Characterizing the cis-regulatory roles of translating downstream open reading frames

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Anuj Chaturvedi

20201260

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

Date: April, 2025

Under the guidance of

Supervisor: Dr. Ariel Bazzini,

Associate Investigator

Stowers Institute for Medical Research, Kansas City, USA

From June 2024 to March 2025

Certificate

This is to certify that this dissertation entitled **“Characterizing the cis-regulatory roles of translating downstream open reading frames”** towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune, represents work carried out by **Anuj Chaturvedi** at Stowers Institute for Medical Research, USA under the supervision of **Dr. Ariel Bazzini,** Associate Investigator, during the academic year 2024-2025.

Anuj Chaturvedi
20201260

Dr. Ariel Bazzini
Associate Investigator
Stowers Institute for Medical Research

Thesis Advisory Committee (TAC)

Supervisor: Dr. Ariel Bazzini

Thesis Advisor: Dr. Girish Ratnaparkhi

This thesis is dedicated to my father.

Declaration

I hereby declare that the matter embodied in the report entitled **“Characterizing the cis-regulatory roles of translating downstream open reading frames”** are the results of the work carried out by me at the Stowers Institute for Medical Research under the supervision of Dr. Ariel Bazzini and the same has not been submitted elsewhere for any other degree.

Anuj Chaturvedi

20201260

Date: March 24, 2025

Table of Contents

List of Figures

Abstract

Understanding the coding potential of the genome is a fundamental step for deciphering the molecular framework of living systems. Traditionally, eukaryotic mRNAs were thought to contain only one protein-coding sequence. However, with the advent of high-throughput techniques like ribosome profiling and proteomics, people have identified small translated open reading frames (ORFs) in both the 5′ UTRs and 3′ UTRs called upstream ORFs (uORFs) and downstream ORFs (dORF), respectively. The lab is mainly interested in newly found translation events in 3′ UTR and the unexplored consequences of the presence of actively translating ribosomes in this region. The study explores how the translation of dORFs influences other regulatory processes involving 3′ UTRs.

Using an *in vitro* transcribed mCherry-based reporter system, we investigated the impact of ribosomes translating poly-A sequences on mRNA stability and protein output. We hypothesized that the translation of the poly-A sequence imposed by dORF would trigger transcript decay by the No-Go Decay (NGD) mechanism. To test this, we measured the protein levels as well as transcript levels. Our findings suggest that the translation of multiple adenines within dORFs does not induce large-scale mRNA decay.

Furthermore, we also explored the interaction between the active translating ribosomes and repression by microRNA-AGO complex. We hypothesize that dORFs work as negative regulators of microRNA (miRNA) targeting, such as observed in canonical ORFs, where miRNA target sites are not efficiently responsive to their cognate miRNA. To test this, we designed a mCherry expression vector containing an artificial 3′ UTR with a dORF encompassing two miRNA target sites for miR-1. Protein levels of mCherry were determined by cytometry analysis, and levels of transcripts were determined by qPCR.

Acknowledgments

Choosing to pursue my thesis outside abroad had its fair share of challenges and growth opportunities. Fortunately, I can look back on my decision and be proud of the obstacles I overcame and the person I have become. None of this would have been possible without the support system that I am blessed with. I can not be thankful enough for the impact you all have had on my life.

First and foremost, I would like to thank Ariel for giving me the opportunity to do research for my thesis at Stowers Institute. Your kindness creates an environment where every member of the lab feels valued, much like a parent taking care of their children. Your passion for science and new ideas is infectious and inspiring on many levels. Thank you for mentoring me and giving me a home away from home.

I am immensely grateful to every member of the Bazzini lab. The welcoming atmosphere that you people have fostered in the lab has allowed me to grow as a scientist and, more so, as a human. I owe a great deal to Majo and Luli for helping me get familiarised with the functioning of the lab and the techniques. To =Dani, Hori, and Cam—your invaluable support on my projects, along with your life advice, helped me get over a lot of stuff that felt overwhelming. Gopal, you always treated me like a younger brother and made sure I was not alone. I am so thankful for each and every one of you—Daniel, Eugenia, Ale, Cata, AJ, Amruta, Gabe, Daniel Careno, and Sergio.

I was fortunate enough to come with friends from my college. Subhransu, Nandeesh, and Atharva, I'm particularly grateful for your companionship during this time in the faraway land. To Ian and Lucy, thank you for being an integral part of my life in this country—you guys are a source of constant joy and inspiration. To my friends at IISER—Manav, Yatish, Srivathsa, Anju, Nikhil, Surya, and Giri—I may not have been in touch as much this past year, but I am so excited to see you guys again and share the battle scars we have earned this past year. It has been a blessing to have known each and every one of you.

Special thanks to Sara, our administrative assistant, thanks a lot for handling all the administrative work. The core facilities at the Stowers Institute have been incredibly helpful with the technical and logistical support that made my research possible.

It would have been very hard for me to navigate through life without the support I received from my elder brother; thanks a lot for being a good brother and supporting my choices. Mother, I can not highlight enough how much I admire you; being a single mother and raising two kids all by yourself while going through what you went through is a story people can only find in books. Dad, I may only have a year or two worth of memories with you, but the love I experienced is still a source of inspiration for me.

