CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	A study of mRNA translation with computational analysis of ribosome profiling datasets
Authors	O'Connor, Patrick
Publication date	2017
Original Citation	O'Connor, P. 2017. A study of mRNA translation with computational analysis of ribosome profiling datasets. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2017, Patrick O'Connor. - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2026-09-17 06:04:28
Item downloaded from	https://hdl.handle.net/10468/5431

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

A study of mRNA translation with computational analysis of ribosome profiling datasets

By

Patrick O'Connor

Thesis in fulfilment for the degree of

PhD (Science)

National University of Ireland, Cork

School of Biochemistry and Cell Biology

June 2017

Head of School: Rosemary O'Connor

Supervisor: Pavel Baranov

Declaration

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

Patrick O'Connor

Acknowledgements

This thesis is essentially a collective of scientific papers of which I was a significant contributor. Many people have contributed to the work presented in the papers. Chief amongst these are my supervisor Pavel Baranov and co-first-author Dmitry Andreev. I carried out the data analysis of the high throughput sequencing data and contributed to the interpretation of the results. I also had an appreciable involvement to the writing of the manuscripts.

My scholarship was funded by grant awarded to Pasha from SFI (Science Foundation Ireland).

I thank Pasha for helping to create a work environment conducive for experimentation and for his open-door supervisory style. My scientific productivity would be significantly reduced without his input. I'm indebted to Dmitry for his indefatigable attitude as well as producing excellent data. I thank Audrey for inviting me to contribute to many of her projects and for encouraging mine. I thank Anmol, Svetlana and Martina who are excellent sources of distraction. I also acknowledge the work of Ken Nally and Andrew Firth in reviewing this thesis. This undertaking would be considerably less successful without the numerous people who anonymously contributed both to the open source movement and to forums such as Stackoverflow and Wikipedia.

Table of Contents

Preface **7**

Insights into the mechanisms of eukaryotic translation gained with Ribosome Profiling **13**

ABSTRACT 13

INTRODUCTION 13

INITIATION 16

mRNA recruitment to the 43S preinitiation complex 16

Old and new players that participate in scanning 5' leaders 19

Regulation of translation by upstream open reading frames 22

TCP-Seq: RiboSeq of scanning and initiating ribosomes 24

ELONGATION 25

TERMINATION AND RECYCLING 27

RIBOSEQ ARTIFACTS 31

PERSPECTIVES 35

Translation of 5' leaders is pervasive in genes resistant to eIF2 repression **37**

ABSTRACT 37

INTRODUCTION 37

RESULTS 39

Ribosome profiling 39

Effects of arsenite treatment on transcriptome. 41

Arsenite treatment strongly inhibits global translation, while translation of a few specific mRNAs is resistant. 42

Efficient translation of uORFs combined with inefficient translation of CDS is a predictor of stress resistant mRNAs. 45

Newly discovered cases of resistance to eIF2 inhibition mediated by uORFs. 47

5' leaders of several newly identified mRNAs are sufficient to provide resistance to the translation inhibition. 48

Site-specific mutagenesis confirms the critical role of *IFRD1* and *PPP1R15B* uORFs translation in mediating resistance to eIF2 inhibition. 51

An unstructured leader sequence upstream of the IFRD1 uORF is necessary for stress resistance 52

DISCUSSION 53

METHODS 56

Ribosome profiling 56

Plasmid constructions 58

mRNA preparation 59

Cell culture, western blots and transfection procedures 59
Data analysis 60
Data availability 63
Table 1. Translation response of mRNAs with reported IRES from IRESite. 64
Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes 67
ABSTRACT 67
BACKGROUND 68
RESULTS AND DISCUSSION 69
Energy stress, HIF signaling and pathways regulating translation under OGD. 69
OGD alters translation of ~3,000 mRNAs, while affecting RNA levels of only about a hundred genes 71
Ribosome density increases at mRNAs 5' leaders during OGD 73
OGD effect on translation termination 74
OGD alters local decoding rates. 75
OGD alters the recognition of translation initiation sites. 77
Novel protein isoforms, polycistronic and dual coding mRNAs. 79
CONCLUSIONS 81
METHODS 82
Materials 82
Tissue culture and experimental conditions 83
OGD and OD experiments 83
Generation of ribo-seq and mRNA-seq libraries 84
Protein isolation and Western blotting analysis. 84
ATP measurement 84
Measurement of the mitochondrial membrane potential 84
Initial processing of sequence libraries. 85
Differential gene expression analysis. 85
Analysis of pairwise similarity of ribo-seq profiles of individual mRNAs. 87
Identification of genes with leaky termination. 87
Identification of translated uORFs 87
Production of metagene profile at initiation and termination sites. 88
Analysis of biological pathways 88
rHRE containing and mTOR sensitive mRNAs 88
Data access. 89

rRNA:mRNA pairing alters the length and the symmetry of mRNA protected fragments in ribosome profiling experiments. 90
ABSTRACT 90
INTRODUCTION 90
METHODS 92
RESULTS AND DISCUSSION 95
Comparative survey of the relative impact of mRNA features on local ribosome profiling read density 99
ABSTRACT 99
INTRODUCTION 99
RESULTS 101
Ribo-seq Unit Step Transformation (RUST) 101
Evaluation of normalization methods with simulated data 104
The impact of technical biases varies among datasets 106
Influence of RNA secondary structure and nascent peptide 111
Accurate prediction of experimental footprint densities 113
Comparison of the datasets 114
DISCUSSION 114
METHODS 117
Ribo-seq datasets used in this study and their processing 117
Ribo-seq simulation 118
Determining offset to the A-site 119
Normalization approaches 120
Kullback-Leibler divergence 122
RNA secondary structure analysis 123
Amino acid physicochemical properties 123
Standard score to identify synergistic interactions 123
The comparison of predicted and real footprint densities 124
Code availability 124
Data availability 125
Bibliography 129
Appendices 163
RiboGalaxy: a platform for the alignment, analysis and visualization of ribo-seq data. 163
GWIPS-viz: Development of a ribo-seq genome browser 171

Preface

mRNA translation is the 2nd of the two main stages of protein synthesis. Described simply, it is a process by which a biological machine reads a code/syntax present in a linear macromolecule chain in order to produce another macromolecule chain. The result, a protein, is one of a range of structurally diverse molecules most of which are made up of 20 common building blocks arranged in a specific order in a chain hundreds of units long. The structure of a protein is the product of the biochemical interactions between its building blocks and the environment and this helps to confer it with its useful property (i.e. function follows form). Proteins are the active agents of known life. They (along with Brownian motion) are the cause of essentially every process instigated by life such as transporting of oxygen in the blood, memory formation and photosynthesis.

Protein synthesis is an energetically expensive cellular activity. The requirements for particular proteins in a cell may vary considerably over time and location and cell type. Thus, protein synthesis is a highly regulated process. Only a few cases of translation regulation in humans are very well characterized, some of these processes are systematic while others are specific to a single gene. The best known systematic responses are the sequestration of eIF4E and the inhibition of recycling of eIF2. To some extent these are 'emergency responses', they promote general inhibition of translation in response to a significant cell stress, in order to restore cell homeostasis. Textbook examples of gene specific regulation in humans include the genes *AMD1* (nascent peptide affect in the uORF), *ATF4* (translation of uORFs or CDS influenced by reinitiation), *EIF1* (start codon selection stringency), *PABP* (auto-regulatory binding) and *OAZ1* (programmed ribosome frameshifting). Such known cases of targeted regulation however are quite rare. It is not clear however whether this is a reflection of it occurring in a small number of genes or of our ignorance of the mechanisms of translation regulation. Generally, such mechanisms are found in genes which have a narrow range of optimal expression and perform an important function.

Translational regulation is a product of a complex network of internal cell signalling (driven by various biochemical interactions). Our major challenge with this however does

not lie with an inability to comprehend it (either of the biochemistry or the signalling) but rather in observing the process. Ribosome profiling is currently one of the best tools which we have to 'see' translation. The technique enables locating the positions of actively translating ribosomes within their mRNA. It relies on the property of the ribosome to protect a fragment of mRNA ("footprints") from degradation with RNA nucleases. This approach is typically carried out on a transcriptome wide scale and the fragments produced are converted to cDNA which are sequenced with platforms that can process up to hundreds of millions fragments. It enables for the examination of the process of translation with far greater detail and scale than was previously achieved. Ribosome profiling was developed in the lab of J. Weissman (Ingolia et al. 2009).

The technique is being quickly adopted by the community and has been responsible for significant advances of our knowledge. Its finest accomplishments have been the characterization of changes in translation under a variety of conditions. The most informative studies in terms of increasing our understanding of the fundamental roles of individual factors tend to target a single factor and measure the translational response within a relatively short time period after the perturbation. The targeting of a specific factor enables for clear identification of influence on translation and the short time period limits the occurrence of secondary downstream effects. Such experiments have significantly bettered our understanding of the several important factors of regulation including (mTOR, eIF4A, EFP, Hsp70, Dom34 and Rli1/ABCE1). We have by no means exhausted potential experiments of this type.

Along with my supervisor Pasha, Dmitry Andreev from the Ivan Shatsky lab and others, I took part in a study to examine the translational response to significant reduction of eIF2 recycling by increased levels of eIF2 phosphorylation. This is the key component of the "integrated stress response" a very well characterized pathway which results in a global inhibition of translation. The pathway is activated under a variety of cellular stresses including viral assault and oxidative stress. We examined the response after a short treatment with sodium arsenite, a potent cause of oxidative stress. We reported that the repression of translation is almost universal and that the presence of a highly translated uORF is a characteristic property of those genes that are comparatively insensitive. A parallel study by the Walter lab had similar findings, confirming that the translation

response was similar when instigated by another agent, (tunicamycin) as well as identifying that ISIRB reverses the effects of eIF2 phosphorylation (Sidrauski et al. 2015). It achieves this by activating eIF2B (the protein responsible for eIF2 recycling) (Sekine et al. 2015).

Other ribosome profiling studies may be classified as more observational, in that the objective was to examine the translational aspect of a much larger multifactorial response. Examples of such studies have examined the circadian rhythm, heatshock response and cell meiosis. In a study of a similar objective, Pasha, Dmitry and I in collaboration with A. Zhdanov from the Papkovksy lab explored the immediate translational response of neural cells to the deprivation of oxygen and glucose. This stress is of particular interest owing to its relevance to ischemic stroke. We reported the occurrence of a widespread translational response within 20 minutes of the start of the stress, evidence of altered decoding rates and altered start codon recognition.

Similar to other newly developed techniques, ribosome profiling has associated artifacts or selection bias, which if not understood may lead to the incorrect interpretation of the results. The issue may become of greatest concern when ribosome profiling data are used to study specific aspects of translation such as ribosome decoding rates and start codon selection. This is because it necessitates looking at a much narrower slice of footprints (for example only the footprints found at AUG codons as opposed to footprints aligning to the coding regions). It may be very difficult to detect these biases, to quantify their influence and to prevent their occurrence.

A study in the Weissman lab found that the interaction of the Shine-Dalgarno with the translating ribosome was a significant driver of pausing of ribosome elongation (Li, Oh, and Weissman 2012). I took part in a study which examined whether this ribosome pausing may have been an artifact driven by an atypical read distribution. We suspected this to be the case owing to strong evidence that the Shine-Dalgarno interaction occurs over a range of positions and that it may have a “spring” like influence on the ribosome movement during frameshifting (Larsen et al. 1994). Our study did indeed find that the data were consistent with our conjecture of progressive movement of the ribosome decoding center along mRNA without disruption of ribosome interaction with Shine-Dalgarno sequence. Nevertheless our study still supported the finding that the Shine-Dalgarno significantly

inhibited ribosome elongation. A later study from the Green lab provided evidence that the original study had a selection bias where longer fragments were overrepresented in the sequencing data (Mohammad et al. 2016). As a consequence of the selection bias the ribosome profiling data was overrepresented with footprints that aligned near the Shine-Dalgarno. They suggested that the Shine-Dalgarno interaction doesn't promote a significant level of pausing of the translating ribosome.

Other cases of selection bias include attributing a relatively high footprint density downstream of the start codon to biological reasons, and a significant overestimation of the range of TE (ribo-seq/mRNA). The former was later found to be caused by a protocol which allowed for continued initiation of ribosomes while inhibiting their elongation (Gerashchenko and Gladyshev 2014). The latter was caused by differences in the affinity (and therefore their quantification with sequencing) of different mRNA transcripts to poly(A) selection (Weinberg et al. 2016).

A significant effort has also been made to use ribosome profiling data as a means to examine ribosome decoding rates. In many aspects this is an objective to which ribosome profiling data is immensely suited as the relative read density of footprints at a position in a gene should be inversely related to its decoding rate. One may expect therefore that this would quickly become a solved problem, but it is still an active area of research, owing to difficulties in the confident identification of the relative influence of the factors that affect elongation rates.

We carried out a study with an aim of understanding the factors that influence the read distribution across the mRNA. We reported based on simulated ribosome profiling data that conventional normalization (division by average read density across the transcript) provides a surprisingly accurate measurement of the relative decoding rates in a ribosome profiling experiment. We introduced a modification of this approach to make it less sensitive to atypical high peaks (caused by ribosome pauses or potential artifacts) and applied it across 30 ribosome profiling datasets. Remarkably we found that the codon enrichments over the protected footprints can explain 20-30% of the variation in ribosome profile data. In comparison, the stochastic nature of ribosome profiling means that a replicate profiling experiment describes 50% of the variation. Nevertheless we had

considerable difficulties explaining the differences in codon enrichment from the profiling protocol. We concluded that bias is having a major influence on the distribution of reads across the coding regions. A contemporary study by the W. Press lab revealed that pretreatment with cycloheximide was one of the sources of bias (Hussmann et al. 2015).

In the 8 years since its development ribosome profiling has been used to uncover many different aspects of translational regulation especially focusing on cases of the major translational pathways which respond to significant stresses (oxidative stress, proteotoxic stress, starvation). The properties of an effective response to stress, (it should be rapid and significant) ensures that the response may be more easily delineated. The majority of regulation however is of “housekeeping” and this probably consists of continuous correction in response to relatively slow changes of the cellular environment and for cell maintenance. The mechanisms of such ‘housekeeping’ regulation may reduce the expression of one or many genes by a more limited degree which would correct over a larger time span. The role of translation in this regulation is not clear, and the detection of its effects may be more difficult. However increases in sequencing depth and greater temporal resolution (time series) should offer means to circumvent these issues.

The work in this thesis has been published and the reference is listed below.

Andreev, D. E., O'Connor, P. B. F., 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 Research, 45(2), 513–526.

Andreev, D. E., O'Connor, P. B. F., 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 Biology, 16, 90.

Andreev, D. E., O'Connor, P. B., Fahey, C., Kenny, E. M., Terenin, I. M., Dmitriev, S. E., Cormican, P., Morris, D.W. & Baranov, P. V. (2015). Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. Elife, 4, e03971. Journal Article.

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. Journal Article.

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. Journal Article

Insights into the mechanisms of eukaryotic translation gained with Ribosome Profiling

This chapter has been published as a review in Nucleic Acids Res 2016; 45(2):513-26; PMID:27923997; http://dx.doi.org/10.1093/nar/gkw1190

ABSTRACT

The development of Ribosome Profiling (RiboSeq) has revolutionized functional genomics. RiboSeq is based on capturing and sequencing of the mRNA fragments enclosed within the translating ribosome and it thereby provides a "snapshot" of ribosome positions at the transcriptome wide level. Although the method is predominantly used for analysis of differential gene expression and discovery of novel translated ORFs, RiboSeq data can also be a rich source of information about molecular mechanisms of polypeptide synthesis and translational control. This review will focus on how recent findings made with RiboSeq have revealed important details of the molecular mechanisms of translation in eukaryotes. These include mRNA translation sensitivity to drugs affecting translation initiation and elongation, the roles of upstream ORFs (uORFs) in the response to stress, the dynamics of elongation and termination as well as details of intrinsic ribosome behavior at mRNA transcripts after translation termination. As the RiboSeq method is still at a relatively early stage we will also discuss the implications of RiboSeq artifacts on data interpretation.

INTRODUCTION

It has been known for decades that the ribosome protects the portion of mRNA that is enclosed within its subunits from ribonuclease digestion (Castles and Singer 1969; Lazarowitz and Robertson 1977; Kozak 1977; Steitz 1969; Wolin and Walter 1988). The advent of massively parallel sequencing enabled the development of the RiboSeq technique which is based on the sequencing of these protected mRNA fragments on a large scale. Initially it was demonstrated for yeast cells (Ingolia et al. 2009) and later was applied to other species. Since its inception, ribosome profiling has been used to examine many aspects of translation in humans (Guo et al. 2010) and mice (Ingolia, Lareau, and Weissman 2011), fishes (Bazzini, Lee, and Giraldez 2012), insects (Dunn et al. 2013). Nematodes (Stadler and Fire 2011), as well as plants (Liu et al. 2013), protozoa (Caro et al. 2014), bacteria (Oh et al. 2011) and virus-infected cells (Stern-Ginossar et al. 2012).

The fragments that the elongating ribosome protects are normally 18-34 nucleotides long depending on the organism, the ribosome conformation state and the experimental protocol applied (Oh et al. 2011; Lareau et al. 2014; O'Connor et al. 2013; Rooijers et al. 2013; Ingolia 2016). Usually, the treatment of cytoplasmic extracts with ribonucleases is followed by the isolation of 80S particles which contain Ribosome Protected Fragments (RPFs) or ribosome "footprints". The positions of the mapped reads on the mRNA sequence display a 3 nucleotide periodicity reflecting the stepwise movement of the elongating ribosome over codons. This enables the detection of translated ORFs and their reading-frame transitions (Michel et al. 2012). In the course of RPF preparation, the ribosomes need to be blocked at mRNA. This is frequently done with antibiotics that arrest elongation (cycloheximide, emetine, chloramphenicol, anisomycin) or flash freezing in liquid nitrogen. The isolated RPFs are used to generate cDNAs that are subsequently sequenced using massively parallel sequencing platforms. An important control is to measure in parallel the level of mRNAs present in the cell (RNAseq). To this end, the mRNA is isolated from a proportion of the cytoplasmic extract, prior to nuclease digestion, and then randomly cleaved by alkaline hydrolysis to yield mRNA fragments (typically around the same length as ribosome footprints). The ratio of RPFs to RNAseq fragments mapped to each transcript is often used as an estimate of its relative translation efficiency (TE). The ribosome profiling protocol is described in dedicated publications (Ingolia et al. 2012; Reid, Shenolikar, and Nicchitta 2015) and commercial kits are currently available. Numerous *ad hoc* modifications of the protocols have been made.

RiboSeq is not the first genome wide approach to study translation. Polysome profiling quantifies mRNAs that are bound by multiple ribosomes (polysomes), which are isolated by sucrose gradient centrifugation. Here the RNA quantitation was initially achieved using microarrays (Johannes et al. 1999; Zong et al. 1999; Mikulits et al. 2000), and later using massively parallel sequencing. It is assumed that the presence of mRNAs in polysomes reflects translation of the annotated coding ORFs (acORF), while in reality, ribosomes may translate other ORFs. RiboSeq provides a more direct estimate of protein synthesis as it generates position specific information. This allows a relative measure of the number of ribosomes occupying either a specific mRNA region, such as the acORF (Figure 2.1A), the 5' leader (also known as 5'UTR) (Figure 2.1B), the 3' trailer (also known as 3'UTR)

(Figure 2.1C), or specific elongation pause sites (Figure 2.1D), or start and stop codons (Figure 2.1E).

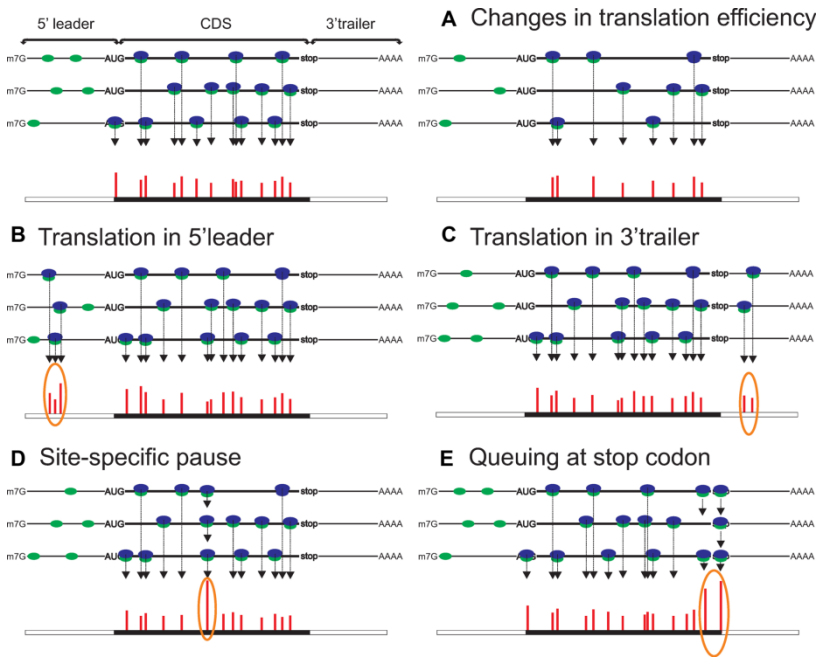


Figure 2.1. Types of alterations in ribosome density observed with ribosome profiling. A. changes in translation efficiency (TE), represented here by a decrease of TE (compare with left part); Left panel represent control conditions and the right panel corresponds to changed conditions B. Translation in 5' leader; C. Presence of ribosomes in 3' trailer; D. Site specific pause originating from ribosomes stalled within acORF at a specific location; E. Ribosomes pause at the stop codon and queued upstream ribosomes; Green and blue shapes represent 40S complexes and 60S subunits respectively, only 80S ribosomes (40S+60S complexes) produce footprints under conventional RiboSeq protocol. Red bars show the number of RPFs aligned to a particular location, due to biases of library preparation and sequencing the height of bars do not fully correspond to actual ribosome occupancies. Orange ovals highlight specific locations of ribosome footprints.

Novel applications of RiboSeq continue to emerge, recently RiboSeq was used for the analysis of translation in subcellular compartments (Jan, Williams, and Weissman 2014; Williams, Jan, and Weissman 2014), in certain cell types in the context of a whole organism (Janich et al. 2015; Gonzalez et al. 2014) and for the detection of translation quantitative trait loci (mRNA components that influence translation rate) (Cenik et al. 2015; Battle et al. 2015). Ribosome profiling is now further empowered by a wide spectrum of computational resources. At RiboSeq.Org, GWIPS-viz (Michel et al. 2014; Michel et al. 2015) provides visualizations of genomic alignments for publicly available RiboSeq data aligned to the

genomes of a dozen of organisms and RiboGalaxy (Michel et al. 2016) allows researchers to analyse their data using a variety of tools from such packages as RUST (O'Connor, Andreev, and Baranov 2016), riboSeqR (Chung et al. 2015) and RiboTools (Legendre et al. 2015). Publicly available RiboSeq data also can be accessed at RPFdb (Xie et al. 2016), sORFs.org (Olexiouk et al. 2016) (ORFs whose translation is supported with RiboSeq data), and TISdb (Wan and Qian 2014)(translation initiation sites predicted from RiboSeq data). Several stand alone tools were developed for differential gene expression analysis (anota (Larsson, Sonenberg, and Nadon 2011), Babel (Olshen et al. 2013), RiboDiff (Zhong et al. 2016), Xtail (Xiao et al. 2016)), and for the prediction of translated ORFs (RiboTaper (Calviello et al. 2016), ORF-Rator (Fields et al. 2015) and riboHMM (Raj et al. 2016)). Ribomap (H. Wang, McManus, and Kingsford 2016) estimates footprint origins in the presence of multiple splice isoforms and repetitive regions and Rfoot (Ji et al. 2016) uses RiboSeq data to predict non-ribosome protected RNA sites. Additional RiboSeq specific pipelines include PROTEOFORMER (Crappe et al. 2015) and RiboProfiling (Popa, Lebrigand, Paquet, et al. 2016).

While the use of ribosome profiling has been reviewed extensively, see (Bartholomaus, Del Campo, and Ignatova 2016; Brar and Weissman 2015; Stern-Ginossar 2015; N T Ingolia 2014; Kuersten et al. 2013; Michel and Baranov 2013; R. Jackson and Standart 2015), this review focuses on how RiboSeq has been used to characterize the mechanisms of initiation, elongation, termination and ribosome recycling specifically in eukaryotic cells.

INITIATION

mRNA recruitment to the 43S preinitiation complex

The 43S pre-initiation complex (PIC), consists of a 40S ribosomal subunit loaded with Met-tRNA*eIF2*GTP (ternary complex, TC) and initiation factors eIF3, eIF1, eIF1A and eIF5. Its recruitment to mRNA is preceded by recognition of the mRNA m^7G cap by initiation factor eIF4E. eIF4E operates as a part of the trimeric complex eIF4F (eIF4E, eIF4G and eIF4A) where eIF4G has a role as a scaffold protein and eIF4A is a helicase that unwinds base

paired nucleotides within 5' leaders of mRNA during 43S PIC scanning (for recent reviews see (Hinnebusch, Ivanov, and Sonenberg 2016; Hinnebusch 2014)).

The recognition of m^7G cap by eIF4F is under stringent control. Some proteins can sequester eIF4E from eIF4G, and thus act as translational repressors. The most well studied translational repressors are mammalian eIF4E binding proteins (4EBPs). Hyper-phosphorylated 4EBPs do not bind to eIF4E, thereby allowing eIF4F to be engaged in translation. The phosphorylation of 4EBP in mammalian cells is mostly carried out by the mTORC1 complex. mTORC1 includes mTOR kinase which plays a central role in proliferation and cell growth. Once mTORC1 is inactivated, rapid dephosphorylation of 4EBPs occurs; 4EBPs then sequester eIF4E from eIF4F thereby inhibiting cap-dependent translation (Fonseca et al. 2014; Raj et al. 2016; Thoreen 2013; Cencic et al. 2009; Wang and Proud 2006).

The interactions of eIF4E with the mRNA cap and eIF4G were considered to be of paramount importance for the translation of almost all cellular mRNAs. Disruption of either interaction is believed to produce a global suppression effect on bulk protein synthesis. However, recent RiboSeq data suggests that disruption of the eIF4E-eIF4G interaction by 4EBP leads to a highly differential effect on translation and that global translation is less inhibited than may have been expected (Shatsky et al. 2014).

Two independent RiboSeq studies (Hsieh et al. 2012; Thoreen et al. 2012) examined the translational response to mTOR inactivation. Mammalian cells were briefly treated with inhibitors PP242, INK128 and Torin-1, which target the catalytic subunit of mTOR and thus activate 4EBPs more efficiently than the drug rapamycin, that acts allosterically (Choi et al. 1996). The global level of translation was significantly reduced, about 40% relative to control conditions (Thoreen et al. 2012). Around 150 mRNAs were inhibited more than 4-8 folds relative to the global translation. The majority of these mRNAs (Perry and Meyuhas 1990) bear a 5' Terminal Oligopyrimidine Tract (5' TOP) motif. In addition to 5' TOP, another motif termed Pyrimidine Tract Responsive Elements (PTREs) was proposed (Hsieh et al. 2012). This element, unlike 5' TOP motifs, possesses an invariant uridine flanked by pyrimidines and, importantly, does not reside at the +1 position of the 5' leader. It was proposed that 5' TOP and PTRE motifs (and perhaps others) may act synergistically. Many

"mTOR sensitive" mRNAs encode components of the protein synthesis apparatus which are in particular demand during active growth and proliferation; consistent with observations that mTOR also regulates both rRNA and tRNA transcription (Mayer and Grummt 2006; Kantidakis et al. 2010; Tsang, Liu, and Zheng 2010). Two other classes of non-TOP "mTOR-sensitive" mRNAs were identified recently (Gandin et al. 2016). The first includes certain mRNAs with long 5' leaders, which mostly encode survival- and proliferation-promoting proteins (e.g., *CCND3*, *ODC1*, *MCL1*, *BIRC5*, *MYC*). The second class is a subset of mRNAs with very short 5' leaders, which mostly encode proteins with mitochondrial functions. Interestingly, these mRNAs are enriched with TISU (Translation Initiator of Short 5' UTR) initiation contexts, which provide unusual requirements for translation initiation components (Elfakess et al. 2011; Sinvani et al. 2015). These TISU elements are not the cause of the mTOR sensitivity however (Gandin et al. 2016). The fact that these "mTOR-sensitive" mRNAs were not identified in prior RiboSeq studies was claimed to be due to technical reasons, including insufficient sequencing depth (Gandin et al. 2016). Continuous translation of mRNAs under conditions of mTOR inhibition can be explained by residual levels of active eIF4F, but reliance on other ribosome recruitment factors is an intriguing possibility.

Although the role of eIF4E has been actively addressed with RiboSeq, many aspects of 43S PIC recruitment have yet to be elucidated. For instance, we do not know the role of the different initiation factor paralogs on the regulation of translation. In humans, there are two different eIF4F scaffolds, eIF4G1 and eIF4G3 (a third isoform, eIF4G2 a.k.a. Dap5, lacks an eIF4E binding site), three eIF4Es, and also three 4EBPs which may have different preferences to each other and to certain mRNAs. We also still have not delineated the importance of other executors of 4EBP phosphorylation, for example cyclin-dependent kinase 1/cyclin B1 was reported to do so during mitosis (Shuda et al. 2015). Finally, apart from "canonical" 4EBPs there are a number of alternative factors which may regulate m7G cap recognition by eIF4E (reviewed in (Topisirovic and Borden 2005)). RiboSeq is ideal for exploring these understudied aspects of 43S PIC recruitment.

Old and new players that participate in scanning 5’ leaders

The 43S PIC scanning along 5’ leaders bearing secondary structures requires helicase activity (for review, see(Parsyan et al. 2011)) and eIF4A is believed to be a key component in mRNA leader structure unwinding.

A few studies addressed the effect of eIF4A inhibition on genome wide protein synthesis in mammalian cells. In two studies, the inhibition was done by silvestrol, which belongs to the flavagline family of plant natural products (Rubio et al. 2014; Wolfe et al. 2014). This drug increases binding of eIF4A to RNA (Cencic et al. 2009) and thus impairs the activity of eIF4F. Genome wide analysis of “silvestrol-sensitive” translationally downregulated mRNAs revealed that their 5’ leaders are usually long and structured. In both studies, the number of these mRNAs was relatively low (284 in case of (Rubio et al. 2014)). Specifically, the 5’ leaders of “silvestrol-sensitive” mRNAs were shown to be enriched in G-quadruplexes. These structures are considered to be efficient inhibitors of translation, and their unwinding upon scanning requires higher helicase activity. Interestingly, many “silvestrol-sensitive” mRNAs encode oncogenes and cell-cycle regulators, which aligns well with the anti-proliferative properties of this drug. A more recent RiboSeq study with another flavagline closely related to silvestrol, rocaglamide A, led to a different conclusion. It was found that this drug specifically clamps eIF4A onto polypurine sequences, and secondary structures within 5’ leaders are only minor determinants for rocaglamide A selectivity (Iwasaki, Floor, and Ingolia 2016). Whether silvestrol and rocaglamide A possess different selectivity is not clear.

Recently a study carried out RiboSeq after treating cells with Pateamine A, which also targets eIF4A (Popa, Lebrigand, Barbry, et al. 2016). The authors report that this treatment potently blocks translation and resulted in a widespread reduction in RPFs on coding sequences with a median inhibition of 22 fold. This effect was independent of the length of the 5' leader. Pateamine A has a different mode of action than flavaglines in that it is responsible for inhibition of the interaction between eIF4A and eIF4G (as opposed to reducing its helicase activity (Low et al. 2005)). It is of interest whether physiological modulation of mammalian eIF4A availability by means of PDCD4, which competes with eIF4G for eIF4A binding (Goke et al. 2002; H. S. Yang et al. 2003), would produce a pattern of regulation similar to these pharmacological treatments.

Whether other helicases and accessory proteins can complement or even substitute for eIF4A is unknown. RiboSeq was applied to yeast strains with mutations in *DED1* (*DDX3* is the mammalian orthologue) or *TIF1* (*EIF4A1* is the mammalian orthologue) (Sen et al. 2015). Both mutations resulted in a significant impairment of general translation as evidenced by a reduction of bulk polysomes. Footprint densities in *DED1* mutants were significantly changed (relative to the average) for 814 transcripts (~17% of expressed genes). The majority of these genes displayed a reduction in TE, indicating that >10% of all mRNAs revealed a hyperdependence on Ded1 for efficient translation. The 5' leaders of Ded1-dependent mRNAs are typically long and predicted to have a greater than average propensity to form secondary structures. In contrast, despite a similarly strong reduction in general translation observed for an eIF4A mutant, translation of only 36 mRNAs was hyperdependent on an eIF4A orthologue, implying that translation of most mRNAs were reduced comparably by inactivation of eIF4A. Further analysis performed with reporter constructs suggests that the eIF4A orthologue, assisted by Ded1, is required for binding of the 43S PIC at the 5' ends of mRNAs harboring cap-proximal RNA secondary structure.

Later this research was substantially developed by the same group to investigate the role of eIF4B, a cofactor of eIF4A which stimulates its helicase activity. To this end, the effect of *TIF3* (eIF4B is the mammalian orthologue) mutation in yeast was investigated (Sen et al. 2016). Similar to Ded1 and TIF1, TIF3 depletion resulted in a strong decrease in bulk translation, with 111 hyperdependent and 48 hypodependent mRNAs identified. The change in gene expression was greater (and considerably different) for the TIF3 mutant than the TIF1 mutant which was unexpected given that TIF3 is a cofactor of TIF1 and that *TIF3* is not an essential gene in yeast. In addition the subsets of eIF4B-hyperdependent (111 mRNA) and eIF4A-hyperdependent mRNAs (36 mRNAs) overlap in only 3 mRNAs. This implies that eIF4B may have an alternative function or acts independently of eIF4A, as many eIF4B-hyperdependent mRNAs have "standard" requirements for eIF4A.

Translation Initiation Site (TIS) selection

Once the 43S PIC is successfully loaded onto mRNA and is in the scanning mode, it searches for an appropriate initiation codon. One of the general conclusions from the very first RiboSeq data was widespread alternative initiation both at AUGs and near-cognate

start codons. This was initially observed by RPFs derived from the mRNA sequences which were annotated as 5' leaders. Later, more specialized approaches were used to precisely map translation initiation sites genome wide. These approaches are based on the properties of particular antibiotics such as harringtonine (Ingolia, Lareau, and Weissman 2011), and later, lactimidomycin (Lee et al. 2012). Both of these drugs block elongation of *de novo* assembled 80S initiation complexes, whereas those 80S already engaged in elongation are believed to be refractory to these treatments. In another approach treatment with puromycin followed by cycloheximide was applied in order to release peptide-elongating ribosomes from mRNAs and then to block ribosomes during the first elongation step (Fritsch et al. 2012). These studies found that widespread non-AUG initiation occurs in eukaryotes. The occurrence of such initiation was not surprising as the ability of the translation initiation apparatus to recognize non-AUG codons as initiation sites is known from pioneering work by M. Kozak (Kozak 1989). However, its apparent frequency was unexpected as analysis using classic molecular biology techniques suggested that such initiation is generally inefficient.

Recent computational estimation of TIS strength from several published RiboSeq datasets provides potential explanations of this discrepancy (Michel, Andreev, and Baranov 2014). Using quantitative estimations of initiation efficiencies, AUG codons were confirmed to be much better initiators than near-cognate codons. The discrepancy between widespread occurrence of non-AUG initiation and its low efficiency is easily resolved once their locations are examined. In the vast majority of cases non-AUG initiation is observed upstream of AUG codons consistent with "leakiness" of the former and "tightness" of the latter.

The influence of nucleotide context surrounding a TIS on the efficiency of initiation is well documented (Noderer et al. 2014; Kozak 1986), especially at positions -3 and +4 with respect to the first (+1) nucleotide of the start codon. However the correlation of the context strength and TIS efficiency in RiboSeq studies was weak (Michel, Andreev, and Baranov 2014). This could be due to technical artifacts (see below), it is also possible that the effect of the nucleotide context is confounded by the other mRNA elements that influence the efficiency of translation initiation. These may include specific secondary structures and RNA binding proteins interfering with the progression of scanning

complexes, obscuring some potentially strong initiation codons and enhancing the weak ones. To conclude, even with the findings of the recent RiboSeq studies, at present we cannot reliably predict the efficiency or even the precise location of translation initiation in particular mRNAs from its sequence alone.

Regulation of translation by upstream open reading frames

Almost all RiboSeq studies reported higher footprint densities for 5' leaders than for 3' trailers indicating translation upstream of acORFs, although some can be attributed to the synthesis of N-terminally extended proteoforms (Ivanov et al. 2011). Upon various stress conditions, the percentage of RPFs at 5' leaders, typically less than 1% (of all footprints aligned to the transcriptome) in mammalian cells, increases relative to acORFs. Translation of uORFs may be expected to be repressive. Sometimes reciprocal changes in translation of uORFs and acORFs can be observed at the level of individual genes. However, this is not a general phenomenon. Alterations in ribosome density at uORFs also occur during conditions, unrelated to stress e.g. during meiosis in yeast (Brar et al. 2012), cell cycle (Stumpf et al. 2013) and circadian oscillations in gene expression (Janich et al. 2015).

The events that trigger alterations in uORF translation are not fully understood. One possibility is that the activity or availability of those translation initiation factors responsible for initiation codon selection changes upon stress. In fact, upon oxygen and glucose deprivation (OGD), extensive translation regulation was observed for mRNAs encoding initiation factors eIF1, eIF5, eIF1A and its paralogs, all of which are known as modulators of initiation codon selection. This may reflect ongoing reprogramming of the translation initiation apparatus (Andreev et al. 2015). It is not well known whether or how the activities of these initiation factors are regulated, except for the fact that both eIF1 and eIF5 are subject to post-translational modifications which may affect initiation codon selection (Homma et al. 2005; Zach, Braunstein, and Stanhill 2014; Majumdar et al. 2002). Intriguingly, both factors regulate their own translation (Ivanov et al. 2010; Loughran et al. 2012).

Much more is known about regulation of eIF2, another principal player in TIS selection, and about the consequences of its regulation on translation. eIF2 forms the ternary complex (TC) with GTP and Met-tRNAi and is loaded onto the 40S ribosome to

enable it to recognize a start codon, after which eIF2*GDP is released. GDP is then recycled to GTP by guanine exchange factor (GEF), eIF2B, to enable another round of initiation. Various stress conditions trigger the integrated stress response (ISR). Upon ISR, any of four kinases, HRI, PKR, PERK or GCN2, phosphorylate the alpha subunit of eIF2 at Ser51 (Baird and Wek 2012). eIF2B forms a stable complex with the phosphorylated eIF2 and is unable to rapidly recycle GTP. Phosphorylation of a modest number of eIF2 molecules rapidly reduces the limited pool of active eIF2B resulting in the general inhibition of protein synthesis. Some mRNAs are known to evade such translation repression, and the translation of their uORFs plays a critical role in this control.

To explore this regulation genome wide, a few research groups applied RiboSeq under conditions of ISR. One study found that upon robust phosphorylation of eIF2, triggered by treatment with arsenite, acORF translation was greatly suppressed except for only a dozen mRNAs, all of which contained translated uORFs (Andreev et al. 2015). Another study found that a very similar set of mRNAs was identified to be resistant to the unfolded protein response (UPR) induced with tunicamycin, a small molecule that blocks N-glycosylation and causes endoplasmic reticulum (ER) stress (Sidrauski et al. 2015), which also results in eIF2 phosphorylation. Strikingly, differential translation upon UPR can be reversed upon treatment with a small molecule ISRIB, which potentiates the activity of eIF2B irrespective of eIF2's phosphorylation status (Sidrauski et al. 2015; Sidrauski et al. 2013; Sekine et al. 2015).

In these studies, the number of preferentially translated mRNAs was low in comparison to all mRNAs with translated uORFs. The common characteristic of the resistant mRNAs is poor translation of acORFs and the presence of at least one efficiently translated uORF under normal conditions (Andreev et al. 2015). Under stress conditions uORF translation is decreased whereas translation of the acORF is modestly enhanced or at least is not changed. Interestingly, in several cases (for example – for *DDIT3, IFRD1* and *PPP1R1B* mRNAs) a single uORF appeared to be sufficient for eIF2 mediated translation control (Andreev et al. 2015; Palam, Baird, and Wek 2011). It is not clear why particular uORFs involved in the translation control are mediating regulation by eIF2 inactivation. It can be speculated that in addition to uORFs, other properties of leaders are responsible for translation control upon eIF2 phosphorylation (Andreev et al. 2015). It appears that TC

deficiency does not influence TIS recognition directly, as no alterations in AUG *vs*. non-AUG codon initiation were observed in general.

A recent computational study of available ribosome profiling data (Chew, Pauli, and Schier 2016) that examined the relationship between the features of the 5' leader and acORF TE concluded that while uORFs are generally repressive they are minor determinants of global acORF TE. Thus it is possible that only a minority of translated uORFs are strong repressors of acORF translation, while the majority are not repressive due to leaky scanning, 43S sliding and/or reinitiation (Hinnebusch 2014; Jackson, Hellen, and Pestova 2012; Terenin et al. 2016). It is also likely that individual uORFs may work as translation regulators sensing specific conditions as exemplified by polyamine levels sensing by a uORFs in *AMD1* (Ruan et al. 1996) and *AZIN1* (Ivanov, Loughran, and Atkins 2008) 5' leaders, and by a uORF-based Mg2+ concentration sensor in a leader of *TRPM7* mRNA, encoding a magnesium channel (Kochetov et al. 2005). Moreover, any alterations in translation elongation, termination and recycling, discussed below, could also affect initiation at mRNAs containing translated uORFs. Given that about half of mammalian leaders possess uAUGs (Kochetov et al. 2005), and non-AUG initiation also occurs, a sophisticated interplay between all stages of translation may occur for uORF containing mRNAs, (see (Hinnebusch, Ivanov, and Sonenberg 2016) for a recent review).

TCP-Seq: RiboSeq of scanning and initiating ribosomes

An important limitation of conventional RiboSeq is the absence of information regarding the intrinsic behavior of scanning ribosomes. To overcome this, Archer and colleagues developed TCP-Seq (Archer et al. 2016), a variation of RiboSeq to detect the footprints protected by 40S ribosome complexes in addition to 80S ribosomes. It is based on formaldehyde crosslinking followed by RNase treatment and isolation of 40S/48S and 80S crosslinked intermediates. TCP-seq was applied to yeast and its results supported a model where multiple ribosomes can scan the 5' leader simultaneously. It was shown that scanning ribosomes pause within 5' leaders without the apparent involvement of secondary structures, and that initiating ribosomes have at least three conformations (protecting fragments of 19, 29 and 37 nt). These footprints share the same 5' end but differ at their 3' ends, reflecting changes at the entry channel following start codon recognition. Furthermore, footprints from lingering 40S subunits at termination codons were also

identified which may prove useful for identifying translation of overlapping ORFs. The development of TCP-Seq for mammalian cells could bring important insights into mammalian translation control, which is especially interesting as mammalian 5' leaders are in general much longer than in yeast.

ELONGATION

Once initiation of translation is accomplished, the ribosome starts to synthesize protein. The rate of protein elongation in living cells has been estimated from experiments with reporter constructs, but until recently it was not possible to investigate it at a global level.

An elegant approach using RiboSeq was applied to calculate the average elongation rate in mammalian cells, where cells were pulse-treated with a drug to selectively stop new initiation followed by RiboSeq at several brief time intervals. This delay creates an area of low ribosome density immediately downstream of start codons whose length is proportional to the interval between the treatments. In this experiment, the mean elongation rate for mammalian cells was estimated to be 5.6 codons per second. This average speed was described to have little variance among all mRNAs irrespective of the length of their coding sequences and TE. Dana and Tuller, however, argued that the rate of elongation is slower at the beginning of coding regions, owing to differences in codon adaption to the tRNA pool (Tuller et al. 2010; Dana and Tuller 2012).

Metagene profiles (an average profile from many transcripts) of many ribosome profiling studies show increased ribosome density at TIS and termination codons, and a consistent elevation of ribosome density in the first few hundred codons after which the density of ribosomes along the coding sequence is often uniform until the end of the acORF. However, under certain conditions this distribution can be altered. Indeed, the accumulation of ribosomes at around codon 65 on most mRNAs during severe heat stress was observed in both mouse and human cells (Shalgi et al. 2013). This was shown to be linked to chaperone (HSP70) function. Thus, the interaction of the chaperone machinery with the nascent peptide is an important factor that modulates the velocity of the ribosome along the mRNA. Similar findings were made for the translation response to proteotoxic stress induced by the amino acid analog L-azetidine-2-carboxylic acid (AZC) which competes

with proline during amino acid incorporation (Liu, Han, and Qian 2013). A case of global deviation from uniform ribosome distribution towards the end of the acORFs was observed in PC12 cells (interestingly – under normal conditions). A metagene profile demonstrated a pattern of RPF close to the stop codon that most likely represents queuing of ribosomes. This queuing disappears upon OGD, which may be the case if either the elongation rate reduces, or the rate of termination/recycling increases, or both (Andreev et al. 2015).

The interaction of the signal recognition particle (SRP) with the nascent polypeptide as it emerges from the 60S ribosome subunit tunnel was speculated to cause pausing of translation. Pechmann et al. found that *in vivo* the co-translational recognition of the signal amino acid sequence (SS) or transmembrane domain (TMD) by the SRP is enhanced when the mRNA encoding the secretory polypeptide contains a cluster of non-optimal codons (~35–40 codons) downstream of the SRP-binding site, i.e. at a distance sufficient to span the ribosomal exit tunnel (Pechmann, Chartron, and Frydman 2014). This led the authors to speculate that the moment of enhanced SRP interaction is controlled by a genetically programmed slowdown in elongation when the SS emerges from the ribosome and this slow-down is determined by a cluster of non-optimal or "slow" codons. In contrast, a later RiboSeq study by the same group (Chartron, Hunt, and Frydman 2016) suggests that in general elongation arrest is not associated with SRP binding. Instead, perhaps in order to avoid such arrest, SRPs are directed to ribosomes translating the mRNA even before translation of the SS/TMD encoding region in a large subset of the mRNAs recognized by the SRP. For some of the mRNAs this recognition was driven through signals contained in their mRNAs 3' trailers (Chartron, Hunt, and Frydman 2016).The pausing of ribosomes however was still observed for mRNAs encoding non secretory proteins that are targeted by the SRP.

It is believed that for most mRNAs initiation is the rate limiting step of translation. Thus, the nature of a force responsible for codon bias is puzzling, it has been suggested, for example, that codon usage optimizes global, rather than local translation (Shah et al. 2013). Yet another possibility was recently provided by Presnyak et al. who showed that in yeast, the codon optimality index is a very strong predictor of mRNA stabilities, and that optimal codons in coding sequences are linked to faster decoding rates and *vice versa* (Presnyak et al. 2015). In the follow-up study, Radhakrishnan et al provided mechanistic insight into how

elongation rates affect mRNA stability in yeast (Radhakrishnan et al. 2016). This was linked to DEAD-box protein Dhh1p, which targets slowly translated mRNAs for decay. Strikingly, the effect of non-optimal codons on mRNA stability was shown to depend on their locations within acORF (the "polarity" effect). In the context of reporter mRNA, the stretch of non-optimal codons shows greater effect when it is placed closer to the 5' end of the acORF. This was explained by the fact that more ribosomes accumulate upstream and thus recruit more Dhh1p, which triggers more efficient degradation. Therefore, prediction of mRNA half-life may be further improved by codon position-specific information.

Whatever the effect of local decoding rates on global translation might be, their alteration can influence cell fitness and phenotype. It was found that the loss of the anticodon wobble uridine (U_{34}) modification (mcm^5s^2U) in a subset of tRNAs leads to ribosome pausing at cognate codons in *S. cerevisiae* and *C. elegans* (Nedialkova and Leidel 2015). Interestingly, upon loss of the U_{34} modification, decoding of only CAA and AAA triplets is affected (becomes slower), whereas other codons that are decoded by U_{34} bearing tRNAs are not affected. This abnormality elicits a widespread protein aggregation *in vivo*.

Availability of certain amino acids (and hence aminoacylated tRNAs) for protein synthesis may be limiting for certain tumors. Using Ribosome Profiling, Loayza-Punch and colleagues (Loayza-Puch et al. 2016) demonstrated that proline levels are restrictive for certain kidney cancers because of insufficient levels of corresponding aminoacylated tRNAs. This is compensated with high levels of PYCR1, the key enzyme of proline biosynthesis pathway, in these tumors. In support of this observation, knocking out *PYCR1* impedes tumorigenic growth.

TERMINATION AND RECYCLING

Recent studies of translation termination and recycling in eukaryotes revealed that, as well as canonical release factors (eRFs), additional players are critically involved (R J Jackson, Hellen, and Pestova 2012; Pisarev, Hellen, and Pestova 2007; Nurenberg and Tampe 2013). Guydosh and Green (Guydosh and Green 2014) used RiboSeq to uncover the function of Dom34 in yeast. Dom34 (*PELO* and *pelota* are the human and fruit fly

orthologues, respectively), is a distant homolog of release factor eRF1. Dom34 was shown to dissociate stalled ribosomes *in vitro* with the participation of Hbs1 and Rli1 (Pisareva et al. 2011). However, its function *in vivo* was not clear. Comparison of RPFs from wild type cells with those from cells with *Dom34* knocked out (dom34Δ) allowed the authors to discover increased ribosome occupancy at one particular position within the *HAC1* mRNA (yeast orthologue of mammalian *XBP1*). *HAC1/XBP1* mRNA is known to be subjected to an unusual event - splicing in the cytoplasm that leads to the synthesis of a protein product encoded in an alternative frame relative to the unspliced variant (Yoshida et al. 2001; Sidrauski, Cox, and Walter 1996). The observation of ribosomes stalled at the very 3' end of the splicing intermediate in *Dom34* knockout cells allowed the authors to conclude that Dom34 normally rescues such stalled ribosomes. Dom34 was found to be unable to rescue ribosomes if they pause within a coding sequence enriched in proline residues or a stretch of histidines when cells are deprived of this amino acid. Another intriguing observation was the presence of ribosomes within the 3' trailers of some mRNAs in the knockout strain. They accumulated mostly at the border of the 3' trailers and the polyA tail and did not reveal any triplet periodicity. Strikingly, only a small percentage (11%) of mRNA species revealed such 3' trailer bound ribosomes, but specific features common to these mRNAs were not found. Thus, it seems likely that Dom34 also participates in termination/recycling on certain intact mRNAs.

These studies were substantially developed in research devoted to another ribosome recycling factor, Rli1 (ABCE1 in mammals) (Young et al. 2015). Rli1 is presumably the principal recycling factor in yeast cells. RiboSeq showed that in Rli1-depleted cells, 80S ribosomes accumulate at the stop codons and within the 3’ trailers of all yeast genes and to a much greater extent than in dom34Δ cells. These ribosomes are very likely engaged in translation, as dozens of endogenous peptides potentially encoded within the 3’ trailers were detected. The authors presented compelling evidence that the ribosome density in 3’ trailers originates from reinitiation events rather than from readthrough of the acORF stop codons,

The phenomenon of ribosomes travelling along 3' trailers remains mysterious. Miettinen and Björklund (Miettinen and Bjorklund 2015) showed that nucleases selected for RiboSeq affect the quantitative analysis of translatability of mRNAs. RPFs at the

beginning of acORF were more abundant in RNase I digests, than in micrococcal nuclease (MN) samples. Also curiously, digestion with MN yielded high amounts of RPFs localizing in 3' trailers. RPFs derived from 3' trailers have a different mean length in comparison to "normal" RPFs suggesting that if these footprints derive from ribosomes, they may traverse 3' trailers in a different conformation.

