DYNAMIC RIBOSOME SCANNING MEDIATED BY RNA HELICASES IN EUKARYOTIC TRANSLATION 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

Yifei Gu

August 2020

DYNAMIC RIBOSOME SCANNING MEDIATED BY RNA HELICASES IN EUKARYOTIC TRANSLATION INITIATION

Yifei Gu, Ph. D

Cornell University 2020

Translation of messenger RNA (mRNA) into protein is a critical process to cell survival and function. mRNA translation is a tightly regulated process, most of which occurs during the initiation stage. Eukaryotic translation initiation primarily relies on a cap-dependent scanning mechanism for start codon selection, facilitated by over a dozen of eukaryotic initiation factors (eIFs) and several RNA helicases.

The current model of eukaryotic translation initiation commences with the attachment of pre-initiation complex (PIC) to the 5' end, followed by 5'→3' uni-directional scanning. Despite the proposed scanning process in start codon selection, the dynamic nature of the scanning PIC is poorly understood. Investigation of start codon recognition using synthetic mRNA reporters with ultra-short 5'UTR or closely spaced AUG triplets suggests that the PIC follows a non-linear scanning mode. I demonstrate the backward excursion of the scanning ribosome using an internal ribosome entry site that enables translation of both upstream and downstream open reading frames. Unexpectedly, the ATP-dependent DEAD-box RNA helicase eIF4A is actively involved in bi-directional oscillations during scanning. Using quantitative profiling of initiating ribosomes (QTI-seq), I show that upregulation of eIF4A ATPase activity globally regulates alternative initiation by stimulating translation from the upstream initiation sites in the 5' untranslated region (5'UTR). Moreover, this study revealed that

eIF4A1 mediated bi-directional ribosome scanning controls the stringency of start codon selection.

During the scanning, the PIC has to migrate along the structured region in 5'UTR, with the help of various RNA helicases. The DEAD-box helicase eIF4A1 was reported to remodel the 5' proximal structures with low stabilities. Another RNA helicase, DHX29, was known to promote unwinding of more stable secondary structures. However, their potential roles in translation of mRNAs with secondary structures in different regions remain elusive. Herein, I reported that the position of stem-loops (SLs) in the 5' UTR did not affect the dependence of mRNA translation on either eIF4A1 or DHX29. Surprisingly, the translation of mRNAs with SLs downstream of the start codon were impaired by DHX29 depletion. DHX29 was found in both 40S and 80S ribosome fractions after polysome fractionation, supporting the presence of DHX29 in 80S. Collectively, these data suggest a novel function for DHX29 in unwinding stem loops during the early stages of elongation.

In summary, this study reported the bi-directional motion of PIC during scanning, which potentially could address the questions regarding the mRNA attachment and the start codon stringency. The novel functions of eIF4A1 in the scanning directionality via ribosome conformation and DHX29 in unwinding mRNA secondary structures during early elongation were demonstrated. This body of work might contribute to better appreciation of the wide-scope functions RNA helicases may have, as well as the comprehensive understanding of the PIC scanning during initiation.

BIOGRAPHICAL SKETCH

Yifei Gu was born December 30, 1992 in Shanghai, China. She earned her Bachelor of Science in Pharmaceutical Engineering from East China University of Science and Technology in 2015. At East China University of Science and Technology, Yifei joined the Undergraduate Innovation and Entrepreneurship Training Program and conducted undergraduate research in the Lab of Dr. Yufang Xu studying the tyrosine kinase inhibitors of epidermal growth factor receptor (EGFR-TKIs). She also interned in the molecular biology labs of Shanghai Fudan-Zhangjiang Bio-Pharmaceutical Co., Ltd. and Novoprotein Scientific Inc. from 2014 to 2015.

In 2015, Yifei was accepted into the Molecular Nutrition doctoral program at Cornell University where she worked with Dr. Shu-Bing Qian on the topic of this dissertation. She presented her research in Translational Control at Cold Spring Harbor in 2018 and submitted her research paper in high-profile journals. Yifei served as a teaching assistant for the Department of Nutritional Sciences from 2015 to 2020. She also supervised undergraduate students' research projects from 2017 to 2020. After earning the Ph.D., Yifei plans to continue with a period of postdoctoral training and then look for a research and development job in the biotechnology and pharmaceutical industry.

ACKNOWLEDGEMENTS

During the five-year journey, I was incredibly lucky to meet many amazing people who supported me while I was working on this dissertation. I would like to thank them for their contributions. In particular, thanks go to my principal investigator, Dr. Shu-Bing Qian, for introducing me into the Qian lab family and mentoring me as his Ph.D student. With deep knowledge and a sharp sense of frontiers in our field, He always provides keen insight and extensive feedback on my research. Only with his expertise was it possible to develop this dissertation. Every time I'm frustrated with the negative results or struggled with technical issues, Shu-Bing always guided me with new ideas and re-inspired my interest in the project from different perspectives. His optimism and enthusiasm in research pushed me to exceed my own expectations.

I would like to express my appreciation to my committee members. Dr. Hojoong Kwak provided insightful suggestions on the model I proposed in this dissertation and had pointed out the potential pitfalls I encountered later. Dr. Jeff Pleiss offered astute ideas on research design, which served as very strong supplementary evidence to validate the conclusions in this dissertation. Dr. Anna Thalacker-mercer not only guided me through this dissertation, but also shaped my critical thinking abilities in bigger pictures. As the instructor of the course I TAed for, she also stimulated my enthusiasm in teaching and taught me a lot of skills in presentation.

Many thanks to all the past and present members of Qian lab. Previous lab members, Joy Zhou, Xiao-Min Liu, and XingQian Zhang, are outstanding postdocs with great expertise on research. With exceedingly patience, they taught me basic biochemistry and cell culture methods, such as in vitro translation, CRISP and protein

purification. To the present lab members, I'd like to thank Leiming Dong for not only teaching me about conducting the library construction but also for renting his nice apartment to me. Longfei Jia guided me to apply Flow cytometry and the newly developed uORF reporter assays. Quanquan Ji helped me with the setting of the qPCR machine. Yuanhui analyzed the sequence data for me. Robert Swanda assisted in the formal and linguistic finalization of this dissertation. Kevin Lin worked hard with me on Chapter 2. I would like to extend my warmest thanks to Erica Shu who helped me with everything I encountered both in and out of the lab.

Outside of the lab, I am grateful to all of my dear families and friends for their constant support and inspiration. My parents, Guoyi Gu and Min Li, nourished my interest in science since my childhood. Although they were worried about their little girl studying abroad by traveling to the other side of earth, they still support every decision I have made on the way. Encouragement from my beloved uncle and aunt further convinced myself to take risks, leave my comfort zone and study at Cornell University. Special thanks to Aunty Alisha and Uncle Joy for inviting me to their wonderful Thanksgiving party every year. A huge thank to my fiancé, Zhexun Sun, for always being by my side. Lastly, I would like to express my sincere gratitude to my forever missed grandpas, whose caring and love on me have ceased in the past two years, yet their spirit will be carried on every step I move forward.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

3'UTR	3' untranslated region
30S	small subunit of prokaryotic 70S ribosome
40S	small subunit of eukaryotic 80S ribosome
4E-BP	4E-binding protein
5' TOP	5' terminal oligopyrimidine
5'UTR	5' untranslated region
50S	large subunit of prokaryotic 70S ribosome
60S	large subunit of eukaryotic 80S ribosome
70S	prokaryotic 70S ribosome
80S	eukaryotic 80S ribosome
A-site	aminoacyl-tRNA site
ABCE1	ATP-binding cassette sub-family E member 1
ADP	adenosine diphosphate
AFM	atomic force microscopy
ALKBH5	AlkB homolog 5
APEX	Ascorbate Peroxidase
ASO	antisense oligonucleotide
ATF4	activating transcription factor 4
aTIS	annotated start sites
ATP	adenosine triphosphate
BioID	Biotin Identification
BSA	bovine serum albumin
CDS	coding region
CHX	cycloheximide
CID	chemical inducers of dimerization
circRNA	circular RNA
Co-IP	co-immunoprecipitation
Cryo-EM	cryogenic electron microscopy
CTD	C-terminal domain
DHX29	DExH-box helicase 29
DMEM	Dulbecco's Modified Eagle's Medium
dORF	downstream open reading frame
dTIS	downstream TIS
E-site	exit site
eEF	eukaryotic elongation factor
eEF1A	eukaryotic translation elongation factor 1A
eEF2	elongation factor 2
EF-Tu	prokaryotic elongation factor thermos unstable
eIF	eukaryotic initiation factor
eIF1	eukaryotic translation initiation factor 1
eIF1A	eukaryotic translation initiation factor 1A
eIF2	eukaryotic translation initiation factor 2
eIF2B	eukaryotic translation initiation factor 2B

eIF2α	α subunit of eukaryotic translation initiation factor 2
eIF2β	β subunit of eukaryotic translation initiation factor 2
eIF2γ	γ subunit of eukaryotic translation initiation factor 2
eIF3	eukaryotic translation initiation factor 3
eIF3a	a subunit of eukaryotic translation initiation factor 3
eIF3b	b subunit of eukaryotic translation initiation factor 3
eIF3c	c subunit of eukaryotic translation initiation factor 3
eIF3d	d subunit of eukaryotic translation initiation factor 3
eIF3e	e subunit of eukaryotic translation initiation factor 3
eIF3f	f subunit of eukaryotic translation initiation factor 3
eIF3g	g subunit of eukaryotic translation initiation factor 3
eIF3h	h subunit of eukaryotic translation initiation factor 3
eIF3i	i subunit of eukaryotic translation initiation factor 3
eIF3j	j subunit of eukaryotic translation initiation factor 3
eIF3k	k subunit of eukaryotic translation initiation factor 3
eIF3l	l subunit of eukaryotic translation initiation factor 3
eIF3m	m subunit of eukaryotic translation initiation factor 3
eIF4A	eukaryotic translation initiation factor 4A
eIF4A1	eukaryotic translation initiation factor 4A isoform 1
eIF4A2	eukaryotic translation initiation factor 4A isoform 2
eIF4A3	eukaryotic translation initiation factor 4A isoform 3
$eIF4A_c$	eIF4A1 in the eIF4F complex
$eIF4A_f$	free eIF4A1
eIF4B	eukaryotic translation initiation factor 4B
eIF4E	eukaryotic translation initiation factor 4E
eIF4F	eukaryotic translation initiation factor 4F
eIF4G	eukaryotic translation initiation factor 4G
eIF4H	eukaryotic translation initiation factor 4H
eIF5	eukaryotic translation initiation factor 5
eIF5A	eukaryotic translation initiation factor 5A
eIF5B	eukaryotic translation initiation factor 5B
EM	electron microscopy
eRF	eukaryotic release factor
eRF1	eukaryotic release factor 1
eRF3	eukaryotic release factor 3
FBS	fetal bovine serum
Fluc	firefly luciferase
$fMet\text{-}tRNA_i^{fMet}$	initiator N-Formylmethionine tRNA in prokaryotes
FTO	obesity-associated protein
GCN4	general control nondepressible 4
GDP	guanosine diphosphate
GST	glutathione
GTP	guanosine triphosphate
H4	histone 4

HalV	Halastavi arva virus
Hippu	Hippuristanol
IRES	internal ribosome entry sites
IRP1	iron-regulatory protein 1
IRP2	iron-regulatory protein 2
ITAF	IRES trans-acting factors
LARP1	RNA-binding protein La-related protein 1
m^6A	N^6-methyladenosine
m^7G	7-Methylguanosine
MARK	mitogen-activated protein kinase
Met-$tRNA_i^{Met}$	initiator methionine tRNA in eukaryotes
METTL14	methyltransferase Like 14 protein
METTL3	methyltransferase Like 3 protein
MFC	multi-factor complex
mRNA	messenger RNA
mTOR	The mammalian target of rapamycin
mTORC1	mTOR complex 1
ncRNA	non-coding RNA
NGS	next generation sequencing
NMD	nonsense-mediated mRNA decay
NTD	N-terminal domain
ORF	open reading frame
P-site	peptidyl-tRNA site
PABP	poly(A) binding protein
PABPC1	poly(A) binding protein cytoplasmic 1
PAR-CLIP	Photoactivatable Ribonucleoside-enhanced crosslinking and immunoprecipitation
PARS	parallel analysis of RNA structures
PatA	Pateamine A
PI3K	phosphoinositide 3-kinase
PIC	pre-initiation complex
PIC-seq	PIC-associated footprints
post-TC	post-termination complex
pre-TC	pre-termination complex
PRF	programed ribosomal frameshifting
QTI-seq	quantitative profiling of initiating ribosomes
RBP	RNA-binding protein
Ribo-seq	ribosome profiling
RNP	ribonucleoprotein
RocA	Rocaglamide A
RP	ribosomal protein
RPS	small ribosomal protein
RPS2	ribosomal protein S2
RPS3	ribosomal protein S3

RRL	rabbit reticulum lysates
rRNA	ribosomal RNA
SD	Shine-Dalgarno
SG	stress granule
SHAPE	selective 2′-hydroxyl acylation analyzed by primer extension
SL	stem-loop
smFISH	single-molecule-resolution fluorescent in situ hybridization
TC	ternary complex
TCP-seq	translation complex profiling
TE	translation efficiency
TIS	translation initiation sites
TISU	translation initiator of short 5′UTR
tRNA	transfer RNA
TSS	transcription start site
TYMV	turnip yellow mosaic virus
uAUG	upstream AUG
uORF	upstream open reading frame
uTIS	upstream start sites
WTAP	wilms' tumor 1-associating protein
YTHDF2	YTH N^6-Methyladenosine RNA Binding Protein 2

PREFACE

The objectives of this dissertation mainly focus on the dynamic ribosome scanning mediated by RNA helicases in eukaryotic translation initiation. This work is composed of two major research projects, emphasizing on the directionality of ribosome scanning and the positioning effects of secondary structures on mRNA translational respectively. Chapter 1 reviews the current knowledge of eukaryotic translation initiation, with emphasis on the ribosome attachment and ribosome scanning. Chapter 2 reports the bi-directional scanning of the ribosome in the canonical cap-dependent translation and elucidates the mechanism underlying eIF4A-mediated scanning directionality. Chapter 3 is a systematic study on the translational effects of secondary structures in different regions of mRNA and identifies the potential role of DHX29 in unwinding secondary structures at early stage of elongation. Chapter 4 describes the major findings of this dissertation and their contributions in the field of mRNA translation, combining several prospective directions for future research.

CHAPTER 1

General Introduction and Literature Review

1.1 mRNA Translation

In the early nineteenth century, studying life science at the molecular level became feasible with the innovation of techniques in chemistry and physics. In the 1930s, a new branch of science - molecular biology was developed and enabled the researchers to appreciate and understand more details of certain life processes using delicate modern techniques[1]. In the 1940s, it was realized that DNA, not protein, was the hereditary material that carries the genetic information[2,3]. The discovery of messenger RNA (mRNA) and transfer RNA (tRNA), together with the deciphering of genetic code clarified how the genetic information in DNA is ultimately expressed in the amino acid sequence of a variety of functional proteins. These findings gave rise to the central dogma of molecular biology, which describes how genetic information is passed from DNA to RNA then to protein in the living cells. The efficient gene expression is mainly achieved by the two-step process: gene transcription (DNA->RNA) and mRNA translation (RNA->Protein). The synthesis of polypeptide, like gene transcription, takes place in three steps: initiation, elongation and termination, with specific factors involved in each step. Given the discovery of recycling factors, ribosome recycling is considered as the fourth step of mRNA translation.

In bacteria, mRNA translation is tightly coupled with transcription. To date, the most well-studied mechanism of translation initiation of prokaryotic mRNAs is Shine-Dalgarno (SD)-mediated initiation. mRNAs containing the SD sequence in 5' untranslated regions (5' UTR) recruit the small ribosomal subunit (30S in bacteria) through base-pairing between the SD sequence and 16S ribosomal RNA (rRNA). It directs the initiator N-Formylmethionine tRNA (fMet-$tRNA_i^{fMet}$) to recognize the start codon of the mRNA in the 30S.

The eukaryotic translation is a more complicated process that occurs in four phases: initiation, elongation, termination, and recycling. The initiation phase of eukaryotic mRNA translation, which involves more than a dozen eukaryotic initiation factors (eIFs)[4], is the rate-limiting step of the translational process (Figure 1-1). The translation on most eukaryotic mRNAs relies on a 7-Methylguanosine (m^7G) cap to load the 40S small ribosomal subunit to the mRNA, followed by a scanning process to position initiator methionine tRNA (Met-$tRNA_i^{Met}$) within the ribosome at the start codon. This canonical cap-dependent scanning pathway commences with the assembly of the 43S pre-initiation complex (PIC), which happens in two steps. The eukaryotic initiation factors (eIFs) eIF1, -1A, -3 and -5 bind to the 40S subunit, followed by binding of initiator Met-$tRNA_i^{Met}$ to the small subunit in a ternary complex (TC) with eukaryotic translation initiation factor 2 (eIF2) and GTP.[5] The resulting 43S PIC attaches to the 5' proximal region of m^7G capped mRNAs in a manner facilitated by the eIF4F complex, which consists of a cap-binding protein eIF4E, a

scaffolding protein eIF4G, and RNA helicase eIF4A[6]. After mRNA attachment, 43S PIC scans the 5'UTR of the mRNA until the appropriate start codon appears in the P-site of 43S PIC. Recognition of the start codon evokes the hydrolysis of the GTP-eIF2 to GDP-eIF2, which dissociates from the ribosome with other scanning components under the catalyzation of eIF5B. Binding of eIF5B-GTP to Met-tRNA$_i^{Met}$ promotes the joining of large ribosomal subunit (60S) to the PIC, and the formation of 80S ribosome is accomplished after all the remaining eukaryotic initiation factors dissociate from the ribosome.

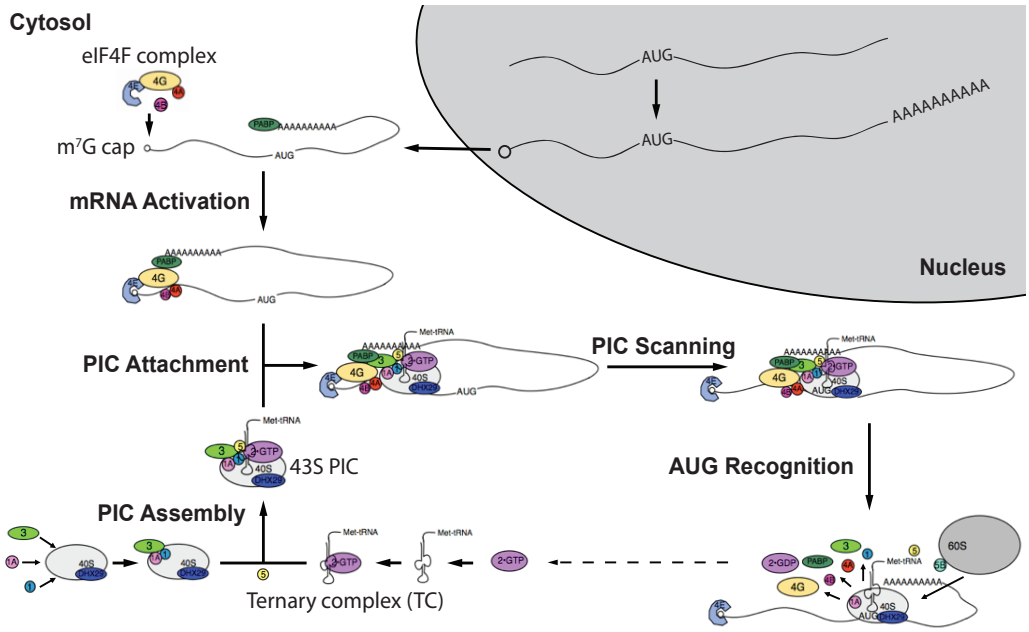

Figure 1-1 Eukaryotic translation initiation via cap-dependent scanning of mRNA.
The eukaryotic translation initiation occurs in the cytosol and starts with the activation of mRNAs by eIF4F complex, through interactions between m^{7}G cap and the cap-binding protein eIF4E. Meanwhile, the pre-initiation complex (PIC) is formed in two steps where eIFs 1, 1A, and 3 bind to the 40S subunit first, followed by the recruitment of TC and eIF5. Once the PIC attaches to the activated mRNA, it scans along the 5'UTR. Recognition of the start codon triggers the release of most eIFs and conversion GTP-eIF2 to GDP-eIF2 to mediate joining of the 60S subunit.

During translation elongation, the ribosome actively moves along the mRNA, using tRNAs to synthesize polypeptides[7]. Elongation involves three

steps: first, the binding of an aminoacyl-tRNA whose anticodon base-pairs with the mRNA codon in the ribosome A-site; second, the formation of a peptide bond by transfer of the amino acid or peptide attached to the tRNA in the P-site to the aminoacyl-tRNA in the A-site; third, the translocation of the newly formed peptidyl-tRNA from the A-site to the P-site, together with the mRNA. Translation elongation is facilitated by a number of eukaryotic elongation factors (eEFs). The eukaryotic translation elongation factor 1A (eEF1A) is activated upon binding to GTP and forms a ternary complex with aminoacyl-tRNA. The aminoacyl-tRNA bound with eEF1A-GTP enters the A-site and base-pairs with mRNA, followed by GTP hydrolysis and the subsequent release of eEF1A-GDP[8]. During the catalysis of peptide bond formation, the eukaryotic initiation factor eIF5A (the ortholog of bacterial elongation factor P) binds to the ribosome in the E-site and interacts with the acceptor arm of the peptidyl-tRNA to induce a favorable positioning of the substrates[9,10,11,12]. Following the peptide formation, the elongation factor 2 (eEF2)-GTP enters A-site and facilitates the translocation of the tRNAs to the E- and P-site by GTP hydrolysis. The release of eEF2-GTP from A-site and the deacylated tRNA from the ribosome allow for the next cycle of elongation[13,14,15].

Termination is triggered when the stop codon reaches the A-site of the ribosome and is recognized by eukaryotic release factors (eRFs) eRF1 and eRF3 (-GTP), forming the pre-termination complex (pre-TC). After GTP hydrolysis by eRF3, the middle domain of eRF1 accommodates the P-site and

induces the release of the nascent polypeptide from peptidyl-tRNA[16]. The resulting post-termination complex (post-TC) is bound by ABCE1, a protein which mediates splitting of post-TC and yields 60S subunit and eRFs. Followed by the dissociation of 40S subunit from mRNA and tRNA, post-TC enables ribosomes and mRNAs to participate in multiple rounds of translation[7,17].

The majority of the cellular transcripts are translated in cytosol which is governed by the cytosolic ribosomes. Ribosome is a complex supramolecular that catalyzes the peptide bonds formation in the order following the mRNA template. Eukaryotic ribosome consists of two unequal subunits with sedimentation coefficients of 40S and 60S and a combined sedimentation coefficient of 80S, while bacterial ribosome consists of 30S and 50S with combined sedimentation coefficient of 70S. Each subunit contains complex ribosomal proteins and at least one rRNAs. The ribosome contains three tRNA binding sites: A-site (Aminoacyl), P-site (Peptidyl), and E-site (Exit). The A-site accommodates the incoming aminoacyl-tRNA to accept the polypeptide during elongation and the termination release factors during termination; The P-site binds to the first incoming initiator Met-tRNA$_i^{Met}$ during initiation and the peptidyl-tRNA during elongation; The E-site binds to the free tRNA after its polypeptide is transferred to a second tRNA. mRNAs traverse ribosomes through the mRNA-binding cleft located on the interface side of 40S subunits. Importantly, the mRNA binding channel includes the decoding site where anticodons of tRNAs recognize the complementary codons of the mRNA. The length of

the mRNA-binding cleft in 40S subunits is about 28-30nt, while the distance of the decoding site from the entry site and the exit site is about 17nt and 12nt respectively. Using RNase protection approach, researchers were able to obtain the footprints of active ribosomes (around 30nt) and monitor the position of decoding sites on the mRNAs.

In a post-genomic era, translational control plays a pivotal role in defining the proteome, maintaining metabolic homeostasis, cell growth and proliferation. Furthermore, mRNA translation is tightly regulated to control gene expression in response to various endogenous or exogenous signals such as nutrient supply, hormones or stress. Therefore, aberrations in its regulation contribute to a number of disease states. In the past decades, the development of the next generation sequencing (NGS) gave rise to a variety of high-throughput sequencing approaches to understand the translation of individual transcripts from different perspectives at the transcriptome and genome wide level.

1.2 Translation Initiation

Translation initiation entails identification of start codon in mRNA by the PIC. Translation initiation differs between eukaryotes and prokaryotes. In bacteria, the lack of a membrane-enclosed nucleus enables the transcription-translation coupling. Ribosome initiates translation at the 5' end of the mRNA before the completion of transcription. Recent studies have shown that the transcriptional rate can be determined by the translational rate via direct

interactions between RNA polymerases and ribosomes[18,19,20,21]. Yet the unveiling of the coupling mechanism remains challenging. Imaging studies showed that RNA polymerases and ribosomes are mainly located in different subcellular regions in bacterial cells, i.e., RNA polymerases localize to the nucleoid area, while ribosomes localize outside nucleoids[22,23]. Thus, how transcription spatially coupled with translation in bacterial cells is more complicated than previously thought. Very little is known about initiation of mRNAs that are engaged in polysomes[24].

Translation initiation of mRNAs containing the Shine-Dalgarno (SD) sequence is extensively studied among different types of mRNA in prokaryotes (Figure 1-2). This consensus sequence is an initiation signal of four to nine purine residues, 8 to 13 nucleotides to the 5' side of the initiation codon. During SD-led initiation, the small subunit (30S in bacteria) is recruited to the ribosome binding site, which spans nucleotides – 20 to +15 around the translation start codon, through base-pairing between the SD sequence and the anti-SD sequence in 16S rRNA. This mRNA-rRNA interaction positions the AUG of the mRNA in the precise position on the 30S subunit where fMet-tRNA$_i^{fMet}$ is bound.

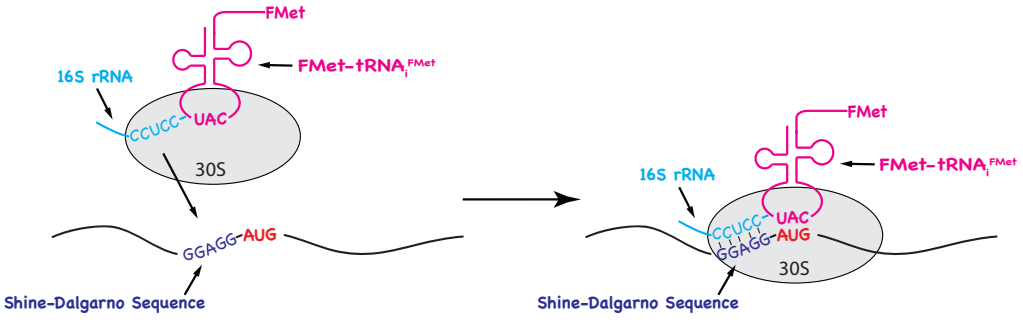

Figure 1-2 Shine-Dalgarno mediated translation initiation in bacteria.
The 3′ tail of 16S rRNA (light blue) in the small subunit of the ribosome, which includes the classical anti-SD motif (3′CCUCC), recognizes a complementary sequence (dark blue) in the 5′ UTR of mRNA (i.e., the SD sequence) by rRNA:mRNA base paring. This interaction places the start codon (red) in the binding site of fMet-tRNA$_i^{fMet}$ (pink).

However, the bacteria and archaea transcripts containing SD motif are less abundant than those without SD sequences. The number of genes with SD motifs among 162 prokaryotic genomes varies from 11.6% to 90%, suggesting the population of non-SD-led genes is significant[25]. Recent ribosome profiling studies in bacteria reported little or no correlation between the ribosome occupancy and the strength of SD motif of a gene[26,27]. The high diversity of the 5′ UTR of prokaryotic genes suggested the mechanisms used for translation initiation in various prokaryotes are more complex and flexible than anticipated.

The mechanism of translation initiation in eukaryotes is more complicated than prokaryotic initiation and is the target of multiple types of regulatory intervention. 5'-cap m^7G structure presents in nearly all eukaryotic cellular mRNAs. The 5'-terminal G with a N^7-methyl group is connected to the first nucleotide of mRNA through a 5'-5' linkage via two pyrophosphoryl bonds. This capping process is achieved by several enzymes, including guanylyltransferase and methyltransferase[28]. The m^7GpppN cap structure, present on most nuclear-transcribed cytoplasmic eukaryotic mRNAs, is critical for splicing, polyadenylation, mRNA stability and translation. As a hallmark of regulation of eukaryotic protein synthesis, it is recognized by different cap-binding proteins that act to impart disparate functional outcomes on gene expression. Cap structure stabilizes mRNA in the nucleus by preventing 5' -> 3' exonucleolytic degradation[29,30,31]. In the cytosol, m^7G cap structure is required for ribosome recruitment to the 5'UTR of mRNA during the initiation through its

interaction with the cap-binding protein (eIF4E)[32], at the meantime, it also participates in the miRNA-mediated suppression of translation.

A great portion of transcripts, except for histone mRNAs[33], bear poly(A) tails at the 3' end. Polyadenylation of eukaryotic pre-mRNAs occurs in the nucleus, which is coupled tightly with transcription termination. Addition of poly(A) tails is a multi-step process initiated by nascent pre-mRNA recognition and binding of the cleavage/polyadenylation machinery at the AAUAAA hexamer (10-30nt upstream of the cleavage site)[34]. Followed by endonucleolytic cleavage of the pre-mRNA, the non-templated poly(A) tail consisting up to 200 to 250 adenosines is added at the 3'end[35]. The polyadenylation of mRNA 3' end serves as a binding site for multiple ribonucleoproteins (RNPs). The poly(A) tail and its associated proteins protect mRNA from enzymatic destruction in the nucleus while mediating mRNA degradation in cytosol. While many prokaryotic mRNAs also acquire poly(A) tails, these tails stimulate decay of mRNA rather than protecting it from degradation. The opposing impact of poly(A) tails on mRNA decay in eukaryotes and prokaryotes is one of the many mechanistic differences between eukaryotic and prokaryotic gene expression.

In addition to its function in mRNA turnover, poly(A) tail also plays essential roles in mRNA translation. The 'close-loop' (circularized mRNA) model of translation initiation in eukaryotes was first proposed based on the discovery of the association between 5' cap structure and 3' poly (A) binding protein[36]. Later studies further validated that this association is achieved through the

interactions between the poly(A) binding protein (PABP) and one component of the cap-binding eIF4F complex, eIF4G[37,38]. The 'close-lope' model not only explains the synergistic effect of cap structure and the poly(A) tail in promoting translation initiation[39] but also allows the direct recycling of 40S from dissociated elongation ribosomes after the termination. The polysome topologies observed by electron microscopy (EM) are mostly consistent with the 'closed-loop' model, although some of the observed shapes suggest additional or alternative interactions between close-spaced ribosomes. However, recent single-molecule-resolution fluorescent in situ hybridization (smFISH) experiments revealed that the bulk of translating mRNAs rarely exhibit colocalized 5' and 3' ends[40,41]. One possible explanation of this observation is that the 3' end probes used in the smFISH targeted at the upstream regions of the poly (A) tails instead of the real 3' end of mRNAs. Whether the eIF4G-PABPC1 interaction is transient in cells and whether the interaction between 5' and 3' ends of mRNAs is stable during the initiation phase remains to be unveiled.

Among the four stages of translation (initiation, elongation, termination, recycling) in eukaryotes, regulation is attained during the initiation, which is the rate-limiting step in the translational process.[42,43] The canonical 'scanning' model of eukaryotic translation cycle can be divided into several steps: formation of 43S PIC, mRNA attachment, 5'UTR scanning, and start codon recognition. Although the current state of knowledge regarding each step of this

model is well established, mechanistic details and alternative mechanisms of ribosome loading, scanning, and start codon selection remain elusive.

1.3 Attachment of PIC

Eukaryotic translation initiation primarily relies on a cap-dependent scanning mechanism. It begins with the loading of small ribosomal subunits to the 5' end of an m^7G-capped mRNAs at the early stage of initiation. This commitment step requires a subset of eIFs to bridge between the 40S and the 5' m^7G cap. The eIF4F complex consists of three core components: eIF4E (a cap-binding protein), eIF4G (a scaffold protein with multiple binding domains for mRNA and other eIFs), and eIF4A (RNA helicase). The cooperative action of the eIF4 group of factors (eIF4F together with eIF4B or eIF4H) is required to pre-activate the 5' end of mRNAs and facilitate the interaction between the 43S PIC and mRNAs. At the same time, Met-$tRNA_i^{Met}$ binds to the native 40S in a ternary complex (TC) with eIF2-GTP, stimulated by eIF1, eIF1A, eIF3 and eIF5. The 43S PIC recruits the mRNA via the interaction between eIF4G in the eIF4F complex and eIF3 in the 43S PIC. In summary, the overall process of PIC attachment can be further divided into three steps: activation of mRNA, assembly of 43S PIC, and 43S PIC loading. Despite a sophisticated knowledge of the roles individual eIFs play in each step, a clear understanding of the temporal order of the steps spanning activation of mRNA to 43S PIC loading is lacking.

1.3.1 Assembly of translation components at the 5' cap

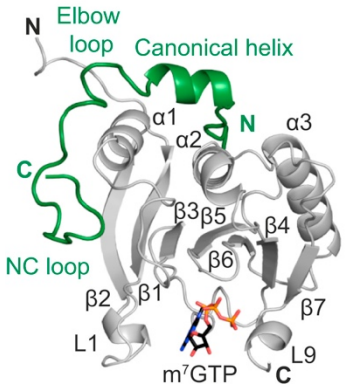

Figure 1-3 Structure of eIF4E bound to m^7G cap.
Crystal structure of human eIF4E (gray) in complex with m7GTP (shown as sticks) and human eIF4G (green). Only eIF4E-binding region of human eIF4G was shown.
Figure from Grüner et al., 2016

Prior to the PIC attachment, mRNAs need to be activated by the eIF4 group of factors. EIF4E, the best-characterized cap-binding protein, recognizes the m^7G cap (Figure 1-3). It was first identified by chemical cross-linking to the viral mRNA cap[32]. A previous study showed that eIF4E occupies mRNA caps most of the time in yeast[45], but its expression levels in most cells are least abundant among the three eIF4F subunits[44]. The binding affinity of eIF4E for the capped mRNA is enhanced with the presence of eIF4G or as part of the eIF4F complex[46,47,48], and further boosted by eIF4G-PABP interactions[49,50,51]. The increased cap-eIF4E association rates were observed on the addition of eIF4G1 and Pab1p (the yeast homolog of mammalian PABP) determined by single-molecule kinetic evidence[50]. In mammals, a well-studied repressor protein regulates the availability of eIF4E for the assembly of the eIF4F complex: 4E-BP[52], which competes with eIF4G for eIF4E and represses the cap-dependent translation in the dephosphorylated form. In Hela cells, the abundance of 4E-BP is similar to eIF4E, indicating that sufficient 4E-BPs are available to compete with eIF4G for binding to eIF4E[53]. Thus, the ratio between eIF4E and eIF4G might vary due to the altered abundance of the factors in mammals[54].

Another core subunit of the eIF4F complex is eIF4A, a weak non-processive helicase that unwinds ~11 base pairs per hydrolyzed ATP molecule[55]. With a prolonged rate (k_{cat}) of ATP hydrolysis on its own, eIF4A acts as a processive helicase in complex with eIF4G (or eIF4F)[56] or its related proteins eIF4B and eIF4H[57,58]. Interestingly, recent studies demonstrated that eIF4A is required for ribosome recruitment by all mRNAs regardless of the 5' UTR structure, suggesting a novel role of eIF4A in the early stage of initiation[59,60]. (Further discussed in Section *1.4.3.1*)

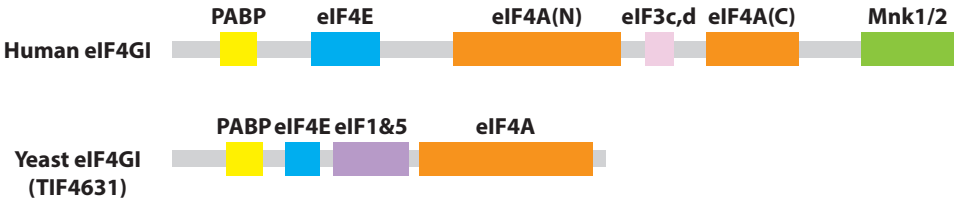

Figure 1-4 Comparison between human and yeast eIF4G1 domain organization.
The identified binding sites for initiation factors are shown above the diagrams. Colored blocks identify homologous regions.

EIF4G is an essential scaffold protein in the eIF4F complex, which possesses multiple binding sites. The N-terminal domain contains the conserved binding sites for eIF4E and PABP[61,62], whereas the middle domain contains the conserved binding site for RNA helicase eIF4A[63,64]. Unlike yeast eIF4G, mammalian eIF4G contains a C-terminal domain with the binding site for eIF4E kinase Mnk[65] and a second eIF4A binding site[66]. As a scaffold, eIF4G links mRNA activation to ribosome loading by its interaction with eIFs in the 43S PIC before it loads onto mRNA. In higher eukaryotes, like mammals, eIF4G facilitates the recruitment of ribosomes via eIF4G affinity to eIF3 mediated by eIF3c, d, and e subunits[67,68]. However, the eIF3-binding site of eIF4G does not

exist in yeast[69]. Instead, yeast eIF4G interacts with eIF1, eIF5[70,71,72], and possibly 40S directly through eIF4G:rRNA binding[73] (Figure 1-4). Thus, 'free' eIF4F might displace eIF4E (or eIF4E:4E-BP) from the m^7G cap as the first step of initiation, potentially with the assistance of mRNA's closed-loop, in the steady-state initiation events. Alternatively, eIF4F could bind the m^7G cap as part of the PIC, as eIF4F can be efficiently isolated from high-salt ribosomal washes but not from the ribosome-free fractions[74].

A 43S PIC complex contains a 40S ribosomal subunit, eukaryotic initiation factors including eIF1, eIF1A, eIF3, and eIF5, and TC (eIF2: GTP: Met-$tRNA_i^{Met}$). Met-$tRNA_i^{Met}$ is delivered to the 40S subunit in the TC with eIF2-GTP. eIF2, a heterotrimeric G protein, consists of the α, β, and γ subunits[75,76]. eIF2γ is the core of the heterotrimeric protein, binding to GTP or GDP and initiator tRNA, and is the connection between α and β units[77]. The sequence of eIF2γ contains all the elements required for GDP or GTP binding. Moreover, the sequence similarity between the elongation factor EF-Tu and eIF2γ suggests that the γ subunit of eIF2 interacts directly with the Met-$tRNA_i^{Met}$ [78,79]. The roles of eIF2α and eIF2β in tRNA binding remains controversial. A number of early studies found little difference between tRNA binding properties of mammalian eIF2 containing the β subunit or not[80,81,82], suggesting the eIF2β may be not involved in tRNA binding. However, recent studies demonstrated the eIF2αγ heterodimers cannot bind initiator tRNA[83] while the eIF2βγ heterodimers did not appear to alter binding of guanine nucleotide or Met-$tRNA_i^{Met}$ ligands by eIF2[84].

These facts together supported the assumption that the eIF2β subunit, rather than eIF2α is indispensable in tRNA binding. Even so, the α subunit of eIF2 plays a critical role in regulating GTP/GDP exchange via phosphorylation at serine 51, which inhibits the eIF2B-catalyzed guanine nucleotide exchange. The affinity of Met-tRNA$_i^{Met}$ is about 10-fold greater for eIF2-GTP than eIF2-GDP, and Met-tRNA$_i^{Met}$ also increases eIF2 affinity for GTP[79,85]. On the other hand, in 43S complexes, the Met-tRNA$_i^{Met}$ anticodon loop is probably not inserted as deeply into the P-site as in the ribosomal complexes with established codon-anticodon base-pairing upon AUG recognition[86].

TC recruitment is facilitated by the eIFs in 43S PIC, like eIF1, eIF1A, and eIF3. These factors bind to 40S subunits directly[87,88] and, at least for eIF1 and eIF1A, cooperatively[89,90]. eIF1 binds to the interface between the platform and initiator tRNA[91]. eIF1A's structured domain resides in the A-site, forming a bridge over the mRNA channel, whereas its N- and C-terminal tails extend into the P-site[92]. Importantly, binding of eIF1 and eIF1A to 40S subunits facilitates mRNA attachment by inducing conformational changes that involve the opening of the mRNA entry channel 'latch' and the establishment of a new head-body connection on the solvent side[93,94]. eIF3 is a multi-subunit complex that consists of 13 subunits (a to m) in mammals[88,95], of which six subunits (a, b, c, g, i, and j) constitute yeast eIF3[96,97]. Cryogenic electron microscopy (Cryo-EM) reconstructions show that the bulk structure of the five-lobed eIF3 complex binds to the solvent side of the 40S ribosome[98], spanning the entry and exit

sites[99,100,101]; whereas the C-terminal of eIF3j localizes in the mRNA channel in the A-site on the interface side[102]. Its j subunit not only enhances 40S binding by eIF3[88,103,87] but facilitates the TC recruitment[89,102,104]. In mammals, eIF3 directly binds to the eIF4G subunit of eIF4F through eIF3e, which are important for binding the 43S complex to the 5′-terminal region of mRNA[68,105]. However, eIF3e is not present in yeast[69], and yeast eIF3 and eIF4G seem not to interact directly[70]. Instead, yeast eIF4G interacts with eIF5[106], and eIF5-CTD can bridge eIF4G interaction with eIF3-NTD to stimulate eIF4G: eIF3 association or bridge eIF4G interaction with 40S directly in yeast[70]. In both mammals and yeast, the main function of eIF5 is to activate the hydrolysis of eIF2-bound GTP during the PIC scanning and to promote the stringency of AUG recognition.

Despite a rich knowledge of individual eIFs in 43S PIC, the sequence of their binding to 40S ribosome or the process of 43S PIC formation remains obscure. The yeast system is better understood due to the discovery of the 'Multi-factor complex' (MFC). Assembly of a MFC (eIF1: TC: eIF3: eIF5 independent of the ribosome and mRNA has been described as an early initiation subcomplex formation in yeast[107,108] To explain the binding of MFC to the 40S subunit, 40S were proposed to directly interact with eIF3j, eIF1, and eIF3c at the mRNA entry site, the P-site, and near the mRNA exit, respectively[109]. This 'subcomplex formation' model has been demonstrated in mammals with different affinities and assembling sequence among factors[110,111,112,113,114,115]. For example, eIF1 and eIF1A can also bind the 40S

independently of the MFC[94]. Also, Met-tRNA$_i^{Met}$ can join the PIC later, either individually or as part of the TC[110]. Furthermore, the timing and extent of binding are unknown for factors which does not have homologous in yeast, such as eIF4B, eIF4H, and DHX29. These observations collectively suggested the existence of multiple routes for factor interacting events whose taking place depends on the type of mRNA and conditions.

1.3.2 Attachment of 43S PIC on mRNA

Despite the rapid development in the field of eukaryotic translation initiation, how the mRNA enters into the ribosome mRNA channel remains unclear. Several models were proposed to explain the mRNA recruitment based on the existing evidences (Figure 1-5). The first hypothesized model was the 'threading' mechanism[116,117], which is postulated that eIF4E first binds to the 43S PIC close to the mRNA entry site where it can insert the 5' end of the mRNA. This process requires an immediate dissociation of eIF4E from the mRNA,

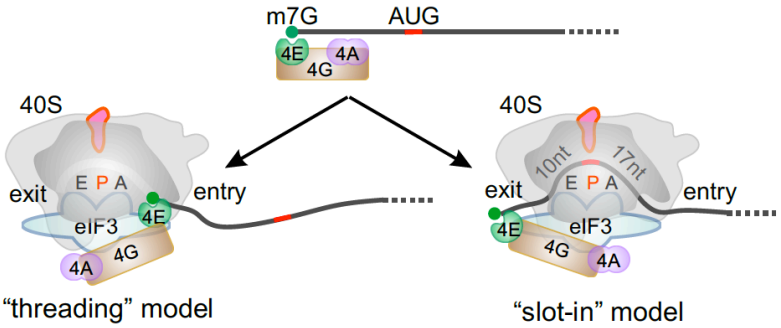

Figure 1-5 Two proposed models for mRNA entering the 40S.
(Left) mRNA 'threads' into the 40S. Cap-bound eIF4E (green) binds the m^{7}G cap, together with eIF4G (yellow) and eIF4A (purple), to recruit the ribosome. The mRNA then detaches from eIF4E and threads into the mRNA-binding cleft of the 40S (gray). This model allows the P-site (red) to inspect the mRNA base-by-base from the 5' end. (Right) Ribosome attaches mRNA via 'slot-in' mode. Following the ribosome recruitment, eIF4E-bound m^{7}G cap first attaches near the mRNA exit site. The 3'ward portion of mRNA is then 'slotted' laterally into the mRNA binding cleft. This model places a portion of mRNA (around 10-12nt) upstream of the P-site.

which allows the mRNA to extrude into the mRNA-binding cleft. In theory, this model permits a base-by-base linear inspection of mRNA from the 5' end, leaving no 'blind spot' during ribosome scanning. The *in vitro* experiment showing that 43S PIC is able to recognize 5' proximal start codon (1nt downstream of the 5' end) is one of the most substantial evidence supporting the 'threading' model[118] by suggesting the absence of an absolute 'blind spot' in 5'UTR[118].

However, the possible position of eIF4E in proximity to E-site is inconsistent with the presumption of the 'threading' model in which eIF4E is placed towards the A-site. Inspired by the discovery of internal ribosome entry sites (IRES), N^6-methyladenosines (m^6A) mediated initiation, and circular RNA's translation, the 'slot-in' model was proposed describing a lateral attachment of mRNA to the 43S PIC. In this model, 5' cap binds eIF4E sits near the exit site of the ribosome, followed by a lateral attachment of downstream mRNA into the mRNA-binding cleft. This hypothesis is supported by the following evidence: 1) eIF4G-binding site for eIF4E is in proximity to eIF3e, 3c, and 3d, all of which are located close to the exit site of mRNA channel[68,67,119]. More importantly, eIF3d is considered as an alternative cap-binding protein[120], implicating the similar position of eIF4E. 2) Cryo-EM structures showed that eIF3 uses eIF4F or the HCV IRES in structurally similar ways to position the mRNA near the exit site of the ribosome[121,98]. 3) The 5' end of mRNA was found to interact with protein

eIF3a, eIF3b, and eIF3d instead of small ribosomal proteins (RPS) revealed by cross-linking[122].

The major conflict between the 'threading' model and the 'slot-in' model seems to be the presence of 'blind spot', as the 'slot-in' model is expected to possess a 'blind spot' between the 5' end and the decoding site (around 12nt downstream). Yet the necessity of the 'blind spot' in 'slot-in' model is debatable. In theory, the lack of 'blind spot' can be explained with a potential non-linear scanning mode of 43S PIC after the lateral attachment of mRNA. If the 43S PIC can perform a bi-directional motion of scanning after the accommodation of mRNA, a linear 5' to 3' inspection from the very 5' end may not be necessary for the recognition of AUGs in the upstream of the decoding site.

Recently, a third 'dragging' mechanism was proposed to reconcile most of the current data[123]. This model entails 40S binding of the eIF4F-mRNA complex near the entry site, followed by a repositioning of eIF4E or eIF4F to the exit site. During the repositioning process, mRNA is dragged into the mRNA channel with a base-by-base inspection from the 5' end through the constant interaction between eIF4E and 5'cap. The small size of eIF4E and the interaction with eIF4G through its dorsal surface[124] make the 'dragging' mechanism plausible, leaving little or no gap between the eIF4E bound 5' end and the inspected portion of 5'UTR sequence by the decoding center. However, the ability of eIF4E to transverse the interface side of the small subunit has yet been identified.

1.3.3 Cap-independent attachment of 43S PIC

IRES-mediated initiation

Although most mRNAs primarily use the cap-dependent scanning pathway, a certain amount of cellular mRNAs use *cis*-acting RNA elements called internal ribosome entry sites (IRES) structure to directly recruit the 40S ribosome, bypass the necessary recognition of the m^{7}G cap[125]. This cap-independent mechanism has been identified in the translation of numerous viral RNAs, and there is evidence showing that certain cellular mRNAs also contain IRES elements[126]. Structurally related IRESs were categorized into different classes based on the distinct mechanisms that depend on particular eIFs for ribosome recruitment[127,128].

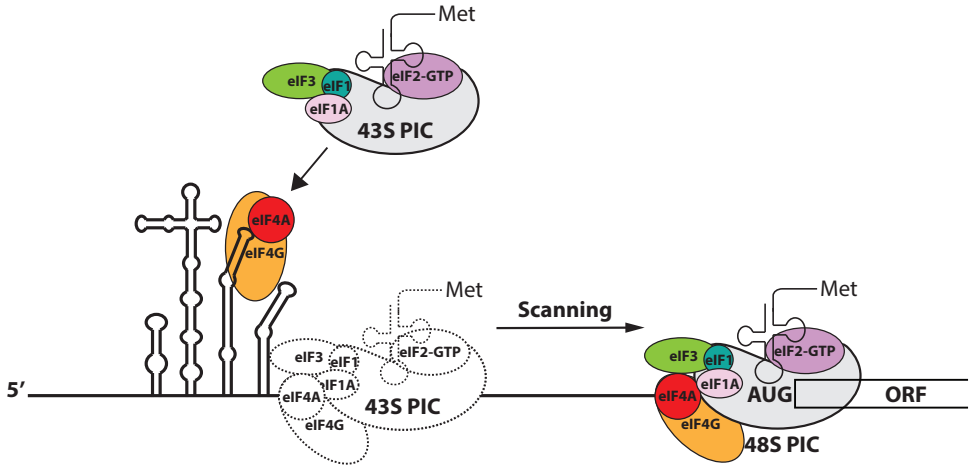

Figure 1-6 Polio (Class I) IRES mediated cap-independent initiation.
The eIF4G (orange)-eIF4A (red) complex recruits 43S complex to the Polio IRES independent of eIF4E, followed by the scanning of downstream region by the PIC until the recognition of start codon. Most eIFs, such as eIF2 (purple), eIF3 (green), eIF1 (teal) and eIF1A (pink) are involved in this process.

Class 1 (e. g., Poliovirus – Figure 1-6) and 2 (e. g., EMCV) IRESs found in the 5'UTR of the viral RNAs in picornaviruses, require almost the entire set of canonical translation initiation factors (excluding eIF4E) for the ribosome attachment[127,129]. Both class 1 and 2 require IRES trans-acting factors (ITAFs)

to stabilize the three-dimensional of conformation-specific IRES[130]. Assisted by ITAF, eIF4G and eIF4A recruit 43S PIC complex to the class 1 and 2 IRESs without the involvement of eIF4E[131,132]. Despite these similarities, class 1 and 2 IRESs differ in the secondary structure and ribosome attachment mechanism. Mediated by the class 1 poliovirus IRES, the 43S PIC is initially recruited to a 'cryptic' AUG codon within the last second domain of the IRES structure, followed by scanning through the downstream untranslated region (including the last domain of the IRES structure) for start codon recognition[133,134]. In the case of the class 2 EMCV IRESs, the second of two closely spaced AUG codons is selected by the 43S PIC immediately after ribosome loading without any scanning process[135]. Initiation on Class 3 (e. g., HCV) IRESs involves their direct binding to the 40S ribosomal subunit and positioning it to the initiation start codon through a scanning-free process independent of eIF4F, eIF4B, eIF1, and eIF1A[121,136]. To initiate translation, class 4 (e. g., CrPV) IRESs bind directly to the 40S ribosomal subunit at the start codon with high affinity independent of Met-tRNA$_i^{Met}$ or any initiation factor (including eIF2)[137,138]. The P-site is occupied by an IRES pseudoknot domain, which mimics the tRNA structure and the mRNA-tRNA codon-anticodon interaction[139,140,141]. In summary, both class 1 and 2 IRESs need most canonical eIFs and other ITAFs for translation initiation to occur, which are not necessary for class 3 and 4. The scanning process is only required for class 1 IRESs, while the other three classes of IRESs recruit 40S ribosomes to the AUG codons through scanning-free pathways. Despite a

rich knowledge of the biochemical characterization of each viral IRES class, cellular IRESs show little structural similarities to each other, and their underlying mechanism remains obscure[130,142]. Moreover, a recent stringent test has questioned the internal initiation capability of many cellular mRNAs containing IRESs[143]. As most cellular IRES-containing mRNAs also bear m^7G cap and can be translated by the canonical initiation mechanism, the intriguing question is how to switch between different modes of initiation for these IRES-containing mRNAs.

m^6A-mediated initiation

Another well-known cap-independent translation mechanism is mediated by methylated adenosine residues in the form of N^6-methyladenosines (m^6A) in the 5′ UTR[144,145]. m^6A is the most abundant post-transcriptional modification in eukaryotic mRNA[146,147]. Directed by its consensus sequence RRACH[148,149], m^6A is installed to the nascent mRNA by methyltransferase complex (m^6A writer) comprising a Methyltransferase Like 3 protein (METTL3), a Methyltransferase Like 14 protein (METTL14), and a Wilms' tumor 1-associating protein (WTAP). It was found that Fat mass and obesity-associated protein (FTO)[150] and AlkB homolog 5 (ALKBH5)[151] are responsible for the demethylation of m^6A (m^6A eraser). m^6A has a wide role in RNA biology, including RNA stability[152], splicing[153], and RNA secondary structure[154]. Cap-independent translation initiation mediated by m^6A was discovered recently[144,145], as exemplified by the

selective translation of heat shock-induced mRNA (e. g., Hsp70) in control of the YTH N^6-Methyladenosine RNA Binding Protein 2 (YTHDF2; m^6A reader)[145]. Interestingly, the observations that m^6A-mediated translation both occurs on fully capped mRNAs and co-exists with eIF4F-mediated cap-dependent translation indicate that the presence of m^6A enables mRNAs to undergo different translational initiation models[155]. In alignment with this notion, the 5'UTR methylation in the form of m^6A has been reported to not only promote cap-independent translation under heat shock stress[145] but also regulate start codon selections under amino acid starvation[155,156]. However, the mechanisms whereby m^6A modification facilitates PIC attachments or regulates TIS selections by PIC remain unclear. eIF3a has been suggested involved in m^6A-mediated translation initiation, as evidenced by its cross-linking with m^6A and by the overlapping between m^6A peaks and eIF3a Photoactivatable Ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) data[144]. Other studies indicated that the ribosome scanning might be influenced by m^6A-induced RNA-protein interactions or RNA structure formation which impact on start codon selections[154,157]. Given the critical role of m^6A modification in physiological processes and the participance of m^6A regulators in pathogenesis[158,151,159,160,161], the molecular mechanism underlying the m^6A-mediated translation and its translational control merits further investigation.

Circular RNA initiation

Covalently closed circular RNAs (circRNAs) were first discovered in the 1970s. Because of their non-linear conformation and the lack of both m^7G cap and polyadenylated tails, it has been suggested that translation of circRNAs relies on a cap-independent manner. Although most endogenous circRNAs were classified as non-coding RNAs (ncRNAs) which are not involved in translation, recent studies have discovered that a small subset of peptides aligned with the sequence of circRNAs corresponding proteins expressed in human[162]. For example, the human circZNF609 was found to be translatable, as evidenced by its association with translating polysomes[163]. It was reported that the translation of circRNAs is achieved via an IRES-mediated mechanism[164,165,166]. IRES structures were embedded within circMbl and circZNF609 to allow the cap-independent translation[163,167]. In addition to IRESs, m^6A modification can also drive circRNA translation, as evidenced by the facts that the translation efficiency was impaired in the absence of m^6A demethylase FTO and promoted by knockdown of the m^6A methyltransferase complex METTL3 and/or 14[168]. Although the functions of the circRNA-encoded proteins remain unknown, circRNA translation was altered in cells under stress conditions. For instance, starvation resulted in increased circMbl translation[167], and heat shock promoted GFP translation from the m^6A-containing circRNA plasmid[168]. Although only a small proportion of circRNAs are associated with polysomes, circRNAs have been discovered to be key regulators that mediate

many fundamental cellular processes[169]. A better understanding of the role and underlying mechanism of circRNAs translation will discover a hidden human proteome and enhance us to appreciate the importance of circRNAs in human diseases.

Poly (A) tract-mediated initiation

Although most viral RNAs translation are driven by IRES structures, mRNA translation in poxviruses must rely on other 5' UTR features as no well-defined IRES structures have been described. Poly (A) sequences, mostly 30-40 nt long, were found immediately upstream of AUG codon in the 5' leaders of highly expressed mRNAs in poxviruses[170]. Translation of the poxvirus mRNAs with a poly (A) tract was later shown to be independent of m^7G cap as well as some translation initiation factors such as eIF4A, eIF4F and eIF3[171,172,173]. Intriguingly, the A-rich elements are also prevalent in a type of low structured eukaryotic cellular IRESs, such as *Drosophila* hsp70[174] and yeast YMR181c. In yeast, the poly (A) tract has been found to promote cap-independent IRES-mediated mRNA translation via Pab1 and eIF4G[175]. The opposite translational dependence on eIF4G from these studies suggest that the IRES-facilitating activity of the internal poly (A) tract is different from the direct ribosome recruiting ability of poly (A) leaders. In the future, more investigations are needed to fully understand the potential role of poly (A) tract in alternative translation initiation modes.

