

ANALYSIS OF THE ROLE OF EIF5A IN MAMMALIAN TRANSLATION

APPROVED BY SUPERVISORY COMMITTEE

Joshua Mendell, M.D., Ph.D.

Thomas Carroll, Ph.D.

Nicholas Conrad, Ph.D.

James Amatruda, M.D., Ph.D.

ACKNOWLEDGMENTS

I have been the recipient of a staggering amount of assistance, encouragement and support throughout my academic career and without it, I would not be writing this today. I would like to thank, first and foremost, Dr. Joshua Mendell for allowing me the opportunity to work and learn in the exceptional research environment that is his laboratory. With his commitment to scientific excellence and infectious intellectual curiosity, he sets a perfect example of what an outstanding scientist should be. Every interaction with him has left me excited, motivated and scientifically inspired, and I am immensely grateful for his guidance.

The Mendell lab fosters a culture where scientists are invested in each other's success and assistance flows freely, and working here has been a wonderfully enriching experience. I have sought out advice from every single person in the group and this work would not have been possible without them. I would like to specifically thank Dr. Guanglu Shi for his experimental guidance during my initial few years in the laboratory. I would also like to acknowledge Dr. Ryan Golden, Dr. Ryan Hunter and Dr. Juliane Braun, for the many hours of scientific discussion that contributed not only towards the progression of my project, but also to my growth as a scientist. In addition, I am grateful to Frank Gillette for his technical assistance with experiments during this last month.

Importantly, this work would not have been possible without the efforts of talented collaborators who worked with me. Dr. He Zhang and Dr. Yang Xie were instrumental in performing all bioinformatics analyses presented here. Dr. Zhang and I spent a lot of time on the more unconventional analyses required of this work, and I am grateful for his

creativity and commitment to this project. I would also like to acknowledge Dr. Yunkun Dang in Dr. Yi Liu's lab, from whom I received advice on bioinformatic analysis of ribosome profiling data that proved critically important for my work. Further, I express my gratitude to Dr. Vanessa Schmid, Rachel Bruce, and Caitlin Eaton at the Next Generation Sequencing Core of the McDermott Center for performing the deep sequencing experiments that comprise such a large part of this study. Dr. Chao Xing, Mohammed Kanchwala and Adwait Sathe provided technical assistance with analyzing NGS data generated. I am also grateful to Dr. Angela Mobley and the members of the UT Southwestern Flow Cytometry Core for assistance in performing FAC sorting experiments. Additionally, I wish to extend my thanks to Stephen Johnson for all software-related assistance and advice.

The members of my thesis committee – Dr. Thomas Carroll, Dr. Nicholas Conrad and Dr. James Amatruda – have been instrumental to my progress and intellectual growth throughout my time in graduate school. I would like to thank them profusely for their time, thoughtful advice and constant support. I would also like to acknowledge the wonderful and supportive scientific communities at UT Southwestern that I have had the privilege to be a part of – the Genetics, Development and Disease graduate program, the Mechanisms of Disease/ Translational Science track, the Department of Molecular Biology and the NA8 superfamily.

I have had the pleasure of having many extraordinary educators in my life, and I owe them all a deep debt of gratitude for indulging my curiosity and encouraging me to continue to do so. I would like to thank Dr. David Berman for being an exceptional

research mentor, introducing me to the fascinating world of cancer molecular biology and for seeing a scientist in me when I didn't see one in myself.

Lastly, I would like to thank my family, my fiancé and my friends for being my source of inspiration, my pillars of strength, and my biggest champions.

ANALYSIS OF THE ROLE OF EIF5A IN MAMMALIAN TRANSLATION

BY

HEMA MANJUNATH

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

May, 2019

Copyright

By

HEMA MANJUNATH, 2019

All Rights Reserved

ANALYSIS OF THE ROLE OF EIF5A IN MAMMALIAN TRANSLATION

HEMA MANJUNATH, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2019

Joshua Mendell, M.D., Ph.D.

MYC is a critical growth-promoting gene that is subject to tight post-transcriptional control. However, the genes and mechanisms that mediate this regulation at the mRNA level are poorly understood. In order to identify regulators of *MYC* that function through the 5' UTR of the transcript, we performed a fluorescent reporter-coupled genome-scale CRISPR/Cas9-mediated loss of function screen. Analysis of screening data identified eukaryotic initiation factor 5A (*EIF5A*) as novel regulator of *MYC* translation.

EIF5A is a highly conserved translation factor that has been demonstrated to relieve ribosome pauses during translation elongation at 'difficult to translate' peptide

sequences in yeast and bacteria. We observed that *EIF5A* regulates protein isoform distribution of *MYC*, and that loss of function of this gene results in enhanced upstream non-canonical translation initiation on this transcript.

Upon performing ribosome profiling in cells where *EIF5A* or its upstream activating enzyme were ablated, we discovered that the protein's function as a ribosome pause relief factor is conserved in mammalian cells. Importantly, analysis of ribosome profiling data under conditions of eIF5A depletion revealed not only evidence of enhanced ribosome pausing within coding sequences at elongation stall sites, but also an increase in non-canonical/sub-optimal translation initiation events in 5' UTRs in both yeast and human cells.

These data lead us to formulate and test the hypothesis that ribosome pausing resulting from loss of *EIF5A* increases non-canonical translation initiation at pause-proximal upstream sub-optimal initiation codons. We present data from ribosome profiling experiments in yeast and human cells, as well as luciferase reporter assays that are consistent with this model. Thus, we propose a novel role for the translation elongation factor *EIF5A* in maintaining appropriate start codon selection during initiation in eukaryotic cells.

TABLE OF CONTENTS

PUBLICATIONS.....	xi
LIST OF FIGURES.....	xii
LIST OF TABLES.....	xiv
LIST OF ABBREVIATIONS.....	xv
CHAPTER ONE: Introduction and Review of Literature.....	1
Post-transcriptional regulation of gene expression.....	1
Peptide bond formation and the ribosome.....	2
The molecular mechanics of translation: initiation, elongation and termination.....	3
Regulation of translation through 5' leaders.....	6
eIF5A and its unique amino acid – hypusine.....	8
Molecular functions of eIF5A.....	10
Project goals and approach.....	15
CHAPTER TWO: EIF5A plays a role in appropriate start codon selection.....	16
Introduction.....	16
Results.....	19
CHAPTER THREE: Discussion.....	73
Review of findings.....	73
A CRISPR/Cas9 screen identifies putative post-transcriptional repressors of MYC.....	73
eIF5A: the exceptionally well conserved ribosome pause relief factor.....	76
A role for eIF5A in translation initiation.....	78
Summary.....	79
CHAPTER FOUR: Methods.....	81
Human cell line culture.....	81

Construction of EGFP reporter constructs	81
Generation of clonal fluorescent reporter cell lines	82
Genome-wide CRISPR/Cas9 screening	83
LentiCRISPR knockout of genes.....	85
Western blotting analysis	86
RNA isolation and qPCR.....	86
Indel analysis	87
Ribosome profiling	87
Bioinformatic analysis of ribosome profiling data	88
Arsenite treatment.....	91
Polyamine treatment.....	92
Luciferase assays	92
APPENDIX	95
BIBLIOGRAPHY	110

PUBLICATIONS

Hunter RW, Liu Y, **Manjunath H**, Acharya A, Jones BT, Zhang H, Chen B, Ramalingam H, Hammer RE, Xie Y, Richardson JA, Rakheja D, Carroll TJ, Mendell JT. Loss of Dis3l2 partially phenocopies Perlman syndrome in mice and results in up-regulation of Igf2 in nephron progenitor cells. *Genes Dev*, 32(13-14): 903-908 (2018).

Golden RJ, Chen B, Li T, Braun J, **Manjunath H**, Chen X, Wu J, Schmid V, Chang TC, Kopp F, Ramirez-Martinez A, Tagliabracci VS, Chen ZJ, Xie Y, Mendell JT. An Argonaute phosphorylation cycle promotes microRNA-mediated silencing. *Nature* 542(7640):197-202 (2017).

Aloisio GM, Nakada Y, Saatcioglu HD, Peña CG, Baker MD, Tarnawa ED, Mukherjee J, **Manjunath H**, Bugde A, Sengupta AL, Amatruda JF, Cuevas I, Hamra FK, Castrillon DH. PAX7 expression defines germline stem cells in the adult testis. *J Clin Invest* 124(9):3929-44 (2014).

Singal AG, **Manjunath H**, Yopp AC, Beg MS, Marrero JA, Gopal P, Waljee AK. The effect of PNPLA3 on fibrosis progression and development of hepatocellular carcinoma: a meta-analysis. *Am J Gastroenterol* 109(3):325-34 (2014).

LIST OF FIGURES

Figure 1. Proline has a unique amino acid structure	13
Figure 2. Schematic of eIF5A binding to the 80S ribosome and function during elongation	14
Figure 3. Schematic of reporter cell line generation and CRISPR/Cas9 screening strategy	20
Figure 4. EGFP fluorescence of reporter cell lines used for screening.....	21
Figure 5. RIGER analysis of screening data.....	25
Figure 6. Validation of screening results by knockout of candidate regulators in reporter cells.....	26
Figure 7. Effect of knockout of candidate regulators on MYC protein levels.....	27
Figure 8. Loss of <i>EIF5A</i> results in an altered isoform distribution of MYC protein	29
Figure 9. Schematic of the human <i>MYC</i> locus	30
Figure 10. Loss of <i>DOHH</i> phenocopies loss of <i>EIF5A</i>	31
Figure 11. eIF5A protein levels in <i>DOHH</i> ^{-/-} HCT116 cells.....	32
Figure 12. <i>MYC</i> transcript levels in <i>EIF5A</i> ^{-/-} and <i>DOHH</i> ^{-/-} HCT116 cells	34
Figure 13. MYC1-selective knockout by CRISPR/Cas9	35
Figure 14. MYC protein upon lentiCRISPR targeting of MYC1.....	36
Figure 15. <i>EIF5AL1</i> is not expressed in HCT116 cells	38
Figure 16. Periodicity of ribosome profiling data.....	40
Figure 17. Ribosome occupancy on ORFs with greater or fewer than one ‘proline-proline’ pause site upon loss of eIF5A.....	42
Figure 18. Metacodon plot of ribosome occupancy surrounding ‘PP’ dipeptides.....	44
Figure 19. Results of a motif analysis to identify sites of ribosome pausing	45
Figure 20. Top 20 tripeptide pause motifs identified in <i>EIF5A</i> ^{-/-} cells.....	47
Figure 21. Metacodon plot of ribosome occupancy surrounding pause sites	48
Figure 22. CDF plot of 5' UTR translation in <i>EIF5A</i> KD and WT yeast.....	50
Figure 23. Increased upstream translation observed in <i>EIF5A</i> KD yeast	51
Figure 24. CDF plot of 5' UTR translation in <i>EIF5A</i> ^{-/-} and <i>DOHH</i> ^{-/-} cells.....	52

Figure 25. Increased 5' UTR translation on the human <i>MYC</i> and <i>CCDC94</i> transcripts	53
Figure 26. Gene ontology analysis of genes with increased 5' UTR translation in <i>EIF5A</i> loss of function conditions	55
Figure 27. Hypusine-modified eIF5A and total eIF5A protein levels upon induction of the integrated stress response	56
Figure 28. phospho-eIF2 α and total eIF2 α protein levels upon lentiCRISPR knockout of <i>EIF5A</i> and <i>DOHH</i>	58
Figure 29. Proximal-pausing model of translation initiation/ start codon selection	61
Figure 30. Transcriptome-wide testing of the pause-proximal translation initiation model (using 'P-P' pause sites)	62
Figure 31. Transcriptome-wide testing of the pause-proximal translation initiation model (using identified pause sites)	64
Figure 32. Transcriptome-wide testing of the pause-proximal translation initiation model in yeast	65
Figure 33. Polyamine-hypusine biosynthesis pathway	67
Figure 34. The effects of modulating polyamine levels on eIF5A and MYC	69
Figure 35. Luciferase reporter assays to test the pause-proximal model of translation initiation	72

LIST OF TABLES

Table 1: Top 50 genes by RIGER analysis of CRISPR-Cas9 screen in HCT116 ^{MYC 5' UTR} cells.....	95
Table 2: Top 50 genes whose 5' UTR translation is increased in <i>EIF5A</i> ^{-/-} yeast, compared to WT yeast.....	98
Table 3: Top 50 genes whose 5' UTR translation is increased in <i>EIF5A</i> ^{-/-} human HCT116 cells, compared to control NT cells	101
Table 4. Oligonucleotide sequences used in this study	104

LIST OF ABBREVIATIONS

A site	Aminoacyl site of the ribosome
aIF	Archaeal initiation factor
ATP	Adenosine triphosphate
CDF	Cumulative distribution function
CDS	Coding sequence
CRISPR	Clustered regularly interspaced short palindromic repeats
DNA	Deoxyribonucleic acid
E site	Exit site of the ribosome
eEF	Eukaryotic elongation factor
EF	Elongation factor (as in EF-Tu, EF-G, EF-P)
EF-Tu	Elongation factor thermo unstable
EGFP	Enhanced green fluorescent protein
eIF	Eukaryotic initiation factor
FACS	Fluorescence activated cell sorting
GTP	Guanosine triphosphate
IRES	Internal ribosome entry site
ISR	Integrated stress response
kDa	Kilo Dalton
Met-tRNA ⁱ	Initiator methionine tRNA
miRNA	MicroRNA
mRNA	Messenger RNA
NGS	Next-generation sequence
NHEJ	Non-homologous end joining
ORF	Open reading frame
P site	Peptidyl site of the ribosome
PIC	Pre-initiation complex
RIGER	RNAi Gene Enrichment Ranking

RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
sgRNA	Single guide RNA
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
T-ALL	T-cell acute lymphoblastic leukemia
TC	Ternary complex
tRNA	Transfer ribonucleic acid
uORF	Upstream open reading frame
UTR	Untranslated region

CHAPTER ONE:

Introduction and Review of Literature

Post-transcriptional regulation of gene expression

The central dogma of molecular biology describes the flow of information that governs all life on Earth – genetic information stored in deoxyribonucleic acid (DNA) is transcribed into ribonucleic acid (RNA), whose message is then translated into protein (10). Importantly, this pipeline highlights not only the molecular processes that execute this transfer of information – transcription and translation – but also the nodes at which these complex processes can be regulated. Transcription, for instance, we now appreciate is the product of combinatorial control by several factors – promoter and enhancer DNA elements, transcription factors, RNA polymerase activity and chromatin organization and modification (11). These well-coordinated regulatory networks are essential for all cellular processes in prokaryotic and eukaryotic organisms alike.

Post-transcriptional regulation of gene expression acts at the level of the RNA, and impacts RNA stability (12), RNA export or localization (13) and translation of messenger RNA (mRNA) by the ribosome into protein (14) (15). Cis-regulatory elements that govern these processes are often located at the 5' or 3' ends of the transcript, within regions that do not encode the amino acid sequence of a protein product known as untranslated regions (UTRs), and exert their effects through the recruitment of trans-regulatory proteins or RNAs (16). Sites complementary to microRNA (miRNA) seed sequences, for instance, lie within the 3' UTRs of mRNA transcripts and dictate transcript destabilization

and repression of translation through the recruitment of a corresponding miRNA and a protein effector complex termed the RNAi Induced Silencing Complex (17). The 5' UTR, or 5' leader sequence, is the primary site of control of translation of a given mRNA transcript and operates through various mechanisms that will be discussed in detail below.

Peptide bond formation and the ribosome

Ribosomes – ribonucleotide-protein complexes measuring millions of Daltons in size – are responsible for synthesizing protein from mRNA in cells. RNA comprises approximately 60% of this macromolecule by weight and interestingly, it is a ribosomal RNA (rRNA) molecule that performs the catalytic function of peptide bond formation between amino acids – i.e. catalyzes the peptidyl transferase reaction. Ribosomes exist in both prokaryotic and eukaryotic cells in the form of large and small subunits (termed 60S and 40S in eukaryotes, respectively) that associate on the mRNA upon productive initiation of translation. The small subunit holds the transcript that is to be decoded into amino acids, and the large subunit is the seat of peptide bond formation and houses transfer RNAs (tRNAs), small RNA molecules that bring amino acids to the site of protein synthesis during translation.

As the ribosome progressively moves along the mRNA, tRNAs, along with the corresponding codon on the mRNA, sequentially occupy three sites within the ribosome – the aminoacyl (A) site, the peptidyl (P) site and the exit (E) site. The tRNA charged with

an activated amino acid enters the ribosome at the A site and binds the appropriate codon on the mRNA. This amino acid is then added to growing polypeptide chain attached to the tRNA in the adjacent P site through the formation of a peptide bond – i.e. a peptidyl transferase reaction. This facilitates the transfer of the polypeptide chain to the A site tRNA, which then progresses into the P site of the ribosome. The uncharged, de-acylated P-site tRNA advances into the E site and is released from the ribosome (18-20).

The molecular mechanics of translation: initiation, elongation and termination

The molecular mechanisms of translation can be divided into three stages – initiation, elongation and termination. At each stage, the general translation machinery – ribosomes, tRNA molecules and mRNA being translated – are bound by accessory factors that promote efficient protein synthesis and/or regulate individual steps in the process. Initiation of translation is the most complex of the three stages and also displays the most divergence between prokaryotes and eukaryotes, perhaps reflecting the inherent differences in mRNA transcript structure in these two kingdoms (21). Prokaryotic transcripts are typically polycistronic with multiple ORFs (open reading frames), each capable of independently initiating translation, with short UTRs between them. In eukaryotes, on the other hand, mRNAs contain longer 5' UTRs upstream of a single ORF. Beyond initiation, the mechanisms of ribosome elongation and protein synthesis, including peptide bond formation (discussed above), are very well conserved between prokaryotic and eukaryotic organisms (19).

Initiation of translation in both prokaryotes and eukaryotes begins with the assembly of the pre-initiation complex (PIC) comprising the initiator Methionine-tRNA (Met-tRNA_i) and the small subunit of the ribosome on the mRNA. In bacteria, the PIC is typically recruited directly onto the site of initiation (AUG or near-AUG codon on the mRNA) through base pairing of a specific upstream mRNA sequence, called the 'Shine Dalgarno' sequence, to rRNA of the small ribosomal subunit. The large subunit associates with the PIC to form a fully functional ribosome, with the Met-tRNA_i positioned within its P site and ready to begin elongation. This process is executed nearly entirely by the mRNA and the ribosome, with only three additional trans-factors to facilitate Met-tRNA_i recruitment in an efficient and site-specific manner (22).

In contrast, initiation on eukaryotic mRNAs occurs through a 'scanning mechanism' and involves over 10 accessory proteins termed eukaryotic initiation factors (eIFs). The complexity that has evolved into this process allows for intricate and robust regulation of protein expression at this step in higher organisms. PIC assembly in eukaryotes begins with the formation of a ternary complex (TC) - comprising the tRNA carrier eIF2 in its GTP bound form and Met-tRNA_i. The TC then binds the small subunit of the eukaryotic ribosome (40s subunit) in a process that is facilitated by a host of eIFs (eIF5, 3, 1 and 1A) to yield the functional PIC.

Following assembly of this multiprotein complex, it is then positioned at the 5' end of an mRNA by accessory factors including the RNA-helicase eIF4A and cap-binding protein eIF4E. Subsequently, the PIC "scans" the mRNA in a 5'-3' ATP-dependent manner for a suitable AUG initiator codon. The ribosome is able to scan through RNA

secondary structures and RNA-binding protein complexes with assistance from the helicase eIF4A, eIF4B (a protein that improves helicase activity of eIF4A) and on occasion, additional accessory factors that are recruited to the initiation complex. Once the ribosomal machinery identifies a start codon within a context favorable for initiation – i.e. once the Met-tRNA_i anticodon binds to an AUG on an mRNA – the PIC changes conformation and releases some initiation factors. This enables the large subunit (60S subunit) to join the complex, resulting in an 80S ribosome primed for elongation, with the Met-tRNA_i in the P site and an A site ready to receive new amino acylated-tRNA (23, 24).

Prokaryotic proteins that bind the ribosome and promote translation elongation – prokaryotic elongation factors – have eukaryotic counterparts that are homologous both in structure and function. Translation elongation progresses in cyclic fashion, starting with the second codon and proceeding until the end of the ORF is reached. First, the eukaryotic translation elongation factor (eEF) 1A and the bacterial elongation factor EF-Tu perform the function of binding an aminoacylated-tRNA and guiding it into the A site of the elongating ribosome. Next, peptide bond formation occurs and is followed by translocation of tRNAs from the P and A sites, to the E and P sites respectively. This movement of the entire ribosomal apparatus forward on the mRNA by one codon requires eEF2 in eukaryotes and its ortholog EF-G in prokaryotic organisms. eEF2/EF-G is then released, resulting in an empty A site and a deacylated-tRNA occupying the E site. Upon binding of a new aminoacyl-tRNA in the A site, the E site empties itself and thus, the cycle of elongation continues (25).

Termination of translation is triggered when the ribosome encounters a stop codon in the A site (UAA, UGA or UAG). This final step in the process of protein synthesis is executed differently in prokaryotes and eukaryotes and seems to have evolved independently in these kingdoms. Briefly, release factors in both systems accomplish the task of recognizing the stop codon, hydrolyzing the peptidyl-tRNA bond and releasing the nascent polypeptide, before dissociating themselves from the ribosome. Recycling factors then facilitate the disassembly of ribosome subunits and tRNAs, to be then used on other mRNAs (22, 25).

Regulation of translation through 5' leaders

5' UTRs govern the assembly of ribosomal subunits and translation machinery, as well as translation initiation, and thus can exert strong effects on protein expression. Cues dictating this control may be embedded within the primary nucleotide sequence of the transcript, or may take the form of secondary or tertiary RNA structures. For example, the Kozak sequence (~10 nucleotides long) on eukaryotic mRNAs places an AUG immediately downstream in a context optimal for initiation. An AUG lacking this hallmark is likely to be skipped by the PIC as it scans through the 5' UTR (26) – a phenomenon referred to as 'leaky scanning' (15).

RNA secondary structures that impede movement of the ribosome can affect translation initiation in primarily two ways. They can either preclude scanning-dependent initiation of protein production downstream (27), or promote initiation upstream at non-

canonical initiation sites (28). The latter effect is attributed to stalling of the 40S subunit at the RNA structure decreasing the probability of leaky scanning through an upstream near-cognate initiation codon (non-AUG) or an AUG in suboptimal context. RNA structures have been also been documented to impede ribosome binding when located proximal to the mRNA 5' cap (29).

Conversely, certain RNA elements have been reported to facilitate ribosome recruitment and initiation of translation in a cap-independent manner. These structures, termed 'Internal Ribosome Entry Sites' or IRES elements, are located in the 5' UTRs of viral transcripts, enabling them to cope with the cessation of canonical translation during viral infection. Several eukaryotic cellular mRNAs are hypothesized to harbor IRESes that promote translation initiation under special circumstances – i.e. in response to an external stimulus, viral infection or stress (29).

Upstream ORFs (uORFs) within 5' UTRs have been documented in a large fraction of the eukaryotic transcriptome (30) and provide an additional layer of control of translation at the mRNA level. The majority of uORFs are thought to be inhibitory in nature, repressing translation of downstream coding sequences (CDS) through leaky scanning mechanisms – either by stalling PICs from reaching canonical initiation sites, or by preventing efficient reinitiation of translation downstream at a canonical AUG (15, 31). Interestingly, while AUG-initiated uORFs seem to function this way, translation of non-AUG uORFs positively correlates with translation of the associated main ORF (32). Typically, initiation at such alternative start codons occurs upon mis-pairing between a near-cognate codon on the mRNA and the Met-tRNA_i within the TC. However, it has been

demonstrated that in some cases, alternative initiation factors that promiscuously bind non-Met tRNAs (eIF2A, among others), can substitute for the canonical Met-tRNAⁱ carrier – eIF2 – and initiate translation at near cognate codons (33).

