

Title	RNA G-quadruplexes and polyamines effects on translation: data analysis and development of software for ribosome pause prediction
Authors	Kumari, Romika
Publication date	2018
Original Citation	Kumari, R. 2018. RNA G-quadruplexes and polyamines effects on translation: data analysis and development of software for ribosome pause prediction. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2018, Romika Kumari. - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2026-09-21 17:08:56
Item downloaded from	https://hdl.handle.net/10468/6257

RNA G-quadruplexes and polyamines effects on translation: Data analysis and development of software for ribosome pause prediction

by

Romika Kumari

**Thesis in fulfilment for the degree of
PhD (Science)**

National University of Ireland, Cork.

School of Biochemistry and Cell Biology

January 2018

Head of School: Professor Rosemary O'Connor
Supervisor: Dr Pavel Baranov

Table of Contents

1. G-quadruplex mediated translation regulation	9
1.1 Introduction	10
1.2 Role of G-quadruplex in cap-dependent translation	13
1.3 Role of G-quadruplex in cap-independent translation	18
1.4 Role of G-quadruplexes in Repeat-associated non-AUG translation	20
1.5 G-quadruplex stabilizing drugs	23
1.6 Conclusion, challenges and perspective	25
2. Analysis of translational response to pharmacological stabilization of G-quadruplexes using ribosome profiling	27
2.1 Introduction	27
2.2 Results	30
2.2.1 Distribution of G-quadruplexes	30
2.2.2 Analysis of ribosome profiling footprint densities in the vicinity of GQs motifs	34
2.2.3 Preliminary analysis of the Ribo-seq and RNA-seq samples	35
2.2.4 Differential gene expression analysis	37
2.2.5 Ribosome pauses prediction	42
2.3 Discussion	43
2.4 Methods	45
3. PausePred and Rfeet: Webtools for inferring ribosome pauses and visualizing footprint density from ribosome profiling data	48
3.1 Introduction	49
3.2 Results	51
3.2.1 PausePred	51
3.2.2 Illustration of PausePred Usage	55

3.2.3 Testing PausePred on ribosome profiling datasets from yeast, rat and human.....	60
3.2.4 Detection of previously reported pauses using PausePred.....	63
3.3 Discussion	66
3.4 Methods	67
4. Translational pauses prediction using polyamine-stimulated ribosome profiling data	70
4.1 Introduction.....	70
4.1.1 Function of polyamines.....	71
4.2 Results	74
4.3 Discussion	84
4.4 Methods	84
5. Bibliography	86
6. Appendices.....	109
Appendix 1: Predicting the role of Rps6 phosphorylation in global translation using ribosome profiling data.	109
Appendix 2: Specific reverse transcriptase slippage at the HIV ribosomal frameshift sequence: potential implications for modulation of GagPol synthesis	
123	
Appendix 3: Identifying the allelic expression in the autism susceptible gene Sprouty3 (<i>SPRY3</i>).....	136
Appendix 4: Ribosome profiling and RNA-seq samples used for analysis in section 2.2.2.....	145

Declaration

This thesis is my own work and has not been submitted for another degree, either at University College Cork or elsewhere.

Signed: _____

Romika Kumari

Abbreviations

Abbreviation	Term
GQ	G-quadruplex
GQs	G-quadruplexes
RAN	Repeat associated non-AUG
IRES	Internal ribosome entry site
CDS	Coding sequence
PCR	Polymerase chain reaction
Ribo-seq	Ribosome profiling
RNA-seq	Randomly fragmented RNA sequencing
UTR	Untranslated region
KEGG	Kyoto Encyclopedia of Genes and Genomes
TE	Translation efficiency
rRNA	Ribosomal RNA
NCBI	National Center for Biotechnology Information
ORF	Open reading frame
SRA	Sequence Read Archive
uORF	Upstream open reading frame
CSV	Comma separated values
EFP	Elongation factor P
WT	Wild-type
BAM	Binary alignment map
HTML	Hypertext Markup Language
GEO	Gene Expression Omnibus
ODC	Ornithine decarboxylase
DFMO	Difluoromethylornithine
EMBOSS	European Molecular Biology Open Software Suite
Nt	Nucleotide
Rps6	Ribosomal protein S6
CB	Cerebellar vermis
SNV	Single nucleotide variant
ASD	Autism spectrum disorder

Acknowledgements

If you are successful, it is because somewhere, sometime, someone gave you a life or an idea that started you in the right direction.

- **Melinda Gates**

To begin with, I wish to thank my family as this thesis would not have been accomplished without their love and support.

I would like to thank my PhD supervisor, Dr Pavel Baranov (aka Pasha) for giving me the opportunity to be a member of LAPTI (The LAboratory of Post-Transcriptional control and bioInformatics) lab. I am grateful to him for providing invaluable advises when required. He has always been there when I needed and provided immediate responses to my queries.

I would like to thank all the members of LAPTI lab for their valuable suggestions and support during course of my PhD. Especially, Dr Audrey Michel for helping me with the PausePred project and for always been there as a great mentor. Also, I would like to thank James Mullan, Patrick O'Connor, Stephen Kiriny and Dmitry Andreev for helping me with different projects. I am grateful to all my collaborators (Dr Christophe Penno, Dr John Atkins, Dr Tom Moore, Dr Seda Yerlikaya, Dr Robbie Loewith) for giving me the opportunity to expand my knowledge by participating in different projects. As well, I am thankful to Science foundation of Ireland for funding my PhD.

I would like to acknowledge various programming related sites available on the google (for e.g. StackOverflow) which allowed me to find quick answers to my questions.

I appreciate the help of my Master's project and NIPGR Lab's mentors (Mr. Harpreet Singh, Dr Andreev Lynn and Dr Rohini Garg) in setting the foundation of my research career. I am grateful to my friends for helping me to get through the difficult times of my PhD and being there when needed. Especially, I appreciate Kanhu Charan Moharana's valuable advises related to programming. And finally, I thank Balaguru Ravikumar for all his love and support.

Abstract

Ribosome profiling (Ribo-seq) technique provides a global snapshot of actively translating ribosomes by capturing the ribosome protected mRNA fragments. Ribo-seq technique can be implemented to study various translation aspects such as differential gene expression at the translation level, ribosomal frameshifting, decoding rate, ribosome pausing etc. The process of translation can be affected locally by the sequence of translated mRNA, e.g. by specific secondary structures as well as globally by cellular conditions, e.g. concentrations of metabolites such as polyamines.

G-quadruplexes is a type of RNA secondary structures that has been implicated in mRNA translation. These G-quadruplex structures have been suggested to play roles in cap-dependent, cap-independent and repeat associated non-AUG (RAN) translation. G- quadruplex structures are also known to stimulate ribosomal frameshifting and ribosome pausing.

In my work, I explored the effect of G-quadruplex on translation by utilizing the publicly available ribosome-profiling datasets to find the relationship between G-quadruplex structures and ribosome footprint density. This analysis revealed the presence of very low ribosome footprint density in the vicinity of the G-quadruplex structures. Further I have analysed the effect of the pharmacological stabilisation of the G-quadruplex on translation using PhenDC3 drug. The treatment of PhenDC3 drug was found to have no significant effect on the translational level. This is described in chapter 2.

The presence of ribosome pauses in the upstream of G-quadruplex regions and the lack of an existing tool to identify ribosome pause sites motivated me to develop a tool for the prediction of the ribosome pauses using ribosome profiling data. This lead to the development of PausePred tool which has been discussed in chapter 3. The PausePred tool can be used to infer ribosome pauses which are scored based on their magnitude relative to the background density within the surrounding area. In addition to the score, PausePred provides the coordinates of the pause location, the footprint density at the pause site and the flanking nucleotide sequence.

Further I utilized PausePred to analyse ribosome pausing using ribosome profiling data obtained under varying concentrations of polyamines (described in chapter 4). I showed that in cells with higher polyamine levels, proline-proline or proline-glycine amino acids are enriched in the E and P-sites of paused ribosomes.

Chapter 1

G-quadruplex mediated translation regulation

If G-quadruplexes form so readily in vitro, Nature will have found a way of using them in vivo - Aaron Klug

G-quadruplex (GQ) is a non-canonical four stranded nucleic acid structures formed by guanine rich nucleotide sequences. RNA secondary structures could affect gene expression at translation level, and G-quadruplexes (GQs) are no exception. RNA GQs are known to function in cap-dependent translation, cap-independent translation and repeat associated non-AUG (RAN) translation. There is evidence suggesting that GQs act as inhibitory elements of cap-dependent translation, through perturbing initiator ribosome complexes scanning from the mRNA 5'-cap towards the translation initiation site. GQs promote cap-independent translation by associating with the internal ribosomal entry site. The RAN translation can be initiated by the direct recruitment of translationally-competent 48S-like pre-initiation complexes to the repeat associated GQs.

The immense enrichment of the GQs in 5' leaders and their high conservation across species signify their regulatory role through their ability to form to a secondary structure. In this chapter, I will critically discuss the role of RNA GQs in translation

focusing on various mechanisms adopted by GQs to inhibit or promote the protein translation.

1.1 Introduction

In the year 1910, it was reported that guanylic acid can lead to gel formation at higher concentration (Bang, 1910). Guanines were later found to have an inherent propensity to self-associate and lead to the formation of a four-stranded helical structure. Analysing this structure using x-ray diffraction technique, confirmed that guanylic acid can assemble into stable tetrameric structures (Gellert, Lipsett, & Davies, 1962). Single stranded nucleic acid sequence containing guanine repeats followed by intermediate loops spontaneously fold into a four stranded structure, wherein adjacent guanines form square planar G-quartets stabilized by hydrogen bonds between them (Bochman, Paeschke, & Zakian, 2012; Murat & Balasubramanian, 2014). Two- three self-stacks of G-quartets leads to the formation of G-quadruplex structure which is further stabilized by the presence of metal cation in the centre cavity (Figure 1.1) (Davis, 2004).

The following sequence pattern represents a potential GQs motif.

(G_X+L_{1-N}G_X+L_{1-N}G_X+L_{1-N}G_X)

Where G stands for guanine, X is number of guanine repeats, L is intermediate loop and N is loop length.

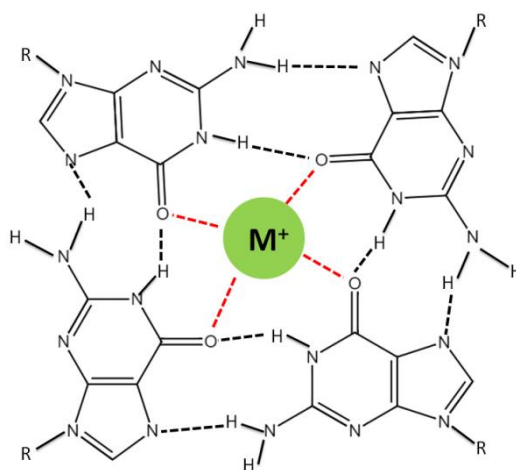


Figure 1.1: The structure of a G-tetrad showing four guanines bonded together through hydrogen bonds, stabilized by the presence of central metal cation.

The stability of a GQs is proportional to the number of G- stacks, and GQs structures with smaller loop lengths tend to be more stable (Huppert, 2010). Stable GQ structures are found to contain a minimum of two G-stacks and with the length of their loop varying from 1-7 nucleotides (Figure 1.2). G-quadruplexes can be formed by DNA, RNA, LNA and PNA, based on the different number of strands involved in the GQ structure formation, it can be either intermolecular or intramolecular. The intermolecular GQ structures can be bimolecular or tetramolecular. The GQs structure which is formed completely from a single strand is termed as intramolecular GQ (Bochman, Paeschke, & Zakian, 2012). G-quadruplex structures can have various topologies depending on the nucleic acid strand orientation and the loop configurations. The same orientation of the all four strands leads to the formation of parallel G-quadruplex structure. As a result, all the guanine's in the G-tetrad will have same glycosidic bond formation. In case of intramolecular parallel GQ structures the loop regions present will be of propeller type (Figure 1.2) (Malgowska, Czajczynska, Gudanis, Tworak, & Gdaniec, 2016). If at least one strand runs in the opposite orientation to the other strands it leads to the formation of an anti-parallel GQ structure. Anti-parallel GQ structures can be of different types depending on the number of strands present in the opposite orientation. Depending on the runs of guanine bases a loop sequence in joining it can be divided in three categories, propeller, lateral and diagonal (Malgowska, Czajczynska, Gudanis, Tworak, &

Gdaniec, 2016) (Figure 1.2). Presence of specific monovalent metal cation in the central cavity of the GQs affects the stability of the structure in following order $K^+ \gg NH_4^+ \gg Na^+ \gg Li^+$ (Guiset Miserachs, Donghi, Borner, Johannsen, & Sigel, 2016; Hardin, Watson, Corregan, & Bailey, 1992; Lane, Chaires, Gray, & Trent, 2008; Wong & Wu, 2003). Further, the stability of GQs structures can be strengthened by using G-quadruplex-interactive drugs such as bisquinolinium compounds, pyridine dicarboxamide derivative 360A, PhenDC3 etc (W. J. Chung, Heddi, Hamon, Teulade-Fichou, & Phan, 2014; De Cian, Delemos, Mergny, Teulade-Fichou, & Monchaud, 2007; Gomez et al., 2010).

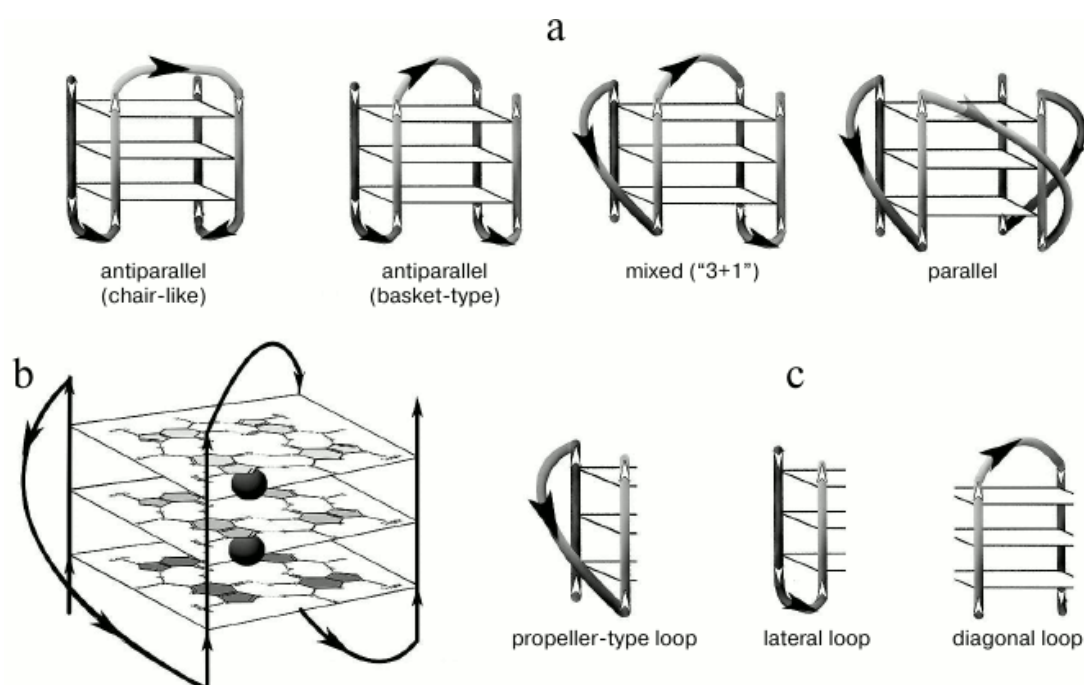


Figure 1.2: Different G-quadruplex topologies. (a) Three G-tetrad stacking to form intramolecular G-quadruplex structures in different topologies. (b) A parallel G-quadruplex structure showing the potassium metal cation (black) in the centre of the structure. (c) Different types of loops present in the G-quadruplex structures. (Ogloblina et al., 2015)

G-quadruplex structures are likely to be found in both DNA and RNA molecules. RNA being single stranded is known to fold into various secondary structure conformations. DNA GQ structures can have both parallel and anti-parallel topologies whereas in case of RNA GQ structures only parallel topology is possible (Tang & Shafer, 2006).

Specific features of RNA molecules lead to increased stability of GQs present in RNA sequences. Presence of additional hydroxyl group in RNAs contributes to enhanced intramolecular interactions within GQs (D. H. Zhang et al., 2010). G-quadruplexes distribution across the genome is non-random and are found to be enriched in the functional regions like promoters, telomeres and 5' leaders (Huppert & Balasubramanian, 2005). Approximately, 3000 human genes are known to contain potential GQs forming sequences in the 5' leader of their encoded mRNA, signifies their role in regulating the gene expression at translational level (Bugaut & Balasubramanian, 2012; Huppert, Bugaut, Kumari, & Balasubramanian, 2008). Several previously reported GQ which play regulatory roles in post-transcriptional processes were found to be conserved in multiple mammalian and non-mammalian species using QGRS-conserve approach (Frees, Menendez, Crum, & Bagga, 2014). The presence of GQs in crucial genomic regions and their strong conservation in multiple species indicate their important role in various biological processes due to their potential of forming a secondary structure (Bugaut & Balasubramanian, 2012; Capra, Paeschke, Singh, & Zakian, 2010; Huppert et al., 2008; Rhodes & Lipps, 2015). The GQs were found to affect protein translation by interacting with the translational machinery. Such interactions have been found to create hindrance in cap binding process of eIF4F complex (Huppert et al., 2008), interfering in recognition of the start codon (Gebauer & Hentze, 2004), inhibiting small ribosomal subunit in scanning and enhancing IRES (Internal ribosome entry site) mediated cap independent translation (Morris, Negishi, Pazsint, Schonhoft, & Basu, 2010; Yu, Teulade-Fichou, & Olsthoorn, 2014).

In this chapter, I will thoroughly discuss the diverse aspects of translational regulation mediated by the presence of G-quadruplex structures.

1.2 Role of G-quadruplex in cap-dependent translation

The genomic information is converted to proteome by the process of translation where messenger RNAs are decoded to amino acids by the translational machinery. The translation has been divided into basic three steps: Initiation, elongation and termination (Gebauer & Hentze, 2004). The majority of the messenger RNA (95-97%

of total cellular mRNAs) initiates translation by the recognition of m⁷G (5') ppp(5')N cap structure at the 5' end of translating mRNA (Sonnenberg & Hinnebusch, 2009). The initiation complex constituting of 40S ribosome, eIF3, eIF2 GTP Met-tRNA_i and eIF4F, binds to the activated mRNA and scans the mRNA to locate the start codon, the 60S subunit later joins with the above complex to form 80S subunit (Jackson, Hellen, & Pestova, 2010; Merrick, 2004). The elongation process involves decoding of an mRNA codon by aminoacyl-tRNA, peptide bond formation and translocation of the tRNA–mRNA complex which presents a new codon at the A-site by moving peptidyl-tRNA from the A site to the P site, leading to the incorporation of amino acids to the elongating peptide. The translation Elongation is advanced by the elongation factors eEF1A and eEF1B (Rodnina & Wintermeyer, 2009). Elongation process is terminated when a ribosome reaches a nonsense codon, during the termination release factors, eRF1, which recognizes all three termination codons, and eRF3, which acts as GTPase, activate the hydrolysis of peptidyl-tRNA at the ribosomal peptidyl transferase centre allowing polypeptide to exit the ribosome (Tate & Brown, 1992). The last step of translation involves the dissociation of 80S ribosome into 40S and 60S, this ribosome subunit splitting is catalysed by eIF3 with the help of eIF3j, eIF1, and eIF1A (Pisarev, Hellen, & Pestova, 2007; Rodnina & Wintermeyer, 2009; Tate & Brown, 1992).

During the translation, the motion of ribosomes along mRNA often coexist with the existence of profound stalls at particular positions (Richter & Collier, 2015). The strong RNA secondary structures whose unwinding is required for ribosome progression and specific nascent peptides interacting with the ribosome are among known causes of such site specific ribosomal halts (Somogyi, Jenner, Brierley, & Inglis, 1993; Tholstrup, Oddershede, & Sorensen, 2012). The presence of free complimentary regions in single stranded mRNA allows it to fold into complex shapes. Such, local nucleotide pairing creates secondary structures such as hairpins, stem–loops and G-quadruplexes (Wan, Kertesz, Spitale, Segal, & Chang, 2011). Guanine repeats with intermediate loops can form a four-stranded structure where guanines interact with each other with Hoogsteen hydrogen bonding and leads to the formation of G-tetrads. Minimum of two of the G-tetrads can stack on top of each other and lead to the formation of a G-quadruplex structure. The stability of the structure is further increased by the presence of a metal cation in the central cavity (Bochman et al., 2012;

Murat & Balasubramanian, 2014). GQs structures are highly enriched in 5' leaders with ~3000 mRNA containing at least one GQs structure in their 5' leader (Kumari, Bugaut, Huppert, & Balasubramanian, 2007). G-quadruplexes conservation in different species and their enrichment in the crucial region of genome signify their important regulatory role.

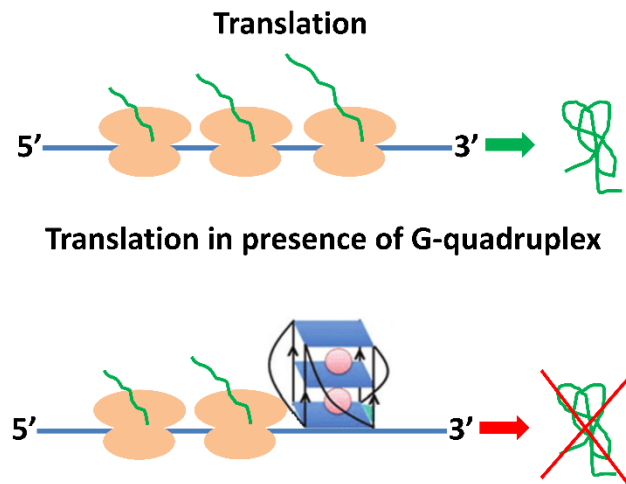


Figure 1.3: Schematic illustration of the translation process in presence and absence of G-quadruplex structure.

Several translation processes such as hindrance in cap binding process of eIF4F complex (Huppert et al., 2008), interfering in recognition of start codon (Gebauer & Hentze, 2004), inhibiting small ribosomal subunit in scanning and enhancing IRES mediated cap independent translation (Morris et al., 2010; Yu et al., 2014) are reported to be affected by the presence of G-quadruplex structures.

During the process of cap dependent translation ATP-dependent DEAD box RNA helicase eIF4A is responsible for unwinding the secondary structures present in the 5' leader which is followed by the recruitment of 43S initiation complex which facilitates the progression of ribosome (Jackson et al., 2010; Merrick, 2004; Sonenberg & Hinnebusch, 2009).

Potent inhibition of translation due to presence of a stable G-quadruplex structure has been observed in number of studies, reviewed in Song et al., 2016 (Song, Perreault, Topisirovic, & Richard, 2016). In the year 2001, it was reported that intramolecular GQs in the 5' leader of Fragile X syndrome linked gene *FMRP* impacts the translation efficiency. The 35 nucleotide (nt) long 5' leader GQ of *FMRP* gene was found to reduce the translational efficiency of a luciferase reporter mRNA by 1.5 folds in comparison to the control mRNA (Schaeffer et al., 2001).

Later in 2007, the translational repression by a GQ was reported in human proto-oncogene gene *NRAS* where presence of GQs in the 5' leader was found to modulate the gene expression at the translational level. The presence of a stable GQ structure in the 5' leader of *NRAS* gene, by acting as a roadblock was found to inhibit the translation progression of ribosome (Kumari et al., 2007).

Inhibition of translation in living eukaryotic cell was first reported in the mRNA of the human *Zic-1* zinc-finger protein. A thermodynamically stable RNA G-quadruplex in the 5' leader was found to repress the protein synthesis by regulating the translation process. By using quantitative RT-PCR it was confirmed that reduced protein levels are due to the translation inhibition (Arora et al., 2008).

In the following years, numerous studies have shown that GQs repress translation in various genes such as *MT3-MMP* (Morris & Basu, 2009), *ESR1* (Balkwill et al., 2009), *BCL-2* (Shahid, Bugaut, & Balasubramanian, 2010), *TRF2* (Gomez et al., 2010), *ADAM10* (Lammich et al., 2011), *TGFβ2* (Agarwala, Pandey, Mapa, & Maiti, 2013) and *EBAG9* (Beaudoin & Perreault, 2010). Table 1.1 lists the various genes and the reported effects of GQs on their translation. Apart from traditional GQs with loop lengths of 1-7nt, there also exist some noncanonical GQs that either consists of surprising lengthy loops structures (>7 nucleotides) or contains a disruption in the G-tracts (Chambers et al., 2015; Huppert & Balasubramanian, 2005). Individual cases of such noncanonical GQs also contribute to hindrance in the translational process (Jodoin et al., 2014). Some genes reported with such noncanonical GQs include *HIRA*, *TOM112* and *APC*, these were found to affect translation significantly which was observed by the decrease in the luciferase activity (Jodoin et al., 2014) (Figure 1.3) (Table 1.1).

Gene name	Role in cap dependent translation	Reference
<i>FMRP</i>	Decrease in translational efficiency by 1.5 folds	(Schaeffer et al., 2001)
<i>NRAS</i>	Translation Inhibition by 80%	(Kumari et al., 2007)
<i>Zic-1</i>	~80% reduction in protein synthesis	(Arora et al., 2008)
<i>MT3-MMP</i>	55% translation inhibition	(Morris & Basu, 2009)
<i>ESR1</i>	6-fold decrease in rate of translation	(Balkwill et al., 2009)
<i>BCL-2</i>	2.3- and 1.9-fold decrease in translation rate.	(Shahid et al., 2010)
<i>TFR2</i>	2.8-fold decrease in the translation	(Gomez et al., 2010)
<i>ADAM10</i>	9.5- to 15-fold decrease the translation.	(Lammich et al., 2011)
<i>TGFβ2</i>	GQs alone- inhibits translation by 56-65% GQs in 5' leader enhances translation by 117%-128%	(Agarwala et al., 2013)
<i>EBAG9</i>	1.8 folds translation inhibition.	(Beaudoin & Perreault, 2010)
<i>HIRA</i>	inhibits translation	(Jodoin et al., 2014)
<i>TOM112</i>	inhibits translation	(Jodoin et al., 2014)
<i>APC</i>	inhibits translation	(Jodoin et al., 2014)

Table 1.1: Examples of G-quadruplex mediated cap-dependent translation. All genes mentioned in the table contains G-quadruplex in 5' leader region of the mRNA.

A transcriptome-wide analysis using ribosome footprinting technique was carried by using eIF4A inhibitor silvestrol in murine T-cell acute lymphoblastic leukaemia (T-ALL) models and primary human T-ALL samples. It was reported that the 12-nucleotide guanine quartet (CGG)₄ motif capable of forming G-quadruplex structures was enriched in the 5' leaders rendering mRNAs surprisingly sensitive to eIF4A, a key factor in cap-dependent translation initiation (Song et al., 2016; Wolfe et al., 2014).

Nevertheless, recently 5' leader G-quadruplex of gene NF-E2-related factor 2 (*NRF2*) and α -synuclein (*SNCA*) have been identified as enhancers of translation. The presence of GQs in the 5' leader of *NRF2* was confirmed by the circular dichroism, Nuclear

Magnetic Resonance (NMR) and dimethylsulfate footprinting analyses. The GQs in 5' leaders were found to be crucial for the *NRF2* gene as its mutation eliminated the oxidative stress activation of *NRF2* 5' leader. This study reported an alternative mechanism of cellular defence incorporating de novo *NRF2* protein translation lead by the EF1a interaction with the G-quadruplex in the *NRF2* 5' leader during oxidative stress (reviewed in Fay et al., 2017 (Fay, Lyons, & Ivanov, 2017)).

The role of *SNCA* long and GC-rich 5' leader capable of folding to a G-quadruplex structure was explored for its role in translation. The cap-dependent translation of *SNCA* attenuated with rapamycin treatment was found to be enhanced by the presence of 5' leader which was further confirmed to contain an IRES element (reviewed in Fay et al., 2017 (Fay, Lyons, & Ivanov, 2017)).

1.3 Role of G-quadruplex in cap-independent translation

In the year 1988, it was reported that picornaviruses lacking 5' mRNA cap are capable of initiating translation from an internal ribosome entry site (IRES) and this mechanism of translation was named as cap-independent translation (Jang et al., 1988; Pelletier & Sonenberg, 1988). This alternative translation mechanism involves internal ribosome entry site (IRES) which is mostly localized in the 5' leaders immediate upstream of the start codon (Pinkstaff, Chappell, Mauro, Edelman, & Krushel, 2001). In few cases, the presence of IRES element has been reported in the coding region leading to the synthesis of truncated protein (Grover, Candeias, Fahraeus, & Das, 2009; Komar et al., 2003). The IRES elements are known to have a complex structure with stem loops and pseudoknots, with no common structural motif. Compared to the viral IRES, the cellular counterparts are of more structural diversity and are less stable (Komar & Hatzoglou, 2011).

Until recently this unconventional translation initiation was not widely recognized in cellular transcripts. Recent reports provide evidences with respect to the physiological functions of IRES elements in supporting the translation initiation of IRES-containing cellular mRNAs, these mRNAs are usually found to have 5' leaders which are either too long or highly structured or contain GC rich regions or encompasses several upstream start codons, making them less efficient for the conventional cap-dependent

mechanism (Park, Lee, Blais, Bell, & Pelletier, 2005). It has also been reported that under stress conditions when the level of cap-dependent translation reduces, IRES mediated translation takes over. These IRES elements in cellular mRNAs are highly enriched in mRNAs coding for proteins related to stress and apoptosis (Mokrejš et al. 2009).

The mechanism of cellular IRES mediated translation initiation is largely hypothetical as extensive systematic studies have not been performed in this field. The speculated mechanism of cap-independent translation initiation begins with positioning of the initiation codon at the ribosomal P-site which is achieved with the help of canonical initiation factors, ITAFs and 40S ribosomal subunits in the absence of ribosomal scanning from the 5' end (Hellen & Sarnow, 2001; Stoneley & Willis, 2004). The 'land and scan' mechanism of IRES elements which is typical in picornavirus has also been reported in certain cellular IRESs (Spriggs et al., 2009). Some cellular IRESs including Gtx and IGF1R have been reported to function through Shine-Dalgarno-like interaction between the IRES and the 18S rRNA (Chappell, Edelman, & Mauro, 2004; Meng, Jackson, Shcherbakov, Choi, & Blume, 2010).

Most examples in the literature describe G-quadruplex structures as translation inhibitors. Whereas, in year 2003, it was reported that IRES element present in the 5' leader of human fibroblast growth factor 2 (*FGF2*) mRNA contains a G-quadruplex structure which constitutes novel structural determinants of IRES function. The *FGF2* IRES element was found to contain two stem-loop regions and an RNA G-quadruplex structure composed of five G-quartet motif which was identified by performing chemical and enzymatic footprinting experiments to probe the structure of the 484-nt *FGF2* mRNA's 5' leader. This was the first study to provide the evidence of a G-quadruplex involved in an IRES element (Bonnal et al., 2003).

The first functional and structural evidence of a G-quadruplex essential for the IRES mediated translation was provided by Morris et al., 2010 (Morris et al., 2010). It was reported that a 5' leader of human vascular endothelial growth factor (*hVEGF*) mRNA contains G-rich sequence which adopts an RNA G-quadruplex structure that is essential for IRES-mediated translation initiation. The ~1000 nt long 5' leader of *hVEGF* contains two separate IRES elements. The 293-nt IRES structure was identified to contain a G-quadruplex structure of 17 nucleotide long using RNase T1

and dimethylsulfate to probe. RNA G-quadruplex formation was found to be crucial for the cap-independent translation initiation in HeLa cells, which was identified by performing site directed mutational analysis of the *hVEGF* IRES-A in the context of a bicistronic dual-luciferase reporter vector. It was found that the mutants having insufficient numbers of guanines to lead to a G-quadruplex structure could completely reduce the activity of the IRES element. It was proposed that G-quadruplex structures can act as switch elements to tune IRES-mediated translation initiation (Morris et al., 2010).

Whereas contradictory to the above-mentioned results, Cammas et al., 2015 (Cammass et al., 2015) reported that G-quadruplex within the *VEGF* IRES is dispensable for its ability of cap-independent translation initiation. This study provides evidence that mutating or replacing the potential GQs forming sequence did not affect the IRES activity. They have mentioned that mutations generated by Morris et al., 2010 (Morris et al., 2010) can introduce an inhibitory upstream open reading frame which can be responsible for the observed inhibition of cap-independent translation. Further they examined that intrinsically stable or ligand-stabilized G-quadruplexes can function as inhibitor of IRES mediated translation, confirming the GQs role as repressors of mRNA translation. It is speculated that cap-independent translation repression is performed by preventing the conformational changes necessary to recruit the ribosome (Cammass et al., 2015).

1.4 Role of G-quadruplexes in Repeat-associated non-AUG translation

Repeat associated non-AUG (RAN) translation mechanism has been observed in some repeat associated disorders such as Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Currently, the process of RAN translation is poorly understood. It was reported that in this non-canonical process, translation is not initiated by formylated methionine, it can occur in all three frames in both sense and antisense strands and is usually provoked by expansion of repeats leading to production of toxic repeat proteins (Pearson, 2011; Zu et al., 2011).

The fragile X mental retardation 1 (*FMR1*) gene was reported to contain CGG repeats in the 5' leader whose expansion has been speculated to cause fragile X-associated tremor/ataxia syndrome (FXTAS) (reviewed in Green et al., 2016 (Green, Linsalata, & Todd, 2016)). It was proposed that RAN translation of *FMR1* gene initiates with a cap-dependent mechanism followed by scanning of 5' leader, ribosomes are speculated to stall at the secondary structure formed by the CGG repeats leading to the non-AUG translation initiation resulting in production of repeat peptides (Kearse et al., 2016).

The expansion of hexameric intronic repeats (GGGGCC)_n of gene *C9ORF72* has been observed in Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) patients. These hexameric repeats are known to translate in all three frames using RAN mechanism leading to the production of toxic dipeptide repeat proteins which were detected in ALS/FTD patients (Ash et al., 2013). These repeats were found to be capable of forming G-quadruplex structures. After predicting that G-rich pattern in *C9ORF72* has higher probability of forming G-quadruplex using in-silico techniques, NMR and CD spectroscopy techniques were used to confirm that the *C9ORF72* hexanucleotide expansion can form a stable G-quadruplex (Fratta et al., 2012). It was postulated that binding of G-quadruplex distorting proteins such as TMPyP4 to the repeats (GGGGCC)_n of gene *C9ORF72* can be potential therapeutic avenue for curing ALS (Zamiri, Reddy, Macgregor, & Pearson, 2014). Overall these discoveries underline the role of small molecules binding to G-rich repeats as possible FTD/ALS therapeutic.

Following the interaction of G-quadruplex with ribosomal proteins and there binding to the 40S ribosomal subunit, it has proposed that r(GGGGCC)_n repeat of the *C9orf72* gene may directly engage translationally competent 48S-like pre-initiation complexes to begin RAN translation and generate dipeptides. This hypothesis is currently under analysis (Fay et al., 2017).

1.5 G-quadruplex binding proteins

For the movement of the either replication, transcription or translation machinery on the corresponding nucleic acid strand a GQs structure must be unfolded. To achieve this specific GQs recognizing and unwinding, there are proteins are present in the cell.

The number of helicases have been reported to unfold the GQs structures during the process of replication, transcription and translation (reviewed in (Mendoza, Bourdoncle, Boule, Brosh, & Mergny, 2016)).

The helicases have been divided into six superfamilies (SF) and PiF1 member of superfamily 1 (SF1) was first helicase reported to be involved in the metabolism of the GQs forming sequences. The G-rich human minisatellite CEB1 inserted in the *Saccharomyces cerevisiae* genome, was reported to form a stable G-quadruplex structure and PiF1 was found to be responsible for the unwinding of these structures (Ribeyre et al., 2009). In following years multiple reports have provided the evidence of Pif1 helicase GQ structure unwinding functionality (reviewed in (Mendoza et al., 2016)). The DNA2 helicase, another member of SF1, was reported to recognize and unfold telomeric GQs structures by cleaving (Lin et al., 2013).

The members of superfamily 2 (SF2) and subfamily Fe-S helicases including FANCI, DDX11 and RTEL1 have also reported to be involved in the unwinding of GQ structures. The FANCI helicase catalyses the unwinding of GQ in an ATP-dependent manner (reviewed in (Mendoza et al., 2016)). The DDX11 helicase is involved in unwinding of two stranded anti parallel DNA GQ structures (Bharti et al., 2014). The mechanism in which members of Fe-S helicase subfamily perform the unwinding of GQ structures has been broadly discussed in (Mendoza et al., 2016). In recent years multiple helicases belonging to superfamilies 3,4 and 5 have been reported to unfold GQ structures (reviewed in (Mendoza et al., 2016)) (Table 1.2).

Superfamily	Subfamily	Helicase name	Substrate	Directionality
SFI		Pif1	DNA	5' to 3'
		DNA2	DNA	5' to 3'
SF2	Fe-S helicases	FANCI	DNA	5' to 3'
		DDX11	DNA	5' to 3'
		RTEL1	DNA	5' to 3'
	RecQ	BLM	DNA	3' to 5'
		WRN	DNA	3' to 5'
		Yeast sgs1	DNA	3' to 5'
		Bacterial RecQ	DNA	3' to 5'
		DHX9	RNA and DNA	3' to 5'
SF3		SV40 T-ag	DNA	3' to 5'
SF4		Twinkle	mtDNA	5' to 3'
SF5		RHAU	RNA and DNA	3' to 5'

Table 1.2: List of helicases which can unwind G-quadruplex structure (taken from (Mendoza et al., 2016))

1.6 G-quadruplex stabilizing drugs

The identification of the G-quadruplexes (GQs) as the novel targets for various diseases including neurodegenerative, neurodevelopmental disorders, diabetes, cardiovascular diseases and cancer has provoked great interest in modelling compounds that can bind with GQs. The GQs are known to be stabilized by the presence of an alkali metal ion located at the centre of the G-tetrad which are involved in electrostatic interactions with the guanine carbonyl groups. The potential modes of ligand binding which G-quadruplex structures include are: loop binding, groove binding and tetrad- stacking (Kaplan, Berber, Hekim, & Doluca, 2016). The chemically distinct environments of the G-tetrad end lead to the structural heterogeneity which constrains the development of selective GQs ligands (Le, Di Antonio, Chan, & Balasubramanian, 2015).

In the year 1999, it was reported that metalloporphyrins can interact with the human telomeric DNA G-quadruplexes and lead to 50% inhibition of the telomerase activity

(Izbicka et al., 1999). It was suggested that the mode of interaction of G-quadruplex and porphyrin-based and porphyrin-like metal complexes was based on the π stacking of the metalloporphyrin system on top of the G-tetrads at the terminal of the G-quadruplex which was favoured by experimental data and computational modelling (Shi, Wheelhouse, Sun, & Hurley, 2001). It was found that synthetic metalloporphyrin are capable of forming electrostatic interaction with the loops and grooves of the G-quadruplex in the negative DNA backbone (Exogenous Ligand Binding Property of a Heme–DNA Coordination Complex). The novel G-quadruplexes stabilizers, funtumine guanylhyazone and peimine, non-planar molecules without aromatic moieties were found to interact with G-quadruplexes through the groove binding mode (Brassart et al., 2007; Li et al., 2009).

In the year 2007, bisquinolinium compounds were claimed to be most potent G-quadruplex binders by exhibiting remarkable efficiency for both stabilization and selectivity (De Cian et al., 2007). The following year, a novel synthetic molecule pyridostanin, was found to have an enhanced stabilization of the human telomeric G-quadruplex with high selectivity relative to double-stranded DNA (Rodriguez et al., 2008). This compound was used for the targeted detection of G-quadruplexes in cellular RNAs (Kwok & Balasubramanian, 2015). A compound called carboxypyridostanin (carbonyPDS) which was identified by *in situ* 'click chemistry' was shown to have a preference for RNA G-quadruplexes as compared to DNA G-quadruplexes (Biffi, Tannahill, McCafferty, & Balasubramanian, 2013).

The ligands TmPyP4, PhenDC3 and PDS were found to have differential chemotype-specific bindings for individual G-tetrads of a model genomic G-quadruplex structure. The quantitative interrogation of mutually exclusive ligand binding interactions at opposed G-tetrads was performed by equilibrium-binding assay mediated by ligand-induced fluorescence quenching of fluorophore-labelled G-quadruplex structures (Le et al., 2015).

A unique collection of the reported G-quadruplex ligands has been provided in the G-quadruplex ligands database (G4LDB) which streamlines the ligand/drug discovery targeting G-quadruplexes (Li et al., 2013).

The finding of these diverse set of synthetic compounds for binding and stabilization of the G-quadruplex structures can be beneficial to produce drugs related to the diseases caused by G-quadruplexes.

1.7 Conclusion, challenges and perspective

Studying the role of G-quadruplexes in the gene expression regulation has evolved as an interesting area of present-day biology. G-quadruplexes are omnipresent in human genome with higher enrichment in the regulatory elements. Different studies have proved the *in-vivo* and *in-vitro* occurrence of GQ structures by using techniques like circular dichroism, NMR, rG4 and using structure specific antibodies. Several reports have shed light on the role of GQs elements as transcriptional and translational repressors. The G-quadruplex secondary structures are speculated to block the transcriptional and translational machinery by acting as roadblocks resulting in reduced gene expressions (reviewed in Fay et al., 2017 (Fay et al., 2017)). Factors such as the nature of cation, sequence and length of loops influence the structure and stability of GQs structures (Guiset Miserachs et al., 2016; Hardin et al., 1992; Lane et al., 2008; Wong & Wu, 2003).

The mRNA is threaded through ribosome in a non-linear fashion which is often accompanied by the presence of strong pauses at specific locations. The non-linearity in the ribosome movement can be caused by presence of G-quadruplex secondary structures whose unwinding is required for ribosome progression. Ribosomal profiling (Ribo-seq) technique provides a snapshot of ribosome locations across all mRNAs by capturing mRNA fragments protected by ribosomes (Ingolia et al., 2009). The ribosome footprint density information can aid in the identification of change in gene expression and pause detection. The ribo-seq technique provides a mean to study genome wide effects on protein translation of various treatments. So far only a single study has utilized ribo-seq approach to study the genome wide effect caused by presence of GQs (Wolfe et al., 2014) by inhibiting the eIF4A helicase activity whereas our in-house analysis revealed this ribo-seq data to be of poor quality. A follow up study it did not find GQs to be involved in sensitivity to eIF4A inhibitors (Iwasaki, Floor, & Ingolia, 2016). The presence of various GQs stabilizing ligands have made it

possible to examine the effect of these structures on the gene expression which can be done at a genome wide scale by employing ribo-seq technique. In chapter 2 we will discuss the translational response to the pharmacological stabilization of the G-quadruplex structures using ribosome profiling technique.

Chapter 2

Analysis of translational response to pharmacological stabilization of G-quadruplexes using ribosome profiling

Like many RNA secondary structures G-quadruplexes (GQs) may interfere with translation when they occur in mRNAs. They have been shown to inhibit cap-dependent translation when they occur in mRNA 5' leaders of specific genes. They probably play a role in repeat associated non-AUG (RAN) translation. GQs also have been shown to stimulate recoding events, such as ribosomal frameshifting. However, their effect on mRNA translation have not been explored at the genome-wide level. Here we utilized publicly available high-throughput datasets to explore the relationship between GQ occurrence and ribosome footprint density. Further we carried out ribosome profiling in HEK293T cells treated with the GQ stabilizing drug, PhenDC3, to examine how GQ stabilization alters the distribution of ribosome footprints across the transcriptome.

2.1 Introduction

RNA secondary structures are known to affect mRNA translation in numerous ways. In bacteria, RNA secondary structures in the vicinity of initiation sites were shown to be the major factor determining translation efficiency (Kudla, Murray, Tollervey, & Plotkin, 2009). In eukaryotes, RNA secondary structures modulate mRNA translation by interfering with the scanning of pre-initiation complexes needed for specialized

helicases (Abaeva, Marintchev, Pisareva, Hellen, & Pestova, 2011; Sen, Zhou, Ingolia, & Hinnebusch, 2015; Svitkin et al., 2001), while at the same time improving recognition of start codons when located downstream (Kozak, 1990), reviewed in Hinnebusch et al., 2016 (Hinnebusch, Ivanov, & Sonenberg, 2016). Strong structures have been shown to serve as roadblocks for elongating ribosomes (Tholstrup et al., 2012). Specific RNA secondary structures are able to alter the decoding process by stimulating ribosomal frameshifting, reviewed in Atkins et al., 2016 (Atkins, Loughran, Bhatt, Firth, & Baranov, 2016) and other recoding events, reviewed in Baranov et al., 2015 (Baranov, Atkins, & Yordanova, 2015). In eukaryotic viruses, specific RNA structures are able to recruit initiating ribosomes to the internal positions of mRNA, reviewed in Filbin et al., 2009, Johnson et al., 2017 and Thompson 2012 (Filbin & Kieft, 2009; Johnson, Grosely, Petrov, & Puglisi, 2017; Thompson, 2012) and are even known to mimic the initiator tRNAs for this purpose, reviewed in Butcher et al., 2016 (Butcher & Jan, 2016).

G-Quadruplexes (GQs) are a specific class of secondary structures where guanine bases interact through hydrogen bonding to form G-tetrads which are further stabilized by the presence of a monovalent cation at the centre of the G-tetrad (Bhattacharyya, Mirihana Arachchilage, & Basu, 2016; Stephen Neidle, 2006). The G-tetrads stack on top of each other and the intervening sequences form loops. Various factors contribute to the stability of the GQ structures including the number of G-stacks and the length of intermediate loops, reviewed in Bochman et al., 2012 (Bochman et al., 2012). The distribution of sequence patterns that could potentially form GQs, such as $(G_3+N_{1-7})_4$ GQs across genome is non-random. G-quadruplexes are significantly enriched in the promoter regions (1kb upstream of transcription start sites) (Huppert & Balasubramanian, 2007). However, they are avoided in protein coding regions of the human genome (Huppert & Balasubramanian, 2005), and their distribution in positive and negative DNA strands is asymmetric suggesting an under-representation of RNA GQs relative to DNA GQs (Huppert et al., 2008). Their distribution in mRNA regions is asymmetric as they occur more frequently towards 5' ends (Huppert et al., 2008). The asymmetric and position specific distribution of GQs suggest that their presence has functional consequences that are utilized in some genomic regions, such as promoters and transcription sites, but are avoided in most of the protein coding regions. Recently a high throughput approach has been developed

for the detection of GQ structures *in vivo* that is based on GQ inducing stalling of reverse transcriptase (Guo & Bartel, 2016; Kwok, Marsico, Sahakyan, Chambers, & Balasubramanian, 2016). The application of this approach revealed the existence of many non-canonical GQ structures with long loops as well as bulged loops. However, *in vivo* RNA secondary structure probing indicated that in mammals they are mostly unfolded in endogenous mRNAs (Guo & Bartel, 2016). This, however, does not necessarily contradict bioinformatics analyses of GQs since it is possible that GQs fold for only short periods of time as argued in Fay et al., 2017 (Fay et al., 2017). This may be sufficient for creating selective evolutionary pressure that shapes their genomic distribution when they exist in the folded form for relatively short periods of time.

Like other RNA secondary structures, GQs are known to affect mRNA translation, reviewed in Bugaut et al., 2012, Fay et al., 2017 and Song et al., 2016 (Bugaut & Balasubramanian, 2012; Fay et al., 2017; Song et al., 2016). The functional consequences of their occurrences in mRNAs are very similar to those of other strong RNA secondary structures. The GQs in mRNA 5' leaders of specific genes have been shown to inhibit translation (Agarwala et al., 2013; Arora et al., 2008; Balkwill et al., 2009; Gomez et al., 2010; Kumari et al., 2007; Lammich et al., 2011; Morris & Basu, 2009; Shahid et al., 2010), presumably by interfering with progression of the scanning pre-initiation complex. The GQs have been also shown to slow down elongating ribosomes (Endoh, Kawasaki, & Sugimoto, 2013) and stimulate ribosomal frameshifting when they occur downstream of shift-prone slippery sequences (Endoh et al., 2013; Yu et al., 2014). Further it is likely that GQs trigger Repeat Associated Non-AUG translation (RAN) (Reddy, Zamiri, Stanley, Macgregor, & Pearson, 2013).

To assess the relationship between G-quadruplex structures and ribosome footprint density I utilized a collection of publicly available ribosome profiling data and recent data on GQ profiling *in vivo* (Kwok et al., 2016). Further to access the role of GQ structures on translation at the genome wide level, the analysis of in-house ribosome profiling data obtained by treating HEK293T cells with the GQ stabilizing drug, PhenDC3 (Figure 2.1) (Halder, Riou, Teulade-Fichou, Frickey, & Hartig, 2012) was performed.

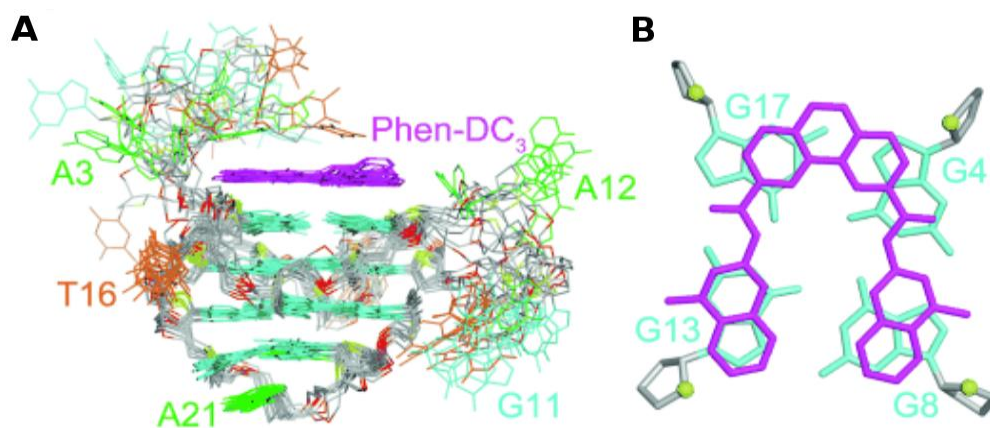


Figure 2.1: (A) Structure of a G-quadruplex bound to the PhenDC3 compound. (B) Top view showing the stacking of PhenDC3 on the top of a G-quadruplex G-tetrad (W. J. Chung et al., 2014).

2.2 Results

2.2.1 Distribution of G-quadruplexes

The genome-wide distribution of G-quadruplexes have been analysed in multiple species including viruses (Biswas, Kandpal, Jauhari, & Vivekanandan, 2016), bacteria (Kaplan et al., 2016), fungus (Hershman et al., 2008), plants (Garg, Aggarwal, & Thakkar, 2016; Mullen et al., 2010; Wang et al., 2015) and human (Huppert & Balasubramanian, 2005).

Using *in-silico* techniques, G-quadruplexes have been found to be highly prevalent throughout the human genome (Huppert & Balasubramanian, 2005). In 2015 Chambers et al., 2015 (Chambers et al., 2015) developed a high-resolution sequencing based method by combining features of the polymerase stop assay with Illumina next-generation sequencing which was used to detect GQs in the human genome. Using B lymphocytes under conditions that either promote or disfavour GQs formation, they reported 716,310 distinct GQs in the human genome. Other studies have provided substantive evidence of the formation of G-quadruplexes in the human genome and transcriptome by visualizing them in human cells using structure specific antibodies

(Biffi, Di Antonio, Tannahill, & Balasubramanian, 2014; Biffi, Tannahill, McCafferty, & Balasubramanian, 2013).

