		Parkinson's disease	Stabilisation of expression of PINK1 splice variant in neurons	(Scheele et al., 2007)
UCHL1-AS	Mouse (uchl1-as)	Parkinson's disease	Stabilisation of UCHL1-AS mRNA and positive regulation of its translation	(Carrieri et al., 2012; Carrieri et al., 2015)
HTT-AS		Huntington's disease	HTT-AS-v1 reduces HTT mRNA levels	(Chung et al., 2011)
PANTR1	Mouse (lincBRN1)		Interaction with HuR to maintain stemness of NSCs	(Carelli et al., 2019)
RMST	frog, mouse (Rmst)		Interaction with SOX2 to promote neuronal differentiation	(Ballas et al., 2005; Ng et al., 2013)
TUNA	zebra fish, mouse (Tuna)		Interaction with PTBP1, hnRNP-K and Nucleolin to mediate transcriptional activation of pluripotency and neuronal differentiation genes	(Lin et al., 2014)
PNKY	Mouse (Pnky)		Interaction with PTBP1 to restrain neuronal commitment and	(Ramos et al., 2015)

			regulate the production of young neurons	
MALAT1	mammals	knockdown of *MALAT1* induces a decrease in synaptic density	Recruits SR splicing proteins to transcription sites to control expression of synaptogenesis-related genes	(Bernard et al., 2010)
Evf2	mouse		Regulation of gene expression of GABAergic interneurons	(Bond et al., 2009)
KCNA2-AS	mouse	Neuropathic pain after nerve injury	Targeting of KCNA2 mRNA and increase of DRG neurons excitability, provoking neuropathic pain.	(Zhao et al., 2013)
BDNF-AS			Negative regulation of BDNF expression by recruitment of EZH2 (a PRC2 core component) in response to neuronal depolarization	(Modarresi et al., 2012)
Sox2OT		Alzheimer's disease, Parkinson's disease	Regulation of *Sox2* gene expression to inhibit neurogenesis	(Arisi et al., 2011; Shimozaki, 2014)

1.6.5 Human neuroblastoma SH-SY5Y cells as a model for neuronal differentiation

The considerable heterogeneity of neuronal and glial cell types as well as the complexity of brain development, make the study of lncRNA interactions with the translation machinery challenging. Poly-Ribo-Seq is a biochemical method that requires a lot of starting material. The acquisition of this material from the differentiation of human ESCs to neurons is quite challenging and the use of mouse primary cells was not an option as most lncRNAs are not conserved from mouse to human. Therefore, the use of a human neuronal cell line, such as human neuroblastoma SH-SY5Y provides a carefully controlled system which allows to generate Poly-Ribo-Seq data.

SH-SY5Y cells are derived through three successive sub-clones of the parental SK-N-SH cell line (Biedler et al., 1973), which was originally isolated from a metastatic bone tumour biopsy from a 4-year old female child suffering from neuroblastoma. Notably, the parental SK-N-SH cell line has been reported to contain two morphologically distinct phenotypes: neuroblast-like cells and epithelial-like cells (Ross et al., 1983). SH-SY5Y cells originate from the neural crest and therefore display phenotypic characteristics of neural progenitors and share few properties with mature neurons (e.g. expressing neuronal markers such as RET) (Forster et al., 2016). Because of the above characteristics, SH-SY5Y are widely used as a model for neuronal differentiation (Agholme et al., 2010; Korecka et al., 2013). A commonly used method of differentiation is by treating the cells with all-trans retinoic acid (RA).

RA is a physiological signalling molecule involved in the regulation of anteroposterior and dorsoventral patterning during embryonic development. It is generated during retinol (vitamin A) metabolism and has long been known to promote the transcription of genes related to differentiation, through the activation of its RAR/RXR receptors in the nucleus (Maden, 2007). RA is being synthesised in the posterior mesoderm and as the RA catabolising enzyme CYP26 is synthesised in the anterior mesoderm, a gradient of RA, which drives the pattern of the neural plate, is formed. Furthermore, RA promotes neuronal differentiation, neuronal plasticity and

regeneration by inducing the expression of pro-survival factors such as BMPs, FGF, IGF2 and neurotrophins, thus stimulating neurite outgrowth (Maden, 2007).

RA is the most widely used method for the differentiation of SH-SY5Y neuroblastoma cells (Constantinescu et al., 2007; Agholme et al., 2010; Korecka et al., 2013; Forster et al., 2016). RA is generated during retinol (vitamin A) metabolism (Figure 1.13). Retinol circulates into the bloodstream bound to Retinol Binding Protein 4 (RBP4) and enters the cell through the interaction with STRA6 receptor. Once in the cytoplasm, retinol is bound to RBP1 and is subsequently transformed into retinal, which is subsequently oxidised by RALDH to generate Retinoic Acid. RA can also be released by the cell and re-enter the cell in an autocrine manner or be received by other cells in a paracrine manner, but the details of those actions remain largely unknown. RA binds CRABP2 (Cellular Retinoic Acid Binding Protein2) which facilitates its import to the nucleus.

Upon entering the nucleus RA binds to the ligand activated transcription factor heterodimer (Figure 1.13: A), comprised by Retinoic Acid Receptor (RAR) and Retinoic X Receptor (RXR) and the complex binds to a DNA region called RA response elements (RARE) (Maden, 2007), inducing the transcription of RA primary target genes such as HOXA-1, HOXB-2, SOX6 and WNT-1 (Huang M, Yu-Chen Y, Shu-Rong C, Jin-Ren C, Lu JX, Zhoa L and LJ, 1998). Notably, there is an alternative, non-genomic signalling pathway, triggered by RA in the cytoplasm (Figure 1.13: B). RA induces the phosphorylation of ERK1/2 kinase, which translocates to the nucleus and leads to direct phosphorylation of transcription factors and to an indirect phoshorylation of c-AMP response element binding protein (CREB). Activation of these factors subsequently stimulates promoters of early genes, that contain CRE, AP-1, or SRE motifs, such as c-Fos (Cañón et al., 2004). These genes activate transcription of late genes involved in differentiation (Cañón et al., 2004). After RA exerts its function it is exported from the nucleus and is further catabolised by CYP26 class of P450 enzymes in the cytoplasm.

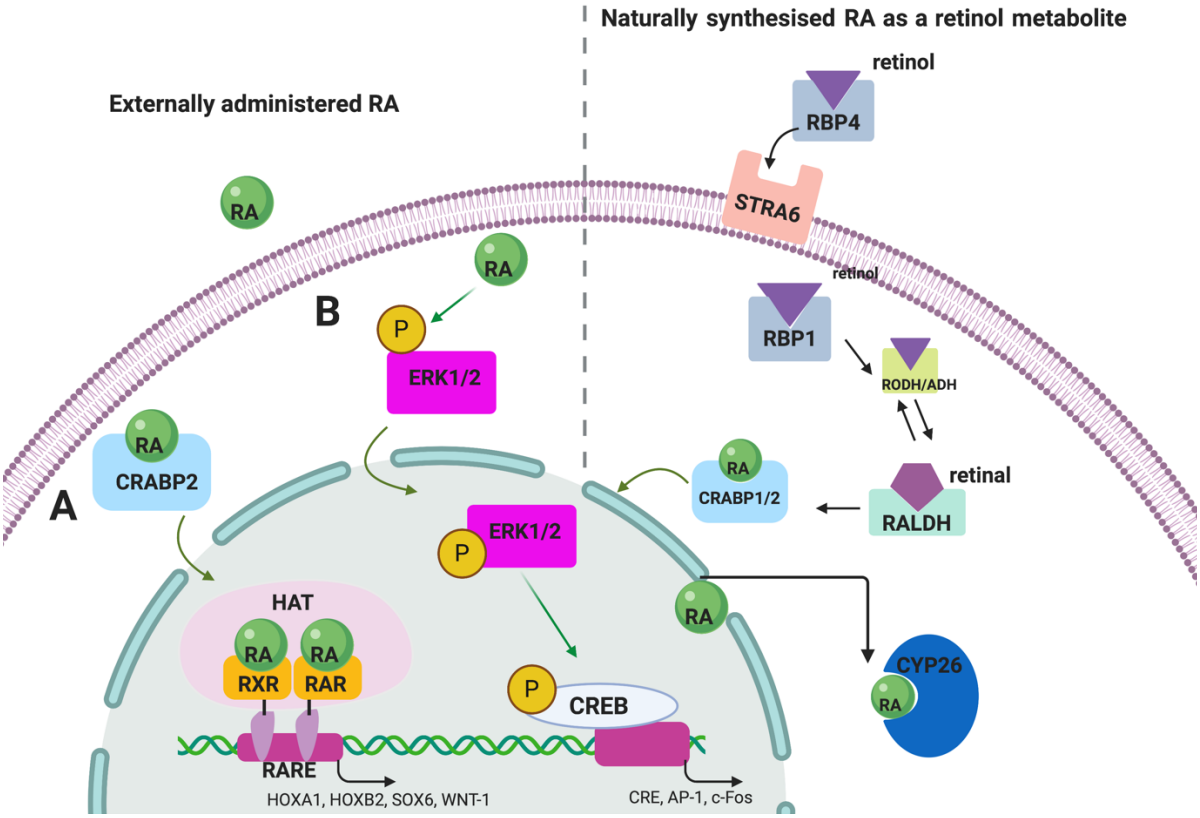


Figure 1.13: Schematic of the RA metabolic pathway. RA binds CRABP2 (Cellular Retinoic Acid Binding Protein2) which facilitates its import to the nucleus. Externally administered RA binds to CRABP2 and once in the nucleus it binds to the ligand activated transcription factor heterodimer, comprised by Retinoic Acid Receptor (RAR) and Retinoic X Receptor (RXR) and then the complex binds to RA response elements (RARE), inducing the transcription of RA primary target genes such as HOXA-1, HOXB-2, SOX6 and WNT-1. Alternatively, RA incduces the phosphorylation of ERK1/2 kinase which, in turn, translocates to the nucleus and leads to an indirect phoshorylation of c-AMP response element binding protein (CREB). Activation of CREB stimulates promoters of early genes, that contain CRE, AP-1, or SRE motifs, such as c-Fos.

1.7 Project Objectives

LncRNAs involved in the regulation of neuronal differentiation are, so far, thought to be restricted in the nucleus and interact with RNA binding proteins and transcription factors to modulate gene expression. However, cytoplasmic lncRNAs play equally important roles in cell homeostasis and their mis-regulation is linked to neurodegeneration. Recent studies have demonstrated that cytoplasmic lncRNAs play pivotal roles in the regulation of myocyte, cardiomyocyte and adipocyte differentiation (Dallagiovanna et al., 2017; Dimartino et al., 2018; Pereira et al., 2020). Hence, it is logical to hypothesise that cytoplasmic lncRNAs may be involved in the regulation of neuronal differentiation.

LncRNAs have been shown to regulate translation, by interacting with ribosomes and being actively translated as well. Regulation of translation is crucial for neuronal differentiation and the roles of lncRNAs in this process still remain elusive. This project aims to address the existing 'gap' in the literature regarding the implication of lncRNA-translation machinery interactions in early human neuronal differentiation. Poly-Ribo-Seq in *D. melanogaster* S2 cells (Aspden et al., 2014) has demonstrated that there are 4 different RNA populations, in the cytoplasm, with regards to their association with translation machinery (Figure 1.14): the actively translated mRNAs, the actively translated lncRNAs, the polysome-associated but not translated lncRNAs and the cytosolic lncRNAs, which have no association with polysomes.

This project aims to characterise the cytoplasmic lncRNA populations, with regards to the translation machinery, and assess their coding potential during human neuronal differentiation, using SH-SY5Y as a model. Specifically, by applying Poly-Ribo-Seq to undifferentiated and differentiated SH-SY5Y cells, I aim to discover alterations in cytoplasmic and polysome-associated lncRNA expression, upon differentiation, to uncover potential differences in polysome enrichment of lncRNAs during differentiation and to detect actively translated smORFs in lncRNAs. Based on the results of data analysis, I aim to select target lncRNAs, to validate their translation and further characterise their role during human neuronal differentiation.

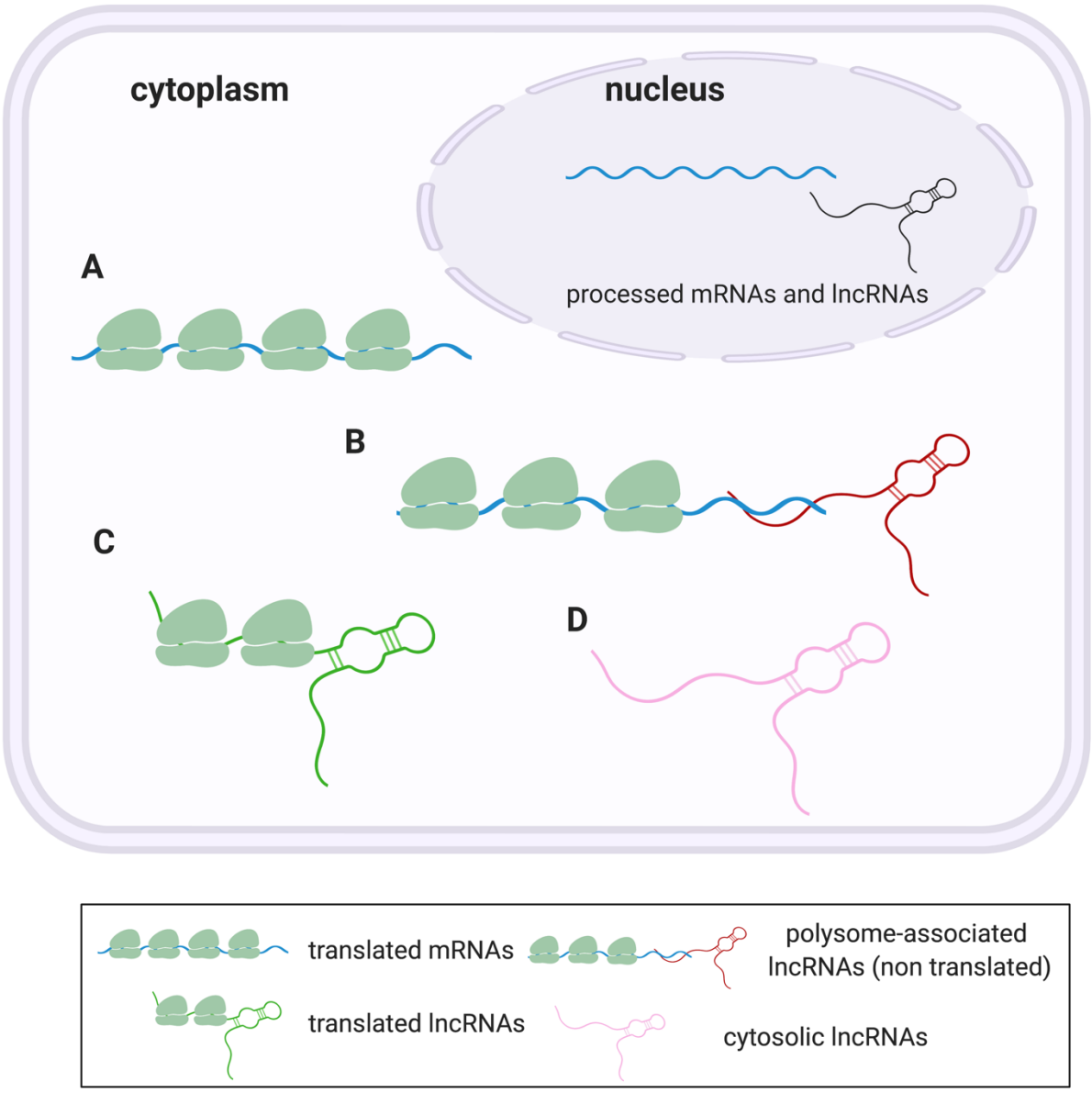


Figure 1.14: Four different RNA populations in the cytoplasm. Cytoplasmic RNAs can be separated into 4 distinct categories, with regards to their interaction with translation machinery: (A) Actively translated mRNAs, (B) Polysome-associated but not translated lncRNAs, (C) Actively translated lncRNAs, (D) cytosolic lncRNAs.

Chapter 2

Materials and Methods

2.1 Cell culture

2.1.1 Cell lines

Experiments were performed in SH-SY5Y (ATCC®) (kindly provided by Dr Eric Hewitt) and in HEK293 (ATCC®) (kindly provided by Dr Iosifina Sampson, Bayliss group) cell lines. SH-SY5Y is a human neuroblastoma cell line, derived through three successive sub-clones of the parental SK-N-SH cell line (Biedler et al., 1973), which was originally isolated from a metastatic bone tumour biopsy from a 4 year-old female child suffering from neuroblastoma. HEK293 cell line is established from a primary embryonic human kidney and transformed with sheared adenovirus type 5 DNA.

2.1.2 Cell culture and passaging

Both SH-SY5Y and HEK293 cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM; 4.5g/L Glucose with L-Glutamine) (Lonza) supplemented with 1% (v/v) Penicillin/Streptomycin (HyClone ™) and 10% Fetal Bovine Serum (FBS) (Sigma). Cells were maintained at 37°C in a humidified 5% CO_2 environment. Medium was changed every few days and cells were passaged at between 80-90% confluence. Cells were passaged by removing the media and washing once in 1 X Phosphate Buffer Saline (PBS) (Lonza). After that, 7mL of trypsin was added to the cells, to detach from the flask and they were incubated at 37°C for 5 min. To deactivate trypsin, full medium was added to the cells (a volume equal to the volume of trypsin) and the cell suspension was centrifuged at 800 xg for 10 min to remove trypsin. Cell pellet was resuspended in fresh media and passaged at a ratio of no lower than 1:10 using fresh complete media.

2.1.3 Trypan Blue Assay

Cell viability was assessed with Trypan Blue assays. Cells were centrifuged at 800 xg for 10 min and cell pellets resuspended in 1 mL PBS. 100 µL of this cell suspension was added to 100 µL Trypan Blue solution (0.4% v/v). The mixture was incubated at room temperature (RT) for 3 min. Next, 10 µL of the cell/Trypan Blue mix were added to a hemocytometer and the number of unstained (viable) and stained (non-viable) cells was counted. To obtain the total number of cells per mL the total number of cells counted in the central rectangle of the hemocytometer was multiplied by 10^4 and corrected for the dilution factor. The percentage of viable cells is

determined by the formula: Viable cells (%) = (number of viable cells/total number of cells) *100.

2.1.4 Cryogenic storage and recovery of cell lines.

After the first passage, 1 x10^6 cells /mL in 1mL freezing media (Table 2.1) were aliquoted in cryo-vials. Vials were frozen overnight at -80^{o}C in an isopropanol-containing vessel and the following day, transferred to liquid nitrogen for long-term storage.

Cells were recovered by quickly thawing at RT and adding to pre-warmed (37^{o}C) complete medium.

Table 2.1: Freezing medium for SH-SY5Y and HEK293 cells

Component	Final concentration	
Cell line	SH-SY5Y	HEK293
DMEM	-	70%
FBS	90%	20%
DMSO	10%	10%

2.1.5 Neural induction using all trans Retinoic Acid.

Neural induction was performed as described previously (Korecka et al., 2013; Forster et al., 2016) with minor alterations. 24h after seeding, All *trans* Retinoic Acid (RA) was added to cells to a final concentration of 30 μM.

2.1.6 Further differentiation using BDNF and N2 supplement.

SH-SY5Y cells were further differentiated following as previously published (Forster et al., 2016). Briefly, neural induction commenced with RA treatment, at a final concentration of 30μM in complete DMEM medium (10% v/v FBS (Sigma), 1% v/v Pen/Step (HyClone) and incubated for 3 days (phase 1 differentiation) before being replaced by Neurobasal medium (ThermoFisher Scientific), supplemented with FBS (10% v/v), N2 (1% v/v), L-glutamine (1% v/v), Pen/Strep 1% v/v) and Brain Derived Neurotrophic Factor (BDNF-Sigma)(50ng/mL). Cells were maintained in culture for another 3 days (phase 2 differentiation).

2.2 Cytoplasmic/Nuclear fractionation of SH-SY5Y cells.

Cells were seeded in 100 mm x 20 mm dishes at a density of 1.5 x10^6 cells per dish. Cells were harvested as described (Methods 2.3.2) and cell pellets washed with 1X PBS prior to lysis. Cells were lysed in whole cell lysis buffer (1X PBS, 1% Tween, 40U/mL RNase inhibitor) (500 μL per 10^6 cells) on ice for 30 min. ~25% of sample was kept as whole cell lysate Nuclei were pelleted at 1,600 xg for 8 min. The cytoplasmic supernatant was transferred to clean tubes. Nuclear and cytoplasmic fractions were subjected to two further clearing steps by centrifugation (3,000 xg and 10,000 xg respectively), to achieve the maximum possible purity. Nuclei were lysed in RIPA buffer (1mL/ 5 x10^6 cells) (150 mM NaCl, 1% IGEPAL, 0.5% Sodium Deoxycholate, 0.1% SDS, 25mM Tris-HCl pH7.4, 1X protease inhibitor cocktail) and lysate was passed through a Medicina IVL05 20mL syringe with a Terumo 23G needle (0.6 x32 mm) to dissolve any clumps. ~10% (~0.3 x10^6 cells) of both nuclear and cytoplasmic lysates and 10% of whole cell lysate (~0.1 x10^6 cells) were used for western blotting and the remainder was subjected to RNA extraction.

2.3 siRNA knockdown

Cells were seeded in 24-well plates at a density of 1x 10^5 cells/well, on Poly-D-Lysine/mouse laminin coated 12 mm round coverslips (Corning BioCoat™ Cellware). LncRNA transcript knockdown was achieved using Lincode siRNA SMARTpool (Dharmacon) that were transfected into cells. This comprised of a pool of siRNA molecules specifically designed against LINC01116 and LINC02143. Lincode Non-targeting Pool was used as scrambled control. Cells were transfected using RNAiMAX lipofectamine (ThermoFisher) as per manufactures's instructions, with minor alterations, the following day. Briefly, lipofectamine and siRNA (10μM stock concentration) were diluted in Opti-MEM® medium, then mixed at a 1:1 ratio and incubated at RT for 25 min. Following that, the mixtures were added to the cells and cells were incubated at 37°C. Knockdown efficiency was assessed by RT-qPCR, using the ΔΔCq method, upto 6 days after transfection. Where differentiation was induced following the knockdown, differentiation was initiated two days post-siRNA transfection, to allow cells to recover from transfection.

2.4 Transient transfections of SH-SY5Y and HEK293 cells

Human neuroblastoma SH-SY5Y and HEK293T cells were transiently transfected using lipofectamine 3000, as per manufacturer's instructions. Cells were seeded 24h prior to transfection on Poly-D-Lysine/mouse laminin coated 12mm round coverslips, one day prior to transfection at a density of 1x 10^5cells/well for 24 well plates and 0.15-0.3 x10^6 (HEK293T) or 0.3-1 x10^6 (SH-SY5Y) cells/well. Lipofectamine was diluted in OptiMEM® medium. Plasmid DNA (200 ng/well for 24well plates; 1 μg/well for 6well plates) was diluted in Opti-MEM® medium with the addition of P3000 reagent (2μL per μg of DNA). Next, the diluted DNA was added to the diluted lipofectamine at a 1:1 ratio, a incubated at RT for 25 min. Following that, the DNA-lipid complexes were added to the cells and cells were incubated at 37° C and analysed after 48h.

2.5 Poly-Ribo-Seq

The Poly-Ribo-Seq protocol (Figure 2.3) described in this chapter is the final protocol after optimisations performed in chapter 3.

2.5.1 Sucrose density gradient preparation

Non-linear Sucrose gradients were generated from 18% to 60% (w/v) (Table 2.2). Cycloheximide, Dithiothreitol (DTT) and protease inhibitor (Roche) were added to pre-made sucrose solutions just before pouring.

Table 2.2: Sucrose solutions

Component	Final concentration
Sucrose	18%-60% (w/v)
Tris-HCl pH8	50 mM
NaCl	150 mM
$MgCl_2$	10 mM
DTT	1 mM
cycloheximide	100 µg/mL
Protease inhibitor	0.33%

Gradients consisted of 0.5 mL of 60%, 2 mL of each of the 50%, 47%, 42% and 1.4 mL of each of the 34%, 26% and 18% sucrose solutions in open-top 12.5 mL POLYCLEAR centrifuge tubes (Seton Scientific), starting with 60% (w/v) and ending with 18% (w/v) sucrose at the top. Each layer was snap-frozen in liquid nitrogen before the next layer was added. Gradients were kept at 4^{o}C and used the following day.

2.5.2 Polysome profiling of human neuronal cells

2.5.2.1 Harvesting

SH-SY5Y cells were seeded in 150 mm x 20 mm dishes (Sarstedt) at a density of 4.5 x10^{6} cells per dish. RA (30 μM final) was added to cells 3 days prior to harvesting, to allow differentiation to occur. Cells were treated with cycloheximide (Sigma) at 100 μg/ml for 3 min at 37^{o}C before being washed (1X PBS, 100 μg/ml cycloheximide) and trypsinised for 5 minutes at 37^{o}C. Cells were then pelleted at 800 x g, washed (1X PBS, 100 μg/ml cycloheximide), and resuspended in ice cold polysome lysis buffer (Table 2.3) for 45 min. Lysates were subjected to centrifugation at 17,000 x g for 5 min, to pellet nuclei. Supernatants, containing cytoplasmic material, were subjected to polysome fractionation. ~20% of the cytosolic material was kept for direct RNA extraction and RNA-Seq.

Table 2.3: Polysome lysis buffer

Component	Final concentration
Tris-HCl pH8	50 mM
NaCl	150 mM
$MgCl_2$	10 mM
DTT	1 mM
IGEPAL	1% (v/v)
cycloheximide	100 µg/ml
Turbo DNase (Thermo™)	24 U/mL
RNasin Plus RNase Inhibitor (Promega)	30 U/mL
cOmplete Protease Inhibitor (Roche)	0.33%

2.5.2.2 Polysome fractionation

Samples were loaded onto sucrose gradients (~70 x10^6 cells per gradient) at 4^{o}C and subjected to ultracentrifugation (121,355 x g_{avg} for 3.5 h, 4^{o}C) in an SW-40Ti rotor (Beckman-Coulter). Gradients were fractionated (1 mL/fraction) using a GRADIENT STATION instrument (Biocomp) and their absorbance at 254 nm was measured using a Biorad UV-vis spectrophotometer.

2.5.3 RNaseI foot-printing

Polysome fractions were pooled and sucrose was diluted down to 10% prior to RNaseI foot-printing in dilution buffer (Table 2.6). RNase was then added (RNaseI EN601, 10 U/µl 0.7-1 U/million cells) (Thermo). Samples were incubated overnight at 4^{o}C or at room temperature (RT) for 1h then RNaseI was deactivated using SUPERase inhibitor (200 U/gradient) (20U/µL; Thermo). Samples were concentrated using 30 kDa molecular weight cut-off columns (Merck), then loaded onto sucrose cushions (1 M sucrose, 50 mM Tris-HCl pH 8, 150 mM NaCl, 10 mM $MgCl_2$, 40 U RNase Inhibitor) in open top thick-wall polycarbonate centrifuge tubes (Seton Scientific). To pellet ribosome protected RNA fragments, tubes were subjected to ultracentrifugation at 204,428 x g_{avg} at 4^{o}C for 4h in a fixed angle TLA110 rotor (Beckman-Coulter). Pellets were resuspended in TRIzol (Ambion,Life technologies) for later RNA purification.

Table 2.4: Dilution buffer

Component	Final concentration
Tris pH 8-HCl	100 mM
NaCl	30 mM
$MgCl_2$	10 mM

2.5.4 RNA purification, PolyA-selection, fragmentation/footprint size selection and rRNA depletion.

2.5.4.1 RNA purification

RNA purification from cytoplasmic cell lysates (2.5.2.1) and RNase footprinted samples (2.5.3) was performed by TRIzol RNA extraction, following the manufacturer's instructions. Briefly, 2 mL TRIzol (ThermoFisher Scientific) was added to whole cell lysate samples. Footprinting samples were resuspended in 1 mL of TRIzol. Samples were incubated for 5 min at RT to permit complete dissociation of nucleoprotein complex before chloroform:isoamyl alcohol (Sigma) was added (at a ratio chloroform:TRIzol; 1:5) and tubes were inverted. Samples were then incubated for 3 min at RT and then centrifuged at 12,000 *x g* for 15 min at 4^{o}C. The aqueous phase was transferred to a sterile tube and RNA was precipitated.

RNA purification from polysome fractions was performed by isopropanol precipitation (2.1, above), followed by TURBO DNase (Thermo 2U/µL) treatment (according to manufacturer's instructions). Samples were incubated in DNase/buffer mix (2x 10^{-6}U of TURBO DNaseI for 10 µg of RNA) at 37^{o}C for 30 min. An equal volume of acidic phenol/chloroform (pH 4.5) (Sigma) was then added to the samples and after brief shaking, were centrifuged at 17,000 *x* g for 10 min at 4^{o}C. The aqueous phase was transferred to a sterile tube and RNA subjected to ethanol precipitation (2.1).

2.5.4.2 Isopropanol/ethanol precipitation

For isopropanol-precipitation, equal volumes of isopropanol (100%, Sigma) were added to samples. For ethanol-precipitation, 2.5X volumes of 100% ethanol was added. For both, NaCl was added at a final concentration of 0.3 M and 1 µL of glycoblue co-precipitant (ThermoFisher Scientific), to facilitate pellet visualisation.

Precipitation was performed at -80^{o}C overnight. RNA was pelleted at 17,000 x g in a benchtop centrifuge for 30-45 min, washed twice with 70% ethanol and air-dried. RNA was resuspended in RNase free ddH_2O and, concentration and quality were assessed using by Nanodrop 8000 instrument equipped with Nanodrop software version 2.3.2a.

2.5.4.3 Poly-A selection

Poly-A selection was performed using oligo (dT) Dynabeads (Invitrogen), according to manufacturer's instructions. Briefly, dT beads (200 µL of dT beads/75-150µg RNA) were washed twice with Binding buffer. RNA samples were heated at 65^{o}C for 2 min to resolve any secondary structures. Following this, RNA was added into beads, resuspended in 200 µL Binding buffer (100 mM Tris-HCl, pH 7.5, 500 mM LiCl, 10 mM EDTA, pH 8, 1% LiDS, 5 mM dithiothreitol (DTT)) and incubated for 10 min at RT on a rotator. Samples were then placed against the magnet and unbound RNA was removed. Beads were washed twice with Buffer A (10 mM Tris-HCl, pH 7.5, 0.15 M LiCl, 1 mM EDTA, 0.1% LiDS) and Buffer B (10 mM Tris-HCl, pH 7.5, 0.15 M LiCl, 1 mM EDTA). 40 µL of 10 mM Tris-HCl pH 7.5 was added to the beads and then polyA RNA eluted at 80^{o}C for 2 min. Poly-A RNA was transferred to a clean tube and was diluted 4-fold in Binding Buffer prior to the second round of Poly-A selection. The same beads were used and the procedure was repeated as described above. Poly-A RNAs were eluted in 40µL of 10mM Tris and transferred to clean microcentrifuge tubes to be further processed by fragmentation.

2.5.4.4 Fragmentation

The poly-A RNA was fragmented by alkaline hydrolysis. Briefly, fragmentation buffer (2 mM EDTA, 10 mM Na_2CO_3, 90 mM $NaHCO_3$) was added to each sample in a 1:1 volume ratio and incubated at 95^{o}C for 10 min. After this, samples were immediately placed on ice and stop solution (0.6 M Sodium Acetate, glycoblue) was added, followed by isopropanol precipitation (Materials and Methods 2.1).

2.5.4.5 Footprint and poly-A RNA size selection and purification

28-34 nt ribosome footprints and 50-80 nt mRNA fragments were gel purified in 10% (w/v) polyacrylamide-TBE-Urea denaturing gel at 300 V for 3.5 hours in 1X TBE. O'RangeRuler 10 bp DNA Ladder was used to aid the band size determination.

Table 2.5: Polyacrylamide Urea denaturing gel

Component	Final concentration
Urea	8M
TBE	1X
Acrylamide	10%
APS	0.1%
TEMED	0.065%

The gel was stained for 30-40 min with SYBER-Gold (Thermo- S11494) and visualised on a UV box. Gel bands between 28 and 34 nt RNA markers for footprint samples and 50-80 nt for total and polysome RNA samples were excised using sterile scalpels. Gel slices were passed through TailorCut Gel Breaker tubes (seqmatic) and subsequently eluted in 750 µL gel elution buffer (Table 2.6) at 4^{o}C rotating overnight. The next day, samples were transferred to a filter column (CoStar®) and centrifuged at 10,000 xg for 1 min. The flow-through that contained the RNA was transferred to a sterile tube and RNA was isopropanol precipitated.

Table 2.6: Gel elution buffer

Component	Final concentration
Tris-HCl pH 7.5	20mM
Sodium Acetate	250mM
EDTA	1mM
SDS	0.25%

After gel elution all samples were subjected to T4 Polynucleotide kinase (PNK) (NEB) reaction in order to re-phosphorylate the 5'-OH end and remove the 3' phosphate terminus that remained after RNase treatment and fragmentation. Briefly, samples were heated at 80^{o}C for 2 min to release secondary structures, and then incubated in T4 PNK reaction buffer, containing 10U T4 PNK, T4 PNK reaction buffer (10X) and 20U (1μL) SUPERase inhibitor, at 37^{o}C for 1h in a thermal cycler. Samples were then heated to 70^{o}C for 10 min, to inactivate PNK and the RNA was isopropanol precipitated.

2.5.4.6 rRNA depletion

Following T4 PNK treatment, ribosome footprint samples were further processed for rRNA depletion (Illumina RiboZero rRNA removal kit, human/mouse/rat RZH1046) following manufacturer's instructions (with some adaptations kindly suggested by Dr Nicholas McGlincy). The probe hybridization step involved incubation of the RNA sample in a PCR tube in a solution with Ribo-Zero reaction buffer and Ribo-Zero removal solution, at a final volume of 40μL. Samples were incubated in a thermal cycler 68^{o}C for 10 min. After this, samples were removed from heat, centrifuged briefly and incubated at RT for 5 min. Magnetic beads were placed against the magnet, liquid was removed, and beads were washed twice with ddH_2O, resuspended by vortex, and placed against the magnet. The supernatant was removed, and the beads were re-suspended in 65 μL of resuspension solution and 100 U RiboGuard RNase inhibitor. The next step was the removal of the probe-bound rRNA, by addition of magnetic beads. Samples were incubated at RT for 5 min and then vortexed for 10 sec. Tubes were then immediately placed on the magnet, and the supernatant, containing rRNA-depleted RNA, was transferred to a fresh tube. RNA was ethanol precipitated.

2.5.5 Library construction and Next Generation Sequencing

5' stranded libraries were constructed for all samples (starting material 100ng-1μg) using NEB Next Multiplex Small RNA Library Prep for Illumina (E7300S), according to manufacturer's instructions. Firstly, 3' end adaptors were ligated on RNA. (Figure 2.1-step 1) Then, the reverse transcription (RT) primer was hybridised to the

3’ end adaptor (Figure 2.1-step 2), prior to the hybridisation of the 5’ end adaptor (Figure 2.1-step 3), to prevent the formation of adaptor dimers. Subsequently, reverse transcription was performed.

At the PCR amplification step (Figure 2.1-step 5), an extra step was added, which slightly deviates from the protocol. PCR amplification of the cDNAs was initially performed in a small scale in order to determine which cycle number was optimal. Components (Tables 2.7, 2.8, 2.10) were mixed and divided in 3 PCR tubes with each amplified for different numbers of cycles (11, 13 and 15).

Table 2.7: small-scale test PCR amplification reaction mix

Component	Vol
Reverse Transcription product	10 µL
LongAmp Taq2X MasterMix	12.5µL
SR Primer for Illumina (5'end primer)	0.625µL
Index primer	0.625µL
ddH_2O	1.25 µL
Total vol	25µL

Table 2.8: PCR protocol:

Step	Temperature	Time	Cycles
Initial denaturation	94°C	30 sec	1
Denaturation	94°C	15 sec	11-15
Annealing	62°C	30 sec	
Extension	70°C	15 sec	
Final extension	70°C	5 min	1
Hold	4°C	□	

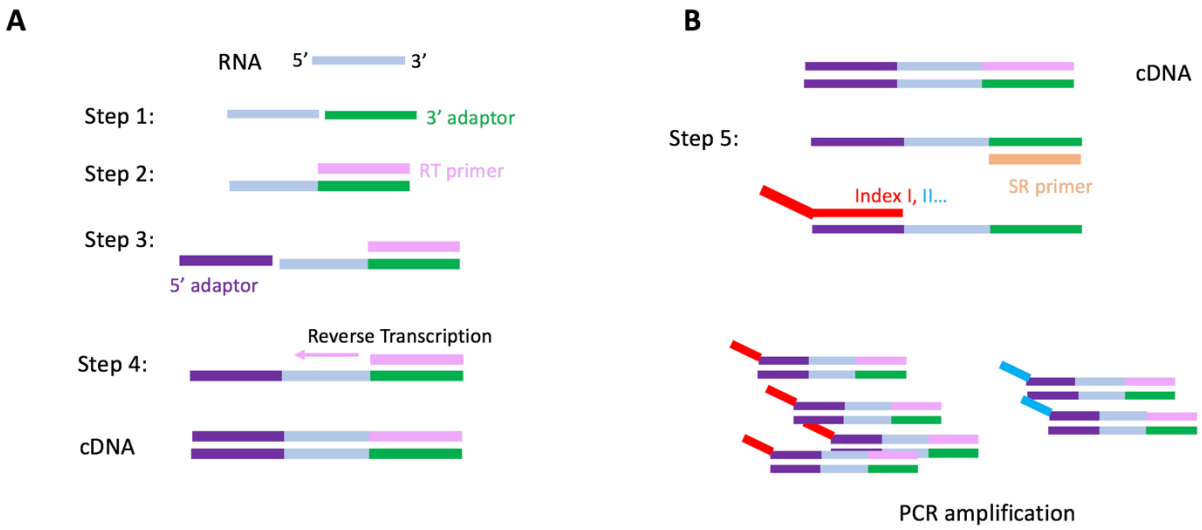


Figure 2.1: The basic steps of library preparation. 3' end adaptor is firstly ligated to the RNA, followed by the ligation of the RT primer. Next, he 5' end adaptor is ligated to the RNA and the reverse transcription is performed. Following that, cDNA is PCR amplified, initially in a small scale, to determine optimal cycle number and subsequently processed for large scale amplification, in order to produce multiple copies for each cDNA.

Samples were run on 8% non-denaturing polyacrylamide gels (Table 2.9) at 200 V for 3 h. Gel were stained for 30-40 min with SYBERGold (Thermofisher) diluted 1 in 10,000 in 1X TBE and visualised on a UV box.

Table 2.9: Non-denaturing (native) polyacrylamide gel

Component	Final concentration
TBE	1X
Acrylamide	8%
APS	0.1%
TEMED	0.065%

After determining the optimal cycle number, PCR was carried out for the rest of the sample.

Table 2.10: Large scale PCR amplification reaction mix

Component	Vol
Reverse Transcription product	30µL
LongAmp Taq2X MasterMix	37.5µL
SR Primer for Illumina (5'end primer)	1.88µL
Index primer	1.88µL
ddH_2O	3.75 µL
Total vol	75µL

Bands ~156 bp for footprint samples, and 150-200 bp for total and polysome RNA samples, were excised from the gel using sterile scalpels, passed through TailorCut Gel Breaker tubes (seqmatic) and subsequently eluted in 750 µL gel elution buffer (Table 2.6) at 4^{o}C rotating overnight. The next day, samples were transferred to a spinX column (Costar) and centrifuged at 15,000 xg for 1 min. The flow-through was transferred to a sterile tube and DNAwas ethanol precipitated overnight at -80^{o}C. RNA precipitants were centrifuged at 17,000 xg for 45 min, washed twice with 70% ethanol, air-dried and resuspended in ddH_2O. The quality of libraries was assessed by Tape-station (Agilent 2200 TapeStation System), and exact concentrations were determined using qubit, by the Next Generation Sequencing facility staff at St James's Hospital. Prior to loading, samples are combined with a fluorescent dye (DNA100 HS kit). Electropherograms and gel-like images are created by the data analysis software for sizing and quantification (Figure 2.2).

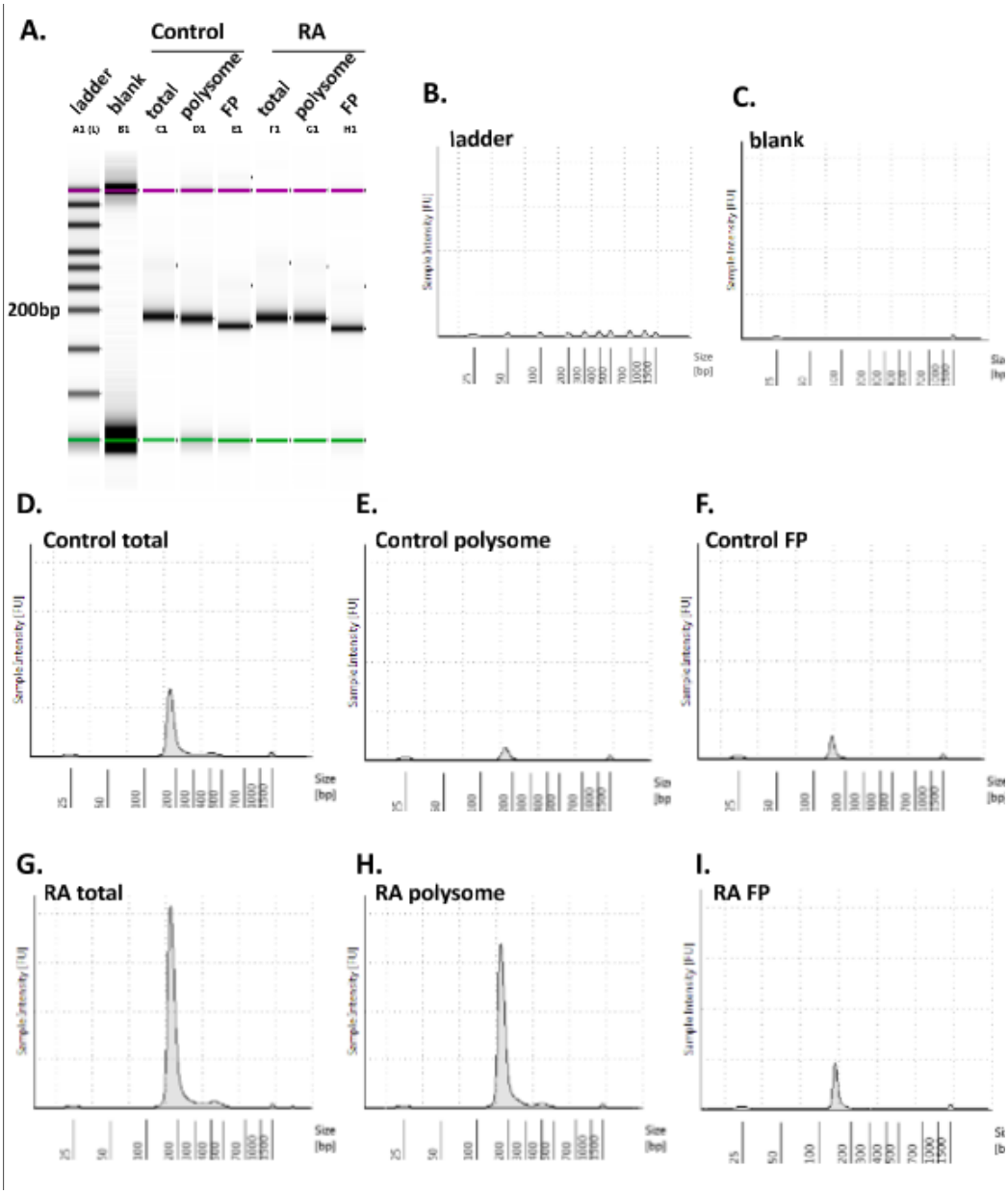


Figure 2.2: Tape-station report for total cytoplasmic, polysome-associated and footprint (FP) samples of Control and RA treated cells. A. Gel image of libraries with size markers (approximately 200 bp for total and polysome and 150 bp for FP). B-I. Quantification of DNA in each lane. On x axis the sizes in bp are depicted and on y axis the quantity of DNA is measured in fluorescent units (FU).

Samples were subjected to 75 bp single end RNA-Seq in a NextSeq500 Illumina sequencer using High Output Kit v2.5 (75 Cycles), that generates ~400 million reads (Next Generation Sequencing Facility, Faculty of Medicine, University of Leeds).

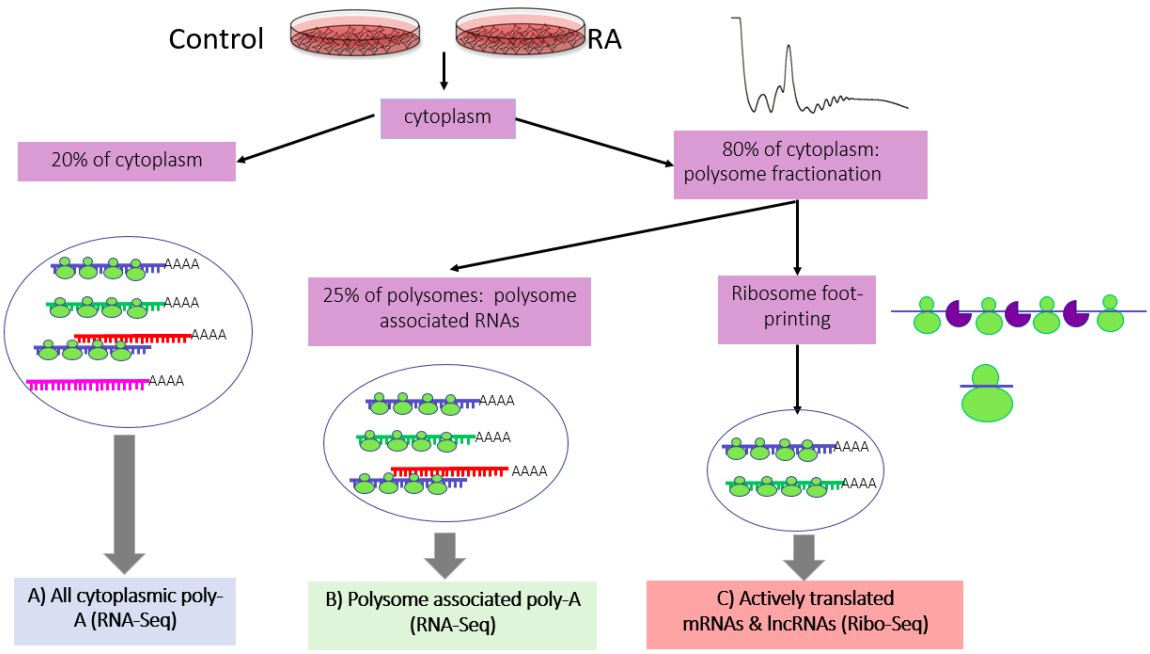


Figure 2.3: Brief schematic of the Poly-Ribo-Seq method. Cells are lysed, nuclei are pelleted, and the cytoplasmic lysate is divided into 3 populations. 20% of the cytoplasm is processed for RNA-Seq (total cytoplasmic RNA). The 80% of the cytoplasm is subjected to sucrose density gradient ultracentrifugation. Following fractionation, polysome fractions from each condition are merged and 25% of this material is also processed for RNA-Seq (polysome-associated RNA). The rest of the polysome fractions is subjected to RNase footprinting and processed for Ribo-Seq.

2.6 RNA purification and quantification

2.6.1 RNA purification from whole cell, nuclear and cytoplasmic lysates.

RNA was purified using Quick RNA miniprep ZYMO R1055 kit, according to manufacturer's instructions eluting in 30 μL ddH_2O. RNA purification from siRNA treated samples was performed using Quick RNA microprep kit (ZYMO R1051).

2.6.2 RNA purification from polysome fractions.

After fractionation, RNA from each fraction was isopropanol-precipitated, treated with TURBO DNase, and then re-purified by acidic phenol/chloroform precipitation as described above (Methods 2.5.4.1).

2.7 Agarose gel electrophoresis

Nucleic acids were typically separated through a 1% (w/v) agarose gels containing 1X Tris-Acetate-EDTA (TAE), or 1x Tris-Borate-EDTA (TBE) (for RNA) buffer (Severn Biotech) and SYBER Safe stain (ThermoFisher Scientific) (0.01% v/v) for DNA or RNA visualisation. Samples were mixed with 6X DNA loading dye (ThermoFisher Scientific) prior to loading on the gel. The presence of a GeneRuler (1kb plus, ThermoFisher Scientific) DNA ladder was used to aid the band size determination. Agarose gels were visualised under UV, using a Chemidoc (BioRad).

2.8 cDNA synthesis

2.8.1 cDNA synthesis on whole cell, nuclear and cytoplasmic lysates processed for qPCR.

Equal amounts of RNA (maximum 1 μg of RNA per reaction) were subjected to cDNA synthesis, using qScript (Quanta Bio/VWR) according to manufacturer's instructions. The following were added in each PCR tube:

Table 2.11: qScript cDNA synthesis protocol

Component	Volume (µL)
RNA	Up to 1µg
qScript reverse transcriptase	1µL
qScript buffer (5X)	1X

Samples were incubated at 22^{o}C for 5 min, followed by 42^{o}C for 30 min and enzyme inactivated at 85^{o}C for 5 min.

2.8.2 cDNA synthesis on polysome fractions processed for qPCR

Equal volumes of RNA were subjected to cDNA synthesis, using qScript (Quanta Bio) as described above. This was done to account for the variation in the quantities of RNA in each fraction, across the gradient.

2.8.3 cDNA synthesis for cloning

cDNA synthesis was performed with Protoscript-II reverse transcriptase (NEB; M0368L- 200,000U/mL) using a mix of random primers and oligo-dT at a 2:5 ratio, to allow for reverse transcription of long polyA transcripts. Briefly, the following were added in each PCR tube (Tables 2.12-2.14):

Table 2.12: Protoscript II cDNA synthesis protocol

Component	Final concentration
RNA	Up to 1µg
Oligo-dT (ThermoFisher Scientific)	3.5µM
Random primer mix (ThermoFisher Scientific)	22.5µM
dNTP	500µM

Samples were incubated at 65ºC for 5 min in a thermal cycler, to denature secondary structure and subsequently transfer on ice for 2 min, centrifuge briefly and add the following components.

Table 2.13: Protoscript II cDNA synthesis reaction mix

Component	Final concentration
Protoscript II buffer	1X
DTT	10mM
Protoscript II-RT	200U/µL
RNAse Inhibitor	40U/µL

Reverse transcription was performed in a thermal cycler following the reverse transcription protocol:

Table 2.14: Protoscript II cDNA synthesis protocol

Step	Temperature	Time
Primer annealing	25ºC	5 min
DNA extension	42ºC	60 min
Reverse transcriptase inactivation	65ºC	20 min
Hold	4ºC	∞

2.9 quantitative Real Time PCR (RT-qPCR)

qPCR was performed using the CFX Connect™ thermal cycler and SYBR Green fluorescent dye (PowerUp™ SYBR™ Green Master Mix, Thermo) according to the manufacturer's instructions. Primers were designed to anneal at 60°C, where possible, and reactions were performed in triplicate in a 96-well qPCR plate (Appleton Woods-BP049). No template and no reverse transcription were employed as negative controls. cDNA was diluted accordingly (usually 1:20) so each reaction contained 12.5 ng cDNA (approximately calculated based on the RNA concentration in the cDNA reaction mix). Components of each well were as follows (Tables 2.15, 2.16):

Table 2.15: qPCR components

Component	Final concentration
PowerUp SYBR Green master mix	1X
Forward Primer	300 nM
Reverse Primer	300 nM
cDNA	2.5 ng/µL

Table 2.16: 1 step PowerUp™SYBR™ Green protocol

Step	Temperature	Time	Cycles
UDG activation	50°C	2 min	hold
Dual lock DNA polymerase	95°C	2 min	hold
Denaturation	95°C	3sec	40
Annealing/extension	60°C*	30sec	
Melt curve	65°C-95°C; increment of 0.5°C	5sec	

*annealing/extension temperature varied from 55°C-60°C for certain transcripts.

Primers (Table 2.17) were designed based on the Ensembl transcript sequences, to specifically anneal at exon-exon junctions, where possible. To assess the properties of the primers I used the IDT oligo analyser (https://www.idtdna.com/pages/tools/oligoanalyzer) and OligoCalc (Kibbe, 2007) tools and the specificity of each pair of primers was verified by *in silico* PCR (https://genome.ucsc.edu/cgi-bin/hgPcr). Expression levels of transcripts were quantified either by absolute or relative quantification. For absolute quantification, the standard curve method was applied, in which a standard curve was created for each primer set, based on the Cq values of serial 10fold dilutions of a pooled sample. For relative quantification the ΔΔCq method was applied and GAPDH was used as a reference gene.

Table 2.17: qPCR primers

gene/ transcript	Forward sequence	Reverse sequence	anneal. temperature °C
CYP26B1	AGTTACCTGCCCAAGATCCAG	CGCCTCCTGGTACACGTTGA	57
MOXD1	GGAAGCCGAAAAGCCAAGTG	TCGAAAATGACGCAGCCTGA	57
NTN4	CGAGTGCAGAACCTGCAAGTGT	CATCTGGAGCTGAGAAGGGTC	60
SCG2	AGCAAAGACCCACTGGCTTG	AGCTTGTTGTCGGAGGTTTTC	56
RET	CCAGCATCTCTACGGCACGTA	CCGCGGTTGCGGACACTGA	60
E2F1	TGGAGCAAGAACCGCTGTTGT	GGGAAAGGCTGATGAACTCCT	57
SOX2	ACATGAACGGCTGGAGCAA	GTAGGACATGCTGTAGGTGGG	57
SOX9	GCTCTGGAGACTTCTGAACGAG	TGGCCTCCTCTGCCTCCG	57
RPS13	ACAGATCGGTGTAATCCTGAGAG	GAAGATCAGGAGCAAGTCCCT	56
RPS5	TGCTGCCAAGGGCTCCTC	GGCTGGGACTGCCCCAAA	60
RPL26	AAGACCGCAAAAAGATCCTCG	TTCCTGCATCTTCTCAATGGTTT	55
RPS28	AGCTAGGTGTAGTGAGCCAGA	GGGGCTGAATCCCACGTTT	57
RPL37	GTCATCGTTTGGAAAGCGTC	ACTCCAGTTATACTTTCTCTTGCG	57
GAPDH	CATCCTGGGCTACACTGAGC	GTCAAAGGTGGAGGAGTGGG	60
LINC01116	TCTAAGAATGGGTCTCACTCTGC	CCAGGCATGGTGGCTCAC	60
LINC02143	AACCTTTGCAGTAGCTCCTGG	GGATGAGGAGACTGAGACTGAGAG	57
NEAT1	TGTTTCCAGGCCTTGCTCAG	CATGGGCTCTGGAACAAGCA	58
AC254633.1	GTGACTCACCTCCCAGACTTC	TGCTGTGCAGCCAGCGTC	57
PART1	CAGGTGATCTGGGGAAAACG	GCCTGCCCTTTGGTTTCT	60
DLGAP1-AS1	TCTGAGAGCCAGCGAACTTT	AGCCTGTTGCGTCATGTGAT	56
DLGAP1-AS2	CCCAGGACACAGACAAGACC	ATGCACGCTCTCTGACAGCA	57
SERPINB9P1	AGTCAGCGAGTGGACAAAGC	GACTCCATGCTGCGGTTTTC	60
AC090001.1	GTGCCCATGAGGGAGAACAC	GACAAGAAGTCAGGAGGTAGACA	57
SNAP25-AS1	AGCCATGGAAGTCAAATGCTG	AGGCATTTTTGCTGTCTTTCCTC	56
SOX9-AS1	TCGATGTGTCTCTTTTTCCCGT	AAGGCTGAACCAGACGACCTT	56
SCAT8	CCCTCACCCTTTCACCTCAA	AGAAGCAACCTGTCATTGGCT	56
TRAF3IP2-AS1	CAGTGTTCGGGGCCGTTTT	CCTCCTCTGCTGGATGTGAA	56
XIST	GGCTCCTCTTGGACATTCTGAG	AGCTTGGCCAGATTCTCAAAG	60

2.10 Immunocytochemistry

Cells were seeded on Poly-D-Lysine/mouse laminin coated 12 mm round coverslips (Corning BioCoat™ Cellware) in 24-well plates at a density of 1x 10^5 cells/well. Fixation was performed with 4% paraformaldehyde (PFA) (Affymetrix) for 20 min at room temperature (RT). A permeabilization step (0.1% TritonX-100 diluted in 1X PBS for 10 min at RT) was performed prior to blocking, followed by blocking at RT in blocking buffer (5% Normal Goat Serum (NGS), 0.1% TritonX, diluted in 1X PBS, or 3% Bovine Serum Albumin (BSA) diluted in1X PBS) for 30 min. Cells were then incubated with primary antibodies (Table 2.22) diluted in 1X PBS, containing TritonX and 0.5% Normal Goat Serum or in 1X PBS, containing 1-3% BSA, depending on the manufacturer's instructions and incubated at 4^o C overnight or at RT for 2h. Cells were washed and labelled with Alexa 488 or Alexa 593 (ThermoFisher Scientific) at 1:500 dilution for 2h at RT in 1X PBS, containing TritonX, 0.5% v/v NGS or 1X PBS, containing 1-3% BSA. Cells were mounted in VECTASHIELD mounting medium, which contains DAPI, for labelling nuclei.

2.11 Microscopy and image analysis

For staining with neuronal and proliferation markers as well as neuronal differentiation assessment after siRNA knockdown, immunocytochemistry slides were imaged using EVOS fluorescent microscope. 'N' stands for separate biological replicates of each experiment and 'n' stands for the total number of cells counted. For FLAG tagging assays, slides were imaged with Confocal LSM700 (Zeiss) at 40X magnification using (Pln Apo 40X/1.3 Oil DICIII). All images were processed and analysed using ImageJ software. Neurite length was measured using the 'segmented line' tool from ImageJ and extending the line from the periphery of the cell soma to the end of each neurite.

2.12 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated by SDS-PAGE. Protein quantification for each sample was performed using the Pierce™ BCA Protein Assay Kit, following the manufacturer's instructions. Samples were diluted in 4X Laemmli sample buffer containing 10%

sodium dodecyl sulphate (SDS) (Biorad) 5% β-mercaptoethanol (Sigma) was added to the buffer prior mixing with the samples. Samples were heated at 95^{o}C for 5 min and loaded on 10% SDS gel along with PageRuler^{TM} pre-stained protein ladder (ThermoFisher Scientific). Gels were run at 150 V in BIO-RAD Mini-PROTEAN$^{™}$ 3 gel electrophoresis system (Bio-Rad Laboratories, Hercules, CA, USA). SDS-PAGE handcast gels were prepared at a final percentage of acrylamide appropriate for detection of the protein of interest (mostly 10%), ranging from 10%-15%. (Table 2.18-2.21). Gels were run in buffer at 90V until samples entered the separating gel and subsequently run at 150V.

Table 2.18: Separating SDS-PAGE solution

Component	Final concentration
Acrylamide 37.5:1	10%
Tris-HCl pH 8.8	375 mM
SDS	0.1%
APS	0.1%
TEMED	0.08%

Table 2.19: Stacking SDS-PAGE solution

Component	Final concentration
Water	
Acrylamide 37.5:1	5%
Tris-HCl pH 6.8	189 mM
SDS	0.1%
APS	0.1%
TEMED	0.1%

Table 2.20: Electrode (running) buffer solution:

Tris- glycine electrophoresis buffer – with SDS

Component	Final concentration
Tris base	3.02g/L
Glycine	18.8g/L
SDS 20%	0.4%

Table 2.21: Transfer buffer solution for nitrocellulose

Component	Final concentration	
	10X	1X
Glycine	144g/L	14.4g/L
Tris base	30.3g/L	3.03g/L

2.13 Western blotting

Following separation by SDS-PAGE, proteins were transferred to Amersham™ Protran™ 0.2 μm nitrocellulose membranes (GE Healthcare) using a BIO-RAD transfer apparatus in transfer buffer (Glycine 14.4g/L, Tris base 3.03g/L). Transfer was carried out at 200 mA for 1.5 h and then membranes was stained with Ponceau Red (VWR). After this, membranes rinsed in 1X PBS and then blocked (5 w/v% fat-free milk powder, 1X PBS ,0.05% Tween-20 (Sigma)) for 1h at RT. Blots were incubated with primary antibodies (Table 2.22), as per manufacturer guidelines, in 1X PBS, containing 0.05% Tween-20 (PBS-T) or in 1X PBS, containing 1% BSA and were incubated at 4^{o}C overnight. Membranes were washed 3 times in PBS-T prior to incubation with secondary antibody (Table 2.22) in PBS-T containing 5% (w/v) non-fat milk for 1.5h at RT. Membranes were then washed 3 times with PBS-T, prior to application of ECL (Biological Industries). Chemiluminescent signal was detected with Chemi-Doc (BIO-RAD). All membranes were probed for β-tubulin to confirm equal loading.

Table 2.22: Antibodies

Protein	Species	Company	Usage (ICC/WB)	Dilution
betaIII-tubulin(TuJ1)	rabbit	ProteinTech	ICC	1:50
ki67	mouse	Dako-Agilent	ICC	1:100
c-Fos	rabbit	Santa Cruz	ICC	1:50
Tubulin B	mouse	DSHB	both	1:5000 for WB, 1:200 for ICC
NXF1	mouse	abcam	WB	1:5000
H3K27me3	mouse	abcam	WB	1:1000
FLAG	mouse	Sigma	both	1:1000
RPS11	rabbit	abcam	both	1:100
hnRNPK	rabbit	abcam	ICC	1:100
Sv2	mouse	DHSB	ICC	1:50
anti-mouse IgG-HRP	goat	New England Biolabs	WB	1:5000
anti-rabbit IgG-HRP	goat	New England Biolabs	WB	1:5000
Alexa anti-mouse IgG-488	goat	Thermofisher	ICC	1:500
Alexa anti-mouse IgG-555	goat	Thermofisher	ICC	1:500
Alexa anti-rabbit IgG-488	goat	Thermofisher	ICC	1:500
Alexa anti-rabbit IgG-555	goat	Thermofisher	ICC	1:500
Alexa anti-rabbit 633	goat	Thermofisher	ICC	1:500

2.14 Plasmid construction for FLAG tagging assays

2.14.1 PCR amplification of inserts

To test translation of smORFs from lncRNA, constructs were cloned that contained putative 5'-UTRs, and smORFs with a C terminal 3xFLAG tag, with no start codon of its own (Figure 2.4: A). Therefore, FLAG signal would be the result of smORF translation. Sequences corresponding to 5'-UTR and smORF CDS were amplified by RT-PCR (Figure 2.4: B) on total RNA from SH5Y5Y cells, using NEB High Fidelity DNA Polymerase (Q5), as per manufacturer's instructions (Tables 2.23-2.24). Annealing times and temperatures were optimised for each lncRNA (Table 2.25).

Table 2.23: PCR components

Component	Final concentrations
Q5 buffer 5X	1X
dNTPs	200 µM
Forward primer	0.5 µM
Reverse primer	0.5 µM
GC enhancer	1X
Q5 polymerase	0.02 U/µl
cDNA	<1000 ng

Table 2.24: PCR cycling conditions

Step	Temperature	Time	Cycles
Initial denaturation	98°C	30sec	
denaturation	98°C	10sec	35
annealing	50°C-72°C	10-30sec	
extension	72°C	1 min/kb	
Final extension	72°C	5 min	
Hold	4°C	-	

Forward primer was designed to anneal at the start of the 5'-UTR and reverse primer to anneal exactly before the stop codon of the smORF at the 3' end. 3xFLAG tag was incorporated into the reverse primer so that the tagged smORF was generated during PCR. (Table 2.25-3xFLAG sequence marked with purple). PCR products were purified prior to restriction digestion, using DNA clean and concentrator kit (ZYMO D4005) as per manufacturer's instructions.

2.14.2 Enzymatic digestion of inserts and vectors

The PCR products were digested with NheI and EcoRV (Table 2.25-restriction sites marked with light blue) and cloned into the NheI and EcoRV restriction sites of the pcDNA3.1-hygro vector (Addgene) kindly provided by Bayliss group. Restriction digestions of the vector and inserts were performed with NheI-HF® (NEB) and EcoRV-HF® (NEB) following manufacturer's instructions. Diagnostic digestions (0.1 U of enzyme/1µg of DNA) and large-scale digestions (10U of enzyme/1µg of DNA) were performed in 1X CutSmart® buffer (NEB) at 37^{o}C for 1h and 3h respectively.

After digestion, the vector was treated with calf intestinal alkaline phosphatase (CIP) (Quick CIP-NEB M0525S (5000U/mL)) (2.9U/µg of plasmid DNA) in CutSmart® buffer at 37^{o}C for 30 min, followed by CIP inactivation at 80^{o}C for 2 min.

2.14.3 Ligation of inserts and vector

Digestion products were separated by agarose gel electrophoresis (Methods 2.2, Figure 2.4: C) and subsequently gel purified using the QIAquick Gel Extraction Kit (QIAGEN 28706). Ligation was performed at 1:0 (negative control), 1:1 and 1:3 vector:insert molar ratios. For each ligation reaction 100 ng of vector were used.

Ligation reaction was performed using T4 ligase (NEB M0202) as per manufacturer's instructions. 400U of enzyme per reaction, incubated at 16^{o}C overnight, followed by T4 ligase inactivation at 65^{o}C for 10 min. 5µL of the ligation reaction were used to transform chemically competent DH5α *E. coli* bacterial cells.

Table 2.25: PCR primers

Primer sequence elements colour code:

Black: complementary to LINC01116 cDNA

Green:sequence for restriction enzyme binding

Light blue: restriction site

Purple: FLAG sequence

Transcript	F sequence	R sequence	anneal. temperature °C
LINC001116-smORF	GGCGGTGCTAGCGCGAGCCACGGGCCTC	GGCGGTGATATCTCACTTGTCATCGTCATCCTTGTAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCGTTTTTAAGCTGACTTGTC	56
LINC01116-whole	GGCGGTGCTAGCGCGAGCCACGGGCCTC	ACCGCCGATATCCAATTACTGCATTCACGTATTCTTC	63
LINC00478	GGCGGTGCTAGCCTCCTGTCGTTAAGATAAATTCTCCA	GGCGGTGATATCTCACTTGTCATCGTCATCCTTGTAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCGTAGTAAAATGCTCTCTGA	56
AL162386	GGCGGTGCTAGCGCTTCGCGGAAGCCGCTG	GGCGGTGATATCTCACTTGTCATCGTCATCCTTGTAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCAGCAGTCCTAAATTCACCA	61
PSMA3-AS1 (ENST00000551597.6)	GGCGGTGCTAGCCTTGTCGGCGCCATTTTGTCTC	GGCGGTGATATCTCACTTGTCATCGTCATCCTTGTAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCGGTGATCCACCCGCCTCGG	70
PSMA3-AS2 (ENST00000554309.1)	GGCGGTGCTAGCGGCGCCATTTTGTCTCGGCAG	GGCGGTGATATCTCACTTGTCATCGTCATCCTTGTAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCTCCCACCTCAGCCTCTTGA	69
MCPH1-AS	GGCGGTGCTAGCGGGAAGGGAGGTGGTGGC	GGCGGTGATATCTCACTTGTCATCGTCATCCTTGTAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCTTTCTTTGGTTCTAGAGAACAGCCGC	59
FLAG-control	GGCGGTGCTAGCATGGACTACAAAGACCATGAC	GGCGGTGATATCTCACTTGTCATCGTCATCCTTG	61
LINC01116- mutant (ATG1)	CTTCTAAGAAAAGGTCTCACTCTGC	CAATTCAGTTGTCTTCTAATAC	57
LINC01116- mutant (ATG2)	GGCACCATCAAAGCTCACTGCAGC	ACCACACTCCAGCCTGGG	66
LINC01116-mutant (both ATG)	GCTGGAGTGTGGTGGCACCATCAAAGCTCACTGCAGCCTTGAA	CTGGGTGATGGCAGAGTGAGACCTTTTCTTAGAAGCAATTCAGTTGTC	61

2.15 Creation of start codon mutants by site directed mutagenesis

The AUG start codon of the putative smORF translated in LINC01116, as well as an addition in-frame AUG start codon, were mutated using Q5® Site-Directed Mutagenesis Kit (NEB), as per manufacturer's instructions with minor alterations. (Tables 2.26-2.28, Figure 2.4: D).

Table 2.26: Site-directed mutagenesis PCR protocol

Reagent	Final concentration
Q5 2X master mix	1X
Fwd primer	0.5 µM
Rev primer	0.5 µM
Template DNA 1-25 ng/µL	1-25 ng

Table 2.27: PCR cycle conditions

Step	Temperature	Time
Initial denaturation	98	30 secs
25 cycles (can increase this if having problems)	98	10 secs
	61 (Ta of primers)	30 secs
	72	4 min (20-30secs/kb of plasmid)
Final extension	72	2 min
Hold	4-10	

PCRs were treated with KLD mix to kinase, circularize and Dpn. 2.5 ng of the original plasmid (template) was treated with KLD -as a negative control. KLD incubation was performed for 1h at RT (Table 2.35):

Table 2.28: KLD treatment buffer

Reagent	Final concentration
PCR product (from step 1)	
2X KLD buffer	1X
10X KLD enzyme mix	1X

Reactions were transformed in DH5α following a standard transformation protocol (Methods 2.18). Miniprep clones were sequenced by Sanger sequencing (Genewiz) with primers T7 promoter forward primer (TAATACGACTCACTATAGGG) and BGH reverse primer (TAGAAGGCACAGTCGAGG).

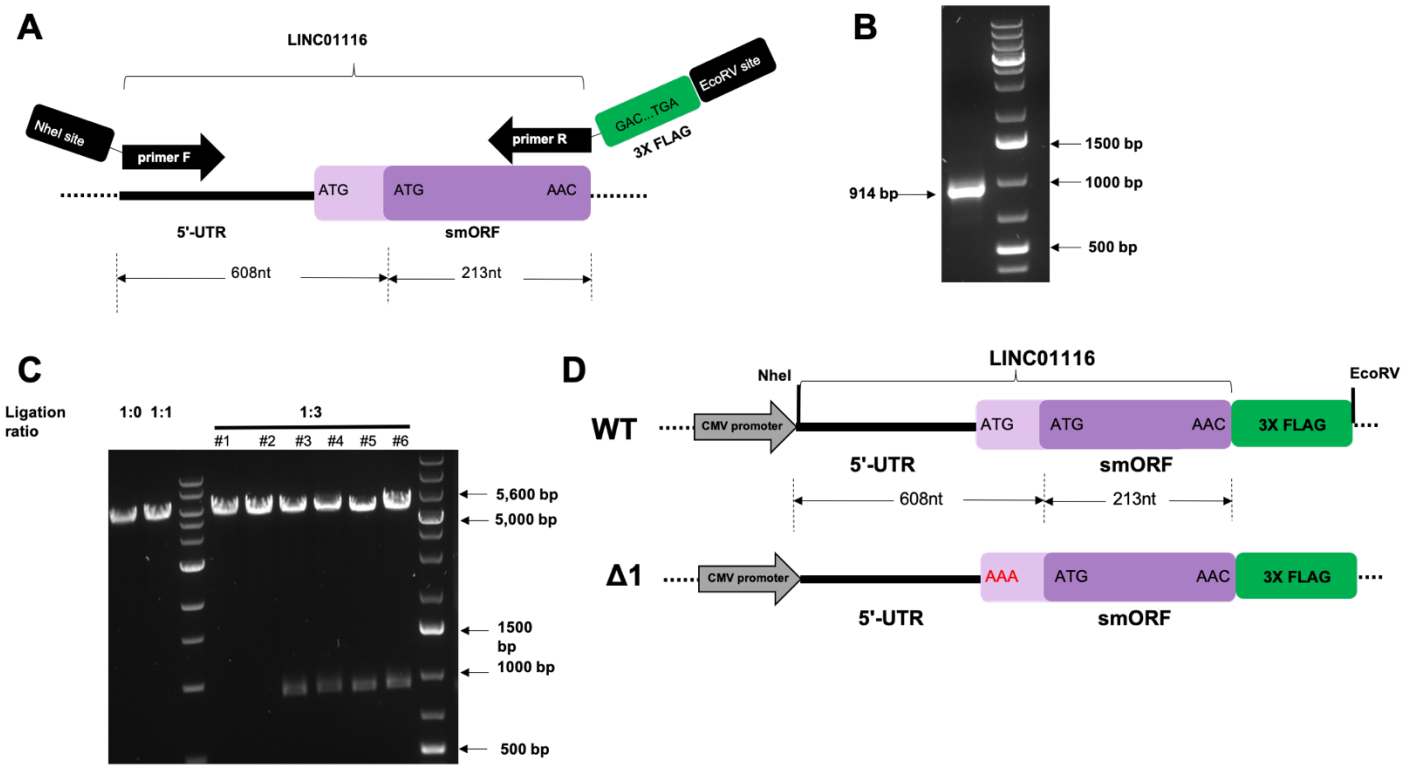


Figure 2.4: Representative example of the cloning process for creation of a FLAG tagged smORF construct and its ATG mutants for LINC01116. (A) Schematic of the PCR amplification for the isolation and simultaneous FLAG-tagging of the insert. (B) Agarose gel image of the PCR product at 914bp. (C) Miniprep DNA from different colonies of DH5α *E.coli* cells resulting from transformation of 1:0 (negative control), 1:1 and 1:3 ligation products, digested with EcoRV and NheI showed that 4 colonies successfully produced the FLAG-tagged construct. PCR and miniprep DNA subject to double digestion were purified from 1% agarose gel and 1kb plus DNA ladder was used to aid band size determination. (D) Schematic of the WT and ATG (Δ1: first ATG mutated) mutant constructs created by site directed mutagenesis.

2.16 Bacterial transformation

Plasmid DNA (up to 1µg) or the ligation reaction (5µL) was added to 50µL of *E.coli* DH5α chemically competent cells. Cells were incubated on ice for 30 min, subsequently heat-shocked at 42^{o}C for 30 sec. and immediately placed on ice for 2 min. 250 µL of LB medium was added cells were incubated at 37^{o}C shaker incubator for 1h. Cells were plated on pre-warmed LB/agar plates containing (100 µg/mL ampicillin or carbenicillin). and were incubated at 37^{o}C overnight.

2.17 Preparation of small (mini-prep) and medium-scale (midi-prep) bacterial cultures

Purification of Plasmid DNA from a small-scale bacterial culture was performed using E.Z.N.A Plasmid DNA Mini Kit I Spin Protocol (Omega -D6942-02). Single colonies were used to inoculate 5 mL LB cultures (100 µg/mL ampicillin or carbenicillin). Culture was incubated at 37^{o}C for 16 h with vigorous shaking (220 rpm). Cultures were centrifuged at 10,000 x g, for 10 min at RT and purification was performed as per manufacturer's instructions. Plasmid DNA concentration and quality were assessed by nanodrop.

Medium-scale Plasmid purification was performed with GeneJET Plasmid Midiprep Kit (Thermo-K0481) following manufacturer's instructions. Starter culture of 5 mL LB (antibiotic) were inoculated with single colonies picked from a fresh selective plate or a glycerol stock and was incubated for approximately 8h at 37^{o}C with vigorous shaking (220 rpm). The 5mL culture was subsequently transferred into 100 mL LB and was further incubated at 37^{o}C for 16 h with vigorous shaking (220 rpm). The principle for the plasmid DNA extraction was the same as in mini prep culture described above, adapted to larger volumes of culture Plasmid DNA was eluted in ~300 µL elution buffer. DNA quality and concentration were determined by nanodrop.

2.18 Bioinformatic analysis

Separate pipelines were used for analysis of the RNA-Seq and Ribo-Seq data. Differential expression analysis of the total and polysome associated RNA-Seq datasets was performed as described below, by Dr Dapeng Wang (LeedsOmics). Ribo-Seq coupled with RNA-Seq analysis was performed by myself, using RiboSeqR pipeline. Further analysis was performed by Isabel Birds (PhD student in the Aspden group) using RiboTaper pipeline (Calviello et al., 2016) with some adaptations.

2.18.1 RNA Seq data analysis and differential expression (performed by Dr Dapeng Wang)

RNA-Seq reads were trimmed with Cutadapt (Martin, 2011) and filtered with fastq_quality_filter (Hannon, 2010) to filter out the reads of low quality (90% of the read to have a phred score above 20). (Figure 2.5) Filtered reads were mapped to human reference genome reference (Li et al., 2009) with Rsubread (Liao et al., 2013) and uniquely mapped reads were reported. Reference consisted of mRNA annotation from UCSC (Haeussler et al., 2019) human genome assembly (hg19) from iGenomes and the lncRNA GENCODE (Frankish et al., 2019a) annotation. Bam file sorting and indexing was performed with SAMtools (Li et al., 2009). Subsequently, summarised read counts for all genes were calculated using featureCounts (Liao et al., 2014). For the sake of normalization, RPKM values for all genes were calculated. The differential expression analysis was conducted between with DESeq2 (Love et al., 2014) employing cut-offs of padj < 0.05 and the absolute value of $\log_2$ Fold-Change > 1. Gene ontology analysis was performed with GOrilla (***G****ene* ***O****ntology en****RI****chment ana****L****ysis and visua****L****iz****A****tion too*l) (Eden et al., 2009) with a cut-off of FDR<0.01.

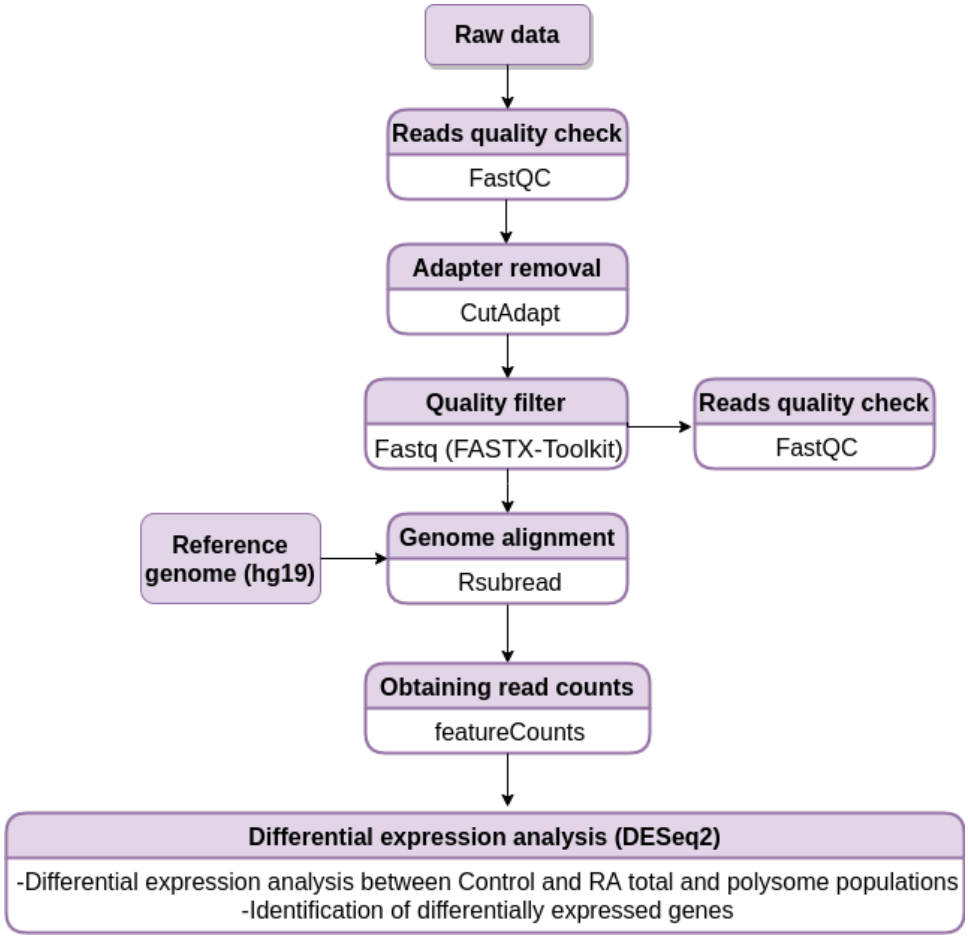


Figure 2.5: Schematic of the RNA-Seq data analysis and differential expression analysis (performed by Dr Dapeng Wang)

2.18.2 Principal Component Analysis (PCA)

PCA analysis (Jolliffe, 2011) was performed on the gene expression matrix of the RNA-Seq datasets (generated by Dr D. Wang, as described in Methods 2.18.1). The matrix contained the rpkm (reads per kilobase per million) values for each gene in all 12 samples (Control total, Control polysome, RA total and RA polysome in three biological replicates). The matrix was filtered so only genes with rpkm>1 were used. Next, a new matrix was generated by calculating the log2 of each of the values and was further subjected to quantile normalisation prior to PCA. Briefly, if rpkm values for each sample are considered as points centred to the origin of a 2D graph, PCA finds the best fitting line between the points, by projecting each point to the line and maximising the sum of the squared distances from those projected points to the origin. This best fitting line accounts for the greatest variation between the samples and is plotted on the 2D graph as PC1. Similarly, the second greatest variation between the data is determined and and plotted it as an axon 90^{o} vertical to PC1.

2.18.3 RNA and Ribo-Seq analysis using RiboSeqR pipeline (performed by myself)

For the analysis of the Ribo-Seq data I utilised a docker version of Ribogalaxy (Michel et al., 2016) that was kindly installed and maintained on the Leeds High Performance Computing (HPC) Unit, by the Leeds HPC Facility. RNA-Seq and Ribo-Seq fastq files were uploaded on Ribogalaxy (Michel et al., 2016) and subjected to quality control using FastQC (Andrews, 2010). Representative results of the key FastQC parameters are shown in Figures 2.6-2.7.

- Per base sequence quality

The per base sequence quality (Figure 2.6) of the datasets is assessed by the accuracy of the base calling at each base of the read. For each position on the read a quality score is assigned (represented by whisker plots). Good quality scores are above 28.