Translational control conferred by these uORFs has been most extensively studied in the context of the integrated stress response (ISR), where they appear to play critical roles in regulating their corresponding downstream ORFs. Upon activation of the stress response, eIF2 is phosphorylated by one of several stress responsive kinases, dramatically impairing TC assembly and resulting in near global shutdown of translation. Some genes escape this inhibition of translation and are translated in a uORF-dependent manner. Translation of stress-response factors *ATF4* and *BIP*, for instance, persists during the ISR and in both cases, is dependent on concurrent translation of non-AUG initiated uORFs in their 5' leaders (34). Another subset of genes including *GCN4* and *GADD34* seem to achieve efficient translation during the ISR through an alternate mechanism whereby leaky scanning is enhanced – leading to reduction in translation of uORFs, and enhanced translation at the principal ORF on the transcript (15).

eIF5A and its unique amino acid – hypusine

Eukaryotic translation initiation factor 5A (eIF5A) is a ubiquitously expressed, 18kDa RNA-binding protein. This protein shows an extraordinary level of conservation across the eukaryotic kingdom, with human and yeast eIF5A sharing >60% amino acid identity (35). The functional equivalency between these two proteins at either end of the

eukaryotic spectrum was demonstrated when human eIF5A protein effectively substituted for the depletion of the yeast homologs (encoded by the genes *HYP2* and *ANB1*) (36, 37). Strikingly, the eukaryotic protein also shares significant homology in both structure and function with its bacterial and archaeal forms (elongation factor P or EF-P, and archaeal IF5A or aIF5A, respectively), and the study of eIF5A homologs in prokaryotes has heavily informed our understanding of the protein in higher organisms.

In humans and yeast, eIF5A comprises a family of two paralogs that are >80% identical in amino acid identity (35). The human isoforms – eIF5A and eIF5A2 – differ only in expression pattern, where the latter is restricted to testis, colorectal adenocarcinoma and brain (38). Remarkably, the eIF5A proteins are the sole proteins in the eukaryotic kingdom to contain the unique amino acid hypusine (39).

Hypusine, or hydroxyputrescine–lysine, is a basic amino acid that is derived from the polyamine spermidine in two steps and post-translationally incorporated into eIF5A, replacing K50 in the amino acid sequence of the protein. Both the modification and the enzymes that catalyze the two sequential reactions – *DHPS* (deoxyhypusine synthase) and *DOHH* (deoxyhypusine hydroxylase) in humans – that result in the production of hypusine are highly conserved from yeast through humans. In fact, upon introduction into yeast, human eIF5A is capable of being appropriately modified (36). Hypusine-modified eIF5A has been shown to be crucial for eIF5A function, and the human and yeast point mutants that are incapable of being modified do not rescue the effects of depletion of the protein (40, 41).

Hypusine is also incorporated into eIF5A, and while EF-P is lacking in this modification, some bacteria harbor a post-translational modification on the corresponding lysine of the protein. This modification, termed lysinylation, renders the amino acid structurally very similar to hypusine and is thought to function in much the same way (42, 43). While EF-P is not essential for bacterial cell growth (44), eIF5A and its hypusine modification are critical for viability in eukaryotic cells (40, 45), and lethal when depleted in mice (46). Loss of function experiments interrogating function of this protein in higher organisms are thus, technically challenging to perform. Such studies typically report the effects of either partial inactivation or short term perturbation of eIF5A function, by employing siRNA knockdown or stimulus-induced rapid degradation of the protein (9).

Molecular functions of eIF5A

eIF5A was purified and initially identified as an initiation factor because it appeared to promote formation of the first peptide bond in an *in vitro* model assay for translation initiation. In this assay, eIF5A (at the time referred to as IF-M2B α , and subsequently as eIF-4D) stimulated the addition of methionine to the aminoacyl-tRNA mimic, puromycin (47). The protein, in the same study, also tested positively in a model assay for elongation factor activity – the initiation-factor dependent, polyU-directed synthesis of phenylalanine. However, upon interrogation of translation initiation on a natural, non-artificial globin mRNA (48), and initiation complex assembly in an *in vitro* system (49), eIF5A failed to display behavior consistent with a role in translation initiation.

We have come to understand that the met-puro assay interrogates general peptide bond formation or peptidyltransferase activity of the ribosome – and thus, is a more effective monitor of elongation, rather than initiation (19). Interestingly, puromycin is an imperfect substrate for peptide bond formation and thus, eIF5A's activity in these assays could be interpreted as catalyzing peptide bond formation in suboptimal contexts. Further, in vitro reconstituted dipeptide synthesis assays that employ true aminoacyl substrates did not require the addition of eIF5A for efficient translation (7). Collectively, these data support the consensus that has slowly been built about the role of eIF5A since its discovery – that it functions as an elongation, not initiation, factor during translation.

Clarity on the molecular function of eIF5A has only truly emerged within the last decade. The protein was shown to be critical for efficient translation of polyproline motifs in yeast and bacteria (7). Proline is the only amino acid that is an 'imino' acid – i.e. contains a cyclic ring side chain (Figure 1) – and this renders it a poor substrate for peptide bond formation (50). In fact, the proline-proline dipeptide bond is the slowest to form amongst all combinations of amino acid dipeptides (51). eIF5A was demonstrated to relieve translational stalling and promote elongation at polyproline stretches that present significant challenges to ribosome progression. Recently however, ribosome profiling experiments in yeast have demonstrated that several peptide sequences may in fact depend on eIF5A for efficient translation elongation through them, resulting in widespread ribosome pausing upon depletion of eIF5A (9, 52). This is consistent with previous data documenting that inactivation of eIF5A in yeast resulted in an increase in both the cellular polysome fraction, and ribosome transit times across mRNAs (53, 54).

Bacteria seem to show more restricted dependence on EF-P, as data has shown that this elongation factor serves to facilitate efficient translation primarily at proline-proline and polyproline motifs (55-58). However, the underlying function of these proteins – to promote peptide bond formation at sites that may otherwise slow or stall the ribosome – seems to be conserved in all life forms. This model is further strengthened by structural data. eIF5A and EF-P loosely resemble the form of a tRNA and have been shown to bind near the E site of the ribosome in such a way that the hypusine/ β -lysine moiety extends into the catalytic site of peptide bond formation and interacts with the peptidyl-tRNA in the P site of the ribosome (Figure 2) (59, 60). This may allow the proteins to facilitate appropriate positioning of tRNAs and amino acid substrates and promote peptide bond formation.

It was also observed in the recent yeast ribosome profiling datasets, for the first time, that eIF5A appeared to play a general role in facilitating appropriate termination of translation (9, 52) by stimulating the activity of a release factor.

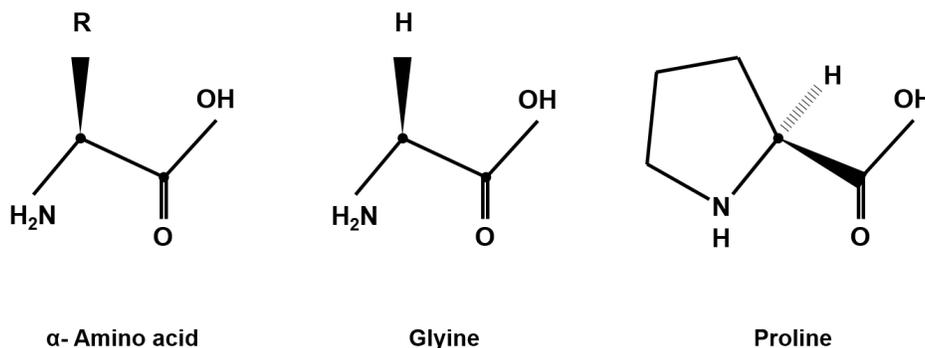


Figure 1. Proline has a unique amino acid structure

A typical amino acid (left) contains a central alpha carbon linked to an amino ($-\text{NH}_2$) group, a carboxylic acid ($-\text{COOH}$) and a variable side chain ($-\text{R}$ group). The R group may be non-polar or non-polar, aliphatic or aromatic and acidic or basic, amongst other classifications. Two amino acids unique in structure are discussed here.

Glycine (middle) has no side chain. The lack of steric constraint afforded by this, along with the side chain's inability to interact with other amino acid side chains imparts flexibility to a peptide chain in which it is incorporated.

Proline (right), on the other hand, is the only amino acid to contain a cyclic ring side chain, and thus is sterically highly constrained and provides rigidity to a peptide. Proline's limited conformational flexibility is also thought to contribute to slow proline-proline peptide bond formation.

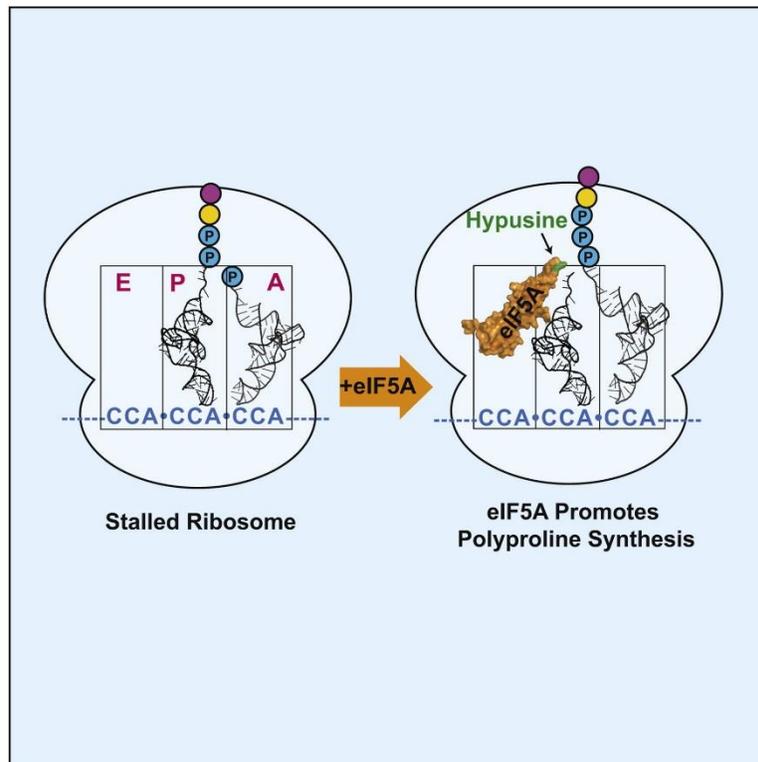


Figure 2. Schematic of eIF5A binding to the 80S ribosome and function during elongation

The hypusine side chain of eIF5A (marked in green) contacts the end of peptidyl-tRNA to stabilize it and its nascent peptide chain, and thereby stimulates peptide bond formation.

Figure adapted from Gutierrez et al. (2013) (7).

Project goals and approach

Genome-wide CRISPR/Cas9-mediated loss of function screening technologies have proven to be an effective means of identifying novel regulators of gene expression (61-63). We sought to employ this approach to interrogate post-transcriptional regulation of *MYC*, a critical growth promoting gene subject to tight control at the mRNA level. While the 5' UTR of *MYC* has been reported to harbor several regulatory cues dictating expression of the protein (64-66), the genes and mechanisms that mediate this regulation are largely unknown. We hypothesized that applying a genome scale-screening strategy to probe a robust fluorescent reporter of *MYC* post-transcriptional regulation would enable us to uncover novel genes that regulate *MYC* and novel regulatory mechanisms that act at the mRNA level.

CHAPTER TWO:

EIF5A plays a role in appropriate start codon selection

Introduction

Untranslated regions (UTRs) of mRNA transcripts are rich in cues that dictate regulation of the protein product encoded between them. These cis-regulatory signals can be encoded in either the primary sequence of the RNA or its secondary structure, and may influence mRNA stability (12), mRNA export or localization (13) and translation (14) (15). For instance, internal ribosome entry sites (IRES) – highly structured RNA elements that that reliably engage translation initiation machinery in a cap-independent manner – have been identified in the 5' UTRs of several viral and cellular transcripts (29).

Chimeric reporter constructs generated by fusing endogenous UTR elements to fluorescent or luminescent reporter ORFs have been used extensively to study mechanisms of post-transcriptional regulation. In order to identify novel genes and mechanisms that control regulation at the mRNA level, we chose to employ a chimeric reporter construct system – coupled with a genome-scale CRISPR/Cas9 screening strategy. CRISPR/Cas9-mediated loss of function screening technologies enable the genetic interrogation of molecular interactions on a large scale and in an unbiased manner. Importantly, when using sgRNAs targeted to a genetic locus, indel formation by non-homologous end joining (NHEJ) results in complete genetic knockout, unlike siRNA or shRNA systems where often only partial loss of function of the gene is achieved.

We chose to use in our studies, as a model system, the *MYC* 5' UTR, due to the large body of evidence indicating the presence of regulatory cues within it. It has been documented for over thirty years that this region of the *MYC* transcript harbors cis-regulatory repressive elements (64, 65), and yet the identity of these elements and mediators of this regulation are unclear. The *MYC* 5' UTR is also thought to harbor a rare eukaryotic cellular IRES element (66), which enables continued protein expression under conditions of impaired cap-dependent translation such as cellular stress or apoptosis (67), and whose regulation also seems to be developmentally controlled (68). While we understand in detail the molecular mechanisms involved in viral IRES translation, eukaryotic IRES elements located on the *MYC* transcript, as well as *BiP*, *FGF2* and potentially up to 10% of the transcriptome, are still poorly characterized (29).

Additionally, the *MYC* transcript produces an N-terminally extended protein isoform of *MYC* upon initiating translation at a CUG initiation codon upstream of the canonical AUG initiation site (69). Despite *MYC1* (CUG-initiated *MYC*) being only 15 amino acids longer than *MYC2* (AUG-initiated *MYC*), the two proteins migrate with apparent molecular weights of 70 and 64kDa respectively. Interestingly, while most non-cognate initiation events result from mismatch between the Met-tRNA_i anticodon and the codon on the mRNA, the translation of *MYC1* was demonstrated to be dependent on the delivery of the amino acid encoded by the CUG codon – leucine, by the corresponding tRNA to the ribosome by an alternative ternary complex not involving eIF2 (33), a mechanism that is employed during antigen presentation on immune cells. It has been suggested that the two isoforms differ in their molecular function as transcription factors

(70, 71) and their expression patterns in certain malignancies (72). This, however, is unclear, as are the molecular mechanisms that regulate this alternative initiation event. Collectively, this large body of work allowed us to reason that the study of post-transcriptional regulators of *MYC* would reveal novel genes and mechanisms involved in this process, not only at the *MYC* locus, but acting on other mRNAs in the transcriptome.

Our screening strategy identified the ubiquitously expressed *EIF5A* as a unique regulator of *MYC* translation – specifically controlling *MYC* isoform distribution. eIF5A has been demonstrated to relieve certain ribosome pauses during translation elongation in yeast and bacterial cells. The ribosome may pause at sites on the mRNA that present challenges to peptide bond formation for several reasons including a sub-optimal amino acid sequence and/or reduced abundance of a certain amino acid. Proline-proline bond formation is particularly inefficient due to this amino acid's cyclic ring side chain, and widespread pausing on mRNAs at sites of di- and poly-proline peptides has been well documented upon loss of eIF5A in yeast and bacteria.

In the following study, we describe the enhanced production of alternatively initiated-MYC1, upon loss of function of *EIF5A* in human cells. We show through ribosome profiling experiments that this function of eIF5A as a translation elongation pause relief factor is conserved in human cells. Further, we provide evidence supporting a model whereby ribosome stalling upon loss of eIF5A can promote alternative initiation events in a pause-proximal manner, in both yeast and human cells. Thus, we propose that eIF5A is a translation elongation factor that serves to also regulate translation initiation and start codon selection.

Results

A genome-scale CRISPR/Cas9 screen identifies EIF5A as a post-transcriptional regulator of MYC

In order to identify novel genes that are involved in 5' UTR-mediated regulation, we employed the fluorescent reporter coupled genome-scale CRISPR/Cas9 screening strategy outlined in Figure 3. We selected the *MYC* mRNA as the model system to use in our screen, guided by data demonstrating that this mRNA was under tight post-transcriptional control dictated by its UTRs. As a screening assay, we generated a chimeric enhanced green fluorescent protein (*EGFP*) construct wherein an *EGFP* ORF was fused to the 5' UTR of *MYC* and thus reported the 5' UTR-mediated regulation of this gene. A control construct similar in all respects, but lacking the 5' UTR was also generated. These constructs were then knocked-in to a safe harbor locus – the *AAVS1* locus – of the stably diploid human colon cancer HCT116 cell line to yield HCT116^{*MYC* 5' UTR} reporter and HCT116^{*EGFP*} control reporter cell lines that were isogenic at all loci except the knock-in allele. Consistent with an extensive body of literature indicating that the *MYC* UTRs harbor repressive regulatory cues, the HCT116^{*MYC* 5' UTR} cells were observed to be dimmer than the HCT116^{*EGFP*} control reporter cells by flow cytometry analysis of EGFP fluorescence (Figure 4).

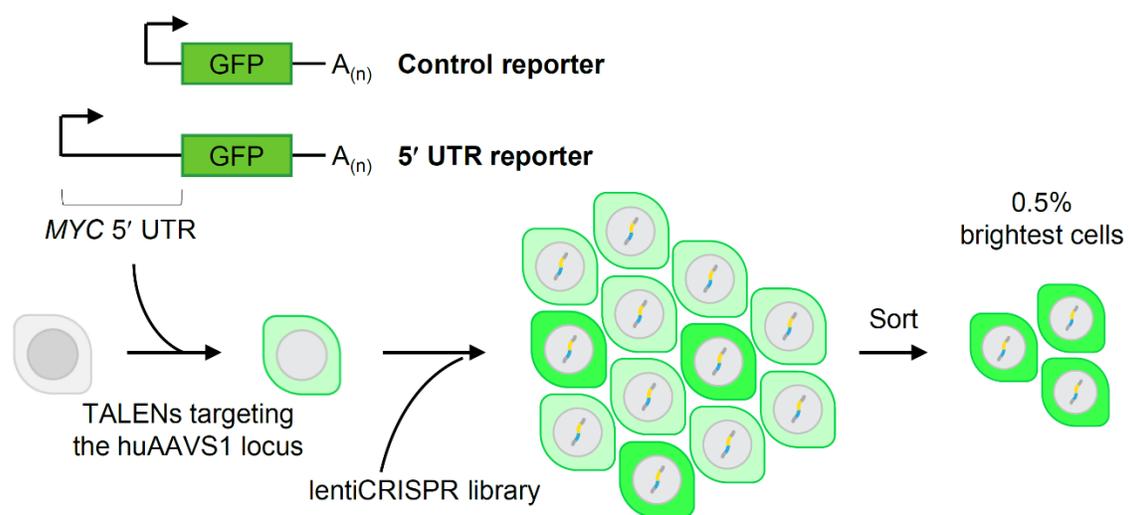


Figure 3. Schematic of reporter cell line generation and CRISPR/Cas9 screening strategy

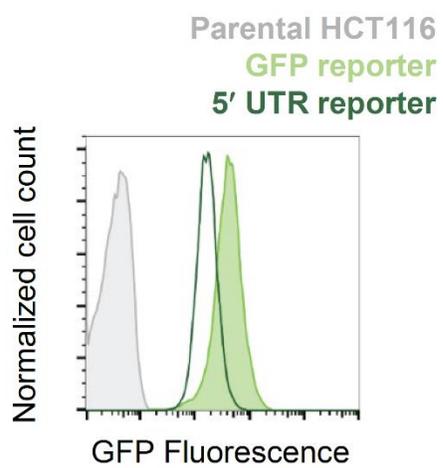


Figure 4. EGFP fluorescence of reporter cell lines used for screening (HCT116^{MYC} 5' UTR reporter and HCT116^{EGFP} control reporter)

Heterozygous knock-in clones of each reporter cell line were subsequently used for genome-scale CRISPR/Cas9-mediated loss of function screening, as previously described (63). Briefly, cells were transduced with a lentiviral library to deliver sgRNAs targeting over 19,000 genes in the human genome, in such a way that each cell received Cas9 and one sgRNA. Selection of these cells in the appropriate antibiotic (puromycin) resulted in the generation of a pool of single gene knockouts that was then subjected to fluorescence activated cell sorting (FACS) to collect the brightest 0.5% of cells. The background population of unsorted cells was also collected and sgRNA representation in the two populations was assessed by next-generation sequencing (NGS). Sorted cells, enriched over the unsorted population for increased EGFP fluorescence, represented cells that received sgRNAs targeting genes that were repressors of *EGFP*. The RNAi Gene Enrichment Ranking (RIGER) algorithm was used to assign a p-value of significance to each gene based on how the collection of sgRNAs targeting each gene performed in the screen. The screen was carried out in exactly the same manner, in replicate, in both HCT116^{MYC 5' UTR} and HCT116^{EGFP} cell lines. Comparison of the screening results from the two cell lines enabled the identification of false positives that regulated the promoter or the *EGFP* ORF directly, and importantly, the selection of UTR-specific hits for further validation (Figure 5).

Chromatin modifiers (*ASH2L*, *BRD4*) and general transcription factors (*TAF1*, *TAF7*, *TAF5*) for instance, might be expected to regulate both control and 5' UTR reporters and accordingly were recovered as significant hits in both screens. Several RNA binding proteins that interact with transcripts via their 5' leaders were identified by the

RIGER algorithm as UTR-specific hits – namely, components of the large subunit of the ribosome (*RPL10*, *RPL10A*) and translation machinery (*EIF3L*, *EIF5A*). Importantly, *CAPRIN1*, an RNA-binding protein previously shown to bind and regulate the *MYC* transcript (73) also emerged as a significant hit in our screen, serving as a positive control and confirming that our screen uncovered biologically relevant post-transcriptional regulators of *MYC*.

Validation of screening results was performed by individually knocking out – using lentiviral delivery of Cas9 and sgRNA, via the lentiCRISPR strategy – genes identified as the most significant, UTR-specific hits, in both HCT116^{MYC^{5' UTR}} and HCT116^{EGFP} reporter cells. Genes that when knocked out, increased fluorescent signal in the HCT116^{MYC^{5' UTR}} cell line, compared to the HCT116^{EGFP} line were carried on for further investigation into their potential effects on *MYC*. (Validation results for *EIF3L* and *EIF5A* shown in Figure 6.) Unfortunately, knockout of genes that were identified by the screen and subsequently validated to regulate fluorescence of the 5' UTR reporter did not increase total *MYC* protein levels in HCT116 cells (Figure 7).

This inability to observe an increase in *MYC* protein levels upon knockout of candidate repressors could be due to the multiple redundant mechanisms that regulate *MYC* protein stability and turnover downstream of the RNA. *MYC* is continuously and rapidly degraded through a process controlled by several regulatory networks (74, 75). Thus, we believe that the regulators identified by the screen hold potential to be true repressors of *MYC*, and any effect they have on *MYC* protein levels might be better observed in other cellular settings. *EIF5A* (Eukaryotic initiation factor 5A), however, not

only successfully validated as a significant hit from our screen – i.e. increased EGFP fluorescence in a *MYC* 5' UTR-specific manner – but also appeared to exert a unique effect on the *MYC* protein.