1.3.4 Cap-dependent but eIF4E-independent attachment of 43S PIC

eIF3d-mediated initiation

A recent study discovered that eIF3 could bind directly to the m^7G cap independent of eIF4F via its eIF3d subunit at the 5' end of c-Jun mRNA. The presence of a specific stem-loop structure within the c-Jun mRNA prevents the recruitment of eIF4F to the 5' end and directly induces the 'open' state of eIF3d for its cap binding[120]. In alignment with this eIF3d: m^7G interaction, evidence from recent studies demonstrated the localization of eIF3d near the mRNA exit channel in the 40S ribosome[176] and its interaction with the upstream mRNA[122]. However, there is a high likelihood that eIF3d functions as an eIF4E cap-binding analog in an mRNA specific manner to promote translation upon allosteric activation, exemplified by the eIF3-binding site in the *JUN* 5′UTR[120].

1.4 Ribosome Scanning

Once the 43S PIC loads onto the eukaryotic mRNAs, it follows the scanning mechanism wherein the 43S PIC migrates on the leader (5'UTR) of mRNAs and continuously inspects the initiation codon (AUG) using complementarity with the anticodon of Met-$tRNA_i^{Met}$. Regulation of translation initiation in eukaryotic mRNAs is primarily conferred by their 5′ UTR features including secondary structures, sequence elements, and length of 5'UTR. Highly structured 5'UTRs have long been known to inhibit scanning; while extremely short 5'UTRs, even without any structure, also dramatically decrease

translation efficiency. Structures that impede threading of mRNA in single-strand form through the 40S subunit are effectively unwound during scanning. Intriguingly, even the scanning of 5'UTR containing weak secondary structures requires the participation of ATP and eIF4F groups including eIF4A, eIF4G, and eIF4B[177]. Other RNA helicases, such as DHX29 and Ded1, have been implicated in scanning on long and highly structured 5'UTRs[178,179]. Ribosome attachment is achieved by interactions among mRNA-40S-eIFs, whereas the fate of each link during the transition from the PIC attachment to the PIC scanning is unclear. In the scanning process, whether the m^7G cap remains attached to the PIC complex has been one of the most vigorous debates in the field. While the widely-accepted classic model of ribosome scanning depicts a 5' -> 3' uni-directional motion of the PIC complex, a non-linear bi-directional motion of the scanning ribosome has been proposed to account for the dynamic nature of scanning ribosome. With above mentioned open questions, the characterization of this highly dynamic process and the molecular mechanism by which factors assist scanning require further investigation.

1.4.1 Features of 5'UTR

1.4.1.1 5'UTR length

In eukaryotes, the average genome length of 5' UTR ranges from ~100 nt to ~200 nt in mammals[180,181] and ~50 nt in yeast[182]. However, the 5'UTR length varies considerably among the transcripts in a genome, ranging from a few base-pairs to several thousand base-pairs[183]. Around 0.25%-0.35% mRNAs in human transcriptome harbor 5'UTR shorter than 10nt (Figure 1-7). It was thought that a minimal length of 20nt is required for efficient recognition of a start codon by the ribosome[184]. Despite that, researchers reported that mRNAs with a very short 5′ UTR or no 5'UTR (e, g,. leaderless) present naturally in a number of primitive eukaryotes[185,186,187]. Also, people found that 5'UTR containing only one nucleotide is able to initiate translation using a mammalian *in vitro* system[188]. However, this discovery has not been examined in mammalian cells because of the rarity of authentic leader sequences shorter than ten nucleotides. With the development of deep sequencing and other biochemical methods, detecting endogenous mRNAs harboring 5'UTR in the range of 0 to 10 nt has become feasible[189]. The key

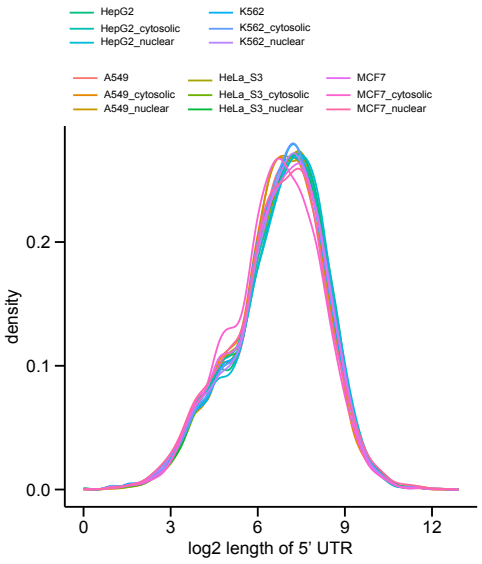

Figure 1-7 Distribution of human 5'UTR length.
Analysis is based on CAGE-seq data from cytosolic, nuclear and total mRNAs across five human cell-lines.

question is how the ribosome could effectively scan such short 5'UTR for accurate start codon selection. This class of 5'UTR could involve a novel translation initiation mode in mammalian cells. Recently, some studies have reported a mechanism by which mRNA with short 5'UTR (~12nt) is translatable. It showed this translation mode is independent of scanning but relies on a restricted consensus sequence for start codon recognition[190]. (Further discussed in section *1.4.6*) However, it remains to be elucidated how mRNAs with 5'UTR in the range of 1 to 10 nucleotides without a conserved sequence are translated *in vivo*.

1.4.1.2 Sequence elements

Kozak sequence

Discovered by Kozak, consensus particular sequences flanking the AUG enhance AUG selection by the scanning 43S PIC, which is known as the Kozak sequence. A 'strong' Kozak consensus sequence is gccRccAUGG, in which the start AUG is surrounded by highly conserved nucleotides (uppercase letters indicate highly conserved nucleotides; lowercase letters indicate less conserved nucleotides; R indicates a purine nucleotide) [191]. This consensus sequence gccRccAUGG is later validated in vertebrates[192]. The consensus sequence is WAMAAMAUGUCY (W: A/U; M: A/C; Y: Pyrimidine) in yeast[193], and AMAAUGGC in higher plants[191]. These sequences were derived from a compilation of relatively abundant, well-translated mRNA. The variations on

consensus sequences may be related to the availability or the abundance of initiation factors in different species. In human, mutations in the Kozak context have the potential to disrupt protein translation levels and impact health[194,195]. For example, several nucleotide context mutations are associated with tumorigenesis[196]. A recent high-throughput sequencing experiment indicated that -3 and the +4 positions are critical in determining TIS efficiency, while the –2, –4, and +5 positions tend are auxiliary[197].[197] Despite its important roles in TIS selection, a large-scale analysis of human mRNAs indicates that only 38% of mRNAs contain an annotated start codon in the optimal Kozak context. Of the remaining mRNAs, 52% and 10% are in suboptimal and non-optimal Kozak contexts, respectively, suggesting lower translation efficiencies of these mRNAs[198]. Currently, it is unclear whether the translation of these suboptimal mRNAs undergoes a leaky scanning of the annotated AUG or follows less-stringent context criteria.

Upstream open reading frames (uORFs)

About 45–50% of mammalian genes (but only ~ 13% of yeast genes) encode mRNAs that have at least one short uORF (typically < 30 codons) upstream of the main coding region[199,200,201]. Ribosome profiling experiments uncovered widespread translation at many of these uORFs[202]. In some cases, uORFs are involved in translational control of annotated ORF. The best-characterized example is the Activating Transcription Factor 4 (ATF4) in

mammals (GCN4 mRNA in yeast). Different from the main ORF translation which is inhibited by uORFs (mainly by uORF2 translation) under normal conditions, ATF4 expression is stimulated by 5-fold upon eIF2α phosphorylation[203,204]. The stimulatory effect is explained by the particular uORF configuration, including a very short upstream uORF1 and a longer uORF2 overlapping the ATF4 (or GCN2) ORF. (Mechanistic details are further discussed in Section *1.4.6.*) A short uORF has also been suggested to stall the cap-recruited ribosome that enhances the downstream IRES-mediated translation[205]. Examples include the uORFs upstream of amino acid transporter cationic 1 (CAT1, also called SLC7A1)[206,207] and fibroblast growth factor 9 (FGF9)[208]. Despite these regulatory effects of uORFs, a recent computational study based on ribosome profiling data examined the relationship between the features of 5'UTR and annotated ORF translation efficiency (TE), and concluded that uORFs are generally repressive but only minor determinants of global annotated ORF translation, while the sequence variation in 5' leader at individual genes may mainly contribute to the expression level[209].

5' terminal oligopyrimidine (5' TOP motif)

The 5'TOP motif is a *cis*-regulatory RNA element located immediately downstream of the m^7G cap of TOP mRNAs that are essential for ribosome biogenesis[210]. Comprising an invariant 5'-cytidine followed by an uninterrupted tract of 4–14 pyrimidine nucleotides, 5' UTRs of TOP mRNAs are relatively

short (40nt on average), and free from stable secondary structures and uORFs[211]. Currently, there are 97 widely accepted TOP mRNAs in higher eukaryotes, which encode mostly translation factors and nearly all ribosomal proteins (RP)[212]. However, a large-scale analyses of mRNA 5′ sequences hinted that thousands of mRNAs may encode TOP sequences[213]. Previous genome-wide studies revealed that the mRNA subsets highly sensitive to the mammalian target of rapamycin complex 1 (mTORC1) signaling consist almost entirely of transcripts with 5'TOP elements[214,215]. In response to nutrients, the activated mTORC1 signaling boosts the production of components of the translation machinery encoded by TOP mRNAs[212,216]. eIF4E, and its repressor proteins, 4E-BP1 and 4E-BP2, have been implicated in the control of TOP mRNA translation downstream of mTORC1[214,215]. Most recent findings indicated that the RNA-binding protein La-related protein 1 (LARP1) may play a central role in mediating translation of 5'TOP mRNAs downstream of mTORC1[217,218]. It has been proposed that, upon mTORC1 inhibition, LARP1 dissociates from mTORC1 and binds to the 5'TOP motif of TOP mRNAs to inhibit the translation[217]. Both crystallography and biochemical analyses revealed that the LARP1-unique domain DM15 specifically binds to the m^7G cap and the invariant first cytidine of TOP mRNAs to prevent the binding of eIF4E to the m^7G cap and block the assembly of the eIF4F complex on TOP mRNAs[219]. A recent crystallography study confirmed that the DM15 pocket of LARP1 involves in TOP-specific m^7GpppC-motif recognition and demonstrated that this

DM15 region can serve as a future pharmacological target[220]. In addition to its translational control on TOP mRNA, LARP1 has been suggested to mediate the stability of TOP mRNAs through its association with poly (A) tail via PABP1[221,222]. Future studies are required for a deeper understanding of the translational control of transcripts with 5' TOP elements and the cellular biology of ribosome biogenesis.

Other 5' UTR elements

Besides the consensus sequence flanking start codons, uORFs, and 5' TOP elements, other 5'UTR elements have also been implicated in the control of translation initiation. For example, the poly (A) tract in 5' leader region has been proposed to induce cap-independent translation (Discussed in Section *1.3.3*); The Translation Initiator of Short 5′UTR (TISU) motif was found to induce the cap-dependent scanning-free translation (Discussed in Section *1.4.6*); RNA modification motifs, like RRACH which induces m^6A modifications (Discussed in Section *1.3.3*), play critical roles in the regulation of translation initiation. With the accumulating knowledge of these 5′ UTR elements, we can expect a better understanding of the diverse mechanisms of translation initiation in eukaryotes.

1.4.1.3 Secondary structures

With the intrinsic propensity to fold and form a higher-order structure quickly and dynamically, mRNA displays a second layer of structural information

specified by the sequence. Secondary structures in the 5'UTR affect the frequency of translation initiation and participate in the regulation of protein synthesis. Primarily in bacteria, RNA secondary structures in the 5'UTR direct multiple functions, including shaping the initiation region (around initiation codon and Shine-Dalgarno sequence), modulating bacterial metabolism through riboswitches and serving as the interacting sites for various RNA-binding proteins (RBPs). While in eukaryotes, structures in the 5'UTR tend to play an inhibitory role in mRNA translation by blocking ribosome loading or impeding the ribosome scanning process. One example is the iron-responsive element (IRE)[223] which represses 43S PIC associating with iron homeostasis related mRNAs (e, g,. ferritin and ferroportin) by binding iron-regulatory protein 1 (IRP1) or IRP2 in low-iron conditions[224,225,226]. However, there are also a few cases where stem-loop structures activate translation by directly binding to eIF3 and recruiting ribosome while bypassing the 5' cap of cellular mRNAs[227].

Parameters such as GC content and negative folding free energy (ΔG) for modeling RNA secondary structures have been used to predict the intricate RNA base-pairing patterns. Whereas, it has difficulty in processing long RNA sequences and predicting distant interactions or complex structures. Recently, by integrating the deep-sequencing and experimental RNA structure probing methodologies, researchers were able to identify genome-wide structural patterns and motifs in coding and non-coding regions of transcripts from various organisms[228]. Studies of high-throughput selective 2′-hydroxyl acylation

analyzed by primer extension (SHAPE)[229], parallel analysis of RNA structures (PARS)[230,231], ds/ssRNA-seq[232,233], DMS-seq[234] and Structure-seq[235] experiments revealed that the relative structural contents vary among 5'UTR, CDS, and 3'UTR in different organisms. Both 5'UTR and 3'UTR are associated with an increased level of RNA secondary structures than CDS in HIV, *Drosophila*, *C.elegans,* and human mRNAs. In contrast, the opposite results were obtained for yeast and *Arabidopsis*, which explained why translation is 90% inhibited by structures with low stability ($\Delta G > -20$ kcal mol^{-1}) in yeast[236,237]. It is not quite clear whether the observed difference in structural pattern among eukaryotes is due to characteristics defined by species or technology issues. Nevertheless, for all eukaryotic species, a significant decrease in mRNA secondary structure contents around the initiation codon has been universally observed[230,231,235,238] (Figure 1-8). And all these genome-wide studies show that mRNA structures in 5'UTR exert a significant translational effect on gene expression.

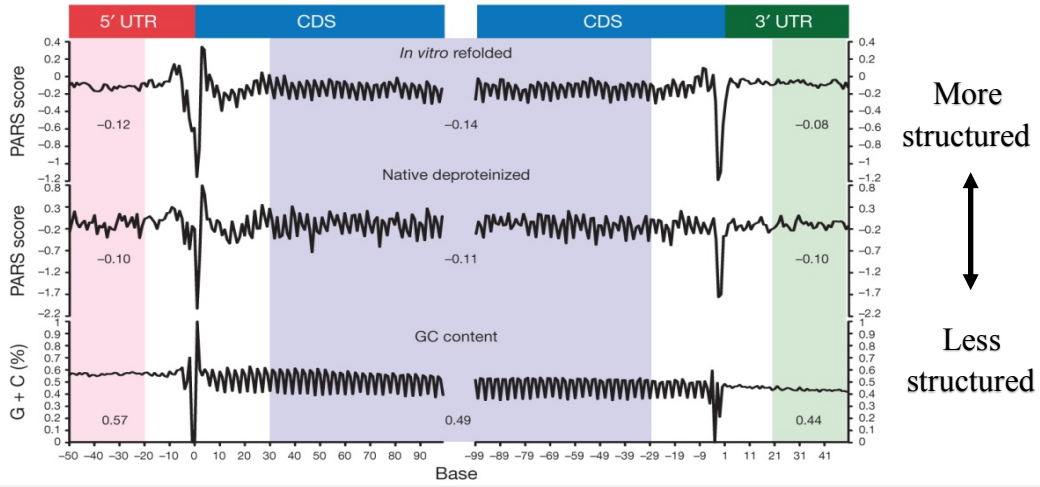

Figure 1-8 Landscape of RNA structure across the human RNA transcriptome.
PARS score (renatured transcripts; native transcripts) and GC content across the 5'UTR, CDS and 3'UTR, averaged across all transcripts, aligned by translational start and stop sites.
Figure is modified from Wan et al., 2014

Secondary structures - Hairpins (Stem loops)

Based on the current knowledge, increasing predicted mRNA thermal stability leads to a decrease in translation efficiency. *In vitro* studies by Kozak et al demonstrated that hairpins with the predicted thermal stability of -30 kcal/mol did not affect mRNA translation, while those of -50 kcal/mol reduced translation by 85%-95%[239]. It was later adjusted to the range from -25 kcal/mol to -35 kcal/mol where the translation efficiency falls off dramatically by a systematic *in vivo* study[240]. Besides the thermal stability, following studies showed that hairpin position is another critical factor on translation efficiency. 1) Translation efficiency decreased dramatically as the distance between the m^7G cap and stem-loop reduced to several nucleotides. Structures with low stability reduced translation when it was placed 12nt from the cap[241]. Later studies confirmed this phenomenon by showing that the cap-proximal hairpins had the greatest inhibitory effects on translation when placed within the first 16nt of the 5'UTR in live cells[240]. Cross-linking studies reported that stem-loop positioned immediately adjacent to the cap within 6nt downstream impairs eIF4E's ability to interact with the cap[242]. 2) Stem-loops in 5'UTR close to the initiation codon can block scanning, although the inhibitory potential of a the cap-proximal stem-loop is greater than those in positions close to the AUG in the eukaryotic systems[236,241,243,244]. Consistently, several structurome studies in both yeast and human revealed a negative correlation between the density of structures at the proximal upstream region of the initiation site and the ribosome density

throughout the transcripts[230,245]. 3) In contrast, when located downstream of, and in proximity to, an initiation codon, a hairpin can increase utilization of the initiation codon by slowing 40S migration and allowing increased codon sampling time[239,246]. A transcriptome-wide *in silico* analysis study of human and mouse mRNAs reported that the third position of the start codon (G) tends to reside in local structures while the first two (A and U) tend to remain unpaired[243,247]. In summary, both the stability and the position of hairpin structures affect initiation and mRNA translation in eukaryotes.

Secondary structures - Antisense oligonucleotides (ASOs)

In the past decades, extensive studies to develop and exploit antisense technologies have been conducted. ASOs are synthetic single strands of 16–20 nucleotides and engineered to regulate mRNA and protein expression with nucleotide specificity through several distinct mechanisms. ASOs act on RNA to modulate splicing, elicit degradation of the mRNAs, and regulate translation through highly specific hybridization to a target RNA. ASOs are chemically modified with increased binding affinity to single-strand RNA, stronger nuclease resistance, and enhanced pharmacological properties. Importantly, several ASO technology-based drugs have been approved by the FDA[248], such as the treatment for homozygous familial hypercholesterolemia and transthyretin amyloidosis[249,250].

Despite a substantial knowledge of the ASOs from the perspectives of synthetic chemistry, pharmacologic, pharmacokinetic, and toxicologic properties, a clear understanding of the basic mechanisms of each type of ASOs remains lacking. For example, morpholinos as one group of ASOs can bind to RNA with high affinity and have limited interaction with proteins. They are designed to bind the sequence in the 5'UTR of the transcripts and sterically block the translation initiation without affecting the mRNA stability. The advantage of this method is that it inhibits the mRNA translation without inducing mRNA degradation pathway or influencing RNA processing. However, recent studies show that using ASOs targeting different regions or elements in the 5'UTR could exhibit opposite effects. Several studies reported that binding the ASOs to the 5' cap proximal region blocked the interaction of 40S to the mRNA and to prevented the formation of the PIC[251]. In alignment with this, the inhibitory effects of ASOs reduce dramatically as target sites are moved downstream within the 5' UTR. It has been reported that the ASOs targeting the region containing AUG show more inhibitory effects than those targeting the upstream or downstream regions of the AUG[252,253]. These observations suggested that ASOs might have similar effects on mRNA translation as the hairpins. Interestingly, recent studies demonstrated that the ASOs increase the translation efficiency by targeting the inhibitory elements in 5'UTR, like uORF or structured region[254,255,256,257]. In conclusion, the activities of the ASOs in modulating translation are position/sequence-dependent.

Secondary structures - unwinding or bypassing

The intriguing question is how a scanning PIC navigates through regions of the mRNA that contain secondary structures. Since the canonical scanning requires ATP[258], the energy is speculated to be consumed by the ATP-dependent DEAD/H-box RNA helicases to overcome the structural impediments. Three RNA helicases, eIF4A, DHX29. and DDX3, have been shown to unfold and rearrange the secondary structures in the mRNA leaders. Their functions in translation initiation are not redundant as they might be specialized for certain types or locations of mRNA structures[178,60]. The rate of cap-binding reduced significantly if the interaction between eIF4E and the cap is sterically hindered by the cap-proximal structures[259,260] (Figure 1-9). To enable PIC loading, eIF4A unwinds the 5' proximal secondary structure with the assistance of eIF4B/eIF4H[261]. However, eIF4F, eIF4A, and eIF4B function poorly in the mammalian system to stimulate scanning through strong SLs of –19 kcal/mol or less[122,262]. DHX29 and the yeast homolog of DDX3 (Ded1) were suggested to stimulate scanning through highly structured regions (-30 kcal/mol or less) distal from the cap[178,262,263,179,264]. Ribo-seq data in yeast Ded1 and eIF4A mutants revealed that Ded1-hyperdependent genes tend to have long and structured 5'UTR, whereas eIF4A contributes more

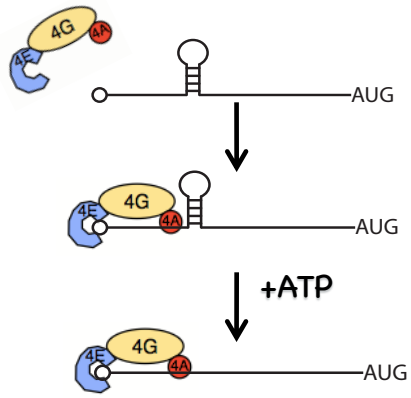

Figure 1-9 eIF4A remodels the cap-proximal secondary structures.
eIF4A (red) in the complex with eIF4G (yellow) and eIF4E (blue) associates 5'end of mRNA and unwinds the cap-proximal secondary structures stimulated by ATP.

equally to the translation of all mRNAs regardless of the degree of structures[265]. Consistently, mammalian homolog DDX3 depletion reduced the expression of mRNA with highly structured 5'UTR[266]. A recent study reported that the mammalian DDX3 cooperates with eIF4A in destabilizing cap-proximal SL before PIC attachment[267]. The synergistic action taken by Ded1 and eIF4A was also confirmed by both unwinding assay[268,269] and the Atomic Force Microscopy (AFM)[270]. Characterization of each RNA helicase and their abilities in unwinding the secondary structures are further discussed in Section *1.4.3*.

Although it is generally accepted that RNA structures with moderate stability are destabilized before entering the mRNA cleft of the 40S subunit, several studies reported the intact stem could be 'bypassed' by the scanning PIC or to some extent penetrates the mRNA-binding cleft. In the absence of DHX29, 48S complexes form on AUG triplets immediately preceding stable stems without unwinding, characterized by aberrant toe-printing +11-12 nt downstream of the stem[262]. Another study demonstrated that ribosomes could efficiently bypass a stable double-stranded RNA bridge without melting the structure[271]. This concept of bypass is similar to the mechanism of ribosome shunting in a way that the encoding RNA signals direct the ribosome to bypass/shunt sections of the 5′ UTR, including highly structured regions and short uORFs, on its way to the AUG start codon. Ribosome shunting was first described in the *Cauliflower mosaic virus* gene[272] and was later reported in several cellular mRNAs. Two of the few examples of this mechanism in the

cellular mRNAs are β-secretase enzyme1 and cellular inhibitor of apoptosis 2, both of which are efficiently translated while containing multiple upstream AUGs (uAUGs) and strong secondary structured regions in 5'UTR[273,274]. The efficient bypassing of stable stems, leading to formation of 48S complexes on AUGs immediately upstream and downstream of the stems, was promoted by eIF4 groups of factors while suppressed by DHX29 and eIF1[262]. Therefore, it was proposed that DHX29, like eIF1, can structurally remodel the PIC (entry site), rather than directly removing the secondary structures[262]. The 'bypassing' or 'shunting' model of the eukaryotic ribosome is still controversial[275,276], and the mechanistic details of the stem 'bypassing' or 'shunting' by the ribosome in eukaryotic 5'UTR is beyond the comprehension.

1.4.2 Scanning-competent PIC

During the transit period of mRNA accommodation into the entry channel of the 43S PIC, the complex undergoes a conformational change to get ready for TC binding, PIC scanning, and start codon selection. Induced by eIF1 and eIF1A, PIC adopts a scanning-competent conformation in which it opens the mRNA binding channel for mRNA accommodation[277,177,278,94]. The omission of eIF1A reduces the intrinsic scanning ability of 43S complexes, whereas the absence of eIF1 almost abrogates it[177]. Importantly, eIF1 antagonizes the conformational changes (from 'open' to 'close') of the 40S subunit which occur upon the codon-anticodon base-pairing. Another function of eIF1 during

scanning is to discriminate non-AUG triplets and AUG triplets in a suboptimal context[177,279,280]. Recent structural studies using cryo-EM revealed that the open (scanning-competent) conformation of the 40S subunit relies on the participants of eIF3, eIF5, and eIF2. The interaction between the N-terminal tail of eIF1A (eIF1A-NTT) and the C-terminal of eIF5 (eIF5-CTD) strengthens binding between eIF1 and eIF5 which retains eIF1 in the 43S PIC[281]. Disruption of this interaction leads to the release of eIF1 and a shift of PIC from the scanning-competent state to the scanning-incompetent state upon AUG recognition[281,282,71]. NMR and complementary binding assay revealed that the N-terminal of eIF3c strengthened the interactions of eIF1 interactions with eIF5 and 43S PIC to modulate the conformational change of PIC during scanning and upon AUG recognition[283]. A recent study using co-immunoprecipitation (Co-IP) and mutation analysis reported that eIF4E indirectly interacts with the eIF1-binding site of eIF4G, suggesting the competitive natures of eIF4E and eIF1 for eIF4G binding. This finding underlined the rearrangement of 43S PIC from the cap-binding mode to the scanning mode during the canonical scanning-dependent initiation[284]. Intriguingly, in the open conformation of the PIC, Met-tRNA$_i^{Met}$ in complex of TC is not fully engaged in the P-site until recognition of AUG, which triggers the transition to the closed conformation with Met-tRNA$_i^{Met}$ fully accommodated in the P-site[285,286]. Moreover, it has been proposed that TC is not absolutely required for mRNA recruitment or PIC scanning, especially when TC availability is reduced in response to stress or stimuli[287,288]. Even so,

the recruitment of TC to the 43S PIC was found to stabilize the open conformation of PIC through the interplay between eIF2β-NTT with eIF1[112]. Additional work is required to understand the dynamic conformational rearrangement of PIC throughout the scanning process, from mRNA attachment up to the AUG recognition.

1.4.3 RNA helicases in scanning

During the initiation phase, the scanning 40S small subunit of the ribosome travels along the 5' UTR until it finds the start codon, where the 60S large subunit is recruited to commence the elongation. The majority of translation regulation occurs during the initiation step which is controlled by initiation factors and affected by various mRNA structures. Structurally, the m^7G cap of most mammalian mRNAs greatly increases the efficiency of translation initiation[289], while hairpins and other secondary structures could impair translation[290]. In particular, stem-loops near the 5' end of the cap inhibit translation initiation by preventing the association of 40S ribosomes[241]. While, equally stable stem-loops 50 to 60 nt downstream of the cap exhibited much less inhibitory effects on the scanning of the 40S ribosome[290], suggesting that helicases within the 40S small subunit are able to unwind the stable stems and access start codons within the structures[239]. This highlights the importance of RNA helicases in PIC scanning during the translation initiation. The DEAD and DExH box helicases involved in translation initiation unwind long duplexes in a

non-processive manner [291,292,293,294]. Their helicase activity is mediated by RNA binding in the presence of ATP, followed by ATP hydrolysis, and the cycles of release and rebinding. eIF4A and DHX29 are the two best-studied DEAD and DExH box RNA helicases involved in mammalian translation initiation.

1.4.3.1 eIF4A

eIF4A - Overview

eIF4A (DDX2) is a DEAD-box protein (contains Asp-Glu-Ala-Asp sequence in conserved motifs) that consists of two conserved helicase core domains, called RecA-like domains[295]. These two domains of eIF4A form a 'clamp'-like structure which switches from inactive 'open' conformation to active 'closed' conformation on transcripts upon the presence of ligand ATP[296] (Figure 1-10). As it recapitulates the characteristic properties of DEAD-box protein, eIF4A exhibits RNA-stimulated ATPase activity and an ATP-dependent RNA unwinding activity[297]. As the paradigm of DEAD-box protein, eIF4A is a non-processive RNA helicase which on its own only unwinds local secondary structure by binding to the duplex region and destabilizing the secondary structure instead of translocating along RNA molecules[292,293].

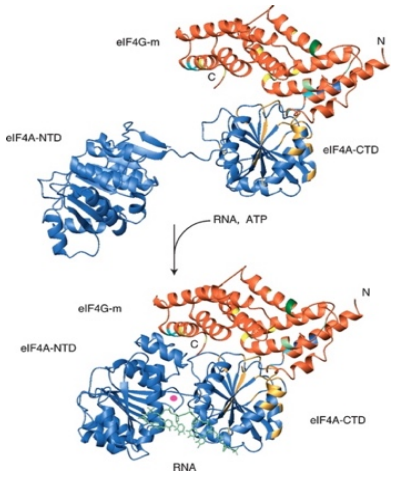

Figure 1-10 Structural models of eIF4A upon RNA and ATP binding.
Figure from Oberer, Marintchev, & Wagner, 2005

eIF4A has three isoforms in mammalian cells: eIF4A1, eIF4A2, and eIF4A3. eIF4A1 and eIF4A2 are involved in translation initiation[113] and share 91% similarity at the protein level. Although they are interchangeable in the eIF4F complex[298] and exhibit similar biochemical activities[299], they do not have identical functions in the translation initiation process.[300,301] Notably, eIF4A1 is more abundant as it is predominantly produced under normal, whereas synthesis of eIF4A2 is induced during growth arrest[302]. eIF4A3 is only 65% identical to eIF4A1 and eIF4A2, yet it is absent in the yeast genome. eIF4A1 and eIF4A2 mainly exist in the cytoplasm to participate in translation initiation[303], while eIF4A3 localizes in both the nucleus and cytoplasm to function in other processes such as mRNA decay[304,305] and exon junction complex after splicing[306]. eIF4A3 also failed to substitute for eIF4A1 in translation and thus is not likely to be involved in the control of initiation[304,307].

eIF4A1 – Potential roles in the initiation

eIF4A1 is the most abundant initiation factor in mammalian cells. More than 90% of eIF4A1 exist in a free form ($eIF4A_f$)[308]. The remaining 10% of eIF4A1 are part of the eIF4F complex ($eIF4A_c$)[53]. As one of the key components of the eIF4F complex, which facilitates the association of 40S ribosomes and mRNAs, eIF4A1 was suggested to remodel the cap proximal secondary structure via its helicase activity[113]. However, other researchers argued that eIF4A1 facilitates ribosome attachment by recruiting additional helicases like

DDX3 instead of directly destabilizing the 5' proximal structures[268]. The ATPase and helicase activity of individual eIF4A1 is low, like other non-processive DEAD-box helicases[55]. In the presence of its accessory factor eIF4B/eIF4H/eIF4G or in the complex of eIF4F, its ATPase and helicase activities were dramatically enhanced. The HEAT-1/MIF4G domain of eIF4G binds both RecA domains of eIF4A1 to induce a 'closed' active conformation and stimulates its ATP hydrolysis and RNA binding activity[309]. eIF4B/4H-mediates stimulation of eIF4A activities likely by a similar mechanism[310,69,58]. In contrast, the antioncogene Pdcd4 inhibits eIF4A activities by binding to both RecA domains in an 'open' inactive conformation[311], as well as competing with eIF4G and RNA binding for eIF4A[312]. In addition to directly promoting the activities of eIF4A, eIF4B/eIF4H/eIF4G each has at least one RNA-binding site, which increases the eIF4A affinity for RNA[313]. While eIF4G bias RNA unwinding activity of eIF4F toward duplexes with 5' overhang[314], eIF4H directed eIF4A to the loop region within the stem-loops rather than to the single-stranded overhangs revealed by sm-FRET[315]. These observations suggest that the accessory factors likely provide the substrate specificity for eIF4A. Last but not least, the eIF4A accessory factors could modulate eIF4A activities in response to different signaling pathways. Phosphoinositide 3-kinase (PI3K)-mTOR signaling pathway promotes the assembly of the eIF4F complex by increasing the availability of eIF4E for eIF4G binding[316] which stimulates eIF4A helicase activity[317]. eIF4B is another downstream effector of major signaling pathways

that control translation initiation such as PI3K/mTOR and mitogen-activated protein kinase (MARK) signaling pathway[318]. Induced S6 kinase phosphorylates Ser406 and Ser422 on eIF4B which promotes its interaction with eIF4A and eIF3[319,318,320]. However, the absence of eIF4H and the second eIF4A binding site on eIF4G[66], as well as the dispensability of the eIF4B RRM domain in yeast[321], suggest that their accessory function might be either dispensable or supplied by other factors. Despite a significant enhancement on its activity, eIF4F/eIF4A/eIF4B still shows limited capability in unwinding stem loops[322] (free energy of -19 kcal/mol *in vivo*[122] and -30 kcal/mol *in vitro*) and promotes initiation of mRNA with relatively short and unstructured 5'UTR.[323,324,177] A single-molecule assay reported that eIF4A displays processivity in the presence of eIF4G and eIF4H/eIF4B which rendered eIF4A the capability of unwinding structures of -150 kcal/mol[55]. The discrepancy in the unwinding activity of eIF4A might be due to different unwinding systems. For example, in the absence of eIF4E or the initiation complex, the binding site of eIF4A on the mRNA will not be limited. Nevertheless, the processivity mechanism used by eIF4A is still unclear.

Based on the observation that eIF4A1, in conjunction with eIF4B or 4H, unwinds RNA duplexes of either orientation (5' -> 3' or 3' -> 5') with similar rates[261,297], eIF4A1 is reported to act as a 'bidirectional helicase[325,326]. Of note, the eIF4F complex preferentially unwinds RNA duplexes with 5'-overhangs facilitated by RNA binding domain of eIF4G[314]. However, all of these studies

examined the unwinding directionality of eIF4A in the *in vitro* systems. Whether eIF4A exhibits 'bi-directional' unwinding *in vivo* and, even more intriguingly, whether its capability of 'bi-directional' unwinding impacts the directionality of the 40S ribosome migration on 5'UTR remain unknown.

In addition to its role in RNA unwinding, eIF4A has been recently proposed to promote ribosome loading onto mRNAs in universal initiation independent of unwinding activity. Both biochemistry assays and genomic-wide analysis revealed that eIF4A is required for mRNA recruitment by all mRNAs regardless of the structural complexity of the 5'UTR in yeast [59,60]. A recent study showed that the eIF4A: ATP binding status modulates the PIC attachment by altering the association between eIF3j and the 43S PIC in the mRNA channel[327]. The eIF4A accessory factor, eIF4B, has been suggested to stimulate ribosome attachment to the mRNA as well[328]. In sum, beyond the established functions of its ATPase and helicase activity, the non-conventional roles of eIF4A in the regulation of translation initiation need further exploration.

eIF4A - Localization

Localization of eIF4A1 relative to 43S PIC is under debate whether it locates at the leading edge or the trailing edge of the 40S subunit. Based on the Cryo-EM reconstruction, the possible location of eIF4G near the mRNA exit channel implied that eIF4A in complex with eIF4F is positioned near the E-site rather than the P-site[98]. Evidence from the crystal structure of human eIF4E

bound to human eIF4G also postulated that the eIF4F complex is located near the SSU platform, close to its E-site[124]. eIF4B was found at the 'head' on the solvent side of the 40S, close to the E-site in yeast by hydroxyl radical mapping[321] Another cryo-EM study revealed that eIF3d, a potent m^7G cap-binding factor like eIF4E, resides around the mRNA exit site, suggesting the position of eIF4F is toward the trailing edge of the 40S subunit[329]. Consistently, cross-linking of eIF3d to mRNA showed that it interacts with outbound mRNA and might participate in extending the exit portion of the mRNA channel[122]. A newly developed approach TCP-seq, which is advanced in detecting the footprint distribution of small ribosomal subunits specifically, uncovered that the additional regions covered by eIFs attached to PIC are on the 5' side of the 40S ribosome[330]. This evidence indicated that eIFs of large size, such as eIF4G or eIF4F, interact with the outbound mRNAs and sit near the exit site of the PIC.

However, other studies showed evidence which is at apparent variance with this notion and supported the location of eIF4A near the mRNA entry site. For example, cross-linking of highly structured incoming 5'UTR with associated eIFs revealed that eIF4A potentially locates in the ES6S region (around the entry site) of the 40S ribosome[331]. Whereas, minimal eIF4A1 was cross-linked to unstructured 5'UTR in the same region[331], suggesting that the presence of eIF4A in the ES6S region is mRNA-specific but not universal. Another supporting evidence came from the newly developed Ascorbate Peroxidase (APEX) proximity proteomics which placed eIF4A1 on the 3' side of the PIC, as

several eIF4A1-proximal proteins are located near the mRNA entry site[332]. However, the eIF4A1-distal proteins identified in the same study, such as RPS2 and RPS3, were in the vicinity of the entry site as well[332]. Moreover, other studies applied similar approaches suggested that eIF4A might be able to interact with both incoming and outbound mRNA. Proximity proteomics Biotin Identification (BioID) revealed that eIF4A1 is in proximity to eIF3 subunits situated in both entry (eIF3b and eIF3I) and exit (eIF3c) sites[333]. Based on the results from biochemistry studies and the structure analysis of human eIF4G, eIF4A could bind to human eIF4G in two different binding sites, which allow eIF4A to be placed at both exit and entry sites of the mRNA channel[334].

eIF4A - Pharmacological targeting

Dysregulation of protein synthesis is considered as a hallmark of cancer and linked to aberrant proliferation, survival, and angiogenesis. Malignant cells exhibit augmented activity on most of the components in the translation machinery. As the key enzymatic core of the eIF4F complex, eIF4A is one of the targets for therapeutic strategies[335,336]. The expression level of eIF4A is abnormally high in several malignancies, such as cervical cancer and leukaemia[337,338]. It has been reported that eIF4A helicase inhibitors can selectively kill tumor cells, and were well tolerated in non-tumor bearing mice without inducing appreciable toxicity[339,338]. High-throughput screens have

identified three natural products, Pateamine A, Hippuristanol, and Silvestrol (one of Rocaglamide family), that selectively target eIF4A[340,341,342].

Hippuristanol is an allosteric inhibitor of eIF4A. It binds to the C-terminal of eIF4A and inhibits RNA interaction by stabilizing eIF4A in a closed inactive conformation[341,343,344]. Almost all the Hippu binding residues are present in mammalian eIF4A1/2 and the yeast eIF4A homolog Tif1/2p but not among other DEAD-Box helicase family members. eIF4A1 and eIF4A2 showed similar sensitivities to Hippuristanol, whereas eIF4A3 required ~10-fold higher concentrations of compound to achieve equivalent inhibition, making it a selective eIF4A inhibitor[343]. Hippuristanol also inhibits eIF4A by abolishing the RNA-binding activity of both $eIF4A_f$ and $eIF4A_c$[344]. As a consequence, it inhibits eIF4A helicase and ATPase activities without affecting the ATP-binding by eIF4A[341,343].

Pateamine A serves as chemical inducers of dimerization (CID) which force engagement of eIF4A and RNA, resulting in sequestration of $eIF4A_f$ from eIF4F complex[340,345,346]. It has been reported that inhibition of eIF4G: eIF4A association by Pateamine A also lead to a reduction in eIF4F formation and impaired ribosome attachment on mRNA[345]. Despite a decreased eIF4F level, the enhanced RNA binding affinity of $eIF4A_c$ by Pateamine A causes a several-fold increase in both ATP hydrolysis activity and helicase activity of $eIF4A_c$[345]. In terms of its specificity, Pateamine A has been reported to inhibit the Nonsense-mediated mRNA decay (NMD) through eIF4A3, suggesting it targets

all three isoforms of eIF4A under similar concentration. However, it is also likely that the inhibition of NMD was secondary to the induction of stress granules (SGs) followed by disruption of eIF4A1 or arrested translation[347].

Silvestrol, one of the best-studied rocaglates, was first discovered from the same high-throughput screen that identified Hippuristanol[342]. In contrast to Hippu that disrupts eIF4A: RNA association, Rocaglates, like Pateamine A, have been reported to stabilize RNA-bound eIF4A, promote both ATP hydrolysis activity and helicase activity of eIF4A, and reduce eIF4A availability for eIF4F complex[342,348]. However, recent studies indicated that rocaglates (specifically Rocaglamide A) impacts translation initiation by a different mechanism where Rocaglamide A (RocA) bound eIF4A clamps on the polypurine motifs in the mRNAs and blocks the scanning 43S PIC[349]. Revealed by the crystal structure of eIF4A: RocA: RNA, C-ring in the rocaglate backbone is important for eIF4A binding, while the A- and B- rings determine the purine selectivity of eIF4A: RocA[350]. Early identified and characterized rocaglate-resistant allele (F163L) in murine cells is consistent with this crystal structure, where F163 was found to interact with C-ring [348,350].

These three compounds have been characterized extensively as inhibitors of eIF4A. Currently, they are explored as potential chemotherapies, and have all demonstrated preclinical efficacy[351]. The major features of each compound are highlighted in Figure 1-11. A number of new small molecules have been recently discovered as the alternative eIF4A targeting modulators

whose effects and biological mechanisms will be revealed with future investigation[352]. Nevertheless, characterizations of these regulators targeting eIF4A and their different impacts on eIF4A or translation initiation will provide more insights to appreciate how eIF4A mediates regulation of the translation initiation.

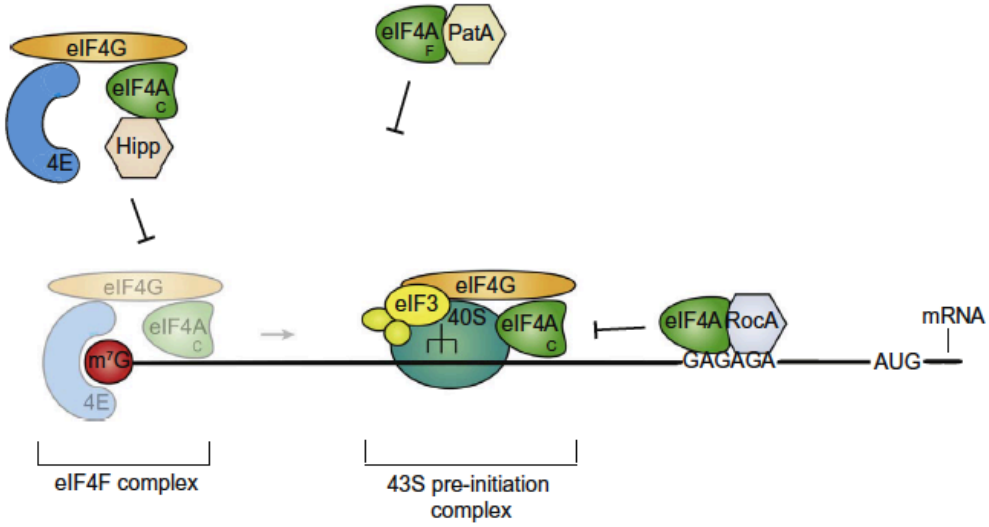

Figure 1-11 Mechanisms of three eIF4A1 regulators.
Hippuristanol (Hippu) binds to the C-terminal of eIF4A and specifically inhibits its RNA binding, ATPase and helicase activity by inducing an inactive conformation. Pateamine A (PatA) stimulates RNA binding activity of free eIF4A ($eIF4A_F$), leading to the sequestration of $eIF4A_F$ from the eIF4F complex. Of the family of Rocaglates, Rocaglamide A (RocA) clamps eIF4A on the polypurine sequences in the mRNAs, hereby blocking the ribosomal scanning.

1.4.3.2 DHX29

DHX29 is a ubiquitously expressed cytoplasmic DEAH-box RNA helicase, found in both yeast and mammals[353]. Different from eIF4A, DHX29 not only contains the conserved sequence Asp-Glu-Ala-His (DEAH), but also carries multiple accessory domains with it to form a complex regulatory network[354]. DHX29 plays a critical role in the PIC scanning during the translation initiation. Silencing of DHX29 has shown reductions in global translation, cell proliferation

and tumoregenesis[353]. Consistently, ribosome profiling studies have shown that the absence of DHX29 caused dissociation of polysomes but an increase in the 80S monosome peak – 'hallmarks' of translation initiation inhibition[353]. Given its major function in the initiation, DHX29 is mostly found in the 40S fraction[353]. It has been reported that DHX29 protects 18S rRNA residues at the top of h16 in 40S from the chemical modification[178]. A dsRNA binding motif in the DHX29 was found critical for its binding to the 43S PIC[354]. DHX29 also contacts eIF3a and eIF3b, and these interactions are important for DHX29's activity during scanning[355]. In agreement with these findings, the recent Cryo-EM analysis of the 43S has found that DHX29 resides at the mRNA entry site[356] (Figure 1-12).

Both *in vitro* and *in vivo* studies have demonstrated that DHX29 is essential for the translation initiation of mRNA sequences with highly structured 5' UTRs[178,263,353]. Due to its position along the leading edge of the mRNA channel, DHX29 was proposed to unwind the stable stems before structured sequence threading into the mRNA channel. Similar as eIF4A, DHX29 alone is a non-processive RNA helicase which requires NTP hydrolysis to function. Whereas, its direct association with the 40S ribosome,

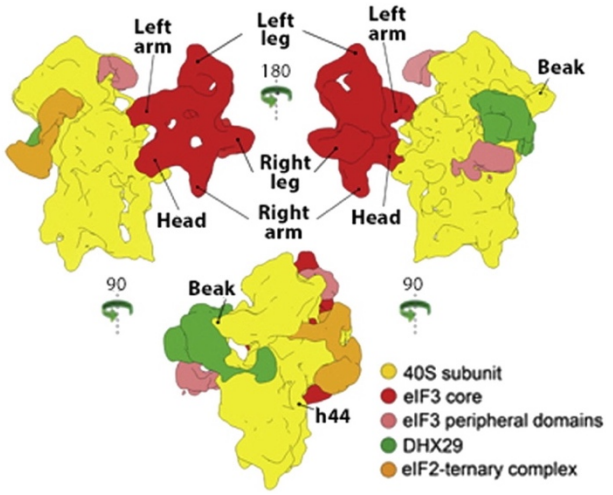

Figure 1-12 Mammalian 43S PIC with DHX29.
Figure from Hashem et al., 2013

but not RNA-binding, enhances its functionality by stimulating the processivity and NTPase activity[178,354].

Recently, DHX29 was suggested to act through remodeling the 40S ribosomal subunits instead of direct RNA unwinding to remove the incoming stable structures. A cross-linking study found that DHX29 produces only weak contacts with the mRNA, and the low efficiency of cross-linking between DHX29 and mRNA sequences suggested that mRNAs in the channel do not pass through the helicase[355]. The addition of DHX29 was found to dissemble the 48S complex which either contains an intact stem in the A-site[262] or pre-assembles on certain viral (CrPV- and CSFV-like HCV IRESs) IRESs[178], indicating that DHX29 could cause conformational changes in 40S subunits.

Beyond the general functions in RNA structure unwinding, DHX29 has also been suggested to participate in the start codon selection from *in vitro* systems. The addition of DHX29 to an *in vitro* system decreased the leaking scanning of upstream AUGs, regardless of the sequence context[357]. In conclusion, despite its role as a key translation initiation factor, the mechanisms of how DHX29 promotes translation and the unwinding of stable stems remains elusive.

1.4.4 Release of the cap from scanning 43S PIC

Several intriguing questions regarding the scanning process remain debatable for decades. The answer to the question of whether m^7G cap remains

attached to or disengages from the scanning 43S PIC is unclear. Two exclusive models were proposed: 'cap-tethered' or 'cap-severed.' In the cap-tethered model, the m^7G cap at the 5' end of the mRNA keeps in contact with the scanning 43 PIC until initiation is completed. After the formation of the 80S, the m^7G cap, as well as other eIFs, dissociate from the ribosome together. This model was once favored for its simplicity as there would be no additional mechanisms required to explain cap detachment from the scanning 43S PIC. However, evidence implied that multiple 43S PIC could transverse single 5'UTR at the same time indicated the interaction between cap and scanning ribosome must be disrupted to allow multiple rounds of 43S PIC loading. For example, reporter mRNAs with increasing 5'UTR length from 43 nt to 1770 nt showed a similar rate of protein expression, suggesting a steady-state initiation frequency[179]. An early *in vitro* biochemical study showed that edeine, a drug that impairs start codon recognition and enables infinite scanning of PIC, significantly induced sequence complexity of the 40S protected fragment, indicating increased 40S binding to a single messenger under edeine treatment[358]. A recently developed translation complex profile sequencing (TCP-seq) method modified canonical ribosome profiling (Ribo-seq) by adding formaldehyde cross-linking and a second round of sucrose gradient separation to enrich footprints of small ribosomal subunit[359]. Correlation between 5'UTR length and 40S occupancy uncovered by TCP-seq showed no negative correlation, consistent with the previous finding[179]. Moreover, the absence of

footprint density near the 5' end in the TCP-seq is another evidence against the 'cap-tethered' model in which a magnified ribosome dwell time at the cap relative to other regions is expected to. Interestingly, TCP revealed a subset of start-codon-associated footprints with extended 30nt length at their 5' end, suggesting the presence of queue-up 40S[330]. These observations of multiple rounds of ribosome loading strongly favored the 'cap-severed' mode.

As the currently accepted model, 'cap-severed' scanning process remained obscure in details such as the location of detachment. Whether cap dissociated from 43S PIC by itself or in the form of cap-eIF4E complex is uncertain according to the current knowledge. The fast dissociation rate of eIF4E from the cap *in vitro* suggests that eIF4E might be released the cap during the scanning process[360]. Biochemical observations supported this model by showing the dramatically reduced cross-linking between m^7G and eIF4E at the late stage of initiation[118]. Another suggestive evidence is that replacement of m^7G cap with non-functional cap analogs impaired de novo initiation but not ongoing translation[361]. Early studies showed that phosphorylation of eIF4E S209 by MNK1, a kinase binding to the eIF4G CTD HEAT domain, decreases eIF4E's affinity for the cap without affecting its binding to 4E-BP[362,363], indicating an MNK-mediated eIF4E release from the cap. These results suggest that m^7G cap detaches from the scanning 43S PIC upon the complete accommodation of mRNA, whereas eIF4E remains bound to the ribosome and does not recycle as efficiently as m^7G does for multiple rounds of ribosome loading.

1.4.5 Directionality of scanning

Another major open question regarding the mechanistic aspect of scanning concerns the directionality of PIC motion. The classic model ribosome scanning model depicts a 5' -> 3' linear, base-by-base motion of PIC complex. Consistent with this, an inverse relationship between 5'UTR length and initial translation rate was observed in cell extracts[179,364]. The linear correlation between the scanning time and the 5'UTR length was demonstrated as the evidence of unidirectional movement of scanning ribosome[364]. Also, where multiple AUGs appear in an mRNA, the first AUG from the 5' end will be chosen as the initiation site[365,366], suggesting a strong 5'-polarity in AUG recognition.

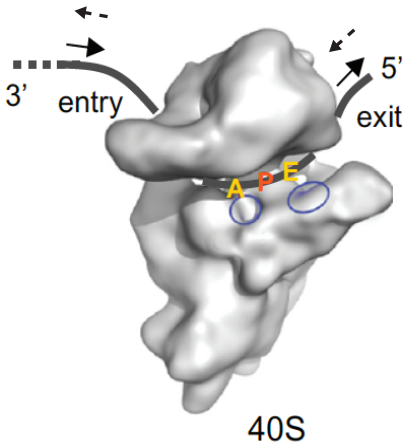

Figure 1-13 Directionality of PIC scanning.
Model of mRNA passing mRNA channel of 40S with A-, P- and E-sites highlighted. The solid and dashed arrows indicate the 5' to 3' and 3' to 5' movements of PIC along the 5'UTR respectively.

However, most of these observations demonstrated a net directionality in the process but not the dynamics of the scanning. It has also been proposed that the scanning process, rather than exclusively uni-directional, may consist of forward (5' to 3') thrusts alternating with limited relaxation in the reverse direction (Figure 1-13). The fact that initiation decisions involving two closely spaced AUGs are not sequential with 5' polarity due to a poor context sequence suggests that ribosome scanning involves a net 5' -> 3' movement with small-amplitude back-and-forth oscillations[367]. In contrast to the classic scanning

model depicting a linear base-by-base movement of the ribosome, the oscillation model permits a non-linear bi-directional migration of the scanning ribosome. Bidirectional scanning allows the start codon recognition within the 'blind-spot' in the scenario of 'slot-in' ribosome attachment (described in Section *1.4.4*). In accordance with this, initiation efficiency at the 5' proximal AUG is reduced by the presence of a nearby downstream AUG[367] in the cells. Consistently, earlier *in vitro* studies demonstrated that the first AUG that resides within 12 nucleotides downstream of the cap is less favored by the scanning ribosome despite the favorable sequence context[184,366]. Although other studies reported the opposite results supporting the classic uni-directional scanning[368], the non-linear scanning mechanism has been suggested in different modes of initiation. Retrograde scanning of the 40S has been proposed in IRES dependent initiation in *Halastavi arva virus*.[369,370] An early study has reported that reinitiation ribosomes reach neighboring initiation codons by lateral diffusion along the mRNA. This retrograde movement of the reinitiating ribosome has a radius of action of more than 40 nucleotides in the eubacteria system[371]. In the eukaryotic system, the bidirectional scanning has also been documented for both post-terminating 80S ribosomes and recycled 40S during translation reinitiation[372,373]. Even though, whether the observed bidirectional scanning is a rare event that happens in an mRNA-specific manner or a ubiquitous action of ribosome that occurs during the canonical scanning process remains unknown.

A more profound question is how the net 5' -> 3' movement of the PIC is achieved on untranslated regions of mRNAs. The canonical translation initiation requires ATP for both mRNA attachment[327] and the scanning process[258]. As the only ATP-binding and ATP-hydrolyzing initiation factors of PIC, the RNA helicases involved in the initiation process (described in *1.4.3*) have been implicated in supplying energy for the migration of the PIC along the 5'UTR. Both DHX29 and DDX3 (or Ded1 in yeast) are found at the leading edge of the 40S ribosome and mainly function on the highly structured 5'UTR to destabilize the structure of incoming mRNAs[374,329,178], which promotes the forward scanning motion of the ribosome. Given an uneven distribution of secondary structure across the entire length of 5' UTRs, it is unwise to assume the same mechanism for all mRNAs, including those with unstructured 5'UTRs. Thus, unwinding the incoming secondary structures by DHX29 or DDX3 ahead of the ribosomes cannot be the only determinant of scanning directionality.

Compared to DHX29 and DDX3X, eIF4A shows a limited capacity of unwinding secondary structures[322]. Intriguingly, as an RNA helicase with high abundance in the cells[375], eIF4A is indispensable for most endogenous transcripts regardless of the degree of structures[60], indicating a more generic role of eIF4A in the translation initiation or scanning process. Therefore, it has been suggested that eIF4A is the key factor for the directionality of the general scanning motion. Two potential roles of eIF4A were proposed to explain the directional scanning of PIC on 5'UTRs: 1. eIF4A localizes in front of the 40S

ribosome and stabilizes the incoming mRNAs through association/unwinding/dissociation cycles[69,376,377,331]; 2. Repeated eIF4A binding to the outbound mRNAs near the exit site prevents the possible backward (3' -> 5') movement of the 40S ribosome[334,98,378]. Both models presume that eIF4A guarantees the forward scanning of PIC on mRNAs. In line with this mechanism, the studies on bidirectional scanning in the context of IRES-mediated initiation and re-initiation reported that the presence of eIF4F complex represses retrograde scanning and imposes 5'-> 3' forward scanning[369,372]. However, both studies applied the *in vitro* toe-printing assay, which might be insufficient to reflect the scanning events in the cells where eIF4A is much more abundant and could exert disparate functions upon binding to different factors. For example, eIF4A bound by eIF4B or 4H, but not eIF4G, can unwind RNA duplex in either direction at the same rate[261,55]. Moreover, there is lack of solid evidence to support the assumption that eIF4A would take different actions based on its position relative to the 40S ribosome. Alternatively, the possibility for eIF4A to unwind secondary structure and stabilize single-strand RNAs near the exit site of the 40S ribosome, or clamp on mRNAs like a roadblock in front of the ribosome could not be eliminated. In either of these cases, eIF4A can block the forward scanning while promoting the backward scanning of PIC on the mRNAs by itself or in conjunction with other binding proteins, The observation that RocA treatment, which clamps eIF4A onto

polypurine sequences, inhibited 5' -> 3' scanning on the 5'UTR bearing downstream polypurine motif is in consistence with this idea[349,350].

Overall, there is no absolute certainty on the molecular mechanism of the scanning directionality. Both duplex separation and active mRNA binding might occur on either side of the 40S ribosome during this highly dynamic process. Mammalian 5'UTRs are longer and more structured, compared to yeast 5'UTRs[379]. Let alone, the scanning ribosomes in mammals involve more factors (eIF3[69]) and cofactors (eIF4H[380]), which could take additional actions (like a second binding site for eIF4A in eIF4G[66] and eIF4B dimerization[380]). The identification of the determinants of scanning directionality is expected to be much more complex in the higher eukaryotes.