A well-known phenomenon that could explain the presence of ribosome footprints in 3' trailers is stop codon readthrough, where tRNAs amino acids are incorporated at stop codons. C-terminal extensions of proteins produced by readthrough may alter protein function (Eswarappa et al. 2014) and localization (Stiebler et al. 2014). Instances of this rare recoding phenomenon have been observed in many organisms (see (Baranov, Atkins, and Yordanova 2015) for a recent review on recoding) including humans (Loughran et al. 2014). However, this phenomenon was shown to be more frequent in *Drosophila* (Dunn et al. 2013) and *Anopheles* mosquitoes (Jungreis et al. 2016). Extensive readthrough was revealed which was consistent with earlier predictions based on phylogenetic analysis of fruit fly genomes (Jungreis et al. 2011). RiboSeq, applied to *Drosophila* embryos and a macrophage like lineage S2 cell line, revealed significant variations in readthrough efficiencies for the same mRNAs, which suggests that readthrough may be controlled in a cell type and development stage specific manner. The cause for the cell type-specific and mRNA specific readthrough is not known.

Stress dependent alterations in readthrough efficiencies were observed in mammalian cells upon OGD (Andreev et al. 2015). For 19 mRNAs, relatively high RPF densities were observed downstream of acORF stop codons under normal conditions, and the triplet periodicity of the signal strongly suggests that ribosomes do translate these regions in the same reading frame. In all these cases the readthrough stop codon is UGA, known as the most prone to readthrough (Stiebler et al. 2014; Loughran et al. 2014; Schueren et al. 2014). Upon OGD, the efficiency of readthrough significantly decreased as early as 20 minutes. Although the cause of readthrough is not clear, it may be speculated that it may relate to hydroxylation of either termination factors or the ribosomal proteins (Feng et al. 2014; Loenarz et al. 2014) which should be reduced in the absence of oxygen.

In some yeast strains, termination factor eRF3 is known to form prion aggregates and this phenotype is described as *[PSI+]*. Baudin-Baillieu et al. took advantage of this and carried out comparative ribosome profiling of [PSI+] and [PSI-] strains to explore the role of eRF3 (Baudin-Baillieu et al. 2014). Readthrough in about 100 genes was identified in [PSI^{+}] cells. Curiously, [PSI^{+}] was also found to affect the accuracy of triplet decoding. Thus, the presence of this conformational form of termination factor eRF3 may have a broader impact on protein synthesis than just an effect on translation termination, however, this also could be due to secondary effects of deregulated protein synthesis.

Variations in genetic codes often arises as a result of stop-codon reassignment (Baranov, Atkins, and Yordanova 2015). Two independent studies (Heaphy et al. 2016; Swart et al. 2016) uncovered strikingly unusual decoding in the ciliate *Condylostoma magnum*, where all stop-codons are decoded as amino acids, except those in close proximity to polyA tails. Swart et al (Swart et al. 2016) carried out ribosome profiling in *C. magnum* and experimentally confirmed that this is indeed the case. Lobanov et al used ribosome profiling to explore global translation in the ciliate *Euplotes* where numerous instances of ribosomal frameshifting have been reported (Swart et al. 2016; R. Wang et al. 2016). It appears that, like *C. magnum*, termination of translation takes place only on stop codons in close proximity to polyA tails, while internal stop codons trigger either +1 or +2 ribosomal frameshifting depending on the preceding codon (Lobanov et al. 2016). Most probably, the proximity of the polyA tail is a prerequisite for termination in these instances. It was speculated that the polyA binding protein (PABP) bound at the polyA tail is needed for the termination and when the stop codons are distant from polyA tails they function as amino acid codons or frameshifting triggers.

At least some ribosomes that successfully accomplish polypeptide synthesis are likely to be engaged in subsequent initiation on the same mRNA. Indeed, growing evidence supports the mRNA closed-loop model which was suggested a long time ago (Gallie 1991; Tarun and Sachs 1995). It is now known that the interaction between eIF4G, bound to m^7G cap via eIF4E, and PABP, bound to polyA tail of mRNA, is responsible for the convergence of both mRNA termini. Intuitively, one feature which may account for the efficiency of closed-loop formation is the length of a particular mRNA. The recent RiboSeq study by Thompson et al (Thompson et al. 2016) provides evidence that the eukaryote-specific 40S ribosomal

protein Asc1/RACK1 is required for efficient translation of mRNAs with short ORFs in *S. cerevisiae*. The effect of Asc1/RACK on length-dependent translation was reproduced with reporter constructs. This specificity to short mRNAs may be explained by the hypothesis that efficiency of short mRNAs translation is more dependent on closed-loop complex formation. Accordingly, depletion of eIF4G mimics the translational defects of ASC1 mutants – where a similar dependence on mRNA length was revealed. How Asc1/RACK1 modulates translation mechanistically remains unclear.

RIBOSEQ ARTIFACTS

Notwithstanding its notable successes, it has become increasingly clear that RiboSeq data do not provide an unbiased representation of translation. Perhaps the most well-known reason for this is the sequence bias introduced during ribosome footprint library generation and its conversion to cDNA and subsequent sequencing. These procedures involve a number of reactions using enzymes with nucleotide sequence specificity. A recent survey of publicly available RiboSeq datasets revealed that in some of the datasets this sequencing bias has a greater influence on the distribution of reads across mRNAs than the identity of codons in the decoding center of the ribosome (O'Connor, Andreev, and Baranov 2016). Certain improvements have been made recently to reduce these biases with the optimization of protocol steps and use of enzymes with reduced sequence specificity.

A key area of ribosome footprint generation is the use of RNases to digest the unprotected mRNA, as these may degrade rRNA to different extents and display sequence specificity. A systematic study of digestion with various nucleases (I,A,S7 and T1) across different species found that T1 best preserved ribosome integrity while thoroughly digesting the polysomes (Gerashchenko and Gladyshev 2016). The weak pairwise correlations of the ribosome density profiles at individual transcripts obtained using 4 different RNases suggests that this could be a major distorting factor in RiboSeq data.

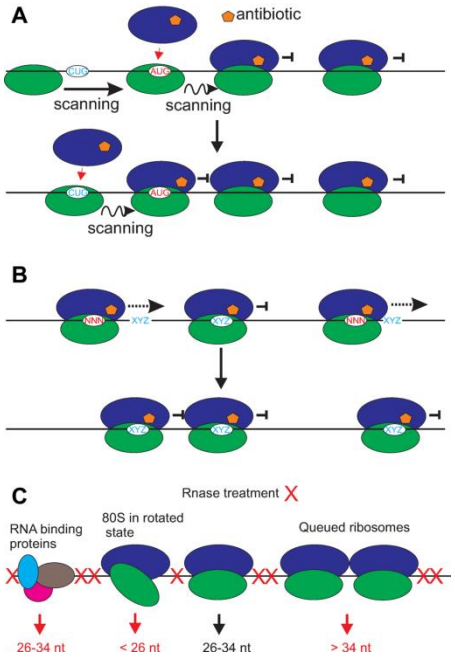


Figure 2.2. Examples of some known and potential artifacts associated with ribosome profiling. A. antibiotic pretreatments may cause increased initiation on upstream translation initiation sites (TISs) because arrested elongating ribosomes prevent scanning. B. Antibiotics may not block elongation completely (on NNN codon in the A-site) but allow slow elongation (until preferred XYZ codon is in the A-site), which affects the estimation of local decoding rates. C. Ribosome protected fragments shorter than 26 and longer than 34 nts may be excluded upon size selection. In contrast, mRNA protected by RNA binding proteins may be selected as RPFs.

The pretreatment of cells with antibiotics, which is used to stall the ribosomes at their locations prior to cell lysis, generates another important artifact. Notably, due to the differences in experimental protocols, drug pretreatment times in different studies range from 1 minute to as long as 30 minutes. Gerashchenko et al. reported that cycloheximide pretreatment was a major contributing factor to the observed increased read density at the start of coding sequences (Gerashchenko and Gladyshev 2014). This is likely an artifact due to cycloheximide inhibiting elongating ribosomes, but not initiating ones. An accumulation of ribosomes shortly downstream of initiation sites due to this phenomenon is therefore expected (Figure 2.2A). However, the increased density is not wholly caused by this artifact as it has also been observed in datasets produced without cycloheximide pretreatment (Weinberg et al. 2016). This increase of footprint density was earlier ascribed as evidence of slower elongation rates at these regions where codon bias was found to be different from other parts of coding regions (Tuller et al. 2010).

Pretreatments with cycloheximide, as well as with elongation inhibitors that preferentially block ribosomes at initiation codons, i.e. lactimidomycin and harringtonine, may have even more serious consequences on the rate of initiation itself and even on the identity of codons identified as translation initiators. The successful initiation at a particular codon requires eIF1 dissociation from the PIC and the initiation rate at a potential start codon increases if the PIC is slowed down at this codon (Hinnebusch 2014). In this case, the arrest of a translating ribosome shortly downstream of a potential start codon would result in pausing of the next scanning PIC and artificially increase the initiation rate at codons which are not normally recognized. At present it is difficult to estimate the significance of this potential artifact, therefore it is advisable to be cautious when interpreting the initiation rate estimates and the initiation sites obtained from ribosome profiling data. This potential artifact may artificially overstate the amount of TISs and translated uORFs. (Figure 2.2A)

Furthermore different antibiotics have a preference for blocking ribosomes at specific codons (Lareau et al. 2014; Requiao et al. 2016). Hussmann et al. have shown that after treatment with cycloheximide, ribosomes may continue elongation in a sequence dependent manner and demonstrated that this is dependent on cycloheximide concentration (Hussmann et al. 2015). Consequently, the read density at codons of the decoding center may not be representative of codon decoding rates in cycloheximide treated samples (Figure 2.2B). Importantly, a high codon specificity is also observed for the "initiation inhibitor" harringtonine – it appears that this antibiotic preferentially arrests ribosomes at Lys, Arg or Tyr codons. This may seriously impact its implication for mapping TISs.

Although such artifacts are undesirable they are generally considered to be relatively unimportant for differential expression analysis as they are thought to occur in both conditions. However, for some conditions this may not be the case. The local footprint density is a product of the local codon decoding time and initiation rate. In the steady state the latter is equal to the rate of protein synthesis and if local decoding rates do not change, changes in footprint density for the entire mRNA equates to changes of TE. However this may not always be the case. Consider a mRNA that is occupied with stalled ribosomes. While the footprint density at such a mRNA would be high, the rate of protein synthesis

would be close to zero. This limitation of footprint density as a proxy for the rate protein synthesis can be observed in real data. In the ciliates *Euplotes*, sites of ribosomal frameshifting are more frequent in lowly expressed genes. Frameshifting is slow and leads to ribosome pauses at the frameshifting sites. Consequently, TE calculated as the number of footprints divided by the number of RNA-seq reads is higher for lowly expressed genes (Lobanov et al. 2016). Thus, it is important to note again that RiboSeq only provides a measurement of the number of elongation ribosomes on an mRNA but does not say anything about the elongation rate of ribosomes. Because of this, using the TE metric as a measure of protein production per mRNA should be used with some caution.

The current RiboSeq workflow usually consists of the exclusive selection of reads of an expected length. Although this has significant advantages in avoiding additional rRNA contamination, such a restricted selection is likely to conceal part of a translation response (Lareau et al. 2014) (Figure 2.2C). For example, it is known that a pair of stalled ribosomes protect longer mRNA fragments from nuclease digestion (Wolin and Walter 1988). Thus, it is possible that selection of only ~30 nt fragments deplete the data from some footprints left by queued stalled ribosomes. The claim that Shine-Dalgarno sequences are a major determinant of ribosome speed in bacteria (Li, Oh, and Weissman 2012) was reported to be an artifact of biased selection for longer footprints (Mohammad et al. 2016) since Shine-Dalgarno interactions with the ribosome lead to longer footprints (O'Connor et al. 2013). Although no such variability has been found in eukaryotes yet, this case signals for caution. In a similar vein, the enrichment of mRNA-seq transcripts by the selection of polyA tails may conceal part of the transcriptional data. mRNAs with no or short polyA tails may be underrepresented as well as those undergoing degradation. This may have significant implications, for example Presnyak et al. calculated the half-life of ADH1 to be 4.2 minutes and 31.7 minutes for polyA selected and rRNA depleted RNA respectively (Presnyak et al. 2015). Thus, rRNA depletion as in (Weinberg et al. 2016) may be advantageous over polyA mRNA selection as it may provide a more accurate TE measure.

In a few studies RNase treatment was applied to polysomal, rather than total ribosomal fractions (Lee et al. 2012; Liu, Han, and Qian 2013; Aspden et al. 2014). This protocol modification is presumably based on the widely-held belief that 80S ribosomes found in monosome fraction are not engaged in translation. However, a recent report from

Heyer and Moore (Heyer and Moore 2016) found that some 80S ribosomes in monosome fractions are engaged in translation in *S. cerevisiae.* They were found mostly at uORFs and short acORFs. Importantly, mRNAs encoding low abundance regulatory proteins tend to be enriched in the monosome fraction. Therefore the exclusion of the monosome fraction may deplete the data of footprints derived from those mRNAs being translated by single 80S.

Finally, not all RNA fragments identified during ribosome profiling may arise due to genuine translation. They may possibly occur because of other heavy RNPs, in which the RNA is protected from ribonucleases (Figure 2.2C). One signature of translation that can be used for discriminating genuine translation from such artifacts is triplet periodicity, however, triplet periodicity might be distorted when overlapping ORFs are translated simultaneously. Ingolia et al (Ingolia et al. 2014) suggested a different approach – analysis of read length distribution. The small number of mRNAs with abnormal read length distribution and the large fraction of reads aligning to acORF suggests the extent of this artifact appears to be relatively low but it may be a significant cause of the reads aligning to RNAs annotated as non-coding (Ingolia et al. 2014).

PERSPECTIVES

Ribosome profiling has opened up a novel way for exploring global protein synthesis *in vivo* at great detail. Despite experimental difficulties, technical artifacts and discrepancies between the studies, the technique has demonstrated its utility and is certain to stay in the arsenal of omics tools. RiboSeq has undergone substantial improvements since its development but it is clearly not yet mature (O'Connor, Andreev, and Baranov 2016). Further optimization to minimize artifacts and starting material (ideally – to single cell level) is achievable.

RiboSeq has been instrumental in addressing the mechanistic aspects of mRNA translation and it will continue to be used for this purpose. For example, we expect that ribosome profiling will be fruitful in combination with recently developed genome editing techniques that enables modifications of known and putative translation factors to explore their roles in translation. RiboSeq is a powerful approach for studying genome wide effects of drug treatments. We expect therefore that it will be routinely used for exploring drug effects on gene expression, especially those that target translation. Finally, RiboSeq may be

significantly enhanced by a combination with other omics approaches such as RNA-protein crosslinking which allows mapping of mRNA binding proteins to the transcriptome (Sauliere and Le Hir 2015). The combination of the two may enable delineation of RNA binding proteins in translational regulation.

Translation of 5' leaders is pervasive in genes resistant to eIF2 repression

This chapter has been published as a research article in Elife, 4, e03971. Journal Article. https://doi.org/10.7554/eLife.03971

ABSTRACT

Eukaryotic cells rapidly reduce protein synthesis in response to various stress conditions. This can be achieved by the phosphorylation-mediated inactivation of a key translation initiation factor, eIF2. However, the persistent translation of certain mRNAs is required for deployment of an adequate stress response. We carried out ribosome profiling of cultured human cells under conditions of severe stress induced with sodium arsenite. Although this led to a ~4.5-fold general translational repression, the protein coding ORFs of certain individual mRNAs exhibited resistance to the inhibition. Nearly all resistant transcripts possess at least one efficiently translated uORF that repress translation of the main coding ORF under normal conditions. Site specific mutagenesis of two identified stress resistant mRNAs (PPP1R15B and IFRD1) demonstrated that a single uORF is sufficient for eIF2-mediated translation control in both cases. Phylogenetic analysis suggests that at least two regulatory uORFs (namely in SLC35A4 and MIEF1) encode functional protein products.

INTRODUCTION

Protein synthesis, as one of the most energy consuming processes in the cell, is under stringent regulation. In eukaryotes, the activity of many components of the translational machinery is modulated by various post-translational modifications in order to adjust either global or mRNA-specific translation. One of the better studied cases of translational control is the phosphorylation of eIF2 (Sonenberg and Hinnebusch 2009).

eIF2 forms the ternary complex (TC) with GTP and Met-tRNAi and is loaded onto the 40S ribosome to enable it to recognize a start codon, after which eIF2*GDP is released. GDP is then recycled to GTP by guanine exchange factor (GEF), eIF2B, to enable another round of initiation. During various stress conditions the cell triggers the integrated stress response (ISR) by activating any of four kinases, EIF2AK1 (also known as (a. k. a.) HRI), EIF2AK2 (a. k.

a. PKR), EIF2AK3 (a. k. a. PERK) or EIF2AK4 (a. k. a. GCN2) that phosphorylate the alpha subunit of eIF2 at Ser51 (Baird and Wek 2012). Instead of a rapid recycling, eIF2B forms a stable complex with phosphorylated eIF2. The concentration of eIF2 is higher than that of eIF2B, therefore even phosphorylation of a modest number of eIF2 molecules rapidly reduces the pool of active eIF2B resulting in the general inhibition of total protein synthesis (Hinnebusch 2014).

While the general suppression of translation conserves cellular resources, the active synthesis of certain factors is required to respond to the consequences of stress. Mammalian genes whose expression is known to evade translational arrest triggered by eIF2 phosphorylation include *ATF4* (Lu, Harding, and Ron 2004; Vattem and Wek 2004), *PPP1R15A* (a. k. a. *GADD34*) (Lee, Cevallos, and Jan 2009), *ATF5* (Watatani et al. 2008; Zhou et al. 2008; Hatano et al. 2013) and *DDIT3* (a. k. a. *CHOP*) (Jousse et al. 2001; Chen et al. 2010).

ATF4 and *ATF5* are believed to be regulated through the mechanism known as delayed reinitiation, initially characterized for the yeast *GCN4* (a functional analogue of *ATF4*) (Hinnebusch 1997). This requires at least two upstream open reading frames (uORFs). In ATF4 mRNA, after translation termination at the first uORF, the 40S resumes scanning albeit without the TC. The distance scanned by this ribosome subunit before it reacquires the TC, depends on TC availability. Under normal conditions most of the 40S are quickly reloaded with TC and therefore can reinitiate at the second uORF. Under stress conditions (i.e. low eIF2 availability) a larger fraction of 40S scan past the second uORF initiation codon before binding of the TC thereby enabling reinitiation at the next ORF (CDS).

A different mechanism of translational resistance, relying on the translation of a single uORF in the 5' leader, has been proposed for *DDIT3* (Palam, Baird, and Wek 2011). A fraction of scanning ribosomes recognize and initiate at the uORF initiation codon in a weak Kozak context. Under normal conditions with a high initiation rate the translation of this uORF inhibits leaky scanning by the obstruction of scanning ribosomes. Under stress conditions, the reduced ribosomal loading results in an alleviation of this obstruction.

For the examples mentioned above translational control is based on the reduced availability of TC. In specific cases initiation can occur without eIF2. Some viral mRNAs harbour internal ribosome entry sites (IRES) that allow translation initiation to take place by

recruiting alternative factors, i.e. eIF5B (Pestova et al. 2008; Terenin et al. 2008), eIF2D, and a complex of MCTS1 (a. k. a. MCT-1) and DENR (Dmitriev et al. 2010; Skabkin et al. 2010), or even initiate without Met-tRNAi and any initiation factors (Wilson et al. 2000). However, the existence of "viral-like" IRESes in mammalian mRNAs remains controversial (Shatsky et al. 2010; Jackson 2013).

The present work uses ribosome profiling (Ingolia et al. 2009) to explore the immediate effect of sodium arsenite treatment (NaAsO2), which results in a rapid phosphorylation of eIF2, on protein synthesis. This technique provides a snapshot of translating ribosomes over the entire transcriptome with subcodon resolution, see the reviews (Michel and Baranov 2013; Ingolia 2014).

RESULTS

Ribosome profiling

In order to generate the most informative conditions for characterizing eIF2-dependent mechanisms of translation regulation it was important to minimize the transcriptional response and induce significant, but not complete inhibition of translation. For this purpose, we chose to treat cells with sodium arsenite for a short time period and to monitor the immediate translational response. Sodium arsenite is a well-known potent inducer of eIF2 phosphorylation that activates EIF2AK1 (McEwen et al. 2005). We examined changes in the phosphorylation status of eIF2 and EIF4EBP1 during arsenite treatment to identify suitable conditions for ribosome profiling. The phosphorylated form of eIF2 (p-eIF2) progressively accumulates during the first 2 h of stress (Figure 3.1A). After 0.5 h of arsenite treatment, p-eIF2 reaches 30-40% of maximal levels of phosphorylation (Figure 3.1A). The dephosphorylation of EIF4EBP1 becomes evident 1 h after treatment and does not revert (Figure 1A). We did not detect changes in phosphorylation of p70 S6 kinase and its substrate RPS6 during arsenite treatment (Figure 3.1A). A robust accumulation of ATF4 is evident 30 min after treatment. These results suggest that 0.5 h post-treatment is likely to be suitable for examining eIF2 inhibition whilst minimising possible arsenite-induced; side effects. Furthermore, the number of ribosomes in polysome fractions was reduced by ~4.5-fold under these conditions (Figure 3.1B).

Therefore, HEK293T cells were treated with 40 µM sodium arsenite for 30 min before harvesting for ribosomal profiling which was carried out according to the Ingolia et al. protocol (Ingolia et al. 2012) with some modifications (see Materials and Methods).

Figure 3.1C-E shows the general characteristics of the ribo-seq and mRNA-seq datasets. As expected, in both conditions most ribo-seq reads were mapped to coding regions. The distribution of ribo-seq 5'-end reads, but not mRNA-seq 5'-end reads (same size randomly fragmented "naked" mRNA isolated from cytoplasmic lysate) exhibits the characteristic triplet periodicity. Over 7,000 mRNA sequences were uniquely mapped with at least 100 ribo-seq reads.

C

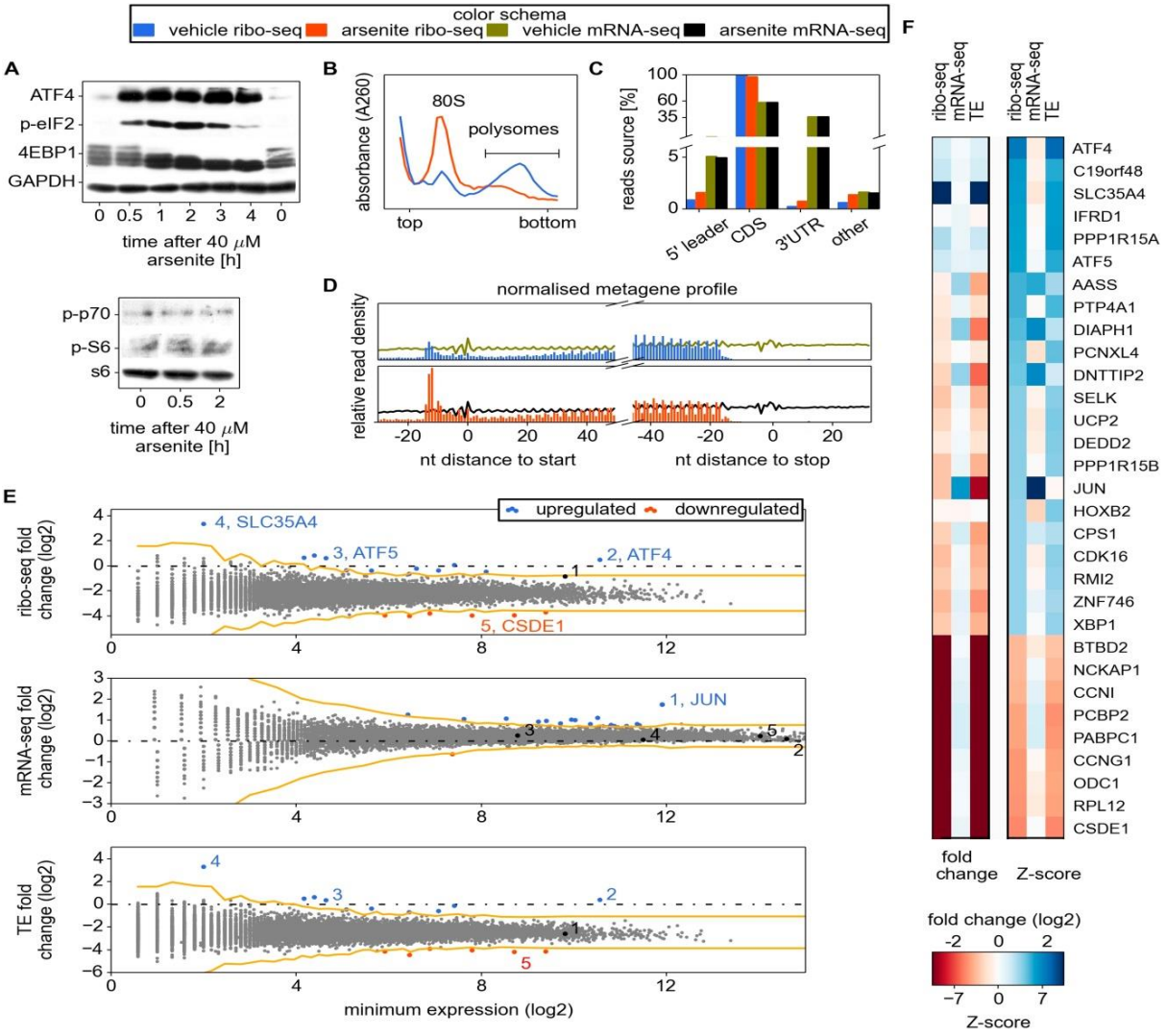


Figure 3.1. Analysis of differential gene expression under conditions of oxidative stress induced with sodium arsenite treatment. (A) Western blotting time series analysis of several protein components of HEK293T lysates after treatment with 40 μM sodium arsenite. (B) Sucrose density gradient profiles of HEK293T cells untreated and treated for 30 min with sodium arsenite at 40 μM. (C) Distribution of raw read counts over mRNA functional regions. (D) Metagene analysis, short reads from all mRNAs are aligned around 5' and 3' ends of CDS, transcript read density (RNA) is shown using curves, ribosome density (number of footprint reads) is shown using columns corresponding to the alignment locations of the read 5' ends. (E) Differential gene expression analysis. Scatter plots compare ribosome occupancy (top), transcript levels (middle) and translation efficiency, TE (bottom) between treated and untreated conditions. To avoid error due to uORF translation the number of ribo-seq reads aligning to the CDS only was used to determine the ribosome occupancy and TE. The *x* axis represents the normalized number of reads corresponding to the experiment/condition of minimal expression, see Experimental Procedures. The threshold used to denote differentially expressed genes (Z-score of 4) is indicated in orange. Certain genes of interest are indicated with numbers, followed by their gene symbols. (F) Two heatmaps displaying the fold change and Z-score for the top 22 most stress resistant and bottom 9 most stress sensitive genes, as estimated based on statistical significance of the change of their ribosome occupancy (ribo-seq Z-score).

Owing to the stochastic nature of massively parallel sequencing, the accuracy of an estimate of the level of expression of a gene is dependent on its sequencing depth. Therefore the estimated expression levels of weakly expressed genes have greater variability in comparison with highly expressed genes. To mitigate this effect we used a Z-score transformation, see the review (Quackenbush 2002). Genes were first ordered based on their lowest read depth (minimum expression). The parameters of the distribution of expression changes for the genes with similar expression levels were used to calculate Z-scores of differential expression for individual genes (see Methods). We used a Z-score of 4 as an arbitrary threshold of statistical significance for differentially regulated genes to minimize the false discovery rate.

Effects of arsenite treatment on transcriptome.

The estimated time required for mRNA maturation (Femino et al. 1998; Audibert, Weil, and Dautry 2002) is comparable to the duration of the arsenite treatment, therefore, we did not expect significant changes in mRNA levels due to a transcriptional response. However, it is conceivable that the arsenite treatment could affect the stability of specific mRNAs. Indeed the treatment was found to significantly alter the transcript levels of 24 genes (Figure 3.1E). The most pronounced effect observed is the accumulation of JUN mRNA, a transcript that is

short-lived under normal conditions (Elkon et al. 2010; Rabani et al. 2011). *JUN* encodes a subunit of the AP-1 transcription factor implicated in response to a myriad of physiological and pathological stimuli, reviewed in (Hess, Angel, and Schorpp-Kistner 2004). AP-1 transcription factors consist of homo- or heterodimers of different subunits and its composition is crucial for its specificity. As displayed in Figure 3.1F, despite a significant decrease in its translational efficiency the overall expression of JUN is almost unaffected upon arsenite treatment because of the increase in its mRNA levels, (Figures 3.1E and 1F). According to previous observations, AP-1 may induce apoptosis upon treatment with arsenite (Huang et al. 1999; Namgung and Xia 2000).

Arsenite treatment strongly inhibits global translation, while translation of a few specific mRNAs is resistant.

A median 5.4-fold reduction of ribosomal occupancy (translational efficiency, TE) was observed with the profiling data, a value that is consistent with the reduced number of ribosomes in the polysome fraction (Figure 3.1B). A relatively small subset of mRNAs displayed exceptional sensitivity to translational inhibition (see Figure 3.1 – source data 1; the most prominent ones are displayed in Figure 3.1F). Among these is *ODC1* which codes for ornithine decarboxylase, the rate-limiting enzyme of the polyamine biosynthesis pathway. An interesting case of potential downregulation is in *EIF2AK2* (a. k. a. *PKR*) that encodes one of the four eIF2 kinases [we refer to it as potential because its TE and ribo-seq Z-scores did not pass the threshold of statistical significance, but are close to it]. Among other extremely sensitive genes are several that encode RNA-binding proteins (*PABPC1, PCBP2, RPL12* and *CSDE1*) and cyclins G1 (*CCNG1*) and I (*CCNI*).

To explore the possible activation of the mTOR signalling axis after the course of 0.5 h arsenite treatment we analysed translation of mRNAs that were reported to be strongly downregulated upon pharmacological inhibition of mTOR (Hsieh et al. 2012). Almost all of them have negative Z-scores with their translational efficiency decrease upon arsenite treatment ~25% greater than the average (Figure 3.1). Thus, while arsenite treatment may affect the mTOR pathway, its impact on translation control is not substantial in comparison with eIF2 inhibition.

Several genes that were previously reported to resist the translation inhibition caused by eIF2 phosphorylation were also found to be resistant in our study (Figure 3.1F). This includes the well-studied *ATF4*, *ATF5* and *PPP1R15A*. We did not observe translational resistance for either *SRC* which was reported to be translated in an eIF2-independent mode (Allam and Ali 2010) nor *PRNP* which was also reported to escape eIF2 associated repression (Moreno et al. 2012). The general repression of eIF2 may be expected to promote translation of cellular IRESes if they enable an eIF2-independent mode of translation as reported for *XIAP* mRNA (Thakor and Holcik 2012). However, we find no evidence of resistance by any genes with putative IRES elements according the to IRESite database (Mokrejs et al. 2010), see Table 1 including XIAP mRNA. It is important to note that many of the genes from the IRESsite are not expressed at levels sufficient for detecting resistance.

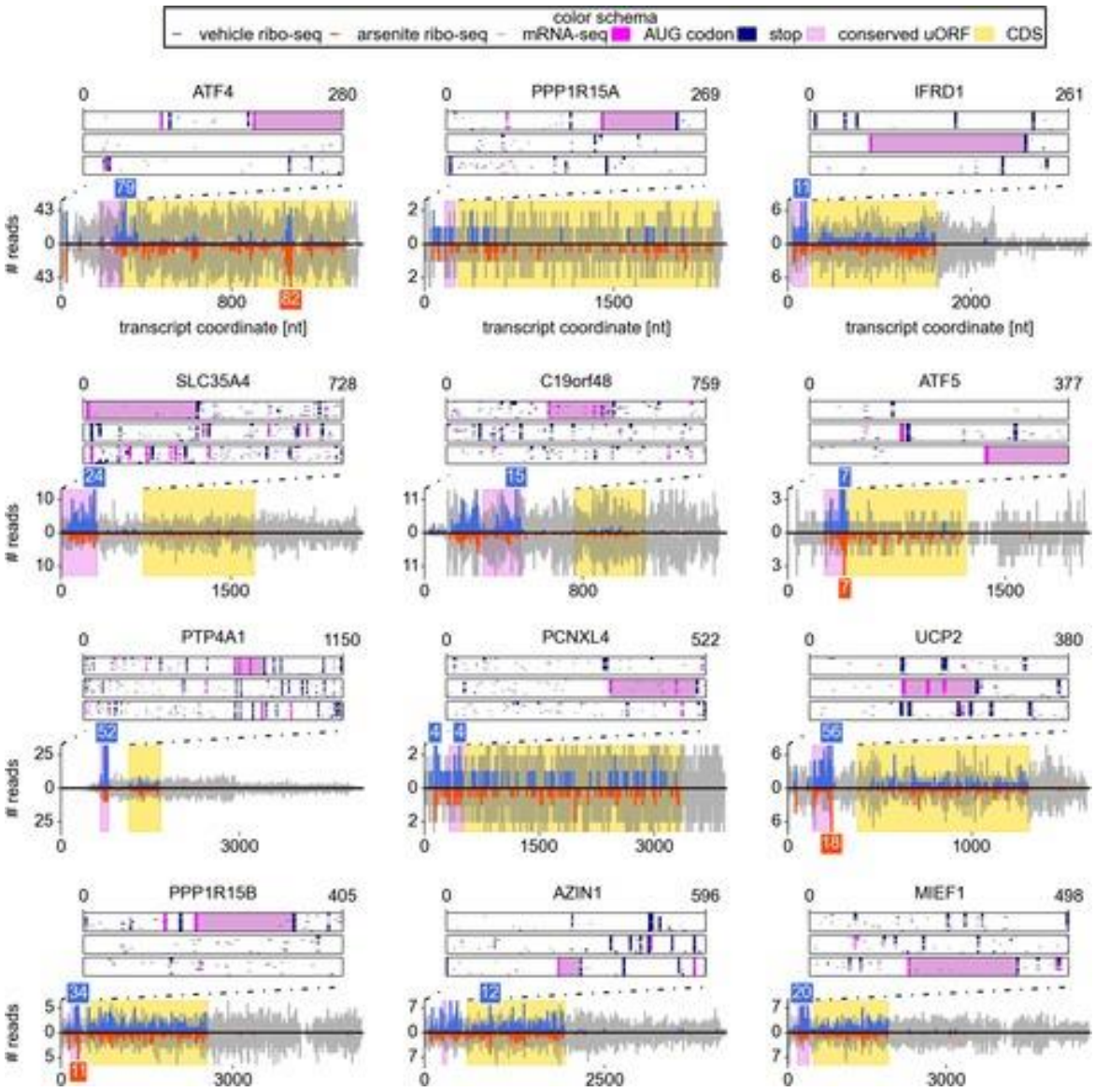


Figure 3.2. uORF conservation and ribosome density for the 8 top most stress resistant mRNAs in terms of their translation efficiency and also for mRNAs of UCP2, PPP1R15B, AZIN1 and MIEF1. Bottom plots for each mRNA show counts of mRNA-seq reads (grey) and ribosome reads (blue and red) as columns (control – positive values, arsenite treatment – negative values). The annotated CDS region is highlighted in yellow. Translated conserved ORFs in the 5' leaders are highlighted in violet. Read counts above cut-off are shown with numbers above corresponding columns. Top plots represent conservation of uORF features within the leaders of the orthologous mRNAs (upstream of annotated CDS) obtained from the analysis of genomic alignments of 46 vertebrates using the human sequence as a reference. Each box corresponds to one of the three reading frames where AUG codons are shown as pink dots and stop codons as navy dots in each of the genomic sequences used in the alignments. Regions of multiple sequence alignment corresponding to translated conserved uORFs are highlighted in violet. Introns and gaps were removed from the alignments.

Efficient translation of uORFs combined with inefficient translation of CDS is a predictor of stress resistant mRNAs.

The mRNAs encoding ATF4, PPP1R15A, SLC35A4, C19orf48, ATF5, HOXB2 were found to be "preferentially translated" (defined as having a TE Z-score >4 and a fold change>1), while mRNAs encoding IFRD1, PTP4A1, PCNXL4 and UCP2 were found to be "resistant" (TE Z-score >4 and a fold change < 1). Due to the small number of preferentially translated and resistant mRNAs, we analysed their properties together and for simplicity we refer to them as resistant for the remainder of this text.

Examination of individual mRNA profiles frequently revealed the presence of extensively translated uORFs in resistant mRNAs (Figure 3.2). Indeed with the exception of a single weakly expressed gene *HOXB2*, all mRNAs found to be resistant (TE Z-score >4) contained uORF(s) that are translated under normal conditions. However, ribosome profiling data suggest that 8% of other expressed mRNAs also have translated uORFs (Figure 3.3A). Hence, the mere presence of translated uORFs is a poor predictor of resistance. Therefore we investigated further the features of uORFs that are associated with stress resistance. As can be seen in Figure 3.3B, uORFs in the resistant mRNAs are usually efficiently translated under normal conditions, though yet again there is a large absolute number of non-resistant mRNAs that contain similarly efficiently translated uORFs. Figure 3.3C shows the relationship between the translation efficiency of the CDS and the resistance: CDSes of the most resistant mRNAs are weakly translated under normal conditions. The ratio of the ribosome densities in uORFs and in CDS provides a much better criterion for discriminating resistant mRNAs from non-resistant ones (Figure 3.3D). The length of all the translated uORFs was found to exceed 20 codons although this is of limited predictive value as there are many long uORFs in non-resistant mRNAs (Figure 3.3E). Based on these findings we expected that upon arsenite treatment the ribosome density for resistant mRNAs would shift from the 5' leaders to CDS. Such a shift is indeed observable (Figure 3.3F).

We also compared various sequence features of 5' leaders and uORFs between the resistant mRNAs and the remaining expressed mRNAs. We explored the nucleotide context surrounding uORF start codons (mostly AUG but also CUG) in resistant mRNAs but found no evidence for selection for a particular context; the frequency of individual heptameric

initiation sequences (-3 to +4) is equally variable across uORFs of resistant and non-resistant mRNAs as well as at annotated starts of CDSes. We found that the average 5' leader length of resistant mRNA is longer (378.5 nt) than of other mRNAs (169.0 nt), but there is significant variation within both distributions. We also found that 5' leaders of resistant mRNAs have lower potential for RNA secondary structure formation within the first 240 nt based on free energy estimates of potential structures predicted with RNAfold (Lorenz et al. 2011). Yet, the difference is small and RNA secondary structure potential does not correlate well with resistance.

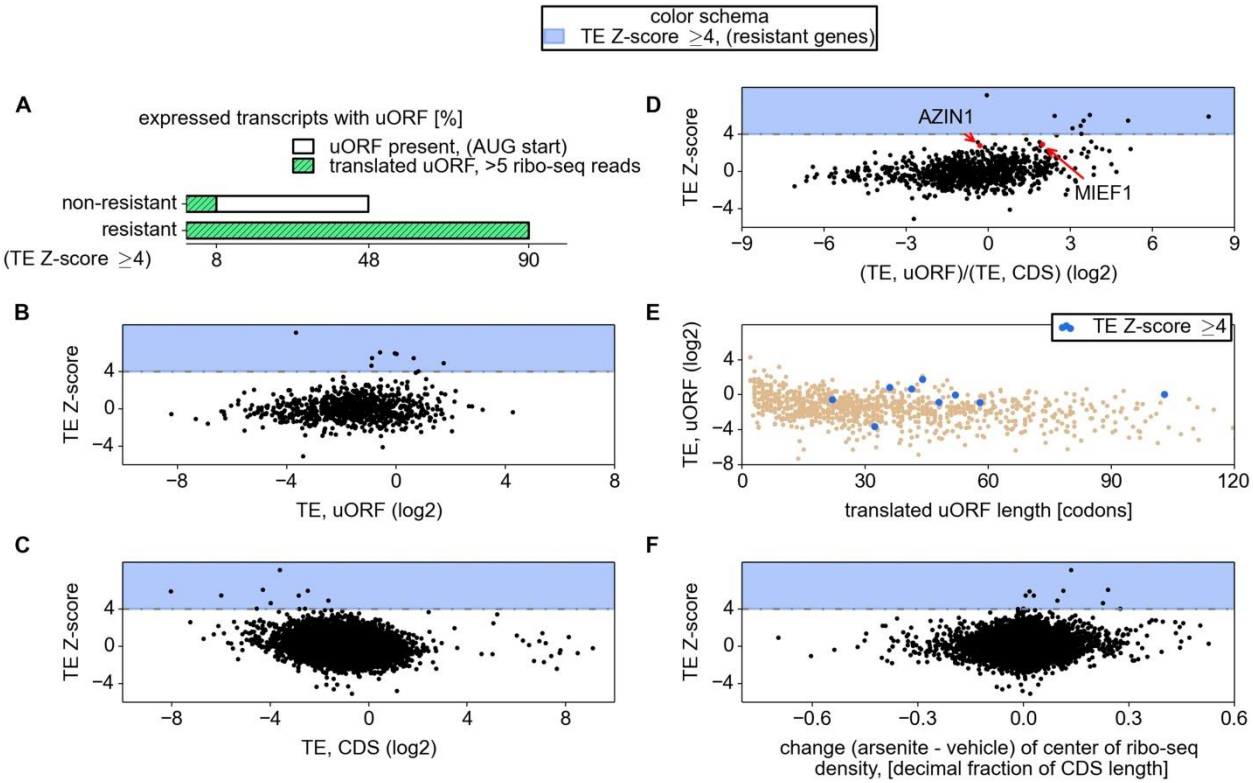


Figure 3.3. Relationship between mRNA stress resistance and uORFs. (A) Frequency of AUG initiating uORF occurrence and their translation in stress resistant and other mRNAs. Relationship between stress resistance (*y* axis) and translation efficiency of uORFs (B) CDS(C), uORF/CDS ratio (D). Translationally resistant genes (shaded in blue) have a high uORF translational efficiency (TE) and a low CDS TE. (E) Relationship between uORF length (*x* axis), their translation efficiency (*y* axis) and the level of stress resistance (differential coloring). (F) Relationship between stress resistance (*y* axis) and shift of ribosome density in the 3' direction. These plots indicate that under normal conditions resistant mRNAs tend to display a high efficiency of uORF translation and a low efficient translation of CDS regions while under stress conditions resistant mRNAs are associated with a shift of ribosome density in the 3' direction owing to a reduced ratio of ribosome density between uORFs and CDS.

Based on this analysis we concluded that uORFs are ubiquitous in all highly resistant mRNAs expressed in HEK293T cells and the efficient (and perhaps inhibitory) translation of these uORFs most likely plays a crucial role in the mechanism of resistance. The mere presence of an uORF and its translation is insufficient to provide the resistance to eIF2 inactivation.

Newly discovered cases of resistance to eIF2 inhibition mediated by uORFs.

We focused our attention on newly identified uORF-bearing mRNAs whose translation was refractory to eIF2 inactivation. For some of these, the regulatory functions of uORFs were described previously, but its implication in eIF2-dependent translational control was not shown. This was true for *IFRD1* (a. k. a. *PC4* and *TIS7*), an interferon-related developmental regulator, that was reported to be a modifier gene for cystic fibrosis (Gu et al. 2009). It has been reported previously that one of the two *IFRD1* transcript variants possesses a 51 codon uORF which triggers mRNA decay upon termination under normal conditions but not under conditions of Unfolded Protein Response mediated by tunicamycin (Zhao et al. 2010). We did not observe a significant change in the transcript level upon arsenite treatment. A second case is *UCP2* which codes for a mitochondrial anion carrier protein which increases the proton conductance of the mitochondrial membrane in response to reactive oxygen species (ROS) production (Baird and Wek 2012). It was shown that expression of *UCP2* is upregulated at the translational level upon oxidative stress (Adam et al. 2006). The *UCP2* 5' leader contains a 36 codon uORF that inhibits translation of *UCP2* mRNA under normal conditions (Hurtaud et al. 2006).

To our knowledge the other newly identified mRNAs have not been shown to be regulated at the translational level before. One of the most unusual examples is the mRNA encoding the probable UDP-sugar transporter protein SLC35A4 (Figure 3.2). Its 5' leader contains 11 AUG codons most of which are not conserved, however, one AUG that initiates a 102 codon uORF is highly conserved across vertebrates. This uORF encodes a peptide sequence containing PFAM domain DUF4535, ID PF15054; moreover the pattern of its conservation and higher frequency of synonymous mutations is consistent with protein coding evolution, suggesting that this uORF likely encodes a functional protein. This alternative protein (EMBL

accession HF548106) was recently detected by mass spectrometry analysis of cultured cells and human tissues (Vanderperre et al. 2013; Kim et al. 2014). We examined translation of this mRNA in other publicly available ribosome profiling datasets using GWIPS-viz (Michel et al. 2014) and found that this uORF is translated in all datasets. How ribosomes reach the 12th AUG codon upon arsenite treatment is unclear and merits further investigation.

Notably, one of the resistant mRNAs found in our study is the *PPP1R15B* gene that encodes a phosphatase that dephosphorylates eIF2, PPP1R15B, (a. k. a. CReP) (Novoa et al. 2001). Sustained translation of PPP1R15B mRNA under conditions of eIF2 inactivation represents a feedback loop for reactivation of eIF2 during recovery from stress.

Although they did not pass our stringent criteria for a resistant gene, other candidates which we identified (based on the gene's function and their profiles) are *AZIN1* (TE Z-score 2.76) and *MIEF1* (TE Z-score 2.88). *AZIN1* encodes an inhibitor of ODC1 (ornithine decarboxylase) antizymes. Antizymes are proteins that target ODC1 for degradation, and AZIN1 is highly similar to ODC1 but lacks ornithine-decarboxylation enzymatic activity. This makes it a competitive inhibitor of antizymes (Murakami, Marumo, and Hayashi 1988). It has been shown that a uORF initiated with a non-cognate AUU-codon mediates sensitivity of *AZIN1* mRNA translation to polyamine levels (Ivanov, Loughran, and Atkins 2008). *MIEF1*, mitochondrial elongation factor 1 is another candidate bicistronic mRNA that we have identified (the other is *SLC35A4,* above). Similar to *SLC35A4,* we observed evidence of protein coding evolution within its uORF. Its translation is also supported by multiple ribo-seq datasets available in GWIPS-viz (Michel et al. 2014). Examination of the sequence encoded by its uORF revealed that it contains a conserved domain that belongs to a PFAM family Complex 1 protein (LYR family), ID PF05347.

5' leaders of several newly identified mRNAs are sufficient to provide resistance to the translation inhibition.

To examine the role of the 5' leaders in modulating the resistance to eIF2 inhibition of resistant mRNAs revealed in this study (*IFRD1*, *PPP1R15B*, *UCP2*, *PTP4A1* and *SLC35A4*) we designed reporter constructs and prepared capped and polyadenylated mRNAs with sequences of 5'leaders upstream of a Firefly luciferase coding region. We used the 5' leader

of *ATF4* mRNA and the HCV IRES as positive controls. As a negative control, we used a non-specific 63 nt leader from the vector pGL3. mRNAs along with a control mRNA encoding *Renilla* luciferase were transfected into HEK293T cells and simultaneously treated with 40uM arsenite or control (Andreev et al. 2009).

Under normal conditions the translation of mRNAs bearing the 5' leaders of *IFRD1*, *PPP1R15B*, *UCP2* and *PTP4A1* was about 7-fold lower than that of the control mRNA with the simple non-specific leader (pGL3), whereas *SLC35A4* was even lower (Figure 3.4A). Arsenite treatment resulted in significant inhibition of pGL3 and control Rluc translation, while translation of other mRNAs did not change considerably and even slightly increased for SLC35A4 and HCV IRES. Similar results were observed in the Huh7 hepatocarcinoma cell line. To address the effect of arsenite treatment on ongoing translation, which may be more relevant in comparison to conditions applied for ribosome profiling, reporter mRNAs were transfected and 1 h later cells were treated with either non-specific translational inhibitor cycloheximide or arsenite. As expected, both inhibitors efficiently blocked translation of control Rluc mRNA, but only cycloheximide was able to arrest translation driven by leaders of resistant mRNAs. Surprisingly, during arsenite treatment, the reporter mRNA with the SLC35A4 5' leader was able to produce fifteen times more luciferase than after treatment with cycloheximide.

For several other reporter mRNAs with different 5' leaders, which possess low translational efficiencies under normal conditions, translation was significantly downregulated upon arsenite treatment (data not shown). This rules out the possibility that low translational efficiency of reporter mRNA results in resistance to stress.

We also measured the kinetics of protein synthesis to rule out the possibility that our observations can be explained by mRNA silencing or destabilization (Andreev et al. 2013). For this purpose cells were treated with arsenite (or a vehicle) immediately after transfection and luciferase activity was measured over time. Both *IFRD1* and *PPP1R15B* reporters showed luciferase activity increases over time with no indication of a plateau that would be expected upon mRNA destabilization (Figure 3.4B).

Treatment with a potent inhibitor of mTOR kinase, torin-1 (Thoreen et al. 2012) led to inhibited translation of these reporters to the same degree as the control pGL3 mRNA.

Thus, *IFRD1* or *PPP1R15B* leaders do not provide translational resistance to the stress response that involves sequestration of the cap-binding protein eIF4E.

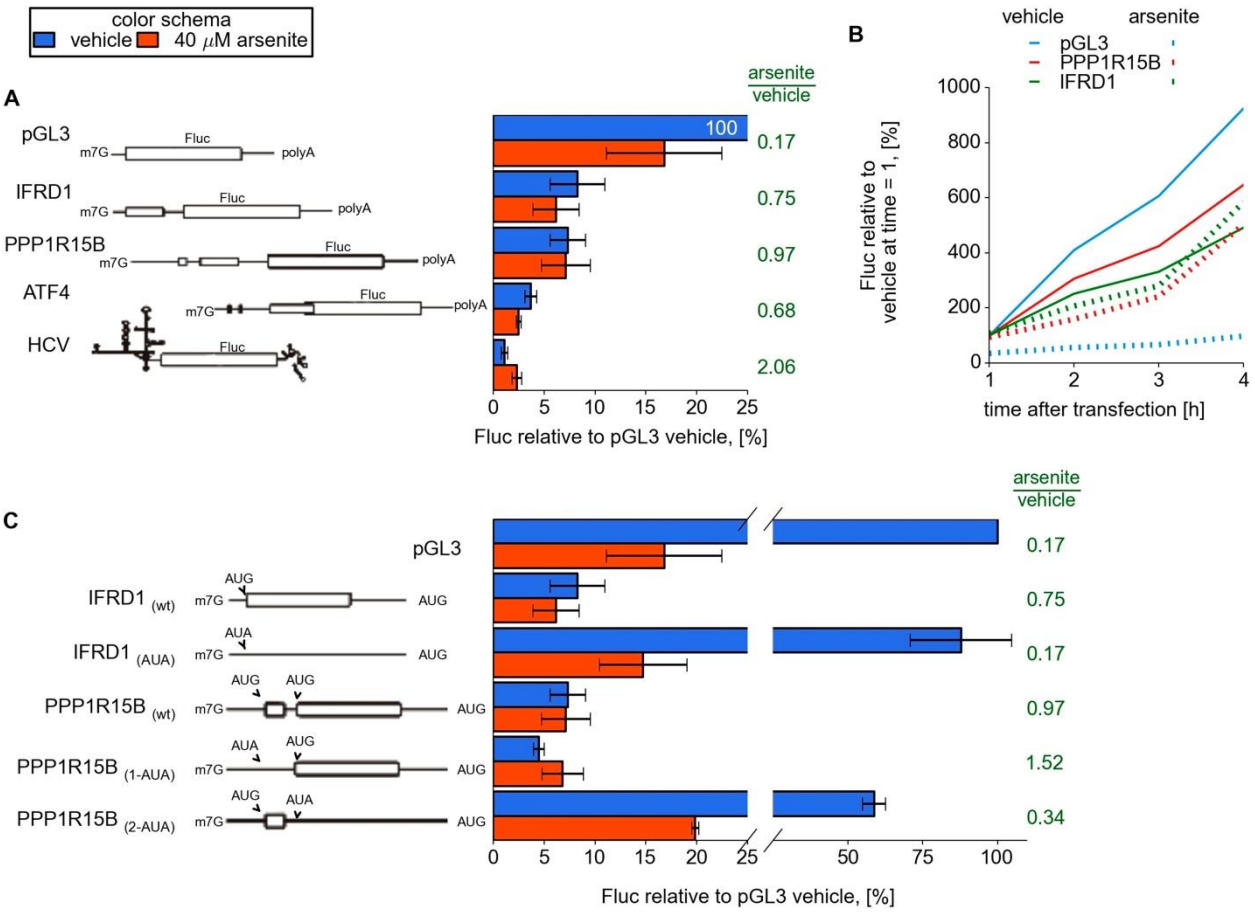


Figure 3.4. uORF involvement in modulation of IFRD1 and PPP1R15B mRNAs stress resistance. (A) Firefly luciferase (Fluc) activity produced by expression of mRNA containing different 5' leaders 2 h after arsenite treatment (red bars) and in untreated cells (blue bars). Relative units correspond to Fluc activity normalized to the median Renilla luciferase (Rluc) activity derived from a co-transfected Rluc mRNA. The green text represents fold change calculated from the same data. The ORF organization of examined mRNAs is outlined on the left. Bars represent standard deviations. (B) The time series analysis of Fluc expression in the cells treated at the time of transfection with sodium arsenite to a concentration of 40 μM (dotted lines) or vehicle (solid lines). Fluc activity of vehicle at 1 h was taken as 100% for each mRNA and experimental condition. (C) The effect of start codon identity in IFRD1 and PPP1R15B 5' leaders on Fluc activity.

