INVESTIGATING TRANSLATIONAL REPROGRAMMING IN CELLULAR STRESS RESPONSE BY ELONGATION PAUSING AND ALTERNATIVE INITIATION

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Botao Liu

May 2014

INVESTIGATING TRANSLATIONAL REPROGRAMMING IN CELLULAR STRESS RESPONSE BY ELONGATION PAUSING AND ALTERNATIVE INITIATION

Botao Liu, Ph.D.

Cornell University 2014

Cell survival in changing environments requires appropriate regulation of gene expression, including translational control. Multiple stress signaling pathways converge on several key translation factors and rapidly modulate mRNA translation at both the initiation and the elongation stages.

Here, I discover that intracellular proteotoxic stress reduces global protein synthesis by halting ribosomes on transcripts during elongation. Deep sequencing of ribosome-protected mRNA fragments reveals an early elongation pausing, roughly at the site where nascent polypeptide chains emerge from the ribosomal exit tunnel. Inhibiting endogenous chaperone molecules by a dominant-negative mutant or chemical inhibitors recapitulates the early elongation pausing, suggesting a dual role of molecular chaperones in facilitating polypeptide elongation and co-translational folding. My results further support that trapped chaperone under stress may prevent the release of elongation factors from ribosomes. My study reveals that translating ribosomes fine-tune the elongation rate by sensing the intracellular folding environment. The early elongation pausing represents a co-translational stress response to maintain the intracellular protein homeostasis.

Correspondingly, repression of global protein synthesis is often accompanied with selective translation of mRNAs encoding proteins that are vital for cell survival and stress recovery. Understanding the selective translational control in gene expression relies on precise and comprehensive determination of translation initiation sites (TIS) across the entire transcriptome. Here, I develop an approach (global translation initiation sequencing, GTI-seq) to achieve simultaneous detection of both initiation and elongation events on a genome-wide scale. With single nucleotide resolution, I show an unprecedented view of alternative translation initiation in mammalian cells. Furthermore, I uncover a robust translational reprogramming of protein catabolic process, in particular the proteasome system, in response to starvation. This regulatory mode of TIS selection indicates that the scope of selective translation under stress conditions is much broader than anticipated.

Collectively, my studies have revealed unprecedented proteome complexity and flexibility through stress-induced translational reprogramming, including ribosome pausing during elongation and wide-spread alternative translation initiation. Elucidation of the regulatory mechanisms underlying translational reprogramming will ultimately lead to the development of novel therapeutic strategies for human diseases.

BIOGRAPHICAL SKETCH

Botao Liu was born December 4, 1987 in Ganzhou, Jiangxi, China. He earned his Bachelor of Science from Zhejiang University in China 2009. During his undergraduate studies, he received numerous honors and awards, including Excellent Graduates with Honors, Outstanding Thesis for Bachelor's Degree, National Scholarship of China, and Excellent Undergraduate Scholarship. He also served as an undergraduate teaching assistant in 2007-2008 and the Director of Academic Department, Student Union, College of Life Sciences in 2006. He travelled to United States first in 2008 during a summer research program at North Carolina State University.

In 2009, Botao Liu joined the doctoral program of Genetics, Genomics and Development at Cornell University. He was awarded the Hsien Wu and Daisy Yen Wu Scholarship/The Liu Memorial Award and Genomics Scholars Award in 2013. He served as a teaching assistant for the Department of Molecular Biology and Genetics in 2013 and a research assistant for the Division of Nutritional Sciences from 2010 to 2014. Outside of academia, Botao Liu interned at the Department of Ion Torrent, Thermo Fisher Scientific Inc in 2013.

Botao Liu presented his research in Molecular Chaperones & Stress Responses Meeting at Cold Spring Harbor in 2012 and published six research and review papers in high-profile journals. Additionally, he is also one of three inventors of a US patent derived from his work. After earning the Ph.D., he plans to continue with a period of postdoctoral training and then look for a faculty position in major institutions or a research and development job in the biotechnology and pharmaceutical industry.

I delicate my dissertation work to my parents, Huazhou Liu and Lanfen Gu, who have given me their unreserved love, support, and trust.

献给我的父母

ACKNOWLEDGMENTS

First, I would like to express sincere appreciation to my supervisor, Dr. Shu-Bing Qian, for all the time and effects that he has put into my graduate studies in the past five years. He has supported me both scientifically and personally. He is very approachable as a supervisor and I still remember all the inspirational discussions during late nights and weekends. His suggestions are always insightful and extremely helpful, which guide me to the effective path. He is open to my crazy ideas and risky proposals at all times. When new discoveries emerge, his excitement and enthusiasm about science encourage and inspire me. When inevitable difficulties occur, his calmness, patience, and persistence keep me away from anxiety and frustration. In addition, he has spent tremendous amount of time on improving my scientific writing and presentation skills. Finally, he indeed cares about the career of students and allows me to explore possibilities in both academia and industry. I would never have been able to finish my dissertation without the guidance from Shu-Bing. He has fully practiced his philosophy of advisor and student relationship - equality, dialogue, communication and understanding.

I would also like to thank Dr. Scott Emr and Dr. Richard Cerione for their research expertise and insightful advice. In my annual committee meetings, I always benefit a lot from their questions, critiques, and suggestions to my project. Their unique perspectives have led me to think about my research in a bigger picture of biology. Scott and Richard have also become role models for me and they have shown me that researchers can be both successful scientists and helpful educators.

In addition to my special committee members, I also appreciate other members of MBG department and scientific community in general for providing a nourishing environment. Specifically, I am thankful for Dr. Dan Barbash, who appreciated my research potential and let me join GG&D field despite of the unsatisfactory GRE and TOEFL scores. I'm also grateful to Drs. Eric Alani, John Parker, Paul Soloway, Jun (Kelly) Liu, Siu Sylvia Lee for their tremendous help and Vicki Shaff for her support throughout my studies. I would also thank all the anonymous reviewers who have read and criticized my studies. Although the review process is actually lengthy and painful, I have learned from peer scientists to be cautious and responsible to the results, interpretations and conclusions.

I feel extremely lucky of spending such a wonderful time with all the former and current Qian lab members. Special thanks go to Dr. Yan Han. It was Yan who convinced me to join Qian lab after my rotations. More importantly, I learned most of the experimental skills used throughout my graduate studies from her, from the basic tissue culture to the cutting-edge ribosome profiling technology. She is also my close friend outside of lab and I consider her as a lovely sister who has warmed my heart in this cold town far away from home. I thank former graduate student Dr. Crystal Conn, who trained me during my rotation and accompanied me during all the nights and weekends in the lab. Her optimism, energetic personality, and kindness have been encouraging me all the time. I am grateful to Drs. XingQian (Ben) Zhang and Xiangwei Gao for all the valuable discussions about science and their help in many experiments. I also appreciate the wonderful and productive collaboration with Drs. Sooncheol Lee and Soohyun Lee, which led to most of the results in Chapter 3. I also

have enjoyed all the scientific and social debates with R. Alex Coots, who has provided me with a broader view of the culture, society and scientific community of US. Additionally, I still miss all the happiness shared with Dr. Jun Sun and Xiaoxing (Alva) Shen. Finally, I really appreciate new perspectives, novel thoughts, and supports brought by recent Qian lab members, including Rada Omanovic, Julianna Magdalon, Drs. Ji Wan, Jun Zhou, and Mridusmita Saikia.

I would never have been able to persist in my adventure of science without the support from all my friends over the world. As classmates in college, Xiang Ma, Zhisong He, Yixing Zhai, Song Feng and I may be far from each other geographically now, but we are close in the journey of pursuing science. We encourage each other in face of difficulties and share the excitement about breakthroughs. We exchange opinions on scientific problems from distinct perspectives. In addition, I would like to specially thank my friends at Cornell, including Lu Huang, Tao Sun, Donghao Li, Mingfei Ji, Xian Qu, Yao Sun, Dr. Zhilong Bao, Dr. Zhen Xue, Dr. Shanshan Zhang, Dr. Liu Yang, Jianfeng Zhang, Li Yao, Xiaowen Shi, and Zijing Zhang. Without you, the life at Cornell would not have been such a fun and joyful experience.

Lastly but most importantly, my sincere gratitude goes to my beloved parents, Huazhou Liu and Lanfen Gu. They have given me all the best things that they have without any reservation. Their unconditional love and complete trust are the inexhaustible source of my courage to believe myself and pursue my own path. It would not have been possible for me to achieve anything without their guidance and support. Thank you for all you've done.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

3'UTR	3' untranslated region
40S	small subunit of eukaryotic 80S ribosome
4EBP	eIF4E-binding protein
5'UTR	5' untranslated region
60S	large subunit of eukaryotic 80S ribosome
80S	eukaryotic 80S ribosome
A-site	aminoacyl-tRNA site
AAR	amino acid response
AdV	adenoviruses
AMPK	AMP-responsive protein kinase
ATA	aurintricarboxylic acid
ATM	ataxia telangiectasia mutated
ATP	adenosine triphosphate
ATF4	activating transcription factor 4
AZC	$_{L}$-Azetidine-2-carboxylic acid
BAG-1	BCL2-associated athanogene
BCL-2	B-cell CLL/lymphoma 2
BiP	binding immunoglobulin protein
BSA	bovine serum albumin
CDS	coding sequence
C/EBP	CCAAT/enhancer binding protein
CHX	cycloheximide
CIRP	cold inducible RNA binding protein
CPEB2	cytoplasmic polyadenylation element binding protein 2
CrPV	cricket paralysis virus
CTP	cytidine triphosphate
DMEM	Dulbecco's Modified Eagle's Medium
DNA	deoxyribonucleic acid
Dox	doxycycline
DSP	Dithiobis (succinimidylpropionate)
dTIS	downstream translation initiation sites
DTT	dithiothreitol
E-site	exit site
eEF1	eukaryotic translation elongation factor 1
eEF1A	eukaryotic translation elongation factor 1 alpha
eEF1	eukaryotic translation elongation factor 2
eEF2K	eukaryotic translation elongation factor 2 kinase
EF-P	elongation factor P
eIF1	eukaryotic translation initiation factor 1
eIF1A	eukaryotic translation initiation factor 1A
eIF2	eukaryotic translation initiation factor 2
eIF2α	eukaryotic translation initiation factor 2 alpha
eIF2B	eukaryotic translation initiation factor 2B

eIF3	eukaryotic translation initiation factor 3
eIF4A	eukaryotic translation initiation factor 4A
eIF4E	eukaryotic translation initiation factor 4E
eIF4F	eukaryotic translation initiation factor 4F
eIF4G	eukaryotic translation initiation factor 4G
eIF5A	eukaryotic translation initiation factor 5A
eIFs	eukaryotic translation initiation factors
EMT	epithelia-to-mesenchymal transitions
eRF1	eukaryotic translation release factor 1
eRF3	eukaryotic translation release factor 3
ER	endoplasmic reticulum
FBS	fetal bovine serum
Fluc	*firefly* luciferase
FKBP	FK506 binding protein
FRB	FKBP12-rapamycin binding domain
HCV	hepatitis C virus
HEK293	human embryonic kidney 293
Hela-tTA	Hela-transactivator
HIAP2	baculoviral IAP repeat containing 2
HIF1A	Hypoxia-inducible factor 1-alpha
hnRNP E1	heterogeneous nuclear ribonucleoprotein E1
HRI	heme-regulated inhibitor kinase
HRT	harringtonine
Hsc70	heat shock cognate 70 kDa protein
HSF1	heat shock factor 1
Hsp70	heat shock 70kDa protein
IP	immunoprecipitation
IRES	internal ribosome entry site
ITAFs	IRES trans-acting factors
GA	geldanamycin
GAP	GTPase-activating protein
GATOR1	GAP activity toward Rags
GCN2	general control nondepressible 2
GCN4	general control nondepressible 4
GDP	guanosine diphosphate
GEF	nucleotide exchange factor
GFP	green fluorescent protein
GTI-seq	global translation initiation sequencing
GTP	guanosine triphosphate
GTPase	guanosine triphosphatases
lincRNAs	large intergenic ncRNAs
LKB1	liver kinase B1
LTM	lactimidomycin
m^5C	5-methylcytidine
m^6A	N6-methyladenine

m7Gppp	5'-7-methyl-guanosine
MEF	mouse embryonic fibroblast
MG132	Z-Leu-Leu-Leu-al
MNase	micrococcal nuclease
mRNA	message RNA
mTORC1	mammalian target of rapamycin complex 1
mtHsp70	mitochondrial Hsp70
NAC	nascent chain-associated complex
Narc	Narciclasine
ncRNA	non-coding RNA
ORFs	open reading frames
P-site	peptidyl-tRNA site
P/M	polysome/ monosome
Pab1	polyA binding protein
PBS	phosphate buffered saline
PERK	PKR-like ER kinase
PES	2-phenylethyenesulfo-namide
PI	pausing index
PI3K	phosphoinositide 3-kinase
PIC	43S pre-initiation complex
PKR	protein kinase RNA
PQC	protein quality control
PSMA7	proteasome α subunit 7
PTB	polypyrimidine tract binding protein
PTC	pepetidyl transferase center
Rag	Ras-related small GTP-binding protein
Rheb	Ras homologue enriched in brain
Rluc	*Renilla* luciferase
RNA	ribonucleic acid
RNC	ribosome-nascent chain
RP	ribosomal protein
RPL4	ribosomal protein large subunit 4
RPS	small subunit ribosomal protein
RpS6	ribosomal protein small subunit 6
RPL	large subunit ribosomal protein
ROS	reactive oxygen species
RPFs	ribosome protected mRNA fragments
RRL	rabbit reticulocyte lysate
rRNA	ribosomal RNA
S6K1,	p70 ribosomal S6 Kinase 1
SeRP	selective ribosome profiling
SG	stress granules
sORFs	short ORFs
TEE	translation-enhancing elements
TGF β	transforming growth factor β

TIS	translation initiation sites
TOP	terminal oligopyrimidine
TSC	tuberous sclerosis complex
tRNA	transfer RNA
uORFs	upstream open reading frames
uTIS	upstream translation initiation sites
UTR	untranslated region
UV	ultraviolet
VEGF	vascular endothelial growth factor
VER	VER-155008
XBP1	X-box-binding protein 1
XIAP	X-linked inhibitor of apoptosis

PREFACE

The main focus of this dissertation is the translational reprogramming induced by cellular stress to maintain protein homeostasis. This work is composed of two major research projects, emphasizing on translational regulation at initiation and elongation stages respectively. Chapter 1 reviews current knowledge of translational reprogramming in cellular stress response, with emphasis on the regulatory modes of global protein synthesis and selective translation. Chapter 2 is the main focus of my graduate work, which examines the global repression of protein synthesis under proteotoxic stress and discovers a novel early elongation pausing response controlled by molecular chaperone. Further dissection of the underlying mechanism is included in the Appendix I as work in progress. Chapter 3 is a collaborative work of developing an approach to achieve simultaneous detection of both initiation and elongation events on a genome-wide scale. Appendix II is the application of this method to profile alternative initiation in response to starvation as part of a manuscript in submission. Chapter 4 summarizes the major findings and the connections under the concept of protein homeostasis, while discussing several open questions still under investigation as well as the potential applications and improvement of current technology.

CHAPTER 1

Translational reprogramming in cellular stress response

This review was submitted July 2013 and first published online as Liu B and Qian SB. **Translational reprogramming in cellular stress response.** *Wiley Interdisciplinary Reviews RNA.* 2013 Dec 23. doi: 10.1002/wrna.1212. Minor modifications have been made for reprint here.

1.1 Abstract

Cell survival in changing environments requires appropriate regulation of gene expression, including translational control. Multiple stress signaling pathways converge on several key translation factors, such as eIF4F and eIF2, and rapidly modulate mRNA translation at both the initiation and the elongation stages. Repression of global protein synthesis is often accompanied with selective translation of mRNAs encoding proteins that are vital for cell survival and stress recovery. The past decade has seen significant progress in our understanding of translational reprogramming in part due to the development of technologies that allow the dissection of the interplay between mRNA elements and corresponding binding proteins. Recent genome-wide studies using ribosome profiling have revealed unprecedented proteome complexity and flexibility through alternative translation,

raising intriguing questions about stress-induced translational reprogramming. Many surprises emerged from these studies, including wide-spread alternative translation initiation, ribosome pausing during elongation, and reversible modification of mRNAs. Elucidation of the regulatory mechanisms underlying translational reprogramming will ultimately lead to the development of novel therapeutic strategies for human diseases.

1.2 Introduction

All living organisms must detect and respond to changing growth conditions and environmental stimuli. Under acute adverse conditions, such as heat shock, hypoxia, nutrient deprivation or DNA damage, gene expression undergoes coordinated changes to ensure cell survival. The past decade has seen significant progress in our understanding of gene regulation in response to stress, including chromatin remodelling, transcriptional regulation, alternative splicing and translational control. Recent advances in next-generation sequencing allow the dissection of gene regulation in an unprecedented scale and resolution.[1] Although transcriptional regulation is essential in mediating the strength of stress response, translational control often provides immediate and effective changes in protein levels.[2] This swift response offers a timely adaptation for cells to maximize survival under stress.[3]

Translation can be divided mechanistically into three stages: initiation, elongation and termination. As the rate-limiting step in translation, initiation is a complex process involving ribosome loading, scanning, and start codon selection

before elongation commitment.[2] Consistent with its critical role in determining the overall rate of translation, initiation is the primary target of regulation under stress. Under various stress conditions, distinct signalling pathways converge to a few initiation regulators resulting in translational inhibition. The two best characterized mechanisms are mRNA cap recognition and ternary complex formation (see below).[4] Although translational control at the initiation stage has been extensively studied,[5] much less is known about the regulatory mechanisms of elongation under stress conditions. Recent development of ribosome profiling technology has reignited the research interest in the translation field.[6, 7] The innovative technique enables monitoring of ribosome dynamics with unprecedented resolution at the genome-wide scale.[8] With this powerful tool, surprising mechanisms at post-initiation stages of translation have been uncovered.[9]

Protein synthesis consumes a lion's share of energy and cellular resources, so translation is generally repressed under most if not all types of stress conditions. However, subsets of mRNAs can bypass the general inhibition and be selectively translated. Most of these mRNAs encode stress response proteins, which protect cells from damages and facilitate the post-stress recovery.[10, 11] The concept of translational reprogramming fits well into the mode of translational control in stress response, allowing selective translation of mRNAs to maintain the expression of stress proteins when general protein synthesis is compromised. Such regulation can be quantitative (all-or-none vs. graded), or qualitative (enabling a single mRNA to produce several different proteins). We argue that translational reprogramming lies at the heart of the stress response and is required for rapid cellular adaptation under stress. Mechanistic

details of translational reprogramming, however, are only beginning to be unfurled. In this review, we discuss mechanisms underlying global repression of translation as well as selective translation in response to stress. Although both processes are tightly coupled during translational reprogramming, for the purpose of clarity, we review each part separately by focusing on mRNA elements as well as corresponding binding proteins. We start with an overview of well-established regulatory mechanisms through initiation and then focus on the recent progress in novel modes of regulation are important in translational reprogramming in stress response.

1.3 Global Repression of Translation During Stress

1.3.1 Overview of Eukaryotic Translation Processes

To better illustrate mechanisms underlying translational reprogramming, it is necessary to briefly revisit what we have learned regarding translation processes in eukaryotic cells. Under normal conditions, eukaryotic cells employ a cap-dependent mechanism to initiate translation for most mRNAs.[12, 13] The 5' end of eukaryotic mRNAs is modified with an m7Gppp cap structure, which is recognized by a eukaryotic initiation factor 4E (eIF4E). eIF4E forms the eIF4F complex by binding to eIF4G (a scaffold protein) and eIF4A (a helicase).[14-16] The cap recognition is the first step that determines which mRNAs are to be translated, and it is not surprising that multiple signalling pathways control this rate-limiting step. Another key step is the formation of a ternary complex, which is composed of a methionine-loaded initiator tRNA and a GTP-coupled eIF2.[17] The ternary complex associates with the 40S small

ribosome subunit and several other initiation factors (eIF1A, eIF3, eIF1) to form the 43S pre-initiation complex (PIC). PIC is then recruited to mRNA via the scaffold eIF4G within the cap-associated eIF4F complex, forming the 48S complex. With the help of eIF4A to unwind mRNA secondary structures, PIC scans the 5' untranslated region (5'UTR) until it encounters an initiation codon.[18, 19] The efficiency of start codon recognition can be influenced by the codon context as well as initiation factors eIF1 and eIF1A, although the precise mechanism remains elusive. The event of start codon recognition is believed to trigger conformational changes of the 48S complex followed by release of the initiation factors. With the help of eIF5 and eIF5B that induce hydrolysis of eIF2-bound GTP, a 60S large ribosome subunit joins the 40S subunit, forming a complete 80S complex ready to proceed to the elongation step.[18]

Translation elongation is mediated by elongation factors eEF1 and eEF2, which delivers amino acid-charged tRNA to the ribosomal A site and catalyses ribosomal translocation, respectively. During elongation, the ribosome does not move at a constant speed but rather in a stop-and-go traffic manner. Both *cis* sequence elements and *trans* regulatory factors contribute to the variations of elongation speed. However, our understanding of elongation control has lagged behind the knowledge of initiation regulation. When the ribosome decoding centre reaches a stop codon, termination occurs via the concerted action of release factors eRF1 and eRF3. Notably, peptide release, tRNA dissociation, and ribosome separation do not take place simultaneously. In some cases, the 40S subunit remains associated with mRNA and could start a second round of translation from the downstream start codon, a process called re-initiation.[12] Strikingly, in a reconstituted *in vitro* translation system,

Skabkin et al found that the post-termination ribosome could migrate bi-directionally to codons cognate to the P-site tRNA.[19] Although it remains to be confirmed whether this radical event occurs *in vivo*, the dynamic ribosome behaviour surrounding termination provides novel mechanistic insights into translation re-initiation.

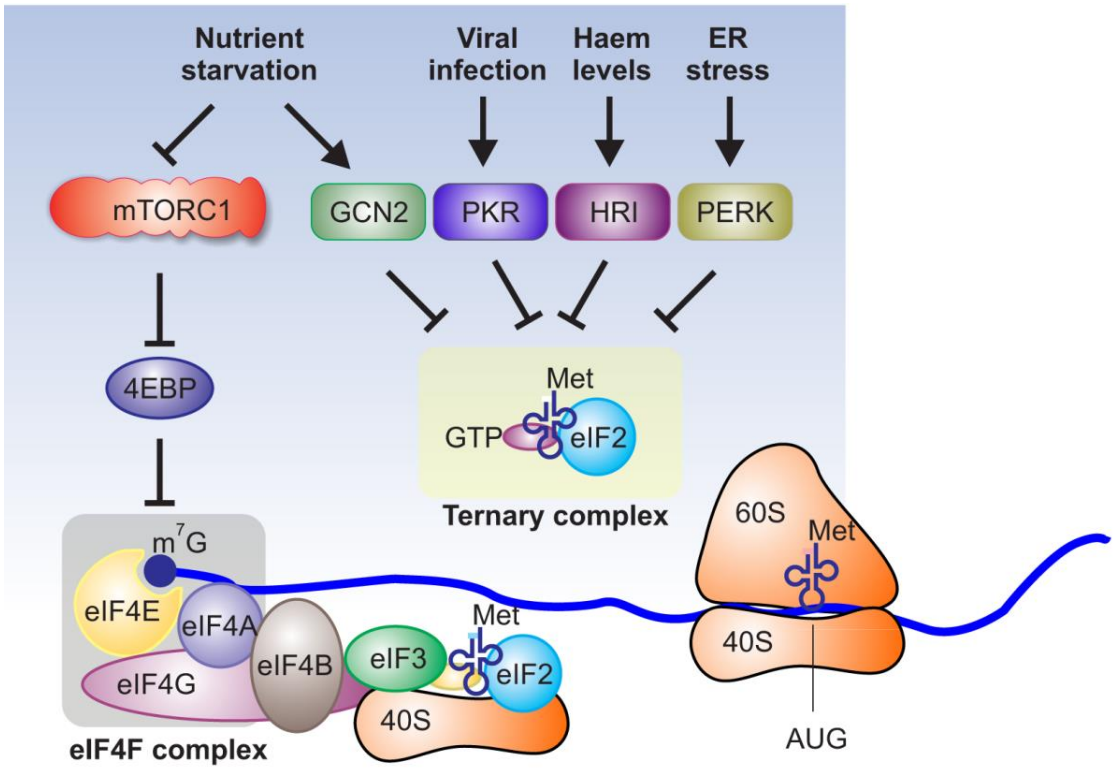


Figure 1-1. Multiple stress signals converge on initiation factors and inhibit global protein synthesis.
Cap-dependent translation initiation requires cap binding, eIF4F complex assembly (light grey square), and ternary complex formation (light yellow square). Nutrient signalling mTORC1 controls eIF4F complex formation by phosphorylating 4EBP, which releases eIF4E for cap binding. Nutrient starvation not only inhibits the mTORC1 signalling pathway, but also triggers GCN2 kinase activity. GCN2 phosphorylates eIF2α that inhibits ternary complex formation. In addition to the GCN2 kinase, other kinases integrate many stress conditions by phosphorylating eIF2α, forming an integrated stress response targeting translation initiation.

1.3.2 Initiation Regulators and Signalling Pathways

eIF4F-mediated 5' cap recognition

A cap-dependent mechanism accounts for the translation of the vast majority of cellular mRNAs. Under stress conditions, a diverse array of signalling pathways control the eIF4F-mediated cap recognition, thereby adjusting the rate of global protein synthesis (Figure 1). One best known regulator is the eIF4E-binding protein (4EBP), which shares a similar structure with eIF4G. By competing with eIF4G, 4EBP acts as a negative regulator of translation initiation by repressing the assembly of eIF4F complexes at the 5' terminus of transcripts. The binding capacity of 4EBP depends on its phosphorylation status. Under normal growth conditions, 4EBP is heavily phosphorylated and has lower affinity with eIF4E.[20] One major signalling pathway that mediates 4EBP phosphorylation is the mammalian target of rapamycin complex 1 (mTORC1).[21, 22] mTORC1 is an evolutionarily conserved serine/threonine kinase that senses extracellular signals as well as the intracellular energy status. Nutritional stresses such as amino acid starvation inhibits global protein synthesis partially through the mTORC1 signalling pathway. mTORC1 senses amino acid levels through a sophisticated system.[23] Recent studies revealed that mTORC1 activation occurs primarily at the surface of the lysosome by heterodimeric RagA/B-RagC/D GTPases.[24] When amino acids are limited, Rag GTPases are inactivated, leading to GDP coupled RagA/B and GTP bound RagC/D, which are unable to recruit mTORC1 to the lysosome membrane.[25] A complex named Ragulator acts as a guanine nucleotide exchange factor (GEF) for RagA and RagB, whereas another complex called GATOR1 has GTPase-activating protein (GAP) activity.[26, 27]

However, the direct intracellular amino acid sensor remains to be characterized. Once recruited to the lysosome surface, mTORC1 is believed to be directly activated by Ras homologue enriched in brain (Rheb).[28, 29] The activated mTORC1 then phosphorylates 4EBP, leading to de-repression of eIF4F and enhanced cap-dependent translation.

At the lysosome surface, Rheb activity is subject to regulation by phosphoinositide 3-kinase (PI3K) pathways. Therefore, both the amino acid sensing system and the insulin signalling pathway converge on mTORC1. Rheb activity is negatively regulated by tuberous sclerosis complex (TSC) 1 and 2, in which TSC2 acts as a GAP towards Rheb.[30, 31] Several stress signals integrate into mTORC1 via TSC. For instance, oxidative stress activates AMP-responsive protein kinase (AMPK) pathway, which suppress mTORC1 by phosphorylating its negative regulator TSC2.[32] In addition, TSC is found to be localized on the peroxisome and inhibit mTORC1 in response to endogenous reactive oxygen species.[33] DNA damage could be sensed in both p53 dependent and independent pathways.[34, 35] The p53-dependent pathway requires the transcriptional activation of Sestrin1 and Sestrin2. Increased Sestrin1 and 2 activate TSC2 through AMPK, eventually repressing mTORC1 activity.[36] For the p53 independent recognition, DNA damage is sensed by a protein kinase ATM (ataxia telangiectasia mutated) and the signal is transduced through liver kinase B1 (LKB1)/AMPK1 to target TSC2 and inhibits mTORC1.[37]

As mentioned above, 4EBP is one of the direct targets of mTORC1. Under suppressed mTORC1 activity during stress, the hypo-phosphorylated 4EBP sequesters eIF4E from the 5'cap of mRNAs, preventing the formation of eI4F complex and the

cap-dependent initiation.[38] Employing the ribosome profiling technique, several recent studies investigated the translational response when mTORC1 was inhibited by chemical inhibitors.[39] Inhibiting mTORC1 activity by Torin significantly reduced the translation of mRNAs containing 5' terminal oligopyrimidine (TOP) motifs or TOP-like motifs.[40, 41] These mRNAs mostly encode ribosomal proteins and translation factors. In addition, several transcripts whose translation is highly regulated by mTORC1 are involved in cell proliferation, metabolism and invasion, confirming the critical role of translational control in cancer progression.[41] Given the widely accepted notion that eIF4F complex formation controls the majority of cap-dependent translation, it is surprising to find that only a subset of mRNAs whose translation is influenced by mTORC1 inhibition. Indeed, in cells lacking both 4EBP1 and 4EBP2, not all mRNA translation is equally upregulated.[42] Interestingly, mRNAs involving cell proliferation are preferentially subjected to translational control by 4EBP. Hence, translation of individual mRNAs has different sensitivity to the perturbation of cap-recognition.

eIF2-controlled ternary complex formation

Many stress conditions trigger the phosphorylation of eIF2α. In mammals, there are four different types of eIF2α kinases activated by different stressors: general control non-derepressible-2 (GCN2) for amino acid starvation, protein kinase RNA (PKR) for double-stranded RNAs during virus infection, PKR-like ER kinase (PERK) for unfolded proteins in ER, and heme-regulated inhibitor kinase (HRI) for heme deprivation.4 eIF2α is a subunit of eIF2 that is part of the ternary complex. As GTP is hydrolysed during translation initiation, eIF2 needs to be recharged by initiator

tRNA. This recharging is accomplished by eIF2B-catalyzed GDP-GTP exchange. Under stress conditions, Ser51 of eIF2α subunit is phosphorylated by stress sensing kinases mentioned above. Phosphorylation of eIF2α inhibits GDP-GTP exchange by reducing the dissociation rate of eIF2B.[43] As a result, ternary complex formation is suppressed and global translation is reduced. Therefore, different types of stress conditions converge on eIF2α, resulting in the inhibition of ternary complex formation (Figure 1). Further supporting this notion, GCN2 also responds to UV exposure and DNA damage response.[44, 45] Moreover, both hypoxia and oxidative stress could activate PERK, resulting in phosphorylation of eIF2α.[46]

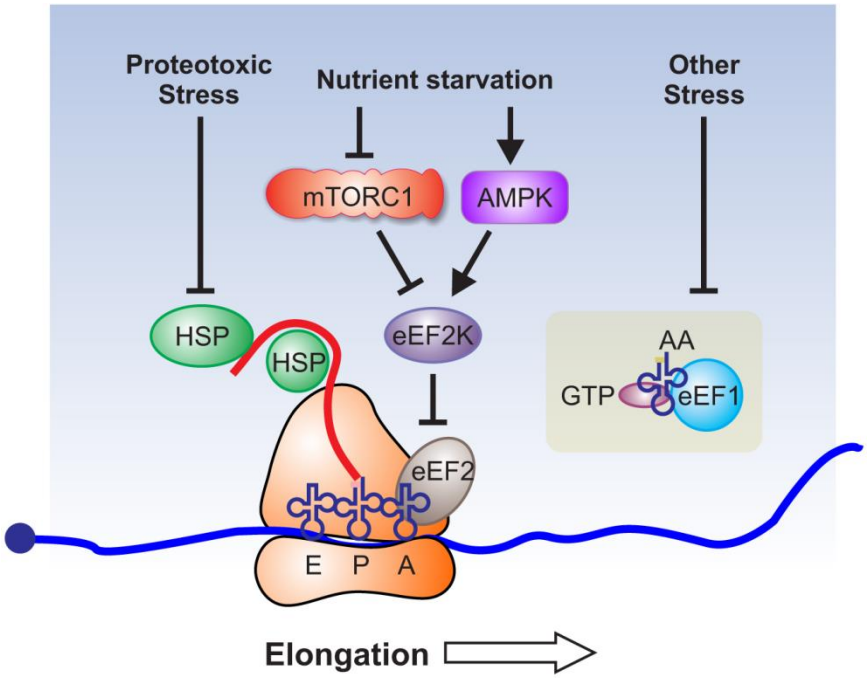


Figure 1-2. Translational regulation at the elongation stage.
Nutrient starvation inhibits mTORC1 and activates eEF2K, which inhibits translation elongation by blocking the function of eEF2. Starvation also activates AMPK that promotes the activation of eEF2K, resulting in elongation inhibition. Many stressors could affect the activity of eEF1, although the underlying mechanism is not completely understood. In addition, ribosome-associated chaperones regulate translation elongation, enabling cells to modulate translational capacity in response to proteotoxic stress.

It is clear that the same type of stress could trigger multiple signalling pathways leading to global protein synthesis inhibition. For instance, amino acid starvation not only suppresses eIF4-mediated cap recognition through aforementioned mTORC1 signalling pathways, but also activates GCN2 via the accumulation of uncharged tRNA.[47] Consequently, both cap-recognition and ternary complex formation are suppressed under nutrient starvation. It seems that both stress signalling pathways act in parallel. However, cells lacking GCN2 blunted the responsiveness of mTORC1 to amino acid deprivation.[48] Much remains to be learned for the crosstalk between GCN2/eIF2α and mTORC1 signalling pathways.

1.3.3 Elongation Modulators and Signalling Pathways

Despite the extensive regulation at the initiation stage, a growing body of evidence suggests that elongation step is subject to more rigorous regulation than is previously assumed (Figure 2).[49] Like some initiation factors, one common regulatory mechanism of elongation factors is phosphorylation. For instance, elongation factor eEF2 undergoes phosphorylation at Thr56 within the GTP-binding domain in response to oxidative stress and this modification interferes with its ability to bind to the ribosome.[50-53] mTORC1 negatively regulates its cognate kinase eEF2K and thereby activates eEF2.[54] Thus, mTORC1 regulates protein translation at multiple stages. The activity of eEF2 can also be regulated by RNA-binding proteins. For instance, cytoplasmic polyadenylation element binding protein 2 (CPEB2) reduces GTP hydrolysis by eEF2.[55] Interestingly, CPEB2 slows down the translation of HIF1A mRNA under normal conditions by binding to the 3'UTR. When cells encounter

hypoxic stress, CPEB2 dissociates from HIF-1α mRNA, leading to rapid synthesis of HIF-1α for hypoxic adaptation. Further supporting the physiological significance of eEF2, eEF2 is repressed by the activation of AMPK-eEF2K-eEF2 pathway under a series of stress conditions, including endoplasmic reticulum stress, hypoxia-induced energy stress, genotoxic stress, and nutrient deprivation.[56-58] Various stress signals trigger the activation of eEF2K by AMPK-mediated phosphorylation on serine 398. Activated eEF2K phosphorylates eEF2 and induces a temporary ribosomal slowdown at the stage of elongation. During recovery stage, eEF2K is degraded by the ubiquitin-proteasome system, allowing the rapid resumption of translation elongation. Remarkably, transformed tumour cells rely on this AMPK-eEF2K axis to survive under nutrient stress conditions. Indeed, expression of eEF2K strongly correlated with overall survival in human medulloblastoma and glioblastoma multiforme.[59] In addition to eEF2, eEF1A also undergoes similar regulation. One example is the role of eEF1A in epithelia-to-mesenchymal transitions (EMT) which occurs in tumour metastasis. This regulation is mediated by transforming growth factor β (TGF β) signalling pathway.[60] In the absence of TGF β signalling, 3'UTRs of specific mRNAs are recognized by a RNA-binding protein heterogeneous nuclear ribonucleoprotein E1 (hnRNP E1), which blocks the translocation of ribosomes by associating with eEF1A. Active TGF β signalling phosphorylates hnRNP E1 and releases eEF1A from ribosomes, allowing the elongation to proceed on mRNAs and promoting EMT. In addition, TGF-β1 also causes dissociation of ribosomal protein RPL26 and eEF1A from p53 mRNA, thereby reducing p53 mRNA translation in response to cellular stress.[61] Finally, eukaryotic initiation factor 5A (elongation factor P in prokaryotes)

has been recently identified to promote elongation of polyproline motifs.[62-67] In bacteria, EF-P influences the stress response of pH receptor CadC and translation of other polyproline-containing proteins, suggesting similar functions of eIF5A in eukaryotes.

In addition to mechanisms regulating the elongation factors, the elongation process itself can cooperate with other stress response pathways to coordinate regulations at various levels. mRNA translation proceeds not at a constant rate but rather in a stop-and-go traffic manner.[68] Variations of elongation speed may result from local stable mRNA structure, or the presence of rare codons.[69-74] Interestingly, nascent chains could also induce translational pausing in a sequence-specific manner. Several recent studies have revealed the importance of elongation pausing in stress response. One example is the splicing of X-box-binding protein 1 messenger RNA (XBP1u mRNA) upon endoplasmic reticulum stress.[75] An evolutionarily conserved peptide module at the carboxyl terminus is responsible for the translational pausing and required for the efficient targeting mRNA-ribosome-nascent chain (RNC) complex to the ER membrane and efficient splicing of the XBP1u mRNA. In addition, ribosomal stalling in the upstream ORF causes mRNA remodeling and formation of an active IRES (discuss in more details below), stimulating the translation of cat-1 Arg/Lys transporter under amino acid starvation.[76] Using ribosome profiling, several recent studies discovered an early ribosome pausing under a variety of stress conditions, including heat shock, proteotoxic stress, and oxidative stress.[77-79] Intriguingly, most of the ribosomes paused within the first 50 codon window of almost all coding sequences, a region corresponding to the length of

nascent chains occupying the ribosomal exit tunnel. Since ribosome-associated chaperone molecules are located near the exit of the tunnel, it is postulated that translation elongation is influenced by chaperone availability. It is still unclear mechanistically how the absence of chaperones brings translation to a halt. This phenomenon nevertheless reveals that translating ribosomes, via associated factors, fine-tune the elongation rate by sensing the intracellular folding environment. The early elongation pausing may represent a co-translational stress response to maintain the intracellular protein homeostasis.

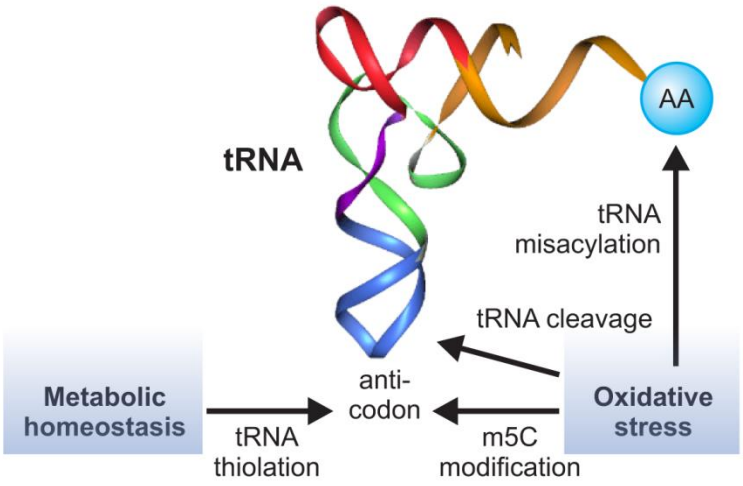


Figure 1-3. Translational regulation by tRNA modification.
tRNA contains many modified nucleobases. Anti-codon modification influences decoding processes and the overall translation capacity. Oxidative stress has multiple effects on tRNA metabolism, including m^5C at the wobble position, tRNA misacylation, and tRNA cleavage. In addition, metabolic homeostasis such as sulfur amino acid levels regulates tRNA thiolation at the wobble position. These tRNA modifications trigger translational reprogramming in response to stress conditions.