Contributions

Contributor name	Contributor role
Bazzini lab	Conceptualization Ideas
Anuj and Bazzini lab members	Methodology
Anuj Chaturvedi	Software
Anuj Chaturvedi	Validation
Anuj Chaturvedi	Formal analysis
Anuj Chaturvedi	Investigation
Anuj and Bazzini lab members	Resources
Anuj Chaturvedi	Data Curation
Anuj Chaturvedi	Writing - original draft preparation
Anuj, Cameron, Hori and Ariel Bazzini	Writing - review and editing
Anuj Chaturvedi	Visualization
Ariel Bazzini	Supervision
Ariel Bazzini	Project administration
Ariel Bazzini	Funding acquisition

Chapter 1

Introduction

The likelihood of offspring often resembling their parents is not merely a coincidence but a byproduct of the tightly controlled transfer of genetic information. The genetic information in parents is encoded in DNA (deoxyribonucleic acid) molecules. The X-ray crystallography work of Rosalind Franklin and Maurice Wilkins enabled James Watson and Francis Crick to propose the double-helical model to describe the structure of DNA. The discovery of DNA's double-helix structure allowed Francis Crick in 1957 to propose the central dogma of molecular biology, which describes the unidirectional flow of genetic information (Pray, L., 2008). According to the proposal, genetic information is first transcribed from DNA into RNA (ribonucleic acid) by RNA polymerases and subsequently translated by ribosomes into proteins. Crick asserted that proteins are the ultimate functional products of this process, as the flow of information is irreversible once it gets decoded into protein (Crick, 1970; Cobb, 2017).

1.1 Dark genome

Out of the 3 billion base pairs that comprise the DNA in humans, 80% are transcribed into RNA molecules, of which only 2% are translated into proteins (International Human Genome Sequencing Consortium, 2004). Some regions of untranscribed DNA can act as promoters, silencers, etc., influencing when and how much a gene is transcribed (Barrett *et al.*, 2012). Compared to the untranscribed genome, a much more significant portion consists of untranslated RNA molecules. Among these are non-coding RNAs (ncRNAs) and untranslated regions (UTRs) within protein-coding transcripts (Venter *et al.*, 2001; Park *et al.*, 2022). The coding sequence (CDS) of an mRNA is defined by an open reading frame (ORF). This nucleotide sequence begins with a start codon and ends with an in-frame stop codon, with a three-nucleotide periodicity, because the information of each amino acid is encoded in 3 nucleotides called codons, to encode a polypeptide. In addition to the ORF, each mRNA contains untranslated regions upstream and downstream of the coding sequence, called the 5′ UTR and 3′ UTR,

respectively. These untranslated regions can play crucial roles in modulating the process of translation (Ingolia *et al.*, 2012; Kute *et al.*, 2022).

1.2 Gene regulation

Gene regulation is the process by which cells control the expression of their genetic material, ensuring the right genes are activated at the correct time, in the suitable cell type, and at the desired levels. (Lambert *et al.*, 2018). Gene regulation enables organisms to adapt to environmental changes, maintain homeostasis, and execute spatial-temporally complex developmental programs (Takahashi and Yamanaka, 2006). Gene regulation is a multi-level process, starting with the transcription of the DNA, which can be modulated by methylation of DNA, specific transcription factors binding to DNA, enhancers, silencers, and chromatin modifications (Lee and Young, 2013). Additional layers of gene expression occur after initiating the transcription of DNA; messenger RNA (mRNA) transcripts undergo splicing, capping at the 5′ end, polyadenylation at 3′, and other mRNA modifications (Moore and Proudfoot, 2009). After mRNA is fully processed, translational controls and post-translational modifications can modulate the translational efficiency of transcripts, protein activity, localization, or stability (Zhong *et al.*, 2023; Walsh *et al., 2005*).

1.3 Role of ribosomes in gene regulation

Ribosomes are macromolecular machinery of ribosomal RNAs (rRNAs) and proteins, which work together to synthesize proteins in the cell (Shah *et al.*, 2024). Ribosomes are made up of two subunits that decode the mRNAs into protein with a periodicity of three nucleotides at a time, using transfer RNAs to convert each codon into an amino acid. These amino acids are joined together by peptide bonds and form proteins (Lareau *et al.*, 2014). Traditionally, ribosomes have been thought of as unchanging macromolecular machinery with no regulatory role in translation (Shah *et al.*, 2024).

Apart from translation, ribosomes also play a crucial role in checking the quality of mRNAs. Ribosomes translating mRNAs with a premature stop codon will cause the transcripts to undergo decay via nonsense-mediated decay (NMD); similarly, translation activity can also result in the degradation of transcripts without a stop codon, and

mRNAs with stalled ribosomes, known as nonstop mediated decay (NSD) and no-go decay (NGD), respectively (Wu and Bazzini, 2018). Such mechanisms are crucial to identify and remove the abnormal mRNAs that could give rise to aberrant proteins. Beyond their role in quality control, ribosomes can also regulate normal mRNAs through a translation-dependent process called codon optimality. mRNAs that are enriched in optimal codons generally exhibit increased stability and higher translational efficiency than mRNAs consisting of relatively non-optimal codons (Wu *et al.*, 2019).

1.4 Role of untranslated regions (UTRs) in regulating gene expression

As mentioned above, a large fraction of the human genome is not translated into proteins (Venter *et al.*, 2001). Most of the remaining part is involved in regulation at various levels. Both 5′ and 3′ UTRs have an important role to play in the post-transcriptional regulation of genes. They influence the transport of mRNA out of the nucleus, translational efficiency, mRNA and protein localization, and mRNA stability (Mignone *et al.*, 2002). The post-transcriptional regulation of UTRs involves the contribution of both the specific sequence as well as secondary structural elements, which can recruit specific RNA-binding proteins and non-coding RNAs (Pesole *et al.*, 2001; Wieder *et al.*, 2024).