The formation of thousands of canonical and non-canonical GQs were reported *in vivo* in HeLa RNA, using the RNA G-quadruplex sequencing (rG4-seq) technique developed by Kwok et al., 2016 (Kwok et al., 2016). This transcriptome wide RNA G-quadruplex profiling method couples rG4-mediated reverse transcriptase stalling with next-generation sequencing and allows the examination of GQs structural motifs (Kwok et al., 2016).

Quadparser is a computational approach for the prediction of potential GQs forming sequence patterns using a simple algorithm of mining sequence patterns of $G_XN_{1-7}G_XN_{1-7}G_X$ (Huppert & Balasubramanian, 2005). I have used this approach developed by Huppert et al., 2005 (Huppert & Balasubramanian, 2005), to predict potential GQs forming sequences in human genome and transcriptome regions (see material and methods). G-quadruplex patterns were found to be enriched in all genomic regions with highest enrichment in the 1000 base pairs upstream region of transcription start sites with a value of 217.90 per Mb (Figure 2.2) (Table 2.1). The higher enrichment of GQs forming sequence patterns in promoters signify their role in gene regulation at the level of transcription. It has been previously reported that nearly 50% of genes contain a GQs forming sequence pattern in their promoter region with higher enrichment in oncogenes (Verma et al., 2008). The effect of a promoter GQs on gene expression was first reported in the human gene *c-MYC* where the stabilization of GQs with ligands found to affect gene expression (Siddiqui-Jain, Grand, Bearss, & Hurley, 2002).

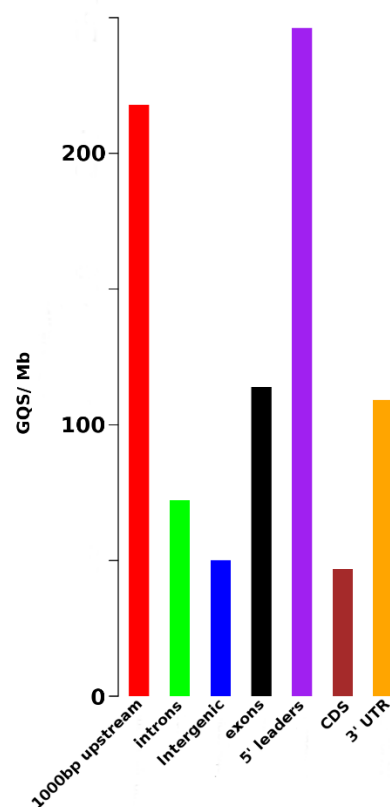


Figure 2.2: Bar plot representation of G-quadruplex density in genomic and transcriptomic regions. Where 1000 bp upstream represents the 1000 nt region 5' of transcription start site.

	Region	GQs/ Mb
Genomic regions	1000 bp upstream	217.90
	Intron	72.31
	Intergenic	50.18
	Exon	113.90
mRNA regions	5' leaders	246.31
	CDS	46.86
	3' UTR	109.19

Table 2.1: The enrichment of potential GQs forming sequences in Genomic and transcriptomic regions. Where 1000 bp upstream represents the 1000 nt region 5' of transcription start site.

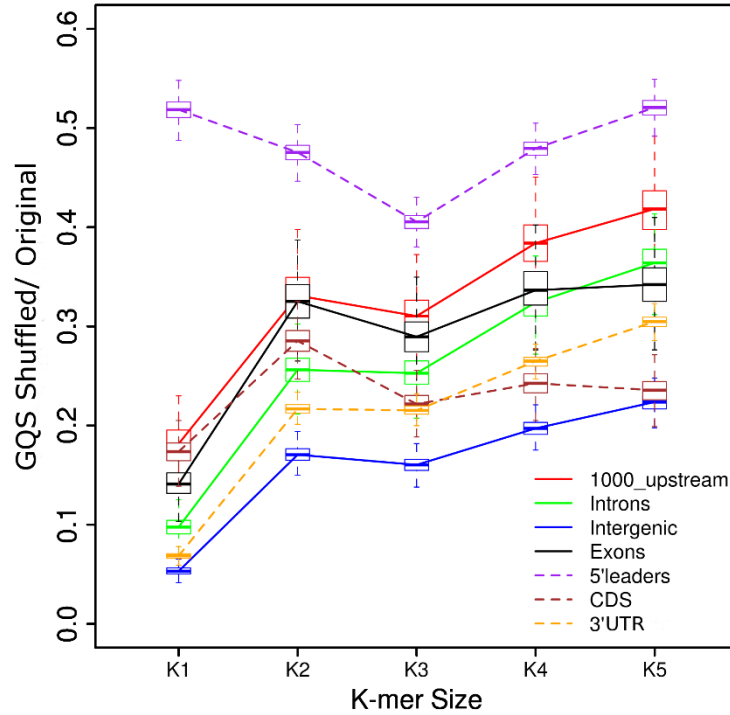


Figure 2.3: The comparison of GQs motifs predicted in original datasets and in-silico generated sequences by preserving the K-mer count from 1 to 5.

In the analysis of different mRNA regions, 5' leaders were found to have highest enrichment with a value of 246.31 per Mb, signifying their role in regulating gene expression at the level of translation. Many reports have examined the role of GQs in repressing translation by creating hindrance in the movement of scanning translational machinery (Table 1.1).

To further explore whether GQs reduce or enhance organism fitness when they occur in different functional regions of the genome (1000 base pair region upstream of translation start sites, intronic, intergenic and exonic regions) and mRNAs (5' leaders, CDS and 3' UTRs), we compared the frequency of sequence patterns that can form GQs to that expected by chance. To achieve this, we generated in-silico sequences from the original datasets by preserving the K-mer count from 1 to 5. We found that potential GQ forming sequences occur more frequently in all genomic regions and the natural mRNA sequences than what would be expected by chance and therefore have a selective pressure in analysed regions of genome and mRNAs (Figure 2.3).

The difference in the original and in-silico sequences was found to be least in case of 5' leaders followed by 1000nt up stream of the transcription start site (Figure 2.3).

2.2.2 Analysis of ribosome profiling footprint densities in the vicinity of GQs motifs

G-quadruplex metagene profiles were generated using global aggregates from 116 ribo-seq untreated samples and 35 RNA-seq untreated samples from published studies (TRIPS-viz) (unpublished) (see appendix 4 Table 5.1 and 5.2). Similar analysis was performed for the ribo-seq and RNA-seq samples generated in-house i.e. control and PhenDC3 treated ribo-seq and the RNA-seq samples. The presence of secondary structures and GC/G rich regions have been considered as barriers to PCR amplification and sequencing (Nelms & Labosky, 2011; Zhu et al., 2016). The formation of a secondary structure can cause resistance to polymerase read-through, which results in DNA sequencing stop and prevents PCR amplification of the secondary structure (Nelms & Labosky, 2011).

Using the global aggregates from a number of studies, we reported that regions in the vicinity of the predicted G-quadruplex regions (predicted using *Quadparser*) (Huppert & Balasubramanian, 2005) are not amplified considerably. The read density was reported considerably low in the predicted G-quadruplex regions. This behaviour was found to be consistent among all the samples used for this analysis (Figure 2.4 (A)). Further, to compare the accuracy of the computational G-quadruplex prediction (Huppert & Balasubramanian, 2005) to the rG4 profiling method (Kwok et al., 2016), we generated metagene profiles showing the read density around the coordinates predicted to form a stable *in vivo* G-quadruplex structures with the rG4 profiling technique (Kwok et al., 2016). The *in vivo* predicted G-quadruplexes were found to show the same behaviour of reduced read density in the G-quadruplex region as reported for the computationally predicted G-quadruplexes (Figure 2.4 (B)), only the depletion in read density was much stronger in the computationally predicted GQs (Figure 2.4 (A)). This analysis reveals that computationally predicted G-quadruplexes are significantly reliable. The comparison of the translation efficiency (TE) of the mRNAs with GQs in 5' leaders, CDS and 3' UTR did not reveal any significant difference in the TE of these mRNAs (Figure 2.4 (C)).

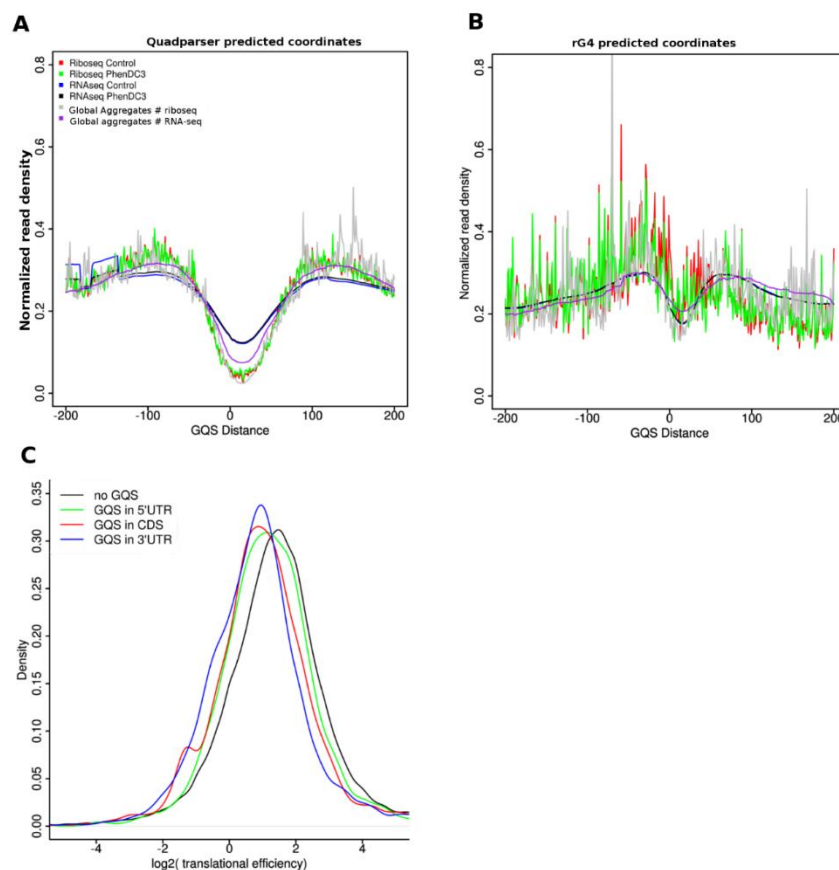


Figure 2.4: Distribution of ribosome footprints in the close vicinity of G-quadruplex start sites. (A) GQs metagene profile for 200 nt regions upstream and downstream of GQs start sites (represented by ‘0’) which were predicted using Quadparser. (B) GQs metagene profile using GQs start sites predicted using the rG4 technique (Kwok et al., 2016). (C) Translation efficiency distribution of the genes with no GQs, GQs in 5’ leader, CDS and 3’ UTR.

2.2.3 Preliminary analysis of the Ribo-seq and RNA-seq samples

Certain compounds increase the stability of GQs (Le et al., 2015), among them is PhenDC3, a drug that interacts with the top G-tetrad of a G-quadruplex structure which leads to increased stability (Figure 2.1) (W. J. Chung et al., 2014). We decided to use it to explore how stabilization of GQs affect mRNA translation *in vivo*. Ribosome profiling (ribo-seq) and mRNA sequencing (RNA-seq) samples were generated using HEK293 cells treated with 0.8uM PhenDC3 for 45 minutes and untreated control cells in two replicates. The preliminary analysis was performed on both replicates to examine the quality of the data and reproducibility of the replicates. The read length distribution of the ribo-seq reads was highest in range of 30 to 34 nucleotides (nt) which

corresponds to the ribosome protected fragment length (Ingolia, Ghaemmaghami, Newman, & Weissman, 2009) (Figure 2.5 (A)). The RNA-seq reads were found to be most prevalent in 27 to 37 read length range (Figure 2.5 (A)). The reads were aligned 30 nt upstream and downstream of the start and stop codon to create a metagene profile, which indicated that the position of the 5' ends of the reads started abruptly 15 nt upstream of the start codon and ended 18 nt upstream of the stop codon (Figure 2.5 (B)). The ribo-seq data was found to have a good triplet periodicity for both replicates indicating the three-nucleotide movement of translating ribosomes (Figure 2.5 (B)). The analysis of the distribution of the reads in different regions of mRNA revealed a slight increase in the read density in the 5' leaders in PhenDC3 treated samples (Figure 2.5 (C)). Reproducibility of the data from all samples was estimated by calculating Pearson's correlation coefficient. A high correlation between all four samples was observed with Pearson's correlation coefficient varying from 0.92 – 0.96 (Figure 2.5 (D)).

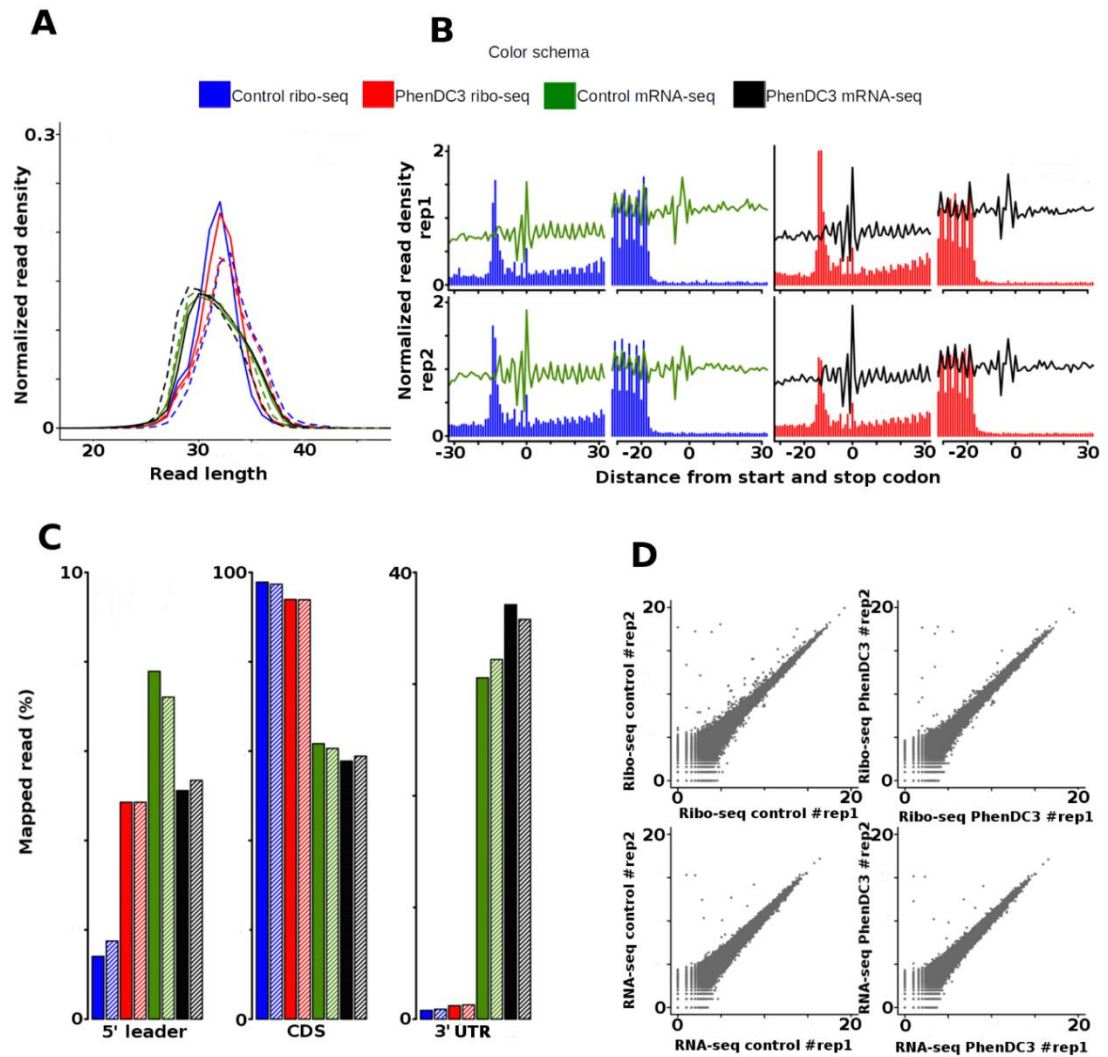


Figure 2.5: Preliminary analysis of both sets of replicates. (A) Read length distribution of ribo-seq and RNA-seq samples. (replicate two shown with dashed lines). (B) Metagene profiles for ribo-seq and RNA-seq samples showing normalized read density from 30 nt upstream and downstream of the start and stop codon. (C) Distribution of reads mapped in 5' leaders, CDS and 3' leaders (where replicate two is shown as dashed bars). (D) Correlation across reads mapped on each gene in both replicates for control and treated samples.

2.2.4 Differential gene expression analysis

The wild-type and PhenDC3 treated samples were compared to detect differential gene expression at the transcriptional and translational level. The fold change of expression values between wild-type and PhenDC3 treated samples were converted to z-score values to filter up and down regulated genes (Figure 2.6 (A) & (B)). Further, a false

discovery rate (FDR) value of 20% was used to set the threshold value of the z-scores to designate a gene as up or down regulated. The z-score values from both replicates were compared to assess the reproducibility (Figure 2.6 (C)). In both replicates, only few genes were found to be differentially regulated between wild-type and PhenDC3 treated samples, at the level of translational. Whereas at the level of transcription, multiple genes from the histone gene family were found to be down-regulated after the PhenDC3 treatment, in both set of replicates (Table 2.2) (Figure 2.6 (C)). For example, the histone family genes *HIST1H3B* and *HIST1H1E* shown in Figure 2.7 were found to be down-regulated at the transcriptional level after PhenDC3 treatment. Further KEGG pathway analysis of down-regulated genes was performed which revealed enrichment of these genes in pathways related to autoimmune disease Systemic lupus erythematosus, alcoholism and the cancer related pathways viral carcinogenesis and transcriptional mis-regulation in cancer (Table 2.3).

Gene	Transcript	Z-score #rep1	Z-score #rep2
<i>HIST1H3B</i>	NM_003537	-5.53	-6.53
<i>HIST1H1E</i>	NM_005321	-6.02	-5.27
<i>HIST2H3A</i>	NM_001005464	-4.87	-5.74
<i>HIST2H3C</i>	NM_021059	-4.72	-5.23
<i>HIST2H4A</i>	NM_003548	-4.52	-4.69
<i>HIST1H3H</i>	NM_003536	-4.4	-4.72
<i>RNU5E-1</i>	NR_002754	-5.84	-3.18
<i>HIST1H1C</i>	NM_005319	-3.49	-4.99
<i>HIST1H4B</i>	NM_003544	-3.55	-4.65
<i>HIST1H2BJ</i>	NM_021058	-3.59	-4.45
<i>HIST2H3D</i>	NM_001123375	-4.01	-3.72
<i>HIST1H2AG</i>	NM_021064	-3.34	-3.85
<i>HIST1H4H</i>	NM_003543	-3.97	-3.21
<i>HIST2H4B</i>	NM_001034077	-4.9	-2.13
<i>HIST2H2AA3</i>	NM_003516	-3.89	-3.11
<i>HIST1H2AH</i>	NM_080596	-3.45	-3.43
<i>HIST1H4D</i>	NM_003539	-3.15	-3.33
<i>HIST1H4C</i>	NM_003542	-3.65	-2.64
<i>HIST1H4J</i>	NM_021968	-3.01	-2.54
<i>HIST1H3A</i>	NM_003529	-2.92	-2.62

Table 2.2: Showing the list of genes down-regulated at the level of transcription.

KEGG Pathway	No. of genes enriched in pathway	P-value
Systemic lupus erythematosus	17	1.6×10^{-028}
Alcoholism	17	1.7×10^{-026}
Viral carcinogenesis	8	0.2×10^{-6}
Transcriptional misregulation in cancer	6	0.3×10^{-4}

Table 2.3: KEGG pathway enrichment of transcriptionally down-regulated genes.

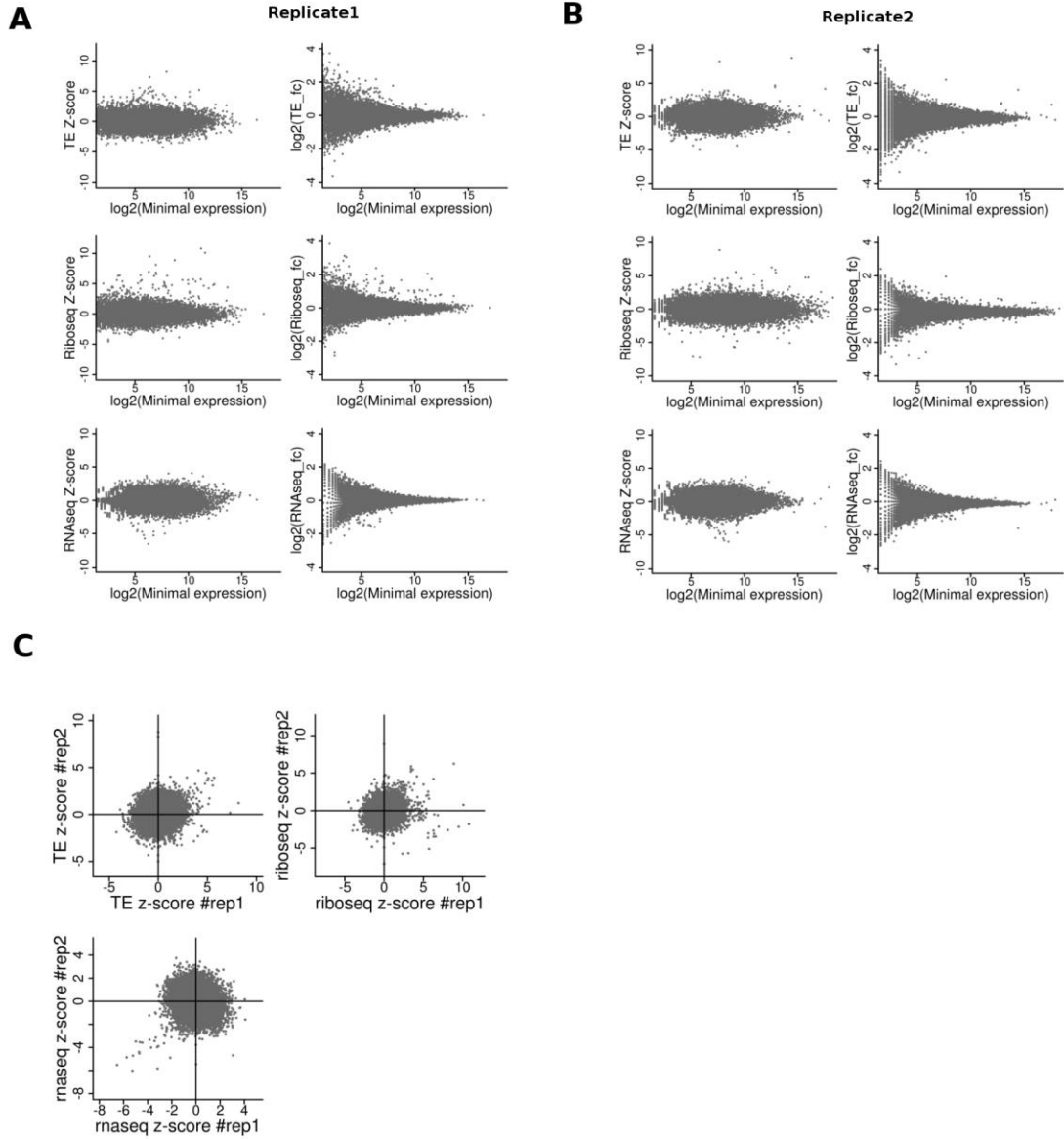


Figure 2.6: Differential gene expression analysis. (A) & (B) First column showing the distribution of z-score (y-axis) obtained from fold-changes of PhenDC3 samples relative to wild-type samples from ribosome occupancy (Ribo-seq), transcript levels (RNA-seq), and translation efficiency (TE). Transcripts are ordered based on their log2 minimal expression signal (x-axis). Second column represents the fold change values (y-axis) of PhenDC3 samples relative to wild-type samples for Ribo-seq, RNA-seq and TE analysis. (C) Representing the correlation between the z-scores obtained for replicate1 (Figure 2.6(A)) and 2 (Figure 2.6(B)).

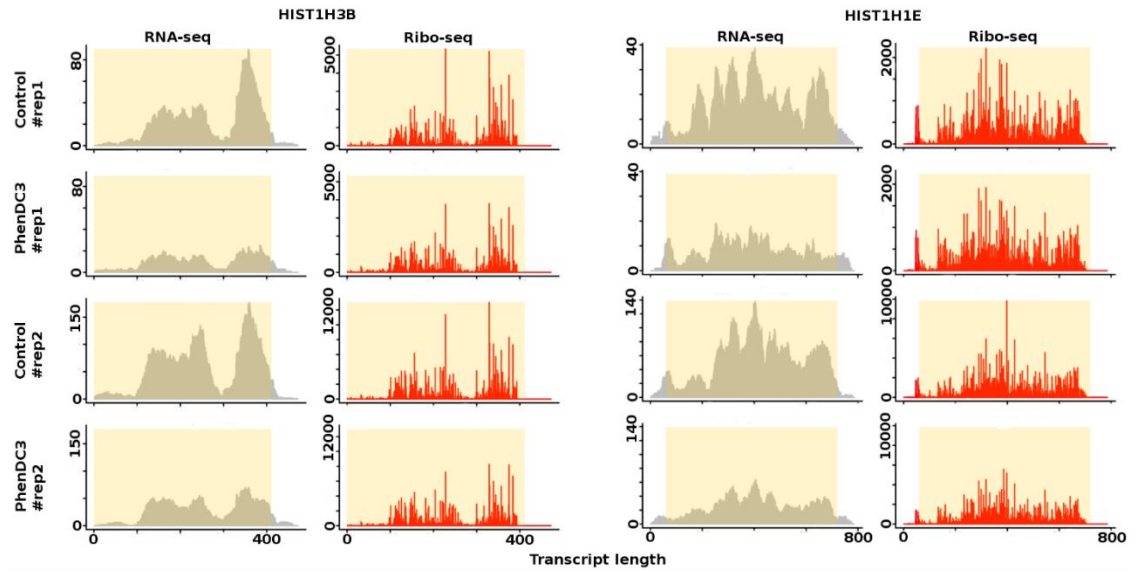


Figure 2.7: Transcriptional down-regulation of histone family genes. The footprint densities profiles of RNA-seq (grey) and ribo-seq (red) for gene *HIST1H3B* and *HIST1H1E* showing the reduction at the transcriptional level after PhenDC3 treatment. The coding region has been highlighted in yellow colour.

2.2.5 Ribosome pauses prediction

It has been reported previously that the presence of secondary structures can act as roadblocks for the scanning translation machinery and lead to the ribosomal pauses (Somogyi et al., 1993). To analyse the effect of the enhanced stability of the GQs structures, following treatment with PhenDC3, on ribosomal pauses, we analysed both wild-type and PhenDC3 treated samples using the PausePred tool (see Methods for parameters) (Chapter 3). The PausePred tool can detect ribosome pauses using the footprint density information from ribosome profiling data. A total of 244021 and 270352 pauses were predicted in control (replicate 1) and PhenDC3 (replicate 1) respectively, of which 64835 pauses were found to occur in both samples with varying strengths. The control and PhenDC3 samples from replicate 2 were found to contain 358718 and 339764 pause sites respectively with 133304 common in both samples (Figure 2.8 (A)). To further determine if PhenDC3 causes an increase in the strength of the pauses, we compared the pause scores of the sites which were common in both the control and PhenDC3 treated samples. Although a good correlation was found in the pause strengths from both samples, there was no significant up-regulation of pauses was reported after PhenDC3 treatment (Figure 2.8 (B)). It was previously

reported that the predicted G-quadruplex regions contain considerably low read density (Figure 2.4 (A) & (B)), which can be caused by the lower PCR amplification of the GC rich regions or presence of the secondary structures (Nelms & Labosky, 2011; Zhu et al., 2016). Based on the analysis shown in the Figure 2.4 (A) & (B), I concluded that the absence of significantly differentially regulated pauses can occur due to reduced read density in the GQs vicinity.

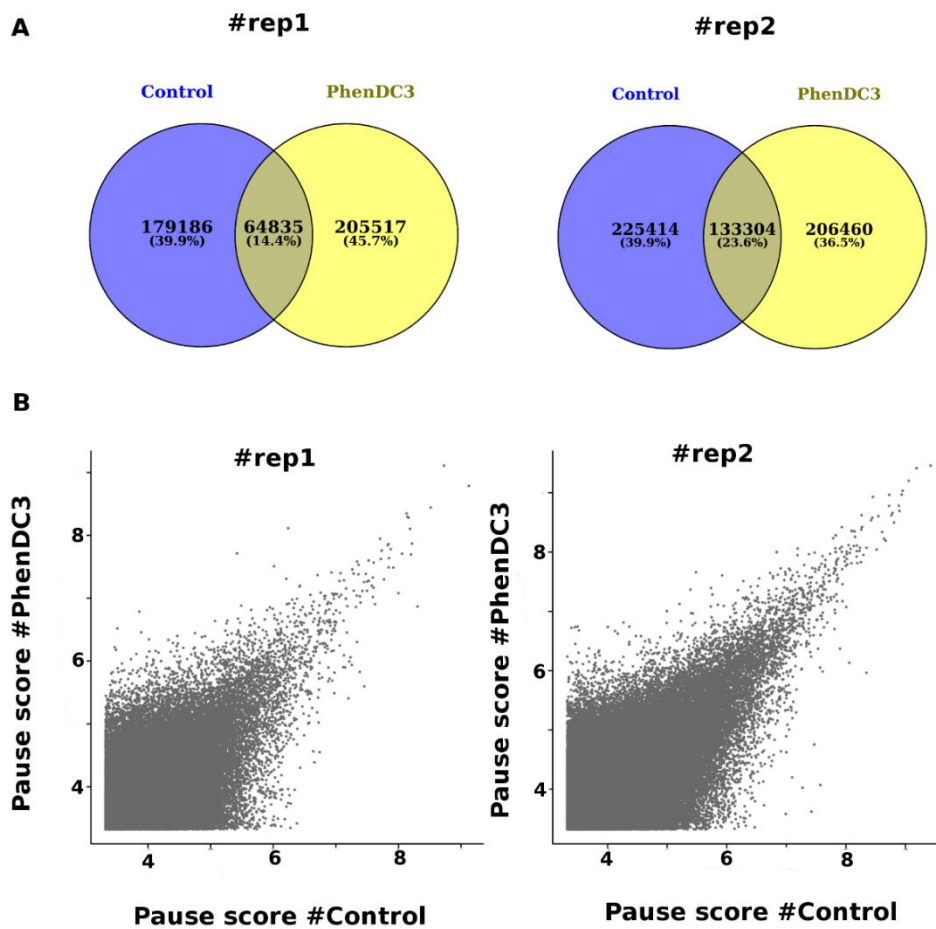


Figure 2.8: Ribosome pause prediction using PausePred (A) Venn diagram representing the overlap in the total number of pauses predicted in control and PhenDC3 treated samples for both replicates. (B) Comparison of the pauses in control and PhenDC3 treated samples for replicate 1 and 2.

2.3 Discussion

G-quadruplexe (GQ) secondary structures can interfere with the movement of translational machinery thus leading to the repression of protein synthesis. Ligands like TmPyP4, PhenDC3, and PDS (Le et al., 2015) can enhance the stability of the G-quadruplex structure and help in studying the role of GQs structures in transcription and translation. The compound PhenDC3 binds on the top of a DNA/RNA G-quadruplex G-tetrad (W. J. Chung et al., 2014), increasing the stability further. The effect of stable RNA G-quadruplex on the translation process can be studied by using the Ribosomal profiling technique which captures the ribosome protected mRNA fragments giving information about actively translating mRNAs (Ingolia et al., 2009).

The GQs structures were found to be enriched in all genomic regions with an over-representation in promoter regions (1000 nt region upstream of transcription start site) and 5' leaders, indicating the role of GQs as regulatory elements for regulating transcription and translation. Further, the comparison of the GQs predicted in actual sequences of different genomic regions to that of *in silico* generated random sequences revealed that GQs sequence patterns are not expected to occur by chance.

The reduced ribosome footprint density found in multiple ribo-seq and RNA-seq datasets around the start sites of predicted GQs point towards two possibilities; the presence of a secondary structure in this vicinity or lower amplification due to GC rich region during sequencing (Nelms & Labosky, 2011; Zhu et al., 2016).

The analysis of the pharmacological stabilization of the GQs using 0.8uM PhenDC3 drug did not reveal any noticeable differences at either the transcriptional or translational levels. A higher concentration of the PhenDC3 drug than used here, has been reported to cause elevated differential transcriptional regulation by Halder et al., 2012 (Halder et al., 2012).

Multiple studies have revealed the role of G-quadruplexes in translational repression. So far most of these studies, have focused on single gene analysis. A genome wide analysis to check the overall effect of GQs on translation similar to our study but with a higher concentration of stabilizing drugs will be a valuable addition to such studies.

Ribosome profiling is an actively emerging technique and since its invention multiple bioinformatics methods have been developed for analysing the data generated by ribosome profiling (Michel et al., 2016). While working on the current project where

prediction of ribosome pauses due to the presence of GQs was of extreme importance, I realized that no devoted software has been developed to infer the ribosome pauses using ribo-seq data. This encourage me to develop a software which can detect ribosome pauses using ribosome profiling data which has been discussed in chapter 3.

2.4 Methods

G-quadruplex metagene analysis using published Ribo-seq and RNA-seq: To analyse the overall behaviour of G-quadruplexes (GQs) in vivo, 116 ribosome profiling control samples and 35 RNA-seq control samples from published studies were obtained from Trips-Viz (unpublished) (see appendix 4 Table 5.1 and 5.2). These samples were processed by Trips-Viz as followed. The adaptor sequences were removed using cutadapt (Martin, 2011). To remove the rRNA contamination, trimmed reads were aligned to human rRNA sequences using Bowtie short read alignment program with parameters (-v 3 --norc -p 6). The reads left after removing the contamination were aligned to gencode v25 of human transcriptome assembly (<https://www.gencodegenes.org/releases/25lift37.html>) using Bowtie (Langmead et al., 2009), with parameters (-a --norc -m 100 -n 2 --seedlen 25 -p 6). Both ambiguous and non- ambiguous alignments were used for generating G-quadruplex metagene profiles.

The GQ start sites used for analysis in Figure 2.4 (A) were predicted using Quadparser (Huppert & Balasubramanian, 2005). The GQs with three stacks were used here (see details in ‘Predicting potential GQ patterns’ section of methods). Whereas for analysis in Figure 2.4 (B) GQs start sites were used from the predictions in (Kwok et al., 2016).

Generation of Ribo-seq and RNA-seq Libraries: The ribosomal profiling technique was carried out as mentioned in Andreev et al., 2015 (Andreev, O'Connor, Zhdanov, et al., 2015). The ribo-seq and RNA-seq libraries were obtained by treating HEK293 cells with 0.8uM PhenDC3 for 45 minutes and untreated control cells in two replicas.

Pre-processing and alignment for in-house Ribo-seq and RNA-seq using PhenDC3 treatment and control: Adaptor sequence (5'-CTGTAGGCACCATCAATAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3') was removed from raw reads using cutadapt (Martin, 2011). To reduce rRNA

contamination trimmed reads were aligned to rRNA sequences using Bowtie short read alignment program and no more than three mismatches were allowed for these alignments (-v 3) (Langmead, Trapnell, Pop, & Salzberg, 2009). Reads aligning to rRNA sequences were removed from the dataset for further analysis. The read length distributions were analysed to determine the most prevalent read lengths among the samples.

The human genome was obtained from NCBI (GCA_000001405.15) and human transcriptome sequences obtained from the RefSeq database from the NCBI (ftp://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens/mRNA_Prot/). Bowtie was used to align reads to genome and transcriptome (Langmead et al., 2009). The parameters used for genome alignments were (-m 100 -v 3) i.e. no more than three mismatches were allowed and maximum of 100 mappings were reported per read. Reads were aligned to transcriptome with the following parameters (-a, -m 100 -norc) reads were aligned to the positive strand allowing no more than three mismatches and no more than 100 mappings per read.

Genes with at least 100 mapped alignments were used for the production of metagene profiles at the location of translation initiation and termination (Figure 2.5 (B)). For the metagene analysis, 30 nucleotides (nt) upstream and downstream of the start and stop codons were used (Figure 2.5 (B)).

Differential gene expression analysis: For differential gene expression analysis, the longest isoforms of transcripts were used. The reads mapped in two conditions were normalized by rescaling the raw counts within the conditions with the higher total count (P_{max}) by a rescale factor F

$$F = \frac{\sum_{i=1}^n P_{max}}{\sum_{i=1}^n P_{min}} \dots\dots\dots (1)$$

where *i* is the number of reads mapped to each gene and *n* corresponds to the total number of the genes.

This was performed separately for ribo-seq and RNA-seq samples. Based on the inferred location of the P-site ribo-seq reads were assigned to mRNA coordinates. For

the reads of length 30 to 34 the P-site codon of the elongating ribosome was assigned at 15 nucleotides downstream of the 5' end of read. A z-score transformation of fold change log ratios was carried out as described earlier (Andreev, O'Connor, Zhdanov, et al., 2015). In brief, genes were grouped into bins of 300 based on the lowest expression signal (ribo-seq or RNA-seq or both across compared conditions). Variation in the distribution of fold change ratios was used to calculate the Z-score. The z-score threshold was selected based on the False discovery rate calculations as described earlier in (Andreev, O'Connor, Zhdanov, et al., 2015). For the differentially regulated genes KEGG pathway enrichment analysis was performed using DAVID (Huang da, Sherman, & Lempicki, 2009).

Predicting potential GQs patterns: The GQ patterns with varying numbers of G-stacks and loop lengths were identified using *Quadparser* (Huppert & Balasubramanian, 2005). The pattern used for this analysis is can be explained with this given formula: $G_X+L_1-NG_X+L_1-NG_X+L_1-NG_X+$, where $X=3$ was used in our search and L representing intermediate loops which can consist of any nucleotide including guanine varying between length 1-7 was used.

Ribosome pause prediction using PausePred: The BAM alignment files were used to predict pauses using PausePred tool (Chapter 3). The pauses were predicted in both control and PhenDC3 treated samples using a fold change value of 10 and a minimum window coverage of 10 (at least 10% of the positions within a window of 1000 contains ≥ 1 read mapped) and offset value of 15.

Venn diagrams were created using Venny 2.1.0 (Oliveros, J.C. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>).

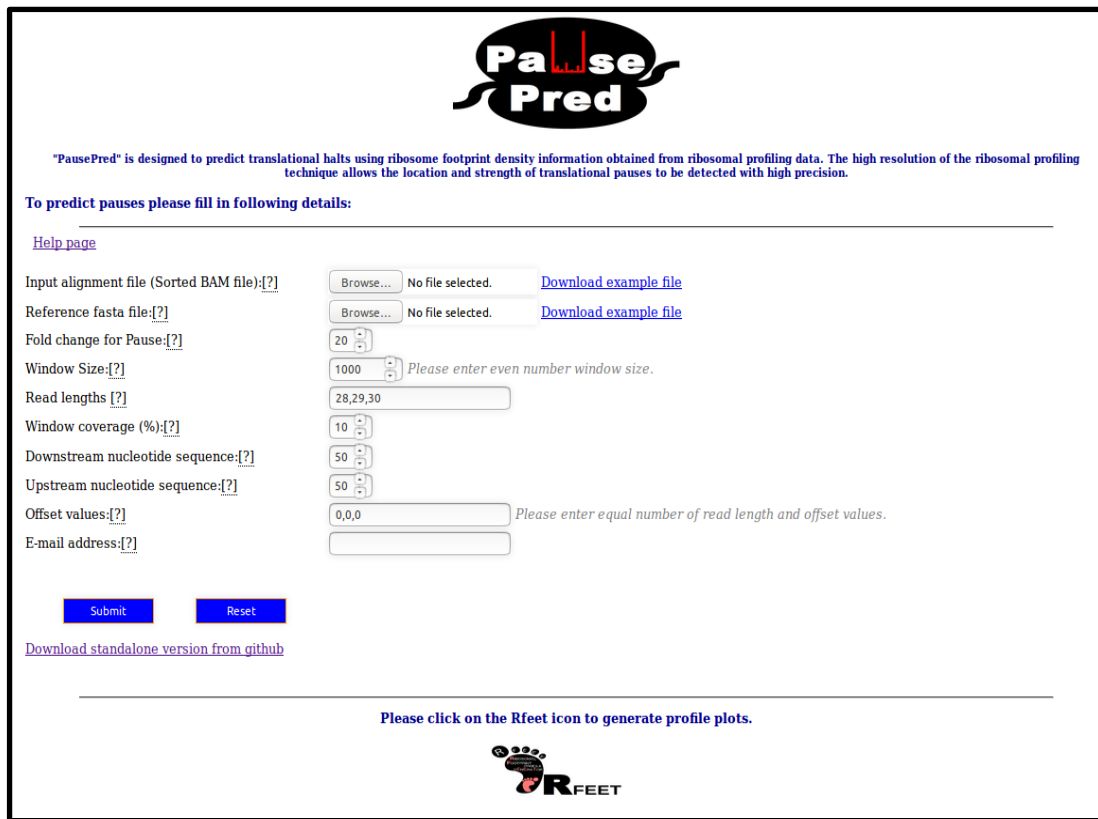
Chapter 3

PausePred and Rfeet: Webtools for inferring ribosome pauses and visualizing footprint density from ribosome profiling data

This chapter has been submitted for publication to RNA journal

The process of translation is characterized by irregularities in the local decoding rates of specific mRNA codons. This includes the occurrences of long pauses that can take place when ribosomes decode certain peptide sequences, encounter strong RNA secondary structures or decode “hungry” codons. Examples are known where such pausing or stalling is used for regulating protein synthesis. This can be achieved at the level of translation *via* direct alteration of ribosome progression through mRNA or by altering mRNA stability *via* NoGo decay. Ribosome pausing has also been implicated in the co-translational folding of proteins. Here we present PausePred (<https://pausepred.ucc.ie/>, https://github.com/romikasaini/Pausepred_offline) which can be used to infer ribosome pauses from ribosome profiling (Ribo-seq) data. Peaks of ribosome footprint density are scored based on their magnitude relative to the background density within the surrounding area. The scoring allows the comparison of peaks across the transcriptome or genome. In addition to the score, PausePred reports the coordinates of the pause, the footprint density at the pause site and the surrounding nucleotide sequence. The pauses can be visualized in the context of Ribo-

seq and RNA-seq density plots generated for specific transcripts or genomic regions with the Rfeet tool.



PausePred

"PausePred" is designed to predict translational halts using ribosome footprint density information obtained from ribosomal profiling data. The high resolution of the ribosomal profiling technique allows the location and strength of translational pauses to be detected with high precision.

To predict pauses please fill in following details:

[Help page](#)

Input alignment file (Sorted BAM file):[?] No file selected. [Download example file](#)

Reference fasta file:[?] No file selected. [Download example file](#)

Fold change for Pause:[?]

Window Size:[?] Please enter even number window size.

Read lengths [?]

Window coverage (%):[?]

Downstream nucleotide sequence:[?]

Upstream nucleotide sequence:[?]

Offset values:[?] Please enter equal number of read length and offset values.

E-mail address:[?]

[Download standalone version from github](#)

Please click on the Rfeet icon to generate profile plots.

RFEET

Figure 3.1: PausePred webtool's home page representing the inputs and parameters user can supply. At the bottom of the page a link to the Rfeet (Ribosomal footprint profile generator) tool has been provided.

3.1 Introduction

The progression of ribosomes along mRNA is often accompanied by the presence of strong pauses at specific locations (Richter & Collier, 2015). Among the known causes of such site-specific pauses are strong RNA secondary structures whose unwinding is required for ribosome progression (Somogyi et al., 1993; Tholstrup et al., 2012) and specific nascent peptides interacting with the ribosome (Tenson & Ehrenberg, 2002). Such peptide-mediated pauses are known to function as part of translation regulators sensing specific molecules as has been demonstrated for the fungal arginine attenuator

peptide (Fang, Spevak, Wu, & Sachs, 2004), erythromycin resistance genes in bacteria (Arenz et al., 2014; Fang et al., 2004) and in the polyamine responsive translation of the *AMD1* mRNA in mammals (Law, Raney, Heusner, & Morris, 2001). Ribosome pausing was recently found to occur at AUG-Stop ORFs in response to boron (Tanaka et al., 2016). Ribosome pausing is also known to affect the stability of mRNA *via* NoGo decay (Doma & Parker, 2006). Differences in local decoding rates have been implicated in co-translational protein folding (Tsai et al., 2008). Ribosomal profiling (Ribo-seq) (Ingolia et al., 2009) provides a snapshot of ribosome locations across all mRNAs by capturing mRNA fragments protected by ribosomes (Andreev et al., 2017; Ingolia, 2016; Michel & Baranov, 2013). The detection of abnormally high peaks of ribosome footprint density can aid in the identification of ribosome pauses (Ingolia, Lareau, & Weissman, 2011; Mohammad, Woolstenhulme, Green, & Buskirk, 2016), and features associated with these pauses (Sabi & Tuller, 2017). Since the invention of ribosome profiling, numerous bioinformatics methods have been developed for the analysis of differential mRNA translation (Oertlin et al., 2017; Olshen et al., 2013; Xiao, Zou, Liu, & Yang, 2016; Zhong et al., 2017), identification of translated open reading frames (Calviello et al., 2016; Chun, Rodriguez, Todd, & Mills, 2016; Fields et al., 2015; Malone et al., 2017; Michel et al., 2012; Ndah et al., 2017; Raj et al., 2016), characterization of global determinants of local decoding rates (O'Connor, Andreev, & Baranov, 2016) as well as general use platforms (Carja, Xing, Plotkin, & Shah, 2017; B. Y. Chung et al., 2015; Crappe et al., 2015; Dunn & Weissman, 2016; Legendre, Baudin-Baillieu, Hatin, & Namy, 2015; Michel et al., 2016) and databases (Michel, Ahern, Donohue, & Baranov, 2015; Olexiouk et al., 2016; Wan & Qian, 2014; Xie et al., 2016). While many ribosome profiling studies involved identification of pauses, no dedicated software for pause detection and analysis is currently available.

To fill this gap, we have developed the PausePred tool for the identification of pauses by analysing ribosome footprints aligned to transcriptome or genome reference sequences (Figure 3.1). PausePred extracts sequences surrounding high density peaks which can be readily used for exploring sequence motifs enriched in the vicinity of the pause sites. PausePred can also be used for analysing changes in pause strength in response to different experimental conditions. The pauses can be visually examined using the Rfeet tool that we have developed to plot Ribo-seq and RNA-seq read

density profiles.

3.2 Results

3.2.1 PausePred

PausePred takes the number of reads r mapped to each position i within a sliding window of length n (step = $n/2$) and normalizes it over the average density within the window. The average of these values across overlapping windows is used as the pause score, S_i :

$$S_i = \frac{n}{2} r_i \frac{\sum_{i=1}^n r_i + \sum_{i=\frac{n}{2}}^{\frac{3n}{2}} r_i}{\sum_{i=1}^n r_i \sum_{i=\frac{n}{2}}^{\frac{3n}{2}} r_i} \quad \{1\}$$

See Figure 3.2 for a more detailed explanation of the approach. Pause scores exceeding a certain threshold are selected for further analysis. The default threshold is 20, which is a highly permissive score that typically reports pauses on 1-3% of coordinates, see Figure 3.3. Highly atypical pauses would normally have a higher score than that. However, in certain cases, a smaller threshold could be more appropriate, e.g. when datasets are small and there is a need to increase the total number of data points as in the example below. Due to differences in ribosome footprint coverage, the statistical significance of the pauses with the same score may differ. To take this into account, PausePred carries out a z-score transformation similar to what has been used for differential gene expression analysis (Andreev, O'Connor, Fahey, et al., 2015), i.e, pauses are grouped into bins of 300 based on footprint coverage within the corresponding windows and z-scores are calculated based on the parameters of the pause score distribution within each bin.

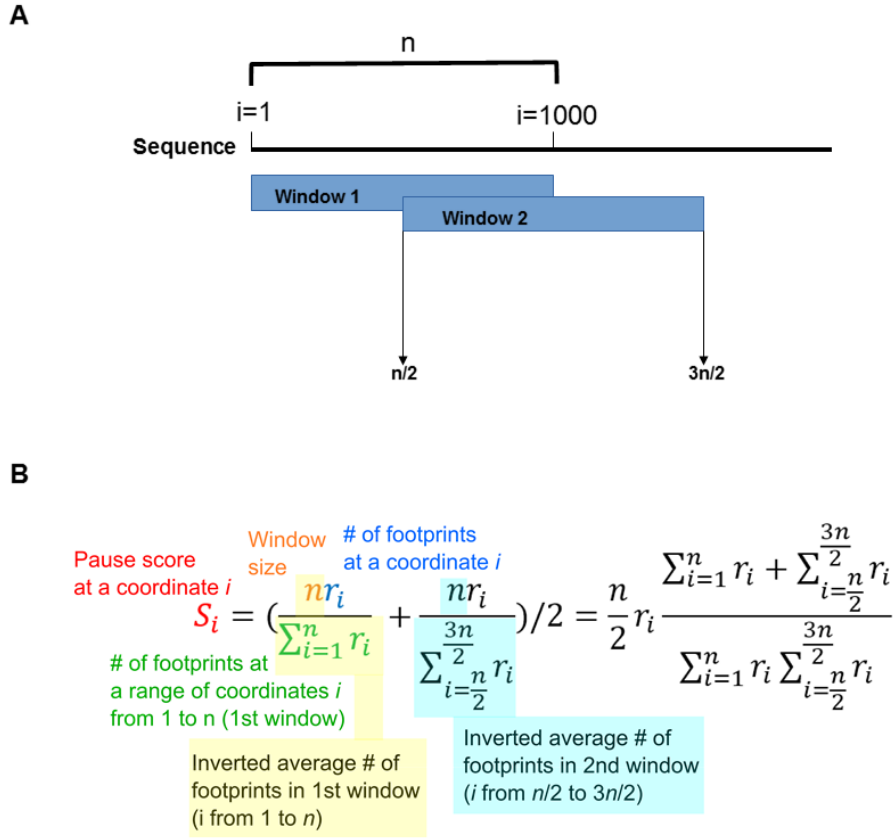


Figure 3.2: PausePred scoring approach. Top panel: The density of each coordinate is compared to the average density within two overlapping sliding windows, 1 to n and $n/2$ to $3n/2$ where n is the size of the window. Bottom panel: Explanation of individual elements of the equation {1}.

The user can specify the footprint read lengths to be considered (28-30 nt by default), the size of the window n for calculating the background density (the default is 1000 nt) and the offset value for inferring the position of the A-site from either the 5' or 3' end of the mapped reads.

The length of footprints varies depending on the organism or organelle used (Oh et al., 2011; Rooijers, Loayza-Puch, Nijtmans, & Agami, 2013) as well as on parameters of the protocols (O'Connor et al., 2016) such as translation inhibitors (Ingolia et al., 2011; Lareau, Hite, Hogan, & Brown, 2014), nuclease digestion (Gerashchenko & Gladyshev, 2017) and ribosome footprint size selection (Mohammad et al., 2016).