Growing ribosome profiling data has enabled computational simulation of the translation process in yeast.[80-82] Consistent with previous studies, initiation and ribosome availability were shown to be the rate-determining factors of translation under normal growth conditions. However, the simulative results suggest that

elongation becomes the limiting step under severe amino acid starvation conditions. The authors argued that reduced initiation rate under stress might increase the free ribosome and tRNA, thereby promoting elongation. Although this hypothesis awaits experimental validation, it supports the importance of elongation regulation under stress.

1.3.4 Stress-Induced RNA Modification

Numerous modifications (>100) have been identified on the four canonical bases in most types of RNA. Some of the RNA modifications serve as sentinels for various stress conditions, while others directly affect the decoding process of translation.[83] Emerging evidence points to a critical role for tRNA and rRNA modifications in the various cellular responses to stress (Figure 3). Using a quantitative system approach, Chan et al reported signature changes in the spectrum of tRNA modifications in S. *cerevisiae* upon oxidative stress.[84] Interestingly, there was an increase in the proportion of $tRNA^{Leu(CAA)}$ containing m^5C at the wobble position. This modification causes selective translation of mRNA from genes enriched in the TTG codon. In addition to tRNA modifications, several recent studies reported that oxidative stress triggers endonucleolytical cleave of tRNAs around the anticodon, giving rise to small RNA species that may participate in various stress signalling pathways.[85-89] The nucleases responsible for stress-induced tRNA cleavage are Rny1 in yeast and angiogenin in mammals. The oxidative-stress activated nucleases cleave within the conserved single-stranded 3'-CCA termini of all tRNAs, thereby blocking their use in translation. This CCA deactivation is reversible and repairable by the

CCA-adding enzyme [ATP(CTP):tRNA nucleotidyltransferase].[90] Through this mechanism the eukaryotic cell dynamically represses and reactivates translation at low metabolic costs. In non-stressed cells, these enzymes cannot gain access to cytosolic tRNAs, suggesting that stress-induced tRNA cleavage is a highly regulated process. However, not all stress conditions can trigger tRNA cleavage. Oxidative stress seems to preferentially affect tRNA biology. Interestingly, up to a tenfold increase of methionine-misacylation occurs at tRNA when cells are exposed to oxidative stress.[91] Likewise, virus infection, treating cells with toll-like receptor ligands or chemicals also induced tRNA mis-acylation. The physiological significance of modified translation fidelity remains unclear. It has been proposed that misincorporation of methionine into cellular proteins could possibly protect cells from reactive oxygen species (ROS)-mediated damage.[91] A recent study reported that thiolation status of tRNA wobble-uridine nucleotides is correlated with the intracellular availability of sulphur amino acids methionine and cysteine.[92] Interestingly, changing tRNA thiolation regulates translational reprogramming and enables cells to modulate translational capacity according to metabolic homeostasis.

In eukaryotic mRNA, different types of methylation modification have been documented. One abundant and conserved mRNA modification is N6-methyladenine (m^6A). The abundance of m6A has been estimated to be 3-5 residues per mRNA on average in HeLa cells.[83] Importantly, the m^6A modification is dynamic and can be reprogrammed under different conditions. Yeast cells have low levels of m^6A modification during regular mitosis growth, but appropriate 50% of mRNAs contain m^6A sites during meiosis.[93] It has been suggested m^6A modification may regulate

translation efficiency. Using m^6A-specific antibodies, two recent studies revealed a wide-spread distribution of m^6A across the mammalian transcriptome.[94, 95] Surprisingly, the mapped m^6A sites were enriched near the stop codons and in the 3'UTRs. Further supporting the dynamic feature of m^6A modification, there was a tissue-specific pattern of m^6A with a dramatic increase during brain development. In addition, the m^6A landscape changes in response to various stimuli. Although the exact function of m^6A in mRNA remains obscure, it is certain that this dynamic modification has important regulatory roles in gene expression, including translational control.

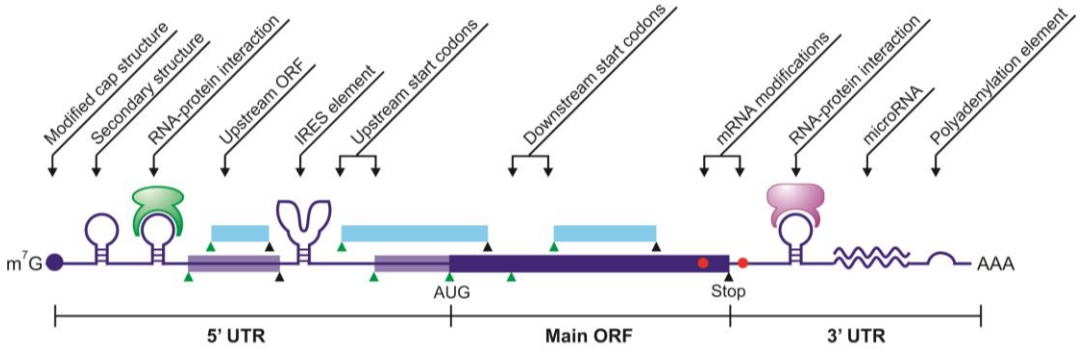


Figure 1-4. Types of *cis*-sequence elements that contribute to translational regulation. mRNA contains multiple start codons (green triangle) and stop codons (black triangle), generating ORFs in-frame (blue box) or out-of-frame (cyan box). Secondary structures are present in 5'UTR and/or 3'UTR, with or without interacting proteins. Reversible mRNA modification could also regulate translational reprogramming in response to stress conditions.

1.4 Selective Translational Regulation During Stress

Repression of global protein synthesis helps reduce the cellular burden during stress conditions. However, subsets of mRNAs undergo selective translation to produce proteins that are vital for cell survival and stress recovery.[4] Cells employ a variety of mechanisms to achieve selective translation, which often involves *cis*

sequence elements on mRNAs and trans regulatory factors recognizing specific mRNA features. Most of the *cis*-elements reside in the untranslated region of mRNAs, including internal ribosome entry sites (IRES), upstream open reading frames (uORFs), motifs with special sequences or secondary structures, and microRNA binding sites (Figure 4). The roles of microRNA in translational regulation during stress have been comprehensively reviewed elsewhere.[96] Here we will focus on other key mechanisms regulating selective translational in response to stress.

1.4.1 Cap-independent Translation Initiation

Not all the mRNAs bear the typical 5' cap structure. The best characterized cap-independent translational mechanism is IRES.[97] Originally discovered in picornavirus mRNAs, the IRES element in the 5'UTR forms complex secondary structures that directly recruit ribosome subunits without the requirement of some or even all initiation factors.[98] In addition to the typical IRES elements found in viral mRNAs, a growing body of evidence suggests that certain cellular mRNAs may use the similar IRES mechanism for cap-independent translation initiation. This non-canonical translation initiation often occurs during special conditions, such as differentiation, apoptosis, and cellular stress.[99] Under genotoxic stress, transcripts encoding c-Myc, p53, X-linked inhibitor of apoptosis (XIAP) and B-cell CLL/lymphoma 2 (BCL-2) are translationally upregulated and these mRNAs are believed to contain IRES at their 5'UTRs.[100-102] During endoplasmic reticulum stress, the inhibitor of apoptosis protein HIAP2 undergoes IRES-mediated translational

induction.[103] In response to hypoxia, translational increase of vascular endothelial growth factor (VEGF) and HIF-1 is also IRES-dependent.[104, 105] Additionally, translation of cold inducible RNA binding protein (CIRP) and heat shock inducible BIP, BAG-1 is also thought to be mediated through IRES.[106, 107] With individual experimental validation, the list of potential IRES-containing mRNAs is expanding rapidly. Using an in vitro selection approach based on mRNA display, a recent study identified over 12,000 random genomic sequences that could act as cap-independent translation-enhancing elements (TEE).[108] Interestingly, the TEE-enriched regions are overrepresented in the 5'UTR, suggesting that cap-independent translational activities might be widespread in the human genome.

Efficient IRES-mediated translation initiation requires RNA binding proteins that are known as IRES trans-acting factors (ITAFs).[109] It is hypothesized that ITAFs may act as RNA chaperones to facilitate the formation of IRES secondary structures.[110] However, our understanding of how IRES-ITAF interaction determines translation initiation is far from complete. For several IRES-containing transcripts mentioned above, such as p53 and BAG-1, polypyrimidine tract binding protein (PTB) functions as the ITAF.[111, 112] During starvation-induced yeast differentiation, an A-rich element in the 5'UTR of some mRNAs involved in invasive growth mediates internal initiation by recruiting polyA binding protein (Pab1).[113] It is likely that different IRES elements and corresponding ITAF factors interplay in distinct manners. However, functional characterization of cellular proteins serving as ITAF has lagged far behind the identification of IRES elements. It remains to be clarified whether the cellular IRES element functions in an exact same manner as the viral IRES.

The presence of both IRES and ITAF does not necessarily guarantee efficient cap-independent translation initiation. Under normal growth conditions, the limiting translation machinery prefers canonical cap-dependent translation. The functional balance between cap-dependent and cap-independent initiation underlies the central translational reprogramming in stress response. Indeed, cap-independent translation dominates only when the general cap-dependent translation is inhibited by cellular stress.[114] This explains why most IRES elements are found in genes whose protein products are involved in cell survival and cell death. Further supporting the coordination between cap-dependent and cap-independent translation, overactivation of nutrient signalling pathway mTORC1 compromises the cap-independent synthesis of stress proteins like Hsp70 and consequently attenuates stress responses.[115] Taken together, cap-independent translation provides an effective means for escaping the global decline in protein synthesis, while permitting the selective translation of specific mRNAs.

1.4.2 Alternative Translation Initiation

Proper selection of the translation initiation site on mRNAs is crucial for the production of desired protein products. In eukaryotes, ribosomal scanning is a well-accepted model for start codon selection.[116] It is commonly assumed that the first AUG codon that the scanning ribosome encounters serves as the start site for translation. However, one or more potential initiation sites could exist upstream of the main start codon, forming upstream open reading frames (uORF).[117] Likewise, many AUG codons downstream of the main start codon could also potentially serve as

initiators. Many factors influence the start codon selection. For instance, the initiator AUG triplet is usually in an optimal context with a purine at position -3 and a guanine at position +4. The presence of mRNA secondary structure at or near the start codon also influences the recognition efficiency. In addition to these *cis* sequence elements, the stringency of start codon selection is also subject to regulation by *trans* acting factors such as eIF1 and eIF1A. Inefficient recognition of an initiator codon results in a portion of 43S PIC continuing to scan and initiating at a downstream site, in a process known as leaky scanning. Many recent studies have uncovered a surprising variety of potential translation start sites in addition to the annotated start codons. Using ribosome profiling coupled with translation inhibitors specifically targeting the initiating ribosomes, several groups have identified multiple initiation sites in almost half of the transcripts in human and mouse transcriptome.[118, 119] Intriguingly, many non-AUG codons, especially CUG, act as alternative start codons for initiating uORF translation.

One expected consequence of alternative translation initiation is an expanded proteome diversity that has not been and could not be predicted by *in silico* analysis of AUG-mediated main ORFs. Indeed, many eukaryotic proteins exhibit a feature of NH2-terminal heterogeneity presumably due to alternative translation. Stress-triggered alternative initiation may generate isoforms with different N-terminus, leading to distinct functions or cellular localization.[79] One well-characterized example is C/EBP, a family of transcription factors that regulate the expression of tissue-specific genes during differentiation. C/EBP mRNA produces protein isoforms with opposite functions according to the level of upstream hormones and signals in a tissue-

specific manner.[120] Alternative start codon selection could also produce functionally distinct protein isoforms. Such a strategy has been widely used by the compact genome in viruses.[121] Comprehensive cataloguing of global translation initiation sites and the associated ORFs is just the beginning in unveiling the role of translational reprogramming in gene expression. The illustration of alternative translation events in response to various stress conditions represents an exciting research field to be fully exploited.

1.4.3 Regulatory uORFs

It has been estimated that about 50% of mammalian transcripts contain at least one upstream open reading frame (uORF).[122] Based on the leaky scanning model, the presence of uORFs is considered to suppress the translation efficiency of main ORFs. Indeed, ribosome profiling results showed a dramatic increase of uORF occupancy under stress conditions such as starvation, oxidative stress, heat shock and proteotoxic stress.[7, 77-79] Interestingly, the ribosome occupancy of uORFs also increased during yeast meiosis and mouse stem cell differentiation.[119, 123] How the up-regulation of uORF translation is achieved under these conditions remains incompletely understood. Despite the inhibitory role of uORF in the translation of most main ORFs, presence of some uORFs could stimulate the translation of mRNAs encoding stress responsive proteins. The best characterized example is GCN4 in yeast or ATF4 in mammals.[124, 125] In the case of ATF4, it contains two uORFs in the 5'UTR: one near the 5' terminus and the other overlapping with the main ORF but in different reading frames. During normal growth conditions, the ternary complex is abundant and ribosome decodes the

first uORF as well as the second uORF. Termination of uORF2 does not allow the initiation of the main ORF because of sequence overlapping. Under stress conditions that trigger eIF2α phosphorylation, reduced ternary complexes formation leads to longer time for the scanning ribosome to acquire a ternary complex. As a result, more ribosomes bypass the second uORF and become available to initiate from the downstream main ORF. It is perplexing to find that uORFs play either stimulating or inhibiting roles in the translation of main ORFs. This conundrum suggests that the uORF number, length, position, and other features might be critical for the overall regulatory effects. Notably, UV-induced DNA damage triggers selective translation of mRNAs containing uORFs in the 5'UTR, indicating that the ATF4-like regulatory mechanism is widely adopted by various stress conditions.[126] It will be desirable to identify stress-specific genes whose mRNA translation depends on specific type of uORFs.

In addition to regulatory roles of uORF mentioned above, the *de novo* translational products of uORF could have direct cellular functions. For instance, small peptides generate by uORFs in fruit fly exert critical functions in development.[127] Given the multiple roles of uORFs in translation control, the importance of UTR region in gene expression cannot be overemphasized. Recent technical advances in capturing 5 termini of transcriptome have uncovered an unexpected heterogeneity of leader sequences in many transcripts.[128] Remarkably, yeast cells produce mRNA isoforms with distinct ends under different growth conditions based on carbon sources.[129] The 5'end heterogeneity in transcripts is

supposed to generate a variety of uORF configuration, further supporting the critical role of uORF in modulating gene expression.

1.4.4 Specialized Ribosomes

As a ribonucleoprotein particle responsible for the catalysis of peptide bond formation, the ribosome has long been considered a "molecular machine" with little intrinsic regulatory potential. A growing body of evidence suggests that ribosome heterogeneity prevails across species, under different developmental stages, and in varied tissues.[130] Variation in ribosome composition, in both rRNA and ribosome proteins, provides a regulatory mechanism to the translation machinery. A clear example is illustrated in E. *coli*, in which a stress-induced endonuclease MazF cleaves the 16S rRNA and removes the anti-Shine-Dalgarno sequence.[131] The resultant "stress ribosome" selectively translates the leaderless mRNAs, a group of transcripts also generated by MazF. Similar to the stress ribosome and transcripts generated by MazF in E. *coli*, eukaryotic cells might also rely on unique interactions between the distinctive component of specialized ribosomes and the *cis*-element on transcripts to achieve functional specificity.[132] In yeast, deletion of RPS25 didn't affect cap-dependent translation but influenced the IRES-mediated translation by hepatitis C virus (HCV) and cricket paralysis virus (CrPV).[133] Whether RPS25 has similar specificity for cellular IRES remains to be elucidated. In plant, RPL24 has been shown to promote re-initiation of ribosomes after completing the uORF translation, thereby promoting the translation of main ORFs.[134]

There are an increasing number of observations that implicate the role of ribosome heterogeneity in selective translation, although mechanistic insight is still lacking. In S. *cerevisiae*, most genes encoding ribosomal proteins have paralogue duplicates and contain introns. A recent study revealed that deleting the intron from one gene copy affected the expression of the other in a nonreciprocal manner.[135] As a result, removing introns within the ribosomal protein genes influenced the cell fitness and growth under stress. These results suggest that ribosomes with distinct composition might form under stress conditions. In mammals, certain ribosome proteins have been found to mediate transcript selectivity during translation. For example, RPL38 is required for translation of Homeobox mRNAs during mouse development.[136] A recent study reported that chicken erythrocytic progenitors transformed by v-erbA oncogene led to the formation of specialized ribosome devoid of PRL11.[137] It remains to be elucidated how specialized ribosomes achieve the selectivity of specific mRNAs. The interplay between specialized ribosomes and the *cis* sequence elements of transcripts adds a novel layer of translational control under stress conditions.

1.5 Conclusions

The field of translational reprogramming has made great progress over the past decade, in large part stemming from technological developments such as ribosome profiling.[8, 9] The next decade should provide both a broader view of translational regulation, as huge data sets of translatome are integrated, and a vastly more detailed

view, as structural studies continuously uncover actions of the translation machinery at the atomic level. The ability of cells to adapt to stress is crucial for their survival. Regulation of global protein synthesis coupled with selective translation allows cells to rapidly respond to a variety of stress conditions. Although accumulating evidence has begun to divulge multiple signalling pathways in the stress response, more questions than answers are brought up by studies of cellular adaptation strategies involving translational reprogramming. For instance, why is the translation of individual mRNA not equally affected by common effectors acting on cap recognition or ternary complex formation? What are the precise mechanisms by which subsets of mRNAs override the repression of protein synthesis? Given the fact that uORFs are frequent in genes with critical biological functions, how does evolution exploit this element for regulatory purposes? With the prevailing mRNA modifications and complex ribosome heterogeneity, how is the imposing goal of coordinating the expression of thousands of transcripts achieved in a cell? It will be exciting to watch the unveiling of answers to these questions and to see the inevitable elegant surprises that will emerge.

As we gain better insight into the mechanisms of translation it is clear that the combination of emerging technologies will paint a multifaceted picture of this paramount cellular process. Elucidating the mechanisms underlying translational reprogramming during stress will not only shed light on the fundamental principles of translation, but also provide deeper insight of the pathophysiology of human diseases.[138, 139] Stress conditions are often an underlying cause of human diseases, including diabetes, neurodegenerative disorders, and cancer. In particular, cancer cells

proliferate rapidly under limited nutrients and are relatively resistant to environmental stress. It is thus critical to understand how abnormal cells alter stress responsive pathways at the translational level. Interestingly, protein translation in cancer cells is coupled to the transcription network centered on heat shock factor 1 (HSF1) and this link supports the anabolic malignant phenotype.[140] Disrupting this linkage using translation initiation inhibitors showed great promise in suppressing tumor growth. A better understanding of translational reprogramming in stress response might ultimately lead to the development of new therapeutic strategies for human diseases.

1.6 Acknowledgements

We apologize to those whose work could not be cited owing to space constraints. We'd like to thank Qian lab members for critical reading of the manuscript. B.L. is a recipient of the Genomics Scholars Award from Center for Vertebrate Genomics at Cornell. S.-B.Q. is supported by grants from National Institutes of Health (1 DP2 OD006449-01, 1R01AG042400-01A1), Ellison Medical Foundation (AG-NS-0605-09), and DOD Exploration-Hypothesis Development Award (W81XWH-11-1-0236).

REFERENCES

1. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008, 26:1135-1145.

2. Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 2009, 136:731-745.

3. Spriggs KA, Bushell M, Willis AE. Translational regulation of gene expression during conditions of cell stress. *Mol Cell* 2010, 40:228-237.

4. Holcik M, Sonenberg N. Translational control in stress and apoptosis. *Nat Rev Mol Cell Biol* 2005, 6:318-327.

5. Aitken CE, Lorsch JR. A mechanistic overview of translation initiation in eukaryotes. *Nat Struct Mol Biol* 2012, 19:568-576.

6. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009, 324:218-223.

7. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* 2012, 7:1534-1550.

8. Michel AM, Baranov PV. Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley Interdiscip Rev RNA* 2013, 4:473-90.

9. Kuersten S, Radek A, Vogel C, Penalva LO. Translation regulation gets its 'omics' moment. *Wiley Interdiscip Rev RNA* 2013,4:617-30.

10. Panniers R. Translational control during heat shock. *Biochimie* 1994, 76:737-747.

11. Richter K, Haslbeck M, Buchner J. The heat shock response: life on the verge of death. *Mol Cell* 2010, 40:253-266.

12. Jackson RJ, Hellen CU, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 2010, 11:113-127.

13. Gebauer F, Hentze MW. Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* 2004, 5:827-835.

14. Marintchev A, Edmonds KA, Marintcheva B, Hendrickson E, Oberer M, Suzuki C, Herdy B, Sonenberg N, Wagner G. Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. *Cell* 2009, 136:447-460.

15. Schutz P, Bumann M, Oberholzer AE, Bieniossek C, Trachsel H, Altmann M, Baumann U. Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc Natl Acad Sci U S A* 2008, 105:9564-9569.

16. Gross JD, Moerke NJ, von der Haar T, Lugovskoy AA, Sachs AB, McCarthy JE, Wagner G. Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell* 2003, 115:739-750.

17. Pisarev AV, Hellen CU, Pestova TV. Recycling of eukaryotic posttermination ribosomal complexes. *Cell* 2007, 131:286-299.

18. Pestova TV, Lomakin IB, Lee JH, Choi SK, Dever TE, Hellen CU. The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* 2000, 403:332-335.

19. Skabkin MA SO, Hellen CUT, PestovaSee TV. Reinitiation and Other Unconventional Posttermination Events during Eukaryotic Translation. *Mol Cell* 2013, 51:249-64.

20. Pause A, Belsham GJ, Gingras AC, Donze O, Lin TA, Lawrence JC, Jr., Sonenberg N. Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function. *Nature* 1994, 371:762-767.

21. Ma XM, Blenis J. Molecular mechanisms of mTOR-mediated translational control. *Nat Rev Mol Cell Biol* 2009, 10:307-318.

22. Zoncu R, Efeyan A, Sabatini DM. mTOR: from growth signal integration to cancer, diabetes and ageing. *Nat Rev Mol Cell Biol* 2011, 12:21-35.

23. Jewell JL, Russell RC, Guan KL. Amino acid signalling upstream of mTOR. *Nat Rev Mol Cell Biol* 2013, 14:133-139.

24. Sancak Y, Peterson TR, Shaul YD, Lindquist RA, Thoreen CC, Bar-Peled L, Sabatini DM. The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. *Science* 2008, 320:1496-1501.

25. Kim E, Goraksha-Hicks P, Li L, Neufeld TP, Guan KL. Regulation of TORC1 by Rag GTPases in nutrient response. *Nat Cell Biol* 2008, 10:935-945.

26. Bar-Peled L, Schweitzer LD, Zoncu R, Sabatini DM. Ragulator is a GEF for the rag GTPases that signal amino acid levels to mTORC1. *Cell* 2012, 150:1196-1208.

27. Bar-Peled L, Chantranupong L, Cherniack AD, Chen WW, Ottina KA, Grabiner BC, Spear ED, Carter SL, Meyerson M, Sabatini DM. A Tumor suppressor complex with GAP activity for the Rag GTPases that signal amino acid sufficiency to mTORC1. *Science* 2013, 340:1100-1106.

28. Inoki K, Li Y, Xu T, Guan KL. Rheb GTPase is a direct target of TSC2 GAP activity and regulates mTOR signaling. *Genes Dev* 2003, 17:1829-1834.

29. Zhang Y, Gao X, Saucedo LJ, Ru B, Edgar BA, Pan D. Rheb is a direct target of the tuberous sclerosis tumour suppressor proteins. *Nat Cell Biol* 2003, 5:578-581.

30. Inoki K, Li Y, Zhu T, Wu J, Guan KL. TSC2 is phosphorylated and inhibited by Akt and suppresses mTOR signalling. *Nat Cell Biol* 2002, 4:648-657.

31. Potter CJ, Pedraza LG, Xu T. Akt regulates growth by directly phosphorylating Tsc2. *Nat Cell Biol* 2002, 4:658-665.

32. Inoki K, Zhu T, Guan KL. TSC2 mediates cellular energy response to control cell growth and survival. *Cell* 2003, 115:577-590.

33. Benjamin D, Hall MN. TSC on the peroxisome controls mTORC1. *Nat Cell Biol* 2013, 15:1135-1136.

34. Braunstein S, Badura ML, Xi Q, Formenti SC, Schneider RJ. Regulation of protein synthesis by ionizing radiation. *Mol Cell Biol* 2009, 29:5645-5656.

35. Connolly E, Braunstein S, Formenti S, Schneider RJ. Hypoxia inhibits protein synthesis through a 4E-BP1 and elongation factor 2 kinase pathway controlled by mTOR and uncoupled in breast cancer cells. *Mol Cell Biol* 2006, 26:3955-3965.

36. Budanov AV, Karin M. p53 target genes sestrin1 and sestrin2 connect genotoxic stress and mTOR signaling. *Cell* 2008, 134:451-460.

37. Alexander A, Cai SL, Kim J, Nanez A, Sahin M, MacLean KH, Inoki K, Guan KL, Shen J, Person MD, et al. ATM signals to TSC2 in the cytoplasm to regulate mTORC1 in response to ROS. *Proc Natl Acad Sci U S A* 2010, 107:4153-4158.

38. Haghighat A, Mader S, Pause A, Sonenberg N. Repression of cap-dependent translation by 4E-binding protein 1: competition with p220 for binding to eukaryotic initiation factor-4E. *EMBO J* 1995, 14:5701-5709.

39. Larsson O, Morita M, Topisirovic I, Alain T, Blouin MJ, Pollak M, Sonenberg N. Distinct perturbation of the translatome by the antidiabetic drug metformin. *Proc Natl Acad Sci U S A* 2012, 109:8977-8982.

40. Thoreen CC, Chantranupong L, Keys HR, Wang T, Gray NS, Sabatini DM. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* 2012, 485:109-113.

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

42. Dowling RJ, Topisirovic I, Alain T, Bidinosti M, Fonseca BD, Petroulakis E, Wang X, Larsson O, Selvaraj A, Liu Y, et al. mTORC1-mediated cell proliferation, but not cell growth, controlled by the 4E-BPs. *Science* 2010, 328:1172-1176.

43. Clemens MJ. Targets and mechanisms for the regulation of translation in malignant transformation. *Oncogene* 2004, 23:3180-3188.

44. Deng J, Harding HP, Raught B, Gingras AC, Berlanga JJ, Scheuner D, Kaufman RJ, Ron D, Sonenberg N. Activation of GCN2 in UV-irradiated cells inhibits translation. *Curr Biol* 2002, 12:1279-1286.

45. Jiang HY, Wek RC. GCN2 phosphorylation of eIF2alpha activates NF-kappaB in response to UV irradiation. *Biochem J* 2005, 385:371-380.

46. Liu L, Wise DR, Diehl JA, Simon MC. Hypoxic reactive oxygen species regulate the integrated stress response and cell survival. *J Biol Chem* 2008, 283:31153-31162.

47. Zhang P, McGrath BC, Reinert J, Olsen DS, Lei L, Gill S, Wek SA, Vattem KM, Wek RC, Kimball SR, et al. The GCN2 eIF2alpha kinase is required for adaptation to amino acid deprivation in mice. *Mol Cell Biol* 2002, 22:6681-6688.

48. Anthony TG, McDaniel BJ, Byerley RL, McGrath BC, Cavener DR, McNurlan MA, Wek RC. Preservation of liver protein synthesis during dietary leucine deprivation occurs at the expense of skeletal muscle mass in mice deleted for eIF2 kinase GCN2. *J Biol Chem* 2004, 279:36553-36561.

49. Kong J, Lasko P. Translational control in cellular and developmental processes. *Nat Rev Genet* 2012, 13:383-394.

50. Ayala A, Parrado J, Bougria M, Machado A. Effect of oxidative stress, produced by cumene hydroperoxide, on the various steps of protein synthesis. Modifications of elongation factor-2. *J Biol Chem* 1996, 271:23105-23110.

51. Kang KR, Lee SY. Effect of serum and hydrogen peroxide on the Ca2+/calmodulin-dependent phosphorylation of eukaryotic elongation factor 2(eEF-2) in Chinese hamster ovary cells. *Exp Mol Med* 2001, 33:198-204.

52. Piwien-Pilipuk G, Ayala A, Machado A, Galigniana MD. Impairment of mineralocorticoid receptor (MR)-dependent biological response by oxidative stress and aging: correlation with post-translational modification of MR and decreased ADP-ribosylatable level of elongating factor 2 in kidney cells. *J Biol Chem* 2002, 277:11896-11903.

53. Shenton D, Smirnova JB, Selley JN, Carroll K, Hubbard SJ, Pavitt GD, Ashe MP, Grant CM. Global translational responses to oxidative stress impact upon multiple levels of protein synthesis. *J Biol Chem* 2006, 281:29011-29021.

54. Ryazanov AG, Davydova EK. Mechanism of elongation factor 2 (EF-2) inactivation upon phosphorylation. Phosphorylated EF-2 is unable to catalyze translocation. *FEBS Lett* 1989, 251:187-190.

55. Chen PJ, Huang YS. CPEB2-eEF2 interaction impedes HIF-1alpha RNA translation. *EMBO J* 2012, 31:959-971.

56. Terai K, Hiramoto Y, Masaki M, Sugiyama S, Kuroda T, Hori M, Kawase I, Hirota H. AMP-activated protein kinase protects cardiomyocytes against hypoxic injury through attenuation of endoplasmic reticulum stress. *Mol Cell Biol* 2005, 25:9554-9575.

57. Liu L, Cash TP, Jones RG, Keith B, Thompson CB, Simon MC. Hypoxia-induced energy stress regulates mRNA translation and cell growth. *Mol Cell* 2006, 21:521-531.

58. Kruiswijk F, Yuniati L, Magliozzi R, Low TY, Lim R, Bolder R, Mohammed S, Proud CG, Heck AJ, Pagano M, et al. Coupled activation and degradation of eEF2K regulates protein synthesis in response to genotoxic stress. *Sci Signal* 2012, 5:ra40.

59. Leprivier G, Remke M, Rotblat B, Dubuc A, Mateo AR, Kool M, Agnihotri S, El-Naggar A, Yu B, Prakash Somasekharan S, et al. The eEF2 Kinase Confers Resistance to Nutrient Deprivation by Blocking Translation Elongation. *Cell* 2013, 153:1064-1079.

60. Hussey GS, Chaudhury A, Dawson AE, Lindner DJ, Knudsen CR, Wilce MC, Merrick WC, Howe PH. Identification of an mRNP complex regulating tumorigenesis at the translational elongation step. *Mol Cell* 2011, 41:419-431.

61. Lopez-Diaz FJ, Gascard P, Balakrishnan SK, Zhao J, Del Rincon SV, Spruck C, Tlsty TD, Emerson BM. Coordinate transcriptional and translational

repression of p53 by TGF-beta1 impairs the stress response. *Mol Cell* 2013, 50:552-564.

62. Saini P, Eyler DE, Green R, Dever TE. Hypusine-containing protein eIF5A promotes translation elongation. *Nature* 2009, 459:118-121.

63. Gutierrez E, Shin BS, Woolstenhulme CJ, Kim JR, Saini P, Buskirk AR, Dever TE. eIF5A promotes translation of polyproline motifs. *Mol Cell* 2013, 51:35-45.

64. Ude S, Lassak J, Starosta AL, Kraxenberger T, Wilson DN, Jung K. Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science* 2013, 339:82-85.

65. Doerfel LK, Wohlgemuth I, Kothe C, Peske F, Urlaub H, Rodnina MV. EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. *Science* 2013, 339:85-88.

66. Scuoppo C, Miething C, Lindqvist L, Reyes J, Ruse C, Appelmann I, Yoon S, Krasnitz A, Teruya-Feldstein J, Pappin D, et al. A tumour suppressor network relying on the polyamine-hypusine axis. *Nature* 2012, 487:244-248.

67. Kaiser A. Translational control of eIF5A in various diseases. *Amino Acids* 2012, 42:679-684.

68. Wolin SL, Walter P. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J* 1988, 7:3559-3569.

69. Fredrick K, Ibba M. How the sequence of a gene can tune its translation. *Cell* 2010, 141:227-229.

70. Gray NK, Hentze MW. Regulation of protein synthesis by mRNA structure. *Mol Biol Rep* 1994, 19:195-200.

71. Elf J, Nilsson D, Tenson T, Ehrenberg M. Selective charging of tRNA isoacceptors explains patterns of codon usage. *Science* 2003, 300:1718-1722.

72. Lavner Y, Kotlar D. Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene* 2005, 345:127-138.

73. Kramer G, Boehringer D, Ban N, Bukau B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol* 2009, 16:589-597.

74. Novoa EM, Ribas de Pouplana L. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet* 2012, 28:574-581.

75. Yanagitani K, Kimata Y, Kadokura H, Kohno K. Translational pausing ensures membrane targeting and cytoplasmic splicing of XBP1u mRNA. *Science* 2011, 331:586-589.

76. Fernandez J, Yaman I, Huang C, Liu H, Lopez AB, Komar AA, Caprara MG, Merrick WC, Snider MD, Kaufman RJ, et al. Ribosome stalling regulates IRES-mediated translation in eukaryotes, a parallel to prokaryotic attenuation. *Mol Cell* 2005, 17:405-416.

77. Liu B, Han Y, Qian SB. Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. *Mol Cell* 2013, 49:453-463.

78. Shalgi R, Hurt JA, Krykbaeva I, Taipale M, Lindquist S, Burge CB. Widespread regulation of translation by elongation pausing in heat shock. *Mol Cell* 2013, 49:439-452.

79. Gerashchenko MV, Lobanov AV, Gladyshev VN. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc Natl Acad Sci U S A* 2012, 109:17394-17399.

80. Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB. Rate-limiting steps in yeast protein translation. *Cell* 2013, 153:1589-1601.

81. Racle J, Picard F, Girbal L, Cocaign-Bousquet M, Hatzimanikatis V. A Genome-Scale Integration and Analysis of Lactococcus lactis Translation Data. *PLoS Comput Biol* 2013, 9:e1003240.

82. Firczuk H, Kannambath S, Pahle J, Claydon A, Beynon R, Duncan J, Westerhoff H, Mendes P, McCarthy JE. An in vivo control map for the eukaryotic mRNA translation machinery. *Mol Syst Biol* 2013, 9:635.

83. Yi C, Pan T. Cellular dynamics of RNA modification. *Acc Chem Res* 2011, 44:1380-1388.

84. Chan CT, Pang YL, Deng W, Babu IR, Dyavaiah M, Begley TJ, Dedon PC. Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat Commun* 2012, 3:937.

85. Thompson DM, Lu C, Green PJ, Parker R. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* 2008, 14:2095-2103.

86. Saikia M, Krokowski D, Guan BJ, Ivanov P, Parisien M, Hu GF, Anderson P, Pan T, Hatzoglou M. Genome-wide identification and quantitative analysis of cleaved tRNA fragments induced by cellular stress. *J Biol Chem* 2012, 287:42708-42725.

87. Thompson DM, Parker R. Stressing out over tRNA cleavage. *Cell* 2009, 138:215-219.

88. Yamasaki S, Ivanov P, Hu GF, Anderson P. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J Cell Biol* 2009, 185:35-42.

89. Ivanov P, Emara MM, Villen J, Gygi SP, Anderson P. Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol Cell* 2011, 43:613-623.

90. Czech A, Wende S, Morl M, Pan T, Ignatova Z. Reversible and rapid transfer-RNA deactivation as a mechanism of translational repression in stress. *PLoS Genet* 2013, 9:e1003767.

91. Netzer N, Goodenbour JM, David A, Dittmar KA, Jones RB, Schneider JR, Boone D, Eves EM, Rosner MR, Gibbs JS, et al. Innate immune and chemically triggered oxidative stress modifies translational fidelity. *Nature* 2009, 462:522-526.

92. Laxman S, Sutter BM, Wu X, Kumar S, Guo X, Trudgian DC, Mirzaei H, Tu BP. Sulfur Amino Acids Regulate Translational Capacity and Metabolic Homeostasis through Modulation of tRNA Thiolation. *Cell* 2013, 154:416-429.

93. Bodi Z, Button JD, Grierson D, Fray RG. Yeast targets for mRNA methylation. *Nucleic Acids Res* 2010, 38:5327-5335.

94. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 2012, 149:1635-1646.

95. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 2012, 485:201-206.

96. Leung AK, Sharp PA. MicroRNA functions in stress responses. *Mol Cell* 2010, 40:205-215.

97. Pelletier J, Sonenberg N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 1988, 334:320-325.

98. Gilbert WV. Alternative ways to think about cellular internal ribosome entry. *J Biol Chem* 2010, 285:29033-29038.

99. Hellen CU, Sarnow P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev* 2001, 15:1593-1612.

100. Dobbyn HC, Hill K, Hamilton TL, Spriggs KA, Pickering BM, Coldwell MJ, de Moor CH, Bushell M, Willis AE. Regulation of BAG-1 IRES-mediated translation following chemotoxic stress. *Oncogene* 2008, 27:1167-1174.

101. Sherrill KW, Byrd MP, Van Eden ME, Lloyd RE. BCL-2 translation is mediated via internal ribosome entry during cell stress. *J Biol Chem* 2004, 279:29066-29074.

102. Yang DQ, Halaby MJ, Zhang Y. The identification of an internal ribosomal entry site in the 5'-untranslated region of p53 mRNA provides a novel mechanism for the regulation of its translation following DNA damage. *Oncogene* 2006, 25:4613-4619.

103. Warnakulasuriyarachchi D, Cerquozzi S, Cheung HH, Holcik M. Translational induction of the inhibitor of apoptosis protein HIAP2 during endoplasmic reticulum stress attenuates cell death and is mediated via an inducible internal ribosome entry site element. *J Biol Chem* 2004, 279:17148-17157.

104. Lang KJ, Kappel A, Goodall GJ. Hypoxia-inducible factor-1alpha mRNA contains an internal ribosome entry site that allows efficient translation during normoxia and hypoxia. *Mol Biol Cell* 2002, 13:1792-1801.

105. Bornes S, Prado-Lourenco L, Bastide A, Zanibellato C, Iacovoni JS, Lacazette E, Prats AC, Touriol C, Prats H. Translational induction of VEGF internal ribosome entry site elements during the early response to ischemic stress. *Circ Res* 2007, 100:305-308.

106. Al-Fageeh MB, Smales CM. Cold-inducible RNA binding protein (CIRP) expression is modulated by alternative mRNAs. *RNA* 2009, 15:1164-1176.

107. Coldwell MJ, deSchoolmeester ML, Fraser GA, Pickering BM, Packham G, Willis AE. The p36 isoform of BAG-1 is translated by internal ribosome entry following heat shock. *Oncogene* 2001, 20:4095-4100.

108. Wellensiek BP, Larsen AC, Stephens B, Kukurba K, Waern K, Briones N, Liu L, Snyder M, Jacobs BL, Kumar S, et al. Genome-wide profiling of human cap-independent translation-enhancing elements. *Nat Methods* 2013.

109. Martinez-Salas E, Ramos R, Lafuente E, Lopez de Quinto S. Functional interactions in internal translation initiation directed by viral and cellular IRES elements. *J Gen Virol* 2001, 82:973-984.

110. Wolin SL, Cedervall T. The La protein. *Annu Rev Biochem* 2002, 71:375-403.

111. Pickering BM, Mitchell SA, Spriggs KA, Stoneley M, Willis AE. Bag-1 internal ribosome entry segment activity is promoted by structural changes

mediated by poly(rC) binding protein 1 and recruitment of polypyrimidine tract binding protein 1. *Mol Cell Biol* 2004, 24:5595-5605.

112. Grover R, Ray PS, Das S. Polypyrimidine tract binding protein regulates IRES-mediated translation of p53 isoforms. *Cell Cycle* 2008, 7:2189-2198.

113. Gilbert WV, Zhou K, Butler TK, Doudna JA. Cap-independent translation is required for starvation-induced differentiation in yeast. *Science* 2007, 317:1224-1227.