1.4.6 Non-canonical scanning

Reinitiation

As described in Section *1.4.1.2*, approximately 40% of eukaryotic mRNAs harbor upstream (u)ORFs[199]. Ribosome profiling experiments uncovered widespread translation at many of these uORFs[202]. In these cases, the ribosomes show the ability to resume scanning following termination at uORFs. Post-termination events at uORF stop codons probably proceed conventionally, with the splitting of 80S ribosomes and the release of 60S subunits and eRFs, while some 40S subunits could remain attached on the mRNA and resume scanning. Although at this stage, 40S subunits are

incompetent for start codon recognition due to lack of TC, they are capable of scanning, during which a new TC can be acquired. This process is termed reinitiation. This resumed scanning after translation termination in the absence of TC in the first place has some difference in mechanistic details compared to the canonical scanning. A subset of eIFs have been suggested to be essential for reinitiation. Reinitation at downstream ORF was found highly dependent on the distance between the ORFs[381]. As TC availability probably determines how far 40S subunits migrate before acquiring one, it has been suggested that a minimal scanning time was required for TC recruitment before the downstream AUG is encountered[382]. Therefore, eIF2, which is responsible for delivering TC to the scanning ribosomes, is considered as one of the most important regulators of reinitiation. The principle of eIF2-mediated reinitiation first explained the translational control of yeast GCN4 mRNA (and mammalian ATF4 mRNAs) by multiple uORFs[383]. Interestingly, it was reported that reinitiation efficiency is inversely correlated with the length of the uORF[381]. Also, stable RNA secondary structures in the uORF negatively influenced reinitiation potential as well[384]. These observations suggest that increased time taken by the 80S ribosome to finish the translation of uORF, due to either increased length of uORFs or stable structures in the uORF, might ultimately lead to the loss of reinitiation potential. Hence, resumption of scanning is restricted to uORFs that are translated in a short period of time. This limitation could be best explained by the high dependence of reinitiation on the eIFs, which remain

attached to the post-termination ribosomes. Therefore, eIFs, which stay with the ribosome for a very short period of time following the start of elongation, were implicated in the reinitiation process[385]. To date, it is unclear when eIF3 and eIF4 group factors dissociate from the ribosomes. It has been suggested that eIF3 (maybe together with eIF4G) might remain attached to the elongating ribosomes for some time, which could be an alternative mechanism to explain the reinitiation after short uORFs[99,386].

Leaky Scanning

Although for some cellular transcripts with at least one uORFs, annotated AUG codons are recognized by the reinitiation mechanism, other annotated AUG codons are reached by the scanning ribosomes via bypassing the uAUGs. This is called leaky scanning. Different from the canonical scanning mechanism whereby ribosomes will initiate at the first AUG codon, leaky scanning results in bypassing the first AUG while selecting a start codon farther downstream. Since the transition from the scanning-competent mode into the arrested mode at the start codon must be a relatively fast process to be compatible with scanning speeds of ~10 nt/s[179,364], sequence context adjacent to the start codon was thought to affect the speed of scanning. In alignment with this notion, leaky scanning was first reported based on the observation that an upstream start codon within a 'weak' Kozak sequence context is unfavored and skipped by the scanning ribosome. Recently, the crystal structures of 48S PIC revealed that

the interaction between eIF1A with the nucleotide next to the P-site (+4 position) is likely to contribute the pausing during scanning[387]. eIF2α interactions with nucleotides in positions -2, -3 were thought to be critical for the start codon recognition[279,388]. These findings suggest that signals from the sequence context slow down the scanning ribosome's speed via interactions with eIFs. Notably, sequence context is not the only determinant of leaky scanning, as 5' proximal AUGs that deviate sufficiently from the optimum context can be bypassed in the event of leaky scanning. As one of the key determinants of the start codon recognition, initiation factors have been implicated in the regulation of leaky scanning. eIF1 and eIF1A were shown to promote leaky scanning by stabilizing 'open', scanning-competent conformation of the ribosome[94]. Accordingly, a genome-wide analysis observed that low eIF1 levels promote translation initiation at upstream start codons, although embedded in a suboptimal nucleotide context[389]. The reduction of TC availability evoked by eIF2α phosphorylation enables rescanning ribosome bypassing the uORFs[204]. eIF3, eIF5 and eIF5B were also implicated in the modulation of start codon selection[100,390,391]. Interestingly, RNA helicase DHX29 has been reported to reduce leaky scanning regardless of its nucleotide context through interaction with eIF1A[357]. More comprehensive studies on these initiation factors could help us to better appreciate the mechanism underlying the leaky scanning under different conditions.

Scanning-independent initiation

A non-canonical translation of mRNAs bearing short 5′ UTR with a median length of 12 nt is governed by the TISU (translation initiator of short 5′ UTRs) element (SAASATGGCGGC) operates in higher eukaryotes[392]. The TISU element was discovered as an important transcription regulatory element by a bioinformatics search for elements overrepresented in the core promoter region. TISU locates downstream to the transcription start site (TSS) from position +5 up to +30 relative to the TSS and is therefore very proximal to the m^7G cap of the mature mRNAs. TISU is present in ~4% of all genes, preferentially in those with 'housekeeping' functions[393]. Like the Kozak sequence, the TISU element contains an AUG codon flanking with unique nucleotides. The AUG of the TISU element serves as the exclusive translation start site, while the flanking nucleotides are very important for resistance to leaky scanning and accurate translation from the 5' proximal AUG[394]. Translation mediated by this element is reported to be cap-dependent, as evidenced by the inhibitory effect of non-functional cap analog incorporation, 5' proximal structures insertion, and disrupted eIF4E:eIF4G interactions on TISU mediated-translation[190]. In agreement with this notion, results from a toe-printing assay indicated that the TISU mediated translation requires eIF4F[395]. However, TISU-directed translation is unaffected by inhibition of the RNA helicase eIF4A suggesting that TISU mediated translation is cap-dependent but acts without scanning[190]. Mechanism of translation initiation driven by TISU elements

requires the release of the eIF4F complex following AUG recognition to prevent the overlapping between the incoming 48S complex and eIF4F complex[395]. A recent study revealed eIF4E indirectly interacts with the eIF1-binding site of eIF4G, suggesting the competing nature of eIF4G binding between eIF4E and eIF1. This finding underlined the rearrangement of 43S PIC from cap-binding mode to the scanning mode during the canonical scanning-dependent initiation, which is absent in the TISU-mediated initiation due to the release of eIF4F complex[284]. And the scanning-arrest state is further stabilized by the sequence-specific contacts with A-site ribosomal protein S3 and S10e enhanced by the presence of eIF1A[394].

Another well-known case of cap-dependent but scanning-free translation is for cellular mRNAs encoding histone proteins, particularly histone 4 (H4) mRNA, which also bear very short 5'UTRs and no poly(A) tail. Interestingly, the m^7G cap is required for H4 mRNA translation but not eIF4E binding on mRNA. The ORF of H4 mRNA contains two structural elements, one of which directly recruits eIF4E in the complex of eIF4F without m^7G as it is sequestered by another structure. Following its binding, eIF4F facilitates ribosome attachment at the internal site and a quick relocation to the region in the vicinity of AUG through the release of m^7G cap from another structure region by eIF4A helicase and the interaction between eIF4E and m^7G cap[396]. This Cap-assisted internal initiation of H4 mRNA is promoted by the purine-rich region of the H4 mRNA, which is complementary to the UUUC of helix h16 of 18S rRNA at the entry site

of the mRNA channel. This element assists in placing the start codon very close to the P-site almost immediately upon its recruitment through the cap-binding complex[397].

1.5 Conclusions

The field of mRNA translation has made tremendous progress in understanding molecular mechanism underlying the canonical eukaryotic translation initiation over the past decade. Biochemical, structural, and genetic studies have enabled the elucidation of the precise nature of interplay among ribosomes, initiation factors, and mRNAs. Nevertheless, the specific interactions that occur during the transition stages and the dynamic actions taken by PIC per se along the pathway remain obscure. The mechanistic details regarding the mRNA recruitment and PIC scanning have yet been fully appreciated. Given the recent progress in single-molecule kinetic analyses and high-throughput sequencing techniques, intensive dissection of these two intricate processes can be anticipated. A thorough understanding of the translation initiation in eukaryotes will ultimately provide insights into the development of new therapeutic strategies for human diseases.

REFERENCE

1. Fry, M. *Landmark experiments in molecular biology*. *Landmark Experiments in Molecular Biology* (2016). doi:10.1016/c2014-0-01907-7
2. Avery, O. T., Macleod, C. M. & Mc Carty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction or transformatioby A Desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* (1994). doi:10.1084/jem.179.2.379
3. HERSHEY, A. D. & CHASE, M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* (1952). doi:10.1085/jgp.36.1.39
4. Hinnebusch, A. G. & Lorsch, J. R. The mechanism of eukaryotic translation initiation - new insights and challenges.pdf. *Cold Spring Harb. Perspect. Biol.* **4**, 1–25 (2012).
5. Hinnebusch, A. G., Ivanov, I. P. & Sonenberg, N. Translational control by 5 ′ - untranslated regions of eukaryotic mRNAs. *Science (80-.).* **352**, 1413–1416 (2016).
6. Hinnebusch, A. G. The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* **83**, 779–812 (2014).
7. Dever, T. E. & Green, R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb. Perspect. Biol.* (2012). doi:10.1101/cshperspect.a013706
8. Andersen, G. R., Valente, L., Pedersen, L., Kinzy, T. G. & Nyborg, J. Crystal structures of nucleotide exchange intermediates in the eEF1A-eEF1Bα complex. *Nat. Struct. Biol.* (2001). doi:10.1038/88598
9. Gutierrez, E. *et al.* eif5A promotes translation of polyproline motifs. *Mol. Cell* (2013). doi:10.1016/j.molcel.2013.04.021
10. Melnikov, S. *et al.* Molecular insights into protein synthesis with proline residues. *EMBO Rep.* (2016). doi:10.15252/embr.201642943
11. Melnikov, S. *et al.* Crystal Structure of Hypusine-Containing Translation Factor eIF5A Bound to a Rotated Eukaryotic Ribosome. *J. Mol. Biol.* (2016). doi:10.1016/j.jmb.2016.05.011
12. Shin, B. S. *et al.* Amino acid substrates impose polyamine, eIF5A, or hypusine requirement for peptide synthesis. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx532
13. Semenkov, Y. P., Rodnina, M. V. & Wintermeyer, W. The 'allosteric three-site model' of elongation cannot be confirmed in a well-defined ribosome system from Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* (1996). doi:10.1073/pnas.93.22.12183
14. Uemura, S. *et al.* Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* (2010). doi:10.1038/nature08925
15. Chen, C. *et al.* Allosteric vs. spontaneous exit-site (E-site) tRNA dissociation early in protein synthesis. *Proc. Natl. Acad. Sci. U. S. A.* (2011). doi:10.1073/pnas.1106999108
16. Frolova, L. Y. *et al.* Mutations in the highly conserved GGQ motif of class I polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA* (1999). doi:10.1017/S135583829999043X
17. Nürenberg, E. & Tampé, R. Tying up loose ends: Ribosome recycling in eukaryotes and archaea. *Trends in Biochemical Sciences* (2013). doi:10.1016/j.tibs.2012.11.003
18. Proshkin, S., Rachid Rahmouni, A., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science (80-.).* (2010). doi:10.1126/science.1184939
19. Burmann, B. M. *et al.* A NusE:NusG complex links transcription and translation. *Science (80-.).* (2010). doi:10.1126/science.1184953
20. Demo, G. *et al.* Structure of RNA polymerase bound to ribosomal 30S subunit. *Elife* (2017). doi:10.7554/eLife.28560

21. Kohler, R., Mooney, R. A., Mills, D. J., Landick, R. & Cramer, P. Architecture of a transcribing-translating expressome. *Science (80-.).* (2017). doi:10.1126/science.aal3059
22. Bakshi, S., Siryaporn, A., Goulian, M. & Weisshaar, J. C. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Mol. Microbiol.* (2012). doi:10.1111/j.1365-2958.2012.08081.x
23. Jin, D. J., Cagliero, C. & Zhou, Y. N. Role of RNA polymerase and transcription in the organization of the bacterial nucleoid. *Chemical Reviews* (2013). doi:10.1021/cr4001429
24. Espah Borujeni, A. & Salis, H. M. Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. *J. Am. Chem. Soc.* (2016). doi:10.1021/jacs.6b01453
25. Chang, B., Halgamuge, S. & Tang, S. L. Analysis of SD sequences in completed microbial genomes: Non-SD-led genes are as common as SD-led genes. *Gene* (2006). doi:10.1016/j.gene.2006.01.033
26. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* (2014). doi:10.1016/j.cell.2014.02.033
27. Schrader, J. M. *et al.* The Coding and Noncoding Architecture of the Caulobacter crescentus Genome. *PLoS Genet.* (2014). doi:10.1371/journal.pgen.1004463
28. Furuichi, Y. & Shatkin, A. J. Caps on Eukaryotic mRNAs. in *Encyclopedia of Life Sciences* (2007). doi:10.1002/9780470015902.a0000891.pub2
29. Furuichi, Y., LaFiandra, A. & Shatkin, A. J. 5′-Terminal structure and mRNA stability. *Nature* (1977). doi:10.1038/266235a0
30. Shimotohno, K., Kodama, Y., Hashimoto, J. & Miura, K. I. Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. *Proc. Natl. Acad. Sci. U. S. A.* (1977). doi:10.1073/pnas.74.7.2734
31. Hsu, C. L. & Stevens, A. Yeast cells lacking 5'-->3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Mol. Cell. Biol.* (1993). doi:10.1128/mcb.13.8.4826
32. Sonenberg, N., Morgan, M. A., Merrick, W. C. & Shatkin, A. J. A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. *Proc. Natl. Acad. Sci. U. S. A.* (1978). doi:10.1073/pnas.75.10.4843
33. Gick, O., Krämer, A., Keller, W. & Birnstiel, M. L. Generation of histone mRNA 3′ ends by endonucleolytic cleavage of the pre-mRNA in a snRNP-dependent in vitro reaction. *EMBO J.* (1986). doi:10.1002/j.1460-2075.1986.tb04362.x
34. Proudfoot, N. J. & Brownlee, G. G. 3′ Non-coding region sequences in eukaryotic messenger RNA. *Nature* (1976). doi:10.1038/263211a0
35. Viphakone, N., Voisinet-Hakil, F. & Minvielle-Sebastia, L. Molecular dissection of mRNA poly(A) tail length control in yeast. *Nucleic Acids Res.* (2008). doi:10.1093/nar/gkn080
36. Tarun, S. Z. & Sachs, A. B. Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *EMBO J.* (1996). doi:10.1002/j.1460-2075.1996.tb01108.x
37. Wells, S. E., Hillner, P. E., Vale, R. D. & Sachs, A. B. Circularization of mRNA by eukaryotic translation initiation factors. *Mol. Cell* (1998). doi:10.1016/S1097-2765(00)80122-7
38. Archer, S. K., Shirokikh, N. E., Hallwirth, C. V., Beilharz, T. H. & Preiss, T. Probing the closed-loop model of mRNA translation in living cells. *RNA Biol.* (2015). doi:10.1080/15476286.2015.1017242
39. Gallie, D. R. The cap and poly(A) tail function synergistically to regulate mRNA

translational efficiency. *Genes Dev.* (1991). doi:10.1101/gad.5.11.2108
40. Adivarahan, S. *et al.* Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. *Mol. Cell* (2018). doi:10.1016/j.molcel.2018.10.010
41. Khong, A. & Parker, R. MRNP architecture in translating and stress conditions reveals an ordered pathway of mRNP compaction. *J. Cell Biol.* (2018). doi:10.1083/jcb.201806183
42. Kapp, L. D. & Lorsch, J. R. The Molecular Mechanics of Eukaryotic Translation. *Annu. Rev. Biochem.* **73**, 657–704 (2004).
43. Gebauer, F. & Hentze, M. W. Molecular mechanisms of translational control. *Nat. Rev. Mol. Cell Biol.* **5**, 827–835 (2004).
44. Hershey, J. W. B. & Merrick, W. C. Pathway and mechanism of initiation of protein synthesis. in *Translational control of gene expressionranslational control of gene expression* (eds. Sonenberg, N., Hershey, J. W. B. & Mathews) 33–88 (Cold Spring Harbor laboratory press, 2000).
45. Von Der Haar, T., Gross, J. D., Wagner, G. & McCarthy, J. E. G. The mRNA cap-binding protein eIF4E in post-transcriptional gene expression. *Nature Structural and Molecular Biology* (2004). doi:10.1038/nsmb779
46. Gross, J. D. *et al.* Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell* (2003). doi:10.1016/S0092-8674(03)00975-9
47. Ptushkina, M. *et al.* Cooperative modulation by eIF4G of eIF4E-binding to the mRNA 5' cap in yeast involves a site partially shared by p20. *EMBO J.* (1998). doi:10.1093/emboj/17.16.4798
48. Yanagiya, A. *et al.* Requirement of RNA Binding of Mammalian Eukaryotic Translation Initiation Factor 4GI (eIF4GI) for Efficient Interaction of eIF4E with the mRNA Cap. *Mol. Cell. Biol.* (2009). doi:10.1128/mcb.01187-08
49. Borman, A. M. Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end. *Nucleic Acids Res.* (2000). doi:10.1093/nar/28.21.4068
50. O'Leary, S. E., Petrov, A., Chen, J. & Puglisi, J. D. Dynamic recognition of the mRNA cap by Saccharomyces cerevisiae eIF4E. *Structure* **21**, 2197–2207 (2013).
51. Wei, C. C., Balasta, M. L., Ren, J. & Goss, D. J. Wheat germ poly(A) binding protein enhances the binding affinity of eukaryotic initiation factor 4F and (iso)4F for cap analogues. *Biochemistry* (1998). doi:10.1021/bi9724570
52. Pause, A. *et al.* Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5′-cap function. *Nature* (1994). doi:10.1038/371762a0
53. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* (2014). doi:10.1038/nmeth.2834
54. Avdulov, S. *et al.* EIF4E threshold levels differ in governing normal and neoplastic expansion of mammary stem and luminal progenitor cells. *Cancer Res.* (2015). doi:10.1158/0008-5472.CAN-14-2571
55. García-García, C., Frieda, K. L., Feoktistova, K., Fraser, C. S. & Block, S. M. Factor-dependent processivity in human eIF4A DEAD-box helicase. *Science (80-.).* (2015). doi:10.1126/science.aaa5089
56. Korneeva, N. L., First, E. A., Benoit, C. A. & Rhoads, R. E. Interaction between the NH2-terminal domain of eIF4A and the central domain of eIF4G modulates RNA-stimulated ATPase activity. *J. Biol. Chem.* (2005). doi:10.1074/jbc.M406168200
57. Rogers, G. W., Komar, A. A. & Merrick, W. C. eIF4A: The godfather of the DEAD box helicases. *Progress in Nucleic Acid Research and Molecular Biology* (2002). doi:10.1016/s0079-6603(02)72073-4

58. Rozovsky, N., Butterworth, A. C. & Moore, M. J. Interactions between eIF4AI and its accessory factors eIF4B and eIF4H. *RNA* (2008). doi:10.1261/rna.1049608
59. Yourik, P. *et al.* Yeast eIF4A enhances recruitment of mRNAs regardless of their structural complexity. *Elife* (2017). doi:10.7554/eLife.31476
60. Sen, N. D., Zhou, F., Ingolia, N. T. & Hinnebusch, A. G. Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. *Genome Res.* **25**, 1196–1205 (2015).
61. Hershey, P. E. C. *et al.* The cap-binding protein eIF4E promotes folding of a functional domain of yeast translation initiation factor eIF4G1. *J. Biol. Chem.* (1999). doi:10.1074/jbc.274.30.21297
62. Imataka, H., Gradi, A. & Sonenberg, N. A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *EMBO J.* (1998). doi:10.1093/emboj/17.24.7480
63. Lamphear, B. J., Kirchweger, R., Skern, T. & Rhoads, R. E. Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. Implications for Cap-dependent and Cap-independent translational initiation. *J. Biol. Chem.* (1995). doi:10.1074/jbc.270.37.21975
64. Neff, C. L. & Sachs, A. B. Eukaryotic Translation Initiation Factors 4G and 4A from Saccharomyces cerevisiae Interact Physically and Functionally . *Mol. Cell. Biol.* (1999). doi:10.1128/mcb.19.8.5557
65. Pyronnet, S. *et al.* Human eukaryotic translation initiation factor 4G (eIF4G) recruits Mnk1 to phosphorylate eIF4E. *EMBO J.* (1999). doi:10.1093/emboj/18.1.270
66. Morino, S., Imataka, H., Svitkin, Y. V., Pestova, T. V. & Sonenberg, N. Eukaryotic Translation Initiation Factor 4E (eIF4E) Binding Site and the Middle One-Third of eIF4GI Constitute the Core Domain for Cap-Dependent Translation, and the C-Terminal One-Third Functions as a Modulatory Region. *Mol. Cell. Biol.* (2000). doi:10.1128/mcb.20.2.468-477.2000
67. Villa, N., Do, A., Hershey, J. W. B. & Fraser, C. S. Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome. *J. Biol. Chem.* (2013). doi:10.1074/jbc.M113.517011
68. Lefebvre, A. K. *et al.* Translation initiation factor eIF4G-1 binds to eIF3 through the eIF3e subunit. *J. Biol. Chem.* (2006). doi:10.1074/jbc.M605418200
69. Marintchev, A. *et al.* Topology and Regulation of the Human eIF4A/4G/4H Helicase Complex in Translation Initiation. *Cell* **136**, 447–460 (2009).
70. Asano, K. *et al.* Multiple roles for the C-terminal domain of elF5 in translation initiation complex assembly and GTPase activation. *EMBO J.* (2001). doi:10.1093/emboj/20.9.2326
71. Nanda, J. S., Saini, A. K., Muñoz, A. M., Hinnebusch, A. G. & Lorsch, J. R. Coordinated movements of eukaryotic translation initiation Factors eIF1, eIF1A, and eIF5 trigger phosphate release from eIF2 in response to start codon recognition by the ribosomal preinitiation complex. *J. Biol. Chem.* (2013). doi:10.1074/jbc.M112.440693
72. He, H. *et al.* The Yeast Eukaryotic Initiation Factor 4G (eIF4G) HEAT Domain Interacts with eIF1 and eIF5 and Is Involved in Stringent AUG Selection. *Mol. Cell. Biol.* (2003). doi:10.1128/mcb.23.15.5431-5445.2003
73. Berset, C., Zurbriggen, A., Djafarzadeh, S., Altmann, M. & Trachsel, H. RNA-binding activity of translation initiation factor eIF4G1 from Saccharomyces cerevisiae. *RNA* (2003). doi:10.1261/rna.5380903
74. Duncan, R., Milburn, S. C. & Hershey, J. W. B. Regulated phosphorylation and low abundance of HeLa cell initiation factor eIF-4F suggest a role in translational control. Heat shock effects on eIF-4F. *J. Biol. Chem.* (1987).
75. Barrieux, A. & Rosenfeld, M. G. Characterization of GTP dependent Met tRNA(f)

binding protein. *J. Biol. Chem.* (1977).
76. Lloyd, M. A., Osborne, J. C., Safer, B., Powell, G. M. & Merrick, W. C. Characteristics of eukaryotic initiation factor 2 and its subunits. *J. Biol. Chem.* (1980).
77. Schmitt, E., Naveau, M. & Mechulam, Y. Eukaryotic and archaeal translation initiation factor 2: A heterotrimeric tRNA carrier. *FEBS Letters* (2010). doi:10.1016/j.febslet.2009.11.002
78. Hannig, E. M., Cigan, A. M., Freeman, B. A. & Kinzy, T. G. GCD11, a negative regulator of GCN4 expression, encodes the gamma subunit of eIF-2 in Saccharomyces cerevisiae. *Mol. Cell. Biol.* (1993). doi:10.1128/mcb.13.1.506
79. Erickson, F. L. & Hannig, E. M. Ligand interactions with eukaryotic translation initiation factor 2: role of the gamma-subunit. *EMBO J.* (1996). doi:10.1002/j.1460-2075.1996.tb01021.x
80. Mitsui, K., Datta, A. & Ochoa, S. Removal of beta subunit of the eukaryotic polypeptide chain initiation factor 2 by limited proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* (1981). doi:10.1073/pnas.78.7.4128
81. Das, A., Bagchi, M. K., Ghosh-Dastidar, P. & Gupta, N. K. Protein synthesis in rabbit reticulocytes. A study of peptide chain initiation using native and beta-subunit-depleted eukaryotic initiation factor 2. *J. Biol. Chem.* (1982).
82. Stringer, E. A., Chaudhuri, A., Valenzuela, D. & Maitra, U. Rabbit reticulocyte initiation factor 2 contains two polypeptide chains of molecular weights 48,000 and 38,000. *Proc. Natl. Acad. Sci. U. S. A.* (1980). doi:10.1073/pnas.77.6.3356
83. Flynn, A., Oldfield, S. & Proud, C. G. The role of the β-subunit of initiation factor eIF-2 in initiation complex formation. *BBA - Gene Struct. Expr.* (1993). doi:10.1016/0167-4781(93)90105-M
84. Nika, J., Rippel, S. & Hannig, E. M. Biochemical analysis of the eIF2βγ complex reveals a structural function for eIF2α in catalyzed nucleotide exchange. *J. Biol. Chem.* (2001). doi:10.1074/jbc.M007398200
85. Kapp, L. D. & Lorsch, J. R. GTP-dependent Recognition of the Methionine Moiety on Initiator tRNA by Translation Factor eIF2. *J. Mol. Biol.* (2004). doi:10.1016/j.jmb.2003.11.025
86. Allen, G. S., Zavialov, A., Gursky, R., Ehrenberg, M. & Frank, J. The cryo-EM structure of a translation initiation complex from Escherichia coli. *Cell* (2005). doi:10.1016/j.cell.2005.03.023
87. Kolupaeva, V. G., Unbehaun, A., Lomakin, I. B., Hellen, C. U. T. & Pestova, T. V. Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. *RNA* (2005). doi:10.1261/rna.7215305
88. Fraser, C. S. *et al.* The j-Subunit of Human Translation Initiation Factor eIF3 Is Required for the Stable Binding of eIF3 and Its Subcomplexes to 40 S Ribosomal Subunits in Vitro. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M312745200
89. Majumdar, R., Bandyopadhyay, A. & Maitra, U. Mammalian translation initiation factor eIF1 functions with eIF1A and eIF3 in the formation of a stable 40 S preinitiation complex. *J. Biol. Chem.* (2003). doi:10.1074/jbc.M210357200
90. Maag, D. & Lorsch, J. R. Communication between eukaryotic translation initiation factors 1 and 1A on the yeast small ribosomal subunit. *J. Mol. Biol.* (2003). doi:10.1016/S0022-2836(03)00665-X
91. Lomakin, I. B., Kolupaeva, V. G., Marintchev, A., Wagner, G. & Pestova, T. V. Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes Dev.* (2003). doi:10.1101/gad.1141803
92. Yu, Y. *et al.* Position of eukaryotic translation initiation factor eIF1A on the 40S ribosomal subunit mapped by directed hydroxyl radical probing. *Nucleic Acids Res.* (2009). doi:10.1093/nar/gkp519

93. Gilbert, R. J. C. *et al.* Reconfiguration of yeast 40S ribosomal subunit domains by the translation initiation multifactor complex. *Proc. Natl. Acad. Sci. U. S. A.* (2007). doi:10.1073/pnas.0606880104
94. Passmore, L. A. *et al.* The Eukaryotic Translation Initiation Factors eIF1 and eIF1A Induce an Open Conformation of the 40S Ribosome. *Mol. Cell* **26**, 41–50 (2007).
95. Unbehaun, A., Borukhov, S. I., Hellen, C. U. T. & Pestova, T. V. Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes Dev.* (2004). doi:10.1101/gad.1255704
96. Phan, L. *et al.* Identification of a Translation Initiation Factor 3 (eIF3) Core Complex, Conserved in Yeast and Mammals, That Interacts with eIF5. *Mol. Cell. Biol.* (1998). doi:10.1128/mcb.18.8.4935
97. Valášek, L., Phan, L., Schoenfeld, L. W., Valášková, V. & Hinnebusch, A. G. Related eIF3 subunits TIF32 and HCR1 interact with an RNA recognition motif in PRT1 required for eIF3 integrity and ribosome binding. *EMBO J.* (2001). doi:10.1093/emboj/20.4.891
98. Siridechadilok, B. Structural Roles for Human Translation Factor eIF3 in Initiation of Protein Synthesis. *Science (80-.).* **310**, 1513–1515 (2005).
99. Cuchalova, L. *et al.* The RNA Recognition Motif of Eukaryotic Translation Initiation Factor 3g (eIF3g) Is Required for Resumption of Scanning of Posttermination Ribosomes for Reinitiation on GCN4 and Together with eIF3i Stimulates Linear Scanning. *Mol. Cell. Biol.* (2010). doi:10.1128/mcb.00430-10
100. ElAntak, L. *et al.* The indispensable n-terminal half of eIF3j/hcr1 cooperates with its structurally conserved binding partner eIF3b/prt1-rrm and with eiF1A in stringent aug selection. *J. Mol. Biol.* (2010). doi:10.1016/j.jmb.2009.12.047
101. Kouba, T., Rutkai, E., Karásková, M. & Valášek, L. S. The eIF3c/NIP1 PCI domain interacts with RNA and RACK1/ASC1 and promotes assembly of translation preinitiation complexes. *Nucleic Acids Res.* (2012). doi:10.1093/nar/gkr1083
102. Fraser, C. S., Berry, K. E., Hershey, J. W. B. & Doudna, J. A. eIF3j Is Located in the Decoding Center of the Human 40S Ribosomal Subunit. *Mol. Cell* (2007). doi:10.1016/j.molcel.2007.05.019
103. Elantak, L., Tzakos, A. G., Locker, N. & Lukavsky, P. J. Structure of eIF3b RNA recognition motif and its interaction with eIF3j: Structural insights into the recruitment of eIF3b to the 40 S ribosomal subunit. *J. Biol. Chem.* (2007). doi:10.1074/jbc.M610860200
104. Peterson, D. T., Merrick, W. C. & Safer, B. Binding and release of radiolabeled eukaryotic initiation factors 2 and 3 during 80 S initiation complex formation. *J. Biol. Chem.* (1979).
105. Korneeva, N. L., Lamphear, B. J., Hennigan, F. L. C. & Rhoads, R. E. Mutually cooperative binding of eukaryotic translation initiation factor (eIF) 3 and eIF4A to human eIF4G-1. *J. Biol. Chem.* (2000). doi:10.1074/jbc.M007525200
106. Mitchell, S. F. *et al.* The 5'-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway. *Mol. Cell* (2010). doi:10.1016/j.molcel.2010.08.021
107. Jivotovskaya, A. V., Valasek, L., Hinnebusch, A. G. & Nielsen, K. H. Eukaryotic Translation Initiation Factor 3 (eIF3) and eIF2 Can Promote mRNA Binding to 40S Subunits Independently of eIF4G in Yeast. *Mol. Cell. Biol.* (2006). doi:10.1128/mcb.26.4.1355-1372.2006
108. Valášek, L., Nielsen, K. H. & Hinnebusch, A. G. Direct eIF2-eIF3 contact in the multifactor complex is important for translation initiation in vivo. *EMBO J.* (2002). doi:10.1093/emboj/cdf563

109. Erzberger, J. P. *et al.* Molecular architecture of the 40S · eIF1 · eIF3 translation initiation complex. *Cell* (2014). doi:10.1016/j.cell.2014.07.044
110. Sokabe, M., Fraser, C. S. & Hershey, J. W. B. The human translation initiation multi-factor complex promotes methionyl-tRNA i binding to the 40S ribosomal subunit. *Nucleic Acids Res.* (2012). doi:10.1093/nar/gkr772
111. Das, S. & Maitra, U. Mutational Analysis of Mammalian Translation Initiation Factor 5 (eIF5): Role of Interaction between the beta Subunit of eIF2 and eIF5 in eIF5 Function In Vitro and In Vivo. *Mol. Cell. Biol.* (2000). doi:10.1128/mcb.20.11.3942-3950.2000
112. Luna, R. E. *et al.* The C-Terminal Domain of Eukaryotic Initiation Factor 5 Promotes Start Codon Recognition by Its Dynamic Interplay with eIF1 and eIF2β. *Cell Rep.* (2012). doi:10.1016/j.celrep.2012.04.007
113. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 113–127 (2010).
114. Aylett, C. H. S. & Ban, N. Eukaryotic aspects of translation initiation brought into focus. *Philosophical Transactions of the Royal Society B: Biological Sciences* (2017). doi:10.1098/rstb.2016.0186
115. Hinnebusch, A. G. Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. *Trends in Biochemical Sciences* (2017). doi:10.1016/j.tibs.2017.03.004
116. Kozak, M. Inability of circular mRNA to attach to eukaryotic ribosomes [26]. *Nature* (1979). doi:10.1038/280082a0
117. Kozak, M. Regulation of translation in eukaryotic systems. *Annual Review of Cell Biology* (1992). doi:10.1146/annurev.cb.08.110192.001213
118. Kumar, P., Hellen, C. U. T. & Pestova, T. V. Toward the mechanism of eIF4F-mediated ribosomal attachment to mammalian capped mRNAs. *Genes Dev.* **30**, 1573–1588 (2016).
119. Zhou, M. *et al.* Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. *Proc. Natl. Acad. Sci. U. S. A.* (2008). doi:10.1073/pnas.0801313105
120. Lee, A. S., Kranzusch, P. J., Doudna, J. A. & Cate, J. H. D. eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. *Nature* **536**, (2016).
121. Hashem, Y. *et al.* Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. *Nature* (2013). doi:10.1038/nature12658
122. Pisarev, A. V., Kolupaeva, V. G., Yusupov, M. M., Hellen, C. U. T. & Pestova, T. V. Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J.* (2008). doi:10.1038/emboj.2008.90
123. Pelletier, J. & Sonenberg, N. The Organizing Principles of Eukaryotic Ribosome Recruitment. *Annu. Rev. Biochem.* (2019). doi:10.1146/annurev-biochem-013118-111042
124. Grüner, S. *et al.* The Structures of eIF4E-eIF4G Complexes Reveal an Extended Interface to Regulate Translation Initiation. *Mol. Cell* **64**, 467–479 (2016).
125. Xue, S. *et al.* RNA regulons in Hox 59 UTRs confer ribosome specificity to gene regulation. *Nature* (2015). doi:10.1038/nature14010
126. Jackson, R. J. The current status of vertebrate cellular mRNA IRESs. *Cold Spring Harb. Perspect. Biol.* (2013). doi:10.1101/cshperspect.a011569
127. Mailliot, J. & Martin, F. Viral internal ribosomal entry sites: four classes for one goal. *Wiley Interdisciplinary Reviews: RNA* (2018). doi:10.1002/wrna.1458
128. Jaafar, Z. A. & Kieft, J. S. Viral RNA structure-based strategies to manipulate

translation. *Nature Reviews Microbiology* (2019). doi:10.1038/s41579-018-0117-x
129. Lozano, G. & Martínez-Salas, E. Structural insights into viral IRES-dependent translation mechanisms. *Current Opinion in Virology* (2015). doi:10.1016/j.coviro.2015.04.008
130. King, H. A., Cobbold, L. C. & Willis, A. E. The role of IRES trans-acting factors in regulating translation initiation. in *Biochemical Society Transactions* (2010). doi:10.1042/BST0381581
131. Sweeney, T. R., Abaeva, I. S., Pestova, T. V. & Hellen, C. U. T. The mechanism of translation initiation on type 1 picornavirus IRESs. *EMBO J.* (2014). doi:10.1002/embj.201386124
132. Kolupaeva, V. G., Pestova, T. V., Hellen, C. U. T. & Shatsky, I. N. Translation eukaryotic initiation factor 4G recognizes a specific structural element within the internal ribosome entry site of encephalomyocarditis virus RNA. *J. Biol. Chem.* (1998). doi:10.1074/jbc.273.29.18599
133. Pilipenko, E. V. *et al.* Prokaryotic-like cis elements in the cap-independent internal initiation of translation on picornavirus RNA. *Cell* (1992). doi:10.1016/0092-8674(92)90211-T
134. Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* (1988). doi:10.1038/334320a0
135. Kaminski, A., Belsham, G. J. & Jackson, R. J. Translation of encephalomyocarditis virus RNA: parameters influencing the selection of the internal initiation site. *EMBO J.* (1994). doi:10.1002/j.1460-2075.1994.tb06431.x
136. Pestova, T. V., Shatsky, I. N., Fletcher, S. P., Jackson, R. J. & Hellen, C. U. T. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initation of hepatitis C and classical swine fever virus RNAs. *Genes Dev.* (1998). doi:10.1101/gad.12.1.67
137. Kerr, C. H. & Jan, E. Commandeering the Ribosome: Lessons Learned from Dicistroviruses about Translation. *J. Virol.* (2016). doi:10.1128/jvi.00737-15
138. Hertz, M. I. & Thompson, S. R. Mechanism of translation initiation by Dicistroviridae IGR IRESs. *Virology* (2011). doi:10.1016/j.virol.2011.01.005
139. Zhu, J. *et al.* Crystal structures of complexes containing domains from two viral internal ribosome entry site (IRES) RNAs bound to the 70S ribosome. *Proc. Natl. Acad. Sci. U. S. A.* (2011). doi:10.1073/pnas.1018582108
140. Butcher, S. E. & Jan, E. tRNA-mimicry in IRES-mediated translation and recoding. *RNA Biology* (2016). doi:10.1080/15476286.2016.1219833
141. Costantino, D. A., Pfingsten, J. S., Rambo, R. P. & Kieft, J. S. tRNA-mRNA mimicry drives translation initiation from a viral IRES. *Nat. Struct. Mol. Biol.* (2008). doi:10.1038/nsmb1351
142. Walters, B. & Thompson, S. R. Cap-independent translational control of carcinogenesis. *Frontiers in Oncology* (2016). doi:10.3389/fonc.2016.00128
143. Baranick, B. T. *et al.* Splicing mediates the activity of four putative cellular internal ribosome entry sites. *Proc. Natl. Acad. Sci. U. S. A.* (2008). doi:10.1073/pnas.0710650105
144. Meyer, K. D. *et al.* 5′ UTR m6A Promotes Cap-Independent Translation. *Cell* (2015). doi:10.1016/j.cell.2015.10.012
145. Zhou, J. *et al.* Dynamic m6 A mRNA methylation directs translational control of heat shock response. *Nature* (2015). doi:10.1038/nature15377
146. Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m 6 A RNA methylation. *Nature Reviews Genetics* (2014). doi:10.1038/nrg3724

147. Meyer, K. D. & Jaffrey, S. R. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nature Reviews Molecular Cell Biology* (2014). doi:10.1038/nrm3785
148. Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* (2014). doi:10.1038/nchembio.1432
149. Ping, X. L. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* (2014). doi:10.1038/cr.2014.3
150. Jia, G. *et al.* N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* (2011). doi:10.1038/nchembio.687
151. Zheng, G. *et al.* ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol. Cell* (2013). doi:10.1016/j.molcel.2012.10.015
152. Wang, X. *et al.* N 6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* (2014). doi:10.1038/nature12730
153. Xiao, W. *et al.* Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* (2016). doi:10.1016/j.molcel.2016.01.012
154. Liu, N. *et al.* N6 -methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* (2015). doi:10.1038/nature14234
155. Coots, R. A. *et al.* m6A Facilitates eIF4F-Independent mRNA Translation. *Mol. Cell* (2017). doi:10.1016/j.molcel.2017.10.002
156. Zhou, J. *et al.* N6-Methyladenosine Guides mRNA Alternative Translation during Integrated Stress Response. *Mol. Cell* (2018). doi:10.1016/j.molcel.2018.01.019
157. Slobodin, B. *et al.* Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. *Cell* **169**, 326-337.e12 (2017).
158. Geula, S. *et al.* m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science (80-.).* (2015). doi:10.1126/science.1261417
159. Zhao, X. *et al.* FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res.* (2014). doi:10.1038/cr.2014.151
160. Lin, S., Choe, J., Du, P., Triboulet, R. & Gregory, R. I. The m6A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol. Cell* (2016). doi:10.1016/j.molcel.2016.03.021
161. Li, Z. *et al.* Suppression of m6A reader Ythdf2 promotes hematopoietic stem cell expansion. *Cell Res.* (2018). doi:10.1038/s41422-018-0072-0
162. Chen, X. *et al.* CircRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Sci. Rep.* (2016). doi:10.1038/srep34985
163. Legnini, I. *et al.* Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cell* (2017). doi:10.1016/j.molcel.2017.02.017
164. Granados-Riveron, J. T. & Aquino-Jarquin, G. The complexity of the translation ability of circRNAs. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* (2016). doi:10.1016/j.bbagrm.2016.07.009
165. Schneider, T. *et al.* CircRNA-protein complexes: IMP3 protein component defines subfamily of circRNPs. *Sci. Rep.* (2016). doi:10.1038/srep31313
166. Wang, Y. & Wang, Z. Efficient backsplicing produces translatable circular mRNAs. *RNA* (2015). doi:10.1261/rna.048272.114
167. Pamudurti, N. R. *et al.* Translation of CircRNAs. *Mol. Cell* (2017). doi:10.1016/j.molcel.2017.02.021
168. Yang, Y. *et al.* Extensive translation of circular RNAs driven by N 6 -methyladenosine. *Cell Res.* (2017). doi:10.1038/cr.2017.31
169. Zhang, M. *et al.* A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. *Oncogene* (2018). doi:10.1038/s41388-017-0019-9
170. Patel, D. D. & Pickup, D. J. Messenger RNAs of a strongly-expressed late gene of cowpox virus contain 5′-terminal poly(A) sequences. *EMBO J.* (1987).

doi:10.1002/j.1460-2075.1987.tb02714.x
171. Mulder, J., Robertson, M. E. M., Seamons, R. A. & Belsham, G. J. Vaccinia Virus Protein Synthesis Has a Low Requirement for the Intact Translation Initiation Factor eIF4F, the Cap-Binding Complex, within Infected Cells. *J. Virol.* (1998). doi:10.1128/jvi.72.11.8813-8819.1998
172. Gudkov, A. T., Ozerova, M. V., Shiryaev, V. M. & Spirin, A. S. 5′-poly(A) sequence as an effective leader for translation in eukaryotic cell-free systems. *Biotechnol. Bioeng.* (2005). doi:10.1002/bit.20525
173. Shirokikh, N. E. & Spirin, A. S. Poly(A) leader of eukaryotic mRNA bypasses the dependence of translation on initiation factors. *Proc. Natl. Acad. Sci. U. S. A.* (2008). doi:10.1073/pnas.0804940105
174. Hernández, G., Vázquez-Pianzola, P., Sierra, J. M. & Rivera-Pomar, R. Internal ribosome entry site drives cap-independent translation of reaper and heat shock protein 70 mRNAs in Drosophila embryos. *RNA* (2004). doi:10.1261/rna.7154104
175. Gilbert, W. V., Zhou, K., Butler, T. K. & Doudna, J. A. Cap-independent translation is required for starvation-induced differentiation in yeast. *Science (80-.).* (2007). doi:10.1126/science.1144467
176. Eliseev, B. *et al.* Structure of a human cap-dependent 48S translation pre-initiation complex. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gky054
177. Pestova, T. V. & Kolupaeva, V. G. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.* **16**, 2906–2922 (2002).
178. Pisareva, V. P., Pisarev, A. V., Komar, A. A., Hellen, C. U. T. & Pestova, T. V. Translation Initiation on Mammalian mRNAs with Structured 5′UTRs Requires DExH-Box Protein DHX29. *Cell* **135**, 1237–1250 (2008).
179. Berthelot, K., Muldoon, M., Rajkowitsch, L., Hughes, J. & McCarthy, J. E. G. Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. *Mol. Microbiol.* (2004). doi:10.1046/j.1365-2958.2003.03898.x
180. Mignone, F., Gissi, C., Liuni, S. & Pesole, G. Untranslated regions of mRNAs. *Genome Biol.* **3**, REVIEWS0004 (2002).
181. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–7 (2009).
182. Cigan, a M. & Donahue, T. F. Sequence and structural features associated with translational initiator regions in yeast--a review. *Gene* **59**, 1–18 (1987).
183. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–9 (2008).
184. Kozak, M. A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. *Gene Expr.* **1**, 111–115 (1991).
185. Li, L. & Wang, C. C. Capped mRNA with a Single Nucleotide Leader Is Optimally Translated in a Primitive Eukaryote, Giardia lamblia. *J. Biol. Chem.* **279**, 14656–14664 (2004).
186. Adam, R. D. The Giardia lamblia genome. *International Journal for Parasitology* **30**, 475–484 (2000).
187. Bruchhaus, I., Leippe, M., Lioutas, C. & Tannich, E. Unusual gene organization in the protozoan parasite Entamoeba histolytica. *DNA Cell Biol* **12**, 925–933 (1993).
188. Hughes, M. J. G. & Andrews, D. W. A single nucleotide is a sufficient 5' untranslated region for translation in an eukaryotic in vitro system. *FEBS Lett.* **414**, 19–22 (1997).
189. Gandin, V. *et al.* NanoCAGE reveals 5' UTR features that define specific modes of translation of functionally related MTOR-sensitive mRNAs. *Genome Res.* **26**, 636–648 (2016).
190. Elfakess, R. *et al.* Unique translation initiation of mRNAs-containing TISU element.

Nucleic Acids Res. **39**, 7598–7609 (2011).
191. Joshi, C. P., Zhou, H., Huang, X. & Chiang, V. L. Context sequences of translation initiation codon in plants. *Plant Mol. Biol.* (1997). doi:10.1023/A:1005816823636
192. Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger rNAS. *Nucleic Acids Res.* (1987). doi:10.1093/nar/15.20.8125
193. Miyasaka, H. The positive relationship between codon usage bias and translation initiation AUG context in Saccharomyces cerevisiae. *Yeast* (1999). doi:10.1002/(SICI)1097-0061(19990615)15:8<633::AID-YEA407>3.0.CO;2-O
194. Kozak, M. Emerging links between initiation of translation and human diseases. *Mammalian Genome* (2002). doi:10.1007/s00335-002-4002-5
195. Wolf, A. *et al.* Single base-pair substitutions at the translation initiation sites of human genes as a cause of inherited disease. *Hum. Mutat.* (2011). doi:10.1002/humu.21547
196. Xu, H. *et al.* Screening of Kozak-motif-located SNPs and analysis of their association with human diseases. *Biochem. Biophys. Res. Commun.* (2010). doi:10.1016/j.bbrc.2010.01.002
197. Noderer, W. L. *et al.* Quantitative analysis of mammalian translation initiation sites by FACS -seq . *Mol. Syst. Biol.* (2014). doi:10.15252/msb.20145136
198. Smith, E. *et al.* Leaky ribosomal scanning in mammalian genomes: Significance of histone H4 alternative translation in vivo. *Nucleic Acids Res.* (2005). doi:10.1093/nar/gki248
199. Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl. Acad. Sci. U. S. A.* (2009). doi:10.1073/pnas.0810916106
200. Resch, A. M., Ogurtsov, A. Y., Rogozin, I. B., Shabalina, S. A. & Koonin, E. V. Evolution of alternative and constitutive regions of mammalian 5′ UTRs. *BMC Genomics* (2009). doi:10.1186/1471-2164-10-162
201. Lawless, C. *et al.* Upstream sequence elements direct post-transcriptional regulation of gene expression under stress conditions in yeast. *BMC Genomics* (2009). doi:10.1186/1471-2164-10-7
202. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
203. Harding, H. P. *et al.* Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol. Cell* (2000). doi:10.1016/S1097-2765(00)00108-8
204. Hinnebusch, A. G. TRANSLATIONAL REGULATION OF GCN4 AND THE GENERAL AMINO ACID CONTROL OF YEAST . *Annu. Rev. Microbiol.* (2005). doi:10.1146/annurev.micro.59.031805.133833
205. Somers, J., Pöyry, T. & Willis, A. E. A perspective on mammalian upstream open reading frame function. *International Journal of Biochemistry and Cell Biology* (2013). doi:10.1016/j.biocel.2013.04.020
206. Yaman, I. *et al.* The zipper model of translational control: A small upstream ORF is the switch that controls structural remodeling of an mRNA leader. *Cell* (2003). doi:10.1016/S0092-8674(03)00345-3
207. Fernandez, J. *et al.* Ribosome stalling regulates ires-mediated translation in eukaryotes, a parallel to prokaryotic attenuation. *Mol. Cell* (2005). doi:10.1016/j.molcel.2004.12.024
208. Chen, T. M. *et al.* Overexpression of FGF9 in colon cancer cells is mediated by hypoxia-induced translational activation. *Nucleic Acids Res.* (2014). doi:10.1093/nar/gkt1286
209. Chew, G. L., Pauli, A. & Schier, A. F. Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nat. Commun.* (2016).

doi:10.1038/ncomms11663
210. Gentilella, A. & Thomas, G. Cancer biology: The director's cut. *Nature* **485**, 50–51 (2012).
211. Levy, S., Avni, D., Hariharan, N., Perry, R. P. & Meyuhas, O. Oligopyrimidine tract at the 5' end of mammalian ribosomal protein mRNAs is required for their translational control. *Proc. Natl. Acad. Sci. U. S. A.* (1991). doi:10.1073/pnas.88.8.3319
212. Meyuhas, O. & Kahan, T. The race to decipher the top secrets of TOP mRNAs. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* (2015). doi:10.1016/j.bbagrm.2014.08.015
213. Yamashita, R. *et al.* Comprehensive detection of human terminal oligo-pyrimidine (TOP) genes and analysis of their characteristics. *Nucleic Acids Res.* **36**, 3707–3715 (2008).
214. Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55–61 (2012).
215. Thoreen, C. C. *et al.* A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109–113 (2012).
216. Chantranupong, L., Wolfson, R. L. & Sabatini, D. M. Nutrient-sensing mechanisms across evolution. *Cell* (2015). doi:10.1016/j.cell.2015.02.041
217. Fonseca, B. D. *et al.* La-related protein 1 (LARP1) represses terminal oligopyrimidine (TOP) mRNA translation downstream of mTOR complex 1 (mTORC1). *J. Biol. Chem.* **290**, 15996–16020 (2015).
218. Philippe, L., Vasseur, J. J., Debart, F. & Thoreen, C. C. La-related protein 1 (LARP1) repression of TOP mRNA translation is mediated through its cap-binding domain and controlled by an adjacent regulatory region. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gkx1237
219. Lahr, R. M. *et al.* La-related protein 1 (LARP1) binds the mRNA cap, blocking eIF4F assembly on TOP mRNAs. *Elife* (2017). doi:10.7554/eLife.24146
220. Cassidy, K. C. *et al.* Capturing the Mechanism Underlying TOP mRNA Binding to LARP1. *Structure* (2019). doi:10.1016/j.str.2019.10.006
221. Aoki, K. *et al.* LARP1 specifically recognizes the 3′ terminus of poly(A) mRNA. *FEBS Lett.* **587**, 2173–2178 (2013).
222. Gentilella, A. *et al.* Autogenous Control of 5′TOP mRNA Stability by 40S Ribosomes. *Mol. Cell* **67**, 55-70.e4 (2017).
223. Hentze, M. W. *et al.* Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science (80-.).* (1987). doi:10.1126/science.3685996
224. Gray, N. K. & Hentze, M. W. Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. *EMBO J.* (1994). doi:10.1002/j.1460-2075.1994.tb06699.x
225. Muckenthaler, M., Gray, N. K. & Hentze, M. W. IRP-1 binding to ferritin mRNA prevents the recruitment of the small ribosomal subunit by the cap-binding complex eIF4F. *Mol. Cell* (1998). doi:10.1016/S1097-2765(00)80282-8
226. Muckenthaler, M. U., Rivella, S., Hentze, M. W. & Galy, B. A Red Carpet for Iron Metabolism. *Cell* (2017). doi:10.1016/j.cell.2016.12.034
227. Lee, A. S. Y., Kranzusch, P. J. & Cate, J. H. D. eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* **522**, (2015).
228. Sun, L. *et al.* RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.* (2019). doi:10.1038/s41594-019-0200-7
229. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* (2009). doi:10.1038/nature08237
230. Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast.

Nature **467**, 103–107 (2010).
231. Wan, Y. *et al.* Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* (2014). doi:10.1038/nature12946
232. Li, F. *et al.* Global Analysis of RNA Secondary Structure in Two Metazoans. *Cell Rep.* (2012). doi:10.1016/j.celrep.2011.10.002
233. Li, F. *et al.* Regulatory impact of RNA secondary structure across the arabidopsis transcriptomeW OA. *Plant Cell* (2012). doi:10.1105/tpc.112.104232
234. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* (2014). doi:10.1038/nature12894
235. Ding, Y. *et al.* In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* (2014). doi:10.1038/nature12756
236. Laso, M. R. V. *et al.* Inhibition of translational initiation in the yeast Saccharomyces cerevisiae as a function of the stability and position of hairpin structures in the mRNA leader. *J. Biol. Chem.* (1993).
237. McCarthy, J. E. G. Posttranscriptional Control of Gene Expression in Yeast. *Microbiol. Mol. Biol. Rev.* (1998). doi:10.1128/mmbr.62.4.1492-1553.1998
238. Incarnato, D., Neri, F., Anselmi, F. & Oliviero, S. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.* (2014). doi:10.1186/s13059-014-0491-2
239. Kozak, M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* (1986). doi:10.1073/pnas.83.9.2850
240. Babendure, J. R., Babendure, J. L., Ding, J. H. & Tsien, R. Y. Control of mammalian translation by mRNA structure near caps. *RNA* (2006). doi:10.1261/rna.2309906
241. Kozak, M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol. Cell. Biol.* (1989). doi:10.1128/mcb.9.11.5134
242. Pelletier, J. & Sonenberg, N. Photochemical cross-linking of cap binding proteins to eucaryotic mRNAs: effect of mRNA 5' secondary structure. *Mol. Cell. Biol.* (1985). doi:10.1128/mcb.5.11.3222
243. Shabalina, S. A., Ogurtsov, A. Y. & Spiridonov, N. A. A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* (2006). doi:10.1093/nar/gkl287
244. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* (2005). doi:10.1016/j.gene.2005.06.037
245. Aw, J. G. A. *et al.* In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Mol. Cell* (2016). doi:10.1016/j.molcel.2016.04.028
246. Kozak, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* (1990). doi:10.1073/pnas.87.21.8301
247. Meyer, I. M. & Miklós, I. Statistical evidence for conserved, local secondary structure in the coding regions of eukaryotic mRNAs and pre-mRNAs. *Nucleic Acids Res.* (2005). doi:10.1093/nar/gki923
248. Bennett, C. F., Baker, B. F., Pham, N., Swayze, E. & Geary, R. S. Pharmacology of Antisense Drugs. *Annu. Rev. Pharmacol. Toxicol.* (2017). doi:10.1146/annurev-pharmtox-010716-104846
249. Toth, P. P. Emerging LDL therapies: Mipomersen - Antisense oligonucleotide therapy in the management of hypercholesterolemia. *J. Clin. Lipidol.* (2013). doi:10.1016/j.jacl.2013.02.004
250. Benson, M. D. *et al.* Antisense oligonucleotide therapy for TTR amyloidosis. *Amyloid* (2011). doi:10.3109/13506129.2011.574354021

251. Baker, B. F. *et al.* 2'-O-(2-methoxy)ethyl-modified anti-intercellular adhesion molecule 1 (ICAM-1) oligonucleotides selectively increase the ICAM-1 mRNA level and inhibit formation of the ICAM-1 translation initiation complex in human umbilical vein endothelial cells. *J. Biol. Chem.* (1997). doi:10.1074/jbc.272.18.11994
252. Braasch, D. A., Liu, Y. & Corey, D. R. Antisense inhibition of gene expression in cells by oligonucleotides incorporating locked nucleic acids: Effect of mRNA target sequence and chimera design. *Nucleic Acids Research* (2002). doi:10.1093/nar/gkf651
253. Doyle, D. F., Braasch, D. A., Simmons, C. G., Janowski, B. A. & Corey, D. R. Inhibition of gene expression inside cells by peptide nucleic acids: Effect of mRNA target sequence, mismatched bases, and PNA length. *Biochemistry* (2001). doi:10.1021/bi0020630
254. Liang, X. H., Shen, W. & Crooke, S. T. Specific increase of protein levels by enhancing translation using antisense oligonucleotides targeting upstream open frames. in *Advances in Experimental Medicine and Biology* (2017). doi:10.1007/978-981-10-4310-9_9
255. Liang, X. H. *et al.* Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. *Nat. Biotechnol.* (2016). doi:10.1038/nbt.3589
256. Liang, X. H. *et al.* Antisense oligonucleotides targeting translation inhibitory elements in 5′ UTRs can selectively increase protein levels. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx632
257. Sasaki, S. *et al.* Steric Inhibition of 5′ UTR Regulatory Elements Results in Upregulation of Human CFTR. *Mol. Ther.* (2019). doi:10.1016/j.ymthe.2019.06.016
258. Kozak, M. Role of ATP in binding and migration of 40S ribosomal subunits. *Cell* (1980). doi:10.1016/0092-8674(80)90356-6
259. Parkin, N. T. *et al.* Mutational analysis of the 5′ non-coding region of human immunodeficiency virus type 1: effects of secondary structure on translation. *EMBO J.* (1988). doi:10.1002/j.1460-2075.1988.tb03139.x
260. Fraser, C. S. Quantitative studies of mRNA recruitment to the eukaryotic ribosome. *Biochimie* (2015). doi:10.1016/j.biochi.2015.02.017
261. Rogers, G. W., Richter, N. J., Lima, W. F. & Merrick, W. C. Modulation of the Helicase Activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J. Biol. Chem.* **276**, 30914–30922 (2001).
262. Abaeva, I. S., Marintchev, A., Pisareva, V. P., Hellen, C. U. T. & Pestova, T. V. Bypassing of stems versus linear base-by-base inspection of mammalian mRNAs during ribosomal scanning. *EMBO J.* **30**, 115–129 (2011).
263. Chiu, W.-L. *et al.* The C-Terminal Region of Eukaryotic Translation Initiation Factor 3a (eIF3a) Promotes mRNA Recruitment, Scanning, and, Together with eIF3j and the eIF3b RNA Recognition Motif, Selection of AUG Start Codons. *Mol. Cell. Biol.* (2010). doi:10.1128/mcb.00280-10
264. Tarn, W.-Y. & Chang, T.-H. The current understanding of Ded1p/DDX3 homologs from yeast to human. *RNA Biol.* **6**, 17–20 (2009).
265. Rubio, C. A. *et al.* Transcriptome-wide characterization of the eIF4A signature highlights plasticity in translation regulation. *Genome Biol.* (2014). doi:10.1186/s13059-014-0476-1
266. Lai, M. C., Lee, Y. H. W. & Tarn, W. Y. The DEAD-box RNA helicase DDX3 associates with export messenger ribonucleoproteins as well as tip-associated protein and participates in translational control. *Mol. Biol. Cell* (2008). doi:10.1091/mbc.E07-12-1264
267. Soto-Rifo, R. *et al.* DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. *EMBO J.* (2012). doi:10.1038/emboj.2012.220

268. Gao, Z. *et al.* Coupling between the DEAD-box RNA helicases Ded1p and eIF4A. *Elife* **5**, 1–22 (2016).
269. Gupta, N., Lorsch, J. R. & Hinnebusch, A. G. Yeast Ded1 promotes 48S translation preinitiation complex assembly in an mRNA-specific and eIF4F-dependent manner. *Elife* (2018). doi:10.7554/eLife.38892
270. Marsden, S., Nardelli, M., Linder, P. & McCarthy, J. E. G. Unwinding Single RNA Molecules Using Helicases Involved in Eukaryotic Translation Initiation. *J. Mol. Biol.* (2006). doi:10.1016/j.jmb.2006.06.016
271. Paek, K. Y., Park, S. M., Hong, K. Y. & Jang, S. K. Cap-dependent translation without base-by-base scanning of an messenger ribonucleic acid. *Nucleic Acids Res.* (2012). doi:10.1093/nar/gks471
272. Fütterer, J., Kiss-László, Z. & Hohn, T. Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA. *Cell* (1993). doi:10.1016/0092-8674(93)90257-Q
273. Rogers, G. W., Edelman, G. M. & Mauro, V. P. Differential utilization of upstream AUGs in the β-secretase mRNA suggests that a shunting mechanism regulates translation. *Proc. Natl. Acad. Sci. U. S. A.* (2004). doi:10.1073/pnas.0308576101
274. Sherrill, K. W. & Lloyd, R. E. Translation of cIAP2 mRNA Is Mediated Exclusively by a Stress-Modulated Ribosome Shunt. *Mol. Cell. Biol.* (2008). doi:10.1128/mcb.01446-07
275. Mihailovich, M., Thermann, R., Grohovaz, F., Hentze, M. W. & Zacchetti, D. Complex translational regulation of BACE1 involves upstream AUGs and stimulatory elements within the 5′ untranslated region. *Nucleic Acids Res.* (2007). doi:10.1093/nar/gkm191
276. De Pietri Tonelli, D. *et al.* Translational regulation of BACE-1 expression in neuronal and non-neuronal cells. *Nucleic Acids Res.* (2004). doi:10.1093/nar/gkh348
277. Pestova, T. V., Borukhov, S. I. & Hellen, C. U. T. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* (1998). doi:10.1038/29703
278. Maag, D., Fekete, C. A., Gryczynski, Z. & Lorsch, J. R. A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol. Cell* (2005). doi:10.1016/j.molcel.2004.11.051
279. Pisarev, A. V. *et al.* Specific functional interactions of nucleotides at key-3 and+4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev.* (2006). doi:10.1101/gad.1397906
280. Lomakin, I. B., Shirokikh, N. E., Yusupov, M. M., Hellen, C. U. T. & Pestova, T. V. The fidelity of translation initiation: Reciprocal activities of eIF1, IF3 and YciH. *EMBO J.* (2006). doi:10.1038/sj.emboj.7600904
281. Luna, R. E. *et al.* The interaction between eukaryotic initiation factor 1A and eIF5 retains eif1 within scanning preinitiation complexes. *Biochemistry* (2013). doi:10.1021/bi4009775
282. Maag, D., Algire, M. A. & Lorsch, J. R. Communication between eukaryotic translation initiation factors 5 and 1A within the ribosomal pre-initiation complex plays a role in start site selection. *J. Mol. Biol.* (2006). doi:10.1016/j.jmb.2005.11.083
283. Obayashi, E. *et al.* Molecular Landscape of the Ribosome Pre-initiation Complex during mRNA Scanning: Structural Role for eIF3c and Its Control by eIF5. *Cell Rep.* (2017). doi:10.1016/j.celrep.2017.02.052
284. Haimov, O. *et al.* Dynamic Interaction of Eukaryotic Initiation Factor 4G1 (eIF4G1) with eIF4E and eIF1 Underlies Scanning-Dependent and -Independent Translation. *Mol. Cell. Biol.* (2018). doi:10.1128/mcb.00139-18
285. Hussain, T. *et al.* Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell* **159**, 597–607 (2014).