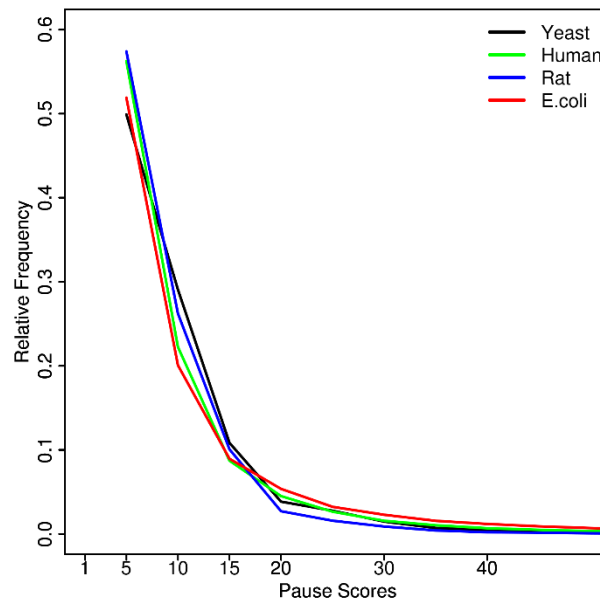


Figure 3.3: Distribution of a fraction of coordinates with a pause score above that at the axis x (*E. coli*- SRA accession number: SRR1734432 (Woolstenhulme *et al.*, 2015); Yeast - control sample (Yerlikaya *et al.*, 2016); Rat – SRA accession number: SRR1557705 (Andreev, O'Connor, Fahey, *et al.*, 2015); Human – SRA accession number: SRR2433794 (Calviello *et al.*, 2016)).

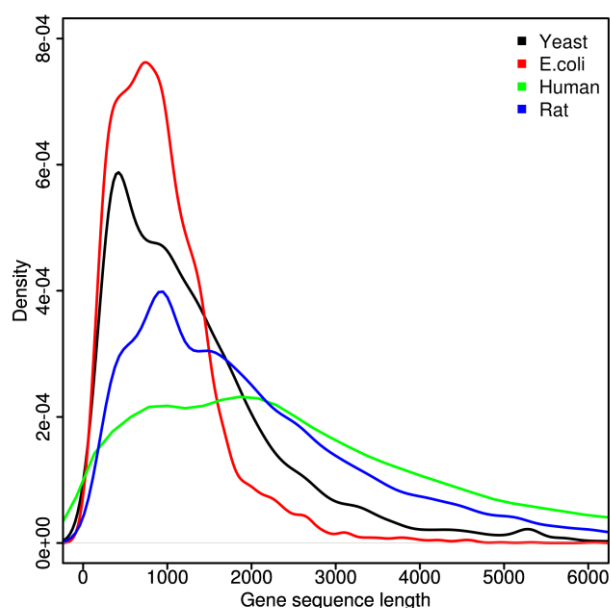


Figure 3.4. Kernel density distributions of the lengths of protein coding ORFs in different organisms based on the transcriptome sequences downloaded from the following resources (Yeast: *Saccharomyces* genome database (Cherry et al., 1997), Strain: S288C; *E. coli*- NCBI (Coordinators, 2017), assembly accession: GCA_000005845.2; Human- Gencode (Harrow et al., 2012) release 25, GRCh37; Rat- Ensembl (Aken et al., 2016), Rnor_6.0 (GCA_000001895.4)).

The default window size of 1000 nt corresponds to the length of a typical protein coding region (Figure 3.4). When using PausePred for the detection of pauses specifically in small ORFs (e.g. in regulatory uORFs) it is advisable to reduce the length of the window size. Obviously, the presence of multiple pauses within the same window would reduce the pause score. Thus, it might be advisable to carry out analysis with smaller window sizes when such a possibility is a concern. The offset for inferring the A-site is an important parameter that is often data specific and should be determined based on the data. Also, due to different asymmetries in ribosome footprint length in bacteria (O'Connor, Li, Weissman, Atkins, & Baranov, 2013), an offset from the 3' end is likely to give more accurate predictions of the A-site position (Mohammad et al., 2016). With PausePred, the user can specify both the offset as well as the end from which it should be applied. The recently developed riboWaltz is an effective approach for determining read length specific A-site offsets (Lauria et al., 2017) and in conjunction with PausePred it can be used for precise pause localisation

using footprints of each specific length separately, since PausePred supports the use of multiple offsets for footprints of different lengths.

For each inferred pause the sequence name/ID, the coordinate of the pause, the number of reads mapped, pause score, average coverage percentage of overlapping windows (where coverage is defined as the percentage of positions within a window having a minimum of one read mapped) and the z-score are output in a csv file. The flanking sequences are also provided (the default is 100 nt with the pause site corresponding to position 51) (Table 3.1). We tested PausePred on ribosome profiling data from different organisms including *E.coli* (Woolstenhulme *et al.*, 2015), yeast (Yerlikaya *et al.*, 2016), rat (Andreev, O'Connor, Fahey, *et al.*, 2015) and human (Calviello *et al.*, 2016) (see Tables 3.1-3.4).

3.2.2 Illustration of PausePred Usage

Here we illustrate PausePred predictions using the Ribo-seq data generated for the *E. coli* MG1655 strain and a mutant strain that lacks elongation factor EFP (Woolstenhulme *et al.*, 2015). After processing the data (see Materials and Methods), the sorted BAM files were uploaded to the PausePred webtool (<https://pausepred.ucc.ie/>). To compare the EFP knockout strain with the wild type, a pause score of 10, read lengths varying from 25-35, a 3' offset of 12 nt and window coverage of 5 (a minimum of 5% of the positions within a window size of 1000 nt had at least one read mapped) were used for pause prediction in both the EFP mutant and the wild type Ribo-seq samples. A lower threshold than the default pause score was used to increase the number of detected pauses. Table 3.1 provides the PausePred output for a subset of the detected pauses and Figure 3.5 shows the Rfeet visualization (see Materials and Methods) of the pause detected in the Ribo-seq density profile of the *ubiD* gene. Of the predicted pauses, 399 candidates were found to have a difference in pause score of at least 50 compared to the corresponding position in the wild-type Ribo-seq alignments (highlighted in red in Fig. 3.6).

PausePred outputs the sequences flanking the pause sites (Table 3.1) which can be explored for pause-related motifs. We analysed the 15 nucleotides upstream of the detected pauses (corresponding approximately to five codons encoding the C-terminal part of the nascent peptide) in the 399 candidates for overrepresented motifs using

Multiple EM for Motif Elicitation (MEME 4.11.2) (Bailey et al., 2009). The most significant motif was found in 88 of these sequences with an e-value of 2.6×10^{-49} (Fig. 3.7). The last six nucleotides of this motif may reflect the enrichment of the polyproline motif reported by Woolstenhulme et al., 2015 (Woolstenhulme *et al.*, 2015) to be associated with pauses in the EFP lacking *E. coli* mutants. The 88 sequences containing the C-rich motif were analysed for a polyproline motif enrichment at the amino acid level by using the corresponding protein sequences obtained from Genbank (Coordinators, 2017) (assembly accession GCA_000005845.2) (Figure 3.8). Further the PP motif was found to be 42.83 times more frequent in the 399 sequences containing the pauses than expected (p-value $< .0001$ Fisher exact test).

Sequence name/ID	Coordinate Position	reads mapped	pause score	coverage (%)	Upstream sequence	Downstream sequence	Z-score
nlpD	174	812	319.93	36.1	AATGCGCCTGCAA ATACTAATTCTGGT ATGTTGATTACGCC GCCGCCGAA	AATGGGGACG ACGTCTACAG CGCAGCAACC GCAAATTCAG CCGGTGCAGC A	8.88
gltD	48	596	312.69	30.6	ATGAGTCAGAATG TTTATCAATTTATC GACCTGCAGCGCG TTGATCC	GCCAAAGAAA CCGCTGAAGA TCCGCAAAAT TGAGTTTGTTG AAATTTACGA	10.08
emtA	88	1113	269.94	24.9	TAGCGGGTTGTTCA TCAAAGCATGACT ATACGAACCCGCC GTGGAACGCG	AAAGTTCGG TGCAACGTGC GATGCAGTGG ATGCCAATAA GCCAGAAAGC C	7.81
proP	178	648	240.17	36.6	TTTATGGTTTTGTT GCTTACGCATTAGG TAAAGTTTTTTTCC CGGGGGCT	GACCCAGCG TGCAGATGGT TGCTGCACTTG CCACTTTCTCC GTCCCTTT	7.96
ubiD	307	524	175.48	32.5	GTGAAGTTGGTAA ATTATTGGCGTTTC TGAAAGAGCCGGA GCCGCCAAAA	GGTTTCCGCG ACCTGTTTGAT AAACTGCCGC AGTTTAAGCA AGTATTGAAC	5.9

Table 3.1: PausePred output: The PausePred output for a sample set of genes using Ribosomal profiling data generated in an *E.coli* strain lacking elongation factor EFP (Woolstenhulme et al., 2015). The genes were arbitrarily picked from a number of cases where the difference in the pause score compared to the wild-type was at least ≥ 50 , with a window coverage higher than 20 and z-score value greater than 3 (see Figure 3.6).

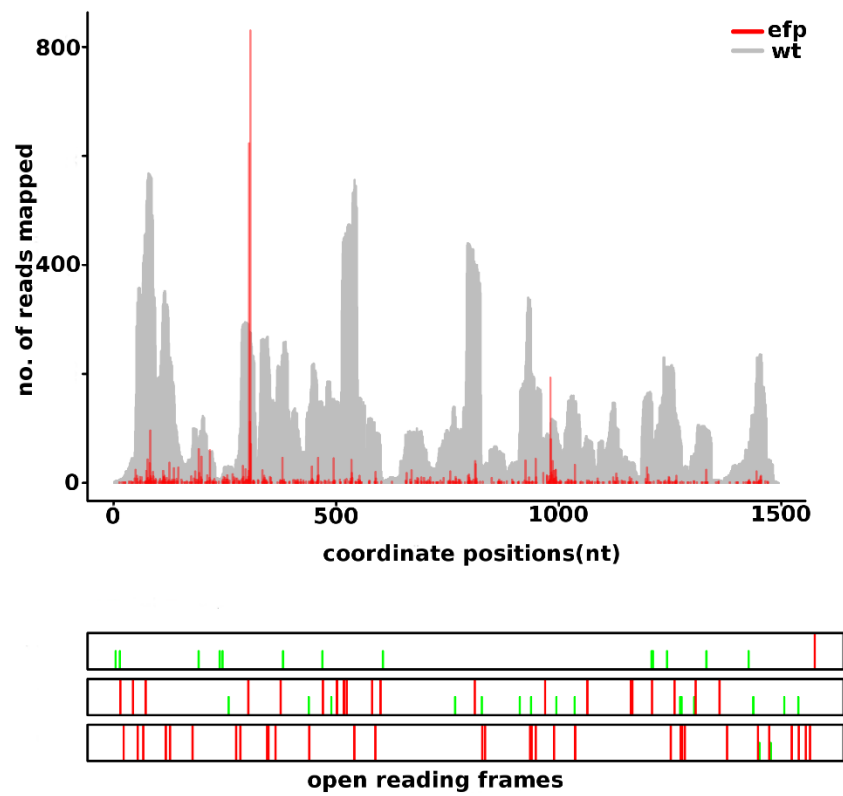


Figure 3.5: Rfeet visualization of the ribosome density profile (red) for the *ubiD* gene containing a ribosome pause detected by PausePred at position 307 of the CDS. A 3' offset of 12 was used to generate the profile using data for an *E. coli* strain lacking elongation factor EFP (Woolstenhulme et al., 2015). The Ribo-seq density for *ubiD* in an *E. coli* wild-type strain is provided as a grey coverage profile, i.e. where the number of sequence reads aligning to each coordinate are displayed. The bottom panel shows the open reading frame (ORF) organization with red lines for stop codons and green lines for AUG codons.

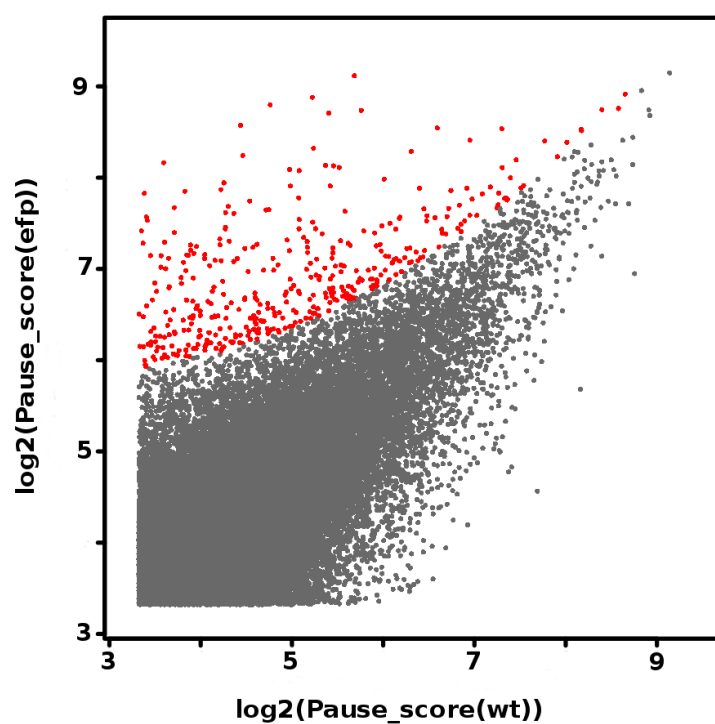


Figure 3.6: Comparison of pause scores predicted by PausePred in *E. coli* mutants lacking elongation factor EFP and the wild-type (WT) sample. A fold change threshold of 10 was used and EFP pauses with a pause score ≥ 50 compared to WT are highlighted in red. The cases provided in Table 3.1 were randomly selected from this set.

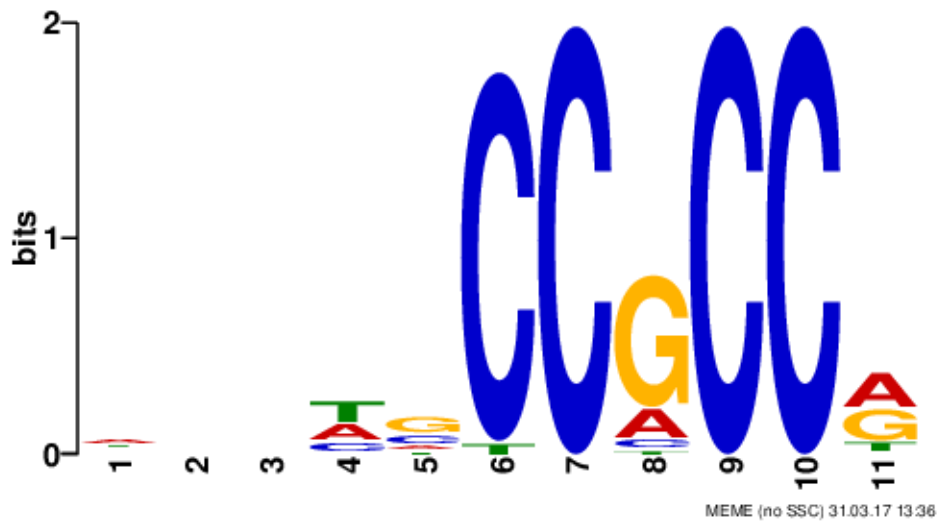


Figure 3.7: The sequence motif found to be overrepresented in 88 cases, using 15 nucleotides upstream of the detected pauses in the filtered 399 cases (Fig. 3.6 highlighted in red). The pause score of these locations is higher than in the wild-type by at least 50.

Enrichment of polyproline motif: The 6 amino acids (protein sequences obtained from NCBI (Coordinators, 2017) assembly accession GCA_000005845.2) upstream of the pause location in the 88 sequences found to contain a C-rich motif (Figure 3.7) were investigated for the presence of a polyproline (PP) motif using Weblogo (Crooks, Hon, Chandonia, & Brenner, 2004). The polyproline motif was found to be overrepresented at the amino acid level (Figure 3.8).

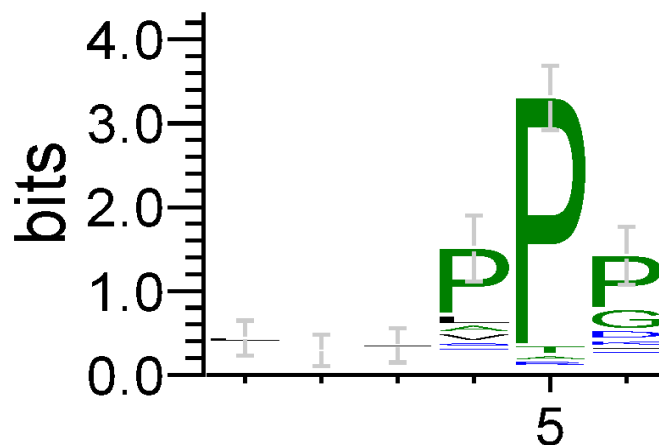


Figure 3.8: The enrichment of the PP motif in the 88 sequences found to contain a C-rich motif at the nucleotide level.

3.2.3 Testing PausePred on ribosome profiling datasets from yeast, rat and human

The PausePred tool was tested on ribosome profiling datasets generated for *E. coli*, yeast, rat and human.

For yeast, we used a control sample from the Yerlikaya et al., 2016 (Yerlikaya et al., 2016). The BAM alignment file was generated using the same approach as for *E. coli* (see Methods section). The offsets for the A-site positions were predicted by creating metagene plots as described in Andreev et al., 2015 (Andreev, O'Connor, Fahey, et al., 2015). The PausePred parameters were set at fold change 10, window size 1000, read length range 28-31 and 5' offset value 15. Table 3.2 provides the output for the top 10 genes (based on their sorted z-score values) predicted to contain pauses.

Systematic ORF name	Standard gene name symbol	Coordinate position	Number of reads mapped	Pause score	Coverage	Z-score
YLR162W-A	RRT15	16	4673	731.3	7.9	78.37
YLR154W-A	-	16	4673	608.86	12.1	64.93
YLR162W	-	334	6249	353.89	11.8	36.94
YOR020C	HSP10	244	41	325.4	5.6	31.84
YLR154W-B	-	133	670	248.06	9.9	25.32
YGR034W	RPL26B	186	16649	231.81	16.1	23.54
YOR207C	RET1	24	18	227.85	5.2	21.89
YPL143W	RPL33A	621	82	204.49	11.5	20.54
YMR251W-A	HOR7	79	56	189.83	7.2	18.01
YDR500C	RPL37B	575	65	170.60	9.7	16.82

Table 3.2. The PausePred output for the top 10 genes based on the sorted z-score values using ribosome profiling data generated for yeast from (Yerlikaya et al., 2016).

Further we tested PausePred tool on a ribosomal profiling dataset generated for rat from Andreev et al., 2015 (accession number: SRR1557705) (Andreev, O'Connor, Fahey, et al., 2015). The parameters used to predict pauses for this dataset include fold change 10, window size 1000, read length range 28-31 and an offset value of 17 for reads of length 28-30 and 18 for reads of length 31. Table 3.3 represents the top 10 genes containing a pause, based on the sorted z-score values.

Transcript name	Gene name	Coordinate position	Number of reads mapped	Pause score	Coverage	Z-score
ENSRNOT00000014382.6	Ppif-201	521	66	474.82	6.1	16.78
ENSRNOT00000079456.1	Ahnak-201	12105	594	310.51	28.9	16.68
ENSRNOT00000087944.1	AABR07000159.2-201	330	186	381.93	8.2	16.67
ENSRNOT00000076692.1	Actr1b-002	1290	116	527.27	6.6	16.60
ENSRNOT00000088387.1	Ahnak-202	3921	501	186.6	40.6	16.49
ENSRNOT00000046799.2	Rpl32-202	102	137	199.71	19.5	16.19
ENSRNOT00000045382.3	Rn50_X_0604.1	147	137	235.80	17.2	16.18
ENSRNOT00000078388.1	AABR07043776.1-201	309	186	267.24	11.9	16.07
ENSRNOT00000011244.2	LOC688684-201	96	137	222.40	16.4	15.99
ENSRNOT00000051890.3	AABR07067274.1-201	90	232	231.31	12.2	15.97

Table 3.3. The PausePred output for the top 10 genes based on the sorted z-score values using ribosome profiling data generated for rat from (Andreev, O'Connor, Fahey, et al., 2015).

To test the PausePred tool on a human ribosome profiling dataset we selected data from Calviello et al., 2016 (accession number: SRR2433794) (Calviello et al., 2016). The parameters used for this dataset were the same as for the (Andreev, O'Connor, Fahey, et al., 2015) rat dataset. Table 3.4 shows the top 10 genes containing a pause based on the sorted z-scores.

Transcript name	Gene name	Coordinate position	Number of reads mapped	Pause score	Coverage	Z-score
ENST00000604031.5	OBSL1	59	169	526.48	8.7	16.58
ENST00000455023.5	VOPP1	457	130	460.99	7.7	16.47
ENST00000436656.5	GAS5	412	310	437.24	15.5	16.29
ENST00000262854.10	HUWE1	7288	275	409.28	14.8	16.19
ENST00000481053.5	CARD9	3951	128	486.69	7.7	16.16
ENST00000540865.5	SNHG1	518	1262	781.42	8.3	16.16
ENST00000371738.3	DNLZ	612	128	396.28	11.9	16.13
ENST00000594841.5	BRD4	1413	219	501.98	9.3	16.11
ENST00000290776.12	CPNE2	1936	103	436.44	7.2	16.11
ENST00000443799.5	GAS5	887	228	439.31	11.3	16.06

Table 3.4. The PausePred output for top 10 genes based on the sorted z-score values using ribosome profiling data generated for human from (Calviello et al., 2016).

3.2.1 Detection of previously reported pauses using PausePred

Ribosome pausing at an upstream open reading frame (MAGDIS) has been reported for the AMD1 gene (Ruan, Shantz, Pegg, & Morris, 1996). Using PausePred and ribosome profiling data generated for human (Calviello et al., 2016), we predicted this pause (mRNA position 47) and additional pauses at positions 837 and 1725 with pause scores 114.98, 59.32 and 373.87 and z-scores of 5.55, 3.89 and 9.03 respectively (see Figure 3.9). In Figure 3.10 we provide the Rfset generated ribosome profile for the AMD1 gene showing the 3 detected pauses.

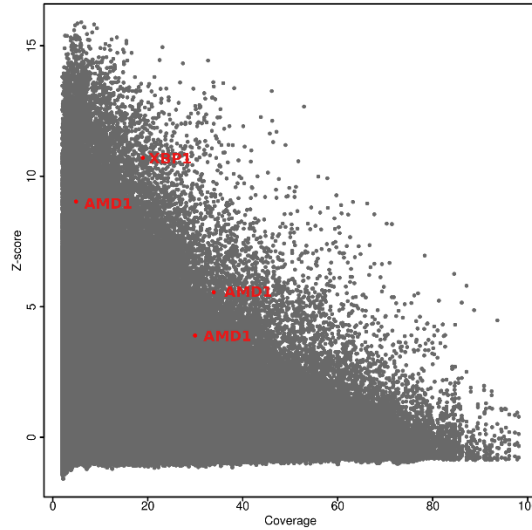


Figure 3.9: The z-score and coverage of all the pauses predicted in the (Calviello et al., 2016) data. The pauses predicted in AMD1 and XBP1 are highlighted in red.

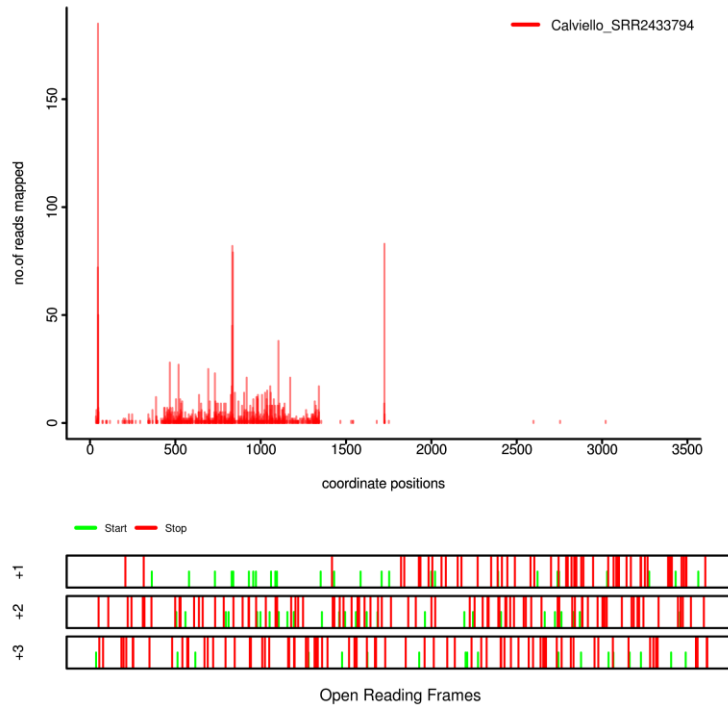


Figure 3.10: Ribosome profile and open reading frame architecture generated with Rfeet showing the footprint density for the human AMD1 gene (transcript ENST00000368885) using data from Calviello et al., 2016 (Calviello et al., 2016). Pauses were detected at positions 47, 837 and 1725 with a pause score of 114.98, 59.32 and 373.87; z-score value of 5.55, 3.89 and 9.03.

PausePred also detected the known pause in XBP1u (Ingolia et al., 2011; Yanagitani, Kimata, Kadokura, & Kohno, 2011) at mRNA position 832 with pause scores 194.6 and z-score value of 10.70 (Figures 3.9 and 3.11).

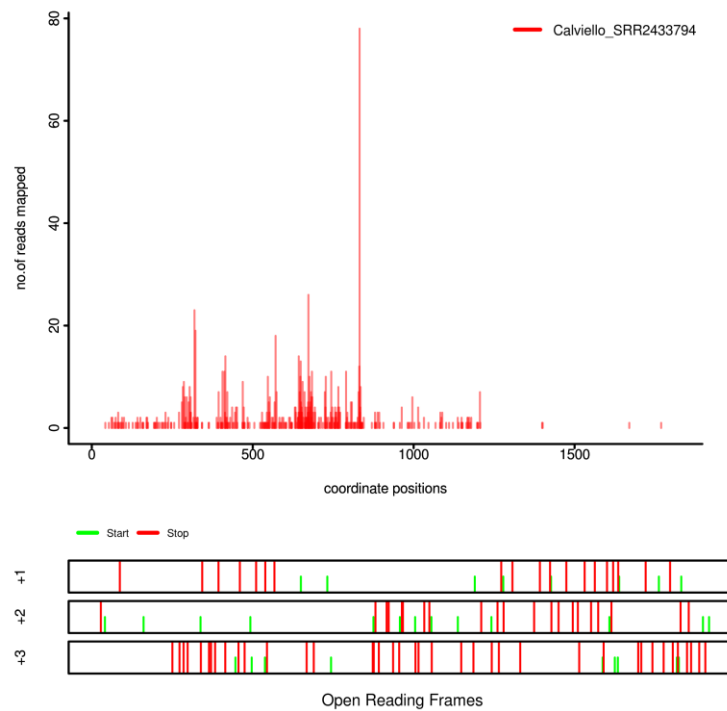


Figure 3.11: Ribosome profile and open reading frame architecture generated with Rfeet showing the footprint density for the XBP1 gene (transcript ENST00000403532) using data from Calviello et al., 2016 (Calviello et al., 2016). A pause was detected in the XBP1 gene at position 832 with pause scores 194.6 and z-score value of 10.70 using PausePred.

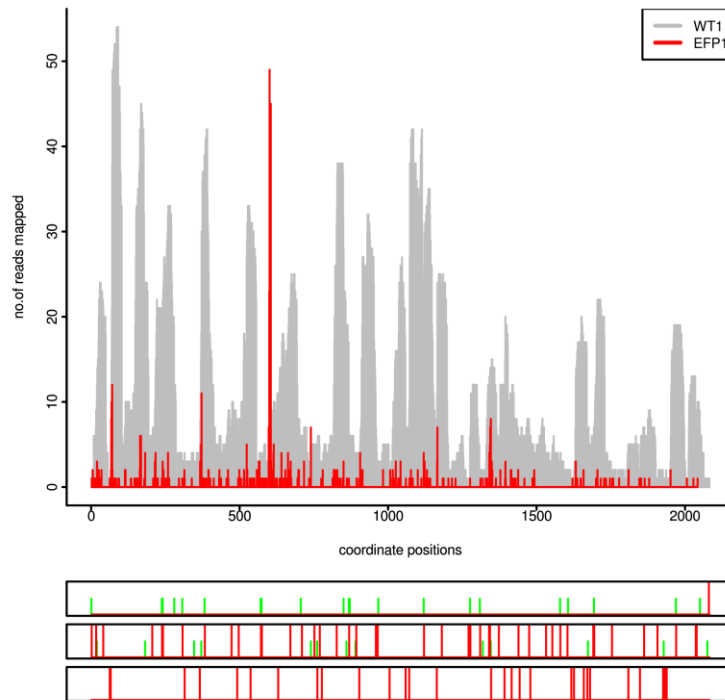


Figure 3.12: Ribosome profile and open reading frame architecture generated with Rfeet showing the footprint density for the gene *recG* using data from Woolstenhulme et al., 2015 (Woolstenhulme et al., 2015). The *recG* gene was found to contain a pause at position 616 with pause score 246.05 and z-score value 6.9 using Pausepred approach.

PausePred also detected many of the pauses predicted by Woolstenhulme et al., 2015 (Woolstenhulme et al., 2015). One such example is the pause at position 616 for the *recG* gene with a pause score 246.05 and z-score value of 6.9 (Figure 3.12).

3.3 Discussion

A number of studies (Ingolia et al., 2011; G. W. Li, Oh, & Weissman, 2012; Matsuo et al., 2017; Mohammad et al., 2016; Woolstenhulme et al., 2015) have relied on in-house scripts to identify pause sites in Ribo-seq data. PausePred provides users with a browser-based translational pause prediction tool. Coupled with the visualization functionality offered by Rfeet, these webtools avoid the need to download software or develop computer programs. For researchers familiar with the command line, stand-alone versions of both tools are also available. To illustrate the utility of these tools, a downloadable example dataset is provided along with step by step help instructions

including screen shots. In future, we plan to enhance the functionality of PausePred by incorporating differential pause analysis.

PausePred and Rfeet are designed for specific purposes in ribosome profiling data analysis, pause detection and visualisation of local footprint density respectively and cannot be used as stand-alone tools for ribosome profiling data analysis which is a multistep process that requires processing of raw reads, alignment to reference sequences, removal of rRNA and tRNA contaminations. There are existing platforms that have been developed for this purpose, e.g. RiboGalaxy (Michel et al., 2016) and Plastid (Dunn & Weissman, 2016) that provide the tools for these steps as well as many additional utilities. We plan to integrate PausePred and Rfeet into RiboGalaxy which should make it readily available for the users of RiboGalaxy.

3.4 Methods

The PausePred and Rfeet tools are written in PERL and can be used *via* a web browser interface at <https://pausepred.ucc.ie/>. A standalone version is available for download on GitHub at https://github.com/romikasaini/Pausepred_offline. The web-based versions are served using the Apache2 web server. The website is designed using HTML and JAVA.

The PausePred workflow requires a reference sequence in FASTA format (genomic or transcriptomic) and Ribo-seq data alignments in sorted BAM format. The *E. coli* Ribo-seq data that we used to illustrate the utility of PausePred were downloaded from NCBI GEO accession number GSE64488 (Woolstenhulme et al., 2015). Raw reads need to be pre-processed to remove adapters sequences and rRNA contaminations followed by the alignment to reference sequences prior to using Pausepred. Raw reads were trimmed using Cutadapt (Martin, 2011) and reads that mapped to rRNA and tRNA sequences were removed using the Bowtie short read alignment program (version 1.1.1) (Langmead et al., 2009) allowing a maximum of three mismatches. The *E. coli* transcriptome obtained from NCBI assembly accession: GCA_000005845.2 and assembly name: ASM584v2 was used as a reference sequence for the alignment of the remaining Ribo-seq reads also using Bowtie version 1.1.1. (Langmead et al., 2009), with no more than three mismatches allowed. The Bowtie SAM outputs were converted to BAM format and sorted using Samtools (Li et al., 2009), The sorted BAM

files were uploaded to the PausePred webtool (<https://pausepred.ucc.ie/>). We used a 3' offset of 12 nt as mentioned in Woolstenhulme et al., 2015 (Woolstenhulme et al., 2015).

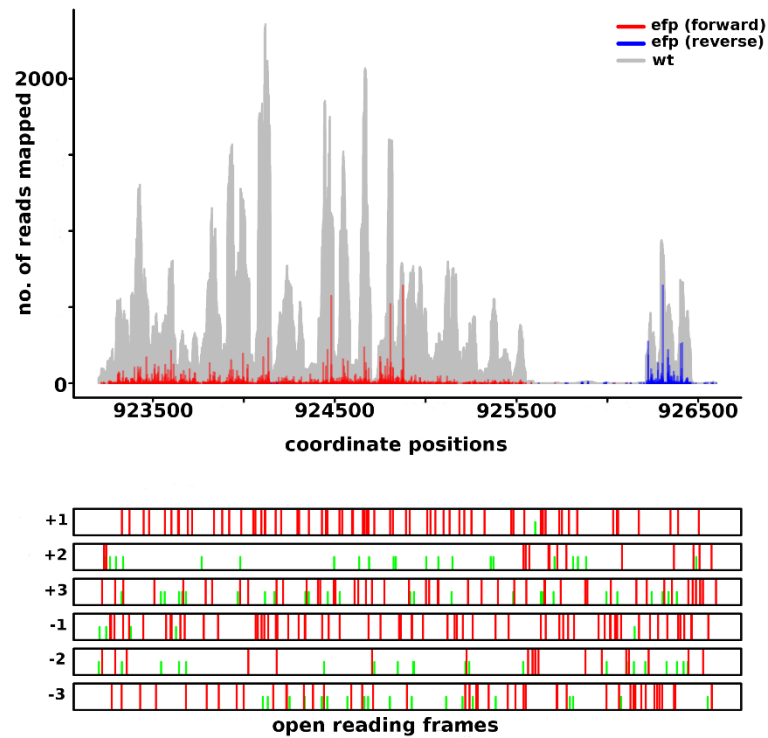


Figure 3.13: Rfeet generated ribosome profiles using data obtained in *E. coli* mutants lacking elongation factor EFP (Woolstenhulme et al., 2015). The region shown corresponds to the *E. coli* genomic segment containing the *clpA* gene (forward strand in red) and the *infA* gene (reverse strand in blue). The background footprint coverage for the wild-type sample is shown in grey in a strand non-specific manner.

Rfeet takes a BAM alignment file (single end reads only) and a reference sequence FASTA file as inputs, and produces Ribo-seq and RNA-seq (if supplied) read density profiles for a sequence name/ID. The nucleotide sequences are fetched with bioperl modules to provide the open reading frame (ORF) architectures for the mRNA (forward strand) or genomic region (forward and reverse strands) with colour delineation for the strand orientation and the start (ATG) and stop (TAG, TAA, TGA) codons. The ribosome footprint density plots, i.e. ribosome profiles, display 5' or 3' footprint end counts or the inferred positions of the decoding center when an offset is

used. The option to use a second BAM alignment file is available for comparing additional RNA-seq or Ribo-seq alignments. For a given profile region (Gene/Transcript/Chr:start-stop), Rfeet provides two plot type options. The “absolute plot” will show the raw counts of the mapped reads and is the default option. Absolute counts are also plotted for the second input file when a coverage plot is selected. A coverage plot provides the number of sequence reads that align to each coordinate. The "normalized plot" is generated by dividing the number of reads mapped on each position by the total number of reads mapped in a particular profile.

$$\frac{r_i}{\sum_{i=1}^n r_i} \quad \{2\}$$

where r is the number of reads mapped to each position i across the profile region of length n .

To generate individual plots for more than one transcript or a genomic region, a comma separated list of IDs should be used. Red and blue colours are used to display ribosome footprints aligned to positive and negative strands respectively (Fig 3.13).

Chapter 4

Translational pauses prediction using polyamine-stimulated ribosomal profiling data

Polyamines play important role in cellular processes like transcription, translation, cellular stress, cell growth and development. To study the role of polyamines in translation regulation, ribosome profiling technique was used to analyse samples with varying concentrations of polyamines. In this chapter, polyamine depleted and polyamine repleted samples were used to predict ribosome pauses which are sensitive to altered polyamine concentration, using PausePred tool (ref. Chapter 3). This analysis revealed enrichment of proline-proline and proline-glycine motif in the E and P-site of the stalled ribosomes using sample treated with 2mM concentration of Spermidine.

4.1 Introduction

Polyamines are ubiquitous organic cations of low molecular weight containing two or more amino groups (Pegg and McCann 1982) (Figure 4.1). In year 1678, Antonie van Leeuwenhoek reported crystalline substances in seminal fluid (reviewed in Bachrach 2010 (Bachrach 2010)). Later in year 1888 Ladenburg and Abel named one of the crystalline substances as Spermine. The early phase of polyamine studies was concluded by Rosenheim in year 1924 by determining the structure of putrescine, spermine and spermidine (reviewed in Bachrach 2010 (Bachrach 2010)).

The effect of the polyamines on the living cells was first studied in bacteria, where it was reported that polyamines stimulate the growth of *Hemophilus parainfluenza*, *Neisseria perflava* and *Pasteurella tularensis*, whereas higher concentration of polyamines lead to the inhibition of gram positive bacteria growth (reviewed in Bachrach 2010 (Bachrach 2010)). The polyamines are present in all organisms in varying concentrations with *E. coli* containing higher concentration of putrescine and many other bacteria and eukaryotes containing higher amount of spermine and spermidine (reviewed in Miller- Fleming et al., 2015 (Miller-Fleming et al. 2015)). The amino acids ornithine and methionine are predominantly used for the polyamine synthesis whereas arginine and lysine are the alternative source of polyamine production (reviewed in Kusano et al., 2008 (Kusano et al. 2008)). The levels of polyamines regulate the efficiency of frameshifting leading to the homeostatic feedback loop (Palanimurugan et al. 2004). Polyamines play an important role in multiple cellular processes, which will be discussed in the next section.

4.1.1 Function of polyamines

Cell growth and development: The requirement of the polyamines for the optimal growth of the cells is absolute. It has been reported that inhibition of polyamines biosynthesis blocks the cell growth (Bitoni and Mccann 1987) and decrease in the concentration of putrescine and spermidine was reported to reduce the cell growth in *E.coli* (Cunningham-Rundles and Maas 1975). In eukaryotic cells a complete arrest of the cell proliferation was reported after reduction in the polyamine levels (Chattopadhyay, Tabor, and Tabor 2002). Also, progressive increase in the polyamine concentrations was reported during the different stages of the cell cycle (Fredlund et al. 1995). In *Saccharomyces cerevisiae*, spermidine was found to be crucial under aerobic conditions implying the role of different polyamines in different metabolic conditions (Balasundaram, Tabor, and Tabor 1991).

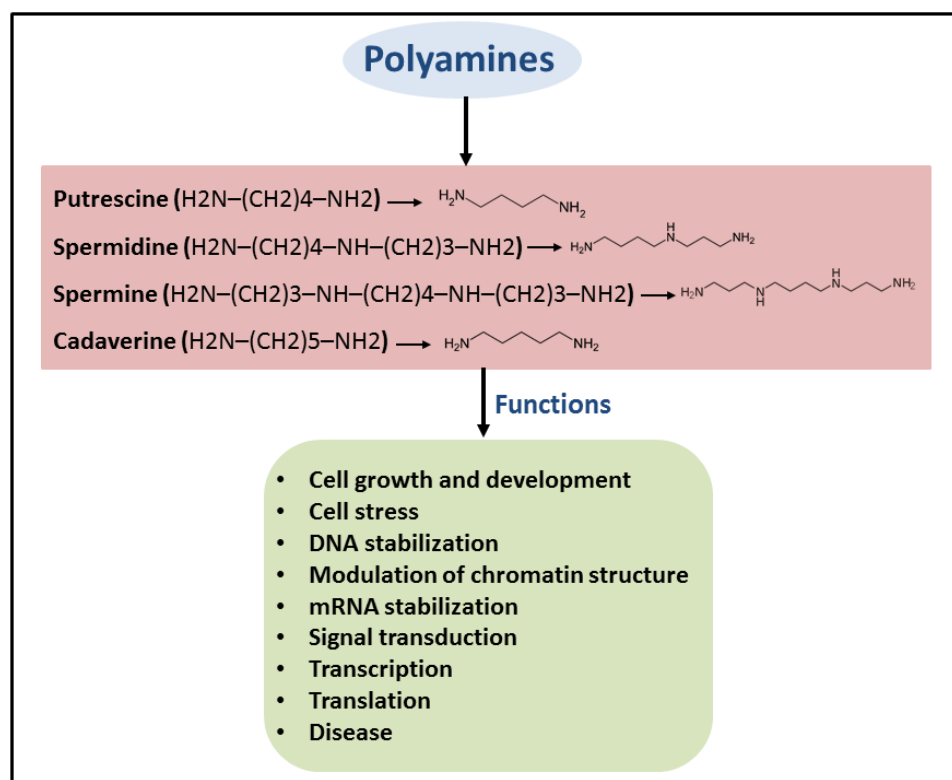


Figure 4.1: The structures and functions of primary polyamines.

Cellular stress: It has been observed that the concentration of the polyamines changes in the cell during its exposure to stress. The cells sensitivity to stress was found to be influenced by modification in the polyamine concentrations (Valdés-Santiago and Ruiz-Herrera 2013). Accumulation of putrescine has been observed in rat hepatoma cell cultures and plants during the osmotic stress. The rat hepatoma cell culture study suggested that regulation of antizyme synthesis provides an effective mechanism for maintaining stable polyamine levels and the mechanism behind the cells steady-state polyamine levels in response to stimulation or stress can be the alteration in the rate of antizyme degradation (Mitchell et al. 1998). Whereas in plants osmosis stress leads to accumulation of the putrescine to a toxic level due to increase in arginine decarboxylase pathways activity (Watson and Malmberg 1996). Some other functions of the polyamines during the stress involve scavenging of ROS, binding to membrane transport proteins and regulating expression of proteins related to the stress response, protecting cell from H_2O_2 by altering mitochondrial permeability transition (reviewed in Miller-Fleming et al., 2015 (Miller-Fleming et al. 2015)).

Gene expression (Transcription and translation): Polyamines play an important role in gene expression by binding to nucleic acids and proteins. Consequently, stabilizing and remodelling the chromatin structure (reviewed in Childs et al., 2003 (Childs, Mehta, and Gerner 2003)). In year 1987, it was reported that spermidine plays a crucial role in controlling the rate of transcription in mycobacteria. It was inferred that RNA synthesis regulation happens by spermidine stabilizing the DNA template and destabilizing the small DNA-RNA hybrid region, consequently helping the RNA polymerase to move on the DNA template easily during the process of chain elongation (Jain and Tyagi 1987). Polyamine mediated DNA structural changes were reported to affect the affinity of the mammalian estrogenic receptor (ER) for its response elements (ERE) (Thomas et al. 1995). Polyamines were reported to play a role in activation of c-MYC transcription which was mediated by altering the quadruplex structure (present in the regulatory sequence of this gene) into an active conformation (Kumar, Basundra, and Maiti 2009). The RNA transcription is first and important step for the gene expression, the role of polyamines at the level of transcription needs to be explored further.

Polyamines were reported to decrease the Mg^{2+} concentration which is necessary for the protein synthesis (Igarashi et al. 1973). A set of various artificial mRNAs was used to test the polyamine stimulation of translation (Igarashi, Watanabe, and Hirose 1975). Further it was found that length of 5' leader of ornithine decarboxylase (ODC) mRNA affects the level of the polyamine stimulation of mammalian ODC synthesis which was studied in a rabbit reticulocyte cell-free system (Ito et al. 1990). Also, polyamines were reported to increase the fidelity of the polypeptide synthesis which occurred at the level of aminoacyl-tRNA binding to ribosomes. It was reported that the assembly of the 30S ribosomal subunit in E. coli was stimulated by polyamines (reviewed in Igarashi et al., 2015 (Igarashi and Kashiwagi 2015)).

Polyamines synthesis: The level of the polyamines is regulated by the enzyme Ornithine decarboxylase (ODC) which catalyses the synthesis of putrescine, spermidine and spermine are further derived from putrescine. The protein antizyme acts as inhibitor of ornithine decarboxylase (ODC). It was reported that there are two overlapping reading frames in the antizyme and +1 frameshifting is required for the synthesis of antizyme. The frameshifting event of the antizyme is effector of an autoregulatory circuit and sensor of polyamine levels. Increased polyamine levels

elevate the frameshifting dynamically. Consequently, elevated levels of antizyme leads to reduced synthesis and cellular import of polyamines (reviewed in Atkins et al., 2016 (Atkins et al. 2016)) (Figure 4.2).

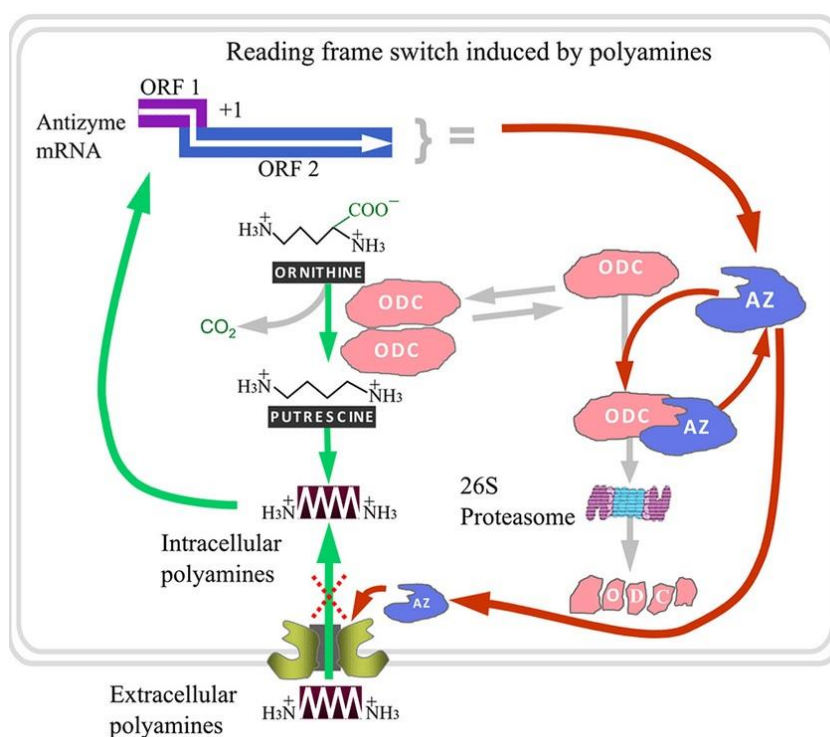


Figure 4.2: Synthesis of Polyamines. Figure illustrating the role of negative regulator, antizyme, in regulating polyamine homeostasis (Figure taken from (Yordanova, Wu, Andreev, Sachs, & Atkins, 2015)).

To explore polyamine regulated translational control, I predicted translational pauses using the ribosome profiling data of polyamine depleted and polyamine replete cells using PausePred tool. The PausePred tool (Chapter 3) identifies translational pauses by scoring the peaks of ribosome footprint density based on their magnitude relative to the background density within the surrounding area.

4.2 Results

The polyamine levels are known to influence the protein translation rate by affecting the various translation mechanisms such as frameshifting (reviewed in Atkins et al., 2016 (Atkins et al., 2016)). The current study describes the effect of polyamine levels on the translational pauses. The ribosome profiling data samples with difluoromethylornithine (DFMO) and 2mM spermidine concentration were used for this analysis with two sets of replicates. To detect the ribosome stalls based on the ribosome footprint density, PausePred tool was used (Chapter 3), see methods sections for parameters used to detect pauses.

Using PausePred tool the total number of pauses predicted in DFMO and 2mM spermidine samples of both replicates were in the range of 960882-1366643 (Figure 4.3). The 2mM spermidine condition was found to contain ~1.35 folds higher number of total predicted pauses than the DFMO condition in both the replicates (Figure 4.3).

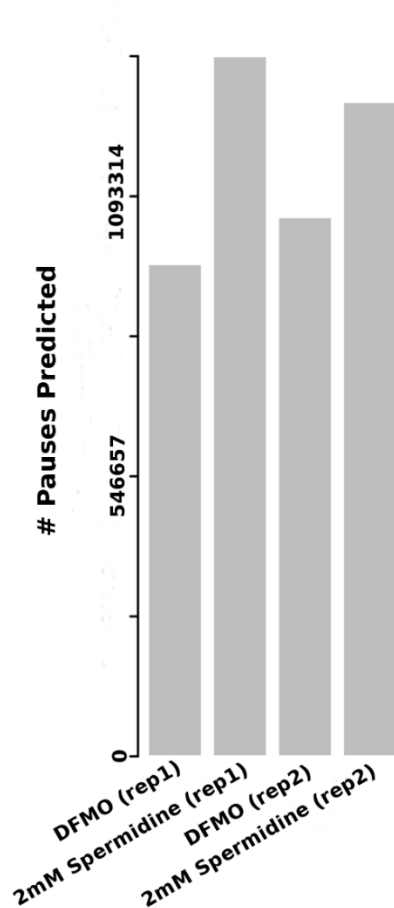


Figure 4.3: The total number of pauses found in the DFMO and 2mM Spermidine conditions for both replicates.

To evaluate the differential regulation of the translational pauses in depleted and repleted polyamine conditions, the square root of the sum of squared pause scores from replicate 1 and 2 were calculated for both DFMO and 2mM Spermidine samples (see equation 1 in section methods). Figure 4.4 represents the comparison of the DFMO and 2mM Spermidine pause scores values calculated for individual coordinate positions (Equation 1). The coordinate positions found to have a pause score difference of ≥ 100 in both conditions were considered as differentially regulated (Figure 4.4).

No real change was recorded in the number of differentially regulated translational pauses in the polyamine depleted (DFMO) and repleted (2mM spermidine) conditions. A total of 300 genes were found to have up-regulated pauses in DFMO condition whereas 237 genes were found to contain up-regulated pauses in 2mM spermidine condition (highlighted in red in Figure 4.4).

Flanking sequences analysis

Dinucleotide frequency: The dinucleotide frequency of the 15nt up and downstream of the differentially regulated pauses was analysed using EMBOSS compseq tool (Rice, Longden, & Bleasby, 2000). This analysis revealed enrichment of CC and GG dinucleotide in the upstream region of the pauses up regulated in 2mM spermidine condition (Figure 4.5). The observed/expected frequency of the CC and GG dimer in the 2mM Spermidine condition was found to be 1.82 and 1.42.