114. Merrick WC. Cap-dependent and cap-independent translation in eukaryotic systems. *Gene* 2004, 332:1-11.

115. Sun J, Conn CS, Han Y, Yeung V, Qian SB. PI3K-mTORC1 attenuates stress response by inhibiting cap-independent Hsp70 translation. *J Biol Chem* 2011, 286:6791-6800.

116. Kozak M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 2002, 299:1-34.

117. Iacono M, Mignone F, Pesole G. uAUG and uORFs in human and rodent 5'untranslated mRNAs. *Gene* 2005, 349:97-105.

118. Lee S, Liu B, Huang SX, Shen B, Qian SB. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A* 2012, 109:E2424-2432.

119. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 2011, 147:789-802.

120. Descombes P, Schibler U. A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. *Cell* 1991, 67:569-579.

121. Stern-Ginossar N, Weisburd B, Michalski A, Le VT, Hein MY, Huang SX, Ma M, Shen B, Qian SB, Hengel H, et al. Decoding human cytomegalovirus. *Science* 2012, 338:1088-1093.

122. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A* 2009, 106:7507-7512.

123. Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* 2012, 335:552-557.

124. Mueller PP, Hinnebusch AG. Multiple upstream AUG codons mediate translational control of GCN4. *Cell* 1986, 45:201-207.

125. Dever TE, Feng L, Wek RC, Cigan AM, Donahue TF, Hinnebusch AG. Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast. *Cell* 1992, 68:585-596.

126. Powley IR, Kondrashov A, Young LA, Dobbyn HC, Hill K, Cannell IG, Stoneley M, Kong YW, Cotes JA, Smith GC, et al. Translational reprogramming following UVB irradiation is mediated by DNA-PKcs and allows selective recruitment to the polysomes of mRNAs encoding DNA repair enzymes. *Genes Dev* 2009, 23:1207-1220.

127. Kondo T, Plaza S, Zanet J, Benrabah E, Valenti P, Hashimoto Y, Kobayashi S, Payre F, Kageyama Y. Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. *Science* 2010, 329:336-339.

128. Arribere JA, Gilbert WV. Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res* 2013, 23:977-987.

129. Pelechano V, Wei W, Steinmetz LM. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* 2013, 497:127-131.

130. Xue S, Barna M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat Rev Mol Cell Biol* 2012, 13:355-369.

131. Vesper O, Amitai S, Belitsky M, Byrgazov K, Kaberdina AC, Engelberg-Kulka H, Moll I. Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in Escherichia coli. *Cell* 2011, 147:147-157.

132. Gilbert WV. Functional specialization of ribosomes? *Trends Biochem Sci* 2011, 36:127-132.

133. Landry DM, Hertz MI, Thompson SR. RPS25 is essential for translation initiation by the Dicistroviridae and hepatitis C viral IRESs. *Genes Dev* 2009, 23:2753-2764.

134. Park HS, Himmelbach A, Browning KS, Hohn T, Ryabova LA. A plant viral "reinitiation" factor interacts with the host translational machinery. *Cell* 2001, 106:723-733.

135. Parenteau J, Durand M, Morin G, Gagnon J, Lucier JF, Wellinger RJ, Chabot B, Elela SA. Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. *Cell* 2011, 147:320-331.

136. Kondrashov N, Pusic A, Stumpf CR, Shimizu K, Hsieh AC, Xue S, Ishijima J, Shiroishi T, Barna M. Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell* 2011, 145:383-397.

137. Nguyen-Lefebvre AT, Leprun G, Morin V, Vinuelas J, Coute Y, Madjar JJ, Gandrillon O, Gonin-Giraud S. V-erbA generates ribosomes devoid of RPL11 and regulates translational activity in avian erythroid progenitors. *Oncogene* 2013.

138. Sonenberg N, Hinnebusch AG. New modes of translational control in development, behavior, and disease. *Mol Cell* 2007, 28:721-729.

139. Gkogkas CG, Khoutorsky A, Ran I, Rampakakis E, Nevarko T, Weatherill DB, Vasuta C, Yee S, Truitt M, Dallaire P, et al. Autism-related deficits via dysregulated eIF4E-dependent translational control. *Nature* 2013, 493:371-377.

140. Santagata S, Mendillo ML, Tang YC, Subramanian A, Perley CC, Roche SP, Wong B, Narayan R, Kwon H, Koeva M, et al. Tight coordination of protein translation and HSF1 activation supports the anabolic malignant state. *Science* 2013, 341:1238303.

CHAPTER 2

Co-Translational Response to Proteotoxic Stress by Elongation Pausing of Ribosomes

This work was submitted April 2012 and first published online on January 03, 2013, doi: 10.1016/j.molcel.2012.12.001. The manuscript was published as Liu B, Han Y, and Qian SB. **Co-translational response to proteotoxic stress by elongation pausing of ribosomes**. *Mol Cell* 2013; 49(3):453-463. Minor modifications have been made for reprint here.

2.1 Abstract

Translational control permits cells to respond swiftly to changing environment. Rapid attenuation of global protein synthesis under stress conditions has been largely ascribed to the inhibition of translation initiation. Here we report that intracellular proteotoxic stress reduces global protein synthesis by halting ribosomes on transcripts during elongation. Deep sequencing of ribosome-protected mRNA fragments reveals an early elongation pausing, roughly at the site where nascent polypeptide chains emerge from the ribosomal exit tunnel. Inhibiting endogenous chaperone molecules by a dominant-negative mutant or chemical inhibitors recapitulates the early elongation pausing, suggesting a dual role of molecular chaperones in facilitating polypeptide elongation and co-translational folding. Our results further support the

chaperone "trapping" mechanism in promoting the passage of nascent chains. Our study reveals that translating ribosomes fine-tune the elongation rate by sensing the intracellular folding environment. The early elongation pausing represents a co-translational stress response to maintain the intracellular protein homeostasis.

2.2 Introduction

Protein misfolding imposes a major risk to the health of cells and organisms. An elaborate protein quality control (PQC) system has been laid down during evolution to maintain protein homeostasis – a delicate balance between protein synthesis, folding, and degradation (Bukau et al., 2006; Frydman, 2001; Hartl et al., 2011). Molecular chaperones are "cellular lifeguards" that govern the integrity of proteome. By interacting with different co-chaperones and co-factors, Hsp70 family proteins actively participate in protein triage decisions from folding, degradation, to aggregation (McClellan et al., 2005; Zhang and Qian, 2011). Most recent studies highlighted the robust network of chaperones acting co-translationally on nascent chains in eukaryotic (del Alamo et al., 2011) as well as prokaryotic cells (Oh et al., 2011). Interestingly, prokaryotes and eukaryotes have evolved distinct ribosome-associated chaperone systems (Kramer et al., 2009). In S. *cerevisiae*, two ribosome-associated systems interact with newly synthesized polypeptides, the nascent chain-associated complex (NAC) and the Hsp70-based Ssb/Ssz/Zuo triad system (Kampinga and Craig, 2010). Both systems are physically located in close proximity at the peptide exit tunnel of ribosomes. The ribosome-associated chaperone system also

exists in mammals, although its functionality is not fully understood (Jaiswal et al., 2011). Despite the wide appreciation of the impact that this chaperone system may have on co-translational folding, little is known about how the ribosome-associated chaperone system regulates the process of translation per se.

mRNA translation can be divided into three stages - initiation, elongation and termination. Regulation of translation occurs predominantly during initiation phase (Sonenberg and Hinnebusch, 2009; Spriggs et al., 2010). The initiation is a complex multi-step process governed by a large number of protein factors and involves mRNA 5'-cap recognition, scanning and start codon recognition (Gray and Wickens, 1998; Jackson et al., 2010). Much attention has been focused on the role of translation initiation factors (eIFs) in the assembly of elongation-competent ribosome complexes. However, after the commitment of polypeptide synthesis, the regulatory steps during elongation remain poorly understood.

Given the fact that translation consumes a lion's share of energy, cells often reduce global protein synthesis under most, if not all, types of adverse conditions. The global repression of protein synthesis not only saves the cellular energy, but also relieves the burden of the PQC system due to the less protein production (Holcik and Sonenberg, 2005). Current models for the mechanism governing this translational attenuation are largely limited to the initiation stage. For instance, eIF4F complex-mediated cap recognition and eIF2-controlled ternary complex formation are key initiation targets in controlling global mRNA translation (Ma and Blenis, 2009; Ron and Walter, 2007). In response to stresses, the shutdown of protein synthesis is, in general, mediated either by the inhibition of 43S complex loading to the 5' end cap

and/or reducing the amount of ternary complex that is available. Despite the well-documented role of these initiation regulators, it remains surprisingly obscure whether the 80S ribosome, once assembled on the mRNA, maintains the responsiveness to protein misfolding during elongation.

Here we report that proteotoxic stress triggers ribosomal pausing during elongation. Remarkably, the pausing occurs primarily near the site where nascent polypeptide chains emerge from the ribosomal exit tunnel. We demonstrate that the early elongation pausing is induced by the sequestration of chaperone molecules by misfolded proteins. Our results expand the critical role of chaperone molecules form co-translational folding to polypeptide elongation. The early elongation pausing of ribosomes thus represents a mechanism of co-translational stress response to maintain intracellular protein homeostasis.

2.3 Results

2.3.1 Proteotoxic Stress Attenuates Global Protein Synthesis

Intracellular accumulation of misfolded proteins is a common feature of a variety of stress conditions. To induce misfolding of newly synthesized polypeptides without massively perturbing cellular functions, we used an amino acid analog $_{L}$-azetidine-2-carboxylic acid (AZC) that competes with proline during amino acid incorporation (Goldberg and Dice, 1974). Once incorporated into proteins in place of proline, AZC potently induces protein misfolding and degradation (Qian et al., 2010; Trotter et al., 2002). Pre-exposure of HEK293 cells to 10 mM AZC resulted in a marked reduction

of [^{35}S] incorporation (Figure 1A). In agreement with the enhanced degradation of AZC-incorporated polypeptides, pulse-chase analysis showed an increased turnover of [^{35}S] labeled proteins in the presence of AZC (Figure S1A). We asked whether proteasome inhibition would prevent the loss of [^{35}S] incorporation by blocking the degradation. To our surprise, adding proteasome inhibitor MG132 further decreased the total amount of [^{35}S] incorporation (Figure 1A). This was not due to the side effects of MG132 because adding this inhibitor alone only partially reduced the level of [^{35}S] incorporation. Since the AZC-induced misfolded polypeptides progressively accumulate under proteasome inhibition, it appears that the intracellular proteotoxic stress triggers a rapid attenuation of translation. To substantiate this finding further, we analyzed the polysome profiles by velocity sedimentation of lysates in sucrose gradients. Treatment with either AZC or MG132 alone had minor effects on the polysome formation (Figure 1B). In contrast, the presence of both AZC and MG132 markedly disassembled the polysomes with an approximately 5 fold decrease in the polysome/ monosome (P/M) ratio.

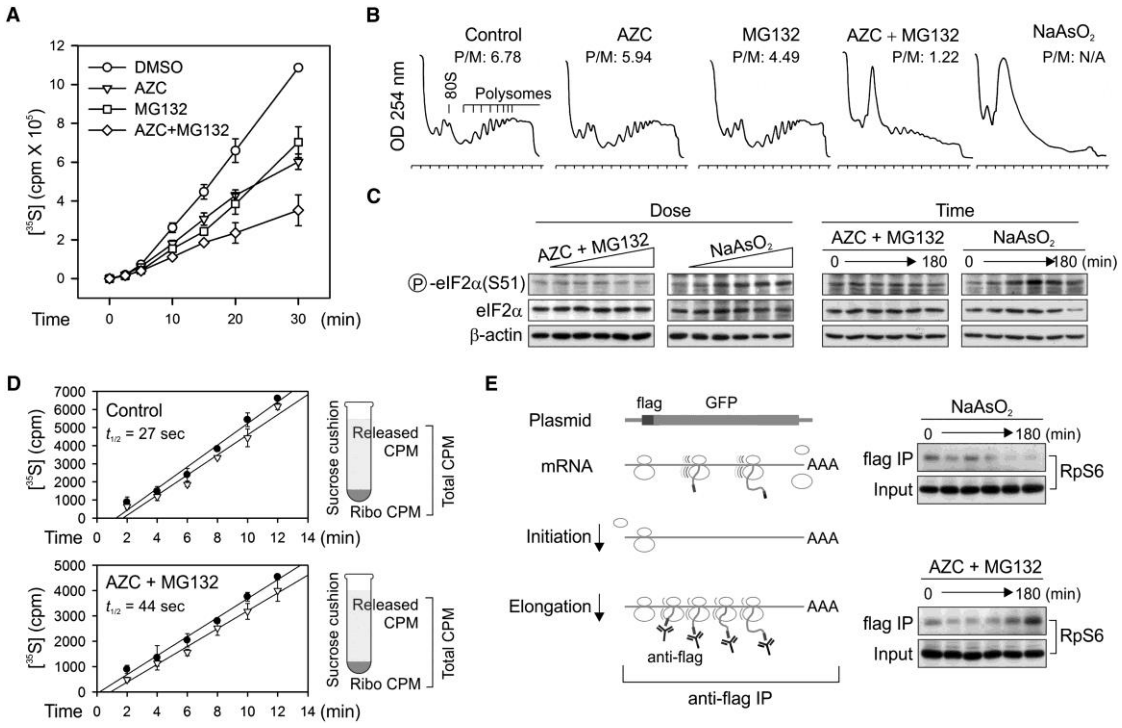


Figure 2-1. Proteotoxic stress attenuates protein synthesis by affecting translation elongation.

(A) Global protein synthesis in HEK293 cells treated with either 10 mM AZC, or 20 μM MG132, or both. [^{35}S] radioactivity of trichloroacetic acid (TCA)-insoluble material was measured at given times. Means ± SEM of four experiments are shown.

(B) Polysome profiles were determined using sucrose gradient sedimentation. HEK293 cells were pre-treated with either 10 mM AZC, or 20 μM MG132, or both for 60 min followed by polysome preparation. P/M ratio is calculated by comparing areas under the polysome and 80S peak.

(C) HEK293 cells were treated with increasing doses of AZC (from 0 to 25 mM with 5-fold dilution) in the presence of 20 μM MG132 for 60 min, or increasing doses of $NaAsO_2$ (from 0 to 1 mM with 2-fold dilution) for 60 min (left two panels), followed by immunoblotting using antibodies as indicated. The right two panels show the immunoblotting results of cells treated with 10 mM AZC and 20 μM MG132 or 500 μM $NaAsO_2$ for various times (0, 10, 30, 60, 120, and 180 min).

(D) The ribosomal half-transit time was determined in the absence or presence of 10 mM AZC and 20 μM MG132. Fitting lines of [^{35}S] incorporation into total (filled circle) and completed (open triangle) protein synthesis are obtained by linear regression. Means ± SEM of three experiments are shown.

(E) Schematic for nascent chain immunoprecipitation assay to differentiate elongation defect from initiation deficiency (left panel). HEK293 cells expressing Flag-GFP were pre-treated with 10 mM AZC and 20 μM MG132 or 500 μM $NaAsO_2$ for various times (0, 10, 30, 60, 120, and 180 min). Immuno-precipitation was performed using anti-Flag antibody-coated beads followed by immunoblotting with anti-RpS6 antibody. The 0 time point serves as the control condition without any drug treatment. See also Figure 2-S1.

2.3.2 Proteotoxic Stress Affects Primarily Translation Elongation

To investigate the mechanisms underlying the proteotoxic stress-induced translational attenuation, we examined the phosphorylation status of eIF2α, a prominent initiation regulator in the unfolded protein response (Ron and Walter, 2007). In contrast to sodium arsenite ($NaAsO_2$), a known inducer of eIF2α phosphorylation, treating cells with both AZC and MG132 at increasing doses and for extended times had little effect on eIF2α phosphorylation (Figure 1C). Additionally, we observed no change in the phosphorylation of S6 and its kinase S6K1, one of the downstream targets of mammalian target of rapamycin complex 1 (mTORC1) (Jackson et al., 2010; Ma and Blenis, 2009) (Figure S1B). Thus, the intracellular proteotoxic stress does not affect primarily the translation initiation regulators, at least in the early stage.

We next examined whether proteotoxic stress inhibits protein synthesis by interfering with post-initiation events, such as elongation. One way to distinguish elongation from initiation is the formation of stress granules (SG). Inhibiting translation initiation triggers SG formation, whereas blocking translation elongation prevents this process (Buchan and Parker, 2009; Kedersha et al., 2000). Unlike sodium arsenite treatment that induced an evident SG formation, adding both AZC and MG132 to cells failed to induce any discernible SG formation (Figure S1C). Thus proteotoxic stress likely affects translation elongation rather than initiation. To assess independently whether the reduced protein synthesis under proteotoxic stress was primarily due to defective elongation, we determined ribosomal transit times in these cells. The ribosomal transit time refers to the time required for a ribosome, after

initiation, to traverse an average-sized mRNA and release the completed polypeptide chain (Nielsen and McConkey, 1980). The estimated half-transit time ($t_{1/2}$) in the presence of both AZC and MG132 (44s) was ~1.6 fold longer than that in control cells (27s) (Figure 1D), confirming that proteotoxic stress significantly reduced the elongation rate of polypeptide synthesis. Additionally, we conducted an elongation chase experiment using a synthesized *firefly* luciferase (Fluc) mRNA in lysates programmed from cells with or without proteotoxic stress. Compared to the control, the stressed cell lysates showed a delayed accumulation of Fluc activity (Figure S1D), further indicating a slowdown of elongation process under proteotoxic stress.

To examine whether the stalled ribosome during elongation was still associated with the newly synthesized polypeptide, we performed nascent chain immunoprecipitation followed by detection of ribosomal small subunit S6 (RpS6). We established a HEK293 cell line stably expressing a GFP reporter with an NH2-terminal Flag-tag (Figure 1E). We enriched the ribosome complexes bearing the partially synthesized GFP by anti-Flag immunoprecipitation. Arsenite treatment led to a progressive loss of the associated RpS6 in a time course-dependent manner (Figure 1E, right top panel), which is consistent with the inhibition of translation initiation. Remarkably, treating cells with both AZC and MG132 resulted in an accumulation of RpS6 in the anti-Flag precipitates, a clear evidence of paused ribosomes on the mRNA during elongation. The prolonged ribosome association with the nascent chain persists in the polysome fractions of these cells (Figures S1E and S1F). Taken together, our results strongly indicate that proteotoxic stress acts at the level of translation elongation to suppress protein synthesis.

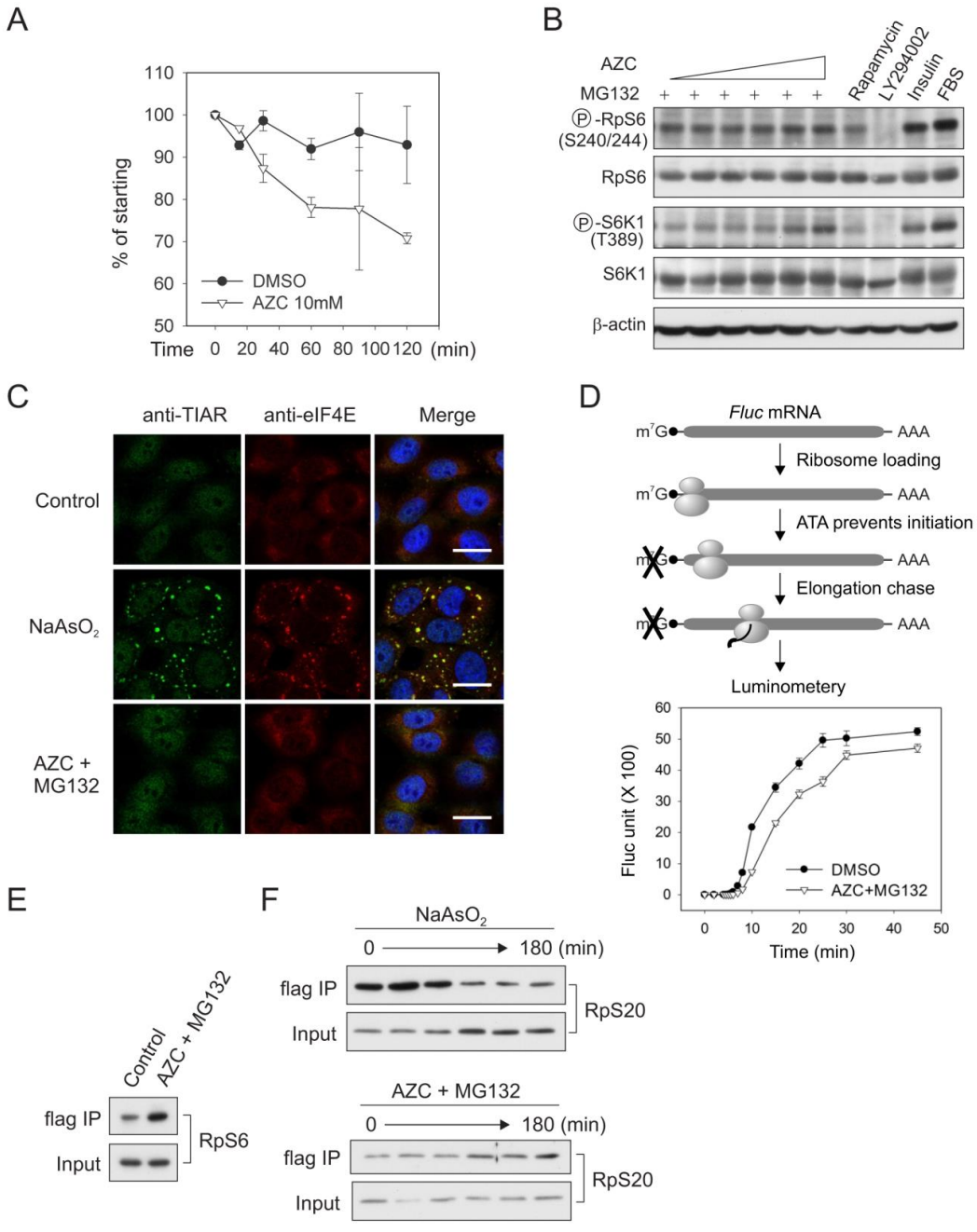

A
% of starting
Time 0 20 40 60 80 100 120 (min)
DMSO
AZC 10mM
B
AZC
MG132
Rapamycin
LY294002
Insulin
FBS
Ⓟ-RpS6 (S240/244)
RpS6
Ⓟ-S6K1 (T389)
S6K1
β-actin
C
anti-TIAR
anti-eIF4E
Merge
Control
NaAsO2
AZC + MG132
D
Fluc mRNA
m7G
AAA
Ribosome loading
ATA prevents initiation
Elongation chase
Luminometery
Fluc unit (X 100)
Time (min)
DMSO
AZC+MG132
E
Control
AZC + MG132
flag IP
Input
RpS6
F
NaAsO2
0
180 (min)
flag IP
Input
RpS20
AZC + MG132
0
180 (min)
flag IP
Input
RpS20

Figure 2-S1. Proteotoxic stress attenuates protein synthesis by affecting translation elongation. Related to Figure 2-1.
(A) Proline analog AZC induces misfolding and degradation of newly synthesized polypeptides. HEK293 cells were pre-treated with 10 mM AZC for 1 h followed by pulse labeling of [^{35}S] Met-Cys for 10 min in the presence of AZC. A chase was performed with unlabeled medium in the absence of AZC. TCA-insoluble radioactivity was measured for cell aliquots at given time points. Means ± SD of two experiments are shown.
(B) Proteotoxic stress does not primarily affect mTORC1 signaling. HEK293 cells were treated with increasing doses of AZC (from 0 to 25 mM with 5-fold dilution) and 20 μM MG132 for 60 min, followed by immunoblotting using antibodies as indicated. Cells were treated with 20 nM rapamycin or 50 μM LY294002 for 3 h as negative controls, whereas 10% FBS re-feeding and 2 μM insulin stimulation for 15 min as positive controls for mTORC1 signaling.
(C) Proteotoxic stress does not trigger the stress granule formation in cells. HeLa cells were treated with 500 μM $NaAsO_2$ or 10 mM AZC and 20 μM MG132 for 60 min followed by immunostaining using anti-TIAR antibody (green channel) and anti-eIF4E (red channel). Nuclei were counterstained with Hoechst. Bar: 10 μm.
(D) Ribosome elongation chase assay in lysates derived from cells with or without proteotoxic stress. Top panel shows a schematic for elongation chase assay using a synthesized Fluc mRNA in an in vitro translation system programmed from cell lysates. Aurintricarboxyllic acid (ATA) is used to inhibit translation initiation after ribosome loading. Bottom panel shows the kinetics of Fluc activity accumulation during in vitro translation in lysates derived from cells with or without proteotoxic stress. Mean ± SD is shown.
(E) HEK293 cells expressing Flag-GFP were treated with 10 mM AZC and 20 μM MG132 for 60 min followed by sucrose gradient sedimentation. Polysome fractions were immune-precipitated using anti-Flag antibody-coated beads followed by immunoblotting of RpS6.
(F) HEK293 cells expressing Flag-GFP were pre-treated with 500 μM $NaAsO_2$ (top panel) or 10 mM AZC and 20 μM MG132 (bottom panel) for various times (0, 10, 30, 60, 120, and 180 min). Immunoprecipitation was performed using anti-Flag antibody-coated beads followed by immunoblotting with anti-RpS20 antibody.

2.3.3 Proteotoxic Stress Triggers Early Elongation Pausing of Ribosomes

A defective translation elongation should result in slower ribosome run-off and the retention of polysomes (Saini et al., 2009). It is surprising to find that the polysomes were largely disassembled in cells treated with both AZC and MG132 (Figure 1B). We considered the possibility that proteotoxic stress primarily induced ribosomal pausing at the early stage of elongation, thereby creating a road block for following ribosomes. To provide a definitive assessment of ribosome positions on mRNAs under proteotoxic stress, we isolated the ribosome protected mRNA

fragments (RPFs) and performed deep-sequencing using methods reported previously (Ingolia et al., 2009). RPF reads obtained from cells with or without proteotoxic stress were of equal quality as evidenced by the similar size distribution and strong 3-nt periodicity after alignment. Notably, AZC and MG132 treatment did not result in global variation in overall ribosome density along each transcript (r = 0.9825) (Figure 2A). To directly visualize the pattern of RPF distribution on individual transcripts, we built a ribosome density map across the entire transcriptome (Figure 2B). Compared to control cells, the presence of both AZC and MG132 led to a clear enrichment of RPF density at the 5' end of coding sequences (CDS) on the vast majority of mRNAs. Meta-gene analysis revealed a pronounced accumulation of RPF reads within the first 50 codon region of transcripts in cells treated with both AZC and MG132 (Figure 2C). We defined the ribosome pausing index (PI) of individual transcript by calculating the normalized ribosome density within a 50-codon window from start codon (5'PI) or stop codon (3'PI) respectively. In cells under proteotoxic stress, the median 5'PI showed more than 2-fold increase as compared to control cells (Figure 2D). Intriguingly, proteotoxic stress also caused an elevation of RPF density in the 5' untranslated region (5'UTR) (Figure 2C), an indication of wide-spread alternative initiation under stress conditions.

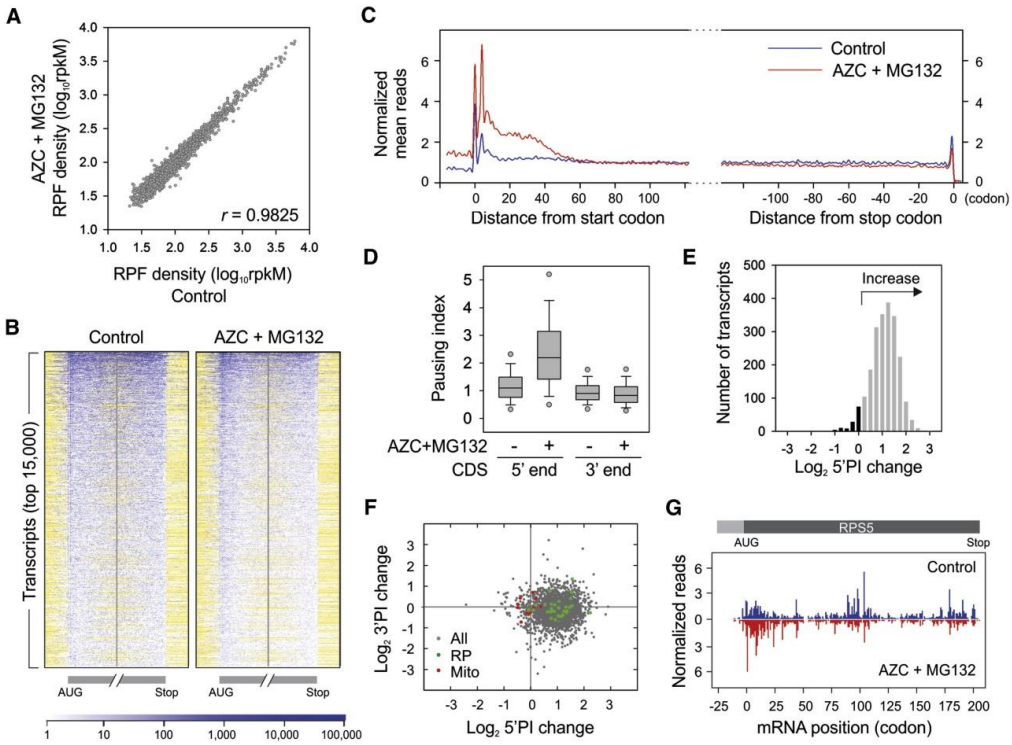


Figure 2-2. Intracellular proteotoxic stress triggers early elongation pausing of ribosomes.

(A) HEK293cells were treated with 10 mM AZC and 20 μM MG132 for 60 min before ribosome profiling. Ribosome densities of cells with or without treatment are plotted for comparison. The density in reads per kilobase of coding sequence per million mapped reads (rpkM) is a measure of overall translation along each transcript.

(B) Ribosome density heat-maps of cells with or without treatment. The entire transcriptome is sorted based on total RPF reads and the top 15,000 transcripts are aligned in row. Both the first and last 160 codon regions of CDS are shown, together with flanking 40 codon untranslated regions. Read density is represented in blue. White indicates regions without reads, whereas yellow indicates regions without sequence. A short 5′ UTR has yellow region before the AUG, whereas a short 3′ UTR has yellow region after the stop codon.

(C) Metagene analysis of early ribosome pausing of cells with or without treatment. Normalized RPF reads are averaged across the entire transcriptome, aligned at either their start (left panel) or stop (right panel) codon, and plotted as smoothed lines.

(D) Ribosome pausing index (PI) is determined in a 50 codon window at the beginning (5′ end) and end (3′ end) of CDS, respectively. Both the 5′ and 3′ PI of each transcript in cells with or without treatment are shown in box plots with single dots as 5th and 95th percentile.

(E) Distribution of 5′PI changes in cells with proteotoxic stress. The $\log_2$ change of 5′PI after AZC and MG132 treatment is plotted, with the increase shown in gray bar and the decrease in black.

(F) Changes of 5′PI and 3′PI after AZC and MG132 treatment. The $\log_2$ change is computed across the entire transcriptome, and presented as a scatter plot with green dots for genes encoding ribosome subunits (RP) and red dots for mitochondria-encoded genes (Mito).

(G) A typical example of early elongation pausing under proteotoxic stress. RPF reads density is shown on the CDS of RPS5 with or without AZC and MG132 treatment. See also Figure 2-S2-S4.

A large portion of mRNAs showed an increased 5'PI in response to proteotoxic stress (Figure 2E, grey bar). However, a small group of transcripts showed less change or even decreased 5'PI (Figure 2E, black bar). At the transcriptome level, neither the CDS length nor the overall translation had any strong correlation with the changes of 5'PI (Figures S2A and S2B). Gene ontology (GO) analysis revealed that genes involving ATP synthesis (e.g., mitochondria-encoded genes) were enriched in the group with decreased 5'PI in response to proteotoxic stress (Figures S2C, S2D and Figure 2F). In contrast, genes with increased 5'PI were involved in cellular processes like RNA metabolism and translation (e.g., ribosomal proteins). As a typical example, proteotoxic stress led to a clear ribosome accumulation near the beginning of RPS5 CDS region (Figure 2G).

Through an independent biological replicate, we confirmed the early elongation pausing of ribosomes in response to proteotoxic stress (Figure S3). The global RPF distribution was highly reproducible across the replicates (Figure S3E). Notably, treatment with either AZC or MG132 alone had little effects on the ribosome dynamics (Figure S4A and S4B). In particular, we saw no unique pausing sites at individual codon in the presence of AZC (Figure S4C). These results argue that the presence of AZC-charged tRNA neither perturbs the intracellular pool of amino acids nor alters the behavior of translating ribosomes. Therefore, it is the accumulation of misfolded proteins that triggers the early elongation pausing.

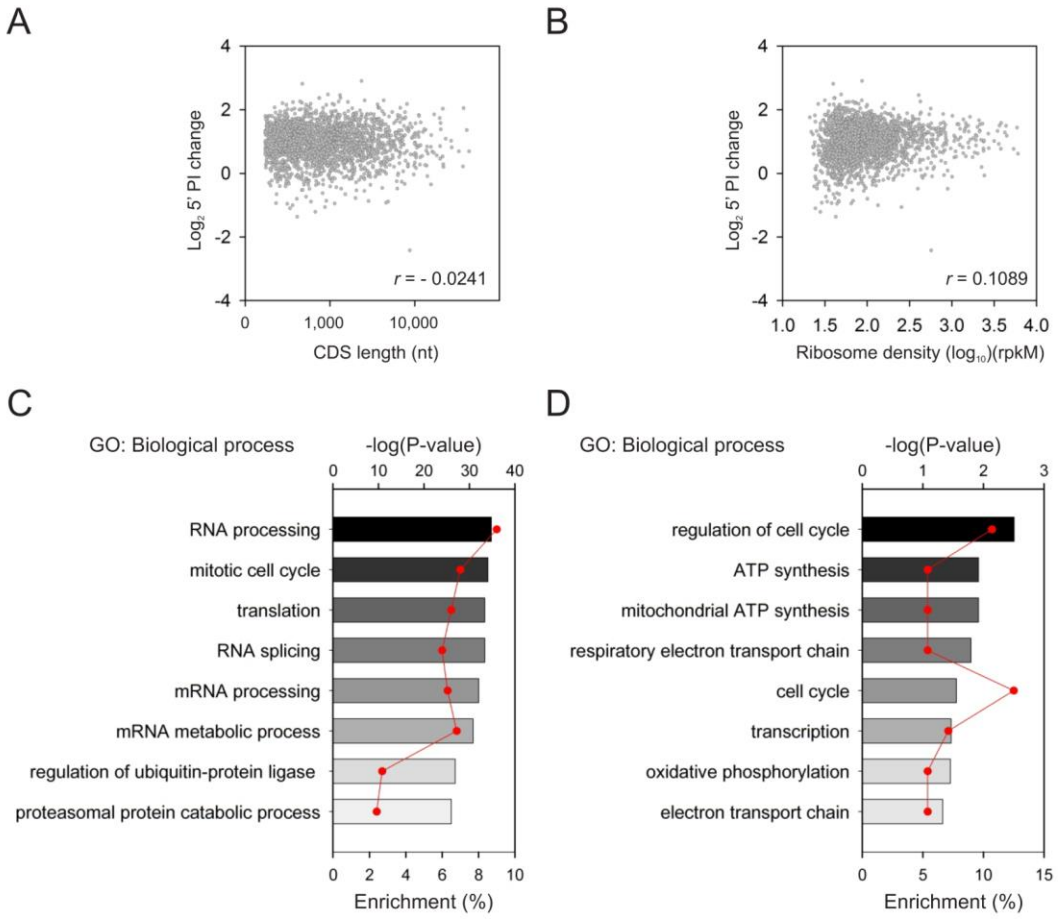


Figure 2-S2. Characterization of early ribosome pausing in response to proteotoxic stress by ribosome profiling. Related to Figure 2-2**.**
(A) The $\log_2$ changes of 5'PI across the entire transcriptome after AZC and MG132 treatment were plotted against the CDS length of each transcript.
(B) The $\log_2$ changes of 5'PI across the entire transcriptome after AZC and MG132 treatment were plotted against the overall translation of each transcript as measured by reads per kilobase million (rpkM).
(C) GO analysis of genes with increased 5'PI (A) and decreased 5'PI (B) in response to proteotoxic stress. The percentage enrichment is plotted on the bottom axis (red line). The significance of the enrichment (bar) is plotted on the top axis as $-\log_{10}$ of P-values; the higher the value, the more significant the enrichment.

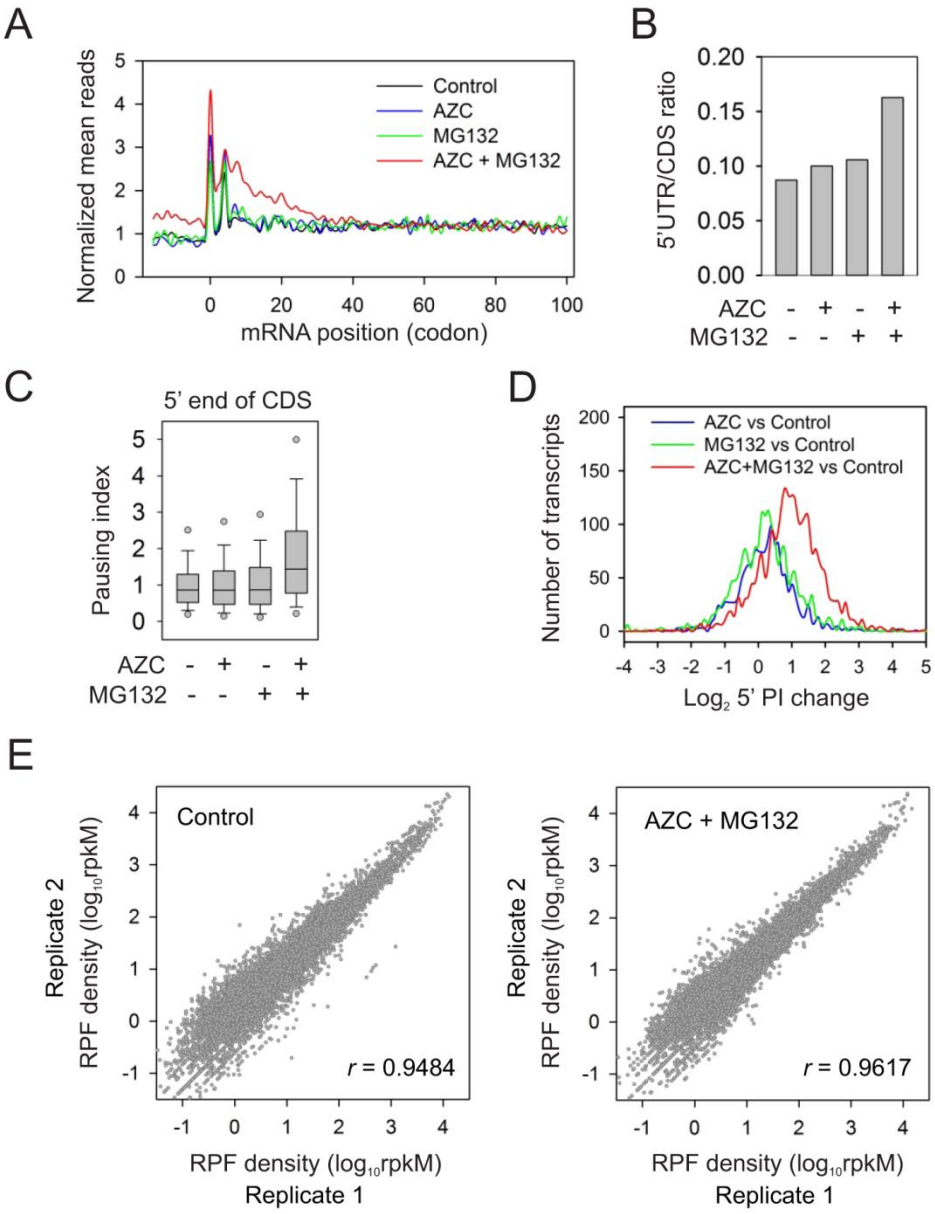


Figure 2-S3. Proteotoxic stress triggers early ribosome pausing in biological replicates. Related to Figure 2-2.