286. Llácer, J. L. *et al.* Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex. *Mol. Cell* (2015). doi:10.1016/j.molcel.2015.06.033
287. Palam, L. R., Baird, T. D. & Wek, R. C. Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M110.216093
288. Sokabe, M. & Fraser, C. S. Toward a kinetic understanding of eukaryotic translation. *Cold Spring Harb. Perspect. Biol.* (2019). doi:10.1101/cshperspect.a032706
289. Shatkin, A. J. Capping of eucaryotic mRNAs. *Cell* (1976). doi:10.1016/0092-8674(76)90128-8
290. Kozak, M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* **266**, 19867–19870 (1991).
291. Rogers, G. W., Richter, N. J. & Merrick, W. C. Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J. Biol. Chem.* (1999). doi:10.1074/jbc.274.18.12236
292. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. Structural Basis for RNA Unwinding by the DEAD-Box Protein Drosophila Vasa. *Cell* **125**, 287–300 (2006).
293. Yang, Q., Del Campo, M., Lambowitz, A. M. & Jankowsky, E. DEAD-Box Proteins Unwind Duplexes by Local Strand Separation. *Mol. Cell* **28**, 253–263 (2007).
294. Yang, Q. & Jankowsky, E. The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat. Struct. Mol. Biol.* (2006). doi:10.1038/nsmb1165
295. Lorsch, J. R. & Herschlag, D. The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry* **37**, 2180–2193 (1998).
296. Oguro, A., Ohtsu, T., Svitkin, Y. V., Sonenberg, N. & Nakamura, Y. RNA aptamers to initiation factor 4A helicase hinder cap-dependent translation by blocking ATP hydrolysis. *RNA* (2003). doi:10.1261/rna.2161303
297. Rogers, G. W., Richter, N. J. & Merrick, W. C. Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J. Biol. Chem.* **274**, 12236–12244 (1999).
298. Yoder-Hill, J., Pause, A., Sonenberg, N. & Merrick, W. C. The p46 subunit of eukaryotic initiation factor (eIF)-4F exchanges with eIF-4A. *J. Biol. Chem.* **268**, 5566–5573 (1993).
299. Conroy, S. C., Dever, T. E., Owens, C. L. & Merrick, W. C. Characterization of the 46,000-Dalton subunit of eIF-4F. *Arch. Biochem. Biophys.* **282**, 363–371 (1990).
300. Galicia-Vázquez, G., Cencic, R., Robert, F., Agenor, A. Q. & Pelletier, J. A cellular response linking eIF4AI activity to eIF4AII transcription. *RNA* **18**, 1373–84 (2012).
301. Meijer, H. A. *et al.* Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. *Science (80-.).* **340**, 82–85 (2013).
302. Williams-Hill, D. M., Duncan, R. F., Nielsen, P. J. & Tahara, S. M. Differential expression of the murine eukaryotic translation initiation factor isogenes eIF4A(I) and eIF4A(II) is dependent upon cellular growth status. *Arch. Biochem. Biophys.* (1997). doi:10.1006/abbi.1996.9804
303. Gingras, A. C., Raught, B. & Sonenberg, N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu Rev Biochem* **68**, 913–963 (1999).
304. Li, Q. *et al.* Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. *Mol. Cell. Biol.* **19**, 7336–7346 (1999).
305. Iwatani-Yoshihara, M. *et al.* Discovery and characterization of a eukaryotic initiation

factor 4A-3-selective inhibitor that suppresses nonsense-mediated mRNA decay. *ACS Chem. Biol.* acschembio.7b00041 (2017). doi:10.1021/acschembio.7b00041
306. Shibuya, T., Tange, T. Ø., Sonenberg, N. & Moore, M. J. eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat. Struct. Mol. Biol.* **11**, 346–351 (2004).
307. Ballut, L. *et al.* The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat. Struct. Mol. Biol.* (2005). doi:10.1038/nsmb990
308. Duncan, R. & Hershey, J. W. Identification and quantitation of levels of protein synthesis initiation factors in crude HeLa cell lysates by two-dimensional polyacrylamide gel electrophoresis. *J Biol Chem* **258**, 7228–7235 (1983).
309. Oberer, M., Marintchev, A. & Wagner, G. Structural basis for the enhancement of eIF4A helicase activity by eIF4G. *Genes Dev.* (2005). doi:10.1101/gad.1335305
310. Bi, X., Ren, J. & Goss, D. J. Wheat germ translation initiation factor eIF4B affects eIF4A and eIFiso4F helicase activity by increasing the ATP binding affinity of eIF4A. *Biochemistry* (2000). doi:10.1021/bi992322p
311. Loh, P. G. *et al.* Structural basis for translational inhibition by the tumour suppressor Pdcd4. *EMBO J.* (2009). doi:10.1038/emboj.2008.278
312. Suzuki, C. *et al.* PDCD4 inhibits translation initiation by binding to eIF4A using both its MA3 domains. *Proc. Natl. Acad. Sci. U. S. A.* (2008). doi:10.1073/pnas.0712235105
313. Andreou, A. Z., Klostermeier, D., Andreou, A. Z. & Klostermeier, D. The DEAD-box helicase eIF4A The DEAD-box helicase eIF4A Paradigm or the odd one out ? **6286**, 19–32 (2016).
314. Rajagopal, V., Park, E. H., Hinnebusch, A. G. & Lorsch, J. R. Specific domains in yeast translation initiation factor eIF4G strongly bias RNA unwinding activity of the eIF4F complex toward duplexes with 5′-overhangs. *J. Biol. Chem.* (2012). doi:10.1074/jbc.M112.347278
315. Sun, Y. *et al.* The eukaryotic initiation factor eIF4H facilitates loop-binding, repetitive RNA unwinding by the eIF4A DEAD-box helicase. *Nucleic Acids Res.* (2012). doi:10.1093/nar/gks278
316. Gingras, A. C. *et al.* Regulation of 4E-BP1 phosphorylation: A novel two step mechanism. *Genes Dev.* (1999). doi:10.1101/gad.13.11.1422
317. Feoktistova, K., Tuvshintogs, E., Do, A. & Fraser, C. S. Human eIF4E promotes mRNA restructuring by stimulating eIF4A helicase activity. *Proc. Natl. Acad. Sci. U. S. A.* (2013). doi:10.1073/pnas.1303781110
318. Shahbazian, D. *et al.* The mTOR/PI3K and MAPK pathways converge on eIF4B to control its phosphorylation and activity. *EMBO J.* (2006). doi:10.1038/sj.emboj.7601166
319. Kroczynska, B. *et al.* Interferon-Dependent Engagement of Eukaryotic Initiation Factor 4B via S6 Kinase (S6K)- and Ribosomal Protein S6K-Mediated Signals. *Mol. Cell. Biol.* (2009). doi:10.1128/mcb.01537-08
320. Holz, M. K., Ballif, B. A., Gygi, S. P. & Blenis, J. mTOR and S6K1 mediate assembly of the translation preinitiation complex through dynamic protein interchange and ordered phosphorylation events. *Cell* (2005). doi:10.1016/j.cell.2005.10.024
321. Walker, S. E. *et al.* Yeast eIF4B binds to the head of the 40S ribosomal subunit and promotes mRNA recruitment through its N-terminal and internal repeat domains. *RNA* **19**, 191–207 (2013).
322. Abaeva, I. S., Marintchev, A., Pisareva, V. P., Hellen, C. U. T. & Pestova, T. V. Bypassing of stems versus linear base-by-base inspection of mammalian mRNAs during ribosomal scanning. *EMBO J.* **30**, 115–29 (2011).
323. Blum, S. *et al.* ATP hydrolysis by initiation factor 4A is required for translation initiation in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 7664–7668

(1992).
324. Svitkin, Y. V *et al.* The requirement for eukaryotic initiation factor 4A (elF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* **7**, 382–394 (2001).
325. Abramson, R. D. *et al.* The ATP-dependent interaction of eukaryotic initiation factors with mRNA. *J. Biol. Chem.* **262**, 3826–3832 (1987).
326. Rozen, F. *et al.* Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. *Mol. Cell. Biol.* **10**, 1134–44 (1990).
327. Sokabe, M. & Fraser, C. S. A helicase-independent activity of eIF4A in promoting mRNA recruitment to the human ribosome. *Proc. Natl. Acad. Sci.* **114**, 6304–6309 (2017).
328. Méthot, N., Pickett, G., Keene, J. D. & Sonenberg, N. In vitro RNA selection identifies RNA ligands that specifically bind to eukaryotic translation initiation factor 4B: The role of the RNA recognition motif. *RNA* (1996).
329. des Georges, A. *et al.* Structure of mammalian eIF3 in the context of the 43S preinitiation complex. *Nature* **525**, 491–5 (2015).
330. Archer, S. K., Shirokikh, N. E., Beilharz, T. H. & Preiss, T. Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature* **535**, 570–4 (2016).
331. Díaz-López, I., Toribio, R., Berlanga, J. J. & Ventoso, I. An mRNA-binding channel in the ES6S region of the translation 48S-PIC promotes RNA unwinding and scanning. *Elife* (2019). doi:10.7554/eLife.48246
332. Padrón, A., Iwasaki, S. & Ingolia, N. T. Proximity RNA Labeling by APEX-Seq Reveals the Organization of Translation Initiation Complexes and Repressive RNA Granules. *Mol. Cell* (2019). doi:10.1016/j.molcel.2019.07.030
333. Youn, J. Y. *et al.* High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Mol. Cell* (2018). doi:10.1016/j.molcel.2017.12.020
334. Nielsen, K. H. *et al.* Synergistic activation of eIF4A by eIF4B and eIF4G. *Nucleic Acids Res.* (2011). doi:10.1093/nar/gkq1206
335. Chan, K. *et al.* eIF4A supports an oncogenic translation program in pancreatic ductal adenocarcinoma. *Nat. Commun.* (2019). doi:10.1038/s41467-019-13086-5
336. Muller, D. *et al.* eIF4A inhibition circumvents uncontrolled DNA replication mediated by 4E-BP1 loss in pancreatic cancer. *JCI insight* (2019). doi:10.1172/jci.insight.121951
337. Liang, S. *et al.* Decreased expression of EIF4A1 after preoperative brachytherapy predicts better tumor-specific survival in cervical cancer. *Int. J. Gynecol. Cancer* **24**, 908–915 (2014).
338. Wolfe, A. L. *et al.* RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**, 65–70 (2014).
339. Cencic, R. *et al.* Antitumor activity and mechanism of action of the cyclopenta[b]benzofuran, silvestrol. *PLoS One* **4**, e5223 (2009).
340. Bordeleau, M.-E. *et al.* Stimulation of mammalian translation initiation factor eIF4A activity by a small molecule inhibitor of eukaryotic translation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10460–10465 (2005).
341. Bordeleau, M.-E. *et al.* Functional characterization of IRESes by an inhibitor of the RNA helicase eIF4A. *Nat. Chem. Biol.* **2**, 213–220 (2006).
342. Bordeleau, M. E. *et al.* Therapeutic suppression of translation initiation modulates chemosensitivity in a mouse lymphoma model. *J. Clin. Invest.* **118**, 2651–2660 (2008).
343. Lindqvist, L. *et al.* Selective pharmacological targeting of a DEAD box RNA helicase. *PLoS One* **3**, (2008).

344. Sun, Y. *et al.* Single-molecule kinetics of the eukaryotic initiation factor 4AI upon RNA unwinding. *Structure* (2014). doi:10.1016/j.str.2014.04.014
345. Low, W. K. *et al.* Inhibition of eukaryotic translation initiation by the marine natural product pateamine A. *Mol. Cell* **20**, 709–722 (2005).
346. Bordeleau, M. E. *et al.* RNA-Mediated Sequestration of the RNA Helicase eIF4A by Pateamine A Inhibits Translation Initiation. *Chem. Biol.* **13**, 1287–1295 (2006).
347. Dang, Y. *et al.* Inhibition of nonsense-mediated mRNA decay by the natural product pateamine a through eukaryotic initiation factor 4AIII. *J. Biol. Chem.* (2009). doi:10.1074/jbc.M109.009985
348. Chu, J. *et al.* CRISPR-Mediated Drug-Target Validation Reveals Selective Pharmacological Inhibition of the RNA Helicase, eIF4A. *Cell Rep.* **15**, 2340–2347 (2016).
349. Iwasaki, S., Floor, S. N. & Ingolia, N. T. Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor. *Nature* **534**, 558–561 (2016).
350. Iwasaki, S. *et al.* The Translation Inhibitor Rocaglamide Targets a Bimolecular Cavity between eIF4A and Polypurine RNA. *Mol. Cell* (2019). doi:10.1016/j.molcel.2018.11.026
351. Malina, A., Mills, J. R. & Pelletier, J. Emerging therapeutics targeting mRNA translation. *Cold Spring Harb. Perspect. Biol.* **4**, (2012).
352. Naineni, S. K. *et al.* A comparative study of small molecules targeting eIF4A. *RNA* (2020). doi:10.1261/rna.072884.119
353. Parsyan, A. *et al.* The helicase protein DHX29 promotes translation initiation, cell proliferation, and tumorigenesis. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 22217–22222 (2009).
354. Dhote, V., Sweeney, T. R., Kim, N., Hellen, C. U. T. & Pestova, T. V. Roles of individual domains in the function of DHX29, an essential factor required for translation of structured mammalian mRNAs. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E3150-9 (2012).
355. Pisareva, V. P. & Pisarev, A. V. DHX29 and eIF3 cooperate in ribosomal scanning on structured mRNAs during translation initiation. 1–12 (2016). doi:10.1261/rna.057851.116.in
356. Hashem, Y. *et al.* XStructure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell* **153**, 1108–1119 (2013).
357. Pisareva, V. P. & Pisarev, A. V. DHX29 reduces leaky scanning through an upstream AUG codon regardless of its nucleotide context. *Nucleic Acids Res.* **44**, 4252–4265 (2016).
358. Kozak, M. & Shatkin, A. J. Migration of 40S ribosomal subunits on messenger RNA in the presence of edeine. *J. Biol. Chem.* (1978).
359. Shirokikh, N. E., Archer, S. K., Beilharz, T. H., Powell, D. & Preiss, T. Translation complex profile sequencing to study the in vivo dynamics of mRNA–ribosome interactions during translation initiation, elongation and termination. *Nat. Protoc.* **12**, 697–731 (2017).
360. Slepenkov, S. V., Korneeva, N. L. & Rhoads, R. E. Kinetic mechanism for assembly of the m 7GpppG·eIF4E·eIF4G complex. *J. Biol. Chem.* (2008). doi:10.1074/jbc.M801786200
361. ASSELBERGS, F. A. M., PETERS, W., van VENROOIJ, W. J. & BLOEMENDAL, H. Diminished Sensitivity of Re-initiation of Translation to Inhibition by Cap Analogues in Reticulocyte Lysates. *Eur. J. Biochem.* (1978). doi:10.1111/j.1432-1033.1978.tb12473.x
362. Scheper, G. C. *et al.* Phosphorylation of eukaryotic initiation factor 4E markedly reduces its affinity for capped mRNA. *J. Biol. Chem.* (2002).

doi:10.1074/jbc.M103607200
363. Zuberek, J. *et al.* Influence of Electric Charge Variation at Residues 209 and 159 on the Interaction of eIF4E with the mRNA 5′ Terminus. *Biochemistry* (2004). doi:10.1021/bi030266t
364. Vassilenko, K. S., Alekhina, O. M., Dmitriev, S. E., Shatsky, I. N. & Spirin, A. S. Unidirectional constant rate motion of the ribosomal scanning particle during eukaryotic translation initiation. *Nucleic Acids Res.* **39**, 5555–5567 (2011).
365. Kozak, M. Selection of initiation sites by eucaryotic ribosomes: Effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. *Nucleic Acids Res.* (1984). doi:10.1093/nar/12.9.3873
366. van den Heuvel, J. J., Bergkamp, R. J. M., Planta, R. J. & Raué, H. A. Effect of deletions in the 5′-noncoding region on the translational efficiency of phosphoglycerate kinase mRNA in yeast. *Gene* (1989). doi:10.1016/0378-1119(89)90094-2
367. Matsuda, D. & Dreher, T. W. Close spacing of AUG initiation codons confers dicistronic character on a eukaryotic mRNA. *RNA* **12**, 1338–49 (2006).
368. Kumar, P., Hellen, C. U. T. & Pestova, T. V. Toward the mechanism of eIF4F-mediated ribosomal attachment to mammalian capped mRNAs. *Genes Dev.* **30**, 1573–1588 (2016).
369. Abaeva, I. S., Pestova, T. V. & Hellen, C. U. T. Attachment of ribosomal complexes and retrograde scanning during initiation on the Halastavi árva virus IRES. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw016
370. Shirokikh, N. E., Dutikova, Y. S., Staroverova, M. A., Hannan, R. D. & Preiss, T. Migration of small ribosomal subunits on the 50 untranslated regions of capped messenger RNA. *Int. J. Mol. Sci.* (2019). doi:10.3390/ijms20184464
371. Adhin, M. R. & van Duin, J. Scanning model for translational reinitiation in eubacteria. *J. Mol. Biol.* (1990). doi:10.1016/S0022-2836(05)80265-7
372. Skabkin, M., Skabkina, O., Hellen, C. T. & Pestova, T. Reinitiation and other unconventional posttermination events during eukaryotic translation. *Mol. Cell* **51**, 249–264 (2013).
373. Gunnery, S., Mäivali, Ü. & Mathews, M. B. Translation of an uncapped mRNA involves scanning. *J. Biol. Chem.* (1997). doi:10.1074/jbc.272.34.21642
374. Guenther, U. P. *et al.* The helicase Ded1p controls use of near-cognate translation initiation codons in 5′ UTRs. *Nature* (2018). doi:10.1038/s41586-018-0258-0
375. Aitken, C. E. & Lorsch, J. R. A mechanistic overview of translation initiation in eukaryotes. *Nature Structural and Molecular Biology* (2012). doi:10.1038/nsmb.2303
376. Marintchev, A. Roles of helicases in translation initiation: A mechanistic view. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* (2013). doi:10.1016/j.bbagrm.2013.01.005
377. Parsyan, A. *et al.* mRNA helicases: the tacticians of translational control. *Nat. Rev. Mol. Cell Biol.* **12**, 235–245 (2011).
378. Spirin, A. S. How does a scanning ribosomal particle move along the 5′-untranslated region of eukaryotic mRNA? Brownian ratchet model. *Biochemistry* (2009). doi:10.1021/bi901379a
379. Tuller, T., Ruppin, E. & Kupiec, M. Properties of untranslated regions of the S. cerevisiae genome. *BMC Genomics* (2009). doi:10.1186/1471-2164-10-391
380. Cheng, S., Sultana, S., Goss, D. J. & Gallie, D. R. Translation initiation factor 4B homodimerization, RNA binding, and interaction with poly(A)-binding protein are enhanced by zinc. *J. Biol. Chem.* (2008). doi:10.1074/jbc.M807716200
381. Kozak, M. Constraints on reinitiation of translation in mammals. *Nucleic Acids Res.* **29**, 5226–32 (2001).
382. Kozak, M. Effects of intercistronic length on the efficiency of reinitiation by eucaryotic

ribosomes. *Mol. Cell. Biol.* (1987). doi:10.1128/mcb.7.10.3438
383. Hinnebusch, A. G. Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. *Microbiol. Mol. Biol. Rev.* (2011). doi:10.1128/mmbr.00008-11
384. Luukkonen, B. G., Tan, W. & Schwartz, S. Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. *J. Virol.* (1995). doi:10.1128/jvi.69.7.4086-4094.1995
385. Pöyry, T. A. A., Kaminski, A. & Jackson, R. J. What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes Dev.* (2004). doi:10.1101/gad.276504
386. Mohammad, M. P., Pondelícková, V. M., Zeman, J., Gunisová, S. & Valásek, L. S. In vivo evidence that eIF3 stays bound to ribosomes elongating and terminating on short upstream ORFs to promote reinitiation. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx049
387. Lomakin, I. B. & Steitz, T. A. The initiation of mammalian protein synthesis and mRNA scanning mechanism. *Nature* (2013). doi:10.1038/nature12355
388. Visweswaraiah, J. & Hinnebusch, A. G. Interface between 40s exit channel protein uS7/Rps5 and eiF2α modulates start codon recognition in vivo. *Elife* (2017). doi:10.7554/eLife.22572
389. Fija-Lkowska, D. *et al.* EIF1 modulates the recognition of suboptimal translation initiation sites and steers gene expression via uORFs. *Nucleic Acids Res.* **45**, 7997–8013 (2017).
390. Saini, A. K. *et al.* Eukaryotic translation initiation factor eIF5 promotes the accuracy of start codon recognition by regulating Pi release and conformational transitions of the preinitiation complex. *Nucleic Acids Res.* (2014). doi:10.1093/nar/gku653
391. Terenin, I. M. *et al.* Sliding of a 43S ribosomal complex from the recognized AUG codon triggered by a delay in eIF2-bound GTP hydrolysis. *Nucleic Acids Res.* (2015). doi:10.1093/nar/gkv1514
392. Dikstein, R. Transcription and translation in a package deal: The TISU paradigm. *Gene* **491**, 1–4 (2012).
393. Elfakess, R. & Dikstein, R. A translation initiation element specific to mRNAs with very short 5???UTR that also regulates transcription. *PLoS One* **3**, (2008).
394. Haimov, O. *et al.* Efficient and Accurate Translation Initiation Directed by TISU Involves RPS3 and RPS10e Binding and Differential Eukaryotic Initiation Factor 1A Regulation. *Mol. Cell. Biol.* (2017). doi:10.1128/mcb.00150-17
395. Sinvani, H. *et al.* Translational tolerance of mitochondrial genes to metabolic energy stress involves TISU and eIF1-eIF4GI cooperation in start codon selection. *Cell Metab.* **21**, 479–492 (2015).
396. Martin, F. *et al.* Cap-Assisted Internal Initiation of Translation of Histone H4. *Mol. Cell* (2011). doi:10.1016/j.molcel.2010.12.019
397. Martin, F. *et al.* Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. *Nat. Commun.* (2016). doi:10.1038/ncomms12622

CHAPTER 2

Nonlinear Ribosome Scanning Controls the Stringency of Start Codon Selection

The contents of this Chapter are in submission for the manuscript to be published as Gu Y, Mao Y, Jia L, Dong L, and Qian SB. **Nonlinear Ribosome Scanning Controls the Stringency of Start Codon Selection.** Minor modifications have been made for reprint.

2.1 Abstract

The fidelity of start codon recognition by ribosomes is paramount during protein synthesis. The textbook knowledge of eukaryotic translation initiation depicts 5'→3' unidirectional migration of the pre-initiation complex (PIC) along the 5'UTR. In probing translation initiation from ultra-short 5'UTR, we report that an AUG triplet near the 5' end can be selected via PIC backsliding. The bi-directional ribosome scanning is supported by competitive selection of closely spaced AUG codons and recognition of two initiation sites flanking an internal ribosome entry site. Transcriptome-wide PIC profiling reveals footprints with a pattern of oscillation near the 5' end and start codons. Depleting the RNA helicase eIF4A leads to reduced PIC oscillations and impaired selection of 5' end start codons. Enhancing the ATPase activity of eIF4A promotes nonlinear PIC scanning and stimulates upstream translation initiation. The eIF4A-mediated PIC conformational switch may provide an operational mechanism that unifies ribosome recruitment, scanning, and start codon selection.

2.2 Introduction

Translation of eukaryotic mRNAs typically begins with the attachment of a 43S PIC to the capped 5' end, which is facilitated by a heterotrimeric complex eIF4F comprised of eIF4E, eIF4A and eIF4G[1,2,3]. While our knowledge of translation initiation is steadily increasing, molecular details of many key steps remain to be hammered out. During the first step of translation, for instance, how mRNA is accommodated into the ribosome remains surprisingly obscure. Studies using in vitro reconstituted translation proposed a 'threading' model[4], whereby the cap-binding protein eIF4E sits near the entry site and feeds the 5' mRNA cap into the mRNA channel (Figure 2-1A). This hypothesis was supported by the observation that 5'UTR has no 'blind spots' during scanning[4]. Although the 'threading' model permits base-by-base inspection of mRNA from the 5' end, it does not explain the poor recognition of start codons near the 5' end. The 'slot-in' model, by contrast, directly places eIF4E near the exit site of the mRNA-binding cleft followed by lateral attachment of mRNA to the 40S (Figure 2-1A). Perhaps the best evidence supporting the slotting mechanism comes from the internal ribosome entry site (IRES)-mediated cap-independent translation[5]. For cap-dependent translation, however, the 'slot-in' model is expected to leave a 'blind spot' of ~12 nucleotides (nt) from the 5' end. Another key difference between these two mutually exclusive pathways lies in the eIF4E kinetics after mRNA loading. For the 'threading' model to occur, eIF4E needs to be detached from the 5' end cap of mRNA. The 'slot-in' model, however, permits

continuous eIF4E engagement at the 5' end cap. While several lines of evidence supported cap-severed scanning[5,6,7], a recent study suggested cap-tethered ribosome scanning in human cells[8]. It remains unclear whether and how eIF4E dissociation from the cap occurs during mRNA accommodation.

Scanning commences once the PIC is loaded onto mRNA. It has been well-documented that stable hairpins in 5'UTR inhibits translation initiation, especially when the secondary structure is close to the 5' end[9]. The DEAD-box RNA helicase eIF4A is thought to prepare a single-stranded region near the 5' end of mRNA, thereby facilitating PIC attachment[10]. However, recent studies suggest that yeast eIF4A enhances recruitment of mRNAs regardless of their structural complexity[11]. Notably, the ATPase activity of eIF4A is markedly stimulated in the presence of the PIC, suggesting a ribosome-dependent function of eIF4A beyond its role in RNA unwinding. Indeed, eIF4A appears to modulate the conformation of the PIC to promote mRNA accommodation in an ATP-dependent but helicase-independent manner[12]. It has been proposed that cycles of ATP hydrolysis of eIF4A lead to rounds of 'closed' and 'open' conformations of eIF4A, which correspond to alterations of 'closed' and 'open' states of the PIC. Despite the fact that eIF4A is the most abundant translation initiation factor[13], structural and biochemical data do not unanimously outline the position and orientation of eIF4A in the PIC complexes. As a result, the exact role of eIF4A in PIC loading and scanning remains incompletely understood.

The scanning process is arguably the most enigmatic stage during translation initiation. A widely accepted view indicates 5'→3' directional movement of the PIC along 5'UTR[14]. However, the underlying driving force remains controversial and the determinants of the scanning directionality have been the subject of much speculation[15]. To complicate things further, noncanonical scanning modes like ribosome jumping or shunting have been documented[16,17,18]. Those cases of nonlinear scanning, albeit exceptional, serve as warnings against the over-simplified scanning model. Due to the highly dynamic nature of the process, a direct characterization of scanning has proven to be challenging. The recent development of translation complex profiling (TCP-seq) revealed heterogeneous footprint sizes for scanning PICs[6]. However, a snapshot of PIC footprints does not reveal the scanning directionality.

Proper recognition of mRNA start codons is crucial to produce desired proteins. It is commonly assumed that the first AUG codon that the scanning PIC encounters servers as the translation initiation site (TIS). This conventional wisdom has been challenged recently by many studies reporting alternative translation events initiated from multiple AUG codons as well as non-AUG triplets[19,20]. The stringency of TIS selection is influenced by cis-sequence elements and trans-acting factors, both of which affect the kinetics of PIC scanning. For instance, impediments to PIC scanning by stable hairpins enhances the recognition of suboptimal start codons that otherwise would be skipped[21,22]. On the other hand, even the most optimal start codon is not

recognized at 100%. The subsequent leaky scanning implies the 5' preference of start codon selection, which has long been attributed to the 3'-ward linear PIC migration. However, such 5' polarity does not always hold true for closely spaced start codons[23], with downstream TIS negatively affects the upstream start codon. The competitive nature of neighboring start codons suggests an intriguing possibility of bi-directional PIC scanning, although the molecular dynamics of nonlinear PIC migration remains poorly understood.

Here, using comprehensive mRNA reporters, we provide additional lines of evidence supporting bi-directional ribosome scanning. A combination of genome-wide PIC and TIS profiling reveals an unexpected role of eIF4A in driving nonlinear PIC scanning, which directly influencing the stringency of start codon selection. The ATPase-mediated conformation switch of ribosomes provides an operational mechanism that unifies lateral PIC recruitment, nonlinear scanning, and alternative start codon selection.

2.3 Results

2.3.1 Recognition of Start Codons Near the 5' End Cap

To address whether eukaryotic translation initiation follows 'threading' or 'slot-in' mechanisms for PIC attachment (Figure 2-1A), it is important to determine the minimal length of 5'UTR. Although in vitro studies have ruled out blind spots for start codon selection[24], whether mRNAs inside cells exhibit the same feature remains unclear. We constructed firefly luciferase (Fluc) reporters

harboring varied length of 5'UTR. To gain an accurate measurement of start codon selection, we (i) mutated the in-frame AUG codons of Fluc to minimize downstream initiation (Figure 2-S1A); (ii) conducted transfection of mRNAs synthesized in vitro to eliminate transcript isoforms (Figure 2-1B); (iii) employed a real-time luminometer to monitor Fluc levels above the background (Figure 2-S1B); (iv) performed RT-qPCR at end points to factor out variations of mRNA stability. Among reporters with increasing 5'UTR length, a binary phase of translation efficiency is evident with the robust translation achieved when the 5'UTR length reaches 20 nt or above (Figure 2-1B, right panel). Intriguingly, Fluc mRNAs with ultra-short 5'UTR (e.g., 2nt) are clearly translatable. The relatively low translation efficiency could be due to the suboptimal context missing the –3 purine.

For mRNA reporters with ultra-short 5'UTR, the observed Fluc activities could be derived from downstream initiation using non-AUG codons. To rule out this possibility, we devised an independent reporter system bearing an upstream open reading frame (uORF) and a downstream GFP (Figure 2-1C). The uORF-encoded tracer peptide (SIINFEKL) is efficiently presented on the cell surface by mouse major histocompatibility complex class I molecules H-2K^{b} [25]. A monoclonal antibody 25D1 recognizes the K^{b}-SIINFEKL complex with exquisite sensitivity[26], thereby offering a direct measurement of the uORF product. Upon transfection into HEK293 cells stably expressing H-2K^{b}, both the tracer peptide and GFP can be quantified via flow cytometry. Like Fluc reporters,

a 2-nt long 5'UTR enables uORF translation as evidenced by the elevated 25D1 signals compared to the GFP control (Figure 2-1C). We obtained similar results in cells transfected with plasmid-based reporters (Figure 2-S1C). The translation of mRNAs with ultra-short 5'UTR is strictly cap-dependent (Figure 2-S1D), excluding the possibility of non-canonical ribosome entry. To confirm the exact positions of start codons selected by PIC, we conducted toe-printing assays using rabbit reticulum lysates (RRL) (Figure 2-1D). Typical start codon-associated peaks appeared at the leading edge of initiating ribosomes immobilized by cycloheximide (CHX). We observed those peaks from the very first AUG triplet of all the mRNA templates regardless of the 5'UTR length. Therefore, start codons near the 5' end can be reached by the PIC in vitro and inside cells, albeit with low efficiency.

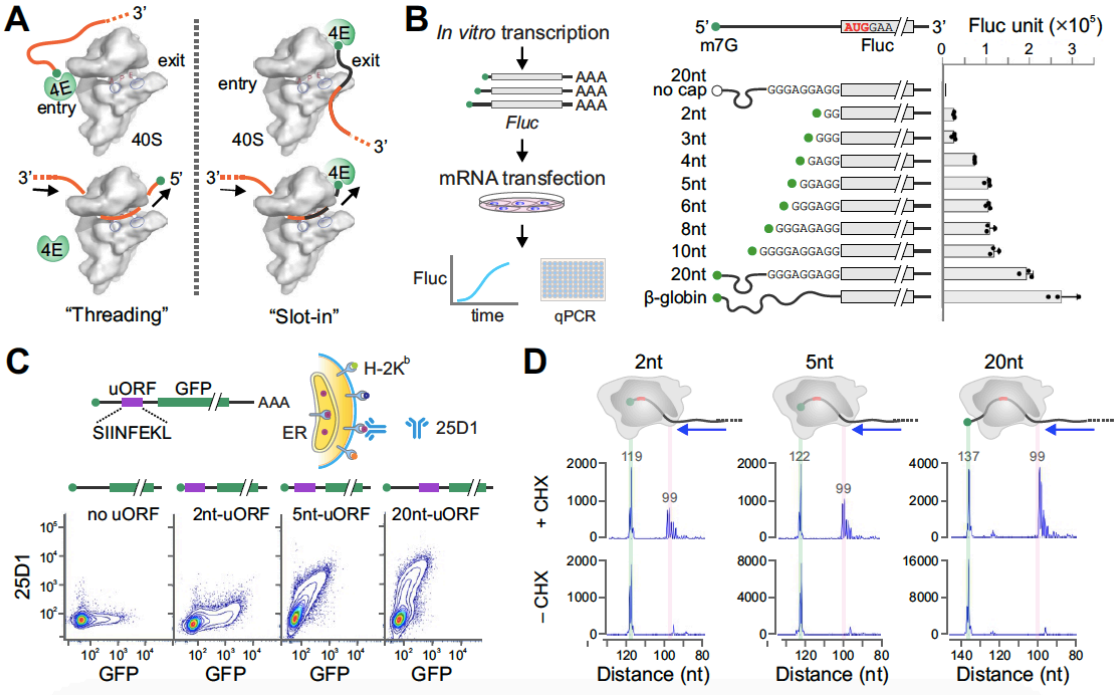

Figure 2-1 Selection of start codons near the 5' end cap of mRNAs.
(A) Schematic of mRNA entering the 40S via the 'threading' mode (left) or the 'slot-in' mode (right). Only eIF4E is shown for simplicity. The black portion of mRNA indicates the 'blind spot'.
(B) Translation efficiency of Fluc mRNA reporters bearing ultra-short 5'UTR. The left panel shows the schematic of Real-time Luciferase Assay. In vitro synthesized mRNAs were transfected into MEF cells and Fluc activities were monitored by real-time luminometry. Fluc activities at 6 hr were normalized to mRNA levels quantified by RT-qPCR. Uncapped mRNAs were included as negative control. Error bars: mean ± SEM; n = 3 biological replicates.
(C) Translation efficiency of uORF mRNA reporters bearing ultra-short 5'UTR. The top panel shows the schematic of uORF reporter assay coupled with FACS. The uORF reporter contains a sequence encoding SIINFEKL (purple) followed by GFP (green). Synthesized mRNA reporters were transfected into HEK293-Kb cells, followed by measurement of 25D1 and GFP using flow cytometry. The bottom panel shows the representative flow cytometry scatterplots of HEK293-Kb cells transfected with uORF reporters with varied length of 5'UTR. mRNAs with no uORF were included as negative control.
(D) Start codon recognition by toe-printing assays. Fluc mRNA reporters with varied length of 5'UTR were incubated in RRL, followed by reverse transcription using a probe downstream of the start codon. Expected positions corresponding to the full length and the ribosome leading edge are highlighted.

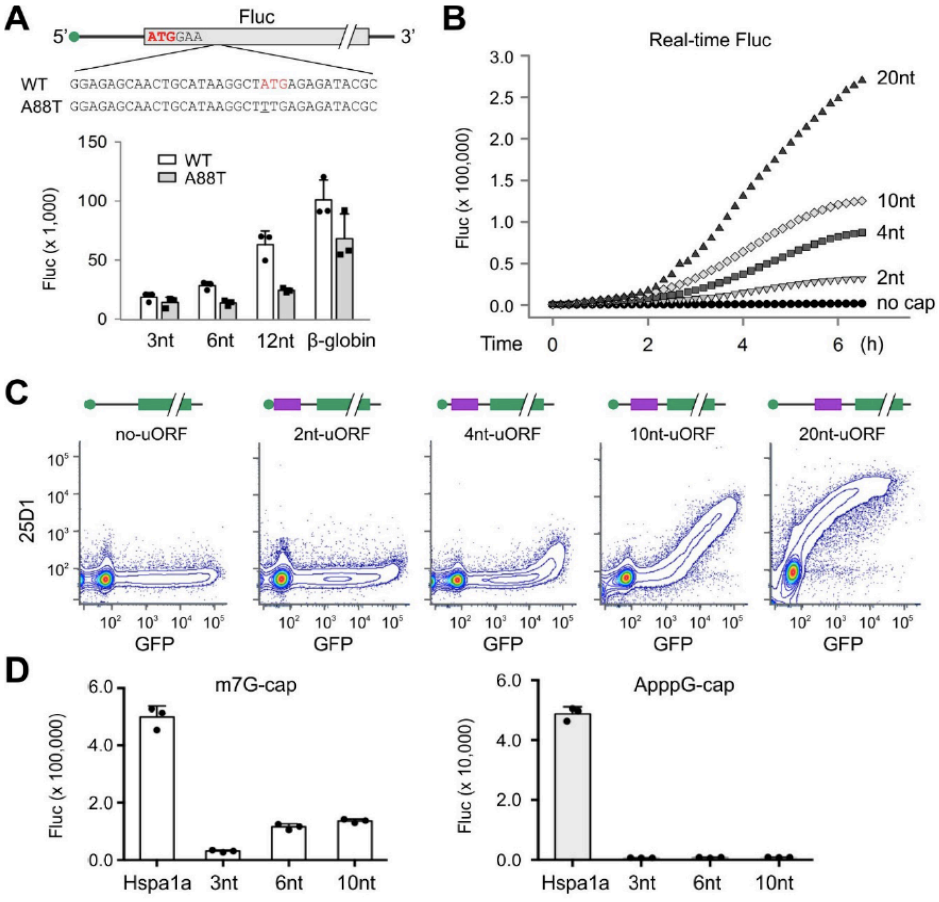

Figure 2-S1 Characterization of Fluc mRNA reporters with ultra-short 5'UTR.
(A) The top panel shows the schematic of the wild type Fluc and the A88T mutant lacking the downstream AUG codon. The bottom panel shows the quantification of Fluc activities of WT or A88T mRNA reporters with varied 5'UTR length in transfected MEF cells. Error bars; mean ± SEM; n=3 biological replicates.
(B) Representative real-time luminometry of MEF cells transfected with Fluc mRNA reporters bearing 5'UTR of the indicated length. Uncapped mRNA with 20-nt 5'UTR was included as negative control.
(C) Representative flow cytometry scatterplots of HEK293-Kb cells transfected with plasmids of uORF reporters with 5'UTR of the indicated length. Plasmid with GFP only was included as negative control.
(D) Fluc mRNA reporters capped with m7G (the left panel) or the cap analog ApppG (the right panel) with varied 5'UTR length were transfected into MEF cells followed by quantification of Fluc activities. Error bars; mean ± SEM; n=3 biological replicates.

2.3.2 Tethering eIF4E Prevents Selection of Start Codons Near the 5' End Cap

The lack of an absolute 'blind spot' in 5'UTR for start codon selection is seemingly incompatible with the 'slot-in' model. If the mRNA follows the 'threading' model for start codon selection, tethering eIF4E to the 5' end m^7G cap would prevent the subsequent PIC attachment. We synthesized a Fluc reporter mRNA capped with m^7s^6G (Figure 2-S2A), which could be crosslinked to its binding protein eIF4E under UV 360 nm (Figure 2-2A)[27]. In the absence of crosslinking, m^7s^6G behaves like the normal m^7G cap as evidenced by the comparable Fluc levels in RRL (Figure. 2-S2B). When purified eIF4E was crosslinked to the m^7s^6G cap, robust Fluc translation was readily observed for mRNAs bearing the β-globin 5'UTR (Figure 2-2B, top panel). This result cannot be explained by the 'threading' model because the constitutive presence of eIF4E at the 5' end would prevent mRNA from entering into the narrow mRNA channel (Figure 2-1A). Intriguingly, the non-dissociable eIF4E nearly abolished the translation of Fluc mRNAs with ultra-short 5'UTR (Figure 2-2C, top panel).

To exclude possible side effects of UV crosslinking, we applied a chemical crosslinking methodology originally designed to identify eIF4E (Figure 2-S2C)[28], and obtained the similar results (Figure 2-2B and 2C, bottom panel). The re-appearance of the 'blind spot' when the 5' cap is tethered with eIF4E suggests that mRNA is loaded into the PIC via 'slot-in'.

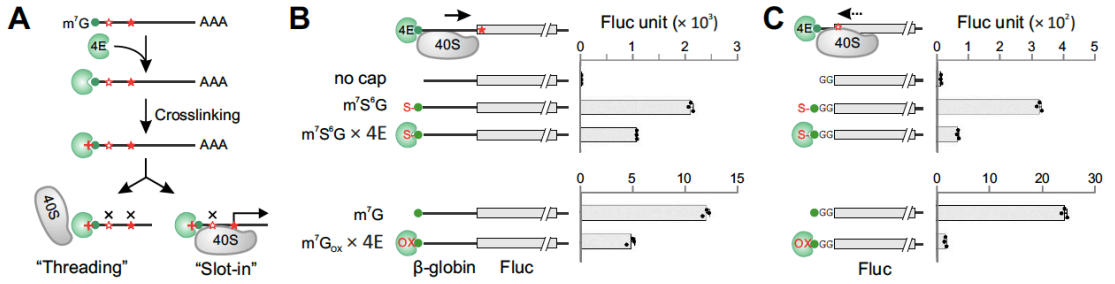

Figure 2-2 Bi-directional scanning of PIC contributes to start codon selection.
(A) Schematic of cross-linking between m7G cap and eIF4E. Tethering eIF4E to m7G cap will prevent start codon selection in the 'threading' model, while permitting recognition of start codon downstream of the 'blind spot' in the 'slot-in' model.
(B) Translational effects of cross-linking between m7G cap and eIF4E using Fluc mRNAs with β-globin 5'UTR. The top panel used m7s6G capped mRNAs followed by cross-linking under UV 360 nm. The bottom panel used redox reactions for cross-linking. In vitro translation was conducted in RRL and Fluc activities were monitored by luminometry. Error bars; mean ± SEM; n=3 biological replicates.
(C) Translational effects of cross-linking between m7G cap and eIF4E using Fluc mRNAs with ultra-short 5'UTR (2 nt). The top and bottom panels are similar to (B). Error bars; mean ± SEM; n = 3 biological replicates.

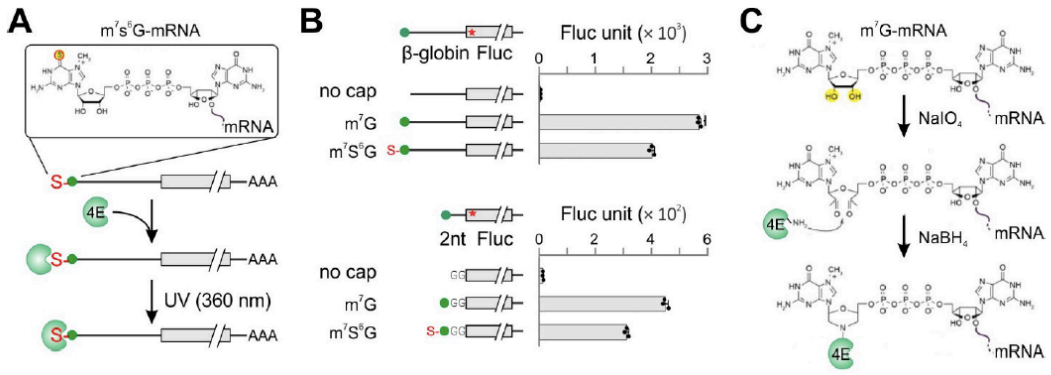

Figure 2-S2 Cross-linking between m^7G cap and eIF4E for Fluc mRNA reporters.
(A) Schematic of cross-linking between m^7s^6G capped mRNAs with purified eIF4E under UV 360 nm. In vitro synthesized mRNAs were capped with 6-Thio-GTP using capping enzyme system. m^7s^6G capped mRNAs were incubated with purified eIF4E in the binding buffer at 4°C under 360 nm UV exposure for 15 min.

(B) Characterization of m^7s^6G capped mRNA reporters. Uncapped mRNA, m^7G or m^7s^6G capped mRNAs with either β-globin (upper) or 2-nt (bottom) 5'UTR were incubated with RRL. Fluc activities were monitored by luminometry. Error bars; mean ± SEM; n=3 biological replicates.
(C) Schematic of cross-linking between m7G capped mRNAs with purified eIF4E via redox reaction. m^7G capped mRNA are oxidized in presence of sodium periodate, followed by ethanol precipitation. Oxidized mRNA was incubated with purified eIF4E in the binding buffer at 4°C for 15 min. Sodium borohydride was added to the mixtures, followed by incubation for 2-3 hr on ice.

2.3.3 Competitive Selection of Start Codons Suggests Bi-directional Scanning of PIC

An immediate question is how the PIC loaded downstream can select an upstream AUG triplet within the 'blind spot'. One possibility is that the PIC relies on 3'→5' movement to scan the 5' end region. The backward excursion of ribosomes has been suggested from experiments using closely spaced AUG codons found in *Turnip yellow mosaic virus* (TYMV)[23]. We conducted similar experiments using toe-printing assays to quantify the relative usage of two AUG codons having identical flanking sequences (Figure 2-3A). Indeed, placing an AUG codon 4 nt downstream reduced the selection of the first AUG. Importantly, with increasing spacer lengths, we observed a trend of ascending recognition of the first AUG and descending recognition of the second (Figure 2-3A). The competitive relationship between neighboring AUG codons is incompatible with the linear scanning model. Only when the scanning ribosome undergoes back-and-forth oscillations, could the downstream AUG codon influence the upstream one. To substantiate this finding further, we placed three AUG codons into the 'blind spot' with a single G in between. This design ensures that all AUG triplets are in different reading frames while retaining the same sequence context

(Figure 2-S3A). It is clear that AUG3 exhibited the most efficient initiation (>3 fold) in transfected cells. We also constructed mRNA reporters with the uORF driven by each AUG codons and obtained the similar results (Figure 2-S3B). Unlike the mRNA 'threading' model that would follow the first-AUG rule in start codon selection, the more favorable recognition of the AUG3 supports reverse scanning of the PIC.

To directly demonstrate that the PIC is capable of bi-directional scanning, we constructed uORF reporters by inserting an internal ribosome entry site (IRES) element between the tracer peptide SIINFEKL and GFP (Figure 2-S3C). The poliovirus IRES was chosen because the class I IRES element follows the canonical initiation mechanism for PIC assembly and scanning[5]. In the absence of IRES, the non-functional cap analog ApppG supported neither uORF nor GFP translation in transfected HEK293-Kb cells (Figure 2-S3D). Remarkably, the presence of the polio- IRES element after the uORF not only mediated downstream GFP translation, but also enabled uORF translation as evidenced by the elevated 25D1 signals (Figure 2-3B). To rule out the possibility that the uORF translation was a result of forward scanning of the PIC looped from IRES, we reversed the orientation of the polio-IRES element (Figure 2-S3C). In spite of the identical sequence, neither uORF nor GFP showed any translational activity (Figure 2-3B). When the distance between uORF and IRES was increased from 6 nt to 18 nt, the translation potential of uORF but not GFP was

reduced. This result supports the backward migration of PIC recruited from the IRES.

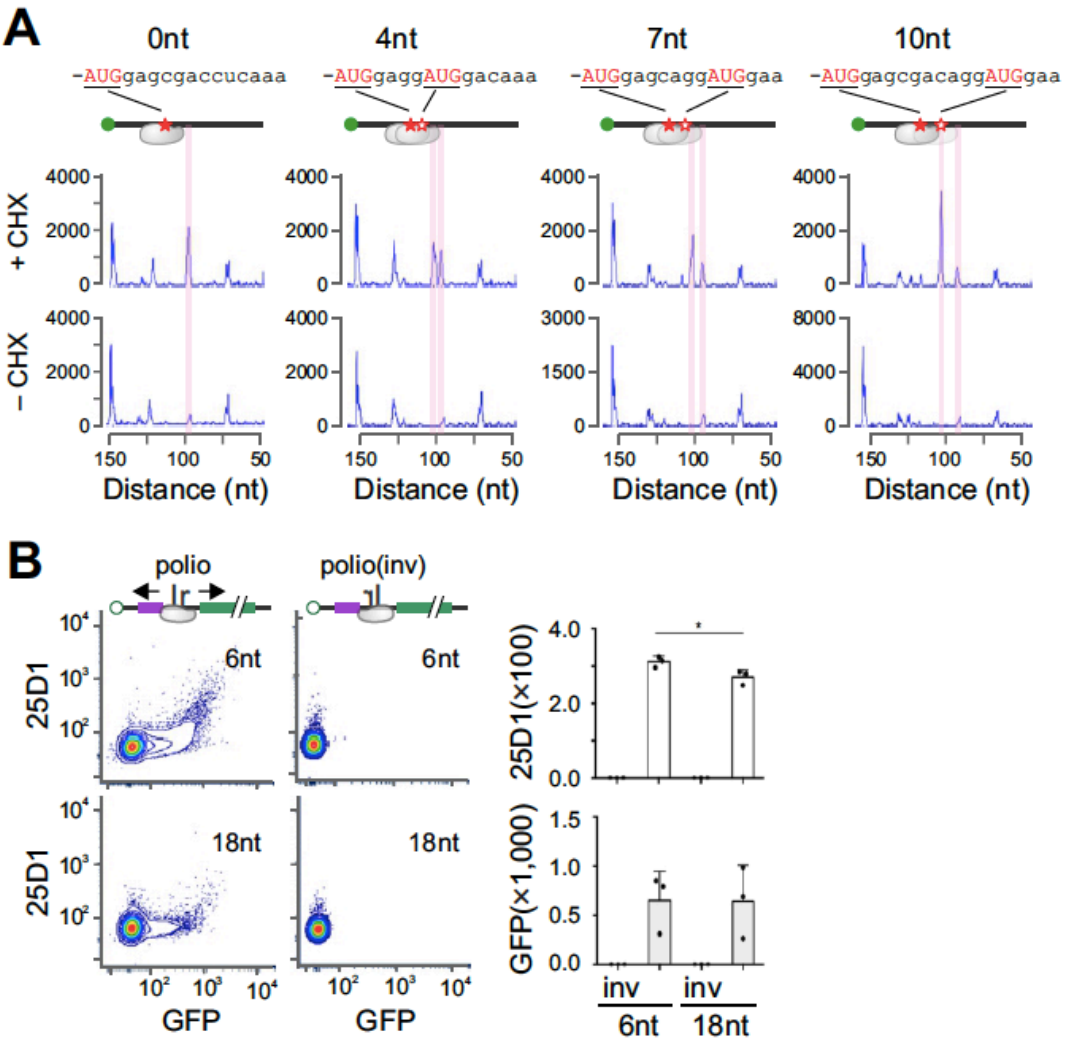

Figure 2-3 Selection of Neighboring Start Codons by Bi-directional PIC Scanning
(A) Selection of closely spaced AUG codons by toe-printing assays. Two AUG codons with identical flanking sequence but different length of the spacer were inserted into Fluc mRNAs with β-globin 5'UTR. Expected positions corresponding to the ribosome leading edge are highlighted. The relative peak heights correspond to the efficiency of start codon selection.
(B) Selection of start codons flanking the IRES. The polio IRES was inserted between the uORF encoding SIINFEKL and GFP with 6 nt or 18 nt downstream of the uORF stop codon. The mRNA reporters were capped with non-functional ApppG. The left panel shows the representative flow cytometry scatterplots of HEK293-Kb cells transfected with mRNA reporters bearing different IRES orientations. The right panel shows the quantification of GFP and 25D1 signals. Error bars: mean ± SEM; n = 3 biological replicates. One-way ANOVA, * $p < 0.05$.