Pauses predicted in
2mM Spermidine
condition only

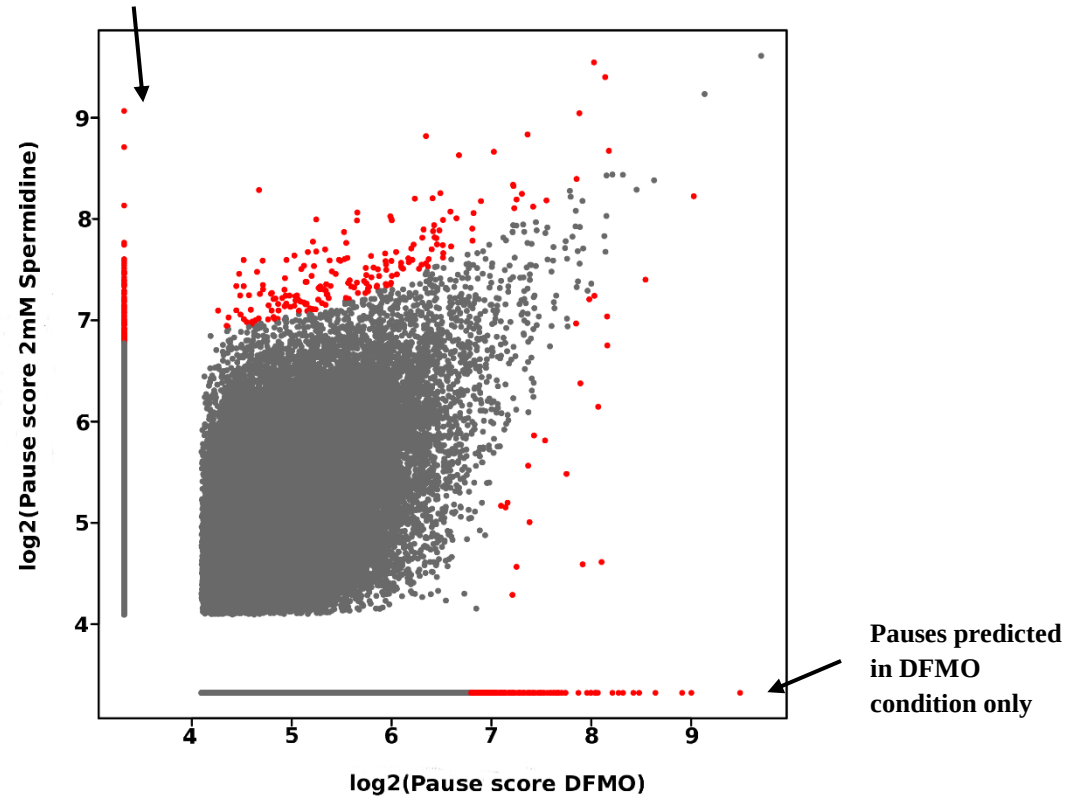


Figure 4.4: The comparison of the pauses scores predicted in DFMO and 2mM Spermidine condition. The ribosome pauses selected as differentially regulated (≥ 100 pause score difference in DFMO and 2mM Spermidine condition) are highlighted in red.

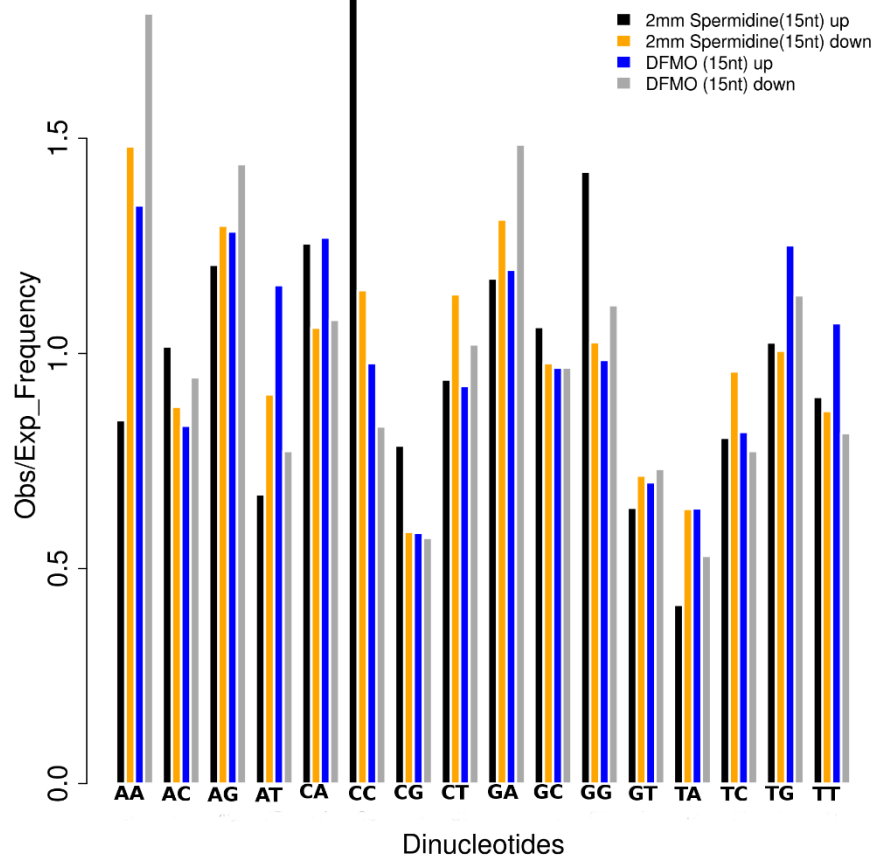


Figure 4.5: Dinucleotide frequency of flanking sequences: The dinucleotide observed/expected frequency count for the 15nt up and downstream sequences of the differentially regulated pauses in DFMO and 2 mM spermidine conditions.

Motif analysis: The 15nt flanking sequences of the differentially regulated pauses in both the conditions were analysed for over-represented motifs using Weblogo (Crooks et al., 2004) representation. There was over-representation of Adenine nucleotide at the +1, +2 and +3 in the differentially regulated pauses in condition DFMO (Figure 4.6 (A)). The 2mM Spermidine sample was found to contain over-representation of a CC rich motif starting at position -6, upstream of the pause location in the differentially regulated pauses (Figure 4.7 (A)).

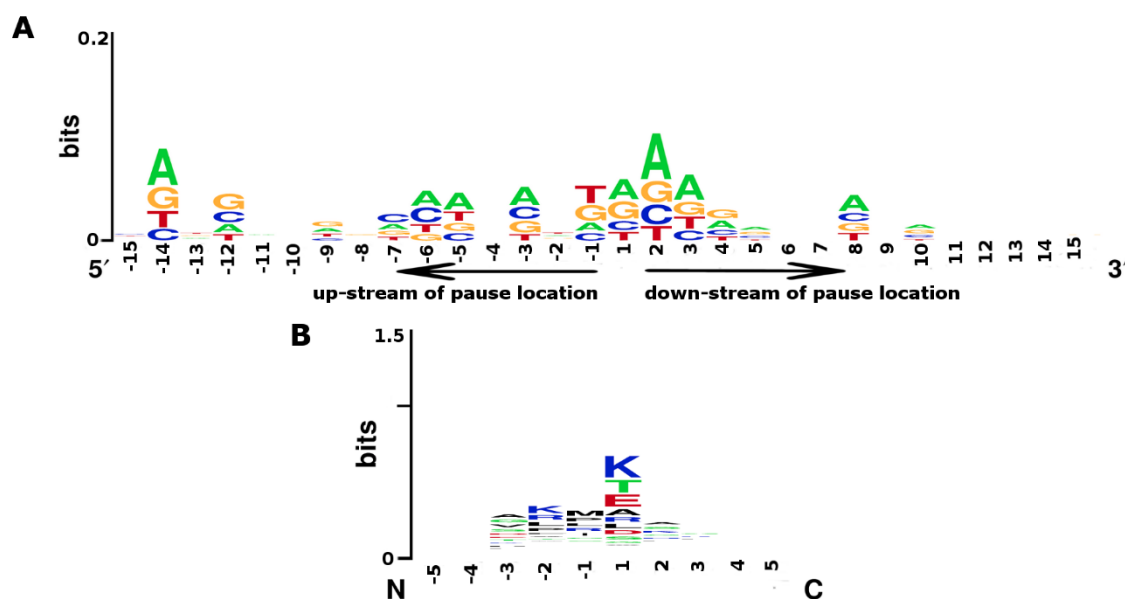


Figure 4.6: Over-represented motifs in differentially regulated pauses in condition DFMO: (A) The over-represented nucleotides in the flanking regions of the pauses, where the nucleotide at the +1 position represents pause location. (B) The amino acid enrichment in the flanking sequence of predicted pauses where +1 represents A-site location of a ribosome and -1 and -2 represent P and E-sites respectively.

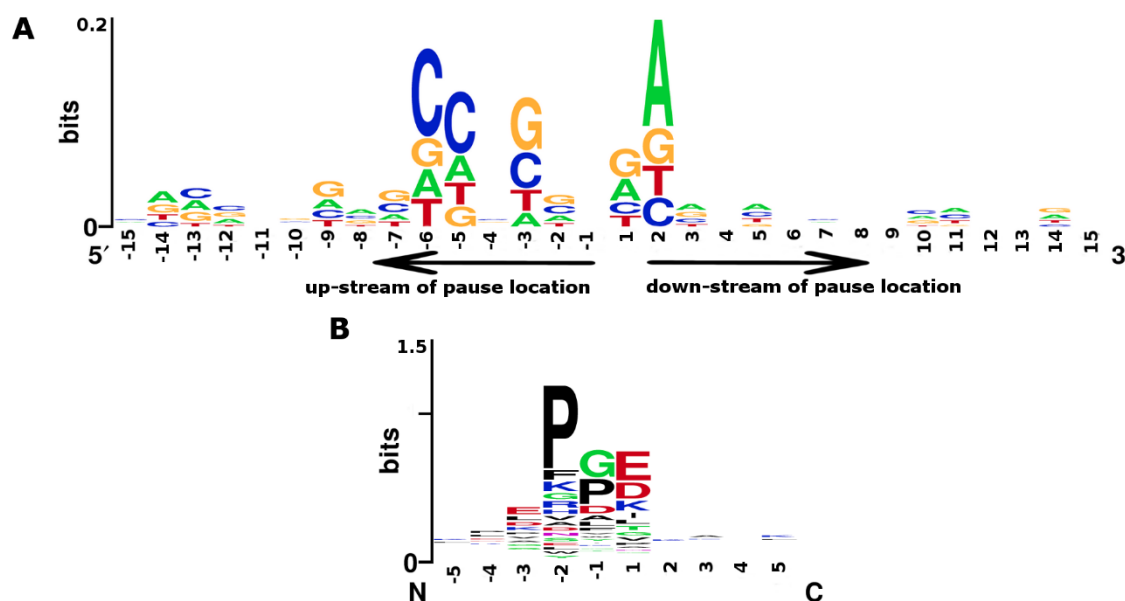


Figure 4.7: Over-represented motifs in differentially regulated pauses in condition 2mM spermidine: (A) The over-represented nucleotides in the flanking regions of the pauses, where the nucleotide at the +1 position represents pause location. (B) The amino acid enrichment in the flanking sequence of predicted pauses where +1 represents A-site location of a ribosome and -1 and -2 represent P and E-sites respectively.

After noticing the enrichment of motifs at the nucleotide level, the flanking sequences of the differentially regulated genes were analysed at amino acid level (see methods). For DFMO sample at the amino acid level, lysine (K) was found to be enriched in the A-site location (Figure 4.6 (B)). For 2mM Spermidine sample, this analysis revealed that the E-site of the stalled ribosome is enriched in proline amino acid (Figure 4.7 (B)). The P-site of the stalled ribosome in 2mM Spermidine condition was found to have over-representation of proline (P) or glycine (G) amino acids. The genes found to contain proline-glycine (PG) (Table 4.1) and proline-proline (PP) (Table 4.2) motif enriched in the E and P-site of ribosome were further analysed by generating ribosome density plots. Figure 4.8 shows the ribosome density plot for gene *ADCK2* and *COX8A* which were predicted to have differentially regulated pauses with proline (P) and proline (P) amino acid in the E and P site of the stalled ribosome (Table 4.2) (Figure 4.8). Figure 4.9 shows an example of genes *IER3IP1* and *SLC35G2* having Proline-Glycine amino acid in the E and P-site of stalled ribosome (Table 4.1).

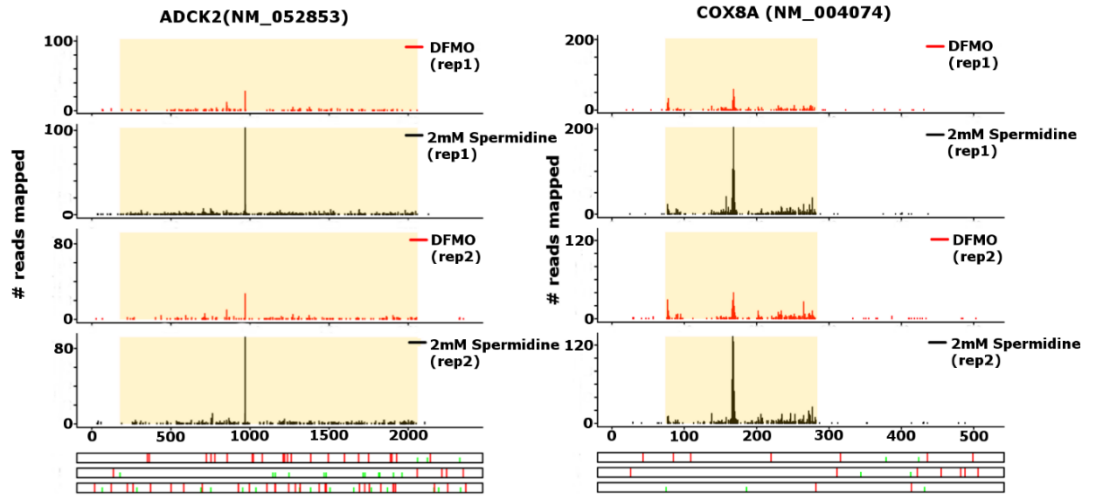


Figure 4.8: The mRNAs containing an up-regulated pause in 2mM Spermidine condition with a proline-proline motif upstream of the pause. Where for each gene top two rows represent DFMO and 2mM Spermidine conditions of replicate 1 and next two rows represent same conditions for replicate 2. The bottom panel for each mRNA provides the reading frame information with start codons shown as green dashes and stop codons shown as red dashes. The CDS region has been highlighted in the yellow colour (see Table 4.2 for details).

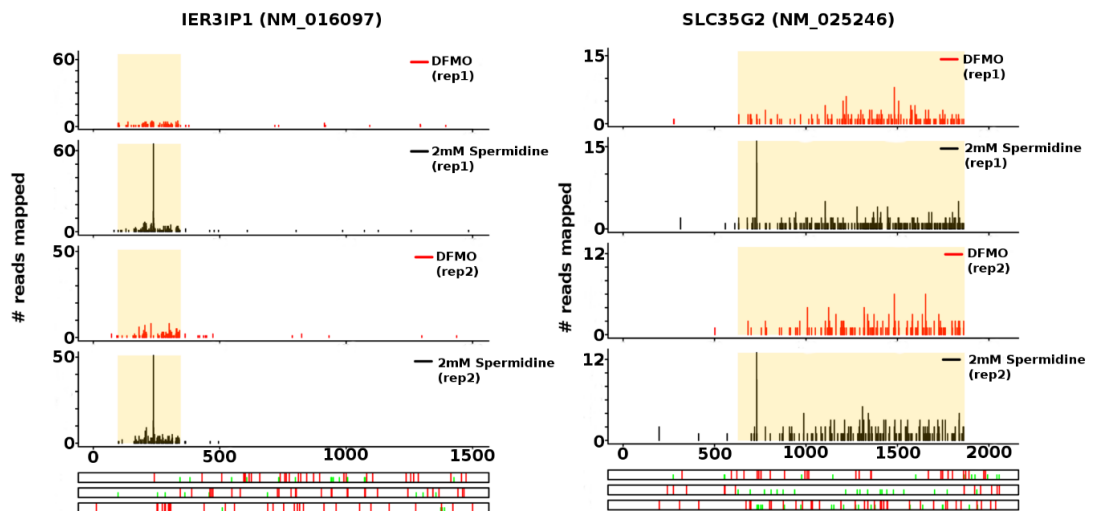


Figure 4.9: Genes containing an up-regulated pause in 2mM Spermidine condition with a proline-glycine motif upstream of the pause. Where for each gene top two rows represent DFMO and 2mM Spermidine conditions of replicate 1 and next two rows represent same conditions for replicate 2 (see Figure 4.8 for notation) (See Table 4.1 for details).

gene symbol	Refseq transcript ID	Coordinate position	Pause score		Flanking protein sequence
			(rep1)	(rep2)	
<i>TUBB2A</i>	NM_001069	300	119.15	46.95	DLEPGTMDSV
<i>AMDHD2</i>	NM_001145815	836	92.635	38.71	HRDPGIVGLL
<i>LGALS1</i>	NM_002305	346	106.8	43.48	PFQPGSVAEV
<i>EFNA4</i>	NM_005227	468	89.74	45.45	EFLPGETYYY
<i>PTPRU</i>	NM_005704	475	41.32	121.21	GHSPGTLGVY
<i>VAV3</i>	NM_006113	1834	125	35.71	QVDPGLPKMQ
<i>ARHGEF12</i>	NM_015313	3786	89.745	55.25	QSTPGEGDND
<i>ICE1</i>	NM_015325	913	82.805	63.28	HRLPGEGRSC
<i>PHF20L1</i>	NM_016018	3137	97.83	69.93	TEDPGDSHLQ
<i>IER3IP1</i>	NM_016097	238	65.29	95.54	GEEPGLIKSQL
<i>BCAS4</i>	NM_017843	275	108.7	98.04	TPEPGAEAKE
<i>ZNF358</i>	NM_018083	518	109.89	48.34	TLSPGDPKVD
<i>SLC35G2</i>	NM_025246	731	68.63	82.97	YPQPGDDGYE
<i>ALKBH7</i>	NM_032306	845	46.585	93.83	TQEPGEWLEL
<i>PRADC1</i>	NM_032319	183	111.11	13.33	VLSPGDIRYI
<i>SLC2A12</i>	NM_145176	1534	88.61	82.57	VTDPGDVPAF
<i>C4orf32</i>	NM_152400	292	67.31	82.64	EWDPGASPRR
<i>MTURN</i>	NM_152793	315	97.83	23.53	YADPGVSFYV
<i>N6AMT2</i>	NM_174928	184	57.14	84.62	QIEPGEDDKY
<i>C7orf55</i>	NM_197964	575	86.33	70.925	PHQPGGKGWE
<i>COL25A1</i>	NM_198721	1483	88.255	72.795	KGEPGESGRP
<i>AFTPH</i>	NM_203437	800	74.38	68.315	SFSPGDFRTN
<i>LYSMD1</i>	NM_212551	777	100	56.625	QLEPGDTLAG
<i>C1orf52</i>	NR_024113	692	44.415	247.67	KVEPGEPAKK
<i>NUDT16</i>	NR_033268	306	108.11	30.53	LGFPGGFVDT
<i>SARS</i>	NR_034072	194	114.29	148.51	FKDPGLVDQL
<i>RNF6</i>	NM_183044	541	43.17	82.935	LGTPEITSE

Table: 4.1: Showing the genes predicted to contain up-regulated pauses in 2mM spermidine condition with a proline-glycine (PG) motif enriched in the E and P-site of the ribosome. The amino acids in E and P–sites have been highlighted in blue colour.

Gene symbol	Refseq transcript ID	Coordinate position	Pause score		Flanking protein sequence
			(rep1)	(rep2)	
<i>SDC1</i>	NM_001006946	485	70	121.43	TNL PP EDQDG
<i>TNRC18</i>	NM_001080495	7197	17.065	99.475	RLL PP DYKIQ
<i>CCHCR1</i>	NM_001105564	2623	112.24	77.38	SPR PP ECsas
<i>C7orf73</i>	NM_001130929	243	95.56	88.44	KKK PP SA*DC
<i>PRRG1</i>	NM_001173489	855	164.56	56.18	LDP PP EYEDI
<i>ZNF706</i>	NM_001267708	667	125.95	32.52	TPL PP ELADV
<i>AMN1</i>	NM_001278412	478	18.52	120.48	KPL PP NIKDR
<i>IRF1</i>	NM_002198	577	117.665	30.3	RML PP LTKNQ
<i>COX8A</i>	NM_004074	168	174.69	168.13	HSL PP EGKLG
<i>KHK</i>	NM_006488	1004	92.235	62.94	TRQ PP EQKIR
<i>PDLIM3</i>	NM_014476	1181	59.41	88.5	RTK PP EGYDT
<i>CAMTA1</i>	NM_015215	316	154.64	111.11	NTV PP IEDDH
<i>RRP7A</i>	NM_015703	243	57.35	88.89	LVN PP YCTEE
<i>LRRC1</i>	NM_018214	507	79.21	74.77	QRL PP EIANF
<i>ZGRF1</i>	NM_018392	689	90.91	44.69	CSM PP LFPTV
<i>CGN</i>	NM_020770	763	83.33	66.995	RML PP EQRKR
<i>SH3RF1</i>	NM_020870	1146	93.02	45.895	WDK PP VPGVD
<i>FAM134A</i>	NM_024293	1691	123.66	84.58	PPK PP DAPPL
<i>SH3BGRL2</i>	NM_031469	345	73.53	124.22	KNV PP EKKPT
<i>IFI27L2</i>	NM_032036	424	56.91	82.28	PPK PP LKSEK
<i>NEURL4</i>	NM_032442	4607	60.61	109.76	LGE PP DPHFS
<i>RITA1</i>	NM_032848	1494	98.59	42.86	KPK PP WK*YS
<i>ADCK2</i>	NM_052853	971	115.41	132.45	GRK PP ENLAD
<i>BRWD3</i>	NM_153252	657	113.99	26.855	RGR PP ELPVN

<i>CSRP2</i>	NR_045072	2109	136.36	75	EDR PP EDSSL
<i>MSTO1</i>	NR_046292	1141	90	135.42	RPE PP VSFPY
<i>GALNT3</i>	NM_004482	901	55.915	101.85	DTR P PECIEQ

Table 4.2: Showing the genes predicted to contain up-regulated pauses in 2mM spermidine condition with a proline-proline (PP) motif enriched in the E and P-site of the ribosome. The amino acids in E and P-sites have been highlighted in blue colour.

4.3 Discussion

Previous studies have indicated that polyamines play an important role in translation regulation by affecting various events of the translation such as initiation, elongation, frameshifting etc (Atkins et al., 2016; Igarashi & Kashiwagi, 2015; Landau, Bercovich, Park, & Kahana, 2010; Mandal, Mandal, Johansson, Orjalo, & Park, 2013). The emergence of ribosome profiling technique has allowed detection of the position of actively translating ribosomes on the mRNA (Ingolia et al., 2009). In this current study we have tried to explore the role of polyamines in influencing the ribosome pauses using the ribosome profiling data. No real change was detected in the differential regulation of the ribosome pauses for DFMO and 2mM Spermidine conditions. The differentially regulated pauses in the higher polyamine concentration were found to be enriched in proline-proline and proline-glycine amino acids in the E and P-site of the stalled ribosome whereas no motif enrichment was recorded in the DFMO condition up-regulated pauses. The mechanism of the translation regulation by the higher concentration of polyamines needs to be explored further.

4.4 Methods

Initial processing: The adapter sequence (5'-CTGTAGGCACCATCAATAGATCGGAAGAGCACACGTCTGAACTCCAGTC A-3') was removed using cutadapt (Martin, 2011). Followed by rRNA contamination removal using bowtie (Langmead et al., 2009) allowing maximum of three mismatches (-v 3). The reads were aligned to human transcriptome sequences obtained from the RefSeq database (O'Leary et al., 2016) (Downloaded on 29th September

2015) from the NCBI (ftp://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens/mRNA_Proc/) using bowtie (Langmead et al., 2009). The reads were aligned to transcriptome with the following parameters (-a, -m 100 -v 2 -norc) i.e. reads were aligned to the positive strand allowing no more than three mismatches and no more than 100 mappings per read.

PausePred analysis: The alignment files obtained from bowtie were further used for pause prediction using PausePred (Chapter 3). The pauses were predicted in DFMO and 2mM spermidine samples for both replicates. The parameters used for PausePred analysis include: window size of 1000, fold change value of 5, read length range: 27-38, window coverage of 5, A-site offset values for individual read lengths (17,16,14,15,16,16,17,17,17,18,18) predicted using riboWaltz (Lauria et al., 2017) and 200 nucleotides (nt) flanking sequence of the predicted pauses. Further the square root of the sum of squared pause scores at a particular position obtained for same condition in both replicates was calculated using following approach:

$$X = \sqrt{(x_1)^2 + (x_2)^2} \quad \text{-----(1)}$$

Where X represent square root of the sum of squared pause scores. And $x_{1\&2}$ represents pause scores predicted for a coordinate position in replicate 1 and 2

Flanking sequence analysis: The flanking nucleotide sequences of the differentially regulated pauses were further analysed for the presence of over-represented motifs using Weblogo (Crooks et al., 2004). To find the enriched motifs in the amino acid sequences encoded by the flanking nucleotide sequences of differentially regulated pauses, the nucleotide sequences were translated to amino acids in all three frames using EMBOSS transeq tool (Rice et al., 2000). The translated sequences were aligned to the human proteome (obtained from the RefSeq database (O'Leary et al., 2016) downloaded on 15 Nov 2017) using BLASTp (Altschul, Gish, Miller, Myers, & Lipman, 1990) to fetch the corresponding sequences for the motif enrichment analysis at the amino acid level. The ribosome density profiles of the genes were created using Rfeet tool (Chapter 3).

Bibliography

- Abaeva, I. S., Marintchev, A., Pisareva, V. P., Hellen, C. U., & Pestova, T. V. (2011). Bypassing of stems versus linear base-by-base inspection of mammalian mRNAs during ribosomal scanning. *EMBO J*, 30(1), 115-129. doi: 10.1038/emboj.2010.302
- Agarwala, P., Pandey, S., Mapa, K., & Maiti, S. (2013). The G-quadruplex augments translation in the 5' untranslated region of transforming growth factor beta2. *Biochemistry*, 52(9), 1528-1538. doi: 10.1021/bi301365g
- Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., . . . Searle, S. M. (2016). The Ensembl gene annotation system. *Database (Oxford)*, 2016. doi: 10.1093/database/baw093
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, 215(3), 403-410. doi: 10.1016/S0022-2836(05)80360-2
- Andreev, D. E., O'Connor, P. B., Fahey, C., Kenny, E. M., Terenin, I. M., Dmitriev, S. E., . . . Baranov, P. V. (2015). Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *Elife*, 4, e03971. doi: 10.7554/eLife.03971
- Andreev, D. E., O'Connor, P. B., Loughran, G., Dmitriev, S. E., Baranov, P. V., & Shatsky, I. N. (2017). Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Res*, 45(2), 513-526. doi: 10.1093/nar/gkw1190
- Andreev, D. E., O'Connor, P. B., Zhdanov, A. V., Dmitriev, R. I., Shatsky, I. N., Papkovsky, D. B., & Baranov, P. V. (2015). Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome Biol*, 16, 90. doi: 10.1186/s13059-015-0651-z
- Arenz, S., Ramu, H., Gupta, P., Berninghausen, O., Beckmann, R., Vazquez-Laslop, N., . . . Wilson, D. N. (2014). Molecular basis for erythromycin-dependent ribosome stalling during translation of the ErmBL leader peptide. *Nat Commun*, 5, 3501. doi: 10.1038/ncomms4501
- Arora, A., Dutkiewicz, M., Scaria, V., Hariharan, M., Maiti, S., & Kurreck, J. (2008). Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex

- motif. *RNA*, 14(7), 1290-1296. doi: 10.1261/rna.1001708
- Ash, P. E., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W. L., Dejesus-Hernandez, M., . . . Petrucelli, L. (2013). Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron*, 77(4), 639-646. doi: 10.1016/j.neuron.2013.02.004
- Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E., & Baranov, P. V. (2016). Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res*, 44(15), 7007-7078. doi: 10.1093/nar/gkw530
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., . . . Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*, 37(Web Server issue), W202-208. doi: 10.1093/nar/gkp335
- Balkwill, G. D., Derecka, K., Garner, T. P., Hodgman, C., Flint, A. P., & Searle, M. S. (2009). Repression of translation of human estrogen receptor alpha by G-quadruplex formation. *Biochemistry*, 48(48), 11487-11495. doi: 10.1021/bi901420k
- Bang, I. (1910). Untersuchungen über die Guanylsäure. *Biochemische Zeitschrift*, 26, 293-311.
- Bachrach, U. (2010). The early history of polyamine research. *Plant Physiol Biochem*, 48(7), 490-495. doi: 10.1016/j.plaphy.2010.02.003
- Balasundaram, D., Tabor, C. W., & Tabor, H. (1991). Spermidine or spermine is essential for the aerobic growth of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, 88(13), 5872-5876.
- Baranov, P. V., Atkins, J. F., & Yordanova, M. M. (2015). Augmented genetic decoding: global, local and temporal alterations of decoding processes and codon meaning. *Nat Rev Genet*, 16(9), 517-529. doi: 10.1038/nrg3963
- Basson, M. A., Akbulut, S., Watson-Johnson, J., Simon, R., Carroll, T. J., Shakya, R., . . . Licht, J. D. (2005). Sprouty1 is a critical regulator of GDNF/RET-mediated kidney induction. *Dev Cell*, 8(2), 229-239. doi: 10.1016/j.devcel.2004.12.004
- Beaudoin, J. D., & Perreault, J. P. (2010). 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res*, 38(20), 7022-7036. doi: 10.1093/nar/gkq557
- Bharti, S. K., Khan, I., Banerjee, T., Sommers, J. A., Wu, Y., & Brosh, R. M., Jr.

- (2014). Molecular functions and cellular roles of the ChlR1 (DDX11) helicase defective in the rare cohesinopathy Warsaw breakage syndrome. *Cell Mol Life Sci*, 71(14), 2625-2639. doi: 10.1007/s00018-014-1569-4
- Bhattacharyya, D., Mirihana Arachchilage, G., & Basu, S. (2016). Metal Cations in G-Quadruplex Folding and Stability. *Front Chem*, 4, 38. doi: 10.3389/fchem.2016.00038
- Biffi, G., Di Antonio, M., Tannahill, D., & Balasubramanian, S. (2014). Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat Chem*, 6(1), 75-80. doi: 10.1038/nchem.1805
- Biffi, G., Tannahill, D., McCafferty, J., & Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem*, 5(3), 182-186. doi: 10.1038/nchem.1548
- Biswas, B., Kandpal, M., Jauhari, U. K., & Vivekanandan, P. (2016). Genome-wide analysis of G-quadruplexes in herpesvirus genomes. *BMC Genomics*, 17(1), 949. doi: 10.1186/s12864-016-3282-1
- Bitonti, A. J., & McCann, P. P. (1987). 12 - Inhibition of Polyamine Biosynthesis in Microorganisms *Inhibition of Polyamine Metabolism* (pp. 259-275): Academic Press.
- Bochman, M. L., Paeschke, K., & Zakian, V. A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet*, 13(11), 770-780. doi: 10.1038/nrg3296
- Bonnal, S., Schaeffer, C., Creancier, L., Clamens, S., Moine, H., Prats, A. C., & Vagner, S. (2003). A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J Biol Chem*, 278(41), 39330-39336. doi: 10.1074/jbc.M305580200
- Brassart, B., Gomez, D., De Cian, A., Paterski, R., Montagnac, A., Qui, K. H., . . . Riou, J. F. (2007). A new steroid derivative stabilizes g-quadruplexes and induces telomere uncapping in human tumor cells. *Mol Pharmacol*, 72(3), 631-640. doi: 10.1124/mol.107.036574
- Bugaut, A., & Balasubramanian, S. (2012). 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res*, 40(11), 4727-4741. doi: 10.1093/nar/gks068
- Butcher, S. E., & Jan, E. (2016). tRNA-mimicry in IRES-mediated translation and

- recoding. *RNA Biol*, 13(11), 1068-1074. doi: 10.1080/15476286.2016.1219833
- Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., . . . Ohler, U. (2016). Detecting actively translated open reading frames in ribosome profiling data. *Nat Methods*, 13(2), 165-170. doi: 10.1038/nmeth.3688
- Cammas, A., Dubrac, A., Morel, B., Lamaa, A., Touriol, C., Teulade-Fichou, M. P., . . . Millevoi, S. (2015). Stabilization of the G-quadruplex at the VEGF IRES represses cap-independent translation. *RNA Biol*, 12(3), 320-329. doi: 10.1080/15476286.2015.1017236
- Capra, J. A., Paeschke, K., Singh, M., & Zakian, V. A. (2010). G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Comput Biol*, 6(7), e1000861. doi: 10.1371/journal.pcbi.1000861
- Carja, O., Xing, T., Plotkin, J. B., & Shah, P. (2017). riboviz: analysis and visualization of ribosome profiling datasets. *bioRxiv*.
- Casci, T., Vinos, J., & Freeman, M. (1999). Sprouty, an intracellular inhibitor of Ras signaling. *Cell*, 96(5), 655-665.
- Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P., & Balasubramanian, S. (2015). High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol*, 33(8), 877-881. doi: 10.1038/nbt.3295
- Chappell, S. A., Edelman, G. M., & Mauro, V. P. (2004). Biochemical and functional analysis of a 9-nt RNA sequence that affects translation efficiency in eukaryotic cells. *Proc Natl Acad Sci U S A*, 101(26), 9590-9594. doi: 10.1073/pnas.0308759101
- Chattopadhyay, M. K., Tabor, C. W., & Tabor, H. (2002). Absolute requirement of spermidine for growth and cell cycle progression of fission yeast (*Schizosaccharomyces pombe*). *Proc Natl Acad Sci U S A*, 99(16), 10330-10334. doi: 10.1073/pnas.162362899
- Cherry, J. M., Ball, C., Weng, S., Juvik, G., Schmidt, R., Adler, C., . . . Botstein, D. (1997). Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature*, 387(6632 Suppl), 67-73.
- Childs, A. C., Mehta, D. J., & Gerner, E. W. (2003). Polyamine-dependent gene

- expression. *Cell Mol Life Sci*, 60(7), 1394-1406. doi: 10.1007/s00018-003-2332-4
- Chun, S. Y., Rodriguez, C. M., Todd, P. K., & Mills, R. E. (2016). SPECtre: a spectral coherence--based classifier of actively translated transcripts from ribosome profiling sequence data. *BMC Bioinformatics*, 17(1), 482. doi: 10.1186/s12859-016-1355-4
- Chung, B. Y., Hardcastle, T. J., Jones, J. D., Irigoyen, N., Firth, A. E., Baulcombe, D. C., & 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), 1731-1745. doi: 10.1261/rna.052548.115
- Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M. P., & Phan, A. T. (2014). Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC(3). *Angew Chem Int Ed Engl*, 53(4), 999-1002. doi: 10.1002/anie.201308063
- Ciccodicola, A., D'Esposito, M., Esposito, T., Gianfrancesco, F., Migliaccio, C., Miano, M. G., . . . D'Urso, M. (2000). Differentially regulated and evolved genes in the fully sequenced Xq/Yq pseudoautosomal region. *Hum Mol Genet*, 9(3), 395-401.
- Coordinators, N. R. (2017). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res*, 45(D1), D12-D17. doi: 10.1093/nar/gkw1071
- Crappe, J., Ndah, E., Koch, A., Steyaert, S., Gawron, D., De Keulenaer, S., . . . Menschaert, G. (2015). PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration. *Nucleic Acids Res*, 43(5), e29. doi: 10.1093/nar/gku1283
- Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res*, 14(6), 1188-1190. doi: 10.1101/gr.849004
- Cunningham-Rundles, S., & Maas, W. K. (1975). Isolation, characterization, and mapping of *Escherichia coli* mutants blocked in the synthesis of ornithine decarboxylase. *J Bacteriol*, 124(2), 791-799.
- D'Esposito, M., Matarazzo, M. R., Ciccodicola, A., Strazzullo, M., Mazzarella, R., Quaderi, N. A., . . . D'Urso, M. (1997). Differential expression pattern of XqPAR-linked genes SYBL1 and IL9R correlates with the structure and

- evolution of the region. *Hum Mol Genet*, 6(11), 1917-1923.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Genomes Project Analysis, G. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158. doi: 10.1093/bioinformatics/btr330
- Davis, J. T. (2004). G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew Chem Int Ed Engl*, 43(6), 668-698. doi: 10.1002/anie.200300589
- De Cian, A., Delemos, E., Mergny, J. L., Teulade-Fichou, M. P., & Monchaud, D. (2007). Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J Am Chem Soc*, 129(7), 1856-1857. doi: 10.1021/ja067352b
- de Maximy, A. A., Nakatake, Y., Moncada, S., Itoh, N., Thiery, J. P., & Bellusci, S. (1999). Cloning and expression pattern of a mouse homologue of drosophila sprouty in the mouse embryo. *Mech Dev*, 81(1-2), 213-216.
- Doma, M. K., & Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*, 440(7083), 561-564. doi: 10.1038/nature04530
- Dunn, J. G., & Weissman, J. S. (2016). Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. *BMC Genomics*, 17(1), 958. doi: 10.1186/s12864-016-3278-x
- Edwin, F., Anderson, K., Ying, C., & Patel, T. B. (2009). Intermolecular interactions of Sprouty proteins and their implications in development and disease. *Mol Pharmacol*, 76(4), 679-691. doi: 10.1124/mol.109.055848
- Endoh, T., Kawasaki, Y., & Sugimoto, N. (2013). Translational halt during elongation caused by G-quadruplex formed by mRNA. *Methods*, 64(1), 73-78. doi: 10.1016/j.ymeth.2013.05.026
- Fang, P., Spevak, C. C., Wu, C., & Sachs, M. S. (2004). A nascent polypeptide domain that can regulate translation elongation. *Proc Natl Acad Sci U S A*, 101(12), 4059-4064. doi: 10.1073/pnas.0400554101
- Fay, M. M., Lyons, S. M., & Ivanov, P. (2017). RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J Mol Biol*, 429(14), 2127-2147. doi: 10.1016/j.jmb.2017.05.017
- Fields, A. P., Rodriguez, E. H., Jovanovic, M., Stern-Ginossar, N., Haas, B. J., Mertins, P., . . . Weissman, J. S. (2015). A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian

- Translation. *Mol Cell*, 60(5), 816-827. doi: 10.1016/j.molcel.2015.11.013
- Filbin, M. E., & Kieft, J. S. (2009). Toward a structural understanding of IRES RNA function. *Curr Opin Struct Biol*, 19(3), 267-276. doi: 10.1016/j.sbi.2009.03.005
- Fratta, P., Mizielinska, S., Nicoll, A. J., Zloh, M., Fisher, E. M., Parkinson, G., & Isaacs, A. M. (2012). C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep*, 2, 1016. doi: 10.1038/srep01016
- Fredlund, J. O., Johansson, M. C., Dahlberg, E., & Oredsson, S. M. (1995). Ornithine decarboxylase and S-adenosylmethionine decarboxylase expression during the cell cycle of Chinese hamster ovary cells. *Exp Cell Res*, 216(1), 86-92. doi: 10.1006/excr.1995.1011
- Frees, S., Menendez, C., Crum, M., & Bagga, P. S. (2014). QGRS-Conserve: a computational method for discovering evolutionarily conserved G-quadruplex motifs. *Hum Genomics*, 8, 8. doi: 10.1186/1479-7364-8-8
- Garg, R., Aggarwal, J., & Thakkar, B. (2016). Genome-wide discovery of G-quadruplex forming sequences and their functional relevance in plants. *Sci Rep*, 6, 28211. doi: 10.1038/srep28211
- Gebauer, F., & Hentze, M. W. (2004). Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol*, 5(10), 827-835. doi: 10.1038/nrm1488
- Gellert, M., Lipsett, M. N., & Davies, D. R. (1962). Helix formation by guanylic acid. *Proc Natl Acad Sci U S A*, 48, 2013-2018.
- Gerashchenko, M. V., & Gladyshev, V. N. (2017). Ribonuclease selection for ribosome profiling. *Nucleic Acids Res*, 45(2), e6. doi: 10.1093/nar/gkw822
- Gomez, D., Guedin, A., Mergny, J. L., Salles, B., Riou, J. F., Teulade-Fichou, M. P., & Calsou, P. (2010). A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res*, 38(20), 7187-7198. doi: 10.1093/nar/gkq563
- Green, K. M., Linsalata, A. E., & Todd, P. K. (2016). RAN translation-What makes it run? *Brain Res*, 1647, 30-42. doi: 10.1016/j.brainres.2016.04.003
- Grover, R., Candeias, M. M., Fahraeus, R., & Das, S. (2009). p53 and little brother p53/47: linking IRES activities with protein functions. *Oncogene*, 28(30), 2766-2772. doi: 10.1038/onc.2009.138
- Guiset Miserachs, H., Donghi, D., Borner, R., Johannsen, S., & Sigel, R. K. (2016).

- Distinct differences in metal ion specificity of RNA and DNA G-quadruplexes. *J Biol Inorg Chem*, 21(8), 975-986. doi: 10.1007/s00775-016-1393-4
- Guo, J. U., & Bartel, D. P. (2016). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, 353(6306). doi: 10.1126/science.aaf5371
- Halder, R., Riou, J. F., Teulade-Fichou, M. P., Frickey, T., & Hartig, J. S. (2012). Bisquinolinium compounds induce quadruplex-specific transcriptome changes in HeLa S3 cell lines. *BMC Res Notes*, 5, 138. doi: 10.1186/1756-0500-5-138
- Hardin, C. C., Watson, T., Corregan, M., & Bailey, C. (1992). Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG). *Biochemistry*, 31(3), 833-841.
- Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., . . . Hubbard, T. J. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*, 22(9), 1760-1774. doi: 10.1101/gr.135350.111
- Helena Mangs, A., & Morris, B. J. (2007). The Human Pseudoautosomal Region (PAR): Origin, Function and Future. *Curr Genomics*, 8(2), 129-136.
- Hellen, C. U., & Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev*, 15(13), 1593-1612. doi: 10.1101/gad.891101
- Hershman, S. G., Chen, Q., Lee, J. Y., Kozak, M. L., Yue, P., Wang, L. S., & Johnson, F. B. (2008). Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res*, 36(1), 144-156. doi: 10.1093/nar/gkm986
- Hinnebusch, A. G., Ivanov, I. P., & Sonenberg, N. (2016). Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science*, 352(6292), 1413-1416. doi: 10.1126/science.aad9868
- Howe, E. A., Sinha, R., Schlauch, D., & Quackenbush, J. (2011). RNA-Seq analysis in MeV. *Bioinformatics*, 27(22), 3209-3210. doi: 10.1093/bioinformatics/btr490
- Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1), 44-57. doi: 10.1038/nprot.2008.211
- Huppert, J. L. (2010). Structure, location and interactions of G-quadruplexes. *FEBS J*, 277(17), 3452-3458. doi: 10.1111/j.1742-4658.2010.07758.x

- Huppert, J. L., & Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic Acids Res*, 33(9), 2908-2916. doi: 10.1093/nar/gki609
- Huppert, J. L., & Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*, 35(2), 406-413. doi: 10.1093/nar/gkl1057
- Huppert, J. L., Bugaut, A., Kumari, S., & Balasubramanian, S. (2008). G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res*, 36(19), 6260-6268. doi: 10.1093/nar/gkn511
- Hwang, H., Buncher, N., Opresko, P. L., & Myong, S. (2012). POT1-TPP1 regulates telomeric overhang structural dynamics. *Structure*, 20(11), 1872-1880. doi: 10.1016/j.str.2012.08.018
- Igarashi, K., Hikami, K., Sugawara, K., & Hirose, S. (1973). Effect of polyamines on polypeptide synthesis in rat liver cell-free system. *Biochim Biophys Acta*, 299(2), 325-330.
- Igarashi, K., Watanabe, Y., & Hirose, S. (1975). Dependency of spermidine stimulation of polypeptide synthesis on the uracil content of messenger ribonucleic acid. *Biochem Biophys Res Commun*, 67(1), 407-413
- Igarashi, K., & Kashiwagi, K. (2015). Modulation of protein synthesis by polyamines. *IUBMB Life*, 67(3), 160-169. doi: 10.1002/iub.1363
- Ingolia, N. T. (2016). Ribosome Footprint Profiling of Translation throughout the Genome. *Cell*, 165(1), 22-33. doi: 10.1016/j.cell.2016.02.066
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 324(5924), 218-223. doi: 10.1126/science.1168978
- Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147(4), 789-802. doi: 10.1016/j.cell.2011.10.002
- Ito, K., Kashiwagi, K., Watanabe, S., Kameji, T., Hayashi, S., & Igarashi, K. (1990). Influence of the 5'-untranslated region of ornithine decarboxylase mRNA and spermidine on ornithine decarboxylase synthesis. *J Biol Chem*, 265(22), 13036-13041
- Iwasaki, S., Floor, S. N., & Ingolia, N. T. (2016). Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor. *Nature*,

- 534(7608), 558-561. doi: 10.1038/nature17978
- Izbicka, E., Wheelhouse, R. T., Raymond, E., Davidson, K. K., Lawrence, R. A., Sun, D., . . . Von Hoff, D. D. (1999). Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Res*, 59(3), 639-644.
- Jackson, R. J., Hellen, C. U., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*, 11(2), 113-127. doi: 10.1038/nrm2838
- Jain, A., & Tyagi, A. K. (1987). Role of polyamines in the synthesis of RNA in mycobacteria. *Mol Cell Biochem*, 78(1), 3-8.
- Jang, S. K., Krausslich, H. G., Nicklin, M. J., Duke, G. M., Palmenberg, A. C., & Wimmer, E. (1988). A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol*, 62(8), 2636-2643.
- Jodoin, R., Bauer, L., Garant, J. M., Mahdi Laaref, A., Phaneuf, F., & Perreault, J. P. (2014). The folding of 5'-UTR human G-quadruplexes possessing a long central loop. *RNA*, 20(7), 1129-1141. doi: 10.1261/rna.044578.114
- Johnson, A. G., Grosely, R., Petrov, A. N., & Puglisi, J. D. (2017). Dynamics of IRES-mediated translation. *Philos Trans R Soc Lond B Biol Sci*, 372(1716). doi: 10.1098/rstb.2016.0177
- Kaplan, O. I., Berber, B., Hekim, N., & Doluca, O. (2016). G-quadruplex prediction in E. coli genome reveals a conserved putative G-quadruplex-Hairpin-Duplex switch. *Nucleic Acids Res*, 44(19), 9083-9095. doi: 10.1093/nar/gkw769
- Kearse, M. G., Green, K. M., Krans, A., Rodriguez, C. M., Linsalata, A. E., Goldstrohm, A. C., & Todd, P. K. (2016). CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell*, 62(2), 314-322. doi: 10.1016/j.molcel.2016.02.034
- Komar, A. A., & Hatzoglou, M. (2011). Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. *Cell Cycle*, 10(2), 229-240. doi: 10.4161/cc.10.2.14472
- Komar, A. A., Lesnik, T., Cullin, C., Merrick, W. C., Trachsel, H., & Altmann, M. (2003). Internal initiation drives the synthesis of Ure2 protein lacking the prion domain and affects [URE3] propagation in yeast cells. *EMBO J*, 22(5), 1199-1209. doi: 10.1093/emboj/cdg103

- Kozak, M. (1990). Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci U S A*, 87(21), 8301-8305.
- Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*, 324(5924), 255-258. doi: 10.1126/science.1170160
- Kumar, N., Basundra, R., & Maiti, S. (2009). Elevated polyamines induce c-MYC overexpression by perturbing quadruplex-WC duplex equilibrium. *Nucleic Acids Res*, 37(10), 3321-3331. doi: 10.1093/nar/gkp196
- Kumari, S., Bugaut, A., Huppert, J. L., & Balasubramanian, S. (2007). An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol*, 3(4), 218-221. doi: 10.1038/nchembio864
- Kusano, T., Berberich, T., Tateda, C., & Takahashi, Y. (2008). Polyamines: essential factors for growth and survival. *Planta*, 228(3), 367-381. doi: 10.1007/s00425-008-0772-7
- Kwok, C. K., & Balasubramanian, S. (2015). Targeted Detection of G-Quadruplexes in Cellular RNAs. *Angew Chem Int Ed Engl*, 54(23), 6751-6754. doi: 10.1002/anie.201500891
- Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S., & Balasubramanian, S. (2016). rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Methods*, 13(10), 841-844. doi: 10.1038/nmeth.3965
- Lammich, S., Kamp, F., Wagner, J., Nuscher, B., Zilow, S., Ludwig, A. K., . . . Haass, C. (2011). Translational repression of the disintegrin and metalloprotease ADAM10 by a stable G-quadruplex secondary structure in its 5'-untranslated region. *J Biol Chem*, 286(52), 45063-45072. doi: 10.1074/jbc.M111.296921
- Landau, G., Bercovich, Z., Park, M. H., & Kahana, C. (2010). The role of polyamines in supporting growth of mammalian cells is mediated through their requirement for translation initiation and elongation. *J Biol Chem*, 285(17), 12474-12481. doi: 10.1074/jbc.M110.106419
- Lane, A. N., Chaires, J. B., Gray, R. D., & Trent, J. O. (2008). Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res*, 36(17), 5482-5515. doi: 10.1093/nar/gkn517
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2.