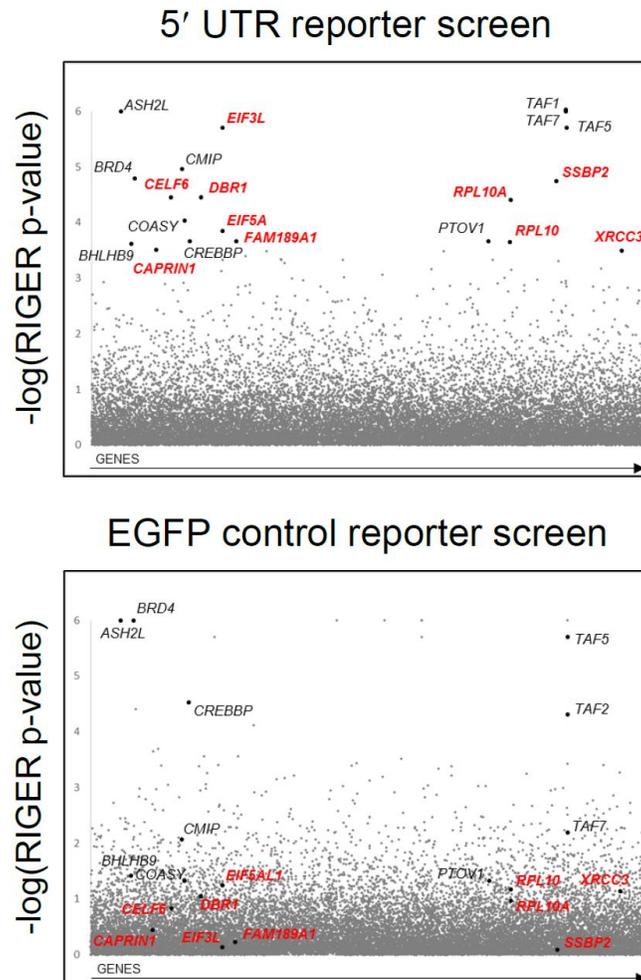


Figure 5. RIGER analysis of screening data from HCT116^{MYC} 5' UTR (top) and HCT116^{EGFP} control reporter cells (bottom), with UTR-specific hits marked in red

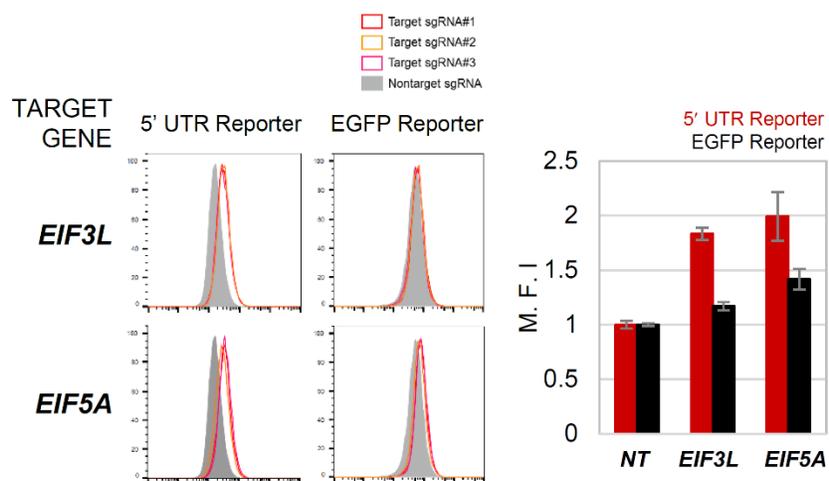


Figure 6. Validation of screening results by knockout of candidate regulators in reporter cells

Flow cytometry analysis of EGFP fluorescence upon knockout of two significant hits identified by the screen in 5' UTR reporter cells (left) and *EGFP* control reporter cells (right); median fluorescence intensity is quantified (right).

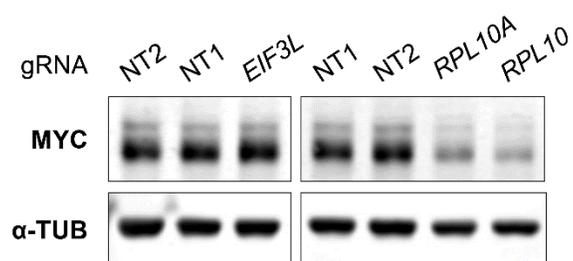


Figure 7. Effect of knockout of candidate regulators on MYC protein levels

Hits identified as post-transcriptional regulators by the screen do not alter total MYC protein levels upon their knockout in HCT116 cells.

Loss of *EIF5A* results in an increase in the N-terminal extension isoform of MYC

We observed a robust and reliable alteration in isoform distribution of MYC in HCT116 cells upon lentiCRISPR knockout of *EIF5A* (Figure 8A). Although *EIF5A* has been well documented to be an essential protein in eukaryotic cells, it is important to note here that we are able to effectively deplete the protein in a population of cells by transduction with lentiCRISPR virus and subsequent selection (for successfully transduced cells) in puromycin for 6 days (Figure 8B). The protein isoform of MYC that appears selectively enhanced in *EIF5A* knockout cells is MYC1 – an N-terminally extended isoform of MYC whose translation has been shown to initiate at a non-canonical CUG initiation codon upstream of the canonical AUG initiation site on the same transcript (Figure 9).

eIF5A is post-translationally modified by the enzyme deoxyhypusine hydrolase (*DOHH*), which catalyzes the conversion of a lysine residue to the unique amino acid hypusine. It is the only protein in the eukaryotic kingdom to contain hypusine and thus, is the only substrate of *DOHH*. Knockout of *DOHH* in HCT116 cells phenocopied the loss of *EIF5A* and resulted in an altered isoform distribution of MYC, with an enhanced production of MYC1 (Figure 10A). Similar to *EIF5A*, *DOHH* is thought to be essential for cellular viability in certain higher organisms, but we are able to achieve near abrogation of protein expression upon lentiCRISPR knockout over a period of 6 days (Figure 10B). Deletion of *DOHH* in the population of cells drastically reduces hypusine-modified, but not total eIF5A levels (Figure 11), indicating that it is the post-translationally modified form of eIF5A that functions to regulate MYC isoform distribution.

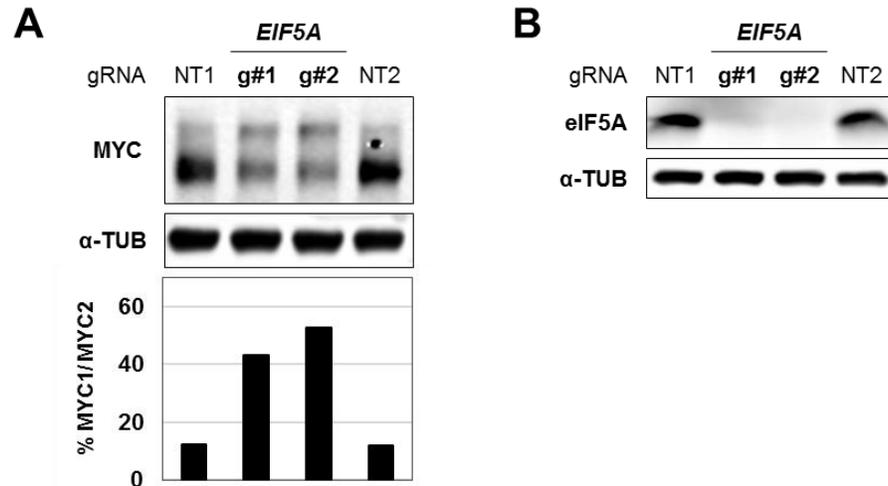


Figure 8. Loss of *EIF5A* results in an altered isoform distribution of MYC protein

(A) MYC protein visualized by Western blot, upon lentiCRISPR knockout of *EIF5A* in HCT116 cells (quantification of MYC1 as a percentage of MYC2, normalized to α -TUB is shown in black). (B) Transduction of HCT116 cells with lentiCRISPR virus targeting *EIF5A* and selection with puromycin for 6 days effectively generates a pool of cells that are depleted of eIF5A protein.



Figure 9. Schematic of the human *MYC* locus showing MYC1 and MYC2 protein isoforms

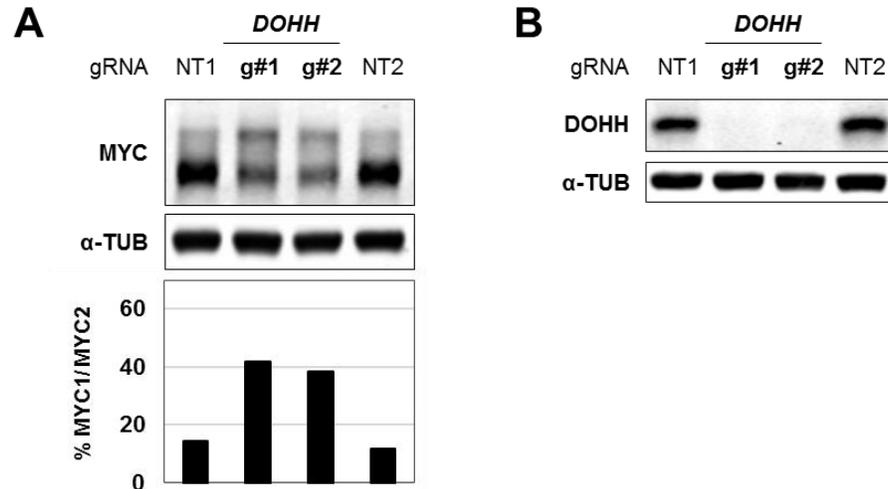


Figure 10. Loss of *DOHH* phenocopies loss of *EIF5A* and alters isoform distribution of MYC in HCT116 cells

(A) MYC protein visualized by Western blot, upon lentiCRISPR knockout of *DOHH* (quantification of MYC1 as a percentage of MYC2, normalized to α -TUB is shown in black). (B) Transduction of HCT116 cells with lentiCRISPR virus targeting *DOHH* and selection with puromycin for 6 days effectively generates a pool of cells that are depleted of DOHH protein.

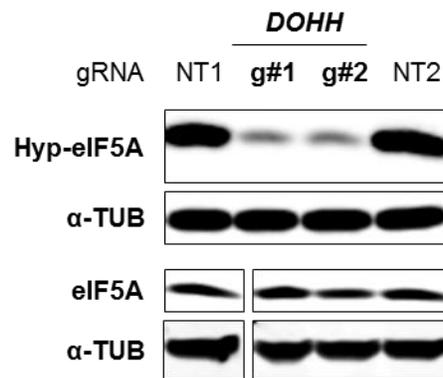


Figure 11. eIF5A protein levels in *DOHH*^{-/-} HCT116 cells (hypusine-modified and total), visualized by Western blot

The change in MYC isoform distribution observed upon loss of hypusine-modified eIF5A appears attributable to impaired or dysregulated translation since MYC1 and MYC2 are protein isoforms that are produced by alternative translation initiation upon the same transcript. This hypothesis is further reinforced by the observation that *MYC* mRNA levels are unchanged upon genetic ablation of *EIF5A* or *DOHH* (Figure 12).

To confirm that the upper band of the MYC protein doublet observed by Western blot is the N-terminally extended MYC1, we took advantage of the fact that MYC1 and MYC2 translation initiation sites are separated in the human genome by a genomic distance of 1.6kb and designed sgRNAs selectively targeting the MYC1 translation initiation site (Figure 13A). Introduction of this sgRNA into HCT116 cells resulted in a reduction of the protein band ascribed to MYC1, due to mutations (insertions/deletions) produced by NHEJ at the translation initiation site of this protein isoform (Figure 14). The reduction in MYC1 expression was further accentuated upon treatment with siRNAs targeting *EIF5A*. NGS confirmed that the MYC1-targeting sgRNAs efficiently generated indels at the desired site near the alternative CUG initiation codon (Figure 13B).

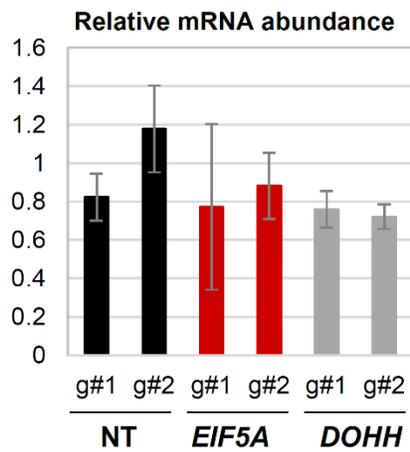


Figure 12. *MYC* transcript levels in *EIF5A*^{-/-} and *DOHH*^{-/-} HCT116 cells assayed by qRT-PCR (18s rRNA used as endogenous normalization control)

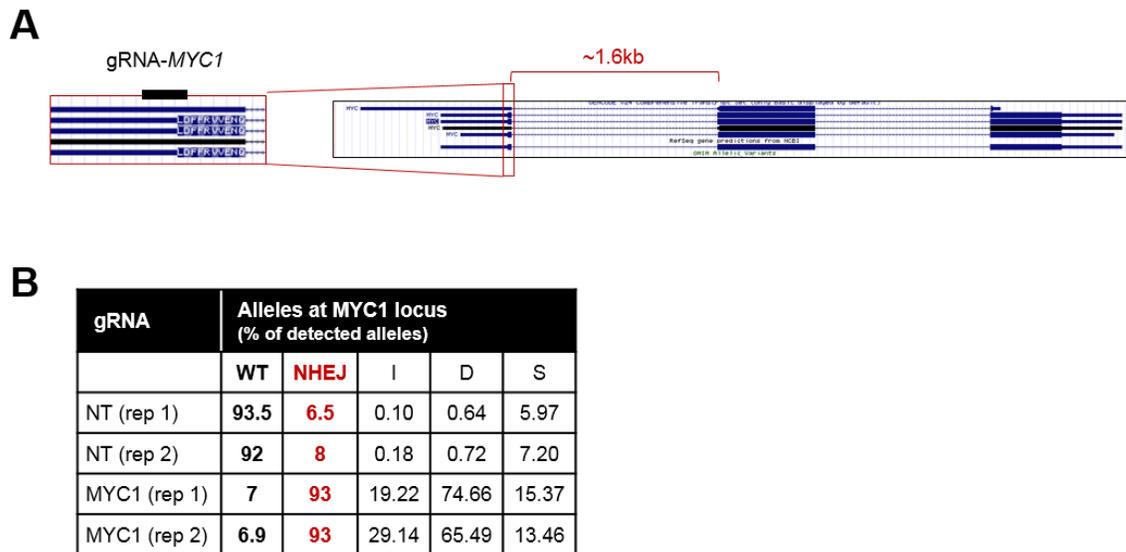


Figure 13. MYC1-selective knockout by CRISPR/Cas9

(A) Human *MYC* genomic locus, with gRNA targeting MYC1 marked in the inset.

(B) Analysis of mutations at the MYC1 genetic locus upon lentiCRISPR transduction with MYC1-sgRNA or NT-gRNA (I = insertions, D = deletions; S = substitutions).

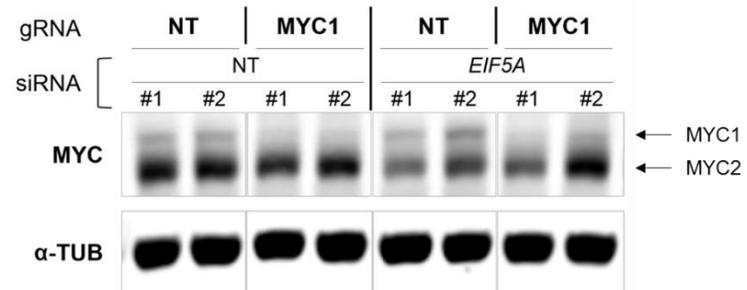


Figure 14. MYC protein upon lentiCRISPR targeting of MYC1, and treatment with siRNAs against *EIF5A*, visualized by Western blot

The gene identified as a UTR-specific, significant hit from our screening data was *EIF5AL1* – a processed pseudogene of *EIF5A* that encodes the same protein. *EIF5AL1* and *EIF5A* differ significantly in transcript structure – with *EIF5AL1* lacking all introns – but have nearly identical amino acid sequence that differs only at 3 positions of the 154 amino acid protein sequence. The sgRNAs enriched in our screening data were annotated as targeting *EIF5AL1*, but were capable of targeting both the *EIF5A* and *EIF5AL1* genes due to their sequence similarity. We were able to conclusively demonstrate that *EIF5A* is the relevant gene in HCT116 cells in two ways. Firstly, qRT-PCR for the two genes showed that the *EIF5AL1* transcript is not expressed in this cell line (Figure 15A). Secondly, sgRNAs designed to selectively target *EIF5AL1* by targeting exon-exon junctions that are absent at the *EIF5A* genomic locus had no effect on eIF5A protein level whereas guides specific for the *EIF5A* locus result in complete loss of eIF5A protein (Figure 15B).

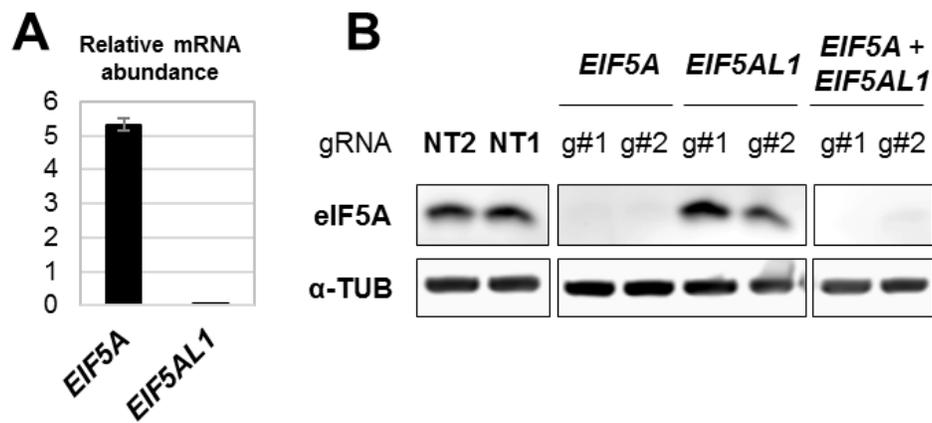


Figure 15. *EIF5AL1* is not expressed in HCT116 cells

(A) qRT-PCR analysis of *EIF5A* and *EIF5AL1* mRNA levels in HCT116 cells (18s rRNA used as endogenous normalization control). (B) eIF5A protein levels upon lentiCRISPR targeting of either *EIF5A*, *EIF5AL1* or both genes.

Ribosome profiling in mammalian cells reveals eIF5A's conserved role as a ribosome pause relief factor

In order to assess whether eIF5A's effect on translation initiation could be observed upon additional transcripts beyond *MYC*, we performed ribosome profiling under *EIF5A* loss of function conditions. We generated pools of *EIF5A* or *DOHH* knockout HCT116 cells using the lentiCRISPR system and following selection in puromycin for 6 days, we harvested the cells and prepared libraries of ribosome footprints from each pool that were then subjected to NGS. The ribosome profiling experiment was carried out in biological replicate – using two independent guides for each experimental and control condition. Upon analysis of the sequencing data, we observed a clear periodicity of the ribosome footprint reads as they mapped to the transcriptome – i.e. a vast enrichment of reads aligned to the correct reading frame of human coding sequences (Figure 16). This is an important hallmark of high-quality ribosome profiling datasets, indicating that ribosome footprints on mRNAs had indeed been effectively isolated and sequenced.

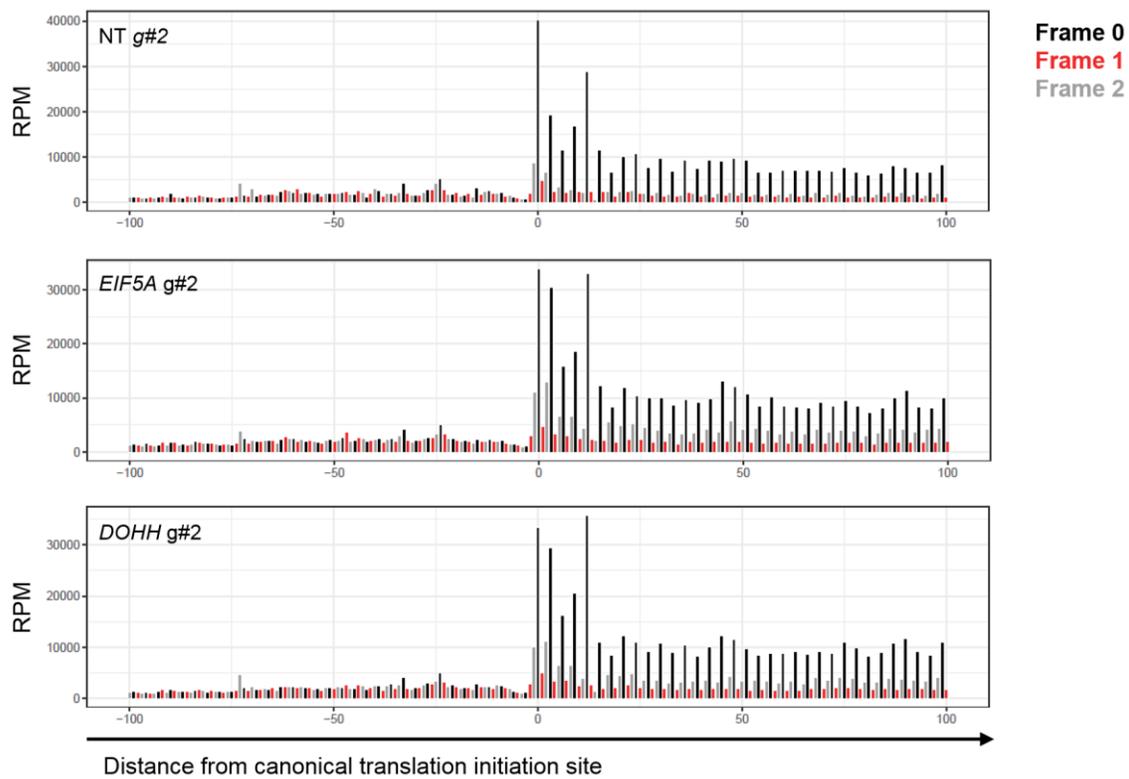


Figure 16. Periodicity of ribosome profiling data (three representative samples shown)

Ribosome profiling experiments provide a snapshot of the position of ribosomes on mRNAs across the transcriptome – enabling the static visualization of elongating ribosomes, but also stationary ribosomes that are stalled on transcripts. The ribosome may pause on a stretch of mRNA for any number of reasons – the occurrence of a suboptimal codon, interaction with the nascent peptide, or sites of slow peptide bond formation (76). It has been demonstrated in yeast and bacteria that eIF5A serves to relieve pauses at these ‘difficult to translate’ peptide motifs by promoting the peptidyl transfer reaction – a classic example being at proline-proline dipeptide motifs. However, the role of mammalian eIF5A in translation pause resolution has not yet been examined, in part because of the technical difficulties in efficiently depleting this essential protein.

Thus, we sought first to utilize our ribosome profiling dataset to assess whether this function of *EIF5A* is conserved in human cells. We discovered that transcripts containing two or more proline-proline dipeptide sequences showed an increased ribosome occupancy in their CDS upon loss of *EIF5A* or *DOHH* (Figure 17), compared to transcripts harboring less than 2 such sites. This increased occupancy is consistent with an increase in ribosomes stalled on proline-proline pause sites within these mRNAs.