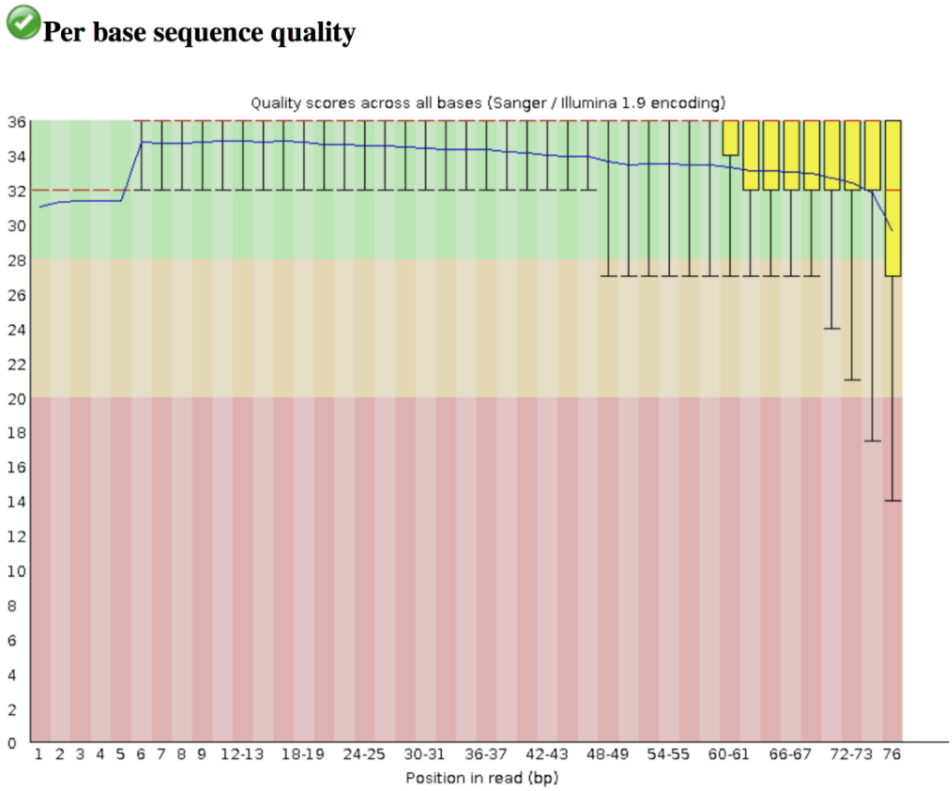


Figure 2.6: Per base sequence quality report indicates correct base-calling. Each position on the read (x axis) is assigned a quality score (y axis) that assesses the accuracy of the base-calling for any given position on the read. The y axis is binned into 3 categories of low, medium and good quality, based on the quality scores. Any quality score above 28 is considered good quality.

- Per sequence quality scores (Phred scores)

This test is used to see if a subset of the sequences has universally low-quality values. In general, Phred quality score (Q) is logarithmically related to the base-calling error probability (P) by the equation: $Q=-10\log_{10}P$. So, if the Phred score at a given base position is 30, it means that the probability of that base being called incorrectly is 1 in 1000. The Phred score for most of the sequences is above 30 (Figure 2.7), therefore the quality of the sequencing is high.

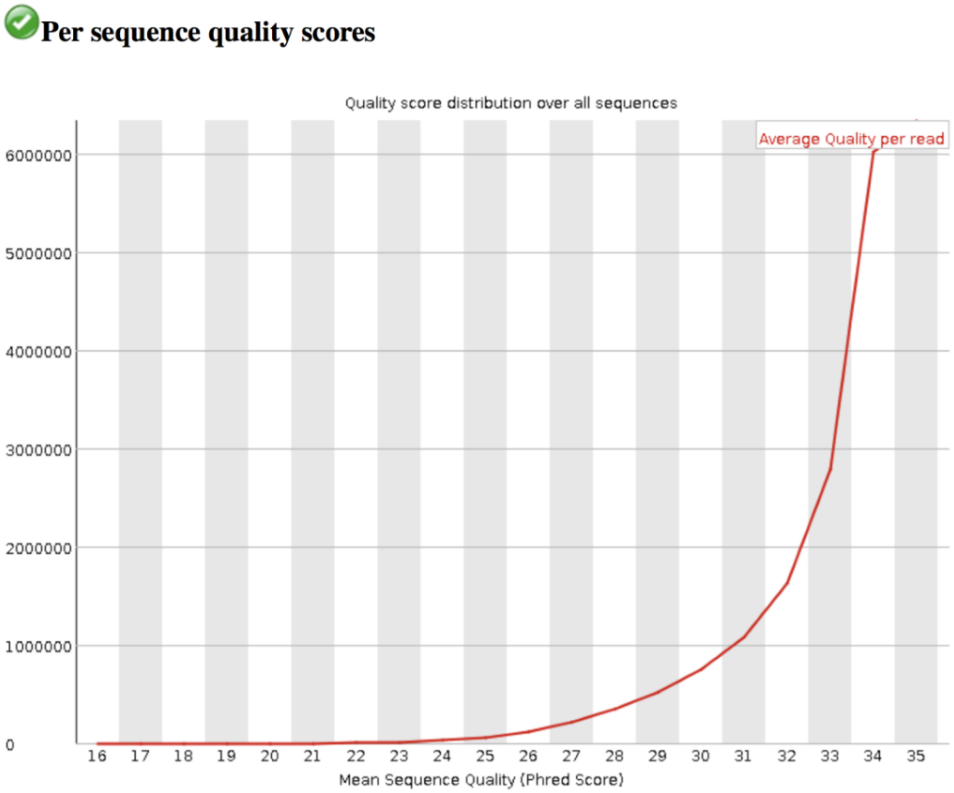


Figure 2.7: Mean sequence Phred score passes 30 for the vast majority of sequences. Sequences on the y axis are assigned a Phred score (x axis) based on the accuracy of the base-calling.

- Sequence length distribution

The samples were subjected to a single end 75bp read Next Gen sequencing run, therefore, the sequencer ‘read’ 75bp of each DNA sequence of the pool. Thus, it is expected that the length distribution plot in figure 3.25 has a peak at ~75bp (figure 2.8).

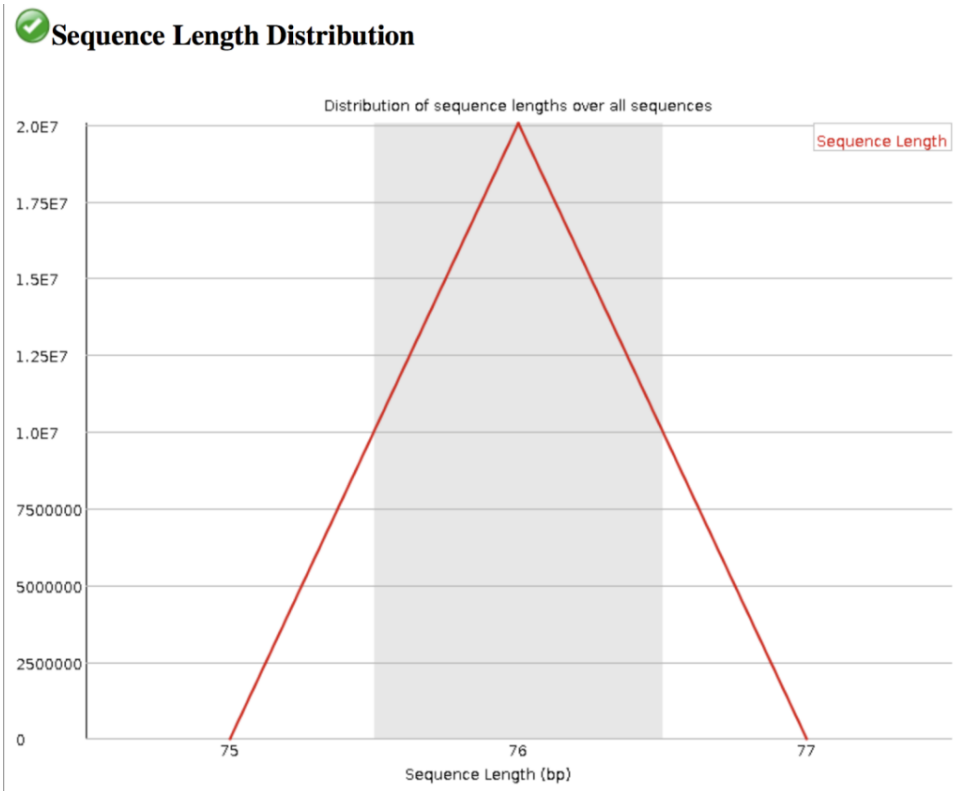


Figure 2.8: Sequence length distribution is as expected in a 75bp run.

Next, 3' end adapter was trimmed from the reads using Cutadapt (Martin, 2011), keeping untrimmed reads in the final output of only the RNA-Seq reads (Figure 2.9). Trimmed reads were further filtered, so that 90% of each read passed the quality threshold Phred score of 20, using the Filter by quality tool (Gordon, 2010) on Galaxy (Afgan et al., 2018). Subsequently, rRNA and tRNA reads were removed, using Bowtie (Langmead et al., 2009) and 1 base trimmed from the 3' end of reads. Next, reads were processed in two different ways: 1) For data visualisation on genome browser, reads were mapped to the human genome (version hg38, Gencode v29) (Frankish et al., 2019b) using Bowtie. Then, bam files (aligned reads) were sorted and were visualised using GoldenHelix software (Golden Helix GenomeBrowser). 2) For assessment of the framing quality of Ribosome profiling, reads were mapped to the human transcriptome (version hg38, Gencode v29) and were subsequently processed with the RiboSeqR pipeline (Hardcastle, 2014). *Step 1* of the pipeline was the formation of an R object (*Prepare RiboSeqR input*) which includes the RiboSeq and RNA Seq aligned sam files. *Step 2*: Triplet periodicity analysis was performed on the output of step 1, in order to assess the precision of the foot-printing and thus the likelihood that footprints represent elongating ribosomes. The analysis was performed on read lengths 25nt-35nt, in order to assess the number of reads of each specific length that are ineach frame. *Step 3*: a metagene analysis step was performed on the reads that display the best triplet periodicity (31 and 33 nt) with parameters for filtering those reads (filterHits parameters) set as: lengths=31, 33; frames=1,2,3; hitMean=50; unqhitMean=10. Plots were generated and the plotCDS (parameters set as: lengths=31, 33; min5p=-100; max5p=100; min3p=-100; max3p=100). In this analysis, reads were globally mapped to 5' and 3' UTRs and coding regions (CDS) and the mean number of reads that is mapped to each region is plotted.

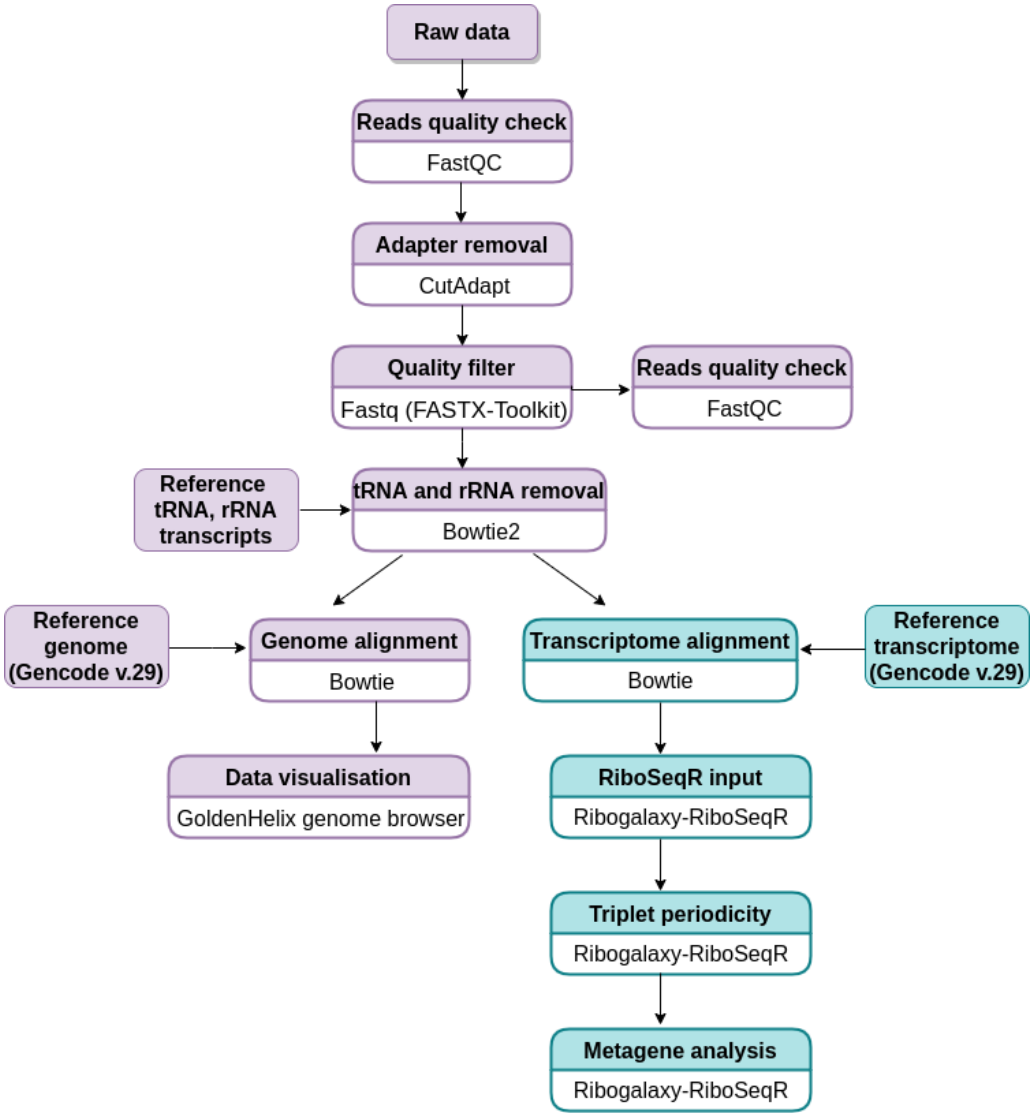


Figure 2.9: Schematic of the RNA-Seq and Ribo-Seq data analysis for data visualisation and assessment of ribosome profiling quality. This analysis was performed by myself using the ribogalaxy platform and RiboSeqR pipeline.

2.18.4 Ribo-Seq analysis using the RiboTaper pipeline (performed by Isabel Birds)

Initial quality reports of Total, Polysome-associated RNA-Seq and Ribo-Seq data were made using Fastqc (v.0.11.8) (Andrews, 2010) (Figure 2.10). Adaptor sequences were trimmed using Cutadapt (v.1.81) (Martin, 2011) with minimum read length of 25bp, and untrimmed outputs only retained for RNA-Seq reads. Low-quality reads (score <20 for 10% or more of read) were then discarded using FASTQ Quality Filter, FASTX-Toolkit (v.0.0.14) (Gordon, 2010). Human rRNA sequences were retrived from RiboGalaxy (Michel et al., 2016), and high confidence hg38 tRNA sequences from GtRNAdb Release 17. One base was removed from 3' end of reads to improve alignment quality, and alignment to rRNA and tRNA performed using Bowtie2 (v.2.3.4.3) (Langmead and Salzberg, 2012).

The splice aware aligner STAR (v2.6.1b) (Dobin et al., 2013) was then used to map remaining reads to human reference genome Release 30 (GRCh38.p12), from Gencode[8]. The STAR (v2.6.1b) genome index was built with a sjdbOverhang of 73. Samtools (v.1.9) (Li et al., 2009) was used to create sorted, indexed bam files of the resulting alignments.

Metaplots of aligned Ribo-seq data were generated using create_metaplots.bash script from Ribotaper (v1.3) (Calviello et al., 2016) pipeline. These show the distance between the 5' ends of Ribo-seq and annotated start and stop codons from CCDS ORFs, allowing the locations of P-sites to be inferred. Read lengths exhibiting the best triplet periodicity were selected for each replicate, along with appropriate offsets.

Actively translated smORFs were then identified using Ribotaper (v1.3) (Calviello et al., 2016). Initially, this requires an exon to contain more than 5 P-sites in order to pass to quality control steps. Identified ORFs were then required to have a 3-nt periodic pattern of Ribo-seq reads, with 50% or more of the P-sites in-frame. In the case of multiple start codons, the most upstream in-frame start codon with a minimum of five P-sites in between it and the next ATG was selected. ORFs for which >30% of the Ribo-seq coverage was only supported by multimapping reads were also subsequently filtered. For a smORF to be considered actively translated in a condition, we required that it be identified in at least two of the three biological replicates for the condition.

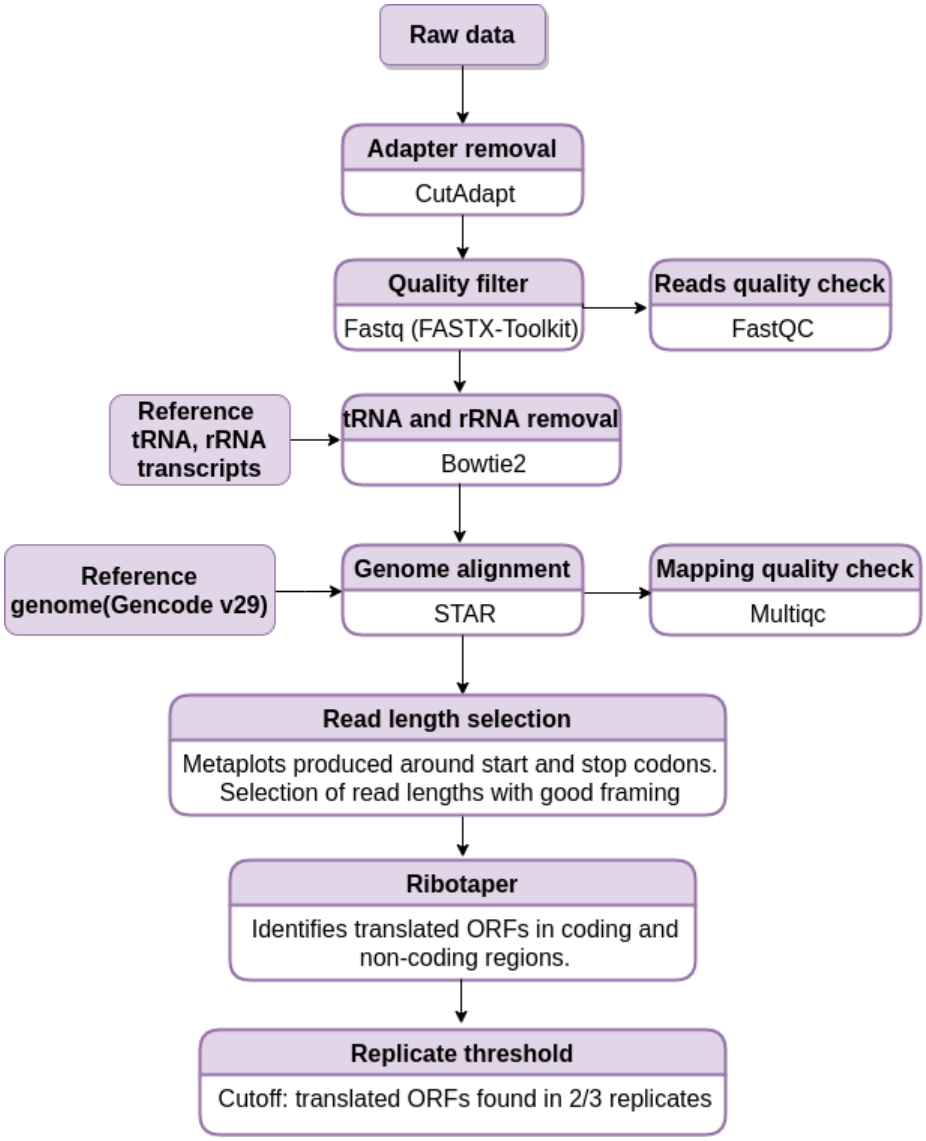


Figure 2.10: Schematic of the further Ribo-Seq data analysis using the RiboTaper pipeline. This analysis was performed by Isabel Birds.

2.19 Statistical analysis

Statistical significance between independent samples with equal variances, student's T test was applied using $p<0.05$ as cut-off.

2.20 Production of graphs and schematics

Graphs were produced using Microsoft Excel or R (R Core Team, 2019) in Rstudio (Rstudio, 2016). Specifically, for data analysis the packages dplyr (Wickham R.; Henry, L.; Müller, K., 2019) and tidyverse (Wickham, 2017) were used. Scatter plots were created using ggplot2 (Wickham, 2016) and volcano plots were produced using EnhancedVolcano (Blighe K Lewis M, 2019). Schematics were produced using Microsoft Power Point or BIORENDER (https://BIORENDER.com/).

Chapter 3

Optimisation of Poly-Ribo-Seq for neuronal cells.

3.1 Introduction

It is generally accepted that differentiation is mediated by changes in gene expression (Albert and Huttner, 2018). Therefore, gene expression at early and late stages of neuronal differentiation needs to be precisely coordinated. Several studies have reported that lncRNAs play pivotal roles in the regulation of neuronal differentiation and fate decision, both in human and mouse (Lin et al., 2014; Winzi et al., 2018; Carelli et al., 2019). Functions of lncRNAs in the nucleus are well characterised and in recent years, many cytoplasmic functions have emerged (Tsagakis et al., 2020). In many cases, their interaction with RNA-binding proteins regulates the balance between stemness and differentiation (Lin et al., 2014; Winzi et al., 2018; Carelli et al., 2019).

Recent evidence suggests that the importance of translational regulation during differentiation has been underestimated. Studies in mouse and human cells have proved that transcript levels do not always correlate with protein levels (Ghazalpour et al., 2011; Edfors et al., 2016) and this has been directly observed in cells and tissues that undergo differentiation (Bulut-Karslioglu et al., 2018; Corsini et al., 2018). During forebrain neuronal differentiation of mouse embryonic stem cells (mESCs), translation levels vary. Specifically, translation levels of genes associated with translation (eg EIF4B, ribosomal protein genes) decrease during the transition to neural progenitor cells (NPCs), while translation levels of transcription associated genes (*YAF2*, ARID1A) increase. Then, after 14 days in neuronal culture, translation levels of genes such as fibronectin FSD1L and calmodulin-dependent protein kinase (CAMK4) are increased (Blair et al., 2017).

The development of ribosome profiling (Ingolia et al., 2009) has greatly contributed to our understanding of post-transcriptional gene regulation. Not only has it allowed us to accurately map the position of ribosomes along ORFs, with a single nucleotide precision, but it has also revealed mechanistic details of translation, and other aspects of translational regulation (Jackson and Standart, 2015), such as the translation of small ORFs within the 5'-UTR or in lncRNA transcripts (Heyer and Moore, 2016; Rodriguez et al., 2019; Chen et al., 2020). Poly-Ribo-Seq (Aspden et al., 2014) has emerged as a more accurate means of measuring active translation compared to ribosome profiling. This is because in Poly-Ribo-Seq the ribosome

protected fragments that are sequenced are derived from the RNase I digestion of only the actively translating polysomes after polysome fractionation. Moreover, Poly-Ribo-Seq allows us to distinguish the actively translated mRNAs amongst those that are associated with the polysomes. Therefore, it sheds more light on mechanisms of translation regulation and helps identify polysome-associated but not translated RNAs. Notably, this technique has the potential to determine if lncRNAs interact with the translation machinery and identify those that are translated.

Although translation regulation during neuronal differentiation has been extensively studied in recent years, the influence of lncRNAs is poorly understood, especially in humans. LncRNAs can interact with the translation machinery, some potentially having a regulatory function, and some possessing actively translated small ORFs that are conserved and encode putative peptides (Aspden et al., 2014; Ruiz-Orera and Albà, 2019).

In this thesis, Polysome profiling and Poly-Ribo-Seq will be utilised to identify lncRNAs that interact with polysomes during human neuronal differentiation and investigate their coding potential. In addition, this analysis will determine whether they are involved in the regulation of differentiation. The first step in assessing potential translation and function of lncRNAs during neuronal differentiation was to develop an appropriate system for experimental analysis.

3.2 Validation of SH-SY5Y RA driven differentiation as a model for the study of translation

Polysome profiling and Poly-Ribo-Seq both require substantial amounts of starting material. Therefore, the human neuroblastoma cell line SH-SY5Y was selected as a good *in vitro* model for the early stages of neuronal differentiation. To drive differentiation in these cells the established model of retinoic acid (RA) treatment was utilised.

RA is a signalling molecule generated during retinol (vitamin A) metabolism and promotes the transcription of genes related to differentiation. Treatment with RA is the most widely used method for the differentiation of SH-SY5Y neuroblastoma cells (Constantinescu et al., 2007; Agholme et al., 2010; Korecka et al., 2013; Forster et al.,

2016). To develop the differentiation system for Poly-Ribo-Seq SH-5Y5Y cells were treated with RA and their phenotypic changes assessed.

3.2.1 Assessment of neurite elongation and neuronal marker expression by immunofluorescence detection of Tuj1 and c-Fos respectively.

To establish differentiation, human neuroblastoma SH-SY5Y cells were treated with Retinoic Acid (RA). Neural induction commenced at passage 4 and was performed as described previously (Korecka et al., 2013; Forster et al., 2016) with minor alterations. To assess the extent of differentiation, several attributes were measured. Brightfield microscopy of undifferentiated SH-SY5Y cells (Control) and cells differentiated for 3 days with 30μM retinoic acid revealed that undifferentiated cells tended to grow in clumps, reminiscent of neurospheres (organised cell clumps in the G_0 phase of the cell cycle that eventually detach from the substrate; (Hämmerle et al., 2013; Piletz et al., 2018)), whereas differentiated cells appear more spread, extending their neurites to form more connections with each other (Figure 3.1).

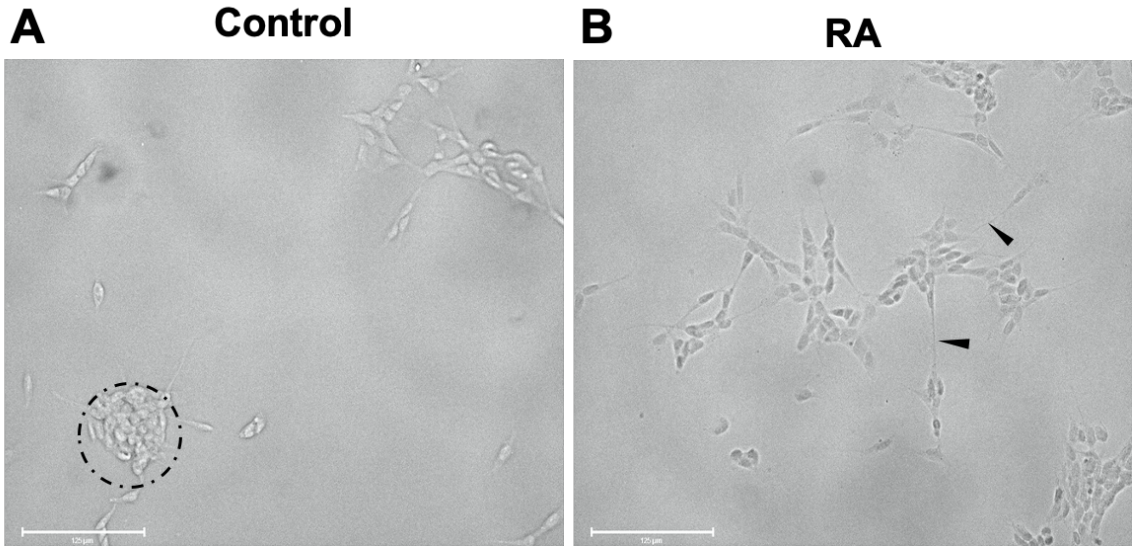


Figure 3.1: SH-SY5Y extend their neurites upon RA treatment. Brightfield image of undifferentiated (Control) and differentiated (RA). **(**A) Undifferentiated cells (treated with DMSO), grow in clumps (dotted circle) that tend to be semi-adherent. (B) Cells treated with RA at a final concentration of 30 μM for 3 days extend their neurites and form more connections. Scale bar 125 μm.

Comparison of the neurite length between undifferentiated and differentiated cells is widely used as a means of assessment of differentiation (Constantinescu et al., 2007; Agholme et al., 2010; Korecka et al., 2013; Forster et al., 2016; Rodriguez et al., 2019). SH-SY5Y express neuronal specific β-III tubulin (TUBB3 or Tuj1), which is the main constituent of microtubules and is involved in axon guidance (Tischfield et al., 2010), both prior to and after RA treatment. Immunocytochemical staining for Tuj1 allows the visualisation of neurites. Therefore, to assess the progress of differentiation, neurite length in both conditions was quantified. RA treated cells have more elongated neurites (see white arrows) compared to Control cells (Figure 3.2: A). The neurite length of more than 100 cells was quantified in each case. A statistically significant 2.25-fold increase in neurite length upon differentiation was observed (Figure 3.2: B).

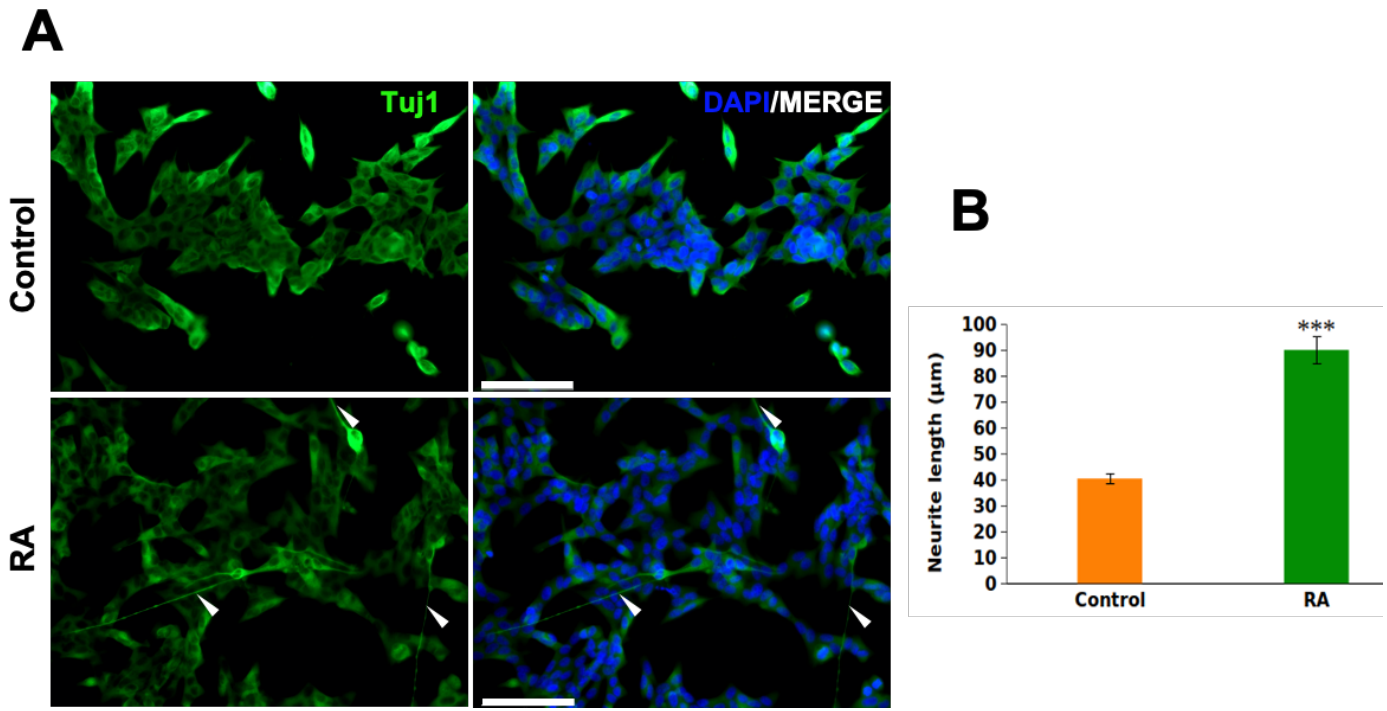


Figure 3.2: Treatment of SH-SY5Y with RA for 3 days induces a significant extension of neurites. (A) Representative fluorescence microscope images of Control and RA treated cells, stained for neuronal marker β-III tubulin (TuJ1) (1:50) and DAPI (nuclei) show that RA treated cells have more extended neurites (white arrows). Scale bar: 100 µm (B) Quantification of neurite length in Control and RA treated cells shows a significant 2.25-fold increase in neurite length upon differentiation (Student's t test, N=3, n>300, $p<0.001$).

One of the target genes of RA signaling cascade is c-Fos, involved in several cellular functions. c-Fos has been known as marker of neuronal activity, appears to play a key role in normal development of Neural Precursor Cells (NPCs) and is essential for neocortex development (Velazquez et al., 2015; Chung, 2015). c-Fos expression is induced upon RA neural induction, in an indirect, CREB dependent manner. Once FOS protein is produced it is transported back to the nucleus, where it interacts with c-Jun to form the heterodimer complex AP-1, which acts as a transcription factor for other differentiation related genes (Cañón et al., 2004; Chung, 2015). To determine if the RA signaling cascade was fully activated c-Fos expression was assessed. Control and RA-treated SH-SY5Y cells were stained with anti-c-Fos antibody after 3 days of differentiation with 30 μM RA (Figure 3.3: A). The proportion of cells expressing c-Fos (c-Fos+) out of the total number of cells (quantified by DAPI nuclear stain) was quantified. The percentage of c-Fos+ cells was significantly increased by RA treatment, by 2.45-fold, as expected, indicating that the cells underwent RA induced neuronal differentiation (Figure 3.3: B).

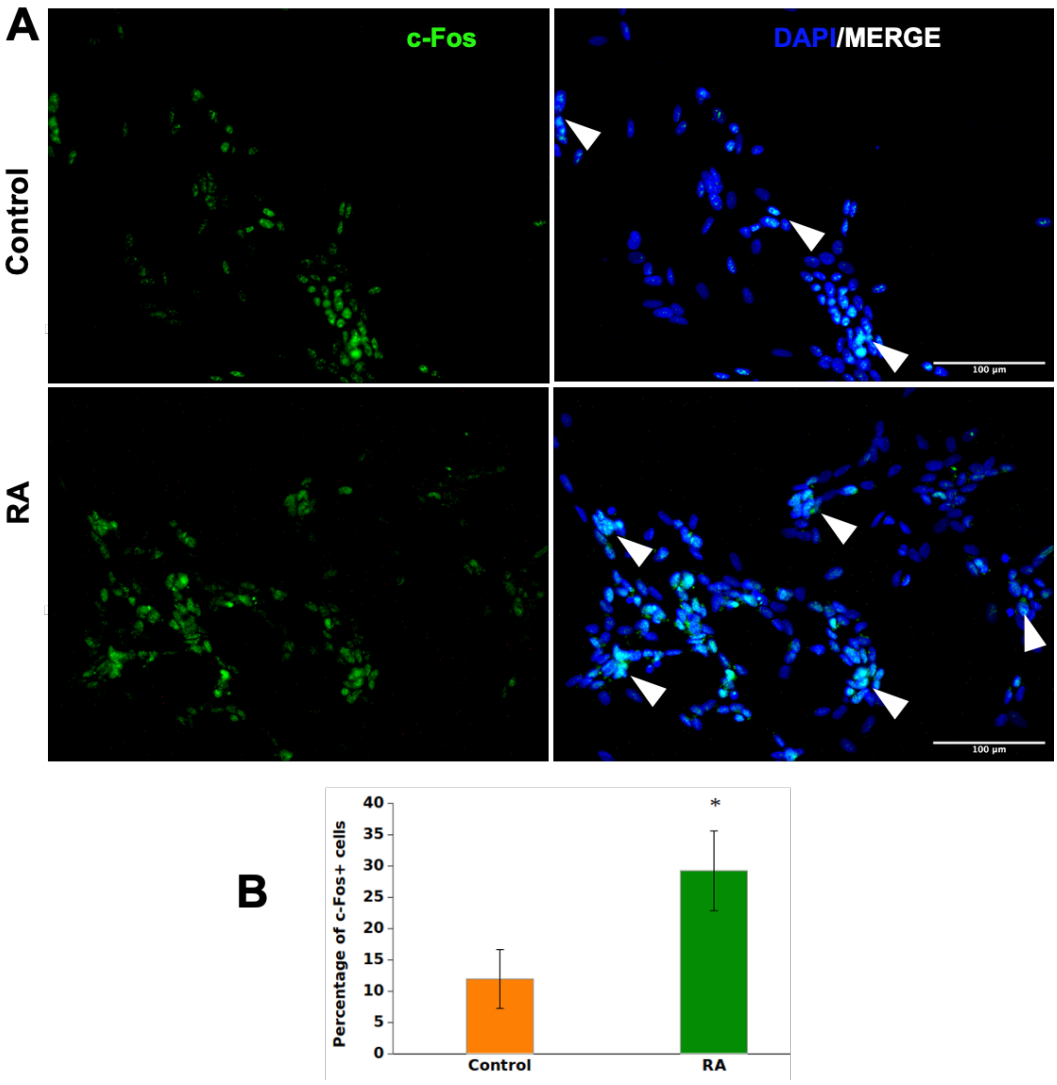


Figure 3.3: Expression of the differentiation marker c-Fos is upregulated upon RA treatment. (A) Representative fluorescence microscope images of Control and RA treated cells stained for neuronal marker c-Fos (1:50) and DAPI (nuclei) show an increased number of c-Fos expressing cells upon treatment with RA. Scale bar: 100µm (B) Quantification of the number of c-Fos+ cells shows that the percentage of cells expressing c-Fos, out of the total number of cells (DAPI+) increases significantly upon differentiation (Student's t-test, N=3, n=300, $p<0.05$, error bars depicting standard error).

3.2.2 Quantification of mRNA levels of the differentiation markers RET and MOXD1 by qPCR.

Since c-Fos expression is directly regulated by RA metabolic pathway it was possible that increased c-Fos expression resulted directly from RA treatment rather than from properly coordinated differentiation. Therefore, to further validate the differentiation of the cells mRNA levels of a selection of differentiation markers were measured. Previous studies show that SH-SY5Y cells express noradrenergic, as well as early and mature dopaminergic markers (Korecka et al., 2013; Forster et al., 2016)

Receptor tyrosine protein kinase (RET) was selected as a target gene because it is a mature marker of all dopaminergic neurons (Arenas et al., 2015) and is involved in axon guidance, cell migration and cell differentiation. RET is the canonical receptor of Glial Derived Neurotrophic Factor (GDNF) family ligands, and plays a pivotal role in the development, maintenance and regeneration of the dopaminergic system. It has been previously reported that RET is upregulated during RA induced differentiation of SH-SY5Y (Forster et al., 2016). To determine the changes in RET level after RA treatment, RET mRNA levels were quantified by RT-qPCR. RET mRNA levels showed a significant 2.75-fold increase upon differentiation, compared to Control (Figure 3.4).

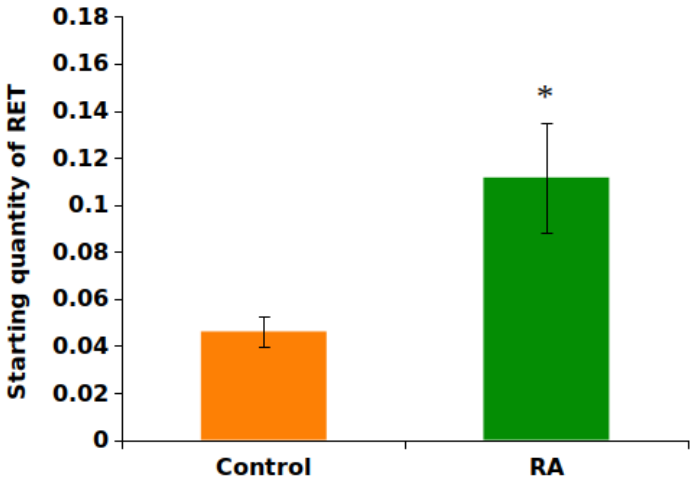


Figure 3.4: Receptor Tyrosine Protein Kinase (RET) mRNA levels increase upon RA induced differentiation. RET mRNA expression was quantified by qPCR, using the absolute quantification method. Starting quantity was estimated in Control and RA treated samples based on a standard curve of RET created from serial dilutions (1:5, 1:50, 1:500 and 1:5000) of pooled sample containing all the samples that were tested. A significant 2.75-fold increase is observed upon differentiation (Student's t test, n=2, $p<0.05$, error bars depicting standard error).

DBH-like monooxygenase protein 1 (MOXD1) was selected as a noradrenergic marker. It belongs to the family of monooxygenases and its catalytic core is homologous to that of Dopamine β-Hydroxylase (DBH). It is expressed in the human brain and is involved in the hydroxylation of dopamine to norepinephrine. It has also been implicated in neural crest development as MOX mRNA levels are increased in the developing chicken embryo neural crest cells (Xin et al., 2004). To examine alterations in MOXD1 mRNA expression levels, RT-qPCR was performed, in Control and RA treated cells. A significant 9-fold increase was observed in differentiated cells, compared to Control (Figure 3.5).

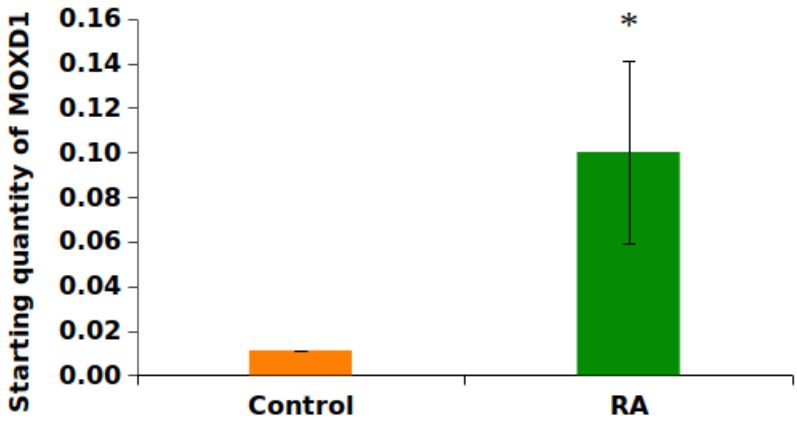


Figure 3.5: DBH-like monooxygenase protein 1 (MOXD1) mRNA levels increase upon RA induced differentiation. MOXD1 mRNA expression was quantified by qPCR, using the absolute quantification method. Starting quantity was estimated in Control and RA treated samples based on a standard curve of MOXD1 created from serial dilutions (1:5, 1:50, 1:500 and 1:5000) of pooled sample containing all the samples that were tested. A significant 9fold increase is observed upon differentiation (Student's t test, n=2, $p<0.05$, error bars depicting standard error).

Taken together, these results indicated that RA administration for 3 days at 30μM induces the differentiation of human neuroblastoma SH-SY5Y cells towards a dopaminergic/noradrenergic fate. This timeframe was suitable for this study's purpose, because the aim is to dissect the interactions of cytoplasmic lncRNAs with the translation machinery during the early stages of differentiation. However, it is important to stress that after 3 days, the cells are still far from being terminally differentiated. This was evident from their morphology, the lack of expression of terminal differentiation markers such as Synaptic vesicle protein 2 (Sv2) (Figure 3.6).

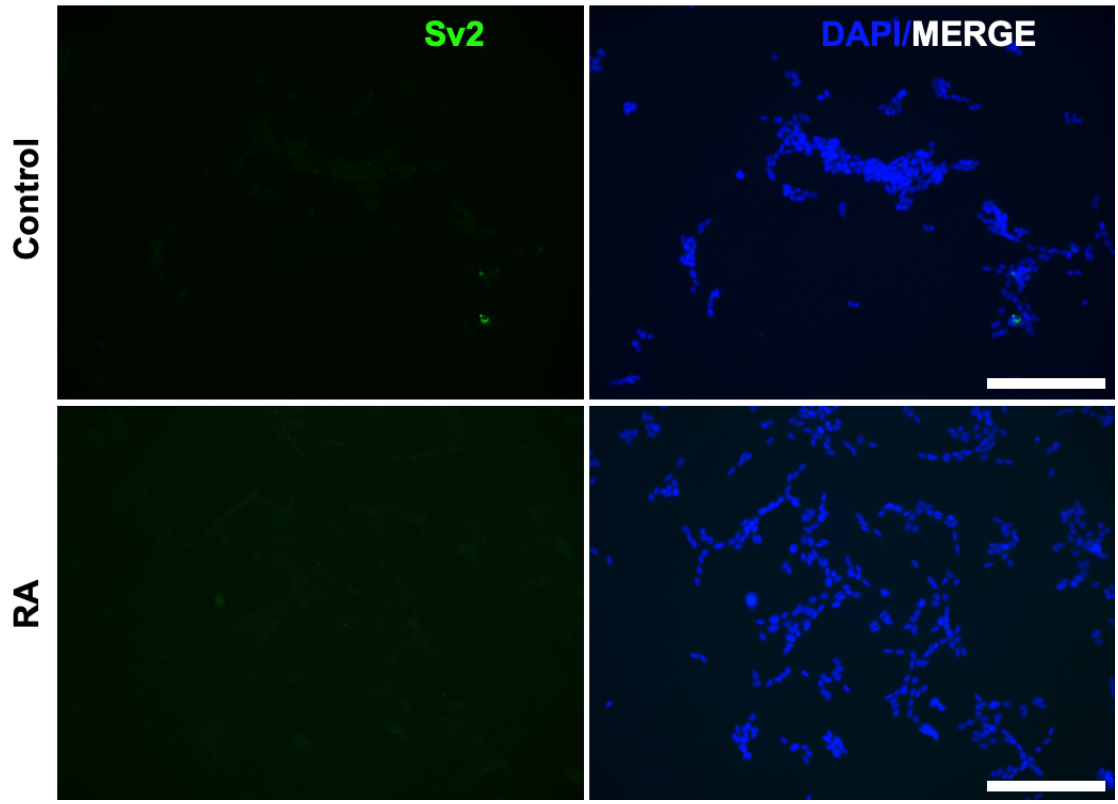


Figure 3.6: Synaptic vesicle protein 2 (Sv2) is not expressed in RA-treated SH-SY5Y cells. Representative fluorescence microscope images of Control and RA treated cells stained for neuronal marker Sv2 (1:50) and DAPI (nuclei) show no expression of Sv2 in Control cells and upon treatment with RA for 3 days. Scale bar: 250µm.

3.2.3 Assessment of cell proliferation and cell viability

An assessment of cell viability and proliferation rate of the cultured cells upon differentiation was important for monitoring the progress of differentiation. To assess the proliferation of SH-SY5Y cells, proliferation marker ki67 was assessed, which is a widely used for the identification of dividing cells because its levels increase specifically during S phase (Miller et al., 2018). Undifferentiated and differentiated cells were stained for ki67 and the number of proliferating cells (ki67+) was quantified. As expected, the proportion of ki67+ cells was substantially higher in undifferentiated cells (Control) compared to differentiated (RA) (Figure 3.7: A). The proportion of proliferating cells decreased 2.7-fold upon differentiation (Figure 3.7: B).

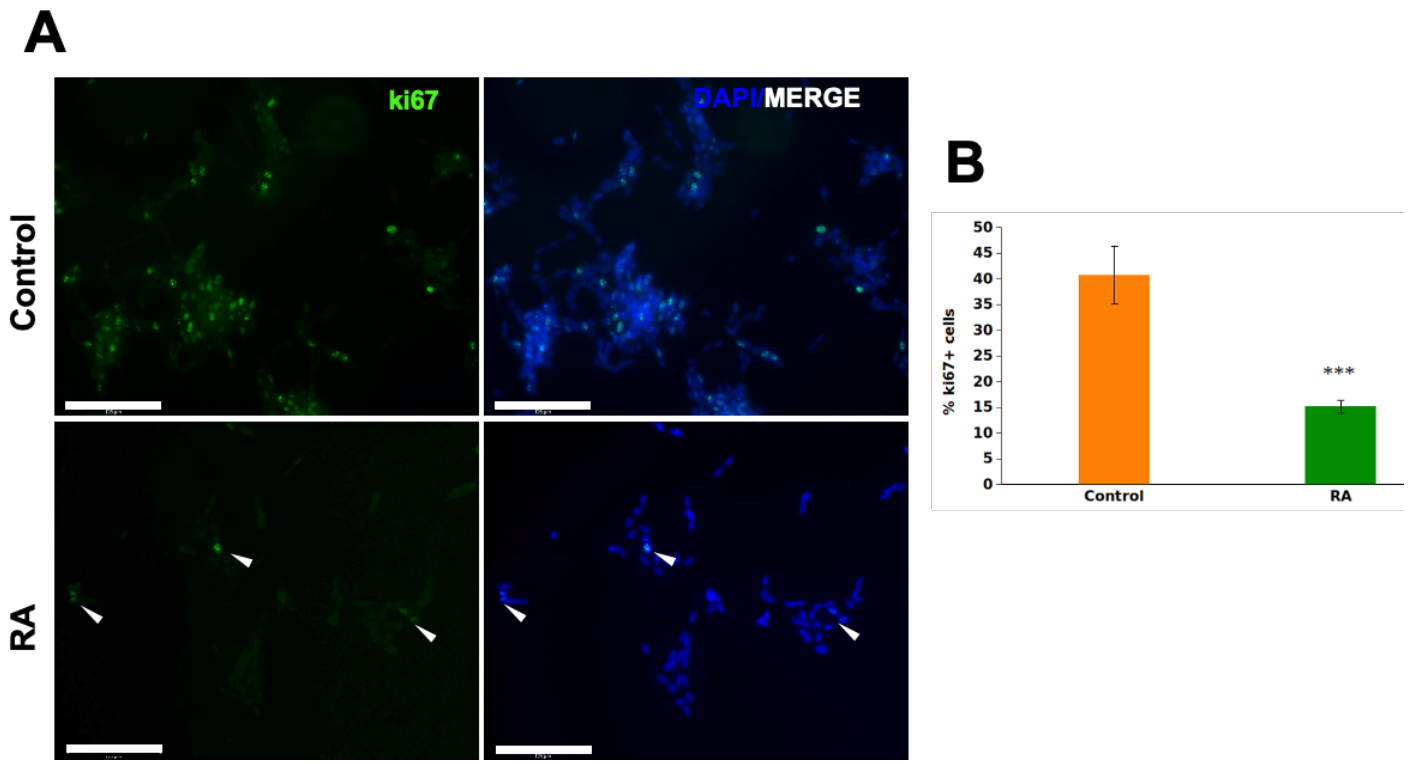


Figure 3.7: RA treatment of SH-SY5Y cells results in a significant reduction of ki67+ cells. (A) Representative fluorescence microscope images of Control and RA treated cells stained for S-phase marker ki67 (1:100) and DAPI (nuclei) show a decreased proportion of ki67 expressing cells upon treatment with RA. (B) Quantification of the ki67+ cells shows that the percentage of cells in S-phase, out of the total number of cells (DAPI+) significantly decreases upon differentiation (Student's t test, N=3, n=1000, $p<0.001$, error bars depicting standard error). Scale bar=125μm.

E2F1 is one of the major regulators of cell cycle progression, promoting the transcription of the genes that govern the G1 to S transition. Therefore, to confirm proliferation levels, RT-qPCR was performed for E2F1 in the same samples used for quantification of RET and MOXD1. Upon differentiation there is a significant reduction in the levels of E2F1 mRNA, consistent with a reduction in the number of proliferating cells (Figure 3.8).

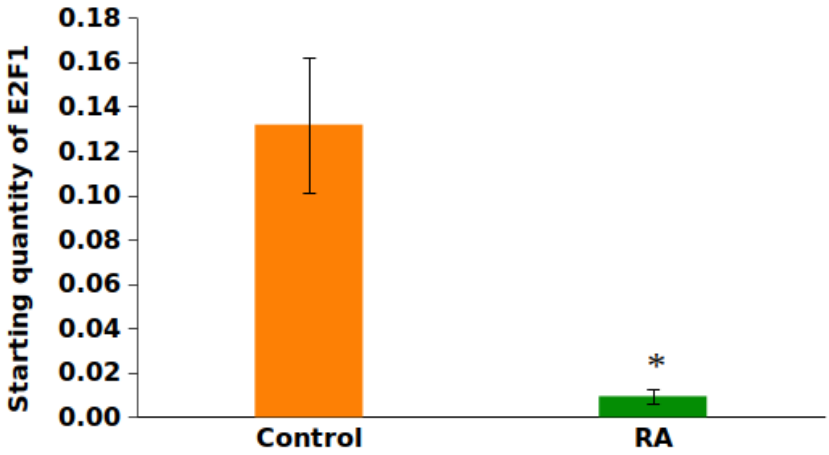


Figure 3.8: E2F1 mRNA levels are reduced upon RA induced differentiation. E2F1 mRNA expression was quantified by RT-qPCR in the same samples used for quantification of RET and MOXD1, using the absolute quantification method. Starting quantity was estimated in Control and RA treated samples based on a standard curve of E2F1 created from serial dilutions (1:5, 1:50, 1:500 and 1:5000) of pooled sample containing all the samples that were tested. A significant 14-fold decrease is observed upon differentiation (Student's t test, n=2, $p<0.05$, error bars depicting standard error).

Based on these results, I confirmed that RA treatment leads to reduction of proliferation and differentiation towards neuronal phenotype. To ensure that the decrease in proliferation was not attributable to any toxicity resulting from RA treatment, cell viability was assessed three days post RA treatment by Trypan Blue Assay in 3 independent replicates (Methods 2.1.3). In Control cultures ~1.7 $x10^6$ cells/mL on average from the 3 replicates, were viable after 3 days. In RA treated cultures, the number of viable cells 3 days post-differentiation was 1.28 $x10^6$ cells/mL on average. The number of dead cells for Control and RA treated cultures was on average 0.25 $x10^6$ cells/mL and 0.2 $x10^6$ cells/mL, respectively. Therefore, in both cases (Control and RA treated) the number of dead cells was similar and the percentage of viable cells post-differentiation in each culture exceeded 85%, indicative of healthy growing cell cultures (Figure 3.9).

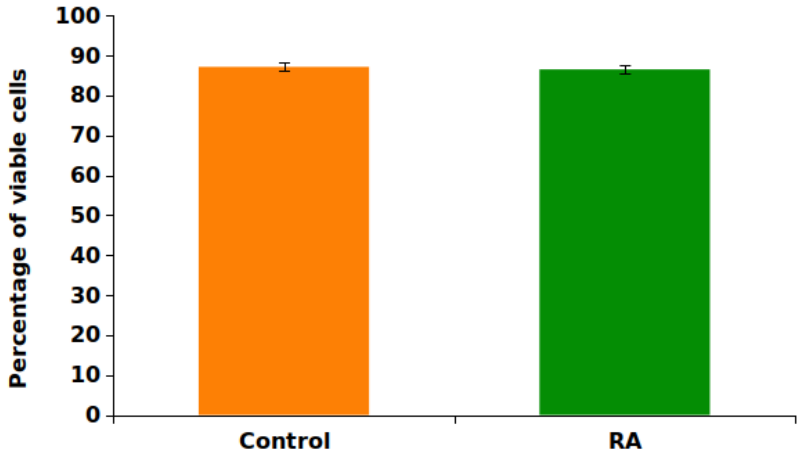


Figure 3.9: RA treatment does not affect cell viability. 84.4% of RA treated and 85.22% of Control (DMSO) treated cells were alive after 3 days of treatment. ~1.7 x10^6 viable cells/mL on average in Control cultures, after 3 days. In RA treated cultures, the number of viable cells 3 days post-differentiation was 1.28 x10^6 cells/mL on average (experiment performed in 3 biological replicates). The percentage of viable cells after 3 days of differentiation is similar to that of undifferentiated cells (Student's t test, n=3, $p>0.05$). Error bars represent standard error.

3.3 Neuronal differentiation induces a reduction in the level of active translation.

SH-SY5Y cells differentiated by RA administration for 3 days showed the expected morphological changes in the neurite length as well as increased expression of three markers of neuronal differentiation. Cell proliferation was reduced, as determined by the number of ki67+ cells and the decrease in the expression of cell cycle promoting transcription factor E2F1. Undifferentiated and differentiated cells showed comparable levels of cell viability. Therefore, results showed that the SH-SY5Y cells provided a useful model for differentiation.

A recent study has shown that active translation is repressed and precisely regulated by 3' and 5' UTRs during neuronal differentiation (Blair et al., 2017). In order to understand how neuronal differentiation might affect translation in SH-SY5Y cells, I performed polysome profiling on undifferentiated and differentiated cells. Polysome profiling is the most widely used method to determine the general translation status of a given cell population and its level of active translation (Murn et al., 2015; Chassé et al., 2016). This approach is based on the differential sedimentation of ribosomal subunits (40S and 60S), monosomes (80S) and polysomes upon sucrose density gradient ultracentrifugation. Fractionation of the sucrose gradients and simultaneous measurement of the optical density (OD) of the RNAs associated with the ribosomes produces the polysome profile (Figure 3.10).

To capture the ribosomes at the point of elongation, Control and RA differentiated cells were treated with cycloheximide, prior to harvesting. Cycloheximide is a eukaryotic protein synthesis inhibitor, which binds to the ribosome and inhibits the eEF2 mediated translocation step (Obrig et al., 1971; Schneider-Poetsch et al., 2010). Cytoplasmic lysates were prepared (Methods 2.5.2) and subjected to ultracentrifugation. In each independent experiment (biological replicate), 3 gradients of Control and 3 gradients of RA treated lysates were analysed (technical replicates). To better study the effect of differentiation on active translation, actively translating polysomes from monosomes need to be separated using sucrose density gradients of 18% to 60%, previously optimised for the better separation of polysomes from monosomes (Aspden et al., 2014).

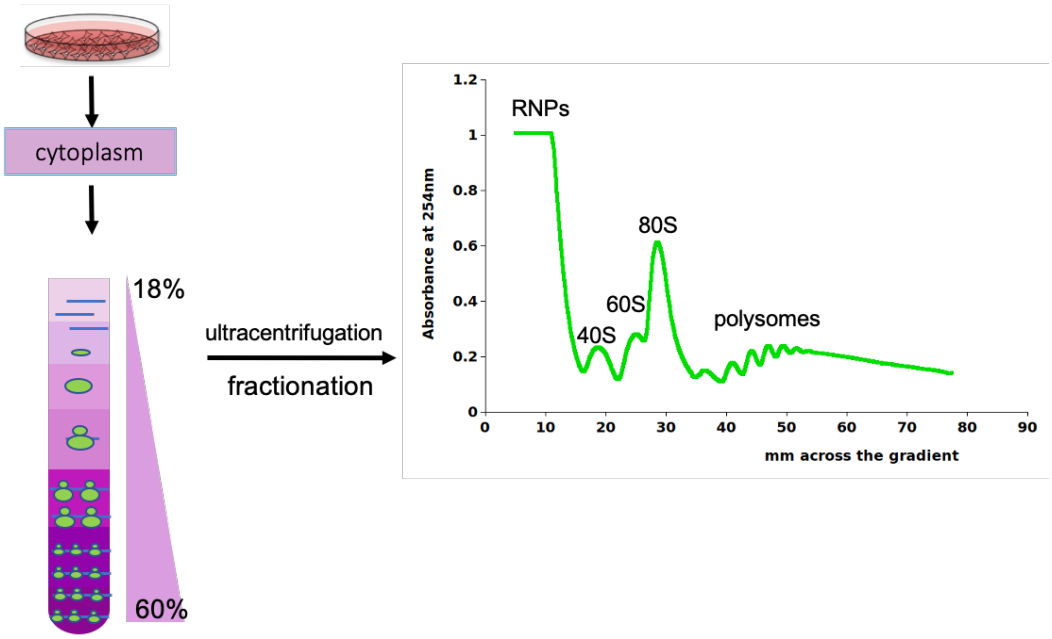


Figure 3.10: Schematic of polysome profiling. Cells were harvested and lysed (Methods 2.5.2). Cytoplasmic lysates were loaded on sucrose gradients (18%-60%) and subjected to ultracentrifugation. Gradients were fractionated, and RNA levels in each fraction were simultaneously quantified by UV absorbance. Ribonucleoprotein complexes (RNPs) occupy the top of the gradient, 40S and 60S subunits migrate ~20-30mm across the gradient, 80S at ~30-35mm and polysomes sediment at the bottom of the gradient.

To examine how differentiation affected translation within the SH-5Y5Y model I compared polysome profiles of differentiated (RA) and undifferentiated (Control) cells. The absorbance at 254nm is measured across the gradient and the peaks reveal the distribution of 40S, 60S subunits, 80S monosomes and polysomes. The UV profile of the differentiated cells showed some substantial differences compared to that of controls. RA treated cells displayed a higher 80S peak and lower polysome peaks (Figure 3.11). To determine whether these differences were significant, the area under the curves (Figure 3.12) for each of the peaks was measured in three technical replicates of three biological replicates for the two conditions (Control and RA).

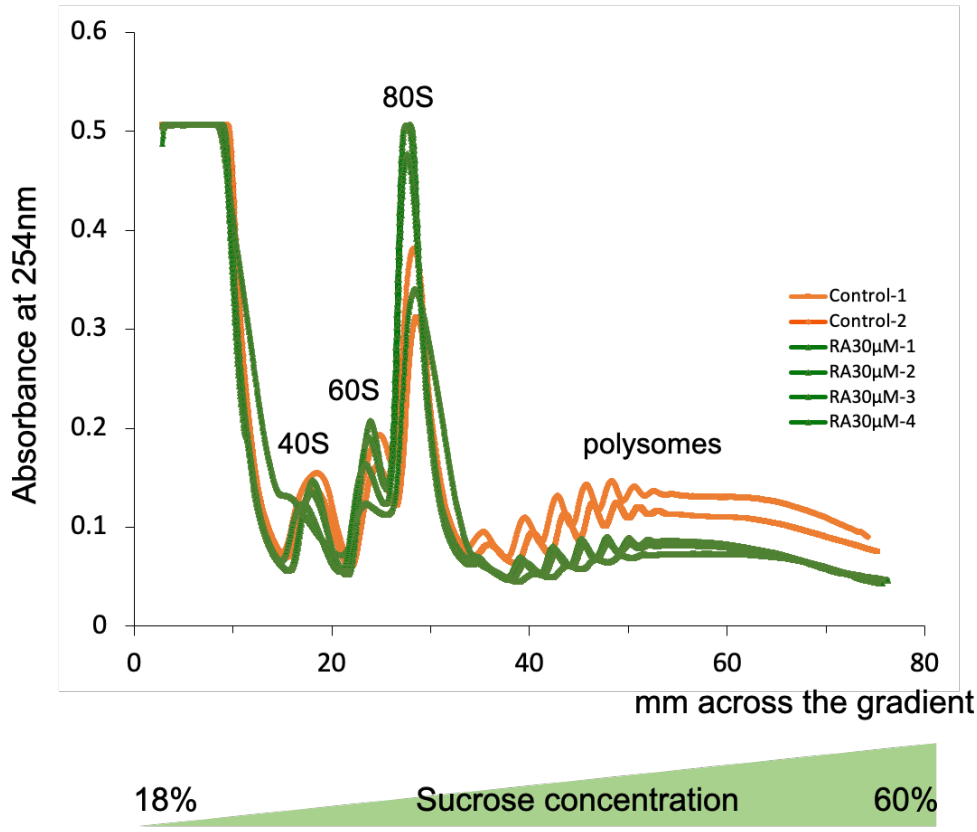


Figure 3.11: Differentiation of SH-SY5Y shows a global reduction in the level of active translation. Polysome profiles of undifferentiated Control (orange) and RA differentiated (green) cells. Cell lysates were loaded on sucrose density gradients of 18%-60% (~70 x10^6 cells per Control gradient, ~42 x10^6 cells per RA gradient) and were subjected to ultracentrifugation (121,355 x g_{avg} for 3.5 h, 4°C). For each biological replicate 2 gradients of Control and 4 gradients of RA treated cells were processed. RNA absorbance at 254nm is measured across the sucrose gradient (mm) and the distinct peaks correspond to ribosomal subunits (40S and 60S), monosome (80S) and polysomes.

Both the decrease in the polysome area and the increase in the 80S area under the curve upon differentiation were statistically significant (Figure 3.12). To my knowledge, this is the first report of this phenomenon in SH-SY5Y cells. A widely used metric of the level of active translation is the ratio of the polysome to monosome area (P/M ratio) (Chassé et al., 2016). Therefore, the P/M ratio in Control and RA treated cells was calculated, which showed a significant reduction of the ratio upon differentiation (Figure 3.13). These results indicate that neuronal differentiation of SH-SY5Y cells leads to a reduction in the proportion of actively translating polysomes. At the same time, the increase in the area of the 80S peak upon RA treatment indicates that more RNAs are engaged by monosomes in differentiated cells compared to Control.

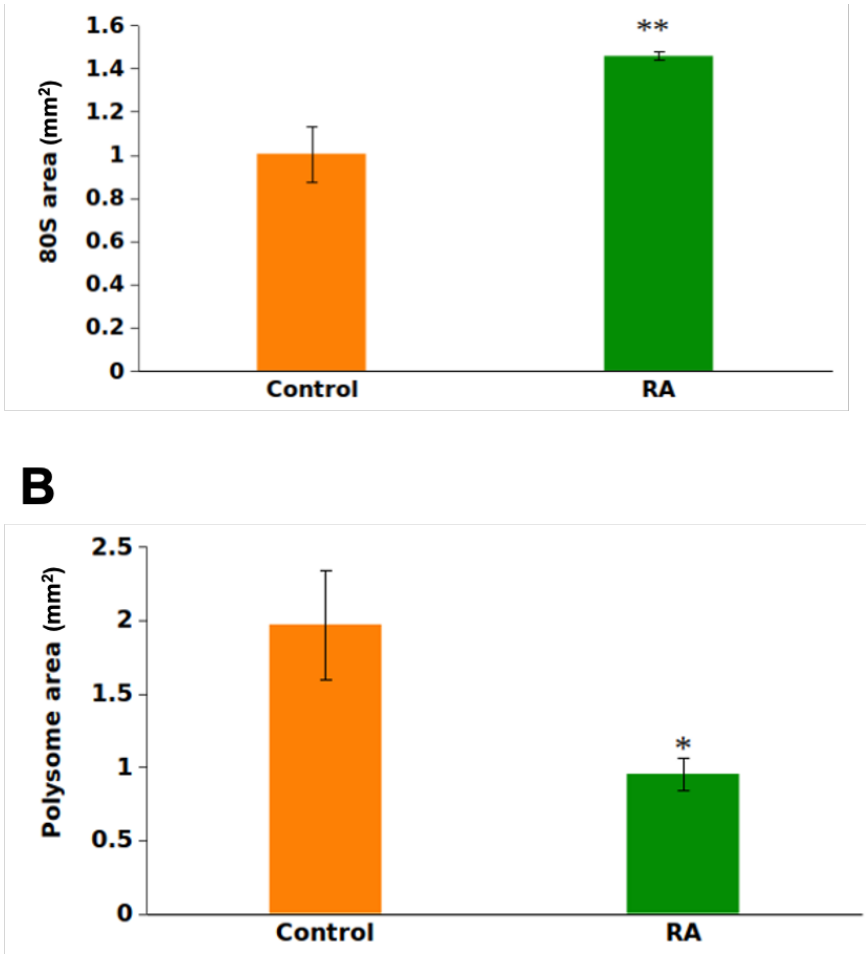


Figure 3.12: Neuronal differentiation induces significant changes in the translation profile of SH-SY5Y cells. (A) 80S area (mm^2) under the curve is significantly increased by 1.5-fold upon differentiation (Student's t test, n=3, $p<0.01$). (B) Polysome area under the curve shows a significant 2-fold reduction in RA treated cells compared to Control (Student's t test, n=3, $p<0.01$). Error bars represent standard error.

The translational status of the 80S has been a subject of contradictory findings. It has been shown that the monosome fraction contains a large number of inactive ribosomes that do not engage on mRNAs to direct active translation (Liu and Qian, 2016). However, more recent ribosome profiling studies in *S. cerevisae* and *Rattus norvegicus* prove that 80S is actually actively translating mRNAs encoding regulatory proteins and synaptic mRNAs, respectively (Heyer and Moore, 2016; Biever et al., 2020). The process of differentiation is associated with tight control of transcription and translation, in order to ensure that the necessary transcripts (eg c-Fos) are translated in adequate level, while reducing the translation level of other non-essential transcripts (eg ki67), at each stage. Potentially, the translation of those non-essential transcripts could be impeded following their association with 80S. A study in yeast has proved that translation can be inhibited after the formation of an 80S complex (Balagopal and Parker, 2011). It is also possible that the increase in the proportion of monosomes is due to the translation of regulatory proteins, a plausible result of translation regulation during the transition to a more differentiated phenotype.

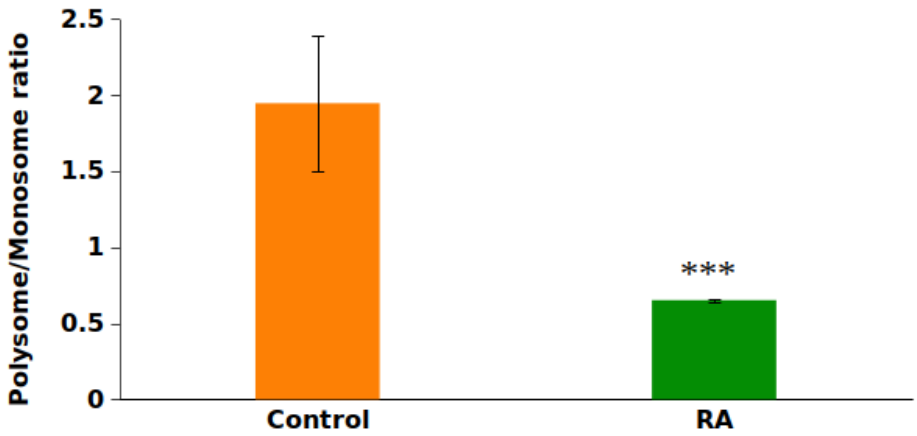


Figure 3.13: Active translation is significantly reduced in differentiated cells. Polysome/Monosome ratio is a proxy for measuring the level of active translation. P/M ratio of the area under the curve is significantly reduced by 3-fold upon differentiation in 3 independent biological replicates (Student's t test, n=3, $p<0.001$). Error bars represent standard error.

3.4 Optimisation of the Poly-Ribo-Seq protocol for SH-SY5Y cells

Having established that neuronal differentiation induces a reduction in active translation, I set out to identify the different populations of cytoplasmic lncRNAs in relation to the translation machinery using Poly-Ribo-Seq. Poly-Ribo-Seq was developed in *D. melanogaster* S2 cells (Figure 3.14). Therefore, it was necessary to optimise certain steps (Figure 3.15) of the original protocol (Aspden et al., 2014) for use with human SH-SY5Y cells.

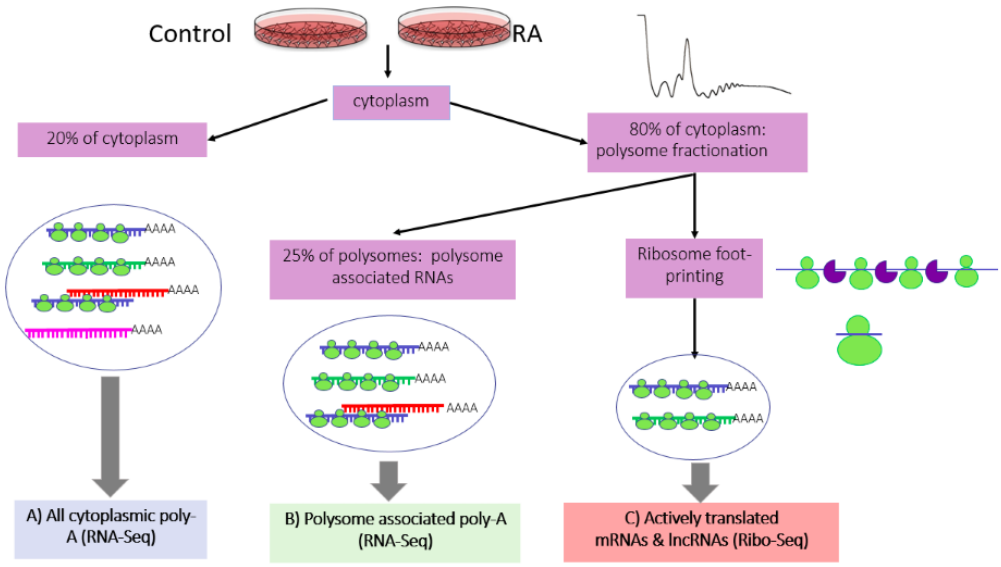


Figure 3.14: Brief schematic of Poly-Ribo-Seq method. Cells are lysed, nuclei are pelleted, and the cytoplasmic lysate is divided into 3 populations. 20% of the cytoplasm is processed for RNA-Seq (total cytoplasmic RNA). The 80% of the cytoplasm is subjected to sucrose density gradient ultracentrifugation. Following fractionation, polysome fractions from each condition are merged and 25% of this material is also processed for RNA-Seq (polysome-associated RNA). The rest of the polysome fractions is subjected to RNase I footprinting and processed for Ribo-Seq.

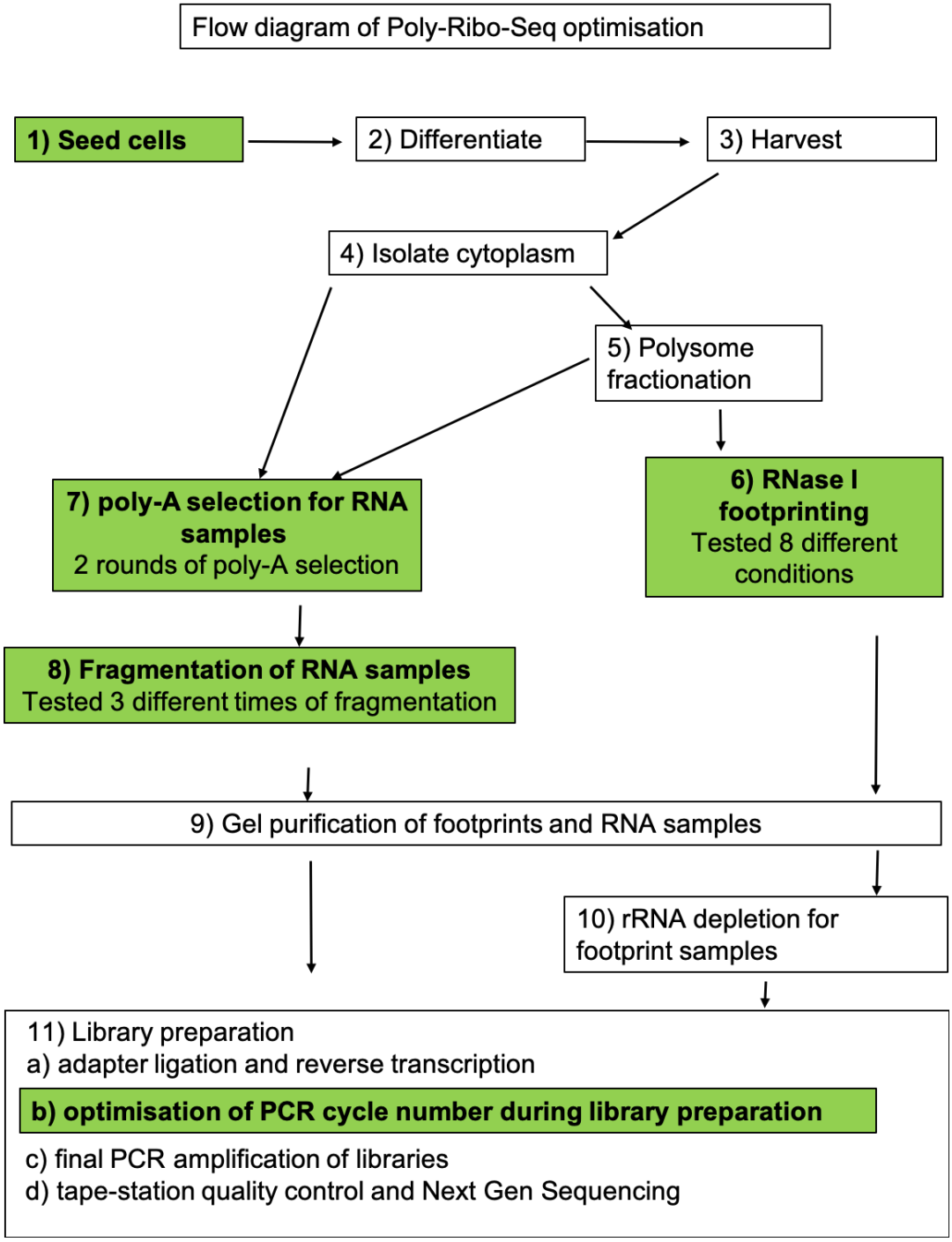


Figure 3.15: Flow-diagram of the basic Poly-Ribo-Seq steps. The steps optimised specifically for SH-SY5Y cells are highlighted in green.

3.4.1 Optimisation of the number of cells seeded prior to the experiment

The Poly-Ribo-Seq protocol has many steps downstream of polysome fractionation, including several ethanol precipitations. In order to have sufficient RNA to construct libraries with, it is necessary to start with a sufficient number of cells. S2 cells exhibit high translation levels and large amounts of polysome material can be obtained from a relatively small number of plates. SH-SY5Y cells have approximately 4 times the volume of S2 cells and display lower levels of translation. Therefore, the same number of culture plates would yield less material. In order to determine the appropriate number of SH-SY5Y cells for Poly-Ribo-Seq (Figure 3.15-step 1) cells were seeded at three different densities (Table 3.1). After three days in culture, cells were harvested, counted the and RNA extracted.

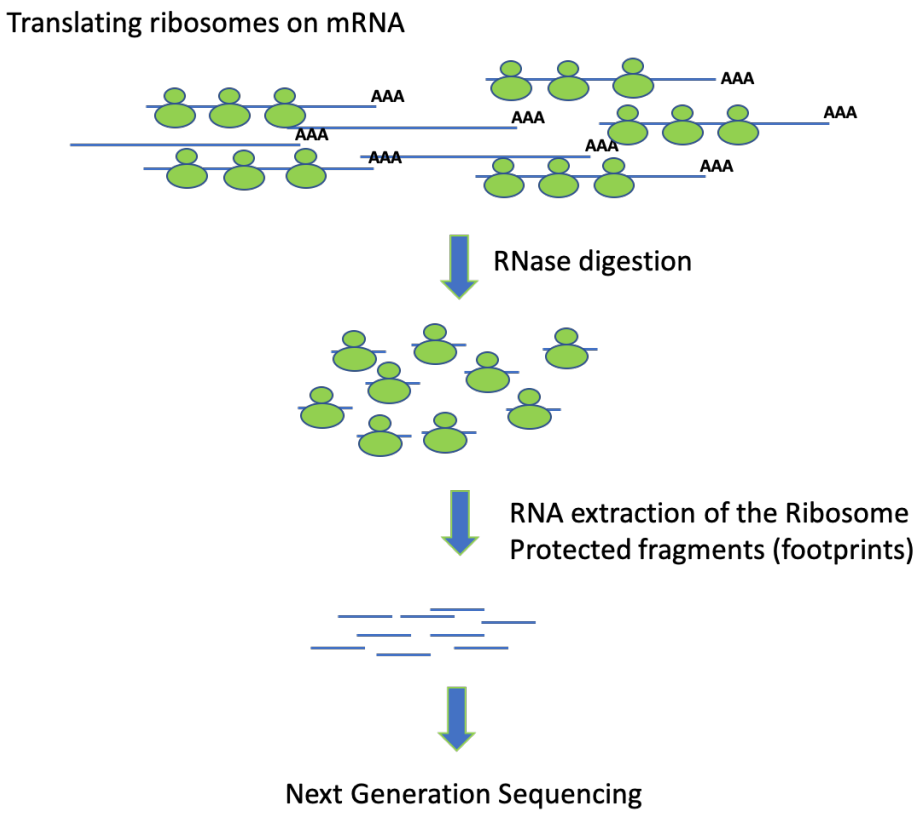
Seeding 4.5 $x10^6$ (medium density) and 6 $x10^6$ (high density) cells yielded approximately the same number of total cells and similar amounts of RNA (Table 3.1). However, the cells in higher density were very dense and therefore more likely to be stressed. It is crucial for the cells not to be stressed as this has a negative impact on the level of translation and the quality of polysome profile (Buchan and Parker, 2009). Thus, optimal number of cells to be seeded per plate was 4.5 $x10^6$. Cell viability assay (3.1.3) has shown that the number of cells, 3 days post-RA treatment is ~1.5-fold lower compared to Control. Therefore, in order to obtain a similar number of cells from both conditions, twice as many plates were seeded for RA treatment.

Table 3.1: Number of cells seeded to test Poly-Ribo-Seq starting material

Initial number of cells plated	Total number of cells harvested	RNA yield (µg)
3 x10^6	6.6 x10^6	12.3
4.5 x10^6	15 x10^6	13.2
6 x10^6	17 x10^6	13.1

3.4.2 Initial optimisation of the RNase I footprinting

Ribosome footprinting is performed to determine the precise position of ribosomes on RNAs (Figure 3.16). This is achieved by RNase I digestion of the polysome complexes derived after the fractionation (Methods 2.5.3). RNase I footprinting precision is critical for the quality of ribosome profiling and the optimal footprinting conditions vary between species (Ingolia et al., 2009; Guo et al., 2010; Ingolia et al., 2013; Duncan and Mata, 2014; Aspden et al., 2014; Chung et al., 2015; Heyer and Moore, 2016; Hsu et al., 2016; McGlincy and Ingolia, 2017).



Adapted from Ingolia, 2016

Figure 3.16: Schematic of RNase I footprinting, adapted from Ingolia, 2016. Translating ribosomes are treated with RNase I to digest the mRNA fragments between ribosomes and leave the ribosome protected fragments (RPFs). RNA is extracted from RPFs and processed for deep sequencing. Reads obtained reflect the precise location of the ribosomes on the mRNA ORF.

Based on literature (Ingolia et al., 2009; Guo et al., 2010; Aspden et al., 2014), I initially decided to test two different RNase I concentrations (10 U/million cells and 20 U/million cells) and two different temperatures (4^{o}C overnight and RT for 1 hour). After fractionation (Figure 3.15: step 6), polysome fractions were combined and dilution buffer was added to dilute sucrose down to 10% prior to the addition of RNase I. After digestion, RNase I was inactivated by the addition of RNase I inhibitor. Samples were concentrated and subjected to ultracentrifugation (204,428 x g_{avg} at 4^{o}C for 4h) to purify ribosome protected fragments (RPFs) from RNA and smaller proteins. To assess the efficiency of footprinting, samples were run on a denaturing polyacrylamide gel and the size and amount of footprinting was monitored (Figure 3.15: step 9). A sample not subjected to RNase I treatment was used as a negative control (Figure 3.17). Ribosome footprints are detected between the 28 and 34nt RNA markers, as the approximate size of a ribosome protected fragment is 30nt. More effective footprinting was achieved with RNase I treatment at room temperature (RT) compared to overnight treatment at 4^{o}C. There was no difference between 10 and 20U. Negative control did not show any footprinting. Therefore, going forward, RNase I treatment was performed with 10 U/million cells of RNase I at RT for 1h.