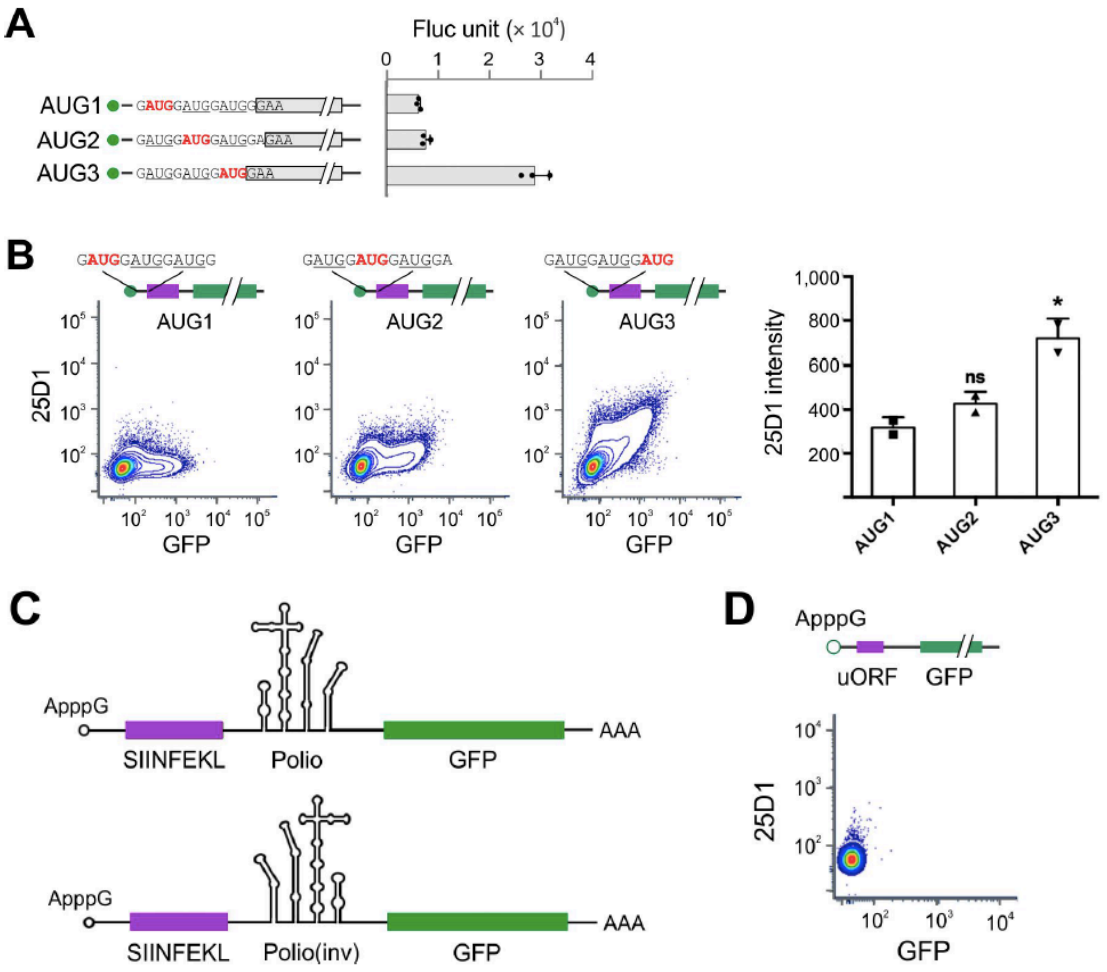

Figure 2-S3 Bi-directional scanning of PIC contributes to start codon selection.
(A) Fluc mRNA reporters containing 3 AUG codons near the 5' end were transfected into MEF cells, followed by quantification of Fluc activities by real-time luminometry. Error bars; mean ± SEM; n=3 biological replicates.
(B) Representative flow cytometry scatterplots of HEK293-Kb cells transfected with uORF reporters containing 3 AUG codons near the 5'end. AUG1 (left), AUG2 (middle) or AUG3 (right) is in the same fame of uORF encoding SIINFEKL (purple). The right panel shows the quantification of GFP and 25D1 fluorescence intensity from transfected HEK293-Kb cells. Error bars: mean ± SEM; n = 3 biological replicates. One-way ANOVA, * $p < 0.05$.
(C) Schematic of the uORF reporters with the polio IRES inserted between uORF and GFP. The bottom panel shows the same reporter with an inversed IRES element.
(D) Representative flow cytometry scatterplots of HEK 293-Kb cells transfected with uORF reporters with neither m7G cap nor the IRES element.

2.3.4 Transcriptome-Wide PIC Profiling Supports Bi-directional Scanning

A previous study developed translation complex profiling (TCP-seq) to track scanning ribosomes in 5'UTR[6]. We reasoned that the pattern of PIC footprints near the 5' end of transcripts might help distinguish different modes of mRNA accommodation. For instance, the 'slot-in' model is expected to leave a uniform width of PIC footprints from the 5' end cap, resulting in parallel distribution of 5' and 3' end of reads (Figure 2-4A). The threading model, by contrast, gives rise to increasing lengths of PIC footprints from the transcription start site (TSS), forming a 5' end peak only. Bi-directional scanning after 'slot-in', however, would form a 3' end bump in addition to the 5' end peak at TSS. In querying the TCP-seq data sets obtained from human cells[29], we noticed that the size of PIC footprint near the mRNA cap (or TSS) is surprisingly broader than that in 5'UTR (Figure 2-4B, left panel). The longer footprint size of the PIC near the TSS argues that the initiation complex loaded via the 5' end cap spans a broader mRNA region. Mapping the 3' end position of PIC footprints revealed a clear bump (Figure 2-4B, right panel), suggesting bi-directional PIC scanning after slot-in. The same feature holds true for the original TCP-seq conducted in budding yeast (Figure 2-S4A). A recent study reported 40S profiling in zebra fish using a similar approach called ribosome complex profiling (RCP-seq)[30]. Despite the similar findings as TCP-seq, the authors concluded that the mRNA recruitment follows the threading model. We found that the same results could be interpreted as bi-directional PIC scanning after slot-in (Figure 2-S4B). To

substantiate this finding further, we explored selective TCP-seq associated with eIF3B[29]. The eIF3-associated PIC footprints near the mRNA 5' end not only showed broader sizes, but also displayed a typical bump for the 3' end of reads (Figure 2-S4C).

TCP-seq relies on sucrose gradient sedimentation to separate crosslinked 40S, which likely excludes initiation complexes with larger sizes. To capture scanning PICs in a more comprehensive manner, we modified TCP-seq by enriching eIF3-associated ribosome complexes from mouse embryonic fibroblast (MEF) cells after formaldehyde crosslinking and RNase I digestion (Figure 2-S4D). Compared to the standard Ribo-seq that shows a typical read length of ~29 nt, eIF3-seq uncovered a broader range of read length especially in 5'UTR (Figure 2-S4D). Like TCP-seq, eIF3 footprints near the TSS showed a broader read length and a unique 3' end bump (Figure 2-4C, top panel).

We next examined the feature of eIF3 footprints near the annotated start codons, which could shed light on PIC dynamics during start codon selection. As expected, Ribo-seq showed single 5' and 3' end peaks corresponding to the boundary of the 80S engaged at start codons (Figure 2-4D, left panel). Intriguingly, eIF3 footprints showed several distinct peaks for 5' and 3' ends of reads (Figure 2-4D, middle panel). Upstream PICs are discernable from 3' end peaks, whereas downstream PICs are evident from peaks formed by the 5' end of reads. The 5' end peak downstream of the start codon was also evident in published TCP-seq data sets using HEK293 cells (Figure 2-S4E). The eIF3

footprints flanking the start codon likely represent oscillating PICs averaging about 12 nt.

It is possible that the eIF3 footprints downstream of the start codon stem from 80S ribosomes in the early stage of elongation[8,29]. To distinguish eIF3:40S from eIF3:80S complexes, we subtracted 80S footprints from the library of total RNA fragments to enrich PIC footprints (Figure 2-S4D). Indeed, PIC-seq uncovered more 5'UTR reads than eIF3-seq with broad read populations ranging from 19 nt to >60 nt. Once again, PIC-seq revealed a broader read length near the TSS and a bump of 3' end read accumulation (Figure 2-4C, bottom panel), echoing the bi-directional scanning of PIC after slot-in. When transcripts are aligned to the start codon, PIC footprints showed more prominent peaks downstream and upstream of the start codon (Figure 2-4D). These results support the oscillating behavior of PIC during start codon recognition, which potentially explains the competitive selection of neighboring start codons.

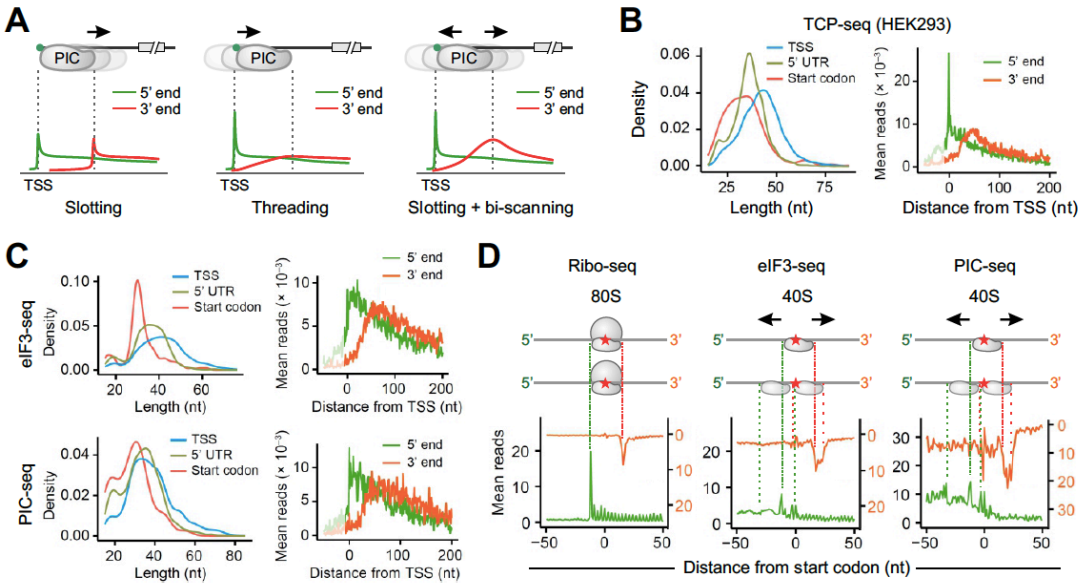

Figure 2-4 Genome-Wide PIC Profiling Supports Bi-directional Scanning.
(A) Schematic representation of 'threading', 'slotting', and 'slotting + bi-scanning' models for PIC loading (top panel) and their corresponding read distribution of 5' end (green) and 3' end (red) of footprints.
(B) Reanalysis of TCP-seq data sets obtained from HEK293 cells (Wagner et al). The left panel shows the length distribution of reads mapped to TSS, 5'UTR, or near start codons. The right panel shows the aggregation plots of 5' end (green) and 3' end (red) of reads on mRNAs aligned to TSS.
(C) MEF cells were used for eIF3-seq (top panels) and PIC-seq (bottom panels) followed by analysis of read length distribution (left panels) and aggregation plots (right panels) of reads on mRNAs aligned to TSS.
(D) Characterization of ribosome footprints from Ribo-seq, eIF3-seq, and PIC-seq data sets in MEF cells. Both the 5′ end (green) and 3′ end (red) of reads were mapped to transcripts aligned to annotated start codons (red star).

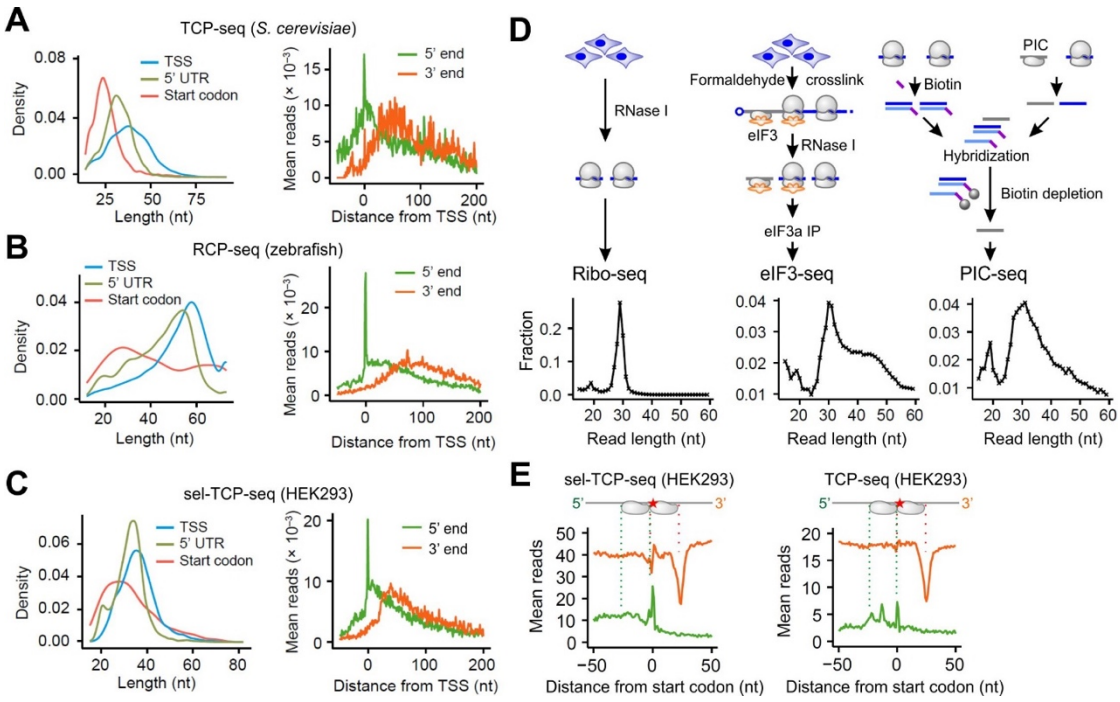

Figure 2-S4 Characterizing PIC Footprints.
(A) Reanalysis of TCP-seq data sets obtained from S. cerevisiae (Archer et al). The left panel shows the length distribution of reads mapped to TSS, 5'UTR, or near start codons. The right panel shows the aggregation plots of 5' end (green) and 3' end (red) of reads on mRNAs aligned to TSS.
(B) Reanalysis of RCP-seq data sets obtained from zebrafish (Giess et al).
(C) Reanalysis of sel-TCP-seq data sets obtained from HEK293T cells (Wagner et al).
(D) Schematic of Ribo-seq, eIF3-seq, and PIC-seq. For Ribo-seq, polysomes were separated from whole cell lysates using sucrose gradients. Collected polysome fractions were digested with RNase I followed by deep sequencing of ribosome-protected mRNA fragments. For eIF3-seq, cells were fixed by formaldehyde followed by polysome separation and RNase I digestion. Collected 40S and 80S fractions were subjected to immunoprecipitation using anti-eIF3a

antibodies. Purified eIF3-associated ribosome footprints were subjected to deep sequencing. For PIC-seq, cells were fixed by formaldehyde followed by polysome separation and RNase I digestion. Purified ribosome footprints from 80S fractions were reverse transcribed using biotinylated primer. The biotin labeled cDNA was hybridized to the purified footprints from 40S fractions. After depletion by streptavidin beads, the PIC-associated ribosome footprints were enriched and subjected to deep sequencing. The bottom panels are read length distribution for 5'UTR reads obtained from Ribo-seq, eIF3-seq and PIC-seq.
(E) Characterization of PIC footprints from sel-TCP-seq and TCP-seq obtained from HEK293T cells (Wagner et al). Both the 5′ end (green) and 3′ end (red) of reads were mapped to transcripts aligned to annotated start codons (red star).

2.3.5 eIF4A Facilitates Selection of Start Codons Near the 5' End Cap and Free Scanning of the PIC

It is unclear whether the bi-directional scanning of PIC is a spontaneous act or a guided event. The ATP-dependent helicase eIF4A has been widely believed to provide scanning directionality by resolving mRNA secondary structures in 5'UTR[31]. We deleted Eif4a1 from MEF cells using CRISPR/Cas9 gene editing and observed a marked disassembly of polysomes (Figure 2-S5A and S5B). Consistently, both cell growth and global protein synthesis were reduced in the absence of eIF4A1 (Figure 2-S5C and S5D). Ribo-seq also revealed global reduction of translation (Wilcoxon-test, P < 2.2Å~10-16) when the total amount of footprints was normalized to mitochondria (Figure 2-5A), whose translation is independent of eIF4A. When individual mRNAs were considered, however, eIF4A1 depletion showed little mRNA specificity (Figure 2-S5E). This finding agrees with the recent proposal that eIF4A promotes ribosome loading to nearly all mRNAs regardless of their structural complexity[11]. We next examined the role of eIF4A1 in the scanning process by measuring the translation of mRNA reporters with varied 5'UTR length. Unexpectedly,

messengers with ultra-short 5'UTR were more sensitive to eIF4A1 depletion than those with long 5'UTR (Figure 2-5B and Figure 2-S5F). We further confirmed that, in mRNA reporters bearing 3 AUG codons near the 5' end, AUG1 was more susceptible to eIF4A1 silencing than AUG3 (Figure 2-5C and Figure 2-S5G). These results suggest that eIF4A1 facilitates recognition of start codons near the 5' end cap.

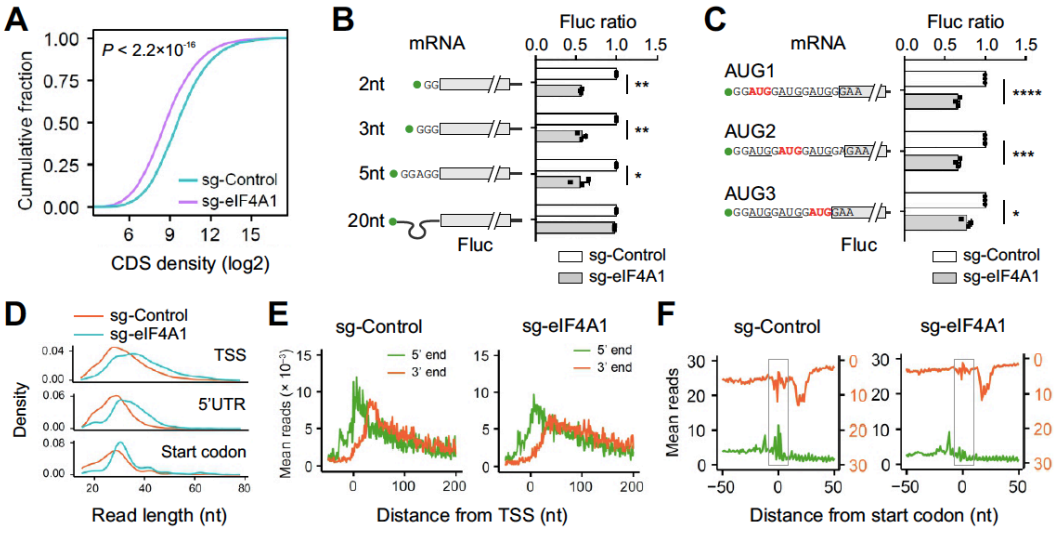

Figure 2-5 The RNA helicase eIF4A1 Promotes Free Scanning of the PIC.
(A) Evaluation of global translation by Ribo-seq in MEF cells with or without eIF4A1 knockdown. Ribosome densities on individual transcripts were normalized to the number of mitochondrial reads and plotted as cumulative fractions.
(B) Translation efficiency of Fluc mRNA reporters with ultra-short 5'UTR in MEF cells with or without eIF4A1 knockdown. Fluc activities in the absence of eIF4A1 were normalized to the control. Error bars: mean ± SEM; n = 3 biological replicates. t-test, * $p < 0.05$; *** $p < 0.001$.
(C) Translation efficiency of Fluc mRNA reporters containing 3 AUG codons near the 5' end in MEF cells with or without eIF4A1 knockdown. Fluc activities in the absence of eIF4A1 were normalized to the control. Error bars: mean ± SEM; n=3 biological replicates. t-test, * $p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$.
(D) Comparison of read length distribution in cells with or without eIF4A1 knockdown. Reads mapped to TSS, 5'UTR, or start codons are plotted separately.
(E) Aggregation plots of 5' end (green) and 3' end (red) of reads on mRNAs aligned to TSS using eIF3-seq data sets obtained from cells with or without eIF4A1 knockdown.
(F) Aggregation plots of 5' end (green) and 3' end (red) of reads on mRNAs aligned to start codons using eIF3-seq data sets obtained from cells with or without eIF4A1 knockdown. Highlighted squares indicate upstream footprints (3' end of reads) and downstream footprints (5' end of reads).

As a DEAD-box RNA-dependent ATPase, eIF4A1 has been shown to unwind RNA duplexes in a bi-directional manner[13,32]. Although the static view of PIC footprints is not suitable to infer the scanning directionality of PIC, the PIC conformation can be assessed by the length of footprints with an open conformation having shorter footprints[6]. Given the coordination between ATPase-dependent eIF4A and ribosome conformations[12], it is possible that the lack of eIF4A1 leads to more closed PIC conformation. To probe PIC conformation in the absence of eIF4A1, we conducted eIF3-seq and PIC-seq. Intriguingly, depletion of eIF4A1 resulted in longer reads in TSS and 5'UTR (Figure 2-5D). This was not due to sample variation because footprints mapped at the start codon have comparable read length. The longer PIC footprints in the absence of eIF4A1, albeit counter-intuitive, is in line with the more closed PIC conformation. For PIC footprints near the TSS, the bump formed by the 3' end of reads was reduced followed by a relatively increased read density downstream (Figure 2-5E). This could be due to reduced PIC backsliding and/or delayed scanning in the absence of eIF4A1. Indeed, PIC-seq revealed higher read density across the 5'UTR in cells lacking eIF4A1 (Figure 2-S5H). The stalled PIC in 5'UTR likely sequester other translation initiation factors like eIF3. This was indeed the case, immunoblotting of ribosome fractions showed a relocation of eIF3a from the ribosome-free fraction to the 40S fraction upon eIF4A1 depletion (Figure 2-S5I).

We next examined whether silencing eIF4A1 affects the PIC behavior in start codon recognition. Compared to the control cells, eIF4A1 depletion nearly eliminated downstream reads (judged from 5' end) and upstream reads (judged from 3' end) relative to the start codon (Figure 2-5F). These results strongly support the notion that eIF4A1 promotes free scanning of PIC by enabling an open conformation during start codon recognition. Without eIF4A1, the closed PIC conformation not only delays scanning, but also limits competitive selection of neighboring start codons.

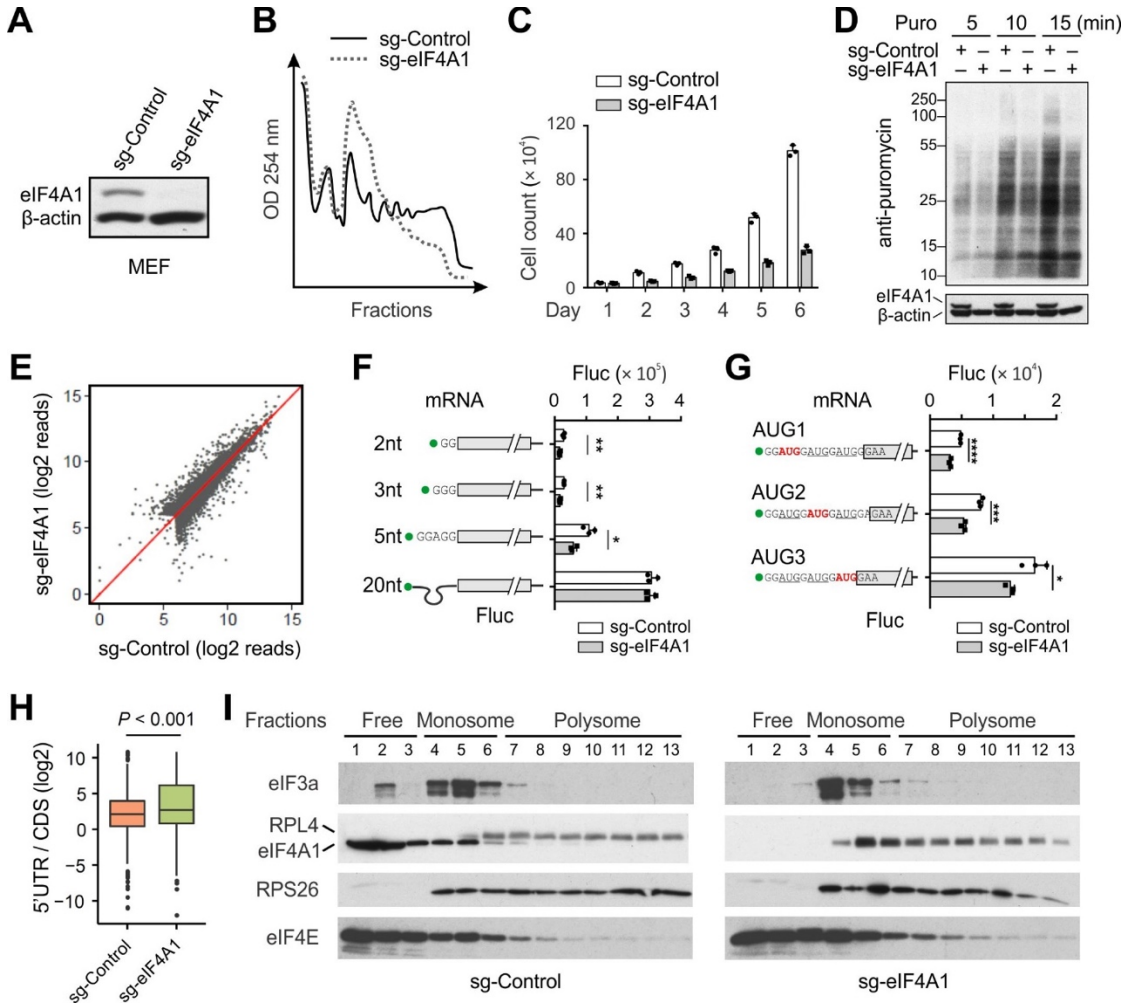

Figure 2-S5 Characterization of MEF cells with eIF4A1 knockdown.
(A) Immunoblotting of MEF cells with or without eIF4A1 knockdown.
(B) Polysome profiles of MEF cells with or without eIF4A1 knockdown were analyzed by sucrose density gradient centrifugation in polysome buffer.
(C) Proliferation rates of MEF cells with or without eIF4A1 knockdown. Cell numbers were normalized to the value obtained on Day 1. Error bars: mean ± SEM; n = 3 biological replicates.
(D) Nascent proteins in MEF cells with or without eIF4A1 knockdown were labeled with 10 μg/ml of puromycin for 5, 10 or 15 mins. Cell lysates were resolved by SDS-PAGE and analyzed by immunoblotting using the indicated antibodies.
(E) Using Ribo-seq data sets, a scatter plot shows the correlation of individual CDS ribosome occupancy between MEF cells with or without eIF4A1 knockdown.
(F) Translation efficiency of Fluc mRNA reporters with ultra-short 5'UTR in MEF cells with or without eIF4A1 knockdown. Error bars: mean ± SEM; n = 3 biological replicates. t-test, * $p < 0.05$; *** $p < 0.001$.
(G) Translation efficiency of Fluc mRNA reporters containing 3 AUG codons near the 5' end in MEF cells with or without eIF4A1 knockdown. Error bars: mean ± SEM; n = 3 biological replicates. t-test, * $p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$.
(H) A box plot of 5'UTR read density over the CDS from PIC-seq data sets obtained from MEF cells with or without eIF4A1 knockdown.
(I) Polysome samples as in (B) were subjected to immunoblotting using antibodies indicated.

2.3.6 eIF4A Promotes PIC Scanning in an ATPase-Dependent Manner

We next seek the possibility of modulating the eIF4A activity using chemical compounds. The natural product Hippuristanol (Hippu) potently inhibits the ATPase activity of eIF4A[33], whereas Pateamine A (PatA) enhances its intrinsic enzymatic activities[34]. Pre-treatment with either Hippu or PatA readily disassembled polysomes (Figure 2-S6A). Not surprisingly, both Hippu and PatA are commonly used as translation inhibitors despite their opposing effects toward the ATPase activity of eIF4A. We monitored the start codon selection from ultra-short 5'UTR in MEFs after exposure to low concentrations of Hippu or PatA. Like eIF4A1 depletion (Figure 2-5A), Hippu treatment repressed initiation of mRNAs with ultra-short 5'UTR (Figure 2-6A, top panel). Remarkably, pre-treatment with PatA promoted recognition of start codons near

the 5' end. The opposing effects of Hippu and PatA were no longer evident in the translation of messengers bearing 20-nt long 5'UTR. We further examined mRNA reporters with 3 AUG codons near the 5' end and found that the selection of those AUG codons from the 'blind spot' was repressed by Hippu but not PatA (Figure 2-6A, bottom panel). The finding that eIF4A enhances recognition of start codons within the 'blind spot' argues against the hypothesis that eIF4A prevents PIC from back-sliding. Rather, eIF4A promotes free scanning of the PIC in an ATP-dependent manner, presumably via an open conformation.

The average length of 5'UTR in mammalian cells is ~200 nt[35]. It is thus unlikely that the bi-directional scanning of PIC is evolved to select start codons near the 5' end. Since PIC oscillation occurs during the entire course of scanning, the back-and-forth excursions is expected to increase the dwell time of PIC in 5'UTR and facilitate the recognition of upstream start sites (uTIS). We conducted Ribo-seq of MEFs exposed to eIF4A modulators for different times. To factor out variations of mRNA levels, we used the read density in the coding region (CDS) as internal control to compute 5'UTR ribosome occupancy. Remarkably, PatA treatment for 2h resulted in a global increase of 5'UTR read density relative to the CDS (Figure 2-6B, top panel). Notably, the overall CDS ribosome occupancy was reduced in the presence of PatA (Figure 2-6C). This effect became more dramatic under prolonged PatA treatment (8h). Typical examples are Rps5 (Figure 2-6D) and *Eef1a1* (Figure 2-S6B). By contrast, Hippu treatment showed minimal changes of ribosome occupancy in 5'UTR

relative to CDS (Figure 2-6B, bottom panel), despite its inhibitory effects on global translation (Figure 2-6C). The opposing effect of Hippu and PatA on 5'UTR read density was further confirmed when the 3'UTR read density was used as internal control (Figure 2-S6C). Since PatA enhances the ATPase activity of eIF4A, the marked accumulation of ribosome footprints in 5'UTR is likely a consequence of increased recognition of uTIS sites.

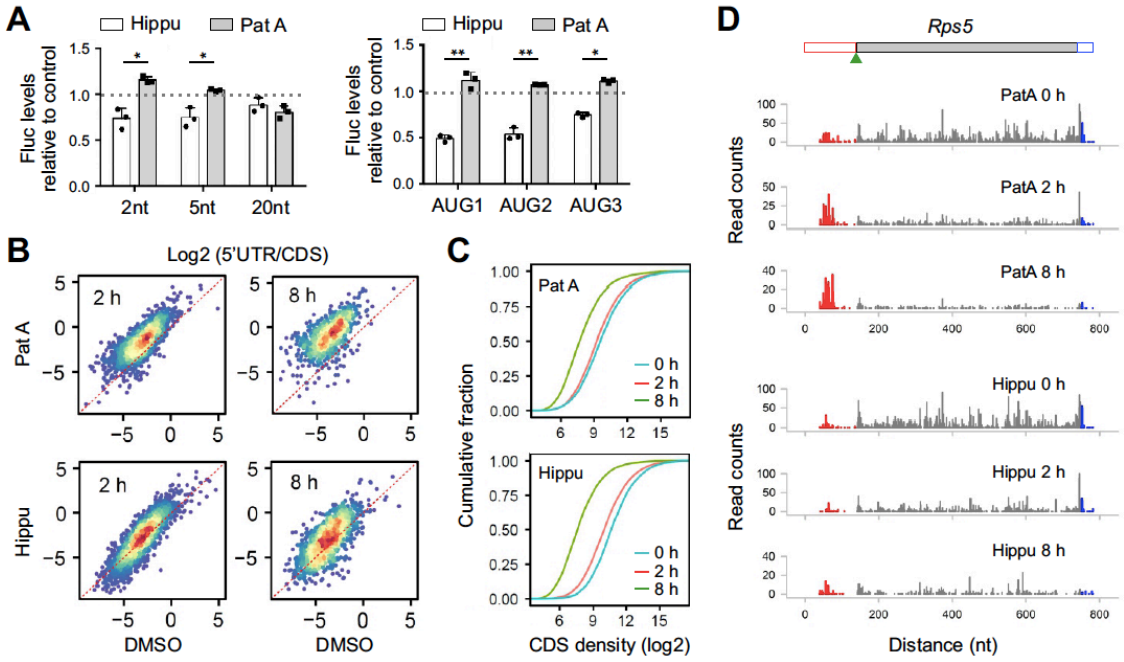

Figure 2-6 eIF4A Promotes PIC Scanning in an ATPase-Dependent Manner.
(A) Effects of Hippuristanol (Hippu) or Pateamine A (PatA) treatment on the translation efficiency of Fluc mRNA reporters with ultra-short 5'UTR (top) or those containing 3 AUG codons near the 5' end (bottom). Transfected MEF cells were treated with 200 nM Hippu or 50 μM PatA and Fluc activities were quantified by real-time luminometry. Error bars: mean ± SEM; n = 3 biological replicates. t-test, * $p < 0.05$; ** $p < 0.01$.
(B) Effects of Hippu or PatA treatment on 5'UTR ribosome occupancy revealed by Ribo-seq. Scatter plots show the correlation of read ratio (5'UTR / CDS) between vehicle control and treatment (200 nM PatA: top; 50 μM Hippu: bottom) for 2 hrs (left) or 8 hrs (right).
(C) Effects of Hippu or PatA treatment on CDS ribosome occupancy revealed by Ribo-seq. From the same data sets as (B), CDS ribosome occupancies on individual transcripts were normalized to the number of mitochondrial reads and plotted as cumulative fractions.

(D) A representative gene (Rps5) shows differential 5'UTR read density (red) relative to the CDS (dark gray) after treatment with eIF4A modulators.

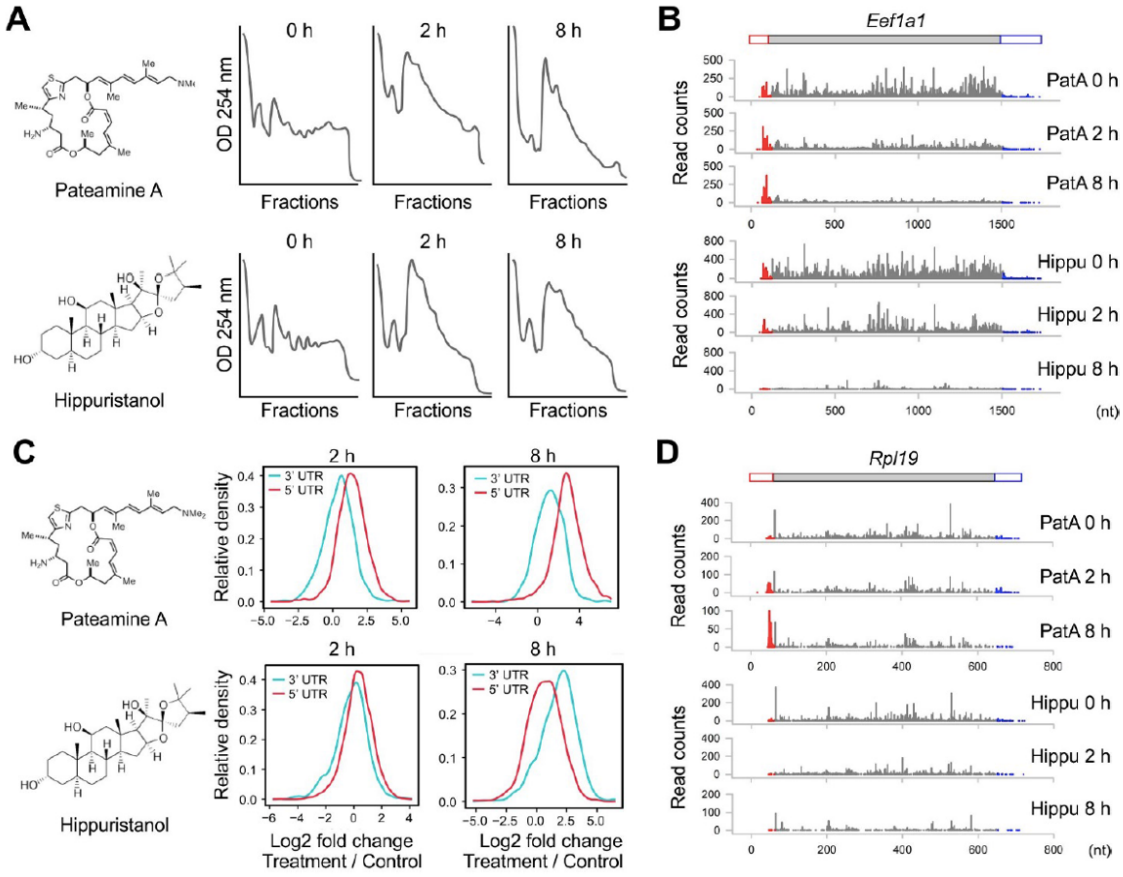

Figure 2-S6 Enhanced the ATPase activity of eIF4A1 promotes alternative translation in 5'UTR.

(A) Polysome profiles of MEF cells treated with eIF4A1 modulators. The upper panels show the polysome profiling of MEF cells treated with 50 μM PatA for different times. The bottom panels show polysome profiling of MEF cells treated with 200 nM Hippu for different times.

(B) A representative example of genes (*Eef1a1*) shows differential ribosome occupancy in 5'UTR after treatment with eIF4A1 modulators.

(C) Distribution of fold changes for 5'UTR ribosome occupancy after treatment with eIF4A1 modulators (red line). 3'UTR read density was used as internal controls (blue line).

(D) A representative example of genes (*Rpl19*) shows newly emerged ribosome occupancy in 5'UTR after treatment with PatA, but not Hippu.

2.3.7 eIF4A Controls Alternative Start Codon Selection in an ATPase-Dependent Manner

We noticed that, for some transcripts with few 5'UTR reads, PatA treatment readily triggered the appearance of new footprints (Figure 2-S6D), representing induced alternative translation. A boost of eIF4A helicase activities thus offers a simple but effective means to uncover hidden uTIS sites across the transcriptome. We previously developed quantitative profiling of initiating ribosomes (QTI-seq) that enables identification of alternative TIS sites at single base resolution[36]. Due to low levels of alternative initiation under physiological conditions, the global TIS map is far from complete. To uncover cryptic TIS sites from 5'UTR, we treated HEK293 cells with 50 nM of PatA for 2 h followed by QTI-seq (Figure 2-S7A). Indeed, PatA treatment not only expanded the scope of identifiable uTIS sites (Figure 2-7A), but also increased the relative uTIS density above the background (Figure 2-S7B). By contrast, the aTIS usage was attenuated by PatA. One typical example is Sod1 that showed only 1 uTIS under the control condition but 3 uTIS peaks were emerged upon PatA treatment (Figure 2-7B). Importantly, eIF4A inhibition by a low dose of Hippu (200 nM) reversed the relative usage between uTIS and aTIS. To gain a better understanding of the cryptic uTIS sites, we compared the sequence features between common uTIS codons and PatA-induced unique uTIS sits. Consistent with previous reports[20], physiological uTIS sites are dominated by the CUG codon with an optimal sequence context (i.e., a purine at position – 3 and a

guanine at position +4) (Figure 2-S7C). Intriguingly, PatA- induced unique uTIS sites showed a less optimal sequence context, especially at position – 3. The underrepresented adenine is also seen at the first nucleotide of start codons. Indeed, codon composition analysis revealed less AUG usage for PatA-induced unique uTIS codons (Figure 2-7C). By contrast, Hippu treatment suppressed the usage of CUG as uTIS codons (Figure 2-7D). Therefore, modulating eIF4A activities controls the stringency of start codon selection with increased helicase activity promoting recognition of suboptimal start codons.

We also treated cells with a high dose of Hippu (1000 nM) that nearly eliminated all the TIS peaks (Figure 2-S8A), presumably as a result of inhibited ribosome loading. This finding echoes the global effect of eIF4A with little mRNA specificity[11]. Intriguingly, a handful of peaks were resistant to the high dose of Hippu, likely representing non-ribosome footprints as most of these peaks were mapped to 3'UTR (Figure 2-S8B). By identifying false-positive TIS peaks from 5'UTR (Figure 2-S8C), high does Hippu treatment offers an orthogonal way to validate uTIS sites identified by QTI-seq.

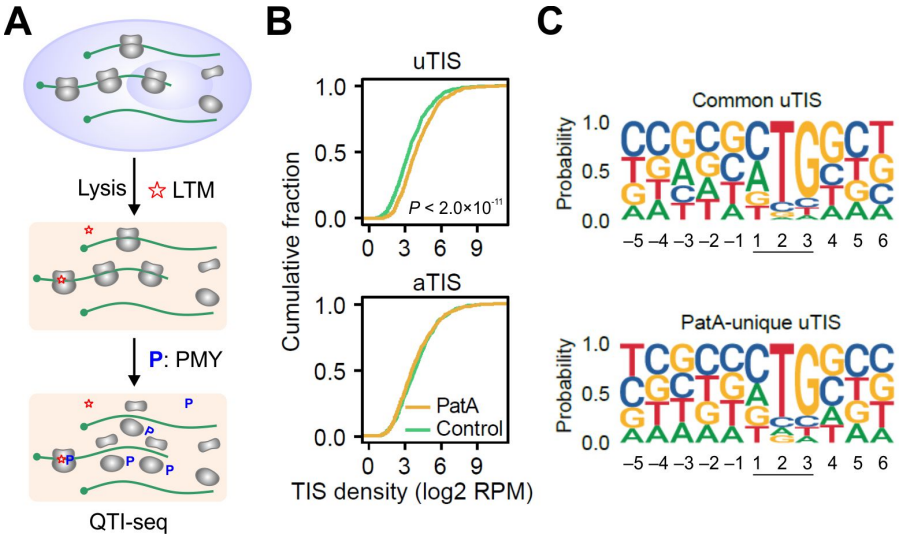

Figure 2-S7 Modulating the ATPase Activity of eIF4A1 Controls the Stringency of uTIS Selection.
(A) A schematic diagram of QTI-seq. Cell lysates were treated with lactinomycin (LTM) to freeze the initiating ribosomes, followed by depletion of elongating ribosomes by puromycin.
(B) Ribosome densities at uTIS (top) and aTIS (bottom) were computed from QTI-seq in HEK293 cells before and after PatA treatment and plotted as cumulative fractions. Wilcoxon test, $P < 2.0 \times 10^{-11}$ (Top); $P = 0.09$ (bottom).
(C) Probabilities for nucleotide sequence surrounding the 'Common uTIS' (upper) or 'PatA-unique uTIS' (bottom) identified from QTI-seq reads data sets.

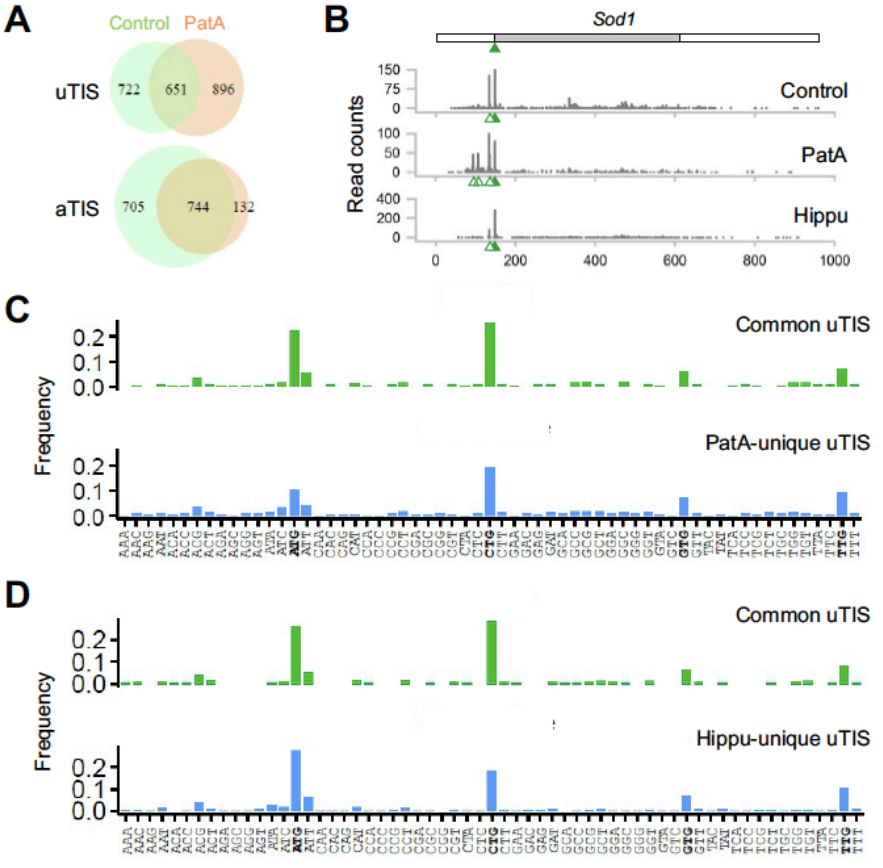

Figure 2-7 eIF4A Controls Alternative Translation Initiation in an ATPase-Dependent Manner.
(A) A Venn diagram shows the number of uTIS (upper) or aTIS (bottom) identified by QTI-seq in HEK293 cells before and after treatment with 50 μM PatA for 2 hrs.
(B) A representative gene (*Sod1*) shows induced uTIS sites by PatA treatment but repressed by Hippu. HEK293 cells were treated with 200 nM Hippu or 50 μM PatA for 2 hrs followed by QTI-seq. Solid green triangle, aTIS; unfilled green triangle, uTIS.
(C) Codon composition analysis of 'Common uTIS' (upper) or 'PatA-unique uTIS' (bottom) identified from QTI-seq reads data sets.
(D) Codon composition analysis of 'Common uTIS' (upper) or 'Hippu-unique uTIS' (bottom) identified from QTI-seq reads data sets.

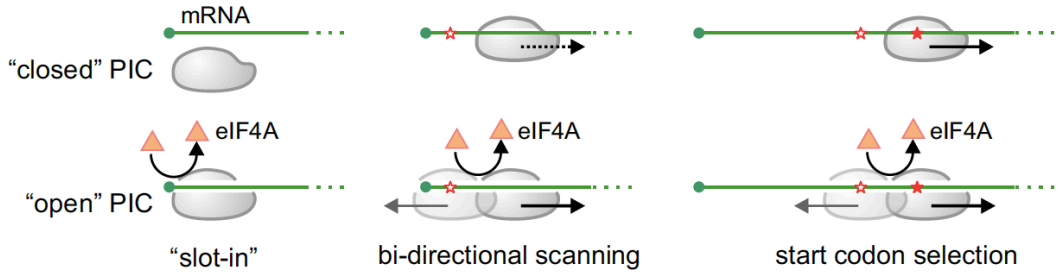

Figure 2-8 Proposed Models for eIF4A-Mediated PIC Conformational Change that Unifies Ribosome Recruitment, Scanning, and Start Codon Selection
In the absence of eIF4A, the closed PIC conformation leads to reduced mRNA loading, delayed PIC scanning, and increased stringency of start codon selection. In the presence of eIF4A (orange triangle), multiple cycles of ATP hydrolysis lead to open conformation of PIC that facilitates 'slot-in' mRNA loading, bi-directional PIC scanning, and selection of suboptimal start codons (empty red star).

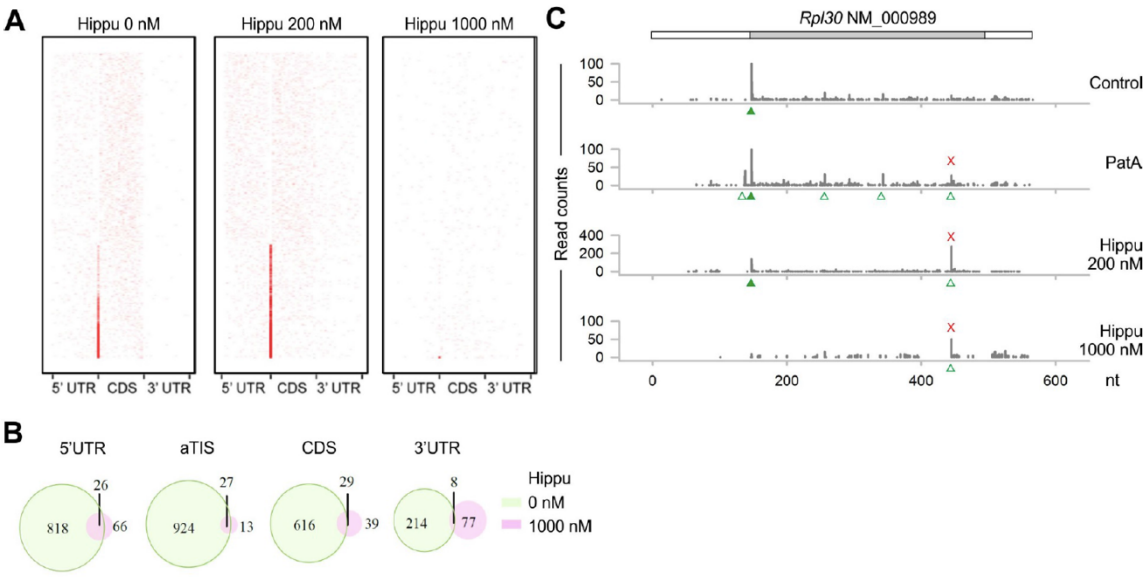

Figure 2-S8 High dose of Hippu eliminates false-positive TIS sites identified from QTI-seq.
(A) Heat-map of TIS density across individual transcripts from HEK293 cells treated with different doses of Hippu.
(B) Venn diagrams show the total number of TIS sites identified by QTI-seq in HEK293 cells before and after treatment with a high dose of Hippu (1000 nM). TIS sites in different regions (5'UTR, aTIS, CDS, and 3'UTR) were counted separately.
(C) A representative example of genes (Rpl30) shows a false-positive TIS peak identified by a high dose of Hippu (1000 nM). Solid green triangle, aTIS; unfilled green triangle, alternative TIS sites.

2.4 Discussion

Eukaryotic translation begins with accommodation of specific mRNAs by the 43S PIC. The proposed mRNA 'threading' mechanism permits base-by-base inspection by the PIC, enabling selection of start codons near the 5' end cap[24]. However, for many transcripts with short 5'UTRs, translation is facilitated by a motif celled translation initiator of short 5'UTR (TISU)[37], which is unlikely functional in the form of linear thread. Another feature of the 'threading' model is the mandatory eIF4E dissociation from the 5' end cap. A recent study using selective 40S footprinting in human cells reported that eIF4E could remain attached on the PIC throughput the[8]. We demonstrated that tethering eIF4E to the 5' end cap does not abolish the translation of mRNAs with long 5'UTR. The 'slot-in' model conforms to mRNA translation mediated by IRES, m^6A residues[38,39], or the poly(A) tract in 5'UTR[40]. The translation potential of circular RNAs (circRNAs) further indicate a more lateral loading of mRNA into the PIC[41]. For cap-dependent mRNA translation, however, the slotting PIC is expected to leave a blind spot on mRNA for start codon selection. In vitro studies have documented that a single nucleotide is a sufficient 5'UTR for translation[42]. Using in vivo mRNA reporters, we showed that a 5'UTR of 2 nt enables mRNA translation albeit with low efficiency. It appears that neither the 'threading' nor the 'slot-in' model could fully explain those experimental observations.

It is possible that 'threading' and 'slotting' modes co-exist inside cells with the latter utilized solely for special types of mRNAs. However, the putative

position of eIF4E in proximity to the E-site is inconsistent with the presumption of the 'threading' model in which the cap-binding protein must be placed near the A-site. Additionally, cross-linking assays revealed that the 5' end of mRNA interacted with eIF3a, eIF3b, and eIF3d instead of small ribosomal proteins[43]. It is also worth noting that the PIC near the 5' end shows a broad range of footprint sizes, which is more consistent with the laterally loaded PIC spanning a sizable mRNA region from the 5' end cap. The major conflict between the 'threading' and the 'slot-in' model lies in the presence of blind spots for start codon selection. If the PIC undergoes a nonlinear motion after accommodation of mRNA, start codons near the 5' end can be reached without 'threading'. Supporting this notion, a blind spot emerged when the eIF4E was crosslinked with the 5' end cap, which prevents PIC backsliding. We propose that PIC undergoes bi-directional migration after slotting, which resolves the fundamental dilemma between 'slot-in' mRNA accommodation and recognition of 5' end start codons.

The scanning process commences once the PIC has been loaded onto the mRNA. Another major open question in translation initiation concerns the scanning directionality. The linear correlation between the scanning time and the 5'UTR length was demonstrated as the evidence supporting the 5'→3' unidirectional movement of PIC[44]. For mRNAs with multiple start codons, the first AUG codon is favored by the scanning PIC also suggests the linear scanning mode[45]. Nevertheless, most of these conclusions considered a net directionality in the process but not the dynamics of the scanning. Despite the

widely accepted scanning model, i.e., 5'→3' linear and base-by-base migration, the nonlinear scanning mechanism has been suggested before. For instance, the IRES element from *Halastavi arva virus* (HalV) recruits PIC immediately downstream of the start codon, which is then picked by retrograde scanning[46]. The bidirectional scanning has also been documented for both post-terminating 80S ribosomes and recycled 40S during translation reinitiation[47,48]. For cap-dependent translation, the back-and-forth oscillations of scanning ribosomes was originally suggested from experiments using closely spaced AUG codons found in *Turnip yellow mosaic virus* (TYMV)[23]. We confirmed the competitive nature of neighboring start codons by demonstrating that downstream start codons negatively regulate upstream ones. Perhaps the most compelling evidence comes from mRNA reporters sandwiched by an IRES. Only when the PIC undergoes bi-directional scanning, could both upstream and downstream start codons be selected.

A more profound question is how the net 5'→3' directionality is achieved on 5'UTR if PIC undergoes back-and-forth migrations. The DEAD-box RNA helicases have long been implicated in the scanning process. The yeast Ded1p catalyzes the forward scanning of PIC by residing at the mRNA entry channel and unwinding mRNA secondary structures[49]. The mammalian specific-DHX29 possibly uses the similar mechanism to assist forward scanning[50,51]. Not surprisingly, translation of mRNAs with complex structures is more susceptible to Ded1p or DHX29 deficiency. As the prototypical member of DEAD-box RNA

helicase, eIF4A is the most abundant translation initiation factor[13]. Unlike DHX29 that is stably positioned at the leading edge of PIC, eIF4A is believed to recycle rapidly during the scanning process[10]. Notably, in vitro studies revealed that eIF4A can unwind RNA duplex in either direction at the same rate[52]. A previous study reported that eIF4A together with Rocaglamide A delays PIC scanning by clamping onto polypurine sequences in an ATP-independent manner[53]. Importantly, eIF4A has been shown to modulate the PIC conformations via ATP hydrolysis[12], suggesting that mRNA is not the only substrate of eIF4A. Indeed, lack of eIF4A leads to longer PIC footprints in 5'UTR, a sign of closed PIC conformation. Consistently, in the absence of eIF4A, the closed PICs tend to stall in traversing 5'UTR. As a result, global translation is severely inhibited. Given the crucial role of eIF4A in driving PIC free scanning, it is conceivable that the nonlinear scanning consists of forward thrusts alternating with limited relaxation in the reverse direction (Figure 2-8).

The scanning process is also crucial for inspection of potential start codons on mRNAs. A growing body of evidence suggests that the stringency of start codon selection is tightly coupled with the scanning kinetics. In general, slow scanning enables recognition of suboptimal start codons, whereas fast scanning increases the fidelity of start site selection. However, the PIC conformation is equally important in controlling the stringency of start codon selection. Intriguingly, eIF4A-dependent changes in the conformation of eIF3j has been implied in the fidelity of start codon recognition[54]. By modulating the

ATPase activity of eIF4A using small molecule compounds, we demonstrate that recognition of cryptic uTIS sites can be induced or repressed. We anticipate that, by adjusting both ATPase and helicase activities of eIF4A, it is possible to fine-tune the stringency of start codon selection and control the translational output. As RNA helicases have been implicated in development and cancer[55], dissecting their distinct roles in the scanning process will accelerate our understanding of translational reprogramming in cellular homeostasis and pathophysiology.

2.5 Material and Methods

Cell lines and reagents

MEF, HEK293 and HEK293-K^b cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum (FBS). Antibodies used in the immunoblotting are listed below: anti-eIF4A1 (Abcam ab31217), anti-puromycin (Developmental Studies Hybridoma Bank PMY-2A4), anti-β-Actin (sigma A5441), anti-eIF3a (Cell signaling 3411S), anti-RPS26 (Protein Tech 14909-1-AP), anti-RPL4 (Protein Tech 11302-1-AP), Alexa Fluor 546 donkey anti-mouse secondary antibody (Invitrogen A10036) and Alexa Fluor 546 donkey anti-rabbit secondary antibody (Invitrogen A10040).