It was conceivable that translational resistance was due to side effects of arsenite treatment rather than its direct effect on eIF2 inactivation. To directly address the impact of eIF2 phosphorylation on reporter mRNA translation during arsenite stress, we carried

out an experiment where cells were pre-transfected with a plasmid encoding the full length human PPP1R15A (a. k. a. GADD34) phosphatase subunit which is able to reverse eIF2 phosphorylation (Brush, Weiser, and Shenolikar 2003). Arsenite-induced eIF2 phosphorylation, as expected, was almost completely alleviated in the presence of GADD34 (residual phosphorylation probably reflects less than 100% efficient plasmid transfection. As a result, the downregulation of control Rluc mRNA was only 2 fold in comparison with an over 6 fold reduction in cells not transfected with the GADD34 plasmid. Translation of the IFRD1 reporter was not affected under either condition. Therefore we concluded that translational inhibition caused by arsenite treatment is predominantly due to the phosphorylation of eIF2. Also, we treated cells with 2.5 mM DTT (another well-known inducer of the Unfolded Protein Response mediated by eIF2 phosphorylation (Prostko, Brostrom, and Brostrom 1993)). We found that the leaders of IFRD1 and PPP1R15B provide translational resistance under these conditions as well.

Site-specific mutagenesis confirms the critical role of *IFRD1* and *PPP1R15B* uORFs translation in mediating resistance to eIF2 inhibition.

While both *IFRD1* and *PPP1R15B* mRNAs possess uORFs with high TE (Figure 3.2), the architecture of their uORFs is markedly different (schematic organisation of the 5' leaders is depicted in Figure 3.4). *IFRD1* mRNA contains a single highly conserved 53 codon uORF that starts 19 nt from the 5' end and ends 43 nt upstream of the main ORF. *PPP1R15B* mRNA contains two in-frame uORFs separated by 21 nt. The first uORF is 8 codons long and is 127 nt from the 5' end. The second, 52 codon long uORF is 75 nt upstream of the CDS.

Substitution of the *IFRD1* uORF AUG with an AUA codon increased the reporter expression 8-fold, but made the translation susceptible to eIF2 inhibition. For *PPP1R15B* substitution of the first uORF AUG with AUA slightly reduced the reporter activity under normal conditions, but surprisingly further increased the reporter resistance to the stress. A similar substitution of the second uORF start-codon significantly reduced the resistance to stress as in the case with *IFRD1* (Figure 3.4C).

We conclude that neither of these genes are regulated by delayed reinitiation (as in *ATF4* and *GCN4*), since in both cases a single uORF is sufficient for eIF2-mediated translational control and that under normal conditions the uORF significantly inhibits translation of the CDS.

The single uAUG in the *IFRD1* mRNA is in a suboptimal initiation context (A in -3 position but U in +4). To explore how the context may affect the resistance we introduced a +4U/G mutation that improves the context. We found a slight inhibition of the *IFDR1* main ORF translation under normal conditions (presumably due to an increased inhibitory effect of the uORF translation). However, this mutation did not alter the sensitivity of the main ORF translation to arsenite stress.

An unstructured leader sequence upstream of the IFRD1 uORF is necessary for stress resistance

We observed earlier that most stress resistant mRNAs possess efficiently translated uORFs. We hypothesized that some features of the 5' leaders upstream of uORFs may be important for resistance. To address this issue, we created two additional reporters based on control pGL3 and IFRD1, where we added a 5' terminal stem-loop of intermediate stability (Figure 3.5). As expected, the addition of this stem-loop resulted in a 3-4 fold decrease of both reporters activity under normal conditions. Interestingly, when arsenite stress was induced, the SL-IFRD1 construct did not exhibit resistance, while translation of SL-PGL3 was downregulated as much as the pGL3 construct. Therefore we propose that efficient loading of preinitiator complexes to the uORF is necessary for stress resistance in IFRD1. Next, we addressed the question of whether the specific sequence upstream of the IFRD1 uORF is required for regulation. We substituted it with an artificial single stranded $(CAA)_6$ sequence of the same length. This mutation did not alter stress resistance. Thus, we hypothesize that for resistant translation, the uORF to be preceded with a leader allowing a high initiation rate at the uORF.

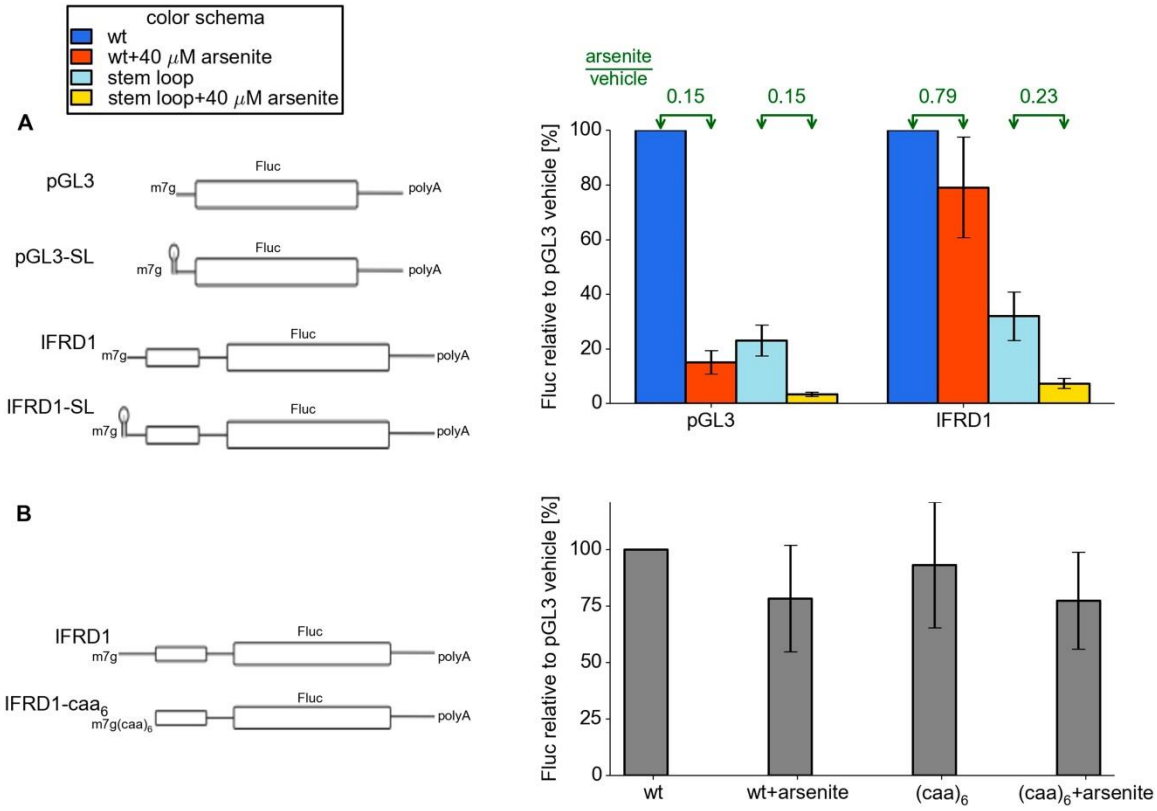


Figure 3.5. Features of the IFRD1 5' leader required for resistance.

(A) Firefly luciferase (Fluc) activity produced by expression of mRNA containing pGL3 and IRFD1 leaders (outlined on the left) with and without an additional stem loop at the 5' end under different conditions. Blue (normal conditions) and red (stress conditions) bars correspond to leaders lacking the stem loop while light blue (normal) and yellow (stress) correspond to leaders with the stem loop. The fold change of Fluc activity in response to stress is indicated above green arrows. (B) The effect of $(CAA)_6$ addition to the 5' leader of IFRD1 on Fluc activity in response to stress.

DISCUSSION

eIF2 phosphorylation is a key event in cell response to various stresses and is also involved in the cell cycle. The global downregulation of protein synthesis triggered by eIF2 inactivation has two main purposes. The first is to conserve cellular resources, the second is to provide a delay to evaluate the severity of the damage and, depending on its level, reprogram gene expression either towards apoptosis or to a pro-survival repair response. This necessarily requires activation of genes involved in the ISR. mRNAs of genes involved in the ISR ought to be translated under conditions of eIF2 inactivation, see the reviews (Luo, Freitag, and Sachs 1995; Baird and Wek 2012).

In order to identify such mRNAs we utilized ribosome profiling, a technique that generates a snapshot of ribosome locations on the entire set of mRNAs (N T Ingolia et al. 2009). We applied this technique to HEK293T cells 30 min after treatment with sodium arsenite, a well-known inducer of eIF2 phosphorylation. This enabled us to study the early stress response at the level of translation. Under our strict criteria of statistical significance (Z score TE>4), translation of ten mRNAs was found to be resistant. Translation of six mRNAs (encoding ATF4, PPP1R15A, SLC35A4, C19orf48, ATF5, HOXB2) was increased and translation of four (encoding IFRD1, PTP4A1, PCNXL4, UCP2) was reduced only slightly in comparison with a global reduction in translation. Seven mRNAs (encoding CCNG1, CCNI, CSDE1, ODC1, PABPC1, PCBP2, RPL12) were found to be particularly sensitive. We confirmed translational resistance of some genes previously reported and identified novel stress resistant genes.

Features found to be common amongst the resistant mRNAs are their low levels of CDS translation (Figure 3.3C) and low levels of transcription, except ATF4 (Fig 3.1F). It is unclear, however, whether these features are common due to the resistance-providing mechanism or due to their function. mRNAs resistant to eIF2 are expected to encode components of cell signalling, e.g. transcription factors, kinases, phosphatases, and therefore are not required in large quantities. We also found that, with one exception, all mRNAs significantly resistant to the stress conditions possess translated uORFs in their 5' leaders. Their number, mutual organization and depth of phylogenetic conservation vary. *IFRD1* has only a single uAUG in its 5' leader while *SLC35A4* has eleven (Figure 3.2). For these mRNAs at least one of the uORFs is efficiently translated under normal conditions and is longer than 20 codons. We confirmed that 5' leaders of some of these mRNAs confer translation resistance to a reporter in synthetic RNA constructs. Owing to the sequence diversity of these 5' leaders the mechanism of uORF mediated resistance may vary and should be studied individually. We chose to explore how the sequence properties of *IFRD1* and *PPP1R15B* 5' leaders affect the resistance of downstream ORFs. For this purpose we carried out site-directed mutagenesis of 5' leader sequences in the constructs containing a luciferase reporter. We found that translation of only one uORF is sufficient to provide resistance to eIF2 inhibition for *IFRD1* and *PPP1R15B*. Therefore it is likely that the resistance for these transcripts is

provided by the mechanism that resembles alleviation of scanning ribosomes obstruction, rather than delayed reinitiation (ATF4-like cases). The 5'leader sequence of IFRD1 is likely to enable rapid ribosome loading at the uORF, since the addition of a 5' terminal stem loop of intermediate stability abolished stress resistance. The substitution of this leader upstream of a uORF with an artificial single stranded sequence does not affect stress resistance, ruling out the requirement for a specific nucleotide sequence. This observation may explain why translation of many mammalian mRNAs possessing uORFs is not resistant to eIF2 phosphorylation. Translation of these uORFs might be inhibited to such an extent that they would be unable to provide any resistance.

While it is possible that uORFs provide mRNAs with stress resistance in the same manner in all resistant mRNAs detected here, we think it is more likely that uORFs are an essential component of diverse mechanisms. We mentioned earlier two such mechanisms, *delayed initiation* and *alleviation of scanning ribosomes obstruction*. However, even an IRES-mediated resistance would likely require translation of a uORF as it would prevent IRES structure melting by the scanning ribosomes.

The genes that we found to be resistant to eIF2 inhibition may participate in the ISR. This is the case with *PPP1R15B*. Similar to *PPP1R15A*, it encodes a subunit of the phosphatase that dephosphorylates eIF2, thus providing feedback preventing complete translation suppression and also enabling recovery from the stress-induced translational arrest (He, Gross, and Roizman 1997; Jousse et al. 2003). Expression of *PPP1R15A* is tightly regulated, its basal level of expression is almost undetectable (Novoa et al. 2001), whereas *PPP1R15B* is constitutively present in cells (Jousse et al. 2003). PPP1R15B is a short-lived protein whose half-life is approximately 45 min (Jousse et al. 2003). Therefore it needs to be continuously synthesized in order to dephosphorylate eIF2. It has been previously found to remain present in the cell upon arsenite and tunicamycin treatments (Jousse et al. 2003).

It is also possible that some of the translationally resistant genes are not directly implicated in the ISR. Their resistance could be related to other eIF2-mediated regulatory mechanisms,

e.g. during cell-cycle progression or development (Datta et al. 1999; Harding et al. 2009). At least one of the newly identified stress resistant mRNAs encoding oncogenic phosphatase PTP4A1 (a. k. a. PRL-1) may be directly implicated in malignant transformation, since it suppresses expression of p53 tumour suppressor (Min et al. 2009).

Translational resistance of some genes to eIF2 phosphorylation may also be a consequence of the ORF organisation of their mRNAs which serves a different purpose. This could be the case for the candidate bicistronic mRNAs identified in this study (*MIEF1* and *SLC35A4)*. The ratio between uORF and main ORF translation changes in both upon eIF2 inactivation. Coding for two functionally related proteins in the same mRNA may be advantageous for coordination of their expression.

To summarize, our work expands the list of mRNAs which are known to be persistently translated under conditions of eIF2 phosphorylation although it suggests that the number of such mRNAs is very low. The analysis of ribosome densities on mRNAs resistant to eIF2 phosphorylation accentuates the vital role of uORFs translation in providing the resistance.

METHODS

Ribosome profiling

The ribosomal profiling technique was carried out according to (Ingolia et al. 2012) but with important modifications described below. HEK293T cells were grown in DMEM supplemented with alanyl-glutamine and 10% FBS, and replated to 150 mm dishes (2 dishes per sample). After cells reach 70%-80% confluency, sodium arsenite (or vehicle) was added at 40 μM and 30 min later cells were harvested: dishes were immediately chilled on ice and washed with PBS+cycloheximide (100 µg/ml). Importantly, cells were not pre-treated with cycloheximide to avoid artificial accumulation of initiation complexes at translation initiation starts (Gerashchenko and Gladyshev 2014). Cells were then lysed with buffer containing 20 mM Tris-HCl (pH 7.5), 250 mM NaCl, 1.5 mM $MgCl_2$, 1mM DTT, 0.5% Triton X-100, 100 µg/ml cycloheximide (Sigma) 20 U/ml TURBO DNAse (Ambion). Note that the buffer contains low magnesium (1.5 mM versus 5 to 15 mM recommended in (Ingolia et al.

2012), since high magnesium concentrations stabilize secondary structures in mRNAs which may hamper digestion by RNase I (RNase I itself does not require divalent cations for its activity). Cell lysates were incubated on ice for 10 min, centrifuged at 16000 g at +4°C for 10 min and the supernatant was divided onto 2 parts, for ribo-seq and "naked" mRNA-seq library preparation. One part of the lysate, usually 10-20 A260, was then treated with RNase I (Ambion) with 100 U per 3.1 A260 of lysate at 23°C for 50 min. The digestion was then stopped with the appropriate amount of SUPERASE inhibitor (Ambion). The treated lysate was then loaded on 10-60% (m/v) sucrose density gradient containing 20 mM Tris-HCl (pH 7.5), 250 mM NaCl, 15 mM $MgCl_2$, 1 mM DTT, 100 ug/ml cycloheximide and centrifuged in SW-41 rotor at 35000 rpm for 3 h. Sucrose density gradients were prepared as described previously (Stone 1974). Briefly, 5.5 ml of 10% sucrose was slowly layered onto the same volume of 60% sucrose, gradient tubes were then sealed with parafilm, slowly placed horizontally for 4 h to allow spontaneous gradient formation and then slowly returned to a vertical position. This method of gradient formation is highly reliable and reproducible and does not require any special equipment. Total RNA from the fractions corresponding to the 80S peak (Figure 3.1B) was then extracted with phenol/chlorophorm followed by ethanol precipitation.

For the mRNA-seq control, the second lysate aliquot was processed with Trizol-LS (Life Technologies) according to the manufacturer's protocol. mRNA from total RNA was isolated using the Oligotex mRNA kit (Qiagen). Two rounds of polyA(+)-mRNA selection instead of one were applied to decrease rRNA contamination to approximately 3%. Purified mRNA was then subjected to alkaline hydrolysis as in (N T Ingolia et al. 2009).

Both ribo-seq and mRNA-seq samples were then loaded on a 15% denaturing urea PAGE (containing 1xTBE, 7M urea and AA:bis-AA in the ratio 20:1). Bands corresponding to nucleic acid fragments of 28-34 nt were excised for both ribo-seq and mRNA-seq samples. RNA was extracted by overnight incubation in a shaker using buffer containing 0.3 M NaOAc (pH 5.1), 1 mM EDTA and 0.1 % SDS followed by precipitation with 1 volume of isopropanol and 2 µl of GlycoBlue (Life Technologies).

The same quantity of both ribo-seq and mRNA-seq fragments (usually – 100 ng) were mixed with 1:10000 of unspecific RNA oligonucleotide 5'AUGUACACGGAGUCGACCCGCAACGCGA 3' which serves as a "spike-in" control. The library preparation was carried out as previously described (Ingolia et al. 2012) with the following modifications. First, the circularization reaction was performed for 2 h. Second, during PCR library amplification, the temperature ramping speed was set as 2.2°C/s, to reduce bias associated with GC content (Aird et al. 2011).

Two independent biological replicates were carried out. Libraries were sequenced either on an Illumina MiSeq genome analyser at the TrinSeq genomic facility (Dublin) or on an Illumina HiSeq 2000 system at the Beijing Genomics Institute (BGI).

Plasmid constructions

Reporter DNA constructs were prepared on the basis of the pGL3R vector (Stoneley et al. 1998). Plasmids containing the 5' leaders of test mRNAs were cloned between *SpeI* and *NcoI*. pGL3R-HCV and pRluc plasmids were described in (Andreev et al. 2009). The 5' leader of *IFDR1* mRNA (NM_001550.3) was shortened by 38 nt at the 5' end to correspond to the location of the likely predominant transcription start based on the analysis of available ESTs for this region. The 5' leader of *PPP1R15B* (NM_032833.3) was extended at the 5' end by one nt which is present in the majority of available EST sequences.

For the same reasons the 5' leader of *UCP2* mRNA (NM_003355.2) was shortened by 16 nt, and the 5' leader of *PTP4A1* mRNA (NM_003463.4) by 547 nt, both 5' leaders were cloned into pGL3R between *Spe1* and *Nco1*. The 5' leader of *SLC35A4* mRNA (NM_080670.2) was extended by 11 nt. To prepare pGL3-ATF4, the human *ATF4* 5' leader was obtained by RT-PCR with primers GGGTAATACGACTCACTATAGGGTTTCTACTTTGCCCGCCCACAG and GGCGCCATGGTTGCGGTGCTTTGCTGGAATCG. The resulting product was digested with *NcoI* (underlined) and inserted into pGL3 vector at SmaI-NcoI sites. pGL3-ATF5 was prepared with leader of ATF5 mRNA (NM_001193646.1) shortened by 62 nt.

The HCV-Fluc plasmid contained the T7-promoter, the entire viral 5' leader and the first 33 codons of viral ORF fused to Fluc without its initiator codon and the entire viral 3' UTR.

The full length human ppp1R15A (GADD34) sequence fused with N-terminal FLAG tag was cloned in to pcDNA 3.1 construct between *HindIII* and *XbaI* to prepare the pcDNA GADD34 construct.

mRNA preparation

mRNA preparation was carried out as described in (Dmitriev et al. 2007). Briefly, PCR products were obtained with a forward primer containing the T7 promoter (either the universal primer which anneals to the vector sequence immediately upstream of insertions, CGCCGTAATACGACTCACTATAGGGAGCTTATCGATACCGTCG or the T7 promoter-containing gene specific primer) and reverse primer containing an oligo(dT) stretch of 50 nt (T50AACTTGTTTATTGCAGCTTATAATGG). To introduce a stem-loop structure, PCR products were obtained with forward primer containing T7 promoter: CGCCGTAATACGACTCACTATAGGGAGTGGACTTCGGTCCACTCCCAGCTTATCGATACCGTCG. To introduce the CAA_6 sequence upstream of the IFRD1 uORF, the following primer was used: CGCCGtaatacgactcactataGGGCAACAACAACAACAACAACAACATGTATCGTTTTCGATCACAGCTC

The PCR products were then purified and used as templates for T7 RNA polymerase using *in vitro* RNA transcription by T7 RiboMAX Large Scale RNA Production kit (Promega). For preparation of m7G-capped transcripts the 3'-*O*-Me-m7GpppG (ARCA cap analogue, NEB) was added to the transcription mix without GTP for 5 min to prime transcripts with cap followed by the addition of GTP (at a ratio of ARCA:GTP 10 to 1). The resulting RNAs were purified by LiCl precipitation and examined for integrity by PAGE.

Cell culture, western blots and transfection procedures

Experiments with mRNA transfection were performed as described in (Andreev et al. 2012). Briefly, the mixture of 0.2 μg of m^7G-capped Fluc mRNA and 0.01 μg of m^7G-capped Rluc

mRNAs per 1 well of a 24-well plate was transfected to the cells at 70-80% confluency either with Lipofectamin 2000 (Invitrogen) or Unifectin56 (Rusbiolink). Simultaneously with transfection cells were treated with either 40 μM sodium arsenite, 2.5 mM DTT or 250 nM Torin-1 (Torics Biosciences). 2 h later (or at the specified time interval), cells were harvested and luciferase activities were analysed with the Dual Luciferase Assay kit (Promega).

For experiments with GADD34 overexpression, cells were transfected with either pcDNA-GADD34 or control pcDNA3.1 one day prior to mRNA transfection. Plasmids were transfected with Fugene 6 (Promega) according to the manufacturer's instructions.

For western blotting, cells were rapidly lysed with buffer containing 1% SDS and 20 mM Tris-HCl pH 6.8 followed by brief sonication of the lysates. This was done to prevent posttranslational modifications of proteins of interest during the lysis. Antibodies used in the study are: rabbit anti-EIF4EBP1 (Chemicon International, AB3251), rabbit anti-GAPDH (Proteintech, PTG10494-1-AP), rabbit anti-IFRD1 (Proteintech, 12939-1-AP), rabbit anti-ATF4 (Proteintech, 10835-1-AP), rabbit anti-phospho-p70 S6 Kinase (Thr389) (Cell Signalling, 9205S), rabbit anti-Phospho-S6 Ribosomal Protein (Ser235/236) (Cell Signalling, 2211S), rabbit anti-S6 Ribosomal Protein (Cell Signalling, 2217S) rabbit anti-phospho eIF2 (S51) (Enzo, SA-405), and anti-FLAG-M2 (Sigma Aldrich). To remove nonspecific binding, phospho-eIF2 antibodies (1:2500) were incubated along with 10% fetal bovine serum in TBS-T.

Data analysis

Cutadapt (Martin 2011) was used to remove the 3' adapter of the reads. Reads that did not map to either the "spike-in" (ATGTACACGGAGTCGACCCGCAACGCGA) or rRNA sequences were aligned to the RefSeq catalogue (Pruitt et al. 2014) downloaded from the NCBI website on the 15 Aug 2013. The alignment was carried out with Bowtie version 1.0.0 (Langmead et al. 2009), with parameters -a -m 100 –norc (all read mappings to the positive strand were taken with exception of those with more than 100 mappings). Reads that mapped to transcripts of more than one gene or multiple times to a transcript were discarded. In order to maximise the genuine ribosome footprints aligning to the transcriptome, ribo-seq reads with a length typical for monosomes (29 to 35 inclusive) were

used for further analysis. In the case of multiple transcript variants, among the transcripts annotated as protein coding, the one with the highest ribo-seq read density in control conditions was brought forward for differential expression analysis.

The raw read count data were rescaled to normalize for the differences in the total number of reads mapped with a rescale factor *F*. For the 1st replicate the rescaled factor *F* for each sample was calculated as the ratio by which the total number of mapped reads exceeds the lowest total number of mapped reads out of two conditions, i.e.

$$F_n = \frac{\sum_i x_{ni}}{\min_i x_{ni}}$$

where x_{ni} is the number of mapped reads from sample n in the condition *i*. This normalization was carried out independently for ribo-seq and mRNA-seq.

The 2nd replicate was sequenced on Illumina Miseq and Hiseq 2000 instruments and obtained sequence reads were aggregated. The number of "spike-in" reads were used to rescale the read counts with a similar approach as for replicate 1. The raw read count of each sample was divided by the rescaling factor *F* calculated as above with the only difference that x_{ni} represents the number of "spike-in" reads from sample *n* in the condition *i*. This rescaling was also implemented for the ribosomal profiles of individual transcripts shown in Figure 3.2.

The normalized read counts of ribo-seq reads aligning to the coding regions (as determined by inferred locations of the A-site codons) and of mRNA-seq reads aligning to the entire transcript were used for the differential expression analysis. For ribo-seq reads the A-site codon of the elongating ribosome was inferred to be the 17th or 18th nt of the read from its 5' end depending on the read length.

Transcripts were binned based on the number of mapped reads (expression/coverage level) in one of the conditions where this value is the minimal. For the analysis of differential translation efficiency the minimum value (referred to as the minimum expression level) was taken from four conditions, while for the analysis of differential RNA level analysis only RNA-seq reads obtained under control and stress conditions were used. With the minimum expression level threshold of 2 reads, transcripts were sorted in ascending order and arranged in bins of size 300. Each bin had transcripts with a similar numbers of mapped reads and was analysed independently. The mean and standard deviation of change in expression of the transcripts within each bin was used to determine a Z-score for each transcript. For the remaining transcripts, of insufficient number to be binned (<300), the mean and standard deviation was obtained from the previous bin. The Z-score determined for each transcript enabled comparison between bins.

The analysis of translation of mRNA leaders was carried out for the transcripts with at least 2 normalized read counts in each of all four experiments/conditions. A uORF was defined as a sequence of sense codons uninterrupted with a stop codon and beginning with an AUG codon located upstream of the annotated CDS. In the case of uORFs overlapping CDS, the 5' end of CDS was considered as the end of the uORF in order to avoid ambiguity in assigning ribo-seq reads to one of the two overlapping ORFs. Nested uORFs (those contained within uORFs in the same frame) were excluded for the same reason. The TE of a uORF was estimated as the average density of ribo-seq reads in the uORF divided by the average density of the mRNA-seq reads for the corresponding mRNA. A uORF was considered to be translated if more than 5 ribo-seq reads aligned to it. For transcripts with more than 1 translated uORF the properties of the uORF with the highest number of aligned ribo-seq reads were used.

For the purpose of the analysis represented in Figure 3.3C the centre of ribosome density was defined as the minimal mRNA coordinate for which the number of ribo-seq reads aligning 5' of the corresponding location is the same or greater than the number of ribo-seq reads aligning 3' of the corresponding location. This value was determined for genes under arsenite and control conditions. The difference in ribosome footprint density was divided

by the CDS length to prevent skewing of results in favour of transcripts with longer coding regions.

The list of human cellular IRESes was obtained from IRESsite (Mokrejs et al. 2010) on 2 Aug 2014.

We identified the most probable translation initiation sites by manually examining the ribo-seq profiles of 8 translationally resistant gene (*PPP1R15A, IFRD1, SLC35A4, C19ORF48, PTP4A1, PCNXL4, UCP2, PPP1R15B*). *ATF4* and *ATF5* were not included as these appeared to be regulated by an alternative method. uORF initiation sites with an AUG or CUG were selected based on its ability to fit with the observed profile upon manual examination. AUG codons were preferred but the surrounding consensus sequence was not considered. A sequence logo of the initiation sites (-4 to +3) was produced with WebLogo (Crooks et al. 2004).

The free energy of leaders was estimated with RNAfold (Lorenz et al. 2011). The first 240 nt was used (transcripts with shorter leader were excluded) as free energy of RNA is related to its length. We chose 240 nt as this was the length of the shortest leader of resistant mRNAs with a translated uORF.

Data availability

Sequences of ribosome profiling libraries have been deposited into the NCBI Gene Expression Omnibus portal under the accession number GSE55195.

Table 1. Translation response of mRNAs with reported IRES from IRESite.

Gene_name	IRES name	ORF #	minimal expression	TE fold change	TE Z-score
AGTR1	AT1R_var1	1	4	0.46	1.12
AGTR1	AT1R_var2	1	4	0.46	1.12
AGTR1	AT1R_var3	1	4	0.46	1.12
AGTR1	AT1R_var4	1	4	0.46	1.12
APAF1	Apaf-1	1	102.9	0.29	1.8
AQP4	AQP4	1	1.5	0.98	2.47
ATAD5	ELG1	1	1083.4	0.22	0.75
BAG1	BAG1_p36delta236nt BAG1_p36	4	1376.5	0.1	-1.31
BCL2	BCL2	1	10	0.18	-0.22
BIRC2	c-IAP1_285-1399 c-IAP1_1313-1462	1	147	0.13	-0.94
CCND1	CCND1	1	213	0.17	-0.15
CDK11A	PITSLRE_p58	1	0	NA	NA
CDKN1B	p27kip1	1	1252	0.17	-0.3
CSDE1	UNR	1	16657.5	0.05	-5.12
DCLRE1A	hSNM1	1	1065.2	0.17	-0.24
EIF4G1	eIF4G	1	12937	0.21	0.65
EIF4G1	eIF4GI-ext	1	12937	0.21	0.65

EIF4G2	DAP5	1	21727.6	0.26	1.56
EIF4G3	eiF4GII	1	1305.6	0.13	-1.46
EIF4G3	eIF4GII-long	1	1305.6	0.13	-1.46
FGF1	FGF1A	1	0	NA	NA
FMR1	FMR1	1	137.8	0.13	-1.35
HSPA1A	hsp70	1	0	NA	NA
HSPA5	BiP_-222_-3	1	1819.3	0.23	1
IGF2	IGF2_leader2	1	0	NA	NA
LAMB1	LamB1_-335_-1	1	1141	0.13	-1.52
LEF1	LEF1	1	248	0.15	-0.8
MNT	MNT_75-267 MNT_36-160	1	144	0.28	1.32
MYB	MYB	1	154.2	0.09	-1.47
MYC	c-myc	2	1946	0.14	-1.2
MYCL1	L-myc	1	0.5	NA	NA
MYCN	n-MYC	1	0	NA	NA
NKRF	NRF_-653_-17	1	1064	0.18	-0.01
PDGFB	PDGF2/c-sis	1	0	NA	NA
PIM1	Pim-1	1	113	0.12	-1.1
RUNX1	AML1/RUNX1	1	10	0.16	-0.38
RUNX1T1	MTG8a	1	163	0.13	-1.08

XIAP	xIAP_5-464	XIAP_305-466	1	2169.6	0.12	-1.67

Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes

This chapter has being published as a research article in Genome Biol. 2015 May 6;16:90.
doi: 10.1186/s13059-015-0651-z.

ABSTRACT

Background. Oxygen and glucose metabolism play pivotal roles in many (patho)physiological conditions. In particular, oxygen and glucose deprivation (OGD) during ischemia and stroke results in extensive tissue injury and cell death.

Results.

Using time-resolved ribosome profiling, we assess gene expression levels in a neural cell line, PC12, during the first hour of OGD. The most substantial alterations are seen to occur within the first 20 minutes of OGD. While transcription of only 100 genes is significantly altered during one hour of OGD, the translation response affects approximately 3000 genes. This response involves reprogramming of initiation and elongation rates, as well as the stringency of start codon recognition. Genes involved in oxidative phosphorylation are most affected. Detailed analysis of ribosome profiles reveals salient alterations of ribosome densities on individual mRNAs. The mRNA-specific alterations include increased translation of upstream open reading frames (uORFs), site-specific ribosome pauses, and production of alternative protein isoforms with N-terminal extensions. Detailed analysis of ribosomal profiles also reveals six mRNAs with translated ORFs occurring downstream of annotated coding regions and two examples of dual coding mRNAs, where two protein products are translated from the same long segment of mRNA, but in two different frames.

Conclusions.

These findings uncover novel regulatory mechanisms of translational response to OGD in mammalian cells that are different from the classical pathways such as Hypoxia Inducible Factor (HIF) signalling, while also revealing sophisticated organization of protein coding information in certain genes.

BACKGROUND

The World Health Organization reports ischemic heart disease and stroke as the leading causes of death in humans (WHO 2014). Both diseases are characterized by the interruption of blood flow, oxygen and glucose supply, which induce severe tissue damage. Metabolic stress, acidosis, energy crisis, cytosolic Na^+ and Ca^{2+} accumulation, over-production of reactive oxygen and nitrogen species and mitochondrial impairment are the hallmarks of the ischemic cascade, which ultimately leads to cell death (Xing et al. 2012). A growing knowledge of the mechanisms of this cascade provides opportunities for its modulation in order to counteract disease-related outcomes (Iadecola and Anrather 2011; Meller and Simon 2013).

The response to oxygen and glucose deprivation (OGD) is known to involve hypoxia inducible factors (HIFs) and 5' AMP-activated protein kinase (AMPK) signaling pathways. HIFs modulate gene expression at the transcriptional level via binding to HIF-responsive elements (HRE) in the promoter regions of the regulated genes (Semenza 2014; Kaelin and Ratcliffe 2008; Ratcliffe 2013). Recently it was shown that in conditions of overall suppression of protein synthesis under hypoxia (achieved by eIF4E sequestration), EPAS1 (also known as (a.k.a.) HIF-2α) can activate translation of mRNAs containing specific RNA secondary structures (rHRE) signals in their 3’ UTRs (Uniacke et al. 2012). AMPK is an important sensor of energy starvation and adaptive response to ischemia in mammalian cells (Nagata, Mogi, and Walsh 2003; Manwani and McCullough 2013). AMPK affects gene expression via phosphorylation of different proteins, including elongation factor 2, eEF2 (Browne, Finn, and Proud 2004), which leads to a general inhibition of protein synthesis.

Whereas HIF-1α levels increase within minutes upon hypoxia onset (Jewell et al. 2001), and reach maximal values *in vivo* within 1-5 h in a tissue specific manner (Stroka et al. 2001), changes in the levels of HIF target proteins (which depend on sequential transcription, maturation and export of new mRNAs, followed by translation) lag behind HIF activation. In contrast, hypoxia-dependent changes in translation or stability of the already existing mRNA pool may provide a quicker adjustment of gene expression, for example via reduced hydroxylation of the components of the translational machinery (Zhdanov et al. 2015) or AMPK activation (Nagata, Mogi, and Walsh 2003; Manwani and McCullough 2013). We reasoned that during OGD, changes in *de novo* protein production in mammalian cells may

occur prior to HIF-mediated reprogramming of gene transcription. To investigate this, we exposed neural PC12 cells to 20, 40 and 60 min of OGD in a hypoxia workstation and examined the dynamics of genome-wide changes in transcription and translation by means of ribosome profiling (ribo-seq). We also closely monitored oxygen levels, mitochondrial polarization and the phosphorylation status of key components of relevant metabolic pathways.

Analysis of ribosomal profiling data revealed ~3000 genes affected at the level of translation. RNA levels of only slightly more than a hundred genes were affected during the same time. The major phase of the response took place within the first 20 minutes. The response involved alteration of general parameters of translation; we observed reduced stringency of start codon selection and increased accuracy of translation termination. Detailed investigations of individual mRNA profiles allowed us to identify over 60 upregulated uORFs, numerous site specific OGD-induced ribosome pauses, synthesis of proteins with terminal extensions and a few translated ORFs located downstream of annotated coding regions. These results suggest that the early OGD mediated reprogramming of gene expression precedes the canonical HIF mediated response and it may significantly affect the latter, as well as the cell fate in general.

RESULTS AND DISCUSSION

Energy stress, HIF signaling and pathways regulating translation under OGD.

After 1 h of OGD cellular ATP levels decreases to almost ~80% of control conditions and it continued to decrease further almost linearly with time (Fig.4. 1A). In line with this, mitochondrial membrane potential (ΔΨm) probes TMRM and JC-1 showed a progressive depolarization of mitochondria in cells exposed to OGD (Fig. 4.). As expected, these effects were not observed when glucose was available: ATP and $\Delta\Psi_m$ levels decreased only slightly due to the continued ATP generation via glycolysis and activity of F1Fo ATP synthase working in reverse mode to maintain mitochondrial polarization. Under energy stress imposed on the cells, only a minor transient elevation in Epas1 (a.k.a. HIF-2α) was observed along with a substantial increase in AMPK phosphorylation (Thr172) during OGD course, contrasting to oxygen deprivation (OD) conditions (Fig. 4.1C,) (Zhdanov et al. 2015).

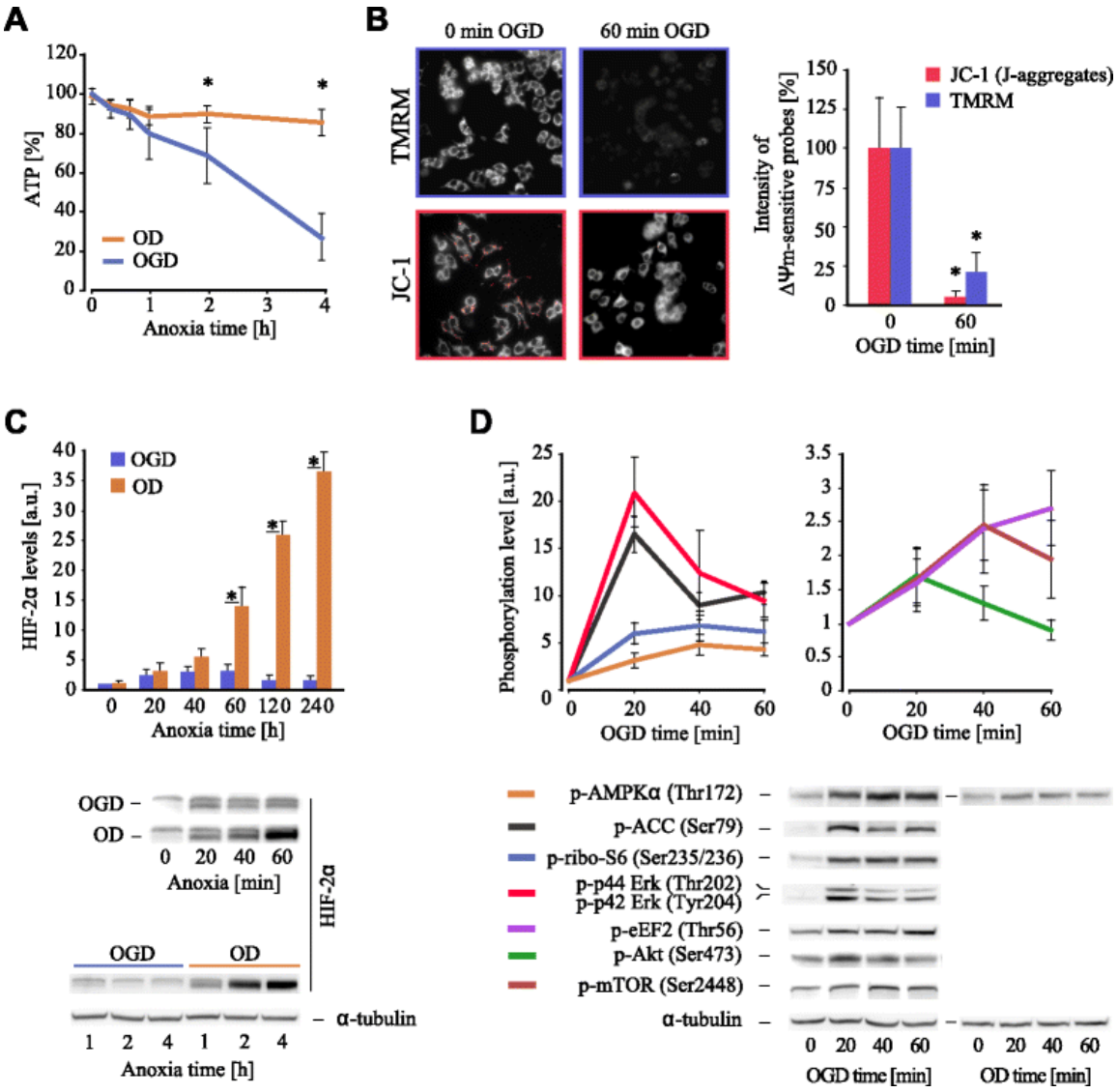


Figure 4.1. In vitro model of ischemia.

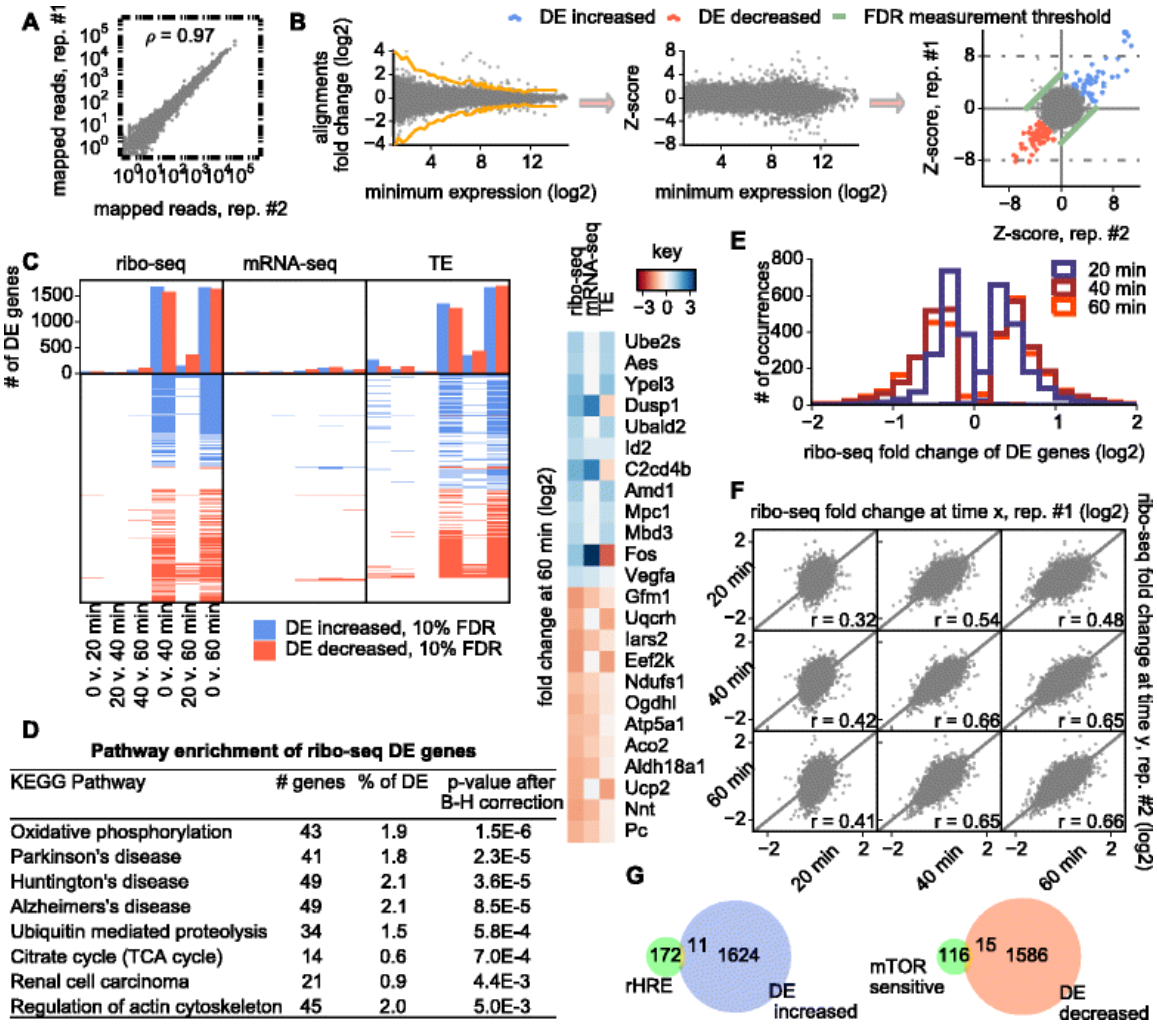
(A) Analysis of cellular ATP during OGD. Reduction in ATP becomes significant after 2 h. In the presence of glucose ATP levels remain unchanged. (B) Decreased intensities of TMRM and J-aggregated form of JC-1 (red) probes under OGD indicate $\Delta\Psi m$ depolarization. (C) Dissimilar to OD, only a minor transient elevation in Epas1 (a.k.a. HIF-2α) levels is seen under OGD. (D) Characteristic OGD induced changes in phosphorylation of the proteins involved in regulation of translation. Minor increase in p-AMPKα under OD suggests smaller effect of this condition on AMPK signaling. Bar is 50 µm. Asterisks demonstrate significant difference. Error bars show SD (A, C and D) or a range in TMRM and JC-1 intensities (B), calculated for 40-50 randomly selected cells.

OGD alters translation of ~3,000 mRNAs, while affecting RNA levels of only about a hundred genes

Ribo-seq and mRNA-seq libraries were obtained in two independent biological replicates (Fig. 4.2A) with between 7 and 17 million non-ribosomal RNA reads mapped from each library. The gene expression exhibited a neuronal signature with high expression of *Th* and *App*. Differentially expressed (DE) genes were identified upon Z-score transformation (Quackenbush 2002) followed by the determination of the 10% FDR (false discovery rate) threshold (see Materials and Methods and Fig. 4.2B). RNA levels of 111 genes were found to be altered after 1 h of OGD (Fig. 4.2C). The function of many of these genes is consistent with a response to the bioenergetic stress as evident from frequently associated gene ontology terms: “response to reactive oxygen species” (*Dusp1*,*Fos*, *Hspd1*), “response to hypoxia” (*Actn4*, *Abat*, *Adm*, *Ddit4*, *Egln3*, *Vegfa*), “oxidative phosphorylation” (*Atp5a1*, *Ndufs1*, *Ndufs2*, *Sdha*, *Uqcrc1*) “ tricarboxylic acid (TCA) cycle” (*Aco2*, *Idh3g*, *Ogdh*, *Ogdhl*, *Pck2, Pc*).

The same statistical procedure uncovered an overwhelming response at the level of translation. Analysis of ribo-seq occupancy revealed 3,202 DE genes after the first 40 min OGD. Analysis of Translation Efficiency (TE, ribo-seq signal normalized over mRNA) revealed 2,556 DE genes (Fig. 4.2C) during the same period. Moreover, the greater part of this response occurred within the first 20 min (Fig. 4.2E). The fold change of gene expression after the first 20 min correlated better than may be expected with that after 40 and 60 min of OGD (Fig. 4.2F). This suggests that the most of the later response is an amplification of the differential expression that occurred during the first 20 min. Note, there seems to be a delayed response in replica #2 compared to replica #1 for 20 min data point. However, the trend of changes is the same across the entire time period. The variance observed between replicas at 20 min of stress may not be coincidental. If changes occur most rapidly around this time point, increased variance in response would be expected for that data point.

The enrichment analysis of the KEGG pathways associated with DE genes (ribo-seq occupancy) revealed an overrepresentation of genes associated with the oxidative phosphorylation pathway and neurological diseases to which oxidative phosphorylation plays a significant role, see (Federico et al. 2012; Youle and van der Bliek 2012) for reviews (Fig. 4.2C). Genes involved in ubiquitin proteolysis and TCA cycle were also overrepresented (Fig. 4.2C).



D

Pathway enrichment of ribo-seq DE genes

KEGG Pathway	# genes	% of DE	p-value after B-H correction
Oxidative phosphorylation	43	1.9	1.5E-6
Parkinson's disease	41	1.8	2.3E-5
Huntington's disease	49	2.1	3.6E-5
Alzheimers's disease	49	2.1	8.5E-5
Ubiquitin mediated proteolysis	34	1.5	5.8E-4
Citrate cycle (TCA cycle)	14	0.6	7.0E-4
Renal cell carcinoma	21	0.9	4.4E-3
Regulation of actin cytoskeleton	45	2.0	5.0E-3

Figure 4.2. Analysis of differential gene expression (DE).

(A) Reproducibility of ribo-seq experiment under control conditions. (B) Z-score transformation and FDR estimation. (C) Heat map diagram of differentially expressed genes. (D) DE genes enrichment over KEGG pathways. (E) The magnitude of ribo-seq occupancy alterations during first 20, 40 and 60 min of OGD for the genes detected as DE after 40 min. (F) Correlations across replicates of ribo-seq fold changes of genes with >32 reads per time point at different time intervals of OGD exposure. (G) Venn diagrams showing the lack of relationship between DE genes, eIF4E sensitivity and the presence of rHRE.

It has been shown previously that upon prolonged hypoxia (24 h) Epas1 (a.k.a. HIF-2α) induces translation of mRNAs bearing rHRE structures in their 3'UTRs (Uniacke et al. 2012). This occurs during general translation suppression by eIF4E sequestration. Therefore we explored the response of the genes containing rHRE elements (Uniacke et al. 2012), as well as of genes whose mRNA translation is mTOR sensitive (Hsieh et al. 2012). Neither group of

genes significantly overlaps with DE genes (Fig. 4.2G) suggesting that these pathways are unlikely to be involved in the early response to OGD.

Ribosome density increases at mRNAs 5' leaders during OGD

In agreement with previous ribosome profiling studies of stress responses (Wiita et al. 2013; Shalgi et al. 2013; B. Liu, Han, and Qian 2013) we observed a widespread increase of ribosome density in mRNA 5' leaders upstream of annotated coding open reading frames (acORFs) (Fig. 4.3A). A metagene profile analysis indicates that this increase spreads throughout the entire 5' leader and the beginning of coding regions (Fig. 4.3B). It has been argued recently that the increase in 5' leader translation may be an artifact of ribosome accumulation in the vicinity of start codons caused by cycloheximide pretreatment that blocks elongating, but not scanning ribosome particles (Gerashchenko and Gladyshev 2014). Therefore, it is important to emphasize that in our procedures we have not pretreated cells with cycloheximide and still observed the increase despite reduced ribosome densities at the beginning of acORFs in all conditions (Fig. 4.3B).