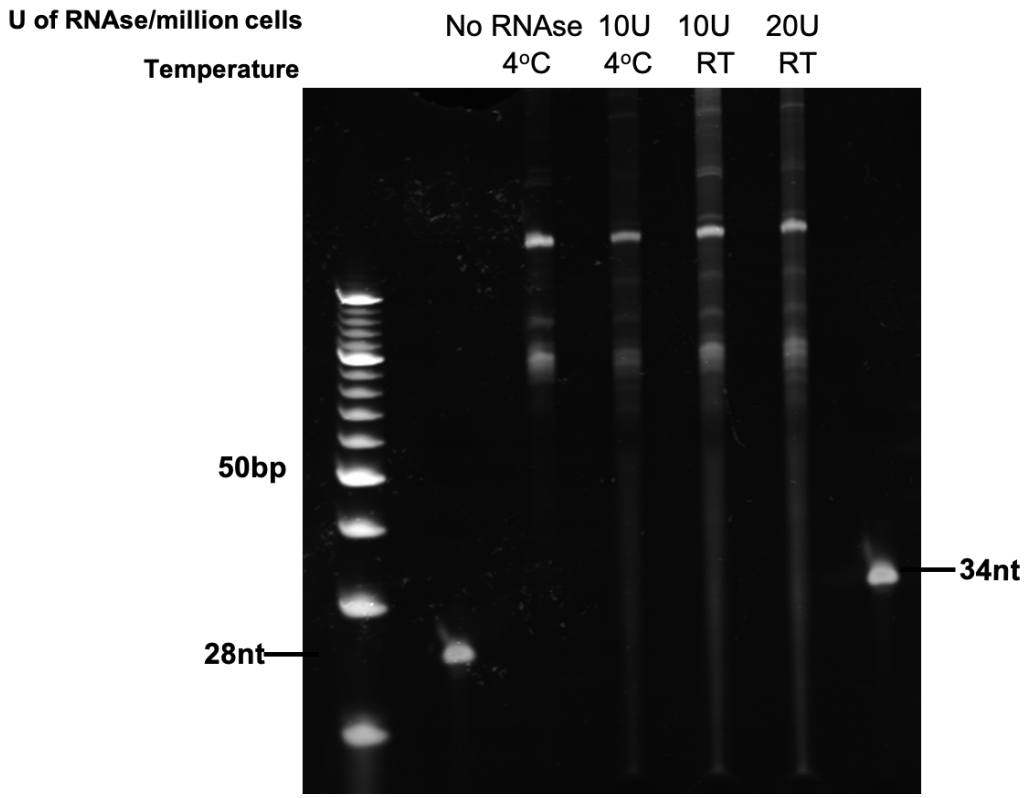


Figure 3.17: Initial optimisation of RNase I footprinting conditions. RNA samples derived from polysome fractions, subjected to RNase I treatment at different enzyme concentrations, temperatures and for different incubation periods to create footprints (visible between 28 and 34nt markers). Lane 1: No RNase I treatment, Lane 2: RNase I treatment with 10U/million cells at 4°C overnight, Lane 3: RNase I treatment with 10 U/million cells at RT for 1h, Lane 4: RNase I treatment with 20 U/million cells at RT for 1h. O'RangeRuler 10 bp DNA Ladder was used to aid the band size determination.

3.4.3 Optimisation of PCR cycle number during library preparation

The final stage of the library preparation (Figure 3.14-step 11) is the PCR amplification of the cDNA for each library. It is important that the number of amplicons is enough to be detectable by the sequencer, however excessive amplification contributes to the accumulation of PCR duplicates in the data. Therefore, the number of the amplification cycles needs to be optimised for each experimental system. To identify the optimal cycle number, PCR was initially performed on a small scale in order to determine which cycle number (11, 13 or 15) was optimal (Figure 3.18).

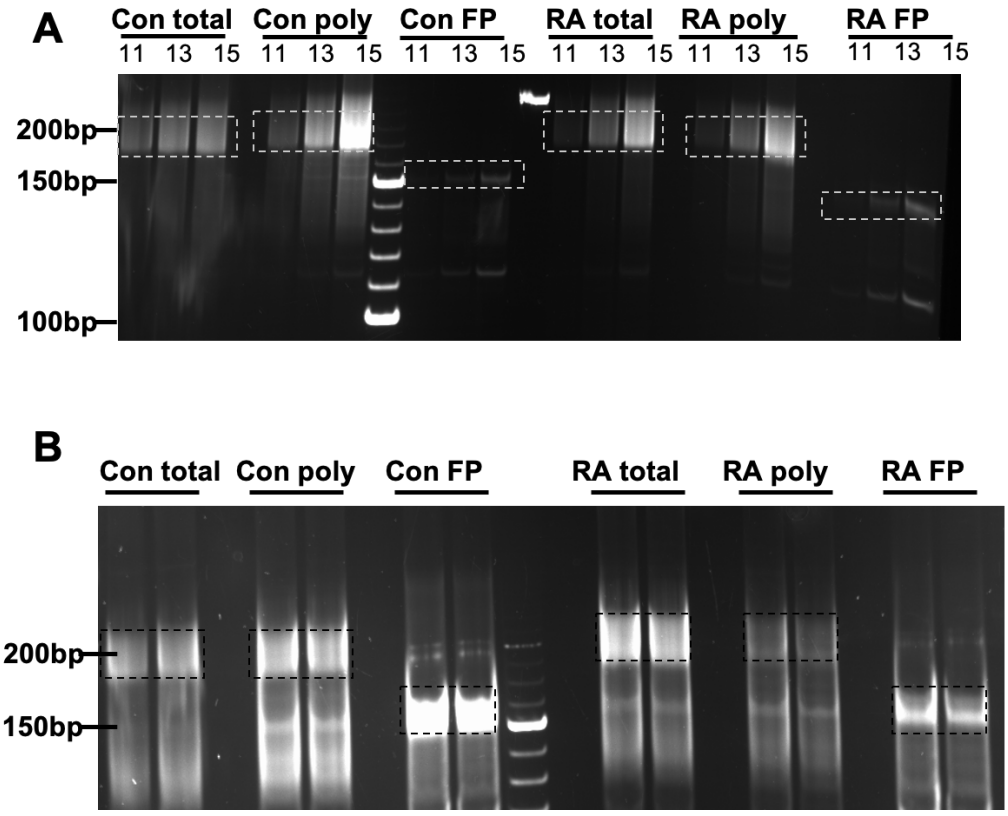


Figure 3.18: Quality control PCR for amplification of the reverse-transcribed footprints (FP) and poly-A selected RNAs from total cytoplasm (total) and polysome fractions (polysome) of undifferentiated (Control) and differentiated (RA) cells. (A) Small-scale PCR to determine the optimal number of cycles. 11, 13 and 15 cycles of PCR amplification were performed for each sample in separate reactions and 15 cycles was chosen as the bands appear brighter (grey dotted rectangles). (B) Final PCR amplification of the reverse transcription products for 15 cycles (black dotted rectangles). Samples were separated by gel electrophoresis in a 10% (w/v) non-denaturing polyacrylamide gel and an O'RangeRuler 10 bp DNA Ladder was used to aid the band size determination.

At this stage the size of the footprints is ~150 bp and the size of the total cytoplasmic and poly-A selected RNAs approximately 200 bp (Figure 3.17). Adapters at either side of the RNA are 21nt (3'end adapter) and 26nt (5'end adapter) each. The SR primer is 49nt long, of which 29nt are overhang allowing the attachment to the flow-cell. The index primer is 64nt long, of which the 43nt are overhang. PCR amplification for 11 and 13 cycles yielded very low and medium intensity bands respectively (Figure 3.18: A). Therefore, amplification for 15 cycles was selected as the optimal, and used for all samples' library preps (Figure 3.18: B). After deep sequencing, data were processed for analysis as previously described (Methods 2.18).

3.4.4 Optimisation of the RNase I footprinting buffer conditions

To assess the quality of the ribosome profiling data, I examined the precision of RNase I footprinting. Ribosome footprints should be around 30nt in length (Ingolia et al., 2009). Therefore, the more precise the foot-printing is, the larger the number of footprint reads ~30nt. The other measure of footprinting quality is triplet periodicity. This is a reflection of how the ribosome moves and decodes in triplets (codons) while translating the mRNA. Ribosome profiling reads are expected to display a triplet periodicity, as opposed to classic RNA-Seq reads that do not display such a pattern. (Figure 3.19).

By mapping the 5' end of each ribosome profiling read to the human transcriptome one can determine the distance between the end of the read and the P site (start codon) or the A site (just before the stop codon) (Ingolia et al., 2009). Therefore, the number of reads that start at each nucleotide of the codon (frame) can be precisely determined. It is expected for ribosome profiling datasets that the majority of reads start at the same frame, reflecting the movement of the ribosome (Figure 3.19). To determine these features, the RiboSeqR (Hardcastle, 2014) pipeline was utilised on Ribogalaxy (Michel et al., 2016).

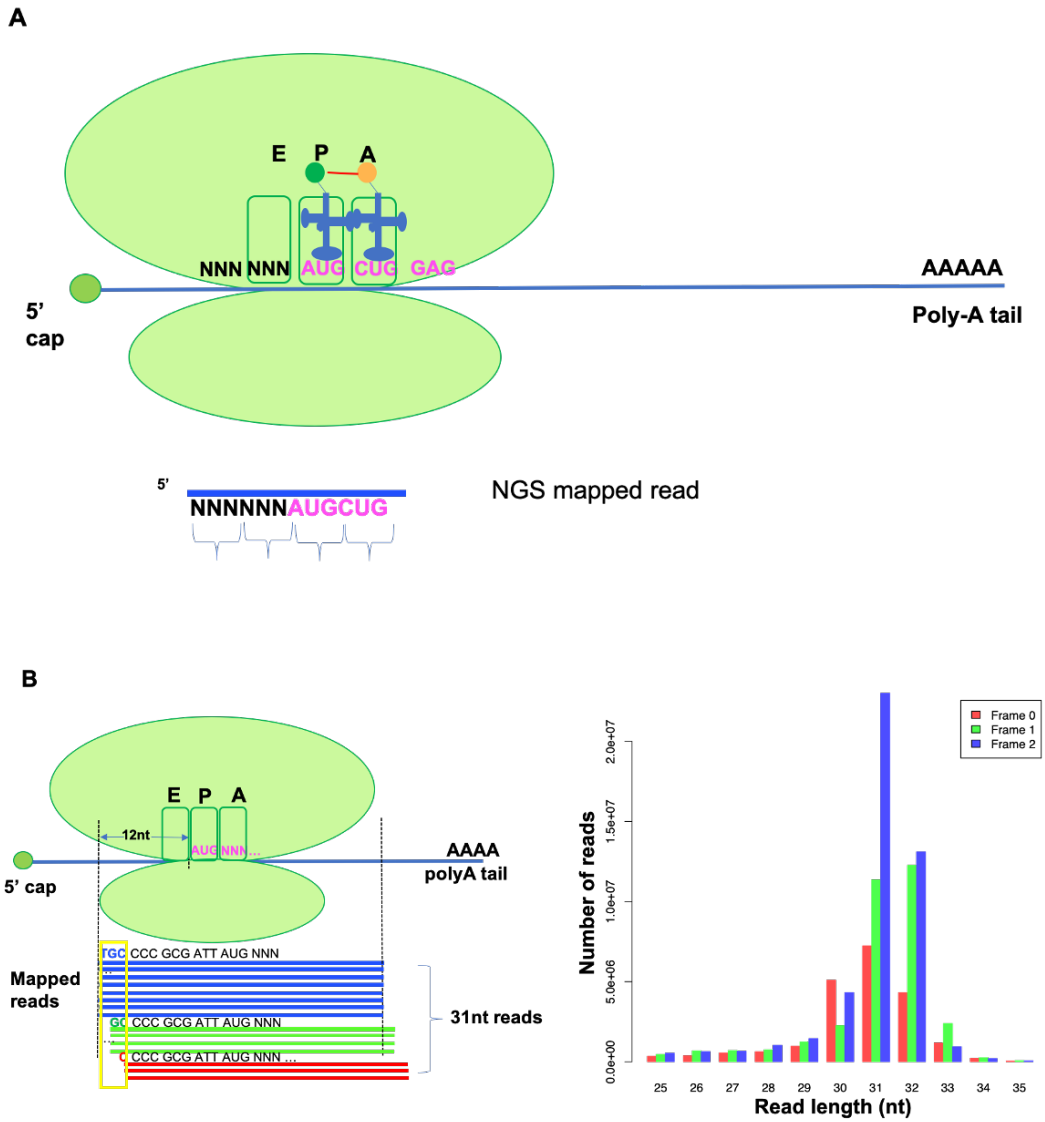


Figure 3.19: 5' ends of Ribosome profiling and Poly-Ribo Seq should display triplet periodicity. Schematic representation of the framing of ribosome profiling reads.

Ribosome profiling reads of the first Poly-Ribo-Seq experiment did not exhibit clear triplet periodicity (Figure 3.20). Read length varied substantially and approximately the same number of reads start at each of the different frames (Figure 3.20). Notably, the reads that are 33nt long show a clearer periodicity pattern than other length. That is to say that not all three frames were represented equally, more reads were in frame 1 than 0 and 2. However, the number of reads that were uniquely mapped to frame 1 was less than 40% of the total 33nt long reads. In RA treated footprint reads, there was no pattern of triplet periodicity (Figure 3.20: B). Therefore, further optimisation of the RNase I footprinting conditions was required.

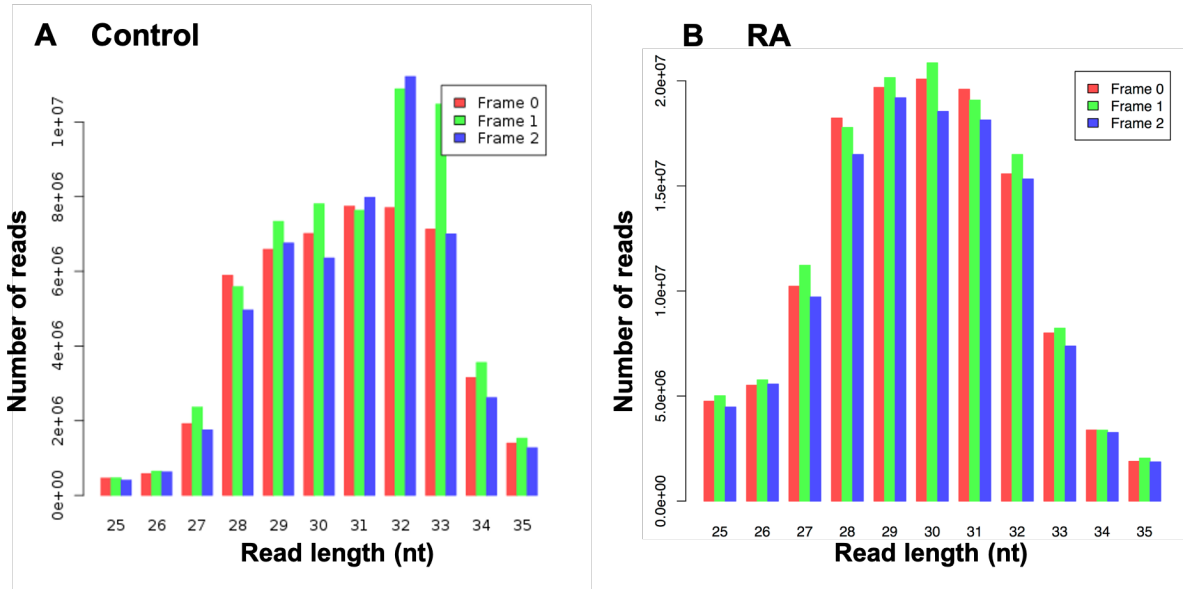


Figure 3.20: Ribosome footprint reads of Control sample do not display clear triplet periodicity. The graphs show the number of reads assigned to each frame (frames are colour-coded) for every read length examined for (A) Control and (B) RA treated cells. The footprint read length varies and for each read length similar number of reads are assigned to all the 3 frames.

For more extensive optimisation of digestion conditions, five critical parameters were considered. The RNase I enzyme, the composition of the buffer, the temperature, the duration of treatment and the temperature at which the cytoplasmic lysate was loaded on the gradients (personal communication with Dr Jose Pueyo-Marques). This experiment was only conducted in undifferentiated cells. In order to optimise the foot printing, I tested 8 different RNase I footprinting conditions (Table 3.2). Polysome profiling was conducted as previously described. After fractionation, different RNase I treatment conditions were applied to each sample (Table 3.2).

Two different RNase I enzymes were also compared, the one previously used (AM2295-ThermoFisher Scientific), and another recommended by recent publication (McGlincy and Ingolia, 2017) (RNase I with a unit definition of 10U/μL (EN0601-ThermoFisher Scientific). The buffer conditions in which RNase I treatment is performed has been shown to have an impact on the footprinting effectiveness. A ribosome profiling study in *Arabidopsis thaliana* reported that ionic strength inhibits the activity of RNase I, thus they used KCl at a final concentration of 40mM and $MgCl_2$ at a final concentration of 20mM (Hsu et al., 2016). The buffer I originally used was 150mM NaCl and 10 mM $MgCl_2$, as previously used in Poly-Ribo-Seq in *D. melanogaster* (Aspden et al., 2014). Therefore, the ionic strength of the buffer was tested by comparing 150mM and 30mM NaCl.

Most ribosome profiling studies in mammalian cells perform RNase I treatment at RT for 30 min-1h (Ingolia et al., 2009; Guo et al., 2010; Chung et al., 2015; Heyer and Moore, 2016). As previously shown (Figure 3.15), gel purification of footprint samples treated with RNase I at RT for 1h showed more effective footprinting, compared to overnight treatment at 4^{o}C so data for RNase I treatment at RT was only acquired. Nevertheless, the band intensity of footprints on the gel is a qualitative, rather than quantitative measure and it doesn't give any information about footprinting precision. Data analysis revealed that footprinting at RT was not effective, so I decided to perform the treatment at 4^{o}C (Aspden et al., 2014) in combination with different buffer conditions this time. Lastly, loading of the lysate was tested on the gradients at 4^{o}C, as at this temperature the positions of the ribosomes are better preserved, and this potentially improves footprinting.

Table 3.2: Summary of the different RNase I treatment conditions tested

Sample	Gradient loading temperature	RNase I	Buffer composition (pH 8)	RNase I treatment temperature	Duration of treatment
1	RT	Thermo (AM2295) 10U/million cells	Tris-HCl 50mM, NaCl 150mM, $MgCl_2$ 10mM	RT	1h
2	RT	Thermo (AM2295) 10U/million cells	Tris-HCl 100mM, NaCl 30mM, $MgCl_2$ 10mM	RT	1h
3	RT	Thermo (EN0601) 0.2U/million cells	Tris-HCl 100mM, NaCl 30mM, $MgCl_2$ 10mM	RT	1h
4	RT	Thermo (EN0601) 0.3U/million cells	Tris-HCl 50mM, NaCl 150mM, $MgCl_2$ 10mM	RT	1h
5	RT	Thermo (AM2295) 20U/million cells	Tris-HCl 100mM, NaCl 30mM, $MgCl_2$ 10mM	RT	1h
6	RT	Thermo (AM2295) 10U/million cells	Tris-HCl 50mM, NaCl 150mM, $MgCl_2$ 10mM	4°C	overnight
7	4°C	Thermo (AM2295) 10U/million cells	Tris-HCl 100mM, NaCl 30mM, $MgCl_2$ 10mM	4°C	overnight
8	4°C	Thermo (EN0601) 0.3U/million cells	Tris-HCl 100mM, NaCl 30mM, $MgCl_2$ 10mM	4°C	overnight

A panel of different footprinting conditions were tested side by side and then compared by a) footprint gel and b) analysis of triplet periodicity. Gel separation of the ribosome footprints shows that the clearest footprint bands are observed in samples treated with RNase I EN0601 (Figure 3.21). The samples treated in lower ionic strength buffer showed stronger footprinting. This suggests that ionic strength is contributing to the footprinting efficacy, but the other factors are equally important. Notably, this time treatment at 4^{o}C in a low ionic strength buffer (Tris-HCl pH 8 100mM, NaCl 30mM, $MgCl_2$ 10mM), regardless of the RNase I enzyme, yielded more effective footprinting (Figure 3.21-last 2 lanes). In general, comparing Figures 3.17 and 3.20, footprint bands of the expected size are much more clearly defined when digestion was performed under the optimised conditions, than when they were performed under the initial conditions.

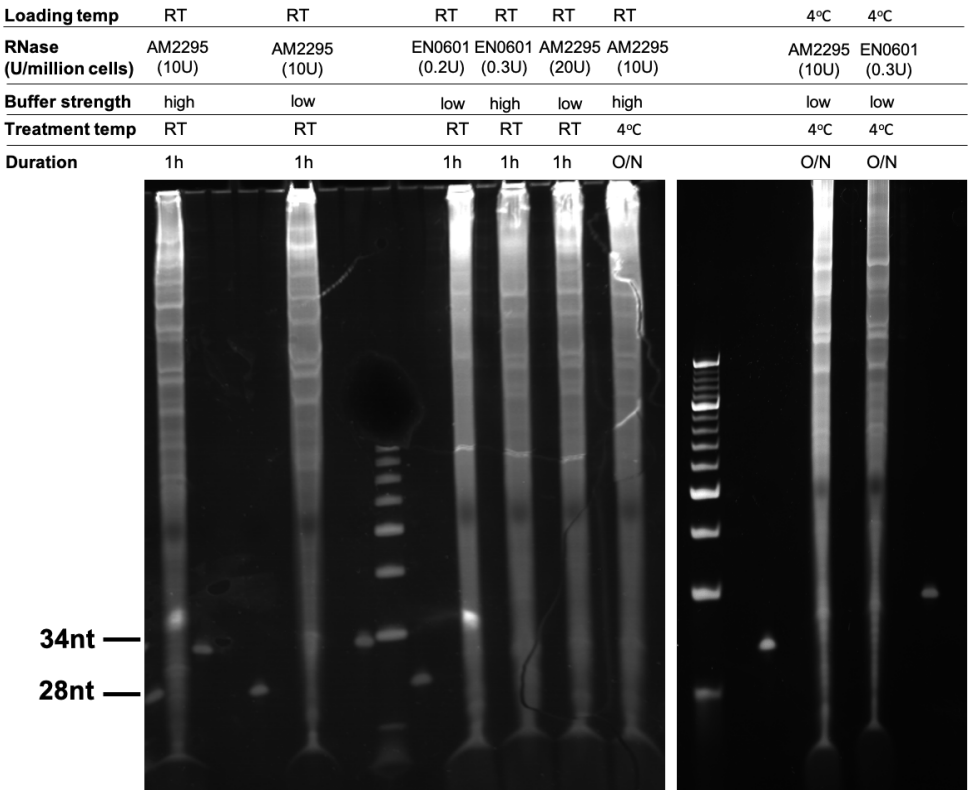


Figure 3.21: Samples treated with RNase I EN0601 in low ionic strength buffer show stronger footprinting. Gel purification of ribosome footprints (FP) of undifferentiated (Control) SH-SY5Y cells from 8 different RNase I footprinting conditions (shown above the gel). A different combination of enzyme, buffer, temperature and duration of treatment was applied to each sample. Samples treated with the RNase I EN0601 in a low ionic strength buffer (100mM Tris-HCl pH 8, 30mM NaCl, 10mM $MgCl_2$) show more effective footprinting compared to samples treated with RNase I AM2295 in a higher ionic strength buffer (50mM Tris-HCl pH 8, 150mM NaCl, 10mM $MgCl_2$). Samples were loaded on 10% (w/v) polyacrylamide-TBE-Urea denaturing gel and visualized using SYBR Gold staining. Markers of 28 and 34nt were loaded in either side of the samples to identify the footprint area (approximately 30nt) and an O'RangeRuler 10 bp DNA Ladder was used to aid the band size determination.

The triplet periodicity for each of the footprint samples was evaluated and four out of the eight samples displayed a clear pattern of periodicity (Figure 3.22). The samples treated with the AM2295 enzyme for 1h at RT in either high (50mM Tris-HCl pH 8, 150mM NaCl, 10mM $MgCl_2$) or low (100mM Tris-HCl pH 8, 30mM NaCl, 10mM $MgCl_2$) ionic strength buffer, showed no signs of triplet periodicity (Figure 3.22-A and B). Increasing the Units of RNase I/million cells (EN0601 or AM2295) seemed to slightly improve the footprinting efficacy, by reducing the number of reads of several read lengths, however, the triplet periodicity was still not satisfactory (Figure 3.22: C and D).

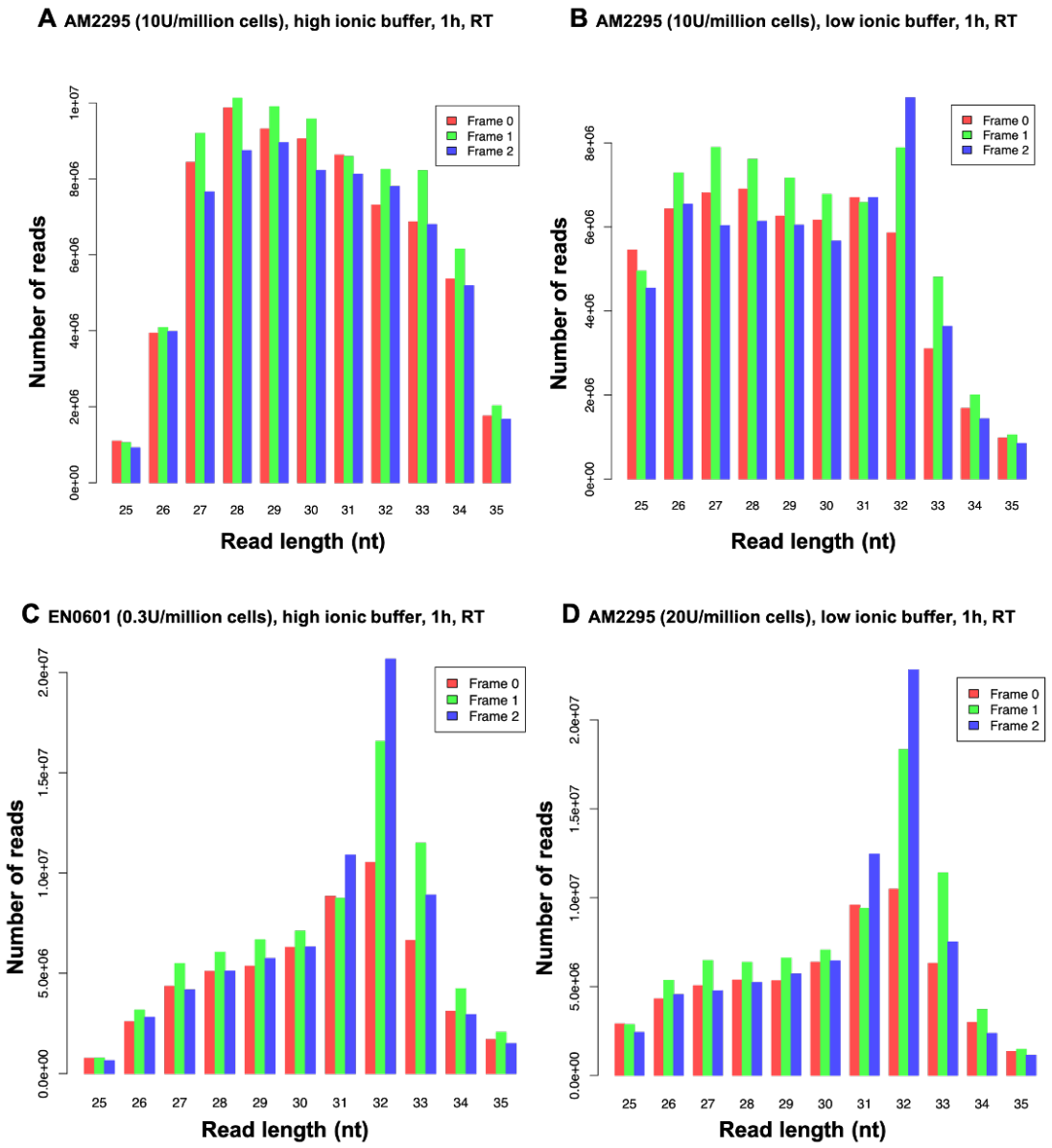


Figure 3.22: RNase I treatment conditions that did not produce good quality footprints. Footprinting was performed: (A) AM2295(10U/million cells) in high ionic strength buffer (Tris-HCl pH8 50mM, NaCl 150mM, $MgCl_2$ 10mM) for 1h at RT , (B) AM2295(10U/million cells) in low ionic strength buffer (Tris-HCl pH8 50mM, NaCl 30mM, $MgCl_2$ 10mM) for 1h at RT, (C) EN0601 (0.3U/million cells) in high ionic strength buffer (Tris-HCl pH8 50mM, NaCl 150mM, $MgCl_2$ 10mM) for 1h at RT and (D) AM2295(20U/million cells) in low ionic strength buffer (Tris-HCl pH8 50mM, NaCl 30mM, $MgCl_2$ 10mM) for 1h at RT did not show clear triplet periodicity.

In the rest of the samples, the precision of the RNase I digestion has substantially increased with a more restricted length distribution, and most reads started at the same frame (Figure 3.23). In general, all the factors were important for the efficiency of the RNase I treatment. Interestingly, temperature and duration of treatment affected the efficiency of the Thermo-AM2295 enzyme - which only worked efficiently at 4^{o}C overnight - but not the efficiency of the Thermo-EN0601 enzyme. The ionic strength of the buffer was important for both enzymes. For example, treatment with the Thermo-EN0601 RNase I for 1h at RT in a buffer with high ionic strength (50mM Tris-HCl pH 8, 150mM NaCl, 10mM $MgCl_2$) did not work efficiently (Figure 3.22: C). However, treatment with the same enzyme using low ionic strength buffer (100mM Tris-HCl pH 8, 30mM NaCl, 10mM $MgCl_2$) yielded much better triplet periodicity (Figure 3.23: A). Overall, between the four conditions that worked well, optimal conditions were: loading the lysate on the gradient at 4^{o}C and RNase I treating using Thermo EN0601 enzyme (0.3U/million cells), at 4^{o}C overnight, in a buffer of lower ionic strength (Tris-HCl pH 8 100mM, NaCl 30mM, $MgCl_2$ 10mM).

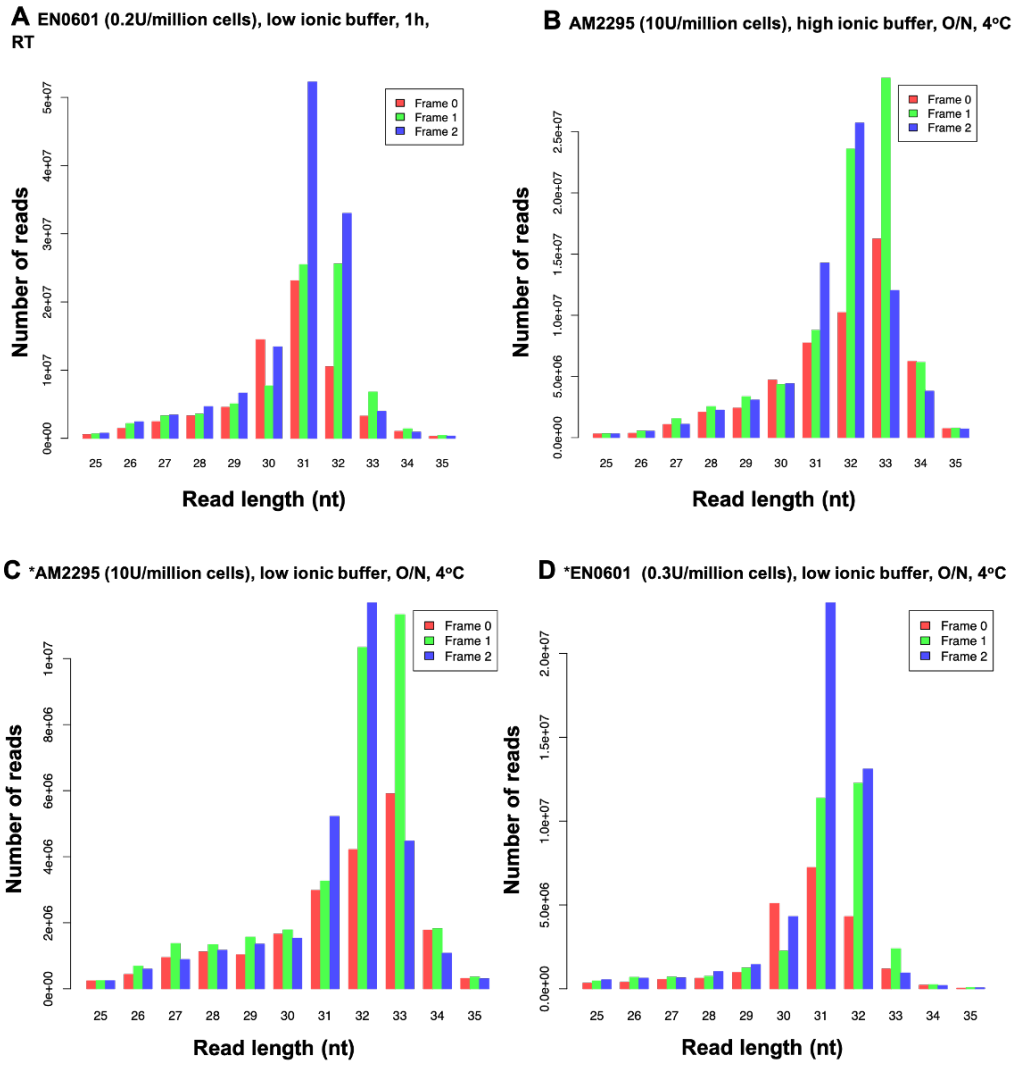


Figure 3.23: RNase I treatment with EN0601 enzyme (0.3U/million cells), at 4°C overnight, in a buffer of lower ionic strength showed the strongest triplet periodicity. Triplet periodicity plots show the number of reads assigned to each frame (frames are colour-coded) for every read length examined. (A) Footprinting was performed with RNase I Thermo-EN0601(0.2U/million cells) for 1h at RT in a low ionic strength buffer (100mM Tris-HCl pH 8, 30mM NaCl, 10mM $MgCl_2$). (B) Footprinting was performed with RNase I Thermo-AM2295 (10U/million cells), at 4°C overnight in a higher ionic strength buffer (50mM Tris-HCl pH 8, 150mM NaCl, 10mM $MgCl_2$). (C) Footprinting was performed with RNase I Thermo-AM2295 (10U/million cells) at 4°C overnight in a lower ionic strength buffer (100mM Tris-HCl pH 8, 30mM NaCl, 10mM $MgCl_2$) (*cell lysate had been loaded on the gradient at 4°C). (D) Footprinting was performed with RNase I Thermo- EN0601 enzyme (0.3U/million cells), at 4°C overnight, in a buffer of lower ionic strength (100mM Tris-HCl pH 8, 30mM NaCl, 10mM $MgCl_2$) (*cell lysate had been loaded on the gradient at 4°C).

To precisely calculate the proportion of reads in frame for the read lengths that exhibited the best triplet periodicity, the number of reads assigned to each frame was extracted from the triplet periodicity analysis (Figure 3.22) and used to calculate the percentage of reads that map to the same frame. The highest percentage of in-frame reads (~56%), corresponds to the footprinting conditions: RNase I treatment with EN0601 (0.3U/million cells) at 4^{o}C overnight, in a buffer of lower ionic strength (100mM Tris-HCl pH 8, 30mM NaCl, 10mM $MgCl_2$) (Figure 3.24).

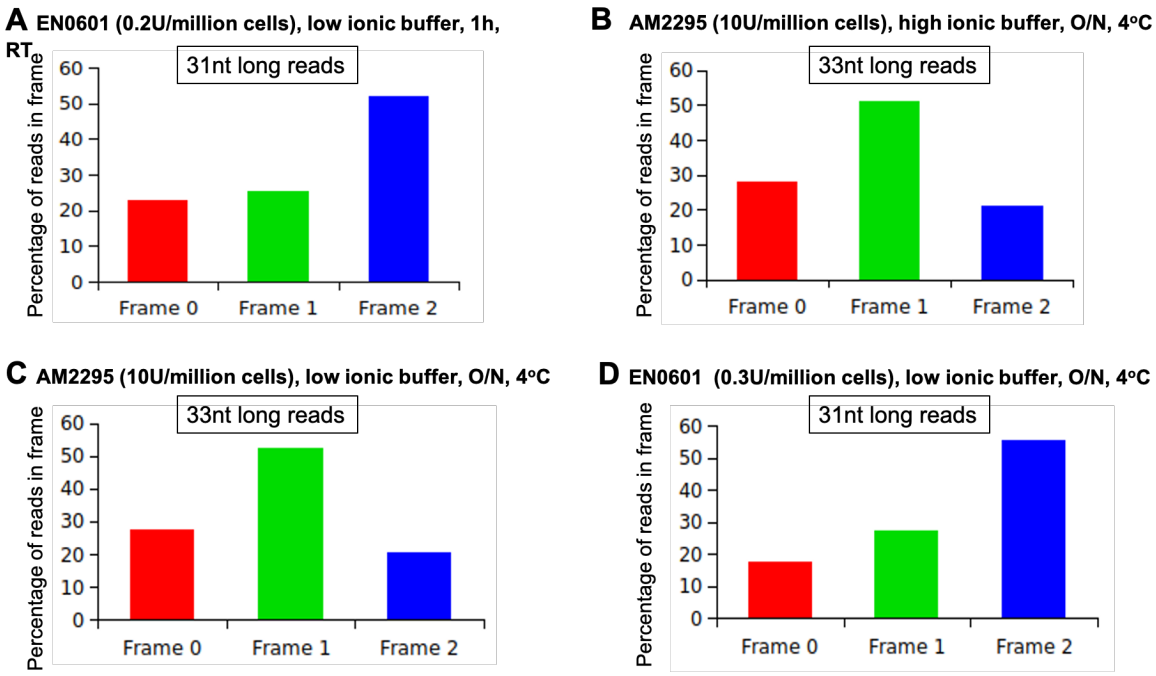


Figure 3.24: More than 53% of the reads in each condition were mapped to the same frame. The percentage of reads mapping to each of the 3 frames (frame 0, frame 1 and frame 2) was calculated for the 31nt (A and D) and 33nt long reads (B and C), that exhibited the best triplet periodicity in each condition. The highest percentage of in-frame reads (~56%), corresponds to the footprinting conditions: (D) RNase I treatment with EN0601 (0.3U/million cells) at 4°C overnight, in a buffer of lower ionic strength (Tris-HCl pH 8 100mM, NaCl 30mM, $MgCl_2$ 10mM).

A metagene analysis was performed using RiboSeqR, in order to examine the global distribution of the reads across the transcriptome (Figure 3.25). This is another key way of assessing ribo-seq data quality. For this analysis, only the read lengths that display the clearest triplet periodicity signal were used. Their distribution over 5'-UTRs, Open Reading Frames (ORFs) and 3'-UTRs is assessed and plotted for the 3 different frames. I selected the 31nt long reads from samples treated with RNase I EN0601 in a low ionic strength buffer at RT and 4^{o}C (Figure 3.25: A and D) and the 33nt long reads from samples treated with RNase I AM2295 at 4^{o}C in low and high ionic strength buffers (Figure 3.25: B and C). Based on the triplet periodicity and length plots these were the lengths with best periodicity.

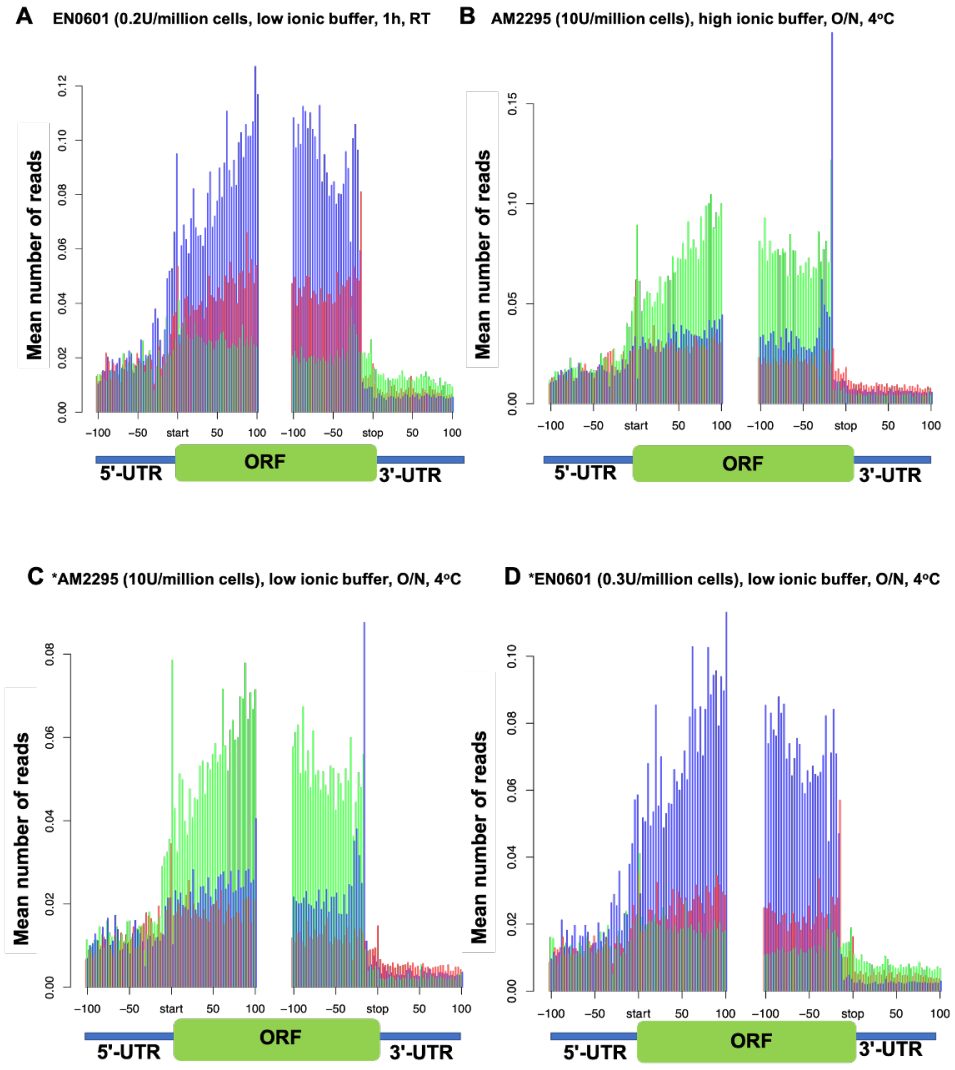


Figure 3.25: Footprints of samples treated with EN0601 enzyme (0.3U/million cells), at 4°C overnight, in a buffer of lower ionic strength map to ORFs. Metagene analysis of footprint reads of the lengths that showed the strongest periodicity from: (A) Sample treated with RNase I Thermo-EN0601(0.2U/million cells) for 1h at RT in a low ionic strength buffer (100mM Tris-HCl pH 8, 30mM NaCl, 10mM $MgCl_2$). (B) Sample treated with RNase I Thermo-AM2295 (10U/million cells), at 4°C overnight in a higher ionic strength buffer (50mM Tris-HCl pH 8, 150mM NaCl, 10mM $MgCl_2$). (C) Sample treated with RNase I Thermo-AM2295 (10U/million cells) at 4°C overnight in a lower ionic strength buffer (100mM,Tris-HCl pH 8 30mM NaCl, 10mM $MgCl_2$) (*cell lysate loaded on the gradient at 4°C. (D) Sample treated with RNase I Thermo- EN0601 enzyme (0.3U/million cells), at 4°C overnight, in a buffer of lower ionic strength (100mM Tris-HCl pH 8, 30mM NaCl, 10mM $MgCl_2$) (*cell lysate loaded on the gradient at 4°C). The plots show the global distribution of the reads across the UTRs and the main ORF of transcripts. Number of reads is plotted against metagene position (100bp up and downstream) to include the 5'- and 3'-UTRs.

For each of the samples the reads of the predominant frame show the expected pattern for ribosome profiling reads, as the majority of the reads map to open reading frames (ORFs) (Figure 3.25). Reads mapping to the other 2 frames did not show such clear pattern. This is why the colours of the peaks are different between samples. It is also expected for some reads to map to 5' and 3'- UTRs as translation of upstream and downstream ORFs has recently been reported (Heyer and Moore, 2016; Rodriguez et al., 2019; Chen et al., 2020). Notably, the accumulation of ribosome footprints at the stop codon is greater than at other codons (Figure 3.25: B and C). This is consistent with previous studies (Schuller et al., 2017) and can be attributed to the slower peptidyl release rate during termination, compared to the peptidyl transfer rate during elongation ((Schuller and Green, 2018)and references therein). Interestingly, the predominant frame of the footprint reads changes just before the stop codon in each of the graphs. This can be explained by the fact that once the release factor eRF1 binds to the stop codon at the A site of the ribosome, it interacts with 18S ribosomal RNA and that stacks on the second and third stop codon bases. This conformational change pulls the next base (1st base of the 3'-UTR) into the A-site (Brown et al., 2015; Matheisl et al., 2015), an event that likely affects the framing at the stop codon. The above evidence is consistent with successful Poly-Ribo-Seq experiment.

3.4.5 Optimisation of poly-A selection of cytoplasmic mRNAs

As part of Poly-Ribo-Seq, I purified both polysome-associated and total cytoplasmic RNA populations (Figure 3.14). It is important to sequence the polysome associated RNAs, in order to distinguish between the lncRNAs that interact with polysomes without being translated and those that are actively translated. Sequencing of the total cytoplasmic RNA population is previously established (Ingolia et al., 2009; Ingolia et al., 2013) and necessary because to calculate translation efficiency, total cytoplasmic mRNA population was also required to be measured.

Poly-A selection of the cytoplasmic and polysome associated RNA samples is an essential step of the protocol that allows for enrichment of poly-A tailed mRNAs and lncRNAs in the RNA-Seq control datasets. The most abundant category of RNA in the cell is rRNA, (~80% of the total RNA), therefore, without selecting for poly-A

RNAs the number of reads that would be mapped to rRNA would be such that there would be a minority of reads left to be mapped to the rest of the genome. In the first Poly-Ribo-Seq experiment poly-A selection of cytoplasmic mRNAs was carried out according to manufacturer's instructions with 1 round of polyA selection. poly-A selected total cytoplasmic and polysome-associated RNAs were gel purified as previously described (Methods 2.5.4.3).

To assess poly-A enrichment, the number of RNA-Seq reads that correspond to rRNA and tRNA rather than polyA mRNA or lncRNA was determined. I took the filtered reads and mapped them to human rRNA and tRNA (.fasta file) using Bowtie (Langmead et al., 2009). The number of reads were calculated for each sample that corresponded to rRNA, tRNA and neither (usable reads). The percentage of rRNA in most of the samples (Figure 3.26) was so high that it constituted more than 50% of the reads in some samples. This considerably reduced the number of reads that could be used for downstream analysis and therefore informative read depth required for robust ribo-seq analysis. Therefore, poly-A selection was not successful enough and further optimisation was required.

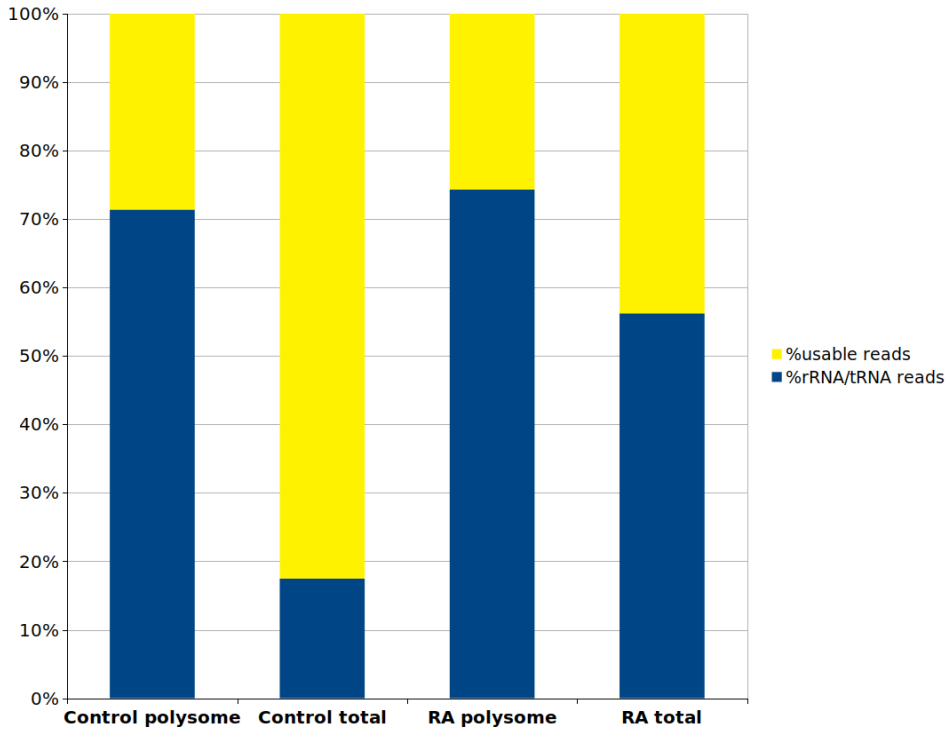


Figure 3.26: rRNA contamination is high in RNA-Seq samples of the first Poly-Ribo-Seq run. The stacked bar-charts depict the read distribution of each sample as a percentage of rRNA, tRNA and usable reads (non-tRNA/rRNA), as a percentage of the high-quality reads that pass the phred score of 20 and were used for analysis.

Several ribo-seq publications used two rounds of poly-A selection so this was tested to see if it would achieve a more complete rRNA and tRNA depletion. Total cytoplasmic poly-A selected RNA after one and two rounds of selection was run on a 10% (w/v) polyacrylamide-Urea-denaturing gel (Figure 3.27: A). After one round of poly-A selection, three distinct bands can be observed on the gel (Figure 3.27: A). The lowest band, at ~150bp likely corresponds to remaining 5.8S rRNA, while the other two might be a contamination. These bands are not detected in the sample that was subjected to two rounds of poly-A selection. This suggests that that two rounds poly-A selection should be performed. 2 rounds of poly-A selection was successful, as shown by the low levels of rRNA in the total cytoplasmic and polysome-associated RNA populations in the next experiment (Figure 3.28).

3.4.6 Optimisation of RNA fragmentation

It was necessary to fragment poly-A selected RNA so that its size (50-80nt) was comparable to that of the footprints. This was because all samples were to be processed together for library preparation and Next Gen Sequencing. Having similar fragment sizes is also important for accurately measuring translation efficiency.

RNA fragmentation was performed by alkaline hydrolysis. The duration of alkaline hydrolysis is dependent on the amount of RNA in the sample; therefore, it was necessary to determine the optimal conditions for this system. In the Poly-Ribo-Seq protocol for S2 cells, alkaline hydrolysis was performed for 15-20min. To establish optimal conditions in SH-SY5Y cells, RNA was extracted from cytoplasmic lysates and performed alkaline hydrolysis for 3 different lengths of time: 5 min, 10 min and 15 min. 15 minutes of alkaline hydrolysis result in over-fragmentation of the RNA (Figure 3.27). Fragmenting for only 5min was insufficient, as most of the RNA migrated above the 80bp band of the ladder. Therefore, 10 min was chosen as the optimal time for fragmentation.

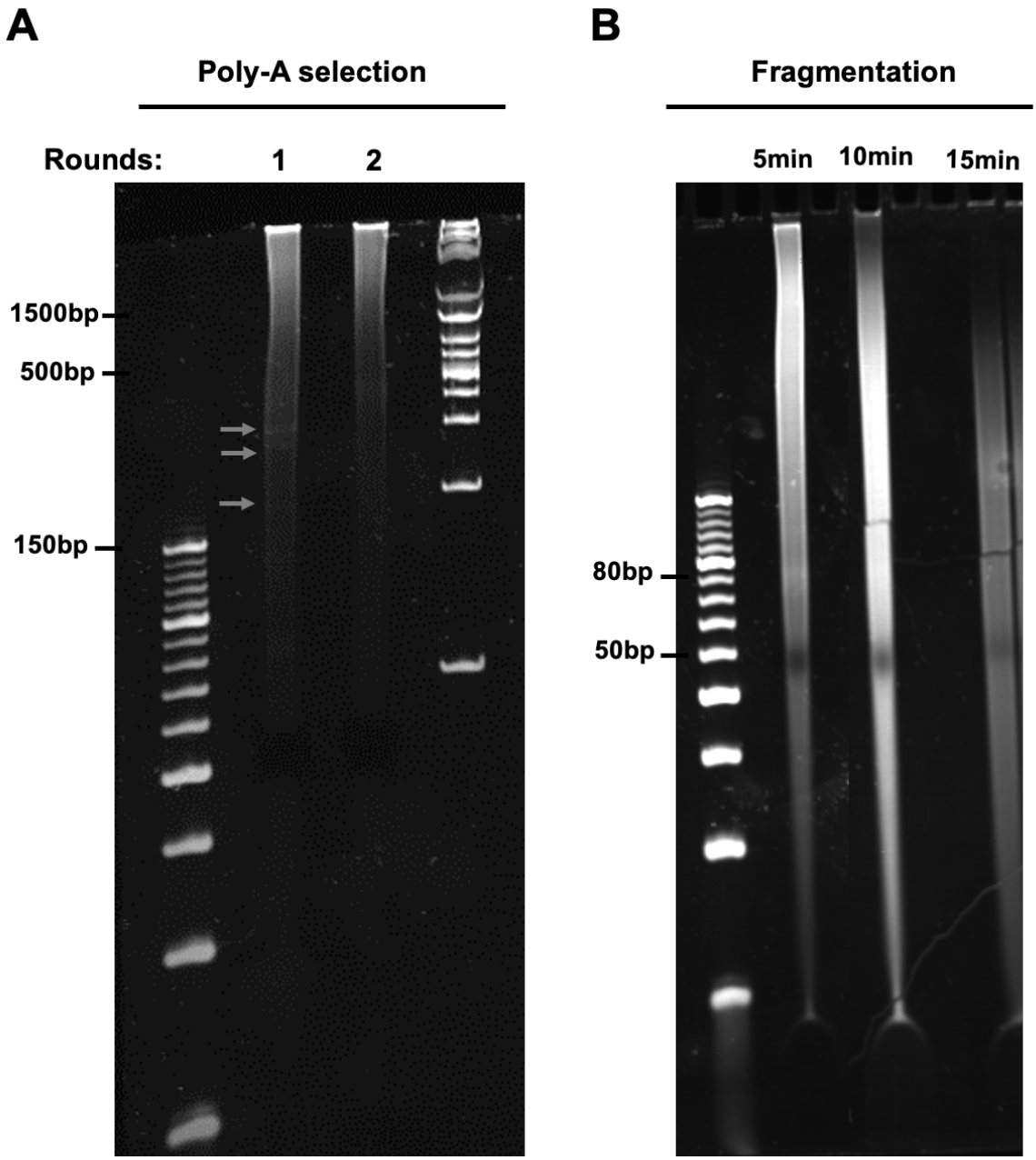


Figure 3.27: Poly-A selection and fragmentation optimisation. (A) Poly-A selection performed in two rounds, reduces the amount of non-specific material. Total RNA samples subjected to poly-A selection for one round (lane 1) and two rounds (lane 2). In lane 1, the smear is brighter, and three distinct bands can be observed (grey arrows), which are absent from lane 2 and correspond to rRNA. (B) Fragmentation of poly-A selected RNA for 10min optimal for SH-SY5Y. Alkaline hydrolysis was performed in three different poly-A selected RNA samples for 5, 10- and 15-min. Samples were run on 10% (w/v) polyacrylamide-TBE-Urea denaturing gel and visualized using SYBR Gold staining.

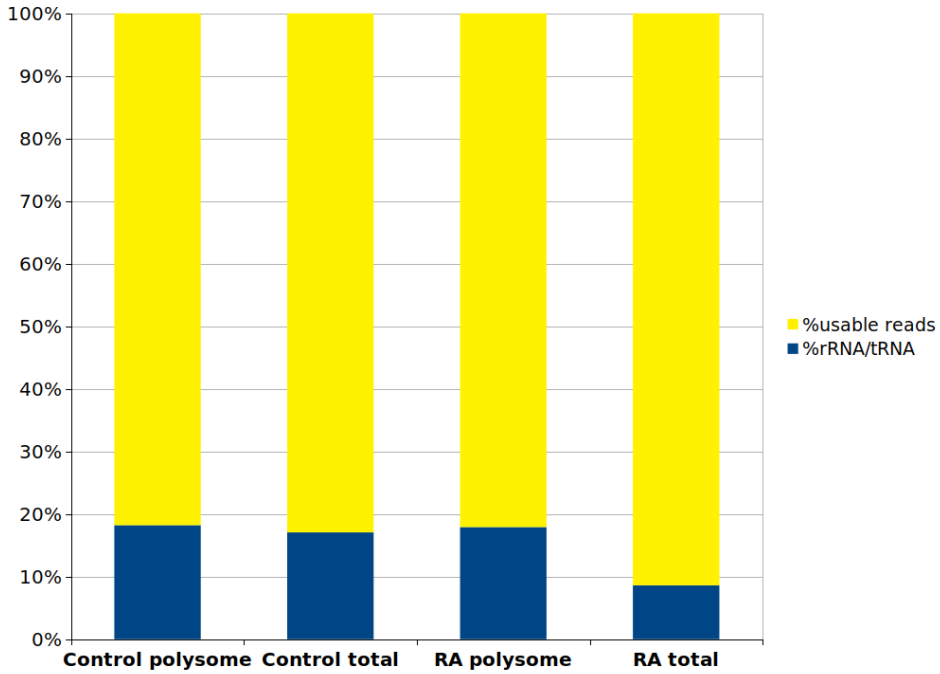


Figure 3.28: rRNA contamination is reduced after 2 rounds of poly-A selection. Representative sample of one of the Poly-Ribo-Seq biological replicates, in which poly-A selection was performed in two rounds and Illumina kit was used for rRNA depletion. The stacked bar-charts depict the read distribution of each sample as a percentage of rRNA, tRNA and usable reads, as a percentage of the high-quality reads that pass the phred score of 20 and were used for analysis.

3.5 Discussion

3.5.1 Neuronal differentiation induces a reduction in the level of active translation

In summary, this chapter reports the validation of my chosen model of human neuronal differentiation and the optimisation of Poly-Ribo-Seq protocol for use with this model. Neuronal differentiation induces a reduction in the level of active translation as shown by the significant reduction level of polysomes (Figure 3.11) and the significant decrease of the polysome/monosome (P/M) ratio (Figure 3.12). This is consistent with other studies showing downregulation of translation machinery, demonstrated by translational repression of ribosomal proteins during differentiation of Human Embryonic Stem cells (hESCs) to Neural Stem Cells (NSCs) (Blair et al., 2017). Notably, upon differentiation a significant increase in the number of monosomes, was observed (Figure 3.11).

The time-window of differentiation must be taken into consideration here as well. SH-SY5Y are harvested three days post-differentiation. This time-point falls within a period of tight spatiotemporal regulation of protein synthesis that ensures the acquisition of the proper cell fate ((Mohammad et al., 2019)and references therein). Therefore, at this transition state, the translation of certain transcripts is stimulated (eg axon guidance gene PLXNA2; (Rodriguez et al., 2019)) whereas the translation of others is inhibited (eg MYC; (Blair et al., 2017)). The latter would be a possible explanation for the increased number of RNA transcripts engaged by the 80S. mRNAs translated by one ribosome (80S) at a time is also a plausible explanation, based on previous studies (Heyer and Moore, 2016; Biever et al., 2020). Monosomes can also initiate translation, when they are engaged to mRNAs, but may require additional signals to continue in elongation or disassemble (Balagopal and Parker, 2011; Kong and Lasko, 2012).

Interestingly, a recent study in *S. cerevisae* sheds more light to the role of the 80S, showing that most monosomes are actively elongating and translating short ORFs, uORFs, NMD targets and mRNAs encoding for low abundance regulatory proteins (Heyer and Moore, 2016). Moreover, another recent study in rat shows that monosomes actively translate key synaptic transcripts in dendrites and axons and more importantly, translation by 80S seems to be preferred in the neurites, compared to the cell soma (Biever et al., 2020). This might mean that the increase in 80S area

is related to an increase in the production of regulatory proteins required for the fine-tuning of differentiation.

3.5.2 Optimisation of ribosome foot printing in SH-SY5Y cells

Poly-Ribo-Seq is an adaptation of ribosome profiling that distinguishes between the cytoplasmic poly-A RNAs that are translated from those that interact with the translation machinery without being translated. The method was established in *D. melanogaster* S2 cells (Aspden et al., 2014) and this is the first time it has been performed in a human neuronal cell type.

The long term aim of my project was to study the interaction of cytoplasmic lncRNAs with the translation machinery and to identify lncRNAs with actively translated small open reading frames (smORFs) during neuronal differentiation.This method was optimised for use in human neuronal cells and therefore addresses this aim, overcoming initial issues of rRNA contamination (figure 3.26) and lack of triplet periodicity (figure 3.22).

The efficiency of the RNase I digestion in footprinting was a key determinant of data quality. Sometimes, the literature indicates that the same conditions seem to work well for different organisms (Ingolia et al., 2009; Guo et al., 2010). However, ribosome profiling studies in *A. Thaliana* (Hsu et al., 2016), *C. Reinhardtii* (Chung et al., 2015), *S. Cerevisiae* (Ingolia et al., 2009; Heyer and Moore, 2016; McGlincy and Ingolia, 2017), *S. pombe* (Duncan and Mata, 2014), *D. melanogaster* (Aspden et al., 2014) and cultured mammalian cells (Guo et al., 2010; Ingolia et al., 2013) perform RNase I footprinting in different ways, using different enzymes. Based on my work, optimisation is recommended for every new sample-type. Notably, there is no definitive way of assessing the efficiency of the RNase I treatment prior to sequencing and the only means of evaluating the quality of the ribosome profiling data is by looking at the triplet periodicity.

After analysing the ribosome profiling data, it was established that the optimal footprinting conditions for SH-SY5Y cells were RNase I treatment with the Thermo EN0601 enzyme (0.3U/million cells), at 4^{o}C overnight, in a buffer of lower ionic strength (100mM Tris-HCl pH8, 30mM NaCl, 10mM $MgCl_2$).

3.5.3 Conclusions

Overall, in this chapter, the use of SH-SY5Y human neuroblastoma cells differentiated with RA was introduced as a model of neuronal differentiation for the study of cytoplasmic lncRNAs in relation to the translation machinery. Results show that translation is repressed during the early stages of SH-SY5Y differentiation. To further study the implication of lncRNAs in this translational regulation, optimisation of the Poly-Ribo-Seq protocol was required. The data presented in this chapter provide the basis for the successful application of Poly-Ribo-Seq in human neuronal cells. Having established the appropriate cell model and the optimal conditions of RNase I footprinting for SH-SY5Y cells, the next chapter will examine the potential role of lncRNAs in the regulation of translation upon neuronal differentiation.

Chapter 4

LncRNAs dynamically interact with the translation machinery during neuronal differentiation

4.1 Introduction

Understanding the relationship of cytoplasmic lncRNAs with the translation machinery is key to discovering more details about their functions. Several studies in fly, mouse and human report the association of lncRNAs with polysome complexes (Carrieri et al., 2012; Aspden et al., 2014; Carlevaro-Fita et al., 2016). LncRNAs can base-pair with mRNAs to facilitate translation of these mRNAs (e.g. UCHL1-AS, lnc-31;(Carrieri et al., 2012);(Dimartino et al., 2018)) affecting a number of processes in the cell, including differentiation (Dimartino et al., 2018). In contrast, lncRNAs can interact with components of the translation machinery, such as eIF4A and Poly-A binding protein (PABP) to inhibit mRNA translation (BC200; (Wang et al., 2002)). Notably, interaction of lncRNAs with polysomes is important for human adipocyte (hASCs) differentiation as well as hESCs differentiation to cardiomyocytes (Dallagiovanna et al., 2017; Pereira et al., 2020).

A number of lncRNAs appear to be actively translated in yeast, fly, mouse and humans, producing small peptides with so far, poorly characterised functions (Duncan and Mata, 2014; Aspden et al., 2014; Ruiz-Orera and Albà, 2019; Minati et al., 2020; Chen et al., 2020). Non-canonical small peptides, involved in the regulation of development, were reported in *D. melanogaster*, more than a decade ago (Kondo et al., 2007; Pueyo and Couso, 2008). A recent study reports the discovery of > 200 novel small peptides encoded by uORFs and lncRNA smORFs in humans and showed evidence of functionality for ~15 of them (Chen et al., 2020). Two micro-peptides (Myoregulin and LEMP) encoded by smORFs within lncRNAs, play important roles in skeletal muscle differentiation (Anderson et al., 2015; L. Wang et al., 2020)and another one (SPAAR) is involved in endothelial cell differentiation (Spencer et al., 2020). Given that neuronal differentiation is a precisely coordinated process in which several lncRNAs are involved (Lin et al., 2014; Winzi et al., 2018; Carelli et al., 2019), it is reasonable to hypothesise that the interaction of lncRNAs with translating polysomes, and potential translation of lncRNAs can have regulatory roles in neuronal differentiation as well.

In the previous chapter, I established that neuronal differentiation of SH-SY5Y cells for 3 days, induces a reduction in the level of active translation and I optimised the Poly-Ribo-Seq protocol for human neuronal cells. In this chapter I focus on the

analysis of Poly-Ribo-Seq data from Control and differentiated cells. I aim to a) discover if there are alterations in cytoplasmic and polysome-associated lncRNA expression, upon differentiation, b) uncover potential differences in polysome enrichment of lncRNAs during differentiation, c) detect actively translated smORFs in lncRNAs, d) select target lncRNAs for further study, based on the data analysis.

4.2 Poly-Ribo-Seq produced high quality data for further analysis of transcription and translation changes upon differentiation.

Having established that RA treated SH-SY5Y cells are an appropriate model for neuronal differentiation and having optimised Poly-Ribo-Seq protocol for SH-SY5Y cells was performed in 3 independent biological replicates. SH-SY5Y cells were cultured to ~70% confluency and then treated with either DMSO (Control) or 30μM RA diluted in DMSO for 3 days. Control and RA treated cells were harvested and lysed (Methods 2.5.2). ~20% of the cytoplasm was processed for RNA-Seq (total cytoplasmic population) and the remaining ~80% of cytoplasmic lysate was subjected to sucrose gradient fractionation. Subsequently, polysome fractions from each condition were combined and ~25% of this material was processed for RNA-Seq (polysome-associated population). The remaining ~75% of polysome-associated RNA sample was 'footprinted' with RNaseI and processed for Ribo-Seq (translated population) (Methods 2.5.2, Figure 2.3). All samples from each replicate were barcoded, pooled, and then subjected to 75bp single-end sequencing on the same lane. Notably, libraries were pooled so that footprint samples were sequenced more deeply, relative to poly-A RNA samples, to provide sufficient read depth for rare lncRNA smORF translation events to be detected.

To check the quality of each dataset standard QC analyses were performed. Initially, the quality of the raw reads of each dataset was assessed, using FastQC (Andrews, 2010) and 3' end adapter was trimmed from the reads using Cutadapt (Martin, 2011) (details in Methods 2.22). The proportion of high-quality reads with Phred score> 20 was between 79%-85% in every sample (Figure 4.1). To assess the proportions of reads for each sample that are derived from rRNA and tRNA sequence, and the proportion of usable reads present in each sample (Table 4.1 for replicate 1

and the others in Appendix-II), the reads were mapped to rRNA and tRNA sequences (Figure 4.2). The reads that mapped to rRNA and tRNA were then removed. ~30-63 million RNA-Seq reads and 65-120 million Ribo-Seq reads (across all samples from all replicates) remained for further analysis (Figure 4.2), hereafter termed as ‘usable’ reads.

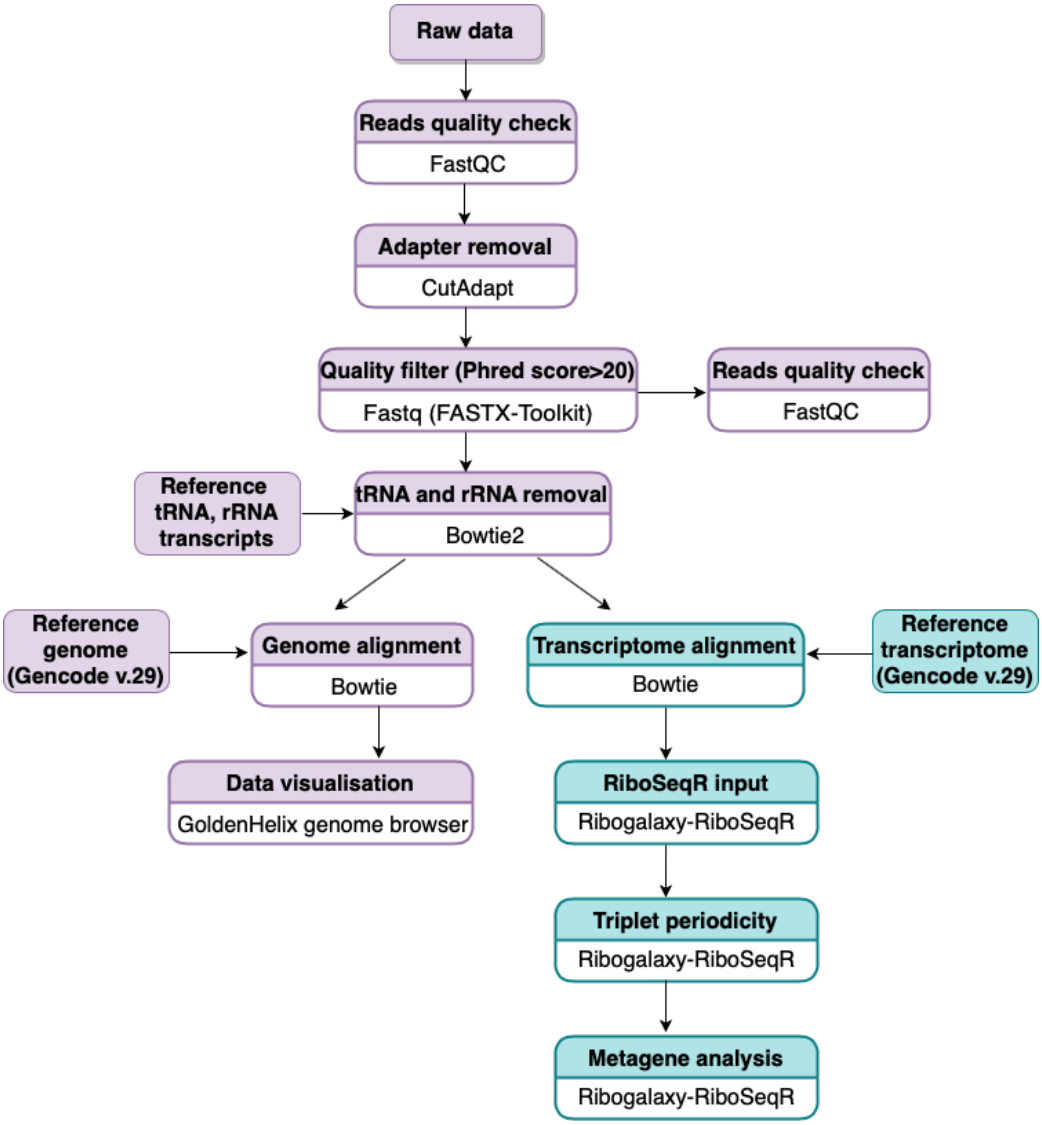


Figure 4.1: RNA-Seq and Ribo-Seq data analysis pipeline for assessment of ribosome profiling quality and data visualisation. This analysis was performed using the ribogalaxy platform and RiboSeqR pipeline. Raw reads were subjected to quality control using FastQC and subsequently 3' end adapter sequences were removed (Cutadapt). Reads were then filtered so that only high-quality reads passing the Phred score of 20 were kept for analysis. Following rRNA/tRNA removal, reads were mapped to the genome, to be processed for visualisation on the genome browser and to the transcriptome to perform the RiboSeqR triplet periodicity and metagene analysis.

The variation between the number of ‘usable reads’ between the samples is consistent with the initial variation in the number of raw reads for each sample. The read depth resulting from the number of usable reads is sufficient for conducting downstream statistical analyses (e.g. Differential expression analysis) according to previous studies (Robinson and Storey, 2014; Bass et al., 2019). Usable reads were then mapped to the most recent version of the human transcriptome (Gencode v.29) (Frankish et al., 2019a) using bowtie (Langmead et al., 2009). The proportion of usable reads that mapped to the transcriptome varied between 73% and 93% in the different samples (Table 4.1 and Appendix-II).

Table 4.1: Summary of the number of reads for biological replicate 1 after each step of the pipeline.

sample	number of raw reads	% of reads after Filtering	number of high-quality reads after rRNA and tRNA removal	% of usable high-quality usable reads mapped to transcriptome	% of high-quality usable reads mapped to lncRNAs	% of high-quality usable reads mapped to protein-coding transcripts	% of protein-coding mapped reads that map to CDS	% of protein-coding mapped reads that mapped to 3'-UTR	% of protein-coding mapped reads that mapped to 5'-UTR
Control FP	210,729,343	83%	119,371,359	93%	3,5%	87%	95%	17%	21%
Control polysome	75,815,765	81%	50,389,179	85%	2%	83%	55%	45%	14%
Control total	74,580,218	80%	49,700,483	84%	2%	80%	55%	44%	13%
RA FP	180,333,341	80%	106,866,877	80%	4.7%	74%	83%	23%	25%
RA polysome	64,468,224	80%	42,224,846	86%	2.4%	83%	49%	50%	13%
RA total	47,091,265	80%	34,776,336	86%	2%	83%	59%	40%	12.7%

More than 80% of the raw reads passed the quality filtering. Following rRNA/tRNA removal a sufficient number of reads remained for further analysis ('usable') and more than 80% of those reads mapped to transcriptome. Between 2% and 4.7% of the 'usable' reads mapped to lncRNAs and between 74% and 87% mapped to protein-coding sequences. 83%-95% of Footprint (FP) reads mapped to coding sequences (CDS), 17%-23% mapped to 3'-UTRs and 21%-25% mapped to 5'-UTRs. 49%-59% of RNA-Seq reads (total cytoplasmic and polysome-associated) mapped to CDS, 40%-50% mapped to 3'-UTRs -as expected since total and polysome-associated RNA samples were subject to poly-A selection and 12.7%-14% mapped to 5'-UTRs.

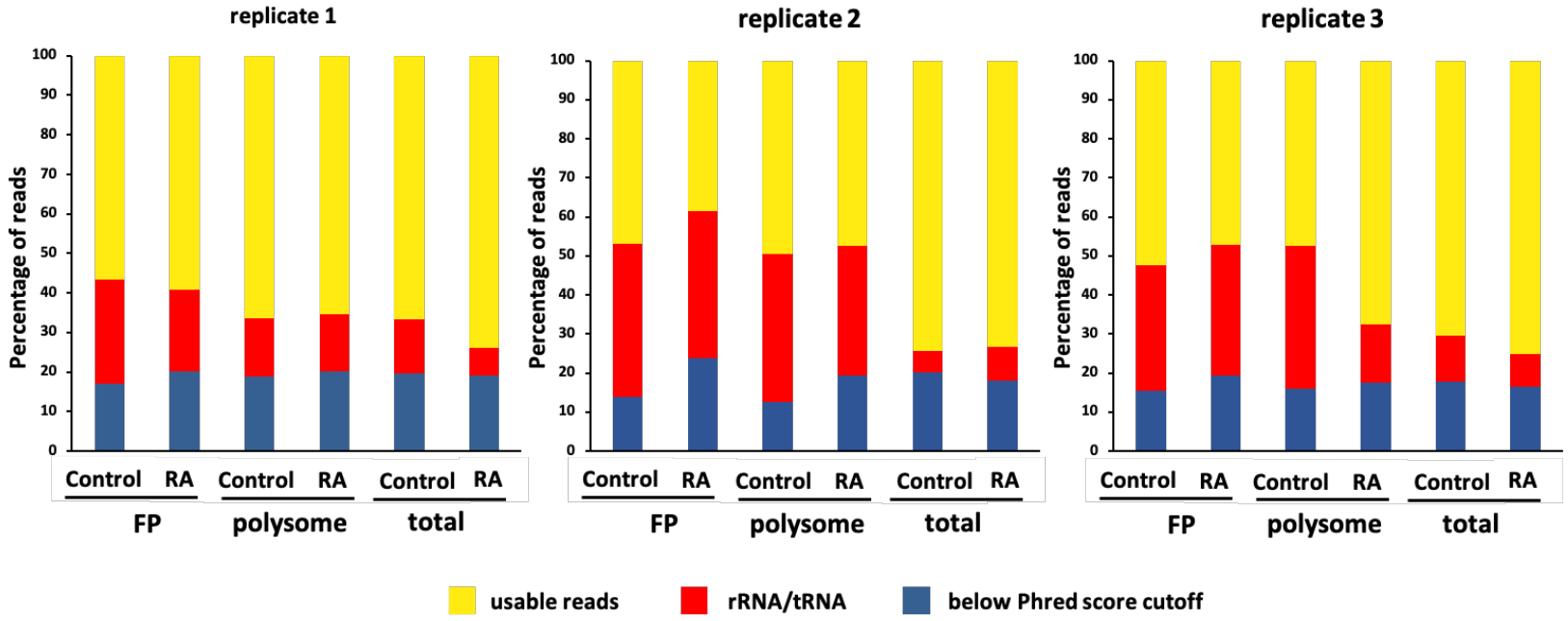


Figure 4.2 Percentage of high quality 'usable', poor quality, rRNA/tRNA reads. Following adapter trimming, raw reads from each biological replicate were filtered so that 'low-quality' reads, which had a Phred score <20 were discarded. High quality reads were subjected to rRNA/tRNA removal. Following that, up to 75% of the raw reads remained as high-quality usable reads.

To assess the distribution of the reads across coding sequences (CDS) and untranslated regions (UTRs), ‘usable’ reads were mapped to each of these reference sequences (Gencode v.29). This analysis provided an indication of whether the foot-printing experiments had been successful, since foot-printing samples are expected to be enriched in reads mapping to CDS, in contrast to RNA-Seq samples, which will be spread across the transcriptome. The numbers and percentages of usable reads mapping to the whole transcriptome (Gencode v.29) and numbers of reads mapping to CDSes, 3’-UTRs and 5’-UTRs were calculated (Table 4.1). The number of reads that do not map to CDS is different from the number of reads that map to UTRs, because a proportion of reads map to more than one feature, e.g. CDS and 3’-UTR (Figure 4.3).

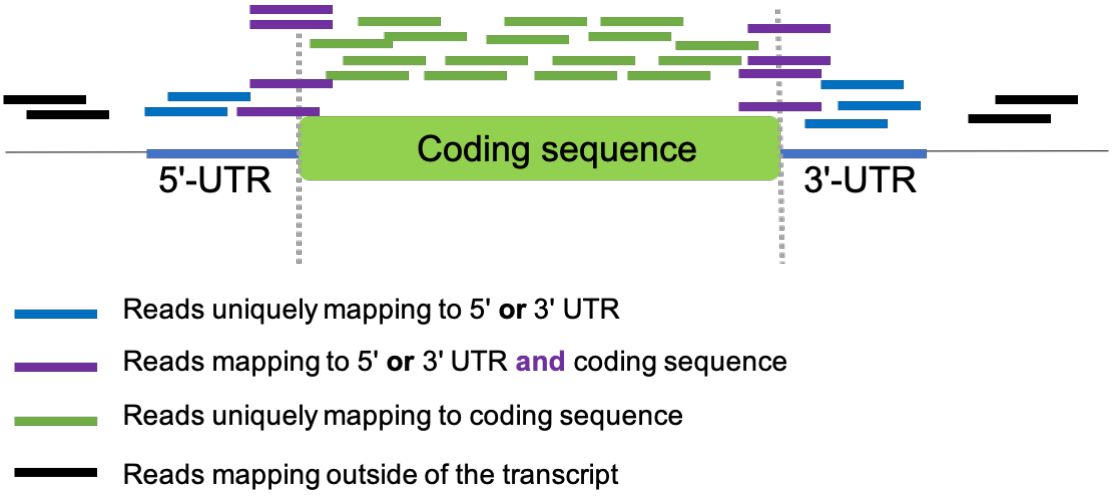


Figure 4.3: Schematic of read alignment to the coding sequences and 5'-and 3'-UTRs. Reads that map partially to the 5' or 3'-UTR and partially to the coding sequence are counted in both alignments.

On average, 54% of ‘total cytoplasmic’ and 52% of ‘polysome-associated’ RNA-Seq reads mapped to CDSes. In contrast ~90-95% of the Ribo-Seq reads map to CDSes (Figure 4.4: A). This pattern is to be expected because RNA-Seq reads will map across whole transcript whilst Ribo-Seq reads should map only to translated regions, the majority of which are annotated CDSes. Interestingly, there was a substantial difference between the proportion of the reads that map to 5’-UTRs and 3’-UTRs. On average, 21% of the Ribo-Seq reads and 13% of RNA-Seq reads map to 5’-UTRs (Figure 4.4: B). Whilst the percentage of Ribo-Seq reads that map to 3’-UTRs is 16% and that of RNA-Seq reads is 46% (Figure 4.4: C). The higher proportion of footprints on 5’-UTRs, than 3’-UTRs is consistent with the previous reports of actively translated uORFs (Heyer and Moore, 2016; Rodriguez et al., 2019; Chen et al., 2020). 3’-UTRs are generally much longer than 5’-UTRs (Hong et al., 2006), which makes it more likely for reads to align to them. RNA-Seq libraries were subject to poly-A selection, therefore, they are enriched for 3’ ends of mRNAs, providing an explanation of why the percentage of RNA-Seq reads mapping to 3’-UTRs is larger than to 5’-UTRs.