Plasmid constructions

The full-length firefly luciferase (Fluc) gene was cloned into pcDNA3.1 vector (Invitrogen) to generate the Fluc/pcDNA3.1 vector. To create the A88T mutant,

site-directed mutagenesis was performed using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). Primers are 5′-GCGTATCTCTTCAAAGCCTTATGCAGTTGCTCTCC -3′ and 5′-GGAGAGCAACTGCATAAGGCTTTGAAGAGATACGC -3′. Mutation was confirmed by Sanger DNA sequencing. The uORF plasmid reporters were constructed by inserting DNA sequences containing 5'UTR (2-nt, 4-nt, 10-nt or 20-nt) and the sequence encoding SIINFEKL into Sac I/BamH I sites of EGFP/pcDNA3.1 vector to generate SIINFEKL-EGFP/pcDNA3.1 vector. To construct the SIINFEKL-6nt-IRES-Fluc/pcDNA3.1 vector, the DNA sequence encoding SIINFEKL was cloned into Nhe I/ Pml I sites of Rluc-IRES-Fluc/ pcDNA3.1 vector. DNA sequence encoding EGFP was amplified from EGFP/pcDNA3.1 vector, and subcloned into BamHI/XohI sites of SIINFEKL-6nt-IRES-Fluc/pcDNA3.1 vector to generate SIINFEKL-6nt-IRES-EGFP/pcDNA3.1 vector. The SIINFEKL-18nt-IRES-EGFP/pcDNA3.1 vector was constructed by inserting 12nt + IRES sequence into Pml I/BamH I sites of SIINFEKL-6nt-IRES-EGFP/pcDNA3.1 vector. The Polio(inv) sequence was synthesized (Top Gene Technologies) and cloned into Pml I/BamH I sites of SIINFEKL-6nt-IRES-EGFP/pcDNA3.1 and SIINFEKL-18nt-IRES-EGFP/pcDNA3.1 vectors to generate SIINFEKL-6nt-IRES(inv)-EGFP/pcDNA3.1 and SIINFEKL-18nt-IRES(inv)-EGFP/pcDNA3.1 vectors, respectively. DNA sequences of all primers used in this study are listed in Appendix III.

In Vitro Transcription

To generate mRNAs suitable for transfection, 3 μg PCR products described above were utilized for in vitro transcription. Transcripts with the normal m^7G cap were generated using the mMESSAGE mMACHINE T7 Transcription Kit (Invitrogen 1344) and transcripts with the non-functional cap analog ApppG (NEB) were synthesized using MEGAscript T7 Transcription Kit (Invitrogen 1333), followed by polyadenylation using Poly(A) Tailing Kit (Invitrogen AM1350M). mRNA products were purified according to the manufacturer's instruction.

Real-Time Luciferase Assay

Cells grown in 35 mm dishes were transfected with in vitro synthesized luciferase reporter mRNAs. Luciferase substrate D-luciferin (1 mM, Regis Tech) was added into the culture medium immediately after transfection. Luciferase activity was monitored and recorded using Kronos Dio Luminometer (Atto).

Real-time quantitative PCR

Total RNA was isolated by TRIzol reagent (Invitrogen) and used for reverse transcription assay via High Capacity cDNA Reverse Transcription Kit (Invitrogen). Real-time PCR analysis was conducted using Power SYBR Green PCR Master Mix (Applied Biosystems) and carried on a LightCycler 480 Real-Time PCR System (Roche Applied Science). Primers for amplifying each target are listed in Appendix III.

Flow cytometry

Transfected HEK293-K^b cells were washed with PBS and harvested by trypsin. Cells were then re-suspended in blocking buffer (1% bovine serum albumin (BSA) in PBS). Cells are aliquoted into a 96-well plate followed by 2000 rpm spinning for 2 min. After removal of blocking buffer, cells were washed one more time followed by staining with 25D1 Alexa 647 antibody (1:1000 in 75 uL solution per well). After incubation in the dark with gentle rocking at 4°C for 30 min, cells were washed three times with 200 uL of the blocking buffer to remove unbound antibodies. Resuspend cells in 300 µl of blocking buffer followed by single cell filtering (Falcon). Cells were analyzed on a BD FACSAria Fusion flow cytometer (BD Biosciences). Cytometry data analysis is conducted using FlowJo.

Toeprinting

The ribosome binding reaction mixture was prepared on ice in a total volume of 10 µl containing 50% Nuclease-treated rabbit reticulocyte lysate (Promega), 20 pmol of primer (5'- 6-FAM-AATTGTTCCAGGAACCAG- 3'), 20 µM amino acid mixture, 0.4 U/µl RNaseOUT (Invitrogen) and 50 mM Tris-HCl pH 7.5. Reactions were treated with 0.5 mg/ml CHX followed by incubation at 37 °C for 5 min. After addition of 0.3 mg of reporter mRNAs, the reaction mixtures were incubated at 30 °C for 20 min to allow the translation machinery to assemble at the start codon. The reverse transcriptase reaction was conducted in a total volume of 20 µl containing the entire ribosome binding reaction, 1x Superscript III reverse transcriptase buffer, 5 mM DTT, 40 mM KCl, 3 mM $MgCl_2$, 50 mM

Tris-HCl pH7.5, 0.5 mg/ml cycloheximide, 0.8 mM of dNTP, 1.5 U/µl RNaseOUT, 5 U/µl Superscript III reverse transcriptase. After incubation at 25 °C for 10 min, the reaction was terminated by nucleic acids extraction by phenol:chloroform and ethanol precipitation. The primer extension products were resuspended in 10 µl of Hi-Di formamide. 2 µl aliquot was run with 0.2 µl GeneScan 500 LIZ dye Size Standard (Fisher) on an ABI 3730xl instrument. Data is analyzed by Peak Scanner 2 software.

Recombinant protein purification

The *Eif4e* coding sequence was cloned into pGEX-6P-1 vector using the following primers:

eIF4E-F, 5′- GCGAATTCATGGCGACTGTCGAACCGGA-3′;

eIF4E-R, 5′-CCGCTCGAGTTAAACAACAAACCTATTTTTAG-3′.

The construct was transformed into the *E. coli* bacteria BL21. GST fusion protein was induced by 0.5 mM IPTG at 20°C for 3-4 hr. The cells were harvested and lysed in the lysis buffer (PBS supplemented with 0.5 mM PMSF, 1 mM DTT, protease inhibitor cocktail (Roche), 0.1% (v/v) Triton X-100) with 10 min sonication. Cell debris was removed by centrifuge at 12,000 rpm for 30 min. The supernatant was mixed with 2 ml equilibrated Pierce glutathione agarose followed by 4 °C incubation for 2-3 hr. The resin was washed five times and the fusion protein was eluted in GST elution buffer (5 mM glutathione, 50 mM Tris-HCl pH 8.0).

UV-crosslinking and chemical crosslinking

mRNAs synthesized with the T7 RNA polymerase system (Invitrogen AM1333) were capped with GTP or 6-Thio-GTP using the capping enzyme system (CellScript C-SCCS1710), followed by polyadenylation with Poly(A) Tailing Kit (Invitrogen AM1350M). For UV-crosslinking, Purified GST-eIF4E protein was incubated with m^7S^6G capped mRNA in binding buffer (10 mM HEPES-KOH pH 7.6, 100 mM KCl, 0.5 mM $MgCl_2$, and protease inhibitors) at 4 °C under 360 nm UV exposure for 15 min. For chemical crosslinking, m^7G capped mRNA are incubated for 2-3 hr on ice in 100 mM sodium acetate pH 5.3, 10 mM EDTA, 0.2 mM sodium periodate. The crosslinked mixture was neutralized by adding glycerol to 2% (v/v) final concentration. After 10 min incubation at room temperature, the crosslinked mRNAs were phenol extracted twice and ethanal precipitated. The RNA pellet was dissolved in Nuclease-free water. Purified GST-eIF4E protein was incubated with oxidized mRNA in binding buffer (10 mM HEPES-KOH pH7.6, 100 mM KCl, 0.5 mM $MgCl_2$, and protease inhibitors) at 4 °C for 15 min. $NaBH_4$ (20 mM) was added and incubation was pursued for 2-3 hr on ice. UV or Chemical Crosslinked Mixtures were then incubated with 70% Nuclease-treated rabbit reticulocyte lysate (Promega), 20 mM amino acid mixture for 1.5 hr at 30 °C. Firefly luciferase activities were measured using Luciferase Assay System (Promega E1500). Briefly, 20 µl of RRL mixture was transferred to each well of a 96-well plate. The 96-well plate was placed into the Plate-reading Luminometer with injector (BioTek). The injector added 100 µl of

Luciferase Assay Reagent per well followed by the measurement of light intensity before the plate was advanced to the next well for a repeat of the inject-then-read process.

Lentiviral sgRNAs

LentiCRISPRv2 plasmids targeting eIF4AI were constructed using methods described previously ([56],[57]). Briefly, complementary oligonucleotides containing the specific sgRNA sequence and overhangs complementary to overhangs generated by BsmBI digestion of LentiCRISPRv2 were annealed to the BsmBI digested LentiCRISPRv2 plasmid to generate the functional transfer vector. Undigested LentiCRISPRv2 plasmid lacking a sgRNA sequence was used for pseudovirus production as a control (Scramble). Lentiviral particles were packaged using Lenti-X 293T cells (Clontech). Virus-containing supernatants were collected at 48 hr after transfection and filtered to eliminate cells. MEF cells were infected by the lentivirus for 48 hr followed by selection with 2 μg/ml puromycin. Selected cells were plated in 96-well plates by serial dilution at 1 cell per well. Single cells were then expanded and analyzed by PCR amplification of genomic DNA flanking the CRISPR-targeted region. sgRNA targeting sequence used in this study is listed in Appendix III.

Puromycin Labeling

Cells at 80-90% confluence were treated with DMEM + 10% FBS 2 hr before media was changed to DMEM + 10% FBS supplemented with 10 μg/ml

puromycin for an additional 5 min, 10 min or 15 min. Cells were washed twice with ice-cold DPBS and lysed, followed by immunoblotting.

Immunoblotting

Cells were lysed on ice in TBS buffer (50 mM Tris pH7.5, 150 mM NaCl, 1 mM EDTA) containing 1% Triton X-100, 2 U/ml DNase and protease inhibitor cocktail tablet. The lysates were incubated on ice for 30 min, followed by heating for 10 min in SDS-PAGE sample buffer (50 mM Tris pH6.8, 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, 10% glycerol). Proteins were separated on SDS-PAGE and transferred to PVDF membranes (Fisher). Membranes were blocked in TBS containing 5% non-fat milk and 0.1 % Tween-20 for 1 hr, followed by incubation with primary antibodies overnight at 4 °C. After incubation with horseradish peroxidase-coupled secondary antibodies at room temperature for 1 hr, immunoblots were visualized using enhanced chemiluminescence (ECLPlus, GE Healthcare).

eIF3-seq and PIC-seq

A total of five 10 cm dishes with 90% confluence of MEF cells were used for either eIF3-seq or PIC-seq. After cell washing using ice-cold DPBS, cells were fixed in 10 ml/dish ice-cold formaldehyde solution (0.5% formaldehyde in DPBS) followed by 10 min incubation at 4°C on a rocker. The cells were then washed with ice-cold DPBS and quenched in 10ml/dish ice-cold buffer (50 mM Glycine, 50 mM Tris-HCl pH 7.5 in Nulease-free water) for 10 min at 4°C on a rocker.

After removal of the quench buffer, cells were washed with polysome buffer (10mM HEPES-KOH pH 7.4, 100 mM KCl and 5 mM $MgCl_2$), and collected in 400 µl lysis buffer (1% Triton-X-100 in polysome buffer) on ice. Whole cell lysates were cleared at 15,000 rpm for 10 min at 4 °C. The supernatant was transferred into a 1.5 ml tube followed by RNA digestion with RNase I (Ambion, 750 U per 100 A260 units). The mixture was incubated for 45 min at 4 °C. During the incubation, the sucrose solutions were prepared in polysome buffer. 10%-40% (w/v) sucrose density gradients were freshly prepared in a SW41 ultracentrifuge tube (Backman) using a Gradient Master (BioComp Instruments). Digested supernatant was loaded onto sucrose gradients followed by centrifugation for 3 hr 30 min at 32,000 rpm 4 °C in a SW41 rotor. Separated samples were fractionated at 1.5 ml/min through an automated fractionation system (Isco) that continually monitors OD254 values. After separation, for eIF3-seq, 40S and 80S fractions were pool together and mixed thoroughly to get 600 µl total sample. 5 µg/mg lysate eIF3a antibody (Cell signaling 3411S) and 0.5 U/ µl SUPERase•In (Invitrogen AM2696) were added to the IP samples, followed by an incubation under gentle rotation at 4 °C for 3 hr. Washed (3 times) Protein A/G beads were added into each IP sample and rock IP samples at 4 °C overnight. Beads were collected by spinning at 1000 rpm for 3 min at 4 °C and the supernatant was removed. Beads were washed 3 times with polysome buffer. After last washing, all supernatant was carefully removed and 400 µl of polysome buffer were added to resuspend the beads. For PIC-seq, 40S and

80S fractions were pool separately to get 400 µl each. For eIF3-seq, RNA was extracted from resuspended beads in polysome buffer. For PIC-seq, RNA was extracted from sucrose of 40S fractions and 80S fractions separately using methods previously described[58]. Briefly, samples were adjusted to 10 mM Tris-HCl pH 7.4, 10 mM glycine, 1% (w/v) SDS and 10 mM EDTA pH 8.0 and incubated at 65 °C for 5 min. Add one volume of acidic phenol/chloroform solution and vortex at maximum speed for 2 min. Place the tubes with mixtures into thermomixer and continue shaking at 1400 rpm for 20 min at 65 °C to reverse the cross-links. Centrifuge the samples at 14000 rpm for 5 min at room temperature. Carefully transfer the aqueous phases to new 1.5 ml tubes. Perform ethanol precipitation of RNA. Purified RNA was used for cDNA library construction and high-throughput sequencing described below.

QTI-seq and Ribo-seq

QTI-seq was conducted using a method described previously ([59]). In brief, cells were collected in 400 ml of ice-cold polysome buffer containing 5 mM LTM. Samples were transferred to a 1.5 ml tube containing Lysing Matrix-D and cells were lysed by vortexing 20 s for six times with a 40 s interval on ice. After removing the debris by centrifugation for 10 min at 13,000 g 4 °C, supernatant was transferred to a new 1.5ml tube and supplemented with 10 mM creatine phosphate, 0.1 mM spermidine, 40 mg/ml creatine phosphokinase, 0.8 mM ATP, 20 mM ATA and 25 mM of puromycin. The mixtures were incubated at 35 °C for 15 min and ready for the sucrose gradient sedimentation. For Ribo-seq, five

10 cm dishes of cells were harvested in 450 μl lysis buffer (1% Triton X-100 in polysome buffer) containing CHX (100 μg/ml), then centrifuged at 12,000 g 4 °C for 10 min. The supernatant was collected and subjected to sucrose gradient sedimentation. Sucrose solutions were prepared in polysome buffer. 15%- 45% (w/v) Sucrose density gradients were freshly prepared in a SW41 ultracentrifuge tube (Backman) using a Gradient Master (BioComp Instruments). Supernatant was loaded onto sucrose gradients followed by centrifugation for 2 hr 30 min at 32,000 rpm 4 °C in a SW41 rotor. Separated samples were fractionated at 1.5 ml/min through an automated fractionation system (Isco) that continually monitors OD254 values. For both QTI-seq and Ribo-seq, ribosome fractions separated by sucrose gradient sedimentation were pooled and digested with E. coli RNase I (Ambion, 750 U per 100 A260 units) by incubation at 4 °C for 1 hr. SUPERase•In (50 U per 100 U RNase I) was then added into the reaction mixture to stop the digestion. Total RNA was extracted using TRIzol LS reagent. Purified RNA was used for cDNA library construction and high-throughput sequencing described below.

cDNA library construction

Fragmented RNAs were separated on a 15% polyacrylamide TBE-urea gel (Invitrogen) and visualized using SYBR Gold (Invitrogen). Selected regions of the gel corresponding to 20-70 nt (for PIC-seq and eIF3-seq) or 25-35 nt (for Ribo-seq and QTI-seq) were excised. The gel slices were disrupted by using centrifugation through the holes at the bottom of the tube. RNA fragments were

dissolved by soaking overnight in 400 µl RNA gel elution buffer (300 mM NaOAc pH 5.5, 1 mM EDTA, 0.1 U/µl SUPERase•In). The gel debris was removed using a Spin-X column (Corning), followed by ethanol precipitation.

Purified RNA fragments were re-suspended in Nuclease-free water, then dephosphorylated for 1 hr at 37 °C in a 15 µl reaction (1× T4 polynucleotide kinase buffer, 10 U SUPERase•In and 20 U T4 polynucleotide kinase). Dephosphorylated RNA fragments were precipitated using ethanol and re-suspended in Nuclease-free water. 0.15 µg linker (rApp/NNNNCTGTAGGCACCATCAAT/3ddC) then was added to the RNA fragments, heated at 70 °C for 90 s and then cooled to room temperature, followed by ligation for 3 hr at at 22 °C in a 20 µl reaction (1× T4 Rnl2 reaction buffer, 10 U SUPERase•In, 15% PEG8000 and 20 U T4 RNA ligase 2 truncated). The reaction was heat inactivated at 80 °C for 10 min and the products were separated on a 10% polyacrylamide TBE-urea gel and selected regions in the gel corresponding to 45-95 nt (for PIC-seq and eIF3-seq) or 60-70 nt (for Ribo-seq and QTI-seq) were excised. RNA fragments were dissolved by soaking overnight in 400 µl RNA gel elution buffer. RNA fragments were Purified from RNA gel elution buffer as described earlier and re-suspended in Nuclease-free water.

For reverse transcription, the following oligos containing barcodes were used:

(Phos)CTANNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCT
CGGTGGTCGC(SpC18)CACTCA(SpC18)TTCAGACGTGTGCTCTTCCGATC
TATTGATGGTGCCTACAG

(Phos)AGCNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCT
CGGTGGTCGC(SpC18)CACTCA(SpC18)TTCAGACGTGTGCTCTTCCGATC
TATTGATGGTGCCTACAG

(Phos)ATTNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCT
CGGTGGTCGC(SpC18)CACTCA(SpC18)TTCAGACGTGTGCTCTTCCGATC
TATTGATGGTGCCTACAG

(Phos)CCGNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCT
CGGTGGTCGC(SpC18)CACTCA(SpC18)TTCAGACGTGTGCTCTTCCGATC
TATTGATGGTGCCTACAG

where Phos represents phosphorylation, NNN represents random sequence, SpC18 represents Hexa-ethyleneglycol spacer.

The linker ligated RNA sample was mixed with 0.5 mM dNTP and 2.5 mM synthesized primer and incubated at 75 ° C for 3 min, followed by incubation on ice for at least 1 min. The reaction master mix was then added with 20 mM Tris (pH 8.4), 50 mM KCl, 5 mM $MgCl_2$, 10 mM DTT, 40 U RNaseOUT and 200 U SuperScript III. Reverse transcription reaction was performed according to the manufacturer's instruction. Reverse transcription products were separated on a 10% polyacrylamide TBE-urea gel. Corresponding region was excised,

which was expected to be approximately 200 nt. The first-strand cDNA products were recovered in DNA gel elution buffer (300 mM NaCl, 1 mM EDTA), then purified and re-suspended in Nuclease-free water as described earlier.

cDNA products were circularized in 20 µl of reaction containing 1×CircLigase buffer, 2.5 mM $MnCl_2$, 1M Betaine and 100 U CircLigase II (Epicentre). Circularization was performed at 60 °C for 1 hr and the reaction was heat inactivated at 80 °C for 10 min.

Deep sequencing

Circular template was amplified by PCR using the Phusion high-fidelity (HF) enzyme (NEB) according to the manufacturer's instructions. The PCR forward primer: 5'-AATGATACGGCGACCACCGAGATCTACAC-3' and reverse primer: 5'-CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCA GACGTGTGCTCTTCCG -3' were used to create products suitable for sequencing. PCR reaction contains 1× HF buffer, 0.2 mM dNTP, 0.5 μM oligonucleotide primers, and 0.25 U Phusion polymerase. PCR was carried out with an initial 30 s denaturation at 98 °C, followed by 12 cycles of 10 s denaturation at 98 °C, 20 s annealing at 65 °C, and 20 s extension at 72 °C. PCR products were separated on a nondenaturing 8% polyacrylamide TBE gel. Expected products around 180 bp were excised and recovered in DNA gel elution buffer, then purified and re-suspended in Nuclease-free water as described earlier ([59]).

After quantification by Agilent BioAnalyzer DNA 1000 assay, equal amounts of barcoded samples were pooled into one sample. Approximately 5 pM mixed DNA samples were used for cluster generation followed by sequencing by using sequencing primer 5'-CGACAGGTTCAGAGTTCTACAGTCCGACGATC-3' (Illumina HiSeq).

Alignment of sequencing reads

The 3' adapter CTGTAGGCACCATCAAT was trimmed by Cutadapt ([60]). The trimmed reads with length shorter than 15 nucleotides or longer than 35 nucleotides were discarded. The remaining reads were mapped to the mouse transcriptome using STAR with default parameters ([61]). Only the reads that were aligned to a unique position with mismatches lower than 2 nucleotides were kept. To construct mouse transcriptome, the annotation file and genome sequences downloaded from ENSEMBL database (GRCm38) were used. For each gene, the mRNA with longest CDS was selected. In the case of equal CDS length, the longest transcript was used.

Ribosome density of transcript

For each transcript, RPKM (reads per kilobase of CDS, per Million mapped reads) was used to estimate the ribosome density of transcript. To exclude the global effect of drug treatment, the count of total reads mapped to cytoplasmic transcriptome was first normalized by the count of reads that were aligned to mitochondrial genome. mRNAs with RPKM value <1 were excluded.

Aggregation Plot of ribosome density

To make the aggregation plot around start and stop codons, for each transcript, Ribo-seq reads at individual positions were normalized by the average reads of transcript. Then, the normalized counts at the same position relative to start or stop codon were averaged. For 5' end aggregation plot, positions of 5' end of Ribo-seq reads were used. For 3' end aggregation plot, positions of 3' end of Ribo-seq reads were used.

TIS identification

We used the same method in Gao et al[59] to identify translation initiation sites (TIS) from QTI-seq, with slightly modifications. Briefly, for each transcript with total QTI-seq reads higher than 16, the distribution of P-site read count is modeled by Zero-Truncated Binomial Negative (ZTNB) distribution. For the sites in CDS and 3' UTR, the sites with FDR adjusted p-value lower than 0.05 were considered as potential downstream TIS (dTIS). For the sites in other regions, a putative TIS is predicted provided it meets the following criteria: 1) p-value is lower than 0.01, 2) ribosome read count (from Ribo-seq) in corresponding ORF is higher than 5, 3) in-frame reads is significantly higher than out-frame reads (Wilcoxon-test). The overlapped TIS within a small window (from -5 to +5 nt) were merged into a single TIS, by selecting the TIS with highest QTI peak density.

Motif analysis

The sequence motifs around TIS were plotted by ggseqlogo[62].

2.6 Acknowledgements

We thank Dr. Jonathan W. Yewdell for providing 25D1 reagents and HEK293-Kb cells and Dr. Jerry Pelletier for the gift of Hippuristanol. We are grateful to Cornell University Life Sciences Core Laboratory Center for sequencing and FACS. This work was supported by US National Institutes of Health (R01GM1222814 and R21CA227917) and HHMI Faculty Scholar (55108556) to S.-B.Q.

REFERENCE

1. Hinnebusch, A. G. The Scanning Mechanism of Eukaryotic Translation Initiation. *Annu. Rev. Biochem.* **83**, 779–812 (2014).
2. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 113–127 (2010).
3. Hershey, J. W. B., Sonenberg, N. & Mathews, M. B. Principles of translational control. *Cold Spring Harb. Perspect. Biol.* (2019). doi:10.1101/cshperspect.a032607
4. Kumar, P., Hellen, C. U. T. & Pestova, T. V. Toward the mechanism of eIF4F-mediated ribosomal attachment to mammalian capped mRNAs. *Genes Dev.* **30**, 1573–1588 (2016).
5. Hellen, C. U. T. & Sarnow, P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes and Development* (2001). doi:10.1101/gad.891101
6. Archer, S. K., Shirokikh, N. E., Beilharz, T. H. & Preiss, T. Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature* **535**, 570–4 (2016).
7. Slepenkov, S. V., Korneeva, N. L. & Rhoads, R. E. Kinetic mechanism for assembly of the m 7GpppG·eIF4E·eIF4G complex. *J. Biol. Chem.* (2008). doi:10.1074/jbc.M801786200
8. Bohlen, J., Fenzl, K., Kramer, G., Bukau, B. & Teleman, A. A. Selective 40S Footprinting Reveals Cap-Tethered Ribosome Scanning in Human Cells. *Mol. Cell* (2020). doi:10.1016/j.molcel.2020.06.005
9. Babendure, J. R., Babendure, J. L., Ding, J. H. & Tsien, R. Y. Control of mammalian translation by mRNA structure near caps. *RNA* (2006). doi:10.1261/rna.2309906
10. Pelletier, J. & Sonenberg, N. The Organizing Principles of Eukaryotic Ribosome Recruitment. *Annu. Rev. Biochem.* (2019). doi:10.1146/annurev-biochem-013118-111042
11. Yourik, P. *et al.* Yeast eIF4A enhances recruitment of mRNAs regardless of their structural complexity. *Elife* (2017). doi:10.7554/eLife.31476
12. Sokabe, M. & Fraser, C. S. A helicase-independent activity of eIF4A in promoting mRNA recruitment to the human ribosome. *Proc. Natl. Acad. Sci.* **114**, 6304–6309 (2017).
13. Andreou, A. Z. & Klostermeier, D. The DEAD-box helicase eIF4A: Paradigm or the odd one out? *RNA Biol.* **10**, 19–32 (2013).
14. Kozak, M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* (2002). doi:10.1016/S0378-1119(02)01056-9
15. Spirin, A. S. How does a scanning ribosomal particle move along the 5′-untranslated region of eukaryotic mRNA? Brownian ratchet model. *Biochemistry* (2009). doi:10.1021/bi901379a
16. Fütterer, J., Kiss-László, Z. & Hohn, T. Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA. *Cell* (1993). doi:10.1016/0092-8674(93)90257-Q
17. Paek, K. Y., Park, S. M., Hong, K. Y. & Jang, S. K. Cap-dependent translation without base-by-base scanning of an messenger ribonucleic acid. *Nucleic Acids Res.* (2012). doi:10.1093/nar/gks471
18. Yueh, A. & Schneider, R. J. Selective translation initiation by ribosome jumping in adenovirus- infected and heat-shocked cells. *Genes Dev.* (1996). doi:10.1101/gad.10.12.1557
19. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes.

Cell **147**, 789–802 (2011).
20. Lee, S. *et al.* Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci.* **109**, E2424–E2432 (2012).
21. Ivanov, I. P. *et al.* Polyamine Control of Translation Elongation Regulates Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing. *Mol. Cell* (2018). doi:10.1016/j.molcel.2018.03.015
22. Kozak, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* (1990). doi:10.1073/pnas.87.21.8301
23. Matsuda, D. & Dreher, T. W. Close spacing of AUG initiation codons confers dicistronic character on a eukaryotic mRNA. *RNA* **12**, 1338–49 (2006).
24. Kumar, P., Hellen, C. U. T. & Pestova, T. V. Toward the mechanism of eIF4F-mediated ribosomal attachment to mammalian capped mRNAs. *Genes Dev.* **30**, 1573–1588 (2016).
25. Starck, S. R. *et al.* Translation from the 5' untranslated region shapes the integrated stress response. *Science (80-.).* **351**, aad3867–aad3867 (2016).
26. Dersh, D., Yewdell, J. W. & Wei, J. A SIINFEKL-based system to measure MHC class I antigen presentation efficiency and kinetics. in *Methods in Molecular Biology* (2019). doi:10.1007/978-1-4939-9450-2_9
27. Martin, F. *et al.* Cap-Assisted Internal Initiation of Translation of Histone H4. *Mol. Cell* (2011). doi:10.1016/j.molcel.2010.12.019
28. Sonenberg, N. & Shatkin, A. J. Reovirus mRNA can be covalently crosslinked via the 5' cap to proteins in initiation complexes. *Proc. Natl. Acad. Sci. U. S. A.* (1977). doi:10.1073/pnas.74.10.4288
29. Wagner, S. *et al.* Selective Translation Complex Profiling Reveals Staged Initiation and Co-translational Assembly of Initiation Factor Complexes. *Mol. Cell* (2020). doi:10.1016/j.molcel.2020.06.004
30. Giess, A. *et al.* Profiling of Small Ribosomal Subunits Reveals Modes and Regulation of Translation Initiation. *Cell Rep.* (2020). doi:10.1016/j.celrep.2020.107534
31. Parsyan, A. *et al.* mRNA helicases: the tacticians of translational control. *Nat. Rev. Mol. Cell Biol.* **12**, 235–245 (2011).
32. Rozen, F. *et al.* Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. *Mol. Cell. Biol.* **10**, 1134–44 (1990).
33. Bordeleau, M.-E. *et al.* Functional characterization of IRESes by an inhibitor of the RNA helicase eIF4A. *Nat. Chem. Biol.* **2**, 213–220 (2006).
34. Low, W. K. *et al.* Inhibition of eukaryotic translation initiation by the marine natural product pateamine A. *Mol. Cell* **20**, 709–722 (2005).
35. Leppek, K., Das, R. & Barna, M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews Molecular Cell Biology* (2018). doi:10.1038/nrm.2017.103
36. Gao, X. *et al.* Quantitative profiling of initiating ribosomes in vivo. *Nat Methods* **12**, 147–153 (2015).
37. Elfakess, R. *et al.* Unique translation initiation of mRNAs-containing TISU element. *Nucleic Acids Res.* **39**, 7598–7609 (2011).
38. Meyer, K. D. *et al.* 5′ UTR m6A Promotes Cap-Independent Translation. *Cell* (2015). doi:10.1016/j.cell.2015.10.012
39. Zhou, J. *et al.* Dynamic m6 A mRNA methylation directs translational control of heat shock response. *Nature* (2015). doi:10.1038/nature15377
40. Gilbert, W. V., Zhou, K., Butler, T. K. & Doudna, J. A. Cap-independent translation is required for starvation-induced differentiation in yeast. *Science (80-.).* (2007). doi:10.1126/science.1144467

41. Yang, Y. *et al.* Extensive translation of circular RNAs driven by N 6 - methyladenosine. *Cell Res.* (2017). doi:10.1038/cr.2017.31
42. Hughes, M. J. G. & Andrews, D. W. A single nucleotide is a sufficient 5' untranslated region for translation in an eukaryotic in vitro system. *FEBS Lett.* **414**, 19–22 (1997).
43. Pisarev, A. V., Kolupaeva, V. G., Yusupov, M. M., Hellen, C. U. T. & Pestova, T. V. Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J.* (2008). doi:10.1038/emboj.2008.90
44. Vassilenko, K. S., Alekhina, O. M., Dmitriev, S. E., Shatsky, I. N. & Spirin, A. S. Unidirectional constant rate motion of the ribosomal scanning particle during eukaryotic translation initiation. *Nucleic Acids Res.* **39**, 5555–5567 (2011).
45. Kozak, M. Selection of initiation sites by eucaryotic ribosomes: Effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. *Nucleic Acids Res.* (1984). doi:10.1093/nar/12.9.3873
46. Abaeva, I. S., Pestova, T. V. & Hellen, C. U. T. Attachment of ribosomal complexes and retrograde scanning during initiation on the Halastavi árva virus IRES. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw016
47. Gunnery, S., Mäivali, Ü. & Mathews, M. B. Translation of an uncapped mRNA involves scanning. *J. Biol. Chem.* (1997). doi:10.1074/jbc.272.34.21642
48. Skabkin, M., Skabkina, O., Hellen, C. T. & Pestova, T. Reinitiation and other unconventional posttermination events during eukaryotic translation. *Mol. Cell* **51**, 249–264 (2013).
49. Guenther, U. P. *et al.* The helicase Ded1p controls use of near-cognate translation initiation codons in 5′ UTRs. *Nature* (2018). doi:10.1038/s41586-018-0258-0
50. des Georges, A. *et al.* Structure of mammalian eIF3 in the context of the 43S preinitiation complex. *Nature* **525**, 491–5 (2015).
51. Pisareva, V. P., Pisarev, A. V., Komar, A. A., Hellen, C. U. T. & Pestova, T. V. Translation Initiation on Mammalian mRNAs with Structured 5′UTRs Requires DExH-Box Protein DHX29. *Cell* **135**, 1237–1250 (2008).
52. Rogers, G. W., Richter, N. J., Lima, W. F. & Merrick, W. C. Modulation of the Helicase Activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J. Biol. Chem.* **276**, 30914–30922 (2001).
53. Iwasaki, S., Floor, S. N. & Ingolia, N. T. Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor. *Nature* **534**, 558–561 (2016).
54. ElAntak, L. *et al.* The indispensable n-terminal half of eIF3j/hcr1 cooperates with its structurally conserved binding partner eIF3b/prt1-rrm and with eiF1A in stringent aug selection. *J. Mol. Biol.* (2010). doi:10.1016/j.jmb.2009.12.047
55. Bhat, M. *et al.* Targeting the translation machinery in cancer. *Nat. Rev. Drug Discov.* **14**, 261–278 (2015).
56. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods* (2014). doi:10.1038/nmeth.3047
57. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (80-.).* (2014). doi:10.1126/science.1247005
58. Shirokikh, N. E., Archer, S. K., Beilharz, T. H., Powell, D. & Preiss, T. Translation complex profile sequencing to study the in vivo dynamics of mRNA–ribosome interactions during translation initiation, elongation and termination. *Nat. Protoc.* **12**, 697–731 (2017).
59. Gao, X., Wan, J. & Qian, S. B. Genome-wide profiling of alternative translation initiation sites. in *Methods in Molecular Biology* (2016). doi:10.1007/978-1-4939-3067-8_19
60. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* (2011). doi:10.14806/ej.17.1.200

61. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (2013). doi:10.1093/bioinformatics/bts635
62. Wagih, O. Ggseqlogo: A versatile R package for drawing sequence logos. *Bioinformatics* (2017). doi:10.1093/bioinformatics/btx469

CHAPTER 3

Roles of RNA Helicases on the Ribosome Scanning of mRNAs with Secondary Structures in Different Regions

The contents of this Chapter are in preparation for the manuscript to be published as Gu Y, Lin K, Mao Y, Dong L, and Qian SB. **Roles of RNA Helicases on the Ribosome Scanning of mRNAs with Secondary Structures in Different Regions.**

3.1 Abstract

Single-stranded RNA molecules tend to form high-order structures via intricate patterns of base pairing. Structural features of mRNA transcripts provide additional layers of gene regulation beyond that of their sequence. However, the specific effects of the secondary structures in different mRNA regions on the eukaryotic translation and the underlying mechanisms remain elusive. Here, we report that the translational effects of stem-loops (SLs) in the 5' untranslated region (5'UTR) depend on their positions. However, the positioning of these SLs did not affect their sensitivities to the RNA helicase eIF4A1 or DHX29. To our surprise, the AUG-proximal SLs in the coding region (CDS) showed high dependence on DHX29, similar to SLs in the 5'UTR. In contrast, all the mRNA reporters showed equal dependence on eIF4A1, regardless of the structural complexity. Also, DHX29 was found in both 40S and 80S monosome fractions after polysome fractionation. Collectively, these findings suggest that, besides its function initiation, DHX29 is likely involved in the early stage of the elongation before it dissociates from the 80S ribosome.

3.2 Introduction

Protein synthesis via mRNA translation is one of the most fundamental cellular processes in eukaryotes. Regulation of mRNA translation is principally directed at the initiation stage in order to achieve rapid, reversible and dynamic control of gene expression. In eukaryotes, the canonical translation initiation starts with the assembly of the ternary complex (TC) which consist of eukaryotic initiation factor 2 (eIF2), GTP and aminoacyl methionine tRNA Met-tRNA (Met-tRNAi^{Met}). TC, in cooperation with eukaryotic initiation factor 3 (eIF3), -1 (eIF1) and -1A (eIF1A), binds to the 40S ribosomal subunit, forming the 43S pre-initiation complex (PIC). Facilitated by the eukaryotic initiation factor 4F (eIF4F) complex which binds to the m^{7}G cap, the PIC attaches to the 5'end of mRNAs and scans the 5'UTR to the downstream appropriate start codon. Upon the start codon recognition, 48S PIC is formed and ready for the joining of the 60S ribosome[1].

Due to the intrinsic propensity to quickly and dynamically fold or form structures, mRNA exhibits a second layer of structural information which shapes the post-transcriptional control of gene expression. However, RNA structures, such as hairpins (SLs) and RNA duplexes, in the 5'UTR tend to play an inhibitory role in translation initiation by blocking the ribosome loading or impeding the ribosome scanning process[2]. Previous studies on the relationship between cap-to-SL distances and translation efficiency revealed that SLs proximal to the 5'end inhibits translation initiation by preventing the association of 40S

ribosomes[3,4], whereas the inhibitory effect of the equally stable SL decreases dramatically as its position was moved downstream (e.g., 16nt) of the cap[2,4]. Similar phenomenon was found when the antisense oligonucleotides (ASOs) complementary to the particular region of 5'UTR were introduced[5]. On the other hand, SL-to-AUG distances are able to determine the translation efficiency as well. SLs present upstream of an initiation codon have been reported to block the translation of endogenous transacripts[6,7]. Intriguingly, for all eukaryotic species, a significant decrease in mRNA secondary structure contents around the initiation codon have been universally observed in recent genome-wide structural studies[6,8,9,10]. Despite the wide appreciation of the impact that secondary structures at various regions may have on translation initiation, little is known about the differences in their mechanistic details.

Ribosomal scanning through structural barriers is known to be regulated through helicase-mediated destabilization of RNA structures. RNA helicases associated with the 43S PIC are able to unwind the stable stems and assist the ribosome to access the downstream start codons. Eukaryotic initiation factor 4A (eIF4A) and DExH/D-box helicase 29 (DHX29) are two prominent RNA helicases which are part of the translation machinery mainly involved in initiation. As one of the key components of the eIF4F complex, which facilitates the association of the 40S ribosome and mRNAs, eIF4A was suggested to remodel the cap-proximal secondary structures via its helicase activity. However, the ATPase and helicase activity of individual eIF4A is low. Despite the

enhancement of its activity by its accessory factors like eIF4G and eIF4B, eIF4A still shows limited capability in unwinding stem loops[11]. Even though, eIF4A is essential for the general translation of transcripts regardless of the structure complexity[12]. eIF4A has been recently proposed to promote the ribosome loading on mRNAs in universal initiation independent of its unwinding activity. Given that, unwinding of stable structures in 5'UTR must rely on other RNA helicases like DHX29.

DHX29 is an RNA helicase which is critical for translation initiation. It is a ubiquitously expressed cytoplasmic protein, and is found in a wide range of human and murine cell lines[13]. The knockdown of DHX29 also caused a dissociation of polysomes but an increase of the monosome in polysome profiling – 'hallmarks' of translation initiation inhibition. Both *in vitro* and *in vivo* studies have reported that DHX29 is necessary for the translation of mRNAs with highly structured 5' UTRs[14,15,13]. Recent cryo-EM structure of 48S ribosome revealed that DHX29 resides along the A-site of the ribosome[16]. Due to its position at the leading edge of the mRNA channel, DHX29 was thought to promote scanning of PIC by directly unwinding the secondary structures before entering into the mRNA binding cleft within the ribosome[17]. Beyond the general function in the RNA structure unwinding, DHX29 has been suggested to promote 48S PIC formation, although the specific mechanism is uncertain.

Despite the well-accepted fact that eIF4A and DHX29 play non-redundant roles in the translation initiation, a comprehensive understanding of

their non-overlapping functions to promote translation and unwind of secondary structures is lacking. In this study, the ability of eIF4A and DHX29 to promote the unwinding of secondary structures in different regions of the mRNA was compared in order to better understand the wide scope of functions these two RNA helicases may have. Unexpectedly, we found that the structures in CDS close to the start codons, like those in the 5'UTR, significantly impaired the translation when DHX29 was depleted. Whereas, all the mRNA reporters showed a uniform decrease in translation efficiency in the absence of eIF4A. Moreover, DHX29 was found in both 40S and 80S monosome fractions after polysome fractionation. Together all, we proposed that DHX29 possesses more functions than just the regulation of translation initiation and may even participate in the early stages of elongation.

3.3 Results

3.3.1 5' proximal SLs close to the AUG abrogate the start codon recognition.

To investigate how the position of SLs in 5'UTR affects the translation initiation, I first validated the inhibitory effect of 5' proximal SLs on translation by comparing the translatability of Firefly luciferase (Fluc) mRNA reporters inserted with a stable SLs (ΔG= -20 kcal mol^{-1}) 20nt, 6nt, or 2nt downstream of the 5' end (Figure 3-1A left panel). As expected, I observed a dramatic (around 4-fold) reduction in translation efficiency when the cap-to-SL distance was shortened from 20nt to 2nt (Figure 3-1A right panel). Although 5' proximal SL has been

suggested to inhibit translation initiation by preventing the association of 40S ribosomes[3,4], it remains unclear whether features other than the cap-to-SL distance could determine the effect of 5' proximal SLs. To examine whether the SL-to-AUG distances also play critical roles in the attenuated translation caused by the 5' proximal SLs, I measured the Fluc activities of mRNA reporters with the same 5' proximal SLs (cap-2nt-SL) while containing an untranslated region of various length (from 20nt to 2nt) between the SL and the downstream AUG. Interestingly, I found that SL-to-AUG distance shorter than 10 nt leads to a complete abrogation of Fluc translation, indicated by the observed Fluc activity comparable to the basal level. In comparison, the 2nt-SL-20nt-AUG reporter exhibited a robust expression level (Figure 3-1B). In Chapter 2, I showed that the length of the 5'UTR determines the position of the decoding site of the PIC relative to the AUG on the mRNA after the mRNA recruitment via 'slot-in'. Here, I hypothesize that, mRNAs with 5' proximal SL are capable to attach to the PIC, while the 5' proximal SL with SL-to-AUG distance less than 10 nt places the start codon within the 'blind-spot' of the PIC and abolishes its recognition (Figure 3-1C). To evaluate this hypothesis, I conducted the toe-printing assay to reveal the exact position of start codons selected by PIC in the rabbit reticulum lysates (RRL). Sharp and strong peaks corresponding to the first AUG downstream of the 5' proximal SL were detected for 2nt-SL-20nt-AUG reporters. However, no start codon-associated peaks were detected for the other reporters with SLs proximal to both 5'ends and start codons (Figure 3-1D). Together the present

results, both *in vitro* and *in vivo* data demonstrated that PICs load onto the mRNA in the downstream region of the 5' proximal SLs and the proximal downstream AUG was placed within the 'blind-spot', which leads to the failure in the start codon recognition by the PIC.

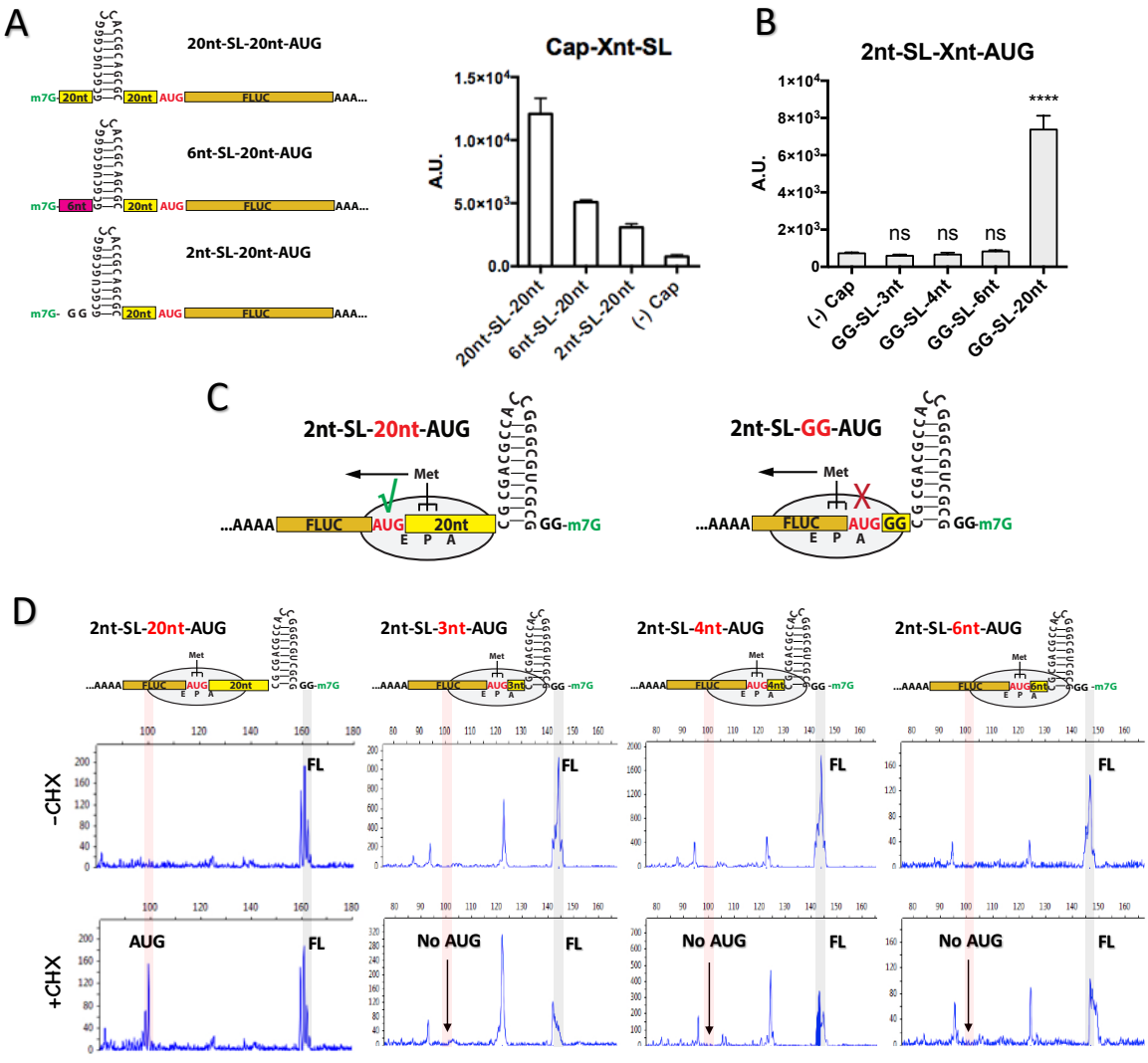

Figure 3-1 Translational effects of 5' proximal SLs.
(A) Translation efficiency of Fluc mRNA reporters bearing SLs proximal to 5'end. (Left) Sequence of Fluc reporter mRNAs. (Right) Fluc activity of indicated mRNAs. In vitro synthesized mRNAs were transfected into MEF cells and Fluc activities were monitored by real-time luminometry. The reporter mRNAs bearing SLs 20nt downstream of the 5' end were included as positive control. Uncapped mRNAs were included as negative control. Error bars: mean ± SEM; n = 3 biological replicates. (B) Translational effects of SL-to-AUG distance on the Fluc mRNA reporters bearing SLs proximal to 5'end. Fluc mRNA reporters, bearing 5' proximal SLs with indicated SL-to-AUG distance, were transfected into MEF cells, followed

by quantification of Fluc activities by real-time luminometry. Error bars: mean ± SEM; n = 3 biological replicates, ns, not significant, ****P<0.0001; One-way ANOVA. (C) Schematic of PIC loading on to the mRNAs bearing SLs proximal to 5'end. (Left) On the 2nt-SL-20nt-AUG reporter, PIC loading placed the AUG downstream of the decoding site and allowed the AUG recognition. (Right) On the 2nt-SL-GG-AUG reporter, PIC loading placed the AUG upstream of the decoding site and disabled the AUG recognition. (D) Start codon recognition by toe-printing assays. Fluc mRNA reporters, bearing 5' proximal SLs with indicated SL-to-AUG distance, were incubated in RRL, followed by reverse transcription using a probe downstream of the start codon. Expected positions corresponding to the full length and the ribosome leading edge are highlighted.

3.3.2 AUG-proximal SLs and AUG-distal SLs in the 5'UTR show similar dependence on DHX29 and eIF4A1.

To further examine the relationship between SL-to-AUG distances and the mRNA translation efficiency, I constructed a set of Fluc reporters with a 5' proximal (6nt) or distal (20nt or 56nt) SL, followed by an AUG either 2nt or 20nt downstream (Figure 3-2A). The 56-nt sequence used here is from the full length of unstructured β-globin 5'UTR[18]. Regardless of the cap-to-SL distances, mRNAs with SLs further upstream from the start codons were always translated more efficiently than those close to the start codons (Figure 3-2A), suggesting that the SL-to-AUG distance does affect the translation efficiency. This observation is in alignment with the previous sequencing data that AUG-proximal SLs in the 5' UTR are associated with decreased ribosome occupancy[6,7]. Notably, this trend is even more significant in the reporters with 5' proximal SLs (6nt-SL-GG-AUG vs 6nt-SL-20nt-AUG), compared to the mRNAs with longer cap-to-SL distances (e.g., 20 nt and 56 nt) (Figure 3-2B). The observed difference agrees with the hypothesis that SLs proximal to both 5' end

and the downstream AUG disrupt the start codon recognition by the PIC as discussed before (Figure 3-1C Right panel).

To elucidate the mechanistic details regarding the translational effects of AUG-proximal SLs in the 5'UTR, I tested whether they are more dependent on particular RNA helicases than AUG-distal SLs. Using the CRISPR/Cas9 gene editing technique, I deleted *Eif4A1* and *Dhx29* individually in the MEF cell-line. The knockdown efficiency was verified through Immunoblotting (Figure 3-2B and 2C left panels). Fluc reporters with relatively long cap-to-SL distances (e.g., 20 nt and 56 nt) were transfected into the eIF4A1 knockdown, DHX29 knockdown, and Scramble MEF cells, and the fold change (Knockdown versus Scramble) in the Fluc activity was compared. EIF4A1 depletion decreased the translation of all the reporter mRNAs, as evidenced by fold change less than 1. However, no difference was observed in the fold change between the structured and unstructured mRNAs (Figure 3-2B right two panels). This data suggested that eIF4A1 depletion impairs the overall translation rather than selectively inhibits the translation of mRNAs with structured 5' UTRs. In contrast, mRNAs with structured 5' UTRs showed a much lower fold change than those with unstructured 5' UTRs upon DHX29 depletion. (Figure 3-2C right two panels) This is expected as DHX29 is known to promote the unwinding of stable SLs during PIC scanning[14]. However, there was no difference in the fold change between the reporters with AUG-proximal SL (SL-2nt-AUG) and the reporters with AUG-distal SL (SL-20nt-AUG) (Figure 3-2C right two panels). This lack of

discrepancy indicates that DHX29 does not selectively enhance the unwinding of SLs closer or further to the start codon in the 5'UTR. Together, I conclude that DHX29 selectively promotes translation of mRNAs with structured 5' UTRs, whereas eIF4A1 does not. However, neither eIF4A1 nor DHX29 mediates the translational effects of AUG-proximal SLs in the 5'UTR on the mRNA translatability.

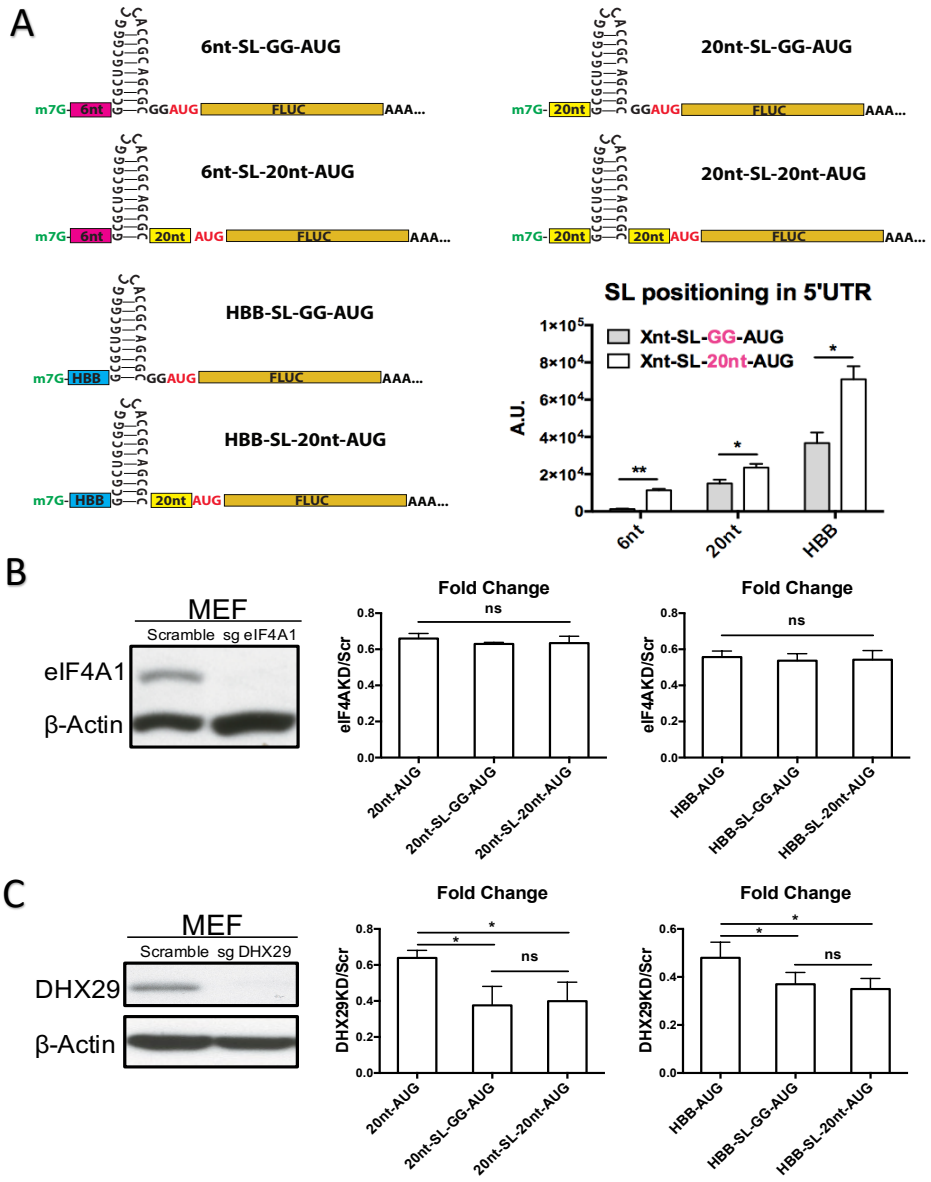

Figure 3-2 Translational effects of SL-to-AUG distance.
(A) Translation efficiency of Fluc mRNA reporters bearing SLs 20nt or 2nt upstream of the start codon. The indicated mRNA reporters were transfected into MEF cells, followed by quantification of Fluc activities by real-time luminometry. Error bars: mean ± SEM; n = 3 biological replicates. *t*-test, * $p < 0.05$; ** $p < 0.01$. Sequence of Fluc reporter mRNAs were illustrated. (B) Translation efficiency of Fluc mRNA reporters with indicated SL-to-AUG distance in MEF cells with or without eIF4A1 knockdown. (Left) Immunoblotting of MEF cells with or without eIF4A1 knockdown. (Middel and Right) Fluc activities in the absence of eIF4A1 were normalized to the control. (C) Translation efficiency of Fluc mRNA reporters with indicated SL-to-AUG distance in MEF cells with or without DHX29 knockdown. (Left) Immunoblotting of MEF cells with or without DHX29 knockdown. (Middel and Right) Fluc activities in the absence of DHX29 were normalized to the control. Error bars: mean ± SEM; n = 3 biological replicates. One-way ANOVA for multiple comparisons and unpaired Student's *t*-test for individual comparisons, ns, not significant; * $p < 0.05$.

3.3.3 DHX29 mediates the unwinding of SLs in the CDS near the AUG.

In addition to the SLs in the 5'UTR, SLs outside the 5' UTR were monitored as well. I constructed another set of Fluc mRNA reporters harboring SLs at different positions within the CDS (Figure 3-3A). Unexpectedly, mRNAs with SLs in CDS near the start codon, but not the stop codon, showed sensitivities to the depletion of DHX29. The dependence of the 20nt-AUG-GG-SL reporter on DHX29 could be explained by unfinished initiation allowing DHX29 to unwind the SLs at the leading edge of the PIC before the decoding site reaches the start codon. However, the SL 20nt downstream of the AUG ought to be unwound by the 80S ribosome during the early elongation phase where DHX29 has been released.

ASOs could inhibit translation through highly specific hybridization to a target mRNA[19,20]. To confirm the dependence on DHX29 of mRNAs with SLs in CDS close to the AUG, I designed 15-nt DNA oligonucleotides (oligos) complementary to different regions of the Fluc reporter mRNA with 56-nt (β-

globin) 5'UTR (Figure 3-3B). DNA oligos were annealed to the mRNA reporters, forming the DNA:RNA hybrids, followed by the transfection of annealing products into the MEF cell-lines. To exclude the impact of the DNA:RNA hybrids on the mRNA stability, Fluc activities were normalized to the mRNA levels at end points. Consistent with the previous finding using ASOs[19,20,21,22,] , DNA oligos binding to the 5'UTR near the AUG ('U20') or to the region including AUG ('AUG') impaired translation more significantly than those targeting the region further downstream of the AUG ('D1000'). Of note, DNA oligos binding to the CDS near the AUG ('D20') showed similar effects as 'U20' (Figure 3-3B). Next, I transfected the hybrids into DHX29 knockdown cells and found that the reporter mRNA was more sensitive to the depletion of DHX29 when targeted by 'U20', 'D20', and 'AUG' DNA oligos, but not 'D1000'. These findings validated that the structure in the CDS near the start codon is likely modulated by DHX29.