1.5 Ribosome profiling

Ribosomes act on mRNAs and synthesize proteins, the primary effector of many cellular processes. While the genetic makeup determines the amount and type of proteins produced by a cell, the information about the DNA and RNA are often not sufficient to predict the protein expression due to the extensive regulation. In 2009, Jonathan Weissman’s lab introduced ribosome profiling, a strategy to directly measure translation genome wide. Ribosome profiling is a technique based on the sequencing of the mRNA fragments, protected by translating ribosomes. The process starts with treating the cells with cycloheximide to immobilize the ribosomes and lysing the cells to retrieve ribosomes with in vivo positions on mRNAs. As ribosomes translate mRNAs, about 30 nucleotides of mRNA sequence enter these molecular machines, protecting them from nuclease digestion, forming ribosome footprints that can be sequenced. Ribosome

footprints are then purified and reverse-transcribed into a DNA library, followed by deep sequencing **(Figure 1)**. This approach is useful in indicating the sequences that are being translated (Ingolia *et al.*, 2009). This method has been instrumental in uncovering previously unrecognized translation events. For instance, ribosome profiling of Zika virus (ZIKV) infected cells revealed novel upstream open reading frames (uORFs) across different viral strains. Peptides encoded by these regions have been shown to modulate viral infection (Lefèvre *et al.*, 2024). While eukaryotic mRNAs were traditionally thought to contain only a single coding sequence (CDS), ribosome profiling has revealed many short ORFs (sORFs) that encode short peptides within regions once considered untranslated (Ingolia et al., 2012; Kute et al., 2022).

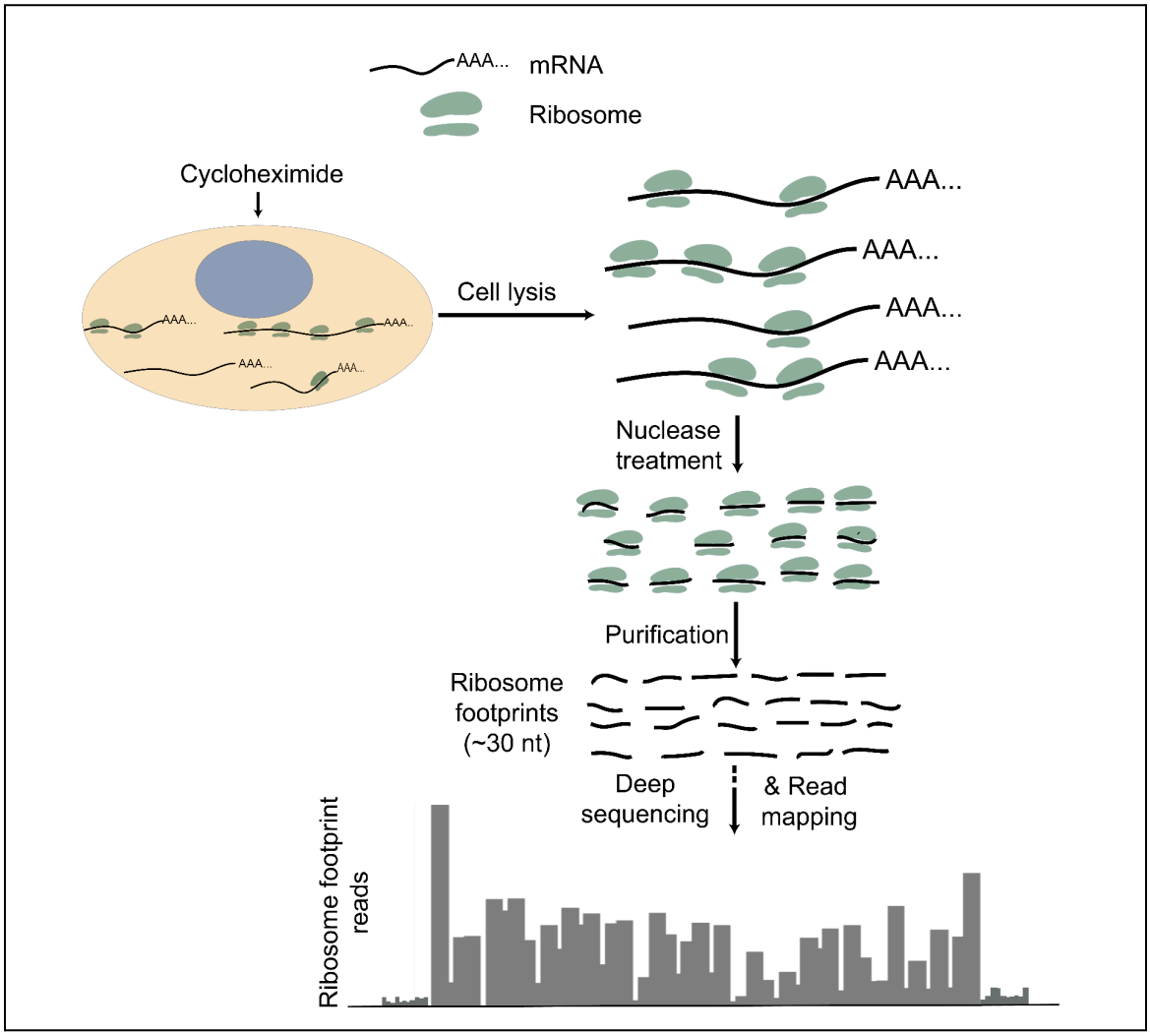

Figure 1: Graphical representation of ribosome footprinting described in Ingolia et al. 2009.

1.6 5′ UTRs

The 5′ UTR is defined as the sequence upstream of the main ORF. Secondary structures within the 5′ UTR can modulate the translational efficiency and transcript turnover rates. In eukaryotes, translation typically begins when the small ribosomal subunit binds to the initiator methionyl-tRNA and scans downstream from the 5′ cap for the complementary AUG start codon. Because this scanning process is impeded by the complex secondary structures, less structured 5′ UTRs generally enhance translation efficiency and increase protein output (Ringnér and Krogh, 2005; Dodbele and Wilusz, 2020). 5′ UTRs also harbor many regulatory elements, including the internal ribosome-entry sites (IRESs), small open reading frames called uORFs, G-quadruplexes, and translation initiation sites (Johnstone *et al.*, 2016).