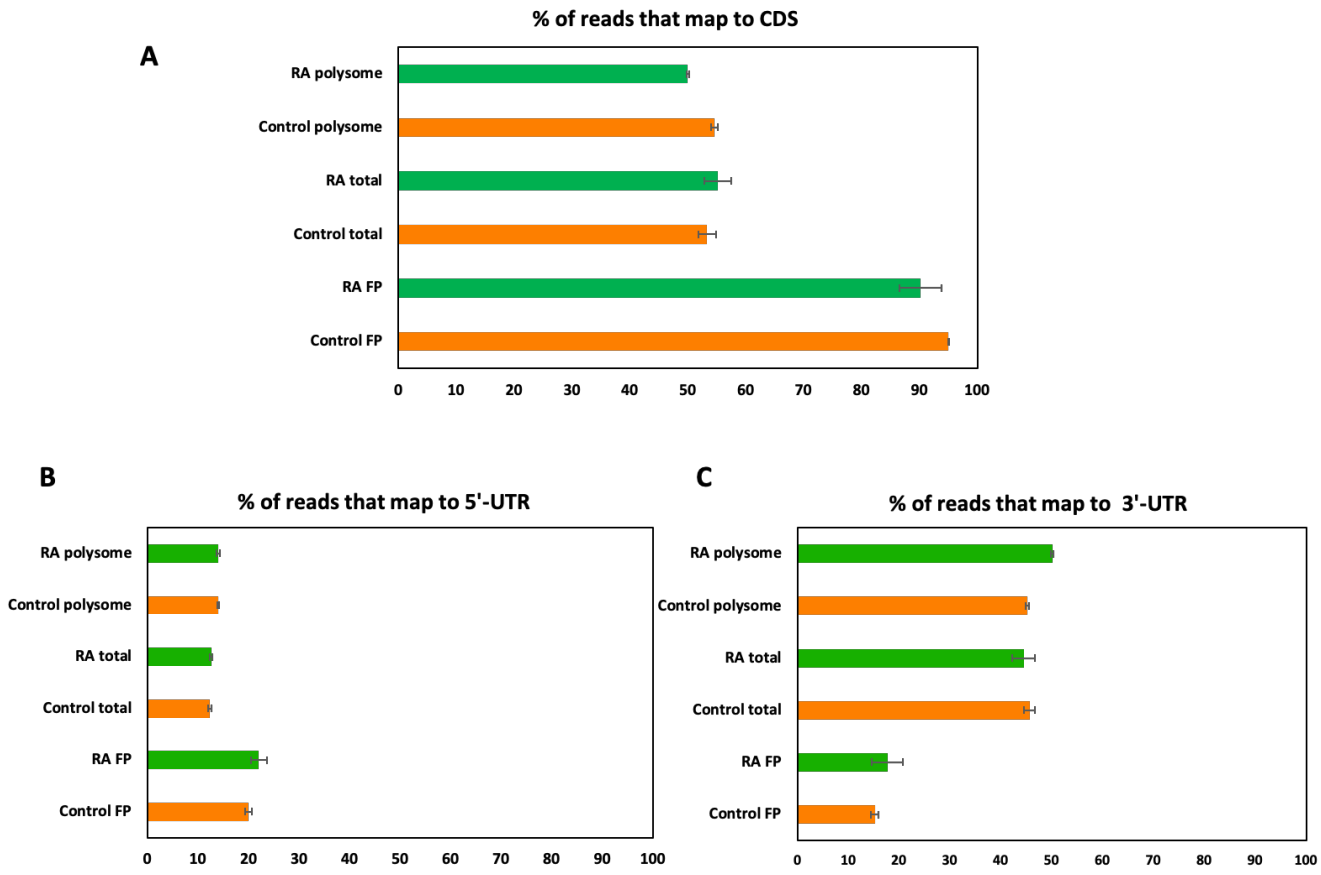


Figure 4.4: The majority of the Ribo-Seq reads map to coding sequences (CDSes). Average percentage of reads from three replicates, mapping to CDSes, 5' and 3'-UTRs (Standard error is plotted). For each sample, the percentages of reads that map to CDSes, 5' and 3'-UTRs do not add up to 100%, due to the overlap of reads that map partially to UTRs and partially to CDS. (A) The percentage of reads that map to CDSs is ~93% for the footprint samples (Control FP and RA FP), 52% for polysome samples (Control polysome, RA polysome) and 54% for total cytoplasmic samples (Control total, RA total). (B) The percentage of the Ribo-Seq reads that map to 5'-UTRs (21% on average) is higher compared to that of RNA-Seq samples (13% on average). (C) RNA-Seq reads of total and polysome samples map equally within 3'-UTRs (46% of the RNA-Seq reads on average) and outside of them (54% of the RNA-Seq reads on average). Only 16% of Ribo-Seq reads maps to 3'-UTRs on average.

Taken together, this quality control analyses showed that Poly-Ribo-Seq data were of high quality and each dataset contained sufficient number of reads to proceed with further analysis. Data from total cytoplasmic and polysome-associated RNAs were broadly comparable, in terms of number and quality. The quality control analyses of the foot-printing (Ribo-Seq) data showed the expected enrichment in reads mapping to CDSs, as well as a higher proportion of reads mapping to 5'-UTRs as compared to 3'-UTRs. This is expected as more than 40% of mammalian mRNAs contain uORFs (Young and Wek, 2016) and specifically in neuroblastoma, 31% of expressed mRNA transcripts contain uORFs and ~5000 uORFs have previously been shown to be translated in neuroblastoma cell lines (Rodriguez et al., 2019). This provided a first indication that the ribosome profiling was successful.

To identify the correlation within and between the independent biological replicates, I used principal component analysis (PCA) (Jolliffe, 2011) (Figure 4.5). PCA identifies variation between data from multiple samples and so the way each sample clusters in relation to the others on a PCA plot provides an indication of likely similarities and differences. PCA was performed on the gene expression matrix of the RNA-Seq datasets (generated by Dr D. Wang, Methods 2.18.1) to determine the correlation between total and polysome RNA populations in Control and upon RA differentiation. Briefly, if we consider RPKM values for each sample as points centred to the origin of a 2D graph, PCA finds the best fitting line between the points, by projecting each point to the line and maximising the sum of the squared distances from those projected points to the origin. This best fitting line accounts for the greatest variation between the samples and is plotted on the 2D graph as PC1 (Figure 4.5 PC1). Similarly, the second greatest variation between the data is determined and plotted it as an axon 90^{o} vertical to PC1 (Figure 4.5 PC2). For the RNA-Seq data, PC1 and PC2 accounted for the 74.8% of the variation between samples (Figure 4.5). Therefore, I can be confident that is the plot provides a good indication of the correlation between them.

On the PCA plot, all Control samples clustered on the left side whereas all RA treated samples clustered on the right side (Figure 4.5). This means that Control total and polysome samples are more similar to each other than to the RA treated samples and vice versa. Furthermore, the 3 replicates of RA total cytoplasmic samples cluster together at the bottom right side of the graph, separately from the RA polysome samples, which cluster together at the top right side. Control total and polysome samples do not form separate clusters. This may suggest that the variation in gene expression between RA total and RA polysome is greater than the variation between Control total and Control polysome. In summary, the PCA indicates that differentiation is associated with broad transcriptional changes. The distribution of control and RA samples at alternate ends of the x-axis did not identify any outliers and suggested that replicate samples are comparable.

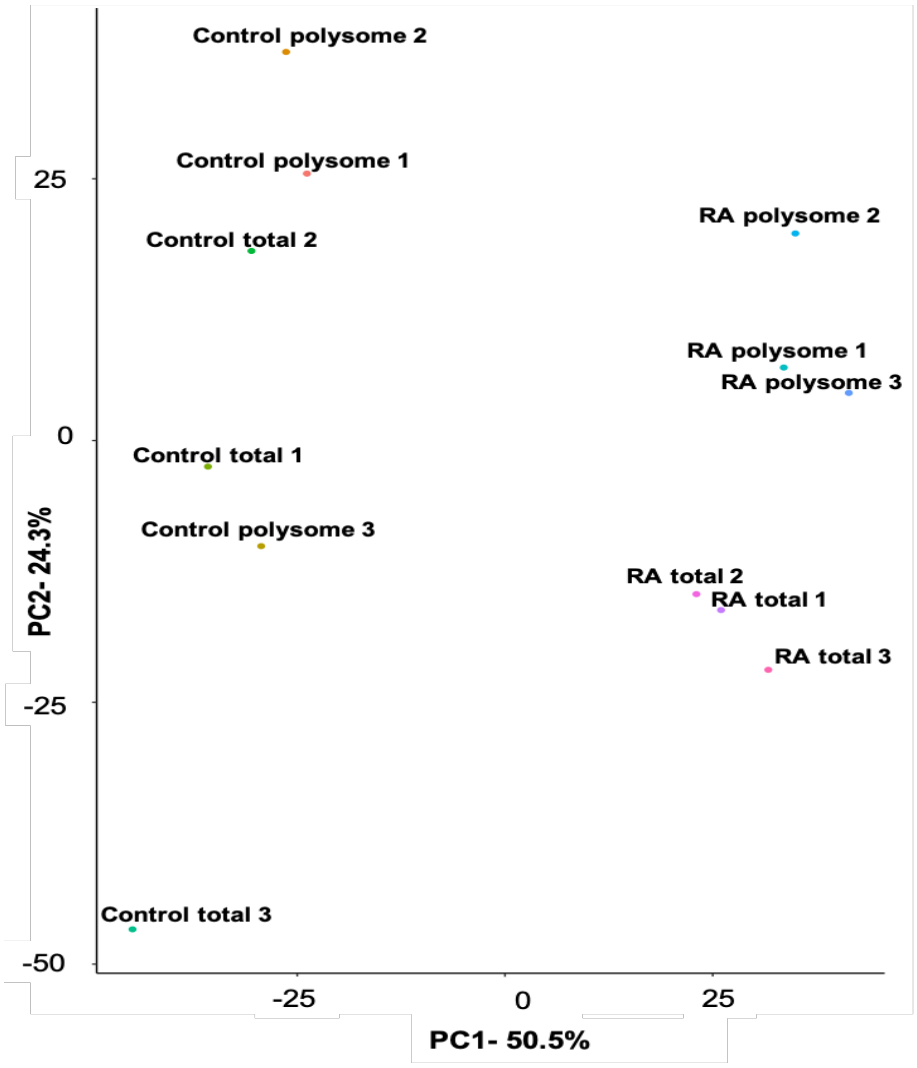


Figure 4.5: PCA of the RNA-Seq datasets shows that Control samples are more similar to each other than to RA treated samples. On the x axis the principal component that accounts for the 50.5% of variation between the data is plotted (PC1). On the y axis the principal component that accounts for the 24.3% of variation between the data is plotted. Control samples cluster separately from RA samples, indicating significant variation in gene expression between Control and differentiated cells. RA total samples cluster separately from RA polysome samples, whereas Control samples do not show different clustering, indicating that the variation in gene expression between RA total cytoplasmic (RA total) and RA polysome-associated (RA polysome) RNA samples is greater than the variation between Control total cytoplasmic (Control total) and Control polysome-associated (Control polysome) RNA samples.

4.3 Neuronal differentiation induces alterations in gene expression and affects the polysome enrichment of protein-coding genes.

To assess the differences suggested by PCA in more detail, I compared gene expression a) between control and RA treated conditions, and b) between total cytoplasmic and polysome-associated RNA populations (Figure 4.6).

4.3.1 Gene expression changes upon neuronal differentiation

Although the main aim was to identify differentially expressed lncRNAs, upon neuronal differentiation, alterations in the expression of protein-coding genes were initially analysed to understand broader gene expression changes related to neuronal differentiation. To this end, a differential expression analysis of protein-coding and lncRNA genes (section 4.3) was performed by Dr Dapeng Wang using DESeq2 (Love et al., 2014) (Methods 2.22). To identify the differentially expressed genes in total cytoplasmic and polysome-associated populations upon differentiation, total cytoplasmic and polysome-associated datasets were analysed in parallel (Table 4.2).

Table 4.2: Separate DESeq2 analyses for total and polysome populations

Purpose of analysis	Input datasets for DESeq2
Identification of total cytoplasmic differentially expressed genes upon differentiation	Control total and RA total
Identification of polysome-associated differentially expressed genes upon differentiation	Control polysome and RA polysome
Identification of genes (transcripts) enriched or depleted from polysomes in undifferentiated cells	Control total and Control polysome
Identification of genes (transcripts) enriched or depleted from polysomes in differentiated cells	RA total and RA polysome

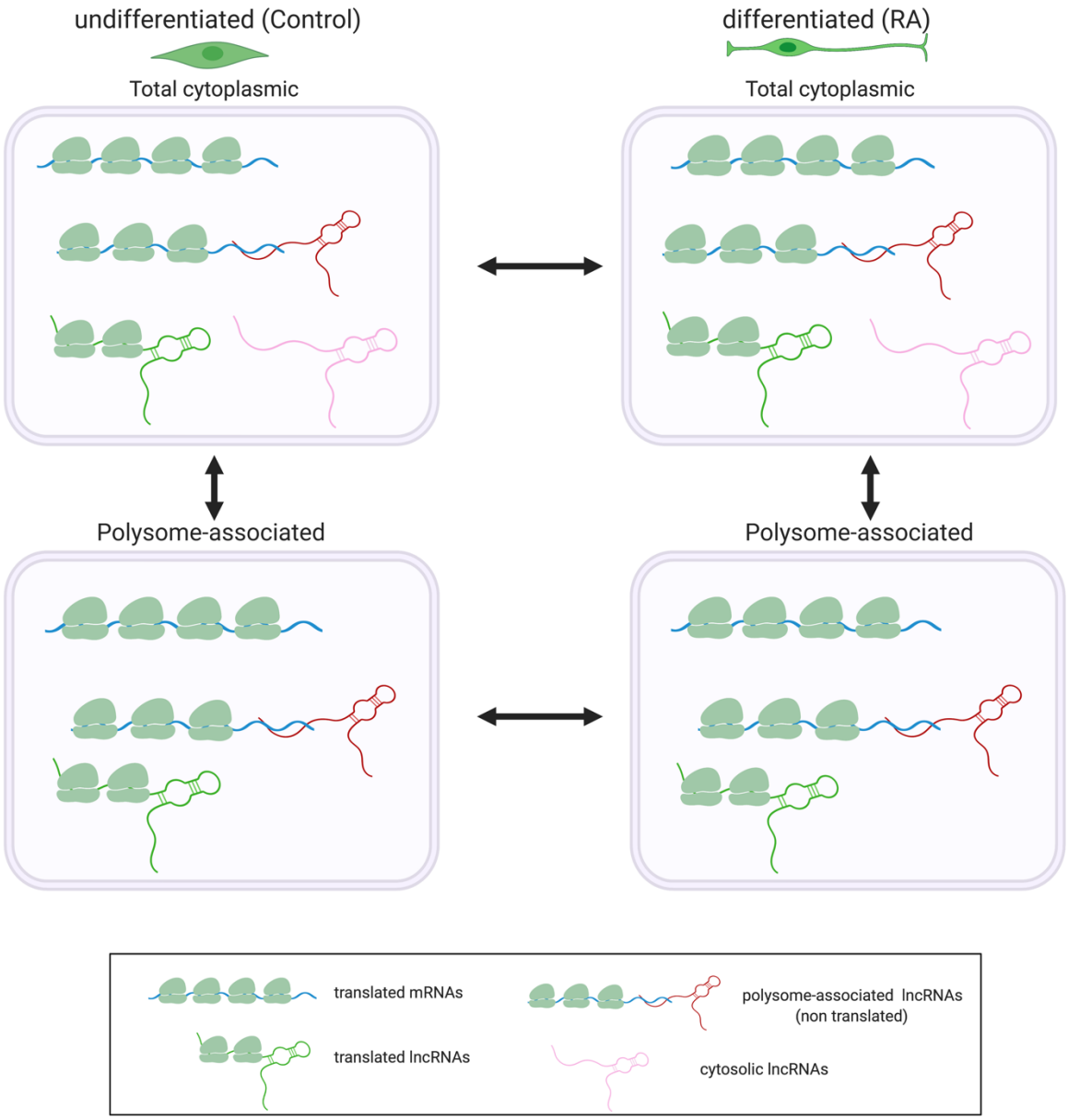


Figure 4.6: Schematic representation of the different RNA populations compared in the differential expression analyses. In the cytoplasm, there are 3 distinct populations of lncRNAs, with respect to their association with the translation machinery. The 'free' cytosolic lncRNAs, which do not interact with polysomes, the polysome-associated but not translated lncRNAs and the lncRNAs that become actively translated. The polysome-associated lncRNAs can be identified by comparing the total cytoplasmic samples to the polysome-associated samples from each condition. Image created using BIORENDER.

Upon RA differentiation, 1628 protein-coding genes were differentially expressed in total RNA-Seq (936 up- and 692 downregulated) and 1629 were differentially expressed (897 up- and 732 downregulated) in polysome-associated RNA-Seq dataset, exhibiting a $\log_2$fold change>1 (Figure 4.7). Among the genes with the highest and most significant upregulation upon RA treatment were several associated with axon growth and neuronal differentiation (NTN4, NCAM2, SLIT2). Genes involved in cell cycle (E2F1) or cancer progression (MYC) were significantly downregulated as expected during differentiation. Most of the differentially expressed genes were in common between the total cytoplasm and polysome populations (Figure 4.7 and Table 4.3). Even though polysome-associated RNA population is a subset of the total cytoplasmic RNA, more genes were detected as differentially expressed. This could be the result of translational regulation during differentiation. Notably, the polysome-associated population, containing a smaller number of transcripts, yielded very similar number of reads, therefore was sequenced more deeply allowing more differences to appear. Overall, the 2 analyses (Control total vs RA total and Control polysome vs RA polysome) had 742 upregulated and 541 downregulated genes in common (Figure 4.8).

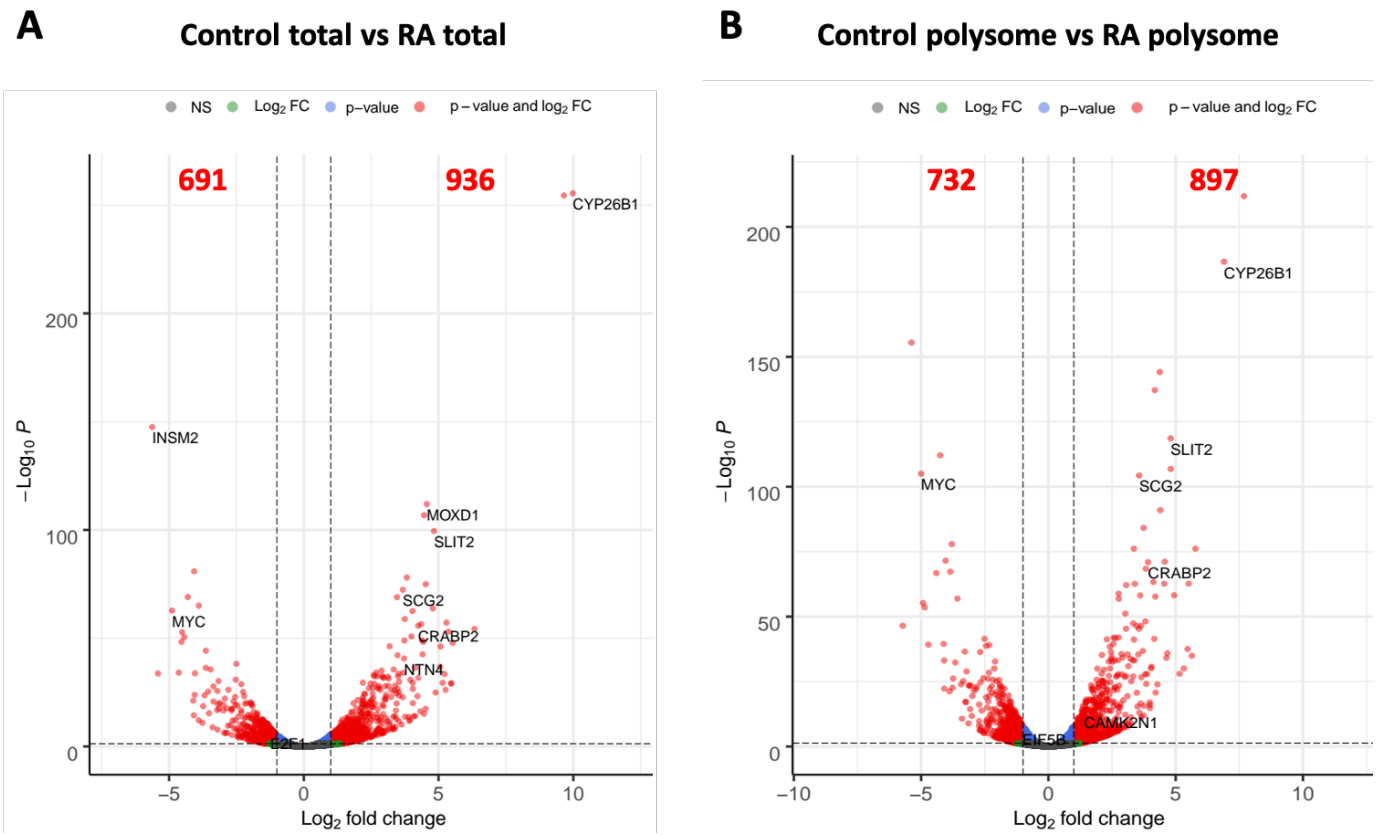


Figure 4.7: Volcano plots of differentially expressed protein-coding genes upon differentiation in total cytoplasm and polysomes. (A) In total cytoplasm 936 genes were significantly upregulated and 691 genes were significantly downregulated with a |log_2fold change|>1 upon differentiation. (B) In polysomes, 897 genes were significantly upregulated and 732 were significantly downregulated with a |log_2fold change|>1 upon differentiation. In both cases the upregulated genes were related to RA metabolism (CRABP2), axon guidance (SLIT2, NTN4) and neuronal differentiation (SCG2, CAMK2N1). Downregulated genes are related to cell cycle progression and cell proliferation (E2F1, MYC). P value cutoff for significance=0.05.

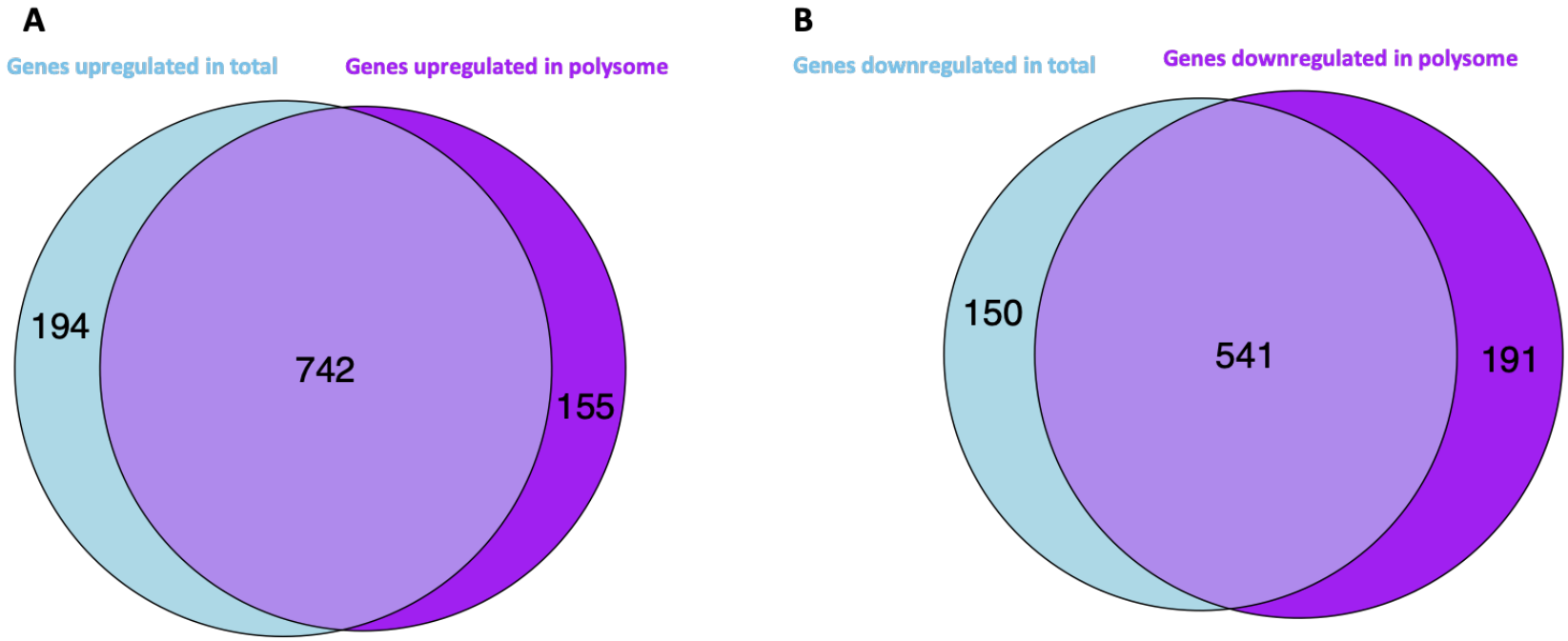


Figure 4.8: Most of the differentially expressed genes upon RA treatment are in common between total cytoplasm and polysomes. Venn diagrams showing the overlap between: (A) the upregulated protein-coding genes in the total cytoplasm (light blue) and the polysomes (purple), upon differentiation, (B) the downregulated protein-coding genes in the total cytoplasm (light blue) and the polysomes (purple) upon differentiation.

Table 4.3: Summary of gene expression changes in protein-coding genes in all different analyses

Differential expression analysis	Genes significantly upregulated with log_2fold change>1	Genes significantly downregulated with log_2fold change<-1	Genes not changed
Control total vs RA total	**936**	**692**	7,131
Control polysome vs RA polysome	**897**	**732**	6,527
Control total vs Control polysome	**15**	**19**	10,657
RA total vs RA polysome	**5**	**55**	9,387

The number of genes that are significantly up or downregulated with a log_2fold change>|1| is reported. The number of genes that do not show any changes in expression is derived by subtraction of the number of significantly differentially expressed genes (regardless of log_2fold change) from the total number of genes detected in each dataset.

To understand the general changes in gene expression, induced by neuronal differentiation, I performed Gene Ontology over-representation analysis (GO-ORA) on the up and downregulated protein-coding genes. The GO terms enriched in both total cytoplasmic (Control total vs RA total) and polysome-associated (Control polysome vs RA polysome) genes upon differentiation, were similar (Figure 4.9). Upregulated genes were enriched in GO terms relevant to differentiation, such as ‘cell migration’, ‘neuron cell-cell adhesion’ and ‘synaptic transmission’ (Figure 4.9: A,C). This is consistent with the phenotypic characteristics of the differentiating SH-SY5Y cells, as they extend their neurites and connect with each other during differentiation. Downregulated genes upon differentiation in total cytoplasmic, as well as in polysome-associated populations were enriched in GO terms relevant to ribosome biogenesis and translation (Figure 4.9: B,D), consistent with the reduction in level of active translation observed following RA treatment (Chapter 3).

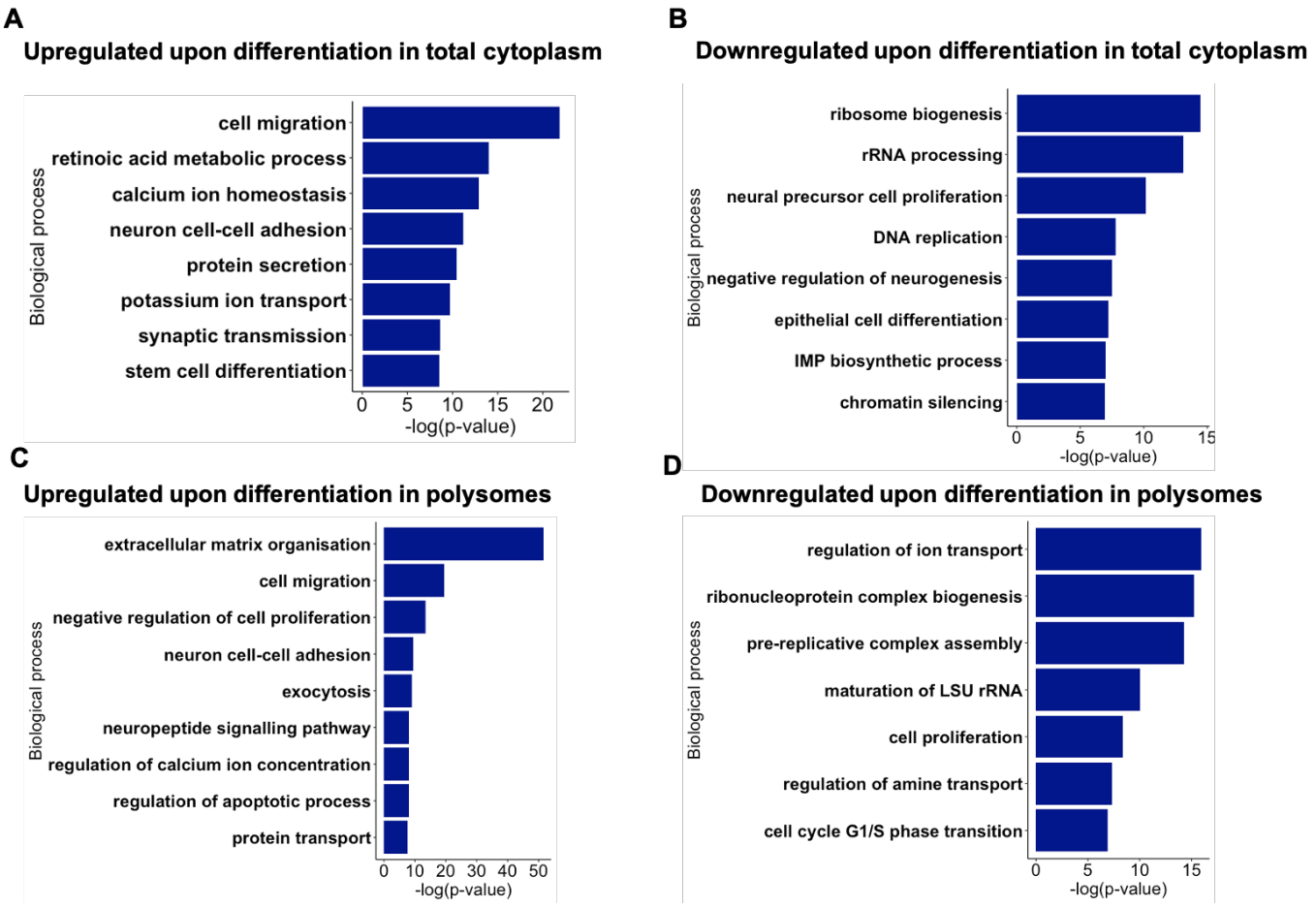


Figure 4.9: GO terms enriched in differentially expressed genes in total cytoplasm and polysomes. Upregulated genes in (A) total cytoplasm and (C) polysome complexes are enriched in terms relevant to neuronal differentiation, axon development, secretion and cell migration. Downregulated genes in (B) total cytoplasm and (D) polysome complexes are enriched in terms relevant to

translation regulation and ribosome biogenesis. GO analysis was performed using GOrilla platform and GO terms were filtered based on p-value cutoff of 0.001.

4.3.2 Polysome association of protein-coding genes is dynamic upon neuronal differentiation

The PCA (Figure 4.5) showed that for the RA treated samples, expression profiles of cytoplasmic and polysome-associated fractions clustered separately, suggesting that differentiation is likely to induce alterations in polysome enrichment of certain mRNAs as a result of translation regulation. To identify those alterations, I first compared the expression levels (RPKM values) between total and polysome-associated protein-coding genes in Control and RA treated cells. This comparison revealed a marked depletion of ribosomal protein mRNAs from polysomes, upon differentiation (Figure 4.10; and Appendix-II for other replicates).

To profile more robustly the changes in polysome enrichment or depletion upon differentiation, separate differential expression analyses were performed between total cytoplasmic and polysome-associated populations, for Control and RA treated cells respectively (Table 4.2–Dr Dapeng Wang). In differentiated cells, 33 out of the 55 protein-coding genes, significantly depleted from polysomes, were ribosomal protein genes (Figure 4.11). In Control cells only 7 ribosomal protein-coding genes appear significantly depleted from polysomes. Taken together, the above results show that ribosomal protein mRNAs are depleted from polysomes upon RA treatment and this is consistent with the reduction in the level of active translation, upon differentiation (Chapter 3) and with the clustering of RA total and polysome samples separately from each other in the PCA plot (Figure 4.5).

Previous GO term analysis (Figure 4.9) had also revealed that genes involved in ribosome biogenesis were downregulated in total cytoplasmic and polysome-associated RNA populations, upon differentiation. Therefore, it is possible that depletion of these mRNAs from actively translating ribosomes represents an additional layer of translational regulation, on top of transcriptional downregulation.

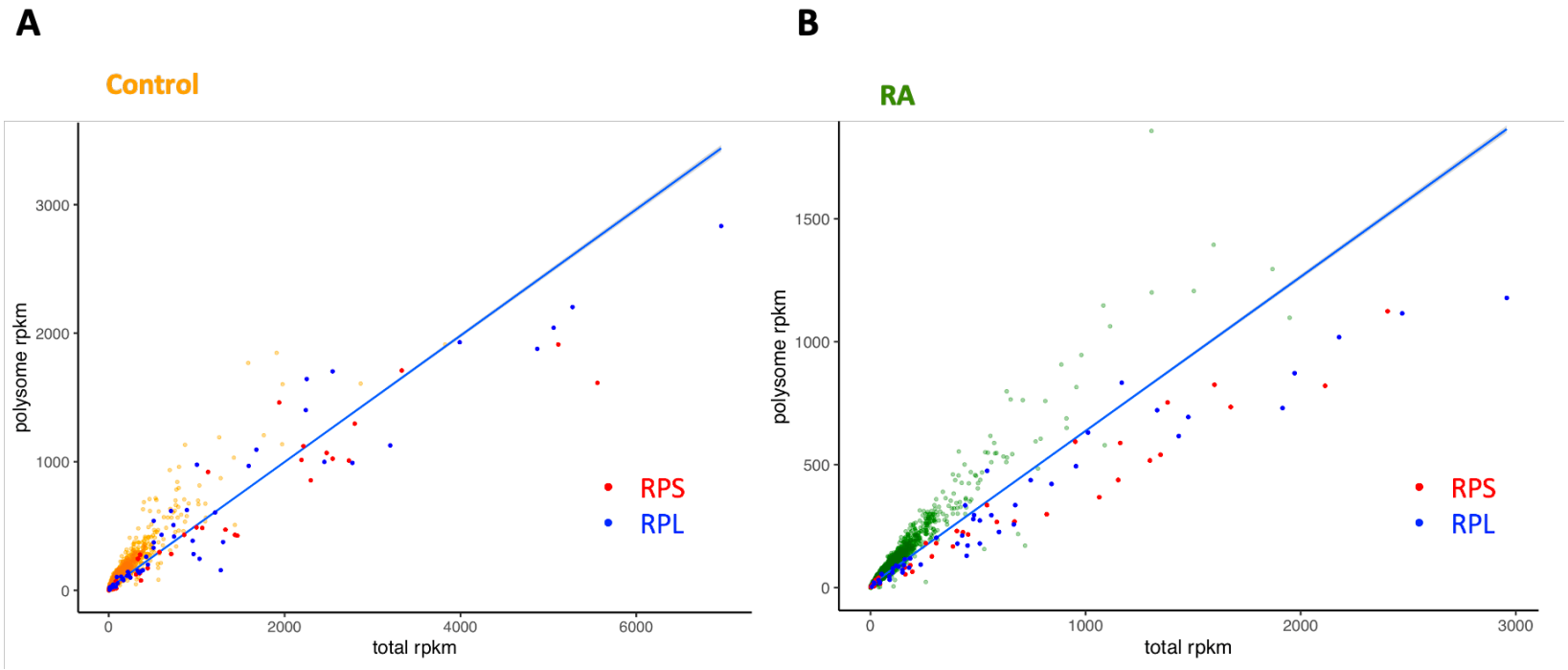


Figure 4.10: Ribosomal protein-coding genes are depleted from polysomes upon differentiation. Scatter plots of total/polysome in (A) Control (orange dots) and (B) RA treated (green dots) cells of replicate 3. Ribosomal protein genes in RA treated cells shift from polysomes to total. Large ribosomal proteins (RPL) are labelled as blue dots and small ribosomal proteins (RPS) as red dots. Blue line is the linear model that fitted the data.

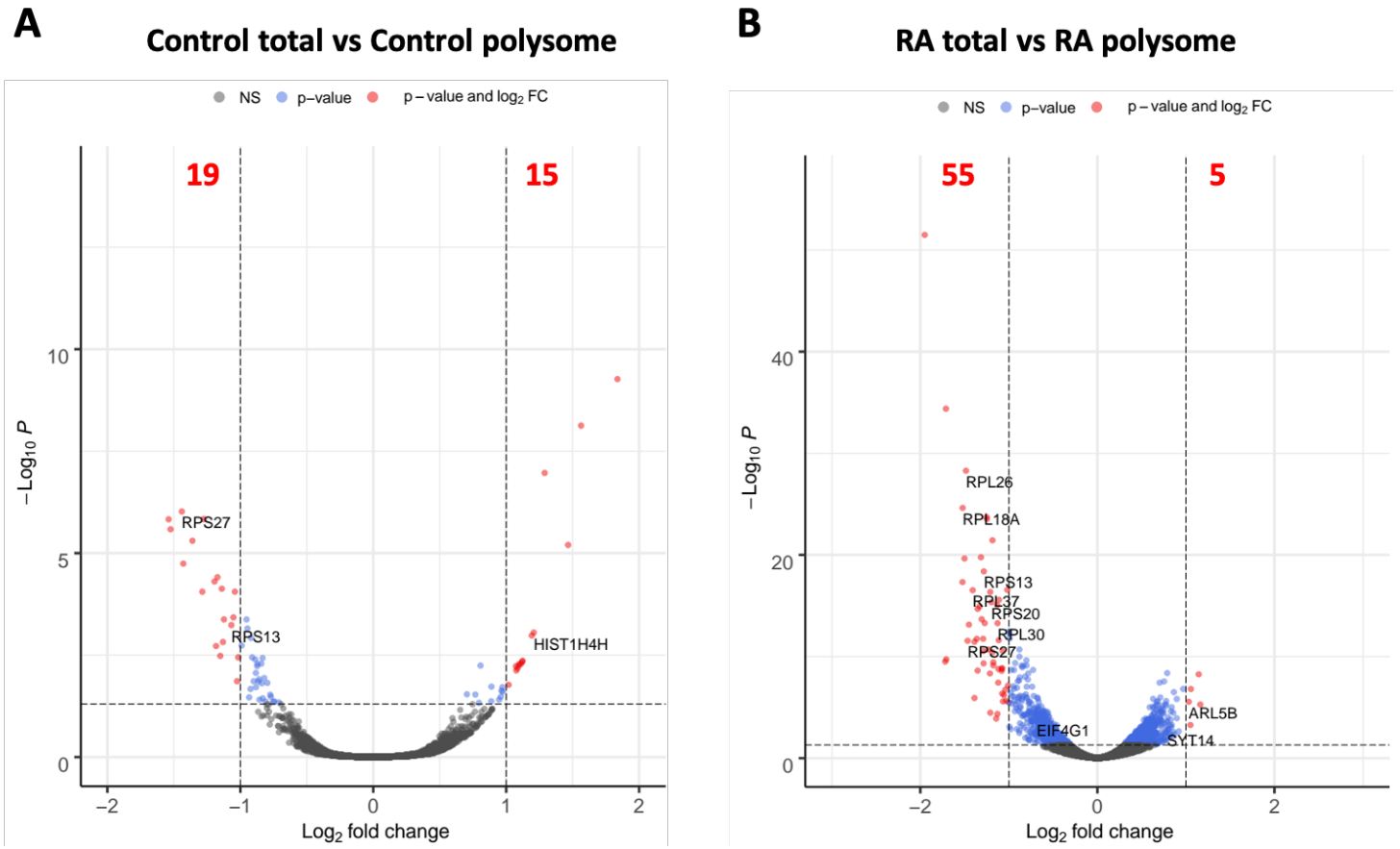


Figure 4.11: Volcano plots of differentially enriched protein-coding genes in polysomes in Control and differentiated cells. (A) In Control, 15 genes were significantly enriched and 19 were significantly depleted from polysomes, among them some ribosomal protein-coding genes. (B) In differentiated cells, 5 genes were significantly enriched and 55 were significantly depleted. Notably almost all of the significantly depleted from polysomes in differentiated cells were ribosomal protein genes. P value for significance=0.05.

To experimentally validate the depletion of ribosomal protein mRNAs from polysomes I assessed their distribution across sucrose gradients. RT-qPCR of specific ribosomal protein mRNAs was performed across fractions corresponding to RNPs+40S+60S, 80S monosomes and polysomes. Cytoplasmic lysates of Control and RA treated cells were subjected to ultracentrifugation (3 gradients per treatment, followed by fractionation and RNA extraction from the fractions corresponding to ribonucleoparticles (RNPs), ribosomal subunits, 80S and polysomes (Figure 4.12), (Methods 2.5.2).

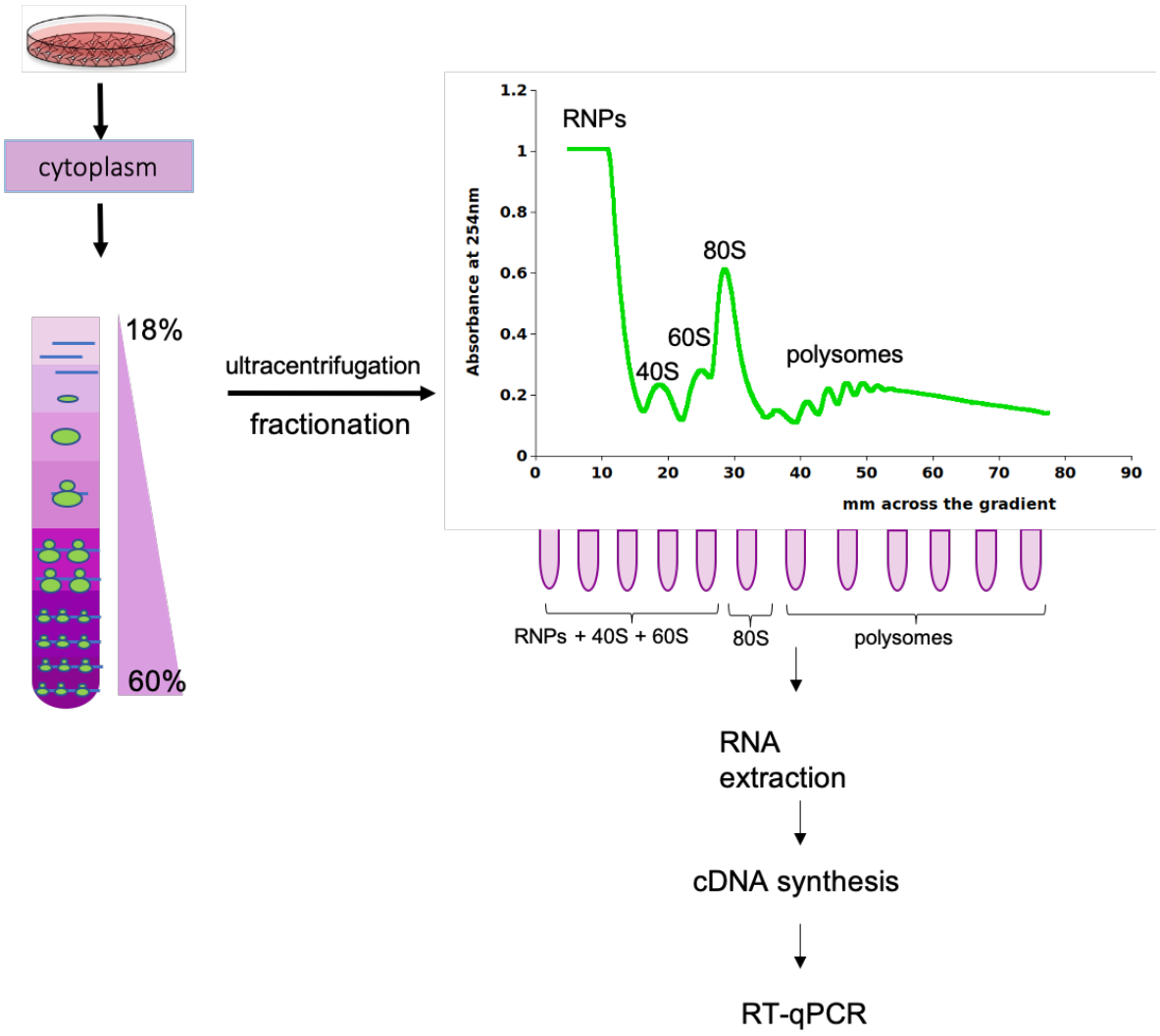


Figure 4.12: Schematic of polysome fractionation for RT-qPCR. Ribosomal subunits, monosomes and polysomes are separated by ultracentrifugation and fractionated. RNPs + 40S + 60S subunits sediment in fractions 1-5, 80S in fraction 6 and polysomes sediment in fractions 7-12. RNA was extracted from each fraction, cDNA synthesis was performed, and samples were processed for RT-qPCR.

The ribosomal protein (RP) transcripts selected for validation were chosen based on their differential enrichment in polysomes. Specifically, RPL26, RPS28 and RPL37 were depleted from polysomes only in RA treated cells. RPS13 was depleted from polysomes both in Control and RA treated cells. RPS5 was not depleted from polysome in any of the conditions and was used as negative control. To assess the relative enrichment of the RP transcripts in the polysomes, and calculated the quantity of mRNA in each fraction as a percentage of the total mRNA was determined using a previously published approach (Chassé et al., 2016). GAPDH mRNA was used as a control mRNA as it showed no change in polysome enrichment upon differentiation.

RT-qPCR indicated that 83% and 70% of the total GAPDH transcripts are present in the polysomes in Control and differentiated cells respectively, showing no significant difference in distribution upon differentiation (Figure 4.13: A), whilst RPL26 and RPL37 mRNAs are both significantly depleted from polysomes upon differentiation (RPL26 RNA-Seq log_2fold change=-1.5, RPL37 RNA-Seq log_2fold change=-1.4) (Figure 4.13: A, B). The proportion of RPL26 transcripts present in polysome fractions (32% of the total on average) of differentiated cells, was significantly lower compared to that of Control cells (57% of the total on average). Concomitantly RPL26 mRNA levels in the RNPs + 40S + 60S fractions significantly increased upon differentiation. Similarly, the proportion of RPL37 transcripts present in polysome fractions (23% of the total on average) of differentiated cells, was significantly lower compared to that of Control (47% of the total on average). RPL37 mRNA levels were increased in the RNPs + 40S + 60S fractions, but it was not statistically significant.

RPS28 mRNA distribution across the gradient showed a similar trend to RPL26 and RPL37 (Figure 4.13: C). However, the change upon differentiation was not as substantial as with RPL26 and RPL37. This fits with the lower log_2fold change of –1.1 for RPS28 in the RNA-Seq differential expression analysis between RA total and RA polysome

As predicted from RNA-Seq analysis, RT-qPCR for RPS5 mRNA across the gradient showed no significant changes in the distribution between Control and differentiated cells. In differentiated cells, only 29% of RPS13 mRNA is present in the polysomes. However, in Control cells, 47% of the total RPS13 transcripts is present in

the polysomes. These data indicate that RPS13 mRNA does shift from polysomes to subunits, but to a lesser and statistically insignificant extent than RPL37 and RPL26. This fits with the RNA-Seq analysis, in which RPS13 is significantly depleted from polysomes with a greater log_2fold change in RA treated cells (RNA-Seq log_2fold change = -1.3) compared to Control (RNA-Seq log_2fold change = -1).

Together, the above results validate that RA treatment induces a depletion of RPL26 and RPL37 mRNAs from the polysomes and support a trend towards proportionate depletion of some RP mRNAs from polysomes and their enrichment in RNPs+40S+60S fractions upon RA induced differentiation, indicating that their translation is repressed during differentiation.

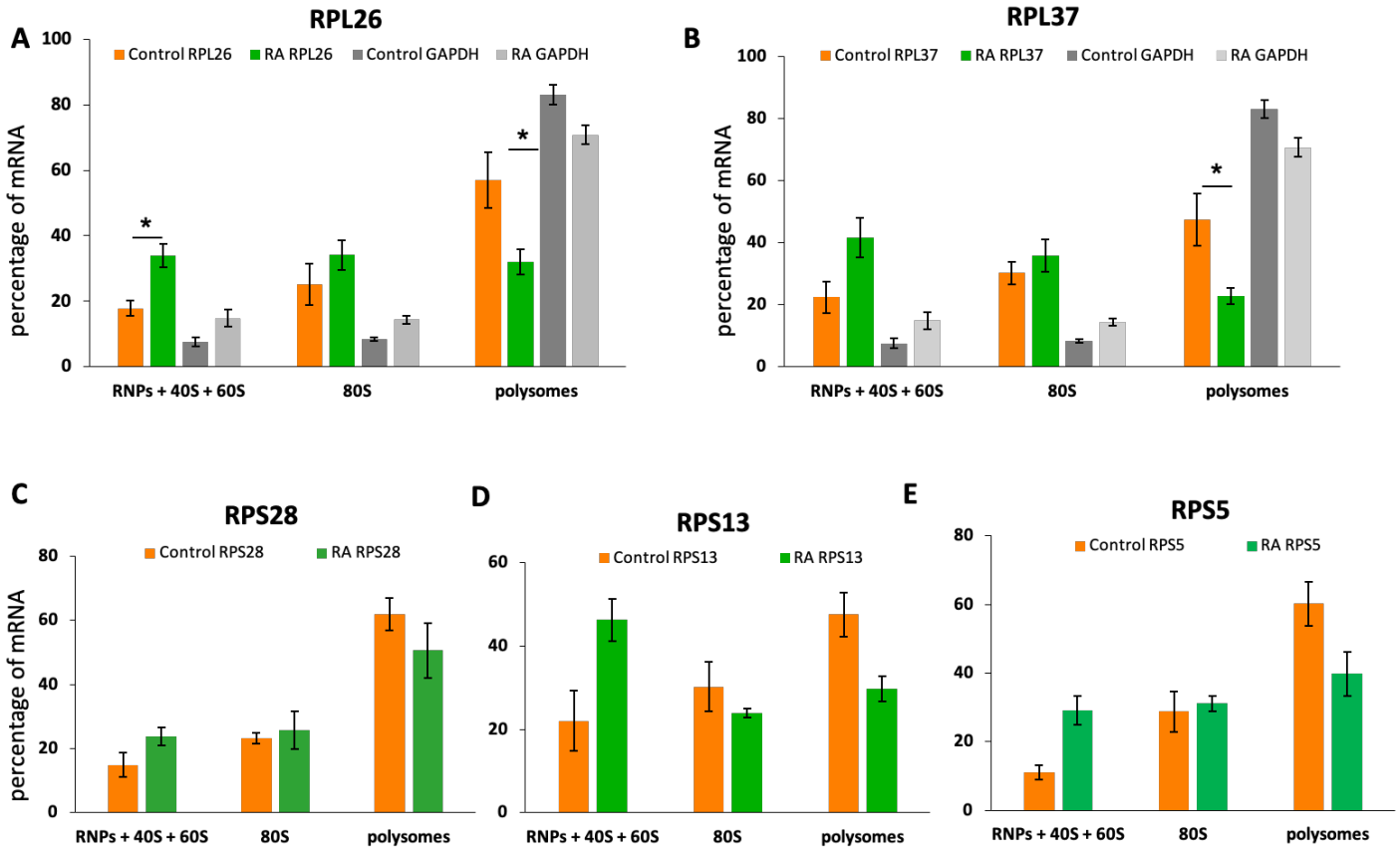


Figure 4.13: Ribosomal protein mRNAs shift from polysomes to ribosomal subunits upon differentiation. Distribution of ribosomal protein mRNAs and GAPDH mRNA (as control, shown in A and B) in the RNPs and ribosomal subunits (40S+60S), the monosomes (80S) and the polysomes is shown as a percentage of the total. (A) Proportion of RPL26 mRNA in the polysomes of RA

treated cells is significantly lower compared to Control and significantly higher in the RNPs+40S+60S subunits, compared to Control (student's t test, n=3, $p<0.05$). (B) RPL37 percentage in the polysomes of RA treated cells is significantly lower compared to control (student's t test, n=3, $p<0.05$). (C, D) Proportion of RPS28 and RPS13 in the polysomes of RA treated cells is lower compared to Control. Similarly, their proportion in the 40S+60S subunits is higher in RA treated cells compared to control but these differences are non-significant (student's t test, n=3, $p>0.05$). (E) Distribution of RPS5 mRNA across the gradient showed no significant changes between Control and differentiated cells (student's t test, n=3, $p>0.05$). GAPDH mRNA distribution shows no significant changes upon differentiation (student's t test, n=3, $p>0.05$). RT-qPCR was performed in samples from 3 biological replicates and error bars show standard error between biological replicates.

4.4 LncRNA gene expression and polysome association changes upon neuronal differentiation.

Having validated that large scale protein-coding gene transcriptional changes are occurring upon differentiation, I aimed to determine if lncRNAs relationship with the translation machinery is affected by differentiation and whether specific lncRNAs potentially regulated this process.

4.4.1 Differentiation induces alterations in lncRNA gene expression

To understand the effect of neuronal differentiation on the expression of lncRNAs, differential gene expression analysis was performed (Dr Dapeng Wang) for lncRNA genes (4.2.1 and 4.2.2). I compared differences in lncRNAs levels between Control and RA treated cells, in both the total cytoplasmic and polysomal RNA populations. Changes in lncRNA expression were considered significant when the p value<0.05. In total cytoplasm, 801 lncRNA genes with an RPKM>1 in Control and 916 lncRNA genes in differentiated cells were detected in all replicates. 178 lncRNA genes were significantly upregulated with a log_2fold change>1 and 100 were significantly downregulated in total cytoplasm. 237 polysome-associated lncRNAs were significantly upregulated with a log2fold change>1 and 82 were significantly downregulated (Table 4.4 - Figure 4.14). Moreover, 126 of the differentially expressed lncRNA genes were in common between the 2 differential expression analyses (Figure 4.15). This suggests that a substantial proportion of the differentially expressed cytoplasmic lncRNAs interact with actively translating polysomes.

Table 4.4: Summary of gene expression changes in lncRNA genes in all different analyses

Differential expression analysis	LncRNA genes significantly upregulated	LncRNA genes significantly upregulated with log_2fold change>1	LncRNA genes significantly downregulated	LncRNA genes significantly downregulated with log_2fold change<-1	LncRNA genes not changed
Control total vs RA total	**230**	**178**	**137**	**100**	325
Control polysome vs RA polysome	**329**	**237**	**122**	**82**	274
Control total vs Control polysome	**3**	**0**	**24**	**12**	657
RA total vs RA polysome	**47**	**0**	**58**	**10**	715

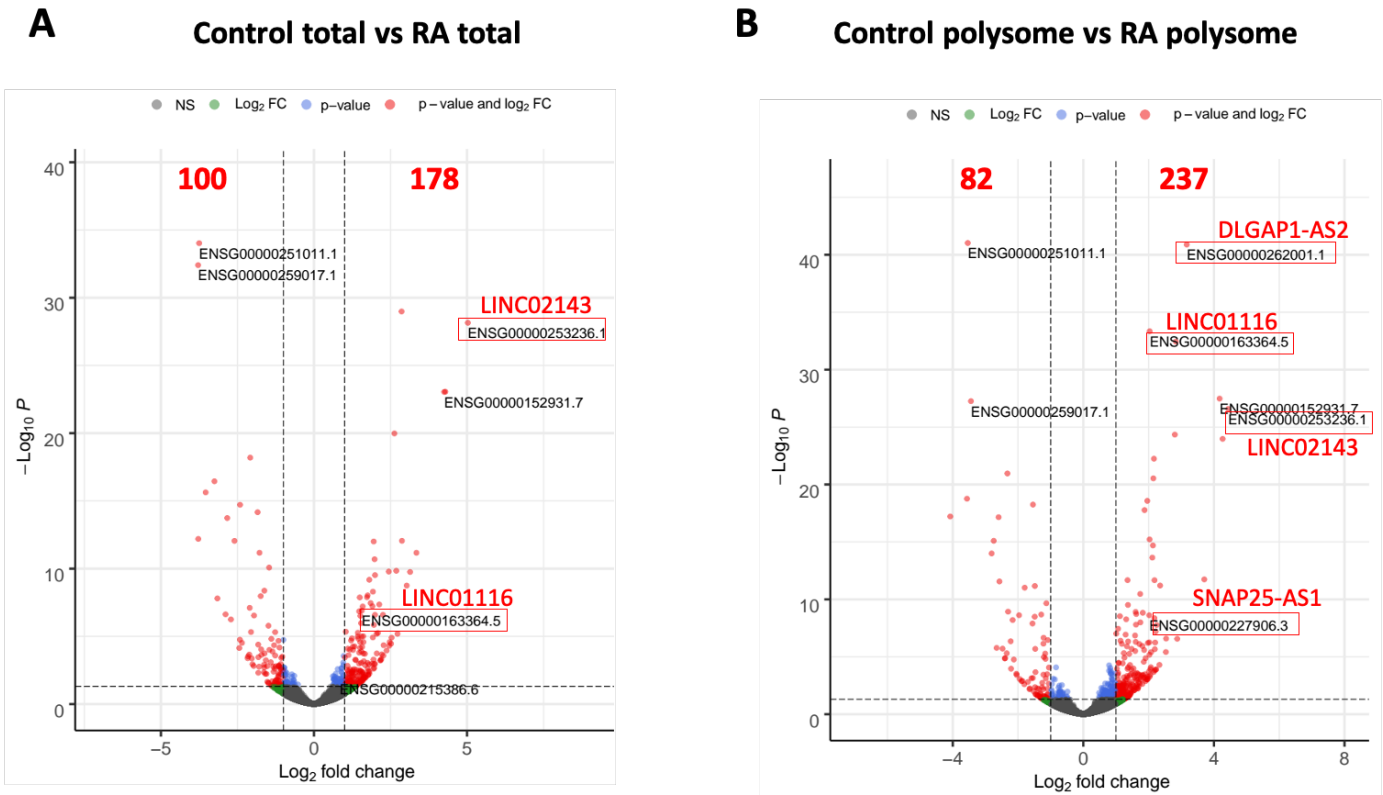


Figure 4.14: Volcano plots of differentially expressed lncRNA genes upon differentiation in total cytoplasm and polysomes. In total cytoplasm (A) 178 lncRNA genes were significantly upregulated (among them LINC01116 and LINC02143) and 100 were significantly downregulated. In polysomes (B) 237 lncRNA genes were significantly upregulated (among them DLGAP1-A2, SNAP25-AS1, LINC01116 and LINC02143) and 82 were significantly downregulated. P value for significance=0.05.

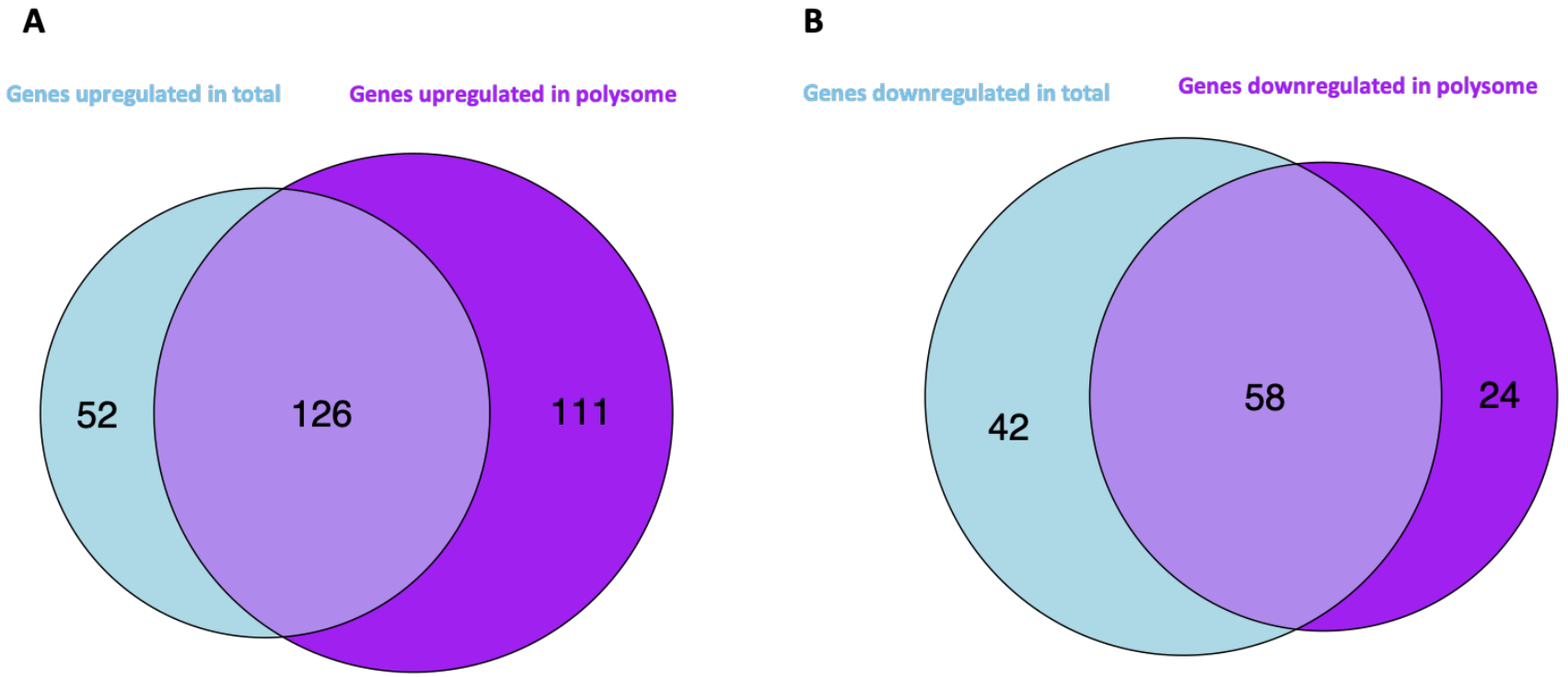


Figure 4.15: Most of the differentially expressed lncRNA genes upon RA treatment are in common between total cytoplasm and polysomes. Venn diagrams showing the overlap between: (A) the upregulated lncRNA genes in the total cytoplasm (light blue) and the polysomes (purple), upon differentiation, (B) the downregulated lncRNA genes in the total cytoplasm (light blue) and the polysomes (purple) upon differentiation.

The functions of lncRNAs are not known, therefore it was not possible to use GO analysis to understand general changes in lncRNA expression patterns. However, to explore any specific global pattern in lncRNA gene expression, I examined the types of lncRNAs that were differentially expressed. In summary, the majority of the lncRNA transcripts that were up or downregulated during differentiation in total cytoplasm and polysomes, belong to the categories of intergenic and anti-sense (Figure 4.16). This is expected and consistent with the fact that the vast majority of lncRNA transcripts are intergenic or anti-sense (Gencode v.29, release GRch38).

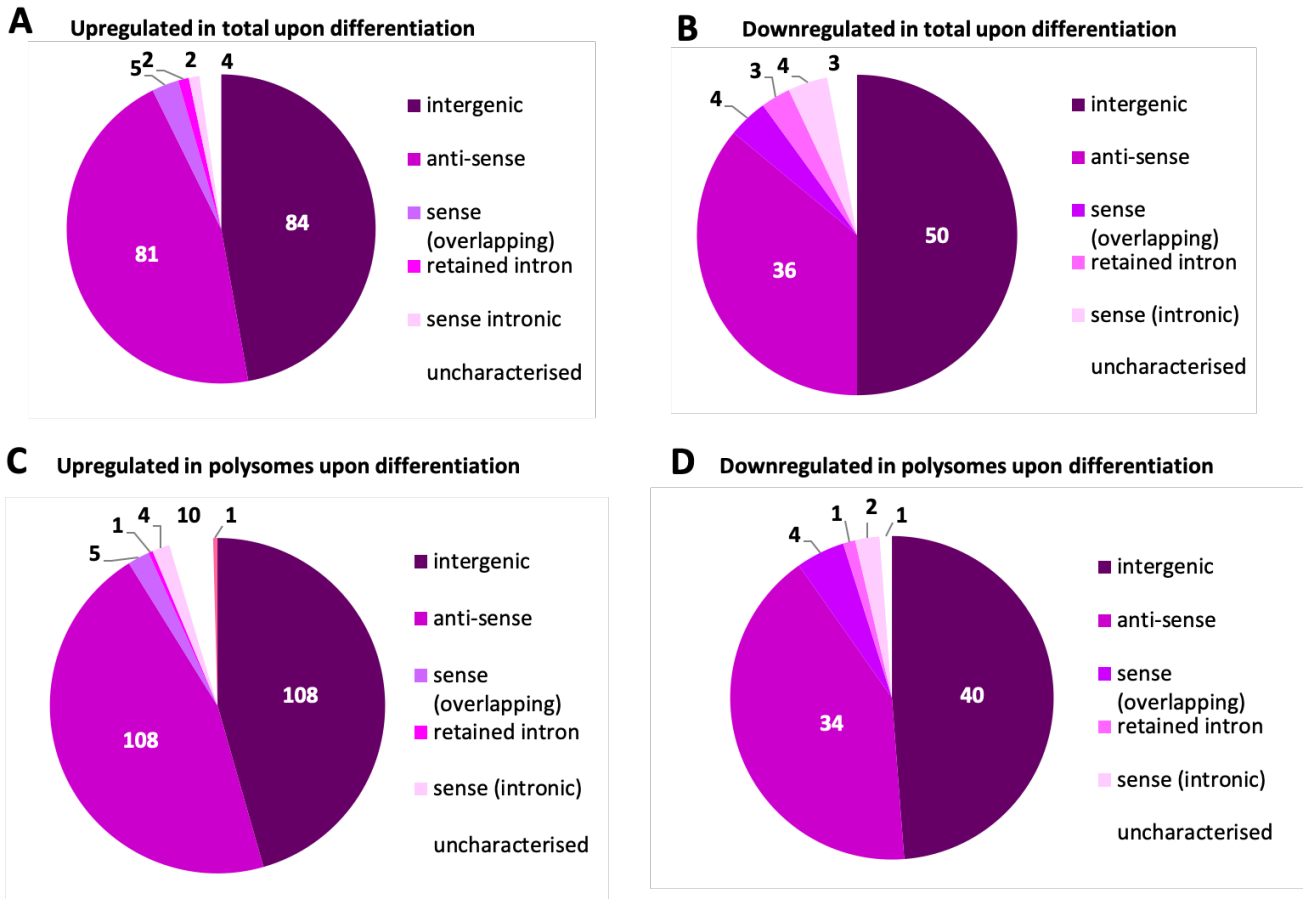


Figure 4.16: The majority of lncRNAs regulated during differentiation are intergenic or anti-sense type lncRNAs. (A) 165 out of 178 upregulated lncRNAs in the total cytoplasm upon differentiation are either intergenic or anti-sense. (B) Half of the downregulated lncRNAs in total cytoplasm upon differentiation are intergenic. (C) 216 out of 237 lncRNA transcripts are intergenic or anti-sense. 1 transcript is mis-annotated as NMD target (D) Almost half of the downregulated polysome-associated lncRNAs upon differentiation are intergenic.

4.4.2 LncRNA association with polysomes is dynamic during neuronal differentiation

To understand how cytoplasmic lncRNAs might interact with the translational machinery I wanted to determine how the interactions of lncRNAs with polysomes might change during differentiation. To determine if lncRNAs have a preference for associating with polysomes the levels of each lncRNA between total cytoplasm and polysome complexes were compared. This was performed in both Control and RA conditions. To achieve this differential expression analysis was performed (Table 4.2-Dr Dapeng Wang). LncRNAs are generally present in polysomes or depleted from polysomes, not enriched (Figure 4.17). Specifically, in Control cells (Control total vs Control polysome), 12 lncRNAs were depleted from the polysomes but no lncRNAs were enriched in polysomes. In RA treated cells (RA total vs RA polysome) 10 lncRNAs were significantly depleted from polysomes with a $|\log_2 \text{fold change}|>1$ but no lncRNAs were enriched in polysomes with a $|\log_2 \text{fold change}|>1$ (Table 4.4). However, there were certain lncRNAs significantly enriched in polysomes, with a $0<\log_2 \text{fold change}<1$. Specifically, in Control, only 3 lncRNA genes were enriched in polysomes at this lower fold difference, whilst in RA treated cells, 47 lncRNAs were enriched in polysomes (Figure 4.17- lncRNAs with the highest enrichment in polysomes labelled). The lncRNAs that were depleted from polysomes in Control but not in RA treated cells are potentially interesting, as their interaction with polysomes might be regulated. 10 such lncRNAs were identified, 5 of them were annotated as small nucleolar RNA host genes, suggesting potential regulation of snoRNAs during differentiation. This is consistent with previous studies that report the differential expression of snoRNAs during neuronal differentiation, as well as the post-transcriptional regulation of neuronal genes by snoRNAs (Skreka et al., 2012; Bratkovič et al., 2018). One is small Cajal body specific RNA 9 and the rest were novel unannotated transcripts.

To identify potential patterns among the lncRNAs depleted from polysomes, I examined the categories of lncRNAs, to which those transcripts belong. The majority of the lncRNAs that were depleted from polysomes in Control and upon differentiation are

intergenic and anti-sense type lncRNAs (Figure 4.18), consistent with the fact that the vast majority of lncRNA transcripts are intergenic (Gencode v.29, release GRch38).

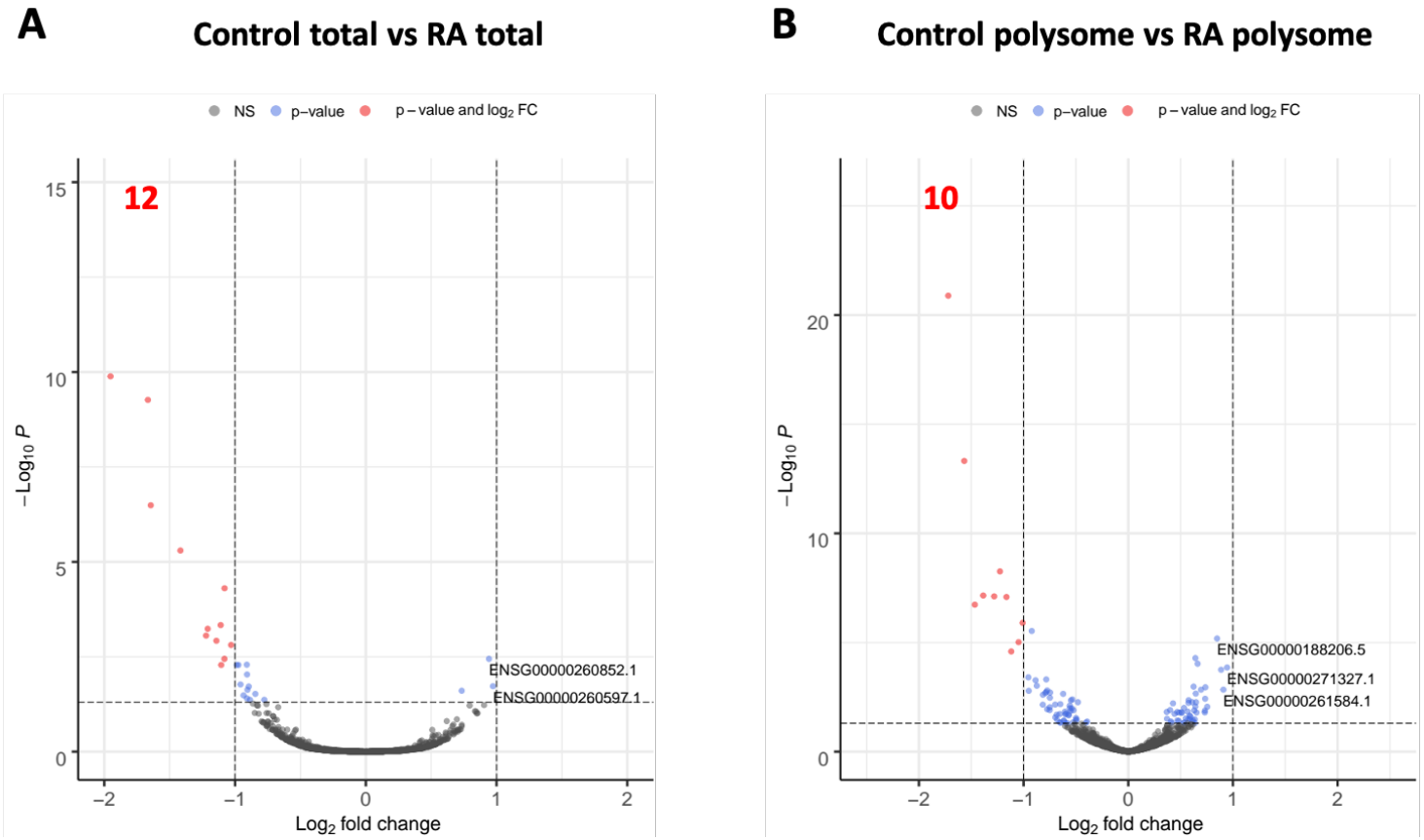


Figure 4.17: Volcano plots of differentially enriched lncRNA genes in polysomes in Control and differentiated cells. (A) In Control cells 12 lncRNAs were significantly depleted from polysomes. Only 3 lncRNA genes were enriched in polysomes but their log2fold change is lower than 1. (B) In RA treated cells 10 lncRNAs were significantly depleted with a $|\log_2 \text{fold change}|>1$. 47 lncRNAs were enriched in polysomes, with a $|\log_2 \text{fold change}|<1$.

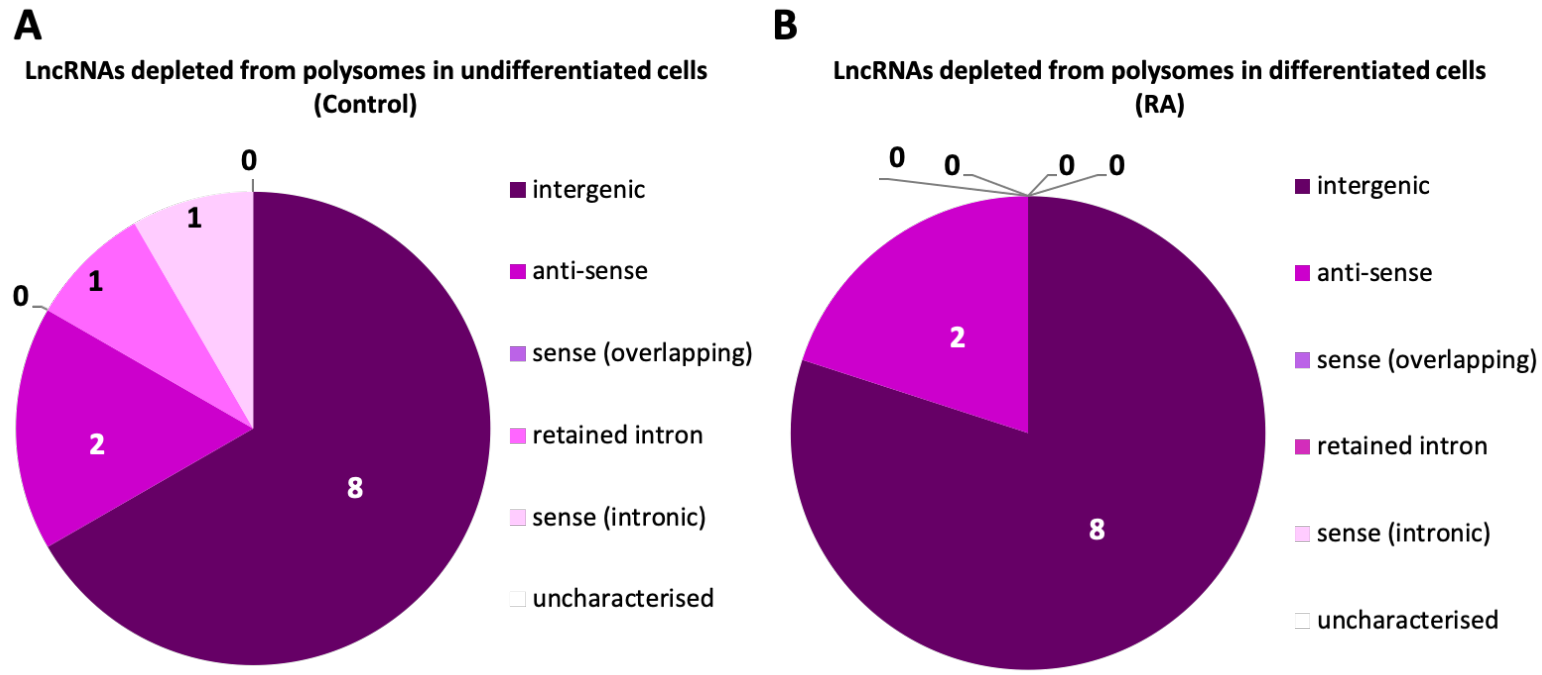


Figure 4.18: The majority of lncRNAs depleted from polysomes in Control and differentiated cells are intergenic. (A) 67% of polysome depleted lncRNAs in Control are intergenic. (B) 80% of the polysome depleted lncRNAs upon differentiation are intergenic.

4.5 *In vitro* validation of upregulation of lncRNA expression upon differentiation

Analyses of differential lncRNA expression upon differentiation identified 178 upregulated and 100 downregulated lncRNAs in the cytoplasm of differentiated cells. I chose to further study the upregulated lncRNAs as those would be more likely to be associated with changes during the progress of differentiation. Specifically, lncRNAs with log_2fold change>1 were selected as targets for further study, based on a recently published functional study of lncRNAs (Perry et al., 2018). To confirm their upregulation upon differentiation 7 candidates lncRNAs (Table 4.5) were selected, based on their levels of upregulated upon RA treatment (Table 4.5).

To validate their expression levels, RT-qPCR was performed. This was done on independent samples to those Poly-Ribo-Seq was performed upon, with 3 biological replicates and 3 technical replicates in each biological replicate. RT-qPCR was performed for LINC01116, LINC02143, DLGAP1-AS1, DLGAP1-AS2, SERPINB9P1, SNAP25-AS1 and AC254633.1 lncRNA transcripts. GAPDH and XIST were used as control protein-coding and lncRNA transcripts, respectively, which exhibited no difference in expression upon differentiation by Poly-Ribo-Seq (Figure 4.19).

The expression of 7/7 candidate lncRNAs was increased in differentiated cells, compared to Control, as was expected. To examine whether the level of upregulation quantified by RT-qPCR was similar to the log_2fold change from the RNA-Seq data, the correlation between the RT-qPCR log_2fold change and the RNA-Seq log_2fold change (Figure 4.20) was assessed and they appeared to show high correlation (Pearson correlation coefficient=0.85).

Table 4.5: LncRNA candidates upregulated in total cytoplasm and polysomes upon differentiation

LncRNA	Log_2fold change in total cytoplasm	Log_2fold change in polysome
SNAP25-AS1	1.9	2.01
AC254633.1	4.29	4.26
LINC01116	1.56	2.03
SERPINB9P1	2.87	2.21
DLGAP1-AS2	2.86	3.17
DLGAP1-AS1	1.97	2.14
LINC02143	5.02	4.43

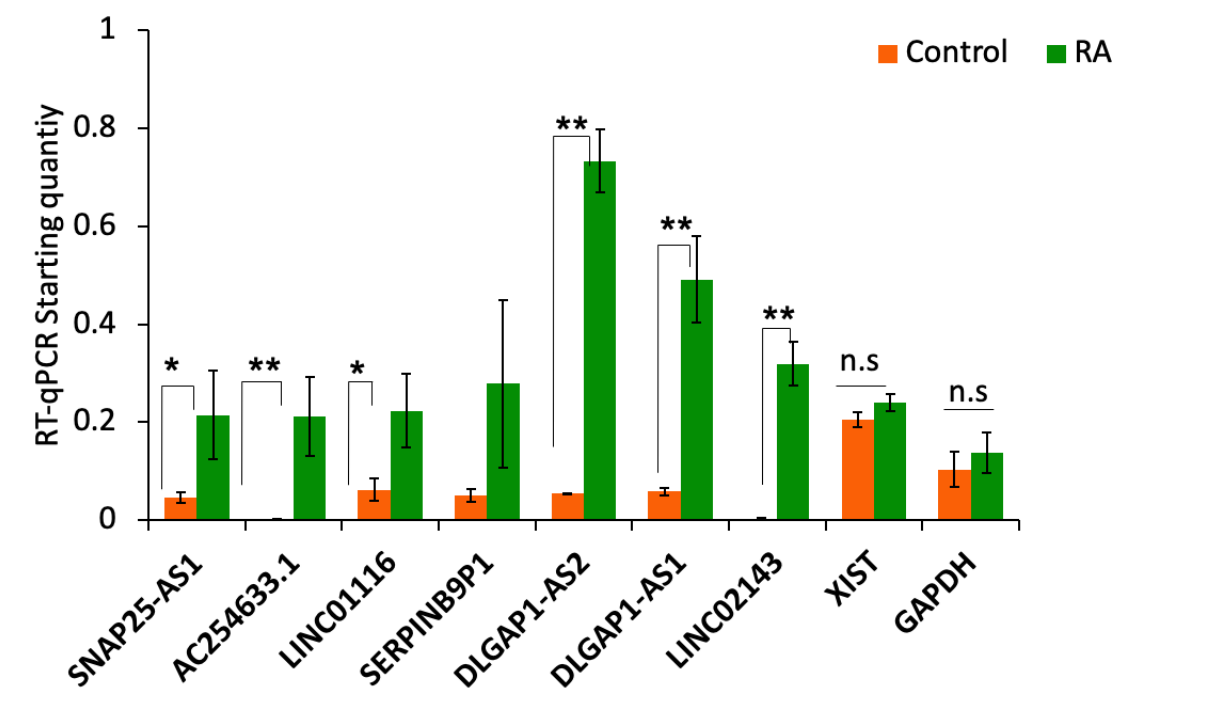


Figure 4.19: Validation of selected lncRNAs are upregulated during neuronal differentiation. RT-qPCR of LINC01116, LINC02143, DLGAP1-AS1, DLGAP1-AS2, SERPINB9P1, SNAP25-AS1 and AC254633.1 lncRNA transcripts in whole cell extract of undifferentiated (Control) and differentiated (RA) cells (3 biological replicates, 3 technical replicates for each sample). 6 out of the 7 candidates were significantly upregulated (student's t test, n=3, *p<0.05, **p<0.01). XIST and GAPDH were used as controls of a lncRNA and a protein-coding transcript that do not exhibit differential expression during differentiation (student's t test, n=3, p>0.05). Experiment performed by Sophie Clayton, undergrad student, Aspden lab, who I supervised.