(A) Meta-gene analysis of early ribosome pausing in HEK293 cells treated with either 10 mM AZC, or 20 μM MG132, or both. Normalized RPF density profiles over the entire transcriptome, aligned at their start codon, are shown for the first 100-codon region.

(B) Changes of RPF density on 5'UTR after proteotoxic stress. RPF reads density on 5'UTR after different drug treatments was shown as a ratio of CDS reads.

(C) Ribosome pausing index at the 5' end of CDS was determined in a 50-codon window and shown in box plots.

(D) The $\log_2$ changes of 5'PI across the entire transcriptome after different treatments were determined by direct comparison to the control.

(E) Scatter plots show the RPF densities of two biological replicates.

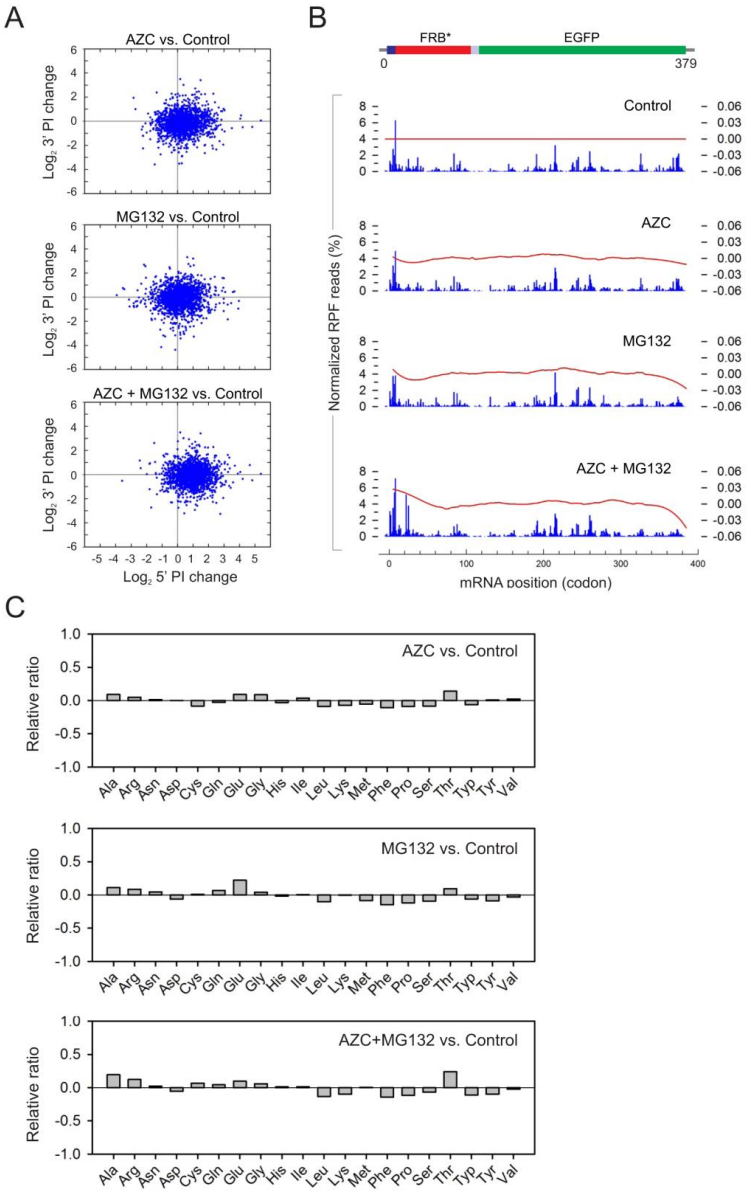


Figure 2-S4. Treatment with AZC or MG132 alone has limited effects on the pattern of RPF reads distribution. Related to Figure 2-2.
(A) The $\log_2$ changes of 5'PI and 3'PI across the entire transcriptome after different treatments were determined by direct comparison to the control and plotted as scatter plots.
(B) Comparison of RPF distribution on FRB*-GFP transcript before and after different drug treatments as indicated. Blue peaks (left axis) represent normalized RPF reads on a codon basis, whereas the red line (right axis) represents the *LOESS* smoothed trend line for single codon peak ratio (sampling proportion = 0.2). *LOESS*: locally weighted scatterplot smoothing.
(C) RPF reads from cells under different treatments were used to determine the codon compositions at the ribosome A-site. Codons encoding same amino acids were summed and the RPF frequency on codons encoding individual amino acids was computed. Compared to control cells, the relative change of RPF frequency after different treatments was plotted.

2.3.4 A Dominant-Negative Hsc70 Mutant Induces Early Elongation Pausing of Ribosomes

The approximate 50 codon region where the elongation pausing occurs under proteotoxic stress corresponds remarkably well to the length of polypeptide needed to fill the exit tunnel of the ribosome (approximately 30-40 amino acids in extended conformation) (Kramer et al., 2009). This raises an intriguing possibility that the changing environment of nascent polypeptides from the ribosome tunnel to the cytosol might influence the dynamics of translating ribosomes. Within the cellular environment, the emerging nascent chains interact with molecular chaperones that guide their folding process. At the forefront is Hsc/Hsp70 that transiently associates with a large fraction of nascent chains (Frydman, 2001; Hansen et al., 1999; Kampinga and Craig, 2010). This led us to hypothesize that the accumulated misfolded proteins titrate out the intracellular chaperone pool and the lack of chaperone association might prevent nascent chains from protruding out of the ribosome exit tunnel. The elongation slowdown at this position likely causes ribosomes to pile up over the first 50-codon region. Supporting the notion that proteotoxic stress sequesters intracellular chaperone molecules, we observed a progressive loss of ribosome associated Hsc70 along with AZC and MG132 treatment (Figure 3A). To test the hypothesis that reduced chaperone availability leads to an early elongation pausing, we first used a dominant-negative mutant Hsc70 (K71M), which sequesters and inactivates the endogenous Hsc70 molecules (Newmyer and Schmid, 2001). The integrated "tet-off" system allows a rapid induction of the transgene expression in HeLa-tTA cells after removal of doxycycline (Dox) (Figure

S5A). After 12h of transgene induction by removal of Dox, [^{35}S] metabolic labeling revealed ~ 40% decrease of the global protein synthesis (Figure 3B). Similar to cells treated with both AZC and MG132, Hsc70(K71M) expression caused disassembly of polysomes with a concomitant increase of 80S peak (Figure 3C).

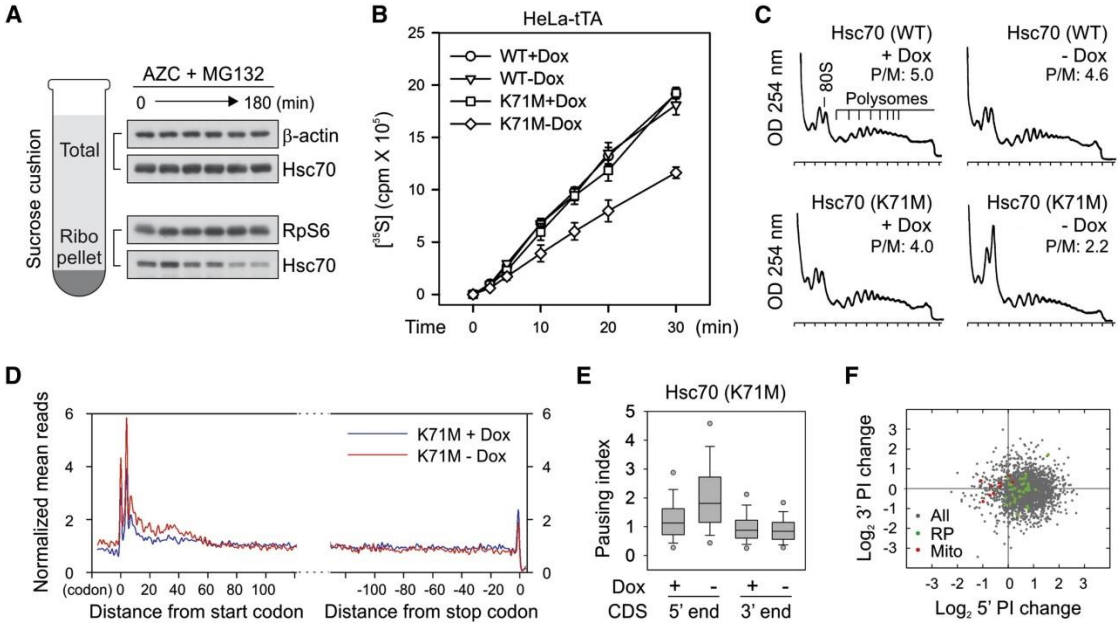


Figure 2-3. Disrupting endogenous Hsc70 recapitulates the effects of proteotoxic stress on early elongation pausing.

(A) Sucrose cushion analysis of ribosome-associated Hsc70 along with AZC and MG132 treatment. Both the total and ribosome pellet were immunoblotted using antibodies as indicated.

(B) Global protein synthesis was analyzed in HeLa-tTA cells infected with adenoviruses expressing Hsc70(WT) and Hsc70(K71M). Transgene expression was induced by 12 h Dox removal. [^{35}S] radioactivity of TCA-insoluble material was measured at given times. Means ± SEM of three experiments are shown.

(C) Polysome profiles were determined from cells as in (B) using sucrose gradient sedimentation.

(D) Meta-gene analysis for early elongation pausing in cells with or without Hsc70(K71M) expression. Normalized RPF reads are averaged across the entire transcriptome, aligned at either their start (left panel) or stop (right panel) codon.

(E) Both the 5' and 3' PI of each transcript in cells with or without Hsc70(K71M) expression are shown in box plots.

(F) Changes of 5'PI and 3'PI after Hsc70(K71M) expression. The $\log_2$ change is computed across the entire transcriptome and presented as a scatter plot with green dots for genes encoding ribosome subunits and red dots for mitochondria- encoded genes. See also Figure 2-S5.

To evaluate whether Hsc70(K71M) expression leads to an early elongation pausing, we performed deep sequencing of RPFs extracted from the polysomes of HeLa-tTA cells with or without transgene induction. Meta-gene analysis revealed a modest excess (~1.6-fold) in density over the initial 50 codons after Hsc70(K71M) expression (Figure 3D and 3E). This is similar in pattern, but of smaller magnitude of early elongation pausing seen in cells treated with both AZC and MG132 (Figure 2B). We repeated the experiment and obtained the similar extent of elongation pausing in the presence of Hsc70(K71M) (Figure S5). Similar to AZC and MG132 treatment, there was an evident separation between genes encoding ribosome subunits and mitochondria proteins in response to the Hsc70(K71M) expression (Figure 3F). The 5'PI changes also showed a good correlation between the two conditions (r = 0.65), although different cell lines were used (Figure S5F). Thus, interfering with endogenous Hsc70 recapitulates the effects of the proteotoxic stress in triggering early elongation pausing.

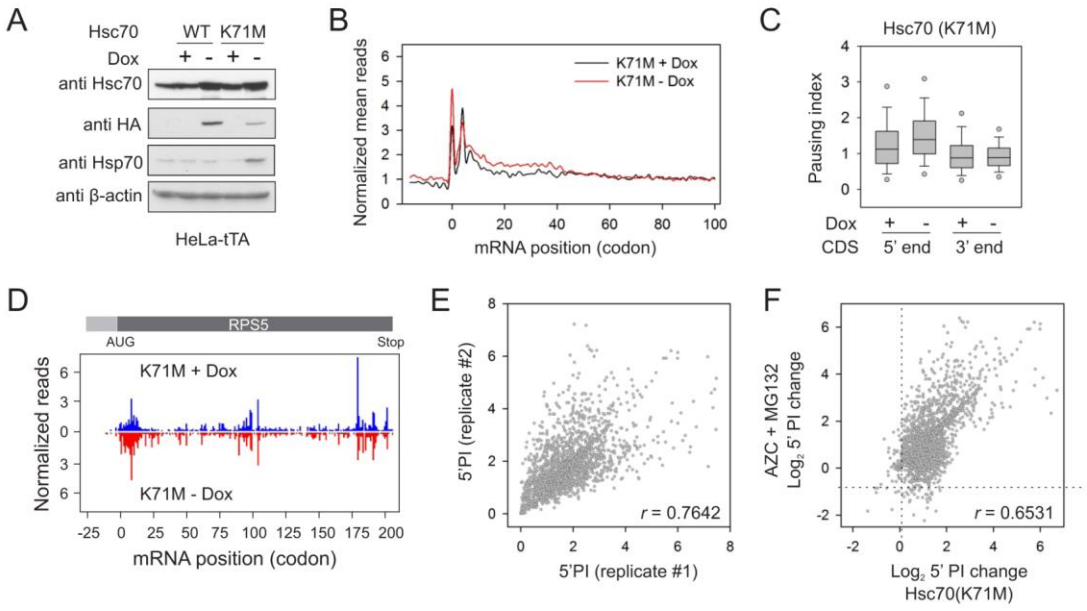


Figure 2-S5. Characterization of early ribosome pausing after inhibiting endogenous Hsc70 by expressing a dominant-negative mutant. Related to **Figure 2-3**.
(A) HeLa-tTA cells were infected with adenoviruses expressing Hsc70(WT) and Hsc70(K71M). The expression levels of both transgene and endogenous genes were determined by immunoblotting after 12 h induction by Dox removal.
(B) Meta-gene analysis of early ribosome pausing in cells with or without Hsc70(K71M) expression. Normalized RPF density profiles over the entire transcriptome, aligned at their start codon, are shown for the first 100-codon region.
(C) Ribosome pausing indices at 5' and 3' end of CDS in cells with or without Hsc70(K71M) expression were shown in box plots.
(D) A typical example of early ribosome pausing under Hsc70(K71M) expression. RPF reads density was shown on the CDS of RPS5 with or without Hsc70(K71M) expression.
(E) Correlation of 5'PI between replicates with Hsc70(K71M) expression.
(F) Scatter plot of 5'PI change (log_2) between HeLa-tTA cells expressing Hsc70(K71M) and HEK293 cells with AZC and MG132 treatment.

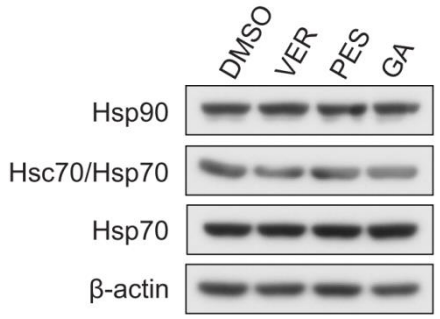


Figure 2-S6. Chaperone levels in cells after treatment with specific chaperone inhibitors. Related to **Figure 2-4**.
HEK293 cells were treated with 100 μM VER, 50 μM PES, or 1 μM GA for 1 h. Whole cell lysates were used for immunoblotting using antibodies indicated.

2.3.5 Direct Hsc/Hsp70 Inhibition Induces Early Elongation Pausing of Ribosomes

The dominant-negative Hsc70(K71M) mutant induced a rather weak elongation pausing when compared to AZC and MG132 treatment. It was likely due to an adaptive stress response under 12 h of Hsc70(K71M) expression, in which the subsequent induction of Hsp70 compromised the early elongation pausing (Figure S5A). In contrast, 1 h of AZC and MG132 treatment did not yet trigger Hsp70 expression due to the time lag. Additionally, the continuous presence of the analog

prevents the production of functional chaperones, if any. To address whether chaperones play a direct role in translation elongation, we applied several specific chaperone inhibitors to HEK293 cells and monitored global protein synthesis (Figure 4A). VER-155008 is a potent inhibitor of the Hsp70 family chaperones (Massey et al., 2010), whereas 2-phenylethyenesulfo-namide (PES) acts as a direct inhibitor of stress-inducible Hsp70 (Leu et al., 2011). We also included a specific Hsp90 inhibitor geldanamycin (GA) to examine the role of different chaperones in ribosome behavior. To minimize the compensatory stress response, we only treated cells with these inhibitors for 1 h. This short treatment allows us to capture direct effects of chaperone inhibition without inducing massive accumulation of misfolded proteins.

Metabolic radiolabeling analysis revealed that both VER and PES potently inhibited [^{35}S] incorporation, whereas the Hsp90 inhibitor GA slightly reduced the level of global protein synthesis (Figure 4B). The extent of translation repression was also reflected in the pattern of polysome profile, in which the inhibitors of Hsp70 family protein, but not Hsp90, disassembled the polysome (Figure 4C). Despite the most severe inhibition of protein synthesis, 1 h treatment of VER resulted in little accumulation of ubiquitin conjugated species in cells (Figure 4D). In addition, the steady state chaperone levels remained unchanged in the presence of these inhibitors (Figure S6), suggesting that the stress response after 1 h of chaperone inhibition was minimal. We next performed deep sequencing of RPFs derived from the cells treated with these chaperone inhibitors. Meta-gene analysis revealed a prominent excess of ribosome density over the first 50-codon region in cells treated with either VER or PES (Figure 4E, 4F). Only minor effect was observed after GA-mediated Hsp90

inhibition. Collectively, these results indicate that direct inhibition of Hsp70 family proteins triggers early elongation pausing of ribosomes.

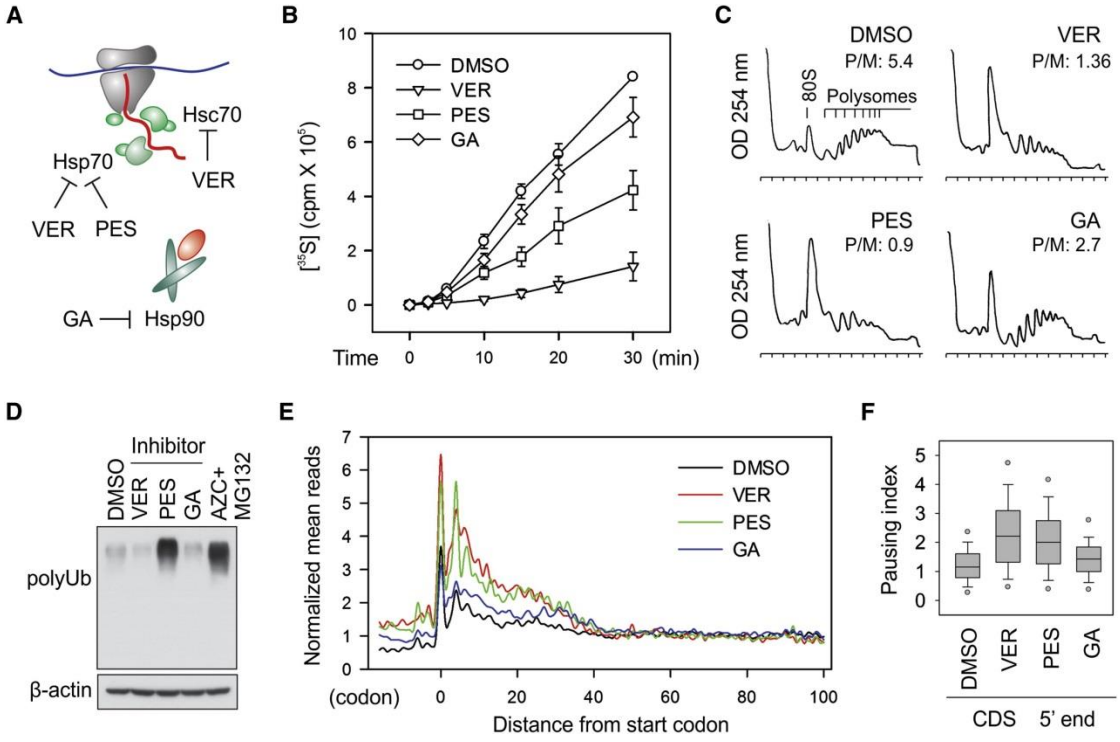


Figure 2-4. Direct Hsc/Hsp70 inhibition induces early elongation pausing of ribosomes
(A) Schematic for chaperone targets of small molecule inhibitors. VER155008 inhibits Hsc70, Hsp70 and Grp78 (not shown); PES selectively inhibits Hsp70; whereas geldanamycin (GA) is a specific inhibitor of Hsp90.
(B) Global protein synthesis was analyzed in HEK293 cells treated with 100 μM VER, 50 μM PES, or 1 μM GA for 1 h. [^{35}S] radioactivity of TCA-insoluble material was measured at given times. Means ± SEM of three experiments are shown.
(C) Polysome profiles were determined from cells treated with chaperone inhibitors as in (B) using sucrose gradient sedimentation.
(D) Immunoblotting of whole cell lysates from cells treated with chaperone inhibitors as in (B).
(E) Meta-gene analysis for early elongation pausing in cells treated with chaperone inhibitors as in (B). Normalized RPF reads are averaged across the entire transcriptome, aligned at their start codon.
(F) The 5' PI of each transcript in cells treated with chaperone inhibitors as in (B) are shown in box plots. See also Figure 2-S6.

2.3.6 Co-translational Interaction of Nascent Chains Influences Elongation Rate

Hsp70 family proteins, including BiP of the endoplasmic reticulum and mtHsp70 of the mitochondrion, are essential for protein translocation across the membrane via unidirectional pulling (Jensen and Johnson, 1999). It is likely that the cytosol Hsc/Hsp70 uses the similar mechanism to pull the emerging polypeptide out of the ribosome exit tunnel. Two models have been proposed to describe how Hsp70 can generate such driving force: "trapping" and "power stroking" (Goloubinoff and De Los Rios, 2007). While both models rely on direct interactions, the latter requires ATP hydrolysis. We hypothesize that the Hsc/Hsp70 "trapping" might be sufficient to exert an entropy pulling force because most nascent chains emerging the exit tunnel are unfolded. To investigate whether co-translational protein interaction would generate the "pulling" force for the ribosome-bound nascent chain, we utilized the hetero-dimerization property of FRB (FKBP12-rapamycin binding domain) and FKBP (FK506 binding protein), whose high affinity binding can be induced by the small molecule rapamycin (Choi et al., 1996; Qian et al., 2009). We constructed a fusion protein FRB-GFP in order to evaluate whether the association of the NH_2-terminal FRB domain with the added FKBP protein during translation would affect the elongation rate of the carboxyl terminal GFP. In an *in vitro* translation system based on rabbit reticulocyte lysate (RRL), we compared the translation efficiency of FRB-GFP after supplementation with the recombinant FKBP protein. Remarkably, upon addition of 1 μM rapamycin to the RRL, the kinetics of FRB-GFP completion showed a significant acceleration (Figure 5A, left panel). We observed the similar effects after

swapping the FRB and FKBP domains (Figure 5A, right panel). Thus, co-translational interaction between nascent chains and specific binding partners promotes the elongation of emerging polypeptides.

In order to extend these findings from RRL to mammalian cells, we utilized a well-characterized rapamycin analog AP21967 (rapalog) and a mutant FRB domain (FRB*) to avoid interfering with the endogenous mTOR function (Klemm et al., 1998). A HEK293 cell line stably expressing FRB*-GFP was transfected with plasmid-borne FKBP. After 60 min of pre-incubation with rapalog, polysome fractions were isolated followed by deep sequencing of RPFs. Notably, the presence of rapalog had little effect on the pattern of RPFs across the entire transcriptome (Figure S7). However, the FRB*-GFP transcript exhibited an altered distribution of RPF reads after rapalog treatment (Figure 5B). When the total reads mapped to the FRB* domain were normalized to be equal, rapalog treatment resulted in a 34% decrease of the average RPF density in the coding region of GFP (Figure 5B, bottom panel). Single codon comparison revealed that the reduction of RPF reads mainly occurred at the ribosome pausing sites of GFP. Since the RPF density on a given codon is proportional to the average ribosome dwell time there, the reduced ribosome density after co-translational interaction between FRB* and FKBP suggests an accelerated elongation for the remaining polypeptide.

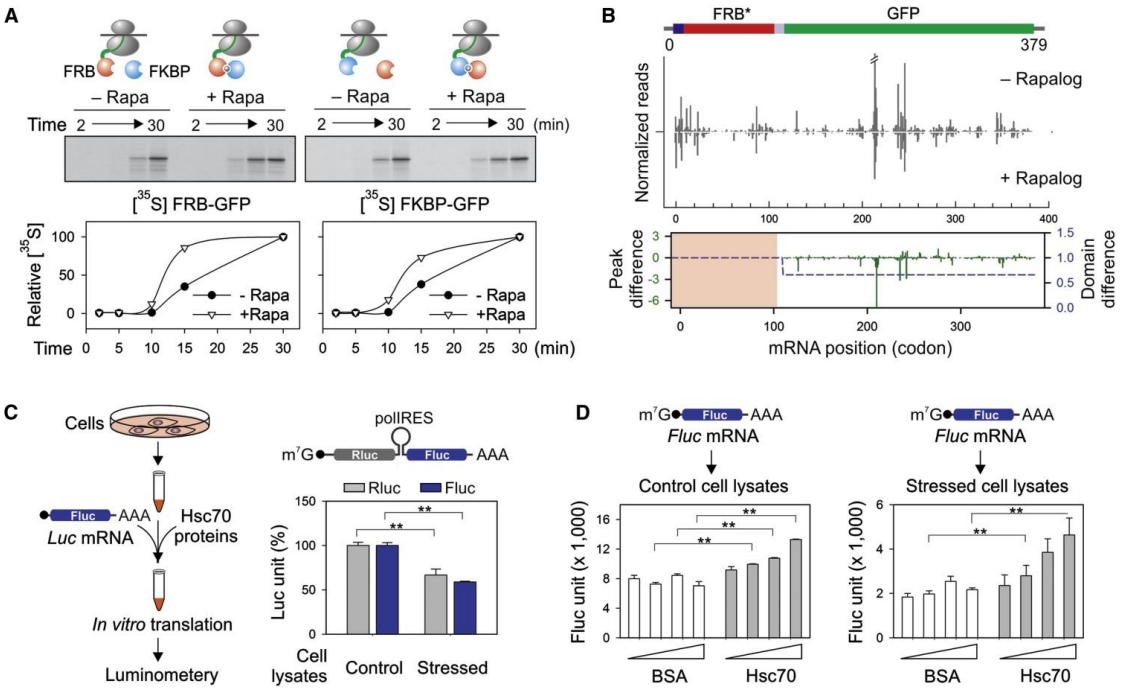


Figure 2-5. Co-translational interaction of nascent chains facilitates the elongation of polypeptides.

(A) Effects of FKBP (blue ball) on the in vitro translation of FRB-GFP (red ball) in the absence or presence of 1 μM rapamycin (left panel). The right panel shows the effects of FRB (red ball) on the *in vitro* translation of FKBP-GFP (blue ball) in the absence or presence of 1 μM rapamycin. Autoradiography of full length GFP fusion protein is quantitated and plotted as a function of time.

(B) HEK293 expressing FRB*-GFP was transfected with the plasmid encoding FKBP. Cells were pre-treated with 1 μM rapalog for 60 min before polysome profiling. The RPF density profiles are shown for the transgene FRB*-GFP with and without rapalog treatment. The RPF reads density is normalized based on the FRB* domain. The average change of RPF density over the entire GFP region (blue dot line) and single codon change (green line) are plotted together (Wilcoxon signed-rank test, p-value = 3×10^{-4}). See also Figure 2-S7.

(C) Schematic of experimental design using recombinant Hsc70 protein to restore translation efficiency using an *in vitro* translation system programmed from cells with or without proteotoxic stress. The right panel shows the relative translation efficiency of a synthesized bicistronic mRNA containing a polio IRES element between Rluc and Fluc. Error bar: SEM. **, $p < 0.001$.

(D) The *in vitro* translation system as (C) was used to translate a synthesized Fluc mRNA in the absence or presence of recombinant Hsc70. Error bar: SEM. **, $p < 0.01$.

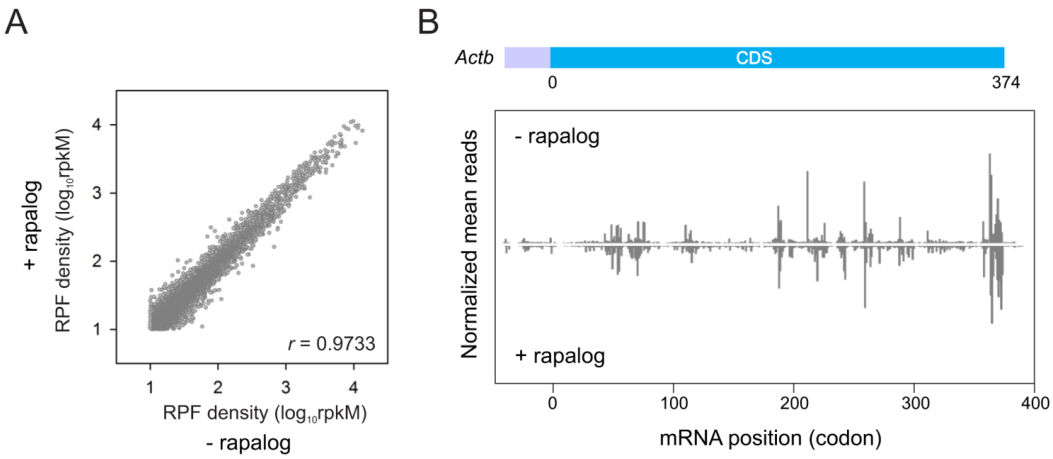


Figure 2-S7. Rapalog treatment has little effects on the distribution of RPF reads. Related to **Figure 2-5**.
(A) RPF densities of cells with or without rapalog were plotted for comparison.
(B) An example of endogenous gene that shows no changes of RPF density pattern in the presence of rapalog. RPF reads density was shown on the CDS of *Actb* with or without rapalog.

2.3.7 Increasing Chaperone Availability Restores Translation Efficiency

The functional connection between chaperone availability and translation elongation underscores the central role of chaperones in protein homeostasis. Based on our results, we expected that chaperone overexpression might prevent the translation inhibition under proteotoxic stress. However, it is inherently difficult to alter the chaperone levels in cells because the chaperone concentration is controlled closely by the heat shock transcription factor 1 (HSF1) (Morimoto, 2008). Overexpression of exogenous chaperone genes inevitably suppresses the endogenous chaperone expression. To circumvent this limitation, we established an *in vitro* translation system programmed from cells with or without proteotoxic stress (Figure 5C, left panel). We first examined the translation efficiency using a synthesized bicistronic mRNA containing the polio internal ribosome entry site (IRES) between

Renilla luciferase (Rluc) and *firefly* luciferase (Fluc). While the synthesis of Fluc is cap-dependent, translation of Rluc is driven by IRES via a cap-independent mechanism (Sun et al., 2011). In lysates derived from stressed cells, the synthesis of both Rluc and Fluc were equally reduced in comparison with the control cell lysates (Figure 5C, right panel). This result further supports the notion that proteotoxic stress does not primarily inhibit cap-dependent initiation.

Next we monitored the translation efficiency of a synthesized Fluc mRNA in cell lysates supplemented with recombinant chaperone molecules. Adding recombinant Hsc70, but not bovine serum albumin (BSA), increased the Fluc activities in a dose-dependent manner (Figure 5D). Co-translational folding of Fluc has been shown to be quite efficient (Kolb et al., 2000), so the translation rate is likely to be the major determinant of luciferase activity in the *in vitro* translation system. Notably, the chaperone-mediated rescue effect was more dramatic in the system derived from the stressed cells than the control. Therefore, increasing chaperone availability restores the translation efficiency.

2.4 Discussion

The journey of a nascent polypeptide starts from the pepetidyl transferase center (PTC) of the ribosome followed by traversing the peptide exit tunnel. Once the nascent chain begins to emerge from the exit tunnel, it faces a drastic environmental change. Surprisingly, most recent ribosome profiling data did not show any specific pausing sites corresponding to this turning point (Guo et al., 2010; Ingolia et al.,

2011). The smooth transition from the inside of the tunnel to the outside ribosome surface is likely due to the presence of ribosome-associated chaperone systems (Kramer et al., 2009). Our study provides strong evidence that the Hsc/Hsp70 family protein plays a crucial role in the passage of nascent chains upon emerging from the ribosome exit tunnel. Reducing chaperone availability by proteotoxic stress or chemical inhibitors unequivocally caused a pileup of ribosomes on the first 50-codon region of transcripts. Notably, we did not observe any specific RPF spikes at specific codon positions, suggesting that the lack of chaperone association slows down rather than stops the elongation. The feature of ribosome stacking at the 5' end of the CDS further indicates that the stress-induced elongation pausing precedes the suppression of translation initiation.

mRNA translation proceeds not at a constant rate but rather in a stop-and-go traffic manner (Fredrick and Ibba, 2010). Variations of elongation speed may result from local stable mRNA structure (Gray and Hentze, 1994), or the presence of rare codons (Elf et al., 2003; Lavner and Kotlar, 2005). Interestingly, nascent chains could also induce translational pausing in a sequence-specific manner (Kramer et al., 2009). Our results uncover an additional layer of elongation regulation mediated by the ribosome-associated chaperone system. The Hsc/Hsp70 family protein, like the ER and mitochondrion counterparts, not only assists co-translational folding, but also accelerates the elongation of nascent polypeptides primarily at the site where the nascent polypeptide emerges from the ribosome exit tunnel. Early studies in *S. cerevisiae* reported a similar function for Ssb, although identifying the elongation pausing sites was beyond the technical ability at that time (Nelson et al., 1992). Since

multiple factors constitute the chaperone network linked to protein synthesis (Albanese et al., 2006), it will be intriguing to determine whether interfering specific chaperone or co-chaperone molecules causes selective elongation pausing on a subset of transcripts.

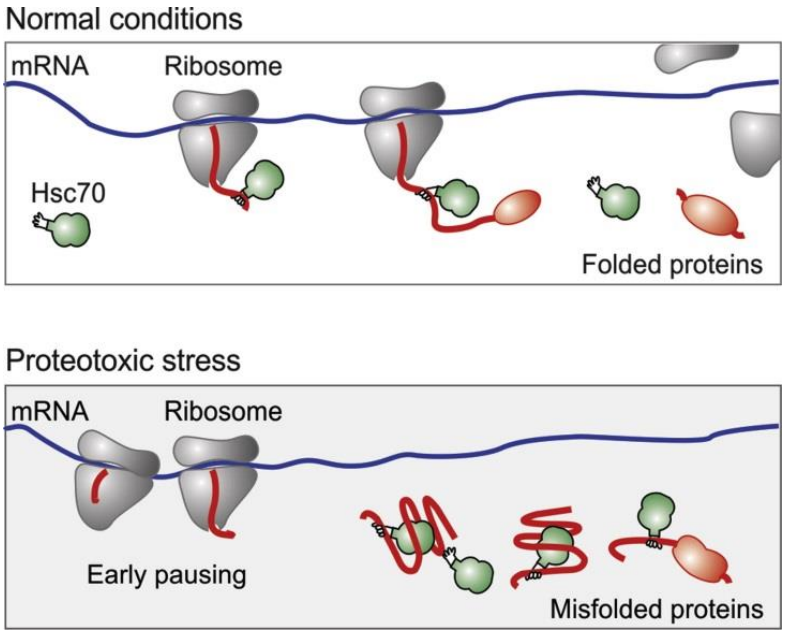


Figure 2-6. A model for co-translational stress response via early ribosome pausing. The cytosolic chaperone molecules, such as Hsc70 (green), not only assist the co-translational folding, but also facilitate the elongation of emerging polypeptides (red). Under the condition of proteotoxic stress, the accumulation of misfolded proteins titrates out molecular chaperones. The lack of co-translational interaction of chaperone molecules leads to early elongation pausing and rapid suppression of global protein synthesis.

Despite the apparent abundance of chaperone molecules in cells, their concentration is titrated closely to the folding requirements within a specific cell type (Morimoto, 2008). Cells exploit chaperone availability as a sensing mechanism to induce stress response. At the level of transcription, reduced chaperone availability triggers the activation of heat shock transcription factor 1 (HSF1) (Morimoto, 1998). As a result, more chaperone molecules will be produced to restore the protein

homeostasis. The functional connection between chaperone availability and translation elongation offers a novel mode of regulation in response to stress conditions (Figure 6). Intracellular accumulation of misfolded proteins, a common feature of a variety of stress conditions, sequesters molecular chaperones and the lack of chaperone association with the ribosome delays nascent chains from emerging. Our data suggest that the ribosome fine-tunes the elongation rate based on the chaperone availability to match protein production with the intracellular folding capacity. This level of control allows a rapid change in the complement of proteins prior to transcriptional regulation. The early elongation pausing under proteotoxic stress thus represents the very first line of protective response for cells to maintain intracellular protein homeostasis.

2.5 Materials & Methods

Cells and Reagents

Human HEK293 and Tet-off HeLa-tTA cells (Clontech) were maintained in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum (FBS). $_{L}$-azetidine-2-carboxylic acid (AZC), Z-Leu-Leu-Leu-al (MG132), sodium arsenite ($NaAsO_2$), cycloheximide (CHX), rapamycin, VER-155008, 2-phenylethyenesulfonamide (PES), geldanamycin (GA), aurintricarboxylic acid (ATA) and bovine serum albumin (BSA) were purchased from Sigma. Anti-RpS6, anti-Phos-RpS6, anti-eIF2α, anti-Phos-eIF2α, anti-S6K1, anti-Phos-S6K1, anti-eIF4E and anti-TIAR antibodies were purchased from Cell signaling. Anti-Hsc70, anti-Hsp70, anti-

Hsp90, anti-polyubiquitin and bovine Hsc70 recombinant protein were bought from Enzo Life Sciences. Anti-HA (Roche), anti-β-actin (Sigma) primary antibodies and secondary antibodies (Sigma) were also acquired. Rapalog AP21967 was provided by Ariad Inc.

Transfection and Adenoviruses Infection

Plasmid transfection was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Recombinant adenoviruses (AdV) expressing a wild-type Hsc70(WT) and a dominant-negative mutant Hsc70 (K71M) were generously provided by Dr. Jonathan Yewdell (NIAID, NIH). AdV infection was carried out in Tet-off HeLa-tTA cells in the presence of 1 μg/ml doxycycline (Dox) to repress transgene expression. Transgene expression was typically induced for 12 h by removing Dox from the media.

[^{35}S] Pulse Chase Assay

Cells were trypsinized and suspended in the regular DMEM (10%FBS) medium and incubated at 37 ℃ for 30 min to achieve equilibrium. After the drug treatment, cells were pelleted and re-suspended in pulsing medium containing 10 μCi [^{35}S] $_{L}$-methionine and [^{35}S] $_{L}$-cystine mix (Perkin Elmer) supplemented with 10% FBS, $_{L}$-Glutamine and drugs if needed. For the pulsing only experiment, an aliquot of cells was withdrawn at each time point and mixed with stop medium (ice-cold DMEM containing 1mg/ml cold $_{L}$-methionine, 1mg/ml cold $_{L}$-cystine and 100 μg/ml cycloheximide) to terminate the protein synthesis and [^{35}S] incorporation. For the pulse-chase experiment, pulsed cells were washed once and re-suspended in chasing medium (DMEM supplemented with 10%FBS, 1mg/ml $_{L}$-methionine and 1mg/ml $_{L}$-

cystine). After incubation at 37 ℃ for various times, aliquots of cells were pelleted and lysed with ice-cold polysome lysis buffer (pH 7.4, 10 mM HEPES, 100 mM KCl, 5 mM $MgCl_2$, 100 μg/ml cycloheximide and 2% Triton X-100). After centrifugation at14,000 rpm for 10 min at 4 ℃, proteins in the supernatant were precipitated with 10% Trichloroacetic acid (Sigma). The mixture was heated for 10 min at 90 ℃ and then chilled on ice for 10 min. The precipitates were collected on GF/C filter membrane (Watman) and the [^{35}S] incorporation was measured by scintillation counting (Beckman).