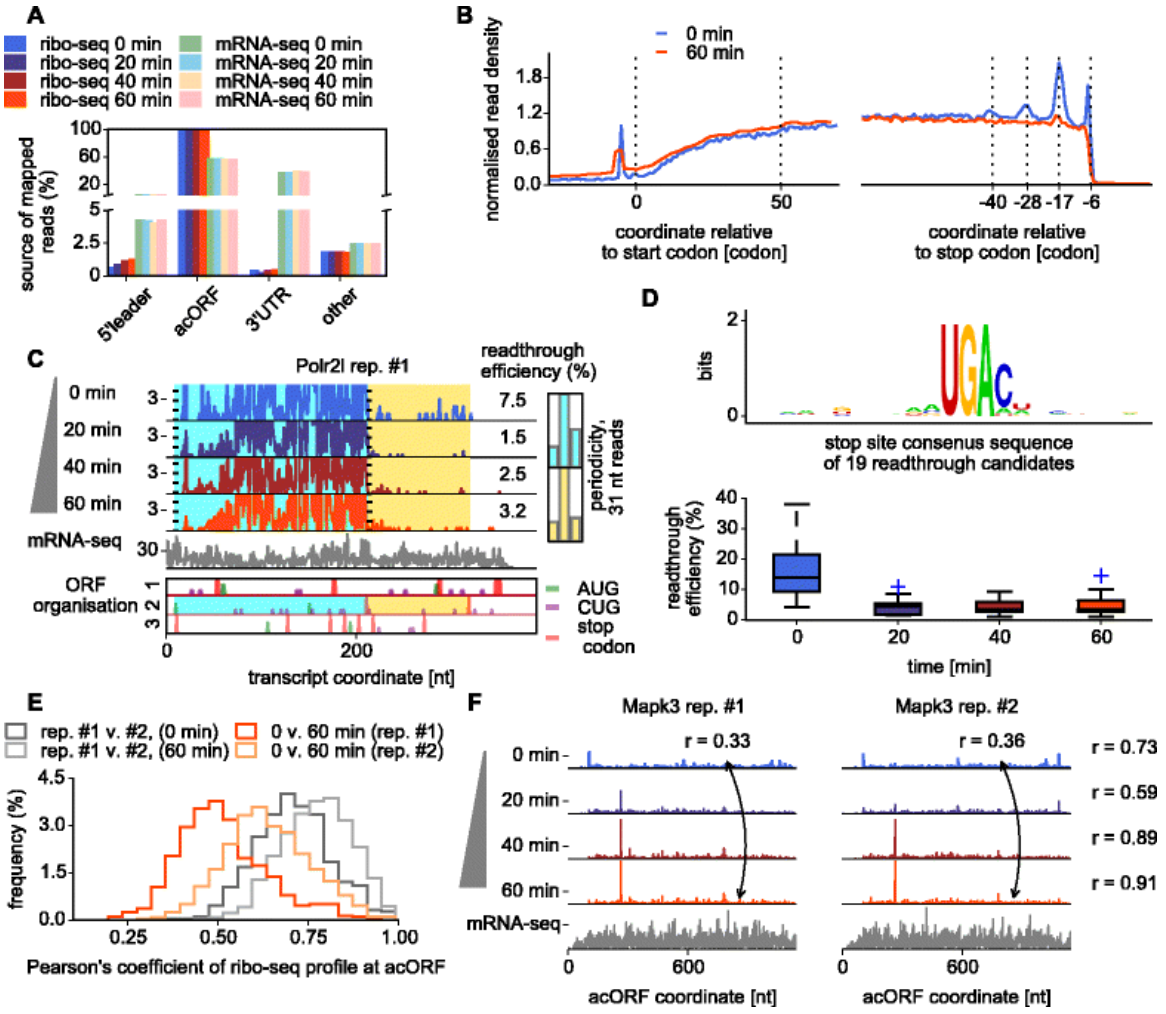


Figure 4.3. OGD induced changes to initiation, elongation and termination.

(A) Distribution of sequenced fragments across functional regions of mRNAs under different conditions. (B) Metagene profile of the ribosome density obtained under normal condition and after 1 h OGD. Positions of footprint 5' ends are shown relative to coordinates of start and stop codons of acORFs which are set to 0. Over 5,000 mRNAs profiles were aggregated to produce the profile. (C) Ribosomal density at *Polr2I* mRNA. The density in the 3'UTR indicates leaky termination (readthrough). (D) Sequence logo of nucleotide context surrounding stop codons in the 19 mRNAs exhibiting high levels of readthrough (top) and the change of stop codon readthrough efficiency during the time course of OGD. (E) Distribution of the Pearson's correlations of ribosome densities for individual mRNA profiles between replicas and across conditions. (F) An example of mRNA (*Mapk3*) with changes in footprint density distribution between conditions and replicas.

OGD effect on translation termination

A striking feature of metagene profiles is the presence of distinct peaks of ribosome density at the 3' ends of coding region under control conditions (Fig. 4.3B). The peaks are separated with 11-12 codons intervals roughly corresponding to the length of ribosome protected

fragment. Therefore we believe that they occur due to ribosome queuing at the ends of coding regions of some mRNAs. This stacking is independent of the identity of stop codons and is observed in metagene profiles constructed based on the identity of stop codons. Indeed, it appears to be widespread in all mRNAs . Presumably the stacking occurs due to a slow rate of translation termination relative to elongation. The stacking disappears under OGD stress either due to an increased rate of translation termination or due to a reduction of elongation rates.

Puzzled by this observation we explored how OGD affects the accuracy of translation termination. We found 19 mRNAs with relatively high density of footprints immediately downstream of annotated stop codons. An example profile for *Polr2l* mRNA is shown in Fig. 4.3C, profiles for all other mRNAs can be found at the mini website accompanying this paper. All of these mRNAs have acORFs terminating with a UGA codon, most with adjacent 3' C (Fig. 4.3D) which is known to be frequent among mammalian genes that were recently discovered to utilize stop codon readthrough in their expression (Stiebler et al. 2014; Schueren et al. 2014; Loughran et al. 2014).

The efficiency of termination at these 19 genes increased considerably by 20 minutes of OGD and remained high during the rest of the time course (Fig. 4.3D). This was surprising in the light of evidence that hydroxylation of eRF1 is known to promote termination efficiency (Feng et al. 2014). However, it has also been shown recently that protein components of the ribosome decoding center in *Saccharomyces cerevisiae* are subject to hydroxylation and the hydroxylation affects stop codon readthrough in opposite ways depending on the stop codon and the nucleotide context surrounding it (Loenarz et al. 2014). The structures of oxygenases are highly similar across all kingdoms of life (Chowdhury et al. 2014) and it is very likely that mammalian ribosomes could be affected in a similar way.

OGD alters local decoding rates.

The phosphorylation of eEF2 (Fig. 4.1D) led us to examine the profile data for changes in translation rates. The density of ribosomes at any specific location of an acORF is expected to negatively correlate with the local rate of ribosome elongation, but to correlate with the rate of translation initiation for that acORF. Therefore, if the translation initiation rate at

the acORF is changed but local elongation rates are unchanged we expect the geometrical shape of the ribosome profile to remain the same. To explore whether this is the case we measured pairwise profile similarities (by calculating Pearson's correlation coefficients) across conditions and between replicas (Fig. 4.3E).

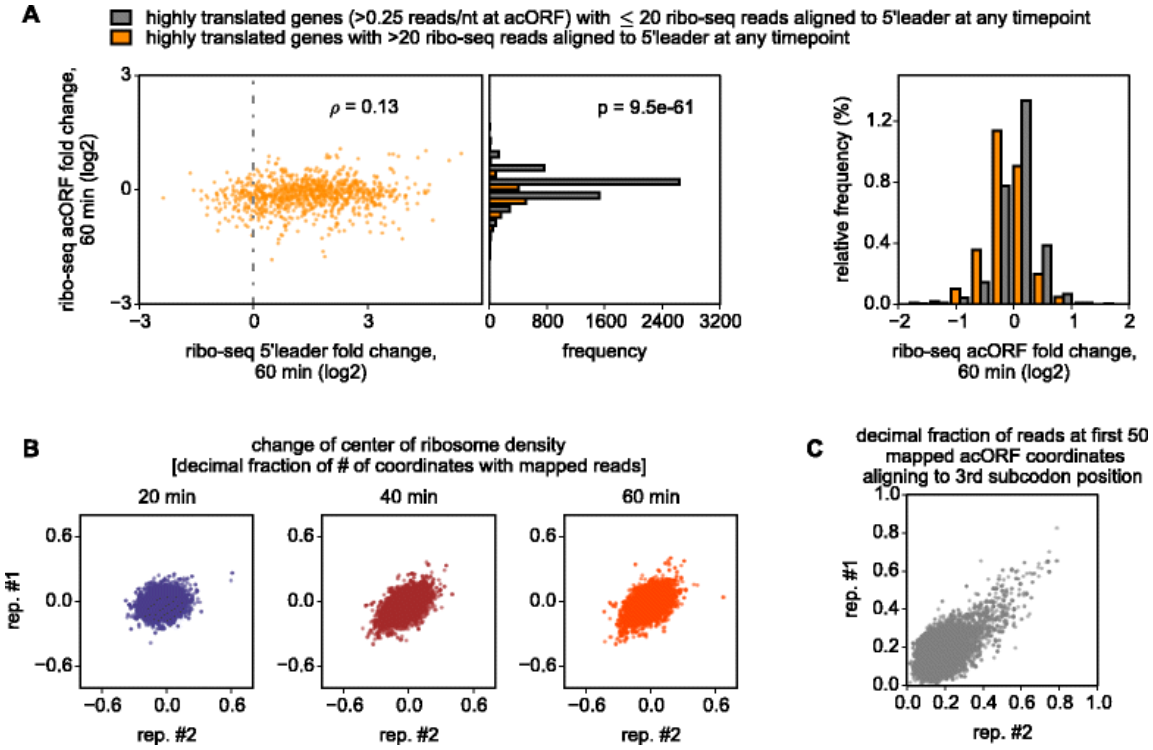


Figure 4.4 The effect of 5' leader translation on translation of acORFs and identification of regulatory uORFs.
(A) Correlation between fold change of ribosome density in the 5' leaders and in acORFs and the change in acORF translation in mRNAs with translated 5' leaders (orange bars) relative to all mRNA (grey): absolute (left) and relative (right) frequencies of genes. **(B)** Change in the position of the centre of ribosome density for each mRNA during the course of OGD. mRNAs with upregulation of inhibitory uORFs are likely to be found in the left bottom corner. Ribosome profiles of corresponding mRNAs were chosen for manual examination. **(C)** acORF regions translated in more than one reading frame are expected to have altered triplet periodicity (high density of footprints with 5' ends at the third subcodon position). This analysis was used to detect overlapping uORFs.

We found that the ribosome density distributions over individual mRNAs correlated better between replicas for the same conditions than within replicas for different conditions indicating that OGD induces alterations in the local elongation rates. We excluded the possibility that differences of mRNA sequencing depth may be a confounding factor by repeating the analysis with pairwise comparison of profiles of equal sequencing depth. The

change of profile shape was caused by both the alleviation of pauses observed under normal conditions and the introduction of new pauses. In some cases, these alterations resulted in reproducible peaks of ribosomal density with their heights increasing as OGD progresses, as exemplified by *Mapk3* mRNA profile in Figure 4.3F. These peaks most likely represent OGD induced ribosome pauses. A simple sequence logo examination (Crooks et al. 2004) of the nucleotide or amino acid context in the vicinity of the strongest pauses did not reveal a common sequence motif associated with them.

OGD alters the recognition of translation initiation sites.

We considered a possibility that the increase in 5' leader ribo-seq reads may have contributed to the translational response by inhibiting translation of acORFs. Although the fold change of 5' leader ribo-seq occupancy does not correlate well with that of the acORF globally, we observed that translation of acORFs in mRNAs with translated 5' leaders was more likely to be downregulated in comparison with the rest (Fig. 4.4A). Therefore it is likely that part of the translational response is due to the inhibitory roles of upregulated uORFs. We screened for such cases by using two metrics, the shift of the center of ribosome density (Andreev et al. 2015) towards the 5' end and the triplet periodicity signal at the start of acORFs (Michel et al. 2012; Bazzini, Lee, and Giraldez 2012), see Fig. 4.4B,C. The triplet periodicity was used to infer the phase of translated frames, while a shift of the center of density is observed when translation is altered differently at the 5' part of the mRNAs (e.g. uORFs) or at the 3' part of the mRNAs (acORF region that does not overlap any uORFs). Inhibitory uORFs should display a shift of density owing to the expected increase of 5' leader occupancy and subsequent decrease of acORF occupancy. If the uORF overlaps with theacORF, its translation would distort the triplet periodicity of ribosome footprints in the region of the overlap (Michel et al. 2012). The screen was followed by manual evaluation of mRNAs and the analysis of ribosome density changes in uORFs relative to acORFs. We identified several dozens of mRNAs whose uORFs translation increased during OGD with an accompanied decrease of acORF translation. Many of these uORFs lack AUG codons. Ribosome profiles of *Eif1a* and *Eef2k* mRNAs exemplify such regulation (Fig. 4.5A The strongest increase in translation of uORFs was observed for *Eif1a* and *Junb* mRNAs. For these genes ribosome density on uORFs increased over 10-fold during OGD from almost

undetectable levels under normal conditions, and greatly exceeded ribosome density at the acORFs (Fig. 4.5A).

The likely mechanism of the increased uORF translation is the altered recognition of translation initiation sites. This is supported by the observed decrease in translation of acORF in *Eif5* as well as in *Eif1a* and its paralogues *Eif1ad* and *Eif1ax*, which are all implicated in start codon selection (Fekete et al. 2007; T V Pestova, Borukhov, and Hellen 1998; Passmore et al. 2007). Interestingly, it is likely that decreased production of these factors is also mediated by uORFs translation (Fig. 4.5A and the mini web site). Synthesis of another important factor controlling stringency of initiation site selection, eIF1 (Nanda et al. 2009), increased about 2-fold (see Additional Data). eIF1 and eIF5 have opposing effects on the stringency of start codon selection (Loughran et al. 2012) and both regulate translation of their own mRNAs via negative control feedback loops. The negative feedback control mechanism is modulated by the initiation context in *Eif1* (Ivanov et al. 2010). In *Eif5* the mechanism involves translation of inhibitory uORFs (Loughran et al. 2012). The inhibition of eIF5 synthesis and increased eIF1 synthesis under OGD are synergistic responses, expected to counteract the reduction of start codon selection stringency.

Another important example of uORF-mediated inhibition of acORF translation is in *Eef2k* mRNA (Fig. 4.5A). The observed reduction of *Eef2k* acORF translation is consistent with a reduction in the levels of eEF2k protein. Once activated by AMPK, eEF2k phosphorylates eEF2 and induces a general suppression of translation (Leprivier et al. 2013; Ryazanov, Shestakova, and Natapov 1988; Carlberg, Nilsson, and Nygård 1990). Thus, the uORF-mediated inhibition of *Eef2k* expression counteracts activation of its protein product and may serve as a break mechanism to avoid complete suppression of translation elongation.

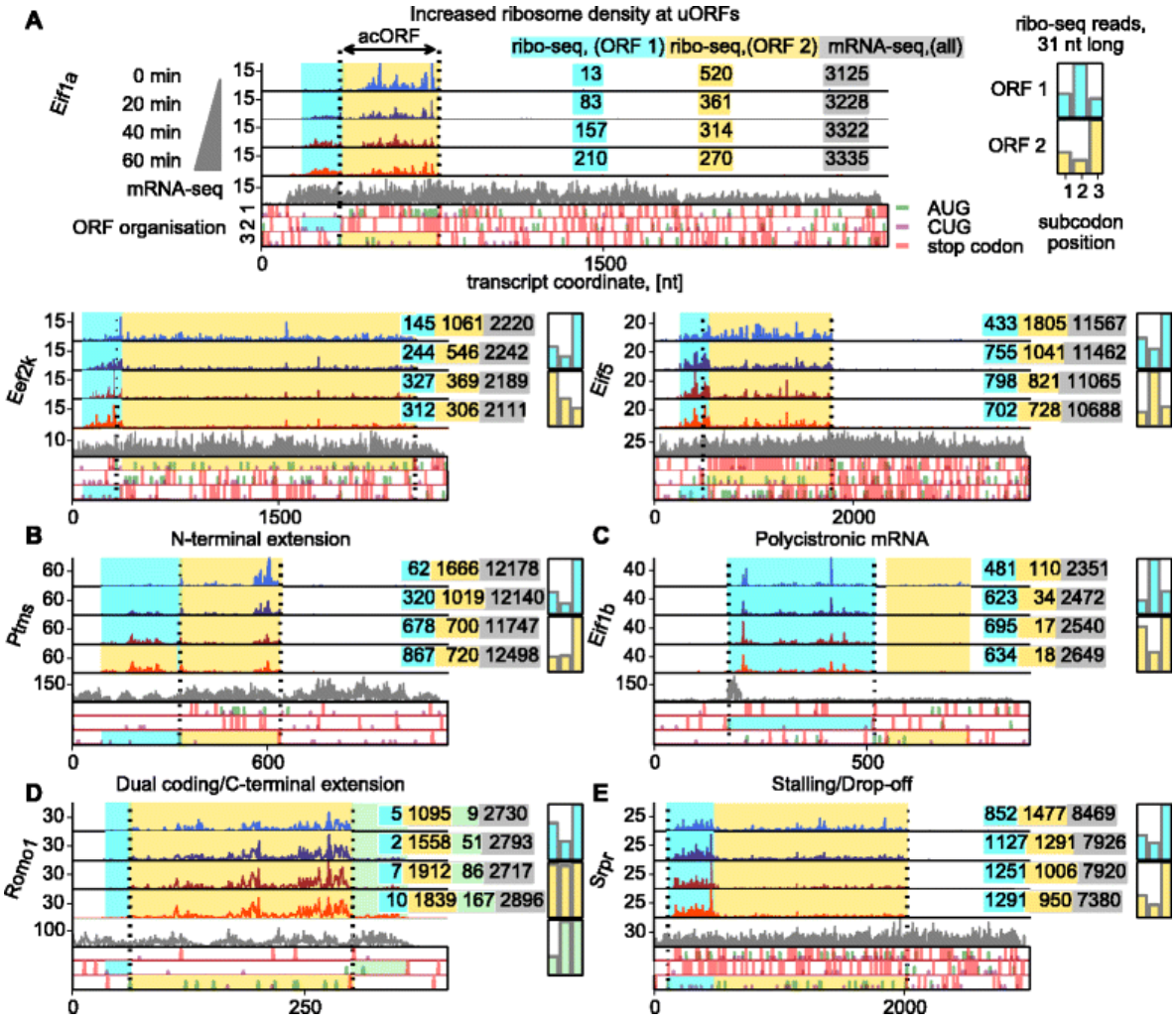


Figure 4.5. Ribosome profiles exhibiting mRNA-specific OGD-induced features.

(A) Selected mRNAs whose ribosome densities increase at uORFs and decrease at acORF during OGD. The ribosome profile for *Eif1a* mRNA is used as the legend. (B) Upregulation of parathymosin isoform with N-terminal extension whose synthesis begins at a non-AUG codon. (C) Increased uORF density on the *Eif1b* mRNA coincides with decreased density at the acORF and non-annotated downstream ORF. (D) Increased ribosome density downstream of Romo1 acORF that may occur as a result of translation of a long overlapping ORF. (E) Ribosome stalling or an abortive translation event at *Srpr* mRNA.

Novel protein isoforms, polycistronic and dual coding mRNAs.

The altered recognition of translation initiation sites can potentially lead to initiation translation at non-AUG codons that are in-frame with acORF, leading to the production of protein isoforms with N-terminal extensions. Evolutionary conserved extended isoforms are known (Ivanov et al. 2011). Some of them have been shown to have important biological functions, e.g. in PTEN (Hopkins et al. 2013). Here we detected translation of OGD-induced

extended protein isoforms with upstream in-frame non-AUG initiation sites of five genes, *Adm*, *Bcl211*, *Fam178b*, *Ptms*, and *Ppp1r2* (Fig. 4.5B and the mini web site). In *Ptms* mRNA, encoding parathymosin, under normal conditions nearly all ribosome footprints are concentrated within the acORF. After 1 h OGD, the ribosome density upstream of acORF increased 15 fold (Fig. 4.5B).

Whereas translation of uORFs in the 5' leader seems to be widespread in mammalian mRNAs, translation of ORFs downstream of the acORF is infrequent. One known example is the bicistronic mRNA *Rpp14* (Autio et al. 2008). The first acORF of this mRNA codes for ribonuclease P subunit (Rpp14) while the second acORF codes for 3-hydroxyacyl-thioester dehydratase 2 (HsHTD2) (Autio et al. 2008). Both appear to be inhibited by OGD-induced uORF translation. Intrigued by these cases we searched for other examples of mRNAs with translated ORFs located in 3'UTRs. In total we identified six candidate bicistronic mRNAs (available at the mini web site), one of the most highly translated downstream ORFs is in the mRNA of the *Eif1b* gene (Fig. 4.5C). While translation of the acORF encoding eIF1b is increased during OGD, translation of the downstream ORF is inhibited (Fig. 4.5C).

In the *Romo1* mRNA encoding a reactive oxygen species modulator protein we observed increased ribosome density immediately downstream of the acORF (Fig. 4.5D). This may indicate a regulated C-terminal extension due to frameshifting, stop codon readthrough or translation of a long non-AUG initiated overlapping ORF. The latter is supported by the lack of clear triplet periodicity of ribosome footprints aligning to the acORF (presumably due to the simultaneous translation of two frames). A similar situation is observed in the mRNA of *Amn*.

Particularly interesting is the OGD-induced pause occurring in *Srpr* mRNA which encodes the β subunit of the signal recognition particle (SRP) receptor. Unlike other position specific pauses that we described earlier (Fig. 4.3F), The *Srpr* pause is accompanied by a decreased ribosome density downstream of the pause site (Fig. 4.5D). This may occur in case of abortive translation due to ribosomal frameshifting or drop-off. Expression of truncated SRPR may influence ER localized translation.

CONCLUSIONS

Using ribosome profiling we explored the immediate response of neural cells gene expression to oxygen and glucose deprivation at the transcription and translation levels. This condition, mimicking ischemia, imposed a dramatic stress on cellular bioenergetics and mitochondrial function (Fig. 4.1A, B). The most surprising finding is the wide spread effect of OGD on translation in comparison with its effect on the transcriptome. While RNA levels of only slightly more than a hundred of genes were affected during the course of 1 h OGD, translation of about a fifth of all expressed genes underwent significant alterations. The other striking observation is the speed of the response: most of the changes occurred rapidly under the first 20 minutes and were slowly amplified during continuation of the exposure to OGD (Fig. 4.2). Rapid translational response (<30 mins) has been described for a number of changing conditions, such as a change in the source of carbon in *S. cerevisiae* growth media (Kuhn et al. 2001) and oxidative stress (Gerashchenko, Lobanov, and Gladyshev 2012). In mammalian cells fast translational response has been observed in response to certain drug treatments, e.g. thapsigargin inducing unfolded protein response (Reid et al. 2014), arsenite inducing eIF2 phosphorylation (Andreev et al. 2015) and tunicamycin inducing Endoplasmic Reticulum stress (Sidrauski et al. 2015). Therefore rapid translational response preceding (and likely shaping) transcriptional response could be a general mode of how cells regulate their gene expression in reaction to various stimuli. The analysis of general translation parameters suggests that there is unlikely a single regulatory mechanism. The involvement of a multitude of factors is evident from global alterations of translation which include: increased translation of 5’ leaders similar to what has been observed in studies of translation response to other stresses (Shalgi et al. 2013; B. Liu, Han, and Qian 2013; Reid et al. 2014); numerous OGD-induced pauses that are indicative of context dependent local changes in elongation rates; reduced stringency of start codon selection in translation initiation sites and improved accuracy of translation termination. The ceasing of ribosome queuing at the ends of mRNAs could be either due to increased translation termination or decreased elongation rates, the latter is anticipated as a result of eEF2 phosphorylation. It would be very interesting to identify individual causing factors responsible for these changes. Our study points to some potential players involved - a number of mRNAs whose translation was altered encode proteins involved in translation regulation, e. g. certain translation initiation factors and eEF2k, the only kinase targeting

eEF2. The changes in the translation of these mRNAs may well be due to autoregulatory mechanisms.

Despite the wide spread effect that OGD produced on gene expression, the response targets functionally relevant genes (particularly at a transcriptional level) as is evident from gene ontology analysis of differentially expressed genes. Oxidative phosphorylation was affected the most along with the closely related TCA cycle, as well as neurodegeneration pathways and ubiquitin dependent proteolysis (Fig. 4.2D).

We also conducted detailed characterization of individual translation response for a number of affected mRNAs. This revealed dozens of regulatory uORFs, (many of them are non-AUG initiated), translation of extended protein isoforms, six polycistronic mRNAs and two novel dual coding mRNAs.

This study suggests that cellular gene expression reacts to OGD prior to (or even instead of) HIF promoted transcriptional response by altering translation of available mRNAs. It is intriguing to speculate that some components of translational machinery are regulated by oxygen availability and trigger specific responses upon oxygen deprivation. Indeed, a number of translation components have been shown to be hydroxylated under normoxic conditions (Singleton et al. 2014; Loenarz et al. 2014). It is likely that some of the alternative translation products and protein isoforms detected in this study have crucial implications in cell response to OGD. Their functional roles under this stress merit further investigation.

METHODS

Materials. TMRM and JC-1 probes were from Invitrogen Life Technologies (Carlsbad, CA and Dun Laoghaire, Ireland). Amersham™ ECL™ Prime Western blotting reagent was from GE Healthcare Life Sciences (Waukesha, WI), pre-made acrylamide gels, running and transfer buffers were from GeneScript (Piscataway, NJ), RIPA buffer, BCA™ Protein Assay kit and Pre-stained Protein Ladder were from Thermo Fisher Scientific (Rockford, Ill). CellTiter-Glo® ATP Assay was from Promega (Madison, WI). PhosphoStop Phosphatase Inhibitor and complete Protease Inhibitor Cocktail Tablets were from Roche (Ireland). Dulbecco's Modified Eagle's medium (DMEM), Roswell Park Memorial Institute (RPMI) media,

cycloheximide and all the other reagents were from Sigma-Aldrich. Primary and secondary antibodies are listed in Table E1. Plastic and glassware were from Sarstedt (Ireland), Corning Life Sciences (Corning, NY), Greiner Bio One (Frickenhausen, Germany) and Pecon (Erbach, Germany).

Tissue culture and experimental conditions. PC12 cells from ATCC (<40 passages) were maintained in suspension in RPMI 1640 medium supplemented with $NaHCO_3$, 2 mM L-glutamine, 10% horse serum (HS), 5% fetal bovine serum (FBS), 100 U/mL penicillin and 100 μg/mL streptomycin (P/S), in a humidified incubator set to 5% CO_2 and 37°C. For all experiments, PC12 cells were seeded at 4-6·10^4 cells / cm^2 on 75 cm^2 flasks, 10 cm or 15 cm Petri dishes and 96 well plates (all coated with collagen IV at 0.01 mg/mL) or 4.2 cm glass coverslips (Pecon), coated with a mixture of collagen IV and poly-D-lysine (Zhdanov et al. 2013). Cells were maintained in the adherent state for up to 3 days. In cell monolayers O_2 is distributed more uniformly than in small clumps, common for PC12 cell suspension (Greene and Tischler 1976), therefore data variability was expected to be lower for adherent cells.

OGD and OD experiments. Media for OGD and O_2 deprivation (OD) were prepared similarly to (Zhdanov et al. 2013) as follows. Powder DMEM (Sigma, cat. No 5030) was reconstituted in deionised water, buffered with 20 mM HEPES (pH 7.2) and filter-sterilised. Using this plain DMEM the experimental media were composed by addition of 10 mM glucose, 2 mM glutamine and 1 mM pyruvate (OD medium) or glutamine and pyruvate only (OGD). No serum was added. Media were equilibrated for 20 h at 0% O_2 using the hypoxia workstation (Coy Laboratory Products, Grass Lake, MI). Media deoxygenation was monitored using pre-calibrated phosphorescent dOxyBead™ sensors (Luxcel Biosciences, Ireland) and OpTech™ O_2 Platinum handheld detector (Mocon, MN).

Adherent cells grown in a regular CO_2 incubator were transferred to the hypoxia workstation (Coy) equilibrated with 95% N_2 and 5% CO_2, (0% O_2) at 37 °C. Growth medium was rapidly replaced with deoxygenated OGD or OD media: 20 mL for 15 cm Petri dish, 10 mL for 10 cm dish, 1 mL for 4.2 cm coverslip, 100 μL for 1 well of 96-well plate. Cells were then incubated in anoxic conditions for 20 min, 40 min, 1 h, 2 h and 4 h. Control cells were exposed to the oxygenated OGD or OD media under normoxia for 1 h. At the time points indicated, cells were rapidly extracted from the workstation and subjected to analysis.

Generation of ribo-seq and mRNA-seq libraries. The ribosomal profiling technique was carried out as in Ingolia et al. (Ingolia et al. 2012) but with a few modifications outlined in Andreev et al. (Andreev et al. 2015). Libraries were sequenced on an Illumina HiSeq 2000 system at the Beijing Genomics Institute (BGI).

Protein isolation and Western blotting analysis. Standard analysis of whole cell lysates prepared using RIPA buffer was performed using Western blotting analysis as in Zhdanov et al (Zhdanov et al. 2013). Quantitative data analysis was conducted with ImageJ program using α-tubulin signals for normalization. Images were processed with Picasa, Photoshop, and Illustrator programs. Experiments were carried out in triplicates.

ATP measurement. Cellular ATP was measured using CellTiter-Glo® Assay (Promega), which allows for the end-point high throughput quantitative analysis of cellular energy state. 100 uL of pre-mixed kit reagents were added to the cells directly in hypoxia workstation. A plate with cell lysates was transferred to normoxia. After intensive shaking the lysates were transferred into white 96-well plates (Greiner Bio One) and their luminescence was measured on a Victor 2 reader.

Measurement of the mitochondrial membrane potential. For monitoring of the mitochondrial membrane potential (ΔΨm) we utilised two commercial fluorescent probes, TMRM and JC-1. When used at non-quenching concentrations (1-20 nM), TMRM rapidly accumulates in polarised mitochondria and is easily released from depolarised mitochondria. TMRM is also sensitive to the changes in the plasma membrane potential; therefore additional controls are required for ΔΨm analysis. Upon the accumulation in the polarised mitochondria, JC-1 exhibits a split to green and red fluorescent spectra. The latter is only produced by highly polarised ('energized') mitochondria, when JC-1 forms J-aggregates. To ensure equal loading conditions, we stained the cells at 21% O_2 for 25 min (1 µM JC-1 and 20 nM TMRM). TMRM probe was maintained in the media at 20 nM during the whole experiment. For imaging, the coverslips with cells were locked in a mini hypoxia chamber in ~1 mL of the oxygenated or deoxygenated medium, using 3 silicon ring gadgets and a second coverslip, which were sandwiched air-tightly together with a metallic base with the help of the screwing cap. Upon incubation of the cells as described in the Results section, ΔΨm then was quickly monitored (within 10 min) on a wide field fluorescence microscope (Zeiss, Germany). TMRM was excited using 590 nm 10 mW LED with emissions

collected at 604-644 nm. JC-1 was excited at 488 nm and 590 nm, while collecting emission at 510-550 nm and 604-644 nm, respectively.

Initial processing of sequence libraries. Cutadapt (Martin 2011) was used for the removal of the adapter sequence (CTGTAGGCACCATCAATAGATCGGAAGAGCACACGTCTGAACTCCAGTCA). The reads were then aligned to rRNA and the introduced "spike-in" (ATGTACACGGAGTCGACCCGCAACGCGA) to remove reads that originated from these sources. Reads were aligned to the rat RefSeq gene catalogue downloaded from NCBI on January 2014 (ftp://ftp.ncbi.nih.gov/refseq/R_norvegicus/mRNA_Prot/) with bowtie (Langmead et al. 2009). The alignment parameters used were (-a -m 100 -v 2 –norc), i.e. reads were aligned to the positive strand allowing no more than two mismatches and no more than 100 mappings per read.

Differential gene expression analysis. For differential expression analysis we calculated the number of reads aligning to an exon union of a gene (Olshen et al. 2013). Any read aligning to any of RefSeq transcripts originating from the same gene was counted once irrespective of the number of isoforms to which it aligns. Such a gene centric approach should provide a more accurate representation of gene expression than alignment to the longest transcript, since even the longest transcript may not contain all exons. ~2% more reads were aligned to the exon unions of the genes in comparison with the alignment to the longest transcripts. For certain genes (including *Nnat*, *My16*, *Cabin1*) we observed more than a two-fold increase in the number of aligned ribo-seq and mRNA-seq reads.

Ribo-seq reads were assigned to mRNA coordinates based on the inferred location of the A-site. The A-site location was set at 17 nucleotides downstream of the 5' end of reads of length 29 to 33 inclusive and 18 for reads of length 34 and 35. Ribo-seq reads of length less than 29 or greater than 35 were discarded. For genes that had transcripts with an annotated coding region we used only the ribo-seq reads aligning to it. Otherwise the ribo-seq reads mapped to the entire transcript were included (note for the mRNA-seq, the reads aligning to the entire transcript were used). Some ribo-seq reads were mapped to more than 1 region of a transcript. We assigned them to unique locations based on the following priority order: acORF, 5' leader, 3' UTR.

Approximately 80% of mapped ribo-seq reads were mapped to a single gene. 9% mapped to two genes and 4% mapped to three. For the mRNA-seq reads these values are 84%, 8% and 3% respectively. Ambiguously aligned reads that aligned to more than three locations were discarded. The weighting of reads aligned to 2 or 3 locations was reduced by 2 and 3 respectively.

Normalisation (the rescaling of read counts to remove the differences due to the total number of mapped reads) and differential analysis were carried out with an approach similar to the one we used previously (Andreev et al. 2015). The read counts aligning to a feature in the sample *k* were divided by the rescaling factor *x[k]/min(x[i])* where *x[k]* is the number of all sequence reads aligning to all features within sample *k*, and *min(x[i])* is the number of all reads aligning to all features in the sample *i* which has the smallest number of reads. This rescaling was carried out independently for the mRNA-seq and ribo-seq data, however both replicates were normalised together to enable for comparison between replicates.

To identify differentially expressed genes we carried out the Z-score transformation of log ratios (Quackenbush 2002). Genes were grouped in bins of 300 based on the minimal level of their expression (minimal number of mRNA-seq, ribo-seq or both depending on the analysis). The parameters of expression fold-change distribution were used to calculate local Z-score for every gene. The gene was classified as DE increased if $(Z_1+Z_2)/2 > T$ and DE decreased if $(Z_1+Z_2)/2 < -T$, where Z_1 and Z_2 are Z-scores for the same gene obtained in two replicas and T is the threshold. The threshold T was set based on the desired false discovery rate (FDR). The number of false positives was estimated as the number of genes for which $(|Z_1|+|Z_2|)/2>T$ when $Z_1Z_2<0$ (see Fig. 4.2B for the graphical illustration of the procedure).

Owing to misalignments, amplification biases and changes in elongation rates between the conditions the total number of reads mapped to a gene is not necessarily an accurate representative of its level of expression. For example a strong condition-dependent ribosome stalling at a specific location could lead to spurious upregulation. A gene was considered robustly regulated only if it was detected as regulated after exclusion of the acORF coordinates corresponding to the three highest density peaks of the ribosome density for each of its transcripts.

Analysis of pairwise similarity of ribo-seq profiles of individual mRNAs. The pairwise similarity between two acORF ribo-seq profiles was assessed by calculating Pearson's correlation coefficients for local footprint densities of the compared profiles. Only unambiguously aligned ribo-seq reads were used for this analysis. The position of the reads aligning to the acORF region was determined at a codon, rather than at a nucleotide level. Only the longest transcript variants of each gene that had an average acORF density exceeding 1 footprint per nucleotide in both compared samples were used in this analysis. The process was repeated using alignments produced by randomly sampling a fixed number of reads for each mRNA from the real data. In each pairwise comparison, for the conditions with the higher sequence coverage ribo-seq reads were sampled from the original alignment until the number of reads equal that for a condition with lower sequencing coverage. The average pairwise Pearson's coefficient of these profiles were used to produce the distribution.

Identification of genes with leaky termination. The profiles of transcripts that had a total number of ribo-seq reads aligning to their 3'UTR greater than 20 were manually examined. Ambiguous alignments were included. The selection was made without prior knowledge of the stop codon identity and nucleotide context. The context of termination sites was visualized using sequence logo produced with weblogo (Crooks et al. 2004). The ratio of ribo-seq read density (alignments at ORF/ length of ORF) at the acORF and downstream ORF was used to measure stop codon readthrough efficiency.

Identification of translated uORFs. The probability that the distributions of highly translated genes with and without a translated uORF (Fig. 4.4A) are sampled from the same distribution was calculated by Wilcoxon rank sum test implemented with the ranksums function in the Scipy library. To enrich the gene pool for those with regulatory uORFs we analysed the shift of the centre of ribosome density (Andreev et al. 2015). To determine the center of ribosome density we first produced "compressed profiles", where every coordinate of an mRNA profile that did not contain an alignment in at least one of the time points was removed. This has the advantage of producing a more robust result with a sparse number of reads. The center of ribosome density was defined as the minimal coordinate of a compressed profile at which the cumulative number of ribo-seq reads aligning upstream of it exceeds the cumulative number of reads downstream of it. The shift

of the center was measured relative to the length of translated regions of mRNAs, i.e. divided by the number of coordinates with at least one ribo-seq read alignment. The change of center of density was obtained for all annotated coding transcripts with greater than 64 ribo-seq read alignments. The profiles of over 200 transcripts with the greatest average shift of density towards the 5' end after 1 h OGD (Fig. 4.4B) were manually evaluated.

As the strongest triplet periodicity was observed with ribo-seq reads of 31 nt in length, these alone were used for the analysis of a triplet periodicity signal. To identify cases with an atypical periodicity we used the alignments to the first 50 coordinates of the acORF. For most transcripts ~20% of reads aligned to the 3rd subcodon position; a higher proportion of ribo-seq reads aligning to the 3rd subcodon position was considered as an indicator of triplet periodicity distortion. Transcripts containing fewer than 50 coordinates with aligned ribo-seq reads were excluded from this analysis.

Production of metagene profile at initiation and termination sites. The transcript used for generating a metagene profile was selected to satisfy the following criteria: 1) it is the longest among other transcript variants for the corresponding gene; 2) it has at least 100 mapped alignments; 3) its length is greater than 600 nt; 4) the length of both, its 5' leader and 3'UTR is greater than 45 nt. Each individual transcript profile from 45 nt upstream of the initiation site to 45 nt downstream of the termination site was normalised by the average gene density before the aggregation by the determination of a position specific average. This was converted to the average number of alignments per codon.

Analysis of biological pathways. The 3000 genes found to have the greatest absolute average ribo-seq Z-score after 60 min OGD (DE increased and DE decreased) were used for GO and KEGG pathway enrichment analysis using DAVID (Huang, Sherman, and Lempicki 2009). The statistical significance of gene enrichment was corrected for multiple testing using Benjamini-Hochberg method.

rHRE containing and mTOR sensitive mRNAs. The list of transationally regulated PP242 responsive genes from supplementary figure 5 of Hsieh et al. (Hsieh et al. 2012) was used as a set of mTOR sensitive genes. The set of transcripts containing rHRE containing

transcripts was built from supplementary data 2 of Uniacle et al. (Uniacke et al. 2012) requiring at least 3 PAR-CLIP reads supporting EPAS1/RBM4 target.

All computations and plots were produced with custom python scripts using Matplotlib library (Hunter 2007).

Data access. Sequences of ribo-seq and mRNA-seq cDNA libraries have been deposited to NCBI Genome Expression Omnibus (GEO) under accession number GSE60752.

rRNA:mRNA pairing alters the length and the symmetry of mRNA protected fragments in ribosome profiling experiments.

This chapter has being published as a research article in Bioinformatics 2013 Jun 15;29(12):1488-91

ABSTRACT

Motivation: Ribosome profiling is a new technique that allows monitoring locations of translating ribosomes on mRNA at a whole transcriptome level. A recent ribosome profiling study demonstrated that internal Shine-Dalgarno (SD) sequences have a major global effect on translation rates in bacteria: ribosomes pause at SD sites in mRNA. Therefore it is important to understand how SD sites effect mRNA movement through the ribosome and generation of ribosome footprints.

Results: Here, we provide evidence that in addition to the pausing effect, internal SD sequences induce a caterpillar like movement of mRNA through the ribosome cavity. Once an SD site binds to the ribosome, it remains attached to it while the ribosome decodes a few subsequent codons. This leads to asymmetric progressive elongation of ribosome footprints at the 3' end. It is likely that internal SD sequences induce a pause not on a single, but on several adjacent codons. This finding is important for our understanding of mRNA movement through the ribosome and also should facilitate interpretation of ribosome profiling data.

INTRODUCTION

Translation initiation sites in bacteria usually consist of a start codon (most frequently AUG, GUG or UUG) and a purine reach region 6-10 nts upstream, termed a Shine-Dalgarno (SD) sequence (Shine and Dalgarno 1975). The SD sequence binds to the small ribosomal subunit through complementary interactions with an anti-Shine-Dalgarno (aSD) sequence located at the 3' end of 16S rRNA (3'-AUUCCUCCAC). The importance of SD sequences for translation initiation is well known; its presence is routinely verified in algorithms for start codon prediction in bacteria (Borodovsky and Lomsadze 2011; Hyatt et al. 2010).

Until recently the known role of internal SD sequences was limited to its stimulatory effect on ribosomal frameshifting in bacteria (Larsen et al. 1994; Weiss et al. 1988; Weiss et al. 1987). Recent single molecule studies revealed that internal SD sites can arrest ribosome

movement along mRNA (Wen et al. 2008). Application of the ribosome profiling technique (Ingolia et al. 2009) to the analysis of protein synthesis in bacteria demonstrated that this effect is pervasive (Li, Oh, and Weissman 2012). It was shown that ribosome density is positively correlated with the presence of strong SD sequences and strong SD-aSD interactions cause pausing of elongating ribosomes. Under conditions of fast growth, the SD site effect on translation rates appears to be more significant than that of other known coding sequence features, such as codon bias (Li, Oh, and Weissman 2012).

However, the pausing effect of SD sequences on elongating ribosomes cannot explain all the empirical observations regarding its stimulation of ribosomal frameshifting. Indeed the effect of an SD site on frameshifting is strictly determined by its position relative to the frameshift site. For stimulation of +1 ribosomal frameshifting, an SD site needs to be located just 3 nucleotides upstream of the P-site codon (Devaraj and Fredrick 2010; Weiss et al. 1987), while stimulation of -1 frameshifting requires a distance of 10-14 nucleotides (Larsen et al. 1994). In bacterial release factor 2 genes, which have a programmed frameshift site, not only the SD site, but its distance from the frameshift site is highly conserved (Baranov, Gesteland, and Atkins 2002). In addition to interference with a tRNA in the E-site (Márquez et al. 2004), the position specificity of the SD site effect on frameshifting could be due to tensions imposed on the bulk of mRNA located between the decoding center and an SD:aSD duplex. A simple analogy for such tensions is a “spring effect” illustrated on Fig. 1A. A short spacer is expected to produce “a stretched spring effect” promoting a movement of codons in the 5’ direction relative to tRNAs (tRNAs move forward). A long spacer is expected to produce the opposite “compressed spring effect”. The model illustrated in Fig. 1A requires the SD site to be able to interact with aSD at a range of distances from the decoded codon. Such interactions would force mRNA to move through a ribosome in a caterpillar-like rippling fashion. This could occur only if ribosomes are capable of accommodating variable amount of mRNA between the SD:aSD duplex and the decoding center. In this case ribosomes would remain bound to the same SD site while decoding several subsequent codons leading to asymmetric progressive extension of ribosome footprints as shown in Fig. 1B.

In this work we analyzed ribosome profiling data obtained for *Escherichia coli* and *Bacillus subtilis* and have shown that the length and distribution of ribosome footprints at

SD site containing mRNA regions is consistent with “the caterpillar model” of progressive footprint extension.

METHODS

All ribosome profiling data are from (Li, Oh, and Weissman 2012). The footprint sequences were aligned to the organisms’ reference genomes, NC_000964.3 for *B. subtilis* and NC_000913.2 for *E. coli*. The reads were aligned allowing for two mismatches. Reads not aligning to a unique location were discarded. To ensure that footprints correspond to elongating and not initiating ribosomes, the only footprints used were those that align downstream of the first 30 nucleotides, but upstream of the 75 last nucleotides of coding ORFs. An SD site was defined as an octamer that is predicted to form a duplex with the aSD (5’CACCUCCU3’) with free energy (ΔG) less than 8 kcal/mol. ΔG was calculated as in (Li, Oh, and Weissman 2012) using RNASubopt tool from Vienna RNA package (Gruber et al. 2008). We found 1206 such SD sites in 704 genes from *B.subtilis* and 1213 in 809 genes from *E.coli*. Of 21 596 339 reads aligning the genes containing strong SD sites, 1 486 956 were found to interact with SD sites (6-fold increase over expected number). For *E.coli*, 1 668 187 reads were found to interact with SD of 56 389 881 reads (3-fold increase over the expected value).

The first nucleotide of the octamer was considered as the first nucleotide of SD site irrespective of whether it forms a base pair with aSD, e.g. the C in 5’CGGAGGUG3’ is considered as the first nucleotide of the SD site.

Distributions of footprint lengths (Fig. 2) were obtained for the 2,122 genes of *E. coli* and 1,468 in *B. subtillis* that had an average density greater than 10 footprints per codon. To avoid unequal effects from differentially expressed genes and pause sites, the frequency of footprint occurrence of a particular length *l* was calculated as follows:

$$f(l) = \frac{1}{\sum_j N_j} \sum_{j,k} \frac{x_{l,j,k}}{\sum_l x_{l,j,k}} \qquad \{1\}$$

Where $x_{l,j,k}$ is the number of footprints in SD j of gene k and N_j is a number of SDs in a gene j. Essentially, distributions for "No SD" footprints were normalized on a per gene basis, while distributions for "SD" footprints were normalized on a per SD basis. SD containing footprints were defined as those that have an SD site within the 3rd and 17th nucleotides from the 5' end of a footprint. "No SD" footprints were defined as those that do not have an octamer that can form a duplex with aSD at ΔG<-2 kcal/mol within the same region of a footprint. The significance of the difference between the two distributions in Fig. 2A was estimated using the Wilcoxon rank sum test.

The correlation analysis determined both the average normalized density and average length of fragments whose 5' end is between 8 to 2 nucleotides upstream for each hexanucleotide sequence (Fig 2B, C). Similar to above, this analysis was carried out for genes with average density greater than 10 footprints per codon. To reduce statistical noise, the plots include data points whose averages were obtained from at least 10 values.

For the plots in Fig. 1D, the *B. subtilis* footprints were mapped to the genome and all internal SD site locations were identified using the criteria described in the first paragraph of this section. Only the regions containing SD sites (50 nts upstream and downstream of the first nucleotide in an SD site) with the average footprint density higher than 1 per codon were used. The densities of the 5' and 3' ends of the footprints aggregated across all SD sites were calculated as follows:

$$d(s) = \frac{1}{N}\sum_i \frac{x_{s,i}}{X_i} \qquad \{2\}$$

Where $d(s)$ is the density of footprint ends at s coordinate relative to SD, e. g. s=1 for the first nucleotide of an SD site. $x_{s,i}$ is the number of footprint ends at the position s for SD site i; X_i is the average number of footprints per nucleotide in the region of SD site i and N is the total number of the SD sites used in the analysis. The plot in Fig. 1C was generated in the same way, but separately for the footprints of the different lengths, i.e. in equation {2}, both x and X values refer to the footprints of a particular length.

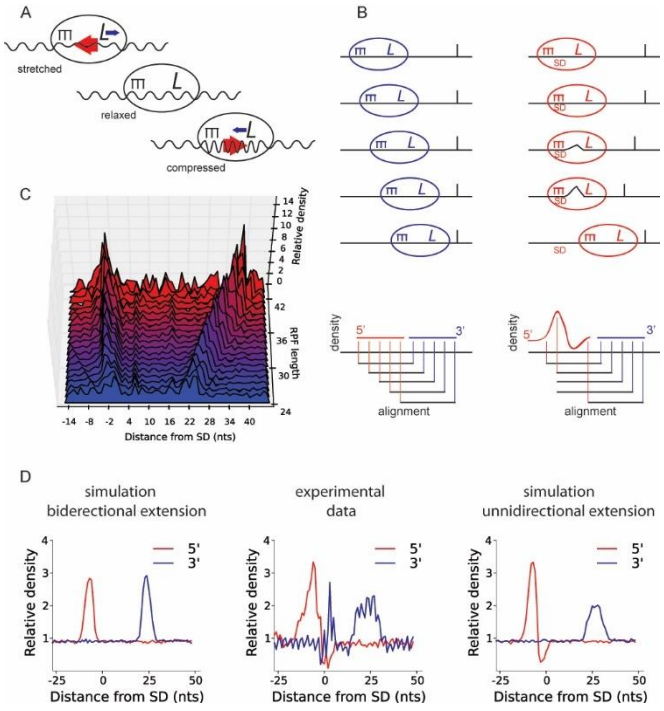


Fig. 1. "The caterpillar model" of SD effect on ribosome movement**. A.** mRNA tensions created by the distance between SD:aSD complex and the decoding center. Red arrows show the direction of the forces generated by the tensions and blue arrows show the direction of the tRNA movement relative to the mRNA during ribosomal frameshifting. **B.** A model of footprint generation for ribosomes moving along mRNA with (red) and without (blue) SD interactions. As ribosomes move at a constant speed, a peak of ribosome density (followed by density displacement) is expected if measured using locations of footprint 5' ends. **C.** Density of footprint ends (both 5' and 3') relative to SD site (see Methods) for footprints of different lengths. It can be seen that the location of the 5' peak remains unchanged, while the location of the 3' peak depends on the length of footprints. **D.** The experimental ribosome profiling data at pause sites match a model of progressive extension of footprints length. Ribosome density in the area of an SD site as estimated with the 5' (red) and the 3' ends (blue). The plot in the center was constructed using *E. coli* experimental data. The plot on the left is a result of a model where SD site induces a pause on a single codon and footprint extensions are uniform at both ends. The plot on the right corresponds to a model where SD site induces a pause at three adjacent codons resulting in asymmetrical extension of footprints relative to the P-site codons. The latter model better reflects the distribution obtained from the experimental data.

To simulate ribosomal profiling data *in silico,* 1000 random 1000 nt long sequences were generated. SD site positions on each sequence were determined based on free energy of the corresponding SD:aSD complex as described above. A series of 600 footprints were

generated for each sequence. Footprints were generated of a varied length (20 to 47 nts) with their centers corresponding to P-site codon locations. The distribution of footprint lengths was modeled based on an empirical distribution (Fig. 2A) of footprint lengths obtained from the experimental data. P-site codon locations of ribosomes were chosen randomly with equal probability $p_i=p$ for rapidly translated codons at coordinates i and $p_x=9p$ for a slowly translated codons at coordinates x. The slowly translated codons were defined as those whose coordinates are located 11 nts downstream of a SD. To model SD site dependent extension of footprint lengths, footprints were generated under a similar model with two alterations. First, to distribute the pause over three adjacent codons, the probability of P-site codon locations were set at $p_x=p_{x+3}=p_{x+6}=3p$. Second, to model asymmetric extension, footprint lengths were extended at their 5'-end for P-site codon locations at x+1 to x+6 by the distance between the P-site and x.

RESULTS AND DISCUSSION

"The caterpillar model" of SD mediated footprint elongation is shown in Fig. 1B. The tight association of an SD:aSD duplex could make disruption of the SD:aSD duplex more energetically expensive than incorporation and folding of extra mRNA into the cavity between ribosome subunits, thus allowing for increased spacer distance between the SD site and the codon in the P-site. In this scenario, footprints containing internal SD sequences in close proximity to their 5' ends would be expected to be longer on average than footprints lacking these sequences (Fig. 1B). To ascertain whether this is the case we compared the length of ribosome footprints obtained by (Li, Oh, and Weissman 2012) at locations with strong internal SD sites with those found at non-SD sites (Fig. 2), see Methods. The distribution of footprints derived from sequences not containing an SD site is skewed to the right. On average, the length of ribosome footprints without an SD site is close to the lower boundary of the distribution. By contrast, the distribution of SD-containing footprints is shifted towards longer footprints and appears more symmetric, suggesting that on average SD-containing footprints are longer, as expected from the model in Fig. 1. Moreover there is a correlation between footprint length and strength of the SD:aSD complex (Fig. 2B). The correlation is not strong (r=0.58 for *B. subtilis* and r=0.53 for *E. coli)*, due to the large length variance of footprints not containing SD sequences (Fig. 1). However, we observed a consistent trend toward an increase in the average footprint length at stronger SD sites. If ribosome pausing and footprint elongation are both the

effects of SD sequences, then a correlation between ribosome density and the length of footprints should exist. Indeed a strong correlation is observed (Fig. 2B), r=0.80 for *B. subtilis* and r=0.71 for *E. coli*. The correlation is nearly as strong as the correlation observed between ribosome density and the strength of SD:aSD duplex, r=0.86 (Li, Oh, and Weissman 2012) and is stronger than correlation between strength of SD:aSD and the length of ribosome footprints. This
raises a question regarding existence of an additional link between the length of footprints and ribosome pausing.