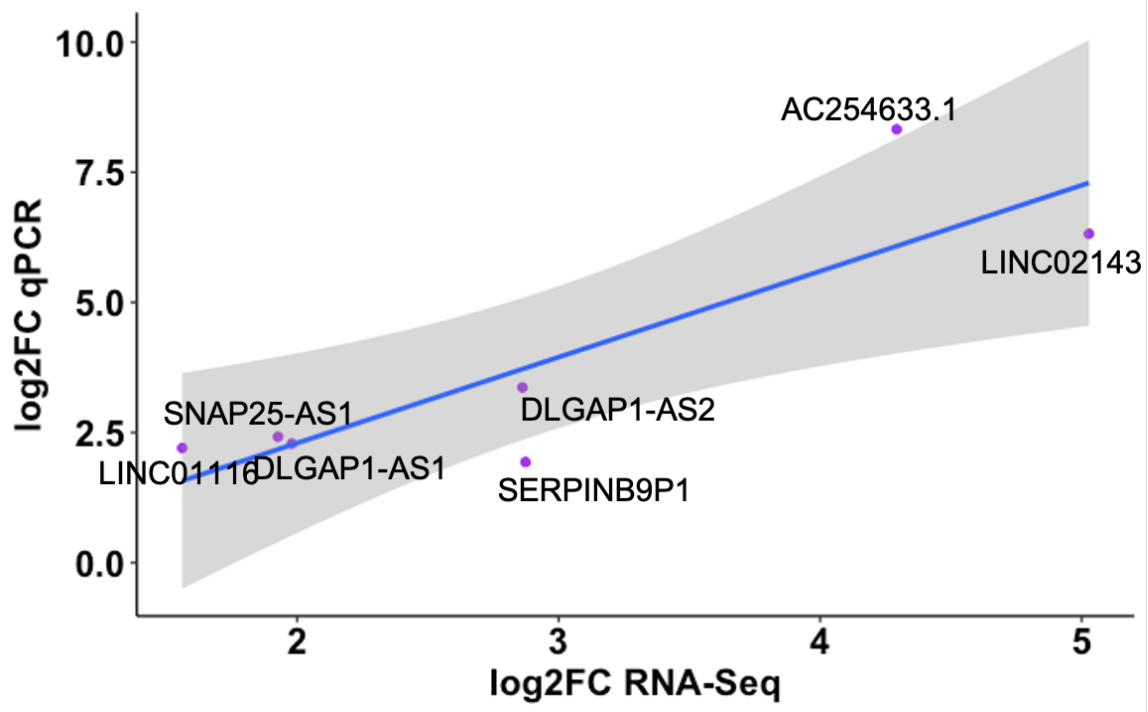


Figure 4.20: Log_2fold change of selected upregulated lncRNAs upon differentiation in RNA-Seq correlates with their log_2fold change in RT-qPCR. Log_2fold change of LINC01116, LINC02143, SNAP25-AS, DLGAP1-AS1 and DLGAP1-AS2 in RNA-Seq (x axis) highly correlates with their respective log_2fold change in RT-qPCR (y axis). (Pearson correlation coefficient=0.85).

4.6 Majority of selected candidate lncRNAs are enriched in cytoplasm

The majority of lncRNAs are thought to be enriched in the nucleus compared to the cytoplasm, and therefore possess nuclear function (Derrien et al., 2012; Djebali et al., 2012). Here, I am interested in the potential for lncRNA with cytoplasmic function. To select candidate lncRNAs for further analysis it was therefore necessary to select those which are enriched in the cytoplasm. It is important to determine whether these lncRNAs detected in the cytoplasm are just there at low levels, and mainly localise in the nucleus or they are cytoplasmic lncRNAs. Therefore, candidate lncRNAs were screened for their propensity for cytoplasmic enrichment. To determine the sub-cellular localisation of these lncRNAs, cytoplasmic/nuclear fractionation of Control and RA treated SH-SY5Y cells in 3 independent biological replicates was performed (Methods 2.2). RNA and protein were extracted from whole cell, cytoplasmic and nuclear fractions. Quality of the fractionation was assessed by western blot, probing for cytoplasmic (β-tubulin) and nuclear (NXF1, H3K27me3) markers (Figure 4.21). This revealed that cytoplasmic and nuclear fractions were well fractionated and generally clear of contamination.

To assess the distribution of lncRNA targets in whole cell, cytoplasmic and nuclear fractions I performed RT-qPCR for the candidate lncRNAs, whose upregulation were previously validated (Section 4.5). The lncRNAs were selected based on their log_2fold change upon differentiation in both Total and Polysome RNA-Seq datasets (Table 4.5). RT-qPCR was performed for LINC01116, LINC02143, DLGAP1-AS1, DLGAP1-AS2, SERPINB9P1, SNAP25-AS1 and AC254633.1 lncRNA transcripts. GAPDH mRNA was used as a control RNA, the majority of which will be present in the cytoplasm. XIST lncRNA was used as a control, which will be enriched in the nucleus. These RT-qPCRs were performed by Sophie Clayton, undergraduate student, whom I supervised. The ratio of cytoplasmic/nuclear RNA levels, or cytoplasmic enrichment was calculated for each lncRNA. As expected XIST was enriched in the nucleus and GAPDH in the cytoplasm (Figure 4.24). 6/7 candidate lncRNAs were enriched in the cytoplasm. DLGAP1-AS2, was the only candidate not enriched in the cytoplasm (Figure 4.22) because its cytoplasm/nucleus ratio is slightly above 1 (log_2(Cyto/Nuc) = 0.67 for Control and 1 for RA), indicating a rather homogeneous distribution between the nucleus and cytoplasm.

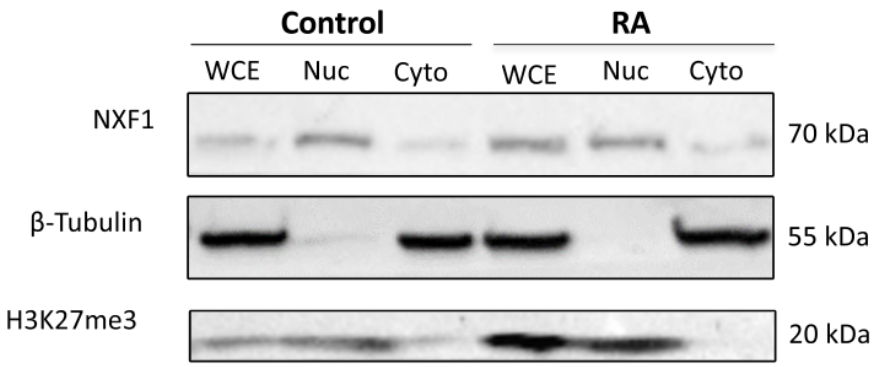


Figure 4.21: Sub-cellular fractionation of SH-SY5Y cells. Western blot for detection of cytoplasmic (β-Tubulin, 1:5000 antibody dilution) and nuclear (NXF1, 1:5000 antibody dilution and H3K27me3, 1:1000 antibody dilution) markers in whole cell extract (WCE), corresponding to ~0.1 $x10^6$ cells, nuclear (Nuc) (~0.3 $x10^6$ cells) and cytoplasmic (Cyto) samples (~0.3 $x10^6$ cells), from Control and differentiated (RA) cells.

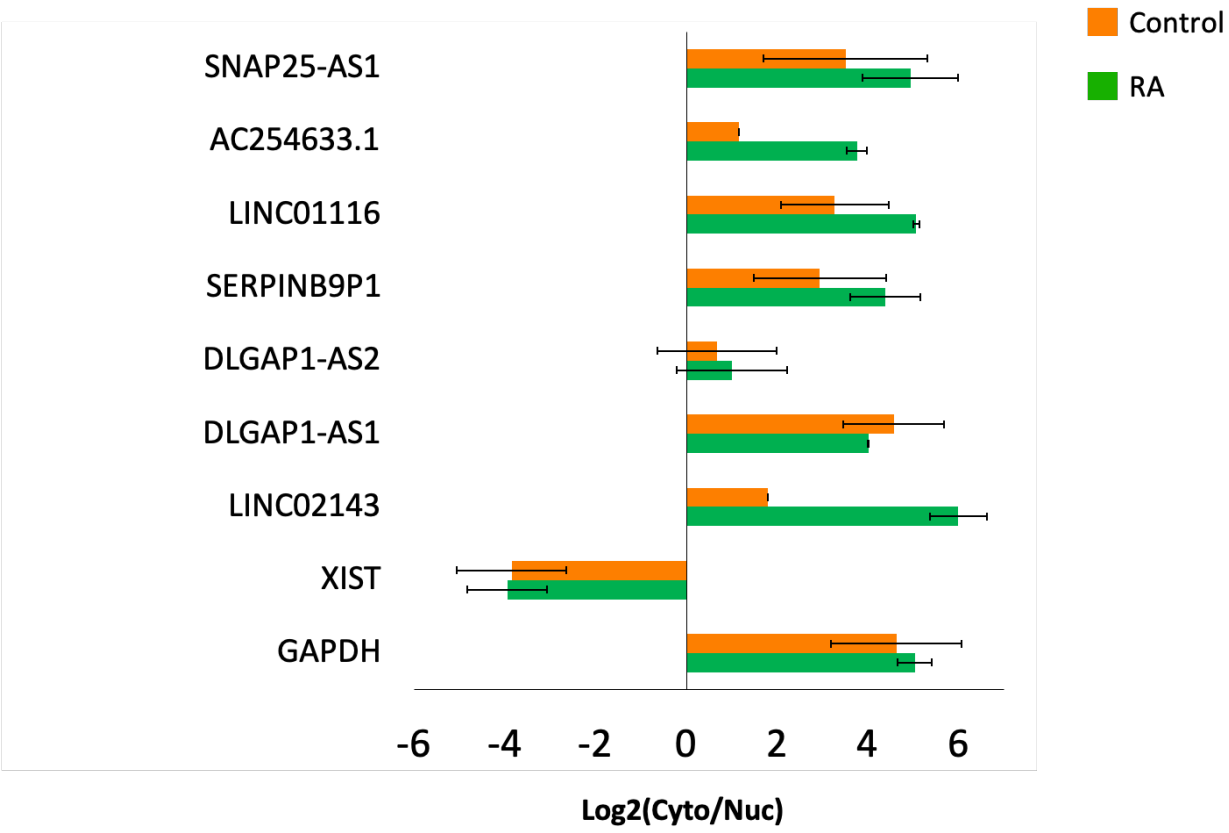


Figure 4.22: Majority of lncRNAs detected in Poly-Ribo-Seq are enriched in the cytoplasm: Log_2(Cyto/Nuclear) ratio of transcript levels indicates whether lncRNAs are enriched in cytoplasm (values 1 upwards) or nuclear (-1 and below) or equally distributed evenly (between 1 and -1). SNAP25-AS1, AC254633.1, LINC01116, SERPINB9P1, DLGAP-AS1 and LINC02143 transcripts are enriched in the cytoplasm, both in Control and upon differentiation. DLGAP1-AS2 Log_2(Cyto/Nuclear) ratio is below 1 therefore, not enriched in the cytoplasm. (n=3, SE plotted). RT-qPCR was performed in samples from 3 biological replicates, in triplicates for each sample. Error bars indicate standard error. (experiment performed by Sophie Clayton, undergrad student, Aspden lab) supervised by me.

4.7 *In vitro* validation of lncRNA interaction with polysome complexes

To understand the polysome association of the candidate lncRNAs identified to be enriched in the cytoplasm, their distribution across sucrose gradients was assessed. I performed RT-qPCR on gradient fractions, collected from Control and RA-treated cells as previously described (Section 4.2.2). These upregulated lncRNAs upon neuronal differentiation, were selected based on their log_2fold change (LINC02143), the extent of their cytoplasmic enrichment (LINC01116, LINC02143, DLGAP1-AS1) and the existing literature associated with them (LINC01116). RT-qPCR were performed by Abigail Byford and Sophie Clayton, both undergraduate students, whom I supervised. To assess the distribution of these lncRNA transcripts across the sucrose gradient and monitor any changes, upon differentiation (as shown by the RNA-Seq data), the quantity of each lncRNA transcript detected in each fraction was calculated and plotted across the gradient (Figure 4.23).

LINC02143 (Figure 4.14) is the lncRNA with the highest log_2fold change upon differentiation by Poly-Ribo-Seq (log_2fold change=5.02 in total, log_2fold change= 4.43 in polysome). It is an intergenic lncRNA with no known function associated to it so far. Another polysome-associated intergenic lncRNA that is upregulated during differentiation, is LINC01116 (or TALNEC2) with log_2fold change=1.5 in total and log2fold change=2 in polysome. This lincRNA has recently been reported to be involved in the progression of glioblastoma (GBM) (Brodie et al., 2017)and overexpression of LINC01116 promoted invasion and migration of gastric cancer cells (Su et al., 2019). This suggests a potential role of LINC01116 in the formation of cell membrane protrusions. Given that the GO terms cell adhesion and cell migration were enriched upon differentiation, I considered LINC01116 as an interesting target to further investigate.

SNAP25-AS1 and DLGAP1-AS1 (Figure 4.14) are upregulated during differentiation and polysome-associated according to Poly-Ribo-Seq. Both transcripts are anti-sense in relation to neuronal genes. Synaptosome associated protein 25 (SNAP25) and Disks large associated protein-1 (DLGAP1) are both synaptic proteins. In fact, SNAP25 was found to be significantly upregulated upon RA treatment, with a log_2fold change of ~0.7, in both total cytoplasmic and polysome-associated RNA populations,

which may suggest that the sense and anti-sense transcripts are co-regulated. DLGAP1 expression was not significantly changed upon differentiation. One might argue that the chromosomal location of these lncRNAs in relation to synaptic genes is not meaningful in this case, because the experiment was conducted on cytoplasmic lysates. However, the fact that these lncRNAs associate with the polysome complexes might be an important indication that they are enriched in the cytoplasm and potentially functional.

When RT-qPCR was performed, LINC01116 is detected in high levels in the 80S (monosome) fraction and in small polysome complexes (Figure 4.23: A). A sharp increase in the amount of LINC01116 in disomes is detected upon differentiation, consistent with the upregulation in the polysomes detected by RNA-Seq, indicating a functional interaction of LINC01116 with polysomes during differentiation. Notably, LINC01116 transcript is mainly detected in polysome peaks consisting of less than 3 ribosomes. This suggests that LINC01116 potentially contains a smORF that is being decoded by 1-3 ribosomes. On average, 66% of the transcript was detected in the polysome fractions in Control and 57% upon differentiation. This difference, however, was not statistically significant.

LINC02143 interacts with actively translating polysomes during neuronal differentiation (Figure 4.23: B). It is clear from RT-qPCR of gradient fractions that LINC02143 is induced upon neuronal differentiation, as its expression is almost undetectable in undifferentiated cells, as indicated by Poly-Ribo-Seq. In differentiated cells, 5% of the transcript was detected in 80S (monosome) fraction and 41% in small polysome complexes. The fact that LINC02143 is mainly detected in the disome fraction and at low levels in larger polysomes, suggests that it interacts with (or it is engaged by) disomes, potentially being decoded.

DLGAP1-AS1 is upregulated during differentiation and it is detected in high levels in the 80S (monosome) fraction and in small polysome complexes as shown by RT-qPCR data (Figure 4.23-C). DLGAP1-AS1 level in 80S is higher in differentiated cells, compared to Control. On average, 63% of the transcript was detected in the polysome fractions in Control and 49% upon differentiation. This confirms DLGAP1-AS1 interacts with actively translated polysomes. Taken together, the above data confirm the interaction between

LINC01116, LINC02143 and DLGAP1-AS1 and 80S and small polysomes, in both Control and differentiated cells. Engagement by monosomes and small polysomes suggests potential decoding of smORFs by small polysome complexes.

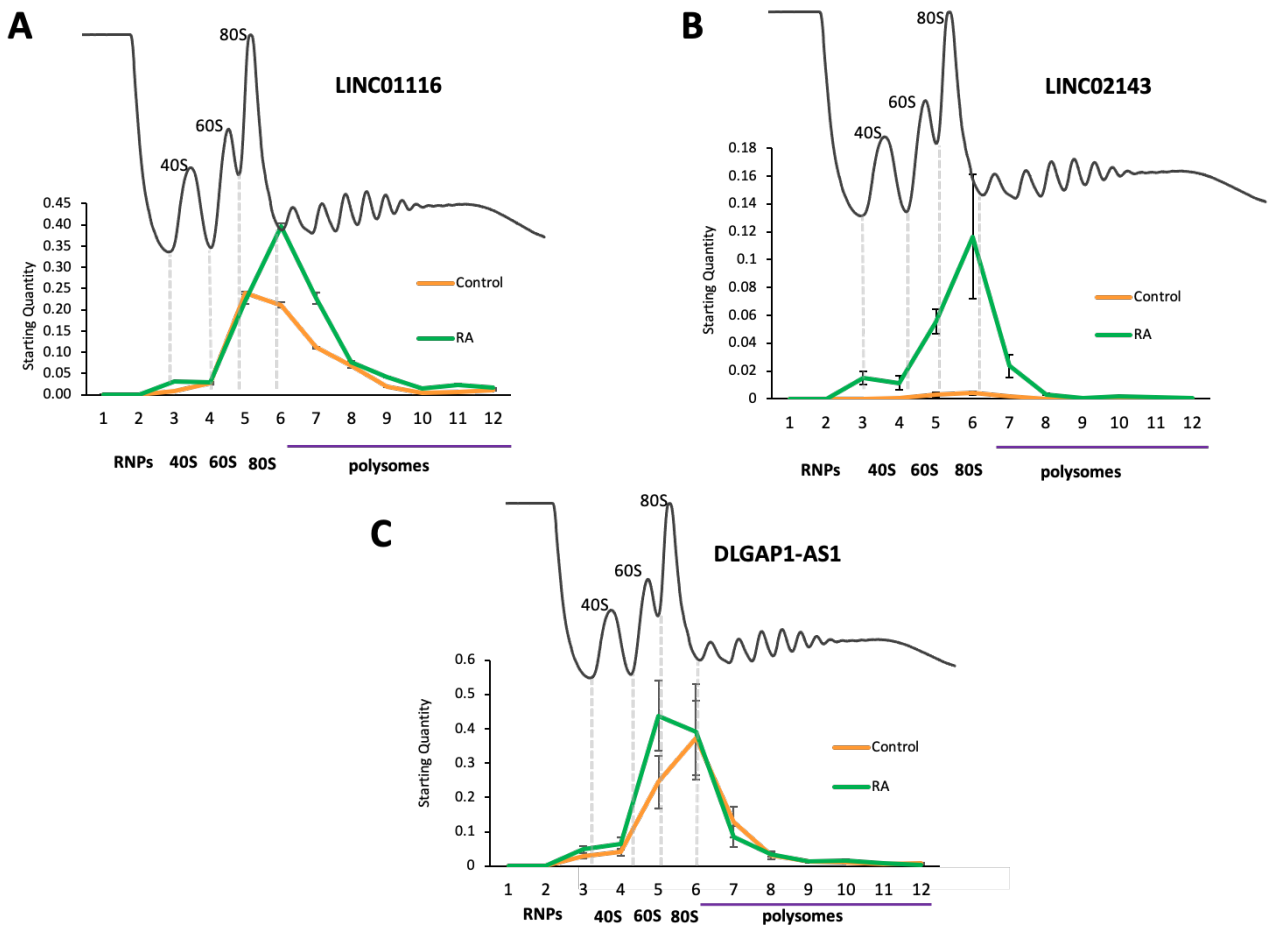


Figure 4.23: Cytoplasmic lncRNAs upregulated upon differentiation interact with actively translating polysomes. RT-qPCR of lncRNAs on sucrose gradient fractions of Control and differentiated (RA) cells. A polysome profile of one of the replicates of Control SH-SY5Y cells is shown for clarity. RT-qPCR was performed on fractions from 3 gradients (n=3) for each condition in technical triplicates for each fraction. Levels of lncRNAs are calculated by absolute quantification RT-qPCR, using a standard curve derived from serial dilutions of a pool of all the fractions from both conditions (n=3, SE plotted corresponds to different gradient replicates). (A) LINC01116 is detected in high levels in the 80S and small polysomes (66% in Control and 57% in RA, on average) and its levels in the disome sharply increase upon differentiation. (B) LINC02143 is almost undetectable in Control but is induced upon differentiation and is detected in high levels in the 80S and in small polysomes (41% in RA). (C) DLGAP1-AS1 is detected in high levels in the 80S and in small polysomes (63% in Control and 49% in RA, on average). (RT-qPCRs performed by Abigail Byford and Sophie Clayton, undergrad students), whom I supervised.

4.8 Assessment of Ribo-Seq data

In order to use the Ribo-Seq data to assess the coding potential of lncRNAs, it was necessary to evaluate the quality of ribosome profiling. As shown previously (Figure 4.4), ~93% of the usable reads of footprint samples (Control and RA) mapped to coding sequences, which is the expected pattern in ribosome profiling datasets. To prove that the Ribo-Seq reads are indeed generated from actively elongating ribosomes, triplet periodicity analysis was performed on the transcriptome mapped reads, using RiboSeqR (Hardcastle, 2014) (Methods 2.18). Ribosome footprinting produces reads of ~30nt, so the triplet periodicity of reads between 25-35nt was examined (Appendix-III). For each read length assessed, the number of reads that were assigned to each frame was calculated. The majority of the transcriptome-mapped reads were 31nt and 33nt long and predominantly mapped to frames 1 and 2 respectively (Figures 4.24: A, B) indicating precise footprinting. Therefore, Ribo-Seq reads correspond to actively elongating ribosomes. The total number of 31nt long reads (from all 3 replicates) in frame is very similar to the total number of 33nt long reads in frame. To assess the global distribution of the reads across the transcriptome, a metagene analysis was performed on the reads that exhibited the best triplet periodicity (31 and 33nt long reads) (Figure 4.14: C and D). The vast majority of these reads map to coding sequences and there is also some signal on the 5' and 3'-UTRs, consistent with previous evidence of uORF and dORF translation (Heyer and Moore, 2016; Rodriguez et al., 2019; Chen et al., 2020).

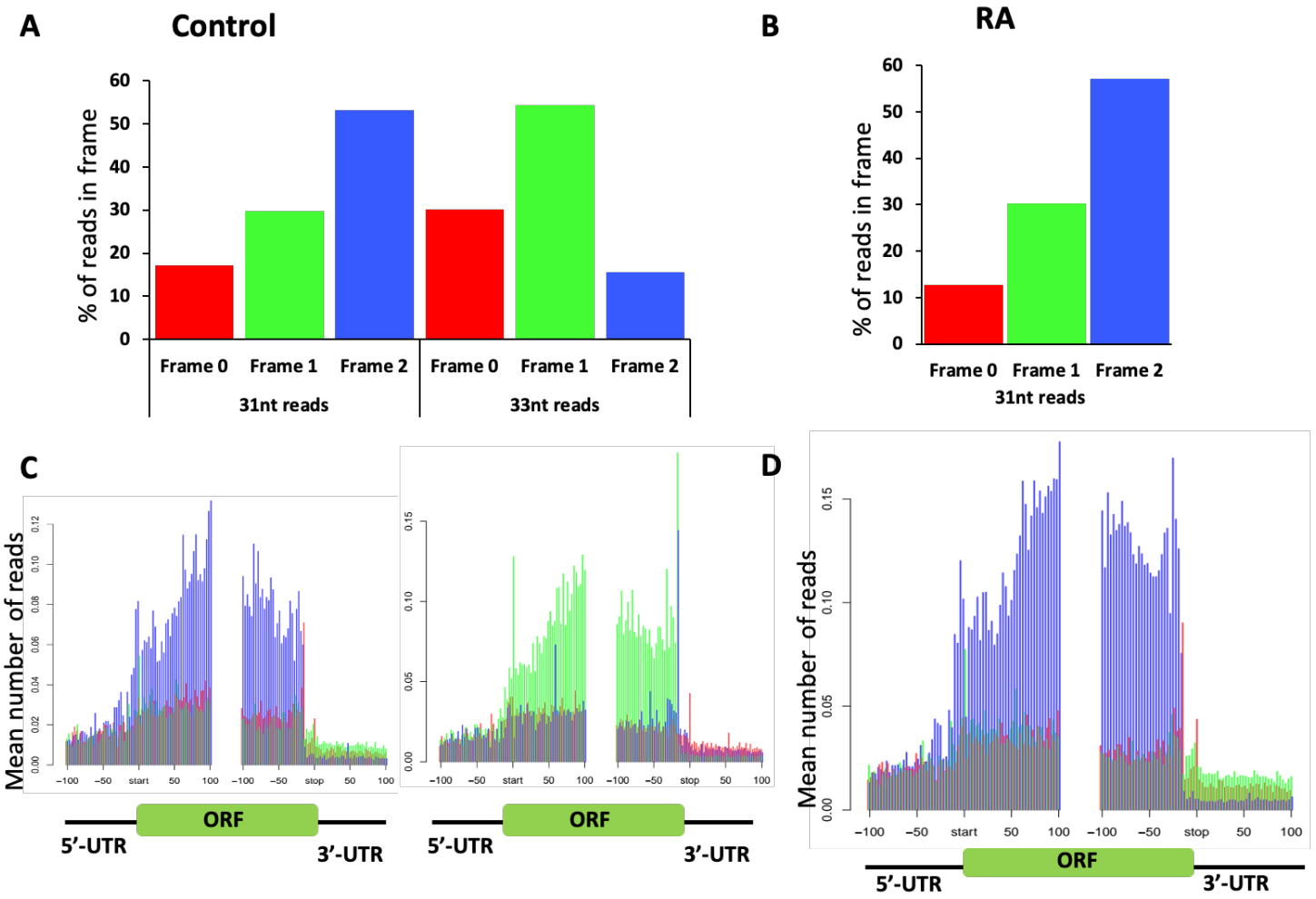


Figure 4.24: Triplet periodicity and metagene analysis for Control and RA sample, replicate 3. (A and B) Triplet periodicity plots show the percentage of reads assigned to each frame (frames are colour-coded) for 31 and 33nt reads for (A) Control and for 33nt reads for (B) RA. (C and D) Metagene analysis of footprint reads of the lengths that showed the strongest periodicity for (C) Control

and (D) RA samples. Reads are globally mapped across transcripts and assigned to 5’-UTR, ORF and 3’-UTR regions (100nt upstream and downstream of start and stop codons plotted). Most of the reads map to the coding sequence (ORF) and fewer reads map to 5’ and 3’-UTRs.

To assess translation in undifferentiated and differentiated cells, I examined the level of translation of specific protein-coding transcripts, as a positive controls, using the RiboSeqR pipeline (Hardcastle*, 2014*) (Methods 2.18). One of them was neuronal specific β-III tubulin (TUBB3 or Tuj1), which is expected to be translated in both Control and differentiated SH-SY5Y (RA). Notably, β-III tubulin transcription is significantly downregulated upon differentiation in total cytoplasm (log_2fold change=-1) but is not depleted from polysomes, consistent with it being translated in both Control and differentiated cells (Figure 4.25).

To identify which ORFs are translated and their translational efficiencies (the level of protein production from mRNA (Hernandez-Alias et al., 2020) the RiboTaper was employed by Isabel Birds (PhD student, Aspden lab) (Calviello et al., 2016) (Methods 2.18). The 31nt and 33nt long reads were used in this pipeline, as they displayed the clearest triplet periodicity in all the 3 replicates (Figure 4.24 and Appendix-III).

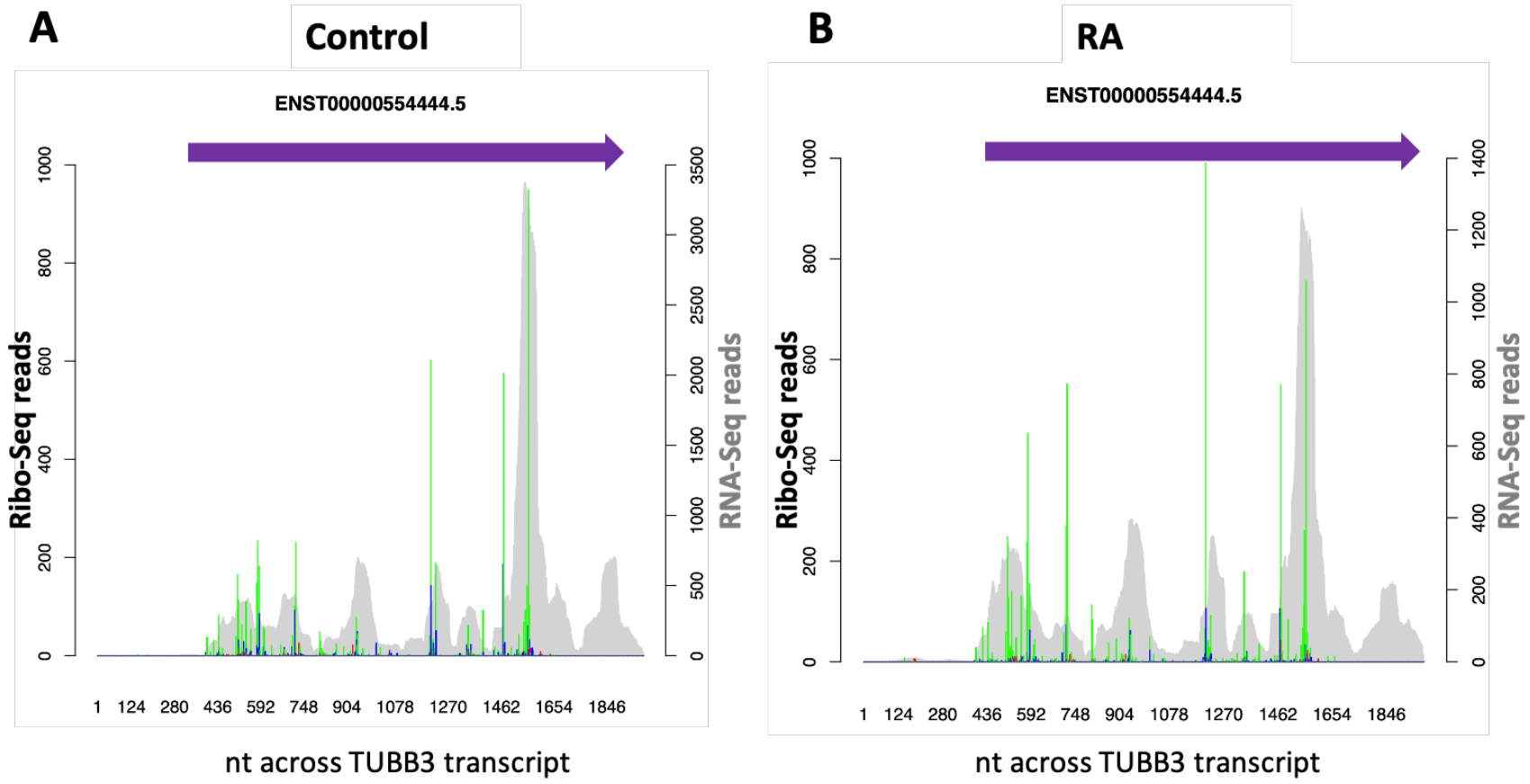


Figure 4.25: β-III tubulin is translated in similar level in both Control and differentiated cells. Transcript plots from Poly-Ribo-Seq replicate 3, showing the number of Ribo-Seq reads (footprints) colour-coded per frame on the left y axis and the number of RNA-Seq reads on the right y axis (grey), across the β-III tubulin (TUBB3 or Tuj1) transcript for (A) Control and (B) RA differentiated cells.

4.9 Actively translated smORFs within lncRNAs are identified by Poly-Ribo-Seq.

Analysis of Total and Polysome RNA-Seq (Section 4.4) revealed that cytoplasmic lncRNAs dynamically associate with the translation machinery during differentiation. Based on previous studies (Carrieri et al., 2012; Dimartino et al., 2018; Chen et al., 2020), these interactions are either the result of the translational regulation of an mRNA by a lncRNA, or active translation of a smORF within the lncRNA.

To determine if lncRNAs are translated, I examined the triplet periodicity of the Ribo-Seq reads that mapped to lncRNAs, using the RiboSeqR pipeline (Hardcastle 2014) as previously described (section 4.7). As with protein-coding genes, the majority of the Ribo-Seq reads, which map to lncRNAs are 31-33nt long and display triplet periodicity (Figure 4.26 and appendix for graphs of all replicates). This pattern of periodicity was highly similar to that seen for the protein-coding transcripts suggesting that these reads represent genuine translation events on some lncRNAs. In particular, 33nt length reads showed the strongest framing in both Control and RA conditions (Figure 4.26).

To identify which small ORFs were translated on lncRNAs, further analysis was performed by Isabel Birds, using RiboTaper. Specifically, 28 actively translated smORFs were identified in lncRNA transcripts in Control cells and 23 in RA treated cells, 7 of them are in common between the two conditions (Table 4.6). The stringent cut-offs that were used detected only the smORFs that exhibited strong periodicity and were called as translated in at least 2 out of 3 replicates to be considered translated lncRNA-smORFs. Therefore, it is not surprising that such a small number of lncRNA smORFs are detected as translated, compared to the number of translated protein-coding genes. To determine the level of translation from smORFs, their translation efficiency (TE) was assessed (analysis conducted by Isabel Birds). The translated smORFs exhibited similar TEs to that of protein-coding genes.

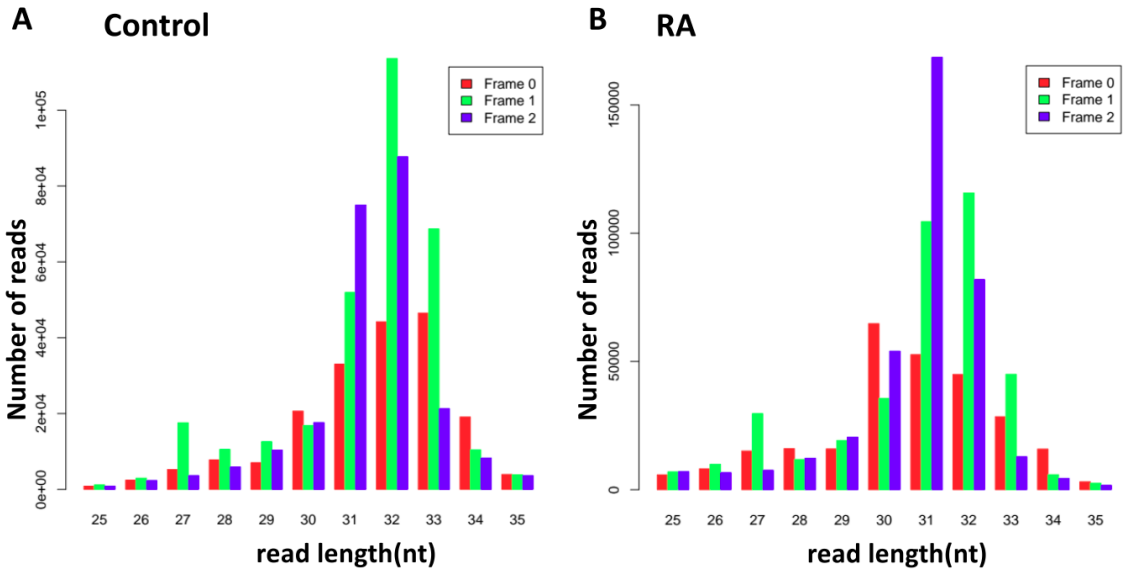


Figure 4.26: Ribo-Seq reads from biological replicate 3 mapping to lncRNAs display triplet periodicity. Framing was calculated for a read length range from 25-35nt and the majority of the reads were 31nt to 33nt long (A) in Control samples and 31-32nt (B) in RA. In control 33nt long reads exhibit highest triplet periodicity and 31nt in RA.

Table 4.6: Translated smORFs detected by Poly-Ribo-Seq

ORF ID	ORF length (excluding stop codon)	Transcript name	Detected in Control	Detected in RA	smORF annotation status
ENST00000608678.5_120_171	51	LINC01515	✓		novel
ENST00000526036.1_1226_2042	816	AP001372.2-201	✓		novel
ENST00000456602.5_704_800	96	RBM26-AS1	✓		novel
ENST00000590750.1_25_283	258	AC020928.1	✓		novel
ENST00000445520.5_206_341	135	MAP4K3-DT	✓		novel
ENST00000295549.8_609_822	213	LINC01116	✓		novel
ENST00000453910.5_151_262	111	LINC00478	✓	✓	novel
ENST00000424349.1_116_296	180	FGD5-AS1	✓	✓	novel
ENST00000504792.6_74_230	156	THAP9-AS1	✓		novel
ENST00000504869.1_132_360	228	THAP9-AS1	✓		novel
ENST00000507322.5_6_189	183	HAND2-AS1	✓	✓	novel
ENST00000515376.5_746_851	105	HAND2-AS1	✓	✓	novel
ENST00000437621.6_279_426	147	PSMG3-AS1	✓		novel
ENST00000518942.1_154_277	123	AC064807.1	✓		novel
ENST00000625445.1_752_881	129	EBLN3P	✓		novel
ENST00000435378.5_73_145	72	AL392172.1	✓		novel
ENST00000442171.5_25_97	72	AL392172.1	✓		novel
ENST00000449419.5_342_636	294	ENTPD1-AS1	✓		novel
ENST00000641571.1_548_773	225	OLMALINC	✓		novel
ENST00000454935.1_477_633	156	OLMALINC	✓		novel
ENST00000603633.2_174_258	84	LINC00221	✓		novel
ENST00000592201.1_41_158	117	AC027097.1	✓		novel
ENST00000597230.2_325_523	198	AC020915.3	✓		novel

ENST00000566220.2_141_570	429	TUG1	✓		novel
ENST00000502941.5_6_222	216	HAND2-AS1	✓	✓	novel
ENST00000510221.5_6_285	279	HAND2-AS1	✓	✓	novel
ENST00000645869.1_2_212	210	HAND2-AS1	✓	✓	novel
ENST00000651163.1_67_343	276	AC244102.3	✓		novel
ENST00000609803.2_330_426	96	AC008124.1		✓	novel
ENST00000441257.1_75_243	168	ZEB1-AS1		✓	novel
ENST00000429940.6_97_595	498	LINC00839		✓	novel
ENST00000547492.1_524_716	192	GIHCG		✓	novel
ENST00000557660.5_42_186	144	PSMA3-AS1		✓	novel
ENST00000551597.6_54_252	198	PSMA3-AS1		✓	novel
ENST00000554309.1_48_240	192	PSMA3-AS1		✓	novel
ENST00000501177.7_136_388	252	CRNDE		✓	Szafron et al., 2015
ENST00000651711.1_64_313	249	AC068616.1		✓	novel
ENST00000602414.5_453_561	108	SNHG8		✓	novel
ENST00000499459.2_92_287	195	AC008966.1		✓	novel
ENST00000545177.4_162_333	171	SERPINB9P1		✓	novel
ENST00000440088.5_140_317	177	APTR		✓	novel
ENST00000419422.1_379_634	255	EMSLR		✓	novel
ENST00000525186.1_60_129	69	MCPH1-AS1		✓	novel
ENST00000652695.1_282_585	303	PVT1		✓	novel
ENST00000442428.1_220_352	132	AL162386.2		✓	novel

Four of the translated lncRNAs identified by Poly-Ribo-Seq have been previously reported to be translated in different tissues. The translated smORFs reported for three of them (TUG1, MCPH1-AS and HAND2-AS1) are different from the smORFs identified by RiboTaper pipeline. TUG1 lncRNAs (translated in Control) is conserved between human and mouse (77% exonic nucleotide conservation level) and encodes an 143aa evolutionarily conserved peptide that regulates mitochondrial membrane potential (Lewandowski et al., 2019). HAND2-AS1 (translated in Control and RA) is translated in human and rodent heart and encodes an integral membrane component of the endoplasmic reticulum (van Heesch et al., 2019). A 9aa peptide was recently detected to be encoded from MCPH1-AS1 (translated in differentiated cells) in a proteogenomic screening for tumour-specific non-canonical HLA binding peptides (Chong et al., 2020). CRNDE, which is translated only upon differentiation, encodes a peptide (CRNDEP) which was previously characterised as nuclear (Szafron et al., 2015).

To my knowledge, the rest of the translated smORFs identified by Poly-Ribo-Seq have not yet been described as translated and therefore, can be considered as novel translated smORFs and some of them will be selected for *in vitro* validation of their translation in the following chapter.

The length of the peptides encoded by those smORFs varies between 17 and 101aa, with the exception of 3 peptides (2 detected in Control and 1 in RA) that exceed 140aa. (Figure 4.27). To explore if there was any specific characteristic in common between the translated smORFs, I examined the types of lncRNAs that they belong to as well as the length of the putative peptides they encode. The majority of the smORFs belong to lncRNAs that fall into 4 categories: long intergenic (lincRNAs), antisense lncRNAs, sense-intronic and bi-directional promoter lncRNAs (Figure 4.28).

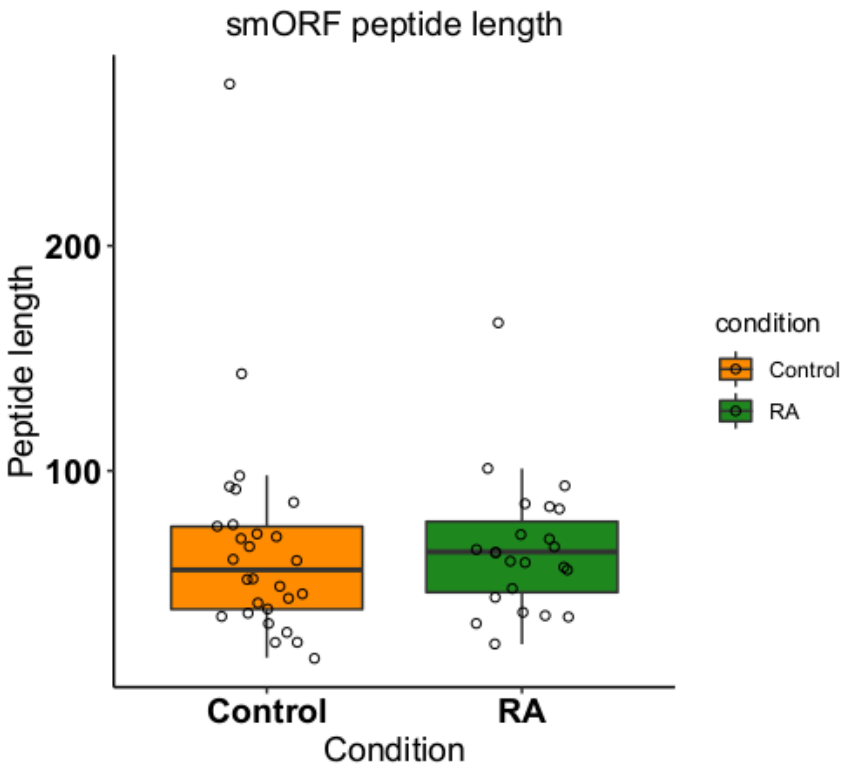


Figure 4.27: LncRNA smORFs in Control and differentiated encode peptides of various lengths. Both in Control and in differentiated cells (RA). Median length is 56aa for Control peptides and 64aa for RA peptides, with 3 exceptions that exceed 140aa in length.

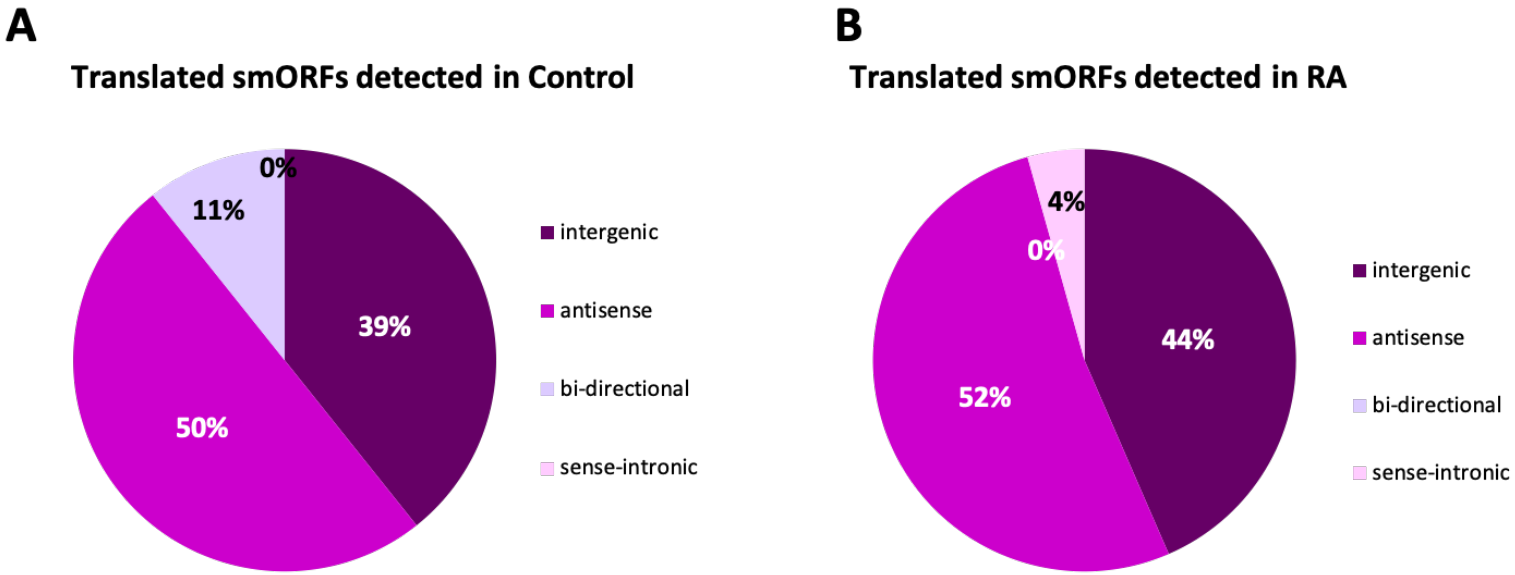


Figure 4.28: The majority of translated smORFs belong to intergenic and antisense lncRNAs. (A) In Control, half of the translated smORFs belong to anti-sense lncRNAs and ~40% belong to intergenic lncRNAs. 3 out of 28 actively translated smORFs belong to bi-directional transcripts. (B) In differentiated cells (RA) 52% of the translated smORFs belong to anti-sense lncRNAs and 44% belong to intergenic lncRNAs. Only 1 out of 23 smORFs belongs to a sense-intronic lncRNA transcript.

4.10 Identification of novel smORF peptide translation event in LINC01116 and LINC00478

One of the lncRNAs, which contains an ORF detected as translated is ENST00000295549.8 or LINC01116 (Figure 4.27). LINC0116 levels were found to be significantly increased during differentiation and it interacts with polysomes (Section 4.5). LINC01116 contains a smORF of 213nt, which exhibits strong triplet periodicity in our Ribo-Seq data and passes our detection threshold for translation in both Control and differentiated cells (Figure 4.29-courtesy of Isabel Birds). The identification of this translated smORF within LINC01116 is consistent with its presence in the disome fraction (Figure 4.25) because based on its length it can be decoded by up to 5-6 ribosomes. LINC00478 transcript, which has previously been reported as tumour suppressor in breast cancer (Gökmen-Polar et al., 2016) was also found to contain an actively translated smORF of 111nt, with strong framing in Control and differentiated cells (Figure 4.30-courtesy of Isabel Birds).

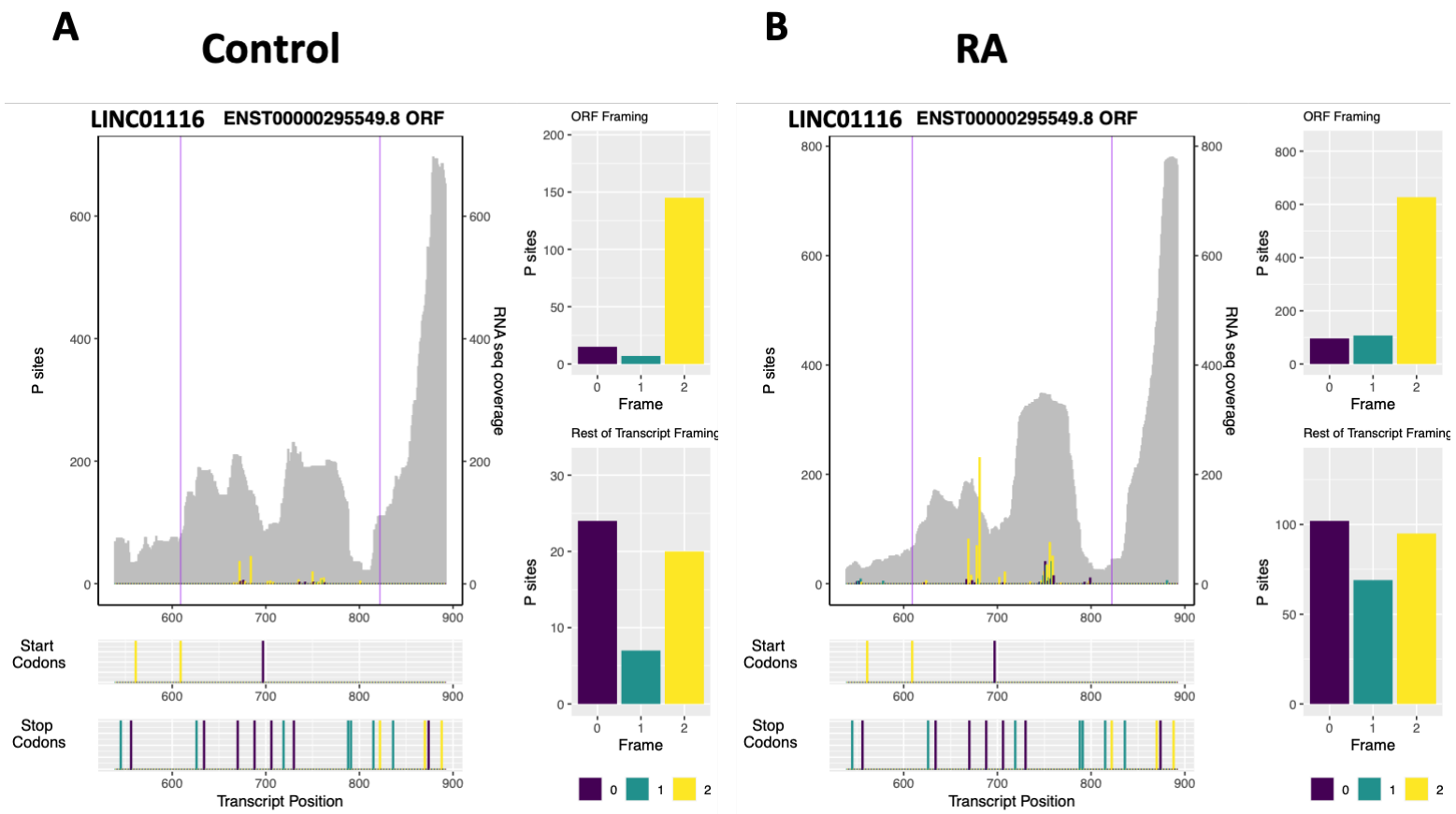


Figure 4.29: LINC01116 contains a 216nt smORF that is actively translated in Control and upon differentiation. Plots show the Ribo-Seq P sites (in all 3 frames) and RNA-Seq coverage across LINC01116 transcript (ORF is indicated by purple vertical lines) in (A) Control and (B) RA treated cells. Positions of start and stop codons in the 3 different frames are marked below the plots. 75-80% of reads in ORF are in specific frame (right upper panel), whilst outside the ORF distribution between 3 frames is far more even (rest of transcript framing, calculated by the number of P-sites outside the smORF) (right lower panel). Figure generated by Isabel Birds, but designed in collaboration.

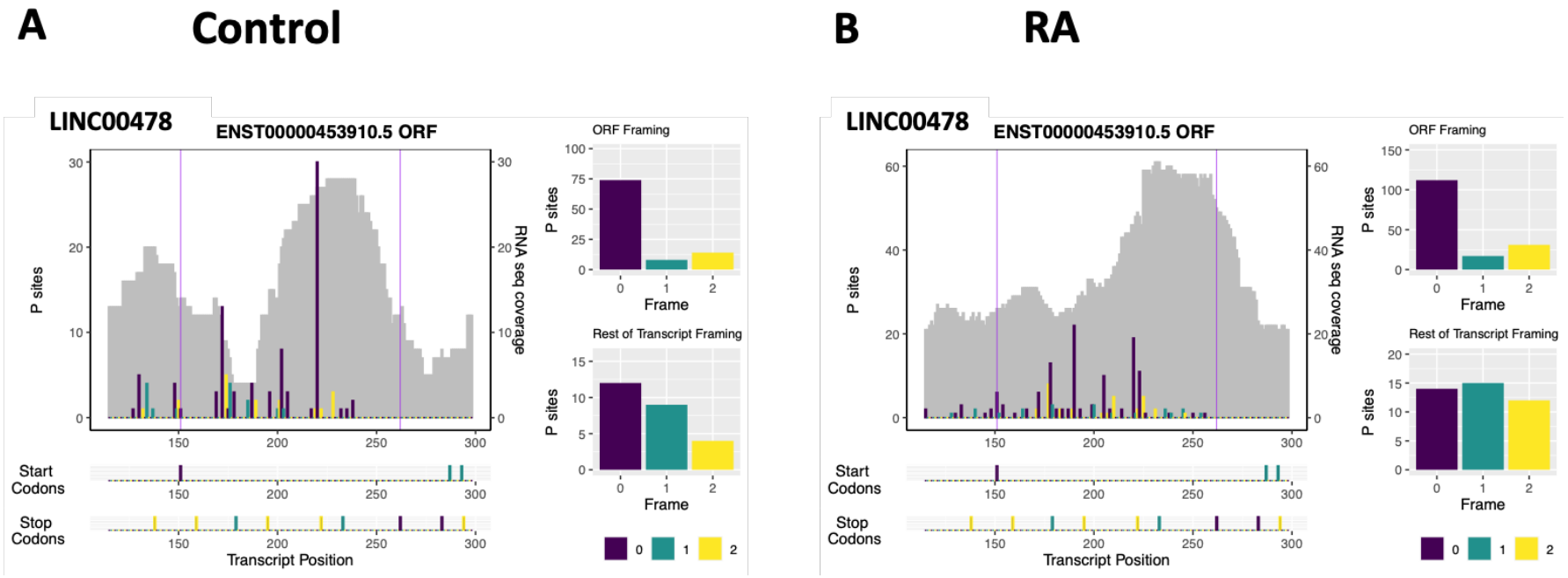


Figure 4.30: LINC00478 contains a 113nt smORF that is actively translated in Control and differentiated SH-SY5Y cells. Plots show the Ribo-Seq P-sites (in all 3 frames) and RNA-Seq coverage across LINC00478 transcript (ORF is indicated by purple vertical lines) in (A) Control and (B) RA treated cells. Positions of start and stop codons in the 3 different frames are marked below the plots. 75-80% of reads in ORF are in specific frame (right upper panel), whilst outside the ORF distribution between 3 frames is far more even (rest of transcript framing, calculated by the number of P-sites outside the smORF) (right lower panel). Figure generated by Isabel Birds, but designed in collaboration.

4.11 Discussion

4.11.1 Expression levels of translation machinery components is downregulated upon neuronal differentiation

This chapter focused on the analysis of the Poly-Ribo-Seq data of Control and differentiated cells. Differential expression analysis of protein-coding transcripts identified large-scale transcriptional changes consistent with neuronal differentiation. Upon differentiation, ribosomal protein (RP) mRNAs were significantly depleted from actively translating polysomes. This was revealed by comparing levels of RP mRNAs in total cytoplasm RNA-Seq to polysome-associated mRNAs, in Control and differentiated cells. RT-qPCR across sucrose gradients confirmed that upon differentiation levels of ribosomal protein transcripts in the polysomes significantly shifted from polysomes to subunits and RNPs, while their levels in RNPs and 40S and 60S subunits significantly increased. All this is consistent with the reduction in the level of active translation, induced upon neuronal differentiation (Chapter 3).

4.11.2 LncRNAs dynamically interact with polysome complexes upon neuronal differentiation

Differential expression analysis revealed a substantial number of lncRNAs differentially expressed upon neuronal differentiation, both in total cytoplasm and in Polysomal populations. Moreover, most of the differentially expressed lncRNA transcripts in the Total cytoplasm were also differentially expressed in the polysomes upon differentiation. The association of these lncRNAs with polysomes suggests that they are either being translated or regulating the translation of the mRNAs with which they interact. The differentially expressed lncRNAs upon differentiation were mostly intergenic or anti-sense transcripts, with very few sense-intronic and sense-overlapping transcripts. Interestingly, certain lncRNAs changed their polysome association upon differentiation, indicating that the association of lncRNAs with the translation machinery is dynamic. 10 lncRNAs were depleted from polysomes in Control but not in RA treated cells, indicating that their interaction with the translation machinery may be suppressed, until the induction of differentiation. Moreover, 47 lncRNAs were moderately but significantly enriched in polysome complexes upon differentiation. This suggests a potential functional role for those interactions of

lncRNAs with actively translating polysomes during neuronal differentiation, for example in regulating the translation level of the mRNAs they are interacting with, as is the case with BC200, BACE1-AS, UCHL1-AS and lnc-31 lncRNAs (Wang et al., 2002; Faghihi et al., 2010; Carrieri et al., 2012; Dimartino et al., 2018) or their translational regulation during differentiation. Further study is required to characterise the functions of those polysome-associated lncRNAs.

4.11.3 LncRNAs of both undifferentiated and differentiated cells contain actively translated smORFs

Triplet periodicity analysis on lncRNA transcripts revealed good framing, suggestive of potential translation events. RiboTaper analysis (performed by Isabel Birds) identified 28 smORFs in Control and 23 smORFs in RA treated cells, exhibiting strong framing. Importantly, translation efficiency (TE) calculation (Isabel Birds) shows that these lncRNA-smORFs are translated with a TE similar to that of protein-coding genes, adding to the evidence of genuine translation of these smORFs. Among the translated smORFs identified by Poly-Ribo-Seq is a previously characterised smORF that encodes a small nuclear peptide called CRNDEP previously characterised by (Szafron et al., 2015). The translated smORFs mostly belong to intergenic and anti-sense lncRNA transcripts, which is expected as these are the most abundant categories of lncRNAs. The peptides encoded by those smORFs have a median length of 56aa in Control and 64aa in RA, with the exception of 3 peptides, longer than 140aa. Those are encoded by: the bi-directional transcripts AP001372.2 and TUG1 in Control and the lincRNA LINC00839 in differentiated cells. TUG1 (Taurine Upregulated gene 1) lncRNA locus is highly conserved, associated with male fertility and encodes an evolutionarily conserved peptide that when overexpressed, affects mitochondrial membrane potential (Lewandowski et al., 2019).

4.11.4 Poly-Ribo-Seq identified promising lncRNA candidates for further study

Following the analysis of the large Poly-Ribo-Seq datasets, I wanted to focus on specific lncRNAs and further investigate their function. The cytoplasmic enrichment of lncRNAs as well as their interaction with the actively translating polysomes was experimentally validated. 6/7 lncRNAs tested were shown to be enriched in the cytoplasm. The 1/7 is DLGAP1-AS2, which exhibits a Cytoplasmic/Nuclear ratio of ~1, indicating its equal distribution between cytoplasm and nucleus. The interaction of lncRNAs with polysomes was validated for LINC01116, DLGAP1-AS1 and LINC02143 transcripts by RT-qPCR across sucrose gradients. LINC02143 is only induced upon neuronal differentiation and is found to interact with monosomes and disomes. However, it does not contain an actively translated smORF, therefore its interaction with polysomes likely results from its association with an mRNA. LINC01116 and DLGAP1-AS1 were detected in high levels in the 80S and small polysome complexes, showing that their association with translation machinery is dynamic during differentiation. Upon differentiation, LINC01116 transcript level was substantially increased in disome fraction, consistent with the presence of a higher number of P sites on the actively translated LINC01116 smORF in RA treated cells, compared to Control. This transcript contains a 216nt long smORF, with high ribosome profiling signal, indicating that it is being actively translated. LINC01116 lincRNA is involved in the progression of glioblastoma (GBM) (Brodie et al., 2017) and was found to be upregulated in gefitinib resistant non-small cell lung cancer cells (Wang He et al., 2020). In both these studies, knockdown of LINC01116 by siRNA decreased the expression of stemness markers (Nanog, SOX2 and Oct4) and reduced cell proliferation (increased number of cells in G0/G1 and decreased number of cells in S phase). This suggests LINC01116 promotes or at least allows cell proliferation, in this context, and is conflicting with my finding of it being upregulated upon differentiation. Interestingly however, knockdown of LINC01116 also inhibited migration of glioma stem cells (GSCs) (Brodie et al., 2017), while overexpression of LINC01116 promoted invasion and migration of gastric cancer cells (Su et al., 2019). This suggests that LINC01116 has a potential role in the formation of cell membrane protrusions, one of the main characteristics of differentiating SH-SY5Y cells, evident by the upregulation of protein-coding genes related to cell adhesion and migration upon differentiation. Based on all above evidence, LINC01116 is an interesting target to further investigate.

4.11.5 Conclusions

Overall, in this chapter I have shown that lncRNAs interact with the translation machinery and this interaction is affected by neuronal differentiation. ~3.5% of the polysome-associated lncRNAs appear to be translated in Control and ~2% in differentiated cells. smORFs within lncRNAs exhibit dynamic translation upon differentiation. Particularly interesting target lncRNAs have been selected from both polysome-associated and translated (e.g. LINC01116), for further study to dissect their function in relation to differentiation and their coding potential in the next chapter.

Chapter 5

Translated lncRNAs contribute to neuronal differentiation

5.1 Introduction

Non-canonical peptides encoded from lncRNAs have emerged as important regulators of organismal physiology. Genes previously annotated as lncRNAs have been found to encode small peptides, some as short as 11 aa. Functions for such peptides in development were first described more than a decade ago (Galindo et al., 2007; Kondo et al., 2007; Pueyo and Couso, 2008), and several more examples have since being characterised. Peptides encoded by 'Tarsal-less' gene locus (also known as 'pri') play pivotal roles in regulating *D. melanogaster* tracheal and leg morphogenesis (Kondo et al., 2007; Pueyo and Couso, 2008). Recently 2 peptides, each less than 30aa long, encoded by the same gene putative noncoding RNA gene pncr003:2L were also discovered in *D. melanogaster*. These were found to regulate Ca^{2+} signalling and regular muscle contraction in the heart (Magny et al., 2013). A human ortholog of these peptides has also emerged as a key regulator of Ca^{2+} metabolism during cardiac contraction in mammals. Myoregulin (MLN) peptide is encoded by LINC00948 and is conserved from *D. melanogaster* to human. It is a transmembrane peptide that binds to SERCA Ca^{2+} ATPase and inhibits Ca^{2+} reuptake by the sarcoplasmic reticulum in a similar manner to Phospholamban (PLN) and Sarcolipin (SLN) (Anderson et al., 2015). Notably, in contrast to SLN and PLN, MLN expression is not restricted only to the heart and slow skeletal muscle; it is expressed in all skeletal muscle (slow and fast) (Anderson et al., 2015).

Functional small peptides encoded from lncRNAs are reported to be involved in a number of different processes including differentiation. SPAAR 75aa long micro-peptide is encoded by a lncRNA LINC00961. It is expressed during embryonic stem cell to endothelial cell (EC) differentiation and its overexpression enhances endothelial network formation, via its interaction with SYNE1 protein. Interestingly, SPAAR peptide and LINC00961 transcript have opposing roles in angiogenesis, the former being pro-angiogenic and the latter anti-angiogenic. LINC00961 was shown to interact with Tβ4 and inhibit Tβ4-mediated angiogenesis (Spencer et al., 2020). MyolncR4 is a conserved lncRNA across vertebrates and is enriched in developing somites. It encodes the micro-peptide LEMP, that localises to the plasma membrane and to mitochondria, and has been shown to be essential for muscle development in mice. Disruption of just the LEMP ORF, without affecting the rest of the transcript showed

reduced expression of myosin heavy chain, confirming that the peptide, rather than the transcript is functional (L. Wang et al., 2020).

I have used Poly-Ribo-Seq to identify 45 lncRNA smORFs that are actively translated (analysis performed by Isabel Birds) in undifferentiated (28 smORFs) and differentiated (23 smORFs) SH-SY5Y cells, 6 of which are in common between the two conditions. One of the 41 novel translated smORFs belongs to LINC01116 lncRNA, which is enriched in the cytoplasm and is significantly upregulated during RA induced differentiation. This chapter aims to validate *in vitro* the translation events occurring on LINC01116 and to investigate the potential function of during neuronal differentiation.

5.2 LncRNAs are actively translated and produce small peptides

In the previous chapter, Poly-Ribo-Seq analysis identified several translated smORFs, in Control and differentiated cells, with translation efficiencies similar to those of known protein coding genes. Based on the triplet periodicity, the ribosome profiling signal on these smORFs corresponds to active translation by ribosomes. However, ribosome decoding doesn't necessarily result in the production of stable functional peptides. To determine whether the identified events result in the production of peptides, I undertook a FLAG-tagging approach.

5.2.1 Generation of FLAG-tagged constructs for lncRNA smORFs

To confirm the translation of smORFs from lncRNAs in an independent system, 2 potential smORFs were cloned into a FLAG-tagged mammalian expression plasmid, pcDNA3.1 (Table 5.1). The 5'-UTR and smORF of each lncRNA of interest, lacking its own stop codon, was inserted upstream of 3X FLAG tag, lacking its own start codon. Upon transient transfection of the FLAG tagged smORF into SH-SY5Y and HEK293 cells, detection of FLAG signal would therefore be the result of translation of the smORF (Figure 5.1). Selection of the smORFs was based on their translational efficiency in Control and RA treated cells, and the length of their 5'-UTR, as smORFs with very long 5'-UTRs would be more challenging to clone. Translation efficiency was

calculated for smORFs that were detected as translated in more than 2 biological replicates for each condition.

LINC00478 (or mir-99a-let-7c cluster host gene) which has previously been reported to be a tumour suppressor in ER+ breast cancer cells (Gökmen-Polar et al., 2016), is a lincRNA that has several alternative transcripts, ranging in length from 426 up to 5,889nt. Transcript ENST00000453910.5 (559nt long) contains a 111nt long smORF that showed robust ribosome profiling signal in both Control and differentiated cells (Chapter 4.9). LINC01116 lncRNA contains a 216nt long translated smORF, with robust ribosome profiling signal in Control (Chapter 4.9). FLAG-tagged constructs of LINC01116 and LINC00478 smORFs were created to be used for transfections in SH-SY5Y and HEK293 cells (Figure 5.1).

Table 5.1: Summary of smORFs selected for FLAG-tagging assay

Transcript name	Transcript ID	ORF length (nt)	Peptide length (aa)	5'-UTR length (nt)	TE in Control	TE in RA
LINC01116	ENST00000295549.9	216	71	608	0.55	0.41
LINC00478	ENST00000453910.5	114	37	150	1.96	3.23

In this table the ORF length, 5'-UTR length and TE (in Control and RA) of the 2 selected smORFs encoded by LINC01116 and LINC00478 are reported. LINC01116 was only detected as translated in one of the 3 RA biological replicates, therefore did not pass the threshold to be called translated in RA, which is why TE was calculated for LINC01116 in differentiated cells.

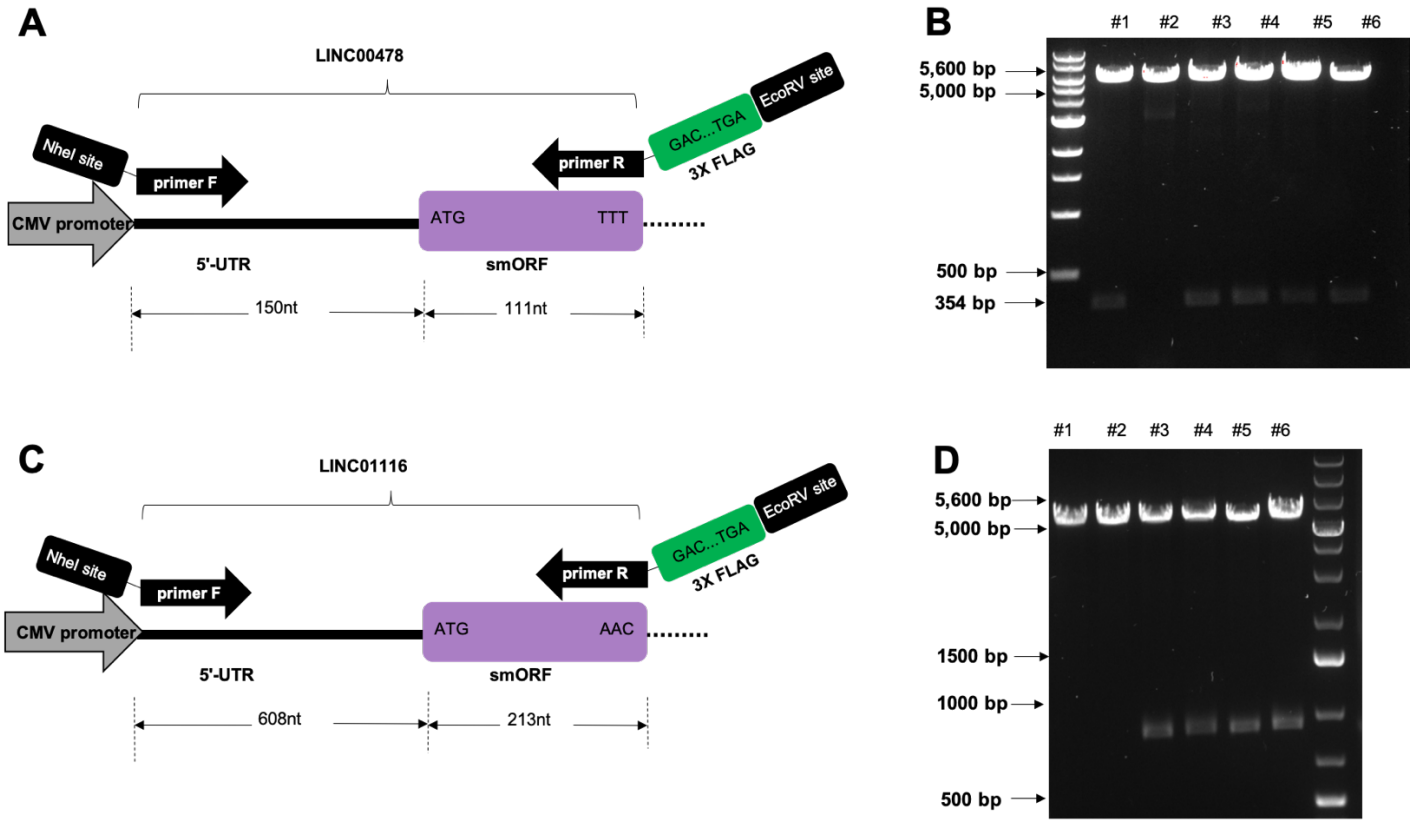


Figure 5.1: FLAG-tagged constructs for LINC01116 and LINC00578 smORFs were successfully generated. (A) Schematic of cloning strategy for LINC00478 smORF (B) Diagnostic digestions of miniprep plasmid DNA from 6 different *E.coli* DH5α colonies transformed with pcDNA-LINC00478smORF-FLAG construct. Colonies #1, #3, #4, #5 and #6 contained the insert (354bp). (C) Schematic of cloning strategy for LINC01116 smORF (D) Diagnostic digestions of miniprep plasmid DNA from 6 different *E.coli* DH5α colonies transformed with WT pcDNA-LINC01116smORF-FLAG construct. Colonies #3, #4, #5 and #6 contained the insert (914bp). Double digestion products were visualized on a 1% agarose gel and 1kb plus DNA ladder was used to aid the band size determination.

5.2.2 Expression of LINC00478 in SH-SY5Y and HEK293 cells

Having successfully cloned FLAG-tagged smORFs into pcDNA3.1, the aim was to assess expression of LINC00478 from the recombinant vector in SH-SY5Y cells. SH-SY5Ys have very low transfection efficiency (Nektaria Andronikou et al., 2015). To optimise transfection conditions a GFP expression plasmid (pLenti-CMV-GFP-puro-DNA, Addgene) was used to facilitate calculation of transfection effciencies, kindly provided by the Whitehouse group (University of Leeds). After optimising the quantity of DNA and the lipofectamine: DNA ratio, from across range manufacturer suggested, only an efficiency of ~30% (Figure 5.2) was achieved, consistent with previous reports (Andronikou et al., 2015).

As a positive control for FLAG translation, the same vector was used but with no lncRNA 5'-UTR and smORF inserted upstream of 3X FLAG-tag sequence but with its own start codon. When this positive control was transfected into SH-SY5Y cells however no FLAG signal was detected by immunofluorescence (Figure 5.3). The sequence of the construct was double checked to be correct, so it is not clear why no FLAG signal was detected. It is possible that FLAG tag is quickly degraded and therefore fails to be detected by immunocytochemistry. To overcome this issue, I used a construct with well-established FLAG expression (kindly provided by the Whitehouse group) as positive control for the detection of FLAG by immunofluorescence. The construct contains the FLAG-tagged sequence for DIM2 (yeast homolog of PNO1, an assembly factor for 40S subunit biogenesis), which localises in the nucleus in characteristic circular structures (Figure 5.4).

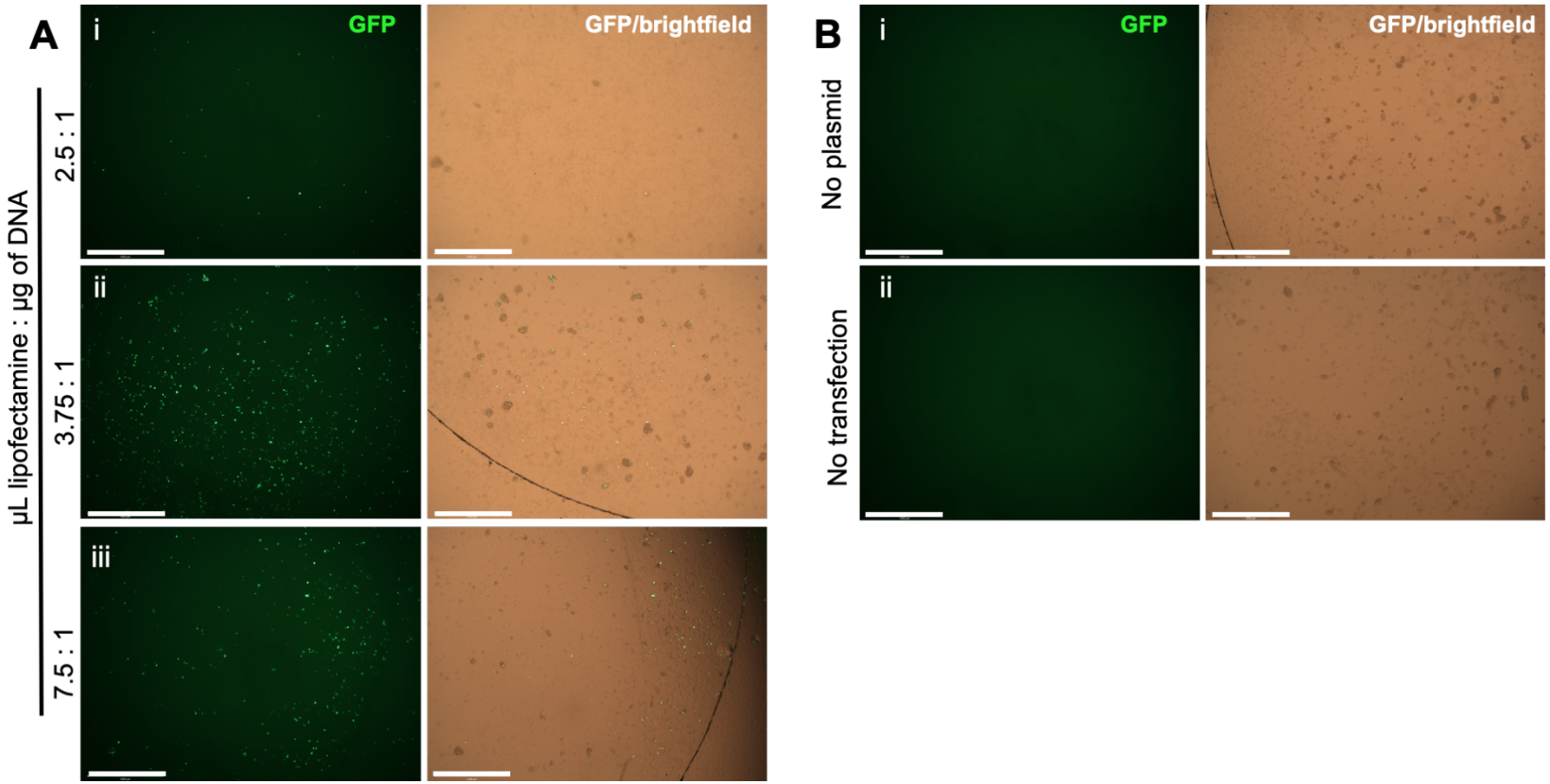


Figure 5.2: Transfection efficiency of SH-SY5Y does not exceed 30%. (A) SH-SY5Y cells, seeded in 6-well plates, on poly-L-lysine/mouse laminin coated coverslips were transfected with 1µg of GFP plasmid (pLenti-CMV-GFP-puro-DNA, Addgene) at ratios (µL of lipofectamine : µg of DNA) 2.5:1 (A:i), 3.75:1 (A:ii) and 7.5:1 (A:iii) (as per manufacturer's instructions). 3.75:1 ratio yielded the best transfection efficiency but did not exceed~30%. (B) Mock transfection with only lipofectamine (No plasmid control) showed that cells are stressed post-transfection, regardless of the plasmid DNA. Non transfected cells were used for comparison of cell density. Scale bar is 1mm.

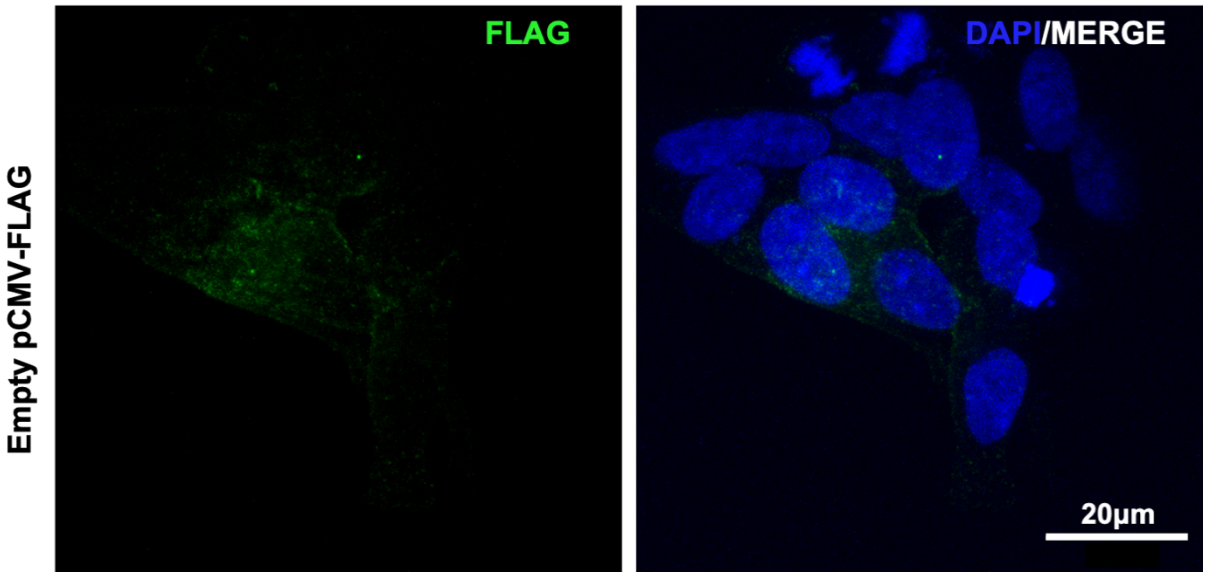


Figure 5.3: Empty pCMV-FLAG construct did not show adequate transfection efficiency. SH-SY5Y cells, seeded in 6-well plates, on poly-L-lysine/mouse laminin coated coverslips were transfected with pCMV-FLAG construct. Cells were incubated with mouse anti-FLAG antibody (1:500 dilution) and FLAG signal was detected by a 488 Alexa Fluor secondary antibody. Images obtained with an LSM700 (Zeiss) confocal microscope at 40X magnification using (Pln Apo 40X/1.3 Oil DICIII). Scale bar is 20µm.

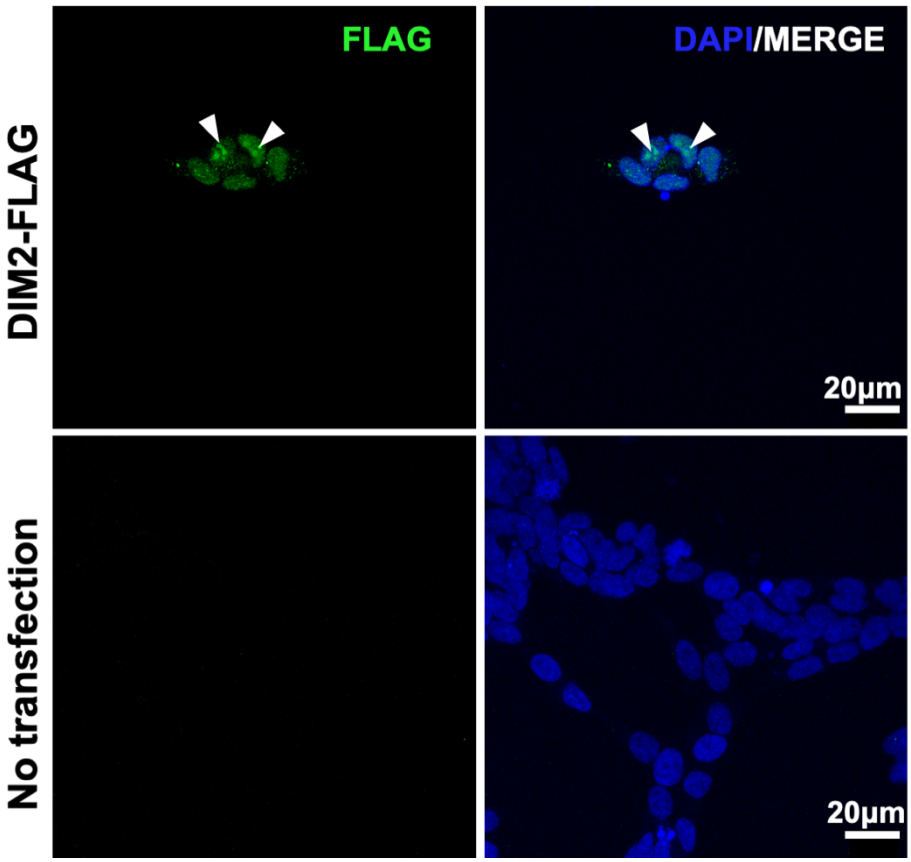


Figure 5.4: FLAG positive control for immunofluorescence. SH-SY5Y cells, seeded in 6-well plates, on poly-L-lysine/mouse laminin coated coverslips were transfected with DIM2-FLAG plasmid. Cells were incubated with mouse anti-FLAG antibody (1:500 dilution) and FLAG signal was detected in the nucleus, as circular structures (arrowheads) by a 488 Alexa Fluor secondary antibody. Untransfected cells were used as a negative control. Images obtained with an LSM700 (Zeiss) confocal microscope at 40X magnification using (Pln Apo 40X/1.3 Oil DICIII). Scale bar is 20µm.