- Nat Methods*, 9(4), 357-359. doi: 10.1038/nmeth.1923
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10(3), R25. doi: 10.1186/gb-2009-10-3-r25
- Lareau, L. F., Hite, D. H., Hogan, G. J., & Brown, P. O. (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife*, 3, e01257. doi: 10.7554/eLife.01257
- Lauria, F., Tebaldi, T., Bernabo, P., Groen, E. J. N., Gillingwater, T. H., & Viero, G. (2017). riboWaltz: optimization of ribosome P-site positioning in ribosome profiling data. *bioRxiv*.
- Law, G. L., Raney, A., Heusner, C., & Morris, D. R. (2001). Polyamine regulation of ribosome pausing at the upstream open reading frame of S-adenosylmethionine decarboxylase. *J Biol Chem*, 276(41), 38036-38043. doi: 10.1074/jbc.M105944200
- Le, D. D., Di Antonio, M., Chan, L. K., & Balasubramanian, S. (2015). G-quadruplex ligands exhibit differential G-tetrad selectivity. *Chem Commun (Camb)*, 51(38), 8048-8050. doi: 10.1039/c5cc02252e
- Leeksa, O. C., Van Achterberg, T. A., Tsumura, Y., Toshima, J., Eldering, E., Kroes, W. G., . . . de Vries, C. J. (2002). Human sprouty 4, a new ras antagonist on 5q31, interacts with the dual specificity kinase TESK1. *Eur J Biochem*, 269(10), 2546-2556.
- Legendre, R., Baudin-Baillieu, A., Hatin, I., & Namy, O. (2015). RiboTools: a Galaxy toolbox for qualitative ribosome profiling analysis. *Bioinformatics*, 31(15), 2586-2588. doi: 10.1093/bioinformatics/btv174
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079. doi: 10.1093/bioinformatics/btp352
- Li, G. W., Oh, E., & Weissman, J. S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, 484(7395), 538-541. doi: 10.1038/nature10965
- Li, Q., Xiang, J., Li, X., Chen, L., Xu, X., Tang, Y., . . . Xu, G. (2009). Stabilizing parallel G-quadruplex DNA by a new class of ligands: two non-planar alkaloids through interaction in lateral grooves. *Biochimie*, 91(7), 811-819.

doi: 10.1016/j.biochi.2009.03.007

- Li, Q., Xiang, J. F., Yang, Q. F., Sun, H. X., Guan, A. J., & Tang, Y. L. (2013). G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res*, 41(Database issue), D1115-1123. doi: 10.1093/nar/gks1101
- Lin, W., Sampathi, S., Dai, H., Liu, C., Zhou, M., Hu, J., . . . Shen, B. (2013). Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *EMBO J*, 32(10), 1425-1439. doi: 10.1038/emboj.2013.88
- Malgowska, M., Czajczynska, K., Gudanis, D., Tworak, A., & Gdaniec, Z. (2016). Overview of the RNA G-quadruplex structures. *Acta Biochim Pol*, 63(4), 609-621. doi: 10.18388/abp.2016_1335
- Malone, B., Atanassov, I., Aeschmann, F., Li, X., Grosshans, H., & Dieterich, C. (2017). Bayesian prediction of RNA translation from ribosome profiling. *Nucleic Acids Res*, 45(6), 2960-2972. doi: 10.1093/nar/gkw1350
- Mandal, S., Mandal, A., Johansson, H. E., Orjalo, A. V., & Park, M. H. (2013). Depletion of cellular polyamines, spermidine and spermine, causes a total arrest in translation and growth in mammalian cells. *Proc Natl Acad Sci U S A*, 110(6), 2169-2174. doi: 10.1073/pnas.1219002110
- Martin, M. (2011). Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet.journal*, 17(1), 10-12.
- Matsuo, Y., Ikeuchi, K., Saeki, Y., Iwasaki, S., Schmidt, C., Udagawa, T., . . . Inada, T. (2017). Ubiquitination of stalled ribosome triggers ribosome-associated quality control. *Nat Commun*, 8(1), 159. doi: 10.1038/s41467-017-00188-1
- Mendoza, O., Bourdoncle, A., Boule, J. B., Brosh, R. M., Jr., & Mergny, J. L. (2016). G-quadruplexes and helicases. *Nucleic Acids Res*, 44(5), 1989-2006. doi: 10.1093/nar/gkw079
- Meng, Z., Jackson, N. L., Shcherbakov, O. D., Choi, H., & Blume, S. W. (2010). The human IGF1R IRES likely operates through a Shine-Dalgarno-like interaction with the G961 loop (E-site) of the 18S rRNA and is kinetically modulated by a naturally polymorphic polyU loop. *J Cell Biochem*, 110(2), 531-544. doi: 10.1002/jcb.22569
- Merrick, W. C. (2004). Cap-dependent and cap-independent translation in eukaryotic systems. *Gene*, 332, 1-11. doi: 10.1016/j.gene.2004.02.051

- Michel, A. M., Ahern, A. M., Donohue, C. A., & Baranov, P. V. (2015). GWIPS-viz as a tool for exploring ribosome profiling evidence supporting the synthesis of alternative proteoforms. *Proteomics*, 15(14), 2410-2416. doi: 10.1002/pmic.201400603
- Michel, A. M., & Baranov, P. V. (2013). Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley Interdiscip Rev RNA*, 4(5), 473-490. doi: 10.1002/wrna.1172
- Michel, A. M., Choudhury, K. R., Firth, A. E., Ingolia, N. T., Atkins, J. F., & Baranov, P. V. (2012). Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res*, 22(11), 2219-2229. doi: 10.1101/gr.133249.111
- Michel, A. M., Fox, G., A, M. K., De Bo, C., O'Connor, P. B., Heaphy, S. M., . . . Baranov, P. V. (2014). GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res*, 42(Database issue), D859-864. doi: 10.1093/nar/gkt1035
- Michel, A. M., Mullan, J. P., Velayudhan, V., O'Connor, P. B., Donohue, C. A., & Baranov, P. V. (2016). RiboGalaxy: A browser based platform for the alignment, analysis and visualization of ribosome profiling data. *RNA Biol*, 13(3), 316-319. doi: 10.1080/15476286.2016.1141862
- Miller-Fleming, L., Olin-Sandoval, V., Campbell, K., & Ralser, M. (2015). Remaining Mysteries of Molecular Biology: The Role of Polyamines in the Cell. *J Mol Biol*, 427(21), 3389-3406. doi: 10.1016/j.jmb.2015.06.020
- Mitchell, J. L., Judd, G. G., Leyser, A., & Choe, C. (1998). Osmotic stress induces variation in cellular levels of ornithine decarboxylase-antizyme. *Biochem J*, 329 (Pt 3), 453-459.
- Mohammad, F., Woolstenhulme, C. J., Green, R., & Buskirk, A. R. (2016). Clarifying the Translational Pausing Landscape in Bacteria by Ribosome Profiling. *Cell Rep*, 14(4), 686-694. doi: 10.1016/j.celrep.2015.12.073
- Mokrejs, M., Masek, T., Vopalensky, V., Hlubucek, P., Delbos, P., & Pospisek, M. (2010). IRESite--a tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Res*, 38(Database issue), D131-136. doi: 10.1093/nar/gkp981
- Morris, M. J., & Basu, S. (2009). An unusually stable G-quadruplex within the 5'-UTR of the MT3 matrix metalloproteinase mRNA represses translation in

- eukaryotic cells. *Biochemistry*, 48(23), 5313-5319. doi: 10.1021/bi900498z
- Morris, M. J., Negishi, Y., Pazsint, C., Schonhoft, J. D., & Basu, S. (2010). An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *J Am Chem Soc*, 132(50), 17831-17839. doi: 10.1021/ja106287x
- Mullen, M. A., Olson, K. J., Dallaire, P., Major, F., Assmann, S. M., & Bevilacqua, P. C. (2010). RNA G-Quadruplexes in the model plant species *Arabidopsis thaliana*: prevalence and possible functional roles. *Nucleic Acids Res*, 38(22), 8149-8163. doi: 10.1093/nar/gkq804
- Murat, P., & Balasubramanian, S. (2014). Existence and consequences of G-quadruplex structures in DNA. *Curr Opin Genet Dev*, 25, 22-29. doi: 10.1016/j.gde.2013.10.012
- Ndah, E., Jonckheere, V., Giess, A., Valen, E., Menschaert, G., & Van Damme, P. (2017). REPARATION: Ribosome Profiling Assisted (Re-)Annotation of Bacterial genomes. *bioRxiv*.
- Nelms, B. L., & Labosky, P. A. (2011). A predicted hairpin cluster correlates with barriers to PCR, sequencing and possibly BAC recombineering. *Sci Rep*, 1, 106. doi: 10.1038/srep00106
- Ning, Z., McLellan, A. S., Ball, M., Wynne, F., O'Neill, C., Mills, W., . . . Moore, T. (2015). Regulation of SPRY3 by X chromosome and PAR2-linked promoters in an autism susceptibility region. *Hum Mol Genet*, 24(18), 5126-5141. doi: 10.1093/hmg/ddv231
- O'Connor, P. B., Li, G. W., Weissman, J. S., Atkins, J. F., & Baranov, P. V. (2013). rRNA:mRNA pairing alters the length and the symmetry of mRNA-protected fragments in ribosome profiling experiments. *Bioinformatics*, 29(12), 1488-1491. doi:10.1093/bioinformatics/btt184
- O'Connor, P. B., Andreev, D. E., & Baranov, P. V. (2016). Comparative survey of the relative impact of mRNA features on local ribosome profiling read density. *Nat Commun*, 7, 12915. doi: 10.1038/ncomms12915
- Ogloblina, A. M., Bannikova, V. A., Khristich, A. N., Oretskaya, T. S., Yakubovskaya, M. G., & Dolinnaya, N. G. (2015). Parallel G-Quadruplexes Formed by Guanine-Rich Microsatellite Repeats Inhibit Human Topoisomerase I. *Biochemistry (Mosc)*, 80(8), 1026-1038. doi: 10.1134/S0006297915080088
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., . . .

- Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, 44(D1), D733-745. doi: 10.1093/nar/gkv1189
- Oertlin, C., Lorent, J., Gandin, V., Murie, C., Masvidal, L., Cargnello, M., . . . Larsson, O. (2017). Genome-wide analysis of differential translation and differential translational buffering using anota2seq. *bioRxiv*.
- Oh, E., Becker, A. H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., . . . Bukau, B. (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell*, 147(6), 1295-1308. doi: 10.1016/j.cell.2011.10.044
- Olexiouk, V., Crappe, J., Verbruggen, S., Verhegen, K., Martens, L., & Menschaert, G. (2016). sORFs.org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res*, 44(D1), D324-329. doi: 10.1093/nar/gkv1175
- Olshen, A. B., Hsieh, A. C., Stumpf, C. R., Olshen, R. A., Ruggero, D., & Taylor, B. S. (2013). Assessing gene-level translational control from ribosome profiling. *Bioinformatics*, 29(23), 2995-3002. doi: 10.1093/bioinformatics/btt533
- Palanimurugan, R., Scheel, H., Hofmann, K., & Dohmen, R. J. (2004). Polyamines regulate their synthesis by inducing expression and blocking degradation of ODC antizyme. *EMBO J*, 23(24), 4857-4867. doi: 10.1038/sj.emboj.7600473
- Panagiotaki, N., Dajas-Bailador, F., Amaya, E., Papalopulu, N., & Dorey, K. (2010). Characterisation of a new regulator of BDNF signalling, Sprouty3, involved in axonal morphogenesis in vivo. *Development*, 137(23), 4005-4015. doi: 10.1242/dev.053173
- Parikshak, N. N., Swarup, V., Belgard, T. G., Irimia, M., Ramaswami, G., Gandal, M. J., . . . Geschwind, D. H. (2016). Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*, 540(7633), 423-427. doi: 10.1038/nature20612
- Park, E. H., Lee, J. M., Blais, J. D., Bell, J. C., & Pelletier, J. (2005). Internal translation initiation mediated by the angiogenic factor Tie2. *J Biol Chem*, 280(22), 20945-20953. doi: 10.1074/jbc.M412744200
- Pearson, C. E. (2011). Repeat associated non-ATG translation initiation: one DNA, two transcripts, seven reading frames, potentially nine toxic entities! *PLoS Genet*, 7(3), e1002018. doi: 10.1371/journal.pgen.1002018

- Pegg, A. E., & McCann, P. P. (1982). Polyamine metabolism and function. *Am J Physiol*, 243(5), C212-221. doi: 10.1152/ajpcell.1982.243.5.C212
- Pelletier, J., & Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature*, 334(6180), 320-325. doi: 10.1038/334320a0
- Pillai, A. (2008). Decreased expression of Sprouty2 in the dorsolateral prefrontal cortex in schizophrenia and bipolar disorder: a correlation with BDNF expression. *PLoS One*, 3(3), e1784. doi: 10.1371/journal.pone.0001784
- Pinkstaff, J. K., Chappell, S. A., Mauro, V. P., Edelman, G. M., & Krushel, L. A. (2001). Internal initiation of translation of five dendritically localized neuronal mRNAs. *Proc Natl Acad Sci U S A*, 98(5), 2770-2775. doi: 10.1073/pnas.051623398
- Pisarev, A. V., Hellen, C. U., & Pestova, T. V. (2007). Recycling of eukaryotic posttermination ribosomal complexes. *Cell*, 131(2), 286-299. doi: 10.1016/j.cell.2007.08.041
- Purcell, P., Joo, B. W., Hu, J. K., Tran, P. V., Calicchio, M. L., O'Connell, D. J., . . . Tabin, C. J. (2009). Temporomandibular joint formation requires two distinct hedgehog-dependent steps. *Proc Natl Acad Sci U S A*, 106(43), 18297-18302. doi: 10.1073/pnas.0908836106
- Raj, A., Wang, S. H., Shim, H., Harpak, A., Li, Y. I., Engelmann, B., . . . Pritchard, J. K. (2016). Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. *Elife*, 5. doi: 10.7554/eLife.13328
- Reddy, K., Zamiri, B., Stanley, S. Y., Macgregor, R. B., Jr., & Pearson, C. E. (2013). The disease-associated r(GGGGCC)_n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *J Biol Chem*, 288(14), 9860-9866. doi: 10.1074/jbc.C113.452532
- Rhodes, D., & Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res*, 43(18), 8627-8637. doi: 10.1093/nar/gkv862
- Ribeyre, C., Lopes, J., Boule, J. B., Piazza, A., Guedin, A., Zakian, V. A., . . . Nicolas, A. (2009). The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS Genet*, 5(5), e1000475. doi: 10.1371/journal.pgen.1000475
- Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*, 16(6), 276-277.

- Richter, J. D., & Collier, J. (2015). Pausing on Polyribosomes: Make Way for Elongation in Translational Control. *Cell*, 163(2), 292-300. doi: 10.1016/j.cell.2015.09.041
- Rodnina, M. V., & Wintermeyer, W. (2009). Recent mechanistic insights into eukaryotic ribosomes. *Curr Opin Cell Biol*, 21(3), 435-443. doi: 10.1016/j.ceb.2009.01.023
- Rodriguez, R., Muller, S., Yeoman, J. A., Trentesaux, C., Riou, J. F., & Balasubramanian, S. (2008). A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J Am Chem Soc*, 130(47), 15758-15759. doi: 10.1021/ja805615w
- Rooijers, K., Loayza-Puch, F., Nijtmans, L. G., & Agami, R. (2013). Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. *Nat Commun*, 4, 2886. doi: 10.1038/ncomms3886
- Ruan, H., Shantz, L. M., Pegg, A. E., & Morris, D. R. (1996). The upstream open reading frame of the mRNA encoding S-adenosylmethionine decarboxylase is a polyamine-responsive translational control element. *J Biol Chem*, 271(47), 29576-29582.
- Sabi, R., & Tuller, T. (2017). Computational analysis of nascent peptides that induce ribosome stalling and their proteomic distribution in *Saccharomyces cerevisiae*. *RNA*, 23(7), 983-994. doi: 10.1261/rna.059188.116
- Sayed, D., Rane, S., Lypowy, J., He, M., Chen, I. Y., Vashistha, H., . . . Abdellatif, M. (2008). MicroRNA-21 targets Sprouty2 and promotes cellular outgrowths. *Mol Biol Cell*, 19(8), 3272-3282. doi: 10.1091/mbc.E08-02-0159
- Schaeffer, C., Bardoni, B., Mandel, J. L., Ehresmann, B., Ehresmann, C., & Moine, H. (2001). The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J*, 20(17), 4803-4813. doi: 10.1093/emboj/20.17.4803
- Sen, N. D., Zhou, F., Ingolia, N. T., & Hinnebusch, A. G. (2015). Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. *Genome Res*, 25(8), 1196-1205. doi: 10.1101/gr.191601.115
- Shahid, R., Bugaut, A., & Balasubramanian, S. (2010). The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry*, 49(38), 8300-8306. doi: 10.1021/bi100957h

- Shi, D. F., Wheelhouse, R. T., Sun, D., & Hurley, L. H. (2001). Quadruplex-interactive agents as telomerase inhibitors: synthesis of porphyrins and structure-activity relationship for the inhibition of telomerase. *J Med Chem*, 44(26), 4509-4523.
- Shim, K., Minowada, G., Coling, D. E., & Martin, G. R. (2005). Sprouty2, a mouse deafness gene, regulates cell fate decisions in the auditory sensory epithelium by antagonizing FGF signaling. *Dev Cell*, 8(4), 553-564. doi: 10.1016/j.devcel.2005.02.009
- Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., & Hurley, L. H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A*, 99(18), 11593-11598. doi: 10.1073/pnas.182256799
- Somogyi, P., Jenner, A. J., Brierley, I., & Inglis, S. C. (1993). Ribosomal pausing during translation of an RNA pseudoknot. *Mol Cell Biol*, 13(11), 6931-6940.
- Sonenberg, N., & Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, 136(4), 731-745. doi: 10.1016/j.cell.2009.01.042
- Song, J., Perreault, J. P., Topisirovic, I., & Richard, S. (2016). RNA G-quadruplexes and their potential regulatory roles in translation. *Translation (Austin)*, 4(2), e1244031. doi: 10.1080/21690731.2016.1244031
- Spriggs, K. A., Cobbold, L. C., Jopling, C. L., Cooper, R. E., Wilson, L. A., Stoneley, M., . . . Willis, A. E. (2009). Canonical initiation factor requirements of the Myc family of internal ribosome entry segments. *Mol Cell Biol*, 29(6), 1565-1574. doi: 10.1128/MCB.01283-08
- Stephen Neidle, S. B. (2006). Quadruplex Nucleic Acids. *Royal Society of Chemistry*, i-302. doi: 10.1039/9781847555298
- Stoneley, M., & Willis, A. E. (2004). Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene*, 23(18), 3200-3207. doi: 10.1038/sj.onc.1207551
- Svitkin, Y. V., Pause, A., Haghighat, A., Pyronnet, S., Witherell, G., Belsham, G. J., & Sonenberg, N. (2001). The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA*, 7(3), 382-394.
- Taketomi, T., Yoshiga, D., Taniguchi, K., Kobayashi, T., Nonami, A., Kato, R., . . . Yoshimura, A. (2005). Loss of mammalian Sprouty2 leads to enteric neuronal

- hyperplasia and esophageal achalasia. *Nat Neurosci*, 8(7), 855-857. doi: 10.1038/nn1485
- Tanaka, M., Sotta, N., Yamazumi, Y., Yamashita, Y., Miwa, K., Murota, K., . . . Fujiwara, T. (2016). The Minimum Open Reading Frame, AUG-Stop, Induces Boron-Dependent Ribosome Stalling and mRNA Degradation. *Plant Cell*, 28(11), 2830-2849. doi: 10.1105/tpc.16.00481
- Tang, C. F., & Shafer, R. H. (2006). Engineering the quadruplex fold: nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. *J Am Chem Soc*, 128(17), 5966-5973. doi: 10.1021/ja0603958
- Taniguchi, K., Ayada, T., Ichiyama, K., Kohno, R., Yonemitsu, Y., Minami, Y., . . . Yoshimura, A. (2007). Sprouty2 and Sprouty4 are essential for embryonic morphogenesis and regulation of FGF signaling. *Biochem Biophys Res Commun*, 352(4), 896-902. doi: 10.1016/j.bbrc.2006.11.107
- Tate, W. P., & Brown, C. M. (1992). Translational termination: "stop" for protein synthesis or "pause" for regulation of gene expression. *Biochemistry*, 31(9), 2443-2450.
- Tenson, T., & Ehrenberg, M. (2002). Regulatory nascent peptides in the ribosomal tunnel. *Cell*, 108(5), 591-594.
- Tholstrup, J., Oddershede, L. B., & Sorensen, M. A. (2012). mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Res*, 40(1), 303-313. doi: 10.1093/nar/gkr686
- Thomas, T., Gallo, M. A., Klinge, C. M., & Thomas, T. J. (1995). Polyamine-mediated conformational perturbations in DNA alter the binding of estrogen receptor to poly(dG-m5dC).poly(dG-m5dC) and a plasmid containing the estrogen response element. *J Steroid Biochem Mol Biol*, 54(3-4), 89-99.
- Thompson, S. R. (2012). Tricks an IRES uses to enslave ribosomes. *Trends Microbiol*, 20(11), 558-566. doi: 10.1016/j.tim.2012.08.002
- Tsai, C. J., Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V., Gottesman, M. M., & Nussinov, R. (2008). Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J Mol Biol*, 383(2), 281-291. doi: 10.1016/j.jmb.2008.08.012
- Valdes-Santiago, L., & Ruiz-Herrera, J. (2013). Stress and polyamine metabolism in fungi. *Front Chem*, 1, 42. doi: 10.3389/fchem.2013.00042
- Verma, A., Halder, K., Halder, R., Yadav, V. K., Rawal, P., Thakur, R. K., . . .

- Chowdhury, S. (2008). Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *J Med Chem*, 51(18), 5641-5649. doi: 10.1021/jm800448a
- Wan, Y., Kertesz, M., Spitale, R. C., Segal, E., & Chang, H. Y. (2011). Understanding the transcriptome through RNA structure. *Nat Rev Genet*, 12(9), 641-655. doi: 10.1038/nrg3049
- Wan, J., & Qian, S. B. (2014). TISdb: a database for alternative translation initiation in mammalian cells. *Nucleic Acids Res*, 42(Database issue), D845-850. doi: 10.1093/nar/gkt1085
- Wang, Y., Zhao, M., Zhang, Q., Zhu, G. F., Li, F. F., & Du, L. F. (2015). Genomic distribution and possible functional roles of putative G-quadruplex motifs in two subspecies of *Oryza sativa*. *Comput Biol Chem*, 56, 122-130. doi: 10.1016/j.compbiolchem.2015.04.009
- Watson, M. B., & Malmberg, R. L. (1996). Regulation of *Arabidopsis thaliana* (L.) Heynh Arginine decarboxylase by potassium deficiency stress. *Plant Physiol*, 111(4), 1077-1083.
- Wolfe, A. L., Singh, K., Zhong, Y., Drewe, P., Rajasekhar, V. K., Sanghvi, V. R., . . . Wendel, H. G. (2014). RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature*, 513(7516), 65-70. doi: 10.1038/nature13485
- Wong, A., & Wu, G. (2003). Selective binding of monovalent cations to the stacking G-quartet structure formed by guanosine 5'-monophosphate: a solid-state NMR study. *J Am Chem Soc*, 125(45), 13895-13905. doi: 10.1021/ja0302174
- Woolstenhulme, C. J., Guydosh, N. R., Green, R., & Buskirk, A. R. (2015). High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Rep*, 11(1), 13-21. doi: 10.1016/j.celrep.2015.03.014
- Xiao, Z., Zou, Q., Liu, Y., & Yang, X. (2016). Genome-wide assessment of differential translations with ribosome profiling data. *Nat Commun*, 7, 11194. doi: 10.1038/ncomms11194
- Xie, S. Q., Nie, P., Wang, Y., Wang, H., Li, H., Yang, Z., . . . Xie, Z. (2016). RPFdb: a database for genome wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res*, 44(D1), D254-258. doi: 10.1093/nar/gkv972

- Yanagitani, K., Kimata, Y., Kadokura, H., & Kohno, K. (2011). Translational pausing ensures membrane targeting and cytoplasmic splicing of XBP1u mRNA. *Science*, 331(6017), 586-589. doi: 10.1126/science.1197142
- Yordanova, M. M., Wu, C., Andreev, D. E., Sachs, M. S., & Atkins, J. F. (2015). A Nascent Peptide Signal Responsive to Endogenous Levels of Polyamines Acts to Stimulate Regulatory Frameshifting on Antizyme mRNA. *J Biol Chem*, 290(29), 17863-17878. doi: 10.1074/jbc.M115.647065
- Yerlikaya, S., Meusbürger, M., Kumari, R., Huber, A., Anrather, D., Costanzo, M., . . . Loewith, R. (2016). TORC1 and TORC2 work together to regulate ribosomal protein S6 phosphorylation in *Saccharomyces cerevisiae*. *Mol Biol Cell*, 27(2), 397-409. doi: 10.1091/mbc.E15-08-0594
- Yu, C. H., Teulade-Fichou, M. P., & Olsthoorn, R. C. (2014). Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Res*, 42(3), 1887-1892. doi: 10.1093/nar/gkt1022
- Zamiri, B., Reddy, K., Macgregor, R. B., Jr., & Pearson, C. E. (2014). TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)_n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. *J Biol Chem*, 289(8), 4653-4659. doi: 10.1074/jbc.C113.502336
- Zhang, C., Chaturvedi, D., Jaggar, L., Magnuson, D., Lee, J. M., & Patel, T. B. (2005). Regulation of vascular smooth muscle cell proliferation and migration by human sprouty 2. *Arterioscler Thromb Vasc Biol*, 25(3), 533-538. doi: 10.1161/01.ATV.0000155461.50450.5a
- Zhang, D. H., Fujimoto, T., Saxena, S., Yu, H. Q., Miyoshi, D., & Sugimoto, N. (2010). Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry*, 49(21), 4554-4563. doi: 10.1021/bi1002822
- Zhong, Y., Karaletsos, T., Drewe, P., Sreedharan, V. T., Kuo, D., Singh, K., . . . Ratsch, G. (2017). RiboDiff: detecting changes of mRNA translation efficiency from ribosome footprints. *Bioinformatics*, 33(1), 139-141. doi: 10.1093/bioinformatics/btw585
- Zhu, X. J., Sun, S., Xie, B., Hu, X., Zhang, Z., Qiu, M., & Dai, Z. M. (2016). Guanine-rich sequences inhibit proofreading DNA polymerases. *Sci Rep*, 6, 28769. doi: 10.1038/srep28769

Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., . . .
Ranum, L. P. (2011). Non-ATG-initiated translation directed by microsatellite
expansions. *Proc Natl Acad Sci U S A*, 108(1), 260-265. doi:
10.1073/pnas.1013343108

Chapter 5

Appendices

Appendices discuss the collaborative projects completed during the course of this PhD thesis.

Appendix 1: Predicting the role of Rps6 phosphorylation in global translation using ribosome profiling data.

This appendix has been published as a research article in Molecular Biology of the Cell (2016) 27(2):397-409.

To study the role of Ribosomal protein S6 (Rsp6) in the global translation and individual mRNA translation, I analysed the ribosome profiling from two different mutants, $ypk1^{as} ypk2 ypk3^{as} RPS6^{wt}$ and $ypk1^{as} ypk2 ypk3^{as} RPS6^{AA}$ with or without 1NM-PP1. Specifically, the profiles of these cells after 20 and 40 min of refeeding with glucose were compared.

TORC1 and TORC2 work together to regulate ribosomal protein S6 phosphorylation in *Saccharomyces cerevisiae*

Seda Yerlikaya^{a,*}, Madeleine Meusburger^{a,*}, Romika Kumari^b, Alexandre Huber^a, Dorothea Anrather^c, Michael Costanzo^d, Charles Boone^d, Gustav Ammerer^c, Pavel V. Baranov^b, and Robbie Loewith^{a,e}

^aDepartment of Molecular Biology and Institute of Genetics and Genomics of Geneva, University of Geneva, CH-1211 Geneva, Switzerland; ^bSchool of Biochemistry and Cell Biology, University College Cork, Cork, Ireland; ^cMax F. Perutz Laboratories, Department of Biochemistry, University of Vienna, A1030 Vienna, Austria; ^dBanting and Best Department of Medical Research, Donnelly Center for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada; ^eSwiss National Centre for Competence in Research Programme Chemical Biology, 1211 Geneva, Switzerland

ABSTRACT Nutrient-sensitive phosphorylation of the S6 protein of the 40S subunit of the eukaryote ribosome is highly conserved. However, despite four decades of research, the functional consequences of this modification remain unknown. Revisiting this enigma in *Saccharomyces cerevisiae*, we found that the regulation of Rps6 phosphorylation on Ser-232 and Ser-233 is mediated by both TOR complex 1 (TORC1) and TORC2. TORC1 regulates phosphorylation of both sites via the poorly characterized AGC-family kinase Ypk3 and the PP1 phosphatase Glc7, whereas TORC2 regulates phosphorylation of only the N-terminal phosphosite via Ypk1. Cells expressing a nonphosphorylatable variant of Rps6 display a reduced growth rate and a 40S biogenesis defect, but these phenotypes are not observed in cells in which Rps6 kinase activity is compromised. Furthermore, using polysome profiling and ribosome profiling, we failed to uncover a role of Rps6 phosphorylation in either global translation or translation of individual mRNAs. Taking the results together, this work depicts the signaling cascades orchestrating Rps6 phosphorylation in budding yeast, challenges the notion that Rps6 phosphorylation plays a role in translation, and demonstrates that observations made with Rps6 knock-ins must be interpreted cautiously.

Monitoring Editor
Sandra Lemmon
University of Miami

Received: Aug 24, 2015
Revised: Oct 27, 2015
Accepted: Nov 9, 2015

INTRODUCTION

The target of rapamycin (TOR) signaling network plays an important role in the adaptation of cells to their changing environment by linking extracellular and cell-intrinsic cues to the major cellular processes that regulate growth. For example, TOR controls cell growth

by regulating gene transcription, translation, ribosome biogenesis, autophagy, lipid synthesis, and actin polarization (Loewith and Hall, 2011). In yeast, there are two TOR genes, *TOR1* and *TOR2*, whereas in higher eukaryotes there is only one. TOR kinases are found in two functionally and structurally distinct multiprotein complexes named TOR complex 1 (TORC1) and TOR complex 2 (TORC2). Rapamycin, an antifungal and immunosuppressant macrolide, inhibits TORC1 but not TORC2. When applied in low millimolar concentrations, caffeine also specifically inhibits TORC1 (Kuranda *et al.*, 2006; Reinke *et al.*, 2006; Wanke *et al.*, 2008; Loewith and Hall, 2011). Pharmacological inhibition of TORC2 in yeast was achieved only recently with the characterization of an imidazoquinoline called NVP-BHS345 (BHS), an ATP-competitive inhibitor of TOR that inhibits both complexes (Shimada *et al.*, 2013; Rispal *et al.*, 2015).

In yeast, as well as in higher eukaryotes, arguably the best-characterized substrates of TOR are members of the AGC family of

This article was published online ahead of print in MBcC in Press (<http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E15-08-0594>) on November 18, 2015.

*These authors contributed equally to this work.

Address correspondence to: Robbie Loewith (robbie.loewith@unige.ch).

Abbreviations used: BHS, NVP-BHS345; Rps6, ribosomal protein S6; TE, translational efficiency; TOR, target of rapamycin.

© 2016 Yerlikaya, Meusburger, *et al.* This article is distributed by The American Society for Cell Biology under license from the author(s). Two months after publication it is available to the public under an Attribution-Noncommercial-Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0/>).

"ASCB®," "The American Society for Cell Biology®," and "Molecular Biology of the Cell®" are registered trademarks of The American Society for Cell Biology.

kinases. These include S6 kinases (S6Ks) and Sch9, which are phosphorylated by mammalian TORC1 (mTORC1) and yeast TORC1, respectively; and Akt/protein kinase B and Sgk1, and Ypk1 and Ypk2, which are phosphorylated by mTORC2 and TORC2, respectively (Pearce *et al.*, 2010; Loewith and Hall, 2011). Phosphorylation of AGC kinases by TOR complexes occurs in distinct, carboxyl-terminal sequences known as the turn motif and hydrophobic motif and increases their activity (Pearce *et al.*, 2010).

Ribosomal protein S6 (Rps6) was one of the first ribosomal proteins found to undergo phosphorylation (Kabat, 1970; Gressner and Wool, 1974). Rps6 phosphorylation is conserved from yeast to mammalian cells. In higher eukaryotes, there are five phosphorylation sites on its carboxyl terminus, whereas only two of these sites (Ser-232 and Ser-233), corresponding to Ser-235 and Ser-236 in higher eukaryotes, are conserved in yeast (Supplemental Figure S1A). Rps6 phosphorylation is sensitive to nutrients, hormones, growth factors, and a variety of stress conditions (Meyuhas, 2008). In higher eukaryotes, S6 kinases 1 and 2 (S6K1 and S6K2) are the major kinases phosphorylating Rps6 on several sites (S235/S236/S240/S244) in response to mTORC1 activity (Pende *et al.*, 2004). Rps6 is additionally phosphorylated by RSKs (90-kDa ribosomal protein S6 kinases) on S235/S236 downstream of mitogen-activated protein kinase signaling (Roux *et al.*, 2007). Members of casein kinase I kinase family were found to phosphorylate Rps6 on Ser-247 in response to mitogenic stimuli (Hutchinson *et al.*, 2011). Intriguingly, both TORC1 and TORC2 regulate Rps6 phosphorylation in fission yeast, via Psk1 and Gad8, respectively (Du *et al.*, 2012; Nakashima *et al.*, 2012). Rps6 dephosphorylation is less well studied but, at least in mammals, appears to require the type 1 protein phosphatase (PP1) (Belandia *et al.*, 1994; Hutchinson *et al.*, 2011).

The major regulators of Rps6 phosphorylation in budding yeast remain unknown. Considering its role in translation and its ability to phosphorylate Rps6 *in vitro*, Sch9 was proposed to be the functional orthologue of S6 kinase in higher eukaryotes (Urban *et al.*, 2007). However, phylogenetic analyses of the TOR signaling network has suggested that Ypk3, rather than Sch9, is more likely to be the S6 kinase orthologue in *S. cerevisiae* (van Dam *et al.*, 2011). Consistently, YPK3 is orthologous to *Schizosaccharomyces pombe* PSK1. In addition, a recent study proposed that Ypk3 activity, but not Sch9 activity, is required for Rps6 phosphorylation *in vivo* in budding yeast (Gonzalez *et al.*, 2015).

Despite years of research, the physiological role of Rps6 phosphorylation remains mysterious. Knock-in mice expressing only the alanine-substituted nonphosphorylatable version of Rps6 exhibited severe phenotypes, including a reduced size, glucose intolerance, and muscle weakness (Ruvinsky *et al.*, 2005). Studies in yeast, however, suggested that Rps6 phosphorylation had no observable effect on the growth of cells under different conditions (Kruse *et al.*, 1985; Johnson and Warner, 1987). Rps6 phosphorylation was once believed to specifically couple mTORC1 signals to the translation of so-called 5'-TOP mRNAs (Fumagalli, 2000), which predominantly encode proteins required for protein synthesis (Levy *et al.*, 1991; Iadevaia *et al.*, 2008). However, another mTORC1 effector, 4E-BP1, is now believed to serve in this capacity (Hsieh *et al.*, 2012; Thoreen *et al.*, 2012). More recently, Rps6 phosphorylation has again been linked to the translation machinery via regulation of the transcription of genes encoding ribosome biogenesis factors (Chauvin *et al.*, 2014).

Here we revisited Rps6 phosphorylation in the model eukaryote *S. cerevisiae*. We found that Rps6 phosphorylation is differentially regulated downstream of TORC1 via Ypk3 and TORC2 via Ypk1 and Ypk2. TORC1 additionally appears to regulate Rps6 dephosphory-

lation via Glc7/Shp1. Cells expressing nonphosphorylatable variants of Rps6 presented phenotypes associated with 40S ribosome biogenesis defects, but these phenotypes were not recapitulated under conditions in which wild-type (wt) Rps6 protein is hypophosphorylated. Cells expressing nonphosphorylatable variants of Rps6 were partially resistant to the loss of Glc7 activity, suggesting that Rps6 hyperphosphorylation is toxic. Curiously, ribosome-footprinting analyses revealed that neither Rps6 phosphorylation nor Ypk1-3 activities are overtly involved in the regulation of gene expression at the level of transcription or translation and thus challenge the present models on the function of Rps6 phosphorylation derived from studies in higher eukaryotes.

RESULTS

Rps6 is differentially phosphorylated on Ser-232 and Ser-233 in a TORC1- and TORC2-dependent manner

Rps6 is encoded by two paralogues, *RPS6A* and *RPS6B*, in *S. cerevisiae*, which encode for identical proteins, Rps6A and Rps6B. Both proteins are phosphorylated on Ser-232 and -233, and knock-ins targeting these phosphorylation sites were made in both *RPS6A* and *RPS6B* genes.

To functionally characterize Rps6 phosphorylation, we initially generated antisera that recognize Rps6 proteins regardless of their phosphorylation status (Figure 1A). We also screened a number of commercially available antisera and found one (which we refer to as Rps6-PP) that cross-reacts with Rps6 when doubly phosphorylated on Ser-232 and -233 (Figure 1A). Immunoreactivity with this antiserum is lost if either or both of these serines are replaced with an alanine (Rps6^{AA}, Rps6^{SA}, or Rps6^{AS}; Figure 1A). We additionally investigated immunoreactivity of Rps6 variants with antiserum recognizing phospho-RXXS*/T* motifs. Rps6^{SS} and Rps6^{SA} but not the Rps6^{AS} variant are recognized by this antiserum (Figure 1A). This result implies that either the antiserum cannot recognize Rps6 phosphorylated on Ser-233 alone or that Ser-233 can only be phosphorylated after prior Ser-232 phosphorylation. We believe that the first hypothesis is correct, as we observe a slower SDS-PAGE migration of the Rps6^{AS} variant compared with the Rps6^{AA} variant (Supplemental Figure S1, B and C).

Using these antisera, we examined how Rps6 phosphorylation responds to TORC1 and TORC2 inhibition. Inhibition of TORC1 with either rapamycin (Figure 1B) or caffeine (Supplemental Figure S1D) triggered rapid dephosphorylation of Ser-233 but not Ser-232. In contrast, inhibition of both TORC1 and TORC2 with BHS triggered rapid dephosphorylation of both serines (Figure 1C). TORC1 (Urban *et al.*, 2007) and TORC2 (Y. Bouza, M. Piccolis, and R. Loewith, unpublished data) are sensitive to carbon quality, and consistent with the BHS result, downshift in carbon quality also triggered rapid dephosphorylation of both serines (Figure 1D). Nitrogen depletion also triggered dephosphorylation of both serines, with slower but otherwise parallel kinetics (Supplemental Figure S1E). Collectively these observations suggested that TORC1 regulates phosphorylation of Ser-233, whereas TORC1 and/or TORC2 regulate phosphorylation of Ser-232.

To distinguish between the two possibilities, we constructed a strain in which TORC2 activity could be inhibited upon addition of the plant hormone auxin, which triggered proteolysis of the essential TORC2 subunit Avo1 fused to an auxin-activated degron (Nishimura *et al.*, 2009; Supplemental Figure S1F). Before this experiment, we first confirmed that TORC1 and TORC2 activities in wild-type cells were not affected by auxin treatment *per se* (Supplemental Figure S1G). Subsequently, we found that inhibition of TORC2 resulted in Ser-232 dephosphorylation only when TORC1

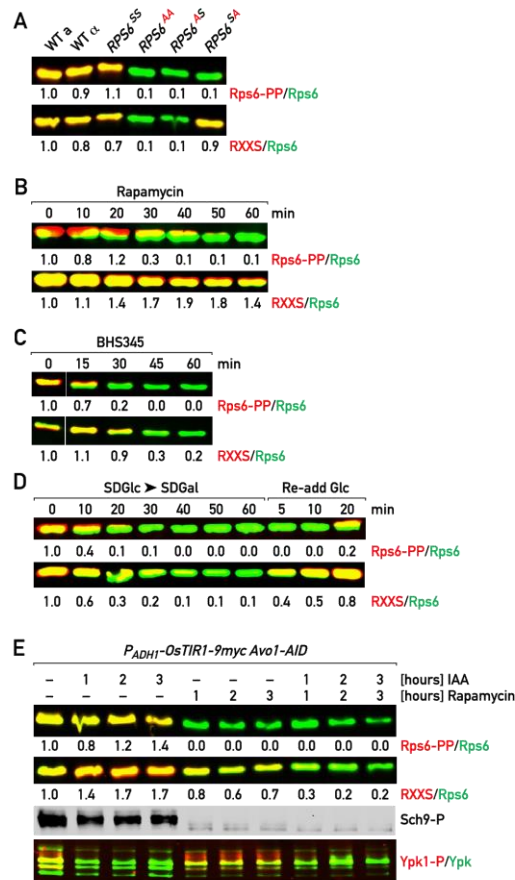


FIGURE 1: Rps6 phosphorylation is differentially regulated on Ser-232 and -233. (A) Western blot of denaturing total protein extracts prepared from yeast cells with the indicated genetic modifications. Membranes were probed with the following antibodies: rabbit anti-phospho-S6 ribosomal protein (S235/S236; Rps6-PP) or rabbit anti-phospho-Akt substrate (RXXS/*/*), together with guinea pig anti-Rps6. Binding of phosphospecific antibodies was revealed with IRDye 680CW-conjugated anti-rabbit immunoglobulin G (IgG; red signal). Binding of anti-Rps6 antibody was revealed with IRDye 800CW-conjugated anti-guinea pig IgG (green signal). (B, C) Exponentially growing wild-type cells were treated with 200 nM rapamycin or 100 μ M BHS. After treatment, the cells were collected at the indicated time points, and protein extracts were processed as described. (D) Rps6 phosphorylation status over time after a shift from glucose-containing medium (SDGlc) to galactose-containing medium (SDGal). Glucose was added 60 min after the carbon downshift, and Rps6 rephosphorylation was followed over the subsequent 20 min. (E) Rps6 phosphorylation status over time after TORC2 inhibition (in cells expressing an Avo1-AID fusion protein treated with 100 μ M indole-3-acetic acid [IAA]) and/or TORC1 inhibition with rapamycin. Sch9 and Ypk1 phosphorylations were additionally assessed as readouts of TORC1 and TORC2 activities, respectively.

was additionally inhibited (Figure 1E). Thus we conclude that TORC1 signals promote phosphorylation of both Ser-232 and Ser-233, whereas TORC2 signals promote phosphorylation of Ser-232 (Figure 4D).

TORC1 regulates Rps6 phosphorylation on Ser-233 via Ypk3

We screened a panel of kinase deletion mutants (Bodenmiller et al., 2010) to identify the TORC1-effector kinase that phosphorylates Rps6. This screen demonstrated that Rps6 is hypophosphorylated on Ser-233 in *ypk3 Δ* cells (Figure 2A and Supplemental Figure S2A). Using an analogue-sensitive, Ypk3-expressing strain, we also found that Ypk3 activity is required for phosphorylation of Rps6 on Ser-233 upon glucose repletion (Figure 2B). In contrast, inhibition of analogue-sensitive protein kinase A (*tpk1^{as} tpk2^{as} tpk3^{as}* cells) did not trigger dephosphorylation of Rps6 (Supplemental Figure S2B), arguing against a role for this related AGC-family kinase in Rps6 phosphorylation.

Because Ypk3 is an AGC-family kinase, we predicted that it could be a direct substrate of TORC1. Consistent with this hypothesis, we found that Ypk3 is hypophosphorylated upon TORC1 inhibition with rapamycin (Figure 2C), carbon downshift (Supplemental Figure S2C), or nitrogen starvation (Supplemental Figure S2D) and that Ypk3 coprecipitates TORC1 in a rapamycin-sensitive manner (Figure 2D). Ypk3 phosphorylation was not obviously affected upon TORC2 inhibition (Supplemental Figure S2E) but was decreased upon direct inhibition of analogue-sensitive Ypk3 with 1NM-PP1 (Figure 2B), suggesting that it autophosphorylates, which was confirmed by in vitro kinase assays (Figure 2E). Addition of TORC1 in the absence but not the presence of wortmannin further increased Ypk3 phosphorylation in vitro. Together, these observations strongly suggest that Ypk3 is a direct target of TORC1.

TOR targets a highly conserved hydrophobic motif at the C-terminal end of AGC kinases. Based on the homology among Ypk1, Ypk2, and Ypk3, we predicted Ser-513 to be the residue phosphorylated within the hydrophobic motif of Ypk3 (Supplemental Figure S2F). Conversion of Ser-513 to Ala altered the SDS-PAGE mobility of Ypk3 but did not affect its catalytic activity (Supplemental Figure S2G). Substitution of this residue with glutamic acid and aspartic acid did not suppress the rapamycin-induced hypophosphorylation of Rps6 (unpublished data), suggesting that either these substitutions do not adequately mimic phospho-Ser-513 or that TORC1 regulates Ypk3 and/or Rps6 phosphorylation through additional mechanisms. To map other rapamycin-sensitive phosphorylation sites on Ypk3, we decided to immunoprecipitate Flag-tagged Ypk3 from untreated and rapamycin-treated cells and analyze the phosphorylation patterns by mass spectrometry. This analysis revealed three rapamycin-sensitive sites in the N-terminus (S86, S92, S94) and three in the C-terminus of the protein (S505, S517, S519; Supplemental Figure S2F). Alanine substitution of the N-terminal residues S86, S92, and S94 did not visibly alter the SDS-PAGE migration pattern or the activity of Ypk3 (Supplemental Figure S2G), and thus we decided to focus on the C-terminal phosphorylation sites. Intriguingly, alanine substitution of S505, S517, and S519 residues in combination with that of S513 (*Ypk3^{4A}*) greatly diminished the phosphorylation level of Rps6 nearly to the level observed in *ypk3* cells (Supplemental Figure S2H). We substituted these four serines with aspartic acid to see whether the phosphomimetic Ypk3 could suppress Rps6 dephosphorylation induced by rapamycin. However, this was not the case (Supplemental Figure S2H), leading us to believe that Ypk3 is not the only factor through which TORC1 regulates Rps6 phosphorylation (see Discussion).

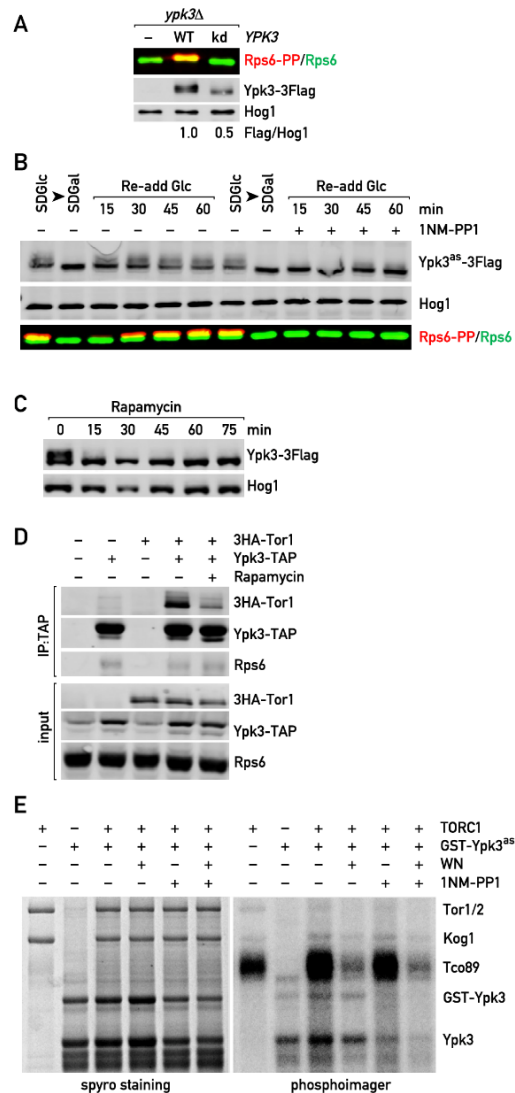


FIGURE 2: TORC1 regulates Rps6 phosphorylation on Ser-233 via Ypk3. (A) Rps6 phosphorylation on Ser-233 is impaired in cells lacking Ypk3 activity. Rps6 phosphorylation in *ypk3* cells expressing an empty vector (-), YPK3-FLAG (WT), or kinase-dead *ypk3*-FLAG (kd) was assessed by Western blot. Hog1 protein levels were assessed as a loading control. (B) A *ypk3^{as}* strain was grown in SD medium with 2% glucose, filtered, and resuspended in SD with 2% galactose. After 1 h of incubation at 30°C, the culture was split into two: one supplemented with 2% glucose and DMSO, and the other with 2% glucose and 500 nM 1NM-PP1 to inhibit Ypk3^{as} activity. Rps6 phosphorylation status and Ypk3-FLAG and Hog1 protein levels were assessed at various time points before, during, and after carbon downshift as indicated. (C) The migration pattern of Ypk3-FLAG collected from exponentially growing cells expressing before and after treatment with 200 nM rapamycin was assessed by Western

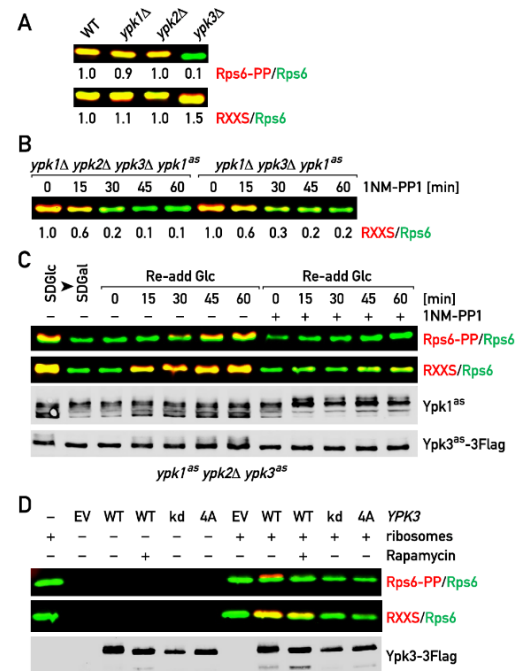


FIGURE 3: TOR complexes 1 and 2 regulate Rps6 phosphorylation on Ser-232 via Ypk1-3. (A) Individual deletion of neither YPK1 nor YPK2 altered Rps6 phosphorylation. (B) Inhibition of Ypk1^{as} with 1NM-PP1 in *ypk2Δ ypk3Δ* and *ypk3Δ* backgrounds triggered loss of Ser-232 phosphorylation. (C) Inhibition of Ypk1^{as} and Ypk3^{as} with 1NM-PP1 prevented Rps6 rephosphorylation upon glucose refeeding. (D) Ypk3 phosphorylates Rps6 in vitro. Rps6 in yeast ribosomes was phosphorylated by wild-type Ypk3-FLAG purified from exponentially growing cells. Rapamycin treatment before purification reduced Ypk3-FLAG activity. Kinase-dead (kd) or nonphosphorylatable Ypk3 variant (4A) also failed to phosphorylate Rps6 in vitro.

TOR complexes 1 and 2 regulate Rps6 phosphorylation on Ser-232 via Ypk1/2/3

Unlike Ser-233, we could still detect phosphorylation of Ser-232 in the absence of YPK3 (Figure 3A). Ypk1 and Ypk2 are bona fide substrates of TORC2 (Kamada et al., 2005; Niles et al., 2012) and are also closely related to Ypk3, suggesting that all three kinases could redundantly mediate Rps6 phosphorylation on Ser-232. Loss of Ypk1 and/or Ypk2 function did not affect the phosphorylation levels of Rps6 (Figure 3A and Supplemental Figure S3A). However, inhibition of Ypk1 and Ypk2 in combination with rapamycin treatment

blot. (D) Ypk3-TAP interacts with Tor1. Protein extracts were prepared from vehicle- or rapamycin-treated (30 min) cells expressing or not Ypk3-TAP and/or HA3-Tor1. Ypk3 was immunoprecipitated using IgG Dynabeads, and associated proteins were detected by Western blot. (E) TORC1 phosphorylates Ypk3 in vitro. GST-Ypk3^{as} purified from yeast was used as a substrate for TORC1 purified from yeast by immunoprecipitating Tco89-TAP. The combination of wortmannin (which inhibits TORC1) and 1NMPP1 (which inhibits GST-Ypk3^{as}) was used to show the specificity of the kinase reaction.

induced hypophosphorylation of Ser-232 (Supplemental Figure S3B). Next we combined the deletion of *YPK3* with the deletion of *YPK2* and an analogue-sensitive allele of *YPK1*. In the absence of the activities of all three kinases, both Ser-232 and Ser-233 were hypophosphorylated (Figure 3B). Rephosphorylation of Rps6 on both Ser-232 and Ser-233 upon refeeding of glucose also appeared to require the activity of all three Ypk kinases (Figure 3C), although we note that the contribution of Ypk2 to Rps6 phosphorylation appears to be minimal.

Wild-type Ypk3 but not inactive variants phosphorylated Rps6 in vitro, strongly suggesting that Rps6 is a direct target of Ypk3 (Figure 3D). Purified ribosomes were used as the source of Rps6 in this assay, as recombinant Rps6 was extremely unstable in our hands. Curiously, we failed to observe phosphorylation of Rps6 by Ypk1 in a similar assay (Supplemental Figure S3C). This negative result prevents us from concluding that Ypk1 directly phosphorylates Rps6. We did, however, observe synthetic lethality in *ypk1* and *ypk3* double mutants (Supplemental Figure S3D). This observation provides genetic support to the idea that Ypk1 and Ypk3 share common targets.

Glc7/Shp1 mediates TORC1/2-dependent phosphorylation of Rps6

Protein phosphatase 1 was recently proposed to mediate dephosphorylation of Ser-247 in Rps6 in mammalian cells (Hutchinson et al., 2011). Thus we investigated the role of PP1 in Rps6 dephosphorylation in yeast, using a loss-of-function mutant of Glc7, the catalytic subunit of yeast PP1 (Baker et al., 1997). Expression of the *glc7-127* allele, which contains two point mutations, K110A and K112A, displays defects in glycogen accumulation and elevated levels of phosphorylation of histone H3 (Baker et al., 1997), blocked the dephosphorylation of Rps6 in response to rapamycin and BHS treatments, and led to higher basal phosphorylation levels under normal conditions (Figure 4A). In yeast, calyculin A is a known inhibitor of PP1 and to a lesser extent PP2A (Hoon et al., 2008). Intriguingly, we found that strains expressing *RPS6^{AA}* exhibited slight resistance to calyculin A treatment compared with the wild-type strains (Figure 4B). Calyculin A sensitivity of these cells could be partially rescued by the overexpression of *GLC7*, supporting the notion that Glc7 is the target of this drug (Supplemental Figure S4A). This observation linked Glc7 activity and Rps6 phosphorylation and suggested that constitutive hyperphosphorylation of Rps6 is toxic, although other, indirect mechanisms could explain the observed growth rescue.