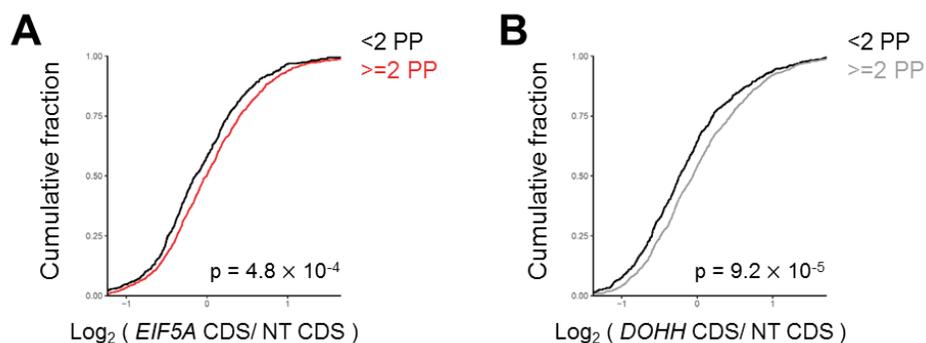


Figure 17. Ribosome occupancy on ORFs with greater or fewer than one ‘proline-proline’ pause site upon loss of eIF5A

Cumulative distribution function (CDF) plot of ribosome occupancy (RPKM) demonstrating that transcripts with more pause sites (≥ 2 ‘PP’ dipeptides in the CDS) see an increased occupancy in (A) *EIF5A*^{-/-} cells and (B) *DOHH*^{-/-} cells compared to NT, when compared to the class of genes with fewer pause sites.

Number of genes included in this analysis: 520 (<2 PP) and 1107 (≥ 2 PP)

In order to visualize whether ribosome occupancy is altered at di-proline motifs that are thought to be strong pause sites for the ribosome, we generated metacodon plots centered on these dipeptide sequences. We observed a clear increase in ribosome occupancy at di-proline peptides in *EIF5A* knockout conditions, compared to non-target cells (Figure 18). A similar effect can also be observed in *DOHH* knockout cells, indicating that loss of the hypusinated-form of eIF5A contributes to this phenotype.

Peptide bond formation between proline-proline residues is thought to be slow and suboptimal due to proline's unique cyclic side chain, resulting in stalling of the ribosome at these locations. However, more recent evidence indicates that several other peptide motifs may also present challenges to peptide bond formation and result in pausing of translation machinery on the mRNA during elongation. We performed a *de novo* motif analysis of peptide sequences where the ribosome stalled on mRNAs in conditions of *EIF5A* or *DOHH* loss of function. To arrive at a probability-derived motif matrix, tripeptides were weighted by their ratio of ribosome occupancy in *EIF5A* knockout cells versus non-target (Figure 19A), or *DOHH* knockout versus non-target (Figure 19B). We observed that ribosomes seem to pause most significantly at proline-proline or proline-glycine dipeptides. This is consistent with several other reports of ribosome stall sequences in yeast and bacteria (9, 52, 55, 57).

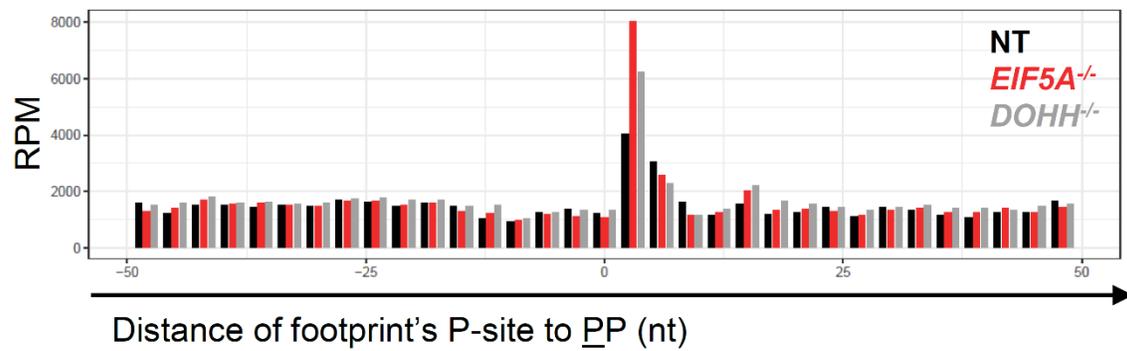


Figure 18. Metacodon plot of ribosome occupancy surrounding 'PP' dipeptides (in a 100nt window) across the transcriptome

Number of 'PP' sites used in this analysis: 14,501

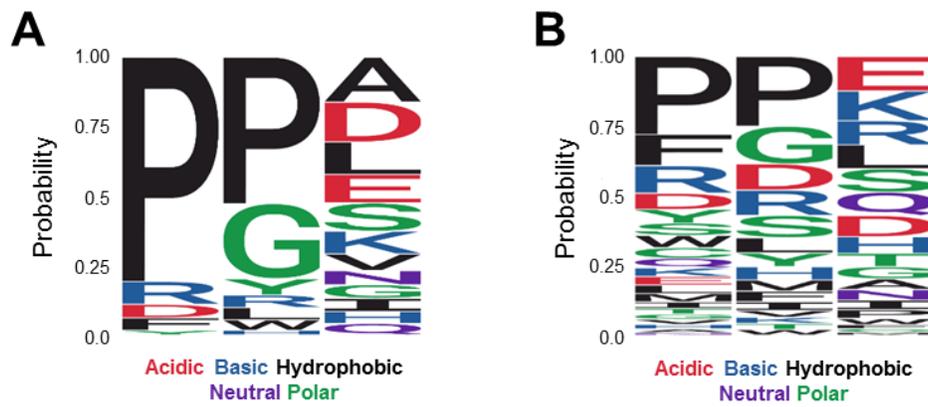


Figure 19. Results of a motif analysis to identify sites of ribosome pausing in (A) *EIF5A*^{-/-} cells and (B) *DOHH*^{-/-} cells

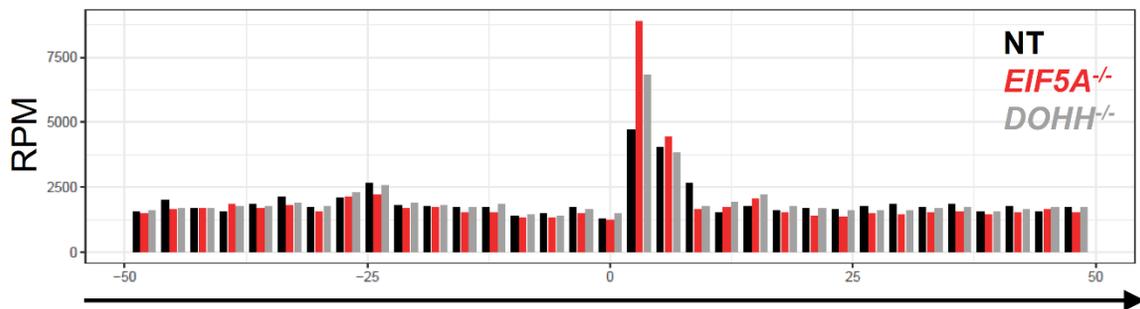
The tripeptide sequences at which ribosomes are most likely to stall in the absence of eIF5A or DOHH (listed in Figure 20) show an overrepresentation of di-proline and proline-glycine residues. Tripeptides in this list are ranked by their pause scores - defined as the ratio of ribosome occupancy in *EIF5A* knockout conditions to control conditions. Importantly, 85% of the motifs identified while comparing *EIF5A* knockout to control were also identified as pause sites in *DOHH* knockout cells, reinforcing the idea that hypusine-modified eIF5A is critical for resolving pauses at P-P and P-G dipeptides. Further, visualizing ribosome occupancy at the identified top common tripeptide stall motifs through a metacodon plot clearly shows that pausing is amplified at these sites in *EIF5A* or *DOHH* knockout cells (Figure 21). Collectively, these data demonstrate that eIF5A's role in relieving ribosome pauses during translation elongation is conserved from bacteria and yeast to humans.

Loss of function of eIF5A results in non-canonical translation initiation events

We attribute the effect loss of hypusinated-eIF5A has on MYC protein to altered translation – specifically, altered initiation characterized by enhanced upstream initiation of translation at a non-cognate codon. This data implies that *EIF5A* plays a role in ensuring appropriate start codon selection on the *MYC* transcript. To investigate the extent to which this hypothesis is true at *MYC* and other loci, we examined upstream translation within the 5' leaders of transcripts in both yeast and human *EIF5A* loss of function systems.

PPD	PGS	LEGEND: PP / PG tripeptides Common to EIF5A ^{-/-} and DOHH ^{-/-}
PPE	PPS	
PPL	DPG	
FRD	PPV	
PPF	PPA	
RLK	KFK	
PGD	KKS	
KPP	PGH	
PPN	EPG	
KMK	PPI	

Figure 20. Top 20 tripeptide pause motifs identified in *EIF5A*^{-/-} cells compared to NT (PP/ PG containing tripeptide motifs marked in red; pause sites identified in both *EIF5A*^{-/-} and *DOHH*^{-/-} cells shaded)



Distance of footprint's P-site to first amino acid of tripeptide pause site (nt)

Figure 21. Metacodon plot of ribosome occupancy surrounding pause sites (in a 100nt window) across the transcriptome (pause sites defined as top 17 motifs listed in Figure 20 identified as pause sites in both *EIF5A*^{-/-} and *DOHH*^{-/-} conditions)

Number of pause sites used in this analysis: 8,359

In 2017, Schuller et al. published ribosome profiling data performed on yeast strains in which eIF5A protein can be induced to rapidly degrade. Our analysis of this data revealed dramatic and widespread increase in upstream translation in yeast with depleted *EIF5A* (compared to WT yeast) as visualized by the cumulative distribution function (CDF) plotting ribosome occupancy in 5' UTRs in *EIF5A* loss of function and WT conditions (Figure 22). These data indicate that a vast number of yeast genes exhibit non-canonical initiation upstream of their CDS upon depletion of eIF5A. Indeed, *GCN4*, *CPA1*, *HAP4* and nearly all classically studied genes known to harbor uORFs show increased ribosome occupancy at these regions (Figure 23).

We next investigated whether ribosome profiling data from HCT116 cells showed the same trend. However, we did not observe a global increase in upstream translation upon knockout of *EIF5A* or *DOHH*, as visualized by a cumulative distribution plot of ribosome occupancy in upstream leader regions (Figure 24). Instead, 5' UTR translation was detected to be enhanced at a subset of genes that includes *MYC* (Figure 25). We thus began to interrogate the features of these genes that resulted in their altered translation upon manipulation of *EIF5A*, and the molecular signatures that might be associated with such an altered translation profile.

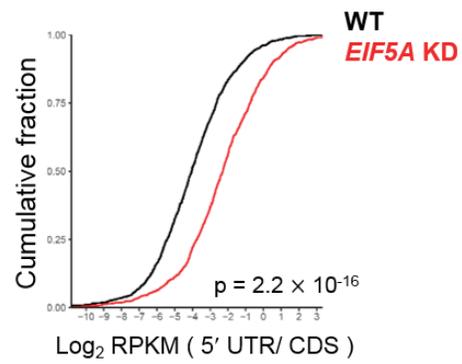


Figure 22. CDF plot of 5' UTR translation in *EIF5A* KD and WT yeast; 5' UTR translation defined as ribosome occupancy (RPKM) in 5' UTR/ CDS

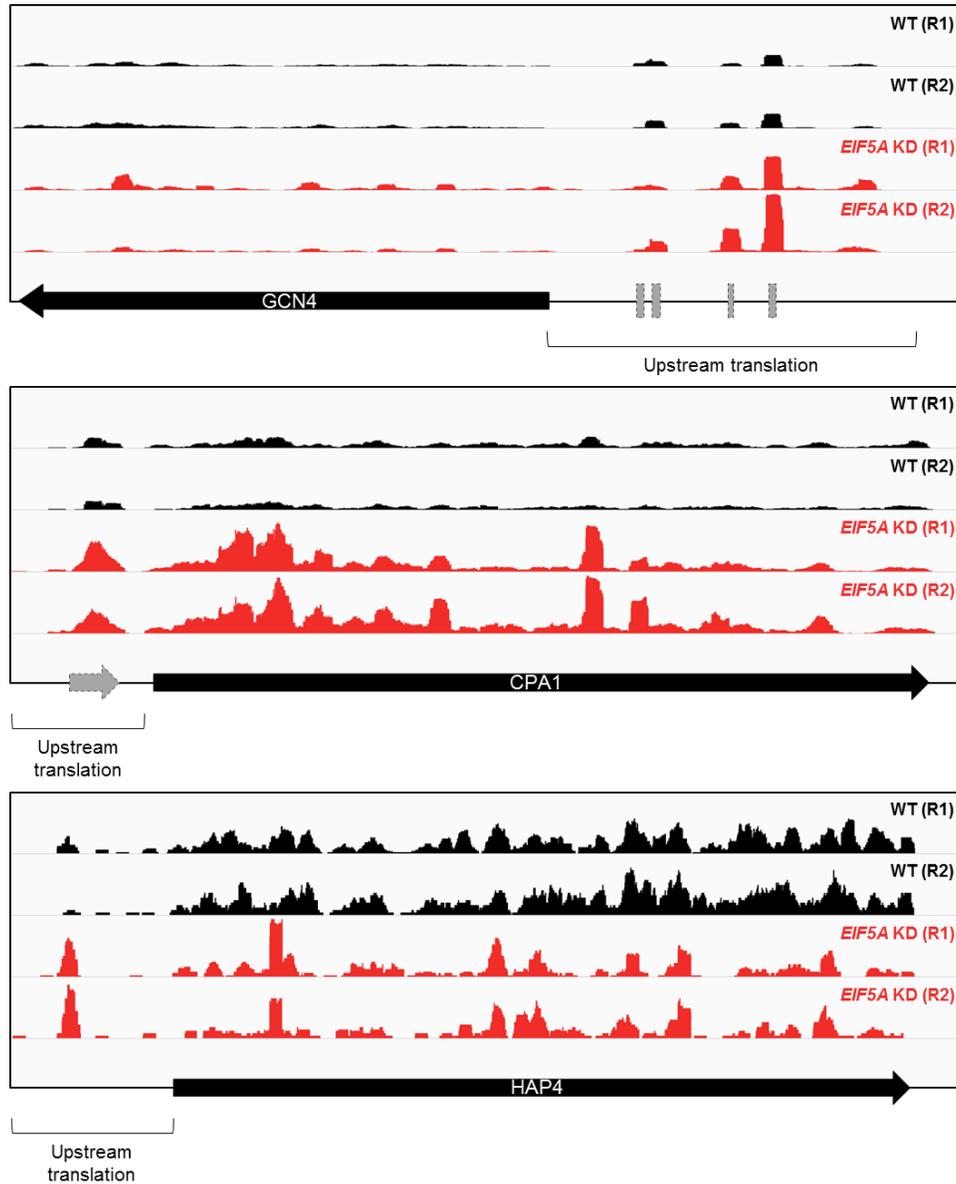


Figure 23. Increased upstream translation observed in *EIF5A* KD yeast compared to WT, on *GCN4*, *CPA1* and *HAP4* transcripts; uORFs previously described associated with *GCN4* and *CPA1* marked in grey (6).

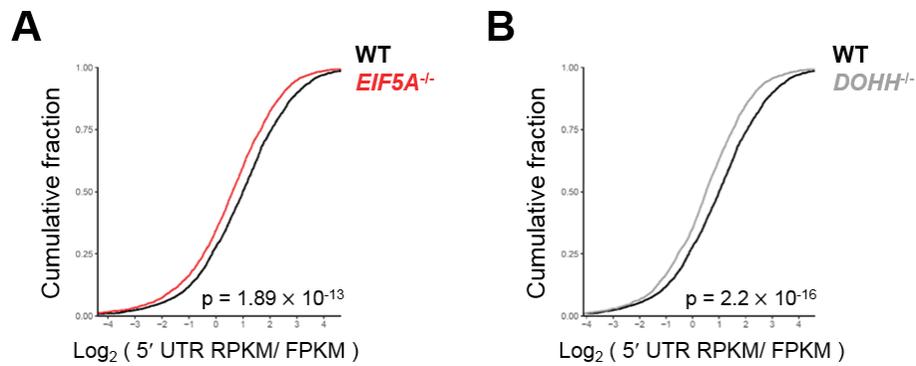


Figure 24. CDF plot of 5' UTR translation in *EIF5A*^{-/-} and *DOHH*^{-/-} cells, marked in red (A) and grey (B) respectively, with 5' UTR translation in control cells marked in black; 5' UTR translation defined as ribosome occupancy in 5' UTR from ribosome profiling/ FPKM from RNAseq

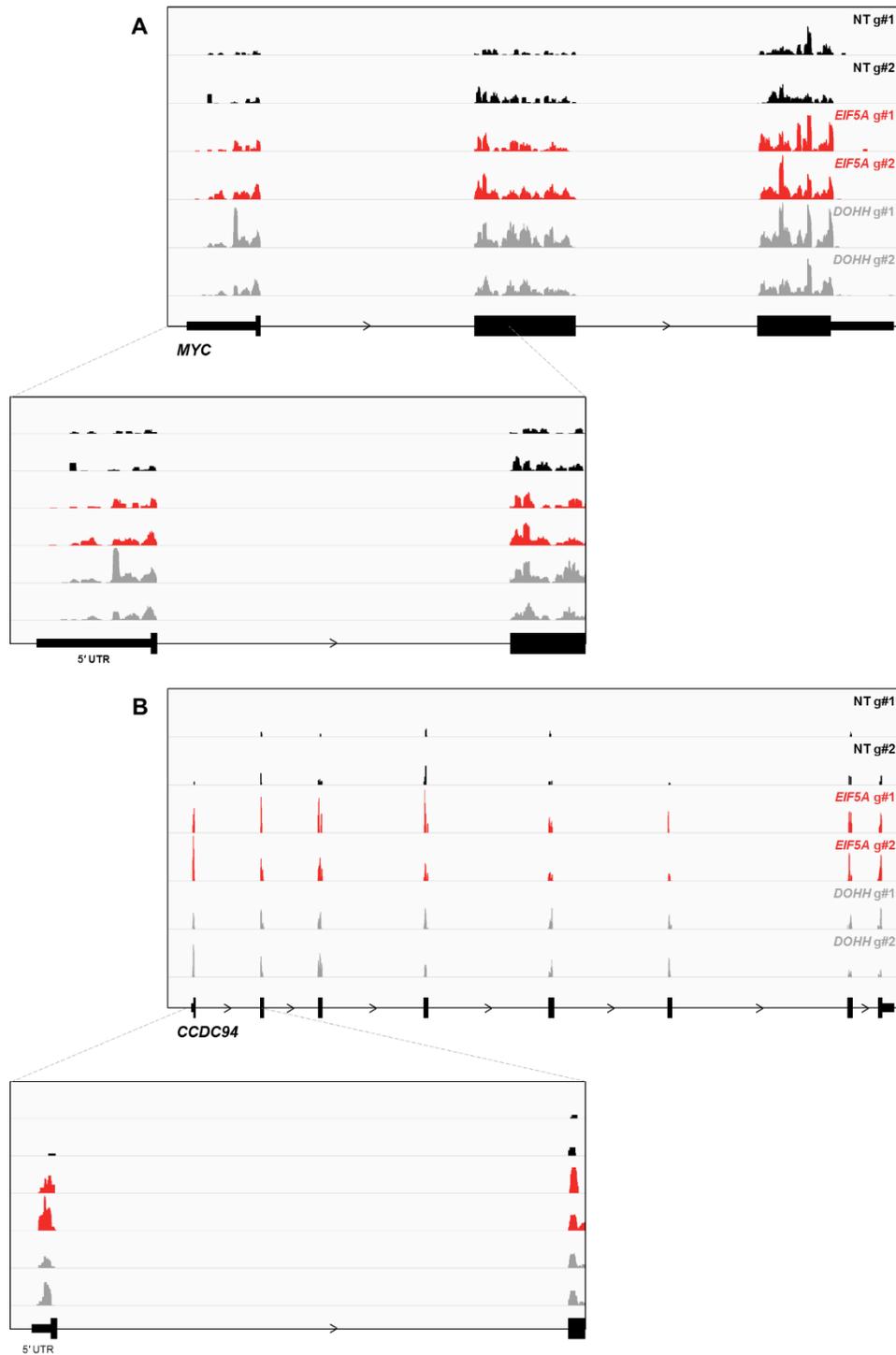


Figure 25. Increased 5' UTR translation on the human *MYC* and *CCDC94* transcripts observed by ribosome profiling in *EIF5A*^{-/-} and *DOHH*^{-/-} cells

Impaired EIF5A function results in a stress-like translational program, that does not interact with the integrated stress response pathway

An examination of the genes within the ribosome profiling dataset whose upstream translation is increased in *EIF5A* knockout HCT116 cells revealed that this class of genes harbored an abundance of stress response genes. The enrichment of this molecular program was calculated to be significant (Figure 26A). Interestingly, this effect was also observed in the yeast ribosome profiling data – potentially indicating the activation of a stress response-like translation program (Figure 26B).

Based on these data, we sought to assess whether signaling through the hypusine pathway and/or eIF5A interacts with the integrated stress response (ISR). To that end, we treated HCT116 cells with sodium arsenite and monitored hypusine-modified and total eIF5A protein levels. We anticipated that should eIF5A be a component of the canonical stress response, we would observe a decrease in hypusine-modified or total eIF5A protein, thus facilitating the altered translation program observed during stress – increased non-canonical, upstream translation and ribosome pausing during elongation. However, our results demonstrated that arsenite-induced stress did not result in a reduction in hypusine-modified or total eIF5A protein (Figure 27).

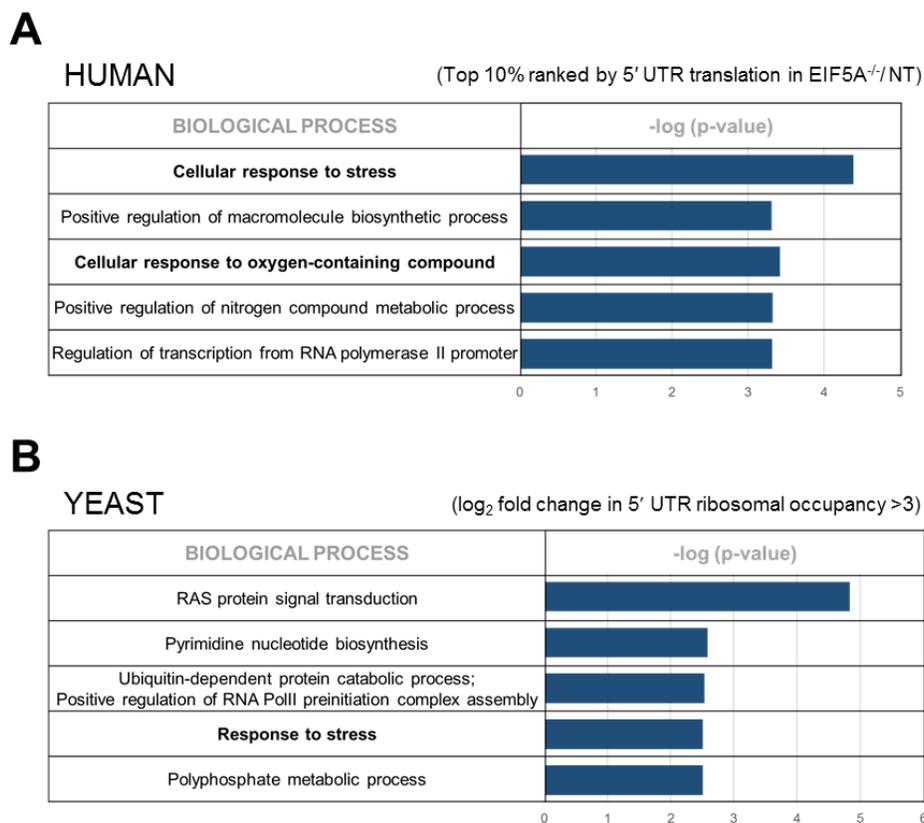


Figure 26. Gene ontology analysis of genes with increased 5' UTR translation in *EIF5A* loss of function conditions compared to control in (A) human and (B) yeast ribosome profiling datasets

(Yeast analysis performed using the Genecodis webtool (1-3); Human analysis performed with the DAVID functional annotation tool (4, 5).)