To determine if LINC00478-smORF is translated, SH-SY5Y cells were transfected with the LINC00478-FLAG construct. When undifferentiated SH-SY5Y cells were transfected, FLAG signal was detected both in the nucleus and in the cytoplasm (~50% of the transfected cells exhibit nuclear signal and ~50% exhibit cytoplasmic signal) (Figure 5.5). Given that LINC00478 exhibited strong ribosome profiling signal upon differentiation as well, validation of its translation in RA treated cells was required. To this end, SH-SY5Y were treated with RA and transfected with LINC00478-FLAG construct 24h later. Cells were processed for immunofluorescence 3 days after RA treatment. Interestingly, whilst the number of transfected cells is similar to control, upon differentiation of SH-SY5Y cells FLAG signal only appears in the nucleus, indicating that LINC00478 peptide is restricted to the nucleus (Figure 5.6).

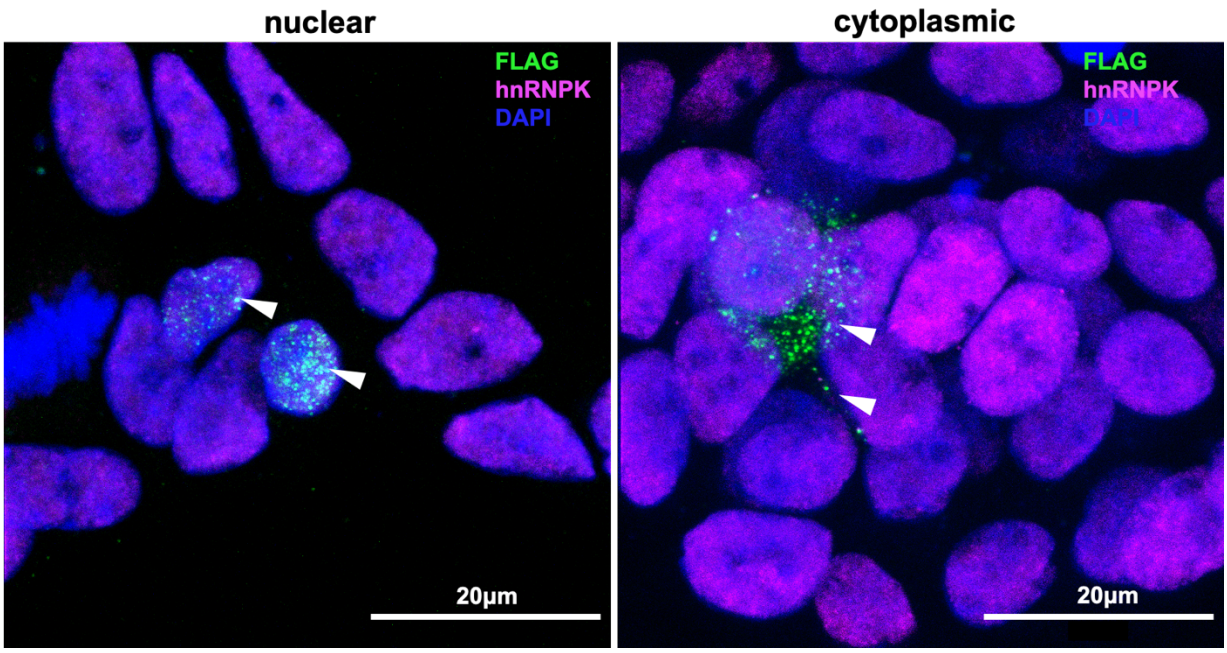


Figure 5.5: LINC00478 peptide localises in both the nucleus and the cytoplasm of SH-SY5Y cells. Representative images of SH-SY5Y cells transiently transfected with pcDNA-LINC00478-smORF-FLAG, were stained with mouse anti-FLAG (1:500 dilution). FLAG was detected by a 488 Alexa Fluor secondary antibody. FLAG signal appears in the nucleus (left panel), co-localising with hnRNPK and DAPI nuclear markers as well in the cytoplasm (right panel). Images are Z-stack overlay projections. Images obtained with an LSM700 (Zeiss) confocal microscope at 40X magnification using (Pln Apo 40X/1.3 Oil DICIII), scale bar= 20µm.

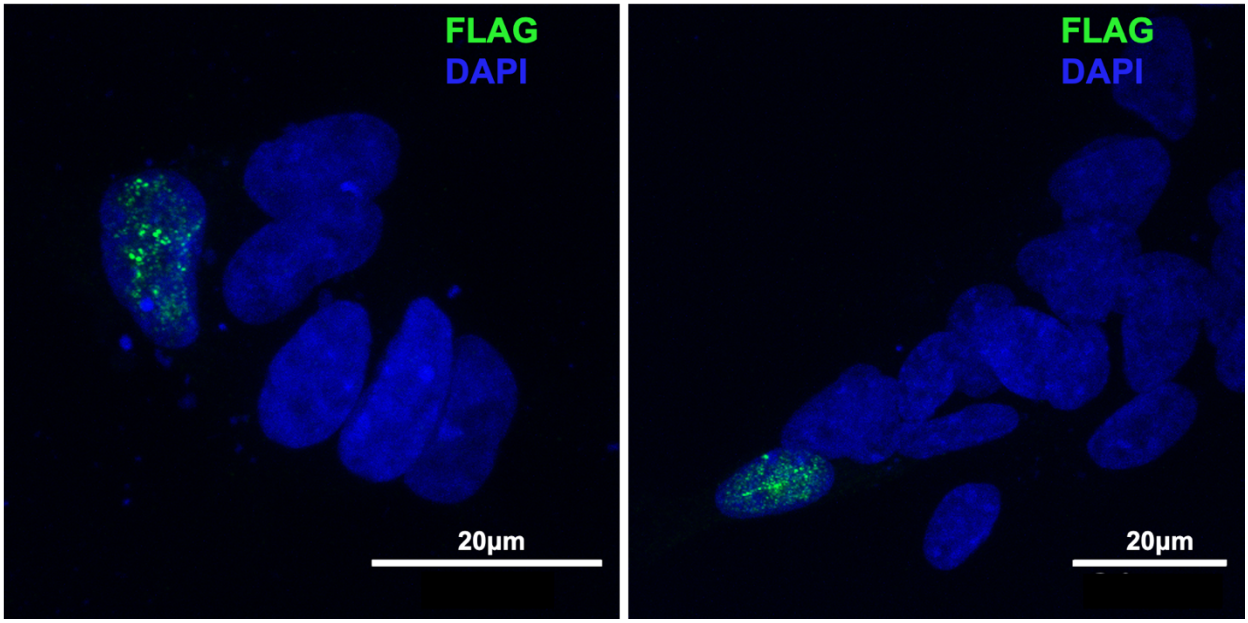


Figure 5.6: LINC00478 peptide localises in the nucleus of RA-treated SH-SY5Y cells. Representative images of differentiated SH-SY5Y cells transiently transfected with pcDNA-LINC00478smORF-FLAG, were stained with mouse anti FLAG (1:500 dilution). FLAG was detected by a 488 Alexa Fluor secondary antibody. FLAG signal appears in the nucleus, co-localising with DAPI nuclear marker. Images are Z-stack overlay projections. Initial images obtained with an LSM700 (Zeiss) confocal microscope at 40X magnification using (Pln Apo 40X/1.3 Oil DICIII). Scale bar is 20µm.

Given that transfection efficiency of SH-SY5Y cells was very low and only a few FLAG positive cells were detected, LINC00478 smORF was also transfected into HEK293 cells, which have a well-established high transfection efficiency. The FLAG-tagged LINC00478 peptide was detected in the nucleus and the cytoplasm of the HEK293 cells (Figure 5.7).

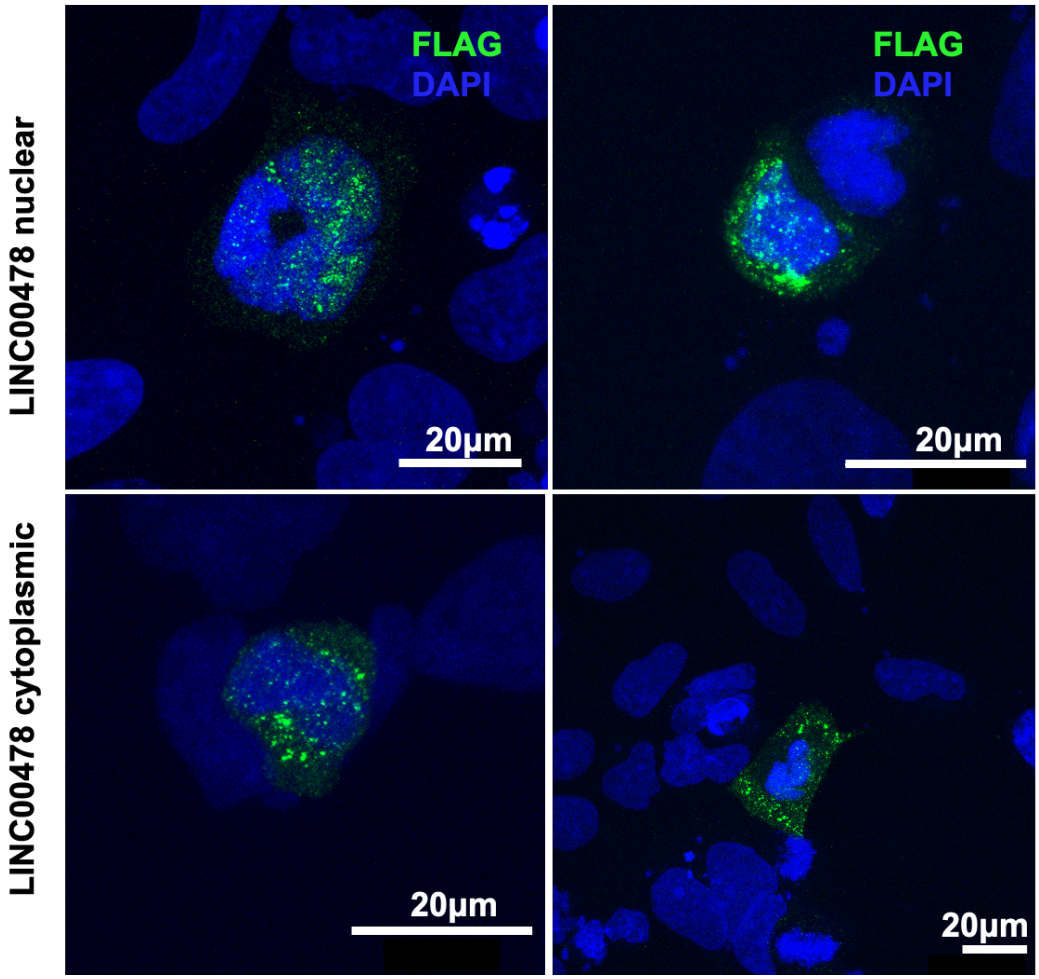


Figure 5.7: LINC00478 peptide localises in the nucleus and cytoplasm of HEK293 cells. Representative images of HEK293 cells transiently transfected with pcDNA-LINC00478smORF-FLAG. Cells were stained with mouse anti FLAG (1:500 dilution). FLAG was detected by a 488 Alexa Fluor secondary antibody. FLAG signal appears as distinct punctae inside the nucleus (upper panel), co-localising with DAPI nuclear marker as well in the cytoplasm (lower panel). Images are Z-stack overlay projections. Initial images obtained with an LSM700 (Zeiss) confocal microscope at 40X magnification using (Pln Apo 40X/1.3 Oil DICIII). Images were magnified 3X. Scale bar is 20μm.

5.2.3 Translation of LINC01116-smORF in SH-SY5Y and HEK293 cells

To assess whether LINC01116-smORF is translated, the FLAG-tagged LINC01116-smORF construct was transfected into SH-SY5Y cells and after 48h of incubation cells were stained with anti-FLAG. FLAG signal was detected from LINC0116-smORF transfection but only in ~10-20 cells per coverslip. (Figure 5.8). This suggests that LINC01116 smORF is translated from this construct, independently validating the results of the Poly-Ribo-Seq. FLAG-tagged LINC01116 peptide localised to punctae within the cytosol (Figure 5.8: A(i-ii)) and specifically close to the cell membrane and cell neuritic protrusions (Figure 5.8: A(iii-iv)) in undifferentiated cells.

Due to the low transfection efficiency of SH-SY5Y cells, LINC01116 smORF translation was further validated in HEK293 cells. Upon transient transfection of LINC01116-FLAG smORF in HEK293, the peptide showed a pattern of punctate cytoplasmic localisation (Figure 5.9). In the transfected cells, the staining appeared enriched close to cell membranes and in protrusions as determined by the localisation of FLAG+ punctae.

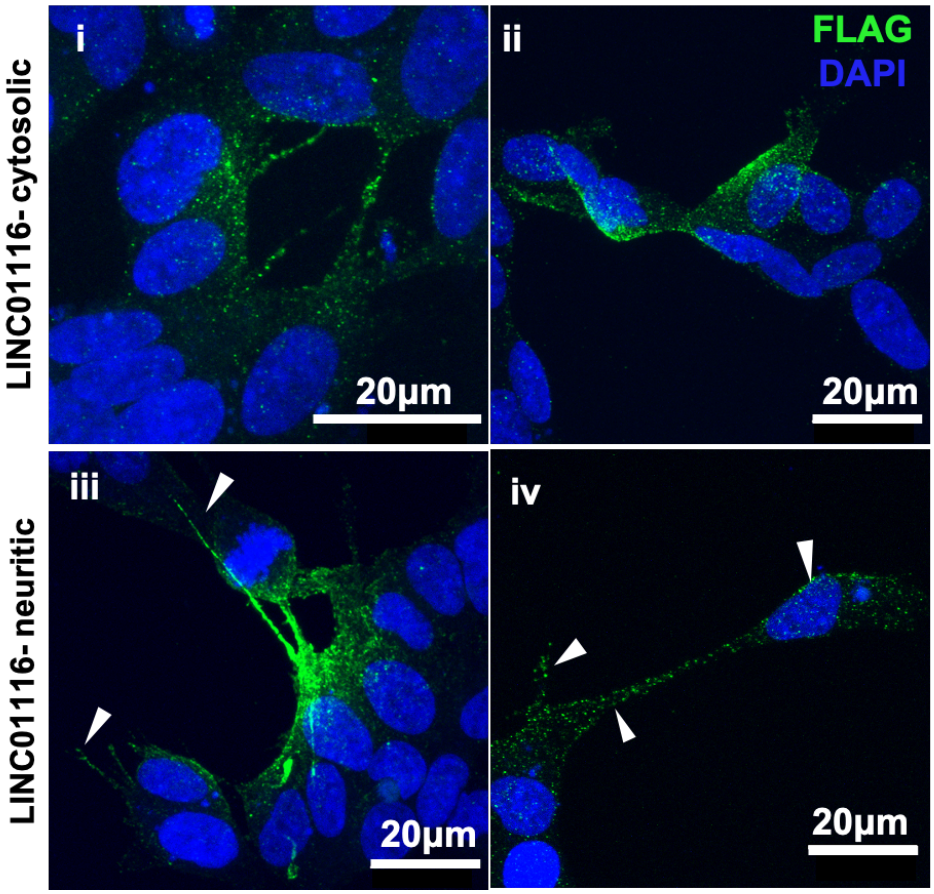


Figure 5.8: LINC01116 smORF produces a peptide that localises to punctae within the cytosol and in neuritic protrusions. SH-SY5Y cells transiently transfected with pcDNA-LINC01116-smORF-FLAG, were stained with mouse anti-FLAG (1:500 dilution). FLAG was detected by a 488 Alexa Fluor secondary antibody, forming punctae in the cytoplasm (i and ii) or in neuritic protrusions (iii and iv). Images are Z-stack overlay projections. Initial images obtained with an LSM700 (Zeiss) confocal microscope at 40X magnification using (Pln Apo 40X/1.3 Oil DICIII). Scale bar is 20µm.

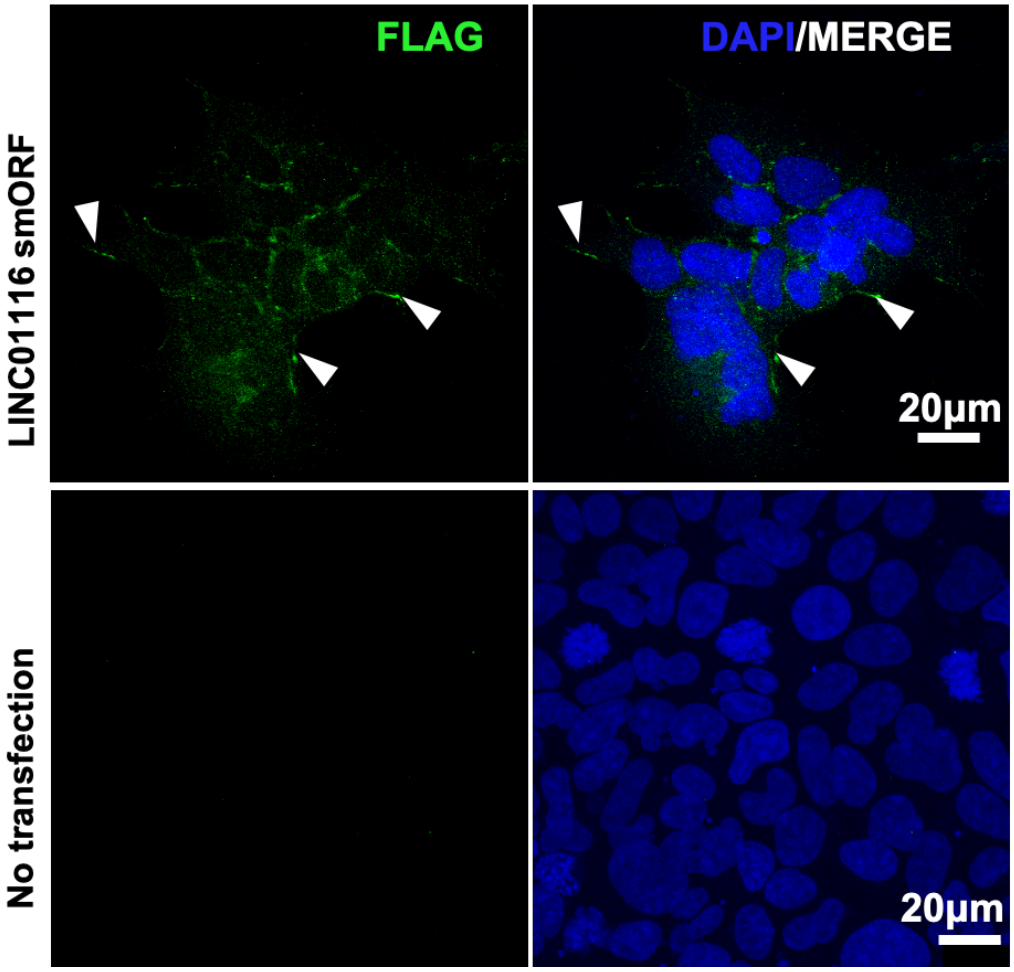


Figure 5.9: LINC01116 smORF is translated in HEK293 cells. (A) HEK293 cells, transfected with WT-LINC01116smORF-FLAG, were stained with mouse anti FLAG (1:500 dilution). FLAG was detected by 488 Alexa Fluor secondary antibody in the cytoplasm and cell membrane protrusions. (B) No transfection (negative control). Images are Z-stack overlay projections. Initial images obtained with an LSM700 (Zeiss) confocal microscope at 40X magnification using (Pln Apo 40X/1.3 Oil DICIII). Scale bar is 20μm.

To test that FLAG signal was the result of translation of the precise smORF predicted from Poly-Ribo-Seq, the ATG start codon was mutated. In fact, within the 5'-UTR of LINC01116 transcript, there are two ATG start codons in frame and in close proximity (45 nucleotides from one another, Figure 5.10: A). No ribosome profiling signal was detected between the upstream ATG and downstream ATG suggesting the second AUG is used for the translation of the smORF. However, that does not exclude the possibility that the first AUG is used as the start codon of the smORF. To help determine which AUG is used, the two start codons were scored according to their similar to the Kozak sequence consensus, using NetStart1.0 (Pedersen and Nielsen, 1997). Both AUGs generated similar scores, (greater than 0.5, which is the cutoff for a probable translation event) to both ATGs (ATG_1= 0.545 and ATG_2= 0.645). Therefore, both are in good context to be used for translation initiation. To investigate whether the upstream ATG is the one used for the translation of the smORF, site directed mutagenesis was used to mutate ATG_1 to AAA (Δ1 mutant) (Figure 5.10: B).

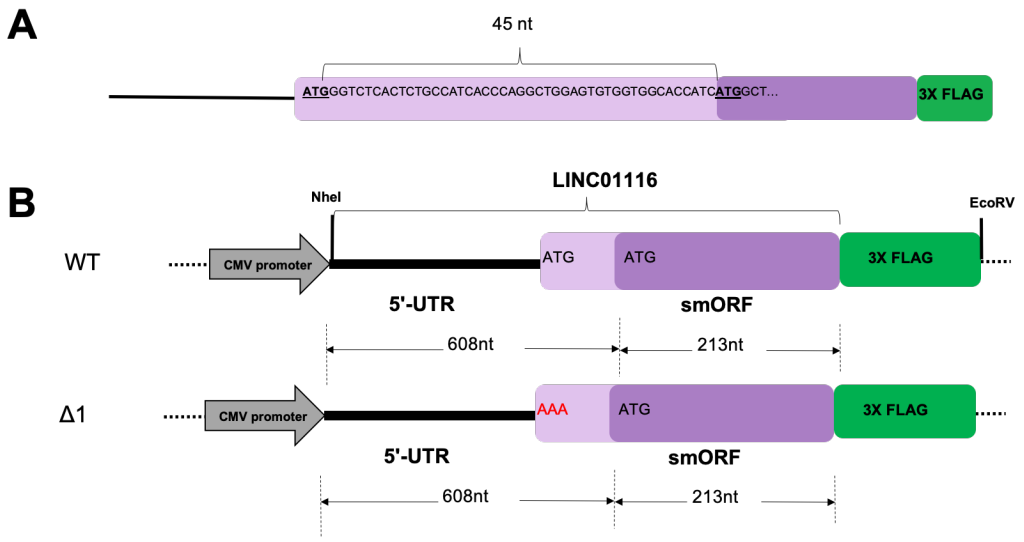


Figure 5.10: Schematic of LINC01116 FLAG-tagged wild type (WT) and first ATG (Δ1) start codon mutants. (A) LINC01116 smORF has 2 ATGs 45nt apart from one another. The first ATG start codon was mutated (Δ1 mutant) to examine whether translation initiates from it.

Transfections with the Δ1 construct produced no signal, suggesting that translation proceeds from the first ATG of the smORF and produces an 87aa peptide (Figure 5.11: A). Δ1 construct was further transfected to HEK293 cells and yielded similar results (Figure 5.11: B). Based on the above results, it seems more likely that the first ATG is the one utilised for translation initiation, however, transfection of a mutant of the second ATG would be required, in order to safely conclude which ATG is used.

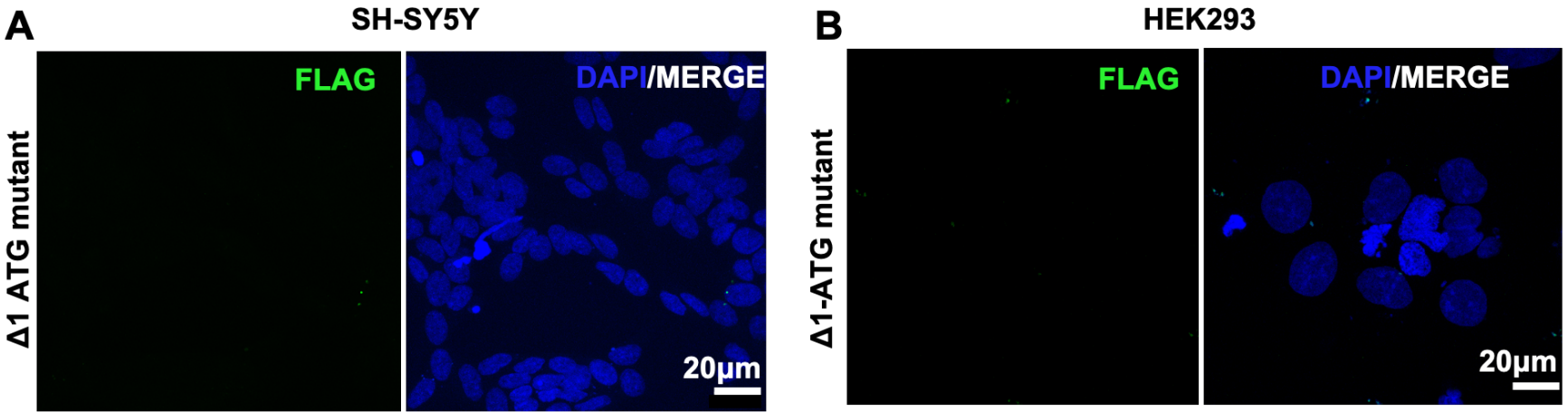


Figure 5.11: Mutations of first AUG start codon abolished FLAG signal from LINC01116-smORF in SH-SY5Y and HEK293 cells. (A) SH-SY5Y and (B) HEK293 cells transfected with Δ1 ATG mutant of LINC01116 smORF-FLAG- tagged construct were stained with mouse anti FLAG (1:500 dilution). FLAG was detected by a 488 Alexa Fluor secondary antibody. No FLAG signal was detected in either of the cell lines, indicating that translation of LINC01116 likely initiates from the first ATG of the transcript Images are Z-stack overlay projections. Initial images obtained with an LSM700 (Zeiss) confocal microscope at 40X magnification using (Pln Apo 40X/1.3 Oil DICIII). Scale bar is 20μm.

To further confirm LINC01116-FLAG and LINC00478-FLAG translation and peptide production Western blots were performed with lysates prepared from transfected cells. However, I was not able to detect a FLAG signal (Figure 5.7). Transfection of a GFP plasmid was also performed as a control for transfection efficiency, and for the Western blotting procedure, and was successful. The fact that FLAG was not detected may be attributed to the small molecular mass of the peptides (LINC01116 peptide ~11kDa, and LINC00478 peptide ~7kDa).

Collectively, the above results confirm the production of small peptides from actively translated smORFs within lncRNA transcripts. The localisation of these peptides in the cell membrane (LINC01116) and in the nucleus (LINC00478) may indicate a potential function for these peptides.

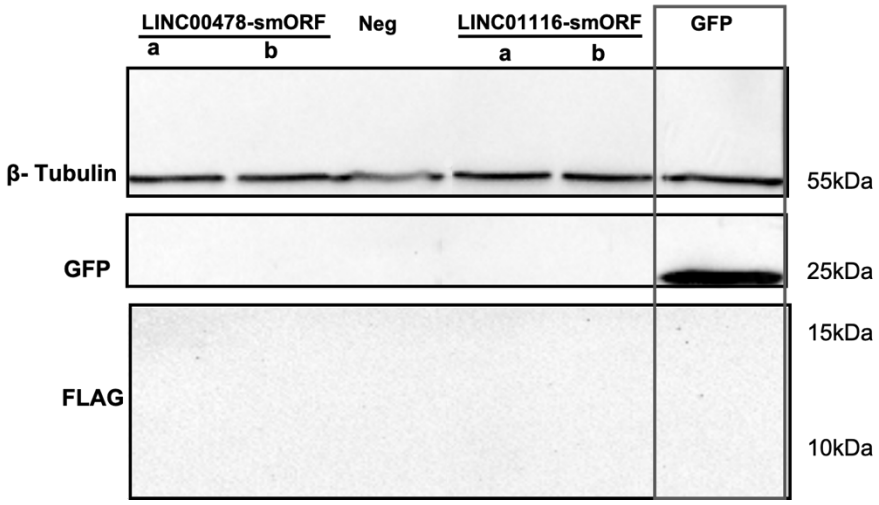


Figure 5.12: LINC01116 and LINC00478 FLAG-tagged peptides were not detected by western blot. Cell lysates from HEK293 transfected with pcDNA-LINC00478smORF-FLAG, pcDNA-LINC00116-smORF(WT)-FLAG (in 2 biological replicates a and b) and GFP plasmid (grey rectangle), as a positive control for transfection efficiency, were run on 15% polyacrylamide gel and protein was transferred to nitrocellulose membrane. Membrane was probed for β-tubulin (55kDa) (1:5000 antibody dilution), GFP (25kDa) (1:1000 antibody dilution) and FLAG (1:500 antibody dilution). No FLAG signal was detected.

5.3 LINC01116 is involved in the regulation of neuronal differentiation

Poly-Ribo-Seq revealed the translation of a smORF in LINC01116, which was independently validated with FLAG-tagged transfection assays in both SH-SY5Y and HEK293 cells. Given that LINC01116 is upregulated upon RA treatment, I hypothesised that it may be involved in the regulation of neuronal differentiation. To elucidate the role of LINC01116 in neuronal differentiation, the aim was to determine its expression profile in different stages of differentiation and undertake a loss of function approach in order to assess its potential function.

5.3.1 Expression profile of LINC01116 during neuronal differentiation

To further study the role of LINC01116, an assessment of its expression levels during neuronal differentiation was required. RNA-Seq analysis showed that LINC01116 is upregulated in SH-SY5Y cells at day 3 post differentiation. To profile its expression during differentiation in more detail, cells were induced for up to 8 days using a published method (Forster et al., 2016). RNA was harvested on days 1, 2, 3 (phase 1) and day 8 (phase 2 differentiated cells) post differentiation (Figure 5.13: A). RT-qPCR showed that LINC01116 expression is upregulated 3.4-fold in the first 24 hours post-differentiation and remains at similar levels until day 3 (Figure 5.13: B). However, after phase 2 differentiation LINC01116 levels drop ~6-fold, almost back down to undifferentiated control levels (Figure 5.13: C). The rapid upregulation of LINC01116, as well as its subsequent sharp decrease suggest that LINC01116 potentially functions in the early stages of differentiation. Therefore, follow up experiments focus on the first 3 days after differentiation is induced.

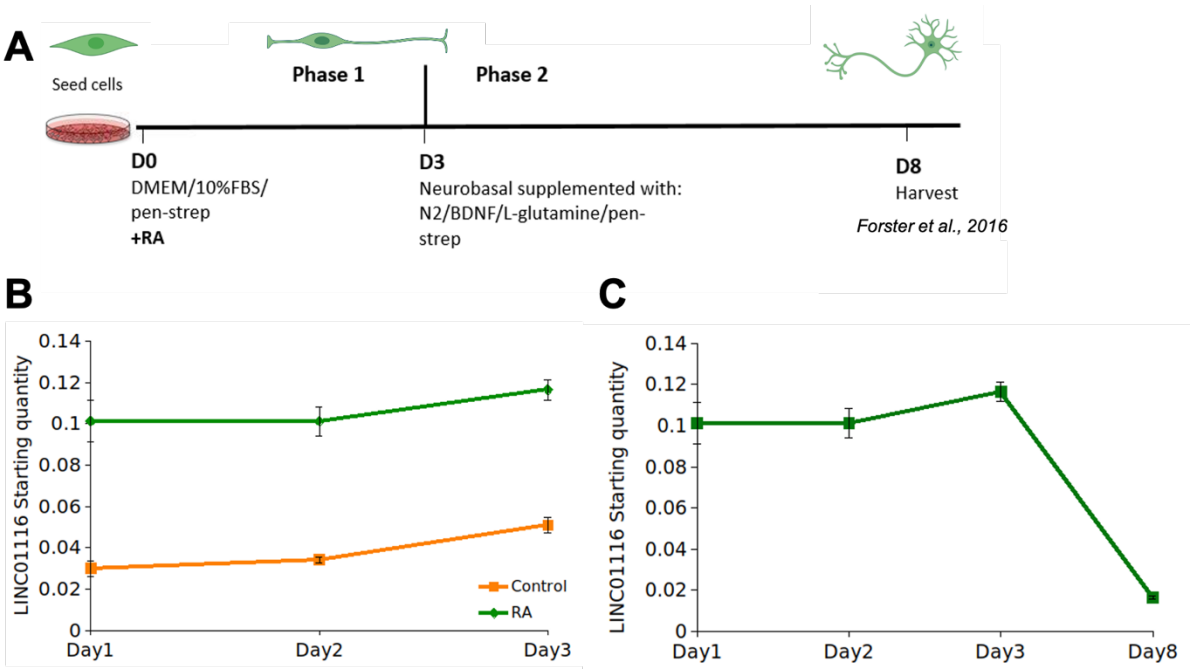


Figure 5.13: LINC01116 is upregulated in the early stages of neuronal differentiation. (A) Phase 1 and phase 2 neuronal differentiation protocol (Forster et al., 2016). (B) RT-qPCR targeting LINC01116 transcript in Control and RA-treated cells at days 1, 2 and 3 post-RA treatment (phase 1) (C) LINC01116 expression levels, in phase 1 and phase 2 differentiated SH-SY5Y cells, measured by RT-qPCR. LINC01116 levels decrease ~6-fold during phase 2 differentiation. LINC01116 expression levels in (B) and (C) were calculated with absolute quantification, based on a standard curve generated from serial dilutions of a pool of Control and RA treated samples.

5.3.2 Knockdown of LINC01116 and effect on viability of SH-SY5Y cells

To investigate the role of LINC01116 lncRNA in neuronal differentiation of SH-SY5Y I took a loss of function approach and knocked its expression down by siRNA. To knockdown LINC01116 a pool of 4 siRNAs specifically designed to target only LINC01116, minimizing off target effects (Lincode-SMARTpool, Dharmacon^{TM}) (Figure 5.14) were used. As a negative control, a pool of Non-targeting oligos were employed (Non-targeting pool, Dharmacon^{TM}). To determine the efficiency of LINC01116 knockdown, the level of LINC01116 RNA was measured in both LINC01116 siRNA and non-targeting siRNA treated cells, relative to a reference gene (GAPDH) using the ΔΔCq method (Haimes and Kelley, 2015). Samples were collected for 3 consecutive days, following transfection of the siRNA. LINC01116 was efficiently knocked-down (78-86%) for all 3 days (Figure 5.15: A).

Having established that knockdown was efficient, possible effects of LINC01116 knockdown on cell viability were assessed. Cell viability was measured, by Trypan Blue assay, for 3 consecutive days following LINC01116 siRNA treatment in 3 independent biological replicates (Figure 5.15: B). On day 1, after knockdown, the number of scrambled treated cells was significantly lower than that of siRNA treated cells. However, in the following days, cells recovered and no significant differences in the number of cells were observed between LINC01116 siRNA treated and scrambled control treated cells. Hence, LINC01116 knockdown did not affect viability of SH-SY5Y cells.

Assessment of the effects of LINC01116 knockdown on differentiation required that reduced expression persists for at least 5 days. Therefore, levels of LINC01116 were measured at 4, 5- and 6-days post knockdown (Figure 5.16) by RT-qPCR and compared to expression level of LINC01116 in the scrambled control sample. This revealed that knockdown of LINC01116 persists up to 6 days post siRNA delivery.

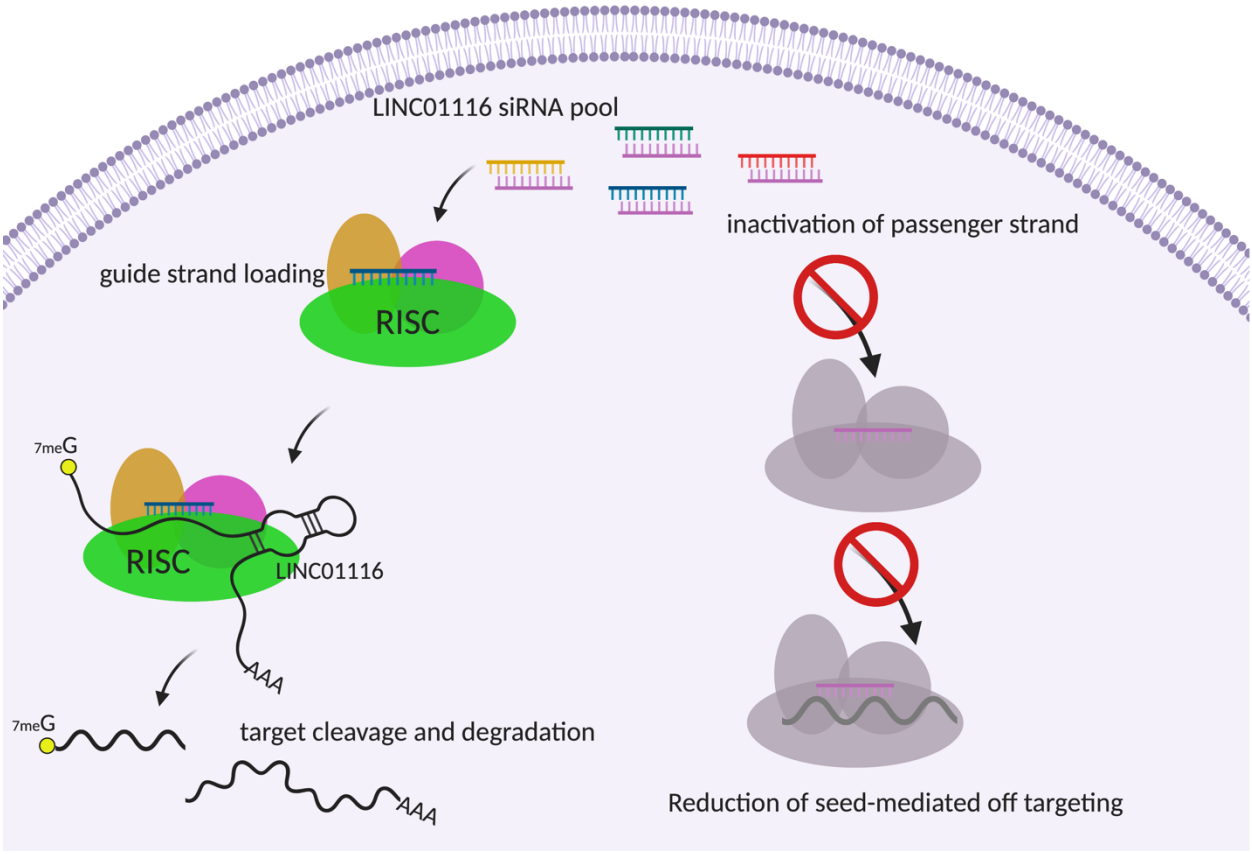


Figure 5.14: Specific LINC01116 knockdown by Lincode SMARTpool of siRNAs designed to minimize off-target effects. siRNA oligos are specifically designed to ensure high-efficiency silencing and also dual-strand chemical modifications to ensure optimal strand loading and disrupt microRNA-like seed activity for the reduction of off-targets. Image was adapted from Horizon Discovery™ and was created using BIORENDER.

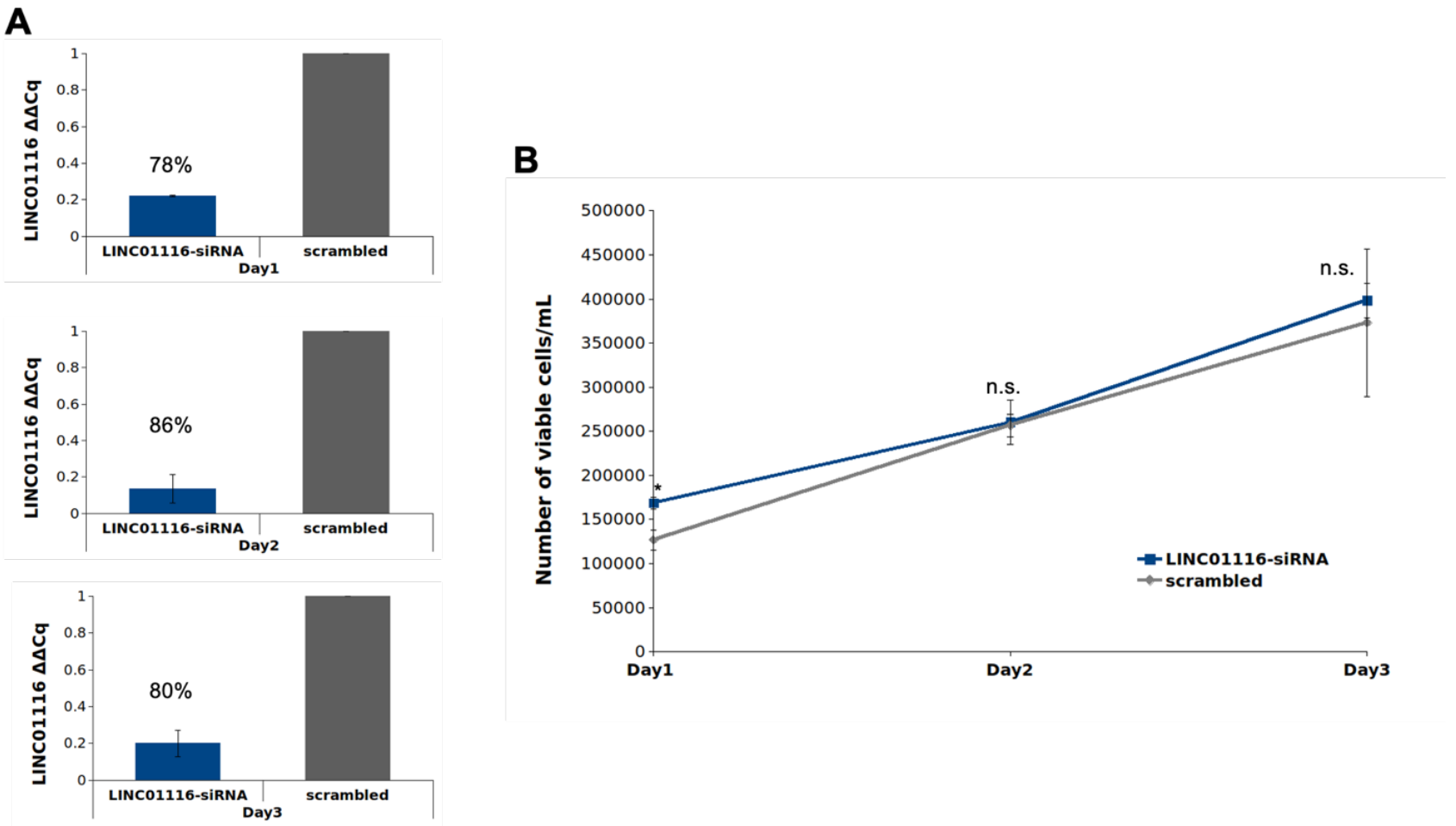


Figure 5.15: LINC01116 knockdown does not affect cell viability. (A) LINC01116 is effectively knocked down (78-86%) from day 1 to day 3 after siRNA delivery, as shown by RT-qPCR quantifying LINC01116 transcript levels relatively to GAPDH using the ΔΔCq method. (B) Cell viability of SH-SY5Y cells was assessed by Trypan Blue assay upon knockdown (LINC01116 siRNA and scrambled

control). On day 1 the number of scrambled treated cells was significantly lower than siRNA treated (student's t test, n=3, $p<0.05$) but in the following 2 days there was no significant difference between siRNA and scrambled treated. Error bars show standard error

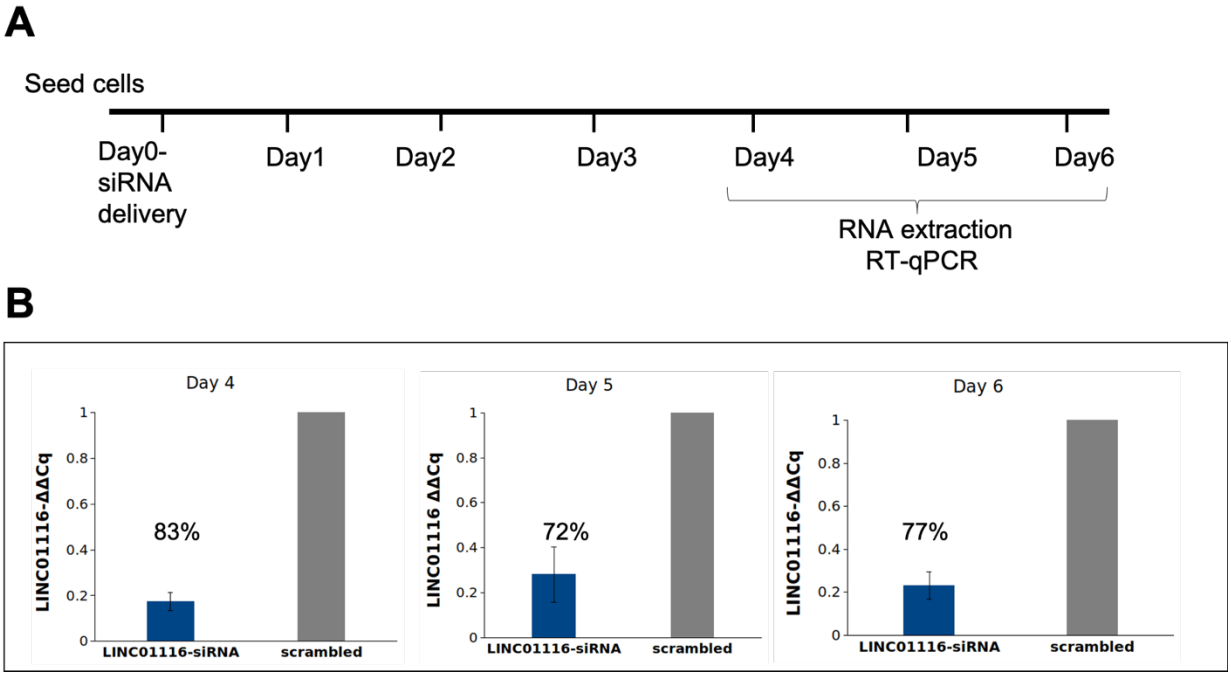


Figure 5.16: LINC01116 knockdown persists 6 days after siRNA transfection. (A) Timeline of knockdown. siRNA oligos targeting LINC01116 were administered to the cells ~18h after seeding and samples were collected in consecutive days to assess the expression level of LINC01116. (B) LINC01116 knockdown (72-83%) is persistent until day 6 after siRNA delivery, as shown by RT-qPCR quantifying LINC01116 transcript levels relatively to GAPDH using the ΔΔCq method.

5.3.3 Knockdown of LINC01116 in differentiated SH-SY5Y induces neurite length reduction but does not affect proliferation

Having established that knockdown of LINC01116 persists for up to 6 days upon siRNA delivery, the next aim was to determine the effect of LINC01116 knockdown on differentiation following by RA treatment. SH-SY5Y cells were seeded 16h prior to siRNA transfections and RA was administered 48h after siRNA delivery, to allow the cells to recover from the stress induced by transfection and restore their physiological proliferation rate (Figure 5.17: A). Knockdown efficiency was assessed as previously described and LINC01116 was shown to be depleted by 94% in differentiated cells as well as undifferentiated cells 86% (Figure 5.17: B).

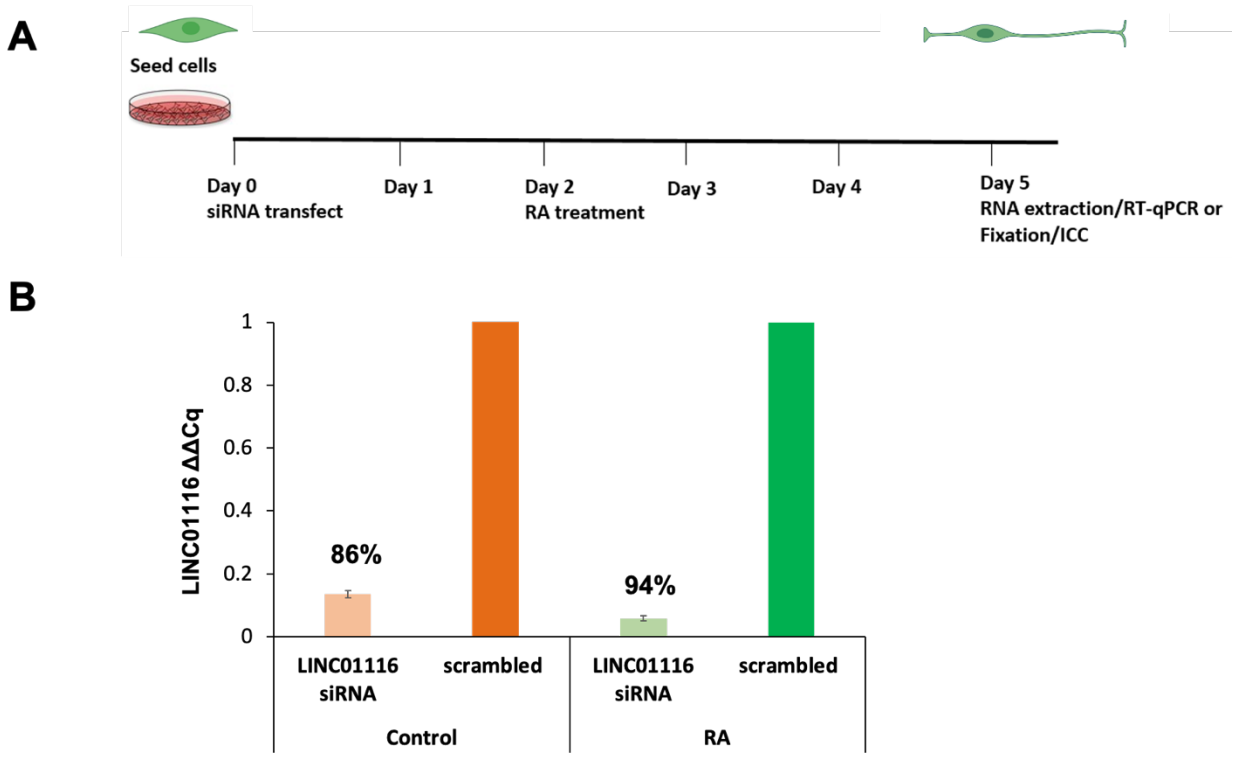


Figure 5.17: LINC01116 is efficiently knocked down during differentiation. (A) Timeline of knockdown and induction of differentiation (RA treatment). (B) LINC01116 is effectively knocked down in Control (86%) and RA treated cells (94%), as shown by RT-qPCR quantifying LINC01116 transcript levels relative to GAPDH using the ΔΔCq method.

To assess the effect of siRNA knock-down of LINC01116 on neuronal differentiation, the neurite lengths of Control and differentiated cells, treated with LINC01116 siRNA or scrambled control were measured. LINC01116 expression was first knocked down, then cells treated to induce differentiation, for 3 days or as control. Immunocytochemical staining for Tuj1 (β-III tubulin), to visualise neurites (chapter 3.1.1) was performed.

No difference in the length of neurites were observed between the LINC01116 siRNA and scrambled siRNA treated undifferentiated cells (control) (Figure 5.18: A). However, upon differentiation, the neurites of LINC01116 knockdown cells were significantly shorter compared to those of scrambled-treated cells (Figure 5.18: A; B). In fact, neurite length of treated cells in which LINC01116 is knocked down is comparable to that seen in untreated controls. This suggests that LINC01116 could play a role in differentiation.

Attenuation of differentiation is often coupled with enhancement of cell proliferation as these events are coordinated (Tsioras et al., 2013). Therefore, the effect of LINC01116 knockdown on cell proliferation was assessed. Following 3 days RA-treatment, no statistically significant difference in the number of proliferative cells, measured as those which are ki67+, between scrambled and siRNA treated cells (Figure 5.19). This was the case both for Control and differentiated cells.

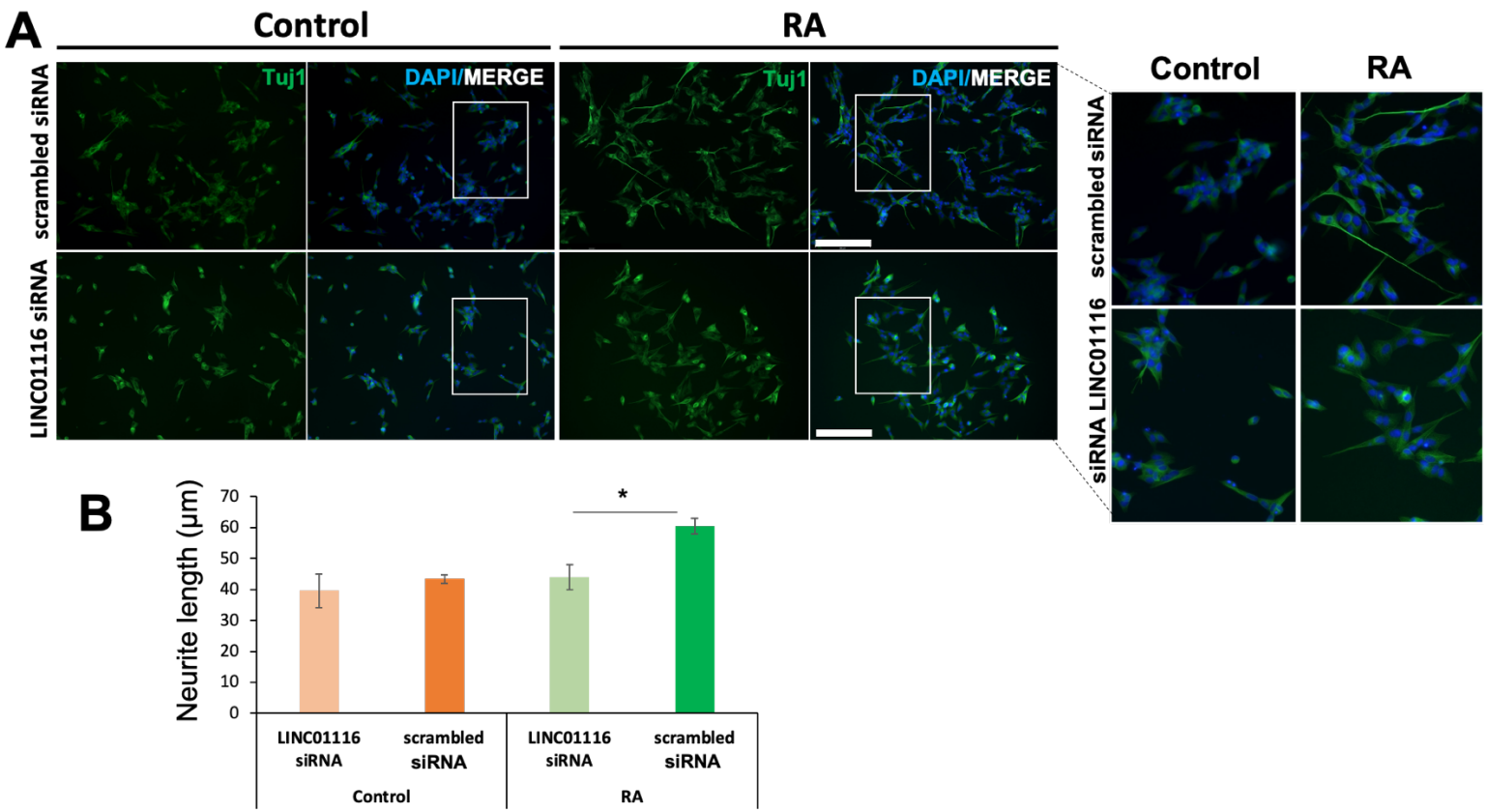


Figure 5.18: Knockdown of LINC01116 results in reduction of neurite length upon neuronal differentiation. (A) Representative immunofluorescence image of Control and RA SH-SY5Y cells, transfected with siRNA targeting LINC01116 and scrambled control, after staining for Tuj (βIII-tubulin; 1:50 antibody dilution) at day 3 post-differentiation. Neurites of RA treated cells transfected with siRNA targeting LINC01116 are shorter compared to those of RA-treated cells transfected with scrambled control (white windows magnified on the right). Scale bar=200µm. (B) Quantification of neurite length in µm shows that knockdown of LINC01116 in differentiated cells results in a significant reduction of the neurite length compared to that of differentiated cells treated with scrambled control (student's t test, N=3, n>300, $p<0.05$). Neurite length of Control (undifferentiated N=3, n>300, $p>0.05$).

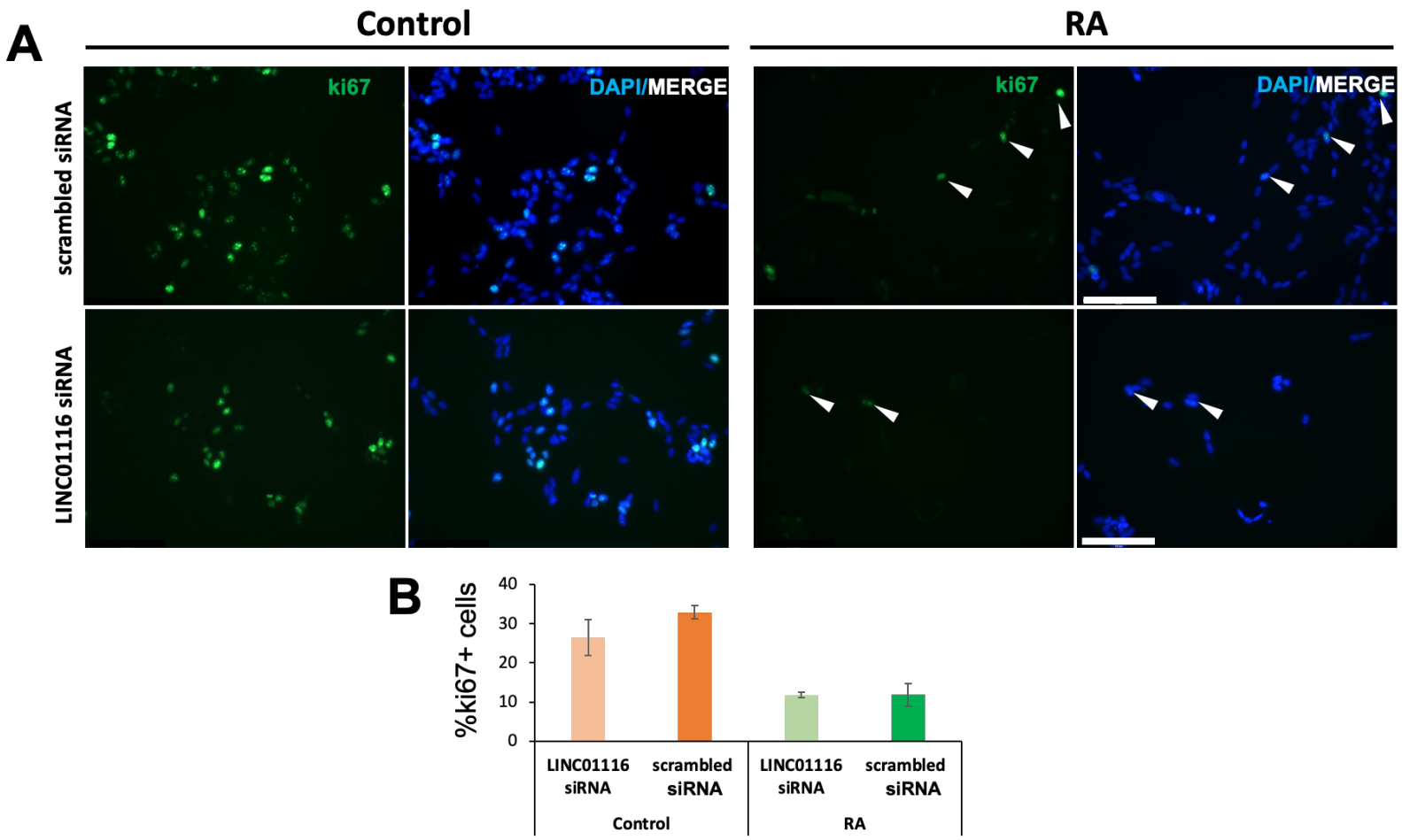


Figure 5.19: Knockdown of LINC01116 does not affect cell proliferation. Representative immunofluorescence image of Control and RA SH-SY5Y cells, transfected with siRNA targeting LINC01116 and scrambled control, after staining for proliferation marker ki67 (1:100 antibody dilution) at day 3 post-differentiation. The number of ki67+ cells is not significantly different between the cells treated with siRNA targeting LINC01116 and cells treated with scrambled siRNA (student's t test, N=3, n>300). Scale bar=100μm.

To further examine effects of LINC01116 knockdown, the expression of several markers was measured by RT-qPCR. MOXD1 and NTN4 were used as differentiation markers and SOX2 and the G1/S phase progression regulator E2F1, as pluripotency markers (Chapter 3.2.3). The levels of these marker mRNAs were normalised to the levels of GAPDH (Figure 5.20).

Upon differentiation there was a statistically significant reduction in the level of MOXD1 in the LINC01116 siRNA treated samples compared to the scrambled control, potentially indicating that LINC01116 knockdown results in a delay of differentiation (Figure 5.20). NTN4 showed a similar reduction, though this was not significant. Neither marker was affected by the knockdown in undifferentiated cells (Figure 5.20).

SOX2 expression and E2F1 expression appeared slightly increased in knockdowns compared to Controls, but these differences were not statistically significant. Taken together, the above data suggest that LINC01116 knockdown may impose a delay on neuronal differentiation and impede axon outgrowth, but that it does not affect proliferation.

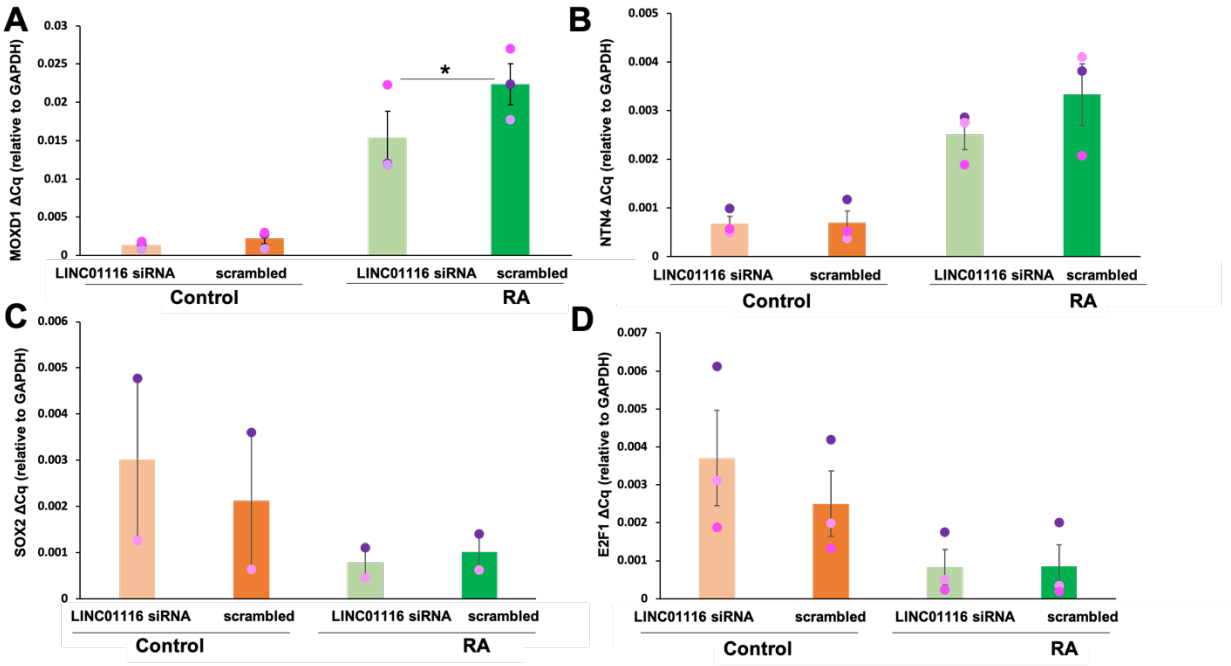


Figure 5.20: LINC01116 knockdown results in reduced differentiation but does not affect proliferation. RT-qPCR of differentiation and pluripotency/proliferation markers in Control and differentiated (RA) cells treated with LINC01116 siRNA or scrambled siRNA control. Relative quantification was performed using the ΔCq method and GAPDH as reference gene. (A) MOXD1 transcript is significantly reduced upon knockdown in differentiated cells (student's t test, n=3, *p<0.05) (B) NTN4 shows a trend of reduction upon knockdown in differentiated cells but the reduction was not significant (student's t test, n=3,). (C) SOX2 and (D) E2F1 expression did not significantly change upon LINC01116 knockdown in Control or differentiated cells (student's t test, n=3, standard error is plotted, biological replicates represented as dots).

5.4 LINC01116 whole transcript overexpression does not affect neurite outgrowth

Given that LINC01116 knockdown resulted in reduced neurite length in RA treated cells, I decided to test whether its exogenous expression would have a reciprocal effect. Additionally, to determine whether the smORF or the whole transcript is responsible for the neurite elongation, WT and double ATG mutant (Δ1Δ2) constructs encoding the whole LINC01116 transcript (with no tags) were transfected into undifferentiated SH-SY5Y cells (Figure 5.21).

Cells were transfected with either WT LINC01116 transcript, or with the Δ1Δ2 mutant that does not produce any peptide (Figure 5.21-schematic). As a marker for the LINC01116 overexpression, cells were co-transfected with a GFP plasmid. GFP+ cells were considered to be LINC01116 overexpressing. Neurite length was also measured in non-transfected cells for comparison (Figure 5.21). Neurite length of cells overexpressing either WT LINC01116 or ATG mutant (Δ1Δ2) LINC01116 did not differ significantly from non-transfected cells.

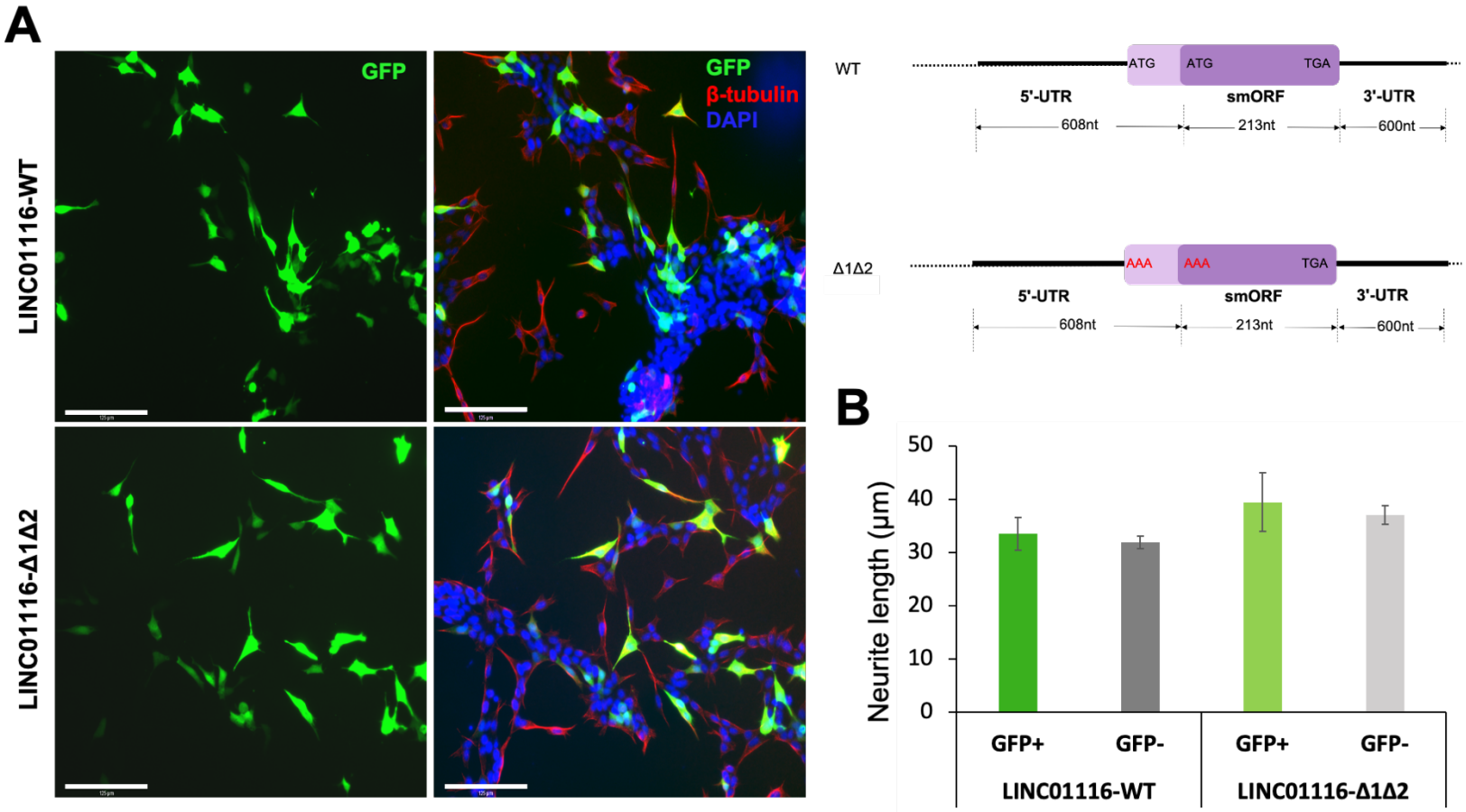


Figure 5.21: LINC01116 WT or ATG mutant overexpression did not significantly change the neurite length of SH-SY5Y cells. (A) SH-SY5Y cells were co-transfected with GFP and LINC01116-WT or LINC01116 ATG-mutant (constructs shown on the right) and stained for GFP (green), β-tubulin (red). Nuclei were stained with DAPI. Scale bar is 125μm. (B) Neurite length quantification for LINC01116 overexpressing (green) and non-transfected cells showed no significant difference between the WT and mutant, or between LINC01116 overexpressing and non-transfected cells (student's t test, N=3, n>100), scale bar= 125μm.

5.5 DISCUSSION

5.5.1 LncRNAs LINC01116 and LINC00478 are translated and produce small peptides.

This chapter followed up on some of the target lncRNAs identified as translated by Poly-Ribo-Seq. Specifically, the production of 2 small peptides from LINC01116 and LINC00478 smORFs, was validated *in vitro* in SH-SY5Y and HEK293 cells, using FLAG tagging assays. LINC01116 peptide was detected in the cytoplasm and appeared to decorate cell membrane protrusions. The LINC01116 smORF appeared to be translated from the first of two in-frame ATG codons. The limited number of FLAG+ cells detected could be attributed to low transcription from the construct, as the selected LINC01116 transcript that was cloned, based on the Ribo-Seq signal, lacks its transcription start site (TSS). This may or may not affect its transcription from the pcDNA3.1 vector. Mass spectrometry of SH-SY5Y cells, or analysis of publicly available Mass spectrometry data (Brenig et al., 2020) would possibly corroborate the existence of a peptide encoded from LINC01116. LINC00478 encodes a 37aa peptide which was detected in the cytoplasm and in the nucleus of both SH-SY5Y and HEK293 cells. Interestingly, upon differentiation LINC00478 peptide was only detected in the nucleus. This is suggestive of a differential regulation of LINC00478 peptide upon RA treatment.

5.5.2 LINC01116 plays a regulatory role in neurite elongation of SH-SY5Y cells.

LINC01116 is upregulated within 24h upon induction of differentiation and its levels are sustained for all of the phase 1 (3 days) but its expression is reduced at a later stage of differentiation (phase 2). This suggests that its function is probably associated with the early stages of neuronal differentiation. Knockdown of LINC01116 upon differentiation resulted in a significant reduction of neurite length as well as reduced levels of the noradrenergic marker MOXD1, without affecting cell proliferation. These data suggest that knockdown may cause a delay in differentiation.

Overexpression of LINC01116 WT and double ATG Δ1Δ2 mutant transcript in undifferentiated SH-SY5Y did not significantly affect neurite outgrowth. Therefore,

based on these data, it may be unlikely that function of LINC01116 in the regulation of neurite elongation is dependent upon its translation, or translation is not enough to drive differentiation. The fact that knockdown of LINC01116 impedes neurite elongation but its overexpression in Control does not induce neurite elongation, suggests that LINC01116 is essential but not sufficient for neurite outgrowth. LINC01116 may be a part of the machinery that orchestrates axon outgrowth in SH-SY5Y cells. Therefore, upon its knockdown during differentiation, the neurite elongation is impeded or delayed, but its sole overexpression in Control cells is not enough, given that the other parts of the machinery remain unaltered. Notably, the transfection efficiency of the LINC01116 construct needs to be taken into consideration. Based on the tagging experiments, the FLAG signal appeared in fewer cells, compared to GFP+. The whole transcript overexpression experiment was preliminary, and more work is needed, in order to establish the level of overexpression. Another interesting approach would be to overexpress the LINC01116 FLAG-tagged smORF, in SH-SY5Y cells in which LINC01116 transcript expression has been knocked-down. This might give a more substantial result in the transfected cells, although the transfection efficiency will still be an issue.

This is the first time LINC01116 has been associated with the progression of differentiation. LINC01116 has been previously associated with progression of glioblastoma (GBM), lung cancer, gastric cancer and is thought to promote cell proliferation (Brodie et al., 2017; Su et al., 2019). Therefore, based on the above data it seems plausible that it is differentially regulated in the context of cancer progression compared to differentiation, induced by RA. Interestingly, overexpression of LINC01116 promoted invasion and migration of gastric cancer cells (Su et al., 2019) suggesting that LINC01116 has a potential role in the formation of cell membrane protrusions. That would be consistent with LINC01116 regulating neurite outgrowth in SH-SY5Y cells.

5.5.3 Conclusions

Overall in this chapter I studied the function of two lncRNAs that were identified as translated in Ribo-Seq data (RiboTaper pipeline, Isabel Birds). LINC01116 is upregulated during the early stages of neuronal differentiation of SH-SY5Y cells and it is shown to be involved in neurite outgrowth. The 87aa peptide encoded by

LINC01116 smORF is detected in the cytoplasm and membrane protrusions of SH-SY5Y and HEK293 cells. Finally, I validated the translation of another small peptide, encoded by a 111nt long smORF within LINC00478 lncRNA. LINC00478 is producing a 37aa peptide detected in the nucleus and cytoplasm of undifferentiated SH-SY5Y and HEK293 cells and in the nucleus of differentiated SH-SY5Y cells.

Chapter 6

General Discussion

6.1 General discussion

During the last decade, lncRNAs have been established as key regulators of cell homeostasis, differentiation and organismal development. Nuclear enriched lncRNAs have long been the focus of research, in the belief that the main function of lncRNAs was the regulation of transcription. However, cytoplasmic lncRNAs have now emerged as equally important regulatory molecules with multiple functions discovered within the cytoplasm, such as mRNA stabilisation via miRNA binding (Faghihi et al., 2010; Carrieri et al., 2012) or interaction with RBPs, regulation of translation (Dimartino et al., 2018) and small peptide production (Pueyo and Couso, 2008; Magny et al., 2013; Aspden et al., 2014). Previous ribosome profiling studies have reported several lncRNAs to be engaged by ribosomes (Ingolia et al., 2013). However, an RNA transcript can be sporadically bound by scanning ribosomes at multiple sites without necessarily undergoing active translation (Guttman et al., 2013). Sequencing RNA that co-sediments with polysome fractions has been proposed as a more accurate means of identifying polysome-associated lncRNAs (Floor and Doudna, 2016; Carlevaro-Fita et al., 2016), however co-sedimentation of a lncRNA with polysomes only proves a potential interaction between the two, not the translation of the lncRNAs.

The development of Poly-Ribo-Seq (Aspden et al., 2014) in *D. melanogaster* S2 cells changed the way cytoplasmic lncRNAs are perceived, in relation to the cell's translation machinery. By separately sequencing polysome-associated RNAs and ribosome-footprinted RNA derived from actively translating polysomes, Poly-Ribo-Seq distinguishes between the lncRNAs that interact with polysomes and those that are being actively decoded (Aspden et al., 2014). Hence, it is thought that there are 3 distinct lncRNA populations within the cytoplasm: the cytosolic lncRNAs, that do not interact with the translation machinery, the polysome-associated but not translated lncRNAs and the actively translated lncRNAs.

Several lncRNAs have been shown to play pivotal roles in the transition of neural stem cells from self-renewal to differentiation. So far, the roles of nuclear lncRNAs during neuronal differentiation have been studied (Ng et al., 2013; Lin et al., 2014; Winzi et al., 2018; Carelli et al., 2019) whilst the roles of cytoplasmic lncRNAs in this process still remain elusive. This PhD thesis examines the interactions of cytoplasmic lncRNAs with the translation machinery during neuronal differentiation of human SH-SY5Y cells.

6.2 Neuronal differentiation of SH-SY5Y for 3 days induces a reduction in the level of active translation

Translation during neuronal differentiation is dynamic and subject to tight regulation to accommodate the needs of differentiating cells. Hence, the regulation of gene expression of the translation machinery components (e.g. ribosomal proteins) is very important during early forebrain development (Fujii et al., 2017; Blair et al., 2017; Chau et al., 2018). Polysome profiling of undifferentiated and differentiated SH-SY5Y human neuroblastoma cells demonstrated a significant reduction in the level of translation upon differentiation (Figure 6.1), as measured by the reduction of the polysome/monosome ratio. In contrast, upon differentiation a significant increase in the level of monosomes, was observed, indicating that more RNAs are engaged by monosomes in differentiated cells compared to undifferentiated. Engagement of an mRNA by the 80S does not necessarily correspond to active translation. In some cases, the 80S monosome fraction is thought to contain a large number of inactive ribosomes that are not engaged on mRNAs (Liu and Qian, 2016). However, recent ribosome profiling studies in *S. cerevisiae* and *Rattus norvegicus* indicate that 80S monosomes can be actively translating synaptic mRNAs and mRNAs encoding regulatory proteins (Heyer and Moore, 2016; Biever et al., 2020). It is therefore possible that the increase in the proportion of monosomes, during differentiation, is due to the translation of regulatory proteins, as a result of translation regulation during the transition to a more differentiated phenotype. A monosome could engage on an mRNA, awaiting further signal to start elongation phase of translation or dissociate. A study in yeast has proved that translation can be inhibited after the formation of an 80S complex (Balagopal and Parker, 2011). In mammals, translation can be inhibited or enhanced, depending on the requirements of cell metabolism. One such example is the regulation of translation of mRNAs involved in iron metabolism by the Iron Regulatory Proteins 1 and 2 (IRP1 and IRP2) (Hentze et al., 1987; Casey et al., 1988; Müllner and Kühn, 1988).

Here, transcriptomic analysis of total cytoplasmic and polysome-associated RNA samples, as well as RT-qPCR across polysome gradients, revealed that ribosomal protein transcripts shift away from polysomes to ribosomal subunits and RNP complexes upon neuronal differentiation, corroborating the observation that

reduced translation level coinciding with differentiation. Notably, these results are also consistent with previous reports of transcriptional and translational downregulation of ribosomal protein transcripts and translation initiation factors during the early stages of forebrain development in human and mouse (Blair et al., 2017; Chau et al., 2018).

6.3 Optimisation of Poly-Ribo-Seq protocol for human neuronal cells

Poly-Ribo-Seq was established in *D. melanogaster* S2 cells (Aspden et al., 2014) and this is the first time it has been performed in a human or neuronal cell type. To achieve the required quality of ribosome profiling and RNA-Seq data, further optimisation of certain steps of the protocol was required. The most important technical aspect of ribosome profiling and Poly-Ribo-Seq is the precision of the RNaseI footprinting. This step is crucial in order to confidently determine the position of the ribosomes on RNA transcripts. Given that the exact length of the ribosome protected fragment (RPF) varies across different organisms (Ingolia et al., 2009; Aspden et al., 2014; Hsu et al., 2016), adaptation of the RNaseI footprinting conditions for SH-SY5Y cells was required. After extensive optimisation, the best footprinting conditions for SH-SY5Y were determined: RNase I treatment with the Thermo EN0601 enzyme (0.3U/million cells), at 4^{o}C overnight, in a buffer of lower ionic strength (Tris-HCl pH8 100mM, NaCl 30mM, $MgCl_2$ 10mM). Previous ribosome profiling studies in *A. Thaliana* (Hsu et al., 2016), *C. Reinhardtii* (Chung et al., 2015), *S. Cerevisiae* (Ingolia et al., 2009; Heyer and Moore, 2016; McGlincy and Ingolia, 2017), *S. pombe* (Duncan and Mata, 2014), *D. melanogaster* (Aspden et al., 2014) and cultured mammalian cells (Guo et al., 2010; Ingolia et al., 2013) have all performed RNaseI footprinting in slightly different conditions, depending on the organism. Thus, based on these and on the work presented here, optimisation is recommended for every cell or tissue type to be subject to Poly-Ribo-Seq.