To investigate directionality of footprint extensions we analyzed aggregated density of both 3’ and 5’ footprint ends relative to SD site position for the footprints of individual lengths. Fig. 1C shows these distributions for footprints obtained from ribosome profiling experiments in *B. subtilis* footprints. Two distinct peaks can be observed for each footprint length, the left peak contains predominantly 5’ ends while the right peak contains predominantly 3’ ends. It is clear that irrespective of the footprint length, 5’-end locations remain unchanged. At the same time, 3’-end locations shift downstream as the length of footprints increases. Therefore SD site containing footprints are extended unidirectionaly at the 3’-ends as expected from the model in Fig. 1.

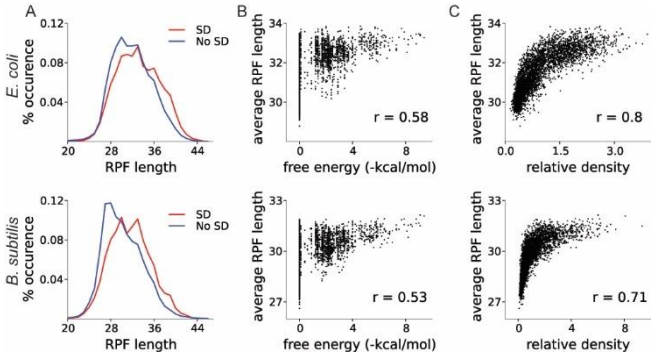


Fig. 2. SD containing footprints have increased length. **A.** Length distribution of footprints (RPFs, ribosome protected fragments) containing strong SD sequences (red) and containing no SD (blue) for *E. coli* (top) and *B. subtilis* (bottom). The difference between distributions is statistically significant; Wilcoxon rank sum test p<2.2e-16 in both cases. **B.** Correlation between the strength of a hexamer SD:aSD duplex and the length of footprints. **C.** Correlation between density of ribosomes on mRNA and the length of footprints.

To verify this observation, we designed two *in silico* models of ribosome pauses (see Methods). The first model assumes that footprints are equally variable at both sides of a P-site codon with SD site inducing a pause at a single codon. The second model assumes that footprints at SD sites could be extended at their 3'-ends, allowing slow translation for three codons. Under the first model, similar distributions of the density would be observed for both 5' and 3' footprint ends with their means separated by the average length of footprints (Fig. 1D, left). In the second model, a dip can be seen shortly downstream from the 5' end peak; also the 3' end peak is wider than the 5' end peak (Fig. 1D, right). The distribution of experimentally obtained footprints matches the behavior of the second model more closely (Fig. 1D, center).

Since development of the ribosome profiling technique (Ingolia et al. 2009), it has become commonly accepted that ribosome density is a direct measure of translation kinetics, e.g. a high density of ribosome footprints indicates slowly decoded mRNA locations (N T Ingolia et al. 2009; Stadler and Fire 2011). However, as we demonstrate here, the footprint extensions are sequence dependent and are not equivalent at the 5' and 3' ends. This implies that a precise position of the A-site codon in a footprint could also be variable in a sequence dependent manner. It is possible that the length of footprints in eukaryotes may also vary due to potential interactions between mRNA and components of

the ribosome. However, if they can happen, it is unlikely that such variations in eukaryotes are as pervasive as in bacteria as evidenced from the sharp triplet periodicity obtained in eukaryotic systems (Guo et al. 2010; Ingolia et al. 2009). The strength of triplet periodicity obtained from alignments of footprint 5'-ends is sufficient to permit detection of translated frames and switches between them (Michel et al. 2012). This suggests that unlike in bacteria the distance between the A-site and the 5'-end of eukaryotic ribosome footprints is predominantly uniform. However, if site specific footprint length variations occur in eukaryotes on specific sequences, it would affect ribosome profiles of individual mRNAs. Therefore it is important to take this possibility into account during interpretation of ribosome profiling data for individual mRNAs.

Comparative survey of the relative impact of mRNA features on local ribosome profiling read density

This chapter has being published as a research article in Nat Commun, 7, 12915. Journal Article. https://doi.org/10.1038/ncomms12915

ABSTRACT

Ribosome profiling (Ribo-seq), a promising technology for exploring ribosome decoding rates, is characterized by the presence of infrequent high peaks in ribosome footprint density and by long alignment gaps. Here, to reduce the impact of the data heterogeneity we introduce a simple normalization method, Ribo-seq Unit Step Transformation (RUST). RUST is robust and outperforms other normalization techniques in the presence of heterogeneous noise. We illustrate how RUST can be used for identifying mRNA sequence features that affect ribosome footprint densities globally. We show that a few parameters extracted with RUST are sufficient for predicting experimental densities with high accuracy. Importantly the application of RUST to 30 publicly available Ribo-seq datasets revealed a substantial variation in sequence determinants of ribosome footprint frequencies, questioning the reliability of Ribo-seq as an accurate representation of local ribosome densities without prior quality control. This emphasizes our incomplete understanding of how protocol parameters affect ribosome footprint densities.

INTRODUCTION

The advent of ribosomal profiling (ribo-seq) has provided the research community with a technique that enables the characterization of the cellular translatome (the translated fraction of the transcriptome). It is based on arresting translating ribosomes and capturing the short mRNA fragments within the ribosome that are protected from nuclease cleavage. The high throughput sequencing of these fragments provides information on the mRNA locations of elongating ribosomes and thereby generates a quantitative measure of ribosome density across each transcript. Accordingly, ribosome profiling data contain information that could be used to infer the properties that affect ribosome decoding (or elongation) rates. Unsurprisingly, a large number of studies analyzing ribosome profiling data for this purpose have been published recently (Artieri and Fraser 2014; Dana and Tuller 2012; Dana and Tuller 2014b; Dana and Tuller 2014a; Gardin et al. 2014; Ingolia,

Lareau, and Weissman 2011; Lareau et al. 2014; Li et al. 2014; Li, Oh, and Weissman 2012; Lopez and Pazos 2015; Charneski and Hurst 2013; Pop et al. 2014; Shah et al. 2013; Stadler and Fire 2011; Tuller et al. 2010; Tuller et al. 2011; Tuller et al. 2010; Woolstenhulme et al. 2013; Yang, Chen, and Zhang 2014; Qian et al. 2012; Weinberg et al. 2016).

There is a considerable discordance among some of the findings in these works that is unlikely to be wholly caused by differences in the biological systems used. It may also be attributed to the computational methods used for estimating local decoding rates which are often based on elaborate models of translation that use certain assumptions regarding the process. The abstraction required for modelling necessitates the generalization of the process across all mRNAs, although we are aware of numerous special cases (Baranov, Atkins, and Yordanova 2015). Even if the generalized models provide an accurate representation of the physical process of translation in the cell, they do not model the ribosome profiling technique itself, which may introduce various technical artefacts. Oft-cited potential artefacts include the methods used to arrest ribosomes (the result is affected by the choice (Hussmann et al. 2015; Lareau et al. 2014) and the timing (Ingolia, Lareau, and Weissman 2011; Weinberg et al. 2016; Gerashchenko and Gladyshev 2014) of antibiotic treatment), the sequence preferences of enzymes involved in the library generation(Artieri and Fraser 2014; Bartholomaus, Del Campo, and Ignatova 2016) and the quality of alignment. These artefacts may distort the output and it may not be easy to disentangle their effects in the presence of biologically functional and sporadic alterations in translation.

Ribosome profiling data are characterized by high heterogeneity caused by alignment gaps and sporadic high density peaks due to technical artefacts and ribosome pauses[4,26]. These fluctuations, even if caused by genuine ribosome pauses, are thought to negatively impact the ability of some methods to accurately characterize factors that influence ribosome read density globally. With this rationale we developed a data smoothing method, that we term RUST (Ribo-seq Unit Step Transformation). We first demonstrate that RUST is resistant to the presence of heterogeneous noise using simulated data and outperforms other normalization techniques in reducing data variance. Then we analyze real data from 30 publicly available ribosome profiling datasets obtained using samples (cells or tissues) from human (Stadler and Fire 2011; Cenik et al. 2015; Andreev et

al. 2015; Guo et al. 2010; Hsieh et al. 2012; Lee et al. 2012; Fabricio Loayza-Puch et al. 2013; Lu and Deutsch 2008; Rooijers et al. 2013; Rubio et al. 2014; Shalgi et al. 2013; Stern-Ginossar et al. 2012; Stumpf et al. 2013), mice(Ingolia, Lareau, and Weissman 2011; Shalgi et al. 2013; Howard et al. 2013; Reid et al. 2014; Thoreen et al. 2012) and yeast (Artieri and Fraser 2014; Gardin et al. 2014; Lareau et al. 2014; Pop et al. 2014; Brar et al. 2012; Ingolia et al. 2009; McManus et al. 2014).

We show that a few parameters extracted with RUST are sufficient to predict experimental footprint densities with high accuracy. This suggests that RUST noise resistance allows accurate quantitative assessments of the global impact of mRNA sequence characteristics on the composition of footprint libraries.

The comparison of RUST parameters among different datasets revealed a considerable discordance in the relative impact of the sequence factors determining frequencies of ribosome footprints in the libraries. This most likely can be attributed to the differences in experimental protocols, suggesting that the variance in the data, rather than in the analytical approaches used is responsible for the current contradictions regarding the sequence determinants of decoding rates.

RESULTS

Ribo-seq Unit Step Transformation (RUST)

The probability of finding a ribosome decoding a particular codon of an mRNA (and by extension the expected number of corresponding ribo-seq reads in a library) depends on three variables: the mRNA expression level, the translation initiation rate for the corresponding open reading frame (ORF) and the time that the ribosome spends at that codon (dwell time). The latter (as an invert) is usually described as a codon elongation rate or a codon decoding rate. Estimating the true decoding rates with ribo-seq is made difficult by the absence of precise measurements of initiation rates. Therefore, studies (including this one) using ribo-seq for this type of analysis typically attempt to measure the relative dwell time of codons instead of the actual dwell time. A frequent and intuitive approach is the normalization of the local ribo-seq signal by the average signal across the coding region (Dana and Tuller 2014b; Li et al. 2014). This approach has been described as conventional

(Dana and Tuller 2014b) and we will refer to it as CN for conventional normalization. It is based on the reasonable assumption that the transcript expression levels and ORF initiation rates are the same for all codons from that ORF. CN is perceived to have two major shortcomings: it is expected to be very sensitive to the high density peaks which frequently occur due to functional ribosome pauses (Dana and Tuller 2014b) (Fig. 5.1a) and it is typically applied only to the transcripts with high ribosome coverage, as the relative impact of a single read alignment on CN is excessive with sparse profile data (Fig. 5.1a).

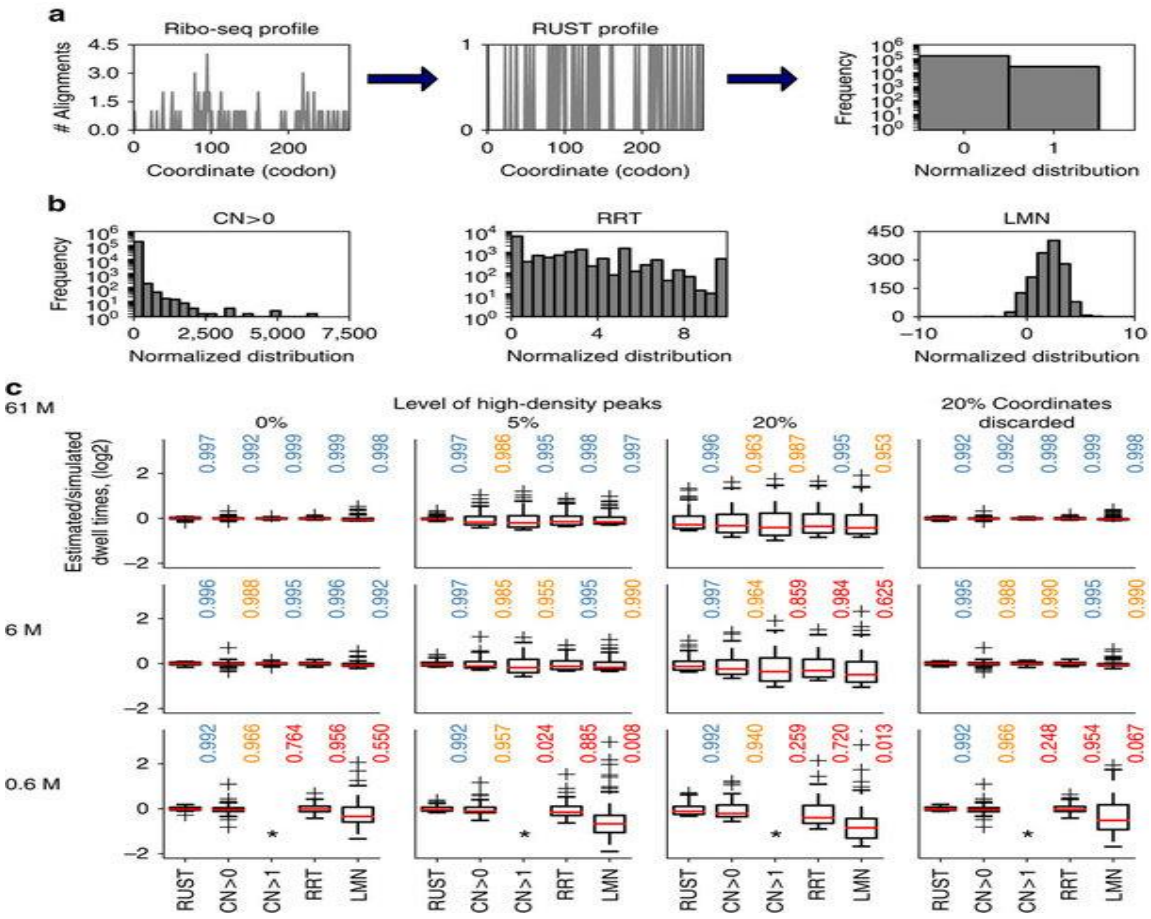


Figure 5.1. Comparison of ribosome profiling normalization approaches

(**a**) A stylized footprint density profile for *MTIF3* gene transcript from 'Andreev' dataset (left) is transformed into a binary function with RUST (center). Each sequence feature, such as AAA codon in the case shown, could be characterized by its frequency as 1 or 0 (right). (**b**) The distributions of normalized codon densities for all AAA codons in 'Andreev' dataset using different approaches, conventional normalization CN (left), ribosome residence time, RRT (centre) and logarithmic mean normalization LMN (right). Note that due to intrinsic differences the scale of possible normalized densities (axes X) varies among the methods and that due to the selection criteria of each approach the number of datapoints used (axes Y) is also variable. (**c**) Performance of five normalization approaches (RUST, CN of transcripts with average gene density >1/nucleotide (CN>1) and CN of all expressed transcripts (CN>0), LMN and RRT) at estimating codon dwell times The box plots show the

distribution of log values of the estimated/simulated dwell times for all 61 codons. The deviations of these values from 0 occur due to under or overestimation of simulated dwell times. The better methods are those that have distributions with a smaller variance. Each subpanel represents a specific scenario. The simulation scenarios differ by coverage that reduces from top to the bottom and the level of noise modelled as high peaks of density that increases from left to right, except for the right-most column where noise is modelled as missing data at 20% of the coordinates. Asterisks used to indicate insufficient data for CN>1.

Various approaches have been tried to reduce the impact of density outliers (Fig. 5.1b). Dana and Tuller removed atypical densities based on expected distribution of densities(Dana and Tuller 2014b). Artieri and Fraser used logarithmic mean instead of the arithmetic mean to produce a "corrected ribo coverage"(Artieri and Fraser 2014) (Fig. 5.1b). Gardin et al.(Gardin et al. 2014) developed an intricate approach for calculating a statistics that they called "ribosome residence time (RRT)". The approach involves CN like sampling, but only from specific segments of RNA that satisfy certain sequence and coverage requirements (Gardin et al. 2014) (Fig. 5.1b). Pop et al. introduced a sophisticated model that is based on the assumption that the ribosome footprint density profile must satisfy flow conservation constraints, i.e. the translation is at steady state and that all ribosomes translated the entire coding region (Pop et al. 2014). While flow conservation constraints may be true for the ribosome densities, they may not hold for footprint densities because of technical artefacts such as sequencing biases and misalignments.

We reasoned that a practical approach for the analysis of ribosome profiling data should be (i) simple; (ii) robust to the presence of heterogeneous noise; (iii) able to use all available data (i.e. no restriction to genes with high read coverage) and (iv) be able to produce statistics that would allow accurate prediction of experimental densities. With this in mind we developed a procedure that we term Ribo-seq Unit Step Transformation (RUST) where the ribosome footprint densities (the number of reads corresponding to the position of the A-site codon) are converted into a binary step unit function (also known as Heaviside step function). Individual codons are given a score of 1 or 0 depending on whether the footprint density at these codons exceeds the average for the corresponding ORF (Fig. 5.1a). In addition to codons, the procedure could be applied to any other potential determinant of read density such as individual nucleotides, encoded amino acids, their

combinations as well as their properties, such as a charge or hydrophobicity of encoded peptides or free energy of RNA secondary structures. The average RUST value for each putative determinant of decoding rates may be compared to the expected RUST value to measure its effect, see Methods. As a result of the transformation the impact of every site has a small influence on the final RUST value. The value is influenced primarily by the consistent presence of reads at numerous sites. For example, no differentiation is made between a stall site where the ribosome density just exceeds the average to one where the average is grossly exceeded. For the details of transformation, see Methods.

Evaluation of normalization methods with simulated data

In order to evaluate RUST performance, we tested its ability to estimate decoding rates from simulated data. We simulated the data under a simplifying assumption that the local decoding rates depend only on the identity of codon in the A-site. To simulate the data we used real transcript sequences and experimental distribution of footprints per transcript, but modeled the distribution of footprints within a transcript by specifying the dwell time of each of 61 codons and introducing different levels of heterogeneous noise (see Methods and below for how the noise was simulated). We compared its performance to the RRT approach, the CN method and to a logarithmic mean normalization (LMN) similar to that obtained with the “corrected ribo coverage” (see Methods). Unlike in the original approach in LMN ribo-seq density is not normalized by the mRNA-seq density. The CN method was used in two modes with filtering requiring a minimal coverage threshold (average transcript footprint density of >1 read/nucleotide) CN>1, and without any threshold, CN>0. The parameters of the simulation were selected either to produce data similar to the experimental data or the data with reduced quality (see Methods). For example, the sequencing depth was either equal or lower than what has been obtained with actual data.

Figure 5.1c compares the performance of the five methods for three different simulated sets of data with different sequencing depth and levels of noise simulated as sporadic high density peaks (3x the value of the highest footprint density for the original simulated profile) or as a loss of density that could arise, for example due to removal of

ambiguous mappings. For these simulations the relative time that ribosome dwell at each of 61 codons t_c was pre-set (see Methods) and the normalization approaches were compared in their ability to accurately detect codon dwell times (t_c) from the simulated data. The estimated-to-simulated dwell time log ratios were obtained for 61 codons. We assessed the performance of each method by showing the distribution obtained using box plots. For accurate methods the values for each codon should be zero, i.e. the observed and simulated values should be the same. We also provide the coefficient of determination, R^2, between the estimated and simulated dwell times as a measure of the normalization approaches accuracy, with values closer to one indicating better accuracy. We find that all approaches estimate relative t_c values very accurately in the absence of noise provided that coverage is high. However, in the presence of noise or under reduced coverage the performance worsens. In this regard RUST appears to be the most resilient to the reduced coverage and both types of noise. While its ability to accurately predict simulated relative dwell times drops under high levels of noise, the combined inferred values still correlate remarkably well with the simulated values (Fig. 5.1c).

We conjectured that the accuracy of the normalization approaches may depend on codon specific properties, such as a relationship between codon usage and dwell times. Therefore, we simulated the data under three different sets of t_c parameters. In the first two simulations the range of t_c values were set to rank-correlate with the codon usage (see Methods), i.e. the lowest t_c was set for the rarest codon and the highest t_c for the most abundant codon. In one set the t_c range spans one order of magnitude and in the other, two orders of magnitude. In the third set, the t_c parameters were set to negatively correlate with the codon usage. For the scenario where the range of decoding rates is increased to span two orders of magnitude the effect of noise on the accuracy of t_c inference is similar.

Interestingly, in all simulations (Fig. 5.1c) the logarithm ratios between estimated and simulated values are not uniform among 61 codons, i.e. the estimations are not equally accurate for each codon. The estimated relative dwell times of quickly decoded codons were found to be consistently overestimated by all methods tested, i.e. inferred as slower. This is more acute when the decoding rates span 2 orders of magnitude but is also observed even when the decoding rates span 1 order of magnitude. We also found that the

R^2 values were consistently lower when the codon usage negatively correlated with the simulated dwell time than when they were positively correlated. However, the difference is small suggesting that relationship between the codon usage and decoding rate appears to have a relatively minor influence on the correct estimation of the relative dwell time.

Counterintuitively in most simulations CN>0 performs similar or even better than CN>1 and the LMN was found to be inferior to both CN normalizations. Under almost all scenarios tested RUST was found to outperform other normalization techniques in the presence of noise.

The impact of technical biases varies among datasets

The velocity of a ribosome could be influenced by the sequence of mRNA in several ways (outlined in the scheme in Fig. 5.2a). Codons in the E-, P- and A-sites of the ribosome determine the identity of corresponding tRNAs (and amino acids) inside the ribosome. The mRNA sequence in the cavity between subunits could affect ribosome movement by directly interacting with its components. In addition, the sequence upstream of the A-site codon (up to 90 nucleotides) could influence the progressive movement of the ribosome through the interactions between the peptide it encodes and ribosome peptide tunnel. Lastly, the sequence downstream of the ribosome could alter its velocity through the formation of stable RNA secondary structures (Kontos, Napthine, and Brierley 2001; Tholstrup, Oddershede, and Sorensen 2012) or the presence of RNA-protein complexes.

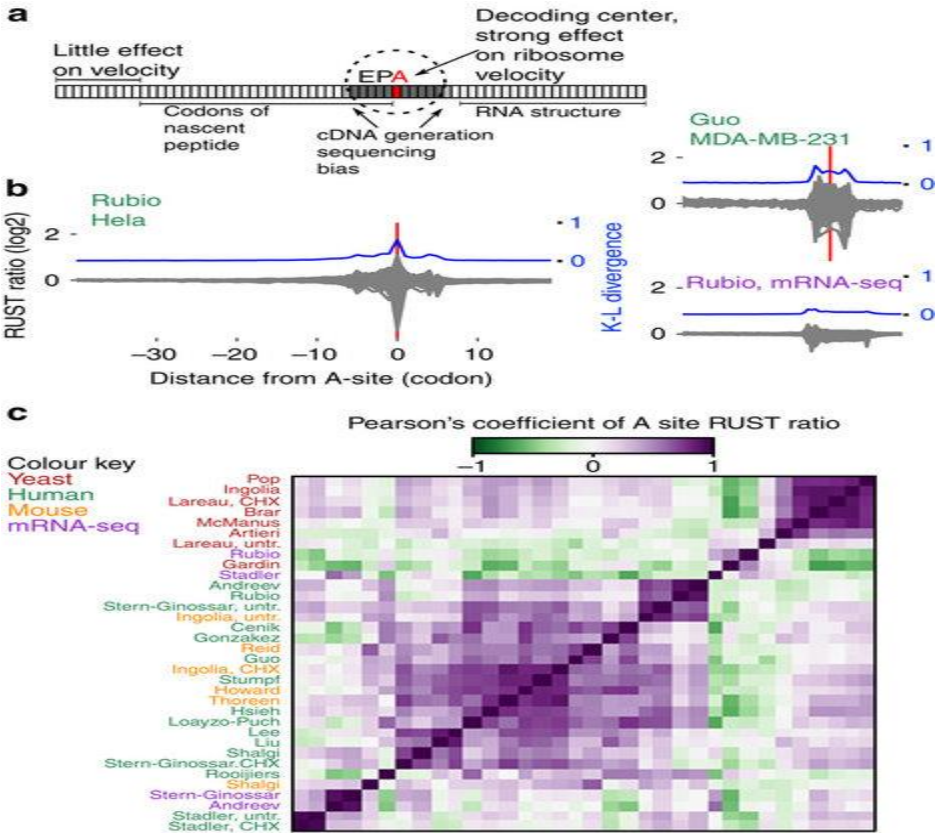


Figure 5.2. Evaluation of ribo-seq datasets with RUST

(**a**) Anatomy of the ribosome footprint displaying position-specific mRNA sequence influence on ribo-seq read density. (**b**) RUST codon metafootprint profiles of selected ribo-seq and mRNA-seq datasets used in this study. The individual RUST ratio values of 61 sense codons across the mRNA are displayed. The resulting grey area is a superposition of each 61 curves. The corresponding Kullback-Leibler divergence (K-L) is shown in blue. The protocol details for each dataset are summarized in Table 1. (**c**) Heatmap displaying the pairwise similarity of codon RUST ratios at the A-site, as measured by the Pearson's correlation, for ribo-seq datasets of human (green), yeast (red) and mouse (orange). Also included are human mRNA-seq data (violet). The datasets are indexed by the name of the first author. The clustering was done with Scipy using the "Euclidean" distance metric with "single" linkage.

In addition to these intrinsic factors affecting ribosome velocities, there are technical factors that influence the distribution of sequencing reads in ribo-seq datasets. First, the drugs used to block elongating ribosomes could act on ribosomes only at a specific conformation[8] or could also alter their distribution along mRNAs (Gerashchenko and Gladyshev 2014; Hussmann et al. 2015). Second, various enzymes used for cleaving mRNA, for generating cDNA libraries and for their sequencing exhibit sequence specificity especially at the boundaries of ribosome footprints (Artieri and Fraser 2014). Third, the

accuracy of alignment step depends on the existence of paralogs and transcript sequence complexity and the way how ambiguous alignments are treated. Fourth, the occurrence of alternative splicing, ribosome drop-off, ribosome stacking and alternative translation initiation may all affect the distribution of reads across individual transcripts.

To analyze how sequence of mRNA effect density of footprints in different locations relative to the A-site in experimental data we used an approach similar to the one used by Artieri and Fraser(Artieri and Fraser 2014). We calculated observed-to-expected RUST ratios for each codon position within a window of 60 codons (see Methods). This window encompasses the ribosome protected fragment (codons -5 to +5), the region encoding the nascent peptide (codons -30 to 0) and the region downstream of the ribosome (+5 to +20), where zero coordinate corresponds to the A-site codon. To measure the contribution of local mRNA positions to the density of footprints correspondingly derived from a ribosome decoding a particular codon, we measured the relative entropy at each position using the Kulback-Leibler (K-L) divergence.

Figure 5.2b shows the relative entropy and normalized observed-to-expected RUST ratios *ro/re* (see Methods) for each individual codon for two of the ribosomal profiling datasets explored in this work. By analogy with metagene profiles we refer to the plots of *ro/re* RUST ratios as metafootprint profiles. The areas of reduced entropy (increased K-L divergence) are mostly contained within a window of 10 codons upstream and downstream of the A-site, approximately matching to the position of the actual ribosome footprint. In almost all cases three local K-L maxima are observed, one corresponds to the decoding center (Fig. 5.2b), the other two maxima roughly correspond to the 5’ and 3’ ends of ribosome footprints. The same procedure carried out on mRNA-seq libraries reveals decreased entropy in the same area with two maxima corresponding to the mRNA fragment ends (Fig. 5.2b). This suggests that the main contributing factors to footprint frequency at the corresponding location are the identity of the codons in the A- and/or P-sites and the sequence-specificity of the enzymes used during library construction. The metafootprint analysis for all human studies explored in this study are available in Figure 5.3. The degree of variation in the relative impact of these factors among different datasets is surprising. In some of the ribo-seq datasets, the density of footprints depends on the identity of the codon at the ends of footprint more than on the identity of the codon in the

A- or P-sites. This is suggestive of a high level of sequencing biases introduced during the cDNA library generation in some of the tested datasets.

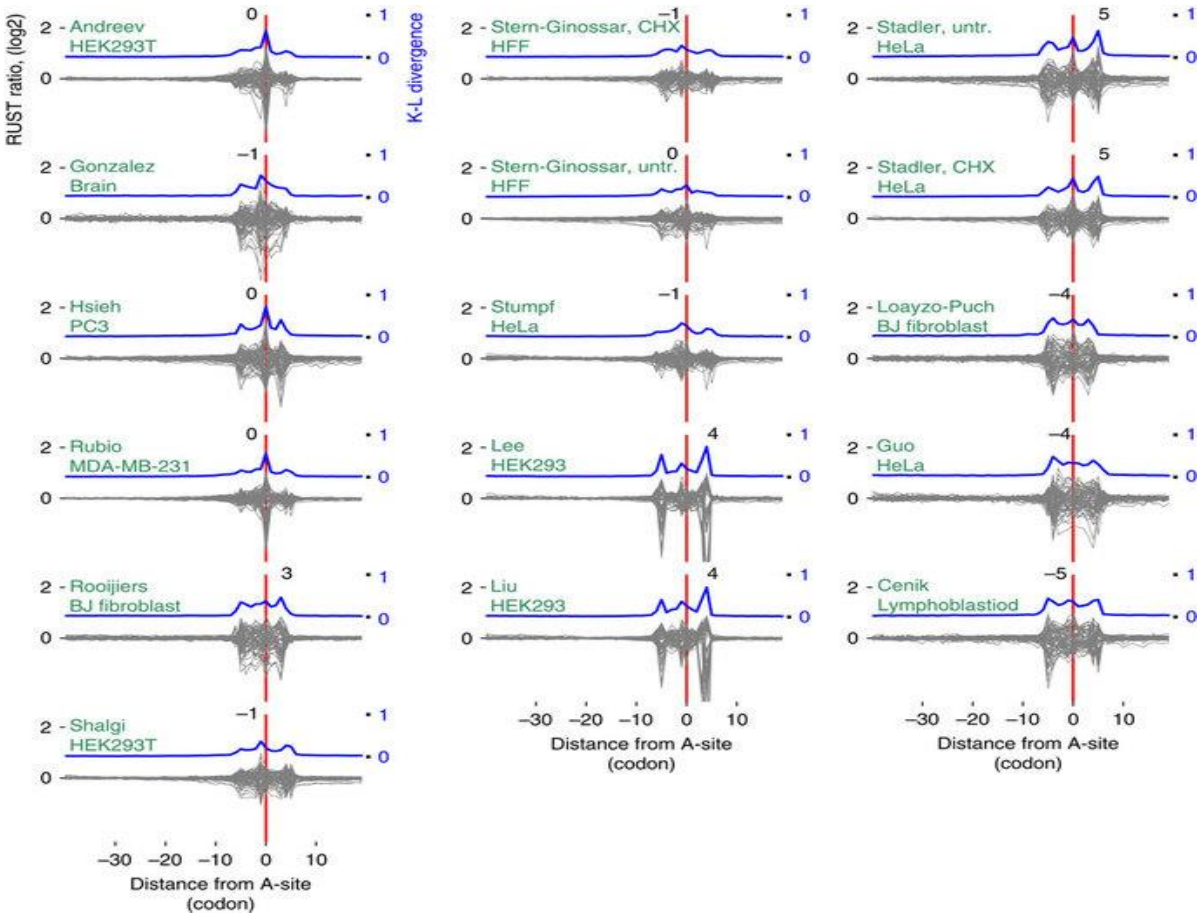


Figure 5.3. RUST metafootprint profiles of the 16 human ribo-seq datasets

Datasets are indexed by the name of the first author followed by drug treatment and source, see Table 1 for more details. The Kulback-Leibler (K-L) divergence is shown in blue, the coordinates of K-L maximum are indicated above the peak in each plot. Zero coordinate corresponds to the inferred position of the A-site and is marked with a red line, coordinates are in codons.

Figure 5.2c shows a heatmap produced as a result of pairwise comparison of observed-to-expected RUST ratios for the 61 codons when they are located in the A-site. Most apparent is the high reproducibility for most ribosomal profiling datasets produced in yeast under cycloheximide pretreatment (Fig 5.2c). The comparison of the protocol conditions (Table 1) points to the consistency in the protocols used in these studies. The variance across the datasets obtained from mammalian sources is more substantial as are the differences in the protocols (Table 1). We found that variance in RUST ratios of

nonsynonymous codons is greater than that of synonymous codons. In other words, the identity of decoded amino acid has a greater influence on read density than the identity of the specific codon. ANOVA revealed that this was statistically significant in 28 of the 30 ribo-seq samples. We carried out similar analysis for mRNA-seq controls for codons located at the same distance from the 5' end as the A-site codons in ribosome footprints. As expected, the degree of variation among all 61 codons was much smaller. However synonymous codons also exhibited statistically significant higher variation. This casts some doubts on biological relevance of this observation.

Some of the studies produced the data with a change to a single parameter: the samples were either pretreated or not with cycloheximide before lysis (Ingolia, Lareau, and Weissman 2011; Lareau et al. 2014; Stadler and Fire 2011; Stern-Ginossar et al. 2012). We found that 'Stadler' (Stadler and Fire 2011) datasets are similar for both types of treatments, while 'Lareau' (Lareau et al. 2014), 'Ingolia' (Ingolia, Lareau, and Weissman 2011) and 'Stern-Ginoassar' (Stern-Ginossar et al. 2012) are different (Fig. 5.2c). Analysis of RUST ratios for 'Lareau' and 'Ingolia' datasets under both conditions, clearly indicating that cycloheximide substantially alters the distribution of footprints on mRNA. This is consistent with the observation that cycloheximide blocks ribosomes in a specific conformation and this ribosome arrest has certain codon preferences (Tuller et al. 2011). A more focused and detailed analysis of this phenomenon (Hussmann et al. 2015) was published while this manuscript was in preparation.

Prior studies explored the effects of different antibiotic treatments in mammalian cells and in yeast (Lareau et al. 2014; Hussmann et al. 2015; Gerashchenko and Gladyshev 2014). The effect of buffer conditions on triplet periodicity was also explored to some extent (Stern-Ginossar et al. 2012; Brar et al. 2012) as well as conditions of nuclease treatments (Miettinen and Bjorklund 2015). We agree with a plea for standardization of ribosome protocols(Bartholomaus, Del Campo, and Ignatova 2016), however, as recently argued (Weinberg et al. 2016) it is clear that a more systematic study of protocol dependency of ribosome profiling data is needed for this.

Influence of RNA secondary structure and nascent peptide

To illustrate RUST capacity at analyzing mRNA features that may affect ribosome velocities we chose three, 'Andreev' (Andreev et al. 2015), 'Rubio' (Rubio et al. 2014), 'Pop' (Pop et al. 2014). These datasets exhibit a low level of K-L divergence at the ends of the footprints and a high K-L divergence at the decoding center, suggesting low sequencing bias at the end of footprints. However, while these datasets are relatively free of sequencing artefacts, the distribution of footprints could still be skewed for other reasons discussed in the previous section and caution needs to be applied in the interpretation of the results described below.

To estimate the effect of RNA secondary structure we calculated the RUST ratios for RNA sequences that can form secondary structures at a particular free energy threshold as calculated with RNAfold (Lorenz et al. 2011), see Methods. The distribution of RUST ratios for RNA secondary structures predicted within 80 nucleotides window with different free energies indicated that sequences predicted to contain stable structures are underrepresented (low RUST ratios) in windows that overlap with sequencing reads. This is observed for both ribo-seq and mRNA-seq reads and therefore is likely to be an artefact related to cDNA library generation and sequencing. This is not explained by a putative nucleotide bias. The distribution of individual nucleotides at the footprint location does not deviate significantly with the exception for the location of the decoding center.

The RUST ratios for individual amino acids and dipeptides do not reveal evidence of universal nascent peptide effect on ribosome velocity from the positions distant from the peptidyl transferase center. Although, such effects can be seen in individual datasets, e.g. strong influence of two Prolines in close proximity to the peptidyl transferase center in 'Andreev' dataset. Such nascent peptide interactions may also be facilitated by specific physicochemical properties of the peptide, as suggested earlier (Charneski and Hurst 2013). In this case the RUST ratio of individual amino acids may not provide an accurate representation of the nascent peptide effect on ribosome movement. Therefore, we measured RUST ratios for peptide fragments (10 residues) with particular physicochemical properties (number of positive charges, net charge and number of hydrophobic amino acids). Under high positive charge we observed deviations for the distributions of these

physicochemical properties in the datasets. However, it is not clear whether they are caused by their direct effects on decoding rates.

We also examined whether tripeptides could affect ribosome velocity differently than may be expected from their individual components. We detect such synergetic effects by comparing the RUST values for tripeptides to what would be expected from independent RUST values of corresponding residues using the standard score (Z-score). We carried out this analysis for adjacent amino acids only and thus explored synergetic effects for 464,000 tripeptides (20x20x20 residues x 58 positions). Approximately 0.2% (~1,000) of the tripeptides were found to have a standard score greater than 4 (S_{ijk}>4 or S_{ijk}< -4) in any individual dataset. These synergistic interactions were found to occur mostly near the decoding center or at the reads termini. They also had a relatively small influence with the majority of interactions having less than a 2 fold change between observed and expected values. In the 'Andreev' dataset the motifs that displayed positive synergetic effects (slower than expected) were overrepresented with Proline. This is a poor substrate for peptide bond formation and therefore a good *a priori* candidate for such synergistic effects. However, there was poor convergence between the results obtained from the thirty datasets, overall 7,854 examples of synergistic interaction were found with the majority (5,850) of candidates found only in a single dataset.

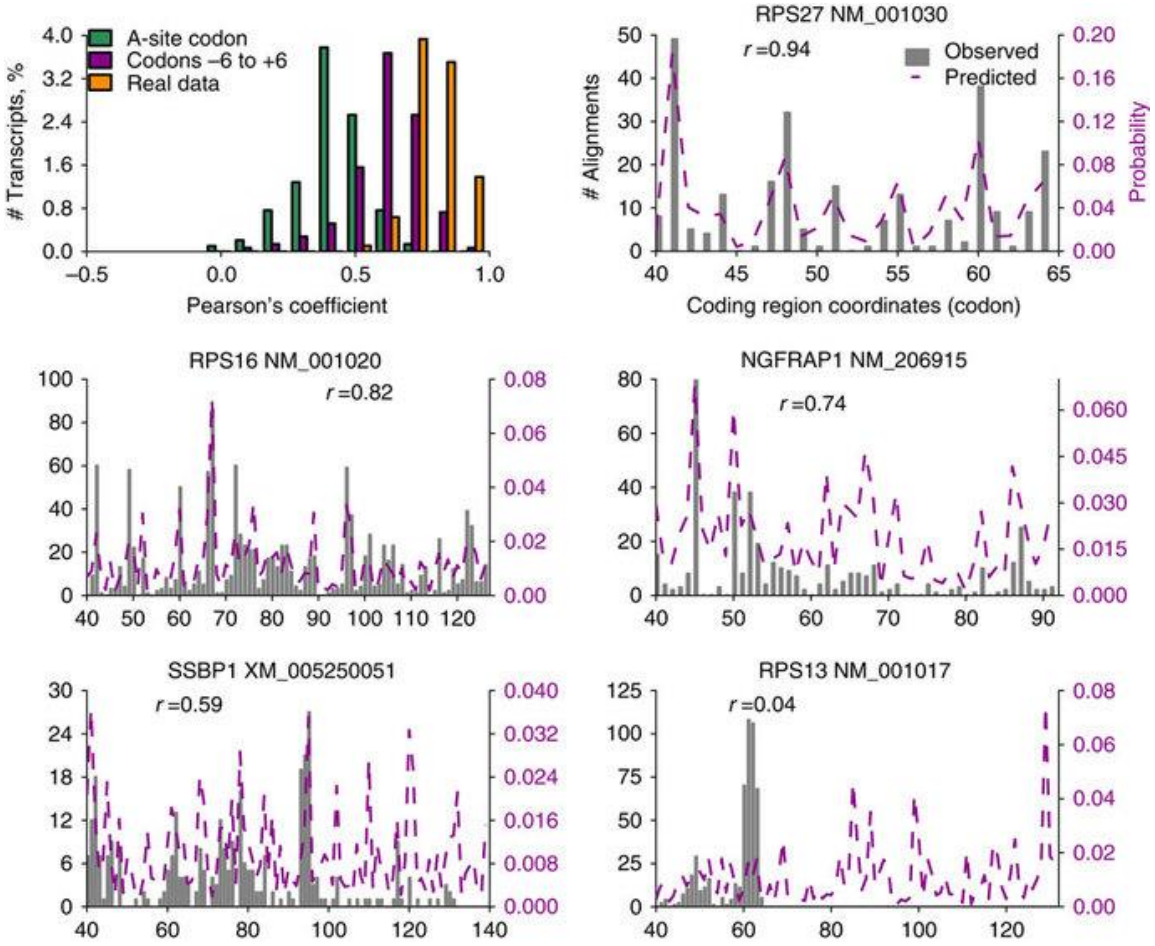


Figure 5.4. RUST accurately predict experimental footprint densities

Top left panel shows distributions of Pearson's correlation coefficients for experimental and predicted footprint densities (green and violet) for individual transcripts as well as correlation between two experimental ribo-seq datasets obtained under the same protocol (orange). Correlations were measured only for coding regions of highly expressed transcripts from 120 nucleotides downstream of the start codons to 60 nucleotides upstream of the stop codons. The other panels show experimental (solid grey) and predicted (based on RUST values for the codons -6 to +6 relative to A-site) ribosome densities (broken purple) for five transcripts, corresponding to the 1st, 10th, 100^{th,}, 500th and 714th strongest correlations, the Pearson's correlation coefficients are indicated. Results displayed are for 'Andreev' dataset.

Accurate prediction of experimental footprint densities

We proceeded to test whether we can reconstruct ribosome densities using RUST ratios obtained for codon positions relative to the decoding center. Figure 5.4 shows the comparison of experimental densities to predicted densities based on RUST ratios for the A-site codon or 12 codons comprising the ribosome footprint and the codons immediately

adjacent to them. Predictions made based only on the A-site RUST values correlate with the real profiles (Pearson's *r*=0.451, Spearman's r=0.503 for Andreev et al dataset[27]). The incorporation of RUST ratios for all codon sites in the footprint improves the predictive power even further, with an average Pearson's *r*=0.62. These values may improve further with increased sequencing depth. Note that this is not an example of overfitting of a model, as the RUST metafootprint profile is relatively unaffected if it is obtained from a subset of genes different from which are used to evaluate the profiles. We also compared the profiles to those obtained from another ribo-seq sample of the same study. This had an average Pearson's r of 0.78, the difference between the samples probably reflects the stochastic nature of RNA-seq. The ability to predict ribosome profiles was replicated for other datasets, with an average Pearson's correlation coefficient greater than 0.5 in 16 of 30th datasets. The accuracy of these predictions support our earlier findings of a limited influence of the nascent peptide, mRNA structure or synergistic effects on read density. Figure 5.4 shows comparison of predicted ribosome profiles with experimental profiles for five mRNAs with different degrees of correlation. It is clear from the example shown that the poor correlation is a result of technical artefacts in the data, rather than poor prediction.

Comparison of the datasets

We designed a web site http://lapti.ucc.ie/rust, that provides detailed characteristics (metafootprint analysis, RUST ratios, triplet periodicity, etc.) of each dataset explored in this study, an example for an individual dataset is shown in Figure 5.5. It also hosts executable scripts to implement the RUST analysis.

DISCUSSION

Here, we described a simple computational technique RUST for the characterization of ribosome profiling data based on a simple smoothing transformation of ribosome density profiles into a binary function. Using simulated data we show that this technique is robust in the presence of sporadic heterogeneous noise (modeled as extra high density and missing data) and outperforms previous methods. Using experimental data, we show that

the characteristics of ribosome profiling data extracted with RUST can explain much of the variation observed in experimental ribosome footprint densities.

We applied this technique to thirty publicly available ribo-seq datasets (obtained from yeast, mammalian cultured cells and tissues) and uncovered substantial variability among them in sequence features that determine footprint frequencies at individual locations. The most similar datasets are those obtained with cycloheximide pre-treatments of yeast cells and no or minimal variations in protocols used. For the datasets obtained in mammalian systems we found substantial variation that is likely to be related to the timing of cycloheximide treatments as well as conditions of buffers used for lysis and nuclease digestion. The position specificity of sequencing biases (they affect the boundaries of ribosome footprints) enabled us to determine their relative impact on composition of footprints in individual datasets.

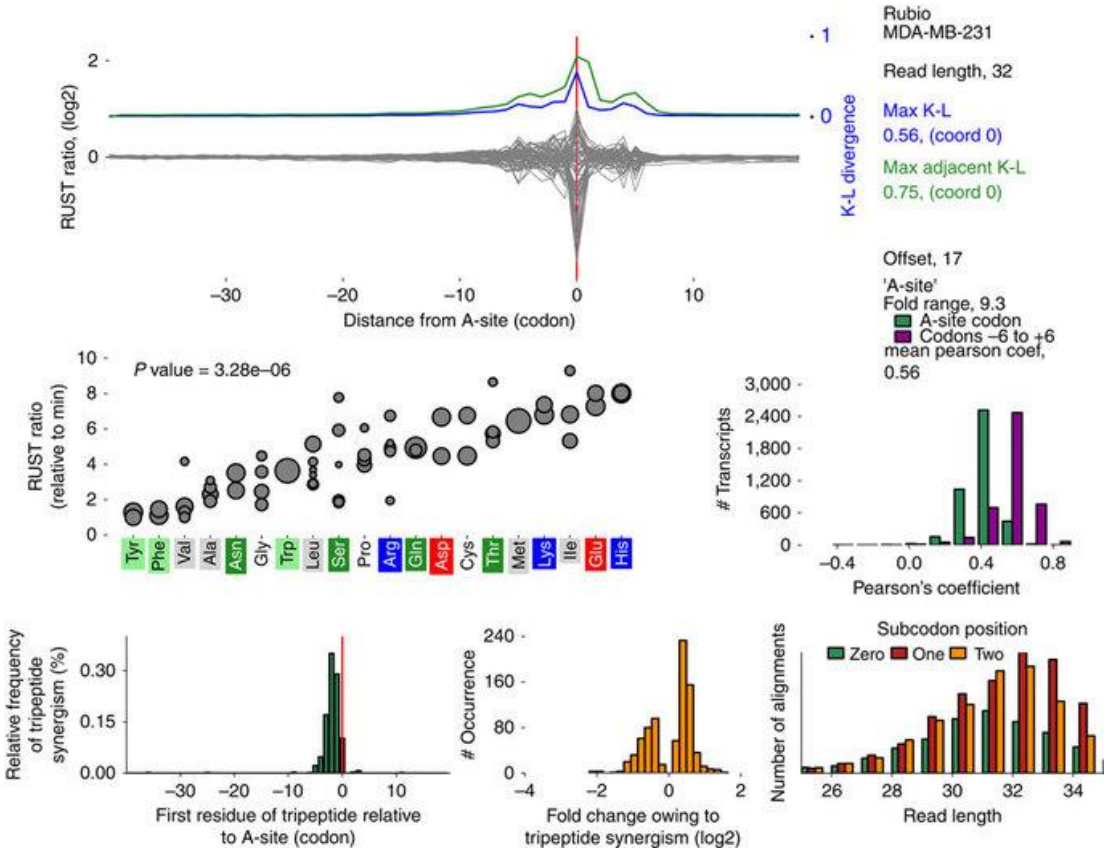


Figure 5.5. An example of information provided for each dataset with RUST Software

RUST metafootprint analysis for codons is shown at the top. Blue indicates Kulback-Leibler (K-L) divergence for individual codons and green for two adjacent codons. Zero coordinate corresponds to the inferred position of the A-site. Middle left plot shows RUST ratios (Y axis) for individual codons ordered by encoding amino acids at the axis X, the size of the

disc indicates the codon usage. Middle right plot illustrates the ability of RUST parameters to reconstruct experimental densities (green – based on identity of the A-site codons and violet is based on all codons within footprints). Bottom left plot shows locations of observed synergistic effects between amino acids that affect decoding rates, the bottom middle plot shows their strength and the bottom right plot illustrates the triplet periodicity signal for footprints of different lengths. This information is available for each dataset used in this study at http://lapti.ucc.ie/rust/.

Our simulations suggest that potential uncharacterized artefacts of the computational analysis of ribo-seq data are unlikely to be a major cause in the current difficulties for the determination of the true ribosome decoding rates. However, it appears that all current approaches including RUST overestimate the dwell time of quickly decoding determinants of elongation (codons in the case of simulations). A number of attempts were made to supersede CN approach. Surprisingly, in this study we find that for many applications, such as the analysis of the enrichment rate of individual codons, the simplest variant CN>0 provides surprisingly accurate results. In our simulation we found that it was only marginally worse than RUST irrespective of the relationship between codon usage and dwell times. For real data CN>0 provided results broadly similar to that obtained with RUST, except that the noise reductions achieved with RUST is counterbalanced with a lower signal. It is likely that the superiority of CN>0 normalization over CN>1 is due to larger volume of data used. While it seems reasonable to filter out lowly expressed genes prior to the analysis because their individual ribosome profiles are unrealistic representations of the real ribosome density, collectively these profiles produce a statistically reliable signal and their analysis is highly informative.

The RUST approach maximizes the chances that detected signal is real in two ways. On one hand it is based on gathering information from all transcriptome coordinates increasing the chance that the signal is not arising due to stochastic reasons. The benefit of this becomes greatest when examining the influence of relatively infrequent determinants, such as certain dipeptides, tripeptides. On the other hand, by reducing the impact of each individual site, RUST ensures that a signal is not a product of a rare outlier (whether due to technical or biological reasons). The smoothing achieved by RUST could also be applied to other high-throughput methods that are characterized by the presence of heterogeneous noise. In this work, for example, we were able to detect that sequencing reads that form

RNA secondary structures are underrepresented not only in ribosome profiling data, but also in mRNA-seq data. Thus RUST could have a broader impact if adopted.

The conversion of regular profile to a binary profile leads to an unavoidable loss of information. The approach is therefore "blind" to individual special cases where infrequent motifs may pause the ribosome for a long period. This, however, can be used to identify such special cases by looking for large discrepancies between the densities in the real data and in simulations based on parameters extracted with RUST. This application, however, is challenged by the presence of technical artefacts as illustrated in Figure 5.5.

We illustrated the applicability of RUST for measurement of mRNA features that impact decoding rates using datasets with lower sequencing bias. The results suggest that sites other than at the decoding center have a relatively minor influence on the decoding rate globally. This observation does not contradict the well characterized pauses modulated by nascent peptide signals and RNA secondary structures at specific locations of individual mRNAs. However, we also showed that in addition to identity of codons in the decoding center of the ribosome, sequences surrounding the ends of footprints are major determinants of footprint densities. The influence of these regions on read density greatly vary among different datasets, in some exceeding that of the sequences in the decoding center. We suggest that this feature could be used for quality assessment of ribosome profiling datasets for the presence of cDNA library construction biases. Cross-platform implementation of RUST is freely available at RiboGalaxy (http://ribogalaxy.ucc.ie).