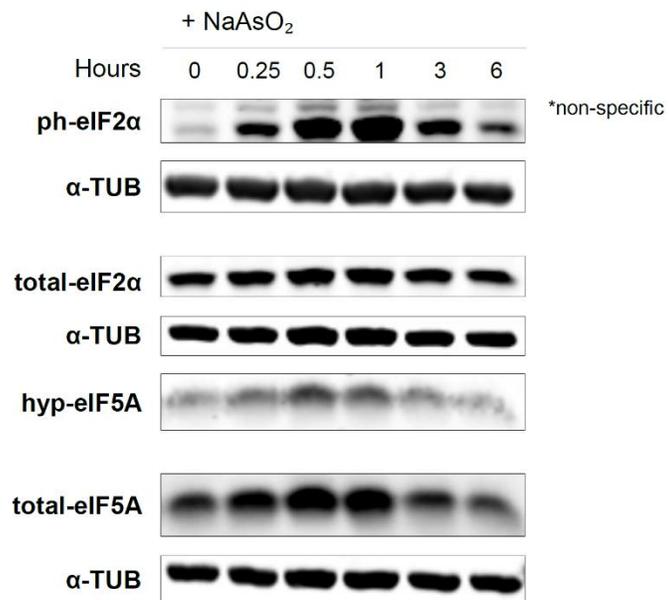


Figure 27. Hypusine-modified eIF5A and total eIF5A protein levels upon induction of the integrated stress response in HCT116 cells by treating with sodium arsenite

Conversely, we theorized that should ablation of *EIF5A* or *DOHH* activate the integrated stress response (ISR), the ensuing phosphorylation of eIF2 α might result in an increase in non-canonical translation. We demonstrated that this is not the case and that HCT116 cells depleted of eIF5A or DOHH protein via the lentiCRISPR system used widely in this study do not have an activated stress response by assaying phosphorylated and total eIF2 α protein levels in these cells (Figure 28). Thus, while loss of *EIF5A* alters translation in a manner similar to the translation stress response, signaling through the hypusine-*EIF5A* pathway is distinct from the integrated stress response.

Proximal-pausing associated non-canonical initiation

We then hypothesized that altered initiation documented upon depletion of hypusine-modified eIF5A protein arises from faulty elongation under these conditions – specifically, enhanced ribosome pausing. We developed a model for pausing associated altered initiation based on two interesting observations. First, we noticed that a ‘proline-proline-alanine’ (PPA) ribosome pause motif was located between the MYC1 and MYC2 initiation sites on the *MYC* transcript. Secondly, it was recently reported that at the *AZIN1* locus, ribosome pausing promoted non-canonical initiation within a short distance upstream of the pause site (77). The authors proposed that this enhanced alternative initiation could be attributed to the queuing of scanning ribosomes upstream of a paused ribosome that consequently enhanced positioning of ribosomes at the non-canonical start site.

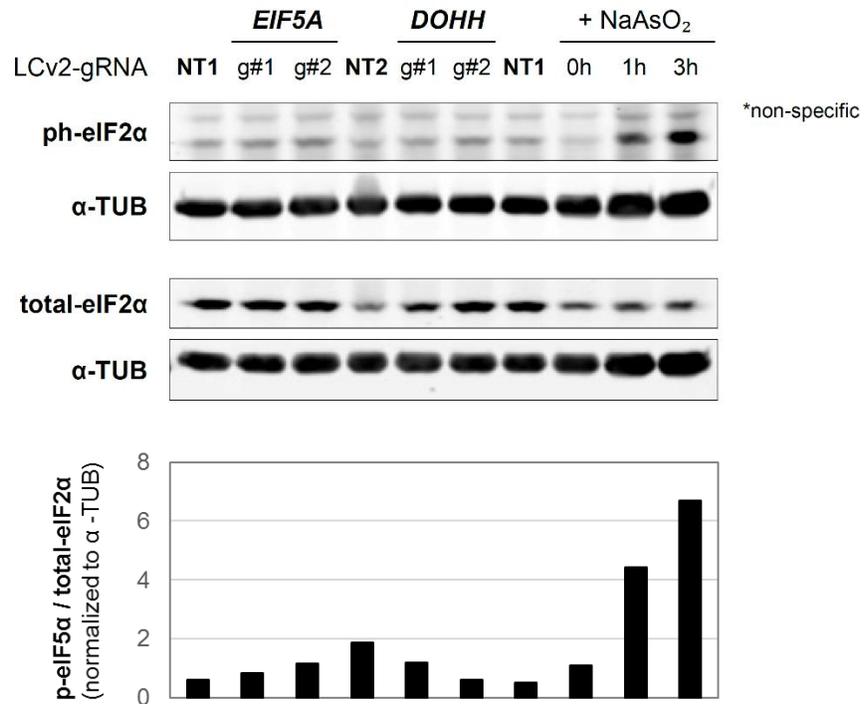


Figure 28. phospho-eIF2 α and total eIF2 α protein levels upon lentiCRISPR knockout of *EIF5A* and *DOHH* in HCT116 cells, monitored by Western blot (lysates from cells treated with sodium arsenite are included as a positive control); quantification of phospho-eIF2 α as a fraction of total is shown (bottom)

This model is consistent with an early report that ribosome stalling (induced not by slow peptide bond formation, but by a stable RNA secondary structure formed by the transcript) downstream of a weak translation initiation site could result in reduction of leaky scanning – i.e. enhanced upstream initiation (28). Thus, we postulated that at the *MYC* locus, ribosome pausing within the leader, that includes a strong pause-inducing tripeptide motif (PPA), drives enhanced usage of the MYC1 translation initiation site upon loss of hypusine-modified eIF5A (Figure 29).

We further posited that this model should hold true at other transcripts that are able to behave in a similar manner, due to the presence of analogous cues in their 5' UTRs – specifically, non-canonical/weak initiation sites upstream of pause motifs. In order to test this hypothesis on a transcriptome-wide scale, we utilized our ribosome profiling dataset. We segregated transcripts into two classes based on their predicted behavior according to our pause-proximal initiation model and then interrogated their 5' UTR translation in *EIF5A* (or *DOHH*) loss of function conditions. The first class comprised genes that would be predicted to show increased 5' UTR translation in *EIF5A*^{-/-} cells, compared to NT – i.e. genes whose transcripts contained a non-canonical initiation site in their 5' UTR and a downstream in-frame 'pause site' within a ~200nt window.

For the purposes of this analysis, we defined a non-canonical initiation site as any 'AUG' trinucleotide with a single mismatch allowed. Importantly, we chose a ~200nt window because of evidence indicating that this was the maximal distance (corresponding to approximately 65 codons) that allowed for effective ribosome queuing-mediated effects

on alternative translation initiation (77). All other genes that did not satisfy these criteria were placed into the second class (Figure 30A).

We examined the cumulative distribution function (CDF) of 5' UTR translation in *EIF5A*^{-/-} cells compared to NT in these two transcript groups. First, in classifying the transcripts, we defined a 'pause site' as the minimal, classically studied *EIF5A*-associated pause motif – a di-proline ('PP') sequence – and interrogated the CDF described above. We observed that genes predicted by the model to have enhanced ribosome occupancy in the leader sequences upon *EIF5A* loss of function, do show enhanced 5' UTR translation as expected (Figure 30B). This trend is also observed when comparing *DOHH*^{-/-} cells to NT (Figure 30C).

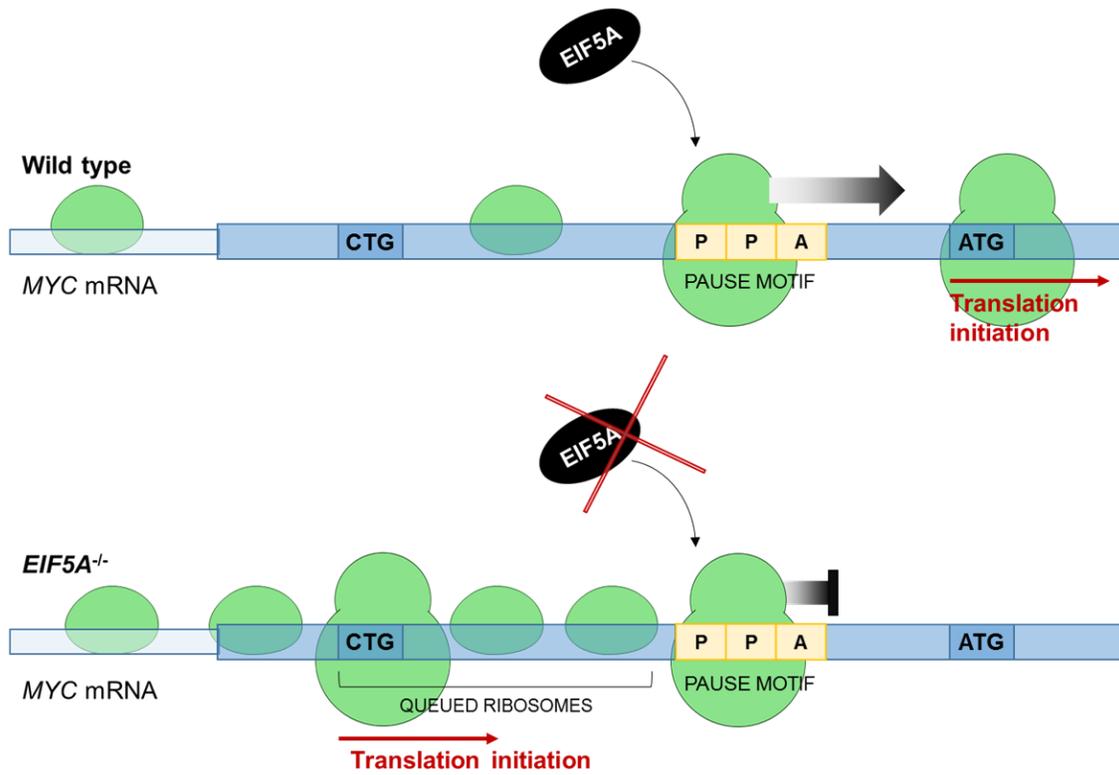


Figure 29. Proximal-pausing model of translation initiation/ start codon selection

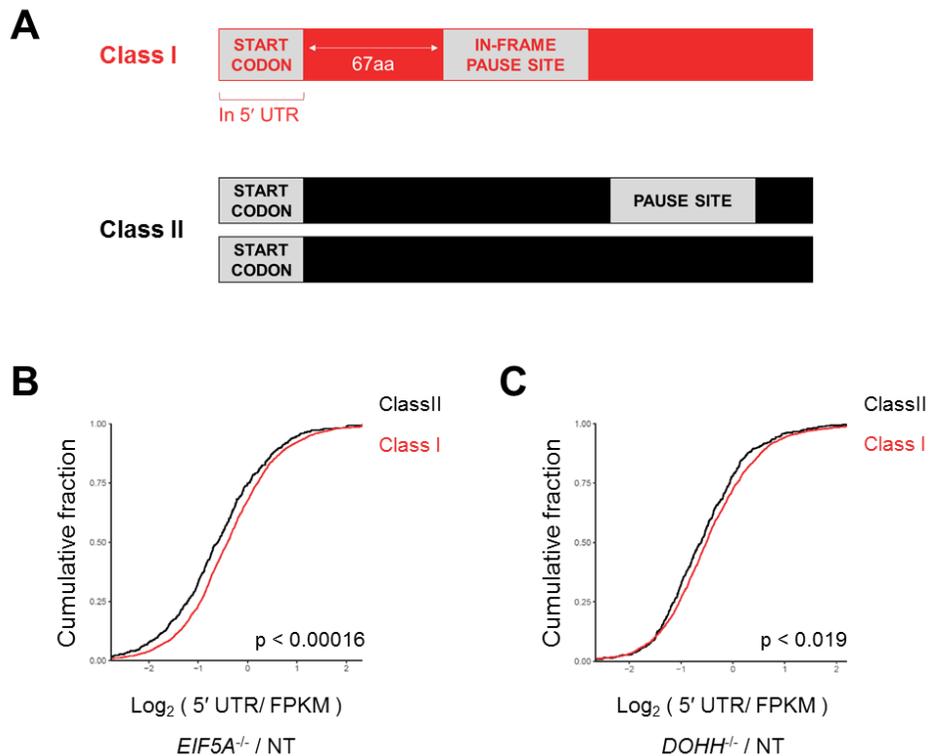


Figure 30. Transcriptome-wide testing of the pause-proximal translation initiation model (using ‘P-P’ pause sites)

(A) Classification of genes based on their predicted behavior according to the pause-proximal model of translation initiation. Genes whose transcript structures indicate an increase in non-canonical upstream translation upon loss of EIF5A are marked in red (Class I), and all other genes in black (Class II). (B) CDF plot demonstrating that Class I transcripts having a non-canonical or weak initiation site in the 5' UTR and a downstream proximal in-frame ‘PP’ dipeptide have increased 5' UTR translation in *EIF5A*^{-/-}, and (C) *DOHH*^{-/-} HCT116 cells.

(2023 Class I genes and 447 Class II genes in this analysis)

Further, when expanding the first class of genes to include those that contain a non-canonical initiation site in the 5' UTR followed by a broader set of in-frame pause sites, as identified by our *de novo* pause motif analysis, we find an even greater increase in upstream translation upon loss of hypusine-modified eIF5A compared to the second class of genes (Figure 31). Lastly, when we apply our model to yeast ribosome profiling data, we see the same trends (Figure 32).

Collectively, these data demonstrate that pause-proximal alternative initiation is a consequence of loss of functional eIF5A protein and provide strong evidence that this mechanism operates on a transcriptome-wide scale, in a manner that is conserved in yeast and human cells.

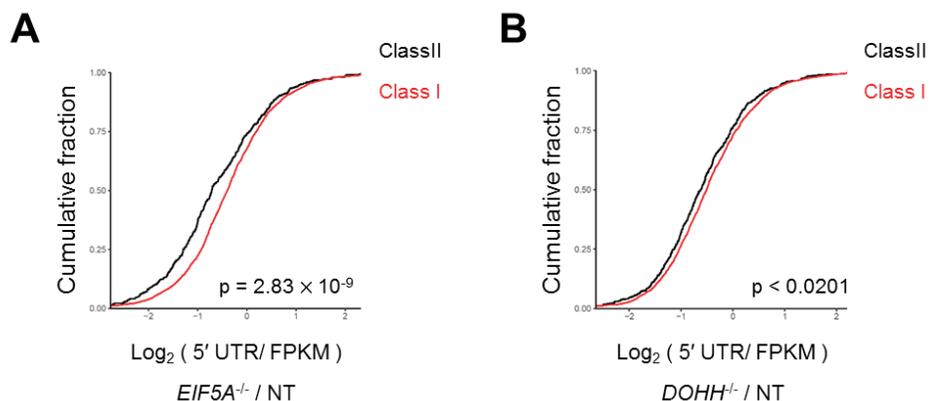


Figure 31. Transcriptome-wide testing of the pause-proximal translation initiation model (using identified pause sites)

CDF plots demonstrating that Class I genes – i.e. genes whose transcripts have a non-canonical or weak initiation site in their 5' UTR and a downstream proximal in-frame pause site – have increased 5' UTR translation in *EIF5A*^{-/-} (A) and *DOHH*^{-/-} (B) cells. Pause sites here are defined as the top 17 common tripeptide pause motifs identified in *EIF5A*^{-/-} and *DOHH*^{-/-} compared to NT HCT116 cells (Figure 20).

(2007 Class I genes and 463 Class II genes in this analysis)

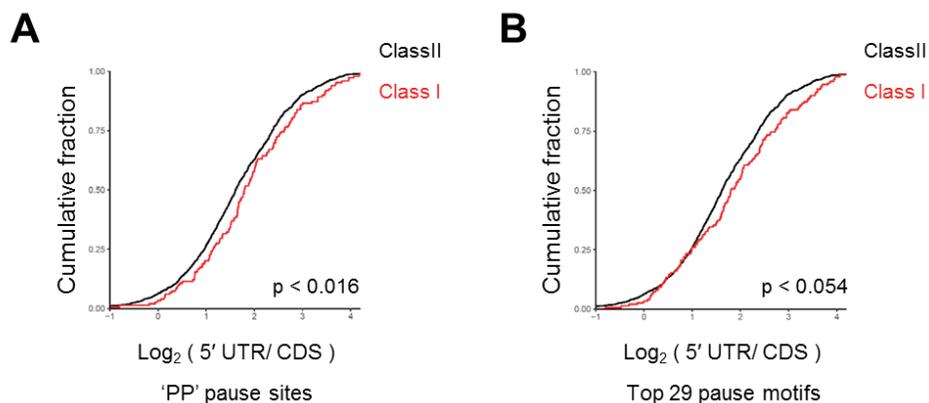


Figure 32. Transcriptome-wide testing of the pause-proximal translation initiation model in yeast

CDF plots demonstrating that consistent with our model, Class I genes in yeast have increased 5' UTR translation in *EIF5A* KD conditions compared to WT yeast, whether pause sites are defined as (A) the minimal classical pause sites - 'PP' dipeptides or (B) the top 29 tripeptide pause motifs identified in *EIF5A* KD yeast in Schuller et al. (2017) (9).

(149 Class I genes and 1189 Class II genes in (A); 150 Class I genes and 1188 Class II genes in (B))

Polyamine modulation of upstream initiation at the *MYC* locus

The study of ribosome queuing-induced upstream translation on the *AZIN1* transcript (77) suggested that this alternative translation from weak initiation sites was modulated by polyamine levels. Polyamines – namely, putrescine, spermine and spermidine – are organic polycations that bind a variety of negatively charged molecules within cells to serve a multitude of functions including modifying chromatin structure and transcription (78), promoting translation (79) and modulating ion channels (80). While the molecular functions of polyamines are not completely understood, it is thought that one of their most critical functions is to generate and incorporate hypusine into the protein eIF5A (81). The regulation of polyamine biosynthesis is intricate, involving unique mechanisms of regulation and multiple feedback loops (Figure 33) (82). *AZIN1* is one such component of the complex polyamine regulatory network and inhibits an inhibitor of ornithine decarboxylase (*ODC*), the enzyme that catalyzes the first step of the polyamine biosynthesis pathway.

AZIN1 translation was shown to be regulated by a uORF whose translation initiation was enhanced in a pause-proximal dependent manner (77). The study put forth a model whereby high levels of polyamines could competitively inhibit eIF5A protein, resulting in enhanced ribosome pausing and pause-proximal non-canonical initiation. Conversely, thus, they proposed that eIF5A protein would function normally to relieve ribosome pauses under conditions of lowered polyamine levels.

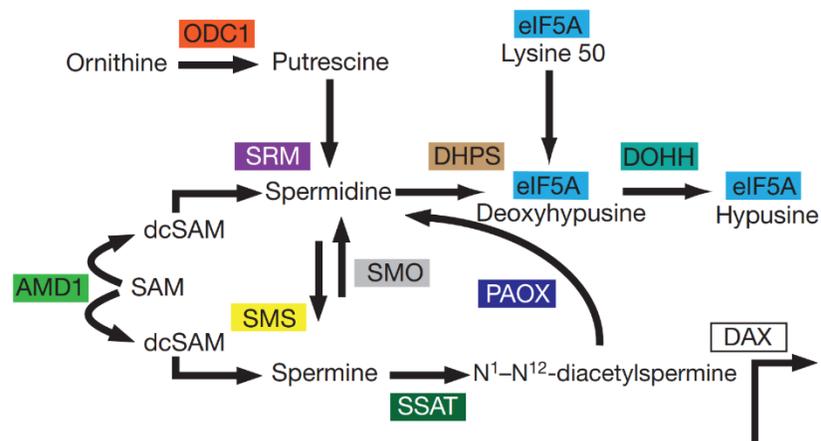


Figure 33. Polyamine-hypusine biosynthesis pathway

Figure from Scuoppo et al. (2012) (8).

Upon testing this model, however, we found that reducing polyamine levels in fact, lowered hypusine-modified, but not total, levels of eIF5A (Figure 34A). This is unsurprising, given that hypusine modification of eIF5A is the ultimate step in the polyamine-biosynthesis pathway (Figure 33). Our work studying ribosome pausing in *EIF5A*^{-/-} and *DOHH*^{-/-} cells illustrates that impairment of the hypusine modification significantly interferes with eIF5A's ability to relieve translation pauses. Taken together, these two pieces of our data indicate an important deviation from the model proposed by Ivanov et al. (2018) (77) – depleting cellular polyamines results in abrogation of *EIF5A* function, and not the converse. This is reflected in the MYC protein isoform distribution observed upon lowering polyamine levels (Figure 34B). It remains possible that polyamines at high concentrations competitively interfere with eIF5A protein function and result in enhanced ribosome pausing.

These experiments highlight that pause-proximal translation initiation is regulated by hypusine-modified eIF5A activity. Further, this unique mode of translation initiation can be physiologically regulated through downregulation of signaling through the polyamine-hypusine biosynthesis axis.

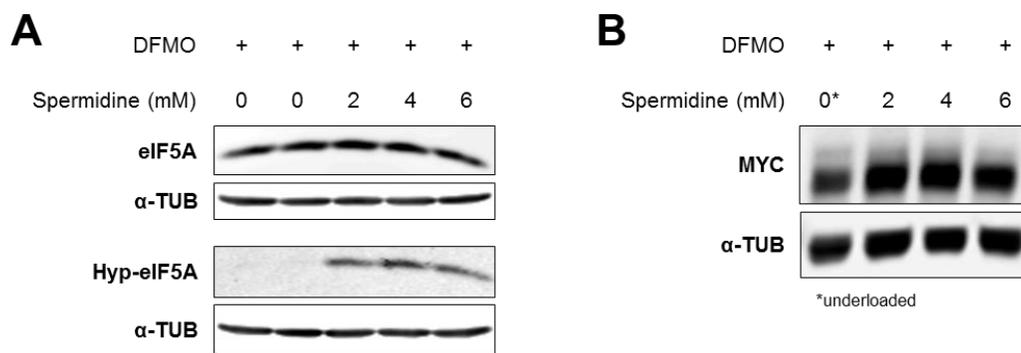


Figure 34. The effects of modulating polyamine levels on eIF5A and MYC

(A) Depletion of polyamines using the ODC inhibitor DFMO in HEK293T cells depletes hypusine-modified eIF5A. (B) Depletion of polyamines in HEK293T cells, and consequently hypusine-modified eIF5A, alters MYC isoform distribution.

The N-terminal amino acid sequence of MYC1 contributes to pause-proximal pausing associated alternative initiation

In order to test whether pause-proximal pausing can induce non-canonical translation initiation on the *MYC* transcript directly, we constructed luciferase reporters that read out MYC1 translation (Figure 35A) and interrogated their behavior under *EIF5A* loss of function conditions. Our model predicts that mutating the ribosome pause-inducing elements within the *MYC* leader – defined here as the 45nt stretch of transcript between MYC1 and MYC2 initiation sites – should reduce enhanced upstream non-canonical initiation upon loss of eIF5A.