Measurement of Ribosomal Half-Transit Time

Cells were pulsed with [^{35}S] $_L$-methionine/$_L$-cystine for various times and lysed with polysome lysis buffer. 100 μl lysates were mixed with 350 μl polysome buffer (pH 7.4, 10 mM HEPES, 100 mM KCl, 5 mM $MgCl_2$) and 450 μl 0.14M sucrose in polysome buffer. 400 μl mixture was saved for measurement of total [^{35}S] incorporation including both nascent and completed polypeptides. Ribosomes were pelleted from the remaining 500 μl mixture by centrifugation at 60,000 rpm for 15 min at 4 ℃ using a Beckman TLA-100.4 rotor. 400 μl of supernatant was taken to measure the [^{35}S] incorporation into the completed polypeptide only by scintillation counting. After linear regression plotting as a function of time, the ribosome half-transit time ($t_{1/2}$) was estimated from the displacement in time between the two lines corresponding to the total and released protein.

Immunoprecipitation and Immunoblotting

For immunoprecipitation of ribosome-associated nascent chains, cells were pre-treated with 100 μg/ml cycloheximide at 37 ℃ for 3mins to stabilize ribosome-

nascent chain complex. Cells were then scraped extensively in polysome lysis buffer supplemented with EDTA-free cocktail protease inhibitor (Roche). After clearance by centrifugation for 10 min at14,000 rpm 4 ℃, the supernatant was collected and incubated with anti-Flag M2 affinity gel (Sigma) at 4 ℃ for 1h. The beads were extensively washed for three times with polysome lysis buffer and the associated proteins were eluted by heating for 10 min in the sample buffer (50 mM Tris-HCl, pH 6.8, 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, 10% glycerol). For immunoblotting, protein samples were resolved on SDS-PAGE and then transferred to Immobilon-P membranes (Millipore). After blocking for 1 hour in TBS containing 5% blotting milk, membranes were incubated with primary antibodies at 4 ℃ overnight. After incubation with horseradish peroxidase–coupled secondary antibodies, immunoblots were developed using enhanced chemiluminescence (GE Healthcare).

Immunofluorescence

For immunostaining of stress granule (SG), HeLa cells were plated on glass coverslips. After the drug treatments, cells were immediately fixed by 4% paraformaldehyde, followed by permeabilization by 0.2% Triton X-100 and blocking in 2% BSA in PBS. Fixed cells were then incubated with primary antibody at 4 ℃ overnight. After washing for three times, cells were incubated with Alexa Fluor–labeled secondary antibodies (Invitrogen) at room temperature for 1 hour. Cell nuclei were counter-stained with Hoechst in PBS for 5 min. After washing for three times, coverslips were mounted onto slides and viewed using a confocal microscope (LSM710, Zeiss).

In vitro Translation

Fluc mRNA was synthesized through *in vitro* transcription using mMESSAGE mMACHINE T7 ULTRA Kit (Ambion) according to manufacturer's instructions. Programmed *in vitro* translation was performed following published protocol (Rakotondrafara A. M. & Hentze M.W. 2011). Briefly, HEK293 cells were pretreated with 10 mM AZC and 20 μM MG132 for 1h at 37 ℃, followed by a wash with ice-cold PBS. Cells were lysed with equal volume of ice-cold lysis buffer (10 mM HEPES, pH7.6, 10 mM potassium acetate, 0.5 mM magnesium acetate, 5 mM DTT, EDTA-free protease inhibitor cocktail) by hypotonic swelling and homogenization through 27G syringe needle. After centrifugation at 10,000 g for 1min, extract supernatant was collected and quantified by Bradford assay. Control and stressed lysates used were adjusted to be equal based on protein concentration. A typical reaction contains cell extract (6-8 mg/ml final protein concentration), 1 × translation buffer (16 mM HEPES, 7.6 pH, 8 mM creatine phosphate, 20 mg/ml creatine kinase, 0.1 mM spermidine, 10 μM each amino acid), 80mM potassium acetate, 1 mM magnesium acetate, 20U RNase inhibitor (Ambion), 0.8 mM ATP(NEB) and 40 μg/ml mRNA template. *In vitro* translation was incubated at 37 ℃ for 2 hour and luciferase substrate (Promega) was added to measure the activity by luminometery. For chaperone rescue experiment, recombinant Hsc70 protein (Enzo Life Sciences) was dialyzed against PBS and the protein concentration was quantified by Bradford assay. Increasing dose of BSA or Hsc70 protein (from 0.4 μM to 3.2 μM with 2-fold dilution) was supplemented in the reaction before the incubation. As to elongation chase assay, reaction above was constituted and incubated at 30 ℃ for 5 min to allow ribosome

loading. Immediately after that, the reaction was paused on ice and 0.5 mM ATA was added to block additional initiation. Subsequently, the reaction was resumed at 30 ℃ and aliquots of samples were withdrawn for luciferase activity measurement at indicated times. For *in vitro* translation of FRB or FKBP assay, TNT Quick Coupled Translation /Transcription system (Promega) based on rabbit reticulocyte lysate (RRL) was used. In brief, pcDNA3 plasmid encoding FRB-GFP or FKBP-GFP was mixed with RRL supplemented with $[^{35}S]_{L}$-methionine and recombinant proteins as indicated. *In vitro* translation was performed in the presence or absence of 1 μM Rapamycin. The products at different time points were collected and resolved on SDS-PAGE. The gel was dried and viewed by phosphor imaging screen (HE healthcare) and the band intensity was quantified using ImageQuant 5.2.

Polysome Profiling

Polysome buffer was used to prepare sucrose solutions. Sucrose density gradients (15%- 45% w/v) were freshly made in SW41 ultracentrifuge tubes (Backman) using a Gradient Master (BioComp Instruments) according to manufacturer's instructions. Cells were pre-treated with 100 μg/ml cycloheximide for 3 min at 37 ℃ to stabilize ribosomes on mRNAs followed by washing using ice-cold PBS containing 100 μg/ml cycloheximide. Cells were then lysed by scraping extensively in polysome lysis buffer. Cell debris were removed by centrifugation at14,000 rpm for 10 min at 4 ℃. 600 μl of supernatant was loaded onto sucrose gradients followed by centrifugation for 100 min at 38,000 rpm 4 ℃ in a SW41 rotor. Separated samples were fractionated at 0.750 ml / min through a fractionation system (Isco) that continually monitored OD_{254} values. Fractions were collected with 0.5 min

interval. The polysome to monosome ratios were calculated by quantifying the area under corresponding peaks in the polysome profiling (NIS-Elements, Nikon). For puromycin sensitivity experiment, 100 μM puromycin was added during the pre-treatment instead of cycloheximide.

Extraction of Ribosome Protected mRNA Fragments

Polysome profiling fractions were mixed and a 200 μl aliquot was digested with 200U *E. coli* RNase I (Ambion) at 4 ℃ for 1h. Total RNA was then extracted by Trizol reagent (Invitrogen). Purified RNA samples were mixed with 1 μl of 10 nM synthetic 28 nt random RNA (5'-AUGUACACGGAGUCGACCCGCAACGCGA-3') as an internal control. Subsequently, RNA molecules were dephosphorylated by 20U T4 polynucleotide kinase (NEB) in the presence of 10 U SUPERase (Ambion) at 37 ℃ for 1 hour. The enzyme was heat-inactivated for 20 min at 65 ℃. The products were then separated on a Novex denaturing 15% polyacrylamide TBE-urea gel (Invitrogen). SYBR Gold (Invitrogen) was utilized to stain the gel and visualize the RNA fragments. Gel bands corresponding to 28-30 nt RNA molecules were excised and physically disrupted by centrifugation through the holes of the tube. Resulting gel debris was soaked overnight in the RNA gel elution buffer (300 mM NaOAc pH 5.5, 1 mM EDTA, 0.1 U/mL SUPERase_In) to recover RNA fragments. The gel debris was filtered out with a Spin-X column (Corning) and RNA was finally purified using ethanol precipitation.

cDNA Library Construction and Deep Sequencing

Poly-A tails were added to the purified RNA fragments by *E. coli* poly-(A)

polymerase (NEB) with 1 mM ATP in the presence of 0.75 U/μL SUPERase_In at 37℃ for 45 min. The tailed RNA molecules were reverse transcribed to generate the first strand cDNA using SuperScript III (Invitrogen) and following oligos containing barcodes:

SCT01:5'-
pCTGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGATT
TTTTTTTTTTTTTTTTTTVN-3';
MCA02: 5'-
pCAGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGAT
TTTTTTTTTTTTTTTTTTVN-3';
LGT03:5'-
pGTGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGATT
TTTTTTTTTTTTTTTTTTVN-3'.
HTC04: 5'-
pTCGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGATT
TTTTTTTTTTTTTTTTTTVN-3'
YAG05:5'-
pAGGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGATT
TTTTTTTTTTTTTTTTTTVN-3';

Reverse transcription products were resolved on a 10% polyacrylamide TBE-urea gel as described above. The expected 92 nt band of first strand cDNA was excised and recovered as above using DNA gel elution buffer (300 mM NaCl, 1 mM

EDTA). Purified first strand cDNA was then circularized by 100U CircLigase II (Epicentre) following manufacturer's instructions. The resulting circular single strand DNA was purified using ethanol precipitation and re-linearized by 7.5 U APE 1 in 1 X buffer 4 (NEB) at 37 ℃ for 1 h. The products were resolved on a Novex 10% polyacrylamide TBE-urea gel (Invitrogen) as described above. The expected 92 nt band was then excised and recovered. Finally, single-stranded template was amplified by PCR using the Phusion High-Fidelity enzyme (NEB) according to the manufacturer's instructions. The primers qNTI200 (5'-CAAGCAGAAGACGGCATA-3') and qNTI201 (5'-AATGATACGGCGACCACCG ACAGGTTCAGAGTTCTACAGTCCGACG-3') were used to create DNA library suitable for sequencing. The PCR reaction contains 1 × HF buffer, 0.2mM dNTP, 0.5 μM primers, 0.5U Phusion polymerase. PCR was carried out with an initial 30 s denaturation at 98 ℃, followed by 12 cycles of 10 s denaturation at 98 ℃, 20 s annealing at 60 ℃, and 10 s extension at 72 ℃. PCR products were separated on a non-denaturing 8% polyacrylamide TBE gel as described above. Expected 120 bp band was excised and recovered as described above. After quantification by Agilent BioAnalyzer DNA 1000 assay, equal amount of barcoded samples were pooled into one sample. 3 ~ 5 pmol mixed DNA samples were typically used for cluster generation followed by sequencing using sequencing primer 5'-CGACAGGTTCAGAGTTC TACAGTCCGACGATC-3' (Illumina Genome Analyzer 2 or HiSEQ, Cornell University Life Sciences Core Laboratories Center).

Deep Sequencing Data Analysis

The deep sequencing data of ribosome footprints was processed and analyzed using a collection of custom Perl scripts. The barcoded multiplex sequencing output files were separated into individual sample datasets according to the first two-nucleotide barcodes. Second, the 3'polyA tails allowing 1 mismatch were identified and removed. The high quality reads of length ranging from 25 to 35 nt were then retained while other reads were excluded from the downstream analysis.

21,822 sequences of the longest transcript isoform for each human gene were downloaded from Ensembl database (www.ensembl.org) to construct a human transcriptome reference. In addition, the plasmid sequence of FRB*-GFP was also included as the reference. The trimmed reads were aligned to the reference by SOAP 2.0 allowing up to 2 mismatches and all multiple equal best hits were retained. The 5' end positions of aligned reads were mapped into the coding frame and the number of reads was counted at each codon ranging from -20 codon 5'UTR to the stop codon for the downstream analysis.

The meta-gene analysis of ribosome footprint distribution across the transcriptome was carried out by calculating the normalized mean reads (NMR) density at each codon position. Only transcripts with coding sequence longer than 150 codons and with more than 50 reads were included in the analysis. r_{ij} stands for the number of reads at codon position *i* of transcript *j* and is normalized to the average density on the entire transcript with length of l_j codons. T_i means all the transcripts available at position *i*. Finally, the normalized mean reads densities were plotted against the codon position to reveal the pattern of reads distribution.

$$\mathrm{NMR}(i) = \frac{\sum_{j}^{T_i} \frac{l_j \times r_{ij}}{\sum_{i=1}^{i=l_j} r_{ij}}}{\sum_{j}^{T_i} 1}$$

For transcripts with coding sequence longer than 150 codons and at least 50 total reads, 25 reads in the first 50 codon window, we defined the ribosome pausing index (5'PI) to evaluate the pausing level at the 5' end CDS for individual transcript. It equals to the ratio between the read density in the first 50 codon window and the immediate following 100 codon region. Besides, a similar pausing index at the 3' end CDS (3'PI) was also calculated.

$$5'\ \mathrm{PI}(j) = \frac{2 \times \sum_{i=1}^{i=50} r_{ij}}{\sum_{i=51}^{i=150} r_{ij}} \qquad 3'\ \mathrm{PI}(j) = \frac{2 \times \sum_{i=-1}^{i=-50} r_{ij}}{\sum_{i=-51}^{i=-150} r_{ij}}$$

To identify significantly enriched functions in groups of genes, the online web tool DAVID (www.david.abcc.ncifcrf.gov) gene functional classification system was employed. Ensembl gene ID list of interested genes was uploaded to the website and gene ontology analysis was performed with the *Biological_Processes_FAT* GO terms. Top non-redundant significantly enriched terms ranked by P-value were selected and further processed in Excel.

Accession Numbers

Sequencing data were deposited in the SRA database with the accession number SRA061778.

2.6 Acknowledgements

We'd like to thank Qian lab members for helpful discussion and Drs. Patsy Brannon and William Brown for critical reading of the manuscript. We also thank Dr. Yewdell (NIH) for providing adenoviruses expressing Hsc70(WT) and Hsc70(K71M), and Cornell University Life Sciences Core Laboratory Center for performing deep sequencing. This work was supported by grants to S.-B.Q. from National Institutes of Health (1 DP2 OD006449-01), Ellison Medical Foundation (AG-NS-0605-09), and DOD Exploration-Hypothesis Development Award (TS10078).

REFERENCES

Albanese, V., Yam, A.Y., Baughman, J., Parnot, C., and Frydman, J. (2006). Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells. Cell *124*, 75-88.

Buchan, J.R., and Parker, R. (2009). Eukaryotic stress granules: the ins and outs of translation. Mol Cell *36*, 932-941.

Bukau, B., Weissman, J., and Horwich, A. (2006). Molecular chaperones and protein quality control. Cell *125*, 443-451.

Choi, J., Chen, J., Schreiber, S.L., and Clardy, J. (1996). Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. Science *273*, 239-242.

del Alamo, M., Hogan, D.J., Pechmann, S., Albanese, V., Brown, P.O., and Frydman, J. (2011). Defining the specificity of cotranslationally acting chaperones by systematic analysis of mRNAs associated with ribosome-nascent chain complexes. PLoS Biol *9*, e1001100.

Elf, J., Nilsson, D., Tenson, T., and Ehrenberg, M. (2003). Selective charging of tRNA isoacceptors explains patterns of codon usage. Science *300*, 1718-1722.

Fredrick, K., and Ibba, M. (2010). How the sequence of a gene can tune its translation. Cell *141*, 227-229.

Frydman, J. (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. Annu Rev Biochem *70*, 603-647.

Goldberg, A.L., and Dice, J.F. (1974). Intracellular protein degradation in mammalian and bacterial cells. Annu Rev Biochem *43*, 835-869.

Goloubinoff, P., and De Los Rios, P. (2007). The mechanism of Hsp70 chaperones: (entropic) pulling the models together. Trends Biochem Sci *32*, 372-380.

Gray, N.K., and Hentze, M.W. (1994). Regulation of protein synthesis by mRNA structure. Mol Biol Rep *19*, 195-200.

Gray, N.K., and Wickens, M. (1998). Control of translation initiation in animals. Annu Rev Cell Dev Biol *14*, 399-458.

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.

Hansen, W.J., Cowan, N.J., and Welch, W.J. (1999). Prefoldin-nascent chain complexes in the folding of cytoskeletal proteins. J Cell Biol *145*, 265-277.

Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. Nature *475*, 324-332.

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. Nat Rev Mol Cell Biol *6*, 318-327.

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.

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.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol *11*, 113-127.

Jaiswal, H., Conz, C., Otto, H., Wolfle, T., Fitzke, E., Mayer, M.P., and Rospert, S. (2011). The chaperone network connected to human ribosome-associated complex. Mol Cell Biol *31*, 1160-1173.

Jensen, R.E., and Johnson, A.E. (1999). Protein translocation: is Hsp70 pulling my chain? Curr Biol *9*, R779-782.

Kampinga, H.H., and Craig, E.A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat Rev Mol Cell Biol *11*, 579-592.

Kedersha, N., Cho, M.R., Li, W., Yacono, P.W., Chen, S., Gilks, N., Golan, D.E., and Anderson, P. (2000). Dynamic shuttling of TIA-1 accompanies the recruitment of mRNA to mammalian stress granules. J Cell Biol *151*, 1257-1268.

Klemm, J.D., Schreiber, S.L., and Crabtree, G.R. (1998). Dimerization as a regulatory mechanism in signal transduction. Annu Rev Immunol *16*, 569-592.

Kolb, V.A., Makeyev, E.V., and Spirin, A.S. (2000). Co-translational folding of an eukaryotic multidomain protein in a prokaryotic translation system. J Biol Chem *275*, 16597-16601.

Kramer, G., Boehringer, D., Ban, N., and Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat Struct Mol Biol *16*, 589-597.

Lavner, Y., and Kotlar, D. (2005). Codon bias as a factor in regulating expression via translation rate in the human genome. Gene *345*, 127-138.

Leu, J.I., Pimkina, J., Pandey, P., Murphy, M.E., and George, D.L. (2011). HSP70 inhibition by the small-molecule 2-phenylethynesulfonamide impairs protein clearance pathways in tumor cells. Mol Cancer Res *9*, 936-947.

Ma, X.M., and Blenis, J. (2009). Molecular mechanisms of mTOR-mediated translational control. Nat Rev Mol Cell Biol *10*, 307-318.

Massey, A.J., Williamson, D.S., Browne, H., Murray, J.B., Dokurno, P., Shaw, T., Macias, A.T., Daniels, Z., Geoffroy, S., Dopson, M., *et al.* (2010). A novel, small molecule inhibitor of Hsc70/Hsp70 potentiates Hsp90 inhibitor induced apoptosis in HCT116 colon carcinoma cells. Cancer Chemother Pharmacol *66*, 535-545.

McClellan, A.J., Tam, S., Kaganovich, D., and Frydman, J. (2005). Protein quality control: chaperones culling corrupt conformations. Nat Cell Biol *7*, 736-741.

Morimoto, R.I. (1998). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes Dev *12*, 3788-3796.

Morimoto, R.I. (2008). Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. Genes Dev *22*, 1427-1438.

Nelson, R.J., Ziegelhoffer, T., Nicolet, C., Werner-Washburne, M., and Craig, E.A. (1992). The translation machinery and 70 kd heat shock protein cooperate in protein synthesis. Cell *71*, 97-105.

Newmyer, S.L., and Schmid, S.L. (2001). Dominant-interfering Hsc70 mutants disrupt multiple stages of the clathrin-coated vesicle cycle in vivo. J Cell Biol *152*, 607-620.

Nielsen, P.J., and McConkey, E.H. (1980). Evidence for control of protein synthesis in HeLa cells via the elongation rate. J Cell Physiol *104*, 269-281.

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.

Qian, S.B., Waldron, L., Choudhary, N., Klevit, R.E., Chazin, W.J., and Patterson, C. (2009). Engineering a ubiquitin ligase reveals conformational flexibility required for ubiquitin transfer. J Biol Chem *284*, 26797-26802.

Qian, S.B., Zhang, X., Sun, J., Bennink, J.R., Yewdell, J.W., and Patterson, C. (2010). mTORC1 links protein quality and quantity control by sensing chaperone availability. J Biol Chem *285*, 27385-27395.

Rakotondrafara, A.M., and Hentze, M.W. (2011). An efficient factor-depleted mammalian in vitro translation system. Nat Protoc *6*, 563-571.

Ron, D., and Walter, P. (2007). Signal integration in the endoplasmic reticulum unfolded protein response. Nat Rev Mol Cell Biol *8*, 519-529.

Saini, P., Eyler, D.E., Green, R., and Dever, T.E. (2009). Hypusine-containing protein eIF5A promotes translation elongation. Nature *459*, 118-121.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell *136*, 731-745.

Spriggs, K.A., Bushell, M., and Willis, A.E. (2010). Translational regulation of gene expression during conditions of cell stress. Mol Cell *40*, 228-237.

Sun, J., Conn, C.S., Han, Y., Yeung, V., and Qian, S.B. (2011). PI3K-mTORC1 attenuates stress response by inhibiting cap-independent Hsp70 translation. J Biol Chem *286*, 6791-6800.

Trotter, E.W., Kao, C.M., Berenfeld, L., Botstein, D., Petsko, G.A., and Gray, J.V. (2002). Misfolded proteins are competent to mediate a subset of the responses to heat shock in Saccharomyces cerevisiae. J Biol Chem *277*, 44817-44825.

Zhang, X., and Qian, S.B. (2011). Chaperone-mediated hierarchical control in targeting misfolded proteins to aggresomes. Mol Biol Cell *22*, 3277-3288.

CHAPTER 3

Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution

This collaborative work was submitted May 2012 and published on September 11, 2012 as Lee S*, Liu B*, Lee S*, Huang SX, Shen B, and Qian SB. **Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution**. *Proc Natl Acad Sci USA*. 2012; 109(37):E2424-32. Minor modifications have been made for reprint. *Authors contributed equally to this work.

3.1 Abstract

Understanding translational control in gene expression relies on precise and comprehensive determination of translation initiation sites (TIS) across the entire transcriptome. The recently developed ribosome profiling technique enables global translation analysis, providing a wealth of information about both the position and density of ribosomes on mRNAs. Here we present an approach (global translation initiation sequencing, GTI-seq), by applying in parallel ribosome E-site translation inhibitors lactimidomycin (LTM) and cycloheximide (CHX), to achieve simultaneous detection of both initiation and elongation events on a genome-wide scale. With single nucleotide resolution, we show an unprecedented view of alternative translation

initiation in mammalian cells. Systemic analysis of TIS positions supports the ribosome linear scanning mechanism in TIS selection. The alternative TIS positions and the associated open reading frames identified by GTI-seq are conserved between human and mouse cells, implying physiological significance of alternative translation. Our study establishes a practical platform for uncovering the hidden coding potential of transcriptome and offers a greater understanding of the complexity of translation initiation.

3.2 Introduction

Protein synthesis is the final step in the flow of genetic information and lies at the heart of cellular metabolism. Translation is principally regulated at the initiation stage and there has been significant progress over the last decade in dissecting the role of initiation factors (eIFs) in the assembly of elongation-competent 80S ribosomes (1-3). However, mechanisms underlying start codon recognition are not fully understood. Proper selection of the translation initiation site (TIS) on mRNAs is crucial for the production of desired protein products. A fundamental and long-sought goal in understanding translational regulation is the precise determination of TIS codons across the entire transcriptome.

In eukaryotes, ribosomal scanning is a well-accepted model for start codon selection (4). During cap-dependent translation initiation, the small ribosome subunit (40S) is recruited to the 5' end of mRNA (the m^7G cap) in the form of a 43S pre-initiation complex (PIC). The PIC is thought to scan along the message in search for

the start codon. It is commonly assumed that the first AUG codon that the scanning PIC encounters serves as the start site for translation. However, many factors influence the start codon selection. For instance, the initiator AUG triplet is usually in an optimal context with a purine at position -3 and a guanine at position +4 (5). The presence of mRNA secondary structure at or near the TIS position also influences the recognition efficiency (6). In addition to these *cis* sequence elements, the stringency of TIS selection is also subject to regulation by *trans* acting factors such as eIF1 and eIF1A (7, 8). Inefficient recognition of an initiator codon results in a portion of 43S PIC continuing to scan and initiating at a downstream site, in a process known as leaky scanning (4). However, little is known about the frequency of leaky scanning events at the transcriptome level.

Many recent studies have uncovered a surprising variety of potential translation start sites upstream of the annotated coding sequence (CDS) (9, 10). It has been estimated that about 50% of mammalian transcripts contain at least one upstream open reading frame (uORF) (11, 12). Intriguingly, many non-AUG triplets have been reported to act as alternative start codons for initiating uORF translation (13). Since there is no reliable way to predict non-AUG codons as potential initiators from *in silico* sequence analysis, there is an urgent need to develop experimental approaches for genome-wide TIS identification.

Ribosome profiling, based on deep sequencing of ribosome-protected mRNA fragments (RPF), has proven to be powerful in defining ribosome positions on the entire transcriptome (14, 15). However, the standard ribosome profiling is not suitable for TIS identification. Elevated ribosome density near the beginning of CDS does not

allow for unambiguous identification of alternative TIS positions, in particular the TIS positions associated with overlapping ORFs. To overcome this problem, a recent study used an initiation-specific translation inhibitor harringtonine to deplete elongating ribosomes from mRNAs (16). This approach uncovered an unexpected abundance of alternative TIS codons, in particular non-AUG codons in the 5'UTR. However, since the inhibitory mechanism of harringtonine on the initiating ribosome is unclear, it remains to be confirmed whether the harringtonine-marked TIS codons truly represent physiological translation initiation sites.

Here, we develop global translation initiation sequencing (GTI-seq) by utilizing two related but distinct translation inhibitors to effectively differentiate ribosome initiation from elongation. GTI-seq has the potential to reveal a comprehensive and unambiguous set of TIS codons at near single nucleotide resolution. The resulting TIS maps provide a remarkable display of alternative translation initiators that vividly delineates the variation in start codon selection. This allows for a more complete assessment of the underlying principles that specify start codon usage *in vivo*.

3.3 Results

3.3.1 Experimental Design

Cycloheximide (CHX) has been widely used in ribosome profiling of eukaryotic cells because of its potency in stabilizing ribosomes on mRNAs. Both the biochemical (17) and structural studies (18) revealed that CHX binds to the exit (E)-

site of the large ribosomal subunit, close to the position where the 3' hydroxyl group of the deacylated tRNA normally binds. CHX thus prevents the release of deacylated tRNA from the E-site and blocks subsequent ribosomal translocation (Fig. 1A, left panel). A new family of CHX-like natural products isolated from *Streptomyces* were recently characterized, including lactimidomycin (LTM) (19, 20). Acting as a potent protein synthesis inhibitor, LTM uses a similar but not identical mechanism as CHX (17). With its 12-membered macrocycle, LTM is significantly larger in size than CHX (Fig. 1A). As a result, LTM cannot bind to the E-site when a deacylated tRNA is present. Only during the initiation step, in which the initiator tRNA directly enters into the peptidyl (P)-site (21), is the empty E-site accessible to LTM. Thus, LTM preferentially acts on the initiating ribosome but not the elongating ribosome. We reasoned that ribosome profiling using LTM side-by-side in comparison with CHX should allow for a complete segregation of the ribosome stalled at the start codon from the one in active elongation (Fig. 1B).

We designed an integrated GTI-seq approach and performed the ribosome profiling in HEK293 cells pretreated with either LTM or CHX. While CHX slightly stabilized the polysomes when compared to the no-drug treatment (DMSO), 30 min of LTM treatment led to a large increase in monosome accompanied by a depletion of polysomes (Fig. S1). This is in agreement with the notion that LTM halts translation initiation while allowing elongating ribosomes to run off (17). After RNase I digestion of the ribosome fractions, the purified RPFs were subjected to deep sequencing. As expected, CHX treatment resulted in an excess of RPFs at the beginning of ORFs in addition to the body of CDS (Fig. 1C). Remarkably, LTM

treatment led to a pronounced single peak located at the -12 nucleotide (nt) position relative to the annotated start codon. This position corresponds to the ribosome P-site at the AUG codon when an offset of 12 nt is considered (14, 15). LTM treatment also eliminated the excess of ribosomes seen at the stop codon in untreated cells or in the presence of CHX. Therefore, LTM efficiently stalls the 80S ribosome at the start codons.

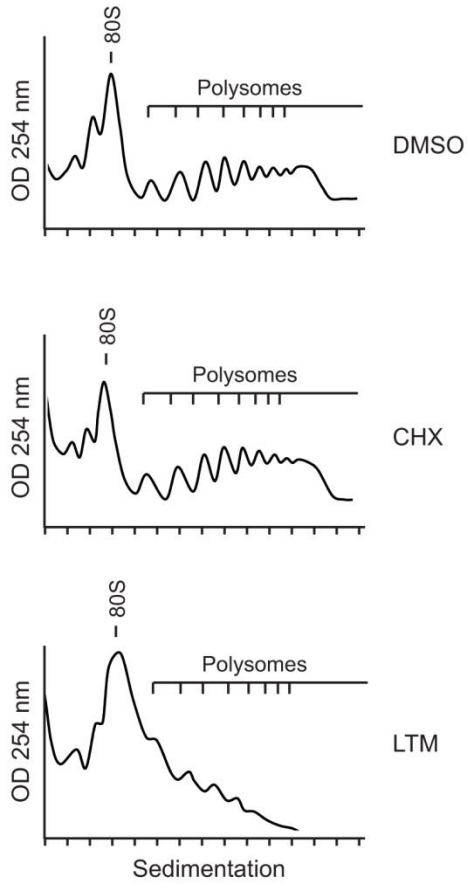


Figure 3-S1. Polysome profile analysis in cells treated with ribosome E-site translation inhibitors.
HEK293 cells were pre-treated with equal volume of DMSO, 100 μM CHX or 50 μM LTM for 30 min followed by sucrose gradient sedimentation. Both 80S monosome and polysome peaks are indicated.

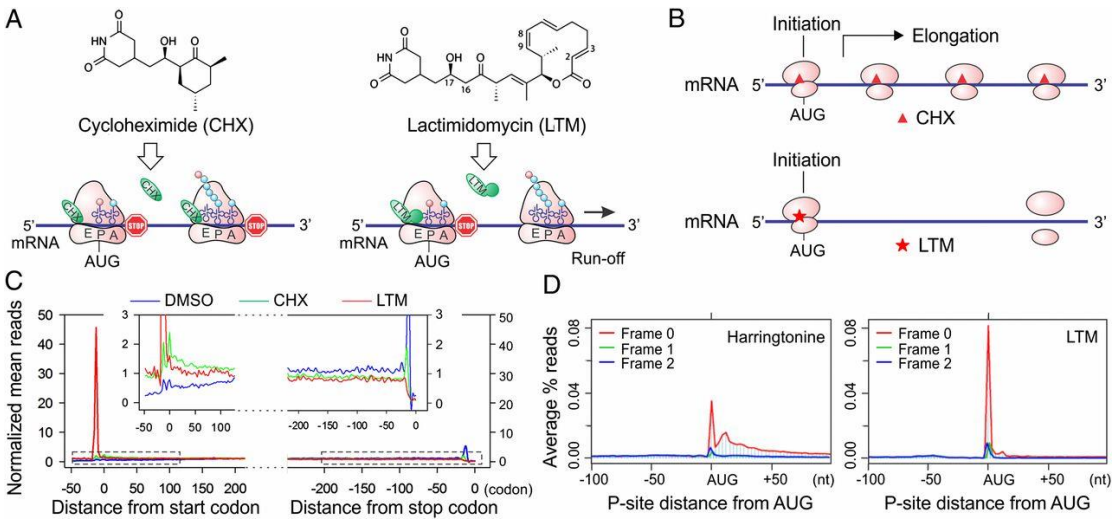


Figure 3-1. Experimental strategy of GTI-seq using ribosome E-site translation inhibitors.
(A) Schematic diagram of the experimental design for GTI-seq. Translation inhibitors CHX and LTM bind to the ribosome E-site, resulting in inhibition of translocation. While CHX binds to all translating ribosomes (left panel), LTM preferably incorporates into the initiating ribosomes when the E-site is free of tRNA (right panel).
(B) Ribosome profiling using CHX and LTM side-by-side allows distinguishing the initiating ribosome from the elongating one.
(C) HEK293 cells were treated with either DMSO, 100 μM CHX, or 50 μM LTM for 30 min before ribosome profiling. Normalized RPF reads are averaged across the entire transcriptome, aligned at either their start site or stop codon.
(D) Metagene analysis of RPFs obtained from HEK293 cells treated with either harringtonine (left panel) or LTM (right panel). All mapped reads are aligned at the annotated start codon AUG, and the reads density at each nt position is averaged using the P-site of RPFs.

During the course of our study, Ingolia *et al* reported a similar TIS mapping approach using harringtonine, a different translation initiation inhibitor (16). One key difference between harringtonine and LTM is that the former drug binds to free 60S subunits (22), whereas LTM binds to the 80S complexes already assembled at the start codon (17). We compared the pattern of RPF density surrounding the annotated start codon between the published datasets (16) and the LTM results (Fig. S2). It appears that a considerable amount of harringtonine-associated RPFs are not exactly located at the annotated start codon. To directly compare the TIS mapping accuracy between LTM and harringtonine, we performed ribosome profiling in HEK293 cells treated

with harringtonine using the same protocol as the LTM treatment. Similar to the previous study, harringtonine treatment caused a substantial fraction of RPFs accumulated in regions downstream of the start codon (Fig. 1D). The relaxed positioning of harringtonine-associated RPFs after prolonged treatment leaves uncertainty in TIS mapping. In contrast, GTI-seq using LTM largely overcomes this deficiency and offers a high precision in global TIS mapping with single nucleotide resolution (Fig. 1D)

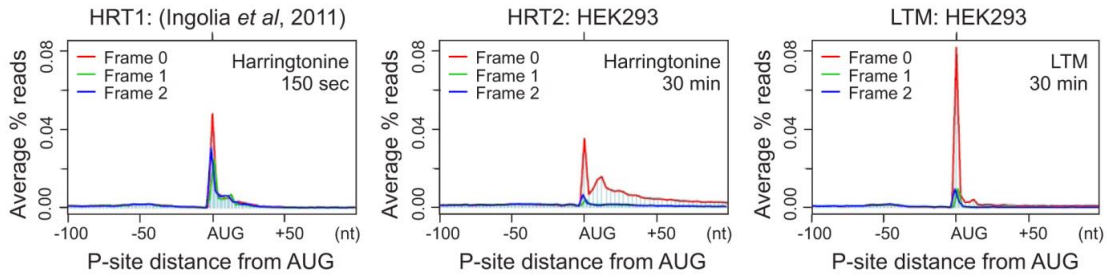


Figure 3-S2. Metagene analysis of RPFs obtained using different approaches.
RPF reads reported by Ingolia *et al* (2011) using harringtonine in mouse embryonic stem cells were re-plotted after P-site adjustment based on the original report (HRT1, left panel). RPF reads obtained from HEK293 cells treated with either harringtonine (HRT2, middle panel) or LTM (right panel) were plotted by applying a 12 nt offset to reads with a length range of 26 ~ 29 nt. All mapped reads are aligned at the annotated start codon AUG, and the reads density at each nt position is averaged using the P-site of RPFs.

3.3.2 Global TIS Identification by GTI-seq

One of the advantages of GTI-seq is its ability to analyze LTM data in parallel with CHX. Due to the structural similarity between these two translation inhibitors, the LTM background reads resembled the pattern of CHX-associated RPFs (Fig. 2A). This feature allows us to further reduce the background noise of LTM-associated RPFs by subtracting the normalized CHX reads density at every nucleotide position from that of LTM. A TIS peak is then called at a position in which the adjusted LTM reads density is well above the background (Fig. 2A, red asterisk, see Methods for

detail). From ~4,000 transcripts with detectable TIS peaks, we identified a total of 16,231 TIS sites (Table S1, online). Codon composition analysis revealed that more than half of the TIS codons used AUG as the translation initiator (Fig. 2B). GTI-seq also identified a significant proportion of TIS codons employing near-cognate codons that differ from AUG by a single nucleotide, in particular CUG (16%). Remarkably, nearly half of the transcripts (42%) contained multiple TIS sites (Fig. 2C), suggesting that alternative translation prevails even under physiological conditions. Surprisingly, about a third of the transcripts (32.4%) showed no TIS peaks at the annotated TIS position (aTIS) despite clear evidence of translation (Table S1). While some of them could be false negative due to stringent threshold cutoff for TIS identification (Fig. S3), others were likely attributed to alternative translation initiation (see below). However, it is also possible that some cases represent mis-annotation. For instance, the translation of *CLK3* clearly starts from the second AUG, although the first one was annotated as the initiator in the current database (Fig. 2D). We found 50 transcripts have possible mis-annotation in their start codons (Table S2, online). However, it is possible that some mRNAs might have alternative transcript processing. In addition, we could not exclude the possibility that some of these genes might have tissue-specific translation initiation sites.

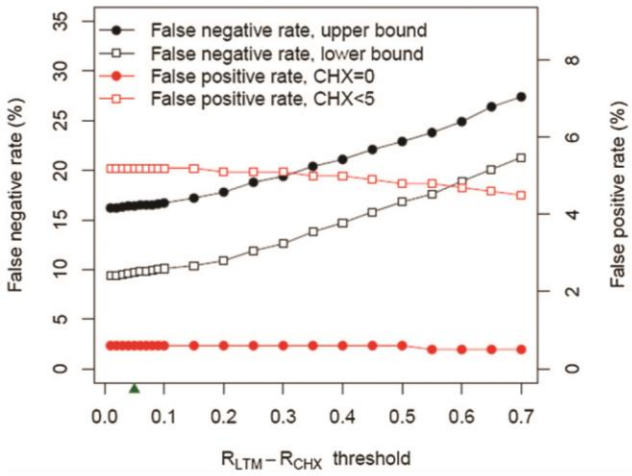


Figure 3-S3. False positive and false negative rates at various R_{LTM}-R_{CHX} thresholds. False negative rate is computed as the percentage of undetected aTIS among the top 10% translated aTIS codons based on CHX reads within 5 codons downstream of the aTIS. The lower and upper bounds of false negative rate are determined by either including or excluding the cases having a dTIS within 5 codons and/or a uORF overlapping with aTIS. False positive rate is computed as the percentage detected among strictly untranslated aTIS codons with either no CHX reads (CHX=0) or less than 5 CHX reads (CHX<5) within 5 codons downstream of the aTIS.

3.3.3 Characterization of Downstream Initiators

In addition to validating initiation at the annotated start codon, GTI-seq revealed clear evidence of downstream initiation on 39% of the analyzed transcripts with TIS peaks (Table S1). As a typical example, *AIMP1* showed three TIS peaks exactly at the first three AUG codons in the same reading frame (Fig. 3A). Thus, the same transcript generates three isoforms of AIMP1 with varied NH_2-terminus, which is consistent with the previous report (23). Of the total TIS positions identified by GTI-seq, 23% (3,741/16,231) were located downstream of aTIS codons, which we termed dTIS, and nearly half of the identified dTIS codons utilized AUG as the initiator (Fig. 3B).

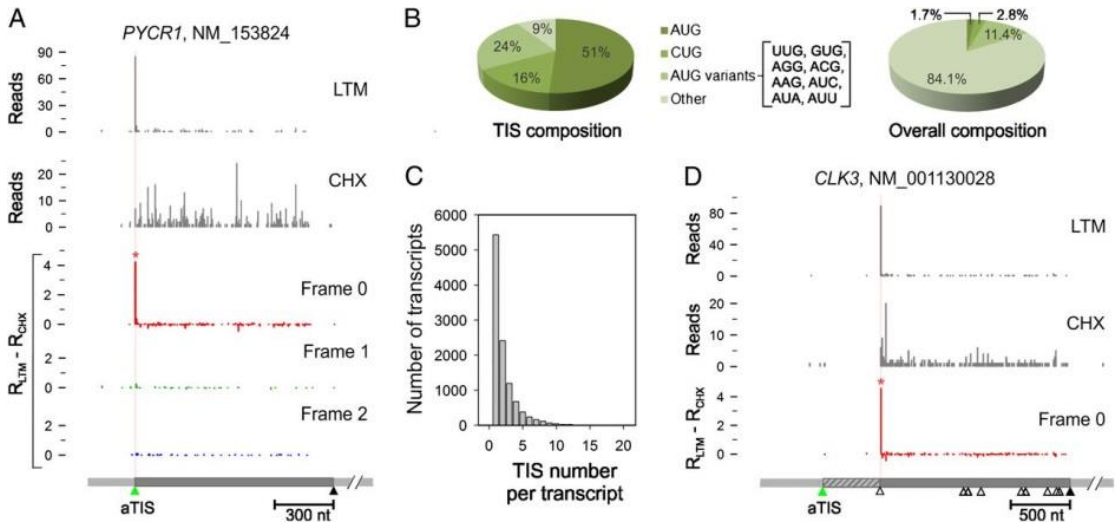


Figure 3-2. Global identification of TIS by GTI-seq.
(A) TIS identification on the *PYCR1* transcript. Both LTM and CHX reads are plotted as grey bar graph. TIS identification is based on normalized LTM reads density subtracted by CHX reads density. All three reading frames are separated and presented as distinct colors. Identified TIS position is marked by red asterisk and highlighted by colored shade. The annotated coding region is illustrated by start codon (green triangle) and stop codon (black triangle).
(B) Codon composition of all TIS codons identified by GTI-seq (left panel) is shown in comparison to the overall codon distribution over the entire transcriptome (right panel).
(C) Histogram showing the overall distribution of TIS number identified on each transcript.
(D) Mis-annotation of the start codon on the *CLK3* transcript. The annotated coding region is illustrated by start codon (green triangle) and stop codon (black triangle). AUG codons on the body of coding region are also shown as empty triangles. Only one reading frame is shown for clarity.