Glc7 activity is regulated by numerous regulatory subunits. To identify the regulatory subunit involved in Rps6 dephosphorylation, we screened deletion mutants of genes encoding all known nonessential Glc7-interacting proteins for effects on Rps6 phosphorylation. This approach identified Shp1 as a protein required for the dephosphorylation of Rps6 upon rapamycin treatment (Supplemental Figure S4B). Deletion of *SHP1* suppressed the effect of rapamycin and BHS on Rps6 phosphorylation to an extent comparable to that of the *glc7-127* allele (Figure 4C). These results functionally link Shp1 and Glc7 in the regulation of Rps6 dephosphorylation downstream of TORC1 and TORC2. The kinase and phosphatase pathways governing Rps6 phosphorylation described here are summarized in Figure 4D.

Physiological characterization of Rps6 phosphomutants

To identify the physiological significance of Rps6 phosphorylation, we first compared the growth rates of cells expressing either wild-type Rps6 protein or the nonphosphorylatable variant *RPS6^{AA}*. In our

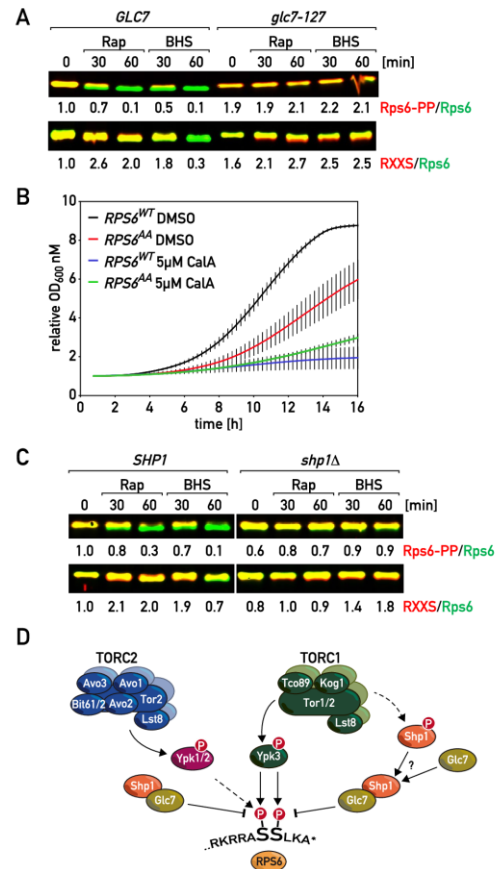


FIGURE 4: Glc7/Shp1 mediates TORC1/2-dependent dephosphorylation of Rps6. (A) Exponentially growing cells of the indicated genotypes were treated with 200 nM rapamycin or 100 μM BHS for the indicated times before analysis of Rps6 phosphorylation. The *glc7-127* allele suppressed both the rapamycin- and the BHS-induced dephosphorylation of Rps6. (B) Strains of the indicated genotypes were diluted into medium containing drug vehicle (DMSO) or 5 μM calyculin A, and growth (optical density at 600 nM) was measured every 15 min for the following 16 h. (C) Like the *glc7-127* allele, deletion of *SHP1* also suppressed both the rapamycin- and the BHS-induced dephosphorylation of Rps6. (D) Cartoon depicting the TORC1- and TORC2-regulated signaling pathways governing Rps6 phosphorylation.

hands, expression of the *RPS6^{AA}* alleles resulted in an almost 30% decrease in the proliferation rate (Figure 5A and Supplemental Figure S5A). The *RPS6^{AA}* cells also had lower protein content than the wild-type cells as assessed by fluorescein isothiocyanate (FITC) staining and fluorescence-activated cell sorting (FACS) analysis (Figure 5B and Supplemental Figure S5B).

To explore the effect of Rps6 phosphorylation on global translation, we monitored the abundance of free 40S and 60S subunits, 80S monosomes, and actively translating ribosomes (polysomes) in

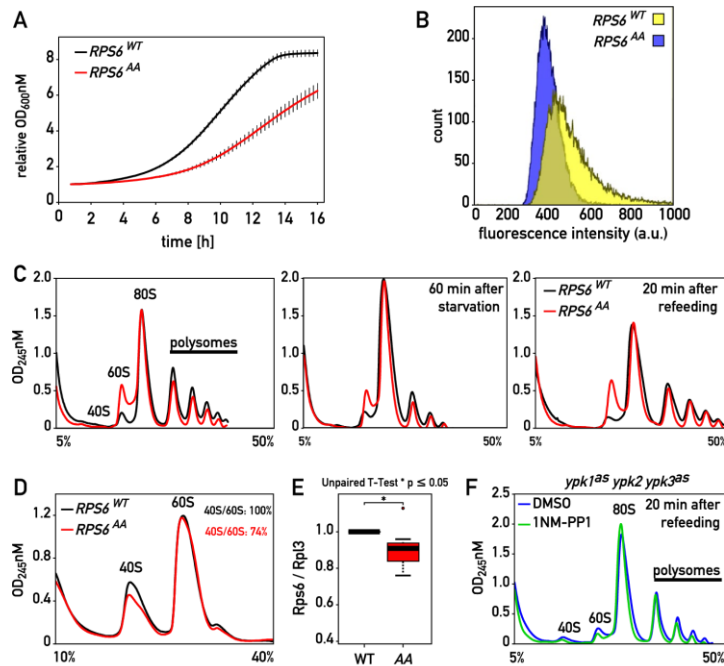


FIGURE 5: Physiological characterization of Rps6 phosphorylation. (A) Growth (OD₆₀₀) of wt and *RPS6^{AA}* cells in rich medium at 30°C was recorded every 15 min for 16 h. (B) The protein content of exponentially growing wt and *RPS6^{AA}* cells was measured by FITC staining and FACS analysis. (C) Polysome profiles of wt and *RPS6^{AA}* cells. Cells were grown to exponential growth phase in SD medium containing 2% glucose. After filtering, cells were resuspended in SD medium lacking a carbon source and incubated at 30°C for 1 h. Subsequently, cells were refed with 2% glucose. Aliquots of cells were treated with cycloheximide before and after 60 min of starvation and after 20 min of refeeding. Polysome profiles from these three time points are shown. (D) Ribosomes extracted from wt and *RPS6^{AA}* cells were incubated in the presence of 800 mM KCl. Free 40S and 60S subunits were analyzed by sucrose gradient centrifugation. (E) The relative amount of Rps6 normalized to Rpl3 in protein extracts prepared from exponentially growing wt and *RPS6^{AA}* cells was assessed by Western blot. (F) *ypk1^Δ ypk2 ypk3^Δ* cells starved of a carbon source for 60 min were subsequently refed with glucose in the presence of DMSO or 1NM-PP1. Twenty minutes later, cells were treated with cycloheximide, and polysome profiles were generated.

wild-type and *RPS6^{AA}* cells using sucrose gradient centrifugation. The ratios of polysomes to 80S monosomes were the same in these two strains (Figure 5C). However, the mutant cells exhibited an elevated 60S peak, indicating an imbalance between the free subunits. Next we analyzed the polysome profiles of wild-type and *RPS6^{AA}* cells upon glucose starvation and refeeding with glucose. Although the *RPS6^{AA}* cells displayed an increase in the 60S peak in all of the conditions we tested, we did not detect other overt changes in the profiles compared with the wild-type cells under these conditions (Figure 5C and Supplemental Figure S5C).

To further investigate whether the elevated 60S peak is due to a defect in the production of the small subunit, we dissociated the monosomes and polysomes into free subunits in the presence of high salt. We observed that the level of 40S subunits was indeed lower in the mutant strain than in the wild type strain (Figure 5D). Quantitative Western blots confirmed that the ratio of Rps6 to a large-subunit protein, Rpl3, is reduced in *RPS6^{AA}* cells (Figure 5E

and Supplemental Figure S5D). Furthermore, by Northern blot, we observed a delay in rRNA maturation in *RPS6^{AA}* cells (Supplemental Figure 5E). In summary, *RPS6^{AA}* cells display a 40S biogenesis defect but no obvious defect in global translation. These phenotypes are similar to what is observed in the absence of one of the two copies of the *RPS6* genes (Pachler et al., 2004), and thus it is unclear whether the 40S biogenesis defect results from Rps6 hypophosphorylation or instead is due to reduced levels of Rps6^{AA} protein (see Discussion).

To probe for a role of Ypk kinases in translation, we compared polysome profiles generated from wt cells to those generated from cells lacking Ypk1–3 activities. Because chronic inhibition of these kinases was not an option (the cells die), we chose to use an experimental setup in which we acutely activate translation in the presence or absence of Ypk1–3 activity. Specifically, *ypk1^Δ, ypk2^Δ, ypk3^Δ* cells were starved of glucose for 1 h and then refed with glucose with or without the analogue (1NM-PP1). Analysis of the abundance of the free subunits (40S and 60S), the monosomes (80S), and the polysomes collected after refeeding with or without 1NM-PP1 revealed that acute inhibition of Ypks had no significant effect on the profiles under any of the conditions tested despite the expectant hypophosphorylation of Rps6 (Figure 5F and Supplemental Figure S5, F and G). These results are in accordance with a recently published study, which showed that the polysome profiles from wild-type and *S6K1;S6K2^{-/-}* liver cells exhibited no significant differences under starvation or refeeding conditions (Chauvin et al., 2014). Unlike *RPS6^{AA}* cells, *ypk3* cells displayed no defects in proliferation, protein accumulation, or 40S biogenesis (Supplemental Figure 5E; unpublished data).

Our polysome profiling results thus far demonstrate that Rps6 phosphorylation plays no role in global translation but do not query for a potential role in the translation of individual messages. To probe for such a role, we performed ribosome profiling on *ypk1^Δ ypk2 ypk3^Δ RPS6^{wt}* and *ypk1^Δ ypk2 ypk3^Δ RPS6^{AA}* cells with or without 1NM-PP1 (protocol in Supplemental Figure S6A). Specifically, we compared the profiles of these cells after 20 and 40 min of refeeding with glucose. Under these conditions, we observed expected patterns of Rps6, Ypk1, and Ypk3 phosphorylation (Supplemental Figure S6B). The majority of ribosome footprints were 28–32 nucleotides (nt) long, and metagene analysis showed the expected triplet periodicity within coding regions (Supplemental Figure S6, C and D). For both RNA sequencing (RNA-seq) and ribo-seq reads, the ribosome profiling data exhibited good correlations for coverage between the replicates (Supplemental Figure S6E). We observed no strain-dependent differences in global mRNA levels (RNA-seq) or global translational efficiencies (TEs; Figure 6A), demonstrating

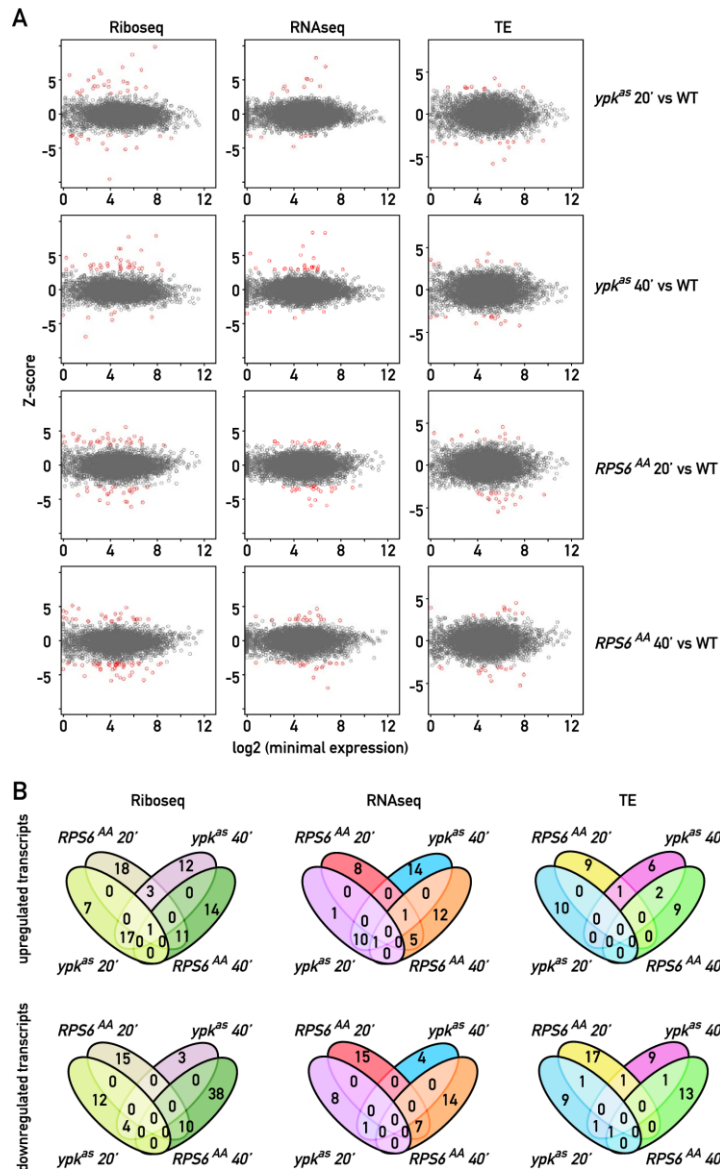


FIGURE 6: Differential gene expression analysis. (A) Scatter plots showing Z-scores (y-axis) of fold changes relative to wild-type control of ribosome occupancy (Riboseq), transcript levels (RNAseq), and translation efficiency (TE). Transcripts are ordered based on their log2 minimal expression signal (x-axis). Transcripts with $|Z| \geq 3$ are shown in red. (B) Venn diagrams showing the number of transcripts whose changes in ribosome occupancy, transcript levels, or TE levels exceed a $|Z| \geq 3$ threshold vs. wild-type controls in the indicated conditions.

that, under these conditions, Rps6 phosphorylation does not strongly affect gene regulation at the level of either transcription or translation. To identify potential differential gene expression, we used Z-score transformation as described earlier (Andreev et al.,

2015). Only five (*SED1*, *YCR007C*, *RPS21A*, *UIP5*, *Q0050*) of 6692 transcripts displayed TE levels with absolute Z-scores ≥ 3 relative to the wild type in both *Rps6^{AA}* and Ypk-inhibited cells (Figure 6B and Supplemental Figure S7). We note that five is close to the expected number of transcripts to be selected by chance with this Z-score cutoff. Indeed, the profiles of *SED1* and *UIP5* when assessed manually did not display obvious differences between conditions (Figure 7A). We also assessed in detail the transcript profiles of a number of known housekeeping genes (e.g., *ACT1*), as well as ribosomal protein genes (e.g., *RPS6B*) and Ribi genes (e.g., *TEF4*) whose expressions are known to be strongly dependent on TORC1 activity (Lippman and Broach, 2009; Huber et al., 2011). None of the selected transcripts exhibited a striking change in profile across the different conditions (Figure 7A). Neither *Rps6^{AA}* mutation nor Ypk activity affected translational efficiency of transcripts possessing introns (e.g., *ACT1*, *RPS6B*, *TEF4*; Figure 7A) and/or transcripts whose translation is regulated by upstream open reading frames (ORFs; e.g., *GCM4*). These observations do not support a role of Rps6 phosphorylation in the pioneer round of translation and/or in initiation.

Next we compared ribosome occupancy of 5' leaders and 3' untranslated regions (UTRs) across the different conditions. The number of footprints mapping to 3' UTRs remained constantly low regardless of the phosphorylation status of Rps6 (Figure 7B). At 20 min, the number of footprints mapping to 5' leaders was lower in *RPS6^{AA}* cells in both replicates; however, the difference was not higher than replicate variation. We hypothesized that the changes in 5' leader densities may occur in specific mRNAs and analyzed the changes in the relative density between 5' leaders and coding regions. The observed changes between different conditions did not exceed those observed for the replicates obtained for the same conditions (Supplemental Figure S7B). A similar, less robust trend was observed in Ypk-inhibited cells.

Finally, we explored a possible role for Rps6 phosphorylation in frameshifting. During ribosomal frameshifting, a proportion of ribosomes shift their reading frame to produce a protein product that is encoded by two ORFs. The ribosome density profiles for such mRNAs are normally characterized by a drop of density downstream of ORF1 and

a change in the phase of the triplet periodicity between nonoverlapping regions of ORF1 and ORF2, which is consistent with the direction of frameshifting (Michel et al., 2012). The drop of density should correlate with frameshifting efficiency and thus could be

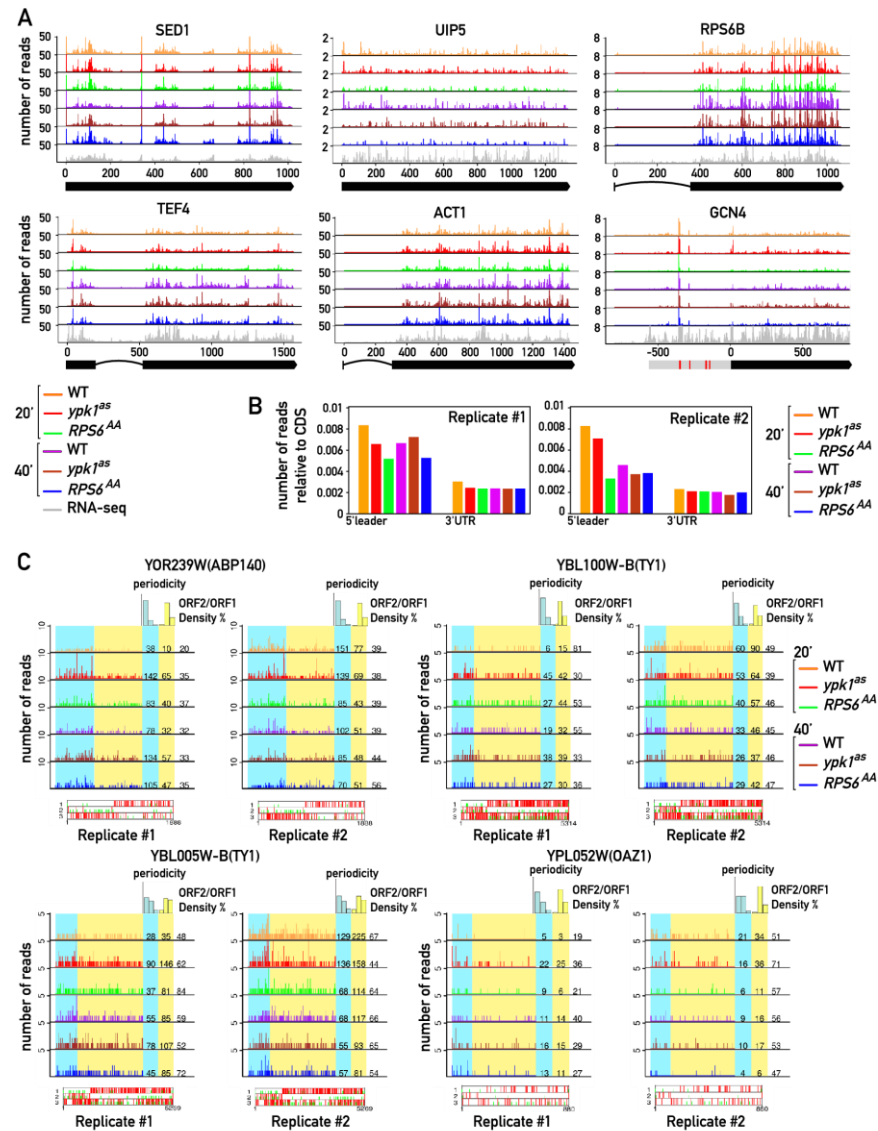


FIGURE 7: Alternative translational processes occur independently of *Rps6* phosphorylation status. (A) Ribo-seq and RNA-seq densities obtained from the six different conditions/time points mapped to genomic loci of individual genes. Ribo-seq densities are indicated with different colors, depending on the condition (indicated at the bottom). Combined RNA-seq densities of all six conditions are shown in gray. Locations of coding regions are indicated with black bars under each profile, introns with black lines, and upstream open reading frames with red bars. Genomic coordinates are given for the sense strand relative to the start codons, with the first nucleotide of the start codons as the zero coordinate. (B) The number of footprints mapped to 5' leaders and 3' UTRs normalized to the number of footprints mapped to coding sequences (CDS) are shown for each Ribo-seq replicate. (C) Analysis of ribosome footprint densities on mRNAs for the genes that require ribosomal frameshifting for their expression. Footprint densities are shown separately for all conditions and two replicates. The ORF plot for each gene is shown below the density plots, ORF1 is highlighted in blue, and a nonoverlapping part of ORF2 is highlighted in yellow. The total number of reads aligning to each of the two areas is indicated on the right side of each plot, along with footprint density within ORF2 relative to ORF1. Histograms above represent periodicity within both ORFs. It can be seen that the phase of periodicity shifts according to the expected +1 directionality of ribosomal frameshifting.

used to assess differences in frameshifting efficiencies between samples.

To obtain information on *S. cerevisiae* genes that require programmed ribosomal frameshifting for their expression (*ABP140*, *EST3*, *OAZ1*, and *TY1* transposable element), we used the Recode-2 database (Bekaert et al., 2010). The *EST3* mRNA had insufficient coverage for the analysis. Profiles of ribosome density for *ABP140*, *OAZ1*, and two arbitrary chosen copies of *TY1* element are shown in Figure 7C. Because the ambiguous alignments were used for frameshifting analysis (see *Materials and Methods*), profiles for *TY1* elements represent mRNA translation of several *TY1* copies that share strong sequence similarity.

We observed a change of triplet periodicity consistent with the known directionality of frameshifting in all cases; however, the observed variation in the drop of ribosome density was insignificant among various samples in comparison to the variation observed between replicates (Figure 7C).

In summary, consistent with previous work (Chauvin et al., 2014), we found that Rps6 phosphorylation plays no role in global protein synthesis upon refeeding as assessed by polysome profiling. To probe for potential effects on the translation of individual transcripts, we exploited a recent technological advance in the translation field—ribosome profiling. Specifically, using ribosome profiling, we assessed changes in RNA levels and translation efficiencies of individual mRNAs at the scale of the whole transcriptome (Supplemental Figure S7) under conditions in which Rps6 was or was not phosphorylated. We used a relaxed threshold of statistical significance to call for differential expression ($|\log_2\text{-score}| \geq 3$), for which 0.25% of analyzed genes are expected to be classified as differentially expressed by chance under the assumption that changes in coverage due to technical factors follow a normal distribution. Even under these relaxed thresholds, we failed to identify differentially expressed genes that would be shared between Rps6 mutants and Ypk-inhibited cells at all time points (Figure 6B).

DISCUSSION

The pathways regulating Rps6 phosphorylation in budding yeast have remained ambiguous and, in higher eukaryotes, surprisingly complex. Furthermore, the physiological role of Rps6 phosphorylation has remained mysterious for the past four decades. Rps6 is a constitutive component of the 40S small subunit of the ribosome, and, since its phosphorylation is promoted downstream of anabolic stimuli, this modification has long been assumed to promote protein synthesis. However, experimental evidence supporting such a role has not been forthcoming. In this study, we defined the signaling cascades that regulate Rps6 phosphorylation in budding yeast. We demonstrated that nonphosphorylatable Rps6 knock-ins do not accurately mimic the function of hypophosphorylated protein. Finally, using ribosome profiling, we found little evidence to support a role for Rps6 phosphorylation in the regulation of translation per se. We are thus compelled to conclude that whatever role Rps6 phosphorylation plays, this role is not directly related to mRNA translation.

Both TORC1 and TORC2 influence Rps6 phosphorylation

In the present work, we demonstrate for the first time that the two phosphorylation sites on the C-terminus of Rps6 are differentially regulated in yeast and that this regulation is dependent on both TORC1 and TORC2. Specifically, we show that TORC1, via Ypk3, regulates Ser-232 and Ser-233 phosphorylation, whereas TORC2, via Ypk1 and, to a lesser extent, Ypk2, regulates Ser-232 phosphorylation. Rps6 phosphorylation thus appears to be similarly regulated in budding and fission yeast (Du et al., 2012; Nakashima et al., 2012).

This conservation across the distantly related yeasts *S. cerevisiae* and *S. pombe* predicts that TORC2 will similarly be important for S6 phosphorylation in higher eukaryotes. Curiously, although we observed direct phosphorylation of Rps6 by Ypk3 *in vitro*, we failed to observe Rps6 phosphorylation by Ypk1 in the same assay. This was surprising, given the dependence of Rps6 phosphorylation on the activity of these kinases *in vivo* and the strong homology among all three Ypks and mammalian S6K1/2. Perhaps phosphorylation of Rps6 by Ypk1/2 occurs in a cellular context that is not recapitulated in the *in vitro* assay.

Although Ypk1 and 2 are well-established effectors of TORC2, direct phosphorylation of Ypk3 by TORC1 is novel and extends previous work (Gonzalez et al., 2015). Specifically, we found that Ypk3 physically interacts with TORC1 *in vivo* in a rapamycin-sensitive manner. We mapped four rapamycin-sensitive phosphorylation sites on the C-terminal end of Ypk3 and demonstrated that they are indispensable for its activity as an S6 kinase. However, expression of the phosphomimetic mutant of Ypk3 failed to suppress rapamycin-induced Rps6 dephosphorylation, leading us to hypothesize that TORC1 might regulate not only Rps6 phosphorylation, but also its dephosphorylation.

In this vein, we found that Glc7/PP1 is the phosphatase responsible for dephosphorylation of both Ser-232 and Ser-233. We further demonstrated that the Glc7 regulator Shp1 is also necessary for Rps6 dephosphorylation. Shp1, in complex with the AAA-ATPase Cdc48, promotes the assembly of Glc7–Sds22–Ypi1 complexes (Cheng and Chen, 2015). However, the mechanism by which TOR complexes interfere with Glc7 and/or Shp1 activities remains to be elucidated. Intriguingly, preliminary data show that Shp1 is phosphorylated in a rapamycin-sensitive manner (unpublished data).

Previously we proposed that Sch9, another AGC kinase and TORC1 effector, could serve as the yeast S6K (Urban et al., 2007). Consistently, we found that expression of a rapamycin-insensitive variant of Sch9 delays rapamycin-induced Rps6 dephosphorylation (unpublished data). However, unlike *YPK3* deletion, *SCH9* deletion did not compromise Rps6 phosphorylation (unpublished data). Thus we no longer believe that Sch9 is truly an S6 kinase. At present, we favor the idea that Sch9 signals block Rps6 dephosphorylation. This could be direct, via Shp1 phosphorylation, or, indirect, via its influence on translation initiation (Urban et al., 2007).

What function does Rps6 phosphorylation serve?

Previous attempts to identify the physiological importance of Rps6 phosphorylation in yeast led to the observation that Rps6 phosphorylation had no obvious effect on growth under a variety of stress conditions or during sporulation (Kruse et al., 1985; Johnson and Warner, 1987). Unexpectedly, in our yeast background (TB50a), we found that the *RPS6^{AA}* mutations result in a lower proliferation rate and protein content than with the wild-type cells, which is likely due to reduced levels of Rps6^{AA} protein. Furthermore these cells present a 40S ribosome biogenesis defect evident at both the protein level and rRNA processing level. These phenotypes are not apparent under conditions in which wild-type Rps6 protein is hypophosphorylated. From these results, we conclude that one must be extremely cautious when assigning phenotypes obtained from cells expressing nonphosphorylatable variants of Rps6 to lack of Rps6 phosphorylation.

To complement our biochemical efforts, we also used synthetic genetic array (SGA) technology (Tong et al., 2001) to probe for synthetic interactions in triple mutants containing *RPS6^{AA}*, *RPS6B^{AA}*, and each nonessential gene deletion. The two validated hits that we obtained from this assay were *dom34* and *hbs1*, each of which

displayed a synthetic sick phenotype in combination with *RPS6^{AA}*, *RPS6^{BA}* (unpublished data). Dom34 and Hbs1 function together in no-go mRNA decay (Doma and Parker, 2006; Harigaya and Parker, 2010), but *dom34* and *hbs1* display synthetic sick interactions with most deletion mutants of 40S ribosomal protein paralogues (Costanzo *et al.*, 2010). This observation again suggests that it is the reduced levels of *Rps6^{AA}* protein and not the lack of *Rps6* phosphorylation that underlies this synthetic interaction. There are now ample data to suggest that many ribosomal proteins perform non-ribosome-related tasks (Warner and McIntosh, 2009). If *Rps6* phosphorylation were involved in such a function, one might hope that this could have been discovered through this SGA approach.

We also made the interesting observation that *RPS6^{AA}* mutation increases resistance to the PP1 inhibitor calyculin A. The simplest interpretation of this result is that hyperphosphorylation of *Rps6* is toxic. Consistently, *RPS6^{DD}* cells were found to be inviable, which, unfortunately, makes their further characterization an intriguing but challenging prospect.

Recent structural studies showed that *Rps6* is located far from the decoding center, accessible from the solvent side of the 40S subunit (Ben-Shem *et al.*, 2011; Khatter *et al.*, 2015). According to these structures, the C-terminal end of Rpl24 relocates and interacts with *Rps6* during translocation. The functional significance of this interaction remains to be elucidated. Unfortunately, the C-terminus of *Rps6* is missing from these structures, making it challenging to define a function for the phosphorylation of this domain from these structural studies.

Previous studies addressing the physiological function of *Rps6* phosphorylation in higher eukaryotes suggested that *Rps6* phosphorylation slows translation elongation (Ruvinsky *et al.*, 2005; Chauvin *et al.*, 2014). These are elegant studies, but because they use deletion of the *S6* kinases and knock-in alleles of *Rps6*, they cannot account for the likely cellular adaptation that occurs in response to these chronic deficiencies, as the authors acknowledge. Here we also queried whether *Rps6* phosphorylation plays a role in translation in yeast. To this end, we generated both polysome profiles and ribosome profiles from wt, *RPS6^{AA}*, and Ypk-inhibited cells after refeeding with glucose to stimulate translation. Of importance, our ability to acutely inactivate the kinases responsible for *Rps6* phosphorylation avoids potential confounding effects arising from cellular adaptation. After extensive analyses, an overt role for *Rps6* phosphorylation could be ascribed to neither global translation nor translational regulation of specific subsets of mRNAs. Specifically, we sought changes in global transcription that would have been apparent in the polysome profiling, whereas transcript-specific effects would have been observable in the ribosome profiling. We found no changes in the density of ribosome-protected fragments across conditions, implying that *Rps6* phosphorylation does not affect the rate of protein synthesis of individual transcripts. In addition, alternative translational processes, including pausing, frameshifting of specific messages, reinitiation on *GCN4* transcripts, and global stop-codon readthrough, appeared to occur independently of the phosphorylation status of *Rps6*. Finally, and in contrast to previous studies, our experimental setup was chosen to potentially expose a role of *Rps6* phosphorylation in the pioneering round of translation of individual transcripts. Specifically, we assumed that refeeding of cells after 1 h of glucose starvation would trigger inaugural translation of newly transcribed messages, which would be visualized at one of the two ribosome profiling time points. Thus, under the experimental conditions used, we find no evidence to support a role for *Rps6* phosphorylation in any aspect of translation. One exception to this general conclusion is the reduction of ribosomes on 5' leaders in

RPS6^{AA} cells and, to a lesser extent, in cells lacking Ypk1–3 activities. Ribosomes accumulate on 5' leaders in starved cells (Ingolia *et al.*, 2009), but the physiological role underlying this phenomenon is unknown. The observed reduction requires further study to assess its statistical significance under the tested conditions.

Recent work implicated *S6* kinases and *Rps6* phosphorylation in the ribosome biogenesis transcriptional program in mouse liver (Chauvin *et al.*, 2014). Although such a function could potentially be conserved in yeast, after analysis of our mRNA-seq data used for the ribosome profiling, we failed to detect *Rps6* phosphorylation-dependent transcriptional regulation of any mRNA subgroups. Thus, like translation elongation, the influence of *Rps6* phosphorylation on transcription appears to be restricted to higher eukaryotes.

MATERIALS AND METHODS

Yeast cultures and inhibitor concentrations

S. cerevisiae strains and plasmids used in the study are listed in Supplemental Tables S1 and S2. Standard protocols were followed to obtain the strains. Cells were grown in YPD (1% yeast extract, 2% peptone, 2% glucose) at 30°C if not indicated otherwise.

Rapamycin was used at 200 ng/ml, wortmannin at 10 µg/ml, and cycloheximide at 25 µg/ml.

Cells for carbon starvation experiments were grown in complete synthetic medium with glutamine (Gln) as the nitrogen source at 30°C, filtered, and resuspended in the same medium lacking a carbon source or with galactose instead of glucose (Glc). At the indicated times after the medium change, either 2% Glc or an equal amount of water was added into the cultures. After further incubation for the indicated times, aliquots of cells were collected for protein extraction.

Cells for nitrogen starvation experiments were grown in yeast nitrogen base (YNB) (-N) + 2% Glc, 0.2% Gln, filtered, and resuspended in YNB (-N) + 2% Glc. At 20 min later, either 0.2% Gln or an equal amount of water was added to the cell cultures.

For spot assays, cells grown overnight were diluted to OD_{600 nm} = 0.1 and grown to exponential phase. Serial dilutions at 1:10 were spotted after normalization of cell numbers. The plates were incubated for 2–3 d at 30°C.

For growth assays, exponentially growing cells were diluted to OD_{600 nm} = 0.005 in the indicated media. A 200-µl amount of cells along with 2 µl of drug solution or drug vehicle alone per well was dispensed in 96-well plates, and the OD_{600 nm} of the wells was recorded every 15 min for 16–24 h at 30°C. The average and SD values were obtained from at least three independent experiments.

Antibodies

For total *Rps6* antibody, guinea pig polyclonal antiserum was raised against the peptides RVFFDKRIGQEV DGE and QRALKVRNAQA-QREA, and affinity purified (Eurogentec, Seraing, Belgium). Supplemental Figure S1b shows the specificity of the antibodies. Other antibodies used in this study were rabbit anti-phospho-S6 ribosomal protein (S235/S236) at 1:2000 (Rps6-PP; 4858S; Cell Signaling Technology, Danvers, MA), rabbit anti-phospho-Akt substrate (RXXS*/T*) at 1:5000 (9614S; Cell Signaling), rabbit anti-Hog1 at 1:1000 (y-215; Santa Cruz Biotechnology, Dallas, TX), mouse anti-Rpl3 at 1:5000 (Developmental Studies Hybridoma Bank, Iowa City, IA), mouse anti-Flag at 1:1000 (F1804; Sigma-Aldrich, St. Louis, MO), rabbit anti-Ypk1 pT662 at 1:500, rabbit anti-Sch9 pT737 at 1:500, goat anti-Ypk1 at 1:1000 (yl-15; Santa Cruz Biotechnology), and the corresponding IRDye infrared dye-labeled secondary antibodies at 1:10,000 (Li-Cor Biosciences, Lincoln, NE). The signals were detected by Odyssey Imaging Systems.

Denaturing protein extraction

Protein extraction was performed based on the trichloroacetic acid-urea extraction method as previously described (Urban *et al.*, 2007). PPI: 10 mM NaF, 10 mM *p*-nitrophenylphosphate, 10 mM Na₂P₂O₄, and 10 mM β -glycerophosphate. PI: 1 $\times$ Roche protease inhibitor cocktail and 1 mM phenylmethylsulfonyl fluoride.

Extraction of ribosomes

Exponentially growing cells were treated with 0.1 mg/ml cycloheximide (CHX). After 10 min of incubation, cells were harvested, washed, and lysed with ribosome lysis buffer (10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 30 mM MgCl₂, 100 μ g/ml CHX, 200 μ g/ml heparin). The extract was centrifuged for 10 min at 13,000 rpm at 4°C. The OD₂₆₀ of 1:100 dilutions of the supernatant was measured. After the addition of glycerol with the final concentration of 10%, the samples were stored at -70°C.

Preparation of recombinant proteins and kinase assays

Purification of TORC1 from RL170-2c cells was performed as described previously for TORC2 (Gaubitz *et al.*, 2015). Glutathione S-transferase (GST)-Ypk3 was expressed under a galactose-inducible promoter in yeast from a 2- μ plasmid. After a 4-h galactose induction, yeast extract was prepared, and the fusion protein was purified using Glutathione Sepharose 4B from GE Healthcare Life Sciences (Little Chalfont, United Kingdom) according to the manufacturer's instructions. The eluted protein was dialyzed, aliquotted, and frozen at -70°C. TORC1 kinase reactions were performed in kinase buffer (2.5 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate [CHAPS], 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid [HEPES], 150 mM KCl, 5 mM MgCl₂, and 1 mM dithiothreitol [DTT]) in a final volume of 30 μ l. Reactions were started with the addition of 300 μ M ATP, 125 mM MgCl₂, and 10 μ Ci of [γ -³²P] ATP and then shaken for 20 min at 30°C. After the addition of 6 μ l of 6 $\times$ SDS-PAGE sample buffer, samples were heated to 65°C for 10 min. Samples run in SDS-PAGE gels were stained with SYPRO Ruby Protein Gel Stain (Thermo Fisher Scientific, Waltham, MA) and analyzed using a Bio-Rad Molecular Imager.

For Ypk3 and Ypk1 kinase assays, ribosomes were used as substrates. The ribosomal extract prepared as described was loaded on top of a 10% sucrose cushion prepared in ribosome lysis buffer and ultracentrifuged for 3.5 h at 35,000 rpm. The pellet was resuspended in the corresponding kinase buffer.

A yeast lysate with Ypk3-FLAG fusion protein was prepared, and the fusion protein was affinity purified using anti-Flag antibody-coupled magnetic beads (Dynabeads M-280 Sheep Anti-Mouse IgG; Thermo Fisher Scientific). The fusion protein was eluted in 1 $\times$ phosphate-buffered saline (PBS) with 100 μ g/ml of 3 $\times$ FLAG Peptide (F4799; Sigma-Aldrich). Ypk3 kinase reactions were performed in kinase buffer (1 $\times$ PBS, 20% glycerol, 0.5% Tween-20, 4 mM MgCl₂, 10 mM DTT, and 600 μ M ATP) in a final volume of 30 μ l for 20 min at 30°C.

GST-Ypk1 was purified according to the cited protocol for GST-Ypk3 and eluted in the following buffer: 2.5 mM CHAPS, 25 mM HEPES, 150 mM KCl, 5 mM MgCl₂, and 1 mM DTT. Ypk1 kinase reactions were performed in kinase buffer (2.5 mM CHAPS, 25 mM HEPES, 150 mM KCl, 5 mM MgCl₂, 1 mM DTT, and 600 μ M ATP) in a final volume of 30 μ l for 20 min at 30°C.

Sucrose density gradient centrifugation-polysome profiling

Ribosomal extracts were prepared as described. Volumes of supernatant containing 10 OD₂₆₀ units were loaded on top of a 7–50% sucrose gradient prepared in gradient buffer (50 mM Tris-AcO,

pH 7.0, 50 mM NH₄Cl, 12 mM MgCl₂, 1 mM DTT). Gradients were ultracentrifuged for 3.5 h at 35,000 rpm, and the ultraviolet profile was recorded using an IG gradient fractionator.

Ribosome profiling

The *ypk1^{as} ypk2 ypk3^{as} RPS6^{wt}* and *ypk1^{as} ypk2 ypk3^{as} RPS6^{AA}* cells were grown in minimal synthetic medium with 2% Glc and 0.2% Gln at 30°C. Once they reached exponential growth, the cells were filtered and resuspended in minimal synthetic medium with 0.2% Gln without any carbon source. After a 1-h incubation at 30°C, cells were treated with 500 nM 1N-PP1 or dimethyl sulfoxide (DMSO). After a 5-min incubation with the drug/vehicle, the cells were refed with 2% glucose. After 20 or 40 min, cells were collected by filtration, resuspended in polysome buffer, and frozen rapidly in liquid nitrogen.

Ribosome footprint and total RNA libraries were prepared using the ARTseq Ribosome Profiling Kit (Yeast, RPYSC12116; Epicentre, Madison, WI) according to the manufacturer's protocol with slight changes. For the preparation of ribosome-protected RNA fragments, 20 μ g of nuclease-digested RNA samples were run on a 12% urea-polyacrylamide gel before rRNA depletion. The gel slices corresponding to fragment sizes of 28–30 nt were excised, and samples purified from these slices were further used for rRNA depletion and library preparation. The libraries were sequenced on an Illumina HiSeq2500 machine on the Genomics Platform (Geneva, Switzerland).

Initial processing and alignment of sequence reads

Adapter sequence (5'-AGATCGGAAGAGCACACGTCT-3') was trimmed from raw reads using Cutadapt (Martin, 2011). To reduce contamination from noncoding RNAs (such as rRNAs, tRNAs, small nuclear RNAs, and small nucleolar RNAs), trimmed reads were aligned to relevant sequences downloaded from the *Saccharomyces* Genome Database (Cherry *et al.*, 2012; downloads.yeastgenome.org/sequence/S288C_reference/rna/rna_coding.fasta.gz), using the Bowtie short read alignment program (Langmead *et al.*, 2009), and no more than three mismatches were allowed for these alignments ($\sim$ 3). Reads aligning to these sequences were removed from the data set for further analysis. Read-length distributions were analyzed to determine the most prevalent read lengths among the samples. Reads of <20 or >80 nt were discarded from further analyses.

The *S. cerevisiae* genome was obtained from the National Center for Biotechnology Information (GCA_000146055.2, dated December 2013; Engel *et al.*, 2014), to which reads were aligned using Bowtie (Langmead *et al.*, 2009). Parameters used for Bowtie alignments were -m 1 -v 3, that is, no more than three mismatches were allowed, and nonunique mappings were discarded. To evaluate the number of reads mapped on each gene, gene annotations were procured from the *Saccharomyces* Genome Database (Cherry *et al.*, 2012). Reproducibility of the data from all the replicates was evaluated by calculating Pearson's *r*. Specifically, the number of raw reads mapped on each gene for ribosomal profiling (two replicates) and total RNA-seq (three replicates) were used for calculating a Pearson's *r*.

Metagene analysis

Genes with at least 100 mapped alignments were used for the production of metagene profiles at the location of translation initiation and termination. To analyze the read density around the initiation and termination sites, 30 nt upstream and downstream of the start and stop codons were used, respectively.

Differential gene expression analysis

Normalization of mapped reads was performed by dividing the total number of reads mapped in a sample by the total number of the reads mapped for the sample containing the lowest number of mapped reads. This was performed separately for ribo-seq and RNA-seq samples. Ribosomal profiling reads were assigned to gene coordinates based on the inferred location of the A-site. The A-site codon of the elongating ribosome was assigned at 15 nt downstream of the 5' end of reads of length 28–32.

Average read counts for each gene for ribosomal profiling data (two replicates) and RNA-seq data (three replicates) were used for differential gene expression calculation. The read counts were compared in a pairwise manner between mutants and wild type and between different time points for the same treatments. The Z-score transformation of change fold log ratios was carried out as described previously (Andreev *et al.*, 2015). In brief, genes were grouped into bins of 300 based on the lowest expression signal (ribo-seq or RNA-seq or both across compared conditions). Variation in the distribution of change fold ratios was used to calculate the Z-score. Profiles for differentially regulated genes were constructed by calculating the average of normalized reads mapped on each position of the gene based on the inferred location of the A-site.

Assigning mapped reads to 5' leaders, 3' UTRs, and coding sequence regions

Genomic alignments were used to assign reads to 5' leaders, 3' UTRs, and coding sequence (CDS) regions. The 5' leaders and 3' UTR annotations were obtained from Nagalakshmi *et al.* (2008). Reads were assigned to 2755 genes that have both 5' leaders and 3' UTRs annotated, and genes with overlapping features were discarded for this analysis. End positions of 5' leaders were modified to 17 nt upstream of the first base of the CDS, and start positions of 3' UTRs were modified to 13 nt upstream of the last base of the CDS. Upstream ORFs were analyzed for differential expression using the same procedure as for CDS regions.

Frameshift analysis

Ribo-seq reads aligning to no more than 100 locations were used for generating ribosome density profiles for *S. cerevisiae* genes (ABP140, EST3, OAZ1) and the TY1 transposable element, which are known to use +1 frameshifting in their expression (Bekaert *et al.*, 2010). For analysis of triplet periodicity, only 29-nt-long reads were used. The fraction of reads aligning to each subcodon position of the A-site (using 15-nt offset from its 5' end) from all conditions was calculated to assess triplet periodicity.

ACKNOWLEDGMENTS

We thank Peggy Janich at the University of Lausanne for assistance in setting up the ribosome profiling protocol, Patrick O'Connor for valuable advice on ribosome profiling data analysis, and Kelly Tatchell for kindly providing plasmid DNA. We are very grateful to the Loewith laboratory members for providing us with plasmids, strains, and purified proteins and also for discussions. R.L. acknowledges support by the European Research Council, the Swiss National Science Foundation, and the Canton of Geneva. This work was partially supported by Science Foundation Ireland Grant 12/IA/1335 to P.V.B.

REFERENCES

Andreev DE, O'Connor PB, Zhdanov AV, Dmitriev RI, Shatsky IN, Papkovsky DB, Baranov PV (2015). Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome Biol* 16, 90.

Baker SH, Frederick DL, Bloecher A, Tatchell K (1997). Alanine-scanning mutagenesis of protein phosphatase type 1 in the yeast *Saccharomyces cerevisiae*. *Genetics* 145, 615–626.

Bekaert M, Firth AE, Zhang Y, Gladyshev VN, Atkins JF, Baranov PV (2010). Recode-2: new design, new search tools, and many more genes. *Nucleic Acids Res* 38, D69–D74.

Belandia B, Brautigan D, Martin-Perez J (1994). Attenuation of ribosomal protein S6 phosphatase activity in chicken embryo fibroblasts transformed by Rous sarcoma virus. *Mol Cell Biol* 14, 200–206.

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* 334, 1524–1529.

Bodenmiller B, Wanka S, Kraft C, Urban J, Campbell D, Pedrioli PG, Gerrits B, Picotti P, Lam H, Vitek O, *et al.* (2010). Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Sci Signal* 3, rs4.

Chauvin C, Koka V, Nouchi A, Mieulet V, Hoareau-Aveilla C, Dreazen A, Cagnard N, Carpentier W, Kiss T, Meyuhas O, Pende M (2014). Ribosomal protein S6 kinase activity controls the ribosome biogenesis transcriptional program. *Oncogene* 33, 474–483.

Cheng YL, Chen RH (2015). Assembly and quality control of the protein phosphatase 1 holoenzyme involves the Cdc48-Shp1 chaperone. *J Cell Sci* 128, 1180–1192.

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, *et al.* (2012). *Saccharomyces Genome Database: the genomics resource of budding yeast*. *Nucleic Acids Res* 40, D700–D705.

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, *et al.* (2010). The genetic landscape of a cell. *Science* 327, 425–431.

Doma MK, Parker R (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* 440, 561–564.

Du W, Halova L, Kirkham S, Atkin J, Petersen J (2012). TORC2 and the AGC kinase Gad8 regulate phosphorylation of the ribosomal protein S6 in fission yeast. *Biol Open* 1, 884–888.

Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS, *et al.* (2014). The reference genome sequence of *Saccharomyces cerevisiae*: then and now. *G3* 4, 389–398.

Fumagalli SA, Thomas G (2000). S6 phosphorylation and signal transduction. In: *Translational Control of Gene Expression*, ed. N Sonenberg, JWB Hershey, and MB Mathews, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 695–717.

Gaubitz C, Oliveira TM, Prouteau M, Leitner A, Karuppasamy M, Konstantinidou G, Rispal D, Eltschinger S, Robinson GC, Thore S, *et al.* (2015). Molecular basis of the rapamycin insensitivity of target of rapamycin complex 2. *Mol Cell* 58, 977–988.

Gonzalez A, Shimobayashi M, Eisenberg T, Merle DA, Pendl T, Hall MN, Moustafa T (2015). TORC1 promotes phosphorylation of ribosomal protein S6 via the AGC kinase Ypk3 in *Saccharomyces cerevisiae*. *PLoS One* 10, e0120250.

Gressner AM, Wool IG (1974). The phosphorylation of liver ribosomal proteins in vivo. Evidence that only a single small subunit protein (S6) is phosphorylated. *J Biol Chem* 249, 6917–6925.

Harigaya Y, Parker R (2010). No-go decay: a quality control mechanism for RNA in translation. *Wiley Interdiscip Rev RNA* 1, 132–141.

Hoon S, Smith AM, Wallace IM, Suresh S, Miranda M, Fung E, Proctor M, Shokat KM, Zhang C, Davis RW, *et al.* (2008). An integrated platform of genomic assays reveals small-molecule bioactivities. *Nat Chem Biol* 4, 498–506.

Hsieh AC, Liu Y, Edlind MP, Ingolia NT, Janes MR, Sher A, Shi EY, Stumpf CR, Christensen C, Bonham MJ, *et al.* (2012). The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* 485, 55–61.

Huber A, French SL, Tekotte H, Yerlikaya S, Stahl M, Perepelkina MP, Tyers M, Rougemont J, Beyer AL, Loewith R (2011). Sch9 regulates ribosome biogenesis via Stb3, Dot6 and Tod6 and the histone deacetylase complex RPD3L. *EMBO J* 30, 3052–3064.

Hutchinson JA, Shanware NP, Chang H, Tibbetts RS (2011). Regulation of ribosomal protein S6 phosphorylation by casein kinase 1 and protein phosphatase 1. *J Biol Chem* 286, 8688–8696.

Iadevaia V, Calderola S, Tino E, Amaldi F, Loreni F (2008). All translation elongation factors and the e, f, and h subunits of translation initiation factor 3 are encoded by 5'-terminal oligopyrimidine (TOP) mRNAs. *RNA* 14, 1730–1736.

- Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223.
- Johnson SP, Warner JR (1987). Phosphorylation of the *Saccharomyces cerevisiae* equivalent of ribosomal protein S6 has no detectable effect on growth. *Mol Cell Biol* 7, 1338–1345.
- Kabat D (1970). Phosphorylation of ribosomal proteins in rabbit reticulocytes. Characterization and regulatory aspects. *Biochemistry* 9, 4160–4175.
- Kamada Y, Fujioka Y, Suzuki NN, Inagaki F, Wulfschleger S, Loewith R, Hall MN, Ohsumi Y (2005). Tor2 directly phosphorylates the AGC kinase Ypk2 to regulate actin polarization. *Mol Cell Biol* 25, 7239–7248.
- Khatte H, Myasnikov AG, Natchiar SK, Klaholz BP (2015). Structure of the human 80S ribosome. *Nature* 520, 640–645.
- Kruse C, Johnson SP, Warner JR (1985). Phosphorylation of the yeast equivalent of ribosomal protein S6 is not essential for growth. *Proc Natl Acad Sci USA* 82, 7515–7519.
- Kuranda K, Leberre V, Sokol S, Palamarczyk G, Francois J (2006). Investigating the caffeine effects in the yeast *Saccharomyces cerevisiae* brings new insights into the connection between TOR, PKC and Ras/cAMP signalling pathways. *Mol Microbiol* 61, 1147–1166.
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.
- Levy S, Avni D, Hariharan N, Perry RP, Meyuhas O (1991). Oligopyrimidine tract at the 5' end of mammalian ribosomal protein mRNAs is required for their translational control. *Proc Natl Acad Sci USA* 88, 3319–3323.
- Lippman SI, Broach JR (2009). Protein kinase A and TORC1 activate genes for ribosomal biogenesis by inactivating repressors encoded by *Dot6* and its homolog *Tod6*. *Proc Natl Acad Sci USA* 106, 19928–19933.
- Loewith R, Hall MN (2011). Target of rapamycin (TOR) in nutrient signaling and growth control. *Genetics* 189, 1177–1201.
- Martin M (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17, 10.
- Meyuhas O (2008). Physiological roles of ribosomal protein S6: one of its kind. *Int Rev Cell Mol Biol* 268, 1–37.
- Michel AM, Choudhury KR, Firth AE, Ingolia NT, Atkins JF, Baranov PV (2012). Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res* 22, 2219–2229.
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344–1349.
- Nakashima A, Otsubo Y, Yamashita A, Sato T, Yamamoto M, Tamanoi F (2012). Psk1, an AGC kinase family member in fission yeast, is directly phosphorylated and controlled by TORC1 and functions as S6 kinase. *J Cell Sci* 125, 5840–5849.
- Niles BJ, Mogri H, Hill A, Vlahakis A, Powers T (2012). Plasma membrane recruitment and activation of the AGC kinase Ypk1 is mediated by target of rapamycin complex 2 (TORC2) and its effector proteins Slm1 and Slm2. *Proc Natl Acad Sci USA* 109, 1536–1541.
- Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* 6, 917–922.
- Pachler K, Karl T, Kolmann K, Mehler N, Eder M, Loeffler M, Oender K, Hochleitner EO, Lottspeich F, Bresgen N, et al. (2004). Functional interaction in establishment of ribosomal integrity between small subunit protein rpS6 and translational regulator rpl10/Grc5p. *FEMS Yeast Res* 5, 271–280.
- Pearce LR, Komander D, Alessi DR (2010). The nuts and bolts of AGC protein kinases. *Nat Rev Mol Cell Biol* 11, 9–22.
- Pende M, Um SH, Mieulet V, Sticker M, Goss VL, Mestan J, Mueller M, Fumagalli S, Kozma SC, Thomas G (2004). S6K1(-/-)/S6K2(-/-) mice exhibit perinatal lethality and rapamycin-sensitive 5'-terminal oligopyrimidine mRNA translation and reveal a mitogen-activated protein kinase-dependent S6 kinase pathway. *Mol Cell Biol* 24, 3112–3124.
- Reinke A, Chen JC, Aronova S, Powers T (2006). Caffeine targets TOR complex I and provides evidence for a regulatory link between the FRB and kinase domains of Tor1p. *J Biol Chem* 281, 31616–31626.
- Rispoli D, Eltschinger S, Stahl M, Vaga S, Bodenmiller B, Abraham Y, Filipuzzi I, Movva NR, Aebersold R, Helliwell SB, Loewith R (2015). Target of rapamycin complex 2 regulates actin polarization and endocytosis via multiple pathways. *J Biol Chem* 290, 14963–14978.
- Roux PP, Shahbazian D, Vu H, Holz MK, Cohen MS, Taunton J, Sonenberg N, Blenis J (2007). RAS/ERK signaling promotes site-specific ribosomal protein S6 phosphorylation via RSK and stimulates cap-dependent translation. *J Biol Chem* 282, 14056–14064.
- Ruvinsky I, Sharon N, Lerer T, Cohen H, Stolovich-Rain M, Nir T, Dor Y, Zisman P, Meyuhas O (2005). Ribosomal protein S6 phosphorylation is a determinant of cell size and glucose homeostasis. *Genes Dev* 19, 2199–2211.
- Shimada K, Filipuzzi I, Stahl M, Helliwell SB, Studer C, Hoepfner D, Seiber A, Loewith R, Movva NR, Gasser SM (2013). TORC2 signaling pathway guarantees genome stability in the face of DNA strand breaks. *Mol Cell* 51, 829–839.
- Thoreen CC, Chantranupong L, Keys HR, Wang T, Gray NS, Sabatini DM (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* 485, 109–113.
- Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 294, 2364–2368.
- Urban J, Souillard A, Huber A, Lippman S, Mukhopadhyay D, Deloche O, Wanke V, Anrather D, Ammerer G, Riezman H, et al. (2007). Sch9 is a major target of TORC1 in *Saccharomyces cerevisiae*. *Mol Cell* 26, 663–674.
- van Dam TJ, Zwartkruis FJ, Bos JL, Snel B (2011). Evolution of the TOR pathway. *J Mol Evol* 73, 209–220.
- Wanke V, Cameroni E, Uotila A, Piccolis M, Urban J, Loewith R, De Virgilio C (2008). Caffeine extends yeast lifespan by targeting TORC1. *Mol Microbiol* 69, 277–285.
- Warner JR, McIntosh KB (2009). How common are extraribosomal functions of ribosomal proteins? *Mol Cell* 34, 3–11.

Appendix 2: Specific reverse transcriptase slippage at the HIV ribosomal frameshift sequence: potential implications for modulation of GagPol synthesis

This project was in continuation with article “Stimulation of reverse transcriptase generated cDNAs with specific indels by template RNA structure: retrotransposon, dNTP balance, RT-reagent usage.” which has been published in Nucleic acid research (2017) 45(17):10143-10155. This appendix section has been published in Nucleic acid research (2017) 45(17):10156-10167.

To study the effect of aspects such as sequence context, RNA secondary structures and varying dNTP concentration in stimulating the reverse transcriptase slippage deep sequencing technique RNA-seq was used and I analysed the samples for the presence of insertions and deletions in the slippery site.

Specific reverse transcriptase slippage at the HIV ribosomal frameshift sequence: potential implications for modulation of GagPol synthesis

Christophe Penno^{1,2,3}, Romika Kumari¹, Pavel V. Baranov¹, Douwe van Sinderen^{2,3} and John F. Atkins^{1,2,4,*}

¹School of Biochemistry, University College Cork, Cork, Ireland, ²School of Microbiology, University College Cork, Cork, Ireland, ³Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland and ⁴Department of Human Genetics, University of Utah, Salt Lake City, UT 84112-5330, USA

Received March 13, 2017; Revised July 19, 2017; Editorial Decision July 21, 2017; Accepted July 24, 2017

ABSTRACT

Synthesis of HIV GagPol involves a proportion of ribosomes translating a U6A shift site at the distal end of the *gag* gene performing a programmed -1 ribosomal frameshift event to enter the overlapping *pol* gene. *In vitro* studies here show that at the same shift motif HIV reverse transcriptase generates -1 and +1 indels with their ratio being sensitive to the relative concentration ratio of dNTPs specified by the RNA template slippage-prone sequence and its 5' adjacent base. The GGG sequence 3' adjacent to the U6A shift/slippage site, which is important for ribosomal frameshifting, is shown here to limit reverse transcriptase base substitution and indel 'errors' in the run of A's in the product. The indels characterized here have either 1 more or less A, than the corresponding number of template U's. cDNA with 5 A's may yield novel Gag product(s), while cDNA with an extra base, 7 A's, may only be a minor contributor to GagPol polypeptide. Synthesis of a proportion of non-ribosomal frameshift derived GagPol would be relevant in efforts to identify therapeutically useful compounds that perturb the ratio of GagPol to Gag, and pertinent to the extent in which specific polymerase slippage is utilized in gene expression.

INTRODUCTION

Moloney Murine Leukemia retrovirus *gag* gene is 5' adjacent to, and in the same frame as its *pol* gene, whereas the Rous Sarcoma and HIV counterpart *pol* genes are in the -1 frame with respect to their *gag* genes. Following the evidence that synthesis of Moloney Murine Leukemia Virus GagPol involves in-frame readthrough of the *gag* termination codon (1,2), attention focused on Rous Sarcoma and

HIV GagPol synthesis. A ribosomal -1 frameshift event was shown to occur in the decoding of a sequence near the 3' end of their *gag* genes that overlaps their *pol* genes (3,4). The demonstration for such frameshifting was a major discovery, and much subsequent work has confirmed the involvement of a -1 ribosomal frameshift event in GagPol synthesis (5,6). The HIV-1 *gag-pol* shift site, U6A, is AAU-UUU-UUA-GGG with the codons shown in the *gag* reading frame and the underlining highlighting the first two codons following realignment to the -1 frame, the *pol* reading frame.

The initial evidence for the frameshifting occurring at the translation level involved cell-free synthesis experiments. The argument against a small proportion of the mRNA used in the mammalian cell-free protein synthesis experiments having *gag* and *pol* in the same reading frame was the reported finding that *Escherichia coli* cell-free translation did not yield the fusion product (3). However, later work showed that *E. coli* ribosomes perform counterpart -1 ribosomal frameshifting at the same sequence with 40% efficiency of mammalian ribosomes (5,7). This ruled out the argument against some GagPol potentially being derived from reverse transcriptase slippage followed by standard in-frame decoding. During the subsequent 28 years, the possibility of reverse transcriptase slippage at the retroviral *gag-pol* frameshift sites has not been investigated, though deep sequencing has revealed heterogeneity at several sites in retroviral sequences (8). Numerous studies, some ongoing, seek compounds that specifically modulate HIV ribosomal frameshifting efficiency (9–15). The therapeutic goal of this work is expected to be dependent not only on their specificity and effectiveness in modulating ribosomal frameshifting, but also on whether a proportion of GagPol is derived from the alternative process of specific polymerase slippage. The present study is confined to *in vitro* experiments, but it points to the need for further work to assess the origin(s) of GagPol for retroviruses whose *pol* is in the

*To whom correspondence should be addressed. Tel: +353 214 205 420; Fax: +353 238 855 147; Email: atkins@genetics.utah.edu or j.atkins@ucc.ie

© The Author(s) 2017. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

–1 frame with respect to their *gag* coding sequence. In addition to practical considerations, historical aspects of the issues faced by one of the two groups that initially studied retroviral ‘frameshifting’ (16), the wider issue is the extent of functional utilization of specific polymerase slippage in gene expression. Early studies on mechanistically distinct utilization in the expression of paramyxoviruses (17,18), the filovirus Ebola (19), and a bacterial DNA polymerase gene (20) have been broadened by several studies including those on bacterial IS elements (21–23), Ebola virus (23–26), several secretion apparatus genes of *Shigella flexneri* (27–29), *Citrobacter rodentium* and *Yersinia pseudotuberculosis* (30), certain high genomic AT content endosymbiotic bacteria of insects (31,32) and a cell movement gene of the major plant virus family of Potyviruses (33–36). Despite these discoveries, the present work points to the need for further work on the relatively understudied phenomenon of specific transcriptional slippage in cellular and viral gene expression. One aspect relevant to searches for unknown cases of slippage utilization is context effects that may inhibit the propensity for polymerase slippage. This is investigated here by analysis of the GGG sequence 3′ adjacent to the ribosomal frameshift site and previously shown to be important for the ribosomal frameshifting (37). Another relevant aspect is potential stimulatory effects of ‘partial road-blocking’ RNA structures ahead of polymerase that could stimulate specific slippage at sites including those not recognizable as likely shift-prone. Such structures can stimulate reverse transcriptase slippage at a nearby 5′ sequence at which it would not otherwise detectably occur (38). Retroviral *gag-pol* ribosomal frameshifting is stimulated by RNA structural recoding elements 3′ of the shift site (16). This raises the possibility that the same structures, or components thereof, may also act to stimulate reverse transcriptase slippage.

Studies addressing potential, or demonstrably, functionally utilized specific polymerase slippage build on earlier fidelity investigations of different polymerases, including of reverse transcriptases. There have been a number of such studies with HIV-1 RT. One of the notable findings was several fold higher fidelity with RNA than DNA templates (39), and another discerned the effects of dNTP pool imbalances on frameshift fidelity with DNA templates (40). This was also studied in a comparative analysis of slippage-mediated product indel formation by several different reverse transcriptases (38). Furthermore, substrate imbalance is also well known to influence *E. coli* and *S. cerevisiae* DNA-dependent RNA polymerase slippage (23). The degree of influence of the ratio of the cognate substrate for the slippage site to that of the substrate base for the 5′ adjacent base, also features in the present study of specific HIV RT slippage at a site region that is also responsible for the ribosomal frameshifting that mediates GagPol synthesis. In distinction to these earlier studies, the present study involves the natural context of the relevant sequence, since it permits potential limiting or stimulatory features to be discerned. Also the deep sequencing employed permits a more refined analysis.

MATERIALS AND METHODS

RNA template constructs

Double-stranded DNA templates, for T7 RNA polymerase, were generated by an overlapping DNA oligonucleotide strategy (Supplementary Table S1) and 3′ extended with Vent DNA polymerase. Preparation of RNA template (200 nt HIV) was generated by T7 RNA polymerase *in vitro* in 20 µl reaction volumes with a Promega kit (T7 RiboMAX™ large scale RNA production system) according to the manufacturer’s instructions. Reaction products were treated with 2 units of DNase I turbo (Ambion) in 50 µL reaction volumes at 37°C for 30 min. 500 µl Trizol (Invitrogen) was added to the reaction and incubated for 2 min. Then, 100 µl chloroform was added and incubated for 5 min at room temperature, before centrifugation, 12 000 × g for 15 min at 4°C. The upper aqueous phase was extracted, and an equal volume of 100% ethanol was mixed with it before it was loaded on a silica column (Anachem PCR purification kit). After centrifugation, 12 000 × g for 30 s, the column was washed with 600 µl of a low salt wash buffer (Anachem’s ‘PCR clean up’ kit wash buffer to which ethanol was added to 80% concentration). The column was then treated with DNase I turbo, 10 units in a 100 µl volume for 1 h incubation at 37°C. This was followed by 600 µl of a high salt wash buffer (Anachem’s ‘PCR clean up’ kit wash buffer to which ethanol was added to 40% concentration) and centrifugation. It was then washed twice, via centrifugation, with the low salt buffer, 600 and 300 µL. The final centrifugation was at 12 000 × g for 2 min to dry the column before elution with 100 µl of RNase free water.

Chemically synthesized RNA templates and DNA oligonucleotides were from IDT-DNA (Supplementary Table S1).

Reverse transcription

HIV-1 reverse transcriptase enzyme was purchased from Abcam. RT reactions involved a pre-annealing step of the RNA template (100 ng): DNA Primer (2 pmol) (Supplementary Table S1), in 1× SuperScript™ III buffer (50 mM Tris–HCl, pH 8.3 at 25°C, 75 mM KCl, 3 mM MgCl₂, 5 mM DTT), in the presence of the dNTP substrate (with the specific concentrations of each indicated in the main text) and, where present, antisense (100 pmol), in a 10 µl reaction volume. With the HIV-1 200 nt cassette, the annealing step was at 95°C for 30 s with a 10% temperature decrease (from 95 to 16°C) every 30 s. With RT reactions involving the HIV-1 chemically synthesized RNA cassette combinations (Figure 3), the annealing step was at 65°C for 5 min before chilling on ice. Then, a 10 µl mix (1× SuperScript™ III buffer, 4 units HIV-1 RT enzyme, 20 mM DTT) was added and incubated at 37°C for 50 min, followed by 85°C for 5 min.

Polymerization chain reaction

Each specific cDNA was amplified using the corresponding set of forward and reverse primers (Supplementary Table S1). Standard PCR reactions were 50 µL volume and contained: 1× Thermo buffer (Biolabs), 2 µl cDNA or 4 nM DNA template oligo, 200 µM each dNTP (Biolabs),

500 nM each specific primer, and 0.8 unit Taq DNA polymerase (Biolabs). The PCR cycle was: denaturation at 94°C for 5 min, then 25 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 30 s and elongation at 72°C for 30 s. This was followed by a final elongation at 72°C for 1 min.

Sample preparation for deep sequencing

A first round of amplification of the cDNA product was with a set of forward and reverse primers with an NGS adapter sequence at their 5' ends and sequence complementary to the cDNA product at their 3' ends. This set of primers included the common forward primer, HIV_NGS_F1697, whose 3' sequence is complementary to the 3' end of the cDNA. The reverse primer used for set I ('long RNA: long cDNA') was HIV_NGS_R1695, and for sets 2 and 3 ('long RNA: short cDNA' and 'short RNA: short cDNA', respectively) it was HIV_NGS_R1696. PCR products were then purified using Zymo-Spin™ I column (Zymo Research) with buffers from the PCR clean-up kit (Roche). Subsequently, a limited cycle PCR using Nextera XT indices (Illumina) was performed to incorporate sequencing adaptors and dual-index barcodes to each PCR product. Products were then purified with Ampure XP beads (Beckman Coulter), quantified using the Qubit fluorimeter (Thermo Fisher) and pooled at an equimolar concentration. The mix of 80 samples was sequenced on a MiSeq (Illumina) using a MiSeq Reagent Nano Kit (300 cycles) with V2 chemistry (Illumina).

Deep sequencing analysis

The paired-end cDNA libraries were obtained for 80 samples with approximate yield of ~10 000/reads per sample. The reads were trimmed using Cutadapt (41). Adapter sequences were clipped from forward (5'-CTGTCTCTTATACACATCTCCGAGCCACGAGAC-3') and reverse (5'-CTGTCTCTTATACACATCTGACGCTGCCGACGA-3') reads. Trimmed reads were aligned to the corresponding reference gene sequence using bowtie2 short read alignment program (42). Local alignments of paired end reads were carried out with default parameters. Sorted BAM alignment files were obtained using SAMtools version 1.3.1 (43) which were used for prediction of SNPs and INDELs. Variant calling was performed using SAMtools mpileup. Parameters used for mpileup predictions were (-g -uf -output-tags DP, AD -L 1000000 -d 1000000) which generated output in BCF format. BCFtools (44) view option was used to convert variant calls to the Variant Call Format (VCF). For graphical analysis of substitution rates in the slippery sites custom script based on seqLogo (45) R package was used to plot sequence logos. Heatmaps were generated using MeV (Multiple Experiment Viewer) (46).

RESULTS

Strategy for testing the effect of a 200 nt HIV1 *gag-pol* cassette

The sequence 3' of the 5'-U6A-3' HIV *gag-pol* frameshift site can form alternative structures. Though most work on ribosomal frameshifting has been done with cassettes with

65 nt 3' of the shift site (47), additional flanking sequences may form other relevant structures (48–51). This was tested using a 140 nt cassette with 71 nt 3', and 63 nt 5', of the shift site. Structures similar to those in template RNA may also form in the nascent cDNA and are potentially relevant to reverse transcriptase slippage. Conceptual precedent comes from the finding that formation of a nascent RNA stem-loop structure can be a strong stimulator for DNA-dependent RNA polymerase slippage (23). To assess potential HIV RT enzyme slippage on the *gag-pol* U6A frameshift motif with putative stimulation by the flanking sequence context, the 140 nt RNA used in the ribosomal frameshifting studies plus 60 nt for primer annealing (total 200 nt), sequence was generated *in vitro* by T7 RNAP (Materials and Methods). This RNA had 70 nt 5' of the shift site, the U6A shift site and 124 nt 3' of it (Figure 1).

The cDNA products were amplified by polymerase chain reaction (PCR) using specific primers. As a control for DNA polymerase fidelity, a counterpart PCR experiment was performed with a chemically synthesized DNA whose sequence was the reverse complement of the 200 nt RNA template used above. The resulting RT-PCR and PCR products were quantitatively analyzed by next-generation sequencing (NGS) (Materials and Methods). For each sample, single nucleotide polymorphism (SNP) and nucleotide insertion/deletion (indel) frequency were determined at each nt location by comparison to their respective cDNA reference sequence. For indel analysis at the 5'-U6A-3' *gag-pol* derived sequence, NGS reads were selected only when the sequence derived from the RNA template run of U's was free of substitutions.

Effect of the HIV1 *gag-pol* 140 nt frameshifting element RNA on RT slippage

With an A6 slippage motif HIV RT slippage directionality is sensitive to an imbalance of the concentrations of the cognate substrate dNTPs specified by a template slippage motif and the base 5' adjacent to the motif (38). To ascertain if counterpart effects occur with 5'-AU6A, an imbalanced ratio of dATP, cognate for the U6 tract, and of dTTP, cognate for the template base 5' adjacent to the U6, was tested. RT reactions were performed with 3 dNTP ratio conditions for the substrates: (i) all 4 dNTPs at 500 μ M, (ii) dTTP 5 μ M, other 3 at 500 μ M, (iii) dATP at 5 μ M other 3 at 500 μ M.

At the 5'-U6A-3' *gag-pol* derived sequence, the NGS data shows a correlation of indel frequency for the cDNA base A with the [dATP]:[dTTP] ratio in the RT reactions. With equimolar dNTP, base omission is 2.1% and base addition 1.4%; with the [dATP]_{500 μ M}: [dTTP]_{5 μ M} ratio, base omission is 1.7% and addition 2.0%; with the [dATP]_{5 μ M}: [dTTP]_{500 μ M} ratio, base omission is 2.6% and addition is 1.1%. The control for DNA polymerase slippage shows a lower frequency (1.2%) of base omission, and a background level (0.1%) for base addition (Figure 2). These NGS results show that HIV RT undergoes specific slippage-mediated base addition (especially with [dATP] > [dTTP]) and omission (especially with [dATP] < [dTTP]). Importantly, the natural *gag-pol* nucleotide context 5' and 3' to the U6 motif does not abolish the slippage propensity of the HIV-1 RT enzyme.

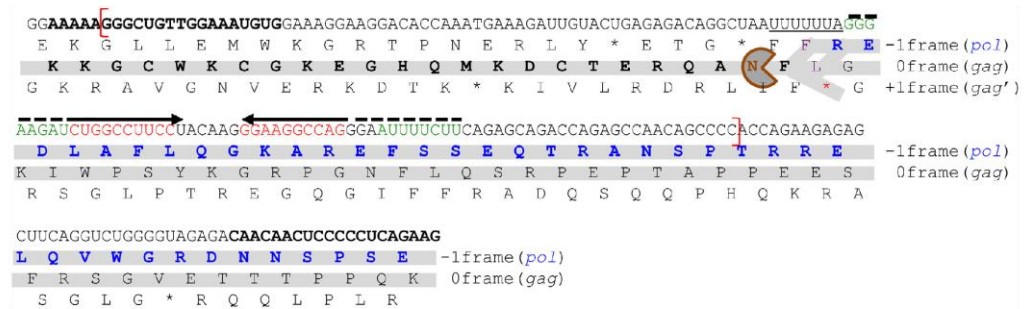


Figure 1. The HIV *gag-pol* ribosomal frameshift site also mediates HIV reverse transcriptase slippage. The *gag-pol* frameshift site, UUUU.UAA (underlined) is shown with 30 nt 5' of the shift site and 48 nt 3' (for which one of the potential pairing possibilities is indicated in red and green). The 5' and 3' borders of an RNA sequence involved in a potential and relevant large alternative structure revealed in a prior study (48), are indicated by red brackets. RNA three translation is shown with the 0 and -1 reading frames being those of *gag* and *pol* respectively. The highlighted grey box indicates the amino-acid sequence of the ribosomal frameshift products and the corresponding sequence specified only by the original zero frame. The 5' and 3' nt sequences in bold indicate the nt annealing region of the primers used in the (RT)-PCR reactions.

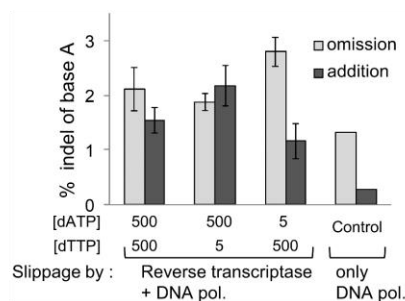


Figure 2. NGS analysis of indels in the run of A's derived from the 200 nt HIV sequence. Frequency of insertion, or omission of base A, derived from cDNA generated at three different relative dNTP concentration conditions. The standard deviation is for 4 independent reverse transcription reactions. The control reflects only DNA polymerase slippage.

The other type of fidelity error, base mis-incorporation was next analyzed to enable its comparison with indel formation. The SNP frequency is less than 0.5%. Its frequency at locations corresponding to each base in the cDNA run of A's (5'-A₁A₂A₃A₄A₅A₆-3', with the number in subscript indicating the position of the base in the homopolymeric run) is similar whether all dNTPs are equimolar, or with the [dATP]_{500μM}:[dTTP]_{5μM} ratio. In contrast, with the [dATP]_{5μM}:[dTTP]_{500μM} ratio, the SNP frequency strongly correlates with the cDNA A-tract base position. It increases from 1.3% corresponding to A₁ to 7.1% at A₆ with a preference for mis-incorporation of the base G (from ≈ 1.1% at A₁ to 4.6% at A₆) (Supplementary Results). These results show that depending on the relative dNTP concentrations, SNP errors can be less significant than indel errors for the 5'-U6A-3' derived sequence. [A graphical comparison of the frequency of indels and SNPs in the cDNA reference sequence 5'-TCCCTAAAAAATTAGCCTC-3' for 4 repeated experiments is shown in Supplementary Dataset

S1. A heatmap representation of the SNP frequency is also shown for 25 nt 5' to the A-tract of the reference cDNA (Supplementary Figure S3). Supplementary Dataset S1 and Supplementary Figure S3 also show the repeat experiments for the constructs described below]. Responsiveness to alterations of the ratios of the different dNTPs is relevant to discrimination against effects of potential infidelity of the T7 RNA polymerase used to generate the template for this work.

The same derived RT-PCR and PCR products used in the NGS analysis were also analyzed by Limited Primer Extension (LPE) (Supplementary Information; Supplementary Figure S1, panel A). This has the advantage of allowing more conditions to be quickly and cheaply tested. Though it permits analysis of only a specific part of the template sequence, the segment analyzed here is the most relevant. LPE analysis yielded qualitatively similar results to NGS for both base indel and SNP, for products derived from the 5'-U6A-3' containing sequence (Supplementary Results and Supplementary Figure S1, panel B). To explore a possible template road-blocking effect on HIV RT reverse transcriptase, an experiment was performed with a 10 nt antisense RNA complementary to the HIV RNA sequence 5' adjacent to the U6-tract. LPE analysis revealed that presence of the antisense favors slippage-mediated base addition (Supplementary Results, Supplementary Figure S1, panel B) (LPE analysis was used for this characterization since it was also used for multiple reaction conditions with the templates described below).

Strategy for testing nucleic acid scaffolds for RT reactions with chemically synthesized RNA templates

Chemically synthesized RNA avoids potential complications due to T7 RNA polymerase infidelity. We explored slippage with synthetic RNAs close to the maximum practical length of single synthesis RNA (84 nt). The two RNAs contain 30 HIV nt 5' of the shift site, the U6 tract, and, in RNA₁₆₅₄, 48 HIV nt, and in RNA₁₆₅₅, 25 HIV nt. 1654 is referred to below as 'WT long RNA' and 1655 as 'WT

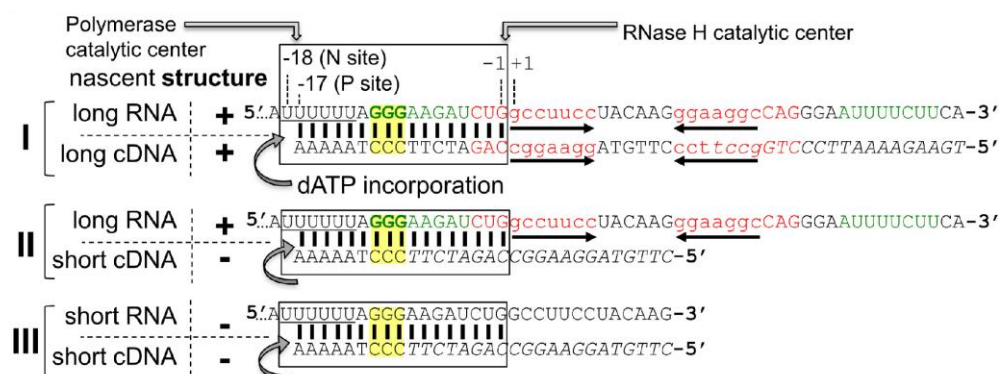


Figure 3. HIV *gag-pol* RT slippage stimulators. Strategy to assess possible stimulatory or inhibitory effects on slippage by potential structure formation in the RNA template and/or in its derived nascent cDNA, during reverse transcription of the U6A motif. Sequence set I shows the chemically synthesized RNA ('long RNA') and a potential for stem-loop formation (red and green). When the RT transcribes as far as the 5' U of the U6 tract (to generate 'long cDNA'), the parts of the RNA or cDNA free behind the 18 bp hybrid, and potentially free to participate in stem-loop formation, are indicated by inverted arrows. Sequence set II shows the same 'long RNA' template reverse transcribed from a primer whose sequence is in italics at the 5' end of the 'short cDNA'. Sequence set III has 'short RNA' without the potential for structure formation involving the sequence underlined by inverted arrows in sets I and II. The RNA triplet GGG present in both 'long and short RNA' templates, which was substituted by the sequence CCC (designated 'mut'), is highlighted in yellow.

short RNA'. In 'WT long RNA' at least 35 nt 3' of the U6 tract have potential to be involved in stem-loop structure formation. In 'WT short RNA' the potential for stem-loop structure is limited (Figure 3).

HIV RT slippage may be stimulated by a structure in the nascent cDNA that is a counterpart of an RNA template structure. To investigate this possibility two different sites were used for priming, only one of which is in 'WT short RNA'. The primer for this site, R_1648, yields a 25 HIV nt nascent cDNA sequence 5' of the A-tract, 'short cDNA', with limiting stem-loop structure formation (Figure 3). The additional primer for 'WT long RNA', (R_1645) yields a 48 HIV nt nascent cDNA sequence 5' of the A-tract, 'long cDNA', with the potential for structure formation (Figure 3). Three different 'RNA template: Primer' combinations were tested (Figure 3).

The first combination involved 'WT long RNA: long cDNA' with both having structure potential, (Figure 3, combination I). The second combination involved 'WT long RNA: short cDNA' with only the template having relevant structure potential (Figure 3, combination II). The third combination involved 'short RNA: short cDNA' with neither having structure potential (Figure 3, combination III).

Due to the effect of dNTP imbalance on RT slippage directionality (resulting in base addition or omission), RT reactions were performed with 9 dNTP concentration conditions. In all conditions dCTP and dGTP were constant at 500 μ M. The concentrations of dATP (the substrate specified by the U6 motif) and dTTP (the substrate specified by the RNA template base 5' adjacent to the U6) were either at 5, 50 or 500 μ M.

Run of U's derived indels studied with several template:primer combinations

For each sample, NGS reads were selected for slippage analysis only when sequence derived from the RNA template run of U's was free of substitutions. For the 3 'RNA template: Primer' combinations (Figure 3), sequence derived from the run of U's exhibits both types of indel due to slippage in both directions, and with a higher frequency of base omission than base addition (Figure 4). This means that potential to form a small cDNA structure adjacent, or almost adjacent, to the hybrid does not influence product indel formation. Also, potential by the 'long RNA' template containing the GGG to form structure does not, at least, significantly influence the efficiency of indel generation. However, specific dNTP ratios are seen to influence the slippage. With cognate dATP at any of 5, 50 or 500 μ M concentrations and the dTTP concentration increasing from 5 to 50 to 500 μ M, the frequency of specific base omission increased from 0.5% up to 2.5%. With the same substrate concentrations, the efficiency of base addition is less than 0.5% except with the highest dATP:dTTP ratios (Figure 4). Qualitatively similar results were also obtained by LPE analysis (Supplementary Results and Supplementary Figure S2A and C).

Does the highly conserved GGG 3' adjacent to the U6A motif influence RT slippage?

The highly conserved GGG triplet 3' adjacent to the U6A motif (37) is in the RNA:DNA hybrid when HIV RT reverse transcribes the U-tract. This GGG is potentially relevant to hybrid stability and realignment.

We used a similar strategy as in the previous section to test for possible effects of substituting GGG with CCC (Figure 3, highlighted in yellow). The three combinations, 'mut

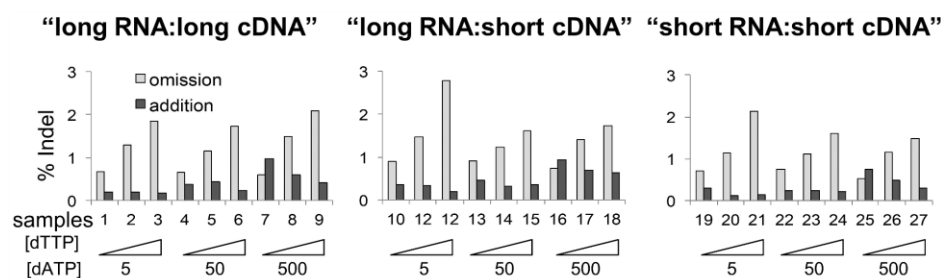


Figure 4. NGS analysis of 'WT' HIV cDNA derived sequence: indel distribution. Frequency of base addition, or omission, in the cDNA run of A's. This is shown for the 3 'RNA template: Primer' combinations indicated above each graph. Sample identity and dNTP concentration ratio conditions are indicated at the bottom. Indel results are from representative samples 1–27 (SI Results & SI Dataset 1).

long RNA: long cDNA', 'mut long RNA: short cDNA' and 'mut short RNA: short cDNA' were tested with nine dNTP concentration conditions. In all conditions dCTP and dGTP were at 500 μM. The concentrations of dATP (specified by U6) and dTTP (specified by the template base 5' adjacent to U6) were either at 5, 50 or 500 μM. The 27 different RT-PCR products derived from cDNA generated by the nine dNTP concentration conditions for each of the 3 'RNA template: Primer' combinations, were analyzed by LPE (Supplementary Results and Supplementary Figure S2B and D). Nine of these different RT-PCR products were also analyzed by NGS; they correspond to the 3 'RNA template: Primer' combinations tested with the three following dNTP ratio conditions (i) all four dNTPs at 500 μM, (ii) dTTP 5 μM, other three at 500 μM, (iii) dATP at 5 μM other three at 500 μM.

With the two combinations, 'mut long RNA: short cDNA' and 'mut short RNA: short cDNA', slippage-mediated base addition occurs at a higher frequency than with their WT counterpart (Figure 5A). With equimolar dNTP concentration, base addition is ~3 times more efficient with the mut combination compared with the WT equivalent. Interestingly, the efficiency of base addition of the mut combinations at equimolar dNTPs is even higher than the efficiency of base addition for the corresponding WT combination at the dNTP concentration ratio which favors base addition, i.e. [dATP]_{500μM}: [dTTP]_{5μM}. In contrast, mut combinations exhibit a similar frequency of product base omission as their WT counterparts (Figure 5A).

With the mut sequence, base substitutions are more prominent at each position of the cDNA run of A's (5'-A₁A₂A₃A₄A₅A₆-3'), especially with 5 μM dATP and the other three dNTPs at 500 μM. This SNP difference between mut and WT can be seen with sequence logo illustrations (Figure 5, panel B and Supplementary Dataset 1). Compared to the 'WT RNA: cDNA' combinations, the corresponding mut combinations, there is a marked SNP increase at cDNA base position A₂ and this high proportion is maintained at a similar level for positions A₃–A₆. In contrast, with the corresponding WT combination, there is a gradual increase of substitutions from positions A₁ to A₆ in the cDNA. This result suggests that mutating the conserved RNA template GGG triplet to CCC, influences RT

polymerase fidelity at least at the template base position U₂. As judged only from mutating the conserved RNA template GGG to CCC, the identity of the bases at this position in WT serves to constrain the frequency of base A errors in the run of A's in the product cDNA.

DISCUSSION

Reverse transcriptase slippage is shown to mediate generation of DNA which when transcribed would yield RNA with the sequence AAU-UUU-UUU-AGG. The underlined U is absent in the product, AAU-UUU-UUA-GG, generated without slippage. Standard translation of the slippage derived RNA would yield Asn-Phe-Phe-Arg, identical to the minor product from -1 ribosomal frameshifting (Figure 6) [two GagPol products are generated by -1 ribosomal frameshifting, 80% with Asn-Phe-Leu-Arg and 20% with Asn-Phe-Phe-Arg specified by the frame junction sequence (4,5,52)]. Despite all the work on retroviral ribosomal frameshifting, it is only relatively recently that a product from ribosome shifting into the alternative +1 frame has been detected. The frameshifting involved was detected with reporter cassettes in reticulocyte lysate translation experiments, and the shift involved is -2 rather than +1 (Figure 6). It was considered likely insignificant as its occurrence during HIV translation would lead to a Gag product with the C-terminal sequence Asn-Phe-Leu-stop (termination being specified by the +1 frame UAG above) (53). Reverse transcriptase slippage is here shown to yield DNA sequence whose derived RNA with the sequence AAU-UUU-UAG would be translated Asn-Phe-stop, i.e. the same as the -2 ribosomal frameshift derived product except for the absence of one amino acid, Leu (Figure 6). WT HIV Gag is proteolytically cleaved after its amino acid Asn encoded by the AAU whose U is the 5' U of the homopolymeric component of the ribosomal frameshift site. In WT Gag multiple amino acids follow Asn and presumably cleavage does not occur when only two AA are C-terminal to Asn. If so, and if the novel termination product is synthesized in virus-infected cells, its generation could be significant. There are five sequential cleavage sites (54–59) and alteration of the 3rd cleavage site may have an effect on the fourth and fifth proteolytic cleavages of the Gag precursor protein. Irrespective, its potential functional role merits investigation, in

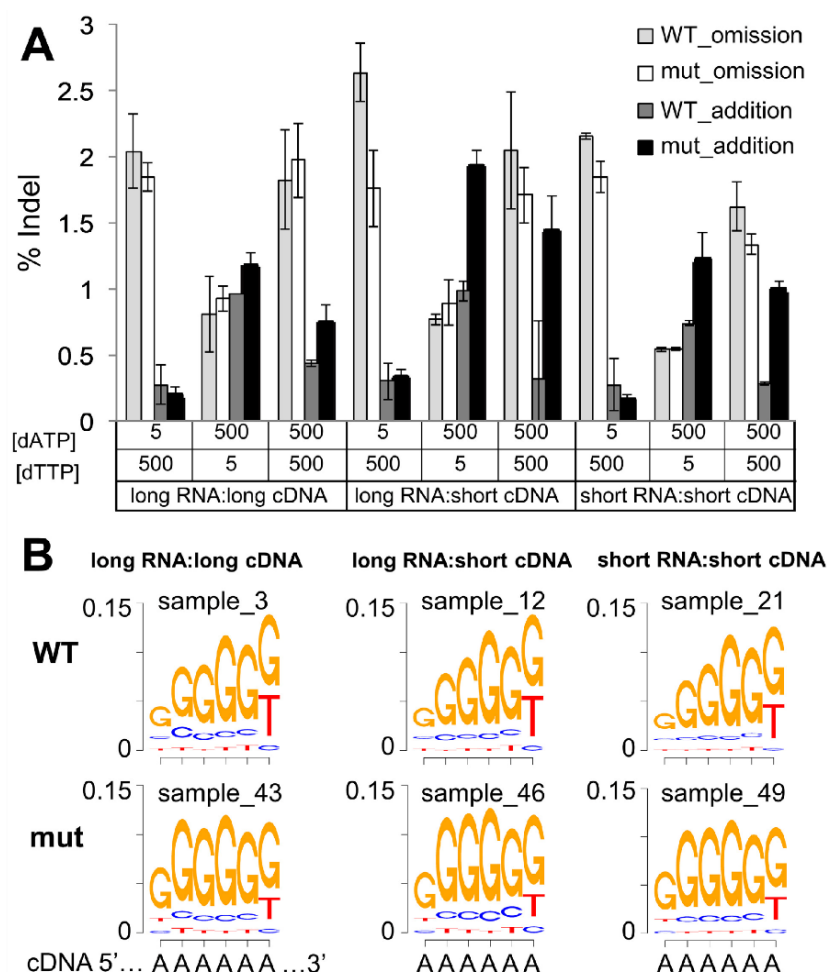


Figure 5. Effect of the nt triplet one nt 3' to the run of U's on HIV RT fidelity. (A) Frequency of slippage-mediated base addition, or omission, for the WT and mut combinations tested with RT reactions involving three different dNTP concentrations conditions. SEM error bars are indicated for two independent experiments. (B) Sequence logo of the substituted base in the run of A's. The cognate substrate was omitted to enhance visualization of the substituted bases in the run of A's in the cDNA. The identity of the 'RNA template: Primer' combination and the dNTP concentration conditions, are indicated above the WT (top row), and below the corresponding mut combination (bottom row). The name of the representative sample is indicated; its corresponding dataset is in SI Dataset S1.

part because of possible relevance to resistance mechanisms against drugs targeting the protease involved.

Protease mutants

In some protease resistance mutants there is a compensatory C-terminal amino acid substitution of the NC-SP2-P6 Gag-derivative product (60). The mutations involved may enhance reverse transcriptase slippage potential. For instance, at the proteolytic cleavage site of NC/SP2, in the mutation that causes an A431V sub-

stitution, the codon GCU, 5' adjacent to the *gag* sequence AAU.UUU.UUA.GGG that contains the ribosomal frameshifting site, is changed to GUU (60). This mutation may increase potential formation of a weak GNRA capped stem-loop 5' adjacent to the U-tract (Supplementary Figure S6). At the proteolytic cleavage site for SP2/P6, in the mutation that causes the L449F substitution, the codon CUU (bold) in the *gag* sequence AAU.UUU.CUU.CAG is replaced by one of the Phe codons, which are UUU and UUC (61,62). Either would

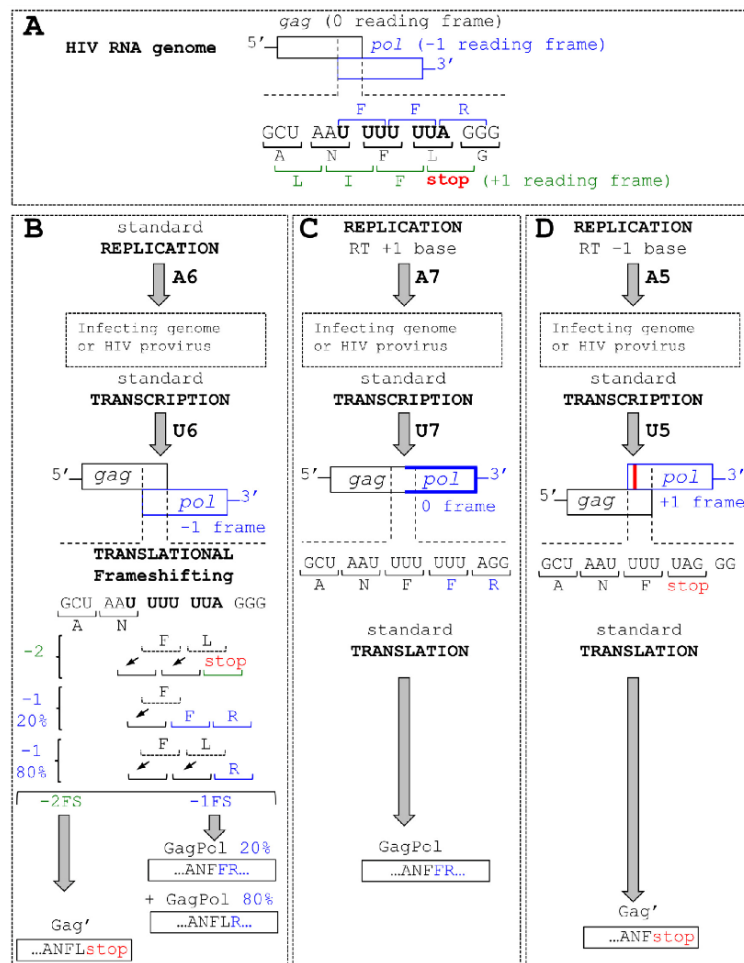


Figure 6. Expression of HIV *gag-pol*. (A) Genetic organization of the *gag* and *pol* open reading frames. With the nomenclature in common use, standard translation of RNA with an extra base is the counterpart of -1 ribosomal frameshifting on WT mRNA (and from the protein product perspective, base omission is the counterpart of +1 ribosomal frameshifting). The *gag-pol* ribosomal frameshift sequence is indicated in bold with the corresponding amino-acid sequences indicated for the *gag* reading 0 frame (black), the *pol* reading -1 frame (blue) and the +1 reading frame (green). (B) Ribosomal frameshifting yield to Gag' and GagPol products. The alternative amino-acid sequences in prior analyses of -1 ribosomal frameshift products with the major transframe encoded Phe-Leu containing GagPol product being 80% of the total and its Phe-Phe counterpart constituting the remaining 20% (16). Brackets with dashed lines under codons indicate initial tRNA anticodon pairing prior to realignment. Potential transcription slippage counterparts due to RT slippage-mediated one base addition (C) or one base omission (D) are considered in the Discussion.

create the potential for an additional slippage motif with a 5' sequence with potential to form an RNA structure that could act as a 'partial roadblock' RT slippage stimulator.

Substrate ratios

dUTP incorporation by RT HIV has a protective affect against host chromosome integration (63). The effect of dUTP on reverse transcriptional slippage was not explored

in the present manuscript. However, this could also be relevant with the natural target of HIV, as CD4 T cells and macrophages contain a 1:4 ratio of dUTP:dTTP in contrast to other cells (1:100 ratio) (63 and refs therein)

dNTP depletion is part of the host's response to viral infection. The deoxynucleotide triphosphohydrolase SAMHD1 depletes dNTP pools which interferes with RT function in myeloid cell (64). Lentiviruses, such as HIV-2, have developed a *vpx* gene to counteract this effect (65-67).

[Contrastingly, in many colon cancer cells SAMHD1 is directly debilitated by mutants in its coding sequence with consequences for dNTP levels and ratios (68); mutator phenotypes of colon cancer cells due to DNA polymerase mutants also affects dNTP levels and ratios (69).] Future work is needed to explore the potential effect on RT HIV slippage of the relative intracellular dNTP concentration in virus-infected cells. The potential selective advantage of base insertion or omission in the *gag-pol* region of chromosomally integrated HIV virus sequence may also be relevant. Such an investigation could involve *gag-pol* sequencing of specific tissues of latent phase infected patients.

Slippage versus ribosomal frameshifting: hybrid lengths

A big distinction between ribosomal frameshifting and polymerase slippage is that with frameshifting at each position a much shorter template sequence is involved in base pairing. The RNA:cDNA hybrid in reverse transcriptase is ~18 bp (70,71). It extends from the RNase H catalytic center to the polymerase catalytic center. The length of the RT slippage U6 motif investigated here is much shorter than the length of the hybrid, suggesting that only a portion of the 18 bp are involved in the re-alignment. RT modelling to identify locations that could accommodate potential hybrid bulge formation (due to an extra helical base), pointed to such potential only corresponding to a hybrid region whose cDNA component is 5–6 bp from its 3' end (72). However, it is possible that the portion of the hybrid between the slippage motif and the RNase H catalytic center may also affect slippage, perhaps due to its helical topology. A constraint on the latitude of sequence variation relevant to both frameshifting and slippage is retention of an Asn codon (overlapping the 5' end of the shift motif) since Asn is important for proteolytic cleavage (55,73). While this restricts the number of possible heptanucleotide frameshift sequences that could have been utilized at this site, there are still some that are unlikely to mediate reverse transcriptase slippage.

Other features of comparative importance of the shift/slippage cassette are also noteworthy. The highly conserved triplet, GGG, separated by 1 base from the 3' end of the U-tract, is significant for frameshifting. This sequence has generally been considered not to have reached the ribosomal A-site when frameshifting occurs. However, a recent study has proposed that it is in a (perhaps distorted) A-site when the ribosomal frameshifting occurs (37). In the present work when GGG was substituted with CCC, efficient RT slippage still occurred but it was qualitatively different. There was also an effect on product base substitutions and further work may reveal relevance of the GGG for shift/slippage site preservation. Formally structural RNA recoding signals can potentially stimulate frameshifting or slippage due to effects on their initial unwinding/unpairing or when the sequence involved (re)forms structure on or after, exit from the ribosome/polymerase. With the WT GGG just discussed, structure involving at least most of this ribosomal frameshift relevant sequence only minimally affects slippage. However, when the GGG is substituted with CCC, the further 3' ribosomal frameshift relevant sequence has also a marked effect on slippage. However, to

avail of the advantages of using homogeneous chemically synthesized—though necessarily shorter - template RNA, slippage studies with the full extent of the 3' recoding signal important for ribosomal frameshifting, were limited. Accordingly, deductions from the part of the work involving the 3' RNA structure need to be treated with caution. Future experiments could employ RNA construct variants with increased pairing potential 3' of the U-tract and at different distances from it, or a similar strategy involving complementary antisense oligonucleotides.

RNA structure effects

The potential for RNA structure formation ahead a slippage motif can strongly influence the directionality of HIV reverse transcriptase slippage (38). Interestingly, an RNA antisense specific to the sequence 5' adjacent to the HIV *gag-pol* U6 tract also influences, though moderately, HIV-1 RT polymerase slippage. Testing whether native *gag* RNA template structure formation 5' to the run of U's influence HIV GagPol synthesis via reverse transcriptase slippage is outside the scope of the present work.

During synthesis of HIV double-stranded DNA from single-stranded cDNA, before the polymerase reaches 3'-TA6TCCC-5', it encounters the complement of the sequence that forms the ribosomal frameshift stimulatory structure. Studies with a model stem-loop 5' adjacent to an A6 RNA motif have shown the potential for 'road-blocking' RNA structure to enhance slippage-mediated base addition (38) raising the possibility that HIV RT undergoes structure stimulated slippage during DNA second strand synthesis. In contrast to RT slippage at 5'-AU6AGGG-3' in the generation of single-stranded cDNA, a higher ratio of dATP compared to dTTP is expected to positively influence slippage-mediated base omission on second strand DNA synthesis.

While any conclusion about whether the presence of a cassette that can mediate both ribosomal frameshifting and reverse transcriptase slippage is fortuitous or has been selected to facilitate both processes, is premature, dual function may not be coincidental.

Perspective

One type of recoding and splicing are alternatively used in different occurrences of one set of orthologs (74–76), in several instances one type of recoding has been replaced by a different type (16). One example of the latter is *dnaX* where in *E. coli* expression of a second protein product involves a ribosomal frameshift event and in other bacteria including *Thermus thermophilus*, synthesis of the counterpart product involves transcription slippage (20,77–79). The present work raises the possibility that several retroviruses use a single cassette for productive ribosomal frameshifting and polymerase slippage. Such dual function could be relevant to studies aiming to ameliorate retroviral infection with compounds that alter only the efficiency of one of the two mechanisms involved.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Drs G. Loughran for facilitating the work; M. O'Connell Motherway, T. Cross and E. Dillan for help with LiCor sequencing; F. Crispie and T. Cotter in the APC Microbiome Institute for Illumina sequencing and advice; and Alan J. Herr and several anonymous reviewers for constructive comments on the ms.

FUNDING

Irish Research Council fellowship to C.P.; Science Foundation Ireland grants [12/IP1492, 13/1A/1853 to J.F.A., 12/1A/1335 to P.V.B., 12/RC/2273 to D.v.S.J.]. Funding for open access charge: Science Foundation Ireland.
Conflict of interest statement. None declared.