Upon transfection of a wild type ‘MYC1’ reporter (containing the wild type *MYC* transcript nucleotide sequence, with initiation codon (ATG) of MYC2 mutated to the non-initiating TTT) into HCT116 cells treated with LCv2-*EIF5A* or LCv2-NT gRNA, we observed enhanced CTG translation in *EIF5A*^{-/-} cells (Figure 35B). In this assay, we measure relative CTG translation – i.e. signal from a luciferase ORF initiated from a CTG codon, normalized to signal from a luciferase ORF initiated from an ATG codon in the identical context – to account for any effects on translation elongation and ensure that we read out the effects on initiation. Introducing a frameshift in the reporter in such a way that amino acids around, and downstream of, the ‘PPA’ pause motif are altered significantly abrogates this effect in *EIF5A* knockout cells. Mutating only the ‘PPA’ motif to ‘PAA’ partially reduced the enhanced CTG translation observed in *EIF5A*^{-/-} conditions.

These data agree with our model that pausing on the *MYC* leader sequence facilitates an increase in upstream initiation at the MYC1 ‘CUG’ initiation codon.

Luciferase reporter data from the frameshifted reporter suggests that pausing on the leader is mediated by not only the 'PPA' pause motif, but also other N-terminal amino acids.

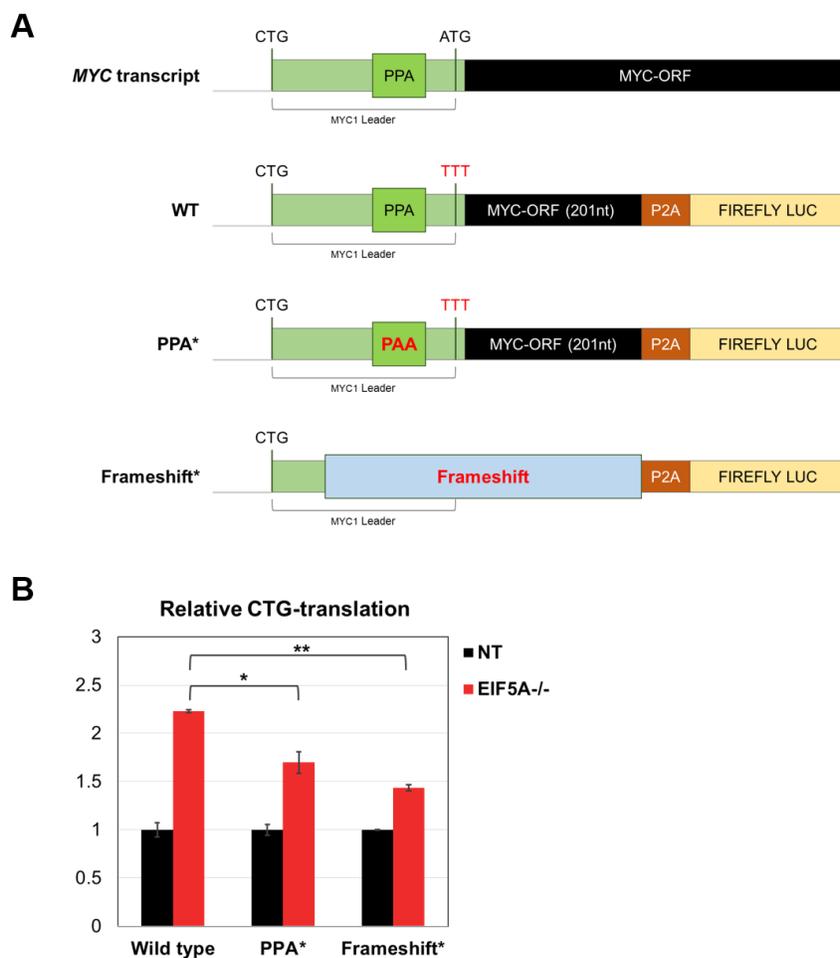


Figure 35. Luciferase reporter assays to test the pause-proximal model of translation initiation

(A) *MYC* transcript structure and reporter constructs used for luciferase assay

(B) Firefly luciferase signal – i.e. measurement of *MYC1* (CTG)-translation

(relative to ATG-translation in corresponding identical context) – upon

transfection of the corresponding firefly reporters into HCT116 cells treated with

LCv2-*EIF5A* or LCv2-NT gRNA. The ‘PPA’ pause motif was mutated to ‘PAA’ to

generate the PPA* reporter, and Frameshift* reporter was constructed by

introducing a single nucleotide deletion 6aa upstream of the ‘PPA’ pause motif.

Firefly luciferase was normalized to co-transfected control renilla luciferase

signal. (** p-value < 0.001; * p-value < 0.02)

CHAPTER THREE:

Discussion

Review of findings

The mechanisms that govern post-transcriptional regulation of *MYC* are poorly understood. In this study, we uncovered a hitherto unappreciated layer of translational control of the *MYC* protein by the ribosome pause relief factor eIF5A. We documented an increase in usage of an upstream non-canonical initiation codon on the *MYC* transcript under conditions of enhanced ribosome pausing resulting from depletion of functional eIF5A protein. Further, we present evidence that similar translation pausing facilitates non-canonical upstream initiation events widely across the transcriptome, and propose a pause-proximal model of alternative translation initiation.

Our data demonstrate – for the first time – not only that the translation factor eIF5A serves to relieve ribosome pauses during translation elongation in human cells, but also that its loss leads to enhanced pausing and non-canonical initiation. Importantly, we establish that these molecular functions of eIF5A are conserved throughout the eukaryotic kingdom - from yeast to human.

A CRISPR/Cas9 screen identifies putative post-transcriptional repressors of MYC

MYC is a critical growth-promoting gene that is subject to tight control at all levels of the central dogma. It is also a potent oncogene in multiple cell and tissue types, whose overexpression or activation is thought to drive tumorigenesis in over 50% of all human

cancers (83). We reasoned that our comprehensive genome-scale interrogation of regulators of *MYC* would reveal genes that play important roles in oncogenesis and present potential novel therapeutic avenues for exploration. The fluorescent reporter-coupled CRISPR/Cas9 screening strategy we designed and employed allowed the recovery of genes that are post-transcriptional repressors of *MYC*, and thus represent potential tumor suppressors.

RPL10 and *RPL10A* were recovered from our screening data, and are interesting candidate tumor suppressors. Heterozygous inactivating mutations or deletions of ribosomal genes often predispose to an increased risk of developing malignancies (84, 85). *RPL10* is particularly well studied in this context and recurrent mutations in this gene have been reported in 9% of pediatric T-ALLs (T-cell acute lymphoblastic leukemias) (86). Whether enhanced signaling through *MYC* contributes to tumorigenesis in patients harboring *RPL10* mutations warrants further inquiry. *EIF3* complex members were also identified by our screen as putative post-transcriptional repressors of *MYC*, which is intriguing in light of data suggesting that specialized eIF3 complexes may dictate the translational regulation of subsets of cellular transcripts – specifically, growth promoting transcripts (87). While the mechanism of translation repression by these specialized eIF3 complexes is unclear, these data allow us to hypothesize that such a complex including the subunits eIF3L and eIF3H (identified by our screen), binds to and regulates translation of the *MYC* transcript.

From the list of candidate repressors of *MYC* revealed by screen, we chose to focus on, for the purposes of this study, the gene *EIF5A*. We demonstrated that depletion

of eIF5A alters the protein isoform distribution of MYC – resulting in the enhanced production of the N-terminally extended isoform MYC1. While some data suggests that MYC1 and MYC2 may differ in their transcriptional targets (70, 88), a clear understanding of the functional differences between these two proteins is lacking, both under conditions of normal physiology and malignancy. This line of examination may prove particularly interesting given that *EIF5A* is recurrently lost in several cancers.

EIF5A is located at the frequently-deleted 17p locus in the human genome that famously contains within the region, the tumor suppressor *TP53*. Approximately 50% of all human cancers have inactivated *TP53*, and a majority of those malignancies contain at least one large chromosomal 17p deletion, thus concurrently ablating *EIF5A* (89). Elegant experiments by Scott Lowe's group demonstrated that loss of function of *EIF5A* independently accelerates *MYC*-driven lymphomagenesis in a mouse model, proposing that the gene is a *bona fide* tumor suppressor and the loss of *EIF5A* is not merely a 'passenger' event associated with genetic lesions incapacitating *TP53* (89). Loss of additional members of the polyamine-hypusine pathway was shown to have the same effect, underscoring the implication that the polyamine-hypusine pathway is anti-oncogenic (8).

Here, we established that eIF5A protein serves to relieve ribosome pauses during translation elongation in human cells and that its loss results in enhanced pausing events across the transcriptome. Further, we provide evidence that these effects on pausing upon depletion of the protein are accompanied by enhanced 5' UTR translation. Strikingly, a recent report suggested that alternative translation initiation within 5' leaders of

transcripts drives oncogenesis in a model of squamous cell carcinoma (90). It is possible that the ability of tumors to adopt these alternative translation programs characterized by upstream leader translation may provide transformed cells with a competitive growth advantage. In that case, the function of *EIF5A* in maintaining appropriate start codon selection could contribute to its mechanism of tumor suppression.

In contrast with the *in vivo* reports that *EIF5A* acts as a tumor suppressor, a significant body of data suggests that flux through the polyamine pathway is tumorigenic. Polyamines levels are increased in proliferating cells. This is largely attributed to the fact that multiple enzymes in the polyamine synthesis pathway are direct transcriptional targets of *MYC* (91), including *ODC1*, the enzyme that catalyzes the first, and rate-limiting, step in polyamine synthesis. *ODC1* has also been shown to be a potent oncogene (92, 93). Lastly, overexpression of *EIF5A* and *EIF5A2* have been widely documented in several cancers (94). These inconsistencies in identifying whether *EIF5A* and the polyamine pathway act in oncogenic or tumor suppressor capacity may explain some of the difficulty in translating discoveries in this field into chemotherapies (95). We can conclude without a doubt, however, that in order to effectively manipulate this pathway in the clinic, it will be important to carefully dissect the oncogenic and tumor suppressor forces acting through it.

eIF5A: the exceptionally well conserved ribosome pause relief factor

eIF5A is well established as a translation factor that functions to relieve ribosome pauses in bacteria and yeast. Studying the effects of loss of function of eIF5A in

eukaryotes has presented technical challenges because the gene is essential, and this has contributed to the lack of data on the protein's role in translation in higher organisms. In this study, we were able to achieve complete loss of eIF5A and DOHH protein using the lentiCRISPR viral system to target these genes in human cells. We performed ribosome profiling on *EIF5A*^{-/-} and *DOHH*^{-/-} pools of cells and documented enhanced ribosome pausing under both these conditions. These experiments establish for the first time, that the molecular function of *EIF5A* is universally conserved throughout all living organisms – in both prokaryotic and eukaryotic systems. That the molecular function of *EIF5A* is so well-conserved is a direct reflection of the remarkable degree to which the homologs of these unique proteins are conserved across these organisms.

Collectively, these data suggest that the hypusine-eIF5A pathway evolved early in evolutionary time to overcome inherent challenges in protein synthesis – i.e. peptide bond formation in non-optimal amino acid contexts. We identified that these contexts include proline-proline and proline-glycine dipeptides by documenting an increase in ribosome occupancy at these sites in the human transcriptome upon depletion of *EIF5A*, or *DOHH*. These are consistent with pause site signatures observed upon performing similar loss of function experiments in bacteria and yeast (52, 57). Recent data from yeast, however, indicates that several amino acids – beyond proline and glycine – can also present challenges to efficient elongation (9, 52). The extent to which this is true in mammalian cells remains to be investigated.

Several cues on a transcript can induce pausing or stalling of the ribosome including the nascent peptide sequence, mRNA codon usage and relative abundance of

various tRNAs, and/or cis-elements that recruit specialized RNA binding proteins (76, 96). We are far from comprehensively understanding the mechanics of this important facet of translation and the full scope of eIF5A's role in it.

It is thought that a stalled ribosome is characterized by an empty E-site, from which a deacylated-tRNA has diffused away. Thus, it is conceivable (and perhaps even likely) that eIF5A binds to any stalled ribosome where peptide bond formation has slowed, and facilitates the peptidyl transferase reaction. However, it is also possible that eIF5A serves to relieve pauses induced by a subset of elongation challenges, and alternative pause relief mechanisms function on a separate subset of stalled ribosomes. Translation pause events present lucrative opportunities to regulate protein synthesis under specific physiological or pathological conditions, and advances in the field will certainly reveal exciting examples of this.

A role for eIF5A in translation initiation

In this study, we proposed a model whereby ribosome stalling at positions on transcripts proximal to non-canonical initiation sites within 5' leaders can promote enhanced alternative upstream initiation. We documented this effect using ribosome profiling data upon loss of *EIF5A*, when such pausing is enhanced across the transcriptome, in both yeast and human cells.

Further, through reporter assays we showed that in the context of *MYC*, mutating an appropriately located proline-proline pause motif and surrounding amino acids can

significantly abrogate this effect. Mutating only the proline-proline motif did not have as strong an effect, suggesting that on this transcript, surrounding amino acids are additive in challenging ribosome elongation. It is also well known that the *MYC* 5' UTR contains stable RNA secondary structures and the contribution of these elements to ribosome pausing, and subsequent effects on alternative initiation are yet to be investigated.

Several fascinating questions arise at this juncture – including whether and how ribosome queuing events upstream of stalled ribosomes mediate this link between pausing and non-canonical initiation. We hypothesize that our pause-proximal model of alternative initiation would be most relevant in circumstances where the pause site was located within a short-distance of the non-canonical initiation site, in such a way that ribosome queuing effects would be maximal. What this distance is, whether certain transcripts are likely to be more or less conducive to such ribosome queuing, and whether other proteins are involved in recognizing stalled or queued ribosomes in this context are all fascinating potential future avenues for investigation.

Summary

We designed and performed a genome-scale loss of function screen to identify novel post-transcriptional regulators of the growth-promoting gene *MYC*. We discovered that the translation elongation factor eIF5A functions to regulate protein isoform distribution of *MYC*, and that loss of function of this protein resulted in enhanced upstream translation initiation on this transcript.

Ribosome profiling experiments revealed that eIF5A's function as a ribosome pause relief factor – previously only documented in bacteria and yeast – is well conserved in mammalian cells. Further, analysis of ribosome profiling data under conditions of eIF5A depletion revealed not only evidence of enhanced pausing within coding sequences at difficult-to-translate peptide sequences, but also an increase in alternative translation initiation events in 5' UTRs in both yeast and human cells.

These data allowed us to formulate and test the hypothesis that enhanced pausing resulting from depletion of *EIF5A* enhances non-canonical translation initiation at pause-proximal upstream sub-optimal initiation codons. Thus, we propose a role for the translation elongation factor *EIF5A* in maintaining appropriate start codon selection during initiation in eukaryotic cells.

CHAPTER FOUR:

Methods

Human cell line culture

Cell lines used in this study were cultured under standard conditions recommended by ATCC, and confirmed to be free of mycoplasma contamination.

Construction of EGFP reporter constructs

The following plasmids were obtained: AAVS1-Donor-PGK-Puro plasmid (Addgene plasmid #22072), pMSCV-Hygro and pEGFP-N1, and *EGFP* reporter constructs were subsequently cloned in four steps. First, 'PGK-Puro' was excised from AAVS1-Donor-PGK-Puro by HindIII digestion, and 'PGK-Hygro' (amplified from pMSCV-Hygro with primer pair HM284/ HM285) was cloned in to replace it using the In Fusion HD Cloning Kit (Clontech, now Takara) to generate the AAVS1-Donor-PGK-Hygro plasmid. During this cloning, FseI and MluI sites were introduced 5' to the PGK promoter, and Sall, EcoRV and HindIII sites were introduced 3' of the hygromycinB resistance gene. Next, a synthetic polyA signal (97) was synthesized, and inserted downstream of the hygromycinB resistance gene using the Sall and EcoRV sites. Third, the CMV-EGFP-SV40polyA cassette was amplified from pEGFP-N1 (using primer pair HM280/ HM281) and cloned into the AAVS1-Donor-PGK-Hygro-pA plasmid using FseI and MluI sites, to generate the AAVS1-Donor-Hygro-EGFP plasmid. Lastly, the *EGFP* ORF was amplified (using the primer pair HM320/ HM321) and cloned back into this vector (AAVS-Donor-

Hygro-EGFP) using KpnI and XbaI sites to generate the final *EGFP* control reporter donor vector (AAVS-EGFP-DONOR). This was done to allow easy in-frame cloning upstream of the *EGFP* ORF in this vector.

The *MYC* 5' UTR was amplified from human genomic DNA (using primers HM322/HM324) and cloned into AAVS-EGFP-DONOR digested with KpnI using the In Fusion HD Cloning kit to generate the *MYC* 5' UTR reporter donor plasmid (AAVS-5' UTR-EGFP-DONOR).

Generation of clonal fluorescent reporter cell lines

A published TALEN pair targeting the human *AAVS1/PPP1R12C* locus (98) were used, in combination with either the *EGFP* control reporter donor vector (AAVS1-EGFP-DONOR) or the *MYC* 5' UTR reporter donor vector (AAVS1-5' UTR-EGFP-DONOR) to generate the following *EGFP* knock-in cell lines stably expressing the corresponding reporters – HCT116^{*MYC* 5' UTR} and HCT116^{*EGFP*}. The three plasmids were mixed in the following molar ratio – Left-TALEN : Right-TALEN : HDR-donor = 1:1:8, and transfected into HCT116 cells using FugeneHD transfection reagent (Promega Corporation).

Two days post-transfection, the cells were split into 0.5mg/ml hygromycinB to allow for selection of cells that underwent successful recombination. The cells were selected for at least 7 days, before plating into 96-well plates at single cell density. Colonies arising

from single cells were picked and expanded, and genomic DNA was extracted using the DNEasy Blood and Tissue Kit (Qiagen). Clonal cell lines were then genotyped by PCR to detect the *AAVS1* wild type and the *EGFP* knock-in alleles, using the primer pairs HM356/HM357 and HM356/ HM359 respectively. Heterozygous knock-in clones were used for all experiments.

Genome-wide CRISPR/Cas9 screening

Fluorescent-reporter coupled CRISPR/Cas9-mediated loss of function screening was carried out as previously described in detail (63). A brief overview is provided here.

The human GeCKO v2 libraries A and B (Addgene pooled libraries ##1000000048, #1000000049) were amplified in bacteria and purified, and then used to prepare lentiviral library in HEK293T cells. Genome-wide screening was performed in the HCT116^{MYC5' UTR} and HCT116^{EGFP} cell lines, in replicate for each library. Reporter cells were transduced with lentivirus at a multiplicity of infection between 0.2 – 0.4. Two days later, the cells were seeded into 1µg/ml puromycin to select for cells that had been successfully transduced. After passaging for 13 days in puromycin, cells were subjected to FAC sorting on a MoFlo cell sorter (Beckman Coulter). The brightest fraction of cells – corresponding to 0.5% of the population as assessed by EGFP fluorescence – were collected by FACS, pelleted and frozen at -80°C.

gRNA representation was maintained at 500X during infection and at all subsequent stages of screening. 120 x10⁶ cells were infected per library and at least 100

10^6 cells were maintained per library during selection. 60-80 10^6 cells were sorted per replicate per library, and 1.5×10^5 – 2.5×10^5 cells were collected post sorting. 40×10^6 unsorted cells were also pelleted and frozen at -80°C .

Genomic DNA was extracted from the sorted and unsorted cell pellets using either a gDNA isolation method that has been previously outlined (63) for the sorted cells or the DNEasy Blood and Tissue Kit (Qiagen) for the unsorted cells. Two sequential rounds of PCR were performed using the isolated genomic DNA as template and the primers listed in Table 5 (HM400/ HM401: PCR1 reaction; HM415-50/ HM414: various PCR2 reactions) to generate PCR amplicon libraries. Library DNA was then purified using AMPure XP beads (Agencourt).

DNA concentration of the libraries was quantified by using the Qubit dsDNA BR assay kit (Thermo Fisher Scientific), as well as by qPCR using the KAPA Library Quantification Kit for Illumina platforms. Amplicon library size and integrity were also assessed via the Bioanalyzer High Sensitivity DNA Analysis Kit (Agilent). Following these quality checks, library amplicons were sequenced on an Illumina HiSeq 2500 or a NextSeq 500 with 75 bp single-end reads. Approximately 20×10^6 reads were obtained per sample, and data was analyzed as described previously (63). In our screening data, we detected read contamination (arising from cloning of individual gRNAs into the LCv2 vector) from gRNAs targeting *NFE2L2* and thus, this gene was removed from our analysis.

LentiCRISPR knockout of genes

Generation of lentiCRISPR virus

gRNAs used to target genes of interest, as well as non-target (NT) gRNAs were cloned into the lentiCRISPRv2 (LCv2) vector (Addgene plasmid #52961) according to the LCv2 cloning protocol (99). All gRNAs used in this study are listed in Table 6.

LCv2-gRNA virus was generated in HEK293T cells as follows. Cells were seeded at a density of 600,000 cells per well of a 6-well dish. The next day, cells were transfected with 0.5ug LCv2-gRNA plasmid, 0.3ug psPAX2 (Addgene #12260) and 0.2ug pMD2.G (Addgene #12259) lentiviral packaging plasmids using FugeneHD transfection reagent (Promega Corporation). Media was changed 24 hours post-transfection, and viral collection was performed on the two subsequent days after that. Media containing virus was filtered through a 0.45 μm SFCA sterile filter and either utilized immediately for transduction or frozen at -80°C in single use aliquots.

Infection of cells and generation of knockout pools

Cells to be infected were seeded in fresh media containing virus at a ratio of 1:1, and polybrene (EMD Millipore) at a final concentration of 8 $\mu\text{g}/\text{ml}$. Cells were infected at densities such that they achieved confluency 48 hours post-transduction. Media was changed the day after infection, and the following day, cells were split into media containing puromycin at a concentration of 1 $\mu\text{g}/\text{ml}$.

For all experiments using *EIF5A*^{-/-} and *DOHH*^{-/-} pools, cells transduced with the appropriate LCv2-gRNA virus were harvested for analysis after 6 days of selection in

puromycin. Control pools used in these experiments were cells transduced with LCv2-NT gRNA virus, selected in puromycin for 6 days and harvested in the same manner as *EIF5A*^{-/-} and *DOHH*^{-/-}.

Western blotting analysis

Cell lysates were prepared by harvesting in ice cold 1X RIPA buffer. Proteins were visualized using an infrared fluorescent antibody detection system (LI-COR). Antibodies used for Western blotting in this study include MYC (Cell Signaling, #5605), α -TUB (Millipore Sigma, #T6199), eIF5A (Abcam, #ab32407), DOHH (Millipore Sigma, #HPA041953), Hypusine (Millipore Sigma, #ABS1064), phospho-eIF2 α (Cell Signaling, #9721) and total-eIF2 α (Cell Signaling, #9722).

RNA isolation and qPCR

Cells at subconfluent densities (~80% confluent) were harvested in Trizol (Thermo Fisher Scientific). RNA was isolated using the Direct-zol RNA miniprep kits (Zymoresearch). The on-column DNase digestion step was performed according to manufacturer's instructions.

SuperScript III First-strand Synthesis SuperMix (Thermo Fisher Scientific) was used to synthesize cDNA from RNA, using 1ug of RNA per reaction. All qPCR assays were performed in technical triplicate, using the SybrGreen 2X PCR Master Mix (Applied Biosystems). Primers used for qRT-PCR are listed in Table 5.

Indel analysis

Generation of MYC1 KO pools and siRNA treatment

HCT116 cells were infected with LCv2-gRNA virus selectively targeting MYC1 (Table 6) and selected in puromycin as described above. On the third day of selection, cells infected with LCv2-MYC1 gRNA and LCv2-NT gRNA were reverse transfected with siRNA (either targeting *EIF5A*, or NT controls) at a concentration of 20nM, using Lipofectamine RNAiMAX (Thermo Fisher Scientific) according to the manufacturer's instructions. Cells were harvested for gDNA isolation and Western blot analysis at 72 hours post-transfection of siRNA (corresponding to 6 days of puromycin selection).