What are the possible factors influencing downstream start codon selection? We classified genes with multiple TIS codons into three groups based on Kozak consensus sequence of the first AUG. The relative leakiness of the first AUG codon was estimated by measuring the fraction of LTM reads at the first AUG over the total reads recovered on and after this position. The AUG codon with a strong Kozak sequence context showed the highest initiation efficiency (or lowest leakiness) in comparison to the one with weak or no consensus sequence (Fig. 3C, $p = 1.12 \times 10^{-142}$). These results indicate the critical role of sequence context in start codon recognition. To substantiate this conclusion further, we performed a reciprocal

analysis by grouping genes according to whether an initiation peak was identified at the aTIS or dTIS positions on their transcripts (Fig. 3D). A survey of the sequences flanking the aTIS revealed a clear preference of Kozak sequence context for different gene groups. In the gene group with aTIS initiation, but no detectable dTIS, we observed the strongest Kozak consensus sequence (Fig. 3D, bottom panel). This sequence context was largely absent in the group of genes lacking detectable translation initiation at the aTIS (Fig. 3D, top panel). Thus ribosome leaky scanning tends to occur when the context of an aTIS is suboptimal.

Cells use the leaky scanning mechanism to generate protein isoforms with changed subcellular localizations or altered functionality from the same transcript (24). In addition to genes that have been reported to produce protein isoforms via leaky scanning, GTI-seq revealed many more cases than previously reported (Table S1). To independently validate the novel dTIS positions identified by GTI-seq, we cloned the gene *CCDC124* whose transcript showed several initiation peaks above the background (Fig. 3E). One dTIS is in the same reading frame of the aTIS, which allows us to use a COOH-terminal tag to detect different translational products in transfected cells. Immunoblotting of transfected HEK293 cells showed two clear bands whose molecular weights correspond to the full length of CCDC124 (28.9 kDa) and the NH_2-terminally truncated isoform (23.7 kDa), respectively. Intriguingly, the relative abundance of both isoforms matched well to the corresponding LTM reads density, suggesting that GTI-seq might provide quantitative aspects of translation initiation.

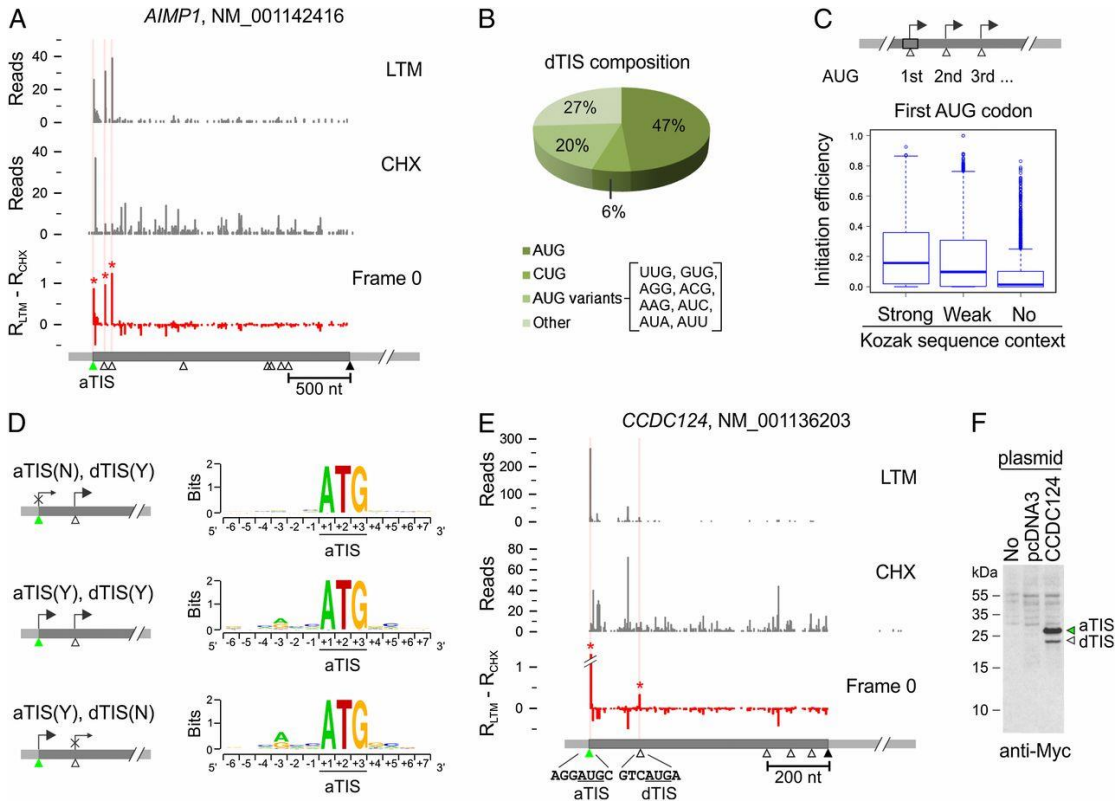


Figure 3-3. Characterization of downstream TIS (dTIS).
(A) Identification of multiple TIS codons on the *AIMP1* transcript. Only one reading frame is shown for clarity.
(B) Codon composition of total dTIS codons identified by GTI-seq.
(C) Relative initiation efficiency at the first AUG codon with different sequence context (one-tailed Wilcoxon Ranksum test, Strong vs. Weak: $p = 7.92 \times 10^{-24}$; Weak vs. No-Kozak: $p = 1.34 \times 10^{-75}$).
(D) Genes are grouped according to the identified initiation at either aTIS, dTIS, or both. Sequence context surrounding the aTIS is shown as Sequence Logos. Chi-square test, $p = 2.57 \times 10^{-100}$ for -3 position and $p = 3.95 \times 10^{-18}$ for +4 position.
(E) Identification of multiple TIS codons on the *CCDC124* transcript.
(F) Validation of *CCDC124* TIS codons by immunoblotting. The DNA fragment encompassing both the 5`UTR and the CDS of *CCDC124* was cloned and transfected into HEK 293 cells. Whole cell lysates were immunoblotted using c-myc antibody.

3.3.4 Characterization of Upstream Initiators

Sequence-based computational analyses predicted that about 50% of mammalian transcripts contain at least one uORF (11, 12). In agreement with this notion, GTI-seq revealed 54% of transcripts bearing one or more TIS positions upstream of the annotated start codon (Table S1). These upstream TIS (uTIS) codons, when out of the aTIS reading frame, are often associated with short ORFs. A classic example is *ATF4*, whose translation is predominantly controlled by several uORFs (25-27). This feature was clearly captured by GTI-seq (Fig. 4A). In addition to the two known uORFs proximal to the aTIS, we identified another extremely short uORF at the beginning of the *ATF4* mRNA. Intriguingly, the AUG start codon is immediately followed by a UAG stop codon. This one-codon uORF was clearly marked by both LTM and CHX-associated RPFs. As expected, the presence of these uORFs efficiently repressed the initiation at the aTIS as evidenced by few CHX reads along the CDS of *ATF4*. Despite the low enrichment of LTM reads at the aTIS of *ATF4*, a specific LTM peak was still distinguishable above the background (Fig. 4A, insert). This example highlights the remarkable sensitivity of GTI-seq in capturing TIS codons with low initiation efficiency.

Of the total TIS positions identified by GTI-seq, nearly half of them were uTIS (7,936/ 16,231). In contrast to the dTIS, which utilized AUG as the primary start codon (Fig. 3B), the majority of uTIS (74.4%) were non-AUG codons (Fig. 4B). Among these AUG variants, CUG was the most prominent one with the frequency even higher than AUG (30.3% vs. 25.6%). In a few well-documented examples, the CUG triplet was reported to serve as an alternative initiator (13). To experimentally

confirm the alternative initiators identified by GTI-seq, we cloned the gene *RND3* that showed a clear initiation peak at a CUG codon in addition to the aTIS (Fig. 4C). The two initiators are in the same reading frame without a stop codon in between, which permits us to detect different translational products using an antibody against the fused COOH-terminal tag. Immunoblotting of transfected HEK293 cells showed two protein bands corresponding to the CUG-initiated long isoform (34 kDa) and the main product (31 kDa) (Fig. 4C). Once again, the levels of both isoforms were in accordance with the relative densities of LTM reads, further supporting the quantitative feature of GTI-seq in TIS mapping.

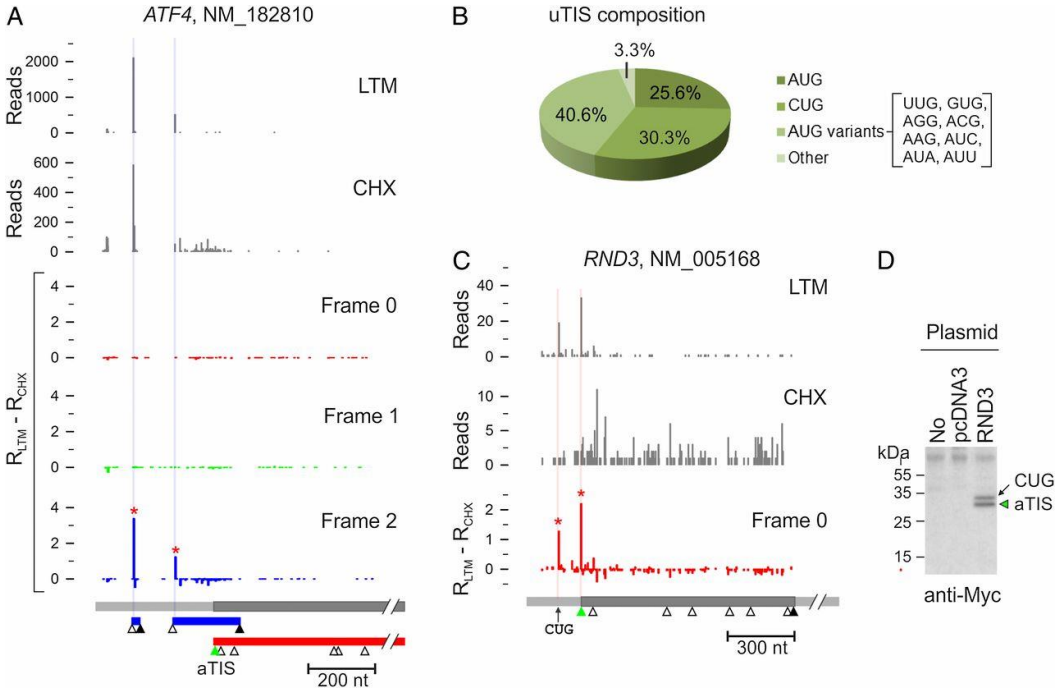


Figure 3-4. Characterization of upstream TIS (uTIS).
(A) Identification of multiple TIS codons on the *ATF4* transcript. Different ORFs are shown in boxes with colors matching different reading frames.
(B) Codon composition of total uTIS codons identified by GTI-seq.
(C) Identification of multiple TIS codons on the *RND3* transcript.
(D) Validation of *RND3* TIS codons by immunoblotting. The DNA fragment encompassing both the 5`UTR and the CDS of *RND3* was cloned and transfected into HEK 293 cells. Whole cell lysates were immunoblotted using c-myc antibody.

3.3.5 Global Impacts of uORFs on Translational Efficiency

Initiation from an uTIS, and the subsequent translation of the short uORF, negatively influences the main ORF translation (10, 11). To find possible factors governing the alternative TIS selection in the 5'UTR, we categorized uTIS-bearing transcripts into two groups according to whether initiation occurs at the aTIS and compared the sequence context of uTIS codons (Fig. 5A). For transcripts with initiation at both uTIS and aTIS positions [aTIS(Y)], the uTIS codons were preferentially composed of non-optimal AUG variants. In contrast, the uTIS codons identified on transcripts with repressed aTIS initiation [aTIS(N)] showed a higher percentage of AUG with Kozak consensus sequences ($p = 1.74 \times 10^{-80}$). These results are in agreement with the notion that the accessibility of an aTIS to the ribosome for initiation depends on the context of uTIS codons.

Recent work showed a correlation between secondary structure stability of local mRNA sequences near the start codon and mRNA translation efficiency (28-30). To examine whether the uTIS initiation is also influenced by local mRNA structures, we computed the free energy associated with secondary structures from regions surrounding the uTIS position (Fig. 5B). We observed an increased folding stability of the region shortly after the uTIS in transcripts with repressed aTIS initiation (Fig. 5B, blue line). In particular, more stable mRNA secondary structures were present on transcripts with less optimal uTIS codons (Fig. 5B, right panels). Therefore, when the consensus sequence is absent from the start codon, the local mRNA secondary structure has a stronger correlation with the TIS selection.

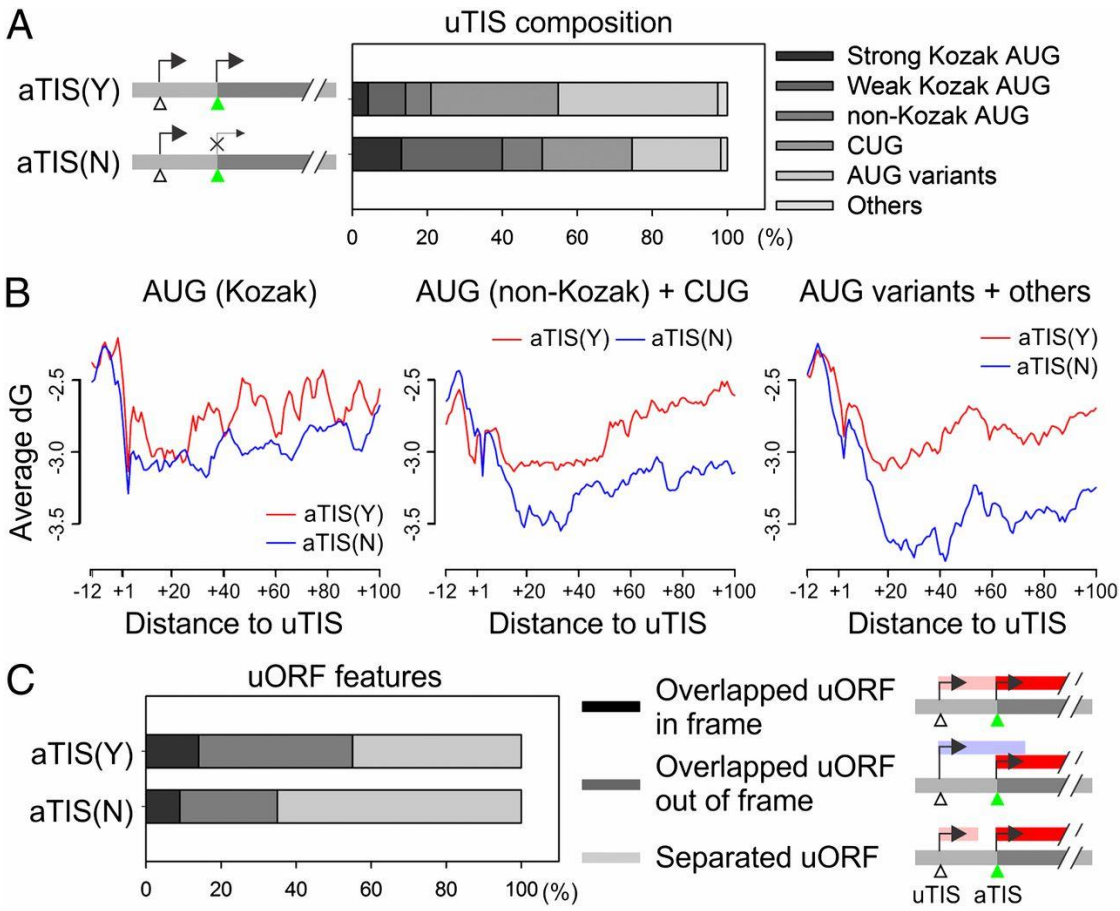


Figure 3-5. Impact of uORF features on translational regulation.
(**A**) The sequence composition of uTIS codons for genes with or without aTIS initiation. Genes are classified into two groups based on aTIS initiation, and the uTIS sequence composition is categorized based on the consensus features shown on the right.
(**B**) The contribution of mRNA secondary structure to TIS selection. Genes are grouped based on uTIS codon features listed in (A). For each group, the transcripts with (red line) or without (blue line) aTIS initiation are analyzed for the averaged ΔG value in regions surrounding the identified uTIS codons.
(**C**) The composition of uORFs in gene groups with or without aTIS initiation on their transcripts. Different ORF features are shown on the right.

Depending on the uTIS positions, the associated uORF can be separated from or overlapped with the main ORF. These different types of uORF could use different mechanisms to control the main ORF translation. For instance, when the uORF is short and separated from the main ORF, the 40S subunit can remain associated to the mRNA after termination at the uORF stop codon and resumes scanning, a process

called reinitiation (2). When the uORF overlaps with the main ORF, the aTIS initiation solely relies on the leaky scanning mechanism. We sought to dissect the respective contributions of reinitiation and leaky scanning to the regulation of aTIS initiation. Interestingly, we found a higher percentage of separated uORFs in transcripts with repressed aTIS initiation [aTIS(N) group] (Fig. 5C, $p = 3.52 \times 10^{-41}$). This result suggests that the re-initiation is generally less efficient than leaky scanning, which is consistent with the negative role of uORFs in translation of main ORFs.

3.3.6 Cross-species Conservation of Alternative Translation Initiators

The prevalence of alternative translation re-shapes the proteome landscape by either increasing the protein diversity or modulating translation efficiency. The biological significance of alternative initiators could be preserved across species if they are of potential fitness benefit. We applied GTI-seq to a mouse embryonic fibroblast (MEF) cell line and identified TIS positions across the mouse transcriptome, including uTIS and dTIS (Table S3, online). Compared to HEK293 cells, MEF cells showed remarkable similarity in overall TIS features (Fig. S4). For example, uTIS codons utilized non-AUG, especially CUG, as the dominant initiator. Additionally, about half of the transcripts in MEF cells exhibited multiple initiators. Thus, the general features of alternative translation are well conserved between human and mouse cells.

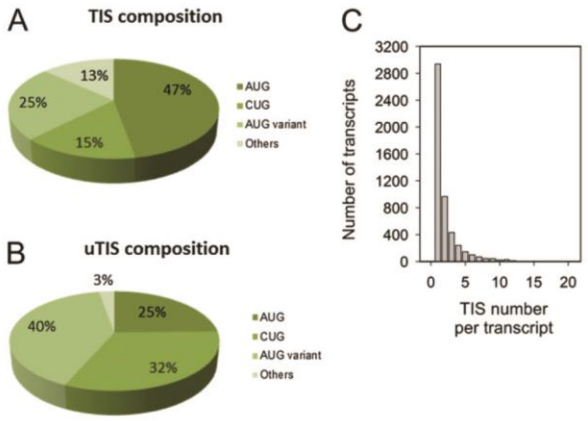
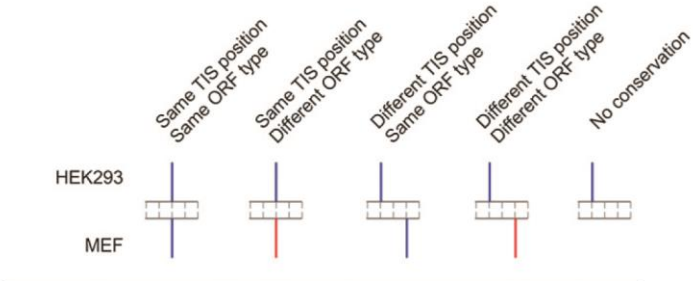


Figure 3-S4. Global TIS identification in MEF cells.
(**A**) Codon composition of all TIS codons identified by GTI-seq in MEF cells.
(**B**) Codon composition of uTIS codons identified by GTI-seq in MEF cells.
(**C**) Histogram showing the overall distribution of the number of TIS positions identified on each transcript from MEF cells.

To analyze conservation of individual alternative TIS position on each transcript, we chose a total of 12,949 human-mouse orthologous mRNA pairs. We analyzed separately the 5'UTR and CDS regions in order to measure the conservation of uTIS and dTIS positions, respectively (Fig. 6A). Each group was classified into two subgroups based on their sequence similarity. For genes with high sequence similarity, 85% of the uTIS and 60% of dTIS positions were conserved between human and mouse cells. Some of these alternative TIS codons were located at the same positions on the aligned sequences (Fig. S5). As an example, *RNF10* in HEK293 cells showed three uTIS positions, which were also found in MEF cells at the identical positions on the aligned 5'UTR sequence of the mouse homolog (Fig. 6B). Remarkably, genes with low sequence similarity also displayed high TIS conservation across the two species (Fig. 6A). For instance, the 5'UTR of *CTTN* gene has low sequence identity between human and mouse homologs (alignment score = 40.3) (Fig. 6C). However, a clear uTIS was identified in both cells at the same position on the

aligned region. Notably, the majority of alternative ORFs conserved between human and mouse cells were of the same type, i.e., either separated from or overlapped with the main ORF (Fig. 6A and Fig. S5). The evolutionary conservation of those TIS positions and the associated ORFs is a strong indication of functional significance of alternative translation in the regulation of gene expression.



	Sequence similarity					
uTIS	H	19%	1%	58%	7%	15%
	L	8%	0%	47%	9%	36%
dTIS	H	15%	0%	32%	13%	39%
	L	9%	0%	29%	8%	54%

Figure 3-S5. Conservation of alternative TIS positions between human and mouse cells. Alternative TIS positions identified on human mRNAs are classified based on whether the position, sequence context or ORF type are conserved in the mouse orthologous mRNAs (same color represents same type). The TIS site with a mouse counterpart at the identical position or with a similar local sequence context on the aligned orthologous sequences are merged. Both uTIS and dTIS positions are each classified into two subsets according to the global alignment score of sequences (5'UTR for uTIS and CDS for dTIS). Percentage values are presented in the table.

3.3.7 Characterization of ncRNA Translation

The mammalian transcriptome contains many non-protein-coding RNAs (ncRNAs) (31). ncRNAs have gained much attention recently due to their emerging role in a variety of cellular processes including embryogenesis and development (32). Motivated by the recent report about the possible translation of large intergenic ncRNAs (lincRNAs) (16), we sought to explore the possible translation, or at least

ribosome association, of ncRNAs in HEK293 cells. We selected RPFs uniquely mapped to ncRNA sequences to exclude the possibility of spurious mapping of reads originated from mRNAs. Of 5,763 ncRNAs annotated in RefSeq, we identified 169 ncRNAs (about 3%) that were associated with RPFs marked by both CHX and LTM (Fig. 6D and Table S4, online). Compared to protein-coding mRNAs, most ORFs recovered from ncRNAs were very short with a median length of 82 nt (Fig. 6E). Several ncRNAs also showed alternative initiation at non-AUG start codons as exemplified by *LOC100499177* (Fig. 6F).

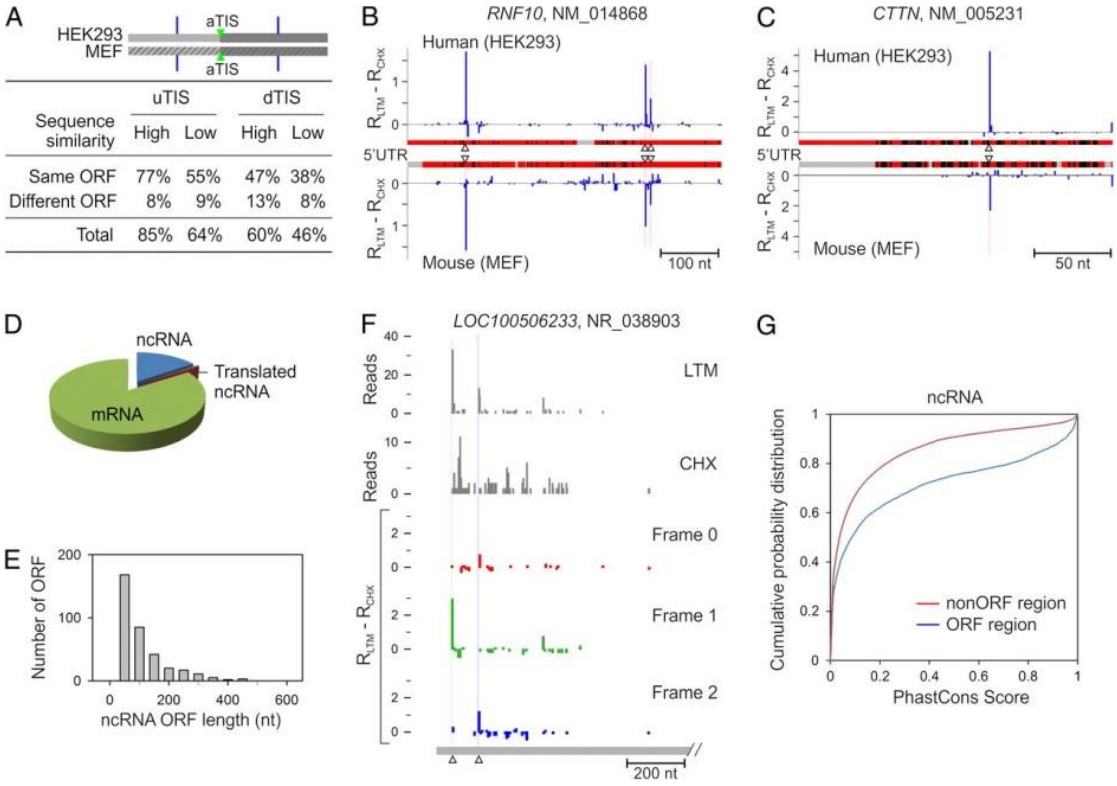


Figure 3-6. Cross-species conservation of alternative TIS positions and identification of translated ncRNA.
(**A**) Evolutionary conservation of alternative TIS positions identified by GTI-seq in HEK293 and MEF cells. Alternative uTIS and dTIS positions identified on human-mouse ortholog mRNA pairs are each classified into two subsets according to the alignment score of relevant sequences (5'UTR for uTIS and CDS for dTIS). Each subset is further divided based on types of alternative ORFs. Percentage values are presented in the table.

(**B**) Conservation of uTIS positions on the *RNF10* transcript with high sequence similarity of 5'UTR between HEK293 and MEF cells. Red region indicates matched sequences, black for mismatched ones, and grey for sequence gaps. Identified uTIS positions are indicated by triangles.
(**C**) Conservation of uTIS positions on the *CTTN* transcript with low sequence similarity of 5'UTR between HEK293 and MEF cells.
(**D**) Pie chart showing the relative percentage of mRNA, ncRNA and translated ncRNA identified by GTI-seq.
(**E**) Histogram showing the overall length distribution of ORF identified in ncRNAs.
(**F**) Identification of multiple TIS positions on the ncRNA *LOC100499177*.
(**G**) Evolutionary conservation of ORF region on ncRNAs identified by GTI-seq. PhastCons scores are retrieved from the primate genome sequence alignment.

Comparative genomics reveals that the coding regions are often evolutionarily conserved elements (33). We retrieved the PhastCons scores for both coding and non-coding regions of ncRNAs and found that the ORF regions identified by GTI-seq indeed showed a higher conservation (Fig. 6G). Some ncRNAs showed a clear enrichment of highly conserved bases within the ORFs marked by both LTM and CHX reads (Fig. S6). Despite the apparent engagement by the protein synthesis machinery, the physiological functions of the coding capacity of these ncRNAs remain to be determined.

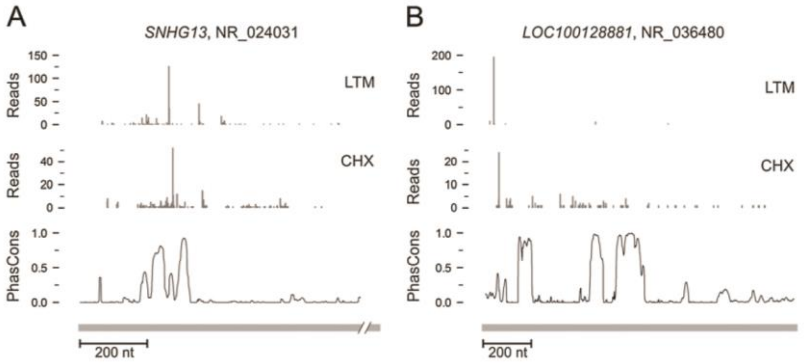


Figure 3-S6. ORF conservation in ncRNAs.
(**A**) Translation in ncRNA *SNHG13* is illustrated by LTM and CHX-associated RPF reads. PhastCons scores retrieved from the primate genome sequence alignment is also plotted (bottom panel).
(**B**) Translation in ncRNA *LOC100128881* is illustrated by LTM and CHX-associated RPF reads. PhastCons scores retrieved from the primate genome sequence alignment is also plotted (bottom panel).

3.4 Discussion

The mechanisms of eukaryotic translation initiation have received increasing attention owing to their central importance in diverse biological processes (1). The use of multiple initiation codons in a single mRNA contributes to protein diversity by expressing several protein isoforms from a single transcript. Distinct ORFs defined by alternative TIS codons could also serve as regulatory elements in controlling the translation of the main ORF (10, 11). Although we have some understanding of how ribosomes determine where and when to start initiation, our knowledge is far from complete. GTI-seq provides a comprehensive and high-resolution view of TIS positions across the entire transcriptome. The precise TIS mapping offers mechanistic insights into the start codon recognition.

3.4.1 Global TIS Mapping at Single Nucleotide Resolution by GTI-seq

Traditional toeprinting analysis showed heavy ribosome pausing at both the initiation and the termination codons of mRNAs (34, 35). Consistently, deep sequencing-based ribosome profiling also revealed the highest RPF density at both the start and the stop codons (14, 15). Although this feature enables approximate determination of decoded mRNA regions, it does not allow for unambiguous identification of TIS positions especially when multiple initiators are utilized. Translation inhibitors specifically acting on the first round of peptide bond formation allow the run-off of elongating ribosomes, thereby specifically halting ribosomes at the initiation codon. Indeed, harringtonine treatment caused a profound accumulation of RPFs in the beginning of CDS (16). A caveat of using harringtonine is that this

drug binds to free 60S subunits and the inhibitory mechanism is unclear. In particular, it is not known whether harringtonine completely blocks the initiation step. We observed that a significant fraction of ribosomes still passed over the start codon in the presence of harringtonine.

The translation inhibitor LTM bears several features in achieving the high resolution of global TIS identification. First, LTM binds to the 80S ribosome already assembled at the initiation codon and permits the first peptide bond formation (17). Thus, the LTM-associated RPF more likely represents physiological TIS positions. Second, LTM occupies the empty E-site of initiating ribosomes and thus completely blocks the translocation. This feature allows the TIS identification at single nucleotide resolution. With this precision, different reading frames become unambiguous, thereby revealing different types of ORFs within each transcript. Third, owing to the similar structure and the same binding site in the ribosome, LTM and CHX can be applied side-by-side to achieve simultaneous assessment of both initiation and elongation for the same transcript. With the high signal/noise ratio, GTI-seq offers a direct TIS identification approach with a minimal computational aid. From our analysis, the uncovering of alternative initiators allows us to probe mechanistic insights of TIS selection. We also experimentally validated different translational products initiated from alternative start codons, including non-AUG. Further confirming the accuracy of GTI-seq, a sizable fraction of alternative start codons identified by GTI-seq exhibited high conservation across species. The evolutionary conservation strongly suggests a physiological significance of alternative translation in gene expression.

3.4.2 Diversity and Complexity of Alternative Start Codons

GTI-seq revealed that the majority of identified TIS positions belong to alternative start codons. The prevailing alternative translation was corroborated by the finding that nearly half of the transcripts contained multiple TIS codons. While dTIS codons use the conventional AUG as the main initiator, a significant fraction of uTIS codons are non-AUG with the CUG as the most frequent one. In a few well-documented cases, including *FGF2* (36), *VEGF*(37), and *Myc* (38), the CUG triplet was reported to serve as the non-AUG start codon. With the high resolution TIS map across the entire transcriptome, GTI-seq greatly expanded the list of hidden coding potential of mRNAs not visible by sequence-based *in silico* analysis.

By what mechanisms are alternative start codons selected? GTI-seq revealed several lines of evidence supporting the linear scanning mechanism for start codon selection. First, the uTIS context, such as the Kozak consensus sequence and the secondary structure, largely influenced the frequency of aTIS initiation. Second, the stringency of an aTIS codon negatively regulated the dTIS efficiency. Third, the leaky potential at the first AUG was inversely correlated with the strength of its sequence context. Since it is less likely for a preinitiation complex to bypass a strong initiator to select a downstream suboptimal one, it is not surprising that most uTIS codons are not canonical whereas the dTIS codons are mostly conventional AUG. In addition to the leaky scanning mechanism for alternative translation initiation, ribosomes could translate a short uORF and reinitiate at downstream ORFs (2). After completing termination of a uORF, it was assumed that some translation factors remain associated with the ribosome, which facilitates the reinitiation process (39). However, this

mechanism is widely considered to be inefficient. From the GTI-seq data set, about half of the uORFs were separated from the main ORFs. Compared to transcripts with overlapping uORFs that must rely on leaky scanning to mediate the downstream translation, we observed repressed aTIS initiation in transcripts containing separated uORFs. It is likely that the ribosome reinitiation mechanism plays a more important role in selective translation under stress conditions (27).

3.4.3 Biological Impacts of Alternative Translation Initiation

One expected consequence of alternative translation initiation is an expanded proteome diversity that has not been and could not be predicted by *in silico* analysis of AUG-mediated main ORFs. Indeed, many eukaryotic proteins exhibit a feature of NH_2-terminal heterogeneity presumably due to alternative translation. Protein isoforms localized in different cellular compartments are typical examples, because most localization signals are within the NH_2-terminal segment (40, 41). Alternative TIS selection could also produce functionally distinct protein isoforms. One well-established example is C/EBP, a family of transcription factors that regulate the expression of tissue-specific genes during differentiation (42).

When an alternative TIS codon is not in the same frame as the aTIS, it is conceivable that the same mRNA will generate unrelated proteins. This could be particularly important for the function of uORFs, which are often separated from the main ORF and encode short polypeptides. Some of these uORF peptide products directly control the ribosome behavior, thereby regulating the translation of the main ORF. For instance, the translation of *S*-adenosylmethionine decarboxylase is subject

to the regulation by the six amino acid product of its uORF (43). The alternative translational products could also function as biologically active peptides. A striking example is the discovery of short ORFs (sORFs) in noncoding RNAs of *Drosophila* that produce functional small peptides during development (44). However, both computational prediction and experimental validation of peptide-encoding short ORFs within the genome are challenging. Our study using GTI-seq represents a potential new addition to the expanding ORF catalog by including novel ORFs from ncRNAs.

3.4.4 Perspective

The enormous biological breadth of translational regulation has led to an enhanced appreciation of its complexities. Yet, the current endeavors aiming to understand protein translation have been hindered by technological limitations. Comprehensive cataloging of global translation initiation sites and the associated ORFs is just the beginning in unveiling the role of translational control in gene expression. More focused studies will be needed to decipher the function and regulatory mechanism of novel ORFs individually. A systematic, high-throughput method like GTI-seq offers a top-down approach, in which one can identify a set of candidate genes to study intensively. GTI-seq is readily applicable to broad fields of fundamental biology. For instance, applications of GTI-seq in different tissues will facilitate the elucidation of the tissue-specific translational control. The illustration of altered TIS selection under different growth conditions will set the stage for future investigation of translational reprogramming during organismal development as well as in human diseases.

3.5 Materials & Methods

Cell Culture and Drug Treatment

Human HEK293 and mouse embryonic fibroblast (MEF) were maintained in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum (FBS). Cycloheximide (CHX) was purchased Sigma and harringtonine from LKT Laboratories. Lactimidomycin (LTM) was previously described (45). All drugs were dissolved in DMSO. Cells were treated with 100 μM CHX, 50 μM LTM, 2 μg/ml (3.8 μM) harringtonine or equal volume of DMSO at 37 ℃ for 30min.

Polysome Profiling

Sucrose solution was prepared in polysome buffer (pH 7.4, 10 mM HEPES, 100 mM KCl, 5 mM $MgCl_2$). Sucrose density gradients (15%- 45% w/v) were freshly made in SW41 ultracentrifuge tubes (Backman) using a Gradient Master (BioComp Instruments) according to manufacturer's instructions. Cells were washed using ice-cold PBS containing 100 μg/ml CHX and then lysed by scraping extensively in polysome lysis buffer (pH 7.4, 10 mM HEPES, 100 mM KCl, 5 mM $MgCl_2$, 100 μg/ml CHX and 2% Triton X-100). For DMSO control, the CHX was omitted in both PBS and polysome lysis buffer. Cell debris were removed by centrifugation at14,000 rpm for 10 min at 4 ℃. 600 μl of supernatant was loaded onto sucrose gradients followed by centrifugation for 100 min at 38,000 rpm 4 ℃ in a SW41 rotor. Separated samples were fractionated at 0.750 ml / min through a fractionation system (Isco) that continually monitored OD_{254} values. Fractions were collected with 0.5 min interval.

Purification of Ribosome Protected mRNA Fragments (RPF)

The general procedure of RPF purification was based on the previously reported protocol (14) with some modifications. In brief, polysome profiling fractions were mixed and a 140 µl aliquot was digested with 200 U *E. coli* RNase I (Ambion) at 4 ℃ for 1h. Total RNA was then extracted by Trizol reagent (Invitrogen) followed by dephosphorylation with 20 U T4 polynucleotide kinase (NEB) in the presence of 10 U SUPERase_In (Ambion) at 37 ℃ for 1 hour. The enzyme was heat-inactivated for 20 min at 65 ℃. The digested RNA products were then separated on a Novex denaturing 15% polyacrylamide TBE-urea gel (Invitrogen). The gel was stained with SYBR Gold (Invitrogen) to visualize the digested RNA fragments. Gel bands around 28 nt RNA molecules were excised and physically disrupted by centrifugation through the holes of the tube. The gel debris was soaked overnight in the RNA gel elution buffer (300 mM NaOAc pH 5.5, 1 mM EDTA, 0.1 U/mL SUPERase_In) to recover the RNA fragments. The gel debris was filtered out with a Spin-X column (Corning) and RNA was purified using ethanol precipitation.

cDNA Library Construction and Deep Sequencing

Poly-A tails were added to the purified RNA fragments by *E. coli* poly-(A) polymerase (NEB) with 1 mM ATP in the presence of 0.75 U/µL SUPERase_In at 37 ℃ for 45 min. The tailed RNA molecules were reverse transcribed to generate the first strand cDNA using SuperScript III (Invitrogen) and following oligos containing barcodes:

SCT01:5’-

pCTGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGATT

TTTTTTTTTTTTTTTTTTVN-3';

MCA02: 5'-

pCAGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGAT

TTTTTTTTTTTTTTTTTTVN-3';

LGT03:5'-

pGTGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGATT

TTTTTTTTTTTTTTTTTTVN-3'.