Ribosome profiling has led to the identification of ribosome-protected fragments in the 5′ UTRs of genes, challenging the longstanding notion of one CDS per gene, and questioning what truly is an ‘untranslated region.’ Based on this data, approximately 40% of all mRNAs in humans contain ORF in the 5′ UTR, called upstream open reading frames (uORFs) (Somers *et al.*, 2013) **(Figure 2)**. uORFs add another facet of post-transcriptional regulation as the translation of uORFs reduces the protein expression from the main CDS, likely due to the competition for translation initiation. Differences in uORF length, number, and distance from the mRNA cap, as well as overlap with the canonical ORF, can influence the regulatory effect of uORFs on the main ORF (Somers *et al.*, 2013; Johnstone *et al.*, 2016).

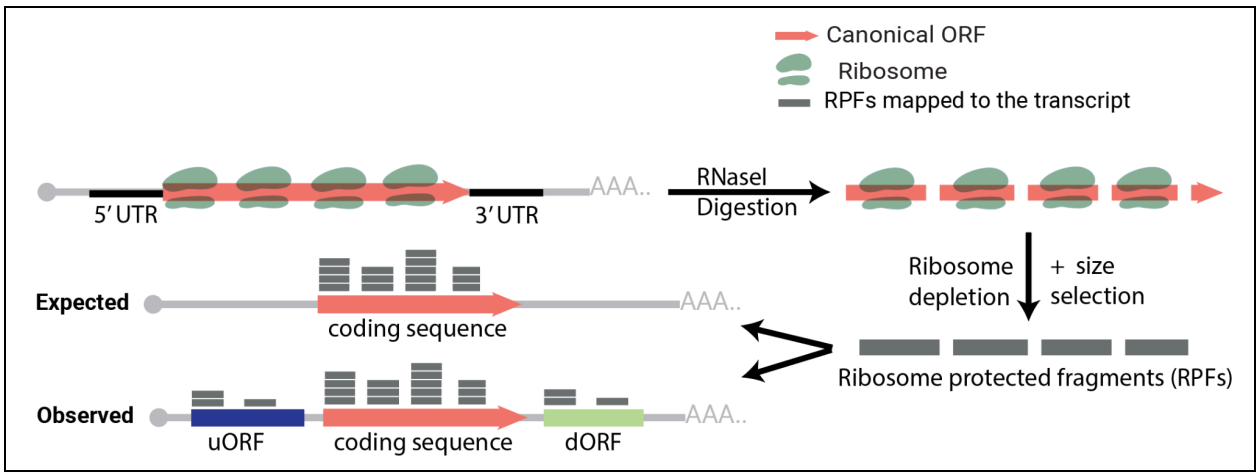

Figure 2: Graphical representation of how ribosome profiling leads to the identification of small open reading frames (Ingolia et al., 2014).

1.7 3′ UTR

An mRNA's 3′ UTR plays a very crucial role in post-transcriptional gene regulation. 3′ UTRs contain sequence motifs that can interact with proteins to regulate mRNA stability, influence nuclear export, sub-cellular localization, and translation efficiency, thus modulating protein synthesis (Matoulkova *et al.*, 2012).

One set of motifs often found in 3′ UTRs, Adenine-uracil-rich elements (AU-rich elements), are bound by heterogeneous nuclear ribonucleoproteins (hnRNPs) that promote mRNA deadenylation and decay, decreasing mRNA stability. These AU-rich elements are enriched in transcripts that encode for oncoproteins, cytokines, and transcription factors (Zubiaga *et al.*, 1995; Peng *et al.*, 1998). During the embryonic development of Drosophila, mRNAs of morphogens like bicoid and nanos are localized to the anterior and posterior poles of the embryo, respectively, due to the motifs present in 3′ UTRs. Nanos protein, which is required for abdomen formation in drosophila, becomes mislocalized when its 3′ UTR is switched with that of bicoid and results in embryos with two abdomens with mirror symmetry (Wang and Lehmann, 1991; Gavis and Lehmann, 1992).

Genes can produce mRNAs with different 3′ UTRs due to alternative 3′ cleavage and polyadenylation. In eukaryotes, RNA transcribed from the DNA needs to be processed before exporting to the nucleus. To create the 3′ end of mRNAs, newly transcribed RNAs are endonucleolytically cleaved and polyadnylated. Regulation of the cleavage site results in more than 50% of mammalian genes containing mRNA isoforms that differ in how much 3′ sequence is added. These cleavage and polyadenylation signals can occur in introns of the mRNA or in 3′ UTR, giving rise to transcripts with different 3′ UTR lengths. These isoforms that differ only in the 3′ UTR sequence can influence transcript localization, stability, and translation. It can lead to truncated protein-coding transcripts if the intronic site is used (Matoulkova *et al.*, 2012; Kowalski *et al.*, 2024).

Trans-acting factors also recognize elements in 3′ UTR to regulate protein output. MicroRNAs (miRNAs) are ~22 nucleotide-long RNAs belonging to the class of noncoding RNAs that interact with argonaute (AGO) proteins and form a silencing complex. miRNAs guide the complex to the specific binding sites, called 'seed' target sites primarily present in 3′ UTRs. The miRNA and AGO complex binding results in decreased protein output from the targeted mRNA due to reduced translation and increased mRNA decay (Gu *et al.*, 2009; Dueck *et al.*, 2012; Matoulkova et al., 2012; McGeary et al., 2019). Similar to uORFs, ribosome profiling has identified small open reading frames in the 3′ UTR, called downstream open reading frames (dORFs), that also have a role to play in gene regulation (Bazzini *et al.*, 2014). The discovery of translated elements in 3′ UTRs is very recent, and little is known about the role of dORF peptides and how these elements can affect the translation of the main CDS. It was shown recently that the translation of the dORFs can enhance the translation efficiency of the canonical ORF, which does not depend on the peptide sequence of the dORFs but rather on the number of dORFs in the 3′ UTR **(Figure 3)** (Wu *et al.*, 2020).