REFERENCES

- Philipson, L., Andersson, P., Olshevsky, U., Weinberg, R., Baltimore, D. and Gesteland, R. (1978) Translation of MuLV and MSV RNAs in nuclease-treated reticulocyte extracts: enhancement of the Gag-Pol polypeptide with yeast suppressor tRNA. *Cell*, **13**, 189–199.
- Yoshinaka, Y., Katoh, I., Copeland, T.D. and Oroszlan, S. (1985) Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 1618–1622.
- Jacks, T. and Varmus, H.E. (1985) Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. *Science*, **230**, 1237–1242.
- Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J. and Varmus, H.E. (1988) Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. *Nature*, **331**, 280–283.
- Weiss, R.B., Dunn, D.M., Shuh, M., Atkins, J.F. and Gesteland, R.F. (1989) *E. coli* ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *New Biol.*, **1**, 159–169.
- Li, G.D. and De Clercq, E. (2016) HIV genome-wide protein associations: a review of 30 years of research. *Microbiol. Mol. Biol. Rev.*, **80**, 679–731.
- Yelverton, E., Lindsley, D., Yamauchi, P. and Gallant, J.A. (1994) The function of a ribosomal frameshifting signal from human immunodeficiency virus-1 in *Escherichia coli*. *Mol. Microbiol.*, **11**, 303–313.
- Flynn, W.F., Chang, M.W., Tan, Z.Q., Oliveira, G., Yuan, J.Y., Okulicz, J.F., Torbett, B.E. and Levy, R.M. (2015) Deep sequencing of protease inhibitor resistant HIV patient isolates reveals patterns of correlated mutations in Gag and protease. *PLoS Comput. Biol.*, **11**, e1004249.
- Brakier-Gingras, L., Charbonneau, J. and Butcher, S.E. (2012) Targeting frameshifting in the human immunodeficiency virus. *Expert Opin. Ther. Targets*, **16**, 249–258.
- Ofori, L.O., Hilimire, T.A., Bennett, R.P., Brown, N.W. Jr, Smith, H.C. and Miller, B.L. (2014) High-affinity recognition of HIV-1 frameshift-stimulating RNA alters frameshifting *in vitro* and interferes with HIV-1 infectivity. *J. Med. Chem.*, **57**, 723–732.
- Cardno, T.S., Shimaki, Y., Sleebs, B.E., Lackovic, K., Parisot, J.P., Moss, R.M., Crowe-McAuliffe, C., Mathew, S.F., Edgar, C.D., Kleffmann, T. et al. (2015) HIV-1 and Human PEG10 frameshift elements are functionally distinct and distinguished by novel small molecule modulators. *PLoS One*, **10**, e0139036.
- Hilimire, T.A., Bennett, R.P., Stewart, R.A., Garcia-Miranda, P., Blume, A., Becker, J., Sherer, N., Helms, E.D., Butcher, S.E., Smith, H.C. et al. (2016) N-methylation as a strategy for enhancing the affinity and selectivity of RNA-binding peptides: application to the HIV-1 frameshift-stimulating RNA. *ACS Chem. Biol.*, **11**, 88–94.
- Dinman, J.D., Ruiz-Echevarria, M.J. and Peltz, S.W. (1998) Translating old drugs into new treatments: ribosomal frameshifting as a target for antiviral agents. *Trends Biotechnol.*, **16**, 190–196.
- Hung, M., Patel, P., Davis, S. and Green, S.R. (1998) Importance of ribosomal frameshifting for human immunodeficiency virus type 1 particle assembly and replication. *J. Virol.*, **72**, 4819–4824.
- Marcheschi, R.J., Tonelli, M., Kumar, A. and Butcher, S.E. (2011) Structure of the HIV-1 frameshift site RNA bound to a small molecule inhibitor of viral replication. *ACS Chem. Biol.*, **6**, 857–864.
- Atkins, J.F., Loughran, G., Bhatt, P.R., Firth, A.E. and Baranov, P.V. (2016) Ribosomal frameshifting and transcriptional slippage: from genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.*, **44**, 7007–7078.
- Jacques, J.P., Hausmann, S. and Kolakofsky, D. (1994) Paramyxovirus messenger RNA editing leads to G deletions as well as insertions. *EMBO J.*, **13**, 5496–5503.
- Kolakofsky, D. (2016) Paramyxovirus RNA synthesis, mRNA editing, and genome hexamer phase: a review. *Virology*, **498**, 94–98.
- Sanchez, A., Trappier, S.G., Mahy, B.W.J., Peters, C.J. and Nichol, S.T. (1996) The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 3602–3607.
- Larsen, B., Wills, N.M., Nelson, C., Atkins, J.F. and Gesteland, R.F. (2000) Nonlinearity in genetic decoding: homologous DNA replicase genes use alternatives of transcriptional slippage or translational frameshifting. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 1683–1688.
- Baranov, P.V., Hammer, A.W., Zhou, J.D., Gesteland, R.F. and Atkins, J.F. (2005) Transcriptional slippage in bacteria: distribution in sequenced genomes and utilization in IS element gene expression. *Genome Biol.*, **6**, R25.
- Baranov, P.V., Fayet, O., Hendrix, R.W. and Atkins, J.F. (2006) Recoding in bacteriophages and bacterial IS elements. *Trends Genet.*, **22**, 174–181.
- Penno, C., Sharma, V., Coakley, A., Motherway, M.O., van Sinderen, D., Lubkowska, L., Kireeva, M.L., Kashlev, M., Baranov, P.V. and Atkins, J.F. (2015) Productive mRNA stem loop-mediated transcriptional slippage: crucial features in common with intrinsic terminators. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E1984–E1993.
- Mehedi, M., Falzarano, D., Seebach, J., Hu, X.J., Carpenter, M.S., Schnittler, H.J. and Feldmann, H. (2011) A new ebola virus non structural glycoprotein expressed through RNA Editing. *J. Virol.*, **85**, 5406–5414.
- Shabman, R.S., Jabado, O.J., Mire, C.E., Stockwell, T.B., Edwards, M., Mahajan, M., Geisbert, T.W. and Basler, C.F. (2014) Deep sequencing identifies noncanonical editing of Ebola and Marburg virus RNAs in infected cells. *MBio*, **5**, e02011.
- Volchkova, V.A., Dolnik, O., Martinez, M.J., Reynard, O. and Volchkov, V.E. (2015) RNA editing of the GP gene of Ebola virus is an important pathogenicity factor. *J. Infect. Dis.*, **212**, S226–S233.
- Penno, C., Sansonetti, P. and Parsot, C. (2005) Frameshifting by transcriptional slippage is involved in production of MxiE, the transcription activator regulated by the activity of the type III secretion apparatus in *Shigella flexneri*. *Mol. Microbiol.*, **56**, 204–214.
- Penno, C. and Parsot, C. (2006) Transcriptional slippage in *mxiE* controls transcription and translation of the downstream *mxiD* gene, which encodes a component of the *Shigella flexneri* type III secretion apparatus. *J. Bacteriol.*, **188**, 1196–1198.
- Penno, C., Hachani, A., Biskri, L., Sansonetti, P., Allaoui, A. and Parsot, C. (2006) Transcriptional slippage controls production of type III secretion apparatus components in *Shigella flexneri*. *Mol. Microbiol.*, **62**, 1460–1468.
- Gueguen, E., Wills, N.M., Atkins, J.F. and Cascales, E. (2014) Transcriptional frameshifting rescues *Citrobacter rodentium* type VI secretion by the production of two length variants from the prematurely interrupted *tssM* gene. *PLoS Genet.*, **10**, e1004869.
- Tamas, I., Wernegreen, J.J., Nystedt, B., Kauppinen, S.N., Darby, A.C., Gomez-Valero, L., Lundin, D., Poole, A.M. and Andersson, S.G.E. (2008) Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 14934–14939.
- Wernegreen, J.J., Kauppinen, S.N. and Degan, P.H. (2010) Slip into something more functional: selection maintains ancient frameshifts in homopolymeric sequences. *Mol. Biol. Evol.*, **27**, 833–839.
- Olspert, A., Chung, B.Y.W., Atkins, J.F., Carr, J.P. and Firth, A.E. (2015) Transcriptional slippage in the positive-sense RNA virus family Potyviridae. *EMBO Rep.*, **16**, 995–1004.
- Untiveros, M., Olspert, A., Artola, K., Firth, A.E., Kreuze, J.F. and Valkonen, J.P. (2016) A novel sweet potato potyvirus ORF is expressed via polymerase slippage and suppresses RNA silencing. *Mol. Plant. Pathol.*, **17**, 1111–1123.

35. Mingot, A., Valli, A., Rodamilans, B., San Leon, D., Baulcombe, D.C., Garcia, J.A. and Lopez-Moya, J.J. (2016) The *PIN-PISPO* trans-frame gene of sweet potato feathery mottle potyvirus is produced during virus infection and functions as an RNA silencing suppressor. *J. Virol.*, **90**, 3543–3557.
36. Hagiwara-Komoda, Y., Choi, S.H., Sato, M., Atsumi, G., Abe, J., Fukuda, J., Honjo, M.N., Nagano, A.J., Komoda, K., Nakahara, K.S. *et al.* (2016) Truncated yet functional viral protein produced via RNA polymerase slippage implies underestimated coding capacity of RNA viruses. *Sci. Rep.*, **6**, doi:10.1038/srep21411.
37. Mathew, S.F., Crowe-McAuliffe, C., Graves, R., Cardno, T.S., McKinney, C., Poole, E.S. and Tate, W.P. (2015) The highly conserved codon following the slippery sequence supports -1 frameshift efficiency at the HIV-1 frameshift site. *PLoS One*, **10**, e0122176.
38. Penno, C., Kumari, R., Baranov, P.V., van Sinderen, D. and Atkins, J.F. (2017) Stimulation of reverse transcriptase generated cDNAs with specific indels by template RNA structure: Retrotransposon, dNTP balance, RT-reagent usage. *Nucleic Acids Res.*, doi:10.1093/nar/gkx689.
39. Boyer, J.C., Bebenek, K. and Kunkel, T.A. (1992) Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 6919–6923.
40. Bebenek, K., Roberts, J.D. and Kunkel, T.A. (1992) The effects of dNTP pool imbalances on frameshift fidelity during DNA replication. *J. Biol. Chem.*, **267**, 3589–3596.
41. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, doi:10.14806/ej.17.1.200.
42. 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 Biol.*, **10**, R25.
43. Li, H., Handsaker, B., Wysocki, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Proc. G.P.D. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
44. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
45. Bembom, O. (2016) SeqLogo: Sequence logos for DNA sequence alignments. *SeqLogo package version*. 1.40.0 (1990):1–5.
46. Howe, E.A., Sinha, R., Schlauch, D. and Quackenbush, J. (2011) RNA-Seq analysis in MeV. *Bioinformatics*, **27**, 3209–3210.
47. Dulude, D., Baril, M. and Brakier-Gingras, L. (2002) Characterization of the frameshift stimulatory signal controlling a programmed-1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res.*, **30**, 5094–5102.
48. Low, J.T., Garcia-Miranda, P., Mouzakis, K.D., Gorelick, R.J., Butcher, S.E. and Weeks, K.M. (2014) Structure and dynamics of the HIV-1 Frameshift element RNA. *Biochemistry*, **53**, 4282–4291.
49. Huang, X., Yang, Y., Wang, G., Cheng, Q. and Du, Z. (2014) Highly conserved RNA pseudoknots at the *gag-pol* junction of HIV-1 suggest a novel mechanism of -1 ribosomal frameshifting. *RNA*, **20**, 587–593.
50. Mouzakis, K.D., Dethoff, E.A., Tonelli, M., Al-Hashimi, H. and Butcher, S.E. (2015) Dynamic motions of the HIV-1 frameshift site RNA. *Biophys. J.*, **108**, 644–654.
51. Qiao, Q., Yan, Y., Guo, J., Du, S., Zhang, J., Jia, R., Ren, H., Qiao, Y. and Li, Q. (2016) A review on architecture of the *gag-pol* ribosomal frameshifting RNA in human immunodeficiency virus: a variability survey of virus genotypes. *J. Biomol. Struct. Dyn.*, **35**, 1629–1653.
52. Liao, P.Y., Choi, Y.S., Dimman, J.D. and Lee, K.H. (2011) The many paths to frameshifting: kinetic modelling and analysis of the effects of different elongation steps on programmed -1 ribosomal frameshifting. *Nucleic Acids Res.*, **39**, 300–312.
53. Lin, Z., Gilbert, R.J. and Brierley, I. (2012) Spacer-length dependence of programmed -1 or -2 ribosomal frameshifting on a U6A heptamer supports a role for messenger RNA (mRNA) tension in frameshifting. *Nucleic Acids Res.*, **40**, 8674–8689.
54. Erickson-Vitonen, S., Manfredi, J., Vitonen, P., Tribe, D.E., Tritch, R., Hutchison, C.A. 3rd, Loeb, D.D. and Swanson, R. (1989) Cleavage of HIV-1 gag polyprotein synthesized *in vitro*: sequential cleavage by the viral protease. *AIDS Res. Hum. Retroviruses*, **5**, 577–591.
55. Henderson, L.E., Bowers, M.A., Sowder, R.C., Serabyn, S.A., Johnson, D.G., Bess, J.W., Arthur, L.O., Bryant, D.K. and Fenselau, C. (1992) Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings, and complete amino acid sequences. *J. Virol.*, **66**, 1856–1865.
56. Pettit, S.C., Moody, M.D., Wehbie, R.S., Kaplan, A.H., Nantermet, P.V., Klein, C.A. and Swanson, R. (1994) The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J. Virol.*, **68**, 8017–8027.
57. Wieggers, K., Rutter, G., Kottler, H., Tessmer, U., Hohenberg, H. and Krausslich, H.G. (1998) Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J. Virol.*, **72**, 2846–2854.
58. Muller, B., Anders, M., Akiyama, H., Welsch, S., Glass, B., Nikovics, K., Clavel, F., Tervo, H.M., Keppler, O.T. and Krausslich, H.G. (2009) HIV-1 Gag processing intermediates trans-dominantly interfere with HIV-1 Infectivity. *J. Biol. Chem.*, **284**, 29692–29703.
59. Ning, J.Y., Erdemci-Tandogan, G., Yufenyuy, E.L., Wagner, J., Himes, B.A., Zhao, G.P., Aiken, C., Zandi, R. and Zhang, P.J. (2016) *In vitro* protease cleavage and computer simulations reveal the HIV-1 capsid maturation pathway. *Nat. Commun.*, **7**, doi:10.1038/ncomms13689.
60. Dam, E., Quercia, R., Glass, B., Descamps, D., Launay, O., Duval, X., Krausslich, H.G., Hance, A.J., Clavel, F. and ANRS 109 study group (2009) Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. *PLoS Pathog.*, **5**, e1000345.
61. Doyon, L., Croteau, G., Thibeault, D., Poulin, F., Pilote, L. and Lamarre, D. (1996) Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J. Virol.*, **70**, 3763–3769.
62. Knops, E., Brakier-Gingras, L., Schuster, E., Pfister, H., Kaiser, R. and Verheyen, J. (2012) Mutational patterns in the frameshift-regulating site of HIV-1 selected by protease inhibitors. *Med. Microbiol. Immunol.*, **201**, 213–218.
63. Yan, N., O'Day, E., Wheeler, L.A., Engelman, A. and Lieberman, J. (2011) HIV DNA is heavily uracilated, which protects it from autointegration. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 9244–9249.
64. Lahouassa, H., Daddacha, W., Hofmann, H., Ayinde, D., Logue, E.C., Dragin, L., Bloch, N., Maudet, C., Bertrand, M., Gramberg, T. *et al.* (2012) SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat. Immunol.*, **13**, 223–228.
65. Hrecka, K., Hao, C., Gierszewska, M., Swanson, S.K., Kesik-Brodacka, M., Srivastava, S., Florens, L., Washburn, M.P. and Skowronski, J. (2011) Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature*, **474**, 658–661.
66. Laguerre, N., Sobhian, B., Casartelli, N., Ringard, M., Chable-Bessia, C., Segal, E., Yatim, A., Emiliani, S., Schwartz, O. and Benkirane, M. (2011) SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature*, **474**, 654–657.
67. Lenzi, G.M., Domaal, R.A., Kim, D.H., Schinazi, R.F. and Kim, B. (2015) Mechanistic and kinetic differences between reverse transcriptases of vpx coding and non-coding lentiviruses. *J. Biol. Chem.*, **290**, 30078–30086.
68. Rentoft, M., Lindell, K., Tran, P., Chabes, A.L., Buckland, R.J., Watt, D.L., Marjavaara, L., Nilsson, A.K., Melin, B., Trygg, J. *et al.* (2016) Heterozygous colon cancer-associated mutations of SAMHD1 have functional significance. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 4723–4728.
69. Mertz, T.M., Sharma, S., Chabes, A. and Shcherbakova, P.V. (2015) Colon cancer-associated mutator DNA polymerase δ variant causes expansion of dNTP pools increasing its own infidelity. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E2467–E2476.
70. Lapkowski, M., Tian, L., Miller, J.T., Le Grice, S.F.J. and Yang, W. (2013) Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. *Nat. Struct. Mol. Biol.*, **20**, 230–236.
71. Nowak, E., Miller, J.T., Bona, M.K., Studnicka, J., Szczepanowski, R.H., Jurkowski, J., Le Grice, S.F.J. and Nowotny, M.

- (2014) Ty3 reverse transcriptase complexed with an RNA-DNA hybrid shows structural and functional asymmetry. *Nat. Struct. Mol. Biol.*, **21**, 389–396.
72. Hamburgh, M.E., Curr, K.A., Monaghan, M., Rao, V.R., Tripathi, S., Preston, B.D., Sarafianos, S., Arnold, E., Darden, T. and Prasad, V.R. (2006) Structural determinants of slippage-mediated mutations by human immunodeficiency virus type 1 reverse transcriptase. *J. Biol. Chem.*, **281**, 7421–7428.
 73. Baril, M., Dulude, D., Gendron, K., Lemay, G. and Brakier-Gingras, L. (2003) Efficiency of a programmed -1 ribosomal frameshift in the different subtypes of the human immunodeficiency virus type 1 group M. *RNA*, **9**, 1246–1253.
 74. Blaby-Haas, C.E., Padilla-Benavides, T., Stube, R., Arguello, J.M. and Merchant, S.S. (2014) Evolution of a plant-specific copper chaperone family for chloroplast copper homeostasis. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E5480–E5487.
 75. Meydan, S., Klepacki, D., Karthikeyan, S., Margus, T., Thomas, P., Jones, J.E., Khan, Y., Briggs, J., Dinman, J.D., Vazquez-Laslop, N. *et al.* (2017) Programmed ribosomal frameshifting generates a copper transporter and a copper chaperone from the same gene. *Mol. Cell*, **65**, 207–219.
 76. Atkins, J.F., Loughran, G. and Baranov, P.V. (2017) A [Cu]rious ribosomal profiling pattern leads to the discovery of ribosomal frameshifting in the synthesis of a copper chaperone. *Mol. Cell*, **65**, 203–204.
 77. Flower, A.M. and Mchenry, C.S. (1990) The gamma subunit of DNA Polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 3713–3717.
 78. Tsuchihashi, Z. and Kornberg, A. (1990) Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 2516–2520.
 79. Antonov, I., Coakley, A., Atkins, J.F., Baranov, P.V. and Borodovsky, M. (2013) Identification of the nature of reading frame transitions observed in prokaryotic genomes. *Nucleic Acids Res.*, **41**, 6514–6530.

Appendix 3: Identifying the allelic expression in the autism susceptible gene *Sprouty3* (*SPRY3*).

Manuscript under preparation.

5.3.1 Introduction

Mammals are known to contain four paralogs of gene *Sprouty*, *SPRY1-4* (de Maximy et al., 1999; Leeksa et al., 2002), which play multiple roles in developmental pathways by interacting with various other proteins (Casci, Vinos, & Freeman, 1999; Edwin, Anderson, Ying, & Patel, 2009). In mouse *SPRY1,2* and *4* are found to be involved in the various organogenesis processes such as kidney development (Basson et al., 2005). The role of the *Sprouty* genes in the development was further confirmed by null mutants. Various mouse *Sprouty* null mutants were found to have defects in the developmental processes such as defects in organ of Corti causing oesophageal achalasia, impaired hearing and enteric neuronal hyperplasia, mandibular defects, defect in formation of the temporomandibular joint, cardiovascular defects, defects in limb, lung and craniofacial morphogenesis (Purcell et al., 2009; Shim, Minowada, Coling, & Martin, 2005; Taketomi et al., 2005; Taniguchi et al., 2007). In case of human, modified expression of one or more *Sprouty* genes (*1,2* and *4*) has been detected in multiple tumours, cardiovascular disease and psychiatric disease (Edwin et al., 2009; Pillai, 2008; Sayed et al., 2008; C. Zhang et al., 2005).

The comparatively less explored *SPRY3* was recently found to be involved in reduction axonal branching in *Xenopus* (Panagiotaki, Dajas-Bailador, Amaya, Papalopulu, & Dorey, 2010). The pseudo-autosomal region 2 (PAR2) located *SPRY3* is found to undergo an unusual form of dosage compensation involving random X-inactivation in females and inactivation of the Y-linked locus in males and is expected to exhibit X-linked genetic transmission (Ciccodicola et al., 2000; D'Esposito et al., 1997; Helena Mangs & Morris, 2007). Another study found that *SPRY3* inhibits growth and branching of ganglion cells during postnatal development and in the adult. This study described complex transcriptional regulation of *SPRY3* by X-linked and

pseudo-autosomal promoters, and provided evidence that *SPRY3* is a susceptibility gene for autism (Ning et al., 2015).

The objective of this work is to detect the deregulation of the *SPRY3* in cerebellum section of the autism brains and if the normally silenced Y-linked copy of *SPRY3* in males is reactivated. The expression of the two allelic variants of *SRPY3* was analysed using RNA-seq libraries of cerebellar vermis (CB) region of the brain containing 33 Autism and 38 control samples.

5.3.2 Results

If two allelic variants of *SPRY3* gene are expressed within the same individual, we expect that *SPRY3* transcripts would contain nucleotide variants in RNA-seq data matching specific Single nucleotide variants (SNVs). To detect the presence of such variants in *SPRY3* transcripts in each individual we used RNA-seq samples from cerebellar vermis (CB) region of the brain (Parikshak et al., 2016). The *SPRY3* transcript was found to have total mapped read count from 653 to 4033 depending on the samples (Figure 5.15 shows an alignment for a randomly selected sample).

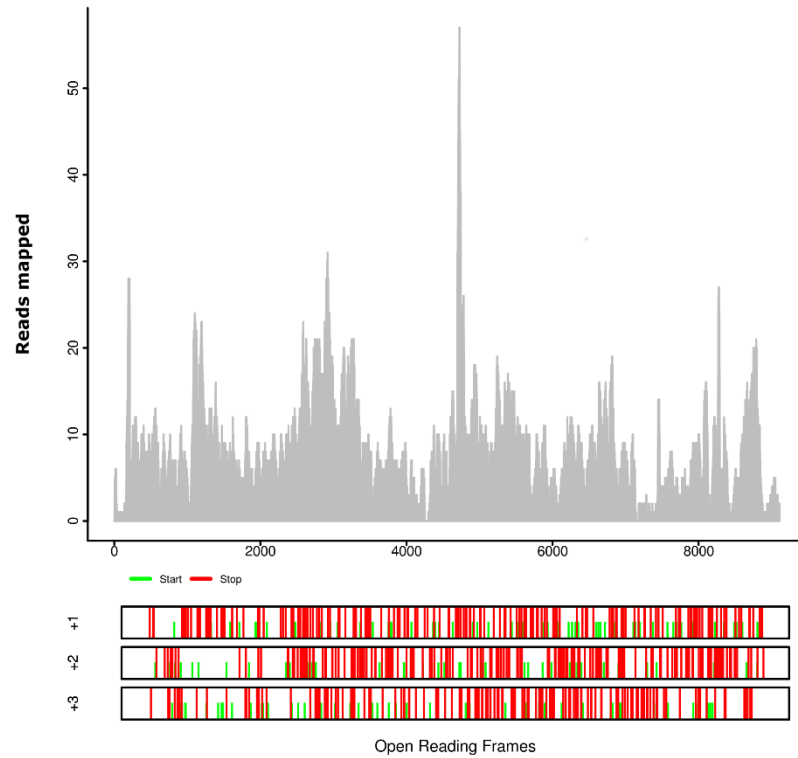


Figure 5.3.1: The top panel is RNA-seq coverage plot showing the number of reads mapped (axis y) to each position (axis x) of *SPRY3* (RefSeq ID - NM_005840.2). The bottom is an open reading frames plot showing positions of start (green); and stop (red) codons in three different reading frames.

Further the composition of different nucleotides mapping each coordinate of *SPRY3* transcript was analysed. For heterozygous genotypes it is expected that the probability of finding a nucleotide matching reference sequence at SNV is 0.5. While for homozygous it is either 0 or 1. It is also expected that the actual frequencies would differ from the expected distributions due to systematic biases and errors in RNA-seq library generation and sequencing as well as due to variations in allele specific expression levels in case of heterozygosity. To control for this variation empirically we analysed the distributions of matching nucleotides in the transcripts where it is known whether one or two alleles are expressed.

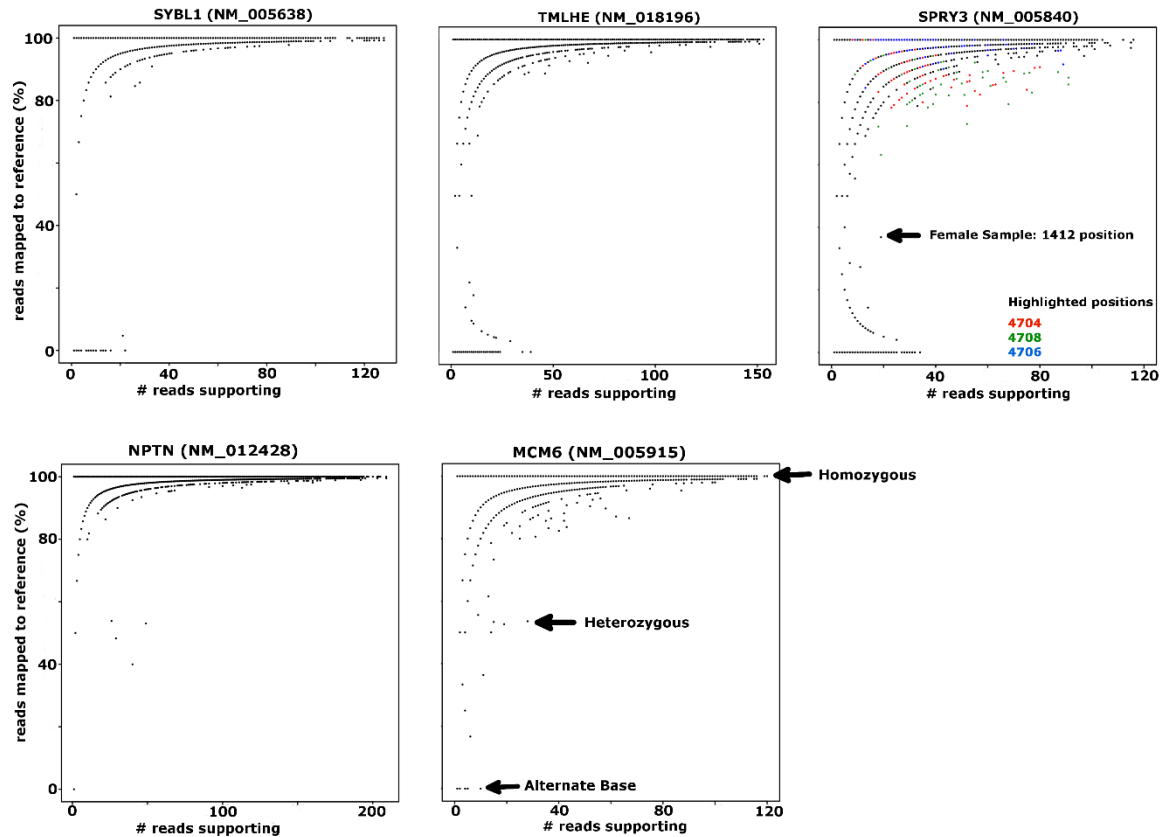


Figure 5.3.2: The percentage of reads supporting reference nucleotide identities (axis y) at each coordinate position in each of 71 samples. Each plot corresponds to a gene transcript as indicated. When two allelic variants are expressed it is expected that about half of mapped reads aligned to SNV would support the reference identity. The axis x indicates the number of reads aligned to each specific position in each sample and this increases the statistical significance. The data points supporting ~15% alternative bases are shown using the same colours, i.e. red for 4704, green for 4708 and blue for 4706 (coordinate numbering is according to NM_005840.2).

X chromosome encoded genes (*SYBL1* and *TMLHE*) were used as control examples of a single allelic variant expression for samples obtained in males, whereas autosomal genes (*NPTN* and *MCM6*) were used as control examples of two alleles expression. As can be seen in Figure 5.16 transcripts from autosome encoded genes contain samples with some coordinates where a large number of aligned reads support the reference identity about 50% of times. *SPRY3* was found to contain some positions supporting dual base composition (~80% supporting reference) (Figure 5.16); these positions were analysed individually to check how they behave in different samples. It appears that high frequency of alternative base occurs at the same coordinates across

multiple samples (Figure 5.16). This is not what would be expected for heterozygous genotypes as only some individuals are expected to be heterozygous. More likely these positions are either prone to sequencing errors or misalignments either because of RNA modifications or due to a similar transcript reads generated deriving from a different locus.

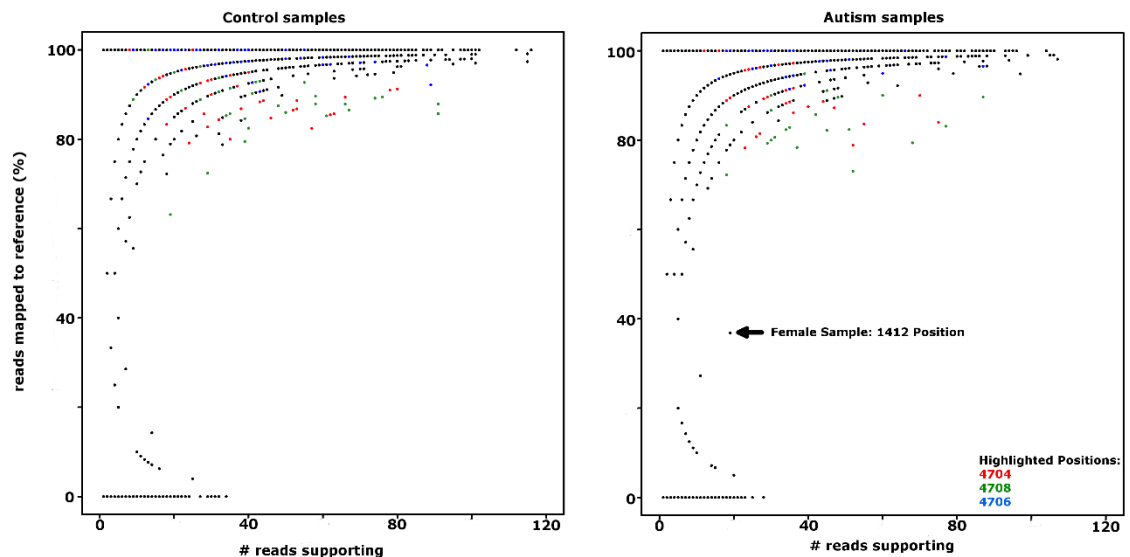


Figure 5.3.4: Comparison of the ASD and Control Samples plots representing the percentage of reads supporting reference and alternative nucleotide at each coordinate position for the 71 samples. The data points supporting ~15% alternative bases are shown using the same colours, i.e. red for 4704, green for 4708 and blue for 4706 (coordinate numbering is according to NM_005840.2).

There was no significant difference observed in the heterozygosity plots of Autism spectrum disorder (ASD) samples versus control samples (Figure 5.17). No significant proof of transcription of T4 and T7 transcripts was found using the reads spanning across the exon junctions (Figure 5.18).

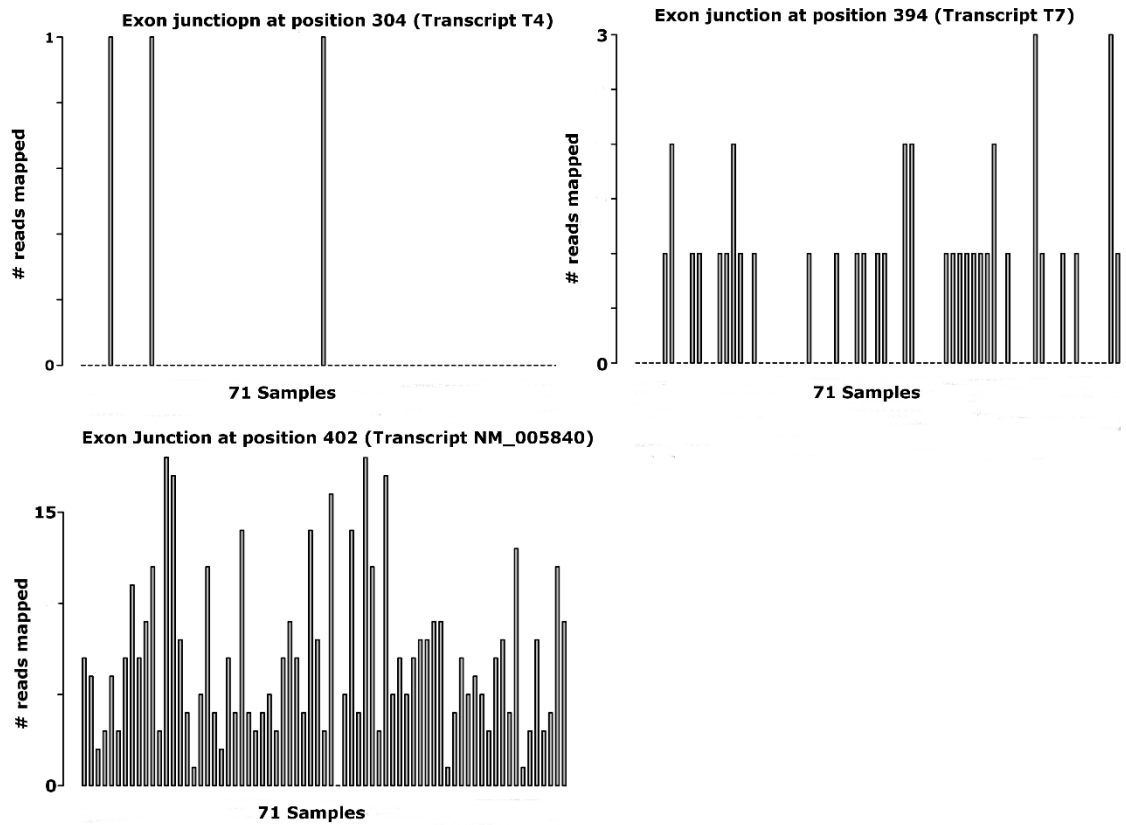


Figure 5.3.5: Represents the total number of reads spanning the exon junctions (axis y) for transcripts T4, T7 and *SPRY3* at positions 304, 394 and 402 respectively (as mentioned in (Ning et al., 2015)) at each sample (axis x), where *SPRY3* 402 position is used as positive control.

5.3.3 Conclusion

Using RNA-seq data from cerebellar vermis (CB) region of autism samples did not provide any significant evidence for expression of two allelic variants of *SPRY3* gene. As well, no significant proof for the expression of alternative transcripts T4 and T7 (Ning et al., 2015) was found. It is speculated that expression of *SPRY3* can be affected by various factors such as age of the patient, section of the cerebellum from where tissue sample was extracted and time from the death when tissue sample was extracted, and it is speculated that modifying these factors can alter the expression of *SPRY3* gene.

5.3.4 Methods

The paired-end RNA-seq libraries of cerebellar vermis (CB) region of the brain containing 33 Autism and 38 control samples were obtained from PsychENCODE (Parikshak et al., 2016). The samples contained read count varying between 50 to 200 million. Human transcriptome sequence was obtained from the RefSeq database (O'Leary et al., 2016) (NCBI Homo sapiens Annotation Release 108). Raw reads were aligned to reference transcriptome sequences using bowtie2 short read alignment program (Langmead & Salzberg, 2012). Default parameters (maximum 2 mismatches allowed) were used to carry out local alignments. Reads mapped to only one location in the transcriptome were fetched by removing the alignments with 'XS:i' bowtie2 tag which represents reads having more than one possible mapping to the reference (Langmead & Salzberg, 2012). SAMtools version 1.3.1 was used to obtain the sorted BAM alignment files which were further used to predict the heterozygosity in *SPRY3* (Li et al., 2009). SAMtools mpileup and BCFtools were used to obtain information on identity of nucleotides of mapped reads at each position of the *SPRY3* transcript (Li et al., 2009; Danecek et al., 2011).

To check the transcription levels of T4 and T7 transcript isoforms, reads mapped on the exon junctions were analysed (Ning et al., 2015). The annotated exon junction of *SPRY3* transcript NM_005840 at position 402 was used as a positive control.

References

- Basson, M. A., Akbulut, S., Watson-Johnson, J., Simon, R., Carroll, T. J., Shakya, R., . . . Licht, J. D. (2005). Sprouty1 is a critical regulator of GDNF/RET-mediated kidney induction. *Dev Cell*, 8(2), 229-239. doi:10.1016/j.devcel.2004.12.004
- Casci, T., Vinos, J., & Freeman, M. (1999). Sprouty, an intracellular inhibitor of Ras signaling. *Cell*, 96(5), 655-665.
- Ciccodicola, A., D'Esposito, M., Esposito, T., Gianfrancesco, F., Migliaccio, C., Miano, M. G., . . . D'Urso, M. (2000). Differentially regulated and evolved genes in the fully sequenced Xq/Yq pseudoautosomal region. *Hum Mol Genet*, 9(3), 395-401.
- D'Esposito, M., Matarazzo, M. R., Ciccodicola, A., Strazzullo, M., Mazzarella, R., Quaderi, N. A., . . . D'Urso, M. (1997). Differential expression pattern of XqPAR-linked genes SYBL1 and IL9R correlates with the structure and

- evolution of the region. *Hum Mol Genet*, 6(11), 1917-1923.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Genomes Project Analysis, G. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158. doi:10.1093/bioinformatics/btr330
- de Maximy, A. A., Nakatake, Y., Moncada, S., Itoh, N., Thiery, J. P., & Bellusci, S. (1999). Cloning and expression pattern of a mouse homologue of drosophila sprouty in the mouse embryo. *Mech Dev*, 81(1-2), 213-216.
- Edwin, F., Anderson, K., Ying, C., & Patel, T. B. (2009). Intermolecular interactions of Sprouty proteins and their implications in development and disease. *Mol Pharmacol*, 76(4), 679-691. doi:10.1124/mol.109.055848
- Helena Mangs, A., & Morris, B. J. (2007). The Human Pseudoautosomal Region (PAR): Origin, Function and Future. *Curr Genomics*, 8(2), 129-136.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4), 357-359. doi:10.1038/nmeth.1923
- Leeksa, O. C., Van Achterberg, T. A., Tsumura, Y., Toshima, J., Eldering, E., Kroes, W. G., . . . de Vries, C. J. (2002). Human sprouty 4, a new ras antagonist on 5q31, interacts with the dual specificity kinase TESK1. *Eur J Biochem*, 269(10), 2546-2556.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079. doi:10.1093/bioinformatics/btp352
- Ning, Z., McLellan, A. S., Ball, M., Wynne, F., O'Neill, C., Mills, W., . . . Moore, T. (2015). Regulation of SPRY3 by X chromosome and PAR2-linked promoters in an autism susceptibility region. *Hum Mol Genet*, 24(18), 5126-5141. doi:10.1093/hmg/ddv231
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciuffo, S., Haddad, D., McVeigh, R., . . . Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, 44(D1), D733-745. doi:10.1093/nar/gkv1189
- Panagiotaki, N., Dajas-Bailador, F., Amaya, E., Papalopulu, N., & Dorey, K. (2010). Characterisation of a new regulator of BDNF signalling, Sprouty3, involved in axonal morphogenesis in vivo. *Development*, 137(23), 4005-4015. doi:10.1242/dev.053173

- Parikshak, N. N., Swarup, V., Belgard, T. G., Irimia, M., Ramaswami, G., Gandal, M. J., . . . Geschwind, D. H. (2016). Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*, 540(7633), 423-427. doi:10.1038/nature20612
- Pillai, A. (2008). Decreased expression of Sprouty2 in the dorsolateral prefrontal cortex in schizophrenia and bipolar disorder: a correlation with BDNF expression. *PLoS One*, 3(3), e1784. doi:10.1371/journal.pone.0001784
- Purcell, P., Joo, B. W., Hu, J. K., Tran, P. V., Calicchio, M. L., O'Connell, D. J., . . . Tabin, C. J. (2009). Temporomandibular joint formation requires two distinct hedgehog-dependent steps. *Proc Natl Acad Sci U S A*, 106(43), 18297-18302. doi:10.1073/pnas.0908836106
- Sayed, D., Rane, S., Lypowy, J., He, M., Chen, I. Y., Vashistha, H., . . . Abdellatif, M. (2008). MicroRNA-21 targets Sprouty2 and promotes cellular outgrowths. *Mol Biol Cell*, 19(8), 3272-3282. doi:10.1091/mbc.E08-02-0159
- Shim, K., Minowada, G., Coling, D. E., & Martin, G. R. (2005). Sprouty2, a mouse deafness gene, regulates cell fate decisions in the auditory sensory epithelium by antagonizing FGF signaling. *Dev Cell*, 8(4), 553-564. doi:10.1016/j.devcel.2005.02.009
- Taketomi, T., Yoshiga, D., Taniguchi, K., Kobayashi, T., Nonami, A., Kato, R., . . . Yoshimura, A. (2005). Loss of mammalian Sprouty2 leads to enteric neuronal hyperplasia and esophageal achalasia. *Nat Neurosci*, 8(7), 855-857. doi:10.1038/nn1485
- Taniguchi, K., Ayada, T., Ichiyama, K., Kohno, R., Yonemitsu, Y., Minami, Y., . . . Yoshimura, A. (2007). Sprouty2 and Sprouty4 are essential for embryonic morphogenesis and regulation of FGF signaling. *Biochem Biophys Res Commun*, 352(4), 896-902. doi:10.1016/j.bbrc.2006.11.107
- Zhang, C., Chaturvedi, D., Jaggar, L., Magnuson, D., Lee, J. M., & Patel, T. B. (2005). Regulation of vascular smooth muscle cell proliferation and migration by human sprouty 2. *Arterioscler Thromb Vasc Biol*, 25(3), 533-538. doi:10.1161/01.ATV.0000155461.50450.5a

Appendix 4: Ribosome profiling and RNA-seq samples used for analysis in section 2.2.2

SRA filename	Sample name
SRR3306583.shelf	S_phase_ribosome_profiling_replicate_1
SRR3306584.shelf	S_phase_ribosome_profiling_replicate_2
SRR3306585.shelf	M_phase_ribosome_profiling_replicate_1
SRR3306586.shelf	M_phase_ribosome_profiling_replicate_2
SRR3306587.shelf	M_phase_ribosome_profiling_synchronized_by_shake_off
SRR3306588.shelf	Asynchronous_cell_ribosome_profiling_1
SRR3306589.shelf	Asynchronous_cell_ribosome_profiling_2
SRR1916542.shelf	C_rp_rep1
SRR1916544.shelf	C_rp_rep2_lane1
SRR1916545.shelf	C_rp_rep2_lane2
SRR2086029.shelf	ribo_ND3_shControl
SRR2086026.shelf	ribo_hES_shControl_replicate2
SRR2086025.shelf	ribo_hES_shControl_replicate1
SRR1610255.shelf	ribo_ND6_shControl_replicate2
SRR1610254.shelf	ribo_ND3_shControl_replicate2
SRR1610253.shelf	ribo_ND1_shControl_replicate2
SRR1610247.shelf	ribo_ND6_shControl_replicate1
SRR1610246.shelf	ribo_ND3_shControl_replicate1
SRR1610245.shelf	ribo_ND1_shControl_replicate1
SRR1610244.shelf	ribo_hES_shControl_replicate1
SRR1610252.shelf	ribo_hES_shControl_replicate2
SRR1562539.shelf	NormalHumanRiboProfilingA
SRR1562540.shelf	NormalHumanRiboProfilingB
SRR1562541.shelf	NormalHumanRiboProfilingC
SRR2075937.shelf	ribo_DMSO_replicate4_Sample_13_
SRR2075926.shelf	ribo_DMSO_replicate2_Sample_2_
SRR2075925.shelf	ribo_DMSO_replicate1_Sample_1_
SRR2075936.shelf	ribo_DMSO_replicate3_Sample_12_
SRR1598971.shelf	mock_transfected_ribosome_profiling
SRR1585537.shelf	GM19152
SRR1585533.shelf	GM19141
SRR1585544.shelf	GM19192
SRR1585527.shelf	GM19128
SRR1585557.shelf	GM19257
SRR1585499.shelf	GM18519
SRR1585540.shelf	GM19160
SRR1585500.shelf	GM18520
SRR1585497.shelf	GM18516

SRR1585529.shelf	GM19131
SRR1585528.shelf	GM19130
SRR1585520.shelf	GM19101
SRR1585530.shelf	GM19137
SRR1585534.shelf	GM19143
SRR1585487.shelf	GM18498
SRR1585505.shelf	GM18855
SRR1585517.shelf	GM19093
SRR1585513.shelf	GM18912
SRR1585538.shelf	GM19153
SRR1585526.shelf	GM19127
SRR1585493.shelf	GM18507
SRR1585509.shelf	GM18862
SRR1585543.shelf	GM19190
SRR1585519.shelf	GM19099
SRR1585524.shelf	GM19116
SRR1585531.shelf	GM19138
SRR1585525.shelf	GM19119
SRR1585546.shelf	GM19200
SRR1585503.shelf	GM18852
SRR1585508.shelf	GM18861
SRR1585548.shelf	GM19204
SRR1585514.shelf	GM18913
SRR1585489.shelf	GM18501
SRR1585512.shelf	GM18909
SRR1585495.shelf	GM18510
SRR1585491.shelf	GM18504
SRR1585550.shelf	GM19209
SRR1585492.shelf	GM18505
SRR1585521.shelf	GM19102
SRR1585506.shelf	GM18856
SRR1585532.shelf	GM19140
SRR1585523.shelf	GM19114
SRR1585536.shelf	GM19147
SRR1585490.shelf	GM18502
SRR1585518.shelf	GM19098
SRR1585515.shelf	GM18916
SRR1585535.shelf	GM19144
SRR1585496.shelf	GM18511
SRR1585507.shelf	GM18858
SRR1585498.shelf	GM18517
SRR1585494.shelf	GM18508
SRR1585545.shelf	GM19193
SRR1585555.shelf	GM19238

SRR1585554.shelf	GM19225
SRR1585501.shelf	GM18522
SRR1585522.shelf	GM19108
SRR1585553.shelf	GM19223
SRR1585511.shelf	GM18907
SRR1585551.shelf	GM19210
SRR1585486.shelf	GM18486
SRR1585488.shelf	GM18499
SRR1585504.shelf	GM18853
SRR1585552.shelf	GM19222
SRR1585547.shelf	GM19201
SRR1585542.shelf	GM19172
SRR1585539.shelf	GM19159
SRR1585549.shelf	GM19207
SRR1585516.shelf	GM19092
SRR1585502.shelf	GM18523
SRR1585541.shelf	GM19171
SRR1585510.shelf	GM18870
SRR1585556.shelf	GM19239
SRR5227294.shelf	vehicle_10_min_rep_1
SRR5227295.shelf	vehicle_10_min_rep_2
SRR5227296.shelf	vehicle_10_min_rep_3
SRR5227303.shelf	vehicle_60_min_rep_1
SRR5227304.shelf	vehicle_60_min_rep_2
SRR5227305.shelf	vehicle_60_min_rep_3
SRR5227310.shelf	vehicle_60_min_rep_4_riboseq
SRR5227312.shelf	vehicle_60_min_rep_5_riboseq
SRR1333393.shelf	Cycloheximide_treated_HCT116
SRR2433794.shelf	ribosome_profiling
SRR5382423.shelf	WT1_RPF
SRR5382424.shelf	WT2_RPF
SRR1173905.shelf	Control_riboseq
SRR1173909.shelf	Control_riboseq
SRR1173910.shelf	Control_riboseq

Table 5.1: The ribosome profiling samples used for analysis in chapter 2 section

2.2.2

SRA filename	Sample name
SRR3306574.shelf	S_phase_RNA_seq_oligo_dT_enriched_replicate_1
SRR3306575.shelf	S_phase_RNA_seq_oligo_dT_enriched_replicate_2
SRR3306576.shelf	S_phase_RNA_seq_ribozero_treated
SRR3306577.shelf	M_phase_RNA_seq_oligo_dT_enriched_replicate_1
SRR3306578.shelf	M_phase_RNA_seq_oligo_dT_enriched_replicate_2
SRR3306579.shelf	M_phase_RNA_seq_ribozero_treated
SRR3306580.shelf	M_phase_RNA_seq_synchronized_by_shake_off
SRR3306581.shelf	Asynchronous_cell_RNA_seq_1
SRR3306582.shelf	Asynchronous_cell_RNA_seq_2
SRR2086032.shelf	rna_hES_shControl_replicate1
SRR1610263.shelf	rna_ND6_shControl_replicate1
SRR1610260.shelf	rna_hES_shControl_replicate1_0
SRR1610273.shelf	rna_ND1_shKBTBD8_replicate2
SRR2086036.shelf	rna_ND3_shControl
SRR1610268.shelf	rna_hES_shControl_replicate2_0
SRR2086033.shelf	rna_hES_shControl_replicate2
SRR1610271.shelf	rna_ND6_shControl_replicate2
SRR1610262.shelf	rna_ND3_shControl_replicate1
SRR1610261.shelf	rna_ND1_shControl_replicate1
SRR1610269.shelf	rna_ND1_shControl_replicate2
SRR1562545.shelf	NormalHumanRNA_SeqB
SRR1562544.shelf	NormalHumanRNA_SeqA
SRR1562546.shelf	NormalHumanRNA_SeqC
SRR2075930.shelf	rna_DMSO_replicate1_Sample_6_
SRR2075931.shelf	rna_DMSO_replicate2_Sample_7_
SRR1598970.shelf	mock_transfected_tRNA_rRNA_depleted_RNA_seq
SRR1598955.shelf	mock_transfected_poly_A_selected_RNA_seq
SRR5227311.shelf	vehicle_60_min_rep_4_mRNAseq
SRR5227313.shelf	vehicle_60_min_rep_5_mRNAseq
SRR5382428.shelf	WT1_mRNA
SRR5382429.shelf	WT2_mRNA
SRR5382430.shelf	WT3_mRNA
SRR1173906.shelf	Control_mRNAseq
SRR1173911.shelf	Control_mRNAseq
SRR1173912.shelf	Control_mRNAseq

Table 5.2: The RNA-seq profiling samples used for analysis in chapter 2 section 2.2.2

Project Outcomes

Publications

- Penno, C., **Kumari, R.**, Baranov, P. V., van Sinderen, D., & Atkins, J. F. (2017a). Specific reverse transcriptase slippage at the HIV ribosomal frameshift sequence: potential implications for modulation of GagPol synthesis. *Nucleic Acids Research*. doi: 10.1093/nar/gkx690
- Penno, C., **Kumari, R.**, Baranov, P. V., van Sinderen, D., & Atkins, J. F. (2017b). Stimulation of reverse transcriptase generated cDNAs with specific indels by template RNA structure: retrotransposon, dNTP balance, RT-reagent usage. *Nucleic Acids Research*. doi: 10.1093/nar/gkx689
- Kumari, R.**, Michel, A. M., M., Baranov, P. V. (2017). PausePred and Rfeet: Webtools for inferring ribosome pauses and visualizing footprint density from ribosome profiling data (*Manuscript Submitted*).
- Yerlikaya, S., Meusburger, M., **Kumari, R.**, Huber, A., Anrather, D., Costanzo, M., . . . Loewith, R. (2016). TORC1 and TORC2 work together to regulate ribosomal protein S6 phosphorylation in *Saccharomyces cerevisiae*. *Mol Biol Cell*, 27(2), 397-409. doi: 10.1091/mbc.E15-08-0594

International conferences

- RNA structure meets functions by European Molecular Biology Organization, June 12-15, 2016, Stockholm Archipelago, Sweden. (Poster presentation)
- 4th and 5th Irish Next generation sequencing conference, Trinity Biomedical Sciences Institute, Dublin, Ireland.