METHODS

Ribo-seq datasets used in this study and their processing

The datasets (and SRA repository accession numbers) are summarized in Supplementary Table 1. For simplicity these datasets are indexed in the text using the first author name of the original article. The processing of the reads consisted of clipping the adapter sequence and removal of ribosomal RNA reads followed by the alignment of the mammalian reads to the RefSeq transcriptome (Pruitt et al. 2014) and the yeast reads to the *Saccharomyces cerevisiae* genome (sacCer3 assembly). The weakly updated human RefSeq catalogue was downloaded on the 13th Aug 2014 from the NCBI ftp website

ftp://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens/ and the mouse RefSeq catalogue was downloaded on the 18th March 2014 from ftp://ftp.ncbi.nlm.nih.gov/refseq/M_musculus/. The yeast genome (sacCer3 assembly) and annotation data were downloaded on 13th Aug 2014 from the UCSC genome browser (Karolchik et al. 2014) website, http://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/bigZips/sacCer3.2bit (genome), http://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/database/sgdGene.txt.gz (annotations).

Bowtie version 1.0.0 (Langmead et al. 2009) was used to carry out the alignments. The reads were aligned using Bowtie to the entire human or mouse catalogue with the following parameters (-a, -m 100 -norc). Except where otherwise indicated the reads that mapped unambiguously to a gene (but not necessarily to a single transcript) were brought forward for further analysis. For the yeast datasets reads were aligned to the yeast genome allowing only unambiguous alignments (-a, -m 1).

Ribo-seq simulation

The simulated alignment data were modelled using real human mRNA sequences obtained from the RefSeq database and with the average transcript read density similar to that of real ribosome profiling data. We simulated the data under the simplistic model where the local decoding rate depends exclusively on the identity of a decoded codon (A-site codon). The number of footprints at each codon position was determined by sampling from the following Poisson probability mass function:

$$p_{m,c,d} = \frac{\left(\frac{t_c D_m}{\sum_{C=(AAA\ldots TTT)} n_{c,m} t_c}\right)^d e^{\frac{-t_c D_m}{\sum_{C=(AAA\ldots TTT)} n_{c,m} t_c}}}{d!} \quad [1]$$

Where $p_{m,c,d}$ is the probability of finding d number of footprints at a specific location at mRNA m at a codon c from the set of 61 sense codons C. D_m is the total number of footprints aligning to mRNA m; $n_{c,m}$ is the number of codons c in the coding region of mRNA m; and t_c is the relative dwell time for the codon c. The dwell times t_c for the 61 sense codons were set to span either a ~10 or ~100 fold range with equal increments of 0.15 or 1.5, (the fastest codon was given a score of 1, the slowest was 10.15 or 92.5). To model the

noise arising from high density peaks the number of reads at a certain percentage of randomly selected coordinates (irrespective of where it originally contained a mapped read) was substituted with a 3x the value of the highest footprint density for the original simulated profile. The number of codons selected was calculated as a percentage (either 5% or 20%) of the number of codons with a mapped read. To model the absence of mapped reads because of discarding of ambiguous alignments reads were removed from 20% of codons with mapped reads. The selection of codons was carried out using a probability distribution, therefore for individual mRNA density profiles the number of altered codons may differ from 5% to 20%.

The normalisation approaches were used to estimate relative dwell times as described below. The normalised estimated/simulated values obtained for all 61 sense codons were used to produce the boxplots in Figure 5.1 and Supplementary Figure 2. The normalisation consisted of dividing each of 61 values by their mean to enable comparison between values from datasets and normalisation approaches. The coefficient of determination between the estimated and simulated values was also used as a measure of the accuracy of the approaches.

To explore how the data specific factors (e.g. coverage, sequencing biases) affect performance of different normalization approaches we carried out simulations using Hseih et al[31] (4 million mapped reads of which 530,051 reads passed selection criteria) and Rubio et al[36] (61 million mapped reads of which 6,470,387 reads passed selection criteria). The simulations on Figure 5.1 are based on Rubio et al. data, those in Supplementary Figure 2 were carried out on the Hseih dataset.

Determining offset to the A-site

An important factor for the analysis described in this work is the application of the correct offset for inferring the position of the A-site codon relative to the footprint 5' end. This is typically estimated with a metagene profile of either initiating or terminating ribosomes. This may not always allow for a precise estimation of the offset and it is possible that initiating or terminating ribosomes do not protect mRNAs in the same way as elongating ones because of conformational differences, e.g. when release factor eRF1 binds to the ribosome[53]. With the premise that the combined A-site and P-sites should have the

greatest influence on decoding rates we set out to estimate the offset using RUST codon metafootprint profiles with the largest K-L divergence at adjacent offsets. We carried out three RUST metafootprint profiles (using the same approach described below) at multiple offsets (usually 16, 17, 18 nucleotides). For these profiles we determined the combined K-L divergence from two adjacent codons. The codon pair in any of the profiles with the largest K-L (that was not at the ends of the reads) was assumed to correspond to the P and A-sites. It was necessary to take the combined K-L divergence from two adjacent sites as in some datasets the divergence of the P-site was greater than that of the A-site. For one of the datasets (with low sequencing bias) we confirmed that the maximal K-L divergence nucleotides corresponded to the A-site offset determined with initiating ribosomes. The offsets used for each dataset are listed in the Supplementary Table 1.

Normalization approaches

For this analysis the alignment data to the longest coding transcript of every expressed gene were used. Owing to possible atypical translation at the beginning or the end of coding regions, the analysis was carried out on coding regions with the A-site position within 120 nucleotides (40 codons) downstream of the annotated start codon and 60 nucleotides upstream of the annotated stop codon. With exception to one of the 'Lareau' datasets the analysis was carried using reads of the predominant length. An offset to the A-site was determined as described earlier. The exclusive selection of reads of one length was necessary to minimise the effect of variation in a distance between footprint ends and the A-site. In this analysis reads were used irrespective of the subcodon position to which they aligned. The exclusive selection of reads that align to a particular subcodon position may further improve the signal. Because of these criteria ~15% of total (non rRNA) mapped reads were used to produce metafootprint profiles (Supplementary Table 1). To check whether exclusion of ambiguously aligned reads had a large influence on the result we repeated the analysis with the unambiguous reads; the obtained results are nearly the same**.**

The first step of "RUST phase" is the conversion of ribosome density profile to a binary profile based on whether the number of alignments at each determinant (codon,

nucleotide, amino acid) exceeds the gene average. The RUST value at each location *l* is denoted as (ro_{cl}), *c* stands one of 61 codons (when codons are examined as determinants). For each sequence determinant the expected value re_{cl} is also obtained. re_{cl} is obtained by averaging local RUST values across a single coding region. For lowly expressed genes it is expected to be close to 0 and for highly expressed genes it is substantially higher. Normalisation over expected values is carried out to control for the non-random distribution of codons (or other determinants) across the genes with different expression levels. If all codons had the same dwell times, their unnormalized RUST values would be higher for codons that are more frequent in highly expressed genes. This analysis is carried out for all mRNA sequences in the translatome. In order to check for an enrichment of reads at a particular determinant the obtained RUST value is compared to the expected RUST value. To produce a metafootprint profile we used a sliding window approach. For the analysis of codons as a determinant of footprint density the window of 61 codons is moved with a step size of one codon. The center of the window is considered to be the A-site codon. The RUST values are calculated for each codon relative to the A-site and represented in the form of a metafootprint profile. The procedure used for other determinants such as nucleotides, amino acids, peptide properties and RNA secondary structures is conceptually the same.

CN normalisation consisted of an initial normalisation of the individual read density profiles by the average read density specific to individual coding regions. This followed by determination of average normalized values for each of 61 codons across the entire dataset. For the generation of metafootprint profiles average codon values were calculated for specific locations within the sliding window similarly to for RUST. For CN>0 all mRNA transcripts were used while for the CN>1 only coding regions with an average read density > 1 read/nucleotide were used.

We carried out “Ribosome residence time” RRT similar to that described by Gardin et al. (Gardin et al. 2014) The analyses was carried out independently on windows of 19 codons in length that satisfy the following requirements: (1) greater than 19 aligned reads, (2) less than 3 codons with no alignments and (3) if the codon at the position 10 occurred only once in the window. For each window the decimal fraction of reads aligning to each codon (relative to the total number of reads in the window) was determined. The average

obtained for each codon at all 19 codons was then used to produce the metafootprint profile.

As the other normalisation procedures do not use mRNA-seq data, we could not carry out an equitable comparison with the "corrected ribo coverage" analysis (Pop et al. 2014). Therefore, instead of using the footprint density normalized by RNA-seq density, we used only footprint densities. We refer to this approach as LMN for logarithmic mean normalisation. Similar to the original approach only coordinates with mapped reads are used and footprint densities are first normalised by the algebraic average read density. The algebraic average of their $\log_2$ values are then calculated across all coding regions (first term in equation [2] below). Further the average of all 61 codons is calculated (second term in equation [2] below) and subtracted from the codon specific value. The procedure can be summarized in the following equation

$$LMN_c = \frac{\sum_L \log_2 d_{cl}}{N_c} - \sum_{C=(AAA\ldots TTT)} \frac{\sum_L \log_2 d_{cl}}{61 N_c} \quad [2]$$

Where LMN_c is LMN value for the codon c (from a set of 61), N_c is the total number of *c* codon occurrences with non 0 footprint densities and d_{cl} is footprint density for the codon *c* at the location *l* normalized by the average footprint density for the corresponding mRNA. We carried out the analysis on coding regions with an average read density > 1 read/nucleotide.

The "aov" function in R was used to calculate the p-values with ANOVA for assessing statistical significance of the difference between the variation among synonymous codons and variation among all codons at the A-site.

Kullback-Leibler divergence

The Kulback-Leibler divergence was used to calculate relative entropy in the RUST metafootprint profiles and was calculated as the following

$$D_l = \sum_c \frac{ro_{cl}}{\sum_{C=(AAA\ldots TTT)} ro_{cl}} log_2 \left(\frac{ro_{cl} / \sum_{C=(AAA\ldots TTT)} ro_{cl}}{re_c / \sum_{C=(AAA\ldots TTT)} re_c} \right) \quad [3]$$

Where D_l is the K-L at location *l*, ro_{cl} is the observed RUST value for codon *c* at location *l* and re_c is the expected RUST value for codon *c*. The higher the K-L, the less uniform the distribution of RUST values is in the corresponding position. Thus, K-L indicates how much the corresponding position contributes to the abundance of footprints.

RNA secondary structure analysis

The computational prediction of RNA secondary structure free energy was performed using RNAfold in the ViennaRNA package (Lorenz et al. 2011). Using a sliding window of 80 nucleotides with a step size of 10 nucleotides the minimal free energy for potential RNA secondary structures was estimated across each transcript. For human data the threshold free energy for the most stable RNA secondary structures was found to be-40.1 kcal/mol for the top 1st percentile, -32.8 kcal/mol for the 5th and -29.0 kcal/mol the 10th percentile.

Amino acid physicochemical properties

In this study Histidine, Lysine, Arginine, were considered to be positively charged. Aspartic acid, Glutamic acid as negatively charged. Alanine, Valine, Isoleucine, Leucine, Methionine, Phenylalanine, Tyrosine, Tryptophan were considered to be hydrophobic.

Standard score to identify synergistic interactions

To identify synergistic interactions, we compared the difference in fold changes between observed and expected metafootprint profiles. The fold change at each position was normalised to the background fold change as follows

$$S_{ijk} = \frac{ro_{ijk}/re_{ijk} - ro_i ro_j\, ro_k/re_i\, re_j re_k}{std} \quad [4]$$

Where $S_{ijk}$are synergy indexes for tripeptide *ijk* and *ro/re* are corresponding RUST ratios. *std* is the standard deviation of the differences observed at regions from -40 to + 18 relative to the A-site.

The comparison of predicted and real footprint densities

When information from all footprint codons, plus two surrounding ones (-6 to +6 relative to the P-site/A-site boundary) was used to model ribo-seq densities, the predicted profile can be represented as a discrete probability density function

$$p_k = \frac{\prod_{i=1}^{N} \frac{ro_{ik}}{r_{eik}}}{\sum_{j=1}^{M}\left(\prod_{i=1}^{N} \frac{ro_{ij}}{ro_{ij}}\right)} \qquad [5]$$

Where p_k is the probability of finding a footprint at the position k of the mRNA coding region consisting of M codons. ro_{ik}/ro_{ik} is the RUST ratio for the codon at the site i (relative to the codon k) from the total of N sites used. For instance, if RUST ratios of AAA in the P-site and A-site are 0.339 and 1.646 respectively, the expected RUST ratio for di-codon AAA-AAA is 0.557 (0.339*1.646). Instead of di-codons in our simulation the RUST ratio is obtained with 12 codons, this corresponds to the numerator in equation [5]. The denominator corresponds to the sum of RUST ratios across the coding region and remains constant for all codons of each transcript.

The comparison between the expected and experimental profiles was carried out on transcripts with a density greater than 1 read/nucleotide. (Transcripts with a lower density were not used as they have insufficient data to correlate with the predicted profile).

The python package matplotlib (Hunter 2007) was used to produce the figures.

Code availability

Supplementary Software is a compressed archive of user friendly executable scripts to run RUST (version 1.2). Its source code is accessible and it includes several implementations of RUST that search for enrichment of codons, amino acids, dipeptides, tripeptides and nucleotides. “rust_synergy” searches for synergistic effects between adjacent amino acids. “rust_predict_profiles” returns a csv file that records the Pearson’s and Spearman’s correlation coefficient between the observed and predicted footprint densities for individual transcripts. “rust_plot_transcript” plots the observed and predicted footprint densities. This (and updated versions in the future) are also available at http://lapti.ucc.ie/rust/ . In addition, Supplementary Software includes “RUST_script.py” a 2nd shorter, non-executable version of the RUST implementation on codon enrichment. This script is a pseudocode intended as an explanatory aid for understanding RUST algorithm. .

RUST is also available via the RUST package at RiboGalaxy (Michel et al. 2016) (http://ribogalaxy.ucc.ie).

Data availability

The NCBI SRA accessions numbers for the datasets processed in this study is listed in Table 1

Table 1. Ribosome profiling protocol conditions for the studies described in this work

Description	PMID	SRA accession	Biological source	Lysis buffer				Drugs	Lysis method	RNase	Separation
				CHX (mins)	Mg^{2+}mM	M^{+} mM					
			Human								
Andreev	25621764	SRR1173909 SRR1173910	HEK293T	no	1.5	250	NaCl	CHX	Detergent	I	GR
Cenick	26297486	SRR1803149	LCL	no	5	150	NaCl	CHX	Freeze	A,T4	CS
Gonzalez	25122893	SRR1562539	Brain	no	15	250	NaCl	CHX	Dounce homogenizer, freeze	I	GR
Guo	20703300	SRR057512	HeLa	8	5	100	KCl	CHX	Detergent	I	GR
Hseih	22367541	SRR403883	PC3	?	?	?		?	?	?	?
Lee	22927429	SRR618771	HEK293	30	5	100	KCl	CHX	Detergent	I	GR
Liu	23290916	SRR619083	HEK293	1	5	100	KCl	CHX	Detergent	I	GR
Loayzo-Puch	23594524	SRR627620	BJ fibroblast	8	10	100	KCl	CHX	Detergent	I	GR
Rooijiers	24301020	SRR935448	BJ fibroblast	5	10	100	KCl	CHX	Detergent	I	GR
Rubio	25273840	SRR1573934	MDA-MB-231	no	15	220	NaCl	CHX	Detergent	I	CS
Shalgi	23290915	SRR648667	HEK293T	5	5	100	KCl	No	Freeze	I	CS
Stadler	22045228	SRR407637	HeLa	no	1.5	140	KCl	CH	Freeze	I	GR

CHX.							X			
Stadler untr.	22045228	SRR407643	HeLa	no	1.5	140 KCl	No	Freeze	I	GR
Stern-Ginossar, CHX	23180859	SRR609197	human foreskin fibroblasts	1	15	250 NaCl	CHX	Detergent	I	CS
Stern-Ginossar, untr	23180859	SRR592961	human foreskin fibroblasts	no	15	250 NaCl	No	Detergent	I	CS
Stumpf	24120665	SRR970561	Hela	2	5	?	CHX	Detergent	I	CS
			Mouse							
Howard	23696641	SRR826795	Liver	no	10	300 KCl	CHX	Homogenizer	I	CS
Ingolia, CHX	22056041	SRR315601	Embryonic stem cell	1	15	250 NaCl	CHX	Detergent	I	CS
Ingolia, untr.	22056041	SRR315616	Embryonic stem cell	no	15	250 NaCl	No	Detergent	I	CS
Reid	25215492	SRR1066893	Embryonic fibroblast	no	15	100 KoAc	CHX	Detergent (digitonine)	MN	CS
Shalgi	23290915	SRR648667	3T3	5	5	100 KCl	No	Freeze	I	CS
Thoreen	22552098	SRR449467	Embryonic fibroblast	5	7.5	300 KCl	CHX	Detergent	I	GR
			Yeast							
Artieri	25294246	SRR104909		2	1.5	140 KCl	CH	Freeze	I	GR

		3				X			
Brar	22194413	SRR387871	2	1.5	140 KCl	CHX	Freeze	I	GR
Gardin	25347064	SRR1506632	no	ARTseq	ARTseq	CHX	Freeze	I	SC
Ingolia	19213877	SRR014374 SRR014375 SRR014376	2	1.5	140 KCl	CHX	Freeze	I	GR
Lareau, CHX	24842990	SRR1363415 SRR1363416	yes	1.5	140 KCl	CHX	Freeze	I	GR
Lareau, untr.	24842990	SRR1363412 SRR1363413 SRR1363414	no	1.5	140 KCl	No	Freeze	I	GR
McManus	24318730	SRR948555	5	1.5	140 KCl	CHX	Freeze	I	CS
Pop	25538139	SRR1688547	2	1.5	140 KCl	CHX	Freeze	I	CS

CHX – cycloheximide, MN – micrococcal nuclease, GR – sucrose gradient, CS – sucrose cushion, SC – spin-column chromotography. Additional information is provided in Supplementary Table 1. ? – not known.

Bibliography

Adam, Alexander C, Carsten Bornhovd, Holger Prokisch, Walter Neupert, and Kai Hell. 2006. "The Nfs1 Interacting Protein Isd11 Has an Essential Role in Fe/S Cluster Biogenesis in Mitochondria." Journal Article, Research Support, Non-U.S. Gov't. *The EMBO Journal* 25 (1). England: 174–83. doi:10.1038/sj.emboj.7600905.

Aird, Daniel, Michael G Ross, Wei-Sheng Chen, Maxwell Danielsson, Timothy Fennell, Carsten Russ, David B Jaffe, Chad Nusbaum, and Andreas Gnirke. 2011. "Analyzing and Minimizing PCR Amplification Bias in Illumina Sequencing Libraries." Journal Article, Research Support, N.I.H., Extramural. *Genome Biology* 12 (2). England: R18. doi:10.1186/gb-2011-12-2-r18.

Allam, Heba, and Naushad Ali. 2010. "Initiation Factor eIF2-Independent Mode of c-Src mRNA Translation Occurs via an Internal Ribosome Entry Site." Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't. *The Journal of Biological Chemistry* 285 (8). United States: 5713–25. doi:10.1074/jbc.M109.029462.

Andreev, D E, S E Dmitriev, I M Terenin, and I N Shatsky. 2013. "Cap-Independent Translation Initiation of Apaf-1 mRNA Based on a Scanning Mechanism Is Determined by Some Features of the Secondary Structure of Its 5' Untranslated Region." Journal Article, Research Support, Non-U.S. Gov't. *Biochemistry. Biokhimiia* 78 (2). United States: 157–65. doi:10.1134/S0006297913020041.

Andreev, D E, P B O'Connor, C Fahey, E M Kenny, I M Terenin, S E Dmitriev, P Cormican, D W Morris, I N Shatsky, and P V Baranov. 2015. "Translation of 5' Leaders Is Pervasive in Genes Resistant to eIF2 Repression." Journal Article. *Elife* 4: e03971. doi:10.7554/eLife.03971.

Andreev, Dmitri E, Sergey E Dmitriev, Ilya M Terenin, Vladimir S Prassolov, William C Merrick, and Ivan N Shatsky. 2009. "Differential Contribution of the m7G-Cap to the 5' End-Dependent Translation Initiation of Mammalian mRNAs." Journal Article, Research Support, Non-U.S. Gov't. *Nucleic Acids Research* 37 (18). England: 6135–47. doi:10.1093/nar/gkp665.

Andreev, Dmitry E, Patrick B F O'Connor, Alexander V Zhdanov, Ruslan I Dmitriev, Ivan N Shatsky, Dmitri B Papkovsky, and Pavel V Baranov. 2015. "Oxygen and Glucose Deprivation Induces Widespread Alterations in mRNA Translation within 20 Minutes."

Journal Article, Research Support, Non-U.S. Gov't. *Genome Biology* 16 (May). England: 90. doi:10.1186/s13059-015-0651-z.

Archer, S K, N E Shirokikh, T H Beilharz, and T Preiss. 2016. "Dynamics of Ribosome Scanning and Recycling Revealed by Translation Complex Profiling." Journal Article. *Nature* 535 (7613): 570–74. doi:10.1038/nature18647.

Artieri, Carlo G, and Hunter B Fraser. 2014. "Accounting for Biases in Riboprofiling Data Indicates a Major Role for Proline in Stalling Translation." Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't. *Genome Research* 24 (12). United States: 2011–21. doi:10.1101/gr.175893.114.

Aspden, Julie L, Ying Chen Eyre-Walker, Rose J Phillips, Unum Amin, Muhammad Ali S Mumtaz, Michele Brocard, and Juan-Pablo Couso. 2014. "Extensive Translation of Small Open Reading Frames Revealed by Poly-Ribo-Seq." Journal Article, Research Support, Non-U.S. Gov't. *eLife* 3 (August). England: e03528. doi:10.7554/eLife.03528.

Audibert, A, D Weil, and F Dautry. 2002. "In Vivo Kinetics of mRNA Splicing and Transport in Mammalian Cells." Journal Article, Research Support, Non-U.S. Gov't. *Molecular and Cellular Biology* 22 (19). United States: 6706–18.

Autio, Kaija J, Alexander J Kastaniotis, Helmut Pospiech, Ilkka J Miinalainen, Melissa S Schonauer, Carol L Dieckmann, and J Kalervo Hiltunen. 2008. "An Ancient Genetic Link between Vertebrate Mitochondrial Fatty Acid Synthesis and RNA Processing." *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* 22 (2): 569–78. doi:10.1096/fj.07-8986.

Baird, T D, and R C Wek. 2012. "Eukaryotic Initiation Factor 2 Phosphorylation and Translational Control in Metabolism." Journal Article. *Adv Nutr* 3 (3): 307–21. doi:10.3945/an.112.002113.

Baranov, Pavel V, John F Atkins, and Martina M Yordanova. 2015. "Augmented Genetic Decoding: Global, Local and Temporal Alterations of Decoding Processes and Codon Meaning." Journal Article, Research Support, Non-U.S. Gov't, Review. *Nature Reviews. Genetics* 16 (9). England: 517–29. doi:10.1038/nrg3963.

Baranov, Pavel V, Raymond F Gesteland, and John F Atkins. 2002. "Release Factor 2 Frameshifting Sites in Different Bacteria." *EMBO Reports* 3 (4). European Molecular Biology Organization: 373–77. doi:10.1093/embo-reports/kvf065.

Bartholomaus, A, C Del Campo, and Z Ignatova. 2016. "Mapping the Non-Standardized

Biases of Ribosome Profiling." Journal Article. *Biol Chem* 397 (1): 23–35. doi:10.1515/hsz-2015-0197.

Battle, A, Z Khan, S H Wang, A Mitrano, M J Ford, J K Pritchard, and Y Gilad. 2015. "Genomic Variation. Impact of Regulatory Variation from RNA to Protein." Journal Article. *Science* 347 (6222): 664–67. doi:10.1126/science.1260793.

Baudin-Baillieu, Agnes, Rachel Legendre, Claire Kuchly, Isabelle Hatin, Stephane Demais, Claire Mestdagh, Daniel Gautheret, and Olivier Namy. 2014. "Genome-Wide Translational Changes Induced by the Prion [PSI+]." Journal Article, Research Support, Non-U.S. Gov't. *Cell Reports* 8 (2). United States: 439–48. doi:10.1016/j.celrep.2014.06.036.

Bazzini, A A, M T Lee, and A J Giraldez. 2012. "Ribosome Profiling Shows That miR-430 Reduces Translation before Causing mRNA Decay in Zebrafish." Journal Article. *Science* 336 (6078): 233–37. doi:10.1126/science.1215704.

Borodovsky, Mark, and Alex Lomsadze. 2011. "Gene Identification in Prokaryotic Genomes, Phages, Metagenomes, and EST Sequences with GeneMarkS Suite." In *Current Protocols in Bioinformatics*, Chapter 4:Unit 4.5.1-17. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/0471250953.bi0405s35.

Brar, G A, and J S Weissman. 2015. "Ribosome Profiling Reveals the What, When, Where and How of Protein Synthesis." Journal Article. *Nat Rev Mol Cell Biol* 16 (11): 651–64. doi:10.1038/nrm4069.

Brar, G A, M Yassour, N Friedman, A Regev, N T Ingolia, and J S Weissman. 2012. "High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling." Journal Article. *Science* 335 (6068): 552–57. doi:10.1126/science.1215110.

Browne, Gareth J, Stephen G Finn, and Christopher G Proud. 2004. "Stimulation of the AMP-Activated Protein Kinase Leads to Activation of Eukaryotic Elongation Factor 2 Kinase and to Its Phosphorylation at a Novel Site, Serine 398." Journal Article, Research Support, Non-U.S. Gov't. *The Journal of Biological Chemistry* 279 (13). United States: 12220–31. doi:10.1074/jbc.M309773200.

Brush, Matthew H, Douglas C Weiser, and Shirish Shenolikar. 2003. "Growth Arrest and DNA Damage-Inducible Protein GADD34 Targets Protein Phosphatase 1 Alpha to the Endoplasmic Reticulum and Promotes Dephosphorylation of the Alpha Subunit of Eukaryotic Translation Initiation Factor 2." Journal Article, Research Support, U.S.

Gov't, Non-P.H.S., Research Support, U.S. Gov't, P.H.S. *Molecular and Cellular Biology* 23 (4). United States: 1292–1303.

Calviello, L, N Mukherjee, E Wyler, H Zauber, A Hirsekorn, M Selbach, M Landthaler, B Obermayer, and U Ohler. 2016. "Detecting Actively Translated Open Reading Frames in Ribosome Profiling Data." Journal Article. *Nat Methods* 13 (2): 165–70. doi:10.1038/nmeth.3688.

Carlberg, U, A Nilsson, and O Nygård. 1990. "Functional Properties of Phosphorylated Elongation Factor 2." *European Journal of Biochemistry* 191 (3): 639–45. http://www.ncbi.nlm.nih.gov/pubmed/2390990.

Caro, F, V Ahyong, M Betegon, and J L DeRisi. 2014. "Genome-Wide Regulatory Dynamics of Translation in the Plasmodium Falciparum Asexual Blood Stages." Journal Article. *Elife* 3. doi:10.7554/eLife.04106.

Castles, J J, and M F Singer. 1969. "Degradation of Polyuridylic Acid by Ribonuclease II: Protection by Ribosomes." Journal Article. *J Mol Biol* 40 (1): 1–17. http://www.ncbi.nlm.nih.gov/pubmed/4903359.

Cencic, R, M Carrier, G Galicia-Vazquez, M E Bordeleau, R Sukarieh, A Bourdeau, B Brem, et al. 2009. "Antitumor Activity and Mechanism of Action of the Cyclopenta[b]benzofuran, Silvestrol." Journal Article. *PLoS One* 4 (4): e5223. doi:10.1371/journal.pone.0005223.

Cenik, C, E S Cenik, G W Byeon, F Grubert, S I Candille, D Spacek, B Alsallakh, et al. 2015. "Integrative Analysis of RNA, Translation, and Protein Levels Reveals Distinct Regulatory Variation across Humans." Journal Article. *Genome Res* 25 (11): 1610–21. doi:10.1101/gr.193342.115.

Charneski, Catherine A, and Laurence D Hurst. 2013. "Positively Charged Residues Are the Major Determinants of Ribosomal Velocity." Journal Article, Research Support, Non-U.S. Gov't. *PLoS Biology* 11 (3). United States: e1001508. doi:10.1371/journal.pbio.1001508.

Chartron, J W, K C Hunt, and J Frydman. 2016. "Cotranslational Signal-Independent SRP Preloading during Membrane Targeting." Journal Article. *Nature* 536 (7615): 224–28. doi:10.1038/nature19309.

Chen, Yi-Jiun, Bertrand Chin-Ming Tan, Ya-Yun Cheng, Jin-Shin Chen, and Sheng-Chung Lee. 2010. "Differential Regulation of CHOP Translation by Phosphorylated eIF4E under

Stress Conditions." Journal Article, Research Support, Non-U.S. Gov't. *Nucleic Acids Research* 38 (3). England: 764–77. doi:10.1093/nar/gkp1034.

Chew, G L, A Pauli, and A F Schier. 2016. "Conservation of uORF Repressiveness and Sequence Features in Mouse, Human and Zebrafish." Journal Article. *Nat Commun* 7: 11663. doi:10.1038/ncomms11663.

Choi, J, J Chen, S L Schreiber, and J Clardy. 1996. "Structure of the FKBP12-Rapamycin Complex Interacting with the Binding Domain of Human FRAP." Journal Article. *Science* 273 (5272): 239–42. http://www.ncbi.nlm.nih.gov/pubmed/8662507.

Chowdhury, Rasheduzzaman, Rok Sekirnik, Nigel C. Brissett, Tobias Krojer, Chia-hua Ho, Stanley S. Ng, Ian J. Clifton, et al. 2014. "Ribosomal Oxygenases Are Structurally Conserved from Prokaryotes to Humans." *Nature*, May. doi:10.1038/nature13263.

Chung, B Y, T J Hardcastle, J D Jones, N Irigoyen, A E Firth, D C Baulcombe, and I Brierley. 2015. "The Use of Duplex-Specific Nuclease in Ribosome Profiling and a User-Friendly Software Package for Ribo-Seq Data Analysis." Journal Article. *RNA* 21 (10): 1731–45. doi:10.1261/rna.052548.115.

Crappe, J, E Ndah, A Koch, S Steyaert, D Gawron, S De Keulenaer, E De Meester, et al. 2015. "PROTEOFORMER: Deep Proteome Coverage through Ribosome Profiling and MS Integration." Journal Article. *Nucleic Acids Res* 43 (5): e29. doi:10.1093/nar/gku1283.

Crooks, Gavin E., Gary Hon, John-Marc Chandonia, and Steven E Brenner. 2004. "WebLogo: A Sequence Logo Generator." Journal Article, Research Support, Non-U.S. Gov't, Research Support, U.S. Gov't, Non-P.H.S., Research Support, U.S. Gov't, P.H.S. *Genome Research* 14 (6). United States: 1188–90. doi:10.1101/gr.849004.

Dana, Alexandra, and Tamir Tuller. 2012. "Determinants of Translation Elongation Speed and Ribosomal Profiling Biases in Mouse Embryonic Stem Cells." Journal Article, Research Support, Non-U.S. Gov't. *PLoS Computational Biology* 8 (11). United States: e1002755. doi:10.1371/journal.pcbi.1002755.

———. 2014a. "Properties and Determinants of Codon Decoding Time Distributions." Journal Article, Research Support, Non-U.S. Gov't. *BMC Genomics* 15 Suppl 6. England: S13. doi:10.1186/1471-2164-15-S6-S13.

———. 2014b. "The Effect of tRNA Levels on Decoding Times of mRNA Codons." Journal Article, Research Support, Non-U.S. Gov't. *Nucleic Acids Research* 42 (14). England: 9171–81. doi:10.1093/nar/gku646.

Datta, B, R Datta, S Mukherjee, and Z Zhang. 1999. “Increased Phosphorylation of Eukaryotic Initiation Factor 2alpha at the G2/M Boundary in Human Osteosarcoma Cells Correlates with Deglycosylation of p67 and a Decreased Rate of Protein Synthesis.” Journal Article, Research Support, U.S. Gov’t, P.H.S. *Experimental Cell Research* 250 (1). United States: 223–30. doi:10.1006/excr.1999.4508.

Devaraj, Aishwarya, and Kurt Fredrick. 2010. “Short Spacing between the Shine-Dalgarno Sequence and P Codon Destabilizes Codon-Anticodon Pairing in the P Site to Promote +1 Programmed Frameshifting.” *Molecular Microbiology* 78 (6): 1500–1509. doi:10.1111/j.1365-2958.2010.07421.x.

Dmitriev, Sergey E, Dmitri E Andreev, Ilya M Terenin, Ivan A Olovnikov, Vladimir S Prassolov, William C Merrick, and Ivan N Shatsky. 2007. “Efficient Translation Initiation Directed by the 900-Nucleotide-Long and GC-Rich 5’ Untranslated Region of the Human Retrotransposon LINE-1 mRNA Is Strictly Cap Dependent rather than Internal Ribosome Entry Site Mediated.” Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov’t. *Molecular and Cellular Biology* 27 (13). United States: 4685–97. doi:10.1128/MCB.02138-06.

Dmitriev, Sergey E, Ilya M Terenin, Dmitri E Andreev, Pavel A Ivanov, Jacov E Dunaevsky, William C Merrick, and Ivan N Shatsky. 2010. “GTP-Independent tRNA Delivery to the Ribosomal P-Site by a Novel Eukaryotic Translation Factor.” Journal Article, Research Support, Non-U.S. Gov’t. *The Journal of Biological Chemistry* 285 (35). United States: 26779–87. doi:10.1074/jbc.M110.119693.

Dunn, J G, C K Foo, N G Belletier, E R Gavis, and J S Weissman. 2013. “Ribosome Profiling Reveals Pervasive and Regulated Stop Codon Readthrough in Drosophila Melanogaster.” Journal Article. *Elife* 2: e01179. doi:10.7554/eLife.01179.

Elfakess, R, H Sinvani, O Haimov, Y Svitkin, N Sonenberg, and R Dikstein. 2011. “Unique Translation Initiation of mRNAs-Containing TISU Element.” Journal Article. *Nucleic Acids Res* 39 (17): 7598–7609. doi:10.1093/nar/gkr484.

Elkon, Ran, Eitan Zlotorynski, Karen I Zeller, and Reuven Agami. 2010. “Major Role for mRNA Stability in Shaping the Kinetics of Gene Induction.” Journal Article, Research Support, Non-U.S. Gov’t. *BMC Genomics* 11 (April). England: 259. doi:10.1186/1471-2164-11-259.

Eswarappa, Sandeepa M, Alka A Potdar, William J Koch, Yi Fan, Kommireddy Vasu, Daniel

Lindner, Belinda Willard, Linda M Graham, Paul E DiCorleto, and Paul L Fox. 2014. "Programmed Translational Readthrough Generates Antiangiogenic VEGF-Ax." Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't. *Cell* 157 (7). United States: 1605–18. doi:10.1016/j.cell.2014.04.033.

Federico, Antonio, Elena Cardaioli, Paola Da Pozzo, Patrizia Formichi, Gian Nicola Gallus, and Elena Radi. 2012. "Mitochondria, Oxidative Stress and Neurodegeneration." *Journal of the Neurological Sciences* 322 (1): 254–62. doi:10.1016/j.jns.2012.05.030.

Fekete, Christie A, Sarah F Mitchell, Vera A Cherkasova, Drew Applefield, Mikkel A Algire, David Maag, Adesh K Saini, Jon R Lorsch, and Alan G Hinnebusch. 2007. "N- and C-Terminal Residues of eIF1A Have Opposing Effects on the Fidelity of Start Codon Selection." *The EMBO Journal* 26 (6): 1602–14. doi:10.1038/sj.emboj.7601613.

Femino, A M, F S Fay, K Fogarty, and R H Singer. 1998. "Visualization of Single RNA Transcripts in Situ." Journal Article, Research Support, U.S. Gov't, P.H.S. *Science (New York, N.Y.)* 280 (5363). United States: 585–90.

Feng, Tianshu, Atsushi Yamamoto, Sarah E Wilkins, Elizaveta Sokolova, Luke A Yates, Martin Münzel, Pooja Singh, et al. 2014. "Optimal Translational Termination Requires C4 Lysyl Hydroxylation of eRF1." Journal Article, Research Support, Non-U.S. Gov't. *Molecular Cell* 53 (4). United States: 645–54. doi:10.1016/j.molcel.2013.12.028.

Fields, A P, E H Rodriguez, M Jovanovic, N Stern-Ginossar, B J Haas, P Mertins, R Raychowdhury, et al. 2015. "A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian Translation." Journal Article. *Mol Cell* 60 (5): 816–27. doi:10.1016/j.molcel.2015.11.013.

Fonseca, B D, E M Smith, N Yelle, T Alain, M Bushell, and A Pause. 2014. "The Ever-Evolving Role of mTOR in Translation." Journal Article. *Semin Cell Dev Biol* 36: 102–12. doi:10.1016/j.semcdb.2014.09.014.

Fritsch, C, A Herrmann, M Nothnagel, K Szafranski, K Huse, F Schumann, S Schreiber, et al. 2012. "Genome-Wide Search for Novel Human uORFs and N-Terminal Protein Extensions Using Ribosomal Footprinting." Journal Article. *Genome Res* 22 (11): 2208–18. doi:10.1101/gr.139568.112.

Gallie, D R. 1991. "The Cap and poly(A) Tail Function Synergistically to Regulate mRNA Translational Efficiency." Journal Article. *Genes & Development* 5 (11). United States: 2108–16.

Gandin, V, L Masvidal, L Hulea, S P Gravel, M Cargnello, S McLaughlan, Y Cai, et al. 2016. "nanoCAGE Reveals 5' UTR Features That Define Specific Modes of Translation of Functionally Related MTOR-Sensitive mRNAs." Journal Article. *Genome Res* 26 (5): 636–48. doi:10.1101/gr.197566.115.

Gardin, Justin, Rukhsana Yeasmin, Alisa Yurovsky, Ying Cai, Steve Skiena, and Bruce Futcher. 2014. "Measurement of Average Decoding Rates of the 61 Sense Codons in Vivo." Journal Article, Research Support, N.I.H., Extramural, Research Support, U.S. Gov't, Non-P.H.S. *eLife* 3 (October). England. doi:10.7554/eLife.03735.

Gerashchenko, Maxim V, and Vadim N Gladyshev. 2014. "Translation Inhibitors Cause Abnormalities in Ribosome Profiling Experiments." Journal Article, Research Support, N.I.H., Extramural. *Nucleic Acids Research* 42 (17). England: e134. doi:10.1093/nar/gku671.

———. 2016. "Ribonuclease Selection for Ribosome Profiling." Journal Article. *Nucleic Acids Research*, September. England. doi:10.1093/nar/gkw822.

Gerashchenko, Maxim V, Alexei V Lobanov, and Vadim N Gladyshev. 2012. "Genome-Wide Ribosome Profiling Reveals Complex Translational Regulation in Response to Oxidative Stress." *Proceedings of the National Academy of Sciences of the United States of America* 109 (43): 17394–99. doi:10.1073/pnas.1120799109.

Goke, A, R Goke, A Knolle, H Trusheim, H Schmidt, A Wilmen, R Carmody, B Goke, and Y H Chen. 2002. "DUG Is a Novel Homologue of Translation Initiation Factor 4G That Binds eIF4A." Journal Article. *Biochem Biophys Res Commun* 297 (1): 78–82. http://www.ncbi.nlm.nih.gov/pubmed/12220511.

Gonzalez, C, J S Sims, N Hornstein, A Mela, F Garcia, L Lei, D A Gass, et al. 2014. "Ribosome Profiling Reveals a Cell-Type-Specific Translational Landscape in Brain Tumors." Journal Article. *J Neurosci* 34 (33): 10924–36. doi:10.1523/JNEUROSCI.0084-14.2014.

Greene, L A, and A S Tischler. 1976. "Establishment of a Noradrenergic Clonal Line of Rat Adrenal Pheochromocytoma Cells Which Respond to Nerve Growth Factor." *Proceedings of the National Academy of Sciences of the United States of America* 73 (7): 2424–28. http://www.ncbi.nlm.nih.gov/pubmed/1065897.

Gruber, A. R., R. Lorenz, S. H. Bernhart, R. Neubock, and I. L. Hofacker. 2008. "The Vienna RNA Websuite." *Nucleic Acids Research* 36 (Web Server): W70–74. doi:10.1093/nar/gkn188.

Gu, YuanYuan, Isaac T W Harley, Lindsay B Henderson, Bruce J Aronow, Ilja Vietor, Lukas A Huber, John B Harley, et al. 2009. “Identification of IFRD1 as a Modifier Gene for Cystic Fibrosis Lung Disease.” Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov’t, Twin Study. *Nature* 458 (7241). England: 1039–42. doi:10.1038/nature07811.

Guo, H, N T Ingolia, J S Weissman, and D P Bartel. 2010. “Mammalian microRNAs Predominantly Act to Decrease Target mRNA Levels.” Journal Article. *Nature* 466 (7308): 835–40. doi:10.1038/nature09267.

Guydosh, N R, and R Green. 2014. “Dom34 Rescues Ribosomes in 3’ Untranslated Regions.” Journal Article. *Cell* 156 (5): 950–62. doi:10.1016/j.cell.2014.02.006.

Harding, Heather P, Yuhong Zhang, Donalyn Scheuner, Jane-Jane Chen, Randal J Kaufman, and David Ron. 2009. “Ppp1r15 Gene Knockout Reveals an Essential Role for Translation Initiation Factor 2 Alpha (eIF2alpha) Dephosphorylation in Mammalian Development.” Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov’t. *Proceedings of the National Academy of Sciences of the United States of America* 106 (6). United States: 1832–37. doi:10.1073/pnas.0809632106.

Hatano, Masaya, Mariko Umemura, Natsumi Kimura, Takashi Yamazaki, Hitoshi Takeda, Haruo Nakano, Shigeru Takahashi, and Yuji Takahashi. 2013. “The 5’-untranslated Region Regulates ATF5 mRNA Stability via Nonsense-Mediated mRNA Decay in Response to Environmental Stress.” Journal Article, Research Support, Non-U.S. Gov’t. *The FEBS Journal* 280 (18). England: 4693–4707. doi:10.1111/febs.12440.

He, B, M Gross, and B Roizman. 1997. “The gamma(1)34.5 Protein of Herpes Simplex Virus 1 Complexes with Protein Phosphatase 1alpha to Dephosphorylate the Alpha Subunit of the Eukaryotic Translation Initiation Factor 2 and Preclude the Shutoff of Protein Synthesis by Double-Stranded RNA-Activa.” Journal Article, Research Support, U.S. Gov’t, P.H.S. *Proceedings of the National Academy of Sciences of the United States of America* 94 (3). United States: 843–48.

Heaphy, Stephen M, Marco Mariotti, Vadim N Gladyshev, John F Atkins, and Pavel V Baranov. 2016. “Novel Ciliate Genetic Code Variants Including the Reassignment of All Three Stop Codons to Sense Codons in Condylostoma Magnum.” Journal Article. *Molecular Biology and Evolution* 33 (11). United States: 2885–89.

doi:10.1093/molbev/msw166.

Hess, Jochen, Peter Angel, and Marina Schorpp-Kistner. 2004. “AP-1 Subunits: Quarrel and Harmony among Siblings.” Journal Article, Research Support, Non-U.S. Gov’t, Review. *Journal of Cell Science* 117 (Pt 25). England: 5965–73. doi:10.1242/jcs.01589.

Heyer, Erin E, and Melissa J Moore. 2016. “Redefining the Translational Status of 80S Monosomes.” Journal Article, Research Support, Non-U.S. Gov’t. *Cell* 164 (4). United States: 757–69. doi:10.1016/j.cell.2016.01.003.

Hinnebusch, A G. 1997. “Translational Regulation of Yeast GCN4. A Window on Factors That Control Initiator-Trna Binding to the Ribosome.” Journal Article, Review. *The Journal of Biological Chemistry* 272 (35). United States: 21661–64.

———. 2014. “The Scanning Mechanism of Eukaryotic Translation Initiation.” Journal Article. *Annu Rev Biochem* 83: 779–812. doi:10.1146/annurev-biochem-060713-035802.

Hinnebusch, A G, I P Ivanov, and N Sonenberg. 2016. “Translational Control by 5’-untranslated Regions of Eukaryotic mRNAs.” Journal Article. *Science* 352 (6292): 1413–16. doi:10.1126/science.aad9868.

Homma, M K, I Wada, T Suzuki, J Yamaki, E G Krebs, and Y Homma. 2005. “CK2 Phosphorylation of Eukaryotic Translation Initiation Factor 5 Potentiates Cell Cycle Progression.” Journal Article. *Proc Natl Acad Sci U S A* 102 (43): 15688–93. doi:10.1073/pnas.0506791102.

Hopkins, Benjamin D, Barry Fine, Nicole Steinbach, Meaghan Dendy, Zachary Rapp, Jacquelyn Shaw, Kyrie Pappas, et al. 2013. “A Secreted PTEN Phosphatase That Enters Cells to Alter Signaling and Survival.” *Science (New York, N.Y.)* 341 (6144): 399–402. doi:10.1126/science.1234907.

Howard, Michael T, Bradley A Carlson, Christine B Anderson, and Dolph L Hatfield. 2013. “Translational Redefinition of UGA Codons Is Regulated by Selenium Availability.” Journal Article, Research Support, N.I.H., Extramural, Research Support, N.I.H., Intramural, Research Support, Non-U.S. Gov’t. *The Journal of Biological Chemistry* 288 (27). United States: 19401–13. doi:10.1074/jbc.M113.481051.

Hsieh, A C, Y Liu, M P Edlind, N T Ingolia, M R Janes, A Sher, E Y Shi, et al. 2012. “The Translational Landscape of mTOR Signalling Steers Cancer Initiation and Metastasis.” Journal Article. *Nature* 485 (7396): 55–61. doi:10.1038/nature10912.

Huang, C, W Y Ma, J Li, and Z Dong. 1999. “Arsenic Induces Apoptosis through a c-Jun NH2-Terminal Kinase-Dependent, p53-Independent Pathway.” Journal Article, Research Support, Non-U.S. Gov’t, Research Support, U.S. Gov’t, P.H.S. *Cancer Research* 59 (13). United States: 3053–58.

Huang, Da Wei, Brad T Sherman, and Richard A Lempicki. 2009. “Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources.” *Nature Protocols* 4 (1): 44–57. doi:10.1038/nprot.2008.211.

Hunter, J. D. 2007. “Matplotlib: A 2D Graphic Environment.” *Computing in Science & Engineering* 9 (3): 90–95. doi:10.1109/MCSE.2007.55.

Hurtaud, C, C Gelly, F Bouillaud, and C Levi-Meyrueis. 2006. “Translation Control of UCP2 Synthesis by the Upstream Open Reading Frame.” Journal Article, Research Support, Non-U.S. Gov’t. *Cellular and Molecular Life Sciences : CMLS* 63 (15). Switzerland: 1780–89. doi:10.1007/s00018-006-6129-0.

Hussmann, Jeffrey A, Stephanie Patchett, Arlen Johnson, Sara Sawyer, and William H Press. 2015. “Understanding Biases in Ribosome Profiling Experiments Reveals Signatures of Translation Dynamics in Yeast.” Journal Article, Research Support, N.I.H., Extramural. *PLoS Genetics* 11 (12). United States: e1005732. doi:10.1371/journal.pgen.1005732.

Hyatt, Doug, Gwo-Liang Chen, Philip F Locascio, Miriam L Land, Frank W Larimer, and Loren J Hauser. 2010. “Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification.” *BMC Bioinformatics* 11 (1): 119. doi:10.1186/1471-2105-11-119.

Iadecola, Costantino, and Josef Anrather. 2011. “Stroke Research at a Crossroad: Asking the Brain for Directions.” Journal Article, Research Support, N.I.H., Extramural. *Nature Neuroscience* 14 (11). United States: 1363–68. doi:10.1038/nn.2953.

Ingolia, N T. 2014. “Ribosome Profiling: New Views of Translation, from Single Codons to Genome Scale.” Journal Article. *Nat Rev Genet* 15 (3): 205–13. doi:10.1038/nrg3645.

———. 2016. “Ribosome Footprint Profiling of Translation throughout the Genome.” Journal Article. *Cell* 165 (1): 22–33. doi:10.1016/j.cell.2016.02.066.

Ingolia, N T, G A Brar, S Rouskin, A M McGeachy, and J S Weissman. 2012. “The Ribosome Profiling Strategy for Monitoring Translation in Vivo by Deep Sequencing of Ribosome-Protected mRNA Fragments.” Journal Article. *Nat Protoc* 7 (8): 1534–50. doi:10.1038/nprot.2012.086.

Ingolia, N T, S Ghaemmaghami, J R Newman, and J S Weissman. 2009. “Genome-Wide

Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling." Journal Article. *Science* 324 (5924): 218–23. doi:10.1126/science.1168978.

Ingolia, N T, L F Lareau, and J S Weissman. 2011. "Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes." Journal Article. *Cell* 147 (4): 789–802. doi:10.1016/j.cell.2011.10.002.

Ingolia, Nicholas T, Gloria A Brar, Noam Stern-Ginossar, Michael S Harris, Gaelle J S Talhouarne, Sarah E Jackson, Mark R Wills, and Jonathan S Weissman. 2014. "Ribosome Profiling Reveals Pervasive Translation outside of Annotated Protein-Coding Genes." Journal Article, Research Support, Non-U.S. Gov't. *Cell Reports* 8 (5). United States: 1365–79. doi:10.1016/j.celrep.2014.07.045.

Ivanov, I P, A E Firth, A M Michel, J F Atkins, and P V Baranov. 2011. "Identification of Evolutionarily Conserved Non-AUG-Initiated N-Terminal Extensions in Human Coding Sequences." Journal Article. *Nucleic Acids Res* 39 (10): 4220–34. doi:10.1093/nar/gkr007.

Ivanov, I P, G Loughran, and J F Atkins. 2008. "uORFs with Unusual Translational Start Codons Autoregulate Expression of Eukaryotic Ornithine Decarboxylase Homologs." Journal Article. *Proc Natl Acad Sci U S A* 105 (29): 10079–84. doi:10.1073/pnas.0801590105.

Ivanov, I P, G Loughran, M S Sachs, and J F Atkins. 2010. "Initiation Context Modulates Autoregulation of Eukaryotic Translation Initiation Factor 1 (eIF1)." Journal Article. *Proc Natl Acad Sci U S A* 107 (42): 18056–60. doi:10.1073/pnas.1009269107.

Iwasaki, S, S N Floor, and N T Ingolia. 2016. "Rocaglates Convert DEAD-Box Protein eIF4A into a Sequence-Selective Translational Repressor." Journal Article. *Nature* 534 (7608): 558–61. doi:10.1038/nature17978.

Jackson, R J, C U Hellen, and T V Pestova. 2012. "Termination and Post-Termination Events in Eukaryotic Translation." Journal Article. *Adv Protein Chem Struct Biol* 86: 45–93. doi:10.1016/B978-0-12-386497-0.00002-5.

Jackson, R, and N Standart. 2015. "The Awesome Power of Ribosome Profiling." Journal Article. *RNA* 21 (4): 652–54. doi:10.1261/rna.049908.115.

Jackson, Richard J. 2013. "The Current Status of Vertebrate Cellular mRNA IRESs." Journal Article, Review. *Cold Spring Harbor Perspectives in Biology* 5 (2). United States. doi:10.1101/cshperspect.a011569.

Jan, C H, C C Williams, and J S Weissman. 2014. “Principles of ER Cotranslational Translocation Revealed by Proximity-Specific Ribosome Profiling.” Journal Article. *Science* 346 (6210): 1257521. doi:10.1126/science.1257521.

Janich, P, A B Arpat, V Castelo-Szekely, M Lopes, and D Gatfield. 2015. “Ribosome Profiling Reveals the Rhythmic Liver Translatome and Circadian Clock Regulation by Upstream Open Reading Frames.” Journal Article. *Genome Res* 25 (12): 1848–59. doi:10.1101/gr.195404.115.

Jewell, U R, I Kvietikova, A Scheid, C Bauer, R H Wenger, and M Gassmann. 2001. “Induction of HIF-1alpha in Response to Hypoxia Is Instantaneous.” Journal Article. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* 15 (7). United States: 1312–14.

Ji, Z, R Song, H Huang, A Regev, and K Struhl. 2016. “Transcriptome-Scale RNase-Footprinting of RNA-Protein Complexes.” Journal Article. *Nat Biotechnol* 34 (4): 410–13. doi:10.1038/nbt.3441.