Amplicon generation for next-generation sequencing

Genomic DNA was isolated from cells treated with NT siRNA, and either LCv2-MYC1 gRNA or LCv2-NT gRNA, using the DNEasy Blood and Tissue Kit (Qiagen). Amplicon libraries (of the gRNA target site) were generated by two sequential rounds of PCR using Phusion High-Fidelity DNA Polymerase (NEB). PCR1 was performed using gene specific primers HM685/ HM686, and PCR2 was performed using Illumina TruSeq CD Indexes D508 (Fwd)/ D706-709 (Rev).

Ribosome profiling

HCT116 cells were transduced with LCv2-*EIF5A*, *DOHH* or NT gRNAs and selected for 6 days in puromycin as described above to generate knockout pools. Cells

were harvested and ribosome profiling was performed as described (100), with the following modifications.

1. Samples were not pooled after linker ligation, and the subsequent steps were performed individually for each sample.

2. In order to effectively separate 3' linker-ligated RNA fragments from unligated linker, we utilized gel electrophoresis, followed by gel extraction as described in an earlier version of the protocol (101), instead of an enzymatic depletion.

Bioinformatic analysis of ribosome profiling data

Read mapping

For analysis of reads from the human ribosome profiling dataset, first, the adapter, inline barcode and the random-mer incorporated into the library amplicon were trimmed for each read. Then, reads from non-coding RNAs were removed by mapping to known rRNA (102), tRNA (103) and snRNA (104) sequences using HISAT2 (105). Reads that did not map to these databases were subsequently mapped to the human reference genome (GRCh38) (106) using HISAT2. Only reads between 24 – 31nt in length were used for all final analyses. Annotation of transcripts was based on GENCODE v27 (106) and for each gene with multiple isoforms, only the longest transcript was used.

Reads from yeast ribosome profiling experiment were processed as follows. First, the adapter was trimmed for each read. The reads with quality score less than 20 at any

position were dropped. The remaining reads were mapped to known non-coding RNAs using HISAT2, and the remaining unmapped reads (corresponding to reads from mRNAs) were then mapped to the yeast reference genome (107) using HISAT2. Only reads between 25 – 34nt were used for all final analyses. Annotation of transcripts was based on the dataset from Nagalakshmi et al. (2009) (108).

Periodicity analysis (human dataset)

For reads with the 5' end mapped to the first nucleotide of a codon, the P-site was estimated as nucleotides 13-15 from the 5' end of each mapped read.

Read depth at each position was normalized to total number of reads per sample (in millions) to yield reads per million mapped reads (RPM). Normalized read depth (i.e. RPM) at each position in the transcript was aggregated across all transcripts.

Metacodon plots (human dataset)

Proline-proline ('PP') motifs from all CDS regions were considered, and those located before and within 50nt of another 'PP' motif on the same strand were discarded. The RPM values were aggregated at each position within a 100nt region centered at each individual 'PP' motif.

A similar analysis was performed to generate metacodon plots centered on 'pause motifs' by using tri-amino acid motifs instead of 'PP'.

Pause motif analysis (human dataset)

The occurrence of each tri-amino acid motif was calculated across all CDSs covered by at least 128 reads in the RNA-seq. Only motifs that appeared at least 10 times in the qualified CDSs were then considered for this analysis.

For each tri-amino acid motif, if the average RPM (across biological replicates) from ribosome profiling data across all CDSs was greater than 10, then the 'pause score' of this motif was calculated as the ratio of average RPM value between *EIF5A*^{-/-} and NT samples, or *DOHH*^{-/-} and NT samples. A position weighted matrix was then constructed according to the pause score for each tri-amino acid motif.

CDF plots (human and yeast datasets)

For CDF plots generated from human ribosome profiling data, genes that met the following coverage criteria were used in the analysis.

1. In the ribosome profiling dataset, genes were covered by at least 128 reads in the CDS and at least 16 reads in 5' UTR region in both WT samples or in both eIF5A samples
2. In the RNAseq dataset, at least 128 reads mapped to the transcript in both WT samples or in both eIF5A samples

For CDF plots generated from yeast ribosome profiling data, coverage constraints used for genes included in the analysis are as follows.

1. In the ribosome profiling dataset, the CDS was covered by at least 64 reads in both WT samples or in both eIF5A samples

2. In the ribosome profiling dataset, 5' UTR were covered by at least 6 reads in both WT samples or in both eIF5A samples

For analysis of ribosome occupancy on CDSs with $<$ or \geq 2 'PP' pause sites, ribosome occupancy was first estimated for the CDS of each gene from the ribosome profiling data ($RPKM_{RPF}$). FPKM for each transcript was calculated from the RNAseq dataset ($FPKM_{RNA}$). Normalized ribosome occupancy for each transcript was computed ($RPKM_{RPF}/ FPKM_{RNA}$). The ratio of this normalized ribosome occupancy between EIF5A^{-/-} and NT samples was calculated and used as the cumulative distribution function in the plot.

In assessing 5' UTR translation in the yeast data, 5' UTR translation was calculated as ribosome occupancy in the 5' UTR of a gene, as a fraction of ribosome occupancy in the CDS of the gene ($RPKM_{5'UTR}/ RPKM_{CDS}$). For human samples, 5' UTR translation was estimated as ribosome occupancy in the 5' UTR of a gene, normalized to transcript abundance ($RPKM_{RPF}/FPKM_{RNA}$). The difference in computing 5' UTR translation arise from the fact that RNAseq data complementary to the yeast ribosome profiling dataset were not available. 5' UTR translation was plotted for EIF5A^{-/-} and WT/NT conditions.

Arsenite treatment

800,000 HCT116 cells were seeded in each well of a 6-well dish. The following day, the cells were treated with 250 μ M sodium arsenite for either 15minutes, 30minutes,

1 hour, 3 hours or 6 hours. Cells were then harvested and lysed in ice cold, 1X RIPA buffer for analysis by Western blotting.

Polyamine treatment

4 million HEK293T cells were seeded on a 10cm plate in 2.5mM Difluoromethylornithine (DFMO, Cayman Chemical, #96020-91-6) to inhibit ornithine decarboxylase (ODC). Cells were grown in DFMO and media was changed daily. After 4 days in DFMO, cells were treated with varying doses (2nM, 4nM and 6nM) of spermidine (Millipore Sigma, #S2626-1G) and 1mM Aminoguanidine hydrochloride (AGH, Millipore Sigma, #396494). Aminoguanidine is routinely added to in vitro experiments modulating polyamine levels because it protects cells against extracellular polyamine toxicity by inhibiting amino-oxidases in serum that generate reactive oxygen species (109).

Cells were harvested for Western blot analysis the following day. Experimental regimens and protocols followed for these experiments were identical to those reported in Ivanov et al. (2018) (77).

Luciferase assays

Construction of luciferase reporters

The pGL3-Control (Promega) plasmid was obtained, and the firefly luciferase ORF was modified to contain no N-terminal/ proximal pause sites (pGL3-Control_NP2). This was done in two steps – first, a ‘proline-glycine’ dipeptide (aa 37-38) was mutated to

'proline-alanine' using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) and the primer pair HM673/ HM674 to generate plasmid pGL3-Control _NP1. Subsequently, a 'proline-proline' dipeptide (aa 173-174) in the luciferase ORF in pGL3-Control_NP1 was mutagenized to 'proline-alanine' using the same kit and primers HM677/ HM678 to generate pGL3-Control_NP2. In order to remove any unwanted mutations that may have been incorporated into this plasmid during the two rounds of mutagenesis, the mutagenized region (obtained by the digestion of pGL3-Control_NP2 with HindIII and XbaI) was subcloned into pGL3-Control digested with the same enzymes. The region was then verified by Sanger sequencing to contain only the expected mutations.

PGL3-Control_NP2 was digested with HindIII and NcoI and all the MYC 5' UTR reporter fragments were cloned into this backbone using the NEBuilder HiFi DNA Assembly Master Mix (NEB). The MYC 5' UTR reporter fragments used in the assembly were designed to contain a fragment of the MYC transcript (as depicted in Figure 36), the P2A sequence and appropriate homology arms to facilitate HiFi assembly, and were obtained as gBlocks synthesized by IDT. The gBlock sequences are listed in Table 5.

Luciferase assays

HCT116 cells were infected with LCv2-*EIF5A* or LCv2-*NT* virus as described above. On the fourth day of selection in puromycin, cells were plated in 24-well dishes at densities such that they would be 50% confluent in 24 hours. The following day, *EIF5A*^{-/-}

and NT control cells were transfected with luciferase reporters using FugeneHD transfection reagent (Promega Corporation). 300ng of DNA was transfected/ well – comprising 2 ng of pRL-SV40 (Promega) as a transfection control, 198 ng of empty pcDNA3.1+ and 100 ng of a pGL3-5' UTR reporter. Cells were harvested 24 hours later to assay luciferase enzyme activity using a Dual-Luciferase Reporter Assay System (Promega). Each transfection was performed in biological duplicate.

APPENDIX

Table 1: Top 50 genes by RIGER analysis of CRISPR-Cas9 screen in HCT116^{MYC 5' UTR} cells

Gene	Results from MYC 5' UTR reporter screen			Gene rank in the EGFP control reporter screen
	Gene rank	Normalized enrichment score	Neg log(p value)	
TAF7	3	0.0004657	6	171
TAF1	2	0.0002495	6	13
ASH2L	1	0.00009979	6	7
EIF3L	5	0.0007152	5.698970004	14551
TAF5	4	0.0006487	5.698970004	10
CMIP	6	0.001863	4.958607315	218
BRD4	7	0.002378	4.795880017	1
SSBP2	8	0.002528	4.744727495	16939
DBR1	9	0.002944	4.48148606	1864
CELF6	10	0.002994	4.468521083	3149
RPL10A	11	0.003127	4.420216403	2262
COASY	12	0.005023	4.045757491	1038
EIF5AL1	13	0.006354	3.853871964	1239
FAM189A1	14	0.007651	3.673664139	11707
PTOV1	15	0.007667	3.673664139	1037
CREBBP	16	0.007734	3.66756154	12
RPL10	17	0.007751	3.661543506	1446

BHLHB9	18	0.008216	3.616184634	890
CAPRIN1	19	0.009214	3.519993057	7652
XRCC3	20	0.009347	3.504455662	1603
GPKOW	21	0.009497	3.488116639	563
OR5L2	22	0.00958	3.480172006	14678
CXXC1	23	0.009896	3.446116973	191
DPY30	24	0.01035	3.419075024	9
FGF10	25	0.01058	3.392544977	754
HIST1H2AK	26	0.01116	3.341035157	1077
HIST1H4I	27	0.01124	3.332547047	17156
IGFBP2	28	0.01128	3.331614083	10843
RAB3GAP1	29	0.01133	3.327902142	6417
EIF3H	30	0.01138	3.326058001	4699
STAT2	31	0.01164	3.310691141	16397
HNRNPUL1	32	0.01221	3.272458743	14034
TUBA1A	33	0.01364	3.171984936	6846
NUBP1	34	0.01375	3.166215625	4081
RPLP0	35	0.01387	3.155522824	12892
EID2B	36	0.01435	3.130181792	13052
C1orf65	37	0.01479	3.097997109	1438
LRRC37A	38	0.01535	3.059483515	706
IL12B	39	0.01595	3.022276395	3558
RIBC1	40	0.01625	3.003050752	1511

hsa-mir-146b	53	0.01933	2.996970529	5320
SLC35E2	58	0.01982	2.974694135	20137
hsa-mir-377	59	0.0201	2.960982678	18257
TAF8	41	0.01721	2.953504836	6
SULT1A2	42	0.01775	2.930331903	44
AHCY	43	0.01786	2.925183559	3930
CNOT11	44	0.01798	2.918652692	4075
RBBP4	45	0.01826	2.90517962	311
TAF12	46	0.01833	2.901356274	21
CCT4	47	0.01845	2.894489815	7715

Table 2: Top 50 genes whose 5' UTR translation is increased in *EIF5A*^{-/-} yeast, compared to WT yeast

Gene (Systematic Name)	Gene (Standard name)	5' UTR/ CDS ribosome occupancy (RPKM; average across replicates)		
		WT	eIF5A knockdown (kd)	eIF5A kd / WT
YOR222W	ODC2	0.223795	64.76737	8.176945548
YGR286C	BIO2	0.02132	2.45269	6.846013652
YPL145C	KES1	0.0084	0.25746	4.937815169
YMR006C	PLB2	0.05281	1.271795	4.58991119
YAL017W	PSK1	0.038195	0.898215	4.555605118
YGL206C	CHC1	0.02505	0.553775	4.466417417
YLR130C	ZRT2	0.116835	2.498705	4.418636156
YDR096W	GIS1	0.094365	1.905645	4.335883713
YBR208C	DUR1,2	0.15996	3.21498	4.329026671
YGR089W	NNF2	0.062205	1.24904	4.327645321
YNL046W		0.230815	4.608765	4.319571318
YFR030W	MET10	0.300205	5.926375	4.303130007
YGR061C	ADE6	0.01407	0.267555	4.249141256
YBR023C	CHS3	0.11054	2.05061	4.213412713
YKL125W	RRN3	0.06003	1.08506	4.175947343
YDR414C	ERD1	0.199495	3.358415	4.073356022
YFR021W	ATG18	0.0302	0.504895	4.063362936

YHR159W	TDA11	0.414085	6.85182	4.048488404
YDR395W	SXM1	0.022195	0.36371	4.034481979
YER100W	UBC6	0.008285	0.12886	3.959158992
YOR108W	LEU9	0.04489	0.697545	3.95782029
YGR184C	UBR1	0.166795	2.590705	3.9571968
YDR044W	HEM13	0.01323	0.20238	3.935181757
YDR306C		0.167985	2.5241	3.909364748
YPL204W	HRR25	0.01417	0.209905	3.888824869
YJL156C	SSY5	0.100015	1.46387	3.871499147
YBR121C	GRS1	0.049145	0.719235	3.871346678
YJR103W	URA8	0.014605	0.21275	3.864624869
YDL132W	CDC53	0.04505	0.65199	3.855250826
YDL193W	NUS1	0.0373	0.535545	3.843760269
YPL019C	VTC3	0.040355	0.57814	3.840599552
YDL047W	SIT4	0.02797	0.39487	3.819425605
YBL030C	PET9	0.008395	0.11834	3.817263665
YKL024C	URA6	0.13687	1.885635	3.784172275
YMR043W	MCM1	0.003565	0.04872	3.772540151
YBL061C	SKT5	0.016855	0.22753	3.754808246
YBR154C	RPB5	0.018745	0.251905	3.748302027
YJR139C	HOM6	0.00874	0.11706	3.743471094
YJR116W	TDA4	0.05481	0.731335	3.738021368
YPR072W	NOT5	0.36188	4.755495	3.716012238

YDR017C	KCS1	0.04559	0.595815	3.70807513
YJL130C	URA2	0.00932	0.121795	3.707981143
YDR071C	PAA1	0.013245	0.17176	3.696874347
YCR069W	CPR4	0.012245	0.158765	3.696628224
YGR191W	HIP1	0.02025	0.262185	3.694591336
YKL210W	UBA1	0.00455	0.058695	3.689299161
YPL226W	NEW1	0.01103	0.141325	3.679512001
YDR480W	DIG2	0.02827	0.358575	3.664931118
YGR007W	ECT1	0.254925	3.231345	3.663990023
YLL040C	VPS13	0.07065	0.890005	3.655051975

Table 3: Top 50 genes whose 5' UTR translation is increased in *EIF5A*^{-/-} human HCT116 cells, compared to control NT cells

Gene name	5' UTR ribosome occupancy/ RNA abundance (RPKM/ FPKM; average across replicates)				
	<i>EIF5A</i> ^{-/-}	<i>DOHH</i> ^{-/-}	NT control	<i>EIF5A</i> ^{-/-} / NT	<i>DOHH</i> ^{-/-} / NT
RNASEH1	0.452367	0.3718045	0.0187425	4.593108382	4.310158924
CPEB4	0.709024	0.6406815	0.042113	4.073496905	3.927269775
PLOD3	4.2762395	2.688502	0.3083025	3.793924163	3.124384052
TOB1	1.6849755	0.8072295	0.142514	3.563552058	2.501875248
ZKSCAN8	9.6245565	3.909	0.8161895	3.559744007	2.259823529
CCDC94	7.7062095	9.442348	0.6930155	3.475061882	3.76818613
BCS1L	1.8476495	0.5741215	0.1853695	3.31721531	1.630952196
TBL1XR1	0.644006	0.294279	0.065851	3.289796875	2.15990734
PIK3R4	1.2300865	2.1122815	0.1328995	3.210352188	3.990394531
SEZ6L2	3.708864	4.272766	0.4448195	3.059685427	3.26387837
SLC23A2	2.4077995	1.505146	0.2961865	3.023137472	2.345325646
ITGA2	4.938942	1.570955	0.6217115	2.989884856	1.337324685
VCPIP1	1.5858365	1.4495805	0.213687	2.89167299	2.762064406
KLHL8	0.963515	0.6680425	0.1308915	2.879935714	2.351568475
MTUS1	0.8717795	0.5651485	0.124914	2.8030281	2.177694825
NCOA3	1.5044615	0.6294895	0.232712	2.692629676	1.435636707
IGSF3	2.713648	1.969959	0.4199535	2.691932097	2.229864107
CHMP7	0.5342565	0.273916	0.0831295	2.684100116	1.7203011

TMEM5	1.93136	1.0277385	0.3085815	2.645893626	1.735749751
SREBF1	3.4103395	1.6986835	0.5473265	2.639441752	1.633943459
HNRNPUL2	7.486751	10.6309105	1.208938	2.630599515	3.136453
PSMC3	0.302709	0.5087175	0.0504345	2.585448709	3.334381864
ERP44	2.643701	0.720728	0.460037	2.522737214	0.647704995
MTR	2.9572165	5.094417	0.5441175	2.442249734	3.226926921
TICAM1	4.2489435	2.136187	0.8347855	2.347626713	1.355560498
BAG3	2.4960385	2.2958345	0.504698	2.306147912	2.185526371
MAT2A	2.521692	3.333834	0.5140525	2.294404461	2.697194657
TBP	3.971817	1.550749	0.8181895	2.279292225	0.922458267
PPAT	1.2955345	0.633685	0.2674645	2.276128111	1.244418447
DGCR8	4.0906325	2.324405	0.860543	2.249004744	1.433542275
MLPH	3.5011535	1.1240975	0.7530115	2.217086512	0.578023372
DHRS3	2.3859555	1.33026	0.5239715	2.187006888	1.344148001
GFPT1	1.116017	1.5900235	0.25051	2.155418904	2.666107988
FAM216A	4.013889	1.1754555	0.9142875	2.13428092	0.362500122
EIF3D	0.160247	0.231224	0.0374555	2.097047862	2.626041665
FAR1	0.9694845	0.642354	0.2370475	2.032041655	1.438192405
FBL	0.3415325	0.4165665	0.083899	2.025297351	2.311821303
PSMD12	1.0707765	0.3835175	0.2675255	2.000909064	0.519615997
SYTL3	0.8713745	0.5644995	0.218694	1.994379255	1.368058663
TBC1D30	3.972826	3.472899	1.0009285	1.988826687	1.794801534

NEDD1	1.8799715	1.244722	0.4749595	1.984834387	1.389947158
GPT2	2.793437	1.2286975	0.7067215	1.982827579	0.797916069
TUBB4B	1.3117005	0.992245	0.3383975	1.954647532	1.551977477
WARS	0.098756	0.211079	0.0260825	1.920786245	3.016628991
PTPRA	1.010774	0.806642	0.2730305	1.888326432	1.562866403
PES1	1.401261	1.5350415	0.3785765	1.888068934	2.019620895
PRKD2	5.3677175	5.278192	1.4793335	1.859361416	1.835096502
SAMD1	2.871467	0.7299695	0.7982935	1.846796813	- 0.129083078
TBCD	15.3405045	10.574972	4.274325	1.843577416	1.30688533
AHCY	0.300118	0.3085815	0.085318	1.814607797	1.85472952
SSX2IP	0.5104895	0.355775	0.1485635	1.780801578	1.259885428
TFAP4	2.799803	1.423304	0.8211945	1.76952945	0.793447966

Table 4. Oligonucleotide sequences used in this study

Name	Sequence	Description
HM284	acagtggggcaagctGGCCGGCCTAACG CGTctaccgggtaggggaggcgcttttc	Cloning <i>EGFP</i> reporter constructs_PCR amplification of PGK-Hygro (Fwd)
HM285	gtccctagtaaagctTAGATATCGAGTCGA Cctattcctttgccctcggacgagtg	Cloning <i>EGFP</i> reporter constructs_PCR amplification of PGK-Hygro (Rev)
HM280	ATGCTGGCCGGCCTTAATAGTAATC AATTACGG	Cloning <i>EGFP</i> reporter constructs_PCR amplification of CMV- <i>EGFP</i> -pA (Fwd)
HM281	ATGCTACGCGTAAGATACATTGATG AGTTTG	Cloning <i>EGFP</i> reporter constructs_PCR amplification of CMV- <i>EGFP</i> -pA (Rev)
HM320	ATGCAggtaccATGGTGAGCAAGGGC GAGGAG	Cloning <i>EGFP</i> reporter constructs_PCR amplification of <i>EGFP</i> (Fwd)
HM321	TATGATCTAGAGTCGCGGCCGC	Cloning <i>EGFP</i> reporter constructs_PCR amplification of <i>EGFP</i> (Rev)
HM322	ATTCTGCAGTCGACGgacccccgagctgt gctg	Cloning <i>EGFP</i> reporter constructs_PCR amplification of human MYC 5' UTR (Fwd)
HM324	GCCCTTGCTCACCATcgtcgcgggaggct gctggtttccactaccg	Cloning <i>EGFP</i> reporter constructs_PCR amplification of human MYC 5' UTR (Rev)
HM356	CTCTCCTGAGTCCGGACCACTTTG	Genotyping of AAVS1 WT/ knock-in <i>EGFP</i> allele (Fwd)
HM357	CAAGCTCTCCCTCCCAGGAT	Genotyping AAVS1 WT allele (Rev)
HM339	ATGGGCTATGAACTAATGACCCC	Genotyping of <i>EGFP</i> knock-in allele at AAVS1 locus (Rev)

HM400	AATGGACTATCATATGCTTACCGTA ACTTGAAAGTATTTTCG	NGS library preparation for CRISPR screening_PCR1 (Fwd)
HM401	TCTACTATTCTTTCCCTGCACTGTt gtgggcatgtgctctg	NGS library preparation for CRISPR screening_PCR1 (Rev)
HM414	CAAGCAGAAGACGGCATAACGAGAT GTGACTGGAGTTCAGACGTGTGCT CTTCGGATCTtctactattctttcccctgcactgt	NGS library preparation for CRISPR screening_PCR2 (Rev)
HM415	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTtACGACTCTtcttgtaaag gacgaaacaccg	NGS library preparation_PCR2 (various Fwds)
HM417	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTgatACGACGAGtcttgtaa aaggacgaaacaccg	
HM419	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTtcatACGCATAtcttgtaa aaggacgaaacaccg	
HM421	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTgatgatACGCGACAtcttg gaaaggacgaaacaccg	
HM423	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTacgatgatACGTAGACTctt gtgaaaggacgaaacaccg	
HM424	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTtACGTATCActtgtaaag gacgaaacaccg	
HM425	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTatACGCTAGTtcttgtaa ggacgaaacaccg	