6.4 LncRNAs dynamically interact with polysome complexes upon neuronal differentiation

Dynamic interaction of lncRNAs with polysomes during differentiation has recently been observed during human adipogenesis and cardiomyogenesis (Dallagiovanna et al., 2017; Pereira et al., 2020). In both these systems lncRNAs exhibit differential association with polysomes at early stages of adipocyte and different stages of cardiomyocyte differentiation. More specifically, in quiescent

human adipose derived cells (hASCs), as well as in committed adipocytes, lncRNAs were found associated with polysomes in a similar level to mRNAs, indicating that interaction of lncRNAs with polysomes is not random. Interestingly, a subset of lncRNAs was actively mobilising to polysomes upon the induction and during the early stages of hASCs differentiation (Dallagiovanna et al., 2017). Here, 278 lncRNAs in the cytoplasm (Figure 6.1) and 319 polysome-associated lncRNAs were differentially expressed during early neuronal differentiation of SH-SY5Y. Hence, it is plausible that lncRNAs are involved in the regulation of neuronal differentiation, as well. Most of the differentially expressed lncRNA transcripts are in common between Total cytoplasm and polysome populations. This suggests that the association of these lncRNAs with polysomes is not an arbitrary event and these lncRNAs are either being translated or regulate the translation of the mRNAs with which they interact. Moreover, comparison of the differentially expressed genes between total cytoplasm and polysome fractions, in Control and in differentiated cells revealed that certain lncRNAs are significantly enriched or depleted from polysomes. This is indicative of the dynamic nature of polysome-association during neuronal differentiation and highlights the importance of the interaction of lncRNAs with the translation machinery for neurogenesis. In support of this notion, a recent study revealed that polysome-associated lncRNAs appear to be more important for cardiomyogenic commitment, compared to cytosolic lncRNAs. A larger proportion of polysome-associated lncRNAs were differentially expressed during the transition from mesodermal to cardiac progenitors, compared to cytosolic lncRNAs, suggesting a regulatory role of the lncRNA-polysome interaction (Pereira et al., 2020).

Expression analysis of selected target lncRNAs in nuclear and cytoplasmic fractions of SH-SY5Y cells revealed that polysome-associated lncRNAs are enriched in the cytoplasm, consistent with a previous study showing that the lncRNAs that interact with ribosomes tend to be more enriched in the cytoplasm (Zeng and Hamada, 2018). Furthermore, the interaction of lncRNAs with polysomes was validated for 2 intergenic (LINC01116 and LINC02143) and 1 anti-sense (DLGAP1-AS1) lncRNA transcripts. The association of these lncRNAs with polysomes is suggestive of a potential function in translation regulation or mRNA stabilisation. Anti-sense lncRNAs have traditionally been associated with gene expression regulation in the nucleus, at the pre-transcriptional and transcriptional level, but a number of studies have reported

that cytoplasmic anti-sense lncRNAs are key regulators of gene expression in the cytoplasm. Acting as molecular sponges, cytoplasmic lncRNAs such as MACC1-AS1 protect their sense mRNA from miRNA degradation (X. Zhang et al., 2019). BACE1-AS base-pairs with BACE mRNA and by stabilising it, promotes its translation (Faghihi et al., 2008). UCHL1-AS binds to UCHL1 mRNA and facilitates its translation through an embedded SINEB2 repeat (Carrieri et al., 2012), essential for ribosome recruitment (Yao et al., 2015). This mechanism is also employed by other anti-sense lncRNAs (Carrieri et al., 2012), suggesting a novel potential mechanism of function for anti-sense lncRNA transcripts. DLGAP1-AS1, studied here, is anti-sense to DLGAP1 mRNA, which encodes a member of Disks large proteins that act as a scaffold for post-synaptic density (Rasmussen et al., 2017). DLGAP1 mRNA did not exhibit differential transcription upon differentiation but it would be interesting to examine DLGAP1 mRNA distribution across the gradient as well as DLGAP1 protein levels in Control and differentiated cells. Moreover, to determine whether DLGAP1-AS is responsible for DLGAP1 mRNA stability, RNAi of DLGAP1-AS, followed by assessment of DLGAP1 transcription and translation levels would be a plausible approach.

6.5 Actively translated small ORFs, within lncRNAs, were identified by Poly-Ribo-Seq

RiboSeqR analysis of lncRNA transcripts revealed the active translation of lncRNAs. Further Ribo-Seq data analysis (by Isabel Birds) identified 28 lncRNA smORFs in Control and 23 in differentiated SH-SY5Y, to be actively translated, exhibiting strong framing across the ORF and similar translation efficiencies (TE) to protein-coding transcripts. 6 of them were detected in both conditions. Four of the translated lncRNAs identified by Poly-Ribo-Seq, namely HAND2-AS1, TUG1, MCPH1-AS1 and CRNDE have been previously reported to encode small peptides conserved from rodents to human. HAND2-AS1 (van Heesch et al., 2019) is involved in heart function and TUG1 (Lewandowski et al., 2019) is involved in male fertility. The lncRNA MCPH1-AS1 has previously been found to be translated but at a different ORF to the one I identified by Poly-Ribo-Seq (Chong et al., 2020). I identified a 84 aa peptide translated from CRNDE lncRNA, which was previously reported as translated (CRNDEP), and found to be conserved in primates and regulates oxygen metabolism

(Szafron et al., 2015). The translated smORFs are generally found in intergenic and anti-sense lncRNA transcripts, which is expected as these are the most abundant categories of lncRNAs (Gencode release 29, GRch38). Interestingly, analysis of mass spectrometry data from SH-SY5Y cells (Brenig et al., 2020) has corroborated the translation of 18% of the smORFs I identified by Poly-Ribo-Seq. The peptides encoded by those smORFs have a median length of 56aa in Control and 64aa in differentiated cells.

6.6 LncRNAs LINC01116 and LINC00478 encode small peptides which localise in the nucleus and cytoplasm of SH-SY5Y cells.

FLAG-tagging assay validated the production of 2 novel small peptides encoded by LINC00478 and LINC01116 transcripts both in SH-SY5Y and HEK293 cells. LINC00478 encodes a 37aa long peptide, which was detected in the nucleus and the cytoplasm in undifferentiated SH-SY5Y and HEK293 but only in the nucleus of differentiated SH-SY5Y (Figure 6.1), suggesting differential regulation of LINC00478 peptide upon RA treatment. The amino acid sequence of this peptide has been found to be conserved between mouse and humans (Isabel Birds). Interestingly, LINC00478 lncRNA transcript is highly expressed in the nucleus accumbens area of the human brain (Lonsdale et al., 2013) as well as in several regions of the developing human brain, such as basal ganglia, forebrain and diencephalon (Lindsay et al., 2016).

LINC01116 encodes a 87aa long peptide, localized in the cytoplasm and neuritic protrusions of undifferentiated SH-SY5Y cells (Figure 6.1). Expression of the peptide from a construct containing the 5’ UTR and coding sequence, but not from a construct in which the apparent ATG translation start-site was mutated, confirmed that the translation of the smORF is a genuine event and not an artifact. Notably, very few FLAG-positive cells were identified in both SH-SY5Y and HEK293 cells. A more thorough examination of the LINC01116 locus revealed that the transcript ENST00000295549.9, which was selected based on the existence of a high Ribo-Seq peak, may lack 5’ sequence elements contributing to its efficient transcription and even its stability. However, its expression should be driven by the plasmid’s CMV promoter. Therefore, follow up experiments should include the re-cloning of this lncRNA reporter including this TSSs upstream of LINC01116 5’-UTR.

6.7 LINC01116 plays a regulatory role in neurite elongation of SH-SY5Y cells.

LINC01116 was selected as a target to further investigate, because it was upregulated during differentiation, contained a translated smORF and because it has been previously associated with progression of glioblastoma (GBM), lung cancer, gastric cancer and is thought to promote cell proliferation (Brodie et al., 2017; Su et al., 2019; H. Wang et al., 2020). LINC01116 is expressed in the developing and adult brain and spinal cord amongst other tissues (Lonsdale et al., 2013; Lindsay et al., 2016), but this is the first time it is being studied in the context of neuronal differentiation. Based on the fact that LINC01116 is upregulated within the first 24h after induction of differentiation and its expression decreases at a later stage (Figure 5.13), its function is probably associated with the early stages of neuronal differentiation. Knockdown of LINC01116 upon differentiation resulted in a significant decrease of neurite length as well as a reduction in levels of the noradrenergic marker MOXD1, without affecting cell proliferation. These data suggest that knockdown may cause a delay in differentiation. Interestingly, a recent study reports that overexpression of LINC01116 promotes invasion and migration of gastric cancer cells (Su et al., 2019), suggesting that LINC01116 has a potential role in the formation of cell membrane protrusions. That would be consistent with LINC01116 regulating neurite outgrowth in SH-SY5Y cells.

Expression of LINC01116 (WT and ATG mutant) transcript did not result in any significant effect in neurite outgrowth. Therefore, LINC01116 function in the regulation of neurite elongation may not be dependent upon its translation, or translation is not enough to drive differentiation. The results of those experiments were preliminary and follow-up experiments would be necessary to confidently determine if the transcript or the peptide is involved in the regulation of differentiation. The low number of cells overexpressing LINC01116 transcript does not allow to draw a safe conclusion. However, the fact that knockdown of LINC01116 transcript impedes neurite elongation but its transfection in Control does not promote neurite elongation, suggests that LINC01116 is essential but not sufficient for neurite outgrowth. LINC01116 may be a part of the machinery that regulates axon outgrowth in SH-SY5Y cells (Figure 6.1). Therefore, upon its knockdown during differentiation, the neurite elongation is impeded or delayed, but its sole overexpression in Control cells is not enough to rescue the phenotype.

6.8 Future experiments and perspective

The results discussed in this thesis provide evidence that cytoplasmic lncRNAs dynamically interact with the translation machinery, encode small peptides and have regulatory roles during the early stages of human neuronal differentiation. The data acquired by Poly-Ribo-Seq, as well as the evidence of smORF translation from the FLAG-tagging assays, and LINC01116 involvement in the regulation of differentiation can be the foundation to a number of follow-up experiments towards our further understanding of the role of cytoplasmic lncRNAs interaction with the translation machinery. The approaches that could be undertaken are listed below:

Translation of lncRNAs: Given that Poly-Ribo-Seq was performed at one time-point in early differentiation, lncRNAs expressed during the specific time-window of 3 days were detected. Therefore, lncRNAs with a role in an earlier or later stage of differentiation are missed. In the future, Poly-Ribo-Seq could be performed in a number of different time-points during SH-SY5Y differentiation.

The data analysis in chapter 4 led to the selection of target lncRNAs based on their translation efficiency in Control or differentiated SH-SY5Y cells. Transfection of FLAG-tagged smORFs, followed by immunocytochemistry in Chapter 5 showed that LINC01116 and LINC00478 are translated. However, I was not able to detect either peptides by western blot, likely due to the small size of the peptides or the low level of transfection and expression, given low number of cells FLAG was detected in. Alternatively, to confirm the translation of the smORF, a FLAG pulldown approach, followed by mass spectrometry (Gerace and Moazed, 2015) could be undertaken to further validate the translation of the lncRNA smORFs. Apart from LINC01116 and LINC00478, FLAG-tagged I generated constructs for 4 other lncRNA smORFs (MCPH1-AS1, PSMA3-AS (2 smORFs) and AL162386.2) and these can be expressed using transfection assay to validate the translation of those smORFs. Previously reported translated lncRNA smORFs have functions that are conserved from *D. melanogaster* to human (Magny et al., 2013; Lewandowski et al., 2019; VanHeesch et al., 2019) Therefore, a conservation analysis of the translated smORFs identified here by Poly-Ribo-Seq, is an interesting thing to do, that will also provide insight to whether the peptides encoded by lncRNAs could be functional or not.

Role of LINC01116 in neuronal differentiation: Experiments in Chapter 4 and 5 confirm that LINC01116 is a cytoplasmic lncRNA that interacts with polysomes and encodes a small peptide. However, the exact localisation of LINC01116 lncRNA in the cytoplasm is yet to be determined. Fluorescent *In Situ* Hybidisation (FISH) or single molecule FISH (smFISH) could be utilised to detect LINC01116 expression in the cytoplasm and could be combined with immunofluorescent labelling of ribosomal proteins, to visualise the association of LINC01116 transcript with polysomes in Control and differentiated SH-SY5Y.

LINC01116 transcript is upregulated during the early stages of neuronal differentiation of SH-SY5Y and is shown to have a regulatory role in neurite elongation. These findings are in line with the fact that LINC01116 is moderately expressed in the developing human forebrain and highly expressed in the developing human midbrain and spinal cord (Lindsay et al., 2016). LINC01116 regulatory role is not fully characterised and it is uncertain whether it is mediated by the smORF translation or by the whole transcript. Preliminary findings in Chapter 5 suggest that LINC01116 function in the regulation of neurite elongation may not be dependent upon its translation, or translation is not enough to drive differentiation. However, given the low transfection efficiency of SH-SY5Y, overexpression of the LINC01116 was only indirectly assessed via the co-overexpression of GFP. Therefore, it is not clear to what extent LINC01116 was indeed expressed. To overcome this issue, LINC01116 expression levels can assessed by RT-qPCR, in parallel with the immunofluorescent detection of GFP, and the expression of neuronal markers can be detected by RT-qPCR in the cells that express LINC01116, compared to control. Alternatively, LINC01116 overexpression can be driven by a construct which only contains the ORF sequence, without its 5'-UTR, which contains a GC-rich element.

Finally, given the promising finding that LINC01116 is a potential regulator of neuronal differentiation, it would be intriguing to investigate potential interactors. This could be achieved by performing RNA pulldown specific for LINC01116, followed by mass spectrometry (RAP-MS), to identify RNA binding proteins that interact with LINC01116 and potentially co-regulate neuronal differentiation. Moreover, to discover other lncRNAs or mRNAs that may interact with LINC01116, RAP-Seq could be performed. Depending on the results of the above approaches, it may be interesting to further investigate the role of LINC01116 at different time windows during neuronal

differentiation of human embryonic stem cell (hESC), as this would be a more representative model of human neurogenesis.

6.9 Conclusions

The work presented in this thesis highlights the importance of the interactions of cytoplasmic lncRNAs with the translation machinery during neuronal differentiation of SH-SY5Y cells. Poly-Ribo-Seq was optimised and performed for the first time in a model of human neuronal differentiation (SH-SY5Y cells) (Figure 6.1). Data analysis revealed 278 cytoplasmic lncRNAs differentially expressed upon neuronal differentiation and several lncRNAs to be dynamically associated with actively translating polysomes. Moreover, 45 smORFs were detected to be actively translated, exhibiting clear triplet periodicity. The production of small peptides was validated for 2 smORFs and further investigation of one of them (LINC01116) revealed a regulatory role in the early stages of neuronal differentiation. Further work is needed to fully elucidate the role of LINC01116 in neurogenesis and determine the roles of the other translated smORFs identified in this study.

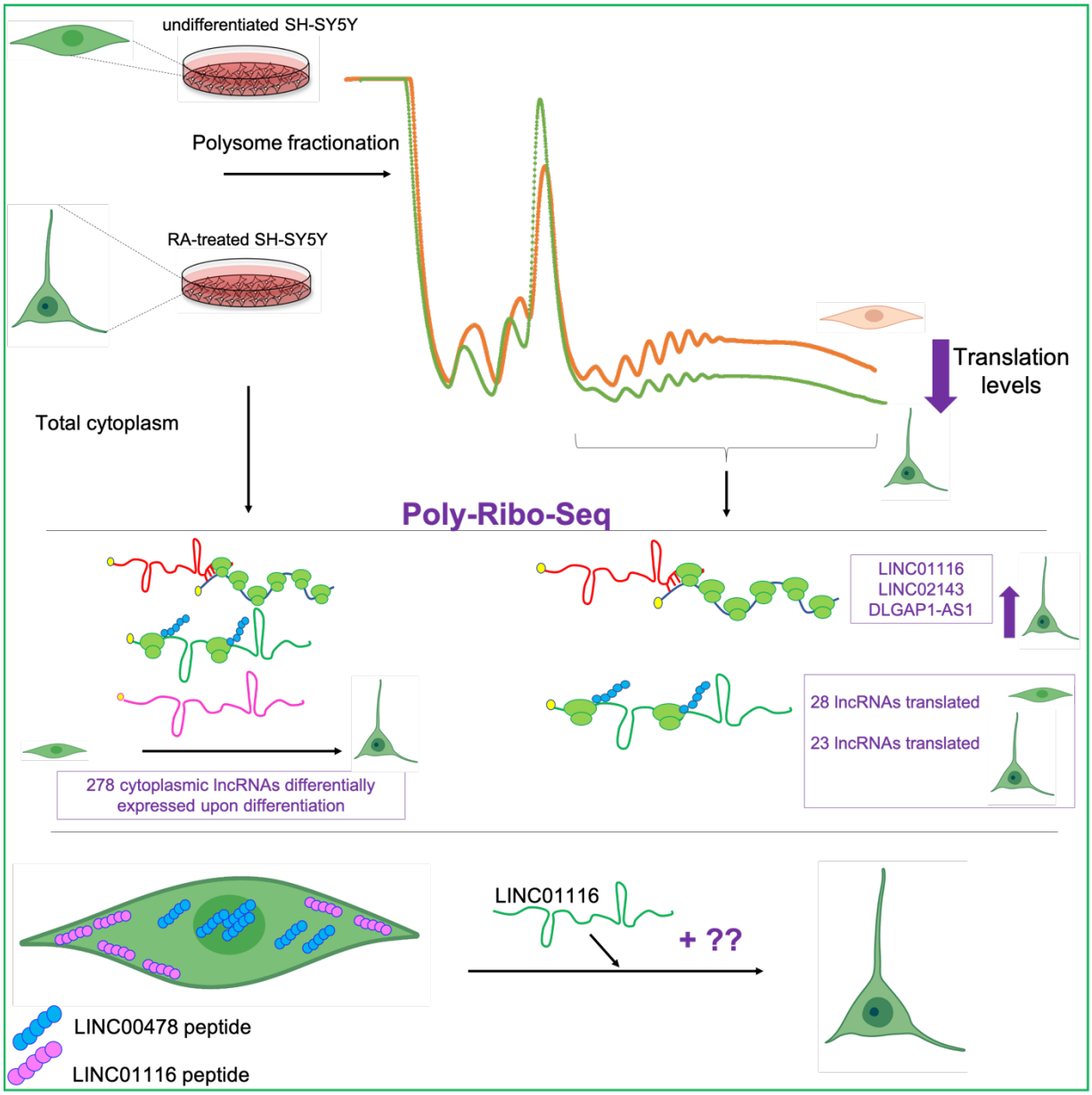


Figure 6.1: Overview of the results presented in this thesis.

Appendix-I

Code and parameters used for RNA-Seq and Ribo-Seq data analysis pipeline performed on ribogalaxy and Galaxy platform: www.ribogalaxy.ucc.ie, www.usegalaxy.org (Michel et al., 2016; Afgan et al., 2018)

- **Pre-processing: Quality control, adapter trimming and transcriptome mapping pipeline parameters:**

1. FastQC – No contaminant list, no submodule and limit specifying file

2. Cutadapt -full parameter list:

Add 3' adapters only
Max error rate: **0.1** (default)
Match times: **1**
Minimum overlap length: **3** (default)
Match read wildcards: **No**
Do not match adapter wildcards: No
Output filtering options (Set filters/No filters)
Discard trimmed reads: No
Minimum length: 25 (default)
Maximum length: 0 (default)
Additional output options: write in different files the rest of read, wildcard file, too short reads, untrimmed reads.
Additional modifications to the reads: No (default)

3. Filter by quality (performed on www.usegalaxy.org)

Quality cutoff value: 20
Percentage of bases that should have quality >/= of the cutoff: 90

4. rRNA/tRNA removal and transcriptome alignment using Bowtie -full parameter list:

Skip the first n reads (-s): 0
Only align the first n reads (-u): -1 for off
Trim n bases from the high-quality left end of the read (-5): 0
Trim n bases from the low-quality right end of the read (-3): 1
Max number of mismatches permitted in the seed (-n): 2 (default) but we can change i
Maximum permitted total of quality values at mismatched read positions (-e): 70(default)
Seed length (-l): 25 (default) minimum is 5
Whether or not to round to the nearest 10 and saturating at 30 (--nomaqround): Round to nearest 10 (default)
Number of mismatches for SOAP-like alignment policy (-v): 3 (or -1 for MAQ-like alignment policy).
**SOAP: short oligonucleotide alignment program- efficient gapped and ungapped alignment of short oligonucleotides onto reference sequences
**MAQ: does not support gapped alignment for single-end reads

Choose whether or not to align against the reverse-complement reference strand (--norc): no
Whether or not to try as hard as possible to find valid alignments when they exist (-y): no
Report up to n valid alignments per read (-k): 1 (default)

Whether or not to report all valid alignments per read (-a): do not report all valid alignments

Suppress all alignments for a read if more than n reportable alignments exist (-m):-1

Write all reads with a number of valid alignments exceeding the limit set with the -m option to a file (--max): no

Write all reads that could not be aligned to a file (--un): yes

Whether or not to make Bowtie guarantee that reported singleton alignments are 'best' in terms of stratum and in terms of the quality values at the mismatched positions (--best): do not use best (default)

Maximum number of backtracks permitted when aligning a read (--maxbts): 125 (default)

Override the off rate of the index to n (-o): -1 (default)

Seed for pseudo-random number generator (--seed): -1 (default)

Suppress the header in the output SAM file: no

- **PCA analysis**

```
source("https://bioconductor.org/biocLite.R")
biocLite("preprocessCore")
install.packages(c("devtools"))
biocLite(c("Biobase"))
library(devtools)
library(Biobase)
library(preprocessCore)

setwd("/home/katerina/Desktop")
data1 <- read.csv("RPKM.csv", header = TRUE, sep = ",")
library(tibble)
data2 <- na.omit(data1)
#To grab rows which pass a filter
data2 <-
data1[which(data1$R_Control_polysome_2>=1)&(data1$R_Control_polysome_3
>=1)&(data1$R_Control_polysome_4>=1)&(data1$R_Control_total_2>=1)&(data
1$R_Control_total_3>=1)&(data1$R_Control_total_4>=1)&(data1$R_RA_polyso
me_2>=1)&(data1$R_RA_polysome_3>=1)&(data1$R_RA_polysome_4>=1)&(d
ata1$R_RA_total_2>=1)&(data1$R_RA_total_3>=1)&(data1$R_RA_total_4>=1),]
names(data2)[2] <- "Con-polysome1"
names(data2)[3] <- "Con-polysome2"
names(data2)[4] <- "Con-polysome3"
names(data2)[5] <- "Con-total1"
names(data2)[6] <- "Con-total2"
names(data2)[7] <- "Con-total3"
names(data2)[8] <- "RA-polysome1"
names(data2)[9] <- "RA-polysome2"
names(data2)[10] <- "RA-polysome3"
```

```
names(data2)[11] <- "RA-total1"
names(data2)[12] <- "RA-total2"
names(data2)[13] <- "RA-total3"

head(data2)
plot(data2)
d.new <- data2
d.new[, 2:13] <- log(data2[2:13], 2)
d.new
rownames(d.new) <- d.new[,1]
data_mat <- data.matrix(d.new[,-1])
head(data_mat)
data_norm <- normalize.quantiles(data_mat, copy = TRUE)
##PCA analysis##
my.PCA <- prcomp(t(data_mat), scale=FALSE)
plot(my.PCA$x[,1], my.PCA$x[,2])
## make a scree plot-which is a graphical representation of the percentages of
variation each PC accounts for (for a good PCA PC1 and PC2 should account for
~90% of variation)##
pca.var <- my.PCA$sdev^2
pca.var.per <- round(pca.var/sum(pca.var)*100, 1)
barplot(pca.var.per, main="Scree Plot", xlab="Principal Component",
ylab="Percent Variation")
## now make a plot that shows the PCs and the variation:
library(ggplot2)
pca.data <- data.frame(Sample=rownames(my.PCA$x),
                       X=my.PCA$x[,1],
                       Y=my.PCA$x[,2])
pca.data
Sample <- pca.data$Sample
ggplot(data=pca.data, aes(x=X, y=Y, label=Sample)) +
  geom_text() +
  xlab(paste("PC1 - ", pca.var.per[1], "%", sep="")) +
  ylab(paste("PC2 - ", pca.var.per[2], "%", sep="")) +
  geom_point(size=1, alpha=3)+
  geom_text(size=4, alpha=3)+
  geom_point(colour="purple")+
  geom_point(aes(color = Sample))+
theme_classic() +
ggtitle("RNA-Seq PCA analysis")

#### alternative plots###
ggplot(data=pca.data, aes(x=X, y=Y, label=Sample)) +
  geom_text() +
  xlab(paste("PC1 - ", pca.var.per[1], "%", sep="")) +
  ylab(paste("PC2 - ", pca.var.per[2], "%", sep="")) +
  geom_point(size=1, alpha=3)+
  geom_text(size=4, alpha=3)+
  geom_point(aes(color = Sample))+
  theme_classic() +
```

```
  ggtitle("RNA-Seq PCA analysis")+
  theme(legend.position='none')

#####################################
ggplot(data=pca.data, aes(x=X, y=Y, label=Sample)) +
  xlab(paste("PC1 - ", pca.var.per[1], "%", sep="")) +
  ylab(paste("PC2 - ", pca.var.per[2], "%", sep="")) +
  geom_point(size=1, alpha=3)+
  geom_point(aes(color = Sample))+
  theme_classic() +
  ggtitle("RNA-Seq PCA analysis")

#### change font size on axis####
plot2 + theme(axis.text.x = element_text(face="bold", color="black",
                                 size=8, angle=90),
          axis.text.y = element_text(face="bold", color="black",
                                 size=7, angle=0),axis.title.x = element_blank(),
          axis.title.y = element_text(color="black", size=7, face="bold"))
```

- **RiboSeqR pipeline** (Hardcastle, 2014)

```
source("https://bioconductor.org/biocLite.R")
#biocLite("riboSeqR")
#biocLite("baySeqR")
#biocLite("Rsamtools")
library(riboSeqR)
library(baySeq)
library(Rsamtools)


humanFasta <- c("/nobackup/bsad/NGS3/UCSC.fasta")

fastaCDS <- findCDS(fastaFile = humanFasta,
                startCodon = c("ATG"),
                stopCodon = c("TAG", "TAA", "TGA"))

#reading the ribosomal files--creates the riboDat variable
ribofileCR <- c("/nobackup/bsad/NGS3/FP/CON_FP.bam",
"/nobackup/bsad/NGS3/FP/RA_FP.bam")

rnafileCR <- c("/nobackup/bsad/NGS3/mRNA/Control_polysome_Sorted.bam",
"/nobackup/bsad/NGS3/mRNA/RA_polysome_Sorted$

riboDat <-readRibodata (ribofileCR, rnafileCR, replicates = c("CON","RA"),
zeroIndexed = FALSE)

#create the triplet periodicity plot
fCs <- frameCounting(riboDat, fastaCDS, lengths=25:35)
write.table(fCs, file="fCs.tsv.", sep = "\t", quote=FALSE)
fS <- readingFrame(rC = fCs, lengths=25:35)
```

```
write.table(fS, file="fS.tsv.", sep = "\t", quote=FALSE)
plotFS(fS, legend.text = c("Frame 0", "Frame 1", "Frame 2"), colour = c("red",
"green", "blue"))

#select best framing and create the metagene analysis plot
ffCs <- filterHits(fCs, lengths = c(31, 32), frames = list(1, 2),
                   hitMean = 50, unqhitMean = 10, fS = fS)
plotCDS(coordinates = ffCs@CDS, riboDat = riboDat, min5p=-100, max5p=100,
min3p=-100, max3p=100, main=c("Metagene Analysis all frames"), lengths = 31, 33)
```

Appendix-II

Supplementary tables to table 4.1 and supplementary figures to Figure 4.10

Table 1 (supplementary to table 4.1): Summary of the number of reads for biological replicate 2 after each step of the pipeline.

sample	number of raw reads	% of reads after Filtering	number of high-quality reads after rRNA and tRNA removal	% of usable high-quality reads mapped to transcriptome	% of high-quality usable reads mapped to lncRNAs	% of high-quality usable reads mapped to protein-coding transcripts	% of protein-coding mapped reads that map to CDS	% of protein-coding mapped reads that mapped to 3'-UTR	% of protein-coding mapped reads that mapped to 5'-UTR
Control FP	172,613,271	86%	81,049,764	92%	3%	89%	95%	15%	20%
Control polysome	251,494,676	87%	124,695,939	65%	2%	62%	54%	45%	14%
Control total	84,553,605	79%	62,957,991	76%	2%	71%	54%	44%	12%
RA FP	206,766,558	76%	79,578,752	73%	3%	70%	92%	16%	22%
RA polysome	63,950,891	80%	30,269,243	56%	2%	52%	50%	50%	15%
RA total	77,512,549	81%	56,862,882	77%	3%	71%	51%	48%	13%

Table 2 (supplementary to table 4.1): Summary of the number of reads for biological replicate 2 after each step of the pipeline

sample	number of raw reads	% of reads after Filtering	number of high-quality reads after rRNA and tRNA removal	% of usable high-quality usable reads mapped to transcriptome	% of high-quality usable reads mapped to lncRNAs	% of high-quality usable reads mapped to protein-coding transcripts	% of protein-coding mapped reads that map to CDS	% of protein-coding mapped reads that mapped to 3'-UTR	% of protein-coding mapped reads that mapped to 5'-UTR
Control FP	123,263,935	85%	64,736,852	74%	2%	72%	95%	14%	19%
Control polysome	74,521,025	84%	35,389,893	79%	3%	76%	55%	44%	14%
Control total	54,946,328	82%	38,652,651	80%	2%	76%	50%	47%	12%
RA FP	189,414,183	80%	89,440,108	87%	2%	84%	95%	13%	19%
RA polysome	69,816,758	82%	47,086,782	82%	3%	79%	50%	49%	14%
RA total	59,056,664	83%	44,314,647	83%	2%	79%	54%	44%	12%

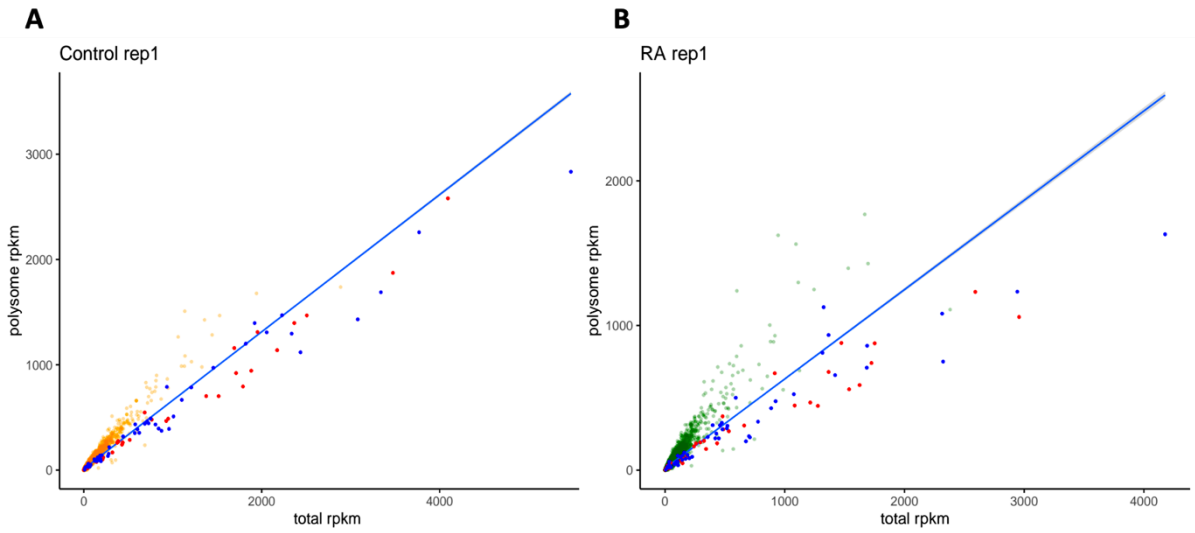


Figure 1 (supplementary to Figure 4.10): Ribosomal protein-coding genes are depleted from polysomes upon differentiation (replicate 1). Scatter plots of total/polysome in (A) Control (orange dots) and (B) RA treated (green dots) cells of replicate 3. Ribosomal protein genes in RA treated cells shift from polysomes to total. Large ribosomal proteins (RPL) are labelled as blue dots and small ribosomal proteins (RPS) as red dots. Blue line is the linear model that fitted the data.

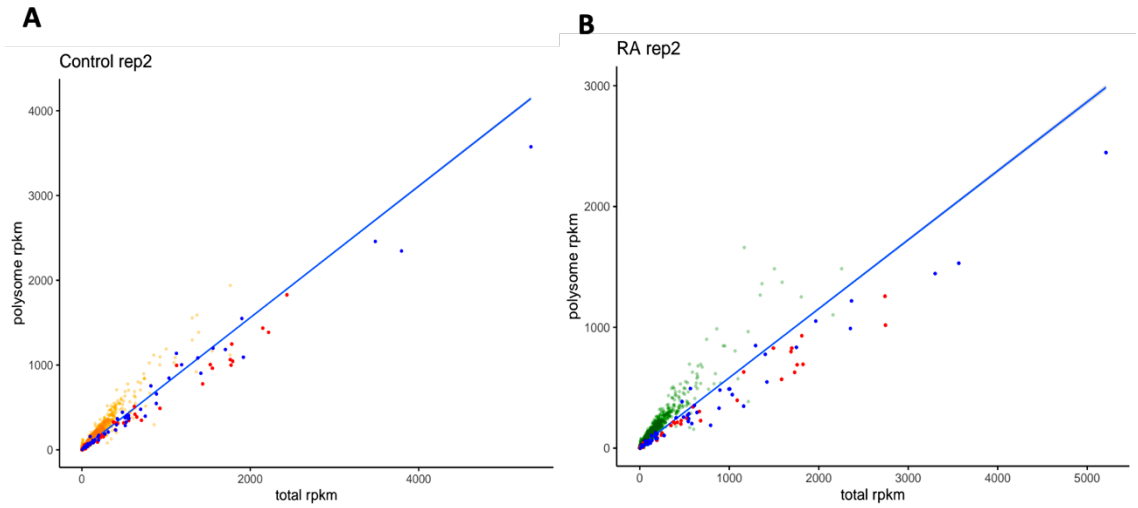


Figure 2 (supplementary to Figure 4.10): Ribosomal protein-coding genes are depleted from polysomes upon differentiation (replicate 2). Scatter plots of total/polysome in (A) Control (orange dots) and (B) RA treated (green dots) cells of replicate 3. Ribosomal protein genes in RA treated cells shift from polysomes to total. Large ribosomal proteins (RPL) are labelled as blue dots and small ribosomal proteins (RPS) as red dots. Blue line is the linear model that fitted the data.

Appendix-III

Supplementary figures to Figures 4.24, 4.25 and 4.26

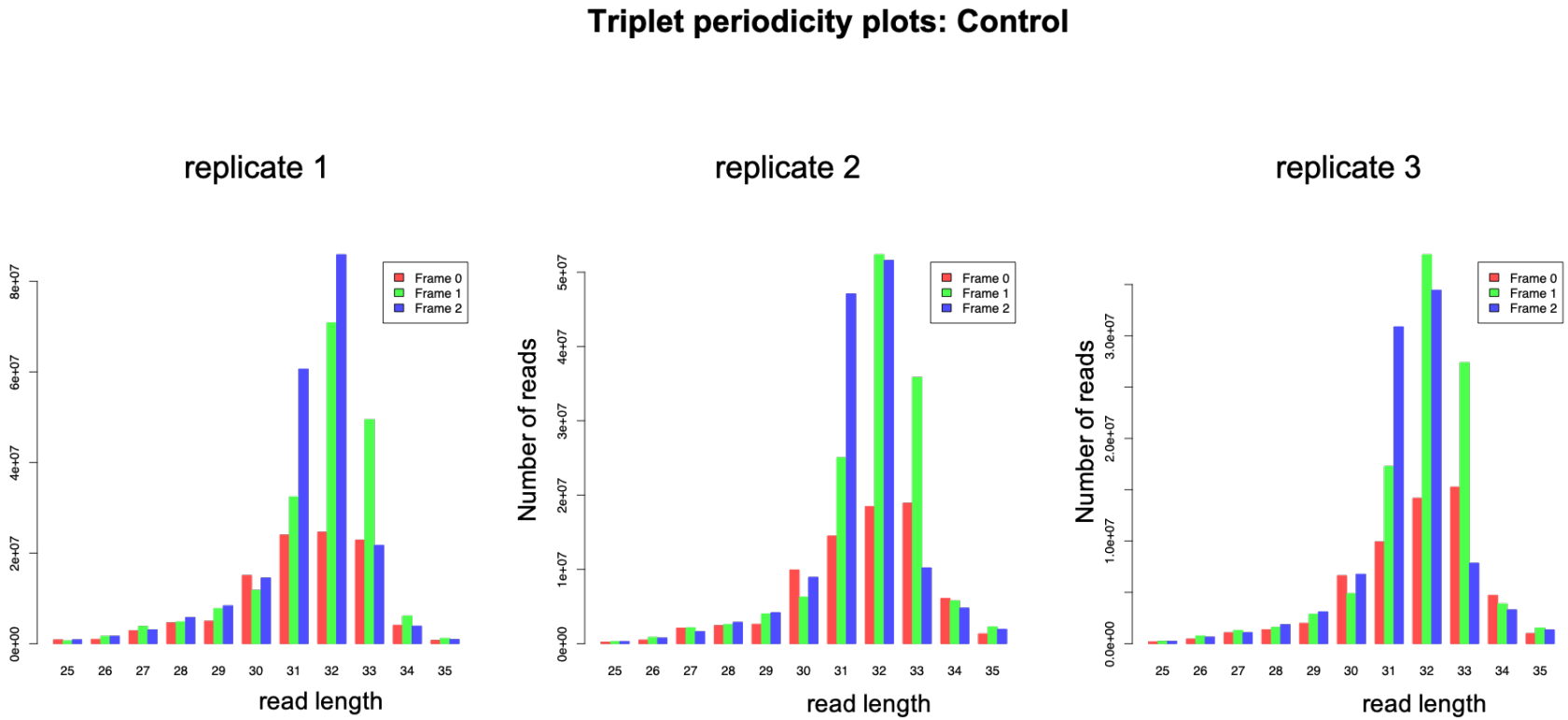


Figure 3 (supplementary to Figure 4.24): Ribo-Seq reads from Control samples display triplet periodicity in all biological replicates. Framing was calculated for a read length range from 25-35nt and the majority of the reads with strong framing were 31nt to 33nt long.

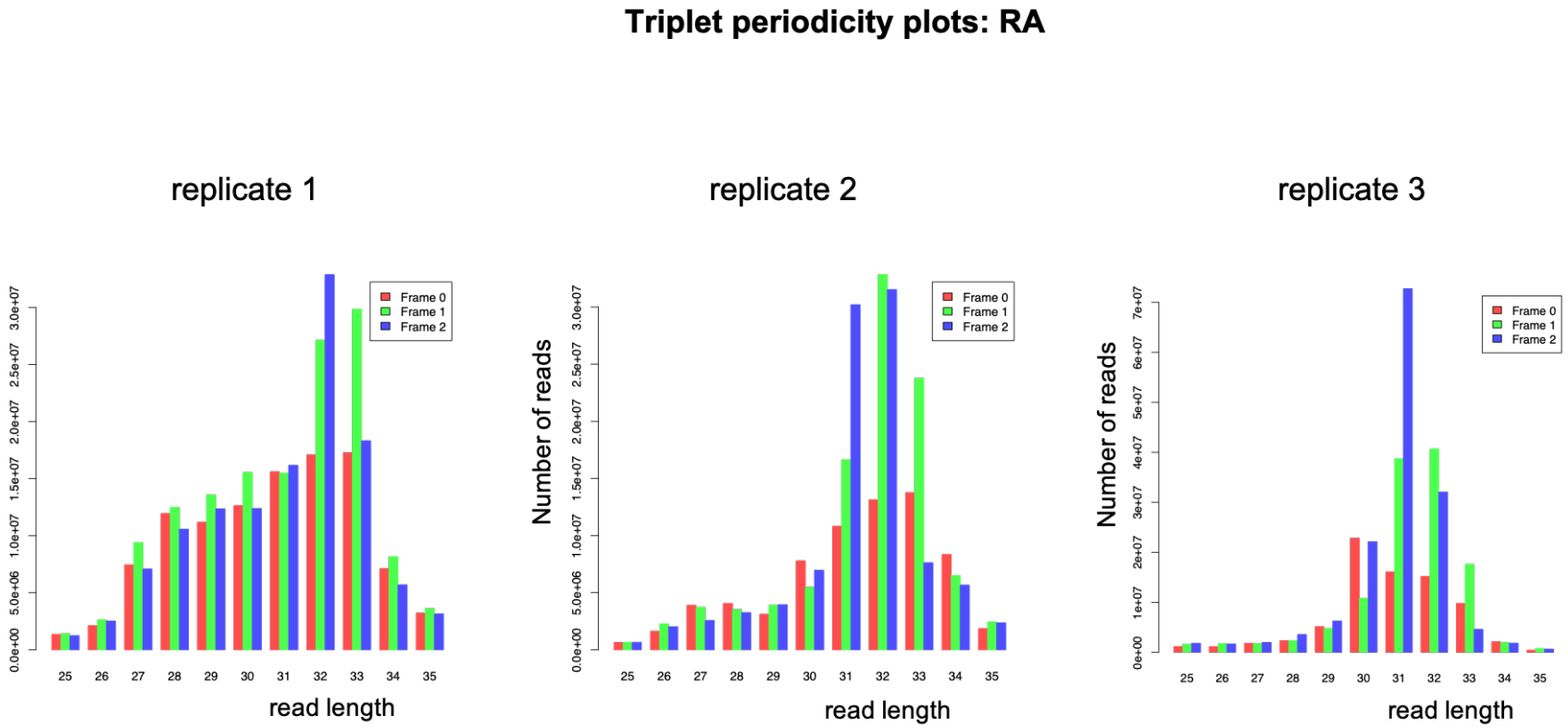


Figure 4 (supplementary to Figure 4.24): Ribo-Seq reads from RA samples display triplet periodicity in all biological replicates. Framing was calculated for a read length range from 25-35nt and the majority of the reads with strong framing were 31nt to 33nt long.

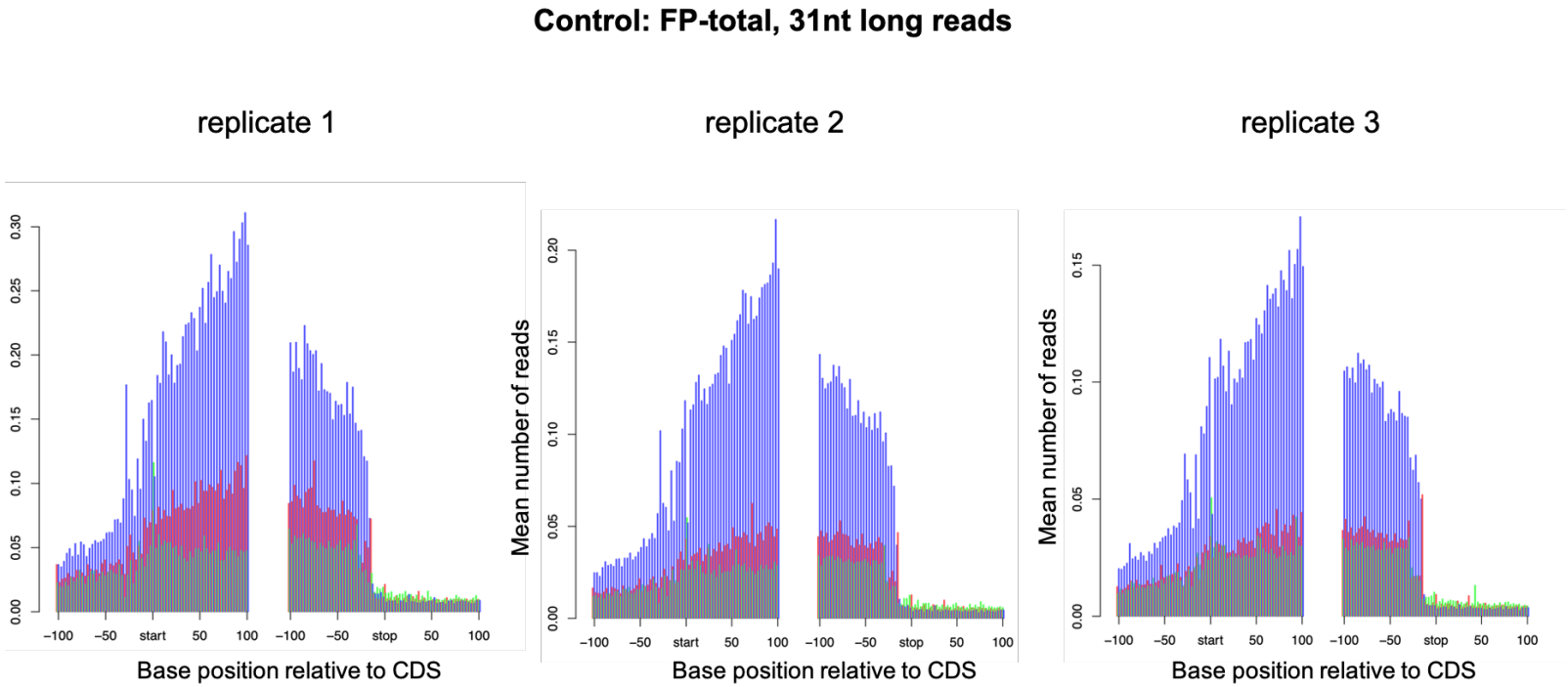


Figure 5 (supplementary to Figure 4.24): Footprint reads map to coding sequences. Metagene analysis of Ribo-Seq coupled with total cytoplasmic RNA-Seq reads of Control samples for 31nt long reads lengths that showed strong periodicity. Reads are globally mapped across transcripts and assigned to 5'-UTR, ORF and 3'-UTR regions (100nt upstream and downstream of start and stop codons plotted). Most of the reads map to the coding sequence (ORF) and fewer reads map to 5' and 3'-UTRs.

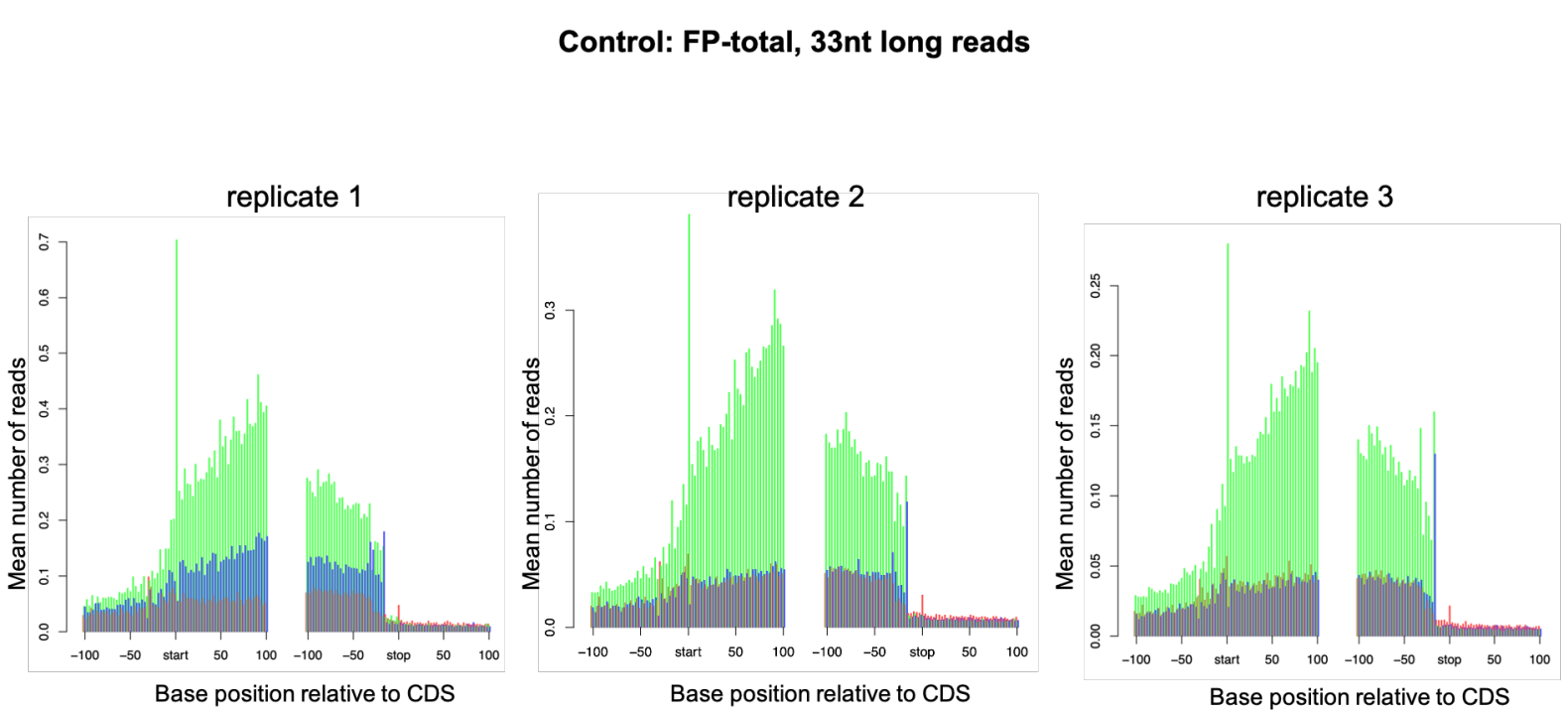


Figure 6 (supplementary to Figure 4.24): Footprint reads map to coding sequences. Metagene analysis of Ribo-Seq coupled with total cytoplasmic RNA-Seq reads of Control samples for 33nt long reads lengths that showed strong periodicity. Reads are globally mapped across transcripts and assigned to 5'-UTR, ORF and 3'-UTR regions (100nt upstream and downstream of start and stop codons plotted). Most of the reads map to the coding sequence (ORF) and fewer reads map to 5' and 3'-UTRs.

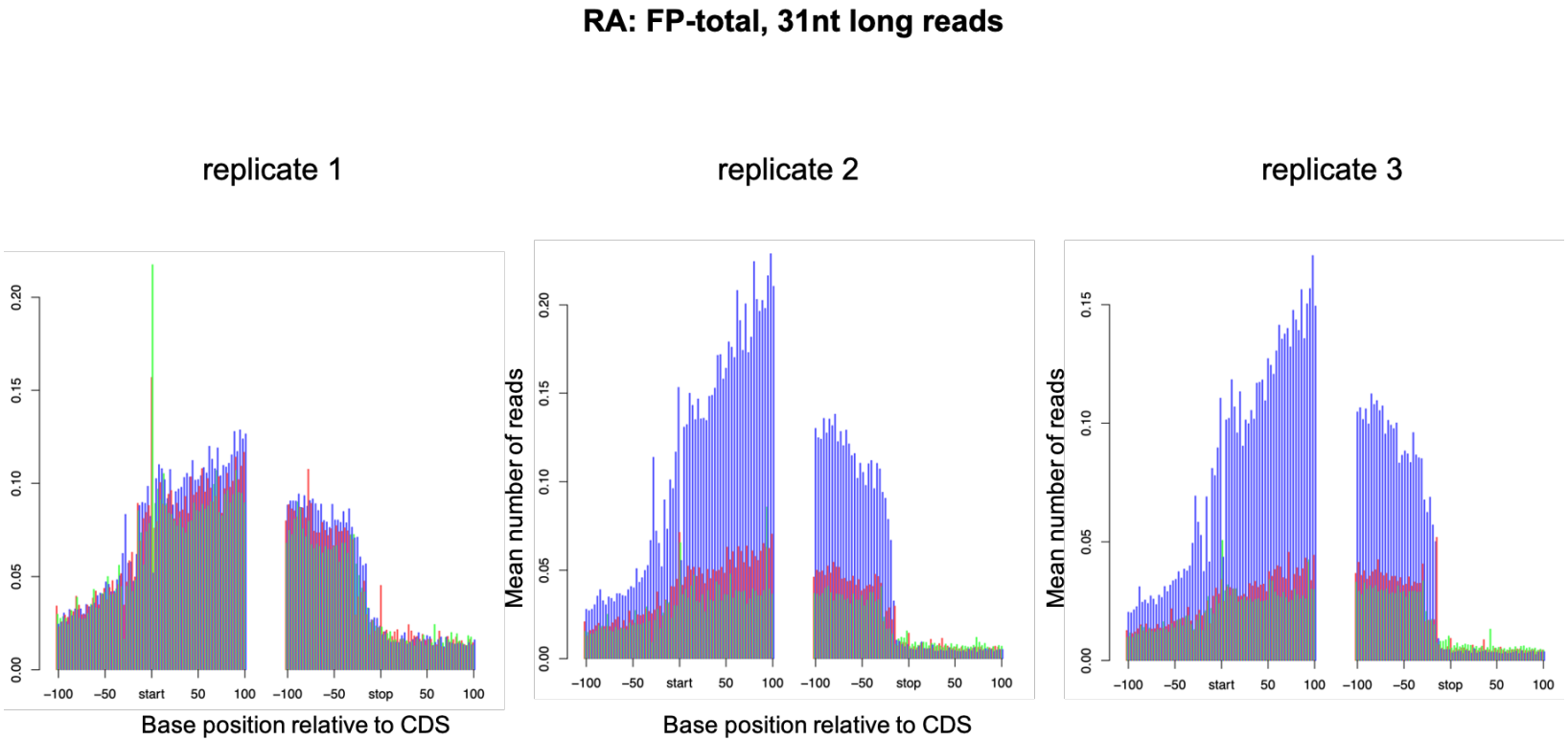


Figure 7 (supplementary to Figure 4.24): Footprint reads map to coding sequences. Metagene analysis of Ribo-Seq coupled with total cytoplasmic RNA-Seq reads of RA samples for 31nt long reads lengths that showed strong periodicity (in replicates 2 and 3). Reads are globally mapped across transcripts and assigned to 5'-UTR, ORF and 3'-UTR regions (100nt upstream and downstream of start and stop codons plotted). Most of the reads map to the coding sequence (ORF) and fewer reads map to 5' and 3'-UTRs.

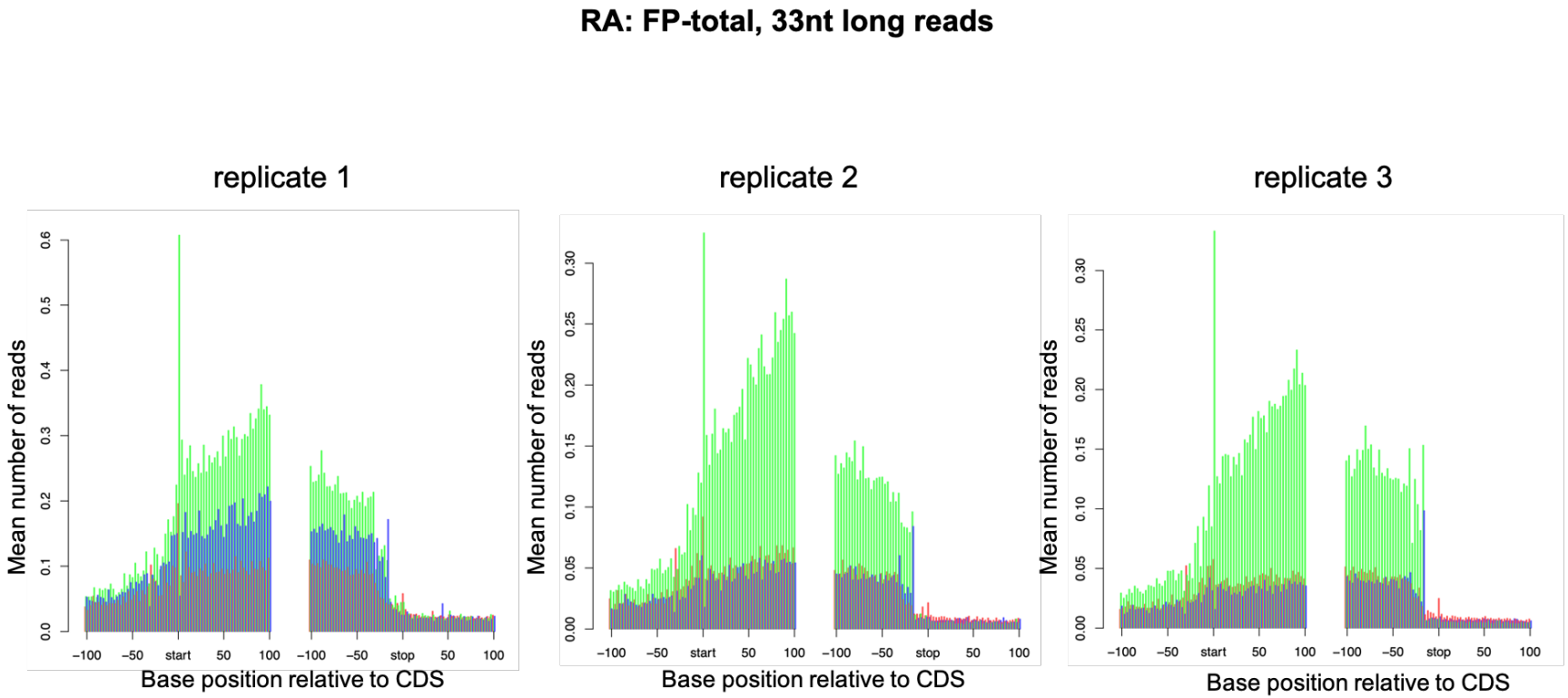


Figure 8 (supplementary to Figure 4.24): Footprint reads map to coding sequences. Metagene analysis of Ribo-Seq coupled with total cytoplasmic RNA-Seq reads of RA samples for 33nt long reads lengths that showed strong periodicity. Reads are globally mapped across transcripts and assigned to 5'-UTR, ORF and 3'-UTR regions (100nt upstream and downstream of start and stop codons plotted). Most of the reads map to the coding sequence (ORF) and fewer reads map to 5' and 3'-UTRs.

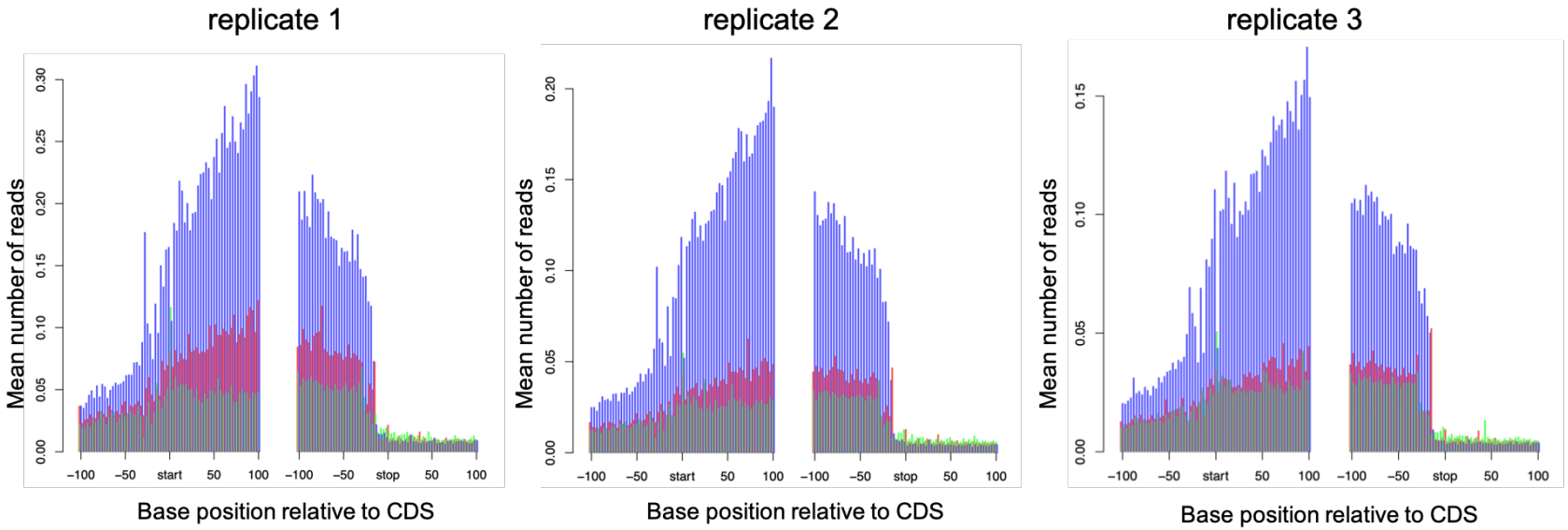


Figure 9 (supplementary to Figure 4.24): Footprint reads map to coding sequences. Metagene analysis of Ribo-Seq coupled with polysome-associated RNA-Seq reads of Control samples for 31nt long reads lengths that showed strong periodicity. Reads are globally mapped across transcripts and assigned to 5'-UTR, ORF and 3'-UTR regions (100nt upstream and downstream of start and stop codons plotted). Most of the reads map to the coding sequence (ORF) and fewer reads map to 5' and 3'-UTRs.

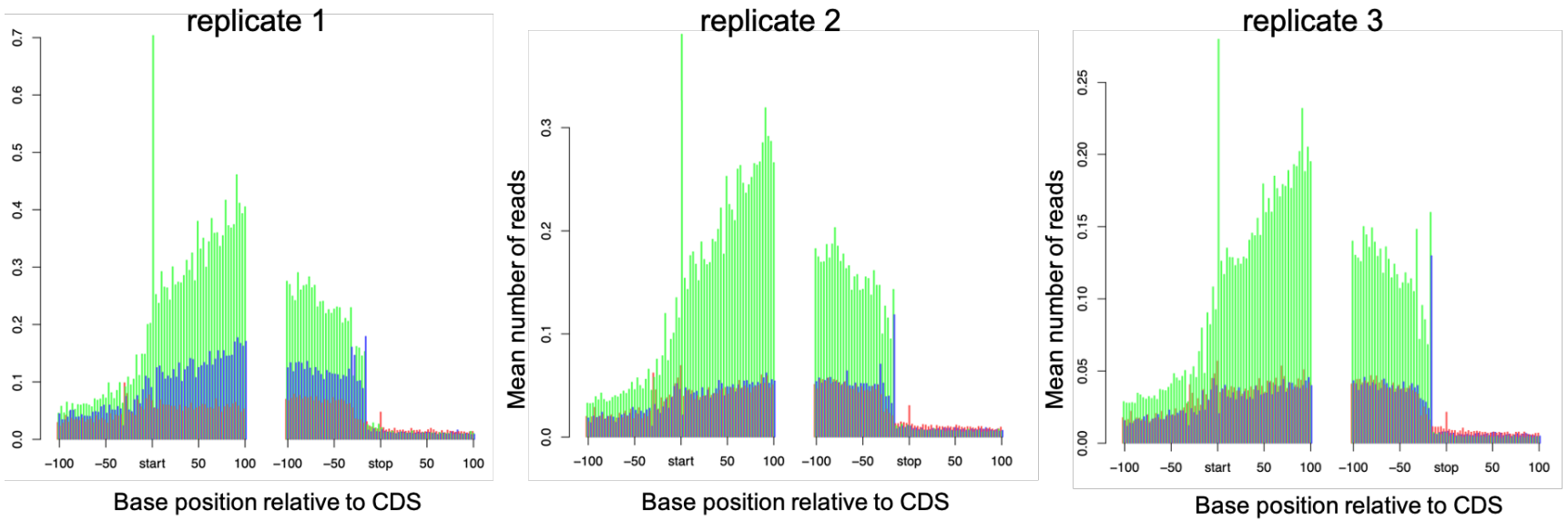


Figure 10 (supplementary to Figure 4.24): Footprint reads map to coding sequences. Metagene analysis of Ribo-Seq coupled with polysome-associated RNA-Seq reads of Control samples for 33nt long reads lengths that showed strong periodicity. Reads are globally mapped across transcripts and assigned to 5'-UTR, ORF and 3'-UTR regions (100nt upstream and downstream of start and stop codons plotted). Most of the reads map to the coding sequence (ORF) and fewer reads map to 5' and 3'-UTRs.

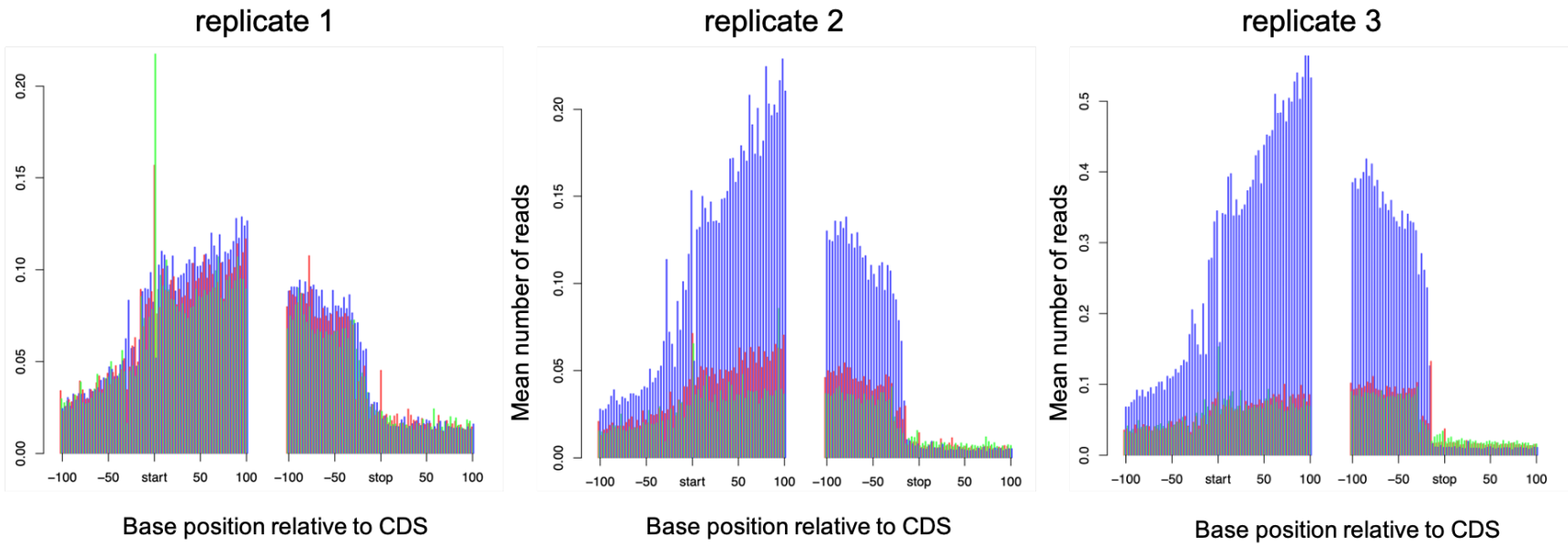


Figure 11 (supplementary to Figure 4.24): Footprint reads map to coding sequences. Metagene analysis of Ribo-Seq coupled with polysome-associated RNA-Seq reads of RA samples for 31nt long reads lengths that showed strong periodicity (replicates 2 and 3). Reads are globally mapped across transcripts and assigned to 5'-UTR, ORF and 3'-UTR regions (100nt upstream and downstream of start and stop codons plotted). Most of the reads map to the coding sequence (ORF) and fewer reads map to 5' and 3'-UTRs.

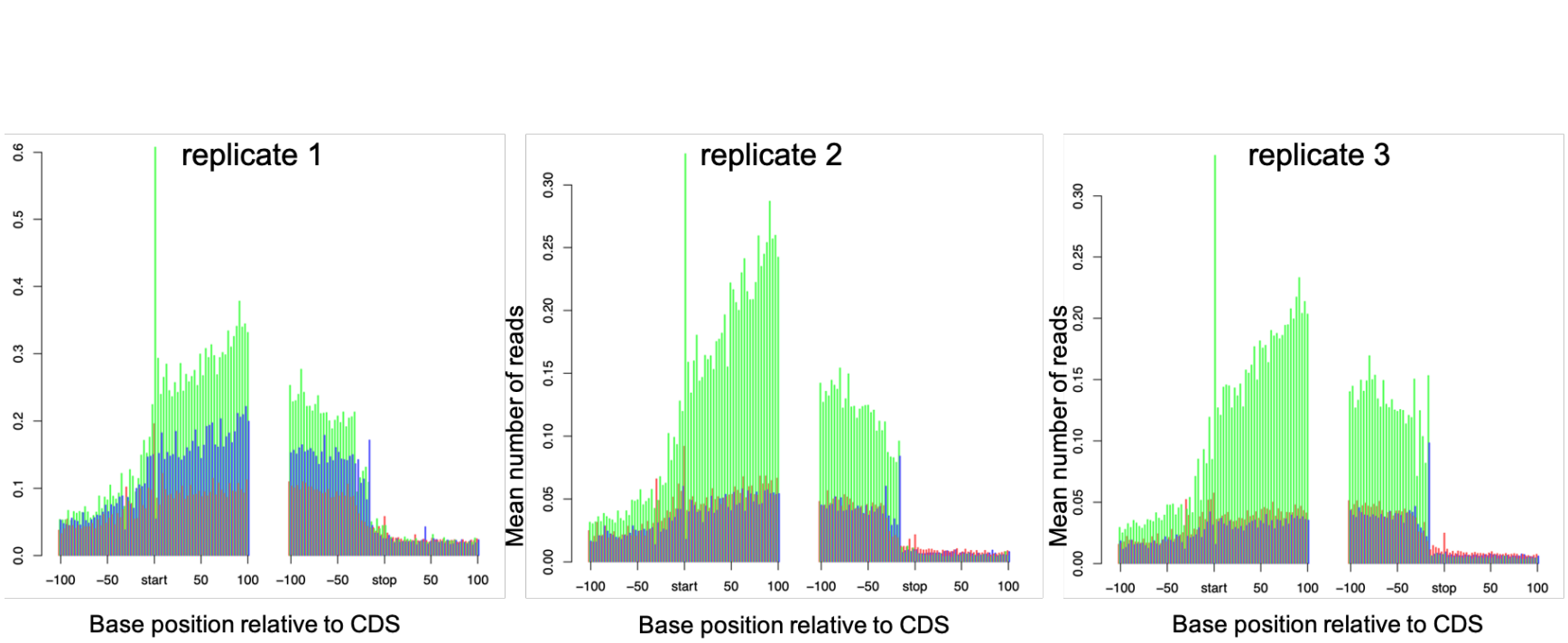


Figure 12 (supplementary to Figure 4.24): Footprint reads map to coding sequences. Metagene analysis of Ribo-Seq coupled with polysome-associated RNA-Seq reads of RA samples for 33nt long reads lengths that showed strong periodicity. Reads are globally mapped across transcripts and assigned to 5'-UTR, ORF and 3'-UTR regions (100nt upstream and downstream of start and stop codons plotted). Most of the reads map to the coding sequence (ORF) and fewer reads map to 5' and 3'-UTRs.

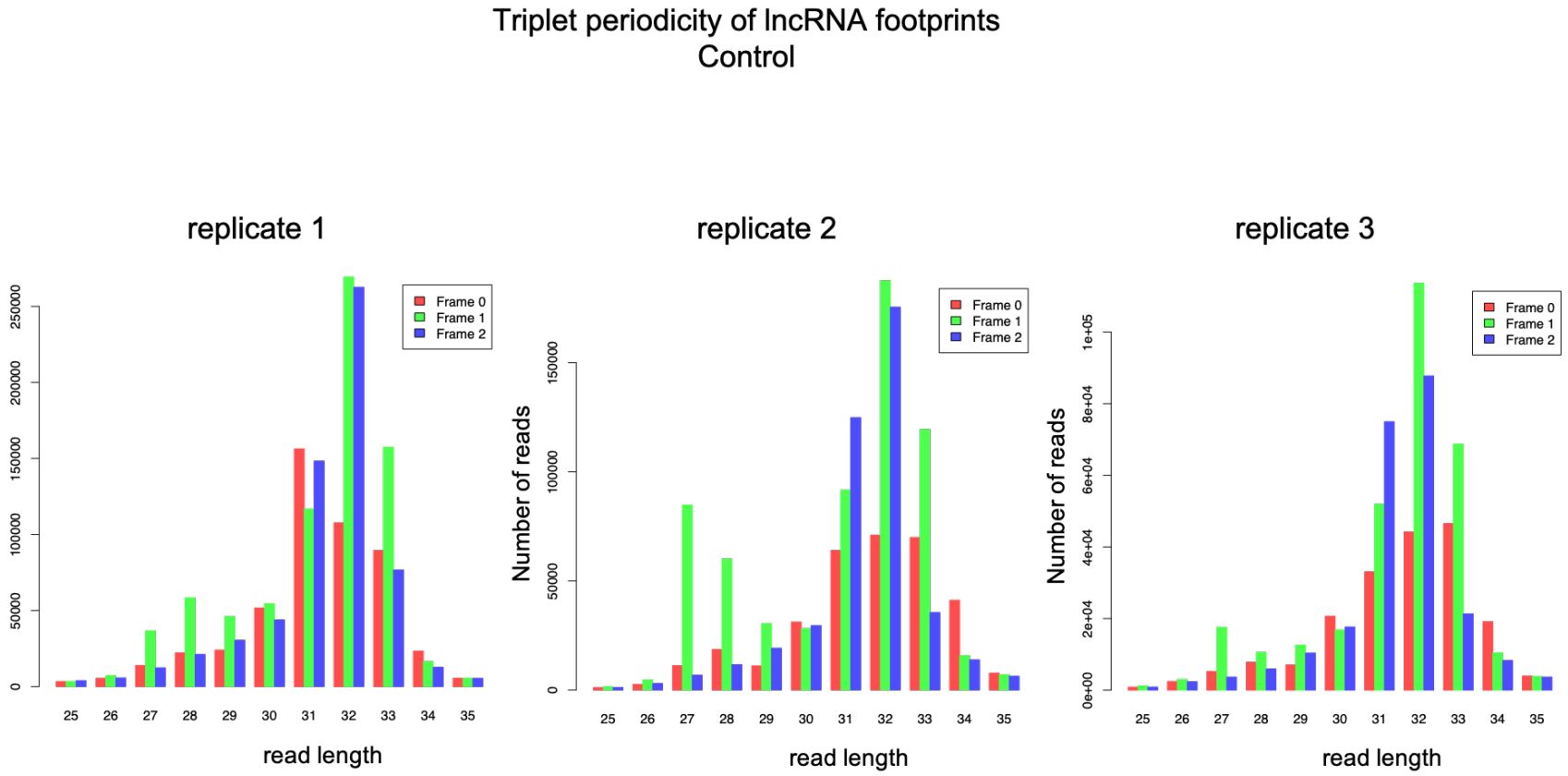


Figure 13 (supplementary to Figure 4.26): Ribo-Seq reads of Control samples mapping to lncRNAs display triplet periodicity. Framing was calculated for a read length range from 25-35nt and the majority of the reads were 31nt to 33nt long.

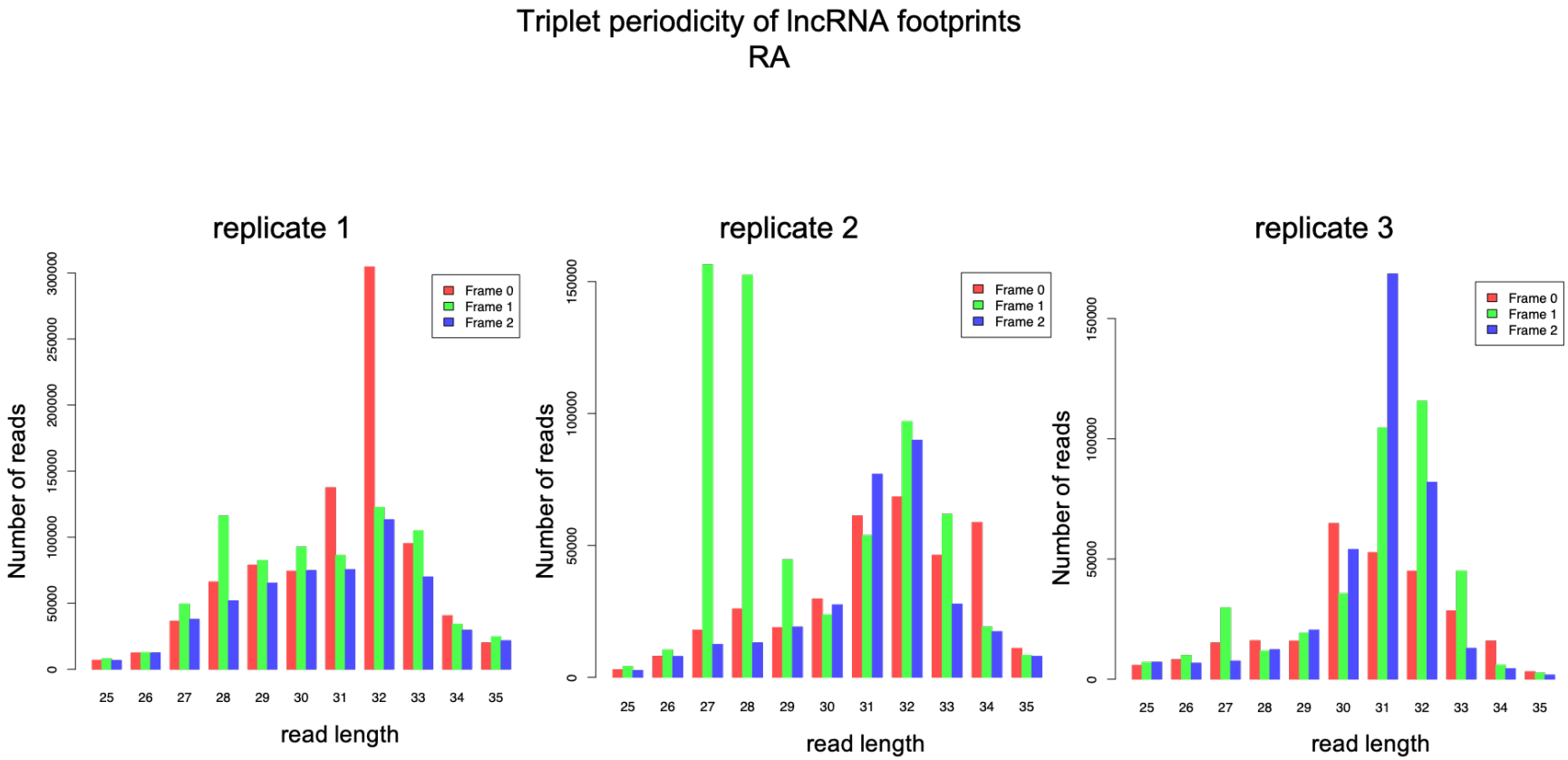


Figure 14 (supplementary to Figure 4.26): Ribo-Seq reads of RA samples mapping to lncRNAs display triplet periodicity. Framing was calculated for a read length range from 25-35nt and the majority of the reads were 27-28 or 31-32nt long.

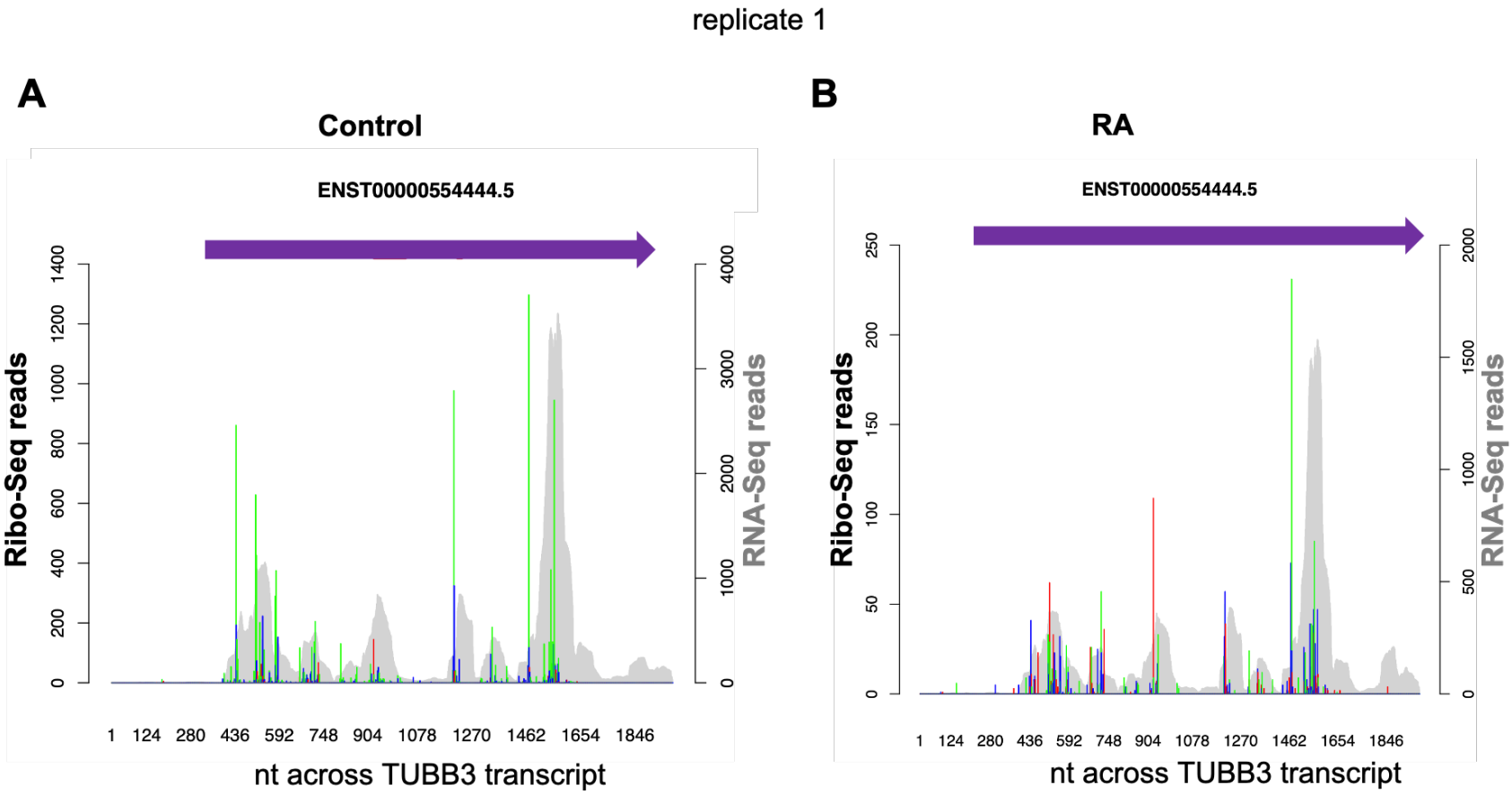


Figure 15 (supplementary to Figure 4.25): Transcript plots from Poly-Ribo-Seq replicate 1, showing the number of Ribo-Seq reads (footprints) colour-coded per frame on the left y axis and the number of RNA-Seq reads on the right y axis (grey), across the β-III tubulin (TUBB3 or Tuj1) transcript for (A) Control and (B) RA differentiated cells.