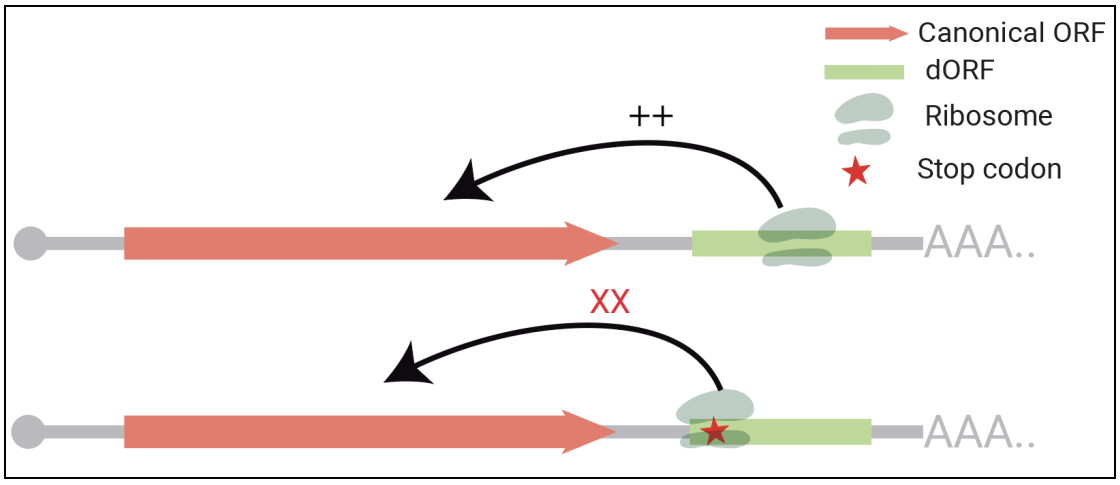

Figure 3: dORFs enhance the translational efficiency of canonical ORF. Without the stop codon, the dORF gets translated and enhances the translation efficiency of the main ORF, introducing a point mutation for the stop codon, preventing dORF translation, and taking away the enhancement effect (Wu et al. 2020).

1.8 Downstream open reading frames

As mentioned above, ribosome-protected fragments have been found to map to 3′ UTRs, with the sequencing reads showing characteristic three-nucleotide periodicity indicative of active translation. The presence of these sORFs within 3′ UTRs has also been independently detected through mass spectrometry. The region between the stop codon of the main CDS and the start codon of the dORF is called iUTR (internal untranslated region). This region is believed to be responsible for recruiting translation factors and driving the translation of dORF, similar to how viral IRESs recruit ribosomes for cap-independent translation (Bazzini *et al.*, 2014; Wu *et al.*, 2020).

A gene regulatory mechanism was described based on the ribosome stalling while translating poly-adenylate tracks. It was shown that the translation of A-rich sequences can tune down the protein output and affect mRNA stability (Arthur *et al.*, 2015).

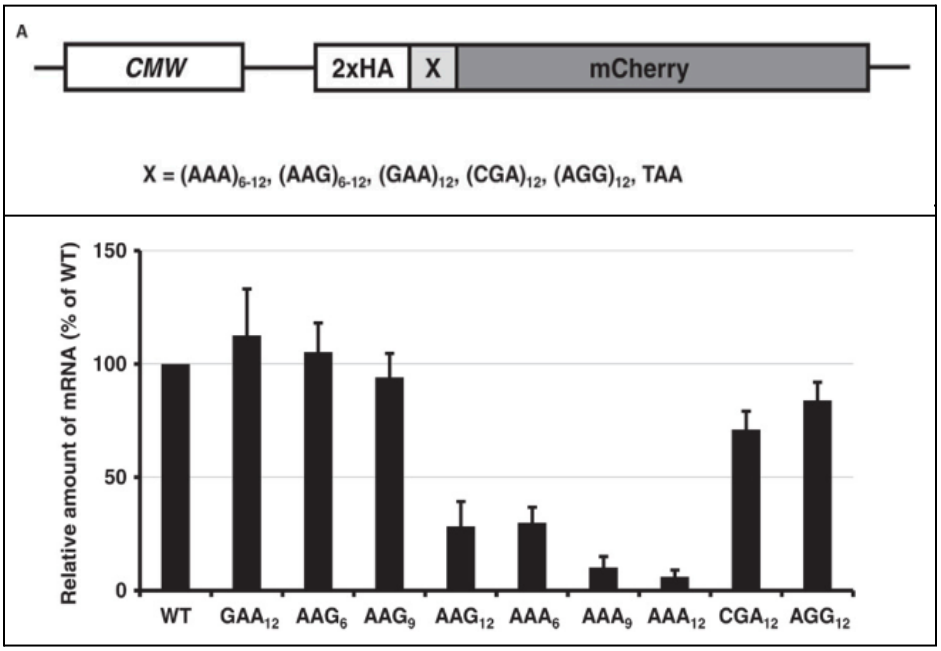

Figure 4: Figure showing the translation of polyA tracks results in reduced stability of the transcripts (Arthur et al. 2015).

Due to unexplored consequences of dORFs in 3′ UTRs is that the translation may proceed into the poly(A) tail if the dORF lacks a stop codon before the site at which 3′ cleavage and polyadenylation occur. For example, suppose a dORF-containing

transcript underwent alternative polyadenylation that removed the stop codon. In that case, this isoform may be degraded or lowly translated due to the translation of the poly(A)tail **(Figure 5)**. To test the effect of dORF translation of the poly(A) stretches, we designed a mCherry expression vector with an artificial 3′ UTR that contains dORFs with 54As in tandem. This study will contribute to our understanding of if there are other regulatory implications of having ribosomes in the 3′ UTR.