HM426	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTgatACGTGCTAtcttggaa aggacgaaacaccg
HM439	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTcgatGACTACTtcttggaa aaggacgaaacaccg
HM441	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTatcgatGACTGCTGtcttgg gaaaggacgaaacaccg
HM443	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTcgatcgatGCAGATAccttgg tggaaaggacgaaacaccg
HM445	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTtGTGTCACAtcttggaaag gacgaaacaccg
HM447	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTgatGCTACTGAtcttggaa aggacgaaacaccg
HM448	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTcgatGAGAGCTCtcttgg aaaggacgaaacaccg
HM449	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTtcgatGCATACTGtcttgg aaaggacgaaacaccg
HM450	AATGATACGGCGACCACCGAGATC TACTCTTTCCCTACACGACGCTC TTCCGATCTatcgatGATCATAGtcttgg gaaaggacgaaacaccg

GS105	CAGTAAGTGCGGGTCATAAGC	qPCR_human 18S (Fwd)
GS106	CAAGTTCGACCGTCTTCTCAG	qPCR_human 18S (Fwd)
HM368	CACCACCAGCAGCGACTCT	qPCR_human MYC (Fwd)
HM369	CTTTTCCACAGAAACAACATCGAT	qPCR_human MYC (Rev)
HM485	GGCAGATGACTTGGACTTCGA	qPCR_human EIF5A (Fwd)
HM486	CCCAGTAAAGATGTCAATACCAACC	qPCR_human EIF5A (Rev)
HM489	GGAGTGTCTTGTGAGGTGCA	qPCR_human EIF5AL1 (Fwd)
HM490	AGAAGCTTCACACGCACTCA	qPCR_human EIF5AL1 (Rev)
HM685	ACACGACGCTCTTCCGATCTGGCA CTTTGCACTGGAACCTTAC	Genotyping MYC1-gRNA targeting site (Fwd)
HM686	CAGACGTGTGCTCTTCCGATCTCAA GTGGACTTCGGTGCTTAC	Genotyping MYC1-gRNA targeting site (Rev)
HM673	GATACGCCCTGGTTCCTGCAACAAT TGCTTTTACAG	Mutagenesis_to generate pGL3- Control_NP1 (Fwd)
HM674	CTGTAAAAGCAATTGTTGCAGGAAC CAGGGCGTATC	Mutagenesis_to generate pGL3- Control_NP1 (Rev)
HM677	CGTCACATCTCATCTACCTGCCGGT TTTAATGAATAC	Mutagenesis_to generate pGL3- Control_NP2 (Fwd)
HM678	GTATTCATTAACCGGCAGGTAGA TGAGATGTGACG	Mutagenesis_to generate pGL3- Control_NP2 (Rev)

gRNA sequences

Name	Sequence	Details
EIF5A gRNA #1 (EIF5A+ EIF5AL1) gRNA #1	AGAGGACCTTCGTCTCCCTG	
EIF5A gRNA #2	TACATACAGGTCCATCTGGT	EIF5A-specific
EIF5A gRNA #3	GTGAAATTCTAACCTTGGCG	EIF5A-specific

EIF5A gRNA #4	TCCTGGATGCCAATCAGCTG	
(EIF5A+ EIF5AL1) gRNA #2	ACATCCATATTATGAGTTGA	
EIF5A gRNA #5	TCTCGACGATCTTACATGGC	EIF5A-specific
EIF3L gRNA #1	CACCTACCATTGCCAACCTG	
EIF3L gRNA #2	ACTTCTTGGCAGTCTTACAG	
EIF3L gRNA #3	CAAGAATACACCTTGGCCCG	
RPL10 gRNA	TATGAGCAGCTGTCCTCTGA	
RPL10A gRNA	TAAGTTCTCTGTGTGTGTCC	
MYC1 gRNA	CCTTGCAGCTGCTTAGACGC	
EIF5AL1 gRNA #1	AGGCTGGCCATGTAAGATCG	
EIF5AL1 gRNA #2	CAACCAGATGGACCTTGGCG	
DOHH gRNA #1	GCGCGGTATCGCTCGAAGAG	
DOHH gRNA #2	GCCTGGTTACCTCGATGACG	
NT gRNA #1	ATCGTTTCCGCTTAACGGCG	
NT gRNA #2	TAGAGATATCCGATCGTGGT	
gBlocks		
Name	Sequence (left homology arm, MYC transcript fragment, P2A)	
WT reporter (CTG)	aggcctaggcttttgcaaaaagcttgacgctgcttagacgctggattttttcgggtagtgaaaac cagcagcctcccgcgacggtttcccctcaacgtagctcaccacaggaactatgacctcgacta cgactcgggtgcagccgattttctactgcgacgaggaggagaacttctaccagcagcagcagca gagcgagctgcagccccggcgcccagcagggatatctggaagaaattcgagctgctgccc ccccggccctgtcccctagccgcccgtccgggGGATCCGGCGCAACAACTTCT CTCTGCTGAAACAAGCCGGAGATGTCTGAAGAGAATCCTGGACCG GAAGACGCCAAAAACATAAA	
WT reporter (ATG)	aggcctaggcttttgcaaaaagcttgacgctgcttagacgAtggattttttcgggtagtgaaaac cagcagcctcccgcgacggtttcccctcaacgtagctcaccacaggaactatgacctcgacta	

BIBLIOGRAPHY

1. R. Nogales-Cadenas *et al.*, GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information. *Nucleic Acids Research* **37**, W317-W322 (2009).
2. D. Tabas-Madrid, R. Nogales-Cadenas, A. Pascual-Montano, GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. *Nucleic Acids Research* **40**, W478-W483 (2012).
3. P. Carmona-Saez, M. Chagoyen, F. Tirado, J. M. Carazo, A. Pascual-Montano, GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biology* **8**, R3 (2007).
4. B. T. Sherman, D. W. Huang, R. A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**, 1-13 (2008).
5. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44 (2008).
6. Z. Zhang, F. S. Dietrich, Identification and characterization of upstream open reading frames (uORF) in the 5' untranslated regions (UTR) of genes in *Saccharomyces cerevisiae*. *Current Genetics* **48**, 77-87 (2005).
7. E. Gutierrez *et al.*, eIF5A Promotes Translation of Polyproline Motifs. *Molecular Cell* **51**, 35-45 (2013).
8. C. Scuoppo *et al.*, A tumour suppressor network relying on the polyamine–hypusine axis. *Nature* **487**, 244 (2012).
9. A. P. Schuller, C. C.-C. Wu, T. E. Dever, A. R. Buskirk, R. Green, eIF5A Functions Globally in Translation Elongation and Termination. *Molecular Cell* **66**, 194-205.e195 (2017).
10. F. Crick, Central Dogma of Molecular Biology. *Nature* **227**, 561-563 (1970).
11. F. Reiter, S. Wienerroither, A. Stark, Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics & Development* **43**, 73-81 (2017).
12. T. Fujiwara, A. Fukao, The coupled and uncoupled mechanisms by which trans-acting factors regulate mRNA stability and translation. *The Journal of Biochemistry* **161**, 309-314 (2016).

13. K. C. Martin, A. Ephrussi, mRNA Localization: Gene Expression in the Spatial Dimension. *Cell* **136**, 719-730 (2009).
14. C. C. Thoreen *et al.*, A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109 (2012).
15. A. G. Hinnebusch, I. P. Ivanov, N. Sonenberg, Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* **352**, 1413 (2016).
16. L. W. Barrett, S. Fletcher, S. D. Wilton, Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences* **69**, 3613-3634 (2012).
17. D. P. Bartel, MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**, 215-233 (2009).
18. G. Cooper. (Sunderland (MA): Sinauer Associates, 2000), chap. Translation of mRNA.
19. T. E. Dever, I. P. Ivanov, Roles of polyamines in translation. *Journal of Biological Chemistry* **293**, 18719-18729 (2018).
20. V. Ramakrishnan, Ribosome Structure and the Mechanism of Translation. *Cell* **108**, 557-572 (2002).
21. M. Kozak, Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187-208 (1999).
22. M. V. Rodnina, Translation in Prokaryotes. *Cold Spring Harbor Perspectives in Biology* **10**, (2018).
23. R. J. Jackson, C. U. T. Hellen, T. V. Pestova, The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology* **11**, 113 (2010).
24. W. C. Merrick, G. D. Pavitt, Protein Synthesis Initiation in Eukaryotic Cells. *Cold Spring Harbor Perspectives in Biology* **10**, (2018).
25. T. E. Dever, R. Green, The Elongation, Termination, and Recycling Phases of Translation in Eukaryotes. *Cold Spring Harbor Perspectives in Biology* **4**, (2012).
26. M. Kozak, Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283-292 (1986).
27. A. G. Hinnebusch, Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. *Microbiology and Molecular Biology Reviews* **75**, 434 (2011).

28. M. Kozak, Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proceedings of the National Academy of Sciences* **87**, 8301 (1990).
29. K. Leppek, R. Das, M. Barna, Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews Molecular Cell Biology* **19**, 158 (2017).
30. Nicholas T. Ingolia *et al.*, Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Reports* **8**, 1365-1379 (2014).
31. G.-L. Chew, A. Pauli, A. F. Schier, Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nature Communications* **7**, 11663 (2016).
32. G. A. Brar *et al.*, High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science* **335**, 552-557 (2012).
33. S. R. Starck *et al.*, Leucine-tRNA Initiates at CUG Start Codons for Protein Synthesis and Presentation by MHC Class I. *Science* **336**, 1719-1723 (2012).
34. S. R. Starck *et al.*, Translation from the 5' untranslated region shapes the integrated stress response. *Science* **351**, aad3867 (2016).
35. T. E. Dever, E. Gutierrez, B.-S. Shin, The hypusine-containing translation factor eIF5A. *Critical Reviews in Biochemistry and Molecular Biology* **49**, 413-425 (2014).
36. V. Magdolen *et al.*, The function of the hypusine-containing proteins of yeast and other eukaryotes is well conserved. *Molecular and General Genetics MGG* **244**, 646-652 (1994).
37. P. M. J. Clement *et al.*, Identification and characterization of eukaryotic initiation factor 5A-2. *European Journal of Biochemistry* **270**, 4254-4263 (2003).
38. Z. A. Jenkins, P. G. Hååg, H. E. Johansson, Human EIF5A2 on Chromosome 3q25–q27 Is a Phylogenetically Conserved Vertebrate Variant of Eukaryotic Translation Initiation Factor 5A with Tissue-Specific Expression. *Genomics* **71**, 101-109 (2001).
39. H. L. Cooper, M. H. Park, J. E. Folk, B. Safer, R. Braverman, Identification of the hypusine-containing protein hy+ as translation initiation factor eIF-4D. *Proceedings of the National Academy of Sciences* **80**, 1854 (1983).
40. J. Schnier, H. G. Schwelberger, Z. Smit-McBride, H. A. Kang, J. W. Hershey, Translation initiation factor 5A and its hypusine modification are essential for cell

- viability in the yeast *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **11**, 3105-3114 (1991).
41. V. S. P. Cano *et al.*, Mutational analyses of human eIF5A-1 – identification of amino acid residues critical for eIF5A activity and hypusine modification. *The FEBS Journal* **275**, 44-58 (2008).
 42. L. Peil *et al.*, Lys34 of translation elongation factor EF-P is hydroxylated by YfcM. *Nature Chemical Biology* **8**, 695 (2012).
 43. D. Rossi, R. Kuroshu, C. F. Zanelli, S. R. Valentini, eIF5A and EF-P: two unique translation factors are now traveling the same road. *Wiley Interdisciplinary Reviews: RNA* **5**, 209-222 (2014).
 44. J. Lassak, D. N. Wilson, K. Jung, Stall no more at polyproline stretches with the translation elongation factors EF-P and IF-5A. *Molecular Microbiology* **99**, 219-235 (2016).
 45. M. H. Park, Y. A. Joe, K. R. Kang, Deoxyhypusine Synthase Activity Is Essential for Cell Viability in the Yeast *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **273**, 1677-1683 (1998).
 46. M. H. Park, E. C. Wolff, Hypusine, a polyamine-derived amino acid critical for eukaryotic translation. *Journal of Biological Chemistry*, (2018).
 47. W. M Kemper, K. W Berry, W. Merrick, *Purification and properties of rabbit reticulocyte protein synthesis initiation factors M2B α and M2B β* . (1976), vol. 251, pp. 5551-5557.
 48. M. H. Schreier, B. Erni, T. Staehelin, Initiation of mammalian protein synthesis: I. Purification and characterization of seven initiation factors. *Journal of Molecular Biology* **116**, 727-753 (1977).
 49. R. Benne, J. W. Hershey, The mechanism of action of protein synthesis initiation factors from rabbit reticulocytes. *Journal of Biological Chemistry* **253**, 3078-3087 (1978).
 50. B.-S. Shin *et al.*, Amino acid substrates impose polyamine, eIF5A, or hypusine requirement for peptide synthesis. *Nucleic Acids Research* **45**, 8392-8402 (2017).
 51. M. Y. Pavlov *et al.*, Slow peptide bond formation by proline and other α -alkylamino acids in translation. *Proceedings of the National Academy of Sciences* **106**, 50-54 (2009).

52. V. Pelechano, P. Alepuz, eIF5A facilitates translation termination globally and promotes the elongation of many non polyproline-specific tripeptide sequences. *Nucleic Acids Research* **45**, 7326-7338 (2017).
53. A. P. B. Gregio, V. P. S. Cano, J. S. Avaca, S. R. Valentini, C. F. Zanelli, eIF5A has a function in the elongation step of translation in yeast. *Biochemical and Biophysical Research Communications* **380**, 785-790 (2009).
54. P. Saini, D. E. Eyler, R. Green, T. E. Dever, Hypusine-containing protein eIF5A promotes translation elongation. *Nature* **459**, 118 (2009).
55. S. Ude *et al.*, Translation Elongation Factor EF-P Alleviates Ribosome Stalling at Polyproline Stretches. *Science* **339**, 82-85 (2013).
56. L. K. Doerfel *et al.*, EF-P Is Essential for Rapid Synthesis of Proteins Containing Consecutive Proline Residues. *Science* **339**, 85-88 (2013).
57. Christopher J. Woolstenhulme, Nicholas R. Guydosh, R. Green, Allen R. Buskirk, High-Precision Analysis of Translational Pausing by Ribosome Profiling in Bacteria Lacking EFP. *Cell Reports* **11**, 13-21 (2015).
58. S. Elgamal *et al.*, EF-P Dependent Pauses Integrate Proximal and Distal Signals during Translation. *PLoS Genetics* **10**, e1004553 (2014).
59. G. Blaha, R. E. Stanley, T. A. Steitz, Formation of the First Peptide Bond: The Structure of EF-P Bound to the 70S Ribosome. *Science* **325**, 966-970 (2009).
60. C. Schmidt *et al.*, Structure of the hypusinylated eukaryotic translation factor eIF-5A bound to the ribosome. *Nucleic Acids Research* **44**, 1944-1951 (2015).
61. O. Shalem, N. E. Sanjana, F. Zhang, High-throughput functional genomics using CRISPR–Cas9. *Nature Reviews Genetics* **16**, 299 (2015).
62. O. Parnas *et al.*, A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162**, 675-686 (2015).
63. R. J. Golden *et al.*, An Argonaute phosphorylation cycle promotes microRNA-mediated silencing. *Nature* **542**, 197 (2017).
64. N. Parkin, A. Darveau, R. Nicholson, N. Sonenberg, cis-acting translational effects of the 5' noncoding region of c-myc mRNA. *Molecular and Cellular Biology* **8**, 2875-2883 (1988).
65. S. D. Fraser, J. Wilkes-Johnston, L. W. Browder, Effects of c-myc first exons and 5' synthetic hairpins on RNA translation in oocytes and early embryos of *Xenopus laevis*. *Oncogene* **12**, 1223-1230 (1996).

66. C. Nanbru *et al.*, Alternative Translation of the Proto-oncogene c-myc by an Internal Ribosome Entry Site. *Journal of Biological Chemistry* **272**, 32061-32066 (1997).
67. M. Stoneley *et al.*, c-Myc Protein Synthesis Is Initiated from the Internal Ribosome Entry Segment during Apoptosis. *Molecular and Cellular Biology* **20**, 1162-1169 (2000).
68. L. Créancier, P. Mercier, A.-C. Prats, D. Morello, c-myc Internal Ribosome Entry Site Activity Is Developmentally Controlled and Subjected to a Strong Translational Repression in Adult Transgenic Mice. *Molecular and Cellular Biology* **21**, 1833-1840 (2001).
69. S. R. Hann, M. W. King, D. L. Bentley, C. W. Anderson, R. N. Eisenman, A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell* **52**, 185-195 (1988).
70. S. R. Hann, M. Dixit, R. C. Sears, L. Sealy, The alternatively initiated c-Myc proteins differentially regulate transcription through a noncanonical DNA-binding site. *Genes & Development* **8**, 2441-2452 (1994).
71. C. Benassayag *et al.*, Human c-Myc Isoforms Differentially Regulate Cell Growth and Apoptosis in *Drosophila melanogaster*. *Molecular and Cellular Biology* **25**, 9897-9909 (2005).
72. H. Saito, A. C. Hayday, K. Wiman, W. S. Hayward, S. Tonegawa, Activation of the c-myc gene by translocation: a model for translational control. *Proceedings of the National Academy of Sciences* **80**, 7476-7480 (1983).
73. S. Solomon *et al.*, Distinct Structural Features of Caprin-1 Mediate Its Interaction with G3BP-1 and Its Induction of Phosphorylation of Eukaryotic Translation Initiation Factor 2 α , Entry to Cytoplasmic Stress Granules, and Selective Interaction with a Subset of mRNAs. *Molecular and Cellular Biology* **27**, 2324-2342 (2007).
74. R. C. Sears, The Life Cycle of C-Myc: From Synthesis to Degradation. *Cell Cycle* **3**, 1131-1135 (2004).
75. A. S. Farrell, R. C. Sears, MYC Degradation. *Cold Spring Harbor Perspectives in Medicine* **4**, (2014).
76. A. R. Buskirk, R. Green, Ribosome pausing, arrest and rescue in bacteria and eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20160183 (2017).

77. I. P. Ivanov *et al.*, Polyamine Control of Translation Elongation Regulates Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing. *Molecular Cell* **70**, 254-264.e256 (2018).
78. A. C. Childs, D. J. Mehta, E. W. Gerner, Polyamine-dependent gene expression. *Cellular and Molecular Life Sciences CMLS* **60**, 1394-1406 (2003).
79. J. F. Atkins, J. B. Lewis, C. W. Anderson, R. F. Gesteland, Enhanced differential synthesis of proteins in a mammalian cell-free system by addition of polyamines. *Journal of Biological Chemistry* **250**, 5688-5695 (1975).
80. Y.-H. Pan *et al.*, The Critical Roles of Polyamines in Regulating ColE7 Production and Restricting ColE7 Uptake of the Colicin-producing *Escherichia coli*. *Journal of Biological Chemistry* **281**, 13083-13091 (2006).
81. M. K. Chattopadhyay, M. H. Park, H. Tabor, Hypusine modification for growth is the major function of spermidine in *Saccharomyces cerevisiae* polyamine auxotrophs grown in limiting spermidine. *Proceedings of the National Academy of Sciences* **105**, 6554-6559 (2008).
82. L. Miller-Fleming, V. Olin-Sandoval, K. Campbell, M. Ralser, Remaining Mysteries of Molecular Biology: The Role of Polyamines in the Cell. *Journal of Molecular Biology* **427**, 3389-3406 (2015).
83. M. Gabay, Y. Li, D. W. Felsher, MYC Activation Is a Hallmark of Cancer Initiation and Maintenance. *Cold Spring Harbor Perspectives in Medicine* **4**, (2014).
84. S. O. Sulima, I. J. F. Hofman, K. De Keersmaecker, J. D. Dinman, How Ribosomes Translate Cancer. *Cancer Discovery* **7**, 1069-1087 (2017).
85. R. Ajore *et al.*, Deletion of ribosomal protein genes is a common vulnerability in human cancer, especially in concert with *TP53* mutations. *EMBO Molecular Medicine* **9**, 498-507 (2017).
86. K. De Keersmaecker *et al.*, Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. *Nature Genetics* **45**, 186 (2012).
87. A. S. Y. Lee, P. J. Kranzusch, J. H. D. Cate, eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* **522**, 111 (2015).
88. E. Batsché, C. Crémisi, Opposite transcriptional activity between the wild type *c-myc* gene coding for c-Myc1 and c-Myc2 proteins and c-Myc1 and c-Myc2 separately. *Oncogene* **18**, 5662 (1999).

89. Y. Liu *et al.*, Deletions linked to TP53 loss drive cancer through p53-independent mechanisms. *Nature* **531**, 471 (2016).
90. A. Sendoel *et al.*, Translation from unconventional 5' start sites drives tumour initiation. *Nature* **541**, 494 (2017).
91. M. D. Hogarty *et al.*, *ODC1* Is a Critical Determinant of *MYCN* Oncogenesis and a Therapeutic Target in Neuroblastoma. *Cancer Research* **68**, 9735-9745 (2008).
92. C. Bello-Fernandez, G. Packham, J. L. Cleveland, The ornithine decarboxylase gene is a transcriptional target of c-Myc. *Proceedings of the National Academy of Sciences* **90**, 7804-7808 (1993).
93. M. Auvinen, A. Paasinen, L. C. Andersson, E. Hölttä, Ornithine decarboxylase activity is critical for cell transformation. *Nature* **360**, 355-358 (1992).
94. M. B. Mathews, J. W. B. Hershey, The translation factor eIF5A and human cancer. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1849**, 836-844 (2015).
95. T. R. Murray-Stewart, P. M. Woster, R. A. Casero, Targeting polyamine metabolism for cancer therapy and prevention. *Biochemical Journal* **473**, 2937-2953 (2016).
96. Joel D. Richter, J. Collier, Pausing on Polyribosomes: Make Way for Elongation in Translational Control. *Cell* **163**, 292-300 (2015).
97. J. Seibler *et al.*, Reversible gene knockdown in mice using a tight, inducible shRNA expression system. *Nucleic Acids Research* **35**, e54-e54 (2007).
98. N. E. Sanjana *et al.*, A transcription activator-like effector toolbox for genome engineering. *Nature Protocols* **7**, 171 (2012).
99. O. Shalem *et al.*, Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**, 84-87 (2014).
100. N. J. Mglincy, N. T. Ingolia, Transcriptome-wide measurement of translation by ribosome profiling. *Methods* **126**, 112-129 (2017).
101. N. T. Ingolia, G. A. Brar, S. Rouskin, A. M. McGeachy, J. S. Weissman, The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols* **7**, 1534 (2012).
102. C. Quast *et al.*, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590-D596 (2012).

103. P. P. Chan, T. M. Lowe, GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research* **44**, D184-D189 (2015).
104. O. Valladares *et al.*, DASHR 2.0: integrated database of human small non-coding RNA genes and mature products. (2018).
105. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**, 357 (2015).
106. A. Frankish *et al.*, GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research* **47**, D766-D773 (2018).
107. S. R. Engel *et al.*, The Reference Genome Sequence of *Saccharomyces cerevisiae*: Then and Now. *G3: Genes/Genomes/Genetics* **4**, 389-398 (2014).
108. U. Nagalakshmi *et al.*, The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science* **320**, 1344-1349 (2008).
109. O. Perez-Leal, C. A. Barrero, A. B. Clarkson, R. A. Casero, S. Merali, Polyamine-Regulated Translation of Spermidine/Spermine-N¹-Acetyltransferase. *Molecular and Cellular Biology* **32**, 1453-1467 (2012).