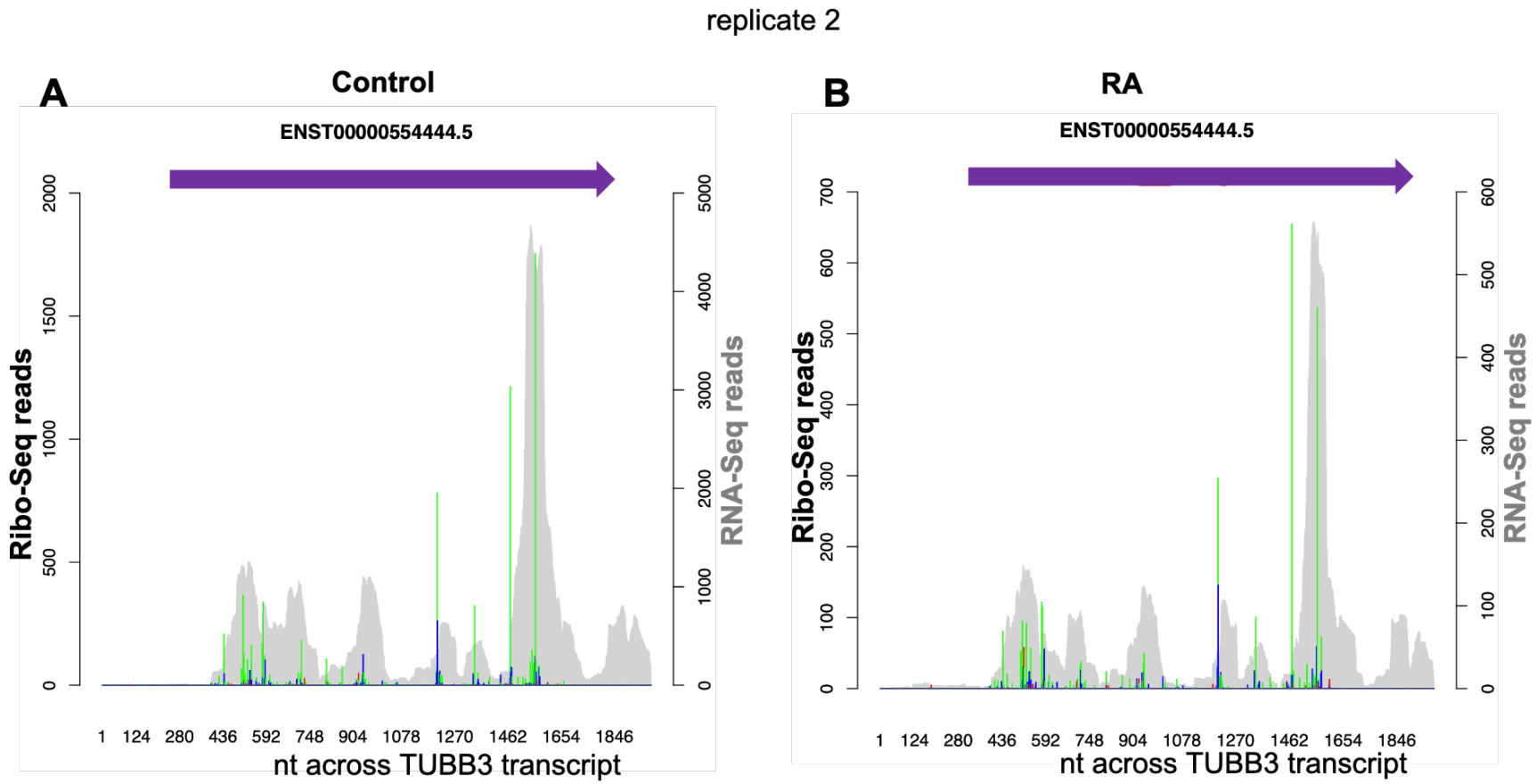


Figure 16 (supplementary to Figure 4.25): Transcript plots from Poly-Ribo-Seq replicate 2, showing the number of Ribo-Seq reads (footprints) colour-coded per frame on the left y axis and the number of RNA-Seq reads on the right y axis (grey), across the β-III tubulin (TUBB3 or Tuj1) transcript for (A) Control and (B) RA differentiated cells.

References

Abdelmohsen, K., Panda, A.C., Munk, R., Grammatikakis, I., Dudekula, D.B., De, S., Kim, J., Noh, J.H., Kim, K.M., Martindale, J.L. and Gorospe, M. 2017. Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1. *RNA Biology*. **14**(3), pp.361–369.

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A. and Blankenberg, D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*. **46**(W1), pp.W537–W544.

Agholme, L., Lindström, T., Kågedal, K., Marcusson, J. and Hallbeck, M. 2010. An In Vitro Model for Neuroscience: Differentiation of SH-SY5Y Cells into Cells with Morphological and Biochemical Characteristics of Mature Neurons. *Journal of Alzheimer's Disease*. **20**(4), pp.1069–1082.

Alarcón, C.R., Lee, H., Goodarzi, H., Halberg, N. and Tavazoie, S.F. 2015. N6-methyladenosine marks primary microRNAs for processing. *Nature*. **519**(7544), pp.482–485.

Albert, M. and Huttner, W.B. 2018. Epigenetic and Transcriptional Pre-patterning—An Emerging Theme in Cortical Neurogenesis. *Frontiers in Neuroscience*. **12**.

An, H., Skelt, L., Notaro, A., Highley, J.R., Fox, A.H., La Bella, V., Buchman, V.L. and Shelkovnikova, T.A. 2019. ALS-linked FUS mutations confer loss and gain of function in the nucleus by promoting excessive formation of dysfunctional paraspeckles. *Acta Neuropathologica Communications*. **7**(1), p.7.

Anderson, D.M., Anderson, K.M., Chang, C.-L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J., Bassel-Duby, R. and Olson, E.N. 2015. A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance. *Cell*. **160**(4), pp.595–606.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.

Arenas, E., Denham, M. and Villaescusa, J.C. 2015. How to make a midbrain dopaminergic neuron. *Development*. **142**(11), pp.1918–1936.

Arisi, I., D'Onofrio, M., Brandi, R., Felsani, A., Capsoni, S., Drovandi, G., Felici, G., Weitschek, E., Bertolazzi, P. and Cattaneo, A. 2011. Gene Expression Biomarkers in the Brain of a Mouse Model for Alzheimer's Disease: Mining of Microarray Data by Logic Classification and Feature Selection. *Journal of Alzheimer's Disease*. **24**(4), pp.721–738.

Aspden, J.L., Eyre-Walker, Y.C., Phillips, R.J., Amin, U., Mumtaz, M.A.S., Brocard, M. and Couso, J.P. 2014. Extensive translation of small open reading frames revealed by poly-ribo-seq. *eLife*. **3**(August2014), pp.1–19.

Balagopal, V. and Parker, R. 2011. Stm1 modulates translation after 80S formation in Saccharomyces cerevisiae. *RNA*. **17**(5), pp.835–842.

Ballas, N., Grunseich, C., Lu, D.D., Speh, J.C. and Mandel, G. 2005. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell*. **121**(4), pp.645–657.

Ban, N. 2000. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 A Resolution. *Science*. **289**(5481), pp.905–920.

Banfai, B., Jia, H., Khatun, J., Wood, E., Risk, B., Gundling, W.E., Kundaje, A., Gunawardena, H.P., Yu, Y., Xie, L., Krajewski, K., Strahl, B.D., Chen, X., Bickel, P., Giddings, M.C., Brown, J.B. and Lipovich, L. 2012. Long noncoding RNAs

are rarely translated in two human cell lines. *Genome Research*. **22**(9), pp.1646–1657.
Bartolomei, M.S., Zemel, S. and Tilghman, S.M. 1991. Parental imprinting of the mouse H19 gene. *Nature*. **351**(6322), pp.153–155.
Basrai, M.A., Hieter, P. and Boeke, J.D. 1997. Small Open Reading Frames: Beautiful Needles in the Haystack. *Genome Research*. **7**(8), pp.768–771.
Bass, A.J., Robinson, D.G. and Storey, J.D. 2019. Determining sufficient sequencing depth in RNA-Seq differential expression studies. *bioRxiv*., p.635623.
Bernard, D., Prasanth, K. V, Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z., Zhang, M.Q., Sedel, F., Jourdren, L., Coulpier, F., Triller, A., Spector, D.L. and Bessis, A. 2010. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *The EMBO Journal*. **29**(18), pp.3082–3093.
Beznosková, P., Wagner, S., Jansen, M.E., von der Haar, T. and Valášek, L.S. 2015. Translation initiation factor eIF3 promotes programmed stop codon readthrough. *Nucleic Acids Research*. **43**(10), pp.5099–5111.
Biedler, J.L., Helson, L. and Spengler, B.A. 1973. Morphology and Growth, Tumorigenicity, and Cytogenetics of Human Neuroblastoma Cells in Continuous Culture. *Cancer Research*. **33**(11), 2643 LP – 2652.
Biever, A., Glock, C., Tushev, G., Ciirdaeva, E., Dalmay, T., Langer, J.D. and Schuman, E.M. 2020. Monosomes actively translate synaptic mRNAs in neuronal processes. *Science*. **367**(6477), p.eaay4991.
Blackburn, E.H. and Gall, J.G. 1978. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *Journal of Molecular Biology*. **120**(1), pp.33–53.
Blair, J.D., Hockemeyer, D., Doudna, J.A., Bateup, H.S. and Floor, S.N. 2017. Widespread Translational Remodeling during Human Neuronal Differentiation. *Cell Reports*. **21**(7), pp.2005–2016.
Blighe K Lewis M, R.S. 2019. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling.
Bodian, D. 1965. A SUGGESTIVE RELATIONSHIP OF NERVE CELL RNA WITH SPECIFIC SYNAPTIC SITES. *Proceedings of the National Academy of Sciences*. **53**(2), pp.418–425.
Bond, A.M., VanGompel, M.J.W., Sametsky, E.A., Clark, M.F., Savage, J.C., Disterhoft, J.F. and Kohtz, J.D. 2009. Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nature Neuroscience*. **12**(8), pp.1020–1027.
Bosch, J.A., Ugur, B., Pichardo-Casas, I., Rabasco, J., Escobedo, F., Zuo, Z., Brown, B., Celniker, S., Sinclair, D., Bellen, H. and Perrimon, N. 2020. Two neuronal peptides encoded from a single transcript regulate mitochondrial function in Drosophila *bioRxiv*., 2020.07.01.182485.
Bratkovič, T., Modic, M., Camargo Ortega, G., Drukker, M. and Rogelj, B. 2018. Neuronal differentiation induces SNORD115 expression and is accompanied by post-transcriptional changes of serotonin receptor 2c mRNA. *Scientific Reports*. **8**(1), p.5101.
Brenig, K., Grube, L., Schwarzländer, M., Köhrer, K., Stühler, K. and Poschmann, G. 2020. The Proteomic Landscape of Cysteine Oxidation That Underpins Retinoic Acid-Induced Neuronal Differentiation. *Journal of Proteome Research*. **19**(5), pp.1923–1940.
Brenner, S., Jacob, F. and Meselson, M. 1961. An Unstable Intermediate Carrying

Information from Genes to Ribosomes for Protein Synthesis. *Nature*. **190**(4776), pp.576–581.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S. and Rastan, S. 1992. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*. **71**(3), pp.515–526.

Brodie, S., Lee, H.K., Jiang, W., Cazacu, S., Xiang, C., Poisson, L.M., Datta, I., Kalkanis, S., Ginsberg, D. and Brodie, C. 2017. The novel long non-coding RNA TALNEC2, regulates tumor cell growth and the stemness and radiation response of glioma stem cells. *Oncotarget*. **8**(19), pp.31785–31801.

Brown, A., Shao, S., Murray, J., Hegde, R.S. and Ramakrishnan, V. 2015. Structural basis for stop codon recognition in eukaryotes. *Nature*. **524**(7566), pp.493–496.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J. and Willard, H.F. 1992. The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*. **71**(3), pp.527–542.

Bulut-Karslioglu, A., Macrae, T.A., Oses-Prieto, J.A., Covarrubias, S., Percharde, M., Ku, G., Diaz, A., McManus, M.T., Burlingame, A.L. and Ramalho-Santos, M. 2018. The Transcriptionally Permissive Chromatin State of Embryonic Stem Cells Is Acutely Tuned to Translational Output. *Cell Stem Cell*. **22**(3), pp.369-383.e8.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. and Rinn, J.L. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*. **25**(18), pp.1915–1927.

Cagnetta, R., Frese, C.K., Shigeoka, T., Krijgsveld, J. and Holt, C.E. 2018. Rapid Cue-Specific Remodeling of the Nascent Axonal Proteome. *Neuron*. **99**(1), pp.29-46.e4.

Cajigas, I.J., Tushev, G., Will, T.J., tom Dieck, S., Fuerst, N. and Schuman, E.M. 2012. The Local Transcriptome in the Synaptic Neuropil Revealed by Deep Sequencing and High-Resolution Imaging. *Neuron*. **74**(3), pp.453–466.

Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., Landthaler, M., Obermayer, B. and Ohler, U. 2016. Detecting actively translated open reading frames in ribosome profiling data. *Nature methods*. **13**(2), pp.165–170.

Cañón, E., Cosgaya, J.M., Scsucova, S. and Aranda, A. 2004. Rapid effects of retinoic acid on CREB and ERK phosphorylation in neuronal cells. *Molecular biology of the cell*. **15**(12), pp.5583–5592.

Carelli, S., Giallongo, T., Rey, F., Latorre, E., Bordoni, M., Mazzucchelli, S., Gorio, M.C., Pansarasa, O., Provenzani, A., Cereda, C. and Di Giulio, A.M. 2019. HuR interacts with lincBRN1a and lincBRN1b during neuronal stem cells differentiation. *RNA Biology*. **16**(10), pp.1471–1485.

Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L.A. and Johnson, R. 2016. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA*. **22**(6), pp.867–882.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., Forrest, A.R.R., Carninci, P., Biffo, S., Stupka, E. and Gustincich, S. 2012. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*. **491**(7424), pp.454–457.

Carrieri, C., Forrest, A.R.R., Santoro, C., Persichetti, F., Carninci, P., Zucchelli, S. and Gustincich, S. 2015. Expression analysis of the long non-coding RNA antisense to Uchl1 (AS Uchl1) during dopaminergic cells' differentiation in vitro and in neurochemical models of Parkinson's disease. *Frontiers in Cellular Neuroscience*. **9**.

Casey, J., Hentze, M., Koeller, D., Caughman, S., Rouault, T., Klausner, R. and Harford, J. 1988. Iron-responsive elements: regulatory RNA sequences that control mRNA levels and translation. *Science*. **240**(4854), pp.924–928.

Cech, T.R. and Steitz, J.A. 2014. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell*. **157**(1), pp.77–94.

Chancellor, T.J., Lee, J., Thodeti, C.K. and Lele, T. 2010. Actomyosin Tension Exerted on the Nucleus through Nesprin-1 Connections Influences Endothelial Cell Adhesion, Migration, and Cyclic Strain-Induced Reorientation. *Biophysical Journal*. **99**(1), pp.115–123.

Chassé, H., Boulben, S., Costache, V., Cormier, P. and Morales, J. 2016. Analysis of translation using polysome profiling. *Nucleic Acids Research*., p.gkw907.

Chau, K.F., Shannon, M.L., Fame, R.M., Fonseca, E., Mullan, H., Johnson, M.B., Sendamarai, A.K., Springel, M.W., Laurent, B. and Lehtinen, M.K. 2018. Downregulation of ribosome biogenesis during early forebrain development. *eLife*. **7**.

Chen, J., Brunner, A.-D., Cogan, J.Z., Nuñez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D. and Weissman, J.S. 2020. Pervasive functional translation of noncanonical human open reading frames. *Science*. **367**(6482), pp.1140–1146.

Cheung, A.F.P., Kondo, S., Abdel-Mannan, O., Chodroff, R.A., Sirey, T.M., Bluy, L.E., Webber, N., DeProto, J., Karlen, S.J., Krubitzer, L., Stolp, H.B., Saunders, N.R. and Molnár, Z. 2010. The Subventricular Zone Is the Developmental Milestone of a 6-Layered Neocortex: Comparisons in Metatherian and Eutherian Mammals. *Cerebral Cortex*. **20**(5), pp.1071–1081.

Chew, G.-L., Pauli, A., Rinn, J.L., Regev, A., Schier, A.F. and Valen, E. 2013. Ribosome profiling reveals resemblance between long non-coding RNAs and 5′ leaders of coding RNAs. *Development*. **140**(13), pp.2828–2834.

Chodroff, R.A., Goodstadt, L., Sirey, T.M., Oliver, P.L., Davies, K.E., Green, E.D., Molnár, Z. and Ponting, C.P. 2010. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biology*. **11**(7), p.R72.

Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B.J., Michaux, J., Bilic, I., Hirsekorn, A., Calviello, L., Simó-Riudalbas, L., Planet, E., Lubiński, J., Bryśkiewicz, M., Wiznerowicz, M., Xenarios, I., Zhang, L., Trono, D., Harari, A., Ohler, U., Coukos, G. and Bassani-Sternberg, M. 2020. Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nature Communications*. **11**(1), p.1293.

Christov, C.P., Trivier, E. and Krude, T. 2008. Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. *British Journal of Cancer*. **98**(5), pp.981–988.

Chung, B.Y., Hardcastle, T.J., Jones, J.D., Irigoyen, N., Firth, A.E., Baulcombe, D.C. and Brierley, I. 2015. The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA*. **21**(10), pp.1731–1745.

Chung, D.W., Rudnicki, D.D., Yu, L. and Margolis, R.L. 2011. A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression. *Human Molecular Genetics*. **20**(17), pp.3467–3477.

Chung, J.-H., Ginn-Pease, M.E. and Eng, C. 2005. Phosphatase and Tensin Homologue Deleted on Chromosome 10 (PTEN) Has Nuclear Localization Signal–Like Sequences for Nuclear Import Mediated by Major Vault Protein. *Cancer Research*. **65**(10), pp.4108–4116.

Chung, L. 2015. A Brief Introduction to the Transduction of Neural Activity into Fos Signal. *Development & Reproduction*. **19**(2), pp.61–67.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A. and Lawrence, J.B. 2009. An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Molecular Cell*. **33**(6), pp.717–726.

Cloonan, N., Forrest, A.R.R., Kolle, G., Gardiner, B.B.A., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., Robertson, A.J., Perkins, A.C., Bruce, S.J., Lee, C.C., Ranade, S.S., Peckham, H.E., Manning, J.M., McKernan, K.J. and Grimmond, S.M. 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods*. **5**(7), pp.613–619.

Cohen, L.D., Zuchman, R., Sorokina, O., Müller, A., Dieterich, D.C., Armstrong, J.D., Ziv, T. and Ziv, N.E. 2013. Metabolic Turnover of Synaptic Proteins: Kinetics, Interdependencies and Implications for Synaptic Maintenance M. Akaaboune, ed. *PLoS ONE*. **8**(5), p.e63191.

Colasante, G., Collombat, P., Raimondi, V., Bonanomi, D., Ferrai, C., Maira, M., Yoshikawa, K., Mansouri, A., Valtorta, F., Rubenstein, J.L.R. and Broccoli, V. 2008. Arx Is a Direct Target of Dlx2 and Thereby Contributes to the Tangential Migration of GABAergic Interneurons. *Journal of Neuroscience*. **28**(42), pp.10674–10686.

Conley, A.B., Miller, W.J. and Jordan, I.K. 2008. Human cis natural antisense transcripts initiated by transposable elements. *Trends in Genetics*. **24**(2), pp.53–56.

Constantinescu, R., Constantinescu, A.T., Reichmann, H. and Janetzky, B. 2007. Neuronal differentiation and long-term culture of the human neuroblastoma line SH-SY5Y *In*: *Neuropsychiatric Disorders An Integrative Approach* [Online]. Vienna: Springer Vienna, pp.17–28. Available from: http://link.springer.com/10.1007/978-3-211-73574-9_3.

Corsini, N.S., Peer, A.M., Moeseneder, P., Roiuk, M., Burkard, T.R., Theussl, H.-C., Moll, I. and Knoblich, J.A. 2018. Coordinated Control of mRNA and rRNA Processing Controls Embryonic Stem Cell Pluripotency and Differentiation. *Cell Stem Cell*. **22**(4), pp.543-558.e12.

Costa-Mattioli, M., Gobert, D., Stern, E., Gamache, K., Colina, R., Cuello, C., Sossin, W., Kaufman, R., Pelletier, J., Rosenblum, K., Krnjević, K., Lacaille, J.-C., Nader, K. and Sonenberg, N. 2007. eIF2α Phosphorylation Bidirectionally Regulates the Switch from Short- to Long-Term Synaptic Plasticity and Memory. *Cell*. **129**(1), pp.195–206.

Costa, R.O., Martins, H., Martins, L.F., Cwetsch, A.W., Mele, M., Pedro, J.R., Tomé, D., Jeon, N.L., Cancedda, L., Jaffrey, S.R. and Almeida, R.D. 2019. Synaptogenesis Stimulates a Proteasome-Mediated Ribosome Reduction in Axons. *Cell Reports*. **28**(4), pp.864-876.e6.

Dallagiovanna, B., Pereira, I.T., Origa-Alves, A.C., Shigunov, P., Naya, H. and Spangenberg, L. 2017. lncRNAs are associated with polysomes during adipose-

derived stem cell differentiation. *Gene*. **610**, pp.103–111.
Deglincerti, A., Liu, Y., Colak, D., Hengst, U., Xu, G. and Jaffrey, S.R. 2015. Coupled local translation and degradation regulate growth cone collapse. *Nature Communications*. **6**(1), p.6888.
Denli, A.M., Tops, B.B.J., Plasterk, R.H.A., Ketting, R.F. and Hannon, G.J. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature*. **432**(7014), pp.231–235.
Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J.B., Lipovich, L., Gonzalez, J.M., Thomas, M., Davis, C.A., Shiekhattar, R., Gingeras, T.R., Hubbard, T.J., Notredame, C., Harrow, J. and Guigo, R. 2012. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*. **22**(9), pp.1775–1789.
Dessaud, E., Ribes, V., Balaskas, N., Yang, L.L., Pierani, A., Kicheva, A., Novitch, B.G., Briscoe, J. and Sasai, N. 2010. Dynamic Assignment and Maintenance of Positional Identity in the Ventral Neural Tube by the Morphogen Sonic Hedgehog M. P. Scott, ed. *PLoS Biology*. **8**(6), p.e1000382.
Dessaud, E., Yang, L.L., Hill, K., Cox, B., Ulloa, F., Ribeiro, A., Mynett, A., Novitch, B.G. and Briscoe, J. 2007. Interpretation of the sonic hedgehog morphogen gradient by a temporal adaptation mechanism. *Nature*. **450**(7170), pp.717–720.
Dimartino, D., Colantoni, A., Ballarino, M., Martone, J., Mariani, D., Danner, J., Bruckmann, A., Meister, G., Morlando, M. and Bozzoni, I. 2018. The Long Non-coding RNA lnc-31 Interacts with Rock1 mRNA and Mediates Its YB-1-Dependent Translation. *Cell Reports*. **23**(3), pp.733–740.
Dinman, J.D. 2012. Mechanisms and implications of programmed translational frameshifting. *Wiley Interdisciplinary Reviews: RNA*. **3**(5), pp.661–673.
Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K., Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Bar, N.S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O.J., Park, E., Persaud, K., Preall, J.B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L.-H., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S.E., Hannon, G., Giddings, M.C., Ruan, Y., Wold, B., Carninci, P., Guigó, R. and Gingeras, T.R. 2012. Landscape of transcription in human cells. *Nature*. **489**(7414), pp.101–108.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**(1), pp.15–21.
Dörrbaum, A.R., Kochen, L., Langer, J.D. and Schuman, E.M. 2018. Local and global influences on protein turnover in neurons and glia. *eLife*. **7**.
Duncan, C.D.S. and Mata, J. 2014. The translational landscape of fission-yeast meiosis and sporulation. *Nature Structural & Molecular Biology*. **21**(7), pp.641–647.

Eden, E., Navon, R., Steinfeld, I., Lipson, D. and Yakhini, Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. **10**(1), p.48.

Edfors, F., Danielsson, F., Hallström, B.M., Käll, L., Lundberg, E., Pontén, F., Forsström, B. and Uhlén, M. 2016. Gene-specific correlation of <scp>RNA</scp> and protein levels in human cells and tissues. *Molecular Systems Biology*. **12**(10), p.883.

Edmonds, M., Vaughan, M.H. and Nakazato, H. 1971. Polyadenylic Acid Sequences in the Heterogeneous Nuclear RNA and Rapidly-Labeled Polyribosomal RNA of HeLa Cells: Possible Evidence for a Precursor Relationship. *Proceedings of the National Academy of Sciences*. **68**(6), pp.1336–1340.

Elisaphenko, E.A., Kolesnikov, N.N., Shevchenko, A.I., Rogozin, I.B., Nesterova, T.B., Brockdorff, N. and Zakian, S.M. 2008. A Dual Origin of the Xist Gene from a Protein-Coding Gene and a Set of Transposable Elements S. Gadagkar, ed. *PLoS ONE*. **3**(6), p.e2521.

Erhard, F., Halenius, A., Zimmermann, C., L'Hernault, A., Kowalewski, D.J., Weekes, M.P., Stevanovic, S., Zimmer, R. and Dölken, L. 2018. Improved Ribo-seq enables identification of cryptic translation events. *Nature Methods*. **15**(5), pp.363–366.

Faghihi, M.A., Modarresi, F., Khalil, A.M., Wood, D.E., Sahagan, B.G., Morgan, T.E., Finch, C.E., St. Laurent III, G., Kenny, P.J. and Wahlestedt, C. 2008. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. *Nature Medicine*. **14**(7), pp.723–730.

Faghihi, M.A., Zhang, M., Huang, J., Modarresi, F., Van der Brug, M.P., Nalls, M.A., Cookson, M.R., St-Laurent, G. and Wahlestedt, C. 2010. Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biology*. **11**(5), p.R56.

Flockhart, R.J., Webster, D.E., Qu, K., Mascarenhas, N., Kovalski, J., Kretz, M. and Khavari, P.A. 2012. BRAF V600E remodels the melanocyte transcriptome and induces BANCR to regulate melanoma cell migration. *Genome Research*. **22**(6), pp.1006–1014.

Floor, S.N. and Doudna, J.A. 2016. Tunable protein synthesis by transcript isoforms in human cells. *eLife*. **5**.

Floquet, C., Hatin, I., Rousset, J.-P. and Bidou, L. 2012. Statistical Analysis of Readthrough Levels for Nonsense Mutations in Mammalian Cells Reveals a Major Determinant of Response to Gentamicin K. M. Flanigan, ed. *PLoS Genetics*. **8**(3), p.e1002608.

Fornasiero, E.F., Mandad, S., Wildhagen, H., Alevra, M., Rammner, B., Keihani, S., Opazo, F., Urban, I., Ischebeck, T., Sakib, M.S., Fard, M.K., Kirli, K., Centeno, T.P., Vidal, R.O., Rahman, R.-U., Benito, E., Fischer, A., Dennerlein, S., Rehling, P., Feussner, I., Bonn, S., Simons, M., Urlaub, H. and Rizzoli, S.O. 2018. Precisely measured protein lifetimes in the mouse brain reveal differences across tissues and subcellular fractions. *Nature Communications*. **9**(1), p.4230.

Forster, J.I., Köglsberger, S., Trefois, C., Boyd, O., Baumuratov, A.S., Buck, L., Balling, R. and Antony, P.M.A. 2016. Characterization of Differentiated SH-SY5Y as Neuronal Screening Model Reveals Increased Oxidative Vulnerability. *Journal of Biomolecular Screening*. **21**(5), pp.496–509.

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S.,

Fiddes, I.T., García Girón, C., Gonzalez, J.M., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O.G., Lagarde, J., Martin, F.J., Martínez, L., Mohanan, S., Muir, P., Navarro, F.C.P., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B.M., Stapleton, E., Suner, M.-M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J.S., Gerstein, M., Guigó, R., Hubbard, T.J.P., Kellis, M., Paten, B., Reymond, A., Tress, M.L. and Flicek, P. 2019a. GENCODE reference annotation for the human and mouse genomes. *Nucleic acids research*. **47**(D1), pp.D766–D773.

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I.T., García Girón, C., Gonzalez, J.M., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O.G., Lagarde, J., Martin, F.J., Martínez, L., Mohanan, S., Muir, P., Navarro, F.C.P., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B.M., Stapleton, E., Suner, M.-M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J.S., Gerstein, M., Guigó, R., Hubbard, T.J.P., Kellis, M., Paten, B., Reymond, A., Tress, M.L. and Flicek, P. 2019b. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research*. **47**(D1), pp.D766–D773.

Fujii, K., Shi, Z., Zhulyn, O., Denans, N. and Barna, M. 2017. Pervasive translational regulation of the cell signalling circuitry underlies mammalian development. *Nature Communications*. **8**(1), p.14443.

Galindo, M.I., Pueyo, J.I., Fouix, S., Bishop, S.A. and Couso, J.P. 2007. Peptides Encoded by Short ORFs Control Development and Define a New Eukaryotic Gene Family A. Martinez Arias, ed. *PLoS Biology*. **5**(5), p.e106.

Geballe, A.P. and Morris, D.R. 1994. Initiation codons within 5′-leaders of mRNAs as regulators of translation. *Trends in Biochemical Sciences*. **19**(4), pp.159–164.

Gerace, E. and Moazed, D. 2015. Affinity Pull-Down of Proteins Using Anti-FLAG M2 Agarose Beads *In*:, pp.99–110. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0076687914000755.

Gezer, U., Özgür, E., Cetinkaya, M., Isin, M. and Dalay, N. 2014. Long non-coding RNAs with low expression levels in cells are enriched in secreted exosomes. *Cell Biology International*., n/a-n/a.

Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Hagopian, R., Mungrue, I.N., Farber, C.R., Sinsheimer, J., Kang, H.M., Furlotte, N., Park, C.C., Wen, P.-Z., Brewer, H., Weitz, K., Camp, D.G., Pan, C., Yordanova, R., Neuhaus, I., Tilford, C., Siemers, N., Gargalovic, P., Eskin, E., Kirchgessner, T., Smith, D.J., Smith, R.D. and Lusis, A.J. 2011. Comparative Analysis of Proteome and Transcriptome Variation in Mouse M. Snyder, ed. *PLoS Genetics*. **7**(6), p.e1001393.

Gil, N. and Ulitsky, I. 2020. Regulation of gene expression by cis-acting long non-coding RNAs. *Nature Reviews Genetics*. **21**(2), pp.102–117.

Gkogkas, C., Sonenberg, N. and Costa-Mattioli, M. 2010. Translational Control Mechanisms in Long-lasting Synaptic Plasticity and Memory. *Journal of Biological Chemistry*. **285**(42), pp.31913–31917.

Glinka, M., Herrmann, T., Funk, N., Havlicek, S., Rossoll, W., Winkler, C. and Sendtner, M. 2010. The heterogeneous nuclear ribonucleoprotein-R is necessary for axonal β-actin mRNA translocation in spinal motor neurons. *Human Molecular Genetics*. **19**(10), pp.1951–1966.

Gökmen-Polar, Y., Zavodszky, M., Chen, X., Gu, X., Kodira, C. and Badve, S. 2016.

Abstract P2-06-05: LINC00478: A novel tumor suppressor in breast cancer *In*: *Poster Session Abstracts* [Online]. American Association for Cancer Research, pp.P2-06-05-P2-06–05. Available from: http://cancerres.aacrjournals.org/lookup/doi/10.1158/1538-7445.SABCS15-P2-06-05.

Gomes, C., Merianda, T.T., Lee, S.J., Yoo, S. and Twiss, J.L. 2014. Molecular determinants of the axonal mRNA transcriptome. *Developmental Neurobiology*. **74**(3), pp.218–232.

Goodman, C.S. and Doe, C.Q. 1993. Embryonic development of theDrosophila, The Development ofDrosophila melanogaster. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY*.

Gordon, A. 2010. FASTQ/A short-reads pre-processing tools.

Goyal, A., Fiškin, E., Gutschner, T., Polycarpou-Schwarz, M., Groß, M., Neugebauer, J., Gandhi, M., Caudron-Herger, M., Benes, V. and Diederichs, S. 2017. A cautionary tale of sense-antisense gene pairs: independent regulation despite inverse correlation of expression. *Nucleic Acids Research*. **45**(21), pp.12496–12508.

Graham, L.D., Pedersen, S.K., Brown, G.S., Ho, T., Kassir, Z., Moynihan, A.T., Vizgoft, E.K., Dunne, R., Pimlott, L., Young, G.P., LaPointe, L.C. and Molloy, P.L. 2011. Colorectal Neoplasia Differentially Expressed (CRNDE), a Novel Gene with Elevated Expression in Colorectal Adenomas and Adenocarcinomas. *Genes & Cancer*. **2**(8), pp.829–840.

Gros, F., Hiatt, H., Gilbert, W., Kurland, C.G., Risebrough, R.W. and Watson, J.D. 1961. Unstable Ribonucleic Acid Revealed by Pulse Labelling of Escherichia Coli. *Nature*. **190**(4776), pp.581–585.

Gu, S., Xie, R., Liu, X., Shou, J., Gu, W. and Che, X. 2017. Long Coding RNA XIST Contributes to Neuronal Apoptosis through the Downregulation of AKT Phosphorylation and Is Negatively Regulated by miR-494 in Rat Spinal Cord Injury. *International Journal of Molecular Sciences*. **18**(4), p.732.

Guarnieri, F.C., de Chevigny, A., Falace, A. and Cardoso, C. 2018. Disorders of neurogenesis and cortical development. *Dialogues in Clinical Neuroscience*. **20**(4), pp.255–266.

Guo, H., Ingolia, N.T., Weissman, J.S. and Bartel, D.P. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*. **466**(7308), pp.835–840.

Gutierrez, E., Shin, B.-S., Woolstenhulme, C.J., Kim, J.-R., Saini, P., Buskirk, A.R. and Dever, T.E. 2013. eIF5A Promotes Translation of Polyproline Motifs. *Molecular Cell*. **51**(1), pp.35–45.

Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S. and Lander, E.S. 2013. Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. *Cell*. **154**(1), pp.240–251.

Haerty, W. and Ponting, C.P. 2013. Mutations within lncRNAs are effectively selected against in fruitfly but not in human. *Genome Biology*. **14**(5), p.R49.

Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N., Gibson, D., Diekhans, M., Clawson, H., Casper, J., Barber, G.P., Haussler, D., Kuhn, R.M. and Kent, W.J. 2019. The UCSC Genome Browser database: 2019 update. *Nucleic acids research*. **47**(D1), pp.D853–D858.

Hämmerle, B., Yañez, Y., Palanca, S., Cañete, A., Burks, D.J., Castel, V. and Font de Mora, J. 2013. Targeting Neuroblastoma Stem Cells with Retinoic Acid and

Proteasome Inhibitor Y. G. Shellman, ed. *PLoS ONE*. **8**(10), p.e76761.
Hannon, G.J. 2010. FASTX-Toolkit.
Hardcastle, T. 2014. riboSeqR: Analysis of sequencing data from ribosome profiling experiments.. R package version 1.0.5.
Harding, H.P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M. and Ron, D. 2000. Regulated Translation Initiation Controls Stress-Induced Gene Expression in Mammalian Cells. *Molecular Cell*. **6**(5), pp.1099–1108.
He, S., Liu, C., Skogerbo, G., Zhao, H., Wang, J., Liu, T., Bai, B., Zhao, Y. and Chen, R. 2007. NONCODE v2.0: decoding the non-coding. *Nucleic Acids Research*. **36**(Database), pp.D170–D172.
van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J.F., Adami, E., Faber, A.B., Kirchner, M., Maatz, H., Blachut, S., Sandmann, C.-L., Kanda, M., Worth, C.L., Schafer, S., Calviello, L., Merriott, R., Patone, G., Hummel, O., Wyler, E., Obermayer, B., Mücke, M.B., Lindberg, E.L., Trnka, F., Memczak, S., Schilling, M., Felkin, L.E., Barton, P.J.R., Quaife, N.M., Vanezis, K., Diecke, S., Mukai, M., Mah, N., Oh, S.-J., Kurtz, A., Schramm, C., Schwinge, D., Sebode, M., Harakalova, M., Asselbergs, F.W., Vink, A., de Weger, R.A., Viswanathan, S., Widjaja, A.A., Gärtner-Rommel, A., Milting, H., dos Remedios, C., Knosalla, C., Mertins, P., Landthaler, M., Vingron, M., Linke, W.A., Seidman, J.G., Seidman, C.E., Rajewsky, N., Ohler, U., Cook, S.A. and Hubner, N. 2019. The Translational Landscape of the Human Heart. *Cell*. **178**(1), pp.242-260.e29.
Hentze, M., Caughman, S., Rouault, T., Barriocanal, J., Dancis, A., Harford, J. and Klausner, R. 1987. Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science*. **238**(4833), pp.1570–1573.
Hernandez-Alias, X., Benisty, H., Schaefer, M.H. and Serrano, L. 2020. Translational efficiency across healthy and tumor tissues is proliferation-related. *Molecular Systems Biology*. **16**(3).
Heyer, E.E. and Moore, M.J. 2016. Redefining the Translational Status of 80S Monosomes. *Cell*. **164**(4), pp.757–769.
Hezroni, H., Ben-Tov Perry, R., Meir, Z., Housman, G., Lubelsky, Y. and Ulitsky, I. 2017. A subset of conserved mammalian long non-coding RNAs are fossils of ancestral protein-coding genes. *Genome Biology*. **18**(1), p.162.
Hezroni, H., Koppstein, D., Schwartz, M.G., Avrutin, A., Bartel, D.P. and Ulitsky, I. 2015. Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Reports*. **11**(7), pp.1110–1122.
Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I. and Zamecnik, P.C. 1958. A soluble ribonucleic acid intermediate in protein synthesis. *The Journal of biological chemistry*. **231**(1), pp.241–257.
Hodnett, J.L. and Busch, H. 1968. Isolation and characterization of uridylic acid-rich 7 S ribonucleic acid of rat liver nuclei. *The Journal of biological chemistry*. **243**(24), pp.6334–6342.
Holt, C.E., Martin, K.C. and Schuman, E.M. 2019. Local translation in neurons: visualization and function. *Nature Structural & Molecular Biology*. **26**(7), pp.557–566.
Hong, X., Scofield, D.G. and Lynch, M. 2006. Intron Size, Abundance, and Distribution within Untranslated Regions of Genes. *Molecular Biology and Evolution*. **23**(12), pp.2392–2404.
Hsu, M.-T. and Coca-Prados, M. 1979. Electron microscopic evidence for the circular

form of RNA in the cytoplasm of eukaryotic cells. *Nature*. **280**(5720), pp.339–340.

Hsu, P.Y., Calviello, L., Wu, H.-Y.L., Li, F.-W., Rothfels, C.J., Ohler, U. and Benfey, P.N. 2016. Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. *Proceedings of the National Academy of Sciences*. **113**(45), pp.E7126–E7135.

Huang M, Yu-Chen Y, Shu-Rong C, Jin-Ren C, Lu JX, Zhoa L, G. and LJ, W.Z. 1998. Use of all trans retinoic acid in the treatment of acute promyelocytic leukemia. *Blood*. (72), pp.567–572.

Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B. and Chess, A. 2007. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*. **8**(1), p.39.

Hüttelmaier, S., Zenklusen, D., Lederer, M., Dictenberg, J., Lorenz, M., Meng, X., Bassell, G.J., Condeelis, J. and Singer, R.H. 2005. Spatial regulation of β-actin translation by Src-dependent phosphorylation of ZBP1. *Nature*. **438**(7067), pp.512–515.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M. and Weissman, J.S. 2013. Genome-Wide Annotation and Quantitation of Translation by Ribosome Profiling *In*: *Current Protocols in Molecular Biology* [Online]. Hoboken, NJ, USA: John Wiley & Sons, Inc., 4.18.1-4.18.19. Available from: http://doi.wiley.com/10.1002/0471142727.mb0418s103.

Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J.S., Jackson, S.E., Wills, M.R. and Weissman, J.S. 2014. Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Reports*. **8**(5), pp.1365–1379.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S. and Weissman, J.S. 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*. **324**(5924), pp.218–223.

Ingolia, N.T., Lareau, L.F. and Weissman, J.S. 2011. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell*. **147**(4), pp.789–802.

Isin, M., Uysaler, E., Ozgur, E., Kozeoglu, H., Sanli, O., Yucel, O.B., Gezer, U. and Dalay, N. 2015. Exosomal lncRNA-p21 levels may help to distinguish prostate cancer from benign disease. *Frontiers in Genetics*. **6**.

Jackson, R. and Standart, N. 2015. The awesome power of ribosome profiling. *RNA*. **21**(4), pp.652–654.

Jackson, R.J., Hellen, C.U.T. and Pestova, T. V. 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*. **11**(2), pp.113–127.

Ji, P., Diederichs, S., Wang, W., Böing, S., Metzger, R., Schneider, P.M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., Thomas, M., Berdel, W.E., Serve, H. and Müller-Tidow, C. 2003. MALAT-1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*. **22**(39), pp.8031–8041.

Jolliffe, I. 2011. Principal Component Analysis *In*: *International Encyclopedia of Statistical Science* [Online]. Berlin, Heidelberg: Springer Berlin Heidelberg, pp.1094–1096. Available from: http://link.springer.com/10.1007/978-3-642-04898-2_455.

Jones, A.N. and Sattler, M. 2019. Challenges and perspectives for structural biology of lncRNAs—the example of the Xist lncRNA A-repeats A. Bindereif, ed. *Journal*

of Molecular Cell Biology. **11**(10), pp.845–859.
Kapp, L.D. and Lorsch, J.R. 2004. The Molecular Mechanics of Eukaryotic Translation. *Annual Review of Biochemistry*. **73**(1), pp.657–704.
Kar, A.N., Lee, S.J. and Twiss, J.L. 2018. Expanding Axonal Transcriptome Brings New Functions for Axonally Synthesized Proteins in Health and Disease. *The Neuroscientist*. **24**(2), pp.111–129.
Kawashima, H., Takano, H., Sugita, S., Takahara, Y., Sugimura, K. and Nakatani, T. 2003. A novel steroid receptor co-activator protein (SRAP) as an alternative form of steroid receptor RNA-activator gene: expression in prostate cancer cells and enhancement of androgen receptor activity. *Biochemical Journal*. **369**(1), pp.163–171.
Kedersha, N.L. and Rome, L.H. 1986. Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. *The Journal of Cell Biology*. **103**(3), pp.699–709.
Kibbe, W. 2007. Kibbe, W.A. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res. 35, W43-W46. *Nucleic acids research*. **35**, pp.W43-6.
Kim, E., Lee, S., Mian, M.F., Yun, S.U., Song, M., Yi, K.-S., Ryu, S.H. and Suh, P.-G. 2006. Crosstalk between Src and major vault protein in epidermal growth factor-dependent cell signalling. *FEBS Journal*. **273**(4), pp.793–804.
Kim, H.H., Lee, S.J., Gardiner, A.S., Perrone-Bizzozero, N.I. and Yoo, S. 2015. Different motif requirements for the localization zipcode element of β-actin mRNA binding by HuD and ZBP1. *Nucleic Acids Research*. **43**(15), pp.7432–7446.
Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P.F., Kreiman, G. and Greenberg, M.E. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature*. **465**(7295), pp.182–187.
Kirk, J.M., Kim, S.O., Inoue, K., Smola, M.J., Lee, D.M., Schertzer, M.D., Wooten, J.S., Baker, A.R., Sprague, D., Collins, D.W., Horning, C.R., Wang, S., Chen, Q., Weeks, K.M., Mucha, P.J. and Calabrese, J.M. 2018. Functional classification of long non-coding RNAs by k-mer content. *Nature Genetics*. **50**(10), pp.1474–1482.
Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S. and Kageyama, Y. 2007. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nature Cell Biology*. **9**(6), pp.660–665.
Kong, J. and Lasko, P. 2012. Translational control in cellular and developmental processes. *Nature Reviews Genetics*. **13**(6), pp.383–394.
Korecka, J.A., van Kesteren, R.E., Blaas, E., Spitzer, S.O., Kamstra, J.H., Smit, A.B., Swaab, D.F., Verhaagen, J. and Bossers, K. 2013. Phenotypic Characterization of Retinoic Acid Differentiated SH-SY5Y Cells by Transcriptional Profiling K.-L. Lim, ed. *PLoS ONE*. **8**(5), p.e63862.
Kowalski, M.P., Dubouix-Bourandy, A., Bajmoczi, M., Golan, D.E., Zaidi, T., Coutinho-Sledge, Y.S., Gygi, M.P., Gygi, S.P., Wiemer, E.A.C. and Pier, G.B. 2007. Host Resistance to Lung Infection Mediated by Major Vault Protein in Epithelial Cells. *Science*. **317**(5834), pp.130–132.
Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research*. **15**(20), pp.8125–8148.
Koziol, M.J. and Rinn, J.L. 2010. RNA traffic control of chromatin complexes. *Current*

Opinion in Genetics & Development. **20**(2), pp.142–148.
Kretz, M., Webster, D.E., Flockhart, R.J., Lee, C.S., Zehnder, A., Lopez-Pajares, V., Qu, K., Zheng, G.X.Y., Chow, J., Kim, G.E., Rinn, J.L., Chang, H.Y., Siprashvili, Z. and Khavari, P.A. 2012. Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes & Development*. **26**(4), pp.338–343.
Lander, E.S., Linton, L.M. and Birren, B., et al. 2001. Initial sequencing and analysis of the human genome. *Nature*. **409**(6822), pp.860–921.
Langmead, B. and Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. **9**(4), pp.357–359.
Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*. **10**(3), p.R25.
Laumont, C.M., Vincent, K., Hesnard, L., Audemard, É., Bonneil, É., Laverdure, J.-P., Gendron, P., Courcelles, M., Hardy, M.-P., Côté, C., Durette, C., St-Pierre, C., Benhammadi, M., Lanoix, J., Vobecky, S., Haddad, E., Lemieux, S., Thibault, P. and Perreault, C. 2018. Noncoding regions are the main source of targetable tumor-specific antigens. *Science Translational Medicine*. **10**(470), p.eaau5516.
Lee, R.C., Feinbaum, R.L. and Ambros, V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*. **75**(5), pp.843–854.
Lerner, M.R. and Steitz, J.A. 1979. Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proceedings of the National Academy of Sciences*. **76**(11), pp.5495–5499.
Leung, K.-M., van Horck, F.P., Lin, A.C., Allison, R., Standart, N. and Holt, C.E. 2006. Asymmetrical β-actin mRNA translation in growth cones mediates attractive turning to netrin-1. *Nature Neuroscience*. **9**(10), pp.1247–1256.
Lewandowski, J.P., Dumbović, G., Watson, A.R., Hwang, T., Jacobs-Palmer, E., Chang, N., Much, C., Turner, K., Kirby, C., Schulz, J.F., Müller, C.-L., Rubinstein, N.D., Groff, A.F., Liapis, S.C., Gerhardinger, C., Hubner, N., van Heesch, S., Hoekstra, H.E., Sauvageau, M. and Rinn, J.L. 2019. The Tug1 Locus is Essential for Male Fertility. *bioRxiv*., p.562066.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*. **25**(16), pp.2078–2079.
Liao, Y., Smyth, G.K. and Shi, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*. **30**(7), pp.923–930.
Liao, Y., Smyth, G.K. and Shi, W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research*. **41**(10), pp.e108–e108.
Lin, N., Chang, K.-Y., Li, Z., Gates, K., Rana, Z.A., Dang, J., Zhang, D., Han, T., Yang, C.-S., Cunningham, T.J., Head, S.R., Duester, G., Dong, P.D.S. and Rana, T.M. 2014. An Evolutionarily Conserved Long Noncoding RNA TUNA Controls Pluripotency and Neural Lineage Commitment. *Molecular Cell*. **53**(6), pp.1005–1019.
Lindsay, S.J., Xu, Y., Lisgo, S.N., Harkin, L.F., Copp, A.J., Gerrelli, D., Clowry, G.J., Talbot, A., Keogh, M.J., Coxhead, J., Santibanez-Koref, M. and Chinnery, P.F.

2016. HDBR Expression: A Unique Resource for Global and Individual Gene Expression Studies during Early Human Brain Development. *Frontiers in Neuroanatomy*. **10**.

Liu, B. and Qian, S.-B. 2016. Characterizing inactive ribosomes in translational profiling. *Translation*. **4**(1), p.e1138018.

Liu, C. 2004. NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Research*. **33**(Database issue), pp.D112–D115.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., Fleming, J., Siminoff, L., Traino, H., Mosavel, M., Barker, L., Jewell, S., Rohrer, D., Maxim, D., Filkins, D., Harbach, P., Cortadillo, E., Berghuis, B., Turner, L., Hudson, E., Feenstra, K., Sobin, L., Robb, J., Branton, P., Korzeniewski, G., Shive, C., Tabor, D., Qi, L., Groch, K., Nampally, S., Buia, S., Zimmerman, A., Smith, A., Burges, R., Robinson, K., Valentino, K., Bradbury, D., Cosentino, M., Diaz-Mayoral, N., Kennedy, M., Engel, T., Williams, P., Erickson, K., Ardlie, K., Winckler, W., Getz, G., DeLuca, D., MacArthur, D., Kellis, M., Thomson, A., Young, T., Gelfand, E., Donovan, M., Meng, Y., Grant, G., Mash, D., Marcus, Y., Basile, M., Liu, J., Zhu, J., Tu, Z., Cox, N.J., Nicolae, D.L., Gamazon, E.R., Im, H.K., Konkashbaev, A., Pritchard, J., Stevens, M., Flutre, T., Wen, X., Dermitzakis, E.T., Lappalainen, T., Guigo, R., Monlong, J., Sammeth, M., Koller, D., Battle, A., Mostafavi, S., McCarthy, M., Rivas, M., Maller, J., Rusyn, I., Nobel, A., Wright, F., Shabalin, A., Feolo, M., Sharopova, N., Sturcke, A., Paschal, J., Anderson, J.M., Wilder, E.L., Derr, L.K., Green, E.D., Struewing, J.P., Temple, G., Volpi, S., Boyer, J.T., Thomson, E.J., Guyer, M.S., Ng, C., Abdallah, A., Colantuoni, D., Insel, T.R., Koester, S.E., Little, A.R., Bender, P.K., Lehner, T., Yao, Y., Compton, C.C., Vaught, J.B., Sawyer, S., Lockhart, N.C., Demchok, J. and Moore, H.F. 2013. The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*. **45**(6), pp.580–585.

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

Ma, L., Bajic, V.B. and Zhang, Z. 2013. On the classification of long non-coding RNAs. *RNA Biology*. **10**(6), pp.924–933.

Maday, S., Twelvetrees, A.E., Moughamian, A.J. and Holzbaur, E.L.F. 2014. Axonal Transport: Cargo-Specific Mechanisms of Motility and Regulation. *Neuron*. **84**(2), pp.292–309.

Maden, M. 2007. Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nature Reviews Neuroscience*. **8**(10), pp.755–765.

Magny, E.G., Pueyo, J.I., Pearl, F.M.G., Cespedes, M.A., Niven, J.E., Bishop, S.A. and Couso, J.P. 2013. Conserved Regulation of Cardiac Calcium Uptake by Peptides Encoded in Small Open Reading Frames. *Science*. **341**(6150), pp.1116–1120.

Marques, A.C. and Ponting, C.P. 2009. Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. *Genome Biology*. **10**(11), p.R124.

Martens, J.A., Laprade, L. and Winston, F. 2004. Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature*. **429**(6991), pp.571–574.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011*. **17**(1), p.3.

Mas-Ponte, D., Carlevaro-Fita, J., Palumbo, E., Hermoso Pulido, T., Guigo, R. and Johnson, R. 2017. LncATLAS database for subcellular localization of long noncoding RNAs. *RNA*. **23**(7), pp.1080–1087.

Massone, S., Vassallo, I., Fiorino, G., Castelnuovo, M., Barbieri, F., Borghi, R., Tabaton, M., Robello, M., Gatta, E., Russo, C., Florio, T., Dieci, G., Cancedda, R. and Pagano, A. 2011. 17A, a novel non-coding RNA, regulates GABA B alternative splicing and signaling in response to inflammatory stimuli and in Alzheimer disease. *Neurobiology of Disease*. **41**(2), pp.308–317.

Matheisl, S., Berninghausen, O., Becker, T. and Beckmann, R. 2015. Structure of a human translation termination complex. *Nucleic Acids Research*. **43**(18), pp.8615–8626.

Matsumoto, A., Pasut, A., Matsumoto, M., Yamashita, R., Fung, J., Monteleone, E., Saghatelian, A., Nakayama, K.I., Clohessy, J.G. and Pandolfi, P.P. 2017. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature*. **541**(7636), pp.228–232.

McGlincy, N.J. and Ingolia, N.T. 2017. Transcriptome-wide measurement of translation by ribosome profiling. *Methods*. **126**, pp.112–129.

Meng, L., Person, R.E. and Beaudet, A.L. 2012. Ube3a-ATS is an atypical RNA polymerase II transcript that represses the paternal expression of Ube3a. *Human Molecular Genetics*. **21**(13), pp.3001–3012.

Mercer, T.R., Neph, S., Dinger, M.E., Crawford, J., Smith, M.A., Shearwood, A.-M.J., Haugen, E., Bracken, C.P., Rackham, O., Stamatoyannopoulos, J.A., Filipovska, A. and Mattick, J.S. 2011. The Human Mitochondrial Transcriptome. *Cell*. **146**(4), pp.645–658.

Michel, A.M., Mullan, J.P.A., Velayudhan, V., O'Connor, P.B.F., Donohue, C.A. and Baranov, P. V 2016. RiboGalaxy: A browser based platform for the alignment, analysis and visualization of ribosome profiling data. *RNA Biology*. **13**(3), pp.316–319.

Miller, I., Min, M., Yang, C., Tian, C., Gookin, S., Carter, D. and Spencer, S.L. 2018. Ki67 is a Graded Rather than a Binary Marker of Proliferation versus Quiescence. *Cell Reports*. **24**(5), pp.1105-1112.e5.

Minati, L., Firrito, C., Piano, A. Del, Peretti, A., Sidoli, S., Peroni, D., Belli, R., Gandolfi, F., Romanel, A., Bernabò, P., Zasso, J., Quattrone, A., Guella, G., Lauria, F., Viero, G. and Clamer, M. 2020. One-shot analysis of translated mammalian lncRNAs with AHARIBO. *bioRxiv*.

Miyagawa, R., Tano, K., Mizuno, R., Nakamura, Y., Ijiri, K., Rakwal, R., Shibato, J., Masuo, Y., Mayeda, A., Hirose, T. and Akimitsu, N. 2012. Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA*. **18**(4), pp.738–751.

Modarresi, F., Faghihi, M.A., Lopez-Toledano, M.A., Fatemi, R.P., Magistri, M., Brothers, S.P., van der Brug, M.P. and Wahlestedt, C. 2012. Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nature Biotechnology*. **30**(5), pp.453–459.

Mohammad, L., Wiseman, J., Erickson, S. and Yang, G. 2019. Protein Synthesis and Translational Control in Neural Stem Cell Development and Neurogenesis *In*: W. Sossin, ed. *The Oxford Handbook of Neuronal Protein Synthesis* [Online]. Oxford University Press. Available from: http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780190686307.001.0001/oxfordhb-9780190686307-e-21.

Molnár, Z., Métin, C., Stoykova, A., Tarabykin, V., Price, D.J., Francis, F., Meyer, G.,

Dehay, C. and Kennedy, H. 2006. Comparative aspects of cerebral cortical development. *European Journal of Neuroscience*. **23**(4), pp.921–934.
Müllner, E.W. and Kühn, L.C. 1988. A stem-loop in the 3′ untranslated region mediates iron-dependent regulation of transferrin receptor mRNA stability in the cytoplasm. *Cell*. **53**(5), pp.815–825.
Murn, J., Zarnack, K., Yang, Y.J., Durak, O., Murphy, E.A., Cheloufi, S., Gonzalez, D.M., Teplova, M., Curk, T., Zuber, J., Patel, D.J., Ule, J., Luscombe, N.M., Tsai, L.-H., Walsh, C.A. and Shi, Y. 2015. Control of a neuronal morphology program by an RNA-binding zinc finger protein, Unkempt. *Genes & Development*. **29**(5), pp.501–512.
Mus, E., Hof, P.R. and Tiedge, H. 2007. Dendritic BC200 RNA in aging and in Alzheimer{\textquoteright}s disease. *Proceedings of the National Academy of Sciences*. **104**(25), pp.10679–10684.
Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. 2008. The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science*. **320**(5881), pp.1344–1349.
Nektaria Andronikou, Yu, X., Wei, J., Sridharan, M., Quintanilla, R., Lakshmipathy, U., Welch, P. and Jeu, X. de M. du 2015. Lipofectamine 3000 reagent—efficient, reproducible transfection for biologically relevant cell models.
Ng, S.-Y., Bogu, G.K., Soh, B.S. and Stanton, L.W. 2013. The Long Noncoding RNA RMST Interacts with SOX2 to Regulate Neurogenesis. *Molecular Cell*. **51**(3), pp.349–359.
Nirenberg, M.W. and Matthaei, J.H. 1961. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences*. **47**(10), pp.1588–1602.
Noller, H., Hoffarth, V. and Zimniak, L. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science*. **256**(5062), pp.1416–1419.
Novikova, I. V., Hennelly, S.P. and Sanbonmatsu, K.Y. 2012. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Research*. **40**(11), pp.5034–5051.
Ntini, E., Louloupi, A., Liz, J., Muino, J.M., Marsico, A. and Ørom, U.A.V. 2018. Long ncRNA A-ROD activates its target gene DKK1 at its release from chromatin. *Nature Communications*. **9**(1), p.1636.
O'Brien, J., Hayder, H., Zayed, Y. and Peng, C. 2018. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology*. **9**.
Obrig, T.G., Culp, W.J., McKeehan, W.L. and Hardesty, B. 1971. The mechanism by which cycloheximide and related glutarimide antibiotics inhibit peptide synthesis on reticulocyte ribosomes. *The Journal of biological chemistry*. **246**(1), pp.174–181.
Okada, C., Yamashita, E., Lee, S.J., Shibata, S., Katahira, J., Nakagawa, A., Yoneda, Y. and Tsukihara, T. 2009. A High-Resolution Structure of the Pre-microRNA Nuclear Export Machinery. *Science*. **326**(5957), pp.1275–1279.
Okamura, K. and Lai, E.C. 2008. Endogenous small interfering RNAs in animals. *Nature Reviews Molecular Cell Biology*. **9**(9), pp.673–678.
Palazzo, A.F. and Lee, E.S. 2015. Non-coding RNA: what is functional and what is junk? *Frontiers in Genetics*. **6**.
Pavlov, M.Y., Watts, R.E., Tan, Z., Cornish, V.W., Ehrenberg, M. and Forster, A.C. 2009. Slow peptide bond formation by proline and other N-alkylamino acids in

translation. *Proceedings of the National Academy of Sciences*. **106**(1), pp.50–54.
Pearson, H., Daouda, T., Granados, D.P., Durette, C., Bonneil, E., Courcelles, M., Rodenbrock, A., Laverdure, J.-P., Côté, C., Mader, S., Lemieux, S., Thibault, P. and Perreault, C. 2016. MHC class I–associated peptides derive from selective regions of the human genome. *Journal of Clinical Investigation*. **126**(12), pp.4690–4701.
Pedersen, A.G. and Nielsen, H. 1997. Neural network prediction of translation initiation sites in eukaryotes: perspectives for EST and genome analysis. *Proceedings. International Conference on Intelligent Systems for Molecular Biology*. **5**, pp.226–233.
Pereira, I.T., Spangenberg, L., Cabrera, G. and Dallagiovanna, B. 2020. Polysome-associated lncRNAs during cardiomyogenesis of hESCs. *Molecular and Cellular Biochemistry*. **468**(1–2), pp.35–45.
Perry, R.B.-T., Hezroni, H., Goldrich, M.J. and Ulitsky, I. 2018. Regulation of Neuroregeneration by Long Noncoding RNAs. *Molecular Cell*. **72**(3), pp.553-567.e5.
Perry, R.B.-T. and Ulitsky, I. 2016. The functions of long noncoding RNAs in development and stem cells. *Development*. **143**(21), pp.3882–3894.
Pielot, R., Smalla, K.-H., Müller, A., Landgraf, P., Lehmann, A.-C., Eisenschmidt, E., Haus, U.-U., Weismantel, R., Gundelfinger, E.D. and Dieterich, D.C. 2012. SynProt: A Database for Proteins of Detergent-Resistant Synaptic Protein Preparations. *Frontiers in Synaptic Neuroscience*. **4**.
Piletz, J.E., Drivon, J., Eisenga, J., Buck, W., Yen, S., McLin, M., Meruvia, W., Amaral, C. and Brue, K. 2018. Human Cells Grown With or Without Substitutes for Fetal Bovine Serum. *Cell Medicine*. **10**, p.215517901875514.
Piper, M., Anderson, R., Dwivedy, A., Weinl, C., van Horck, F., Leung, K.M., Cogill, E. and Holt, C. 2006. Signaling Mechanisms Underlying Slit2-Induced Collapse of Xenopus Retinal Growth Cones. *Neuron*. **49**(2), pp.215–228.
Ponting, C.P., Oliver, P.L. and Reik, W. 2009. Evolution and Functions of Long Noncoding RNAs. *Cell*. **136**(4), pp.629–641.
Pueyo, J.I. and Couso, J.P. 2008. The 11-aminoacid long Tarsal-less peptides trigger a cell signal in Drosophila leg development. *Developmental Biology*. **324**(2), pp.192–201.
R Core Team 2019. R: A language and environment for statistical computing.
Ramos, A.D., Andersen, R.E., Liu, S.J., Nowakowski, T.J., Hong, S.J., Gertz, C.C., Salinas, R.D., Zarabi, H., Kriegstein, A.R. and Lim, D.A. 2015. The Long Noncoding RNA Pnky Regulates Neuronal Differentiation of Embryonic and Postnatal Neural Stem Cells. *Cell Stem Cell*. **16**(4), pp.439–447.
Rashid, F., Shah, A. and Shan, G. 2016. Long Non-coding RNAs in the Cytoplasm. *Genomics, Proteomics & Bioinformatics*. **14**(2), pp.73–80.
Rasmussen, A.H., Rasmussen, H.B. and Silahtaroglu, A. 2017. The DLGAP family: neuronal expression, function and role in brain disorders. *Molecular Brain*. **10**(1), p.43.
Resch, A.M., Ogurtsov, A.Y., Rogozin, I.B., Shabalina, S.A. and Koonin, E. V 2009. Evolution of alternative and constitutive regions of mammalian 5'UTRs. *BMC Genomics*. **10**(1), p.162.
Rinn, J.L. and Chang, H.Y. 2012. Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry*. **81**(1), pp.145–166.
Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A.,

Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E. and Chang, H.Y. 2007. Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell*. **129**(7), pp.1311–1323.

Robinson, D.G. and Storey, J.D. 2014. subSeq: Determining Appropriate Sequencing Depth Through Efficient Read Subsampling. *Bioinformatics*. **30**(23), pp.3424–3426.

Rodriguez, C.M., Chun, S.Y., Mills, R.E. and Todd, P.K. 2019. Translation of upstream open reading frames in a model of neuronal differentiation. *BMC Genomics*. **20**(1), p.391.

Rom, A., Melamed, L., Gil, N., Goldrich, M.J., Kadir, R., Golan, M., Biton, I., Perry, R.B.-T. and Ulitsky, I. 2019. Regulation of CHD2 expression by the Chaserr long noncoding RNA gene is essential for viability. *Nature Communications*. **10**(1), p.5092.

Ross, R.A., Spengler, B.A. and Biedler, J.L. 1983. Coordinate morphological and biochemical interconversion of human neuroblastoma cells. *Journal of the National Cancer Institute*. **71**(4), pp.741–747.

Roundtree, I.A., Luo, G.-Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., Sha, J., Huang, X., Guerrero, L., Xie, P., He, E., Shen, B. and He, C. 2017. YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *eLife*. **6**.

Rstudio, T. 2016. RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.

Ruiz-Orera, J. and Albà, M.M. 2019. Conserved regions in long non-coding RNAs contain abundant translation and protein–RNA interaction signatures. *NAR Genomics and Bioinformatics*. **1**(1), pp.e2–e2.

Ryu, S.J., An, H.J., Oh, Y.S., Choi, H.R., Ha, M.K. and Park, S.C. 2008. On the role of major vault protein in the resistance of senescent human diploid fibroblasts to apoptosis. *Cell Death & Differentiation*. **15**(11), pp.1673–1680.

Sabi, R. and Tuller, T. 2014. Modelling the Efficiency of Codon–tRNA Interactions Based on Codon Usage Bias. *DNA Research*. **21**(5), pp.511–526.

Sagner, A. and Briscoe, J. 2019. Establishing neuronal diversity in the spinal cord: a time and a place. *Development*. **146**(22), p.dev182154.

Sahoo, P.K., Smith, D.S., Perrone-Bizzozero, N. and Twiss, J.L. 2018. Axonal mRNA transport and translation at a glance. *Journal of Cell Science*. **131**(8), p.jcs196808.

Sanbonmatsu, K.Y. 2016. Towards structural classification of long non-coding RNAs. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. **1859**(1), pp.41–45.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.-L. and Natoli, G. 2010. A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers J. S. Mattick, ed. *PLoS Biology*. **8**(5), p.e1000384.

Scheele, C., Petrovic, N., Faghihi, M.A., Lassmann, T., Fredriksson, K., Rooyackers, O., Wahlestedt, C., Good, L. and Timmons, J.A. 2007. The human PINK1 locus is regulated in vivo by a non-coding natural antisense RNA during modulation of mitochondrial function. *BMC Genomics*. **8**(1), p.74.

Schmidt, C., Becker, T., Heuer, A., Braunger, K., Shanmuganathan, V., Pech, M., Berninghausen, O., Wilson, D.N. and Beckmann, R. 2016. Structure of the hypusinylated eukaryotic translation factor eIF-5A bound to the ribosome. *Nucleic Acids Research*. **44**(4), pp.1944–1951.

Schneider-Poetsch, T., Ju, J., Eyler, D.E., Dang, Y., Bhat, S., Merrick, W.C., Green,

R., Shen, B. and Liu, J.O. 2010. Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nature Chemical Biology*. **6**(3), pp.209–217.
Schuller, A.P. and Green, R. 2018. Roadblocks and resolutions in eukaryotic translation. *Nature Reviews Molecular Cell Biology*. **19**(8), pp.526–541.
Schuller, A.P., Wu, C.C.-C., Dever, T.E., Buskirk, A.R. and Green, R. 2017. eIF5A Functions Globally in Translation Elongation and Termination. *Molecular Cell*. **66**(2), pp.194-205.e5.
SF, G. and Sunderland, M. 2000. *Developmental Biology* 6th ed. (S. Associates, ed.).
Shigeoka, T., Jung, H., Jung, J., Turner-Bridger, B., Ohk, J., Lin, J.Q., Amieux, P.S. and Holt, C.E. 2016. Dynamic Axonal Translation in Developing and Mature Visual Circuits. *Cell*. **166**(1), pp.181–192.
Shimamoto, Y., Sumizawa, T., Haraguchi, M., Gotanda, T., Jueng, H.-C., Furukawa, T., Sakata, R. and Akiyama, S.-I. 2006. Direct activation of the human major vault protein gene by DNA-damaging agents. *Oncology reports*. **15**(3), pp.645–652.
Shimozaki, K. 2014. Sox2 transcription network acts as a molecular switch to regulate properties of neural stem cells. *World Journal of Stem Cells*. **6**(4), p.485.
Shukla, C.J., McCorkindale, A.L., Gerhardinger, C., Korthauer, K.D., Cabili, M.N., Shechner, D.M., Irizarry, R.A., Maass, P.G. and Rinn, J.L. 2018. High-throughput identification of <scp>RNA</scp> nuclear enrichment sequences. *The EMBO Journal*. **37**(6).
Siomi, M.C., Sato, K., Pezic, D. and Aravin, A.A. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Reviews Molecular Cell Biology*. **12**(4), pp.246–258.
Sirey, T.M., Roberts, K., Haerty, W., Bedoya-Reina, O., Rogatti-Granados, S., Tan, J.Y., Li, N., Heather, L.C., Carter, R.N., Cooper, S., Finch, A.J., Wills, J., Morton, N.M., Marques, A.C. and Ponting, C.P. 2019. The long non-coding RNA Cerox1 is a post transcriptional regulator of mitochondrial complex I catalytic activity. *eLife*. **8**.
Skreka, K., Schafferer, S., Nat, I.-R., Zywicki, M., Salti, A., Apostolova, G., Griehl, M., Rederstorff, M., Dechant, G. and Hüttenhofer, A. 2012. Identification of differentially expressed non-coding RNAs in embryonic stem cell neural differentiation. *Nucleic Acids Research*. **40**(13), pp.6001–6015.
Smith, C.M. and Steitz, J.A. 1997. Sno Storm in the Nucleolus: New Roles for Myriad Small RNPs. *Cell*. **89**(5), pp.669–672.
Spencer, H.L., Sanders, R., Boulberdaa, M., Meloni, M., Cochrane, A., Spiroski, A.-M., Mountford, J., Emanueli, C., Caporali, A., Brittan, M., Rodor, J. and Baker, A.H. 2020. The LINC00961 transcript and its encoded micropeptide, small regulatory polypeptide of amino acid response, regulate endothelial cell function J. Pearson, ed. *Cardiovascular Research*.
Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T. and Chang, H.Y. 2015. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*. **519**(7544), pp.486–490.
Stadler, P.F., Chen, J.J.-L., Hackermuller, J., Hoffmann, S., Horn, F., Khaitovich, P., Kretzschmar, A.K., Mosig, A., Prohaska, S.J., Qi, X., Schutt, K. and Ullmann, K. 2009. Evolution of Vault RNAs. *Molecular Biology and Evolution*. **26**(9), pp.1975–1991.

Steitz, J.A. and Jakes, K. 1975. How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in Escherichia coli. *Proceedings of the National Academy of Sciences*. **72**(12), pp.4734–4738.

Steward, O. and Levy, W. 1982. Preferential localization of polyribosomes under the base of dendritic spines in granule cells of the dentate gyrus. *The Journal of Neuroscience*. **2**(3), pp.284–291.

Su, X., Zhang, J., Luo, X., Yang, W., Liu, Yanqing, Liu, Yang and Shan, Z. 2019. LncRNA LINC01116 Promotes Cancer Cell Proliferation, Migration And Invasion In Gastric Cancer By Positively Interacting With lncRNA CASC11. *OncoTargets and Therapy*. **Volume 12**, pp.8117–8123.

Szafron, L.M., Balcerak, A., Grzybowska, E.A., Pienkowska-Grela, B., Felisiak-Golabek, A., Podgorska, A., Kulesza, M., Nowak, N., Pomorski, P., Wysocki, J., Rubel, T., Dansonka-Mieszkowska, A., Konopka, B., Lukasik, M. and Kupryjanczyk, J. 2015. The Novel Gene CRNDE Encodes a Nuclear Peptide (CRNDEP) Which Is Overexpressed in Highly Proliferating Tissues R. M. Lafrenie, ed. *PLOS ONE*. **10**(5), p.e0127475.

Tamariz, E. and Varela-EchavarrÃ-a, A. 2015. The discovery of the growth cone and its influence on the study of axon guidance. *Frontiers in Neuroanatomy*. **9**.

Tani, H., Torimura, M. and Akimitsu, N. 2013. The RNA Degradation Pathway Regulates the Function of GAS5 a Non-Coding RNA in Mammalian Cells A. Kanai, ed. *PLoS ONE*. **8**(1), p.e55684.

Tao, Y. and Zhang, S.-C. 2016. Neural Subtype Specification from Human Pluripotent Stem Cells. *Cell Stem Cell*. **19**(5), pp.573–586.

Tennyson, V.M. 1970. THE FINE STRUCTURE OF THE AXON AND GROWTH CONE OF THE DORSAL ROOT NEUROBLAST OF THE RABBIT EMBRYO. *The Journal of Cell Biology*. **44**(1), pp.62–79.

Thebault, P., Boutin, G., Bhat, W., Rufiange, A., Martens, J. and Nourani, A. 2011. Transcription Regulation by the Noncoding RNA SRG1 Requires Spt2-Dependent Chromatin Deposition in the Wake of RNA Polymerase II. *Molecular and Cellular Biology*. **31**(6), pp.1288–1300.

Thomson, D.W., Bracken, C.P. and Goodall, G.J. 2011. Experimental strategies for microRNA target identification. *Nucleic Acids Research*. **39**(16), pp.6845–6853.

Tischfield, M.A., Baris, H.N., Wu, C., Rudolph, G., Van Maldergem, L., He, W., Chan, W.-M., Andrews, C., Demer, J.L., Robertson, R.L., Mackey, D.A., Ruddle, J.B., Bird, T.D., Gottlob, I., Pieh, C., Traboulsi, E.I., Pomeroy, S.L., Hunter, D.G., Soul, J.S., Newlin, A., Sabol, L.J., Doherty, E.J., de Uzcátegui, C.E., de Uzcátegui, N., Collins, M.L.Z., Sener, E.C., Wabbels, B., Hellebrand, H., Meitinger, T., de Berardinis, T., Magli, A., Schiavi, C., Pastore-Trossello, M., Koc, F., Wong, A.M., Levin, A. V., Geraghty, M.T., Descartes, M., Flaherty, M., Jamieson, R. V., Møller, H.U., Meuthen, I., Callen, D.F., Kerwin, J., Lindsay, S., Meindl, A., Gupta, M.L., Pellman, D. and Engle, E.C. 2010. Human TUBB3 Mutations Perturb Microtubule Dynamics, Kinesin Interactions, and Axon Guidance. *Cell*. **140**(1), pp.74–87.

Tollervey, D. and Kiss, T. 1997. Function and synthesis of small nucleolar RNAs. *Current Opinion in Cell Biology*. **9**(3), pp.337–342.

Tomari, Y. 2005. Perspective: machines for RNAi. *Genes & Development*. **19**(5), pp.517–529.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., Blencowe, B.J., Prasanth, S.G. and

Prasanth, K. V. 2010. The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. *Molecular Cell*. **39**(6), pp.925–938.

Tsagakis, I., Douka, K., Birds, I. and Aspden, J.L. 2020. Long non-coding RNAs in development and disease: conservation to mechanisms. *The Journal of Pathology*. **250**(5), pp.480–495.

Tsioras, K., Papastefanaki, F., Politis, P.K., Matsas, R. and Gaitanou, M. 2013. Functional Interactions between BM88/Cend1, Ran-Binding Protein M and Dyrk1B Kinase Affect Cyclin D1 Levels and Cell Cycle Progression/Exit in Mouse Neuroblastoma Cells K. Vekrellis, ed. *PLoS ONE*. **8**(11), p.e82172.

Vanzon, A., Mossink, M., Houtsmuller, A., Schoester, M., Scheffer, G., Scheper, R., Sonneveld, P. and Wiemier, E. 2005. Vault mobility depends in part on microtubules and vaults can be recruited to the nuclear envelope. *Experimental Cell Research*.

Vattem, K.M. and Wek, R.C. 2004. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proceedings of the National Academy of Sciences*. **101**(31), pp.11269–11274.

Vollmar, F., Hacker, C., Zahedi, R.-P., Sickmann, A., Ewald, A., Scheer, U. and Dabauvalle, M.-C. 2009. Assembly of nuclear pore complexes mediated by major vault protein. *Journal of Cell Science*. **122**(6), pp.780–786.

Wang, H., Iacoangeli, A., Popp, S., Muslimov, I.A., Imataka, H., Sonenberg, N., Lomakin, I.B. and Tiedge, H. 2002. Dendritic BC1 RNA: Functional Role in Regulation of Translation Initiation. *Journal of Neuroscience*. **22**(23), pp.10232–10241.

Wang, H., Lu, B., Ren, S., Wu, F., Wang, X., Yan, C. and Wang, Z. 2020. Long Noncoding RNA LINC01116 Contributes to Gefitinib Resistance in Non-small Cell Lung Cancer through Regulating IFI44. *Molecular Therapy - Nucleic Acids*. **19**, pp.218–227.

Wang, L., Fan, J., Han, L., Qi, H., Wang, Y., Wang, H., Chen, S., Du, L., Li, S., Zhang, Y., Tang, W., Ge, G., Pan, W., Hu, P. and Cheng, H. 2020. The micropeptide LEMP plays an evolutionarily conserved role in myogenesis. *Cell Death & Disease*. **11**(5), p.357.

Wei, C.-W., Luo, T., Zou, S.-S. and Wu, A.-S. 2018. The Role of Long Noncoding RNAs in Central Nervous System and Neurodegenerative Diseases. *Frontiers in Behavioral Neuroscience*. **12**.

Weinberg, R.A. and Penman, S. 1968. Small molecular weight monodisperse nuclear RNA. *Journal of Molecular Biology*. **38**(3), pp.289–304.

Wells, D.G. 2006. RNA-Binding Proteins: A Lesson in Repression. *Journal of Neuroscience*. **26**(27), pp.7135–7138.

Wickham R.; Henry, L.; Müller, K., H.. F. 2019. dplyr: A Grammar of Data Manipulation.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. .

Wickham, H. 2017. tidyverse: Easily Install and Load the 'Tidyverse'.

Wightman, B., Ha, I. and Ruvkun, G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*. **75**(5), pp.855–862.

Wilusz, J.E., JnBaptiste, C.K., Lu, L.Y., Kuhn, C.-D., Joshua-Tor, L. and Sharp, P.A. 2012. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes & Development*. **26**(21), pp.2392–2407.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Morgan-Warren, R.J., Carter, A.P.,

Vonrhein, C., Hartsch, T. and Ramakrishnan, V. 2000. Structure of the 30S ribosomal subunit. *Nature*. **407**(6802), pp.327–339.

Winzi, M., Casas Vila, N., Paszkowski-Rogacz, M., Ding, L., Noack, S., Theis, M., Butter, F. and Buchholz, F. 2018. The long noncoding RNA lncR492 inhibits neural differentiation of murine embryonic stem cells. *PLOS ONE*. **13**(1), p.e0191682.

Wu, K.Y., Hengst, U., Cox, L.J., Macosko, E.Z., Jeromin, A., Urquhart, E.R. and Jaffrey, S.R. 2005. Local translation of RhoA regulates growth cone collapse. *Nature*. **436**(7053), pp.1020–1024.

Wu, P., Zuo, X., Deng, H., Liu, X., Liu, L. and Ji, A. 2013. Roles of long noncoding RNAs in brain development, functional diversification and neurodegenerative diseases. *Brain research bulletin*. **97**, pp.69–80.

Wutz, A., Rasmussen, T.P. and Jaenisch, R. 2002. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nature Genetics*. **30**(2), pp.167–174.

Xiang, J.-F., Yin, Q.-F., Chen, T., Zhang, Y., Zhang, X.-O., Wu, Z., Zhang, S., Wang, H.-B., Ge, J., Lu, X., Yang, L. and Chen, L.-L. 2014. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Research*. **24**(5), pp.513–531.

Xin, X., Mains, R.E. and Eipper, B.A. 2004. Monooxygenase X, a Member of the Copper-dependent Monooxygenase Family Localized to the Endoplasmic Reticulum. *Journal of Biological Chemistry*. **279**(46), pp.48159–48167.

Yao, Y., Jin, S., Long, H., Yu, Yingting, Zhang, Z., Cheng, G., Xu, C., Ding, Y., Guan, Q., Li, N., Fu, Suneng, Chen, X.-J., Yan, Y.-B., Zhang, H., Tong, P., Tan, Y., Yu, Yang, Fu, Shushu, Li, J., He, G.-J. and Wu, Q. 2015. RNAe: an effective method for targeted protein translation enhancement by artificial non-coding RNA with SINEB2 repeat. *Nucleic Acids Research*. **43**(9), pp.e58–e58.

Yonath, A., Müssig, J. and Wittmann, H.G. 1982. Parameters for crystal growth of ribosomal subunits. *Journal of Cellular Biochemistry*. **19**(2), pp.145–155.

Yoon, J.-H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G. and Gorospe, M. 2012. LincRNA-p21 Suppresses Target mRNA Translation. *Molecular Cell*. **47**(4), pp.648–655.

Young, R.S., Marques, A.C., Tibbit, C., Haerty, W., Bassett, A.R., Liu, J.-L. and Ponting, C.P. 2012. Identification and Properties of 1,119 Candidate LincRNA Loci in the Drosophila melanogaster Genome. *Genome Biology and Evolution*. **4**(4), pp.427–442.

Young, S.K. and Wek, R.C. 2016. Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response. *Journal of Biological Chemistry*. **291**(33), pp.16927–16935.

Yu, C.-Y. and Kuo, H.-C. 2019. The emerging roles and functions of circular RNAs and their generation. *Journal of Biomedical Science*. **26**(1), p.29.

Zeng, C. and Hamada, M. 2018. Identifying sequence features that drive ribosomal association for lncRNA. *BMC Genomics*. **19**(S10), p.906.

Zhang, B., Gunawardane, L., Niazi, F., Jahanbani, F., Chen, X. and Valadkhan, S. 2014. A Novel RNA Motif Mediates the Strict Nuclear Localization of a Long Noncoding RNA. *Molecular and Cellular Biology*. **34**(12), pp.2318–2329.

Zhang, P., Wu, W., Chen, Q. and Chen, M. 2019. Non-Coding RNAs and their Integrated Networks. *Journal of Integrative Bioinformatics*. **16**(3).

Zhang, X., Zhou, Y., Chen, S., Li, W., Chen, W. and Gu, W. 2019. LncRNA MACC1-AS1 sponges multiple miRNAs and RNA-binding protein PTBP1. *Oncogenesis*.

8(12), p.73.

Zhao, X., Tang, Z., Zhang, H., Atianjoh, F.E., Zhao, J.-Y., Liang, L., Wang, W., Guan, X., Kao, S.-C., Tiwari, V., Gao, Y.-J., Hoffman, P.N., Cui, H., Li, M., Dong, X. and Tao, Y.-X. 2013. A long noncoding RNA contributes to neuropathic pain by silencing Kcna2 in primary afferent neurons. *Nature Neuroscience*. **16**(8), pp.1024–1031.

Zhu, P., Wu, J., Wang, Y., Zhu, X., Lu, T., Liu, B., He, Luyun, Ye, B., Wang, S., Meng, S., Fan, D., Wang, J., Yang, L., Qin, X., Du, Y., Li, C., He, Lei, Ren, W., Wu, X., Tian, Y. and Fan, Z. 2018. LncGata6 maintains stemness of intestinal stem cells and promotes intestinal tumorigenesis. *Nature Cell Biology*. **20**(10), pp.1134–1144.