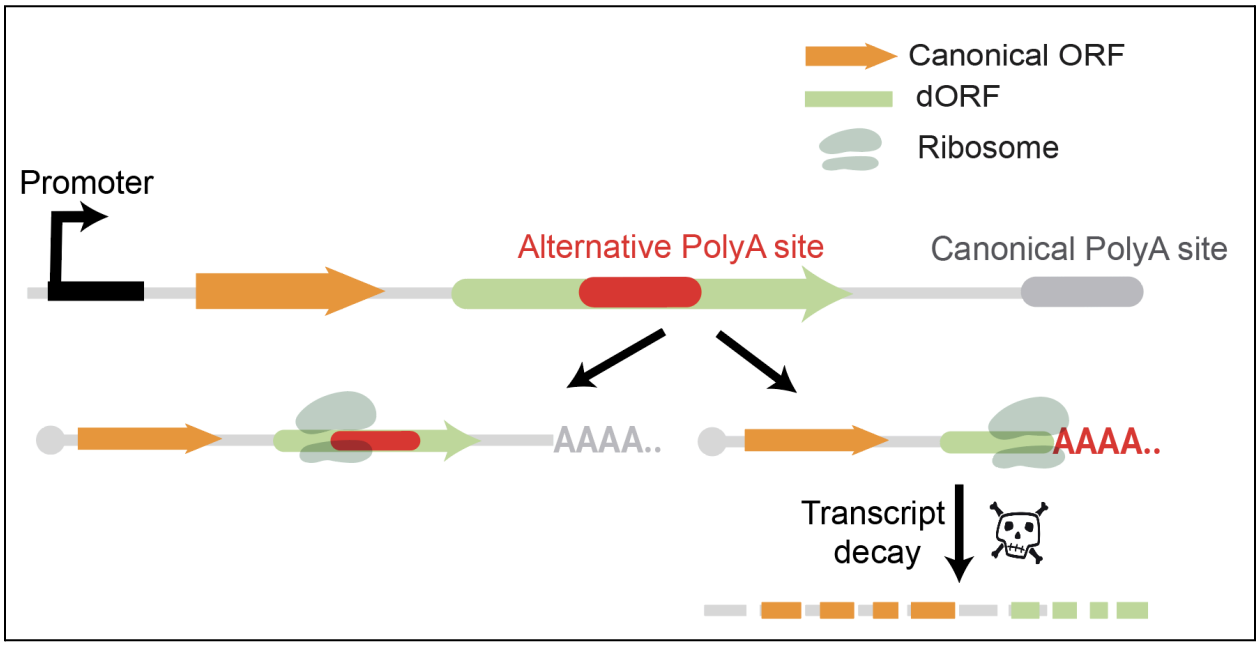

Figure 5: Illustration showing the potential alternative 3' end cleavage resulting in poly-A sequences in the open reading frame of dORF.

As discussed earlier, miRNAs are very important post-transcriptional regulators and usually interact with target seeds in 3′ UTR (Gu *et al.*, 2009). Mark Kay's group 2009 experimentally verified the functional significance of having miRNA target sites located in 3′ UTR and concluded that the active translation sabotages the miRNA-mediated decay of target mRNAs **(Figure 6)**.

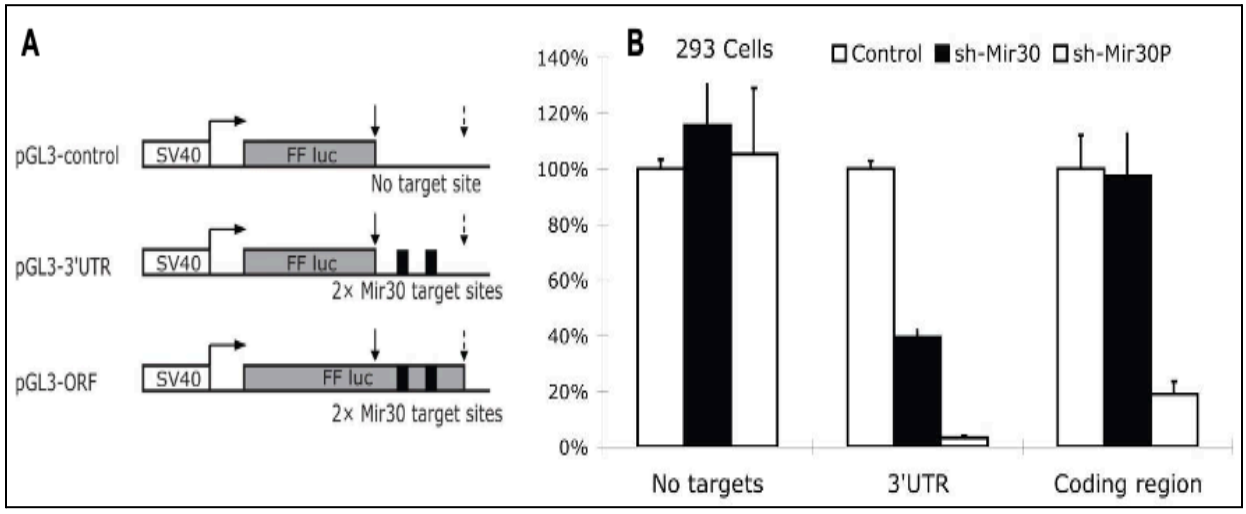

Figure 6: Figure demonstrating experimental design and the repression of the firefly luciferase when the target sites are present in 3' UTR compared to the ORF (Gu *et al.*, 2009).

A: Illustration of the constructs used for the experiment.

B: HEK293 cells were transfected with sh-scramble, sh-miR30 or sh-miR30P. The panel shows the result of the dual-luciferase assay.

Here, we hypothesize that dORFs work as negative regulators of miRNA targeting, as observed in canonical ORFs, where miRNA target sites are not efficiently responsive to their cognate miRNA. To test this hypothesis, we designed a mCherry expression vector containing an artificial 3′ UTR with a dORF encompassing two miRNA target sites from genes previously described to be responsive to miRNA-1 over-expression. We aim to induce miRNA expression by transfection and analyze mCherry expression using cytometry and qPCR. Reporters with translating dORFs will be less sensitive to miRNA overexpression than reporters with truncated dORF, as ribosomes will not impede the miRNA and Ago protein complex. If our hypothesis is correct, the reporters that contain wild-type dORFs (complete dORFs) will be less responsive to miRNA over-expression than reporters with a truncated dORF (dMUT).