Johannes, G, M S Carter, M B Eisen, P O Brown, and P Sarnow. 1999. “Identification of Eukaryotic mRNAs That Are Translated at Reduced Cap Binding Complex eIF4F Concentrations Using a cDNA Microarray.” Journal Article. *Proc Natl Acad Sci U S A* 96 (23): 13118–23. http://www.ncbi.nlm.nih.gov/pubmed/10557283.

Jousse, C, A Bruhat, V Carraro, F Urano, M Ferrara, D Ron, and P Fafournoux. 2001. “Inhibition of CHOP Translation by a Peptide Encoded by an Open Reading Frame Localized in the Chop 5’UTR.” Journal Article, Research Support, Non-U.S. Gov’t. *Nucleic Acids Research* 29 (21). England: 4341–51.

Jousse, Celine, Seiichi Oyadomari, Isabel Novoa, Phoebe Lu, Yuhong Zhang, Heather P Harding, and David Ron. 2003. “Inhibition of a Constitutive Translation Initiation Factor 2alpha Phosphatase, CReP, Promotes Survival of Stressed Cells.” Journal Article, Research Support, Non-U.S. Gov’t, Research Support, U.S. Gov’t, P.H.S. *The Journal of Cell Biology* 163 (4). United States: 767–75. doi:10.1083/jcb.200308075.

Jungreis, Irwin, Clara S Chan, Robert M Waterhouse, Gabriel Fields, Michael F Lin, and Manolis Kellis. 2016. “Evolutionary Dynamics of Abundant Stop Codon Readthrough.” Journal Article. *Molecular Biology and Evolution* 33 (12). United States: 3108–32. doi:10.1093/molbev/msw189.

Jungreis, Irwin, Michael F Lin, Rebecca Spokony, Clara S Chan, Nicolas Negre, Alec Victorsen,

Kevin P White, and Manolis Kellis. 2011. “Evidence of Abundant Stop Codon Readthrough in Drosophila and Other Metazoa.” Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov’t, Research Support, U.S. Gov’t, Non-P.H.S. *Genome Research* 21 (12). United States: 2096–2113. doi:10.1101/gr.119974.110.

Kaelin, William G Jr, and Peter J Ratcliffe. 2008. “Oxygen Sensing by Metazoans: The Central Role of the HIF Hydroxylase Pathway.” Journal Article, Research Support, Non-U.S. Gov’t, Review. *Molecular Cell* 30 (4). United States: 393–402. doi:10.1016/j.molcel.2008.04.009.

Kantidakis, T, B A Ramsbottom, J L Birch, S N Dowding, and R J White. 2010. “mTOR Associates with TFIIIC, Is Found at tRNA and 5S rRNA Genes, and Targets Their Repressor Maf1.” Journal Article. *Proc Natl Acad Sci U S A* 107 (26): 11823–28. doi:10.1073/pnas.1005188107.

Karolchik, Donna, Galt P Barber, Jonathan Casper, Hiram Clawson, Melissa S Cline, Mark Diekhans, Timothy R Dreszer, et al. 2014. “The UCSC Genome Browser Database: 2014 Update.” Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov’t. *Nucleic Acids Research* 42 (Database issue). England: D764-70. doi:10.1093/nar/gkt1168.

Kim, Min-Sik, Sneha M Pinto, Derese Getnet, Raja Sekhar Nirujogi, Srikanth S Manda, Raghothama Chaerkady, Anil K Madugundu, et al. 2014. “A Draft Map of the Human Proteome.” Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov’t. *Nature* 509 (7502). England: 575–81. doi:10.1038/nature13302.

Kochetov, A V, A Sarai, I B Rogozin, V K Shumny, and N A Kolchanov. 2005. “The Role of Alternative Translation Start Sites in the Generation of Human Protein Diversity.” Journal Article. *Mol Genet Genomics* 273 (6): 491–96. doi:10.1007/s00438-005-1152-7.

Kontos, H, S Napthine, and I Brierley. 2001. “Ribosomal Pausing at a Frameshifter RNA Pseudoknot Is Sensitive to Reading Phase but Shows Little Correlation with Frameshift Efficiency.” Journal Article, Research Support, Non-U.S. Gov’t. *Molecular and Cellular Biology* 21 (24). United States: 8657–70. doi:10.1128/MCB.21.24.8657-8670.2001.

Kozak, M. 1977. “Nucleotide Sequences of 5’-terminal Ribosome-Protected Initiation Regions from Two Reovirus Messages.” Journal Article. *Nature* 269 (5627): 391–94. http://www.ncbi.nlm.nih.gov/pubmed/909586.

———. 1986. “Point Mutations Define a Sequence Flanking the AUG Initiator Codon That Modulates Translation by Eukaryotic Ribosomes.” Journal Article. *Cell* 44 (2): 283–92. http://www.ncbi.nlm.nih.gov/pubmed/3943125.

———. 1989. “Context Effects and Inefficient Initiation at Non-AUG Codons in Eucaryotic Cell-Free Translation Systems.” Journal Article. *Mol Cell Biol* 9 (11): 5073–80. http://www.ncbi.nlm.nih.gov/pubmed/2601709.

Kuersten, S, A Radek, C Vogel, and L O Penalva. 2013. “Translation Regulation Gets Its ‘Omics’ Moment.” Journal Article. *Wiley Interdiscip Rev RNA* 4 (6): 617–30. doi:10.1002/wrna.1173.

Kuhn, K M, J L DeRisi, P O Brown, and P Sarnow. 2001. “Global and Specific Translational Regulation in the Genomic Response of Saccharomyces Cerevisiae to a Rapid Transfer from a Fermentable to a Nonfermentable Carbon Source.” *Molecular and Cellular Biology* 21 (3): 916–27. doi:10.1128/MCB.21.3.916-927.2001.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L Salzberg. 2009. “Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome.” Journal Article, Research Support, N.I.H., Extramural. *Genome Biology* 10 (3). England: R25. doi:10.1186/gb-2009-10-3-r25.

Lareau, L F, D H Hite, G J Hogan, and P O Brown. 2014. “Distinct Stages of the Translation Elongation Cycle Revealed by Sequencing Ribosome-Protected mRNA Fragments.” Journal Article. *Elife* 3: e01257. doi:10.7554/eLife.01257.

Larsen, B, N M Wills, R F Gesteland, and J F Atkins. 1994. “rRNA-mRNA Base Pairing Stimulates a Programmed -1 Ribosomal Frameshift.” *Journal of Bacteriology* 176 (22): 6842–51. http://www.ncbi.nlm.nih.gov/pubmed/7961443.

Larsson, O, N Sonenberg, and R Nadon. 2011. “Anota: Analysis of Differential Translation in Genome-Wide Studies.” Journal Article. *Bioinformatics* 27 (10): 1440–41. doi:10.1093/bioinformatics/btr146.

Lazarowitz, S G, and H D Robertson. 1977. “Initiator Regions from the Small Size Class of Reovirus Messenger RNA Protected by Rabbit Reticulocyte Ribosomes.” Journal Article. *J Biol Chem* 252 (21): 7842–49. http://www.ncbi.nlm.nih.gov/pubmed/914843.

Lee, S, B Liu, S Lee, S X Huang, B Shen, and S B Qian. 2012. “Global Mapping of Translation Initiation Sites in Mammalian Cells at Single-Nucleotide Resolution.” Journal Article. *Proc Natl Acad Sci U S A* 109 (37): E2424-32. doi:10.1073/pnas.1207846109.

Lee, Yun-Young, Randal C Cevallos, and Eric Jan. 2009. “An Upstream Open Reading Frame Regulates Translation of GADD34 during Cellular Stresses That Induce eIF2alpha Phosphorylation.” Journal Article, Research Support, Non-U.S. Gov’t. *The Journal of Biological Chemistry* 284 (11). United States: 6661–73. doi:10.1074/jbc.M806735200.

Legendre, R, A Baudin-Baillieu, I Hatin, and O Namy. 2015. “RiboTools: A Galaxy Toolbox for Qualitative Ribosome Profiling Analysis.” Journal Article. *Bioinformatics* 31 (15): 2586–88. doi:10.1093/bioinformatics/btv174.

Leprivier, Gabriel, Marc Remke, Barak Rotblat, Adrian Dubuc, Abigail-Rachele F Mateo, Marcel Kool, Sameer Agnihotri, et al. 2013. “The eEF2 Kinase Confers Resistance to Nutrient Deprivation by Blocking Translation Elongation.” *Cell* 153 (5): 1064–79. doi:10.1016/j.cell.2013.04.055.

Li, Gene-Wei, David Burkhardt, Carol Gross, and Jonathan S Weissman. 2014. “Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources.” Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov’t. *Cell* 157 (3). United States: 624–35. doi:10.1016/j.cell.2014.02.033.

Li, Gene-Wei, Eugene Oh, and Jonathan S Weissman. 2012. “The Anti-Shine-Dalgarno Sequence Drives Translational Pausing and Codon Choice in Bacteria.” Journal Article, Research Support, Non-U.S. Gov’t. *Nature* 484 (7395). England: 538–41. doi:10.1038/nature10965.

Liu, B, Y Han, and S B Qian. 2013. “Cotranslational Response to Proteotoxic Stress by Elongation Pausing of Ribosomes.” Journal Article. *Mol Cell* 49 (3): 453–63. doi:10.1016/j.molcel.2012.12.001.

Liu, M J, S H Wu, J F Wu, W D Lin, Y C Wu, T Y Tsai, H L Tsai, and S H Wu. 2013. “Translational Landscape of Photomorphogenic Arabidopsis.” Journal Article. *Plant Cell* 25 (10): 3699–3710. doi:10.1105/tpc.113.114769.

Loayza-Puch, Fabricio, Jarno Drost, Koos Rooijers, Rui Lopes, Ran Elkon, and Reuven Agami. 2013. “p53 Induces Transcriptional and Translational Programs to Suppress Cell Proliferation and Growth.” Journal Article, Research Support, Non-U.S. Gov’t. *Genome Biology* 14 (4). England: R32. doi:10.1186/gb-2013-14-4-r32.

Loayza-Puch, F, K Rooijers, L C Buil, J Zijlstra, J F Oude Vrielink, R Lopes, A P Ugalde, et al. 2016. “Tumour-Specific Proline Vulnerability Uncovered by Differential Ribosome Codon Reading.” Journal Article. *Nature* 530 (7591): 490–94.

doi:10.1038/nature16982.

Lobanov, Alexei V, Stephen M Heaphy, Anton A Turanov, Maxim V Gerashchenko, Sandra Pucciarelli, Raghul R Devaraj, Fang Xie, et al. 2016. “Position-Dependent Termination and Widespread Obligatory Frameshifting in Euplotes Translation.” Journal Article. *Nature Structural & Molecular Biology*, November. United States. doi:10.1038/nsmb.3330.

Loenarz, Christoph, Rok Sekirnik, Armin Thalhammer, Wei Ge, Ekaterina Spivakovsky, Mukram M Mackeen, Michael A McDonough, et al. 2014. “Hydroxylation of the Eukaryotic Ribosomal Decoding Center Affects Translational Accuracy.” Journal Article, Research Support, Non-U.S. Gov’t. *Proceedings of the National Academy of Sciences of the United States of America* 111 (11). United States: 4019–24. doi:10.1073/pnas.1311750111.

Lopez, Daniel, and Florencio Pazos. 2015. “Protein Functional Features Are Reflected in the Patterns of mRNA Translation Speed.” Journal Article, Research Support, Non-U.S. Gov’t. *BMC Genomics* 16 (July). England: 513. doi:10.1186/s12864-015-1734-7.

Lorenz, Ronny, Stephan H Bernhart, Christian Honer Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. 2011. “ViennaRNA Package 2.0.” Journal Article. *Algorithms for Molecular Biology : AMB* 6 (November). England: 26. doi:10.1186/1748-7188-6-26.

Loughran, Gary, Ming-Yuan Chou, Ivaylo P Ivanov, Irwin Jungreis, Manolis Kellis, Anmol M Kiran, Pavel V Baranov, and John F Atkins. 2014. “Evidence of Efficient Stop Codon Readthrough in Four Mammalian Genes.” Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov’t, Research Support, U.S. Gov’t, Non-P.H.S. *Nucleic Acids Research* 42 (14). England: 8928–38. doi:10.1093/nar/gku608.

Loughran, G, M S Sachs, J F Atkins, and I P Ivanov. 2012. “Stringency of Start Codon Selection Modulates Autoregulation of Translation Initiation Factor eIF5.” Journal Article. *Nucleic Acids Res* 40 (7): 2898–2906. doi:10.1093/nar/gkr1192.

Low, W K, Y Dang, T Schneider-Poetsch, Z Shi, N S Choi, W C Merrick, D Romo, and J O Liu. 2005. “Inhibition of Eukaryotic Translation Initiation by the Marine Natural Product Pateamine A.” Journal Article. *Mol Cell* 20 (5): 709–22. doi:10.1016/j.molcel.2005.10.008.

Lu, Jianli, and Carol Deutsch. 2008. “Electrostatics in the Ribosomal Tunnel Modulate Chain

Elongation Rates." Journal Article, Research Support, N.I.H., Extramural. *Journal of Molecular Biology* 384 (1). England: 73–86. doi:10.1016/j.jmb.2008.08.089.

Lu, Phoebe D, Heather P Harding, and David Ron. 2004. "Translation Reinitiation at Alternative Open Reading Frames Regulates Gene Expression in an Integrated Stress Response." Journal Article, Research Support, Non-U.S. Gov't, Research Support, U.S. Gov't, P.H.S. *The Journal of Cell Biology* 167 (1). United States: 27–33. doi:10.1083/jcb.200408003.

Luo, Z, M Freitag, and M S Sachs. 1995. "Translational Regulation in Response to Changes in Amino Acid Availability in Neurospora Crassa." Journal Article, Research Support, Non-U.S. Gov't, Research Support, U.S. Gov't, P.H.S. *Molecular and Cellular Biology* 15 (10). United States: 5235–45.

Majumdar, R, A Bandyopadhyay, H Deng, and U Maitra. 2002. "Phosphorylation of Mammalian Translation Initiation Factor 5 (eIF5) in Vitro and in Vivo." Journal Article. *Nucleic Acids Res* 30 (5): 1154–62. http://www.ncbi.nlm.nih.gov/pubmed/11861906.

Manwani, Bharti, and Louise D McCullough. 2013. "Function of the Master Energy Regulator Adenosine Monophosphate-Activated Protein Kinase in Stroke." Journal Article, Research Support, N.I.H., Extramural, Review. *Journal of Neuroscience Research* 91 (8). United States: 1018–29. doi:10.1002/jnr.23207.

Márquez, Viter, Daniel N Wilson, Warren P Tate, Francisco Triana-Alonso, and Knud H Nierhaus. 2004. "Maintaining the Ribosomal Reading Frame: The Influence of the E Site during Translational Regulation of Release Factor 2." *Cell* 118 (1): 45–55. doi:10.1016/j.cell.2004.06.012.

Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.journal* 17 (1): 10–12. doi:10.14806/ej.17.1.200.

Mayer, C, and I Grummt. 2006. "Ribosome Biogenesis and Cell Growth: mTOR Coordinates Transcription by All Three Classes of Nuclear RNA Polymerases." Journal Article. *Oncogene* 25 (48): 6384–91. doi:10.1038/sj.onc.1209883.

McEwen, Edward, Nancy Kedersha, Benbo Song, Donalyn Scheuner, Natalie Gilks, Anping Han, Jane-Jane Chen, Paul Anderson, and Randal J Kaufman. 2005. "Heme-Regulated Inhibitor Kinase-Mediated Phosphorylation of Eukaryotic Translation Initiation Factor 2 Inhibits Translation, Induces Stress Granule Formation, and Mediates Survival upon Arsenite Exposure." Journal Article, Research Support, N.I.H., Extramural, Research

Support, U.S. Gov't, P.H.S. *The Journal of Biological Chemistry* 280 (17). United States: 16925–33. doi:10.1074/jbc.M412882200.

McManus, C Joel, Gemma E May, Pieter Spealman, and Alan Shteyman. 2014. "Ribosome Profiling Reveals Post-Transcriptional Buffering of Divergent Gene Expression in Yeast." Journal Article, Research Support, Non-U.S. Gov't. *Genome Research* 24 (3). United States: 422–30. doi:10.1101/gr.164996.113.

Meller, Robert, and Roger P Simon. 2013. "Tolerance to Ischemia - an Increasingly Complex Biology." Journal Article, Research Support, N.I.H., Extramural, Review. *Translational Stroke Research* 4 (1). United States: 40–50. doi:10.1007/s12975-012-0246-x.

Michel, A M, A M Ahern, C A Donohue, and P V Baranov. 2015. "GWIPS-Viz as a Tool for Exploring Ribosome Profiling Evidence Supporting the Synthesis of Alternative Proteoforms." Journal Article. *Proteomics* 15 (14): 2410–16. doi:10.1002/pmic.201400603.

Michel, A M, D E Andreev, and P V Baranov. 2014. "Computational Approach for Calculating the Probability of Eukaryotic Translation Initiation from Ribo-Seq Data That Takes into Account Leaky Scanning." Journal Article. *BMC Bioinformatics* 15: 380. doi:10.1186/s12859-014-0380-4.

Michel, A M, and P V Baranov. 2013. "Ribosome Profiling: A Hi-Def Monitor for Protein Synthesis at the Genome-Wide Scale." Journal Article. *Wiley Interdiscip Rev RNA* 4 (5): 473–90. doi:10.1002/wrna.1172.

Michel, A M, K R Choudhury, A E Firth, N T Ingolia, J F Atkins, and P V Baranov. 2012. "Observation of Dually Decoded Regions of the Human Genome Using Ribosome Profiling Data." Journal Article. *Genome Res* 22 (11): 2219–29. doi:10.1101/gr.133249.111.

Michel, A M, G Fox, M Kiran A, C De Bo, P B O'Connor, S M Heaphy, J P Mullan, C A Donohue, D G Higgins, and P V Baranov. 2014. "GWIPS-Viz: Development of a Ribo-Seq Genome Browser." Journal Article. *Nucleic Acids Res* 42 (Database issue): D859-64. doi:10.1093/nar/gkt1035.

Michel, A M, J P Mullan, V Velayudhan, P B O'Connor, C A Donohue, and P V Baranov. 2016. "RiboGalaxy: A Browser Based Platform for the Alignment, Analysis and Visualization of Ribosome Profiling Data." Journal Article. *RNA Biol* 13 (3): 316–19. doi:10.1080/15476286.2016.1141862.

Miettinen, Teemu P, and Mikael Bjorklund. 2015. “Modified Ribosome Profiling Reveals High Abundance of Ribosome Protected mRNA Fragments Derived from 3’ Untranslated Regions.” Journal Article, Research Support, Non-U.S. Gov’t. *Nucleic Acids Research* 43 (2). England: 1019–34. doi:10.1093/nar/gku1310.

Mikulits, W, B Pradet-Balade, B Habermann, H Beug, J A Garcia-Sanz, and E W Mullner. 2000. “Isolation of Translationally Controlled mRNAs by Differential Screening.” Journal Article. *FASEB J* 14 (11): 1641–52. http://www.ncbi.nlm.nih.gov/pubmed/10928999.

Min, S-H, D M Kim, Y-S Heo, Y-I Kim, H M Kim, J Kim, Y-M Han, I-C Kim, and O-J Yoo. 2009. “New p53 Target, Phosphatase of Regenerating Liver 1 (PRL-1) Downregulates p53.” Journal Article, Research Support, Non-U.S. Gov’t. *Oncogene* 28 (4). England: 545–54. doi:10.1038/onc.2008.409.

Mohammad, Fuad, Christopher J Woolstenhulme, Rachel Green, and Allen R Buskirk. 2016. “Clarifying the Translational Pausing Landscape in Bacteria by Ribosome Profiling.” Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov’t. *Cell Reports* 14 (4). United States: 686–94. doi:10.1016/j.celrep.2015.12.073.

Mokrejs, Martin, Tomas Masek, Vaclav Vopalensky, Petr Hlubucek, Philippe Delbos, and Martin Pospisek. 2010. “IRESite--a Tool for the Examination of Viral and Cellular Internal Ribosome Entry Sites.” Journal Article, Research Support, Non-U.S. Gov’t. *Nucleic Acids Research* 38 (Database issue). England: D131-6. doi:10.1093/nar/gkp981.

Moreno, Julie A, Helois Radford, Diego Peretti, Joern R Steinert, Nicholas Verity, Maria Guerra Martin, Mark Halliday, et al. 2012. “Sustained Translational Repression by eIF2alpha-P Mediates Prion Neurodegeneration.” Journal Article, Research Support, Non-U.S. Gov’t. *Nature* 485 (7399). England: 507–11. doi:10.1038/nature11058.

Murakami, Y, M Marumo, and S Hayashi. 1988. “Ornithine Decarboxylase Antizyme in Kidneys of Male and Female Mice.” Journal Article, Research Support, Non-U.S. Gov’t. *The Biochemical Journal* 254 (2). England: 367–72.

Nagata, Daisuke, Masaki Mogi, and Kenneth Walsh. 2003. “AMP-Activated Protein Kinase (AMPK) Signaling in Endothelial Cells Is Essential for Angiogenesis in Response to Hypoxic Stress.” Journal Article, Research Support, Non-U.S. Gov’t, Research Support, U.S. Gov’t, P.H.S. *The Journal of Biological Chemistry* 278 (33). United States: 31000–6. doi:10.1074/jbc.M300643200.

Namgung, U, and Z Xia. 2000. “Arsenite-Induced Apoptosis in Cortical Neurons Is Mediated by c-Jun N-Terminal Protein Kinase 3 and p38 Mitogen-Activated Protein Kinase.” Journal Article, Research Support, Non-U.S. Gov’t, Research Support, U.S. Gov’t, P.H.S. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 20 (17). United States: 6442–51.

Nanda, Jagpreet S, Yuen-Nei Cheung, Julie E Takacs, Pilar Martin-Marcos, Adesh K Saini, Alan G Hinnebusch, and Jon R Lorsch. 2009. “eIF1 Controls Multiple Steps in Start Codon Recognition during Eukaryotic Translation Initiation.” *Journal of Molecular Biology* 394 (2): 268–85. doi:10.1016/j.jmb.2009.09.017.

Nedialkova, D D, and S A Leidel. 2015. “Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity.” Journal Article. *Cell* 161 (7): 1606–18. doi:10.1016/j.cell.2015.05.022.

Noderer, W L, R J Flockhart, A Bhaduri, A J Diaz de Arce, J Zhang, P A Khavari, and C L Wang. 2014. “Quantitative Analysis of Mammalian Translation Initiation Sites by FACS-Seq.” Journal Article. *Mol Syst Biol* 10: 748. doi:10.15252/msb.20145136.

Novoa, I, H Zeng, H P Harding, and D Ron. 2001. “Feedback Inhibition of the Unfolded Protein Response by GADD34-Mediated Dephosphorylation of eIF2alpha.” Journal Article, Research Support, Non-U.S. Gov’t, Research Support, U.S. Gov’t, P.H.S. *The Journal of Cell Biology* 153 (5). United States: 1011–22.

Nurenberg, E, and R Tampe. 2013. “Tying up Loose Ends: Ribosome Recycling in Eukaryotes and Archaea.” Journal Article. *Trends Biochem Sci* 38 (2): 64–74. doi:10.1016/j.tibs.2012.11.003.

O’Connor, P B, D E Andreev, and P V Baranov. 2016. “Comparative Survey of the Relative Impact of mRNA Features on Local Ribosome Profiling Read Density.” Journal Article. *Nat Commun* 7: 12915. doi:10.1038/ncomms12915.

O’Connor, P B, G W Li, J S Weissman, J F Atkins, and P V Baranov. 2013. “rRNA:mRNA Pairing Alters the Length and the Symmetry of mRNA-Protected Fragments in Ribosome Profiling Experiments.” Journal Article. *Bioinformatics* 29 (12): 1488–91. doi:10.1093/bioinformatics/btt184.

Oh, E, A H Becker, A Sandikci, D Huber, R Chaba, F Gloge, R J Nichols, et al. 2011. “Selective Ribosome Profiling Reveals the Cotranslational Chaperone Action of Trigger Factor in Vivo.” Journal Article. *Cell* 147 (6): 1295–1308. doi:10.1016/j.cell.2011.10.044.

Olexiouk, Volodimir, Jeroen Crapp??, Steven Verbruggen, Kenneth Verhegen, Lennart Martens, and Gerben Menschaert. 2016. “SORFs.org: A Repository of Small ORFs Identified by Ribosome Profiling.” *Nucleic Acids Research* 44 (D1): D324–29. doi:10.1093/nar/gkv1175.

Olshen, A B, A C Hsieh, C R Stumpf, R A Olshen, D Ruggero, and B S Taylor. 2013. “Assessing Gene-Level Translational Control from Ribosome Profiling.” Journal Article. *Bioinformatics* 29 (23): 2995–3002. doi:10.1093/bioinformatics/btt533.

Palam, L R, T D Baird, and R C Wek. 2011. “Phosphorylation of eIF2 Facilitates Ribosomal Bypass of an Inhibitory Upstream ORF to Enhance CHOP Translation.” Journal Article. *J Biol Chem* 286 (13): 10939–49. doi:10.1074/jbc.M110.216093.

Parsyan, A, Y Svitkin, D Shahbazian, C Gkogkas, P Lasko, W C Merrick, and N Sonenberg. 2011. “mRNA Helicases: The Tacticians of Translational Control.” Journal Article. *Nat Rev Mol Cell Biol* 12 (4): 235–45. doi:10.1038/nrm3083.

Passmore, Lori A, T Martin Schmeing, David Maag, Drew J Applefield, Michael G Acker, Mikkel A Algire, Jon R Lorsch, and V Ramakrishnan. 2007. “The Eukaryotic Translation Initiation Factors eIF1 and eIF1A Induce an Open Conformation of the 40S Ribosome.” *Molecular Cell* 26 (1): 41–50. doi:10.1016/j.molcel.2007.03.018.

Pechmann, S, J W Chartron, and J Frydman. 2014. “Local Slowdown of Translation by Nonoptimal Codons Promotes Nascent-Chain Recognition by SRP in Vivo.” Journal Article. *Nat Struct Mol Biol* 21 (12): 1100–1105. doi:10.1038/nsmb.2919.

Perry, R P, and O Meyuhas. 1990. “Translational Control of Ribosomal Protein Production in Mammalian Cells.” Journal Article. *Enzyme* 44 (1–4): 83–92. http://www.ncbi.nlm.nih.gov/pubmed/2133661.

Pestova, Tatyana V, Sylvain de Breyne, Andrey V Pisarev, Irina S Abaeva, and Christopher U T Hellen. 2008. “eIF2-Dependent and eIF2-Independent Modes of Initiation on the CSFV IRES: A Common Role of Domain II.” Journal Article, Research Support, N.I.H., Extramural. *The EMBO Journal* 27 (7). England: 1060–72. doi:10.1038/emboj.2008.49.

Pestova, T V, S I Borukhov, and C U Hellen. 1998. “Eukaryotic Ribosomes Require Initiation Factors 1 and 1A to Locate Initiation Codons.” *Nature* 394 (6696): 854–59. doi:10.1038/29703.

Pisarev, A V, C U Hellen, and T V Pestova. 2007. “Recycling of Eukaryotic Posttermination Ribosomal Complexes.” Journal Article. *Cell* 131 (2): 286–99.

doi:10.1016/j.cell.2007.08.041.

Pisareva, V P, M A Skabkin, C U Hellen, T V Pestova, and A V Pisarev. 2011. “Dissociation by Pelota, Hbs1 and ABCE1 of Mammalian Vacant 80S Ribosomes and Stalled Elongation Complexes.” Journal Article. *EMBO J* 30 (9): 1804–17. doi:10.1038/emboj.2011.93.

Pop, Cristina, Silvi Rouskin, Nicholas T Ingolia, Lu Han, Eric M Phizicky, Jonathan S Weissman, and Daphne Koller. 2014. “Causal Signals between Codon Bias, mRNA Structure, and the Efficiency of Translation and Elongation.” Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov’t, Research Support, U.S. Gov’t, Non-P.H.S. *Molecular Systems Biology* 10 (December). England: 770.

Popa, A, K Lebrigand, P Barbry, and R Waldmann. 2016. “Pateamine A-Sensitive Ribosome Profiling Reveals the Scope of Translation in Mouse Embryonic Stem Cells.” Journal Article. *BMC Genomics* 17: 52. doi:10.1186/s12864-016-2384-0.

Popa, A, K Lebrigand, A Paquet, N Nottet, K Robbe-Sermesant, R Waldmann, and P Barbry. 2016. “RiboProfiling: A Bioconductor Package for Standard Ribo-Seq Pipeline Processing.” Journal Article. *F1000Res* 5: 1309. doi:10.12688/f1000research.8964.1.

Presnyak, V, N Alhusaini, Y H Chen, S Martin, N Morris, N Kline, S Olson, et al. 2015. “Codon Optimality Is a Major Determinant of mRNA Stability.” Journal Article. *Cell* 160 (6): 1111–24. doi:10.1016/j.cell.2015.02.029.

Prostko, C R, M A Brostrom, and C O Brostrom. 1993. “Reversible Phosphorylation of Eukaryotic Initiation Factor 2 Alpha in Response to Endoplasmic Reticular Signaling.” Journal Article, Research Support, U.S. Gov’t, Non-P.H.S., Research Support, U.S. Gov’t, P.H.S. *Molecular and Cellular Biochemistry* 127–128 (November). Netherlands: 255–65.

Pruitt, Kim D, Garth R Brown, Susan M Hiatt, Francoise Thibaud-Nissen, Alexander Astashyn, Olga Ermolaeva, Catherine M Farrell, et al. 2014. “RefSeq: An Update on Mammalian Reference Sequences.” Journal Article, Research Support, N.I.H., Intramural. *Nucleic Acids Research* 42 (Database issue). England: D756-63. doi:10.1093/nar/gkt1114.

Qian, Wenfeng, Jian-Rong Yang, Nathaniel M Pearson, Calum Maclean, and Jianzhi Zhang. 2012. “Balanced Codon Usage Optimizes Eukaryotic Translational Efficiency.” Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov’t. *PLoS Genetics* 8 (3). United States: e1002603. doi:10.1371/journal.pgen.1002603.

Quackenbush, John. 2002. “Microarray Data Normalization and Transformation.” Journal

Article, Research Support, U.S. Gov't, Non-P.H.S., Research Support, U.S. Gov't, P.H.S., Review. *Nature Genetics* 32 Suppl (December). United States: 496–501. doi:10.1038/ng1032.

Rabani, Michal, Joshua Z Levin, Lin Fan, Xian Adiconis, Raktima Raychowdhury, Manuel Garber, Andreas Gnirke, et al. 2011. "Metabolic Labeling of RNA Uncovers Principles of RNA Production and Degradation Dynamics in Mammalian Cells." Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't. *Nature Biotechnology* 29 (5). United States: 436–42. doi:10.1038/nbt.1861.

Radhakrishnan, A, Y H Chen, S Martin, N Alhusaini, R Green, and J Coller. 2016. "The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality." Journal Article. *Cell* 167 (1): 122–132 e9. doi:10.1016/j.cell.2016.08.053.

Raj, A, S H Wang, H Shim, A Harpak, Y I Li, B Engelmann, M Stephens, Y Gilad, and J K Pritchard. 2016. "Thousands of Novel Translated Open Reading Frames in Humans Inferred by Ribosome Footprint Profiling." Journal Article. *Elife* 5. doi:10.7554/eLife.13328.

Ratcliffe, Peter J. 2013. "Oxygen Sensing and Hypoxia Signalling Pathways in Animals: The Implications of Physiology for Cancer." Journal Article, Research Support, Non-U.S. Gov't, Review. *The Journal of Physiology* 591 (8). England: 2027–42. doi:10.1113/jphysiol.2013.251470.

Reid, David W, Qiang Chen, Angeline S-L Tay, Shirish Shenolikar, and Christopher V Nicchitta. 2014. "The Unfolded Protein Response Triggers Selective mRNA Release from the Endoplasmic Reticulum." Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't. *Cell* 158 (6). United States: 1362–74. doi:10.1016/j.cell.2014.08.012.

Reid, D W, S Shenolikar, and C V Nicchitta. 2015. "Simple and Inexpensive Ribosome Profiling Analysis of mRNA Translation." Journal Article. *Methods* 91: 69–74. doi:10.1016/j.ymeth.2015.07.003.

Requiao, Rodrigo D, Henrique Jose Araujo de Souza, Silvana Rossetto, Tatiana Domitrovic, and Fernando L Palhano. 2016. "Increased Ribosome Density Associated to Positively Charged Residues Is Evident in Ribosome Profiling Experiments Performed in the Absence of Translation Inhibitors." Journal Article. *RNA Biology* 13 (6). United States: 561–68. doi:10.1080/15476286.2016.1172755.

Rooijers, K, F Loayza-Puch, L G Nijtmans, and R Agami. 2013. “Ribosome Profiling Reveals Features of Normal and Disease-Associated Mitochondrial Translation.” Journal Article. *Nat Commun* 4: 2886. doi:10.1038/ncomms3886.

Ruan, H, L M Shantz, A E Pegg, and D R Morris. 1996. “The Upstream Open Reading Frame of the mRNA Encoding S-Adenosylmethionine Decarboxylase Is a Polyamine-Responsive Translational Control Element.” Journal Article. *J Biol Chem* 271 (47): 29576–82. http://www.ncbi.nlm.nih.gov/pubmed/8939886.

Rubio, C A, B Weisburd, M Holderfield, C Arias, E Fang, J L DeRisi, and A Fanidi. 2014. “Transcriptome-Wide Characterization of the eIF4A Signature Highlights Plasticity in Translation Regulation.” Journal Article. *Genome Biol* 15 (10): 476. doi:10.1186/s13059-014-0476-1.

Ryazanov, Alexey G., Elena A. Shestakova, and Pavel G. Natapov. 1988. “Phosphorylation of Elongation Factor 2 by EF-2 Kinase Affects Rate of Translation.” *Nature* 334 (6178): 170–73. doi:10.1038/334170a0.

Sauliere, Jerome, and Herve Le Hir. 2015. “CLIP-Seq to Discover Transcriptome-Wide Imprinting of RNA Binding Proteins in Living Cells.” Journal Article. *Methods in Molecular Biology (Clifton, N.J.)* 1296. United States: 151–60. doi:10.1007/978-1-4939-2547-6_14.

Schueren, Fabian, Thomas Lingner, Rosemol George, Julia Hofhuis, Corinna Dickel, Jutta Gartner, and Sven Thoms. 2014. “Peroxisomal Lactate Dehydrogenase Is Generated by Translational Readthrough in Mammals.” Journal Article, Research Support, Non-U.S. Gov’t. *eLife* 3 (September). England: e03640. doi:10.7554/eLife.03640.

Sekine, Y, A Zyryanova, A Crespillo-Casado, P M Fischer, H P Harding, and D Ron. 2015. “Stress Responses. Mutations in a Translation Initiation Factor Identify the Target of a Memory-Enhancing Compound.” Journal Article. *Science* 348 (6238): 1027–30. doi:10.1126/science.aaa6986.

Semenza, Gregg L. 2014. “Oxygen Sensing, Hypoxia-Inducible Factors, and Disease Pathophysiology.” Journal Article, Review. *Annual Review of Pathology* 9. United States: 47–71. doi:10.1146/annurev-pathol-012513-104720.

Sen, N D, F Zhou, M S Harris, N T Ingolia, and A G Hinnebusch. 2016. “eIF4B Stimulates Translation of Long mRNAs with Structured 5’ UTRs and Low Closed-Loop Potential but Weak Dependence on eIF4G.” Journal Article. *Proc Natl Acad Sci U S A* 113 (38):

10464–72. doi:10.1073/pnas.1612398113.

Sen, N D, F Zhou, N T Ingolia, and A G Hinnebusch. 2015. “Genome-Wide Analysis of Translational Efficiency Reveals Distinct but Overlapping Functions of Yeast DEAD-Box RNA Helicases Ded1 and eIF4A.” Journal Article. *Genome Res* 25 (8): 1196–1205. doi:10.1101/gr.191601.115.

Shah, P, Y Ding, M Niemczyk, G Kudla, and J B Plotkin. 2013. “Rate-Limiting Steps in Yeast Protein Translation.” Journal Article. *Cell* 153 (7): 1589–1601. doi:10.1016/j.cell.2013.05.049.

Shalgi, R, J A Hurt, I Krykbaeva, M Taipale, S Lindquist, and C B Burge. 2013. “Widespread Regulation of Translation by Elongation Pausing in Heat Shock.” Journal Article. *Mol Cell* 49 (3): 439–52. doi:10.1016/j.molcel.2012.11.028.

Shatsky, I N, S E Dmitriev, D E Andreev, and I M Terenin. 2014. “Transcriptome-Wide Studies Uncover the Diversity of Modes of mRNA Recruitment to Eukaryotic Ribosomes.” Journal Article. *Crit Rev Biochem Mol Biol* 49 (2): 164–77. doi:10.3109/10409238.2014.887051.

Shatsky, Ivan N, Sergey E Dmitriev, Ilya M Terenin, and D E Andreev. 2010. “Cap- and IRES-Independent Scanning Mechanism of Translation Initiation as an Alternative to the Concept of Cellular IRESs.” Journal Article, Research Support, Non-U.S. Gov’t, Review. *Molecules and Cells* 30 (4). Korea (South): 285–93. doi:10.1007/s10059-010-0149-1.

Shine, J, and L Dalgarno. 1975. “Terminal-Sequence Analysis of Bacterial Ribosomal RNA. Correlation between the 3’-terminal-Polypyrimidine Sequence of 16-S RNA and Translational Specificity of the Ribosome.” *European Journal of Biochemistry* 57 (1): 221–30. http://www.ncbi.nlm.nih.gov/pubmed/809282.

Shuda, M, C Velasquez, E Cheng, D G Cordek, H J Kwun, Y Chang, and P S Moore. 2015. “CDK1 Substitutes for mTOR Kinase to Activate Mitotic Cap-Dependent Protein Translation.” Journal Article. *Proc Natl Acad Sci U S A* 112 (19): 5875–82. doi:10.1073/pnas.1505787112.

Sidrauski, C, D Acosta-Alvear, A Khoutorsky, P Vedantham, B R Hearn, H Li, K Gamache, et al. 2013. “Pharmacological Brake-Release of mRNA Translation Enhances Cognitive Memory.” Journal Article. *Elife* 2: e00498. doi:10.7554/eLife.00498.

Sidrauski, C, A M McGeachy, N T Ingolia, and P Walter. 2015. “The Small Molecule ISRIB Reverses the Effects of eIF2alpha Phosphorylation on Translation and Stress Granule

Assembly." Journal Article. *Elife* 4. doi:10.7554/eLife.05033.

Sidrauski, Carmela, Jeffery S. Cox, and Peter Walter. 1996. "tRNA Ligase Is Required for Regulated mRNA Splicing in the Unfolded Protein Response." *Cell* 87 (3): 405–13. doi:10.1016/S0092-8674(00)81361-6.

Singleton, Rachelle S, Phebee Liu-Yi, Fabio Formenti, Wei Ge, Rok Sekirnik, Roman Fischer, Julie Adam, et al. 2014. "OGFOD1 Catalyzes Prolyl Hydroxylation of RPS23 and Is Involved in Translation Control and Stress Granule Formation." Journal Article, Research Support, Non-U.S. Gov't. *Proceedings of the National Academy of Sciences of the United States of America* 111 (11). United States: 4031–36. doi:10.1073/pnas.1314482111.

Sinvani, H, O Haimov, Y Svitkin, N Sonenberg, A Tamarkin-Ben-Harush, B Viollet, and R Dikstein. 2015. "Translational Tolerance of Mitochondrial Genes to Metabolic Energy Stress Involves TISU and eIF1-eIF4GI Cooperation in Start Codon Selection." Journal Article. *Cell Metab* 21 (3): 479–92. doi:10.1016/j.cmet.2015.02.010.

Skabkin, Maxim A, Olga V Skabkina, Vidya Dhote, Anton A Komar, Christopher U T Hellen, and Tatyana V Pestova. 2010. "Activities of Ligatin and MCT-1/DENR in Eukaryotic Translation Initiation and Ribosomal Recycling." Journal Article, Research Support, N.I.H., Extramural. *Genes & Development* 24 (16). United States: 1787–1801. doi:10.1101/gad.1957510.

Sonenberg, Nahum, and Alan G Hinnebusch. 2009. "Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets." Journal Article, Research Support, N.I.H., Extramural, Research Support, N.I.H., Intramural, Research Support, Non-U.S. Gov't, Review. *Cell* 136 (4). United States: 731–45. doi:10.1016/j.cell.2009.01.042.

Stadler, M, and A Fire. 2011. "Wobble Base-Pairing Slows in Vivo Translation Elongation in Metazoans." Journal Article. *RNA* 17 (12): 2063–73. doi:10.1261/rna.02890211.

Steitz, J A. 1969. "Polypeptide Chain Initiation: Nucleotide Sequences of the Three Ribosomal Binding Sites in Bacteriophage R17 RNA." Journal Article. *Nature* 224 (5223): 957–64. http://www.ncbi.nlm.nih.gov/pubmed/5360547.

Stern-Ginossar, N. 2015. "Decoding Viral Infection by Ribosome Profiling." Journal Article. *J Virol* 89 (12): 6164–66. doi:10.1128/JVI.02528-14.

Stern-Ginossar, N, B Weisburd, A Michalski, V T Le, M Y Hein, S X Huang, M Ma, et al. 2012. "Decoding Human Cytomegalovirus." Journal Article. *Science* 338 (6110): 1088–93.

doi:10.1126/science.1227919.

Stiebler, Alina C, Johannes Freitag, Kay O Schink, Thorsten Stehlik, Britta A M Tillmann, Julia Ast, and Michael Bolker. 2014. “Ribosomal Readthrough at a Short UGA Stop Codon Context Triggers Dual Localization of Metabolic Enzymes in Fungi and Animals.” Journal Article, Research Support, Non-U.S. Gov’t. *PLoS Genetics* 10 (10). United States: e1004685. doi:10.1371/journal.pgen.1004685.

Stone, A B. 1974. “A Simplified Method for Preparing Sucrose Gradients.” Comparative Study, Journal Article. *The Biochemical Journal* 137 (1). England: 117–18.

Stoneley, M, F E Paulin, J P Le Quesne, S A Chappell, and A E Willis. 1998. “C-Myc 5’ Untranslated Region Contains an Internal Ribosome Entry Segment.” Journal Article, Research Support, Non-U.S. Gov’t. *Oncogene* 16 (3). England: 423–28. doi:10.1038/sj.onc.1201763.

Stroka, D M, T Burkhardt, I Desbaillets, R H Wenger, D A Neil, C Bauer, M Gassmann, and D Candinas. 2001. “HIF-1 Is Expressed in Normoxic Tissue and Displays an Organ-Specific Regulation under Systemic Hypoxia.” Journal Article, Research Support, Non-U.S. Gov’t. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* 15 (13). United States: 2445–53. doi:10.1096/fj.01-0125com.

Stumpf, C R, M V Moreno, A B Olshen, B S Taylor, and D Ruggero. 2013. “The Translational Landscape of the Mammalian Cell Cycle.” Journal Article. *Mol Cell* 52 (4): 574–82. doi:10.1016/j.molcel.2013.09.018.

Swart, Estienne Carl, Valentina Serra, Giulio Petroni, and Mariusz Nowacki. 2016. “Genetic Codes with No Dedicated Stop Codon: Context-Dependent Translation Termination.” Journal Article, Research Support, Non-U.S. Gov’t. *Cell* 166 (3). United States: 691–702. doi:10.1016/j.cell.2016.06.020.

Tarun, S Z Jr, and A B Sachs. 1995. “A Common Function for mRNA 5’ and 3’ Ends in Translation Initiation in Yeast.” Journal Article, Research Support, Non-U.S. Gov’t, Research Support, U.S. Gov’t, P.H.S. *Genes & Development* 9 (23). United States: 2997–3007.

Terenin, I M, K A Akulich, D E Andreev, S A Polyanskaya, I N Shatsky, and S E Dmitriev. 2016. “Sliding of a 43S Ribosomal Complex from the Recognized AUG Codon Triggered by a Delay in eIF2-Bound GTP Hydrolysis.” Journal Article. *Nucleic Acids Res* 44 (4): 1882–93. doi:10.1093/nar/gkv1514.

Terenin, Ilya M, Sergey E Dmitriev, Dmitry E Andreev, and Ivan N Shatsky. 2008. “Eukaryotic Translation Initiation Machinery Can Operate in a Bacterial-like Mode without eIF2.” Journal Article, Research Support, Non-U.S. Gov’t. *Nature Structural & Molecular Biology* 15 (8). United States: 836–41. doi:10.1038/nsmb.1445.

Thakor, Nehal, and Martin Holcik. 2012. “IRES-Mediated Translation of Cellular Messenger RNA Operates in eIF2alpha- Independent Manner during Stress.” Journal Article, Research Support, Non-U.S. Gov’t. *Nucleic Acids Research* 40 (2). England: 541–52. doi:10.1093/nar/gkr701.

Tholstrup, Jesper, Lene B Oddershede, and Michael A Sorensen. 2012. “mRNA Pseudoknot Structures Can Act as Ribosomal Roadblocks.” Journal Article, Research Support, Non-U.S. Gov’t. *Nucleic Acids Research* 40 (1). England: 303–13. doi:10.1093/nar/gkr686.

Thompson, Mary K, Maria F Rojas-Duran, Paritosh Gangaramani, and Wendy V Gilbert. 2016. “The Ribosomal Protein Asc1/RACK1 Is Required for Efficient Translation of Short mRNAs.” Journal Article. *eLife* 5 (April). England. doi:10.7554/eLife.11154.

Thoreen, C C. 2013. “Many Roads from mTOR to eIF4F.” Journal Article. *Biochem Soc Trans* 41 (4): 913–16. doi:10.1042/BST20130082.

Thoreen, C C, L Chantranupong, H R Keys, T Wang, N S Gray, and D M Sabatini. 2012. “A Unifying Model for mTORC1-Mediated Regulation of mRNA Translation.” Journal Article. *Nature* 485 (7396): 109–13. doi:10.1038/nature11083.

Topisirovic, I, and K L Borden. 2005. “Homeodomain Proteins and Eukaryotic Translation Initiation Factor 4E (eIF4E): An Unexpected Relationship.” Journal Article. *Histol Histopathol* 20 (4): 1275–84. http://www.ncbi.nlm.nih.gov/pubmed/16136508.

Tsang, C K, H Liu, and X F Zheng. 2010. “mTOR Binds to the Promoters of RNA Polymerase I- and III-Transcribed Genes.” Journal Article. *Cell Cycle* 9 (5): 953–57. doi:10.4161/cc.9.5.10876.

Tuller, T, A Carmi, K Vestsigian, S Navon, Y Dorfan, J Zaborske, T Pan, O Dahan, I Furman, and Y Pilpel. 2010. “An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation.” Journal Article. *Cell* 141 (2): 344–54. doi:10.1016/j.cell.2010.03.031.

Tuller, Tamir, Isana Veksler-Lublinsky, Nir Gazit, Martin Kupiec, Eytan Ruppin, and Michal Ziv-Ukelson. 2011. “Composite Effects of Gene Determinants on the Translation Speed and Density of Ribosomes.” Journal Article, Research Support, Non-U.S. Gov’t. *Genome*

Biology 12 (11). England: R110. doi:10.1186/gb-2011-12-11-r110.

Tuller, Tamir, Yedael Y Waldman, Martin Kupiec, and Eytan Ruppin. 2010. “Translation Efficiency Is Determined by Both Codon Bias and Folding Energy.” Journal Article, Research Support, Non-U.S. Gov’t. *Proceedings of the National Academy of Sciences of the United States of America* 107 (8). United States: 3645–50. doi:10.1073/pnas.0909910107.

Uniacke, James, Chet E Holterman, Gabriel Lachance, Aleksandra Franovic, Mathieu D Jacob, Marc R Fabian, Josianne Payette, Martin Holcik, Arnim Pause, and Stephen Lee. 2012. “An Oxygen-Regulated Switch in the Protein Synthesis Machinery.” Journal Article, Research Support, Non-U.S. Gov’t. *Nature* 486 (7401). England: 126–29. doi:10.1038/nature11055.

Vanderperre, Benoit, Jean-Francois Lucier, Cyntia Bissonnette, Julie Motard, Guillaume Tremblay, Solene Vanderperre, Maxence Wisztorski, Michel Salzet, Francois-Michel Boisvert, and Xavier Roucou. 2013. “Direct Detection of Alternative Open Reading Frames Translation Products in Human Significantly Expands the Proteome.” Journal Article, Research Support, Non-U.S. Gov’t. *PloS One* 8 (8). United States: e70698. doi:10.1371/journal.pone.0070698.

Vattem, Krishna M, and Ronald C Wek. 2004. “Reinitiation Involving Upstream ORFs Regulates ATF4 mRNA Translation in Mammalian Cells.” Journal Article, Research Support, U.S. Gov’t, P.H.S. *Proceedings of the National Academy of Sciences of the United States of America* 101 (31). United States: 11269–74. doi:10.1073/pnas.0400541101.

Wan, Ji, and Shu Bing Qian. 2014. “TISdb: A Database for Alternative Translation Initiation in Mammalian Cells.” Journal Article. *Nucleic Acids Res* 42 (Database issue): D845-50. doi:10.1093/nar/gkt1085.

Wang, H, J McManus, and C Kingsford. 2016. “Isoform-Level Ribosome Occupancy Estimation Guided by Transcript Abundance with Ribomap.” Journal Article. *Bioinformatics* 32 (12): 1880–82. doi:10.1093/bioinformatics/btw085.

Wang, Ruanlin, Zhiyun Zhang, Jun Du, Yuejun Fu, and Aihua Liang. 2016. “Large-Scale Mass Spectrometry-Based Analysis of Euplotes Octocarinatus Supports the High Frequency of +1 Programmed Ribosomal Frameshift.” Journal Article. *Scientific Reports* 6 (September). England: 33020. doi:10.1038/srep33020.

Wang, X, and C G Proud. 2006. “The mTOR Pathway in the Control of Protein Synthesis.” Journal Article. *Physiology (Bethesda)* 21: 362–69. doi:10.1152/physiol.00024.2006.

Watatani, Yujiro, Kenji Ichikawa, Noriko Nakanishi, Maki Fujimoto, Hitoshi Takeda, Natsumi Kimura, Hidenori Hirose, Shigeru Takahashi, and Yuji Takahashi. 2008. “Stress-Induced Translation of ATF5 mRNA Is Regulated by the 5’-untranslated Region.” Journal Article, Research Support, Non-U.S. Gov’t. *The Journal of Biological Chemistry* 283 (5). United States: 2543–53. doi:10.1074/jbc.M707781200.

Weinberg, David E, Premal Shah, Stephen W Eichhorn, Jeffrey A Hussmann, Joshua B Plotkin, and David P Bartel. 2016. “Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation.” Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov’t, Research Support, U.S. Gov’t, Non-P.H.S. *Cell Reports* 14 (7). United States: 1787–99. doi:10.1016/j.celrep.2016.01.043.

Weiss, R B, D M Dunn, J F Atkins, and R F Gesteland. 1987. “Slippery Runs, Shifty Stops, Backward Steps, and Forward Hops: -2, -1, +1, +2, +5, and +6 Ribosomal Frameshifting.” *Cold Spring Harbor Symposia on Quantitative Biology* 52: 687–93. http://www.ncbi.nlm.nih.gov/pubmed/3135981.

Weiss, R B, D M Dunn, A E Dahlberg, J F Atkins, and R F Gesteland. 1988. “Reading Frame Switch Caused by Base-Pair Formation between the 3’ End of 16S rRNA and the mRNA during Elongation of Protein Synthesis in Escherichia Coli.” *The EMBO Journal* 7 (5): 1503–7. http://www.ncbi.nlm.nih.gov/pubmed/2457498.

Wen, Jin-Der, Laura Lancaster, Courtney Hodges, Ana-Carolina Zeri, Shige H Yoshimura, Harry F Noller, Carlos Bustamante, and Ignacio Tinoco. 2008. “Following Translation by Single Ribosomes One Codon at a Time.” *Nature* 452 (7187): 598–603. doi:10.1038/nature06716.