HTC04: 5'-

pTCGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGATT

TTTTTTTTTTTTTTTTTTVN-3'

YAG05:5'-

pAGGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGATT

TTTTTTTTTTTTTTTTTTVN-3';

Reverse transcription products were resolved on a 10% polyacrylamide TBE-urea gel as described above. The expected 92 nt band of the first strand cDNA was excised and recovered using DNA gel elution buffer (300 mM NaCl, 1 mM EDTA). The purified first strand cDNA was then circularized by 100 U CircLigase II (Epicentre) following manufacturer's instructions. The circular single strand DNA was purified using ethanol precipitation and re-linearized by 7.5 U APE 1 in 1 X buffer 4 (NEB) at 37 ℃ for 1 h. The linearized products were resolved on a Novex 10% polyacrylamide TBE-urea gel (Invitrogen). The expected 92 nt band was then excised and recovered.

The single-stranded template was then amplified by PCR using the Phusion High-Fidelity enzyme (NEB) according to the manufacturer's instructions. The primers qNTI200 (5'-CAAGCAGAAGACGGCATA-3') and qNTI201 (5'-AATGATACGGCGACCACCG ACAGGTTCAGAGTTCTACAGTCCGACG-3') were used to create DNA library suitable for sequencing. The PCR reaction contains 1 × HF buffer, 0.2mM dNTP, 0.5 μM primers, 0.5 U Phusion polymerase. PCR was carried out with an initial 30 s denaturation at 98 ℃, followed by 12 cycles of 10 s denaturation at 98 ℃, 20 s annealing at 60 ℃, and 10 s extension at 72 ℃. PCR products were separated on a non-denaturing 8% polyacrylamide TBE gel as described above. Expected 120 bp band was excised and recovered as described above. After quantification by Agilent BioAnalyzer DNA 1000 assay, equal amount of barcoded samples were pooled into one sample. 3 ~ 5 pmol mixed DNA samples were typically used for cluster generation followed by sequencing using sequencing primer 5'-CGACAGGTTCAGAGTTC TACAGTCCGACGATC-3' (Illumina HiSEQ, Cornell University Life Sciences Core Laboratories Center).

Mapping Ribosome Protected mRNA Fragments to RefSeq Transcripts

To remove adaptor sequences, seven nucleotides were cut off from the 3' end of each 50 nt-long Illumina sequence read and a stretch of A's were removed from the 3' end, allowing one mismatch. The remaining insert sequence was separated according to the 2-nucleotide barcode at the 5' end after the barcode was removed. Reads of length between 26 to 29 nt were mapped to the sense strand of the entire human or mouse RefSeq transcript sequence library (release 49), using Bowtie-0.12.7 (46). Reads mapped to the PhiX genome if any were removed beforehand. One

mismatch was allowed in all mappings and in case of multiple mapping, mismatched positions were not used if a perfect match existed. Reads mapped more than 100 times were discarded to remove poly-A-derived reads. Finally, reads were counted at every position of individual transcript by using the 13th nucleotide of the read for the P-site position. Two HEK293 technical replicate controls from the starvation dataset were pooled for most analyses representing HEK293.

Coding Sequence Annotation

The most recent freezes of CCDS (consensus coding sequence) data (47) were downloaded from the NCBI ftp site (Jan 24, 2011 for mouse, Sep 7, 2011 for human) to find annotated translational start and end positions on each mRNA. Each of the CCDS nucleotide sequences were mapped to the associated RefSeq mRNA sequences based on following conditions: (1) the first three nucleotides must be perfectly matched; (2) up to two mismatches are allowed in the first ten nucleotides; (3) up to twenty mismatches are allowed in the full length, with no gaps allowed. The maximum number of mismatches in an accepted alignment was 10.

Read Aggregation Plots

The number of RPF reads aligned to each position of individual transcript was first normalized by the total reads recovered on the same mRNA. The reads counts were then averaged across all mRNAs for each position relative to the annotated start codon. To avoid multiple counting of the same reads mapped to multiple isoforms of the same gene, redundant mRNAs were removed based on the sequence context of -

100 nt ~ +100 nt relative to the annotated TIS. The same approach was used to obtain average read aggregation relative to dTIS or uTIS positions.

Identification of TIS Positions

A peak is defined at the nucleotide level on a transcript. A peak position satisfies the following conditions: (1) The transcript must have both LTM and CHX reads. (2) The position must have at least 10 reads from the LTM data. (3) The position must be a local maximum within 7 nucleotides. (4) The position must have "LTM-CHX" = ($X_{LTM}/N_{LTM} - X_{CHX}/N_{CHX}$) to be at least 0.005, where X_k is the number of reads on that position in data k and N_k is the total number of reads on that transcript in data k. Generally, a peak position is also called a 'TIS'. However, if a peak was not detected on the first position of any AUG or near-cognate start codon but was present at the first position of an immediately preceding or succeeding one of these codons, the position was called a TIS.

Identification of Potentially Misannotated aTIS

Among mRNAs with at least one identified dTIS position, those with no aTIS or uTIS peak were selected. Then, the first dTIS in frame 0 is identified as the potentially correct aTIS (pcaTIS). If this dTIS is not associated with an AUG or near-cognate start codon, it was discarded. Any mRNA with a 5'UTR shorter than 12 nt is excluded, because our method requires at least 12 nt 5'UTR in order to detect the aTIS that would be at the 13^{th} position on a read. To reduce possible false positives, we ensured that: (1) the total CHX reads in the region from position 1 to pcaTIS position -2 on an mRNA must be less than 10; (2) the maximum CHX reads in this region must be less than 2; (3) total LTM reads from position aTIS-1 to aTIS+1 must be 0; (4) the

average CHX read density between pcaTIS-1 and pcaTIS+11 must be higher than 0.1 reads per nucleotide.

Codon Composition Analysis

The number of TIS positions associated with each codon type starting was counted. The enumeration was done after filtering redundant TIS positions based on its flanking sequence context from -30 to +122 nt relative to the TIS position to avoid double counting of the TIS on the common regions of transcript isoforms. The same redundancy filtering was applied in most other analyses and counting described below. Background codon composition was based on all codons in either annotated CDS or 5'UTR of all mRNAs, regardless of reading frame. Redundancy filtering was not performed for background counting.

Ribosomal Leaky Scanning Analysis

Three subsets of aTIS positions were collected based on whether the aTIS has the initiation peak and whether the mRNA has any detectable AUG-associated dTIS (Figure 3D). Sequence logos were drawn using Berkeley Weblogo (48). The uTIS positions with the maximum peak height on an mRNA were grouped according to whether the aTIS has a peak [aTIS(Y)] or not [aTIS(N)] and their Kozak sequence context was analyzed (Figure 5A). For counting the types of uTIS-associated uORFs (Figure 5C), the most downstream uTIS on each mRNA was assigned to one of two groups according to whether the aTIS has a peak [aTIS(Y)] or not [aTIS(N)]. The same uTIS sets collected for the Kozak sequence context analysis were used for measurement of free energy of downstream RNA secondary structures. Each of these

subsets was divided into three groups according to the initiation context – 'AUG (Kozak)', 'AUG (non-Kozak) + CUG' and 'AUG variants + others'. The AUG (Kozak) group includes an AUG with either or both of -3A/G and +4G. AUG (non-Kozak) group is an AUG with neither -3A/G nor +4G. For each TIS position, a window length of 22 nt was moved at step size of 1 nt, starting from -12 nt relative to each uTIS to +100 nt, and the ΔG was calculated for each window using the RNAfold program (49). The ΔG values were averaged for each position relative to the uTIS across all uTIS positions in each set.

TIS Conservation between Human and Mouse

Human and mouse RefSeq protein accessions were extracted from HomoloGene (release 65) (50). Each RefSeq protein accession was matched to the associated mRNA accession, CCDS ID and CCDS amino acid sequence. The amino acid sequence of each homologous protein pair were aligned to each other using Clustalw 2.1 (51), to calculate the alignment score and filter one-to-one orthologous relationships. If two or more proteins from the same species were in the same HomoloGene group, only the single reciprocally best matched pair was used. Likewise, if an orthologous gene has mRNA isoforms, the reciprocally best matched isoform pair was chosen. Any tied matches were removed. The alignment score was computed as [1 - (the number of mismatches and gaps) / (length of human protein)] * 100. Any alignment with an alignment score less than 50 was discarded. The 5'UTR of an orthologous mRNA was considered as an orthologous 5'UTR.

Among the human mRNAs that have a mouse ortholog, 5'UTRs and CDSs were independently grouped into well-aligned and poorly-aligned categories. A 5'UTR with an alignment score less than 50 or with a 30 nt or longer 3'end gap is considered poorly aligned. Likewise, a CDS with a 30 nt or longer initial gap is also considered poorly aligned. Note that a CDS with an alignment score less than 50 was discarded beforehand. Within each category, human uTIS or dTIS were classified into five groups, according to sequence conservation (S0 vs S1)) and subtype conservation (T0 vs T1).

A TIS is conserved in sequence (S1) if there is a mouse TIS peak at the same position on the aligned orthologous mouse sequence or if there is a mouse TIS peak with a similar surrounding sequence. The surrounding sequence is taken from -6 to +24 nt relative to each uTIS. The sequence similarity must be at least 75% identity with no gaps. If a mouse TIS exists in the orthologous 5'UTR or CDS, but not conserved in sequence, it is assigned to the S0 category. If no mouse TIS exists, it is classified as 'N'. If the mouse ortholog has no detectable TIS at all, the pair was removed from the analysis.

A TIS is conserved in subtype (T1) if the corresponding mouse uTIS or dTIS is of the same type. For a uTIS, two subtypes, 'N-terminal extended' *versus* 'overlapped' and 'separated' were considered. For a dTIS, frame 0 *versus* frame 1 and 2 were used as two subtypes. The priority is set in the order of T1S1, T1S0, T0S1, T0S0 and N, in case a TIS belongs to two or more classes.

Identification of Translated ORFs in Non-coding RNA and Conservation Analysis

Human and mouse ncRNAs were collected from the RefSeq (release 49) by extracting the RNAs with an accession beginning with 'NR' and with no mRNA isoforms. To avoid false detection of TIS positions due to spurious mapping of reads sourced from mRNA transcripts, only reads unique to a single ncRNA are used. From the human ncRNAs with at least one identified TIS, PhastCons score for every nucleotide position within either ORF or non-ORF regions was collected. The PhastCons scores were obtained by using the UCSC Table Browser (http://genome.ucsc.edu) (52, 53), from the placental and primate subsets of the 46-way vertebrate genomic alignment. The ncRNAs whose genomic positions were ambiguous (e.g. the ncRNA is not included in the refGene table of the UCSC database or the length of the RNA is different from the refGene record) were excluded from the analysis.

Plasmid Construction and Immunoblotting

cDNA was synthesized by Superscript III RT (Invitrogen) using 1 μg of total RNA extracted from HEK293 cells. *CCDC124* and *RND3* gene encompassing both the 5'UTR and the CDS were amplified by PCR reaction using the following oligo pairs:

ccdc124F: 5'-GGCGCCAAGCTTGGAGGCGCGACCGGGCCGGCGCTGG-3'

ccdc124R: 5'-GGCGCCCTCGAGTTGGGGGCATTGAAGGGCACGGCCC-3'

rnd3F: 5'-GGCGCCAAGCTTCAGTCGGCTCGGAATTGGACTTGGG -3'

rnd3R: 5'-GGCGCCCTCGAGCTATTCTGCACCCTGGAGGCGTAGC-3'.

The PCR fragments were cloned to Hind III and Xho I sites of pcDNA™3.1/myc-His B. Plasmid transfection was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. After 48 hr transfection, cells were lysed by the lysis buffer (Tris-buffered saline, pH 7.4, 2% Triton X-100). The whole cell lysates were heat-denatured for 10 min in NuPAGE® LDS Sample Buffer (Invitrogen). The protein samples were resolved on 12% NuPAGE gel (Invitrogen) and then transferred to Immobilon-P membranes (Millipore). After blocking for 1 hour in TBS containing 5% blotting milk, membranes were incubated with c-*myc* antibodies (Santa Cruz Biotechnology) at 4℃ overnight. After incubation with horseradish peroxidase–coupled secondary antibodies (Sigma), immunoblots were developed using enhanced chemiluminescence (GE Healthcare).

3.6 Acknowledgements

We thank the Qian lab members for helpful discussions during the course of this study. We are grateful to Drs. Chaolin Zhang (Rockefeller University) and Adam Siepel (Cornell University) for critical reading of the manuscript. We also thank Cornell University Life Sciences Core Laboratory Center for performing deep sequencing. This work was supported by grants to B.S. from National Institutes of Health (CA106150) and to S.-B.Q. from National Institutes of Health (1 DP2 OD006449-01), Ellison Medical Foundation (AG-NS-0605-09), and DOD Exploration-Hypothesis Development Award (TS10078).

REFERENCES

1. Sonenberg N & Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136(4):731-745.

2. Jackson RJ, Hellen CU, & Pestova TV (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 11(2):113-127.

3. Gray NK & Wickens M (1998) Control of translation initiation in animals. *Annu Rev Cell Dev Biol* 14:399-458.

4. Kozak M (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 299(1-2):1-34.

5. Kozak M (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem* 266(30):19867-19870.

6. Kozak M (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci U S A* 87(21):8301-8305.

7. Maag D, Fekete CA, Gryczynski Z, & Lorsch JR (2005) A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol Cell* 17(2):265-275.

8. Martin-Marcos P, Cheung YN, & Hinnebusch AG (2011) Functional elements in initiation factors 1, 1A, and 2beta discriminate against poor AUG context and non-AUG start codons. *Mol Cell Biol* 31(23):4814-4831.

9. Iacono M, Mignone F, & Pesole G (2005) uAUG and uORFs in human and rodent 5'untranslated mRNAs. *Gene* 349:97-105.

10. Morris DR & Geballe AP (2000) Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* 20(23):8635-8642.

11. Calvo SE, Pagliarini DJ, & Mootha VK (2009) Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A* 106(18):7507-7512.

12. Resch AM, Ogurtsov AY, Rogozin IB, Shabalina SA, & Koonin EV (2009) Evolution of alternative and constitutive regions of mammalian 5'UTRs. *BMC Genomics* 10:162.

13. Touriol C, *et al.* (2003) Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biol Cell* 95(3-4):169-178.

14. Ingolia NT, Ghaemmaghami S, Newman JR, & Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324(5924):218-223.

15. Guo H, Ingolia NT, Weissman JS, & Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466(7308):835-840.

16. Ingolia NT, Lareau LF, & Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147(4):789-802.

17. Schneider-Poetsch T, *et al.* (2010) Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat Chem Biol* 6(3):209-217.

18. Klinge S, Voigts-Hoffmann F, Leibundgut M, Arpagaus S, & Ban N (2011) Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science* 334(6058):941-948.

19. Ju J, *et al.* (2009) Lactimidomycin, iso-migrastatin and related glutarimide-containing 12-membered macrolides are extremely potent inhibitors of cell migration. *J Am Chem Soc* 131(4):1370-1371.

20. Sugawara K, *et al.* (1992) Lactimidomycin, a new glutarimide group antibiotic. Production, isolation, structure and biological activity. *J Antibiot (Tokyo)* 45(9):1433-1441.

21. Steitz TA (2008) A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol* 9(3):242-253.

22. Fresno M, Jimenez A, & Vazquez D (1977) Inhibition of translation in eukaryotic systems by harringtonine. *Eur J Biochem* 72(2):323-330.

23. Shalak V, Kaminska M, & Mirande M (2009) Translation initiation from two in-frame AUGs generates mitochondrial and cytoplasmic forms of the p43 component of the multisynthetase complex. *Biochemistry* 48(42):9959-9968.

24. Kochetov AV (2008) Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. *Bioessays* 30(7):683-691.

25. Spriggs KA, Bushell M, & Willis AE (2010) Translational regulation of gene expression during conditions of cell stress. *Mol Cell* 40(2):228-237.

26. Harding HP, Calfon M, Urano F, Novoa I, & Ron D (2002) Transcriptional and translational control in the Mammalian unfolded protein response. *Annu Rev Cell Dev Biol* 18:575-599.

27. Vattem KM & Wek RC (2004) Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci U S A* 101(31):11269-11274.

28. Kudla G, Murray AW, Tollervey D, & Plotkin JB (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science* 324(5924):255-258.

29. Kochetov AV, *et al.* (2007) AUG_hairpin: prediction of a downstream secondary structure influencing the recognition of a translation start site. *BMC Bioinformatics* 8:318.

30. Kertesz M, *et al.* (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467(7311):103-107.

31. Mattick JS (2005) The functional genomics of noncoding RNA. *Science* 309(5740):1527-1528.

32. Pauli A, Rinn JL, & Schier AF (2011) Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet* 12(2):136-149.

33. Siepel A, *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15(8):1034-1050.

34. Wolin SL & Walter P (1989) Signal recognition particle mediates a transient elongation arrest of preprolactin in reticulocyte lysate. *J Cell Biol* 109(6 Pt 1):2617-2622.

35. Sachs MS, *et al.* (2002) Toeprint analysis of the positioning of translation apparatus components at initiation and termination codons of fungal mRNAs. *Methods* 26(2):105-114.

36. Vagner S, *et al.* (1996) Translation of CUG- but not AUG-initiated forms of human fibroblast growth factor 2 is activated in transformed and stressed cells. *J Cell Biol* 135(5):1391-1402.

37. Meiron M, Anunu R, Scheinman EJ, Hashmueli S, & Levi BZ (2001) New isoforms of VEGF are translated from alternative initiation CUG codons located in its 5'UTR. *Biochem Biophys Res Commun* 282(4):1053-1060.

38. Hann SR, King MW, Bentley DL, Anderson CW, & Eisenman RN (1988) A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell* 52(2):185-195.

39. Poyry TA, Kaminski A, & Jackson RJ (2004) What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes Dev* 18(1):62-75.

40. Chang KJ & Wang CC (2004) Translation initiation from a naturally occurring non-AUG codon in Saccharomyces cerevisiae. *J Biol Chem* 279(14):13778-13785.

41. Porras P, Padilla CA, Krayl M, Voos W, & Barcena JA (2006) One single in-frame AUG codon is responsible for a diversity of subcellular localizations of glutaredoxin 2 in Saccharomyces cerevisiae. *J Biol Chem* 281(24):16551-16562.

42. Descombes P & Schibler U (1991) A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. *Cell* 67(3):569-579.

43. Hill JR & Morris DR (1993) Cell-specific translational regulation of S-adenosylmethionine decarboxylase mRNA. Dependence on translation and coding capacity of the cis-acting upstream open reading frame. *J Biol Chem* 268(1):726-731.

44. Kondo T, *et al.* (2010) Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. *Science* 329(5989):336-339.

45. Ju J, *et al.* (2009) Lactimidomycin, iso-migrastatin and related glutarimide-containing 12-membered macrolides are extremely potent inhibitors of cell migration. *J Am Chem Soc* 131(4):1370-1371.

46. Langmead B, Trapnell C, Pop M, & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.

47. Pruitt KD, *et al.* (2009) The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res* 19(7):1316-1323.

48. Crooks GE, Hon G, Chandonia JM, & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188-1190.

49. Gruber AR, Lorenz R, Bernhart SH, Neubock R, & Hofacker IL (2008) The Vienna RNA websuite. *Nucleic Acids Res* 36(Web Server issue):W70-74.

50. Sayers EW, *et al.* (2011) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 39(Database issue):D38-51.

51. Larkin MA, *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947-2948.

52. Karolchik D, *et al.* (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32(Database issue):D493-496.

53. Kent WJ, *et al.* (2002) The human genome browser at UCSC. *Genome Res* 12(6):996-1006.

CHAPTER 4

Conclusions and Future Endeavors

Cellular protein homeostasis is maintained by integrated biological pathways that control the biogenesis, folding, trafficking and degradation of proteins. This dynamic balance is critical for the quantity, quality and flexibility of proteome, which is the key for proper cellular functions. The complex network involves several important functional units, including the translational machinery, molecular chaperones, and the ubiquitin proteasome system. In the past decades, great progress has occurred in understanding the structural and biochemical details of these three systems. However, it remains elusive how cells coordinate all these systems to maintain the protein homeostasis in response to intracellular and environmental challenges.

The work presented here focuses on the translational response induced by proteotoxic stress and nutrient starvation. I dissect the complex stress-induced translational reprogramming by inspecting the global protein synthesis as well as selective translation of stress responsive genes. Mechanistically, the translational control at both initiation and elongation stages are investigated with genomic and biochemical approaches to gain a global view across the translatome and a better understanding of underlying mechanisms.

In chapter 2, I discovered that intracellular proteotoxic stress reduces global protein synthesis by halting ribosomes on transcripts during elongation. Ribosome

profiling data reveals an early elongation pausing, roughly at the site where nascent polypeptide chains emerge from the ribosomal exit tunnel. The early elongation pausing response has also been captured by independent labs under several other proteotoxic conditions, including heat shock and oxidative stress. Repression of molecular chaperones by dominant negative mutants and chemical inhibitors recapitulates the early elongation pausing, suggesting cells exploit chaperone availability as a sensing mechanism to induce stress response at translational level. The functional connection between chaperone availability and translation elongation offers a novel mode of regulation in response to stress conditions.

The ribosome complex serves not only as a molecular machine of protein synthesis but also a sorting platform of nascent polypeptide chains for downstream modification, folding, and targeting. The journey of a nascent polypeptide starts from the peptidyl transferase center (PTC) of the ribosome followed by traversing the peptide exit tunnel. Once the nascent chain begins to emerge from the exit tunnel, it faces a drastic environmental change. The nascent chain is sampled and bound by a diverse array of ribosome associated factors, such as signal recognition particle (SRP), nascent polypeptide associated complex (NAC), and Hsc70. Depending on the affinity with the N-terminal peptides of nascent chain, ribosome associated factors compete with each other to bind with newly synthesized proteins and determine their post-synthesis fate. Intriguingly, when one factor is depleted by genetic approaches, the corresponding nascent chain substrates are ectopically bound by other factors, leading to a variety of phenotypes. To explain the regulatory mechanisms underlying the early elongation pausing, I proposed two possible mechanistic models. The first

"depletion" model states that intracellular accumulation of misfolded proteins, a common feature of a variety of stress conditions, sequesters molecular chaperones and the lack of chaperone association with the ribosome delays nascent chains from emerging. The clogged nascent chains in the ribosome exit tunnel might compromise elongation directly. Alternatively, the exposed nascent chains might be bound with other factors, such as Argonaute2, which in turn blocks the release of elongation factors, leading to the early pausing. I proposed this "depletion" model in the published paper of Chapter 2.

However, current available data could also be explained by an alternative "occupation" model, in which nascent polypeptide chains emerging from ribosome exit tunnel may retain associated Hsc70 under proteotoxic stress and trapped Hsc70 prevents the release of eEF1A from ribosome, causing the early elongation pausing. This model is supported by the accumulation of both Hsc70 and eEF1A on the ribosome under direct inhibition of Hsc70 by small molecules as presented in Appendix I. The position information of ribosomes associated with Hsc70 and eEF1A under proteotoxic stress is needed to differentiate the two potential models. Currently, I am developing a selective ribosome profiling approach (SeRP) in mammalian system. Eventually, I am going to rely on the coming SeRP data to further investigate the detailed mechanisms. Since multiple factors constitute the chaperone network linked to protein synthesis, it will also be intriguing to determine whether interfering with specific chaperone or co-chaperone molecules causes selective elongation pausing on a subset of transcripts by taking advantage of SeRP.

The field of translational reprogramming has made great progress over the past decade, in large part stemming from technological developments such as ribosome profiling. Yet, the current endeavors aiming to understand protein translation have been hindered by technological limitations. In Chapter 3 of a collaborative work, I develop an approach (global translation initiation sequencing, GTI-seq), by applying in parallel ribosome E-site translation inhibitors lactimidomycin (LTM) and cycloheximide (CHX), to achieve simultaneous detection of both initiation and elongation events on a genome-wide scale. GTI-seq provides a comprehensive and high-resolution view of TIS positions across the entire transcriptome. The precise TIS mapping offers mechanistic insights into the start codon recognition. GTI-seq reveals that the majority of identified TIS positions belong to alternative start codons. The prevailing alternative translation is corroborated by the finding that nearly half of the transcripts contained multiple TIS codons. One expected consequence of alternative translation initiation is an expanded proteome diversity, including NH_2-terminal heterogenetic protein isoforms with different cellular localization and truncated proteins with distinct functions. After the publication of our method and data, several follow up studies have confirmed many of the isoforms with mass spectrometry and discovered truncated protein variants with novel physiological functions such as in antiviral immunity. In terms of alternative translation at uORF, most of them may serve as regulatory elements in controlling the translation of the main ORF, which has also been consistently discovered in the processes of virus infection, yeast meiosis, and mouse embryonic stem cell differentiation. Additionally, the alternative translational products could also function as biologically active peptides.

Comprehensive cataloging of global translation initiation sites and the associated ORFs is just the beginning in unveiling the role of translational control in gene expression. More focused studies will be needed to decipher the function and regulatory mechanism of novel ORFs individually. For example, a truncated isoform of a mitochondrial ribosomal protein identified from my dataset has been further pursued by Dr. Xingqian (Ben) Zhang in our lab to investigate its stress-induced alternative translation and novel function in regulating the synthesis of Hsp70 during heat shock response.

GTI-seq is also readily applicable to broad fields of fundamental biology. For instance, the illustration of altered TIS selection under different growth conditions will set the stage for future investigation of stress-induced translational reprogramming. In Appendix II, I apply GTI-seq to profile alternative initiation and selective translation in response to starvation. I uncover a robust translational reprogramming of protein catabolic process, in particular the proteasome system, through alternative TIS selection. This regulatory mode of TIS selection indicates that the scope of selective translation under stress conditions is much broader than anticipated.

Despite the great success in mapping translation initiation sites (TISs) with high precision, GTI-seq does not offer reliable quantitative information about differential initiation and cannot be directly applied to different tissues to elucidate the tissue-specific translational control. To overcome these limitations, I help Drs. Xiangwei Gao and Ji Wan to develop an enhanced version that allows an unbiased capture of initiating ribosomes from unperturbed cells. With the application of the improved approach to mouse liver tissue, we identify a liver cell-specific mode of translational

control and reveal a robust translational switch towards the proteasome system in response to fasting, which is consistent with my previous data from the cell culture model. These results further support the coordinated regulation between translation and proteasome systems to maintain protein homeostasis under stress conditions.

The ability of cells to adapt to stress is crucial for their survival. Regulation of global protein synthesis coupled with selective translation allows cells to rapidly respond to a variety of stress conditions. My work here has revealed unprecedented proteome complexity and flexibility through stress-induced translational reprogramming, including ribosome pausing during elongation and wide-spread alternative translation initiation. Elucidating the mechanisms underlying translational reprogramming during stress will not only shed light on the fundamental principles of translation, but also provide deeper insight of the pathophysiology of human diseases. Stress conditions are often an underlying cause of human diseases, including diabetes, neurodegenerative disorders, and cancer. In particular, cancer cells proliferate rapidly under limited nutrients and are relatively resistant to environmental stress. It is thus critical to understand how abnormal cells alter stress responsive pathways at the translational level. A better understanding of translational reprogramming in stress response might ultimately lead to the development of new therapeutic strategies for human diseases.

APPENDIX I

Hsc70 Modulates Elongation Pausing of Ribosomes in Response to Proteotoxic Stress through eEF1A

This section is work in process to further dissect the underlying mechanism of early elongation pausing after the publication of Chapter 2.

AI.1 Results

AI.1.1 eEF1A accumulates on light polysome under proteotoxic stress

In Chapter 2, I described the chaperone controlled early elongation pausing under proteotoxic stress, but the underlying mechanism remains elusive and several questions await further investigation. For instance, what is the difference between early and late elongating ribosomes? How does Hsc70 control the early elongation pausing under proteotoxic stress?

Unsurprisingly, one common regulatory mechanism of translation elongation is through the modulation of elongation factors (Sherman and Qian, 2013). Translation elongation is mediated by eukaryotic translation elongation factor 1 alpha (eEF1A) and eukaryotic translation elongation factor 2 (eEF2), which delivers amino acid-charged tRNA to the ribosomal A site and catalyses ribosomal translocation,

respectively (Dever and Green, 2012). In addition to its canonical role in translation, eEF1A has many other "moonlighting" functions related to the protein quality control system, including the association with Hsp70, nascent polypeptide chains, and ubiquitin-proteasome system, as well as the involvement in protein degradation, aggregation and heat shock response (Chuang and Madura, 2005; Deplazes et al., 2009; Gross and Kinzy, 2007; Mateyak and Kinzy, 2010; Meriin et al., 2012; Mitsui et al., 2002; Shamovsky et al., 2006) (Figure AI-1). Intriguingly, eEF1A has been shown to stall ribosomes at early elongation stage by forming a complex with RNA binding protein PUF and Argonaute protein on selective transcripts (Friend et al., 2012). Notably, the authors actually identified Hsp70 as a potential component of the complex using mass spectrometry, but they didn't investigate its function in details. In addition, during epithelia-to-mesenchymal transitions (EMT), 3'UTRs of specific mRNAs are recognized by a RNA-binding protein heterogeneous nuclear ribonucleoprotein E1 (hnRNP E1), which blocks the translocation of ribosomes by associating with eEF1A. Active Transforming growth factor beta (TGF β) signalling phosphorylates hnRNP E1 and releases eEF1A from ribosomes, allowing the elongation to proceed on mRNAs and promoting EMT (Hussey et al., 2011).

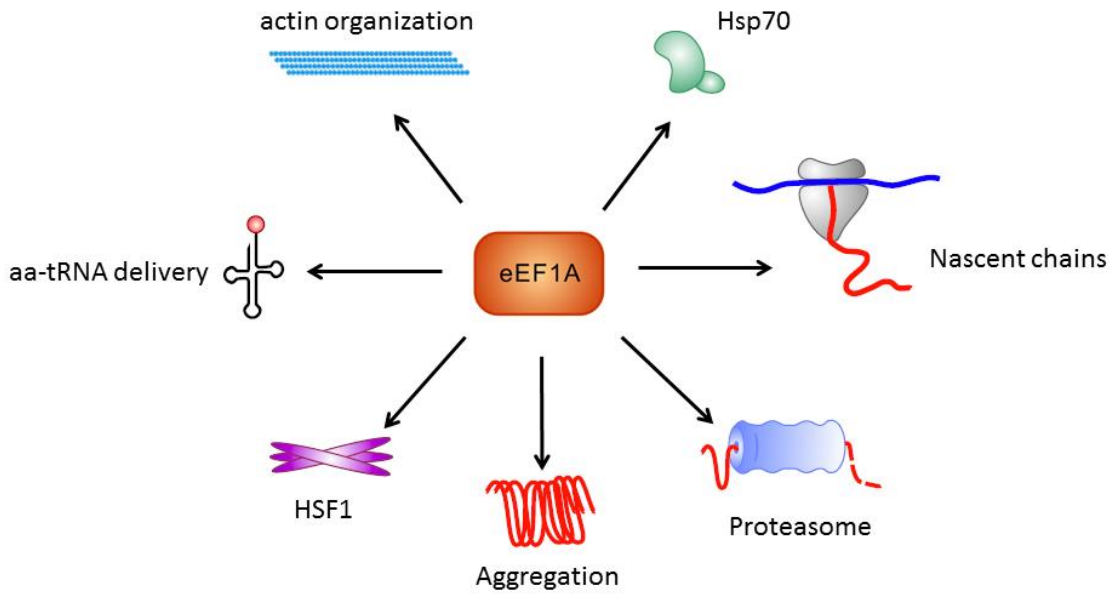


Figure AI-1. Canonical and non-canonical functions attributed to eEF1A.
Figure is adapted from Mateyak, M.K. *et al*. 2010 with modifications to emphasize the "moonlighting" functions related to protein quality control system.

I hypothesized that eEF1A might collaborate with Hsc70 to control early elongation pausing using similar mechanism. To test this hypothesis, I first detected the association of eEF1A and Hsc70 with ribosomes under normal and proteotoxic conditions (Figure AI-2). Consistent with previous observations, Hsc70 and eEF1A molecules were depleted from the heavy polysome that collapsed under stress conditions. However, Hsc70 and eEF1A remained associated with light polysome (2-4 ribosomes) under stress, suggesting potential roles in mediating the specific ribosome dynamics.

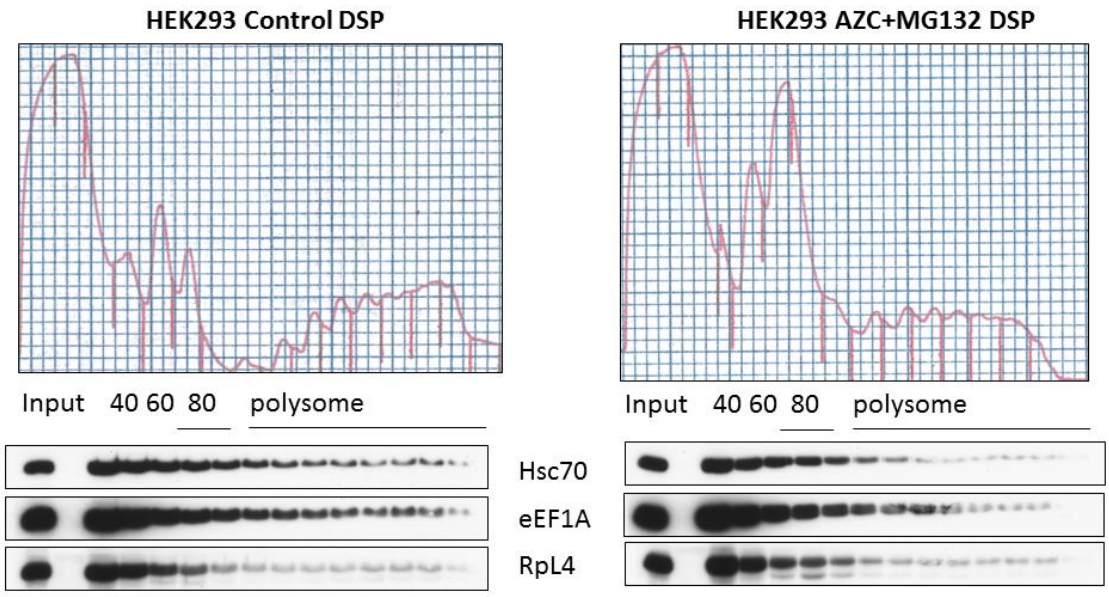


Figure AI-2. eEF1A accumulates on light polysome under proteotoxic stress.
Polysome profiles of HEK293 cells treated with or without 10mM AZC and 20μM MG132 for 1hr. 40S, 60S subunits, 80S monosome and polysome peaks are indicated. The bottom panel shows the distribution of Hsc70, eEF1A, and ribosome proteins in ribosome fractions.

AI.1.2 Direct inhibition of Hsc70 leads to the accumulation of eEF1A on the polysome

In Chapter 2, I showed that direct inhibition of Hsc70 with small molecules such as PES recapitulates the early elongation pausing. To characterize the interaction between Hsc70 and eEF1A under this scenario, I monitored the association of these proteins with ribosomes under Hsc70 inhibition (Figure AI-3). Previously, PES has been shown to bind with C-terminal substrate binding domain of Hsp70, the inducible form of Hsc70, and block its chaperone activity (Leu et al., 2009). Surprisingly, PES dramatically increased the amount of ribosome associated Hsc70 despite of the dissembled polysome, indicating that PES stabilizes the binding between Hsc70 and nascent polypeptide chains on ribosomes. Importantly, with more retained Hsc70 molecules, eEF1A also accumulated on ribosomes, suggesting that Hsc70 trapped with

nascent polypeptide chains might block the release of eEF1A from ribosomes and then cause the early elongation pausing.

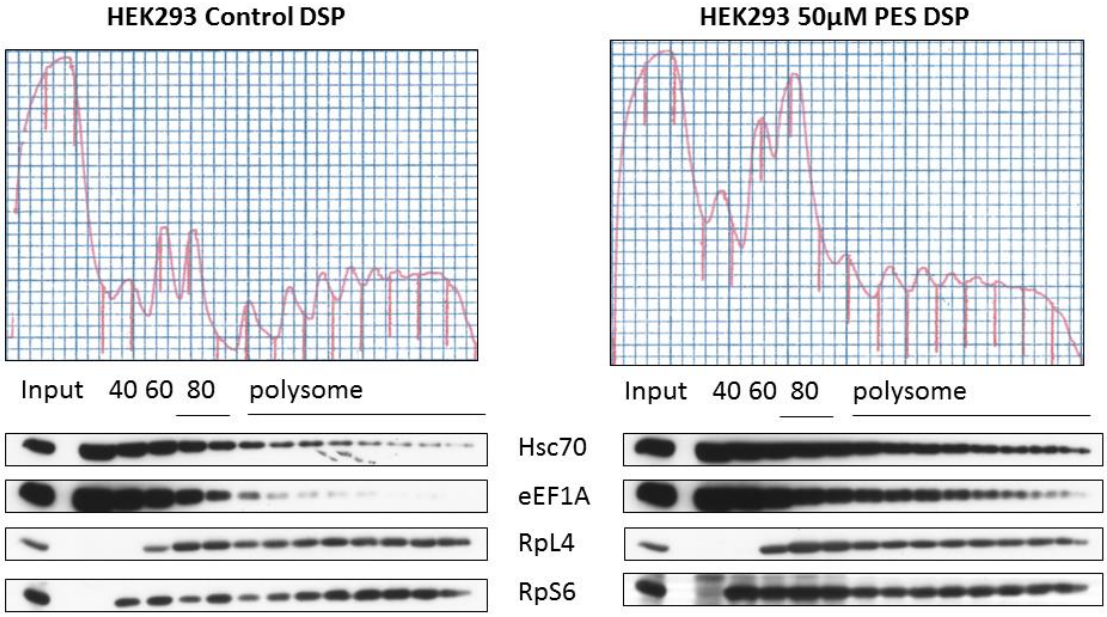


Figure AI-3. Direct inhibition of Hsc70 leads to the accumulation of eEF1A on the polysome.
Polysome profiles of HEK293 cells treated with or without 50μM PES for 1hr. 40S, 60S subunits, 80S monosome and polysome peaks are indicated. The bottom panel shows the distribution of Hsc70, eEF1A, and ribosome proteins in ribosome fractions.

AI.1.3 Direct repression of eEF1A results in a modest elongation pausing

To directly dissect the role of eEF1A in mediating early elongation pausing, I capitalized another chemical inhibitor against eEF1A called Narciclasine (Narc) (Van Goietsenoven et al., 2010) (Figure AI-4A). Indeed, treatment of cells with Narc dramatically inhibited the protein synthesis rate as indicated by the reduced incorporation of [^{35}S] (Figure AI-4B).

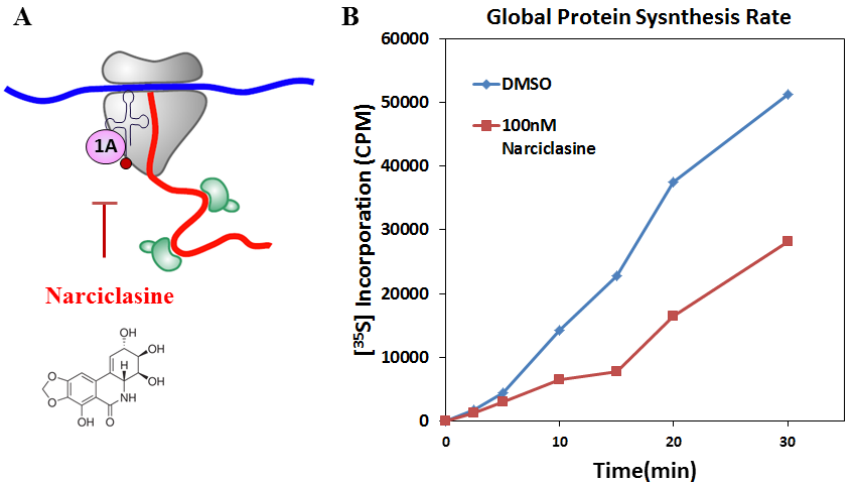


Figure AI-4. eEF1A inhibitor represses global protein synthesis.
(A) Schematic for the inhibitory property of small molecule Narciclasine. Narciclasine inhibits eEF1A (purple). The nascent polypeptide chain is colored in red while Hsc70 in green.
(B) Global protein synthesis was analyzed in HEK293 cells treated with 100 nM Narc for 1 hr. [^{35}S] radioactivity of TCA-insoluble material was measured at given times.