Chapter 2

2.1 Materials

PCR was performed using Phusion Polymerase (NEB, Catalog No. M0530L) with reactions supplemented with dNTPs (Promega, Catalog No. U1515) and nuclease-free water (Ambion, Catalog No. AM9937), using Bio-Rad C-1000 Touch Thermocycler. Gel electrophoresis was performed using 6× Gel Loading Dye (NEB, Catalog No. B7024S), and sizes of the DNA fragments were estimated using the 1kb+ DNA Ladder (Invitrogen, Catalog No. 10488090). PCR purification kit (Qiagen, Catalog No. 28106) was used to clean up DNA fragments post-PCR and other enzymatic reactions. A gel extraction kit (Qiagen, Catalog No. 28704) was used to recover DNA from agarose gels. Restriction enzymes (MluI, PstI, EcoRI, BamHI, XhoI, XbaI, SpeI, HindIII, DraIII, and BsrGI; NEB) were used for restriction cloning, Gibson assembly, cloning confirmations, and linearizing vectors. Ligation reactions were carried out using T4 DNA Ligase (NEB, Catalog No. M0202L), while Gibson assembly was performed using 2x Hi-Fi DNA Assembly Master Mix (NEB, Catalog No. M5520AA). The concentration of nucleic acids was measured using the Thermo Scientific™ NanoDrop™ 1000 Spectrophotometer. High-efficiency DH5α competent cells were used for the transformation of plasmids (NEB, Catalog No. C2987I), and QIAprep Spin Miniprep Kit was used for plasmid purification (Qiagen, Catalog No. 27106). Culture media, including Lysogeny broth (LB), Super Optimal Broth (SOC), and LB-agar plates with ampicillin, were provided by the Media-Prep facility of Stowers Institute. Ampicillin (Catalog no. 76807) was used as the selection marker at 100 µg/mL of working concentration.

For cell culture, DMEM (Corning, Catalog No. 15-013-CV) and 1x PBS (Corning, Catalog No. 1-040-CV) were used. Additional reagents included 0.25% Trypsin-EDTA (Gibco, Catalog No. 25200-056), Fetal Bovine Serum (avantor, Catalog No. 97068-085), Penicillin-Streptomycin (Gibco, Catalog No. 15240-062), and Opti-MEM™ I Reduced Serum Medium (Gibco, Catalog No. 31985-062). HEK-293T cells were obtained from the American Type Culture Collection (ATCC, Identifier #CRL-11268). DNA transfections were performed using Lipofectamine™ 3000 Transfection Reagent (Invitrogen, Catalog

No. L3000001), while RNA transfections were carried out using Lipofectamine™ MessengerMAX™ Transfection Reagent (Invitrogen, Catalog No. LMRNA015). In vitro transcription was conducted using the mMESSAGE mMACHINE™ T3 Transcription Kit (Invitrogen, Catalog No. AM1348), and the resulting RNA was purified using the Qiagen RNeasy Mini Kit (Catalog No. 74104). Flow cytometry analysis was performed using a Bio-Rad Ze5 Flow Cytometer, and data were analyzed with FCS Express 7 Research (Version 7.22.0006). RNA extraction from HEK-293T cells was performed using TRIzol™ Reagent (Invitrogen, Catalog No. 15596018). cDNA synthesis was carried out using the SuperScript™ IV First-Strand Synthesis System (Invitrogen, Catalog No. 18090050). Quantitative PCR (qPCR) was performed using PerfeCTa SYBR Green FastMix Low ROX (Quantabio, Catalog No. 95074-05K) on a QuantStudio 7 Real-Time PCR System. All oligos used were ordered from Integrated DNA Technologies (IDT).

2.2 Methods

2.2.1 Tissue culture

HEK 293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS), 2mM L-Glutamine, 100 units/mL of penicillin, and 100 µg/mL of streptomycin. The cells were obtained from the tissue culture facility from Stowers Institute at a low-level passage, lower than passage 12. All work was done according to the IACUC protocol 2022-144. Cells were washed using phosphate-buffered saline (PBS) before trypsinising them using 0.25% trypsin-EDTA. The cells were incubated at 37℃ supplemented with 5% CO_2. Normal passaging of cells was done in a 10 cm dish when cells reached 80% confluency. Proper aseptic techniques were used to minimize the risk of contamination.

2.2.2 Cloning

To create reporter plasmids, we started with the pCS2 vector available in the lab containing mCherry as the main ORF. A poly-A sequence containing 54 adenines was cloned downstream of the main ORF, with sequences available in between to facilitate

the insertion of iUTRs and dORFs using Gibson assembly, referred to as the parental plasmid. The sequences were taken from endogenous genes and cloned downstream of the main ORF, generating dORF reporter plasmids, which included the iUTR and small peptide coding open reading frame, ensuring the dORF is in frame with the poly-A sequences. For dMUT reporters, we used site-directed mutagenesis using PCR to add a stop codon after the first start codon (ATG) of the dORF sequence. The plasmid was digested with MluI and SpeI, followed by dephosphorylation to prevent the self-ligation. The insert sequence was amplified using PCR and assembled into the digested vector using Gibson assembly. The resulting plasmids were subsequently used for *in vitro* transcription (IVT) reactions.

TargetScanHuman, developed by David Bartel's lab, was used to identify mRNAs predicted to be targeted by miRNA-1 with two 8-mer seed matches. From the screening results, *CORO1C* was selected, and a gBlock with a fragment of its 3′ UTR containing both miR-1 target sites was ordered to preserve endogenous local sequence context. The gBlock was inserted into the plasmid using restriction digestion to generate a parental plasmid, which contained mCherry as the main ORF and the miR-1 target sites downstream. Gibson assembly was used to insert the iUTR and dORF sequences, and site-directed mutagenesis was performed to introduce a stop codon immediately after the first start codon of the dORFs, to create dMUT reporters. The design of the reporter plasmids ensured that the miR-1 target sites were in frame with the downstream open reading frames.