WHO. 2014. “The Top 10 Causes of Death (Fact Sheet N°310).” *2014*. Accessed January 1. http://who.int/mediacentre/factsheets/fs310/en/.

Wiita, Arun P, Etay Ziv, Paul J Wiita, Anatoly Urisman, Olivier Julien, Alma L Burlingame, Jonathan S Weissman, and James A Wells. 2013. “Global Cellular Response to Chemotherapy-Induced Apoptosis.” *eLife* 2 (October). eLife Sciences Publications Limited: e01236. doi:10.7554/eLife.01236.

Williams, C C, C H Jan, and J S Weissman. 2014. “Targeting and Plasticity of Mitochondrial

Proteins Revealed by Proximity-Specific Ribosome Profiling." Journal Article. *Science* 346 (6210): 748–51. doi:10.1126/science.1257522.

Wilson, J E, T V Pestova, C U Hellen, and P Sarnow. 2000. "Initiation of Protein Synthesis from the A Site of the Ribosome." Journal Article, Research Support, Non-U.S. Gov't, Research Support, U.S. Gov't, P.H.S. *Cell* 102 (4). United States: 511–20.

Wolfe, A L, K Singh, Y Zhong, P Drewe, V K Rajasekhar, V R Sanghvi, K J Mavrakis, et al. 2014. "RNA G-Quadruplexes Cause eIF4A-Dependent Oncogene Translation in Cancer." Journal Article. *Nature* 513 (7516): 65–70. doi:10.1038/nature13485.

Wolin, S L, and P Walter. 1988. "Ribosome Pausing and Stacking during Translation of a Eukaryotic mRNA." Journal Article. *EMBO J* 7 (11): 3559–69. http://www.ncbi.nlm.nih.gov/pubmed/2850168.

Woolstenhulme, Christopher J, Shankar Parajuli, David W Healey, Diana P Valverde, E Nicholas Petersen, Agata L Starosta, Nicholas R Guydosh, W Evan Johnson, Daniel N Wilson, and Allen R Buskirk. 2013. "Nascent Peptides That Block Protein Synthesis in Bacteria." Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't. *Proceedings of the National Academy of Sciences of the United States of America* 110 (10). United States: E878-87. doi:10.1073/pnas.1219536110.

Xiao, Z, Q Zou, Y Liu, and X Yang. 2016. "Genome-Wide Assessment of Differential Translations with Ribosome Profiling Data." Journal Article. *Nat Commun* 7: 11194. doi:10.1038/ncomms11194.

Xie, S Q, P Nie, Y Wang, H Wang, H Li, Z Yang, Y Liu, J Ren, and Z Xie. 2016. "RPFdb: A Database for Genome Wide Information of Translated mRNA Generated from Ribosome Profiling." Journal Article. *Nucleic Acids Res* 44 (D1): D254-8. doi:10.1093/nar/gkv972.

Xing, Changhong, Ken Arai, Eng H Lo, and Marc Hommel. 2012. "Pathophysiologic Cascades in Ischemic Stroke." Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't, Review. *International Journal of Stroke : Official Journal of the International Stroke Society* 7 (5). United States: 378–85. doi:10.1111/j.1747-4949.2012.00839.x.

Yang, H S, A P Jansen, A A Komar, X Zheng, W C Merrick, S Costes, S J Lockett, N Sonenberg, and N H Colburn. 2003. "The Transformation Suppressor Pdcd4 Is a Novel Eukaryotic Translation Initiation Factor 4A Binding Protein That Inhibits Translation." Journal

Article. *Mol Cell Biol* 23 (1): 26–37. http://www.ncbi.nlm.nih.gov/pubmed/12482958.

Yang, Jian-Rong, Xiaoshu Chen, and Jianzhi Zhang. 2014. "Codon-by-Codon Modulation of Translational Speed and Accuracy via mRNA Folding." Journal Article, Research Support, N.I.H., Extramural. *PLoS Biology* 12 (7). United States: e1001910. doi:10.1371/journal.pbio.1001910.

Yoshida, H, T Matsui, A Yamamoto, T Okada, and K Mori. 2001. "XBP1 mRNA Is Induced by ATF6 and Spliced by IRE1 in Response to ER Stress to Produce a Highly Active Transcription Factor." Journal Article. *Cell* 107 (7): 881–91. http://www.ncbi.nlm.nih.gov/pubmed/11779464.

Youle, Richard J, and Alexander M van der Bliek. 2012. "Mitochondrial Fission, Fusion, and Stress." JOUR. *Science* 337 (6098): 1062 LP-1065. http://science.sciencemag.org/content/337/6098/1062.abstract.

Young, David J, Nicholas R Guydosh, Fan Zhang, Alan G Hinnebusch, and Rachel Green. 2015. "Rli1/ABCE1 Recycles Terminating Ribosomes and Controls Translation Reinitiation in 3'UTRs In Vivo." Journal Article, Research Support, N.I.H., Intramural, Research Support, Non-U.S. Gov't. *Cell* 162 (4). United States: 872–84. doi:10.1016/j.cell.2015.07.041.

Zach, L, I Braunstein, and A Stanhill. 2014. "Stress-Induced Start Codon Fidelity Regulates Arsenite-Inducible Regulatory Particle-Associated Protein (AIRAP) Translation." Journal Article. *J Biol Chem* 289 (30): 20706–16. doi:10.1074/jbc.M114.547828.

Zhao, Chenyang, Shyamasree Datta, Palash Mandal, Shuqing Xu, and Thomas Hamilton. 2010. "Stress-Sensitive Regulation of IFRD1 mRNA Decay Is Mediated by an Upstream Open Reading Frame." Journal Article, Research Support, N.I.H., Extramural. *The Journal of Biological Chemistry* 285 (12). United States: 8552–62. doi:10.1074/jbc.M109.070920.

Zhdanov, Alexander V, Ruslan I Dmitriev, Anna V Golubeva, Svetlana A Gavrilova, and Dmitri B Papkovsky. 2013. "Chronic Hypoxia Leads to a Glycolytic Phenotype and Suppressed HIF-2 Signaling in PC12 Cells." *Biochimica et Biophysica Acta* 1830 (6): 3553–69. doi:10.1016/j.bbagen.2013.02.016.

Zhdanov, Alexander V, Alicia H C Waters, Anna V Golubeva, and Dmitri B Papkovsky. 2015. "Differential Contribution of Key Metabolic Substrates and Cellular Oxygen in HIF Signalling." Journal Article, Research Support, Non-U.S. Gov't. *Experimental Cell*

Research 330 (1). United States: 13–28. doi:10.1016/j.yexcr.2014.10.005.

Zhong, Y, T Karaletsos, P Drewe, V Sreedharan, D Kuo, K Singh, H G Wendel, and G Ratsch. 2016. “RiboDiff: Detecting Changes of mRNA Translation Efficiency from Ribosome Footprints.” Journal Article. *Bioinformatics*. doi:10.1093/bioinformatics/btw585.

Zhou, Donghui, L Reddy Palam, Li Jiang, Jana Narasimhan, Kirk A Staschke, and Ronald C Wek. 2008. “Phosphorylation of eIF2 Directs ATF5 Translational Control in Response to Diverse Stress Conditions.” Journal Article, Research Support, N.I.H., Extramural. *The Journal of Biological Chemistry* 283 (11). United States: 7064–73. doi:10.1074/jbc.M708530200.

Zong, Q, M Schummer, L Hood, and D R Morris. 1999. “Messenger RNA Translation State: The Second Dimension of High-Throughput Expression Screening.” Journal Article. *Proc Natl Acad Sci U S A* 96 (19): 10632–36. http://www.ncbi.nlm.nih.gov/pubmed/10485877.

Appendices

RiboGalaxy: a platform for the alignment, analysis and visualization of ribo-seq data.

This appendix has being published as a research article RNA Biol. 2016;13(3):316-9. doi: 10.1080/15476286.2016.1141862.

ABSTRACT

Ribosome profiling (ribo-seq) is a technique that uses high-throughput sequencing to reveal the exact locations and densities of translating ribosomes at the entire transcriptome level. The technique has become very popular since its inception in 2009. Yet experimentalists who generate ribo-seq data often have to rely on bioinformaticians to process and analyse their data. We present RiboGalaxy (http://ribogalaxy.ucc.ie), a freely available Galaxy-based web server for processing and analysing ribosome profiling data with the visualization functionality provided by GWIPS-viz (http://gwips.ucc.ie). RiboGalaxy provides researchers with a compact suite of tools specifically tailored for processing ribo-seq and corresponding mRNA-seq data. Researchers can take advantage of the published workflows which reduce the multi-step alignment process to a minimum of inputs from the user. Users can then explore their own aligned data as a custom track in GWIPS-viz and compare their ribosome profiles to existing ribo-seq tracks from published studies. In addition, users can assess the quality of their ribo-seq data, determine the strength of the triplet periodicity signal, generate meta-gene and (sub-codon) ribosome profiles as well as analyse the relative impact of mRNA sequence features on local decoding rates. RiboGalaxy is accompanied by extensive documentation and tips for helping users. In addition we provide a RiboGalaxy forum (http://gwips.ucc.ie/Forum) where we encourage users to post their questions and feedback to improve the overall RiboGalaxy service.

INTRODUCTION

Ribosome profiling, also known as ribo-seq, is a high through-put technique where mRNA fragments protected by ribosomes are isolated and sequenced [1]. This allows the measurement of ribosome density along all mRNA transcripts present in the cell providing genome-wide information on protein synthesis (GWIPS) *in vivo* [2]. The usages and adaptations of the technique have grown rapidly in recent years (see reviews by [3,4]). The

pre-processing and alignment of ribo-seq data requires tools which are primarily designed for command-line usage. To visually explore the ribo-seq alignment profiles other software may be required. If the researcher wishes to compare their ribo-seq alignments to other published ribo-seq data, they would need to download, pre-process and align these data as well. In addition, researchers may wish to determine the strength of the triplet periodicity signal which is particular to ribo-seq data and use the periodicity signature for framing analysis [5–7]. However, researchers who do not have a bioinformatics background need to invest time to learn how to install and use the appropriate software. Furthermore, researchers require access to a robust computational infrastructure and a relatively large storage infrastructure, facilities that are not always available to experimentalists. Undoubtedly all of these challenges can be overcome. But for time-poor researchers who do not have access to computational resources or a bioinformatician, these challenges can be discouraging. To address these challenges, we have developed RiboGalaxy, a Galaxy-based web server [8] where researchers can align, analyse and visually explore their ribo-seq data and corresponding mRNA-seq data without the need to install software tools or build alignment pipelines.

RESULTS

As well as providing free computational infrastructure to researchers, RiboGalaxy allows users to explore their ribo-seq data without the need for command-line tools or the need for a bioinformatics background. Usage of RiboGalaxy does not require user registration. However, to take advantage of certain features such as the published workflows and pages, or upload large datasets using FTP, users will need to create a RiboGalaxy account (an easy procedure where the user only has to provide a username and password).

Using RiboGalaxy

The typical ribo-seq alignment pipeline requires pre-processing the sequence reads to remove adaptor sequences and filter out any sequence reads that correspond to ribosomal RNA (rRNA). The *Pre-processing tool* suite on RiboGalaxy hosts Cutadapt [9] for adaptor sequence removal and bowtie [10] for rRNA removal. RiboGalaxy currently provides pre-built rRNA indices for 10 model organisms saving the researcher the necessity to search for the sequences in public databases. More rRNA indices will be added over time. As

ribosome footprints are derived from mRNA, alignment to a transcriptome is usually appropriate. However, there are cases where alignment to a genome is of interest, for example if a transcriptome is not well characterized. We provide both options on RiboGalaxy. The *Align to transcriptome using Bowtie* option is available under the *Transcriptome Mapping* suite. Note that for alignments to a transcriptome, a reference transcriptome FASTA file is required. One option is to obtain a reference FASTA file from the *UCSC Main table browser* option in *Get Data* (we explain how to do this on the RiboGalaxy Help page). The *Align to genome using Bowtie* is available under the *GWIPS-viz Mapping* suite. Pre-built indices are currently provided for 13 reference genomes. Genomic ribosome profiles can be created for either RNase I or micrococcal nuclease (MNase) generated data and profiles for the entire genome can be viewed directly as a custom track in GWIPS-viz[11].

If the researcher wishes to generate mRNA (sub-codon) ribosome profiles, we recommend mapping with bowtie[10] in the *Transcriptome Mapping* suite followed by *riboplot* (see Fig. 1A). Running the r*ibocount* tool will provide sub-codon profile counts in spreadsheet format for all transcripts to which ribosome footprints reads were mapped. Triplet periodicity and metagene analysis can also be carried out using the *riboSeqR* suite of tools [12] (see Fig. 1B and 1C for examples). The RUST[13] suite of tools allows researchers to check the quality of their ribo-seq data as well as determine the relative impact of mRNA features such as codons ,amino acids, dipeptides and tripeptides on the local ribosome profiling read density (see Fig. 1D and 1E). The RiboTools suite[14] provides functionality for exploring stop codon readthrough events (Fig. 1F) and translation in alternative reading frames. Differential translation expression analysis can be carried out on ribo-seq and corresponding mRNA-seq data using an adaptation of the baySeq[15] tool as part of the riboSeqR package.

We provide *Published pages* (available under the *Shared Data* tab) to illustrate how to use the tools hosted on RiboGalaxy. Datasets from the Andreev *et al*. study[16] were used for this purpose and the raw ribo-seq FASTQ files are available as *Data Libraries* under the *Shared Data* tab for testing purposes.

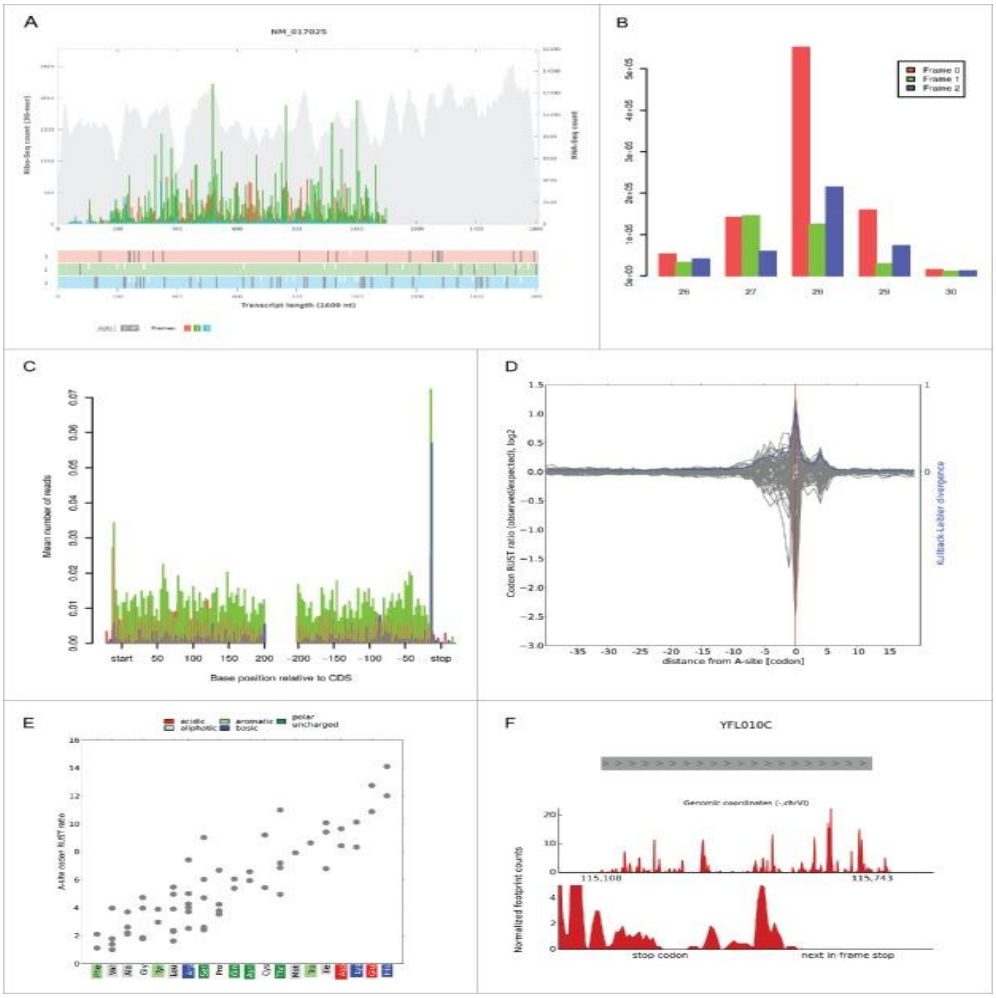


Figure 1. **A.** A sub-codon ribosome profile along with the open reading frame (ORF) organisation for the rat *Ldha* gene. The footprint reads are colour coded (red, green, blue) according to the sub-codon position alignment (1, 2, 3). The background grey alignments represent mRNA-seq data for the corresponding transcript. In this profile, the majority of the footprint reads are green indicating that they originate from an ORF in the second reading frame. **B**. A triplet periodicity plot showing the sub-codon positions of ribo-seq alignments of read lengths 26-30nt to the protein coding regions (CDS) of the zebrafish transcriptome. **C.** A metagene profile of ribosome density relative to the annotated start and stop codons in the zebrafish transcriptome. **D.** A RUST metafootprint profile that reveals the influence of mRNA codons on the relative read density across the entire ribosome and nascent peptide region. The Kullback-Leibler divergence across these sites is also provided. **E.** The RUST ratio for the predicted A-site for individual codons grouped by amino acid. **F.** A ribo-seq profile for the yeast *WWM1* gene (YFL010C) showing the alignment of ribosome footprints downstream of the annotated stop codon. Panel A was generated using the RiboPlot suite on RiboGalaxy using data from the Andreev DE *et al*.[21] study. Panels B and C were generated using the riboSeqR suite[12] on RiboGalaxy using ribo-seq data from the Bazzini AA *et al*.[7] study. Panels C and D were generated using the RUST suite[13] on RiboGalaxy using data from the Andreev DE *et al*.[16] study. Panel F was generated using the RiboTools suite[14] on RiboGalaxy using ribo-seq data from the Baudin-Baillieu A *et al*.[22] study

Using the RiboGalaxy published workflows

In RiboGalaxy we provide *Published workflows* for genome mapping (available under the *Shared Data* tab). Apart from saving time, these published workflows should particularly help researchers who are not familiar with alignment pipelines for ribo-seq data, or who are not familiar with the usage of the individual tools. If the researcher wishes to explore their ribo-seq data as a custom track in GWIPS-viz, with the advantage of comparing their alignment profiles to publicly available ribo-seq and mRNA-seq tracks, then the easiest option is to use the corresponding genome workflow. This will reduce the multi-step pre-processing and alignment process to a minimum of inputs from the user: only the adaptor sequence and the name of the custom track for visualization in GWIPS-viz need to be specified. We currently provide published workflows for the following 11 genome assemblies: *Homo sapiens* (hg19), *Mus musculus* (mm10), *Danio rerio* (danRer7), *Caenorhabditis elegans* (ce10), *Arabidopsis thaliana* (araTha1), *Drosophila melanogaster* (dm3), *Saccharomyces cerevisiae* (sacCer3), *Escherichia coli* (eschcoli_k12), *Bacillus subtilis* (baciSubt2), *Human herpesvirus 5 strain Merlin* (HHV5) and *Bacteriophage lambda* (NC_001416).

On successful completion of an in-built workflow, an email notification is sent to the user. The custom track is visible only to the researcher and not to other public users of GWIPS-viz. All of the alignment results, coverage and profile information can be downloaded from RiboGalaxy while snapshot images of the ribosome profiles can be generated in GWIPS-viz.

We do not currently provide published workflows for transcriptome alignments. The reason is that reference data base annotations are updated regularly and it is advisable to obtain the latest reference annotations. We explain on our Help page how to get annotations in FASTA file format using the *UCSC Main Table Browser* [17] from within RiboGalaxy.

RiboGalaxy Help pages and Forum

RiboGalaxy is publicly available from http://riboseq.org/ and is accompanied by extensive documentation and tips for helping users. We provide help pages specifically for RiboGalaxy usage under the Help section. For each individual tool, we provide short Tips at the top of each tool page to help the user get started as well as more extensive help

support on each tool page. In addition we provide a RiboGalaxy forum (http://gwips.ucc.ie/Forum) where we encourage users to post their questions and feedback to improve the overall RiboGalaxy service.

DISCUSSION

Up to recently, software for the analysis of ribo-seq data were not publicly available. Now, however, several tools have been developed specifically for ribo-seq. We plan to expand the existing repertoire of tools on RiboGalaxy by integrating such tools (e.g. riboDiff[18]) as well as developing more software for the analysis of ribo-seq data.

MATERIALS AND METHODS

RiboGalaxy uses the Galaxy[8] framework for the pre-processing, alignment and analysis pipelines. GWIPS-viz[11] is used for the ribo-seq data visualization as the alignment profiles can be explored in conjunction with publicly available ribo-seq and RNA-seq tracks. Galaxy Trackster[19] is also available for data visualization.

RiboGalaxy runs on Ubuntu 14.04.2 LTS, with Apache 2.4.7 and PostgreSQL 9.3.6. The version of each of the tools hosted on RiboGalaxy is provided at the top of the page when you click on the tool.

As RiboGalaxy is a platform dedicated to ribo-seq data analysis, and not a general Galaxy server platform, only tools that are related to ribo-seq data alignment pipelines and analysis are hosted. The pre-processing and alignment software were downloaded from the Galaxy toolshed[20]. In house python scripts are used for developing the RiboPlot suite and well as for generating ribosome profiles for visualization in GWIPS-viz while xml wrappers were written for the integration of the riboSeqR[12] and the RUST[13] suites of tools. The RiboGalaxy implementations of riboSeqR and RUST are now available in the Galaxy toolshed.

REFERENCES

1. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science (80-) 2009; 324:218–23.
2. Weiss RB, Atkins JF. Translation Goes Global. Science (80-) 2011; 334:1509–10.

3. Michel AM, Baranov P V. Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. Wiley Interdiscip Rev RNA 2013; 4:473–90.
4. Ingolia NT. Ribosome profiling: new views of translation, from single codons to genome scale. Nat Rev Genet 2014; 15:205–13.
5. Michel AM, Choudhury KR, Firth AE, Ingolia NT, Atkins JF, Baranov P V. Observation of dually decoded regions of the human genome using ribosome profiling data. Genome Res 2012; 22(11):2219–29.
6. Gerashchenko M V, Lobanov A V, Gladyshev VN. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. Proc Natl Acad Sci 2012; 109:17394–9.
7. Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J 2014; 33:981–93.
8. Blankenberg D, Taylor J, Nekrutenko A. Online resources for genomic analysis using high-throughput sequencing. Cold Spring Harb Protoc 2015; 2015:pdb.top083667.
9. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 2011; 17:10.
10. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009; 10:R25.
11. Michel AM, Fox G, M Kiran A, De Bo C, O'Connor PBF, Heaphy SM, Mullan JPA, Donohue CA, Higgins DG, Baranov P V. GWIPS-viz: development of a ribo-seq genome browser. Nucleic Acids Res 2014; 42:D859–64.
12. Chung BY, Hardcastle TJ, Jones JD, Irigoyen N, Firth AE, Baulcombe DC, Brierley IAN. The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. RNA 2015; :1–15.
13. O'Connor PB, Andreev DE, Baranov P V. Surveying the relative impact of mRNA features on local ribosome profiling read density in 28 datasets. bioRxiv doi http//dx.doi.org/10.1101/0187622015;
14. Legendre R, Baudin-Baillieu a., Hatin I, Namy O. RiboTools: A Galaxy toolbox for qualitative ribosome profiling analysis. Bioinformatics 2015; :1–3.

15. Hardcastle TJ, Kelly K a. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 2010; 11:422.
16. Andreev DE, O'Connor PB, Fahey C, Kenny EM, Terenin IM, Dmitriev SE, Cormican P, Morris DW, Shatsky IN, Baranov P V. Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. Elife 2015; 4:1–21.
17. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, et al. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res 2014; 43:D670–81.
18. Zhong Y, Karaletsos T, Drewe P, Sreedharan V, Wendel H, Gunnar R. RiboDiff: Detecting Changes of Translation Efficiency from Ribosome Footprints. bioRxiv doi http//dx.doi.org/101101/017111 2015;
19. Goecks J, Coraor N, Nekrutenko A, Taylor J. NGS analyses by visualization with Trackster. Nat Biotechnol 2012; 30:1036–9.
20. Blankenberg D, Von Kuster G, Bouvier E, Baker D, Afgan E, Stoler N, Taylor J, Nekrutenko A. Dissemination of scientific software with Galaxy ToolShed. Genome Biol 2014; 15:403.
21. Andreev DE, O'Connor PB, Zhdanov A V, Dmitriev RI, Shatsky IN, Papkovsky DB, Baranov P V. Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. Genome Biol 2015; 16:1–14.
22. Baudin-Baillieu A, Legendre R, Kuchly C, Hatin I, Demais S, Mestdagh C, Gautheret D, Namy O. Genome-wide Translational Changes Induced by the Prion [PSI+]. Cell Rep 2014; 8:439–48.

GWIPS-viz: Development of a ribo-seq genome browser

This appendix has being published as a research article Nucleic Acids Res. 2014 Jan 1; 42(Database issue): D859–D864. doi: 10.1093/nar/gkt1035

ABSTRACT

Ribosome profiling (ribo-seq) is a recently developed technique that provides Genome Wide Information on Protein Synthesis (GWIPS) in vivo. It is based on the deep sequencing of ribosome protected mRNA fragments which allows the ribosome density along all mRNA transcripts present in the cell to be quantified. Since its inception, ribo-seq has been carried out on a number of eukaryotic and prokaryotic organisms. Due to the increasing interest in ribo-seq, there is a pertinent demand for a dedicated ribo-seq genome browser. Therefore we have developed GWIPS-viz, an online genome browser for viewing ribosome profiling data. GWIPS-viz is based on the UCSC Genome Browser. Ribo-seq tracks coupled with mRNA-seq tracks are currently available for several genomes: Human, Mouse, Zebrafish, Nematode, Yeast, Bacteria (*Escherichia coli* K12, *Bacillus subtilis*), Human Cytomegalovirus and Bacteriophage lambda. Our objective is to continue incorporating published ribo-seq datasets so that the wider community can readily view ribosome profiling information without the need to carry out computational processing.

Database URL: http://gwips.ucc.ie

INTRODUCTION

Ribosome profiling is based on the isolation of mRNA fragments protected by ribosomes followed by massively parallel sequencing of the protected fragments or footprints. This allows the measurement of ribosome density along all mRNA transcripts present in the cell providing genome-wide information on protein synthesis (GWIPS) *in vivo* (1). The ribosome profiling technique, also known as ribo-seq, was first carried out in *Saccharomyces cerevisiae* (2). Since the original publication, the technique has been carried out in many organisms including *Homo sapiens* (3-10) *Mus musculus* (3,7,9,11,12) *Danio rerio* (13), *Caenorhabditis elegans* (4,14), *Saccharomyces cerevisiae* (15,16), *Escherichia coli* (17,18), *Bacillus subtilis* (18), human cytomegalovirus (19) and Bacteriophage lambda (20).

To date, there have been two main strategies of ribosome profiling: ribosome profiling of initiating ribosomes and ribosome profiling of elongating ribosomes. For a review on the usages and advantages of each approach, please see (21).

The majority of published studies using ribosome profiling provide the raw sequencing data in NCBI's Sequence Read Archive (SRA)(22). In addition, most published ribosome profiling experiments have corresponding naked mRNA control, where total mRNA is randomly degraded to yield fragments of a size similar to ribosome protected fragments. For simplicity here we refer to it as mRNA-seq. mRNA-seq is carried out under the same experimental conditions. It helps to take into account the differential abundance of mRNA between experimental conditions and to monitor technical biases associated with cDNA libraries generation and sequencing.

Due to the increasing popularity of the ribo-seq technique, the number of ribosome profiling experiments is expected to increase dramatically in the near future. However, the visualization of ribosome profiling data in a browser first requires pre-processing and aligning the raw sequencing reads. As with any type of next-generation sequencing data (NGS), demands are placed on biomedical researchers in terms of time, data storage, computational knowledge and prototyping of computational pipelines (23). Web-based integrative framework tools such as Galaxy (24) provide centralized platforms for researchers to carry out NGS alignment pipelines. However, due to decreasing costs, the coverage depth of ribo-seq and corresponding mRNA-seq data is continually increasing resulting in ever larger datasets. Consequently the computational resources required to process such data and the computer memory required to store such data may not be available to many biologists. Indeed, the time required to download, pre-process and align the raw data may be the most limiting factor of all for time-poor researchers.

To address these issues, we introduce GWIPS-viz (http://gwips.ucc.ie), a free online browser which is pre-populated with published ribo-seq data. The aim of GWIPS-viz is to provide an intuitive graphical interface of translation in the genomes for which ribo-seq data are available. Users can readily view alignments from many of the published ribo-seq studies without the need to carry out any computational processing. GWIPS-viz is based on a customized version of the UCSC Genome Browser (http://genome.ucsc.edu) (25). Ribo-seq tracks, coupled with mRNA-seq tracks, are currently available for Human, Mouse,

Zebrafish, Nematode, Yeast, two bacterial species (*Escherichia coli* K12 and *Bacillus subtilis*) and two viral genomes (Human Cytomegalovirus and Bacteriophage lambda).

USAGE

In GWIPS-viz, users can search for their gene(s) of interest in the genome(s) for which ribo-seq data is available and view a snapshot of the gene's translation under the conditions of the experiment. Ribosome coverage plots (red) and mRNA-seq coverage plots (green) display the number of reads that cover a given genomic coordinate. Figure 1 provides coverage plots for the *S. cerevisiae* genome locus containing *ABP140*, *MET7*, *SSP2*, and *PUS7* (from Ingolia et al. PMID:19213877) and illustrates how differential translation can be viewed in GWIPS-viz.

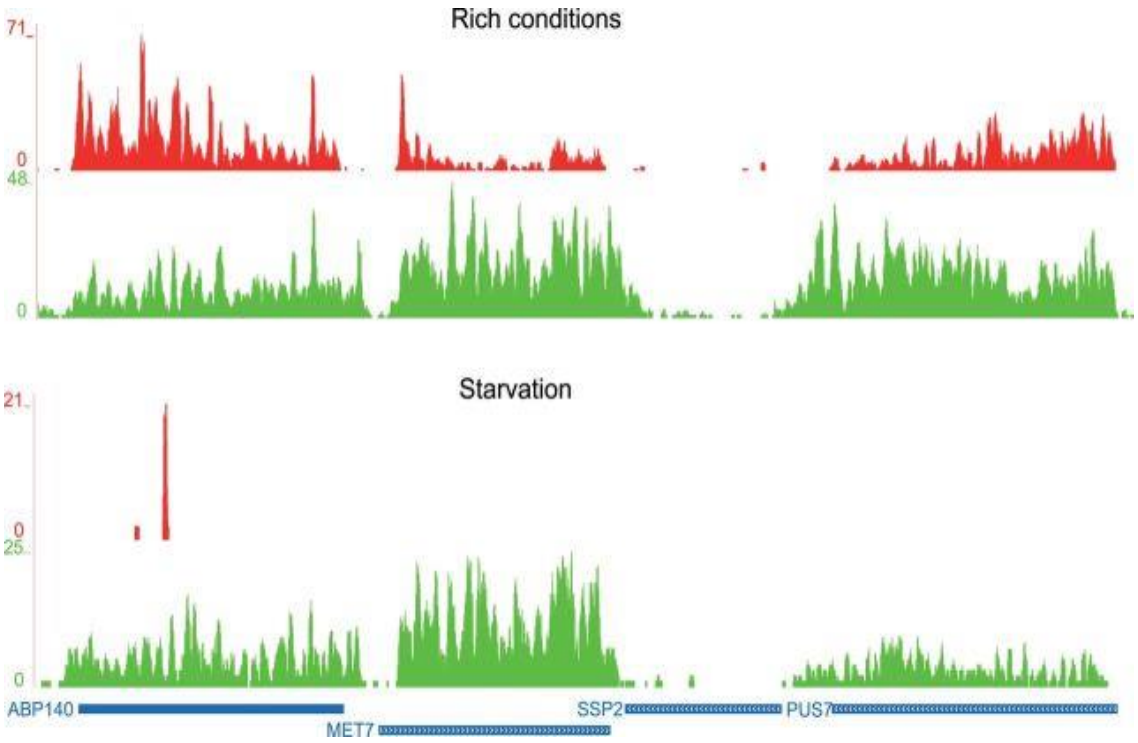


Figure 1. Observing differential translation in GWIPS-viz. Ribo-seq (red) and RNA-seq (green) coverage plots for the *S. cerevisiae* genome locus containing *ABP140*, *MET7*, *SSP2* and *PUS7* genes from (2). Under starvation conditions (right panel), *ABP140*, *MET7* and *PUS7* are transcribed, but not translated.

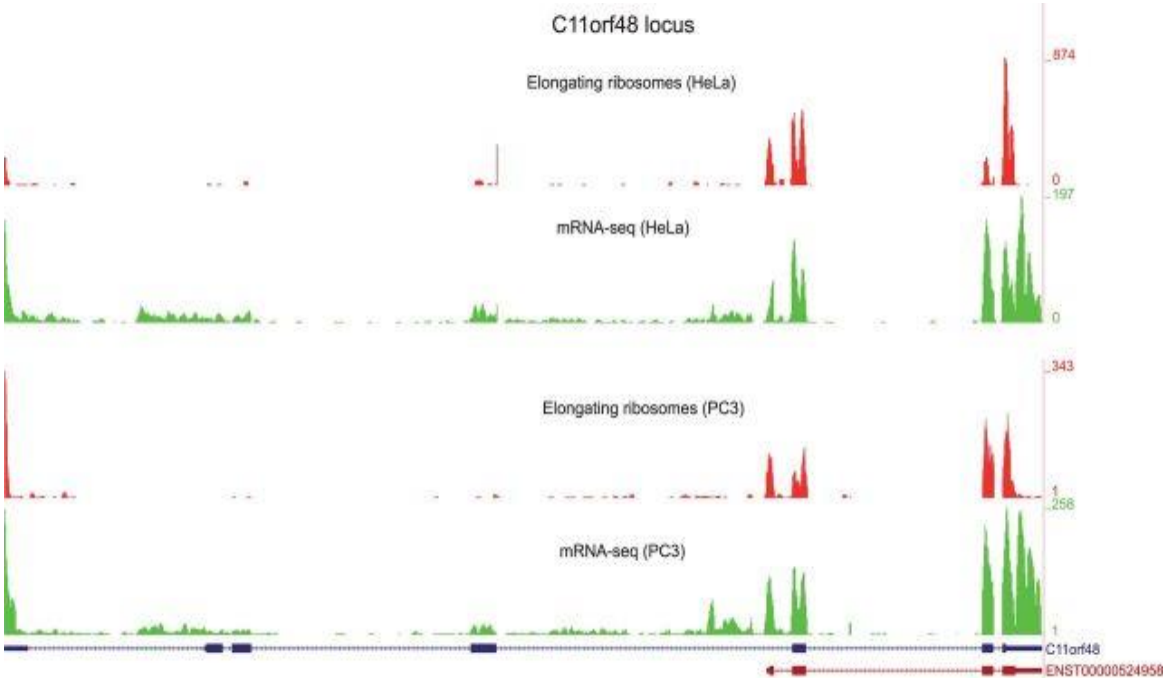


Figure 2. Comparing profiles from independent studies. Data from different studies and different organisms can be compared in GWIPS-viz. The C11orf48 locus in the human genome is shown where translation of an ENSEMBL transcript (brown bars) not annotated in RefSeq (blue bars) has been identified in HeLa cells (26). As can be seen, translation of the Ensembl transcript occurs in both, HeLa (3) and human PC3 cells (6).

Users can visually identify which isoform(s) of a gene is transcribed and translated and also compare translation of the gene between different ribo-seq studies. For example, Figure 2 provides a comparison of two ribo-seq datasets obtained in different tissue-cultured human cells, HeLa (3) and PC3 cells (6). It can be seen that translation of a non-Refseq ENSEMBL transcript, reported based on the analysis of HeLa cell data (26), is observed in both datasets

For the eukaryotic datasets, ribosome profiles display the number of footprint reads at a particular genomic coordinate that align to the A-site (elongating ribosomes) or P-site (initiating ribosomes) of the ribosome, depending on the study. For the prokaryotic datasets, a weighted centred approach (17) is used to indicate the positions of ribosomes. Figure 3 shows ribosome profile densities in a region of the *E. coli* genome that includes the gene *dnaX* (b0470). The ribosome density is scaled relative to the maximum density present within the displayed genomic segment. As a result at the zoom allowing visualization of neighbour genes (top), *dnaX* appears as lowly expressed. However, at a range covering only

the *dnaX* locus, it can be seen that nearly all codons in the dnaX mRNA are covered with footprints. Moreover the coverage is sufficient to allow visual detection of decreased ribosome density downstream of the site of programmed ribosomal frameshifting which is known to causes about 50% of translating ribosomes to terminate prematurely (27,28).

Figure 4 provides an example of how ribo-seq tracks for elongating and initiating ribosomes can be compared. The example illustrates the data obtained in Human HEK293 cells (7) mapped to *TOMM6* and *SFPQ* genes, the latter gene apparently uses two sites of translation initiation for its expression.

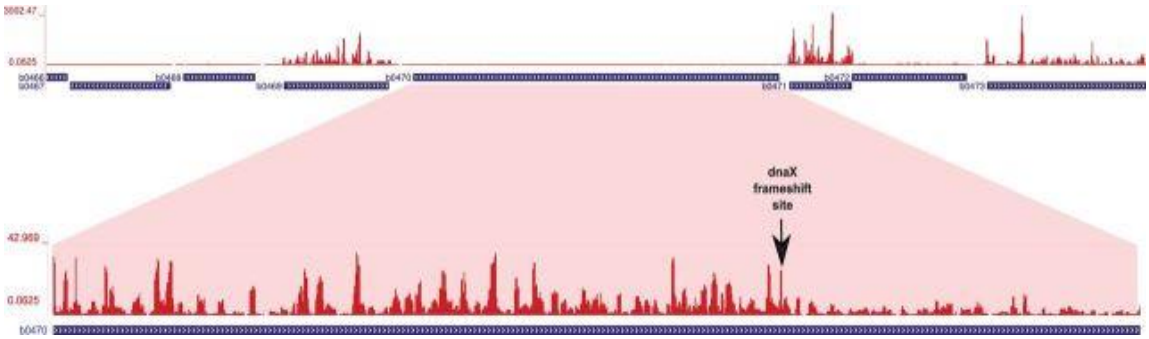


Figure 3. Ribo-seq data for the *dnaX* locus in the *E.coli* genome. The top panel corresponds to a segment containing neighbouring genes. The bottom panel contains the *dnaX* coordinates only. The displayed ribosome density is scaled relative to the maximum density within the selected region. The position of the programmed ribosomal frameshifting site in *dnaX* is indicated with an arrow.

DATABASE DESIGN AND IMPLEMENTATION

GWIPS-viz is a customized version of the UCSC Genome Browser (25) version 269, and runs on Ubuntu Linux version 12.04.1, with Apache version 2.2.22 and MySQL 5.2.24. Static HTML and CSS files of the UCSC Genome Browser were downloaded from http://hgdownload.cse.ucsc.edu/ and rehosted on our local server, while C source code for the CGI executables was downloaded and compiled using gcc 4.6.3. Selected parts of the MySQL databases were synced from the UCSC browser for the majority of organisms included in GWIPS-viz.

Our partial mirror of the UCSC Genome Browser hosted on our server displays tracks for human (hg19), mouse (mm10), *S. cerevisiae* (sacCer3), zebrafish (danRer7), *C. elegans* (ce10), *Escherichia coli* K12 (eschColi_K12), *Bacillus subtilis* (baciSubt2), human cytomegalovirus (Human herpesvirus 5 strain Merlin (HHV5)) and Bacteriophage lambda (NC_001416) assemblies. While several genome assemblies are available for many of the organisms, we chose to include only the most recent genome assembly for each organism.

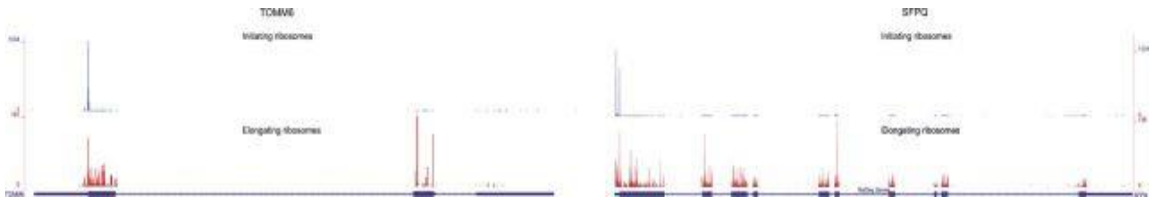


Figure 4. Combining profiles of initiating and elongating ribosomes. Profiles of initiating (blue) and elongating (red) ribosomes generated in Human HEK 293 cells (7). Locations of elongating and initiating ribosomes are consistent with the annotated coding region of the *TOMM6* gene (left). However, ribosome profiles of the SFPQ gene points to the existence of an additional start codon (stronger peak) upstream of the annotated start codon (weaker peak).

Since the goal of GWIPS-viz is to be a browser for ribo-seq data, rather than a mirror of the UCSC browser, some of the functionality of the UCSC browser was removed in order to streamline the interface of GWIPS-viz. For example, the 'clade' menu in the genome selection menu was removed. In the browser window, a link was added in the top bar to allow the user to view the current genome position in the UCSC browser.

Depending on the organism, certain tracks were retained from the UCSC browser (25) and were consolidated into one group called 'Annotation Tracks'. Examples include RefSeq (29), Ensembl (30), CCDS (31), Conservation (32), RepeatMasker (Smit et al., unpublished data, www.repeatmasker.org), Mouse ESTs (33), SGD genes (34), tRNA genes (35).

Ribo-seq and mRNA-seq tracks were added by incorporating the outputs of our RUM (36) alignment pipeline into the MySQL database. These tracks are divided into groups by publication and data type (ribo-seq and mRNA-seq). Tracks generated from uniquely mapping reads are colour coded according to their experiment type (elongating ribosome footprints are red, initiating ribosome footprints are blue, mRNA-seq reads are green).

Raw sequencing data retrieval

Published Ribo-seq and mRNA-seq datasets are downloaded from the NCBI Sequence Read Archive (SRA) (22) and converted to FASTQ format using the fastq-dump utility (SRA Handbook citation, not in PubMed). Data from replicate experiments are consolidated into one dataset so as to have one browser track for each experimental condition. An additional "All" track is generated for each study by aggregating the short reads from all available ribo-seq or mRNA-seq experiments for the given study.

Alignment pipeline

As there are no specific tools as yet for aligning ribo-seq data, RNA-seq tools are used in our pre-processing and alignment pipeline.

Depending on the study, adaptor linker sequence or poly-(A) tails are trimmed from the 3' ends of reads using Cutadapt version 1.1 (37). Trimmed reads shorter than 25 nucleotides are discarded.

Contamination from ribosomal RNA may account for a significant proportion of the raw reads even after depletion by subtractive hybridization during the experiment. Hence it is desirable to remove rRNA reads from the dataset before performing alignments in order to increase the proportion of informative sequences and improve alignment efficiency. To detect reads which are the result of ribosomal RNA contamination, trimmed reads are aligned to rRNA sequences using Bowtie (38). Bowtie version 0.12.8 is run using the -v option allowing three or fewer mismatches between the read sequence and the reference (rRNA) sequence. All reads that align to rRNA are discarded.

In most eukaryotes, a proportion of ribosome footprints will span splice junctions, i.e. the read will span the 3' end of one exon and the 5' end of another. There is the added complexity that ribo-seq reads are typically ~30 nucleotides in length. Hence the short-read alignment program needs to be capable of aligning reads of ~30nt across splice junctions. We use the RNA-seq Unified Mapper (RUM), (current version 2.0.5_05) (36). RUM handles splice junctions by using the short read aligner Bowtie (38) to align sequence reads to both the genome and transcriptome and merging the results, before attempting to map remaining unaligned reads using another existing short-read aligner, BLAT (39).

Due to the relatively short lengths of ribosome footprint reads, a read may align to two or more distinct genomic locations due to sequence similarity. RUM outputs information

separately for uniquely mapping reads and non-uniquely mapping reads (reads which align to several positions in the genome). Currently we provide tracks of uniquely mapping reads only in GWIPS-viz.

RUM's output files include a SAM alignment file showing the alignment(s) for each read, files giving the span of the alignment in genomic coordinates (RUM_Unique and RUM_NU) and coverage files (RUM.cov and RUM_NU.cov) listing the depth of coverage of reads across the genome.

The coverage files generated by the RUM alignment, RUM_Unique.cov and RUM_NU.cov, are in 4 column bedGraph format. The bedGraph data are converted into bigWig format, an indexed binary format that results in higher performance (40).

Ribosome profiles are generated from the RUM_Unique and RUM_NU files by obtaining the number of footprint reads whose 5' ends align at a given genomic coordinate (with an offset of 12nt designating the ribosome P-site for initiating ribosomes or 15nt for the ribosome A-site for elongating ribosomes).

FUTURE PERSPECTIVES

We plan to expand the existing repertoire of ribo-seq tracks by integrating publically available ribosome profiling experiments as they become available.

GWIPS-viz currently displays positions of the ribosomes mapped to the reference genomes. In case of eukaryotic organisms that extensively use RNA splicing, visualization of ribosome positions in GWIPS-viz could be problematic due to a large number of long exons. Therefore, visualization of ribosome positions mapped to individual RNA transcripts is among our top priorities.

We currently provide ribo-seq and mRNA-seq tracks of uniquely mapping reads only. In the future, we wish to provide a differential display that will incorporate non-unique mapping reads (mapping to two or more locations in the genome) with uniquely mapping reads.

We also aim to provide access to the Galaxy platform from within GWIPS-viz so that researchers who generate their own ribo-seq experimental data can pre-process and align their data with the tools provided within Galaxy and then view the alignments in GWIPS-viz.

Our overall objective is to continuously improve the service we provide in GWIPS-viz. As GWIPS-viz is under intensive development, some of the features described in this article could become outdated soon. Hence we encourage users to post their questions, comments and feedback on the GWIPS-viz forum.

REFERENCES

1. Weiss, R.B. and Atkins, J.F. (2011) Molecular biology. Translation goes global. *Science*, **334**, 1509-1510.

2. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. and Weissman, J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218-223.

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

4. Stadler, M. and Fire, A. (2011) Wobble base-pairing slows in vivo translation elongation in metazoans. *RNA*, **17**, 2063-2073.

5. Reid, D.W. and Nicchitta, C.V. (2012) Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. *The Journal of biological chemistry*, **287**, 5518-5527.

6. Hsieh, A.C., Liu, Y., Edlind, M.P., Ingolia, N.T., Janes, M.R., Sher, A., Shi, E.Y., Stumpf, C.R., Christensen, C., Bonham, M.J. *et al.* (2012) The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature*, **485**, 55-61.

7. Lee, S., Liu, B., Huang, S.X., Shen, B. and Qian, S.B. (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E2424-2432.

8. Fritsch, C., Herrmann, A., Nothnagel, M., Szafranski, K., Huse, K., Schumann, F., Schreiber, S., Platzer, M., Krawczak, M., Hampe, J. *et al.* (2012) Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome research*, **22**, 2208-2218.

9. Shalgi, R., Hurt, J.A., Krykbaeva, I., Taipale, M., Lindquist, S. and Burge, C.B. (2013) Widespread regulation of translation by elongation pausing in heat shock. *Molecular cell*, **49**, 439-452.

10. Liu, B., Han, Y. and Qian, S.B. (2013) Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. *Molecular cell*, **49**, 453-463.

11. Ingolia, N.T., Lareau, L.F. and Weissman, J.S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789-802.

12. Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S. and Sabatini, D.M. (2012) A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature*, **485**, 109-113.

13. Bazzini, A.A., Lee, M.T. and Giraldez, A.J. (2012) Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science*, **336**, 233-237.

14. Stadler, M., Artiles, K., Pak, J. and Fire, A. (2012) Contributions of mRNA abundance, ribosome loading, and post- or peri-translational effects to temporal repression of C. elegans heterochronic miRNA targets. *Genome research*, **22**, 2418-2426.

15. Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T. and Weissman, J.S. (2012) High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science*, **335**, 552-557.

16. Gerashchenko, M.V., Lobanov, A.V. and Gladyshev, V.N. (2012) Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 17394-17399.

17. Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G. *et al.* (2011) Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell*, **147**, 1295-1308.

18. Li, G.W., Oh, E. and Weissman, J.S. (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, **484**, 538-541.

19. Stern-Ginossar, N., Weisburd, B., Michalski, A., Le, V.T., Hein, M.Y., Huang, S.X., Ma, M., Shen, B., Qian, S.B., Hengel, H. *et al.* (2012) Decoding human cytomegalovirus. *Science*, **338**, 1088-1093.

20. Liu, X., Jiang, H., Gu, Z. and Roberts, J.W. (2013) High-resolution view of bacteriophage lambda gene expression by ribosome profiling. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 11928-11933.

21. Michel, A.M. and Baranov, P.V. (2013) Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley interdisciplinary reviews. RNA*.

22. Shumway, M., Cochrane, G. and Sugawara, H. (2010) Archiving next generation sequencing data. *Nucleic acids research*, **38**, D870-871.

23. Nekrutenko, A. and Taylor, J. (2012) Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nature reviews. Genetics*, **13**, 667-672.

24. Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J. *et al.* (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome research*, **15**, 1451-1455.

25. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome research*, **12**, 996-1006.

26. Michel, A.M., Choudhury, K.R., Firth, A.E., Ingolia, N.T., Atkins, J.F. and Baranov, P.V. (2012) Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome research*, **22**, 2219-2229.

27. Larsen, B., Gesteland, R.F. and Atkins, J.F. (1997) Structural probing and mutagenic analysis of the stem-loop required for Escherichia coli dnaX ribosomal frameshifting: programmed efficiency of 50%. *Journal of molecular biology*, **271**, 47-60.

28. Tsuchihashi, Z. and Kornberg, A. (1990) Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 2516-2520.

29. Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, **33**, D501-504.

30. Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T. *et al.* (2002) The Ensembl genome database project. *Nucleic acids research*, **30**, 38-41.

31. Pruitt, K.D., Harrow, J., Harte, R.A., Wallin, C., Diekhans, M., Maglott, D.R., Searle, S., Farrell, C.M., Loveland, J.E., Ruef, B.J. *et al.* (2009) The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome research*, **19**, 1316-1323.

32. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. and Siepel, A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research*, **20**, 110-121.

33. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Wheeler, D.L. (2004) GenBank: update. *Nucleic acids research*, **32**, D23-26.

34. Cherry, J.M., Ball, C., Weng, S., Juvik, G., Schmidt, R., Adler, C., Dunn, B., Dwight, S., Riles, L., Mortimer, R.K. *et al.* (1997) Genetic and physical maps of Saccharomyces cerevisiae. *Nature*, **387**, 67-73.

35. Pavesi, A., Conterio, F., Bolchi, A., Dieci, G. and Ottonello, S. (1994) Identification of new eukaryotic tRNA genes in genomic DNA databases by a multistep weight matrix analysis of transcriptional control regions. *Nucleic acids research*, **22**, 1247-1256.

36. Grant, G.R., Farkas, M.H., Pizarro, A.D., Lahens, N.F., Schug, J., Brunk, B.P., Stoeckert, C.J., Hogenesch, J.B. and Pierce, E.A. (2011) Comparative analysis of RNA-Seq

alignment algorithms and the RNA-Seq unified mapper (RUM). *Bioinformatics*, **27**, 2518-2528.

37. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, **17**, 10-12.

38. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, **10**, R25.

39. Kent, W.J. (2002) BLAT--the BLAST-like alignment tool. *Genome research*, **12**, 656-664.

40. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S. and Karolchik, D. (2010) BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*, **26**, 2204-2207.