To further probe the translational effects of SL positioning, I applied an independent uORF reporter assay. As described in Chapter 2, the uORF reporters are transfected into a cell-line stably expressing H-2K^b, the intensity of both AF647 and GFP can be quantified by flow cytometry (Figure 3-3C). I inserted a stable SL at different positions of the reporter mRNAs. The reporter with 20-nt unstructured 5'UTR was included as control (Figure 3-3E upper panel). Similar to the Fluc reporters, SLs in the 5'UTR, the region close to AUG or the region containing AUG impeded the translation of both trace peptide and GFP. However, the SL in the untranslated region between the two ORFs only

impaired the translation efficiency of GFP (Figure 3-3E middle panel). To validate the dependence of DHX29 on translation of mRNAs with SLs in CDS close to AUG, I deleted *DHX29* in HEK293-K^b cell-line using CRISPR/Cas9 (Figure 3-3D) and transfected the uORF reporters into HEK293-K^b cells with DHX29 knockdown. As expected, DHX29 depletion dramatically reduced the AF647 signals from reporters with SLs in the 5'UTR and those with SLs overlapping the start codon, compared to unstructured control (Figure 3-3E lane 1, 2, 3). The SL between the two ORFs only exhibited minimal effects on AF647 signal upon DHX29 depletion, although the GFP signal was decreased apparently (Figure 3-3E lane 5). These findings all fall within the current understanding of DHX29's role in PIC scanning. Intriguingly, the reporter with a SL 15nt downstream of the start codon was highly susceptible to the DHX29 depletion (Figure 3-3E lane 4), similar to those with SLs in the 5'UTR. This data further validated that DHX29 involves in unwinding of SLs outside of 5'UTR, and that DHX29 may be functional beyond translation initiation.

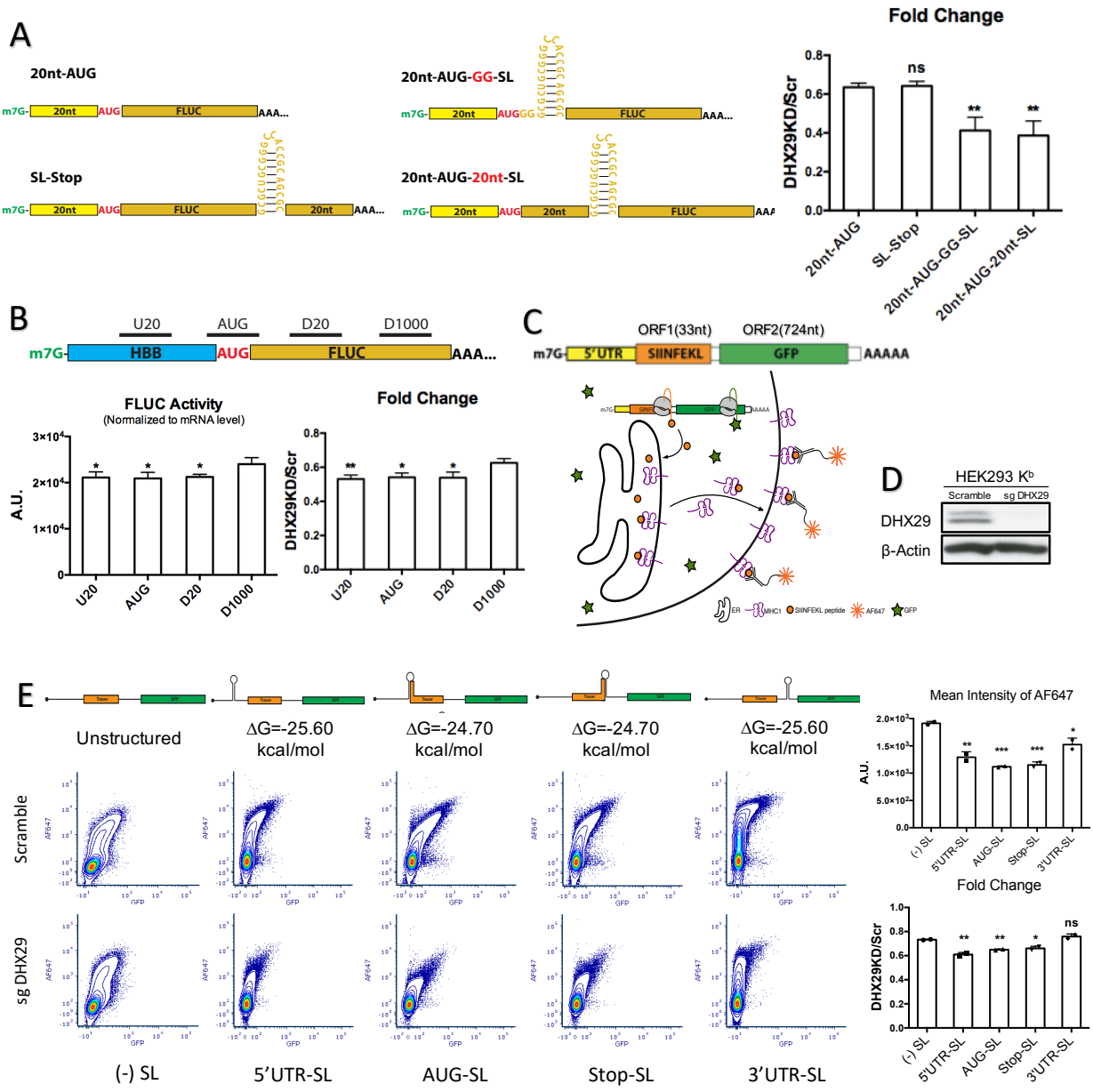

Figure 3-3 DHX29 mediates the unwinding of the CDS SLs close to the AUG.
(A) Translation efficiency of Fluc mRNA bearing CDS SLs with varied AUG-to-SL distance in MEF cells with or without DHX29 knockdown. (Left and Middle) Sequence of Fluc reporter mRNAs. (Right) Fluc activities in the absence of DHX29 were normalized to the control. The mRNAs with no SL inserted was included as positive control. The mRNAs with SL inserted 20nt upstream of the stop codon was included as another positive control. Error bars: mean ± SEM; n = 3 biological replicates. (B) Translation efficiency of Fluc mRNA reporters targeted by DNA oligos in MEF cells with or without DHX29 knockdown. (Left) The reporter mRNAs were annealed to the DNA oligos targeting the indicated regions of the mRNA. Annealing products were transfected into MEF cells and Fluc activities were monitored by real-time luminometry. Fluc activities at 6 hr were normalized to mRNA levels quantified by RT-qPCR. (Right) Fluc activities in the absence of DHX29 were normalized to the control. Error bars:

mean ± SEM; n = 3 biological replicates. (C) The schematic of uORF reporter assay coupled with FACS. The uORF reporter contains a sequence encoding SIINFEKL (orange) followed by GFP (green). Synthesized mRNA reporters were transfected into HEK293-Kb cells, followed by measurement of 25D1 and GFP using flow cytometry. (D) Immunoblotting of HEK293-K^b cells with or without DHX29 knockdown. (E) Translation efficiency of uORF mRNA reporters bearing SLs at different positions in HEK293-K^b cells with or without DHX29 knockdown. (Left)The representative flow cytometry scatterplots of HEK293-K^b Scramble cells or DHX29 knockdown cells transfected with indicated uORF reporters. mRNAs with no SL insertion were included as positive control. (Right upper) The quantification of 25D1 signals in Scramble cells. (Right bottom) 25D1 signals in the absence of DHX29 were normalized to the control. Error bars: mean ± SEM; n = 3 biological replicates. One-way ANOVA, ns, not significant, * $p < 0.05$, **$p<0.01$, ***$p<0.001$.

3.3.4 eIF4A1 is minimally engaged in unwinding SL in the CDS region.

Next, I sought to determine whether eIF4A1 and DHX29 have similar or distinct activities in stimulating unwinding of SLs in the CDS close to the start codons. I repeated the Fuc reporter assays with either SLs inserted or DNA oligos targeted in eIF4A1 knockdown MEF cells and observed a uniform fold change in all constructs (Figure 3-4A and 4B). I also created a HEK293-K^b cell-line with eIF4A1 knockdown using CRISPR-Cas9 gene editing method to conduct the TIS-Selex assay (Figure 3-4C). All the mRNA reporters displayed an equally strong dependence on eIF4A1 including the unstructured control (Figure 3-4D). These observations are in alignment with the recent notion of eIF4A1 that it is essential to the global mRNA translation, regardless of the structural complexity.

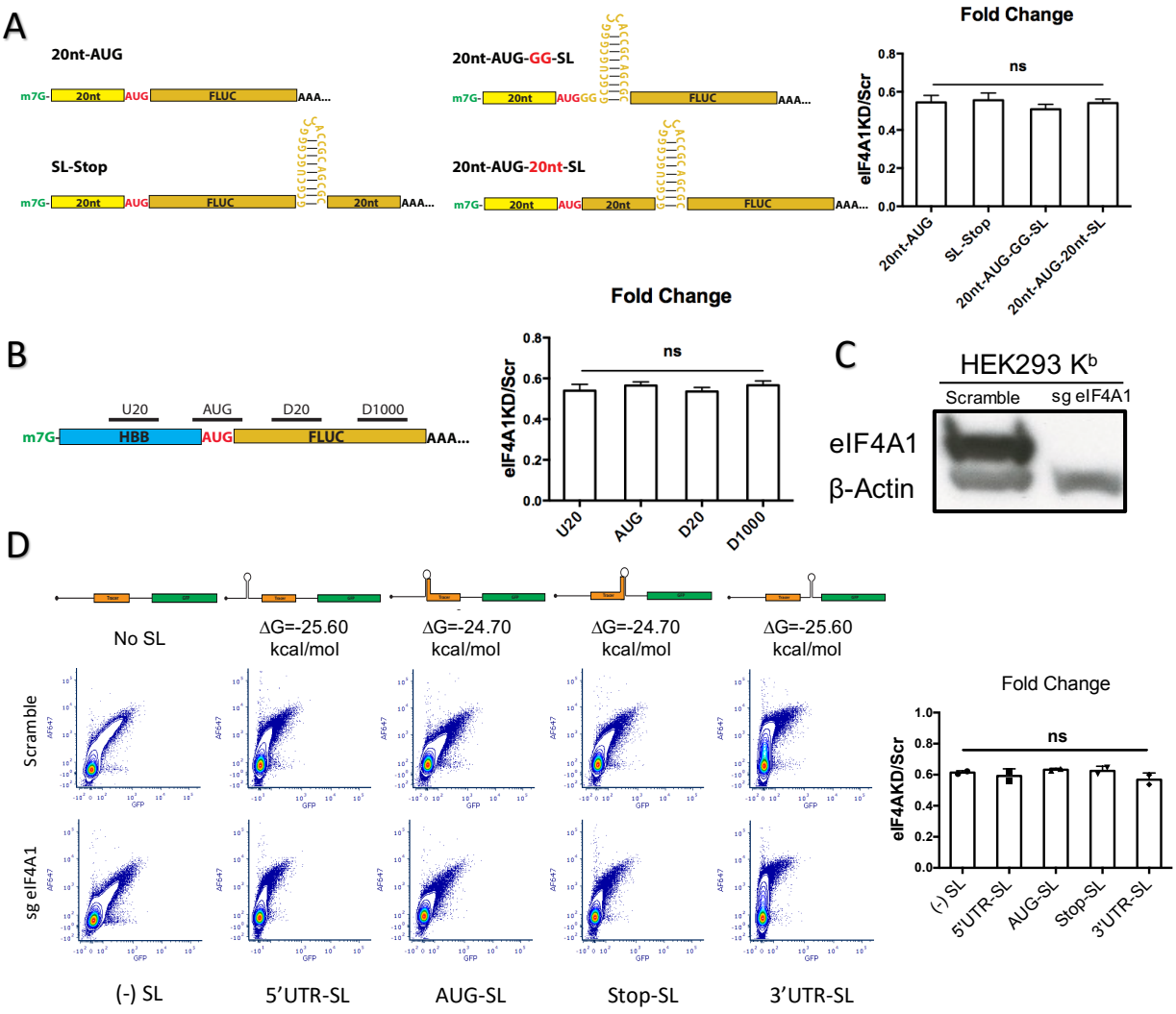

Figure 3-4 eIF4A1 is minimally engaged in unwinding of SLs in CDS near the AUG.
(A) Translation efficiency of Fluc mRNA bearing CDS SLs with varied AUG-to-SL distance in MEF cells with or without eIF4A1 knockdown. Similar to Figure 3-3A (B) Translation efficiency of Fluc mRNA reporters targeted by DNA oligos in MEF cells with or without eIF4A1 knockdown. Similar to Figure 3-3B (C) Immunoblotting of HEK293-K^b cells with or without eIF4A1 knockdown. (D) Translation efficiency of uORF mRNA reporters bearing SLs at different positions in HEK293-K^b cells with or without eIF4A1 knockdown. (Left) The representative flow cytometry scatterplots of HEK293-K^b Scramble cells or eIF4A1 knockdown cells transfected with indicated uORF reporters. Similar to Figure 3-3E. Error bars: mean ± SEM; n = 3 biological replicates. One-way ANOVA, ns, not significant. (Right) 25D1 signals in the absence of eIF4A1 were normalized to the control. Error bars: mean ± SEM; n = 3 biological replicates. One-way ANOVA, ns, not significant.

3.3.5 Impact of DHX29 depletion and eIF4A1 depletion on global translation.

To further study and compare the functional differences between DHX29 and eIF4A1, I measured the relative growth rates of MEF and HEK293-K^b cells depleted with either DHX29 or eIF4A1, compared to the Scramble cell-lines. The depletion of DHX29 minimally reduced the growth rate in both cell lines. Differences in growth rate was not significant ($p > 0.05$) (Figure 3-5A left panels). In agreement with the growth rate, puromycin labeling showed slight difference in the puromycin incorporation into nascent polypeptides between the DHX29 knockdown cell-lines and the Scramble controls (Figure 3-5B, left panel). Consistently, polysome profiling revealed a mild decreased but still robust polysome fractions in both DHX29 knockdown cell-lines (Figure 3-5C, right panels). These findings demonstrated that the depletion of DHX29 minimally affected the overall mRNA translation in both murine and human cell lines.

On the other hand, eIF4A1 depletion displayed a more dramatic impact on both cell growth and global mRNA translation. eIF4A1 knockdown significantly slowed down the growth rate of both MEF and HEK293 K^b cells (Figure 3-5A right panels). Also, the absence of eIF4A1 impaired the synthesis of nascent peptides, revealed by the puromycin labeling (Figure 3-5B right panels). Moreover, eIF4A1 depletion leads to a huge shift from the polysomes to the monosome indicating the remarkable inhibition of translation initiation (Figure 3-5C, middle panels). Taking all these data together, I conclude that

eIF4A1 plays a critical role in both cell growth and global mRNA translation, whereas DHX29 seems to be less essential.

To test the hypothesis that DHX29 may play a role beyond the initiation phase of translation, samples from the fractionation of the wild-type HEK293 K^b cells were collected and subjected to the immunoblotting for DHX29 to monitor its distribution across the polysome fractions. Immunoblotting for RPS26, RPL4 was conducted at the same time as the representatives for 40S and 60S. The Western blot analysis revealed that DHX29 was mostly found in the ribosome free and 40S fractions (Figure 3-5D lane 1, 2, 3), consistent with the previous report that DHX29 was most strongly associated with the 40S[14]. Notably, DHX29 was also found in the monosome and light polysome fractions (Figure 3-5D lane 5, 6, 7), suggesting that DHX29 remains associated with the 80S ribosome beyond initiation at least during the early elongation phase. In comparison, the distribution of eIF4A1 in the polysome fractions showed that eIF4A1 mainly exists in ribosome-free and 40S fractions, while only minimal eIF4A1 appears in the monosome fraction and is absent in the polysome fractions. This observation further supports the hypothesis that DHX29 remains on the 80S and is involved in unwinding SLs within the CDS during the early elongation phase, while eIF4A1 is released from the ribosome upon 80S formation.

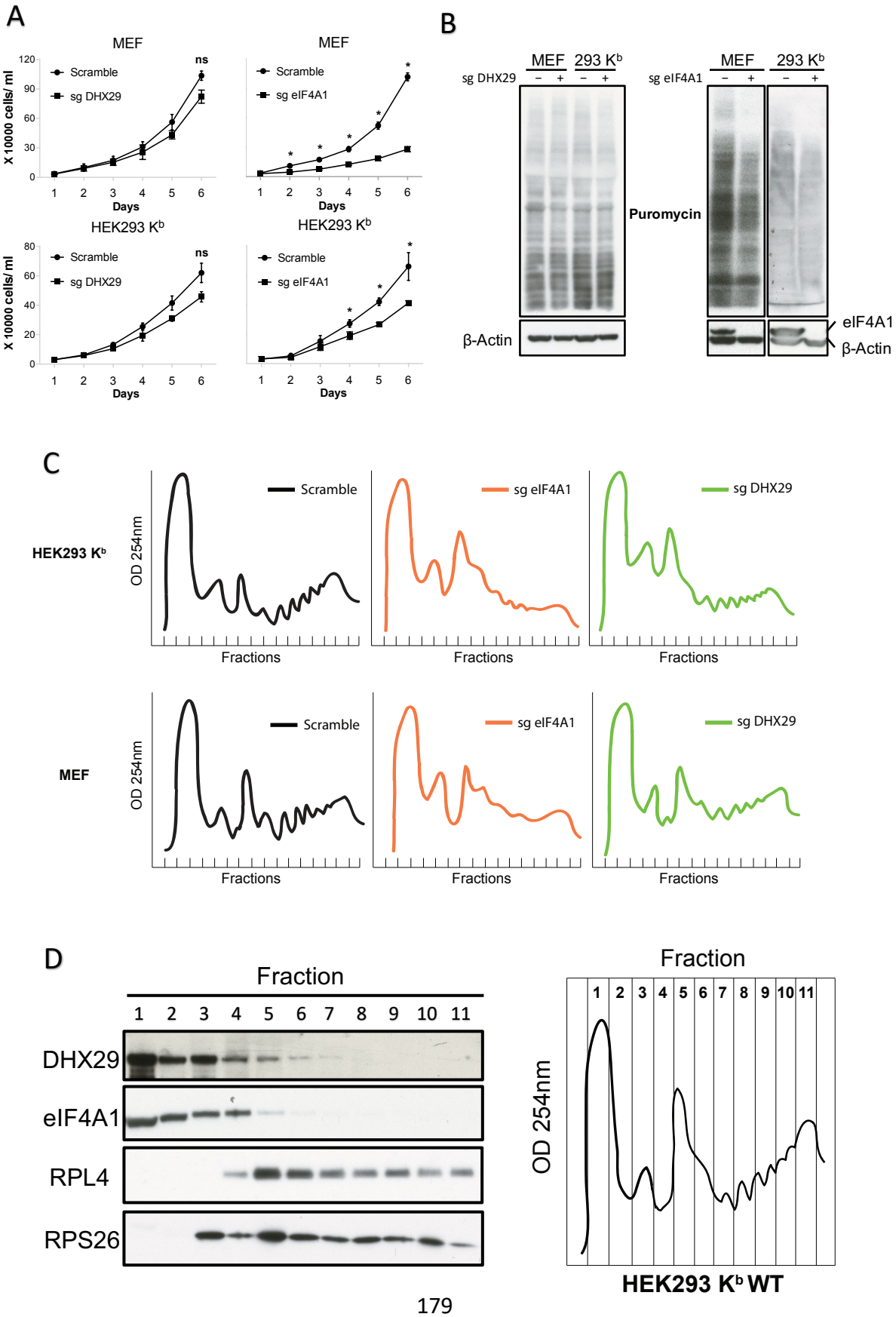
A
MEF
Scramble
sg DHX29
X 10000 cells/ ml
Days
ns
MEF
Scramble
sg eIF4A1
HEK293 Kb
Scramble
sg DHX29
ns
HEK293 Kb
Scramble
sg eIF4A1
B
MEF
293 Kb
sg DHX29
sg eIF4A1
Puromycin
β-Actin
eIF4A1
β-Actin
C
HEK293 Kb
MEF
OD 254nm
Scramble
sg eIF4A1
sg DHX29
Fractions
D
Fraction
1 2 3 4 5 6 7 8 9 10 11
DHX29
eIF4A1
RPL4
RPS26
OD 254nm
HEK293 Kb WT

Figure 3-5 Impact of DHX29 or eIF4A1 depletion on global translation.
(A) Proliferation rate of MEF and HEK293-K^b cells with either DHX29 knockdown or eIF4A1 knockdown. Cell numbers were normalized to the value obtained on Day 1. Error bars, mean ± SD, n = 3 biological replicates; ns, not significant, *P<0.05; Student's t test. (B) Nascent proteins in MEF and HEK293-K^b cells with either DHX29 knockdown or eIF4A1 knockdown were labeled with 10 μg/ml of puromycin for 5, 10 or 15 mins. Cell lysates were resolved by SDS-PAGE and analyzed by immunoblotting using the indicated antibodies. (C) Polysome profiles of MEF and HEK293-K^b cells with either DHX29 knockdown or eIF4A1 knockdown were analyzed by sucrose density gradient centrifugation in polysome buffer. (D) (Light) Distribution of DHX29 and eIF4A1 in polysome fractions. Immunoblotting of DHX29, eIF4A1 RPL4 and RPS26 in indicated fractions from HEK293-K^b wide-type cells. (Right) Polysome profile from HEK293-K^b wide-type cells.

3.4 Discussion

The canonical scanning model of translation initiation proposes that stable secondary structures within 5'UTR can inhibit translation efficiency. In this study, I examined the impact of the location of secondary structures with the same thermal stabilities on mRNA translation. I confirmed that mRNA structure is inhibitory when proximal to the 5' end of mRNA (Figure 3-1A). The previous studies suggested that the inhibitory effect of 5' proximal structures is due to impaired association between the PIC and mRNA. Using the toe-printing assay, I showed that the PIC is still able to attach to the mRNA in presence of the 5' proximal SLs when there is at least 20-nt space between the SL and the downstream AUG (Figure 3-1D). Whereas, the secondary structure proximal to both 5' end and the downstream AUG completely abrogated the start codon recognition by the PIC (Figure 3-1C). I further explored the translational effects of SL-to-AUG distances and found that the secondary structure 2nt upstream of the start codon reduced translation efficiency more dramatically, compared to those 20nt upstream (Figure 3-2A). This observation is in alignment with the

previous transcriptome studies demonstrating that 5'UTR structures close to the start codon impaired the mRNA translatability[6,7]

In addition to the structures within 5'UTR, I also tested the SLs in other regions of the mRNA. I was surprised to notice that the structures located up to 20nt downstream of the start codon exhibited similar inhibitory effects as structures in the 5'UTR (Figure 3-3A). This phenomenon was observed in both Fluc and uORF reporter assays using reporter mRNAs either inserted with SLs or targeted by the complementary DNA oligos in different regions (Figure 3-3B and 3E). The structures 20nt downstream of the start codon ought to be destabilized by the assembled 80S during the elongation, while those more proximal to the AUG (e, g,. 2nt) are likely unwound by the PIC in the late initiation stage.

Several RNA helicases have been implicated in the unwinding secondary structures during PIC scanning. It has been long appreciated that eIF4A1, in the complex of eIF4F, promotes ribosome attachment by unwinding 5' proximal structures[23]. DHX29 was found necessary for the translation of mRNA sequences with highly structured 5' UTRs[14,15,13] In this study, I intended to determine whether either of these two RNA helicases might mediate the translational effects of secondary structures in different positions. As shown in Figure 3-2, the translational efficiency of reporters with structured 5' UTRs show different sensitivity to DHX29 and eIF4A knockdown. Cells with eIF4A1 knockdown demonstrate a uniform reduction in translation efficiency of all the

reporter mRNAs, including unstructured control (Figure 3-2B). This observation agrees with more recent notions that eIF4A1 is required for mRNA recruitment by all mRNAs regardless of the structural complexity[24,25] In contrast, a more significant reduction in the translation of mRNAs with structured 5' UTRs was seen in DHX29 knockdowns (Figure 3-2C). Despite the dependence of structured 5'UTRs on DHX29, the translation of mRNAs with secondary structures in different locations of the 5'UTR show equal sensitivities to DHX29 depletion. These results indicate that the translational effect of the structure positioning is not mediated by DHX29 or eIF4A1.

DHX29 has been long considered to act in the initiation stage of translation. Its strong association with the 40S small subunit suggested that it functioned primarily in the unwinding of secondary structures during the scanning process[13]. In this study, a new function of DHX29 may have been identified. As seen in Figure 3-3 and Figure 3-4, knockdown of DHX29 but not eIF4A1 appeared to impair the translation of reporters with SLs in the CDS near the start codon. In both Fluc and uORF reporter assays, DHX29 depletion affects both the translation of reporters with SLs in the 5'UTR and overlapping the start codon. Unexpectedly, reporters with SLs in the CDS near the start codon were also highly susceptible to the DHX29 depletion (Figure 3-3A and 3E). Reduced translation of mRNAs with SLs in the 5' UTR was expected as DHX29 has been shown to promote the scanning of the 43S ribosome through secondary structures in the 5' UTR[14]. Similarly, secondary structures

overlapping the start codon are also unwound prior to 80S ribosome assembly. However, it came as a surprise that the unwinding of stable stems 20nt downstream of the AUG would also be affected by DHX29 depletion, as the unwinding of these stems would likely occur only in the elongation phase. To mitigate the potential problems posed by the unpredicted structure formation, I further validated this finding by introducing DNA oligos targeting different areas downstream of the start codon in long Fluc reporters. It was shown that the secondary structures within 20nt downstream of the AUG are significantly more sensitive to DHX29 depletion than those more distal from the AUG (Figure 3-3B). In light of this phenomenon, I hypothesize that DHX29 may play a role in the early elongation phase, and not just the initiation phase as previously thought.

The proposed role of DHX29 in the early stages of elongation is also supported by its stronger presence in the 80S monosome fraction and light polysome fractions, but much weaker presence in the heavy polysome fractions, as seen in Figure 3-5. Combined with the findings from the reporter assays, the presence of DHX29 in 80S suggests its potential role in translation beyond just the initiation stage. This notion might run counter to the convention that DHX29 dissociates from 40S complexes[26]. However, the past published polysome profiling data shows that DHX29 can be found in the 80S fraction as well[13]. Also, DHX29 was found to contact eIF3a[27] which was found to remain associated with 80S ribosome during early elongation phase[28]. Taking all this data together, I

propose that DHX29 could be present during the elongation, making it a candidate for secondary structure unwinding during elongation, given its similar role during initiation. While there is still much more research that needs to be done to prove DHX29's role beyond the initiation process, if confirmed, this could suggest a novel function for DHX29.

Previous literature suggested that the two RNA helicases involved in initiation, DHX29 and eIF4A1, may have stimulatory functions or act synergistically to unwind stable secondary structures, while also keeping some overlapping functions[14]. This study found that the absence of eIF4A1 or DHX29 both lead to reduced global translation and cell proliferation, while the reduction of translation was much more dramatic in the absence of eIF4A1 than DHX29. The huge impact of eIF4A1 depletion observed in this study is consistent with the previous notion that eIF4A1 is essential to the cell growth. Despite that, the eIF4A1 knockout was found not lethal in this study. However, previous literature had demonstrated that DHX29 knockdown with shRNAs in HeLa cells showed an average two-fold reduction in global translation levels[13]. However, the reduction of translation observed in this study was not as remarkable as previously described. This discrepancy might be due to differences in cell line, knockdown technique, global translation measurement techniques and growth rate measurement techniques between this study and the previous literature. As only 10% of mRNAs contain atypically long 5' UTRs which have a greater propensity to form stable secondary structures[29], this low prevalence of highly

structured mRNAs may explain the mild decreases in global translation and thus cell growth caused by DHX29 knockdown. In summary, the different influence of two RNA helicases on global translation and cell growth further demonstrate the non-overlapping functions of DHX29 and eIF4A1 in mammalian cells.

3.5 Materials and Methods

Cell lines and reagents

MEF and HEK293-K^b cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum (FBS). Antibodies used in the immunoblotting are listed below: anti-eIF4A1 (Abcam ab31217), anti-puromycin (Developmental Studies Hybridoma Bank PMY-2A4), anti-DHX29 (Cell Signaling 4648S), anti-β-Actin (sigma A5441), anti-RPS26 (Protein Tech 14909-1-AP), anti-RPL4 (Protein Tech 11302-1-AP), Alexa Fluor 546 donkey anti-mouse secondary antibody (Invitrogen A10036) and Alexa Fluor 546 donkey anti-rabbit secondary antibody (Invitrogen A10040).

Plasmid constructions

The full-length firefly luciferase (Fluc) gene was cloned into pcDNA3.1 vector (Invitrogen) to generate the Fluc/pcDNA3.1 vector. The uORF plasmid reporters were constructed by inserting the sequence encoding SIINFEKL into Sac I/BamH I sites of EGFP/pcDNA3.1 vector to generate SIINFEKL-

EGFP/pcDNA3.1 vector. DNA sequences of all primers used in this study are listed in Appendix IV.

In Vitro Transcription

To generate mRNAs suitable for transfection, 3 μg PCR products described above were utilized for in vitro transcription. Transcripts with the normal m^7G cap were generated using the mMESSAGE mMACHINE T7 Transcription Kit (Invitrogen 1344) and transcripts with the non-functional cap analog ApppG (NEB) were synthesized using MEGAscript T7 Transcription Kit (Invitrogen 1333), followed by polyadenylation using Poly(A) Tailing Kit (Invitrogen AM1350M). mRNA products were purified according to the manufacturer's instruction.

Real-Time Luciferase Assay

Cells grown in 35 mm dishes were transfected with in vitro synthesized luciferase reporter mRNAs. Luciferase substrate D-luciferin (1 mM, Regis Tech) was added into the culture medium immediately after transfection. Luciferase activity was monitored and recorded using Kronos Dio Luminometer (Atto).

Real-time quantitative PCR

Total RNA was isolated by TRIzol reagent (Invitrogen) and used for reverse transcription assay via High Capacity cDNA Reverse Transcription Kit

(Invitrogen). Real-time PCR analysis was conducted using Power SYBR Green PCR Master Mix (Applied Biosystems) and carried on a LightCycler 480 Real-Time PCR System (Roche Applied Science). Primers for amplifying each target are listed in Appendix IV.

Flow cytometry

Transfected HEK293-K^b cells were washed with PBS and harvested by trypsin. Cells were then re-suspended in blocking buffer (1% bovine serum albumin (BSA) in PBS). Cells are aliquoted into a 96-well plate followed by 2000 rpm spinning for 2 min. After removal of blocking buffer, cells were washed one more time followed by staining with 25D1 Alexa 647 antibody (1:1000 in 75 uL solution per well). After incubation in the dark with gentle rocking at 4°C for 30 min, cells were washed three times with 200 uL of the blocking buffer to remove unbound antibodies. Resuspend cells in 300 µl of blocking buffer followed by single cell filtering (Falcon). Cells were analyzed on a BD FACSAria Fusion flow cytometer (BD Biosciences). Cytometry data analysis is conducted using FlowJo.

DNA-RNA Oligonucleotide Annealing

The annealing mixture was prepared on ice in a total volume of 20 µl containing 15 picomole of DNA oligonucleotides, 5 picomole of mRNA reporters, 4 μL of 100 mM HEPES, 4 μL of 10 mM MgCl2 and 2 μL of 1M KCl. DNA-RNA. Annealing occurred at 65 ºC for 10 minutes. After 10-minute incubation, the

thermocycler was turned off, and the reactions slowly cool down to the room temperature for 30 minutes. The annealing products were ready for transfection.

Toeprinting

The ribosome binding reaction mixture was prepared on ice in a total volume of 10 µl containing 50% Nuclease-treated rabbit reticulocyte lysate (Promega), 20 pmol of primer (5'- 6-FAM-AATTGTTCCAGGAACCAG- 3'), 20 µM amino acid mixture, 0.4 U/µl RNaseOUT (Invitrogen) and 50 mM Tris-HCl pH 7.5. Reactions were treated with 0.5 mg/ml CHX followed by incubation at 37 °C for 5 min. After addition of 0.3 mg of reporter mRNAs, the reaction mixtures were incubated at 30 °C for 20 min to allow the translation machinery to assemble at the start codon. The reverse transcriptase reaction was conducted in a total volume of 20 µl containing the entire ribosome binding reaction, 1x Superscript III reverse transcriptase buffer, 5 mM DTT, 40 mM KCl, 3 mM $MgCl_2$, 50 mM Tris-HCl pH7.5, 0.5 mg/ml cycloheximide, 0.8 mM of dNTP, 1.5 U/µl RNaseOUT, 5 U/µl Superscript III reverse transcriptase. After incubation at 25 °C for 10 min, the reaction was terminated by nucleic acids extraction by phenol:chloroform and ethanol precipitation. The primer extension products were resuspended in 10 µl of Hi-Di formamide. 2 µl aliquot was run with 0.2 µl GeneScan 500 LIZ dye Size Standard (Fisher) on an ABI 3730xl instrument. Data is analyzed by Peak Scanner 2 software.

Lentiviral sgRNAs

LentiCRISPRv2 plasmids targeting eIF4AI were constructed using methods described previously (30,31). Briefly, complementary oligonucleotides containing the specific sgRNA sequence and overhangs complementary to overhangs generated by BsmBI digestion of LentiCRISPRv2 were annealed to the BsmBI digested LentiCRISPRv2 plasmid to generate the functional transfer vector. Undigested LentiCRISPRv2 plasmid lacking a sgRNA sequence was used for pseudovirus production as a control (Scramble). Lentiviral particles were packaged using Lenti-X 293T cells (Clontech). Virus-containing supernatants were collected at 48 hr after transfection and filtered to eliminate cells. MEF cells were infected by the lentivirus for 48 hr followed by selection with 2 μg/ml puromycin. Selected cells were plated in 96-well plates by serial dilution at 1 cell per well. Single cells were then expanded and analyzed by PCR amplification of genomic DNA flanking the CRISPR-targeted region. sgRNA targeting sequence used in this study is listed in Appendix IV.

Growth Curve

Cell growth rates were compared through cell counting using a hemocytometer. Cells were plated for a 10% confluency into 6-well plates. The first well was counted eight hours post-plating in order to set a Day 1 baseline reading. All other wells were counted 24 hours after Day 1. Relative cell growth was calculated as the cell count for that day normalized to the Day 1 count. Cell

growth was tracked for six days. Cells were plated in triplicates and relative growth rates for each of the replicated were averaged.

Puromycin Labeling

Cells at 80-90% confluence were treated with DMEM + 10% FBS 2 hr before media was changed to DMEM + 10% FBS supplemented with 10 µg/ml puromycin for an additional 5 min, 10 min or 15 min. Cells were washed twice with ice-cold DPBS and lysed, followed by immunoblotting.

Immunoblotting

Cells were lysed on ice in TBS buffer (50 mM Tris pH7.5, 150 mM NaCl, 1 mM EDTA) containing 1% Triton X-100, 2 U/ml DNase and protease inhibitor cocktail tablet. The lysates were incubated on ice for 30 min, followed by heating for 10 min in SDS-PAGE sample buffer (50 mM Tris pH6.8, 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, 10% glycerol). Proteins were separated on SDS-PAGE and transferred to PVDF membranes (Fisher). Membranes were blocked in TBS containing 5% non-fat milk and 0.1 % Tween-20 for 1 hr, followed by incubation with primary antibodies overnight at 4 °C. After incubation with horseradish peroxidase-coupled secondary antibodies at room temperature for 1 hr, immunoblots were visualized using enhanced chemiluminescence (ECLPlus, GE Healthcare).

Polysome Fractionation

Five 10 cm dishes of cells were harvested in 450 μl lysis buffer (1% Triton X-100 in polysome buffer) containing CHX (100 μg/ml), then centrifuged at 12,000 g 4 °C for 10 min. The supernatant was collected and subjected to sucrose gradient sedimentation. Sucrose solutions were prepared in polysome buffer. 15%- 45% (w/v) Sucrose density gradients were freshly prepared in a SW41 ultracentrifuge tube (Backman) using a Gradient Master (BioComp Instruments). Supernatant was loaded onto sucrose gradients followed by centrifugation for 2 hr 30 min at 32,000 rpm 4 °C in a SW41 rotor. Separated samples were fractionated at 1.5 ml/min through an automated fractionation system (Isco) that continually monitors OD254 values.

3.6 Acknowledgements

Many thanks go to Kevin Lin who worked hard with me on this Chapter. He provided huge assistance in cell culture, single-clone selection as well as immunoblotting. I would like to thank Dr. Jonathan W. Yewdell for providing 25D1 reagents and HEK293-K^b cells. I am grateful to Cornell University Life Sciences Core Laboratory Center for FACS. This work was supported by US National Institutes of Health (R01GM1222814 and R21CA227917) and HHMI Faculty Scholar (55108556) to S.-B.Q.

REFERENCE

1. Shirokikh NE, Preiss T. Translation initiation by cap-dependent ribosome recruitment: Recent insights and open questions. *Wiley Interdiscip Rev RNA*. 2018. doi:10.1002/wrna.1473
2. Kozak M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem*. 1991;266(30):19867-19870.
3. Kozak M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol*. 1989. doi:10.1128/mcb.9.11.5134
4. Babendure JR, Babendure JL, Ding JH, Tsien RY. Control of mammalian translation by mRNA structure near caps. *RNA*. 2006. doi:10.1261/rna.2309906
5. Baker BF, Lot SS, Condon TP, et al. 2'-O-(2-methoxy)ethyl-modified anti-intercellular adhesion molecule 1 (ICAM-1) oligonucleotides selectively increase the ICAM-1 mRNA level and inhibit formation of the ICAM-1 translation initiation complex in human umbilical vein endothelial cells. *J Biol Chem*. 1997. doi:10.1074/jbc.272.18.11994
6. Kertesz M, Wan Y, Mazor E, et al. Genome-wide measurement of RNA secondary structure in yeast. *Nature*. 2010;467(7311):103-107. doi:10.1038/nature09322
7. Aw JGA, Shen Y, Wilm A, et al. In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Mol Cell*. 2016. doi:10.1016/j.molcel.2016.04.028
8. Wan Y, Qu K, Zhang QC, et al. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature*. 2014. doi:10.1038/nature12946
9. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*. 2014. doi:10.1038/nature12756
10. Incarnato D, Neri F, Anselmi F, Oliviero S. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol*. 2014. doi:10.1186/s13059-014-0491-2
11. Abaeva IS, Marintchev A, Pisareva VP, Hellen CUT, Pestova T V. Bypassing of stems versus linear base-by-base inspection of mammalian mRNAs during ribosomal scanning. *EMBO J*. 2011;30(1):115-129. doi:10.1038/emboj.2010.302
12. Pestova T V., Kolupaeva VG. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev*. 2002. doi:10.1101/gad.1020902
13. Parsyan A, Shahbazian D, Martineau Y, et al. The helicase protein DHX29 promotes translation initiation, cell proliferation, and tumorigenesis. *Proc Natl Acad Sci U S A*. 2009;106(52):22217-22222. doi:10.1073/pnas.0909773106
14. Pisareva VP, Pisarev A V., Komar AA, Hellen CUT, Pestova T V. Translation Initiation on Mammalian mRNAs with Structured 5′UTRs Requires DExH-Box Protein DHX29. *Cell*. 2008;135(7):1237-1250. doi:10.1016/j.cell.2008.10.037
15. Chiu W-L, Wagner S, Herrmannová A, et al. The C-Terminal Region of Eukaryotic Translation Initiation Factor 3a (eIF3a) Promotes mRNA Recruitment, Scanning, and, Together with eIF3j and the eIF3b RNA Recognition Motif, Selection of AUG Start Codons. *Mol Cell Biol*. 2010. doi:10.1128/mcb.00280-10
16. Hashem Y, des Georges A, Dhote V, et al. Structure of the Mammalian Ribosomal 43S Preinitiation Complex Bound to the Scanning Factor DHX29. *Biophys J*. 2014. doi:10.1016/j.bpj.2013.11.2750
17. Abaeva IS, Marintchev A, Pisareva VP, Hellen CUT, Pestova T V. Bypassing of stems versus linear base-by-base inspection of mammalian mRNAs during ribosomal

scanning. *EMBO J*. 2011;30(1):115-129. doi:10.1038/emboj.2010.302
18. Lockard RE, Currey K, Browner M, Lawrence C, Maizel J. Secondary structure model for mouse β globin mRNA derived from enzymatic digestion data, comparative sequence and computer analysis. *Nucleic Acids Res*. 1986. doi:10.1093/nar/14.14.5827
19. Heasman J, Kofron M, Wylie C. β-catenin signaling activity dissected in the early Xenopus embryo: A novel antisense approach. *Dev Biol*. 2000. doi:10.1006/dbio.2000.9720
20. Faria M, Spiller DG, Dubertret C, et al. Phosphormidate oligonucleotides as potent antisense molecules in cells and in vivo. *Nat Biotechnol*. 2001. doi:10.1038/83489
21. Braasch DA, Liu Y, Corey DR. Antisense inhibition of gene expression in cells by oligonucleotides incorporating locked nucleic acids: Effect of mRNA target sequence and chimera design. *Nucleic Acids Res*. 2002. doi:10.1093/nar/gkf651
22. Doyle DF, Braasch DA, Simmons CG, Janowski BA, Corey DR. Inhibition of gene expression inside cells by peptide nucleic acids: Effect of mRNA target sequence, mismatched bases, and PNA length. *Biochemistry*. 2001. doi:10.1021/bi0020630
23. Jackson RJ, Hellen CUT, Pestova T V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*. 2010;11(2):113-127. doi:10.1038/nrm2838
24. Yourik P, Aitken CE, Zhou F, Gupta N, Hinnebusch AG, Lorsch JR. Yeast eIF4A enhances recruitment of mRNAs regardless of their structural complexity. *Elife*. 2017. doi:10.7554/eLife.31476
25. Sen ND, Zhou F, Ingolia NT, Hinnebusch AG. Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. *Genome Res*. 2015;25:1196-1205. doi:10.1101/gr.191601.115.25
26. Parsyan A, Svitkin Y, Shahbazian D, et al. mRNA helicases: the tacticians of translational control. *Nat Rev Mol Cell Biol*. 2011;12(4):235-245. doi:10.1038/nrm3083
27. Pisareva VP, Pisarev A V. DHX29 and eIF3 cooperate in ribosomal scanning on structured mRNAs during translation initiation. 2016:1-12. doi:10.1261/rna.057851.116.in
28. Szamecz B, Rutkai E, Cuchalová L, et al. eIF3a cooperates with sequences 5′ of uORF1 to promote resumption of scanning by post-termination ribosomes for reinitiation on GCN4 mRNA. *Genes Dev*. 2008. doi:10.1101/gad.480508
29. Pickering BM, Willis AE. The implications of structured 5′ untranslated regions on translation and disease. *Semin Cell Dev Biol*. 2005. doi:10.1016/j.semcdb.2004.11.006
30. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods*. 2014. doi:10.1038/nmeth.3047
31. Shalem O, Sanjana NE, Hartenian E, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (80-)*. 2014. doi:10.1126/science.1247005

CHAPTER 4

Conclusions and Future Endeavors

The selection of proper start codons by ribosomes is a pivotal step in protein synthesis, ensuring that the correct polypeptide is produced from the mRNA. The dynamic start codon selection is coupled to the scanning process of the pre-initiation complex (PIC). One of the most fundamental questions regarding the scanning directionality is still under debate. Several DEAD-box RNA helicases have been implicated in the translation initiation on highly structured mRNA leaders. For instance, the yeast Ded1p catalyzes the forward scanning of PIC by interacting at the mRNA entry channel and unwinding mRNA secondary structures. Reduced Ded1p expression during meiosis contributes to the activation of uORFs by selecting near-cognate start codons. The mammalian specific-DHX29 uses the similar mechanism to assist forward scanning. Despite that the classic linear scanning model (i.e., 5'→3' unidirectional migration of PIC along the 5' leader sequence) is widely accepted, several studies have suggested that this process, rather than exclusively unidirectional, may consist of forward (5' to 3') thrusts alternating with motion the reverse direction (3' to 5') [1,2,3].

In Chapter 2, I aimed to address the question of how the scanning directionality is determined. Using both in vivo reporter assays and quantitative probing of start codon selection in vitro, I provide evidence suggesting that PIC undergoes bi-directional oscillations with a net 5'→3' movement. The

translatability of mRNAs with ultra-short 5'UTR indicates the lack of an absolute 'blind spot' within 5'UTR for start codon selection. Intriguingly, the cross-linking between eIF4E and m^7G cap abolished the translation of mRNA with ultra-short 5'UTR. The dependence of the 'blind spot' on the 5' block suggests a lateral attachment of mRNA ('slot-in' model) during the ribosome loading. The preference on the third AUG among the three closely spaced AUGs in mRNAs with ultra-short 5'UTR hints the 3' to 5' migration of the PIC after slot-in. The backward scanning of PIC was further validated in the scanning-dependent IRES-mediated translation. The bi-directional scanning model unifies the 'slot-in' mode of PIC attachment, cap-severed PIC scanning, and the competitive nature of neighboring start codons for PIC recognition.

The most unexpected finding is the role of eIF4A in bi-directional scanning of PIC. As the prototypical member of DEAD-box RNA helicases, eIF4A is the most abundant translation initiation factor. Similar to prior studies using yeast cells, genetic ablation of eIF4A1 from mammalian cells represses global protein synthesis with little mRNA specificity. In addition to reduced PIC attachment, recognition of start codons near the 5' end is also impaired in the absence of eIF4A1. By manipulating the ATPase activity of eIF4A using small molecule compounds, we demonstrate that it is possible to control the power of the scanning 'motor' as well as its directionality. Unlike DHX29 that is stably positioned at the leading edge of PIC, eIF4A is believed to recycle rapidly during the scanning process. Such action might be required to avoid a futile scanning

process. Instead, it allows the net 5'→3' movement punctuated with bi-directional oscillations. Future studies using the single molecule strategy might elucidate the underlying molecular details. We anticipate that balancing the activity of forward and backward helicases fine-tunes the stringency of start codon selection, which subsequently controls the translational output.

During the canonical scanning of PIC, stable secondary structures within 5'UTR can inhibit the efficiency of start codon recognition and translation initiation. The primary goal of chapter 3 was to investigate the effects of structure positioning on translation efficiency. I first designed Fluc reporter mRNAs with SLs inserted at different positions in 5'UTR. In agreement with published data, I found that the SLs in 5'UTR proximal to either the 5' end or the AUG impede mRNA translation more dramatically than those distal from both ends of 5'UTR. To be noted, results from toe-printing assay revealed that AUG recognition by PIC was completely abrogated when the SL is proximal to both 5' end and AUG. This phenomenon can be explained by the 'slot-in' model of ribosome attachment proposed in chapter 2.

In addition to the SLs in 5'UTR, I also tested the translational effects of SLs outside the 5'UTR. Owing to a very short SIINFEKL uORF, the uORF reporter assay allowed me to systematically examine a wider range of the SL positioning using uORF reporters which contain a GFP dORF serving as the internal control. Interestingly, I observed that the SLs in CDS close to the AUG exhibited similar inhibitory effects on translation efficiency as SLs in 5'UTR did.

It was unexpected as SLs in CDS were thought to have less inhibitory effects, as they were unwound more efficiently by the 80S ribosome during elongation, compared to those in 5'UTR. This finding was validated by the results from Fluc reporter assays using mRNA reporters either inserted with SLs or targeted by the complementary DNA oligos at different regions.

RNA helicases play a key role in promoting the ribosome scanning through 5'UTR during initiation. Among these helicases are DEAD/DEAH-Box helicases such as DHX29 and eIF4A1. The current body of knowledge concludes that DHX29 is required for the initiation of mRNAs with heavily structured 5' UTRs, while eIF4A1 exhibits limited capability of unwinding structures. Upon noticing the distinct translational impacts of SLs located in different 5'UTR regions, I intended to address the question whether SLs in particular regions are more dependent on RNA helicases, like DHX29 and eIF4A1, than those in other regions. Using both Fluc reporter assays and uORF reporter assays, I found that all the constructs displayed the uniform dependence on eIF4A1, regardless of the structure complexity. This is consistent with the notion that eIF4A1 is important for global translation while exhibits little mRNA preference. Different from eIF4A1, DHX29 selectively promoted translation of mRNA with structured 5'UTR, as evidenced by the observed dramatic reduction in the expression level of these mRNAs upon DHX29 depletion. The functional differences between DHX29 and eIF4A1 in

translation of mRNAs with structured 5' UTRs confirmed that these two RNA helicases do not have redundant functions.

However, an unexpected finding was that the absence of DHX29 impeded the translation of mRNAs which contain structures within the range from 5' end up to 20nt downstream of the start codon. Within this range, the position of structures does not affect their sensitivities to DHX29 depletion. A novel function of DHX29 was proposed to explain this phenomenon. The observed impaired translation of reporters with SLs in the CDS upon DHX29 depletion suggests that DHX29 may unwind secondary structures during elongation and it may be involved in the early elongation phase. To validate this hypothesis, I conducted plolysome fractionation and immunoblot DHX29 in the various fractions to show the distribution of DHX29. I found that DHX29 is present not only in 40S fractions but also in the monosome as well as light polysome fractions. These results challenge the current viewpoint that DHX29 dissociates from the ribosome after 48S formation. It may now be considered that DHX29 remains on the ribosome during the early elongation phase. While there are still future investigations to be done in order to confirm the mechanisms through which DHX29 promote the unwinding of stable stems in both the 5' UTR and in the early stages of elongation, these novel findings suggest that DHX29 may play a greater role in the regulation of translation than initially thought and highlights DHX29's importance to protein translation beyond translation initiation stage.

Hairpins located downstream of, and in proximity to, an initiation codon, were reported to increase utilization of the initiation codon by slowing 40S migration and allowing increased codon dwelling time[4,5]. Here, we found that SLs 15nt downstream of AUG significantly reduced translation. To characterize the best position for a downstream hairpin to facilitate start codon recognition, systemic studies of translational effects of SLs located within the proximal downstream region of the initiation site are planned for future endeavors. Also, both chapter 2 and chapter 3 revealed the wide scope of functions of RNA helicases eIF4A1 and DHX29. As many helicases have been implicated in development and cancer, future investigations of their non-overlapping roles in the scanning process will accelerate our understanding of alternative translation in cellular homeostasis and pathophysiology.

REFERENCE

1. Abaeva, I. S., Pestova, T. V. & Hellen, C. U. T. Attachment of ribosomal complexes and retrograde scanning during initiation on the Halastavi árva virus IRES. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw016
2. Skabkin, M., Skabkina, O., Hellen, C. T. & Pestova, T. Reinitiation and other unconventional posttermination events during eukaryotic translation. *Mol. Cell* **51**, 249–264 (2013).
3. Gunnery, S., Mäivali, Ü. & Mathews, M. B. Translation of an uncapped mRNA involves scanning. *J. Biol. Chem.* (1997). doi:10.1074/jbc.272.34.21642
4. Kozak, M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* (1986). doi:10.1073/pnas.83.9.2850
5. Kozak, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* (1990). doi:10.1073/pnas.87.21.8301

APPENDIX I

Start Codon-Associated Ribosomal Frameshifting

This section is work in process to further dissect the mechanism underlying the bi-directional movement of ribosome discussed in Chapter 2, while it was organized as part of a new manuscript in preparation**.**

AI.1 Results

Reading frame maintenance is a critical property of ribosomes during the translation elongation. Ribosome frameshifting during translocation could yield a variety of distinct polypeptides. A number of programed ribosomal frameshifting (PRF) events have been reported. The slippery sequence, a *cis-acting* element in the mRNA sequence, stimulates ribosome frameshifting into the -1 reading frame.[1] The aminoacyl-tRNA limitation was found to induce the alternative ribosome frameshifting pathway.[2] A recent study demonstrated that ribosome collision leads to +1 frameshifting independent of mRNA decay.[3] In this study, a global increased frameshift at the early stage of elongation was revealed by a newly developed high-resolution ribosome profiling approach.

To validate this finding, I first examined the translatability of Fluc ORF in the +1 or +2 reading frame. To factor out the variations of translation products derived from downstream AUGs, I included the reporter mRNAs with the first

start codons in the weak 'Kozak' sequence context (Figure AI-1A). In comparison to the in-frame ORFs ('Strong-0FM' and 'Weak-0FM'), ORFs in the +1 and +2 reading frame showed minimal translation efficiency, regardless of the sequence context (Figure AI-1B). However, a close inspection of the real-time luciferase expression revealed that the translation from the 'Strong-1FM' and 'Strong-2FM' was detectable, evidenced by the increased F0 value. In contrast, the expression of 'Weak-1FM' and 'Weak-2FM' was comparable to the basal level ('Uncapped'), indicating lack of translation from the downstream AUGs (Figure AI-1C).

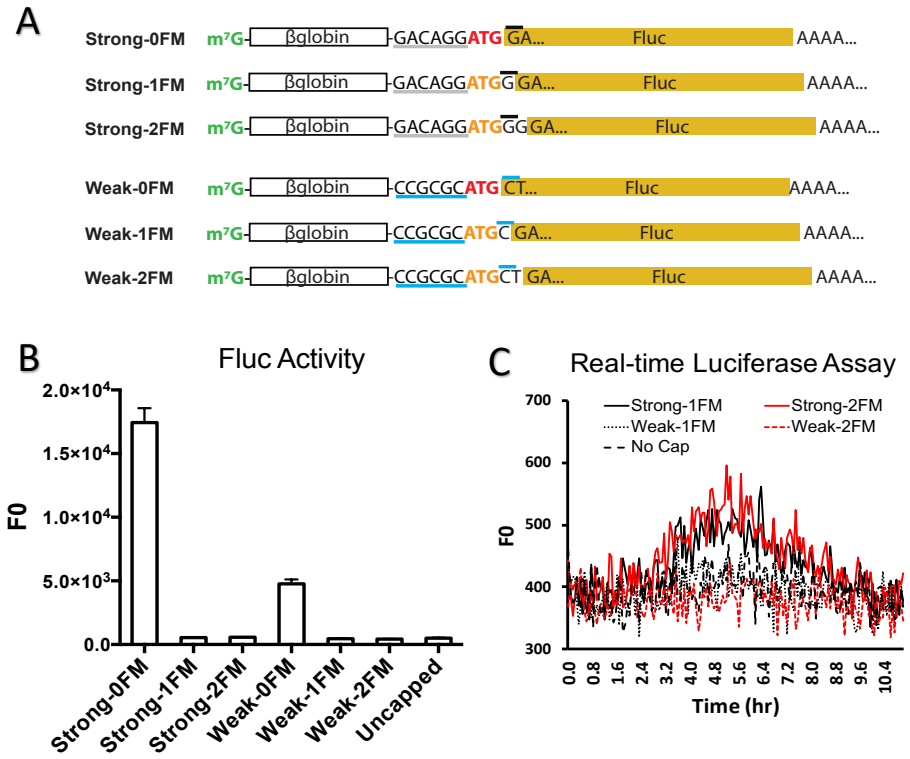

Figure AI-1 Translatability of out-of-frame ORFs.
(A) Sequence of Fluc reporter mRNAs. (B) Fluc activity of indicated mRNAs. In vitro synthesized mRNAs were transfected into MEF cells and Fluc activities were monitored by real-time luminometry. Uncapped mRNAs were included as negative control. Error bars: mean ± SEM; n = 3 biological replicates. (C) Representative real-time luminometry of MEF cells transfected with indicated Fluc mRNA reporters. Uncapped mRNA with 20-nt 5'UTR was included as a negative control.

One possibility of the translatability of these out-of-frame ORFs is the bi-directional movement of the ribosome around the start-codon-associated region. In Chapter 2, I reported the non-linear movement of ribosomes between two closely spaced AUGs by an observed trend of ascending recognition of the first AUG and descending recognition of the second. To substantiate this finding, I designed another set of Fluc reporters with switched reading frames of the neighboring AUG codons (Figure AI-2A). When the space between two AUGs were increased, the first AUG showed promoted translation efficiency (Figure AI-2B) but the second AUG showed the opposite trend (Figure AI-2C), revealed by Fluc reporter assay. I also conducted the Toe-printing assay in the RRL using reporter mRNAs with the first AUG out-of-frame while the second AUG in-frame (Figure AI-3). Despite out-of-frame, the first AUG always showed a strong and sharp peak. As expected, the presence of the downstream adjacent AUG significantly reduced the recognition of the first AUG ('+1FM-7nt' and '+2FM-8nt'), confirming the competitive relationship between closely spaced AUGs. As the dTIS moves further ('+1FM-13nt' and '+2FM-14nt'), the recognition potential of the first AUG was resumed. The back-and-forth oscillation of the ribosome around the start-codon-associated region was further validated using Fluc reporters with three closely spaced AUGs (Figure AI-4).