However, Narc only slightly dissembled polysome and increased monosome peak, with accumulation of eEF1A on ribosomes (Figure AI-5A). Furthermore, ribosome profiling revealed only a modest accumulation of ribosomes at 5' end of CDS under eEF1A inhibition as compared to the pausing level under proteotoxic stress and Hsc70 inhibition (Figure AI-5B).

To further substantiate the findings, I reconstituted the process based on the rabbit reticulocyte lysate (RRL) *in vitro* translation system. Instead of performing high-throughput sequencing, I monitored the positions of ribosomes on a reporter *firefly* luciferase (Fluc) mRNA indirectly through the length of associated nascent polypeptide chains (Janssen et al., 2012) (Figure AI-6A). Consistent with the ribosome profiling data, direct inhibition of eEF1A *in vitro* also leads to a general accumulation of elongating ribosomes on mRNA without drastic preference to the 5'end of CDS (Figure AI-6B).

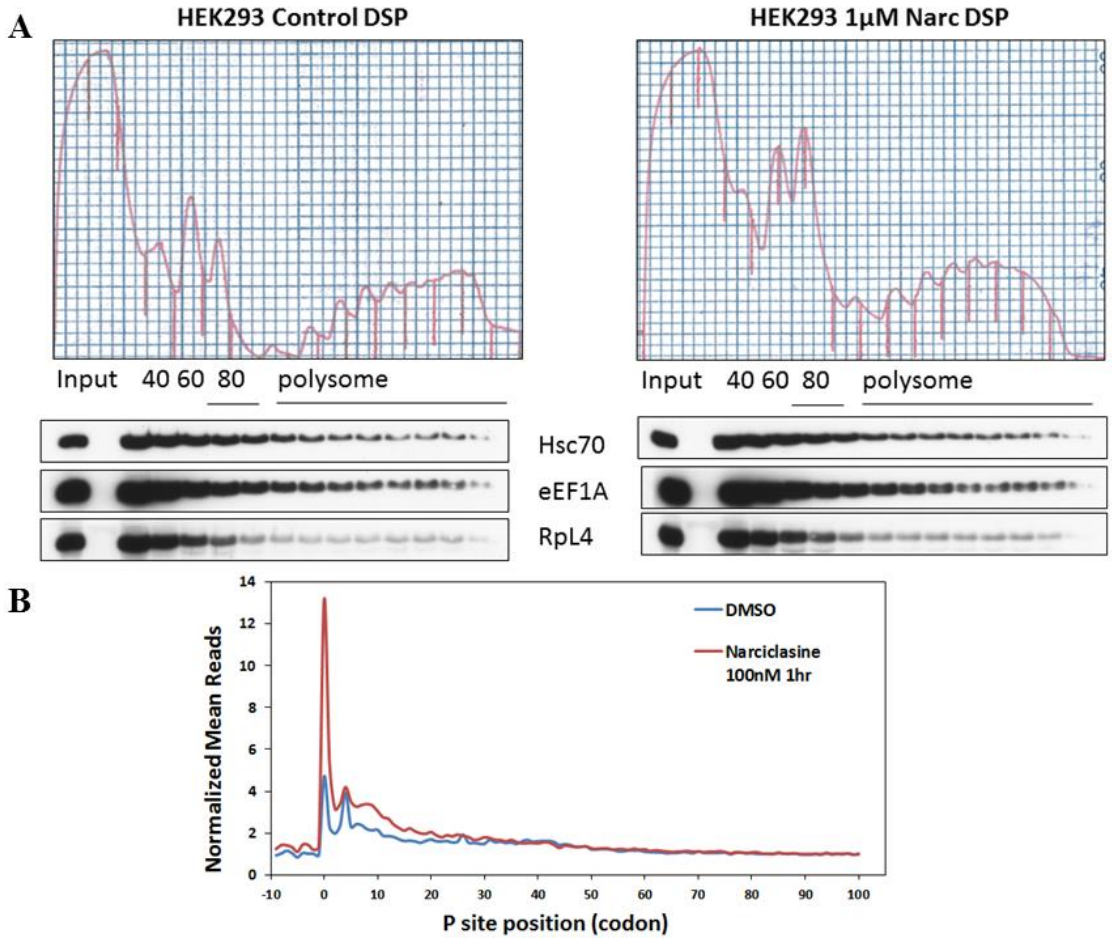


Figure AI-5. Direct inhibition of eEF1A partially recapitulates early elongation pausing. **(A)** Polysome profiles of HEK293 cells treated with or without 1μM Narc for 1hr. 40S, 60S subunits, 80S monosome and polysome peaks are indicated. The bottom panel shows the distribution of Hsc70, eEF1A, and ribosome proteins in ribosome fractions.
(B) Meta-gene analysis for modest early elongation pausing in cells treated with 100 nM Narc for 1hr. Normalized RPF reads are averaged across the entire transcriptome, aligned at their start codon.

These results suggest that repression of eEF1A only is not sufficient to fully recapitulate the early elongation pausing pattern. Instead, the binding of Hsc70 with emerging nascent chains might be required to provide the position information. Therefore, the interplay among nascent chains emerging from ribosome exit tunnel, Hsc70, and eEF1A is necessary to generate the specific early pausing pattern.

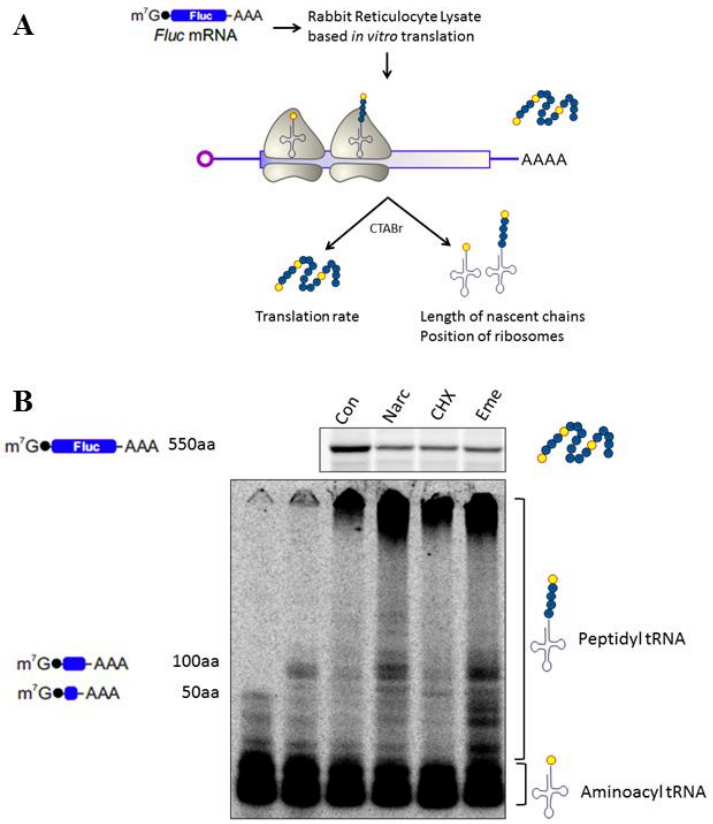


Figure AI-6. Inhibition of eEF1A *in vitro* causes accumulation of elongating ribosomes. **(A)** Schematic of experimental design using the length of associated nascent polypeptide chains to infer ribosome positions on Fluc mRNA *in vitro*.
(B) *In vitro* translation of Fluc mRNA in the presence of DMSO control (Con), 1μM Narciclasine (Narc), or E site elongation inhibitor 10μM Cycloheximide (CHX), or 10μM Emetine (Eme). Reactions were first performed at 30 ℃ for 20 min to allow the engagement of ribosomes on mRNAs before the addition of inhibitors. Full length products are shown on the top, while separated peptidyl tRNA species are shown at the bottom to reflect the distribution of elongating ribosomes. First two lanes are products from templates of N-terminal Fluc fragments used as markers.

AI.1.4 Selective ribosome profiling monitors the association of Hsc70 and eEF1A with elongating ribosomes

Based on all the evidence above, I propose a working model that nascent polypeptide chains emerging from ribosome exit tunnel may retain associated Hsc70 under proteotoxic stress, which in turn prevents the release of eEF1A from ribosome A site and leads to the early elongation pausing.

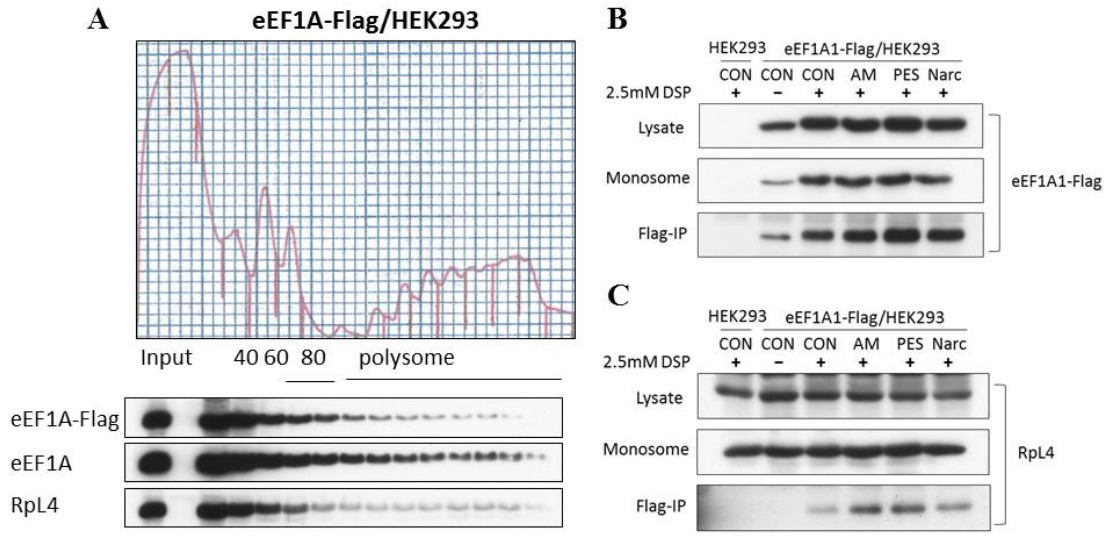


Figure AI-7. Selective ribosome profiling using C-terminal Flag-tagged eEF1A.
(A) Polysome profiles of eEF1A-Flag/ HEK293 cells. 40S, 60S subunits, 80S monosome and polysome peaks are indicated. The bottom panel shows the distribution of exogenous eEF1A-Flag, endogenous eEF1A, and ribosome proteins in ribosome fractions.
(B) HEK293 cells with or without eEF1A-Flag transgene were treated with indicated chemicals for 1hr and lysed with or without DSP crosslinking. Equal amount of lysates were subjected to MNase digestion and generated monosomes were collected from sucrose gradient. anti-Flag IP was performed to selectively purify ribosome complexes associated with eEF1A-Flag . Immunoblotting was conducted using anti-Flag antibody to detect IP efficiency.
(C) Same samples in (B) were blotted with anti-RpL4 antibody to detect selectively purified ribosome complexes.

To test this model, specific maps of ribosomes associated with particular factors are required to determine the ribosome dynamics under proteotoxic stress. Recently, an approach called selective ribosome profiling (SeRP) has been developed to investigate the cotranslational interaction between a bacterial chaperone called trigger factor and ribosome complexes (Becker et al., 2013; Oh et al., 2011). In addition, the substrate binding property of Ssb, the homolog of Hsp70 in yeast, was also studied using a microarray based method (Willmund et al., 2013). Taking advantage of these two published tools, I decide to establish a SeRP method in mammalian system to investigate the cotranslational association of eEF1A and Hsc70 with ribosomes.

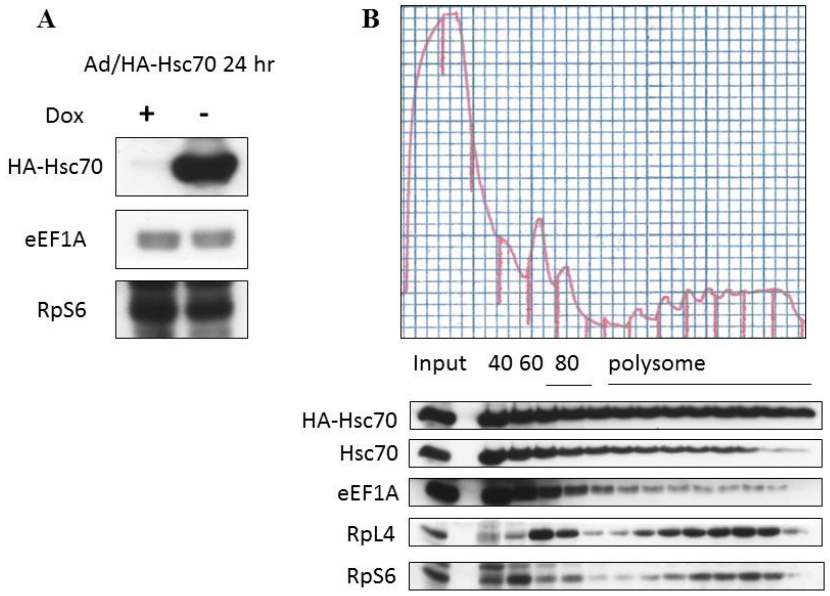


Figure AI-8. HA-Hsc70 is associated with polysome in Hela-tTA cells.
(A) HeLa-tTA cells were infected with adenoviruses expressing HA-Hsc70. The expression levels of both transgene and endogenous genes were determined by immunoblotting after 24 hr induction by Dox removal.
(B) Polysome profiles of HeLa-tTA cells expressing HA-Hsc70. 40S, 60S subunits, 80S monosome and polysome peaks are indicated. The bottom panel shows the distribution of exogenous HA-Hsc70, endogenous Hsc70, eEF1A, and ribosome proteins in ribosome fractions.

To adapt the SeRP protocol to mammalian system, I first established a HEK293 cell line stably expressing C-terminal Flag-tagged eEF1A proteins, so that eEF1A associated ribosome complexes can be specifically purified with high efficiency. The tagged exogenous eEF1A is incorporated into polysome fractions in sucrose gradient and distributes similarly as endogenous proteins in response to various treatment, indicating that the C-terminal tag doesn't interfere with its functions and behaviors in cells (Figure AI-7A and data not shown). Similar assays have also been conducted to make sure the functionality of N-terminal HA-tagged Hsc70 in HeLa- tTA cells (Figure AI-8).

Next, I performed a brief crosslinking *in vivo* with Dithiobis (succinimidylpropionate) (DSP) to stabilize the ribosome complexes with associated factors. Monosomes generated by micrococcal nuclease (MNase) digestion were collected using sucrose gradient approach. Ribosomes associated with particular factors were selectively purified using matrix beads coupled with antibodies against corresponding affinity tags. After extensively optimization of lysis, enzyme digestion, as well as IP and wash conditions, I successfully purified ribosomes-eEF1A complexes with high specificity and efficiency under different treatment (Figure AI-7B&C). Until the dissertation is submitted, I have been constructing the cDNA library from the ribosome protected mRNA fragments for high-throughput sequencing. Based on the expected sequencing data, the binding position of eEF1A and Hsc70 on elongating ribosomes can be inferred and used to differentiate two proposed models.

I conclude that the interaction between Hsc70 and eEF1A links the protein quality control system and the translational machinery, representing a novel co-translational mechanism that maintains the intracellular protein homeostasis.

AI.2 Materials and Methods (in addition to section 2.5)

In vitro translation

For *in vitro* translation of Fluc assay, TNT Quick Coupled Translation /Transcription system (Promega) based on rabbit reticulocyte lysate (RRL) was used. In brief, pcDNA3 plasmid encoding full-length or N-terminal fragments of Fluc mRNA was mixed with RRL supplemented with $[^{35}S]_{L}$-methionine. *In vitro*

translation was first performed at 30 ℃ for 20 min to allow the engagement of ribosomes on mRNAs and then another 40 min in the presence of indicated inhibitors.

Peptidyl tRNA purification and gel electrophoresis

Analysis of peptidyl-tRNA by gel electrophoresis was performed as described before (Janssen et al., 2012). Briefly, 10μl of *in vitro* translation reactions were stopped by chilling on ice and the addition of 35μl lysis buffer (1% SDS, 50mM Tris-acetate [pH 7.0], and 1mM EDTA). 0.5 ml of 2% CTABr solution and 0.5 ml of 0.5M sodium acetate (pH 5.0) solution were added to precipitate aminoacyl- and peptidyl-tRNAs. The solution was mixed, placed in an ice water both for 20 min, and then precipitates were thawed at 30 ℃ for 10 min. tRNA species were pelleted at 13,000 g for 30 min. Supernatants with full-length proteins were collected and precipitated with 10% TCA, washed with 100% acetone. The precipitates were air-dried, re-suspended with SDS sample buffer and resolved on SDS-PAGE. The gel was dried and viewed by phosphor imaging screen (HE healthcare).

On the other hand, tRNA pellets were washed with 1 ml 100% acetone and re-pelleted by centrifuging at 13,000 g for 10 min. Supernatants were carefully removed and pellets were allowed to air-dry. Pellets were re-suspended in 15 μl of AcE gel-loading buffer (8M urea, 10mM sodium acetate [pH 5.0], 1mM EDTA, 0.01% bromophenol blue). tRNA Samples were loaded on 6% acid-urea polyacrylamide gel (6% acrylamide, 8M urea, 100mM sodium acetate [pH 5.0], 1mM EDTA [pH 8.0], 0.16% Ammonium persulfate, 0.16% TEMED) and run at 100 V in a 4 ℃ room for 130 min with sodium acetate running buffer (100 mM sodium acetate [pH 5.0], 1mM EDTA in pre-chilled DEPC-treated water). Gel was fixed in 30% methanol and

shaken at RT for 30min to remove urea. After that, the gel was dried and viewed by phosphor imaging screen (HE healthcare).

Selective ribosome profiling for mammalian cells

Selective ribosome profiling was performed based on published protocol with modifications for cultured mammalian cells (Becker et al., 2013). Briefly, after indicated treatment, medium was aspirated out twice to remove free amino acid residuals as much as possible. Crosslinking was performed with 2.5 mM DSP in PBS (RT) at RT for 1 min and quenched with 50mM Tris (pH7.0) at RT for 1 min. After the aspiration of supernatants, cells were washed with ice-cold PBS (100μg/ml CHX) once and lysed with 3X SeRP lysis buffer (50mM HEPES [pH 7.5], 100mM KCl, 30mM $MgCl_2$, 100μg/ml CHX, 1X EDTA-free protease inhibitor, 0.4% Triton X-100 and 0.1% NP-40), giving lysates in about 10mM $MgCl_2$. After clearance by centrifugation at 13,000g for 10 min at 4℃, supernatants were collected. Protein and RNA concentration were determined by BCA assay (Promega) and Nanodrop. After the adjustment with sucrose gradient buffer (50mM HEPES [pH 7.5], 100mM KCl, 10mM $MgCl_2$, 100μg/ml CHX) to equal amount, lysates were added with 5mM $CaCl_2$ and Superase-In RNase inhibitor (2μl/mg RNA). Lysates with supplements were subjected to MNase (1,500 Gel Unit/mg RNA) digestion at 25℃ for 1hr with constant shaking in a Thermomixer (Eppendorf). The digestion was stopped by chilling on ice and the sequestration of Ca^{2+} by adding 6mM EGTA (pH 8.0). Digested samples were loaded on the top of 10%-45% sucrose gradient and separated by centrifugation in SW40 rotor at 32,000 rpm 4℃ for 4hr. Monosome fractions were collected and combined. An aliquot was saved as input. The rest samples were mixed with beads

coupled with corresponding tag antibody and the IP was performed at 4℃ for 1hr. Beads were washed three times for 5min with wash buffer (50mM HEPES [pH 7.5], 100mM KCl, 10mM MgCl2, 1mM EGTA, 100μg/ml CHX, 1X EDTA-free protease inhibitor, 0.4% Triton X-100 and 0.1% NP-40). RNA samples were extracted with Trizol (Invitrogen) according to manufacturer's instructions and converted to cDNA libraries for Illumina HiSeq sequencing. Protein samples were eluted with SDS sample buffer containing 50mM DTT and the reverse of crosslinking was performed at 37℃ for 30 min. Protein samples were heat denatured and analyzed by SDS-PAGE and immunoblotting.

REFERENCES

Becker, A.H., Oh, E., Weissman, J.S., Kramer, G., and Bukau, B. (2013). Selective ribosome profiling as a tool for studying the interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes. Nat Protoc *8*, 2212-2239.

Chuang, S.M., and Madura, K. (2005). Saccharomyces cerevisiae Ub-conjugating enzyme Ubc4 binds the proteasome in the presence of translationally damaged proteins. Genetics *171*, 1477-1484.

Deplazes, A., Mockli, N., Luke, B., Auerbach, D., and Peter, M. (2009). Yeast Uri1p promotes translation initiation and may provide a link to cotranslational quality control. EMBO J *28*, 1429-1441.

Dever, T.E., and Green, R. (2012). The elongation, termination, and recycling phases of translation in eukaryotes. Cold Spring Harb Perspect Biol *4*, a013706.

Friend, K., Campbell, Z.T., Cooke, A., Kroll-Conner, P., Wickens, M.P., and Kimble, J. (2012). A conserved PUF-Ago-eEF1A complex attenuates translation elongation. Nat Struct Mol Biol *19*, 176-183.

Gross, S.R., and Kinzy, T.G. (2007). Improper organization of the actin cytoskeleton affects protein synthesis at initiation. Mol Cell Biol *27*, 1974-1989.

Hussey, G.S., Chaudhury, A., Dawson, A.E., Lindner, D.J., Knudsen, C.R., Wilce, M.C., Merrick, W.C., and Howe, P.H. (2011). Identification of an mRNP complex regulating tumorigenesis at the translational elongation step. Mol Cell *41*, 419-431.

Janssen, B.D., Diner, E.J., and Hayes, C.S. (2012). Analysis of aminoacyl- and peptidyl-tRNAs by gel electrophoresis. Methods Mol Biol *905*, 291-309.

Leu, J.I., Pimkina, J., Frank, A., Murphy, M.E., and George, D.L. (2009). A small molecule inhibitor of inducible heat shock protein 70. Mol Cell *36*, 15-27.

Mateyak, M.K., and Kinzy, T.G. (2010). eEF1A: thinking outside the ribosome. J Biol Chem *285*, 21209-21213.

Meriin, A.B., Zaarur, N., and Sherman, M.Y. (2012). Association of translation factor eEF1A with defective ribosomal products generates a signal for aggresome formation. J Cell Sci *125*, 2665-2674.

Mitsui, K., Nakayama, H., Akagi, T., Nekooki, M., Ohtawa, K., Takio, K., Hashikawa, T., and Nukina, N. (2002). Purification of polyglutamine aggregates and identification of elongation factor-1alpha and heat shock protein 84 as aggregate-interacting proteins. J Neurosci *22*, 9267-9277.

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.

Shamovsky, I., Ivannikov, M., Kandel, E.S., Gershon, D., and Nudler, E. (2006). RNA-mediated response to heat shock in mammalian cells. Nature *440*, 556-560.

Sherman, M.Y., and Qian, S.B. (2013). Less is more: improving proteostasis by translation slow down. Trends Biochem Sci *38*, 585-591.

Van Goietsenoven, G., Hutton, J., Becker, J.P., Lallemand, B., Robert, F., Lefranc, F., Pirker, C., Vandenbussche, G., Van Antwerpen, P., Evidente, A., *et al.* (2010). Targeting of eEF1A with Amaryllidaceae isocarbostyrils as a strategy to combat melanomas. FASEB J *24*, 4575-4584.

Willmund, F., del Alamo, M., Pechmann, S., Chen, T., Albanese, V., Dammer, E.B., Peng, J., and Frydman, J. (2013). The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis. Cell *152*, 196-209.

APPENDIX II

Profiling of Alternative Initiation in Response to Starvation

This work was originally the last section of the manuscript based on Chapter 3, but it was re-organized as the first section of a new manuscript in submission as Gao X, Wan J, Liu B, and Qian SB. **Quantitative Profiling of Initiating Ribosomes in Mammalian Cells and Liver Tissues Reveals Programmatic Translational Response to Starvation**.

AII.1 Results

For transcripts with multiple TIS positions, the relative LTM reads density reflects the initiation potency under a given growth condition. Thus, changes in TIS selection could be evaluated by measuring the ratio of different TIS codons on the same mRNA. Acute amino acid deprivation triggers an adaptive stress response regulating gene expression at multiple levels, including translation initiation (Hinnebusch, 2005; Kilberg et al., 2005). We examined the changes in TIS selection during starvation by subjecting HEK293 cells to 20 min of incubation in phosphate buffered saline (PBS) followed by GTI-seq. Compared to nutrient-rich growth conditions (DMEM), acute amino acid starvation reduced polysome formation with a concomitant increase of monosome peak (Figure AII-1). In contrast to the previous report using *S. cerevisiae* (Ingolia et al., 2009), we did not find the drastic re-distribution of CHX reads between the UTR and CDS regions after starvation (Figure

AII-2A). This could be due to the fact that the starvation condition we used was not able to trigger the amino acid response (AAR) pathway, at least in HEK293 cells. Consistent with this notion, the eIF2α phosphorylation in HEK293 cells was unaltered after 20 min of PBS incubation (Figure AII-2B). Not surprisingly, we did not observe translational changes of uORFs in the transcript of *ATF4* (Figure AII-2C).

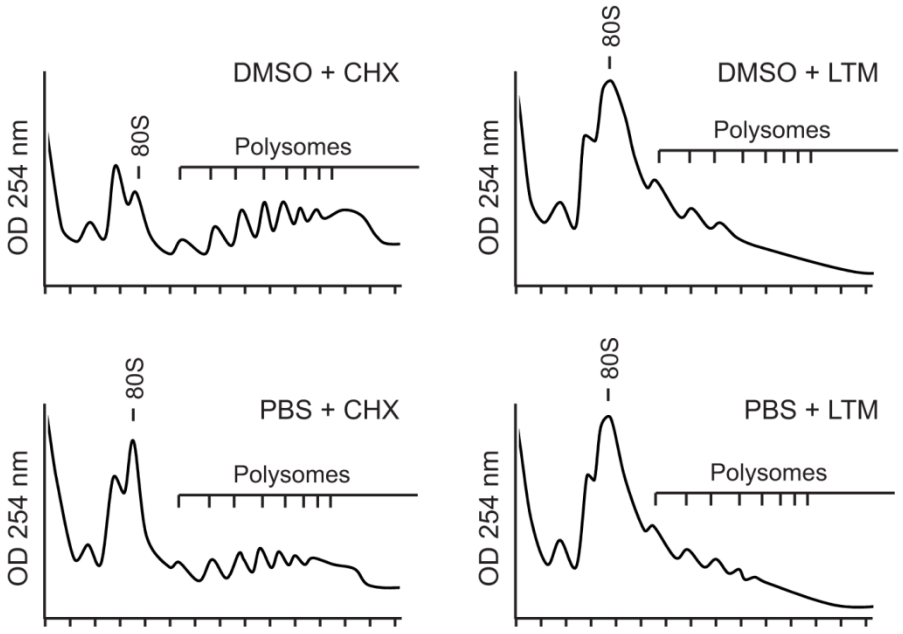


Figure AII-1. Polysome profile analysis in cells with or without amino acid starvation. HEK293 cells were either incubated in nutrition-rich medium DMEM or PBS for 20 min of starvation before the addition of ribosome E-site inhibitors 100 μM CHX or 50 μM LTM for 30 min followed by sucrose gradient sedimentation. Both 80S monosome and polysome peaks are indicated.

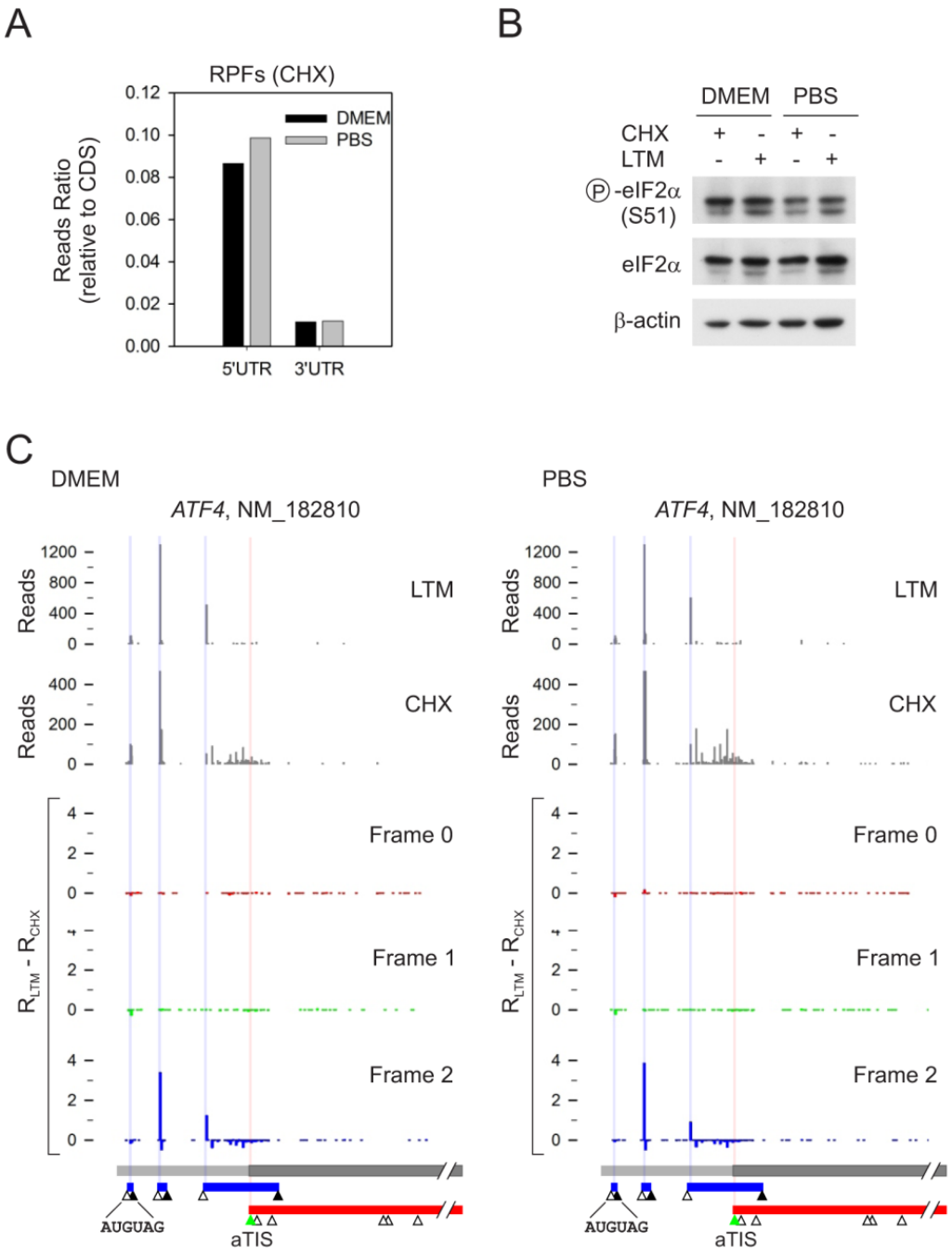


Figure AII-2. PBS incubation of HEK293 cells did not trigger amino acid response pathway.

(A) CHX-associated RPF reads were obtained from HEK293 cells with or without amino acid starvation. Plotted is the ratio of RPF reads mapped to either 5'UTR or 3'UTR relative to the CDS.

(B) eIF2α phosphorylation status in HEK293 cells with or without PBS incubation. The same samples as in (A) were immunoblotted using antibodies as indicated.

(C) Translational regulation of *ATF4* in cells with or without PBS incubation.

To systemically determine whether TIS selection was altered over the entire transcriptome under this starvation condition, we selected genes with both uTIS and aTIS initiation on their transcripts. We analyzed the relative efficiency of these two initiators in cells with or without starvation (Figure AII-3A). We observed a group of transcripts exhibiting reduced aTIS initiation relative to the uTIS (Figure AII-3A, blue dots). Notably, the majority of these gene products were involved in the process of transcription (Figure AII-3B), such as the upstream binding transcription factor *UBTF* (Figure AII-3C). Under nutrient-rich growth conditions, *UBTF* translation was dominated via the aTIS initiation. New uTIS peaks appeared on *UBTF* under amino acid starvation, while the aTIS initiation subsequently reduced. Surprisingly, a certain number of genes showed a relative increase in aTIS initiation on their transcripts, a result of either repressed uORF translation or enhanced main ORF translation (Figure 7A, red dots). Intriguingly, many of these genes were involved in the protein catabolic process, in particular the ubiquitin-proteasome system (Figure AII-3B). For instance, a gene encoding the proteasome α subunit 7 (*PSMA7*) had a CUG-initiated uORF in addition to the main ORF (Figure AII-3D). Amino acid starvation led to a reduced uTIS selection but an increased initiation at the aTIS. The favored production of these gene products under acute starvation condition apparently helps recycling of intracellular amino acids by promoting protein breakdown. This regulatory mode of TIS selection indicates that the scope of selective translation under stress conditions is much broader than anticipated.

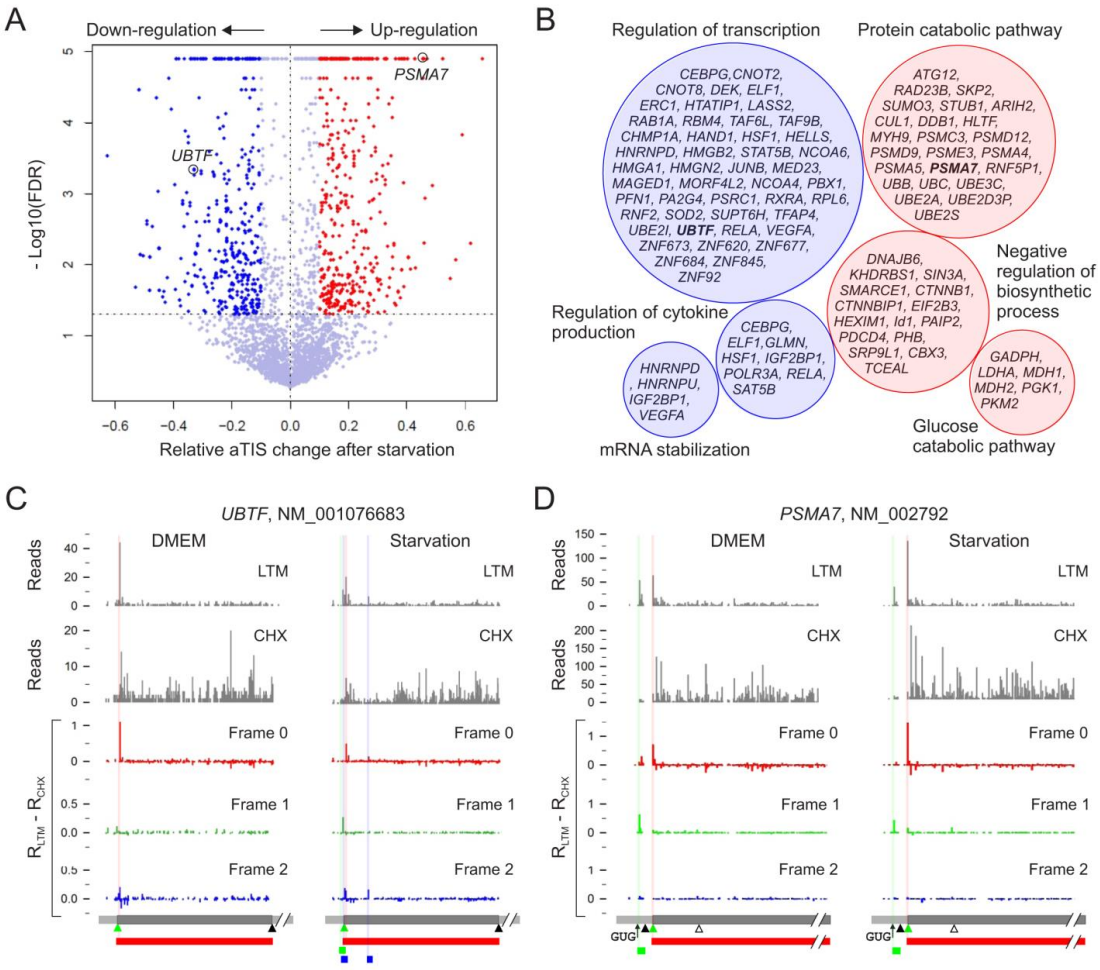


Figure AII-3. Changes of TIS selection under amino acid starvation.
(A) Volcano plot showing aTIS fraction changes after amino acid starvation. GTI-seq was applied to HEK293 cells with or without amino acid starvation. For transcripts with multiple TIS codons, the changes of aTIS initiation over the total TIS positions on each transcript are plotted. Blue dots, decreased aTIS initiation under starvation; red dots: increased aTIS initiation under starvation.
(B) Functional characterization of genes with changed TIS selection under starvation. Blue circles contain genes with decreased aTIS under starvation and red circles for increased aTIS under starvation.
(C) An example of increased uTIS initiation during starvation. Multiple TIS codons on the *UBTF* transcript were shown in HEK293 cells with (right panel) or without (left panel) starvation.
(D) An example of increased aTIS initiation during starvation. Multiple TIS codons on the *PSMA7* transcript are shown in HEK293 cells with (right panel) or without (left panel) starvation.

AII.2 Materials & Methods (in addition to section 3.5)

Starvation Treatment

For the starvation experiment, HEK293 cells were incubated in phosphate buffered saline (PBS) at 37 ℃ for 20 min. After the starvation, 100 μM cycloheximide or 50 μM lactimidomycin were added immediately to the PBS and cells were kept in PBS at 37 ℃ for another 30 min. The control experiment was performed in parallel in DMEM with 10% FBS.

Identification of Genes with Altered TIS Selection under Starvation

The value of aTIS fraction [LTM reads on aTIS / (LTM reads on aTIS + Sum of LTM reads on uTIS)] was computed for each mRNA under each condition (DMEM and PBS). The union of the uTIS codons identified in both conditions was used as the set of uTIS position. N-terminal extension type of uORFs was not included. mRNAs with neither aTIS nor uTIS peak in each condition were excluded. Technical replicates were pooled for each condition. aTIS fraction difference (aTIS fraction of PBS – aTIS fraction of DMEM) was used for statistical testing, which was performed by shuffling the LTM reads at each aTIS or uTIS position. First, the reads from the two conditions at each position were summed and redistributed according to the probability proportional to the total mRNA-mapped reads from the two conditions. Based on the redistributed read counts, aTIS fraction difference was computed. This process was repeated 1 million times. The proportion that the generated aTIS fraction difference was larger (or smaller if the observed aTIS fraction difference is negative) than or equal to the observed value was used as a p-value. To avoid zero p-values, the minimum possible p-value was set to 10^{-6}. To adjust for multiple testing, false

discovery rate was computed based on the Benjamini-Hochberg method using the p.adjust function in R. Up- or down-regulated mRNAs were identified at FDR cut-off 0.05. The Gene Ontology analysis was done at the gene level, rather than the mRNA level. Any gene with both up- and down-regulated mRNA isoforms was removed before the analysis.

REFERENCES

Hinnebusch, A.G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. Annu Rev Microbiol *59*, 407-450.

Kilberg, M.S., Pan, Y.X., Chen, H., and Leung-Pineda, V. (2005). Nutritional control of gene expression: how mammalian cells respond to amino acid limitation. Annu Rev Nutr *25*, 59-85.

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.