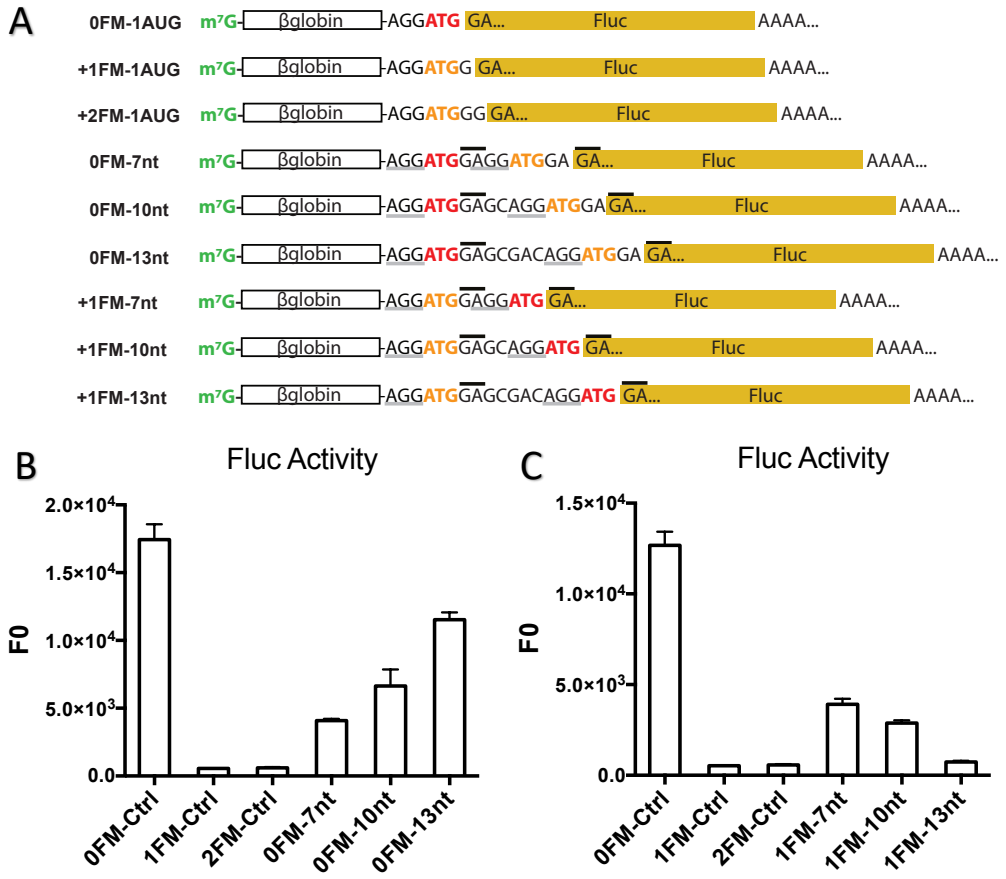

Figure AI-2 Bi-directional movement of ribosomes in start-codon-associated region.
(A) Sequence of Fluc reporter mRNAs. (B) Fluc activity of indicated mRNAs. In vitro synthesized mRNAs were transfected into MEF cells and Fluc activities were monitored by real-time luminometry. The mRNA reporter with in-frame ORF was included as positive control. The mRNA reporters with out-of-frame ORF were included as negative control. (C) Similar as (B). Error bars: mean ± SEM; n = 3 biological replicates.

Revealed by the sequencing data, the presence of uracil at the fifth position downstream of the AUG (+5) could induce the ribosomal frameshifting. To verify this finding, I modified the mRNA reporters used in the toe-printing assay by replacing the 'A' with 'U' at the +5 position of the first AUG (Figure AI-3A). Compared to the unmodified reporters, the modified reporters ('GT') exhibited an increase in the recognition of the second AUG (Figure AI-3B, right two panels), indicating an enhanced frameshifting behavior of the ribosome.

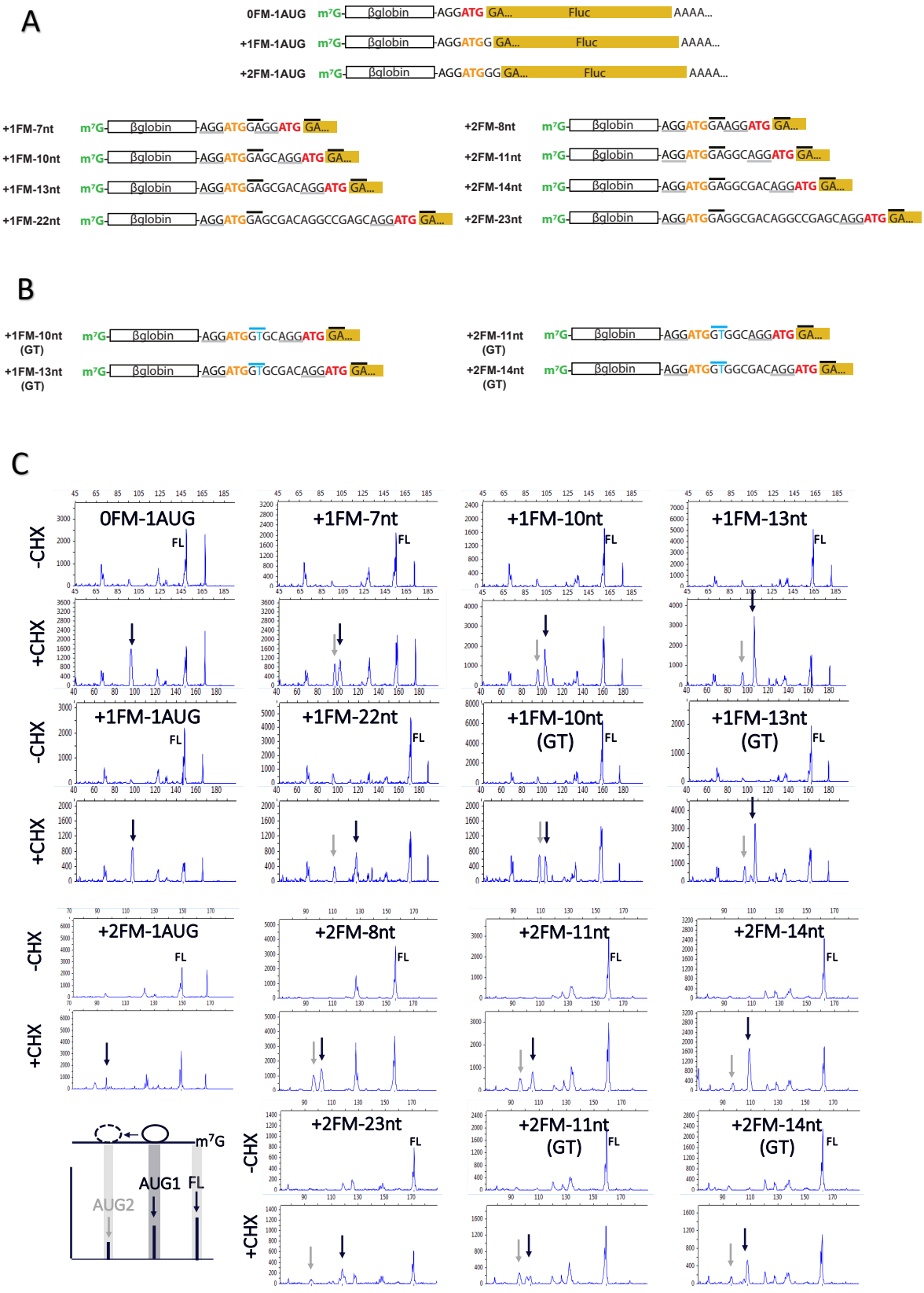
A
0FM-1AUG
+1FM-1AUG
+2FM-1AUG
+1FM-7nt
+1FM-10nt
+1FM-13nt
+1FM-22nt
+2FM-8nt
+2FM-11nt
+2FM-14nt
+2FM-23nt
B
+1FM-10nt (GT)
+1FM-13nt (GT)
+2FM-11nt (GT)
+2FM-14nt (GT)
C
-CHX
+CHX
0FM-1AUG
+1FM-7nt
+1FM-10nt
+1FM-13nt
+1FM-1AUG
+1FM-22nt
+1FM-10nt (GT)
+1FM-13nt (GT)
+2FM-1AUG
+2FM-8nt
+2FM-11nt
+2FM-14nt
+2FM-23nt
+2FM-11nt (GT)
+2FM-14nt (GT)
FL
m7G
AUG1
AUG2

Figure AI-3 Start codon recognition between two closely spaced AUGs.
(A) and (B) Sequence of Fluc reporter mRNAs. (C) Selection of closely spaced AUG codons by toe-printing assays. Two AUG codons with identical flanking sequence but different length of the spacer were inserted into Fluc mRNAs with β-globin 5'UTR. Expected positions corresponding to the ribosome leading edge are highlighted. The first AUG was marked as a black arrow; The second AUG was marked as a grey arrow. The relative peak heights correspond to the efficiency of start codon selection.

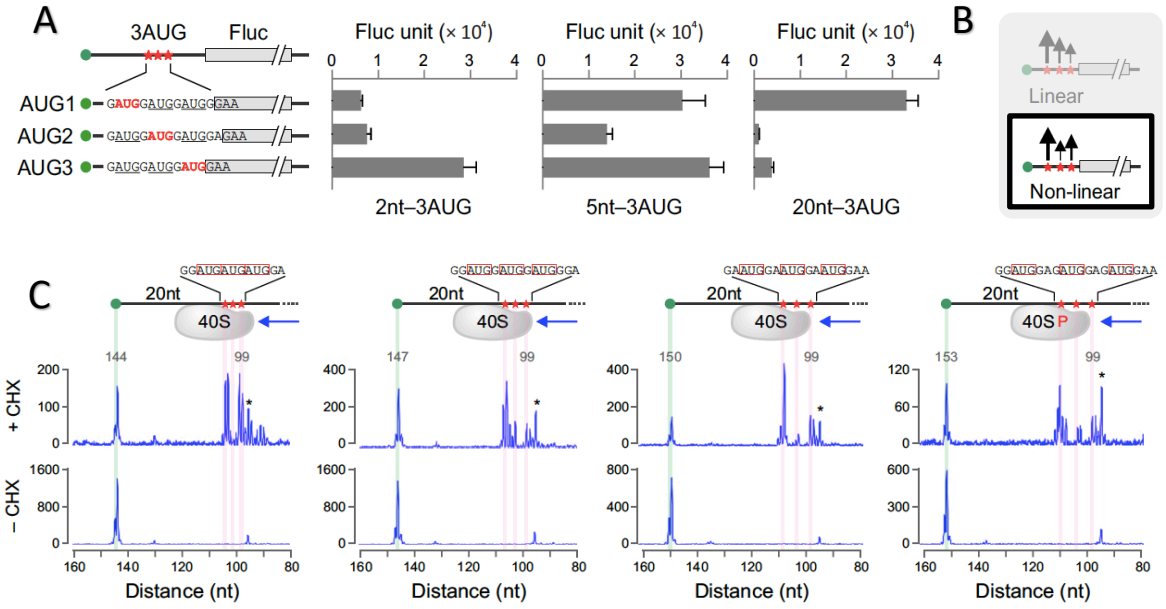

Figure AI-4 Start codon selection among three closely spaced AUGs.
(A) Fluc activity of mRNA reporters containing 3 AUG codons. (Left) Sequence of Fluc reporter mRNAs. (Right) Fluc activity of indicated mRNAs. In vitro synthesized mRNAs were transfected into MEF cells and Fluc activities were monitored by real-time luminometry. Error bars: mean ± SEM; n = 3 biological replicates. (B) Schematic of non-linear movement of ribosome among three AUGs. (C) Selection of closely spaced AUG codons by toe-printing assays. Three AUG codons with various space between neighboring AUGs were inserted into Fluc mRNAs with 20-nt 5'UTR. Expected positions corresponding to the ribosome leading edge are highlighted. The relative peak heights correspond to the efficiency of start codon selection.

In addition to the mRNA features, the sequencing results suggested that conditions of amino acid starvation associated with increased ribosomal frameshifting in the start-codon-associated region. To dissect the underlying mechanism, I conducted the polysome profiling in both MEF and HEK293 K^b cell-lines with either full amino acid or 2-hr amino acid starvation. I also included the eIF4A1 knockdown and DHX29 knockdown samples from both cell-lines to

test their potential functions in the ribosomal frameshifting. The polysome profiling data was shown in the Figure AI-5, and fractionation samples were subjected to the library conduction for the Ribo-seq.

Together all, I conclude that the ribosomal frameshifting around the start-codon-associated region might be due to the bi-directional movement of the ribosome which is stimulated by the sequence element or stress condition like amino acid starvation.

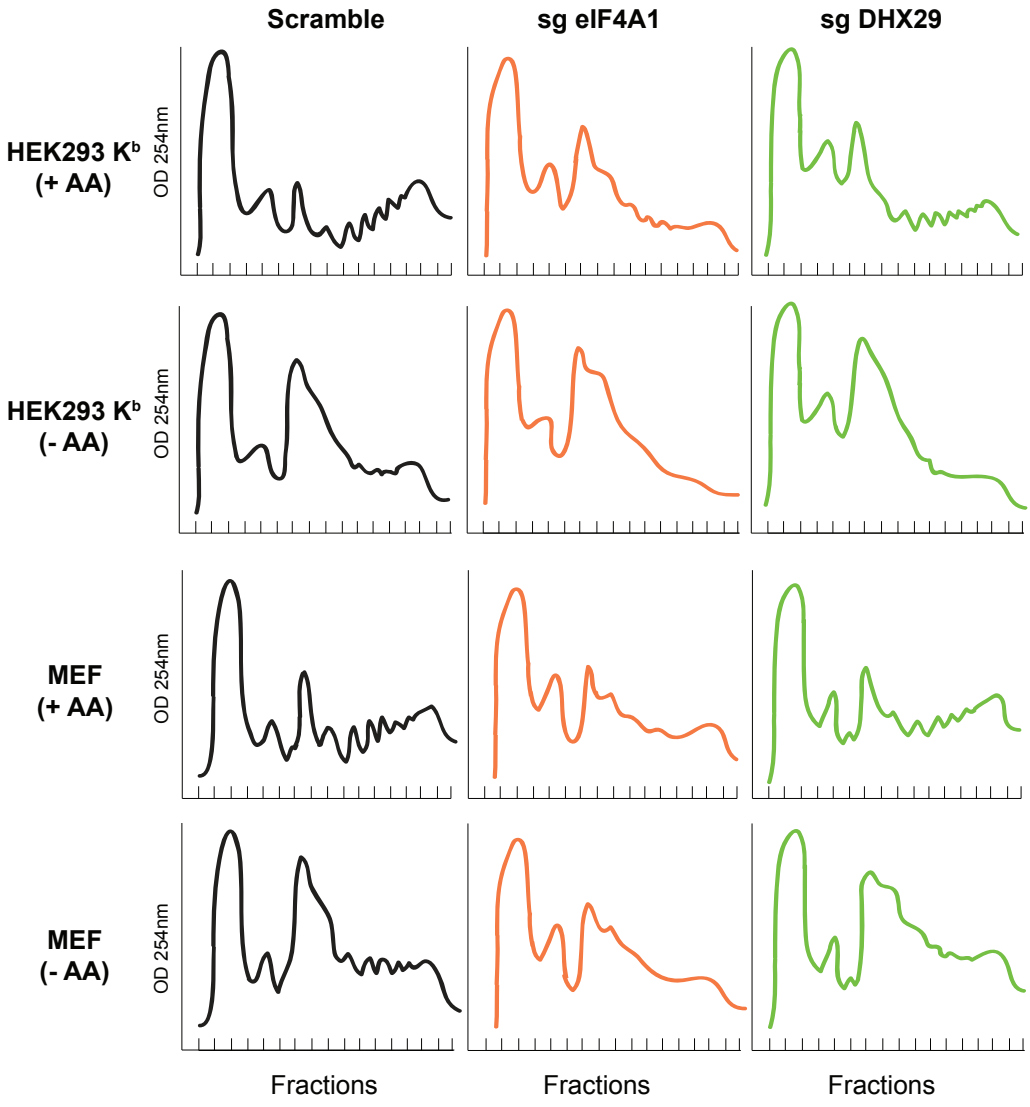

Figure AI-5 Impact of amino acid starvation on mRNA translation upon eIF4A1 or DHX29 depletion.
Polysome profiles of MEF and HEK293-K^b cells with either DHX29 knockdown or eIF4A1 knockdown under 2-hr full amino acid starvation (-AA) or control (+AA) conditions were analyzed by sucrose density gradient centrifugation in polysome buffer.

AI.2 Methods (in addition to section 2.5)

Starvation Treatment

Amino acid starvation treatment was carried out by incubating cells in HBSS buffer (Lonza) with 10% dialyzed FBS (Sigma Aldrich) for two hours at 37 °C. The control experiment was performed in parallel in DMEM with 10% FBS.

Table of primers

Table of primers

Number	Name	Sequence	Description
1	Strong-0FM (Same as 0FM-1AUG)	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGAA GACGCCAAAAAC	Forward primer for constructing Strong-0FM (Same as 0FM-1AUG) reporter
2	Strong-1FM (Same as +1FM-1AUG)	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGGA AGACGCCAAAAAC	Forward primer for constructing Strong-1FM (Same as +1FM-1AUG) reporter
3	Strong-2FM (Same as +2FM-1AUG)	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGGG AAGACGCCAAAAAC	Forward primer for constructing Strong-2FM (Same as +2FM-1AUG) reporter

4	Weak-0FM	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTCCGCGCATGCTCG AAGACGCCAAAAAC	Forward primer for constructing Weak-0FM reporter
5	Weak-1FM	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTCCGCGCATGCGAA GACGCCAAAAAC	Forward primer for constructing Weak-1FM reporter
6	Weak-2FM	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTCCGCGCATGCTGA AGACGCCAAAAAC	Forward primer for constructing Weak-2FM reporter
7	0FM-7nt	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGAG GATGGAGAAGACGCCAAA AAC	Forward primer for constructing 0FM-7nt reporter
8	0FM-10nt	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGAG CAGGATGGAGAAGACGCC AAAAAC	Forward primer for constructing 0FM-10nt reporter
9	0FM-13nt	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGAG CGACAGGATGGAGAAGAC GCCAAAAAC	Forward primer for constructing 0FM-13nt reporter
10	+1FM-7nt	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGAG GATGGAAGACGCCAAAAA C	Forward primer for constructing +1FM-7nt reporter
11	+1FM-10nt	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGAG CAGGATGGAAGACGCCAA AAAC	Forward primer for constructing +1FM-10nt reporter

12	+1FM-13nt	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGAG CGACAGGATGGAAGACGC CAAAAAC	Forward primer for constructing +1FM-13nt reporter
13	+1FM-22nt	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGAG CGACAGGCCGAGCAGGAT GGAAGACGCCAAAAAC	Forward primer for constructing +1FM-22nt reporter
14	+1FM-10nt (GT)	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGTG CAGGATGGAAGACGCCAA AAAC	Forward primer for constructing +1FM-10nt (GT)reporter
15	+1FM-13nt (GT)	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGTG CGACAGGATGGAAGACGC CAAAAAC	Forward primer for constructing +1FM-13nt (GT) reporter
16	+2FM-8nt	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGAA GGATGGAAGACGCCAAAA AC	Forward primer for constructing +2FM-8nt reporter
17	+2FM-11nt	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGAG GCAGGATGGAAGACGCCA AAAAC	Forward primer for constructing +2FM-11nt reporter
18	+2FM-14nt	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGAG GCGACAGGATGGAAGACG CCAAAAAC	Forward primer for constructing +2FM-14nt reporter
19	+2FM-23nt	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGAG GCGACAGGCCGAGCAGGA TGGAAGACGCCAAAAAC	Forward primer for constructing +2FM-23nt reporter

20	+2FM-11nt (GT)	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGTG GCAGGATGGAAGACGCCA AAAAC	Forward primer for constructing +2FM-11nt (GT) reporter
21	+2FM-14nt (GT)	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGGATGGTG GCGACAGGATGGAAGACG CCAAAAAC	Forward primer for constructing +2FM-14nt (GT) reporter
22	2nt-3AUG1	TAATACGACTCACTATA GGATGGATGGATGGGAAG ACGCCAAAAAC	Forward primer for constructing 2nt-3AUG1 reporter
23	2nt-3AUG2	TAATACGACTCACTATA GGATGGATGGATGGAGAA GACGCCAAAAAC	Forward primer for constructing 2nt-3AUG2 reporter
24	2nt-3AUG3	TAATACGACTCACTATA GGATGGATGGATGGAAGA CGCCAAAAAC	Forward primer for constructing 2nt-3AUG3 reporter
25	5nt-3AUG1	TAATACGACTCACTATA GGAGGATGGATGGATGGG AAGACGCCAAAAAC	Forward primer for constructing 5nt-3AUG1 reporter
26	5nt-3AUG2	TAATACGACTCACTATA GGAGGATGGATGGATGGA GAAGACGCCAAAAAC	Forward primer for constructing 5nt-3AUG2 reporter
27	5nt-3AUG3	TAATACGACTCACTATA GGAGGATGGATGGATGGA AGACGCCAAAAAC	Forward primer for constructing 5nt-3AUG3 reporter

28	20nt-3AUG1	TAATACGACTCACTATA GGACACTTGCTTTTGACAG GATGGATGGATGGGAAGA CGCCAAAAAC	Forward primer for constructing 20nt-3AUG1 reporter
29	20nt-3AUG2	TAATACGACTCACTATA GGACACTTGCTTTTGACAG GATGGATGGATGGAGAAG ACGCCAAAAAC	Forward primer for constructing 20nt-3AUG2 reporter
30	20nt-3AUG3(Same as 20nt-3AUG-1sp)	TAATACGACTCACTATA GGACACTTGCTTTTGACAG GATGGATGGATGGAAGAC GCCAAAAAC	Forward primer for constructing 20nt-3AUG3 and 20nt-3AUG-1sp reporter
31	20nt-3AUG-0sp	TAATACGACTCACTATA GGACACTTGCTTTTGACAG GATGGATGGATGGAAGAC GCCAAAAAC	Forward primer for constructing 20nt-3AUG-0sp reporter
32	20nt-3AUG-2sp	TAATACGACTCACTATA GGACACTTGCTTTTGACAG AATGGAATGGAATGGAAG ACGCCAAAAAC	Forward primer for constructing 20nt-3AUG-2sp reporter
33	20nt-3AUG-3sp	TAATACGACTCACTATA GGACACTTGCTTTTGACAG GATGGAGATGGAGATGGA AGACGCCAAAAAC	Forward primer for constructing reporter
34	Fluc_R	TTACACGGCGATCTTTCCG CCCTTC	Reverse primer for constructing all Fluc mRNA reporters

REFERENCE

1. Blinkowa AL, Walker JR. Programmed ribosomal frameshifting generates the Escherichia coli DNA polymerase III γ subunit from within the τ subunit reading frame. *Nucleic Acids Res*. 1990. doi:10.1093/nar/18.7.1725
2. Caliskan N, Wohlgemuth I, Korniy N, Pearson M, Peske F, Rodnina M V. Conditional Switch between Frameshifting Regimes upon Translation of dnaX mRNA. *Mol Cell*. 2017. doi:10.1016/j.molcel.2017.04.023
3. Simms CL, Yan LL, Qiu JK, Zaher HS. Ribosome Collisions Result in +1 Frameshifting in the Absence of No-Go Decay. *Cell Rep*. 2019. doi:10.1016/j.celrep.2019.07.046

APPENDIX II

The Effects of Secondary Structures in Different Regions of mRNAs

This work is in process to further investigate the effects of secondary structures in different regions of mRNA on the basis of Chapter 3**.**

AII.1 Results

AII.1.1 ASOs targeting different regions equally affects mRNA stability.

In Chapter 3, I examined the impact of the location of secondary structures on mRNA translation. As one of the supporting evidences, DNA oligos binding to various regions of the Fluc reporter mRNAs showed different effects on Fluc expression. However, the introduction of antisense oligonucleotides (ASO) could induce the degradation of the targeted mRNA[1]. To rule out the potential impact of DNA:RNA hybridization on mRNA stability, I measured the level of reporter mRNAs after transection. Using RT-qPCR, I observed a uniform reduction (20%) in the mRNA level when the reporter mRNAs were targeted by 15-nt oligos, regardless of the binding region. And the fold changes were more dramatic (50%) but still similar when the targeted regions were extended to 31 nt (Figure AII-1B). This observation further validated that the observed region-specific effects of the secondary structures were primarily at the level of translation.

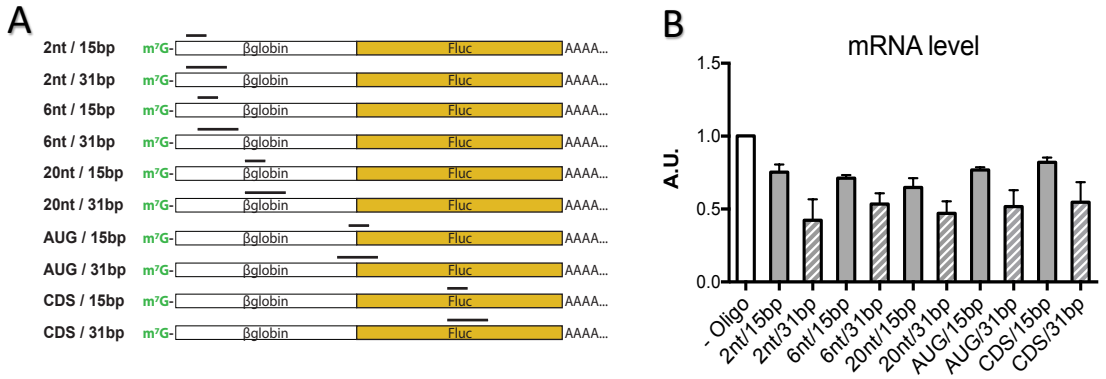

Figure AII-1 ASOs targeting different regions equally affects mRNA stability.
(A) Different regions in the Fluc reporter mRNAs targeted by DNA oligos. (B) Relative mRNA level of reporter Fluc reporters. Total RNA was extracted 6hr after mRNA transfection. The mRNA level was quantified by RT-qPCR, followed by normalization to the endogenous control β-Actin. Error bars: mean ± SEM; n = 3 biological replicates.

AII.1.2 ASOs with overhangs impaired translation independent of eIF4A1.

In addition to the DNA oligos fully complementary to the specific regions in the mRNA, oligos with either 5' overhangs or 3' overhangs were tested as well (Figure AII-2A). Unexpectedly, both oligos with 5' overhang and 3' overhang exhibited a more significant inhibitory effect on the Fluc expression, compared to the oligo with no overhangs (Figure AII-2B). This phenomenon is mainly at the translational level as no difference in the fold change of mRNA level was observed (Figure AII-2C).

A recent study reported that eIF4A1 unwinds RNA duplexes with long overhangs less efficiently. [2] In the light of this finding, I speculated that eIF4A1 could facilitate the translation of mRNAs targeted by DNA oligos with no overhangs but not those with overhangs. To test my hypothesis, I transfected the DNA:RNA hybrids into both Scramble and eIF4A1 knockdown MEF cell-

lines (Figure AII-2D). However, a uniform fold change of the translational efficiency was observed from the mRNA reporters targeted by different oligos (Figure AII-2E). This data suggested that the inhibitory effect of oligos with 5' or 3' overhangs is independent of eIF4A1.

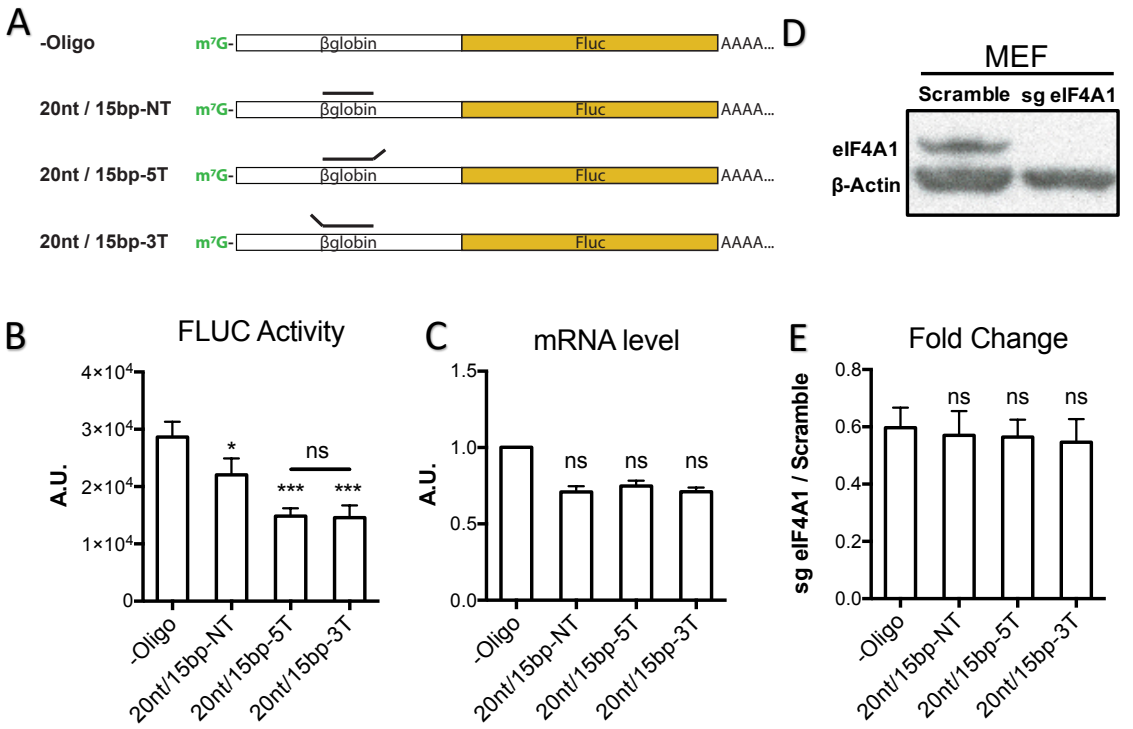

Figure AII-2 ASOs with overhangs impaired translation independent of eIF4A1.
(A) Schematic of Fluc reporter mRNAs targeted by different DNA oligos. (B) Fluc activity of indicated mRNAs. In vitro synthesized mRNAs were transfected into MEF cells and Fluc activities were monitored by real-time luminometry. The mRNA reporter with no oligo binding was included as positive control. (C) Relative mRNA level of reporter Fluc reporters. Total RNA was extracted 6hr after mRNA transfection. The mRNA level was quantified by RT-qPCR, followed by normalization to the endogenous control β-Actin. (D) Immunoblotting of MEF cells with or without eIF4A1 knockdown. (E) Fluc activities in the absence of eIF4A1 were normalized to the control. Error bars: mean ± SEM; n = 3 biological replicates. One-way ANOVA for multiple comparisons and unpaired Student's t-test for individual comparisons, ns, not significant; * $p < 0.05$; ***$p<0.001$.

AII.1.3 SLs downstream proximal to the start codon promote the utilization of the upstream initiation codon.

In Chapter 3, I reported that the secondary structures in the CDS up to 20nt downstream of the start codon exhibited the high dependence of DHX29. It is likely that the AUG-proximal (e,g,. 2nt downstream of AUG) SLs are unwound during the unfinished initiation before the decoding site of PIC reaches the start codon, while the SL 20nt downstream of the AUG ought to be unwound by the 80S ribosome during the early elongation phase. An early study showed that hairpins located downstream of, and in proximity to, an initiation codon, increase utilization of the initiation codon by slowing 40S migration and allowing increased codon dwelling time[34]. To test this proposed mechanism, I conducted the Fluc reporter assays and the Toe-printing assays using Fluc reporters with or without a SL 2nt downstream of the start codon. Interestingly, while the translational efficiency was reduced to half in the presence of the SL (Figure AII-3B), the start codon recognition was slightly increased (Figure AII-3A). This evidence suggested that SLs immediately downstream of the start codon indeed increases the utilization of the upstream initiation codon. However, slowing down 40S migration around the start codon might reduce the initial translation rate of 80S and impair the overall translational efficiency.

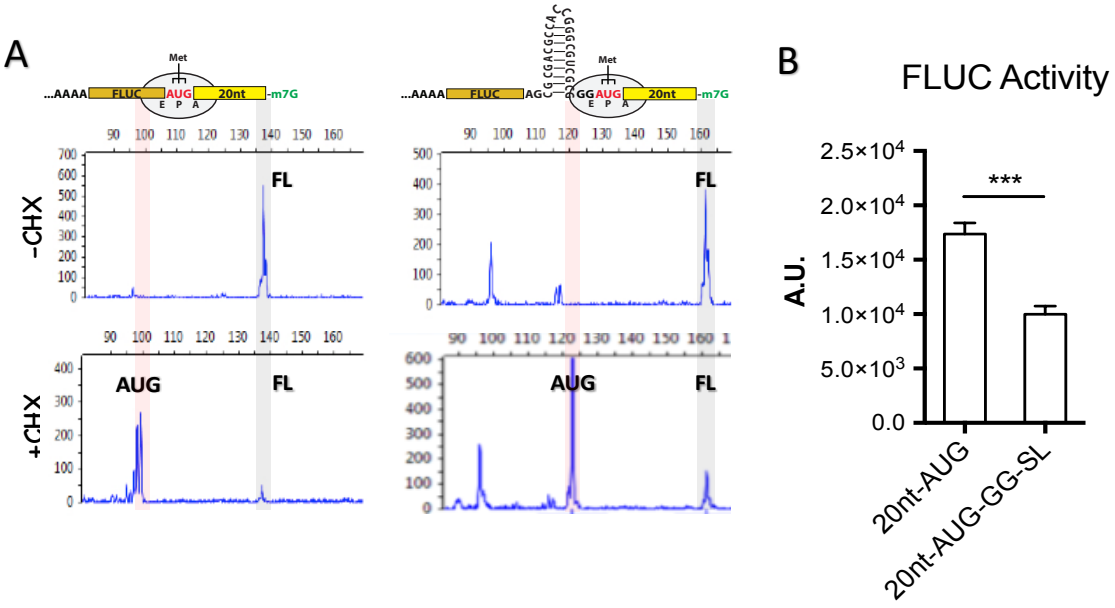

Figure AII-3 SLs downstream proximal to the start codon promote the utilization of the upstream initiation codon.
(A) Selection of AUG immediately upstream of the SL by toe-printing assays. SLs were inserted 2nt downstream of the start codon of the Fluc mRNAs. Expected positions corresponding to the ribosome leading edge are highlighted. The relative peak heights correspond to the efficiency of start codon selection. (B) Fluc activity of mRNA reporters containing a SL 2nt downstream of the start codon. Fluc activity of indicated mRNAs. In vitro synthesized mRNAs were transfected into MEF cells and Fluc activities were monitored by real-time luminometry. The mRNA reporter with no SLs was included as positive control. Error bars: mean ± SEM; n = 3 biological replicates. Unpaired Student's t-test for individual comparison, ***p<0.001.

AII.2 Methods (in addition to section 3.5)

Table of primers

Table of primers			
Number	**Name**	**Sequence**	**Description**
1	T7-βglobin-Fluc_F	TAATACGACTCACTATA GGGAGAACATTTGCTTCTGAC ACAACTGTGTTCACTAGCAAC CTCAAACAGACACC **ATG** GAAGACGCCAAAAAC	Forward primer for constructing βglobin-Fluc mRNA reporter
2	2nt/15bp	GAAGCAAATGTTCTC	15nt of DNA oligo complementary to 2nt downstream region of 5'end
3	2nt/31bp	GAACACAGTTGTGTCAGAAGC AAATGTTCTC	31nt of DNA oligo complementary to 2nt downstream region of 5'end
4	6nt/15bp	GTCAGAAGCAAATGT	15nt of DNA oligo complementary to 6nt downstream region of 5'end
5	6nt/31bp	TAGTGAACACAGTTGTGTCAG AAGCAAATGT	31nt of DNA oligo complementary to 6nt downstream region of 5'end
6	20nt/15bp (Same as 20nt/15bp-NT)	GTGAACACAGTTGTG	15nt of DNA oligo complementary to 20nt downstream region of 5'end
7	20nt/31bp	CTGTTTGAGGTTGCTAGTGAA CACAGTTGTG	31nt of DNA oligo complementary to 20nt downstream region of 5'end

8	AUG/15bp	GTCTTCCATGGTGTC	15nt of DNA oligo complementary to AUG-associated region
9	AUG/31bp	TTTTTGGCGTCTTCCATGGTGTCTGTTTGAG	31nt of DNA oligo complementary to AUG-associated region
10	CDS/15bp	GGCGCCGGGCCTTTC	15nt of DNA oligo complementary to CDS
11	CDS/31bp	TTCCAGCGGATAGAATGGCGCCGGGCCTTTC	31nt of DNA oligo complementary to CDS
12	20nt/15bp-5T	GTGTGTGTGGTGAACACAGTTGTG	15nt of DNA oligo complementary to 20nt downstream region of 5'end with 5' overhang
13	20nt/15bp-3T	GTGAACACAGTTGTGGTGTGTGTG	15nt of DNA oligo complementary to 20nt downstream region of 5'end with 3' overhang
14	T7-20nt-Fluc_F	TAATACGACTCACTATA GGACACTTGCTTTTGACAGG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 20nt-Fluc mRNA reporter
15	T7-20nt-AUG-GG-SL-Luc_F1	**ATG** GGGCGCTGCGGGCCACCGCAG CGC GAA GACGCCAAAAACATAAAGAA AGG	First forward primer for constructing T7-20nt-AUG-GG-SL-Luc reporter

16	T7-20nt-AUG-GG-SL-Luc_F2	TAATACGACTCACTATA GGACACTTGCTTTTGAC AGG **ATG** GGGCGCTGCGGGCCACC	Second forward primer for constructing T7-20nt-AUG-GG-SL-Luc reporter
17	Fluc_R	TTACACGGCGATCTTTCCGCC CTTC	Reverse primer for constructing all Fluc mRNA reporters

REFERENCE

1. Liang XH, Nichols JG, Hsu CW, Vickers TA, Crooke ST. MRNA levels can be reduced by antisense oligonucleotides via no-go decay pathway. *Nucleic Acids Res*. 2019. doi:10.1093/nar/gkz500
2. Andreou AZ, Harms U, Klostermeier D. Single-stranded regions modulate conformational dynamics and ATPase activity of eIF4A to optimize 5′-UTR unwinding. *Nucleic Acids Res*. 2019. doi:10.1093/nar/gkz254
3. Kozak M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc Natl Acad Sci U S A*. 1986. doi:10.1073/pnas.83.9.2850
4. Kozak M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci U S A*. 1990. doi:10.1073/pnas.87.21.8301

APPENDIX III Supplementary Table for Chapter 2

Table of primers and sgRNA targeting sequences			
Number	**Name**	**Sequence**	**Description**
1	T7-2nt-Fluc_F	TAATACGACTCACTATA GG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 2nt-Fluc mRNA reporter
2	T7-3nt-Fluc_F	TAATACGACTCACTATA GGG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 3nt-Fluc mRNA reporter
3	T7-4nt-Fluc_F	TAATACGACTCACTATA GAGG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 4nt-Fluc mRNA reporter
4	T7-5nt-Fluc_F	TAATACGACTCACTATA GGAGG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 5nt-Fluc mRNA reporter
5	T7-6nt-Fluc_F	TAATACGACTCACTATA GGGAGG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 6nt-Fluc mRNA reporter
6	T7-8nt-Fluc_F	TAATACGACTCACTATA GGGAGAGG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 8nt-Fluc mRNA reporter
7	T7-10nt-Fluc_F	TAATACGACTCACTATA GGGGAGGAGG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 10nt-Fluc mRNA reporter

8	T7-20nt-Fluc_F	TAATACGACTCACTATA GGACACTTGCTTTTGACA GG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 20nt-Fluc mRNA reporter
9	T7-βglobin-Fluc_F	TAATACGACTCACTATA GGGAGAACATTTGCTTCT GACACAACTGTGTTCACT AGCAACCTCAAACAGAC ACC **ATG** GAAGACGCCAAAAAC	Forward primer for constructing βglobin-Fluc mRNA reporter
10	Fluc_R	TTACACGGCGATCTTTCC GCCCTTC	Reverse primer for constructing all Fluc mRNA reporters
11	SIINFEKL-EGFP_F1	TCGATCATCAACTTCGAA AAGCTATAG GACACTAGCTTAGGAGG ATCCGGAG	First round forward primer for constructing all SIINFEKL mRNA reporters, SIINFEKL-EGFP/pcDNA3.1, BamHI marked in Magenta
12	T7-2nt-SIINFEKL_F2	TAATACGACTCACTATA GG **ATG** GAATCGATCATCAACTTC G	Second round forward primer for constructing 2nt-SIINFEKL mRNA reporter
13	T7-5nt-SIINFEKL_F2	TAATACGACTCACTATA GG AGG **ATG** GAATCGATCATCAACTTC G	Second round forward primer for constructing 5nt-SIINFEKL mRNA reporter
14	T7-20nt-SIINFEKL_F2	TAATACGACTCACTATA GGACACTTGCTTTTGACA GG **ATG** GAATCGATCATCAAC	Second round forward primer for constructing 20nt-SIINFEKL mRNA reporter; Forward primer for constructing mRNA reporters of SIINFEKL-6nt-IRES-EGFP, SIINFEKL-18nt-IRES-EGFP, SIINFEKL-6nt-Anti-IRES-EGFP and SIINFEKL-18nt-Anti-IRES-EGFP
15	EGFP_R	ACTAGAAGGCACAGTCG AGGCTGAT	Reverse primer for constructing all SIINFEKL mRNA reporters and SIINFEKL-18nt-IRES-EGFP/pcDNA3.1

16	SacI-TSS-2nt-SIINFEKL-_F2	GGGAGCTCTCTGGCTAA GG **ATG** GAATCGATCATCAACTTC GAAAAGCTATAG	Second round forward primer for constructing 2nt-SIINFEKL-EGFP/pcDNA3.1, SacI marked in Turquoise
17	SacI-TSS-4nt-SIINFEKL-_F2	GGGAGCTCTCTGGCTAA GAGG **ATG** GAATCGATCATCAACTTC GAAAAGCTATAG	Second round forward primer for constructing 4nt-SIINFEKL-EGFP/pcDNA3.1, SacI marked in Turquoise
18	SacI-TSS-10nt-SIINFEKL-_F2	GGGAGCTCTCTGGCTAA GGGGAGGAGG **ATG** GAATCGATCATCAACTTC GAAAAGCTATAG	Second round forward primer for constructing 10nt-SIINFEKL-EGFP/pcDNA3.1, SacI marked in Turquoise
19	SacI-TSS-20nt-SIINFEKL-_F2	GGGAGCTCTCTGGCTAA GGACACTTGCTTTTGACA GG **ATG** GAATCGATCATCAACTTC GAAAAGCTATAG	Second round forward primer for constructing 20nt-SIINFEKL-EGFP/pcDNA3.1, SacI marked in Turquoise
20	T7-30nt-1AUG-Fluc_F	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 30nt-Fluc mRNA reporter
21	T7-30nt-2AUG(4nt)-Fluc_F	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGG **ATG** GAGG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 30nt-2AUG(4nt)-Fluc mRNA reporter
22	T7-30nt-2AUG(7nt)-Fluc_F	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGG **ATG** GAGCAGG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 30nt-2AUG(7nt)-Fluc mRNA reporter
23	T7-30nt-2AUG(10nt)-Fluc_F	TAATACGACTCACTATA GGGAGAACTTGGACACTT GCTTTTGACAGG **ATG** GAGCGACAGG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 30nt-2AUG(10nt)-Fluc mRNA reporter

24	NheI-20nt-SIINFEKL-PmlI_F	GCCGCTAGCGGACACTTGCTTTTGACAGG	Forward primer for amplification of SIINFEKL sequence and constructing SIINFEKL-6nt-IRES-EGFP/pcDNA3.1, NheI marked in Tangerine
25	NheI-20nt-SIINFEKL-PmlI_R	GCCCACGTGCTATAGCTTTTCGAAGTTGATGATCGACTC	Reverse primer for amplification of SIINFEKL sequence and constructing SIINFEKL-6nt-IRES-EGFP/pcDNA3.1, PamlI marked in Cantaloupe
26	BamHI-GFP-22nt-XhoI_F	GCCGGATCCATGGTGAGCAAGGGCGAGGAGCTGTTC	Forward primer for amplification of EGFP sequence and constructing SIINFEKL-6nt-IRES-EGFP/pcDNA3.1, BamHI marked in Magenta
27	BamHI-GFP-22nt-XhoI_R	GCCCTCGAGGCTAAGAATTTCGTCATCGCTGCTACTTGTACAGCTCGTCCATGC	Reverse primer for amplification of EGFP sequence and constructing SIINFEKL-6nt-IRES-EGFP/pcDNA3.1, XhoI marked in Green
28	PmlI-12nt-IRES_F	GAATTCCACGTGGACACTAGCTTAGCGGCTAGTACTCCGGTATTGCG	Forward primer for amplification of IRES sequence and constructing SIINFEKL-18nt-IRES-EGFP/pcDNA3.1, PamlI marked in Cantaloupe
29	NheI-SIINFEKL-6nt-Anti-IRES-BamHI_F	GCCGCTAGCGGACACTTGCTTTTGACAGGATGGAATCGATCATCAACTTCGAAAAGCTATAGCACGTGCTATAGAATTGT	Forward primer for amplification of anti-IRES sequence and constructing SIINFEKL-6nt-Anti-IRES-EGFP/pcDNA3.1, NheI marked in Tangerine
30	NheI-SIINFEKL-18nt-Anti-IRES-BamHI_F	GCCGCTAGCGGACACTTGCTTTTGACAGGATGGAATCGATCATCAACTTCGAAAAGCTATAGCACGTGGACACTAGC	Forward primer for amplification of anti-IRES sequence and constructing SIINFEKL-18nt-Anti-IRES-EGFP/pcDNA3.1, NheI marked in Tangerine
31	NheI-SIINFEKL-Anti-IRES-BamHI_R	GCCGGATCCGGGCCCTCTAGATGCGGCCGCCAGTGTG	Reverse primer for amplification of anti-IRES sequence and constructing SIINFEKL-18nt-Anti-IRES-EGFP/pcDNA3.1 and SIINFEKL-18nt-Anti-IRES-EGFP/pcDNA3.1, BamHI marked in Magenta

32	T7-2nt-3AUG(1st)-Fluc_F	TAATACGACTCACTATAGG **ATG**G**ATG**G**ATG**GGAAGACGCCAAAAAC	Forward primer for constructing AUG1-Fluc mRNA reporter
33	T7-2nt-3AUG(2nd)-Fluc_F	TAATACGACTCACTATAGG **ATG**G**ATG**G**ATG**GAGAAGACGCCAAAAAC	Forward primer for constructing AUG2-Fluc mRNA reporter
34	T7-2nt-3AUG(3rd)-Fluc_F	TAATACGACTCACTATAGG **ATG**G**ATG**G**ATG**GAAGACGCCAAAAAC	Forward primer for constructing AUG3-Fluc mRNA reporter
35	T7-2nt-3AUG(1st)-SIINFEKL_F2	TAATACGACTCACTATAGG **ATG**G**ATG**G**ATG**GGAATCGATCATCAACTTCG	Second forward primer for constructing AUG1-SIINFEKL mRNA reporter
36	T7-2nt-3AUG(2nd)-SIINFEKL_F2	TAATACGACTCACTATAGG **ATG**G**ATG**G**ATG**GAGAATCGATCATCAACTTCG	Second forward primer for constructing AUG2-SIINFEKL mRNA reporter
37	T7-2nt-3AUG(3rd)-SIINFEKL_F2	TAATACGACTCACTATAGG **ATG**G**ATG**G**ATG**GAATCGATCATCAACTTCG	Second forward primer for constructing AUG3-SIINFEKL mRNA reporter
38	sg eIF4A1	GGAGCCGGAAGGCGTCATC	sgRNA targeting sequences for eIF4A1

APPENDIX IV Supplementary Table for Chapter 3

Table of primers and sgRNA targeting sequences			
Number	**Name**	**Sequence**	**Description**
1	T7-Xnt-SL-20nt-Luc_F1	GGACACTTGCTTTTGAC AGG **ATG** GAA GACGCCAAAAACATAAA GAAAGG	First round forward primer for constructing 2nt-SL-20nt-AUG and 6nt-SL-20nt-AUG Fluc mRNA reporters
2	T7-GG-SL-20nt-Luc_F2	TAATACGACTCACTATA GG GCGCTGCGGGCCACCGC AGCGC GGACACTTGC	Second round forward primer for constructing 2nt-SL-20nt-AUG Fluc mRNA reporter
3	T7-6nt-SL-20nt-Luc_F2	TAATACGACTCACTATA GGG AGG GCGCTGCGGGCCACCGC AGCGC GGACACTTGC	Second round forward primer for constructing 6nt-SL-20nt-AUG Fluc mRNA reporter
4	T7-Xnt-SL-20nt-AUG-Luc_F1	GCGCTGCGGGCCACCGC AGCGC GGACACTTGCTTTTGAC AGG **ATG** GAAGACGCCAAAAAC	First round forward primer for constructing 20nt-SL-20nt-AUG and 56nt-SL-20nt-AUG Fluc mRNA reporters
5	T7-20nt-SL-Xnt-AUG-Luc_F2	TAATACGACTCACTATA GGACACTTGCTTTTGAC AGG GCGCTGCGGGCCACC	Second round forward primer for constructing 20nt-SL-2nt-AUG and 20nt-SL-20nt-AUG Fluc mRNA reporters
6	T7-GG-SL-Xnt-Luc_F1	GG **ATG** GAAGACGCCAAAAACAT AAAGAAAGG	First round forward primer for constructing 2nt-SL-Xnt-AUG Fluc mRNA reporters
7	T7-GG-SL-2nt-Luc_F2	TAATACGACTCACTATA GG GCGCTGCGGGCCACCGC AGCGC GG **ATG** GAAGAC	Second round forward primer for constructing 2nt-SL-2nt-AUG Fluc mRNA reporter

8	T7-GG-SL-3nt-Luc_F2	TAATACGACTCACTATA GG GCGCTGCGGGCCACCGC AGCGC GGG **ATG** GAA GAC	Second round forward primer for constructing 2nt-SL-3nt-AUG Fluc mRNA reporter
9	T7-GG-SL-4nt-Luc_F2	TAATACGACTCACTATA GG GCGCTGCGGGCCACCGC AGCGC G GGG **ATG** GAA GAC	Second round forward primer for constructing 2nt-SL-4nt-AUG Fluc mRNA reporter
10	T7-GG-SL-5nt-Luc_F2	TAATACGACTCACTATA GG GCGCTGCGGGCCACCGC AGCGC GG AGG **ATG** GAA GAC	Second round forward primer for constructing 2nt-SL-5nt-AUG Fluc mRNA reporter
11	T7-GG-SL-6nt-Luc_F2	TAATACGACTCACTATA GG GCGCTGCGGGCCACCGC AGCGC GGG AGG **ATG** GAA GAC	Second round forward primer for constructing 2nt-SL-6nt-AUG Fluc mRNA reporter
14	T7-GG-SL-8nt-Luc_F2	TAATACGACTCACTATA GG GCGCTGCGGGCCACCGC AGCGC GGG AGA GG **ATG** GAA GAC	Second round forward primer for constructing 2nt-SL-8nt-AUG Fluc mRNA reporter
15	T7-GG-SL-10nt-Luc_F2	TAATACGACTCACTATA GG GCGCTGCGGGCCACCGC AGCGC GGGG AGG AGG **ATG** GAA GAC	Second round forward primer for constructing 2nt-SL-10nt-AUG Fluc mRNA reporter
16	T7-Xnt-SL-GG-AUG-Luc_F1	GAC AGG GCGCTGCGGGCCACCGC AGCGC GG **ATG** GAAGACGCCAAAAACAT AAAGAAAGG	First round forward primer for constructing 20nt-SL-2nt-AUG and 56nt-SL-2nt-AUG Fluc mRNA reporters
17	T7-56nt-SL-Xnt-Luc_F2	TAATACGACTCACTATA GGGAGAACATTTGCTTC TGACACAACTGTGTTCA CTAGCAACCTCAAACAG ACACC GCGCTGCGGGCCACC	Second round forward primer for constructing 56nt-SL-2nt-AUG and 56nt-SL-20nt-AUG Fluc mRNA reporters

18	T7-20nt-Fluc_F	TAATACGACTCACTATA GGACACTTGCTTTTGAC AGG **ATG** GAAGACGCCAAAAAC	Forward primer for constructing 20nt-Fluc mRNA reporter
19	T7-20nt-AUG-GG-SL-Luc_F1	**ATG** GG GCGCTGCGGGCCACCGC AGCGC GAA GACGCCAAAAACATAAA GAAAGG	First round forward primer for constructing 20nt-AUG-2nt-SL and 20nt-AUG-20nt-SL Fluc mRNA reporters
20	T7-20nt-AUG-GG-SL-Luc_F2	TAATACGACTCACTATA GGACACTTGCTTTTGAC AGG **ATG** GG GCGCTGCGGGCCACC	Second round forward primer for constructing 20nt-AUG-2nt-SL mRNA reporter
21	T7-20nt-AUG-20nt-SL-Luc_F2	TAATACGACTCACTATA GGACACTTGCTTTTGAC AGG **ATG** GAAGACGCCAAAAACAT AGG GCGCTGCGGGCCACC	Second round forward primer for constructing 20nt-AUG-20nt-SL mRNA reporter
22	T7-Luc-SL-20nt-Stop_R	CTTACACGGCGATCTTT CCGCCCT GCGCTGCGGTGGCCCGC AGCGC CACGGCGATCTTTCCGC CCTTCTTGGCC	Reverse primer for constructing SL-Stop Fluc mRNA reporter
23	T7-βglobin-Fluc_F	TAATACGACTCACTATA GGGAGAACATTTGCTTC TGACACAACTGTGTTCA CTAGCAACCTCAAACAG ACACC **ATG** GAAGACGCCAAAAAC	Forward primer for constructing βglobin-Fluc mRNA reporter
24	Fluc_R	TTACACGGCGATCTTTC CGCCCTTC	Reverse primer for constructing all Fluc mRNA reporters, except for SL-Stop
25	T7-20nt-SIINFEKL_F2	TAATACGACTCACTATA GGACACTTGCTTTTGAC AGG **ATG** GAATCGATCATCAAC	Forward primer for constructing unstructured uORF mRNA reporter

26	T7-20nt-SL-Tracer-GFP_F1	CACAGACAAGGCATTGTCTGTGAGTAAGGGACACTTGCTTTTGACAGG **ATG**GAATCGATC	First round of forward primer for constructing 20nt-SL-Tracer-GFP uORF mRNA reporter
27	T7-20nt-SL-Tracer-GFP_F2	TAATACGACTCACTATAGGACACTTGCTTTTGACAGGCTTACTCACAGACAAGGC	Second round of forward primer for constructing 20nt-SL-Tracer-GFP uORF mRNA reporter
28	T7-20nt-Tracer(SL+AUG)-GFP_F1	GACAGGGATGATCGATTCCATGCCA**ATG**GAATCGATCATCAACTTCGAAAAGC	First round of forward primer for constructing 20nt-Tracer(SL+AUG)-GFP uORF mRNA reporter
29	T7-20nt-Tracer(SL+AUG)-GFP_F2	TAATACGACTCACTATAGGACACTTGCTTTTGACAGG GATGATCGATTCC	Second round of forward primer for constructing 20nt-Tracer(SL+AUG)-GFP uORF mRNA reporter
30	T7-20nt-Tracer(18nt+SL)-GFP_F1	CGATCATCAACTTCGAAAAGCTATAGGCCACTATAGCTTTTCGAAGACACTAGCTTAGGA	First round of forward primer for constructing 20nt-Tracer(SL+18nt)-GFP uORF mRNA reporter
31	T7-20nt-Tracer(18nt+SL)-GFP_F2	TAATACGACTCACTATAGGACACTTGCTTTTGACAGG **ATG**GAATCGATCATCAAC TTCG	Second round of forward primer for constructing 20nt-Tracer(SL+18nt)-GFP uORF mRNA reporter
32	T7-20nt-Tracer-SL-GFP_F1	CTAGCTTAGGACTTACTCACAGACAAGGCATTGTCTGTGAGTAAGGGAGGAGGAGGATCA	First round of forward primer for constructing 20nt-Tracer-SL-GFP uORF mRNA reporter
33	T7-20nt-Tracer-SL-GFP_F2	TAATACGACTCACTATAGGACACTTGCTTTTGACAGG **ATG**GAATCGATCATCAACTTCGAAAAGCTATAGGACACTAGCTTAGGA C	Second round of forward primer for constructing 20nt-Tracer-SL-GFP uORF mRNA reporter

34	EGFP_R	ACTAGAAGGCACAGTCG AGGCTGAT	Reverse primer for constructing all uORF mRNA reporters
35	sg eIF4A1 (Mouse)	GGAGCCGGAAGGCGTCA TC	sgRNA targeting sequences for eIF4A1 in Mouse
36	sg eIF4A1 (Human)	AAACATCTTGATGTATT TGG	sgRNA targeting sequences for eIF4A1 in Human
37	sg DHX29 (Mouse)	AAAGTAAATTTACCTCA ACA	sgRNA targeting sequences for DHX29 in Mouse
38	sg DHX29 (Human)	ACCAATACACTCTTATA TGG	sgRNA targeting sequences for DHX29 in Human
39	qPCR-Fluc_F	GCCAAAAGCACTCTGAT TGAC	Forward primer for Firefly luciferase qPCR
40	qPCR-Fluc_R	CCATATCCTTGCCTGAT ACCTG	Reverse primer for Firefly luciferase qPCR
41	qPCR-βactin-h_F	AGCCTCGCCTTTGCCGA	Forward primer for human β-actin qPCR

42	qPCR-βactin-h_R	GCGCGGCGATATCATCATC	Reverse primer for human β-actin qPCR
43	qPCR-βactin-m_F	GCGGGCGACGATGCT	Forward primer for mouse β-actin qPCR
44	qPCR-βactin-m_R	TGCCAGATCTTCTCCATGTCG	Reverse primer for mouse β-actin qPCR