2.2.3 *in vitro* transcription and RNA transfection

The plasmids containing the dORF reporters were amplified using a forward primer to incorporate the T3 promoter sequence and a reverse primer containing sixty adenine nucleotides, serving as the poly(A) tail. The amplified products were extracted from the 1% agarose gel and used as a template for *in vitro* transcription reaction using mMESSAGE mMACHINE™ T3 Transcription Kit. The resulting mRNA was purified using the Qiagen RNeasy Mini Kit, and the concentration was quantified using the Nanodrop spectrophotometer. 100,000 HEK-293T cells/well were seeded in a 24-well plate 24 hours before transfection. RNA transfections were performed using the

Lipofectamine™ MessengerMAX™ Transfection Reagent, following the manufacturer's protocol. Each well was transfected with a total of 125 ng of mRNA, comprising 62.5 ng of dORF reporter mRNA and 62.5 ng of GFP mRNA as a transfection control. Cells were collected 24 hours post-transfection for cytometry and RNA extraction.

2.2.4 DNA transfection

DNA transfections in HEK-293T cells were performed using the Lipofectamine™ 3000 Transfection Reagent, following the manufacturer's protocol. 100,000 cells per well were seeded in a 24-well plate 24 hours before the transfection experiments. Each well was transfected with a total of 125 ng of DNA, consisting of 25 ng of dORF reporter plasmid, 25 ng of GFP expressing plasmid to use as transfection control, and 75 ng of plasmids expressing miR-1 or miR-122. Cells were collected 36 hours post-transfection for cytometry and RNA extraction.

2.2.5 Cytometry analysis

The transfected cells were trypsinized and resuspended in DMEM supplemented with 10% FBS. The fluorescent reporter intensity of the cells was quantified using the Ze5 flow cytometer with mCherry (587/610) and GFP (488/510) lasers. Cells were analyzed without fixation. Cytometry data were saved as .fcs files and processed using FCS Express 7 software. Fluorescence intensity was represented as the median intensity of the cell population.

2.2.6 RNA extraction and RT-qPCR

Following trypsinization, transfected cells were resuspended in complete DMEM. A portion of the cells was used for cytometry analysis, while the remaining cells were collected, pelleted, and processed for RNA extraction. RNA was extracted using TRIzol™ Reagent according to the manufacturer's protocol, and its concentration was measured using the Nanodrop spectrophotometer. SuperScript™ IV Reverse Transcriptase kit from Invitrogen was used for the cDNA (complementary DNA) synthesis. RT-qPCR was performed using the PerfeCTa SYBR Green with Low ROX. All

qPCR primers were assay-validated in advance, and only those with amplification efficiencies between 90% and 110% were selected for experiments.

2.2.7 qPCR analysis

Thermo Fisher Connect™ was used to analyze the qPCR data, and Microsoft Excel was used to determine expression levels using the ΔΔCT analysis method.

2.2.8 Data analysis and plotting

The calculations for the data obtained from FCS express software and Thermo Fisher Connect™ were done in Microsoft Excel. Jupyter Notebook was used to plot the data. One-tailed t-test was used to calculate the significance of the data.

Chapter 3

Results

3.1 Investigating the effect of incorporation poly(A) sequence in dORFs reporter transcript expression

3.1.1 Engineering of dORF reporters containing translated polyAs

To investigate if the presence of PolyA sequences in the coding sequence of dORF would have any effect on the transcripts' half-life or protein output. The translation of poly-A tracks reduces transcript stability through ribosome stalling, leading to no-go decay (Arthur *et al.*, 2015). We designed reporters with mCherry as the main ORF and included dORF sequences with polyAs. To specifically assess the mRNA decay and the protein output of our reporters, we transfected *in vitro* transcribed mRNAs into HEK-293T cells to eliminate the contribution by transcription. mRNA of dORF reporters was co-transfected with mRNA of GFP as a normalization control. After 24 hours, cells were harvested, and a subset was taken for cytometry using mCherry and GFP lasers; only the cells positive for both the fluorophores were analyzed to assess the protein output. The transfected cells were collected and lysed for RNA extraction and RT-qPCR (reverse transcription-quantitative polymerase chain reaction) to evaluate reporter transcript stability.

For ease of data visualization and representation, the relative mCherry/GFP expression in dMUT reporters was set to 1, and the expression levels of the corresponding reporter were calculated relative to the dMUT reporter. Untransfected HEK-293T cells were used as the negative control for cytometry to define gating and optimize the flow cytometer's settings. For RT-qPCR, $\Delta\Delta$CT values were calculated with respect to the dMUT reporters as well.

To establish that we can observe the No-Go Decay (NGD) due to poly-A with our reporter assay, a P2A (porcine teschovirus-1 2A) peptide was cloned immediately after the main ORF, mCherry to allow for translation of polyA sequence in main ORF **(Figure**

7A i). The P2A sequence induces a ribosome-skipping event, resulting in two unfused proteins: the protein upstream of the 2A (mCherry) and the protein downstream of the 2A (containing poly-A tail translated as lysines) (Liu *et al.*, 2017). We compared the expression of a reporter where a stop codon was introduced immediately after the P2A sequence to determine the effect of ribosomes translating poly-A tracks **(Figure 7A ii)**. We observed that the activity of ribosomes over the multiple adenine bases resulted in a ~250% reduction in protein levels (t-test, $P<0.01$) **(Figure 7D)** and ~170% reduction in the transcript level **(Figure 7E)** in our positive control containing the P2A sequence. These findings are consistent with the results reported in the literature (Arthur *et al.* 2015), suggesting that our system can be used to test our hypothesis.

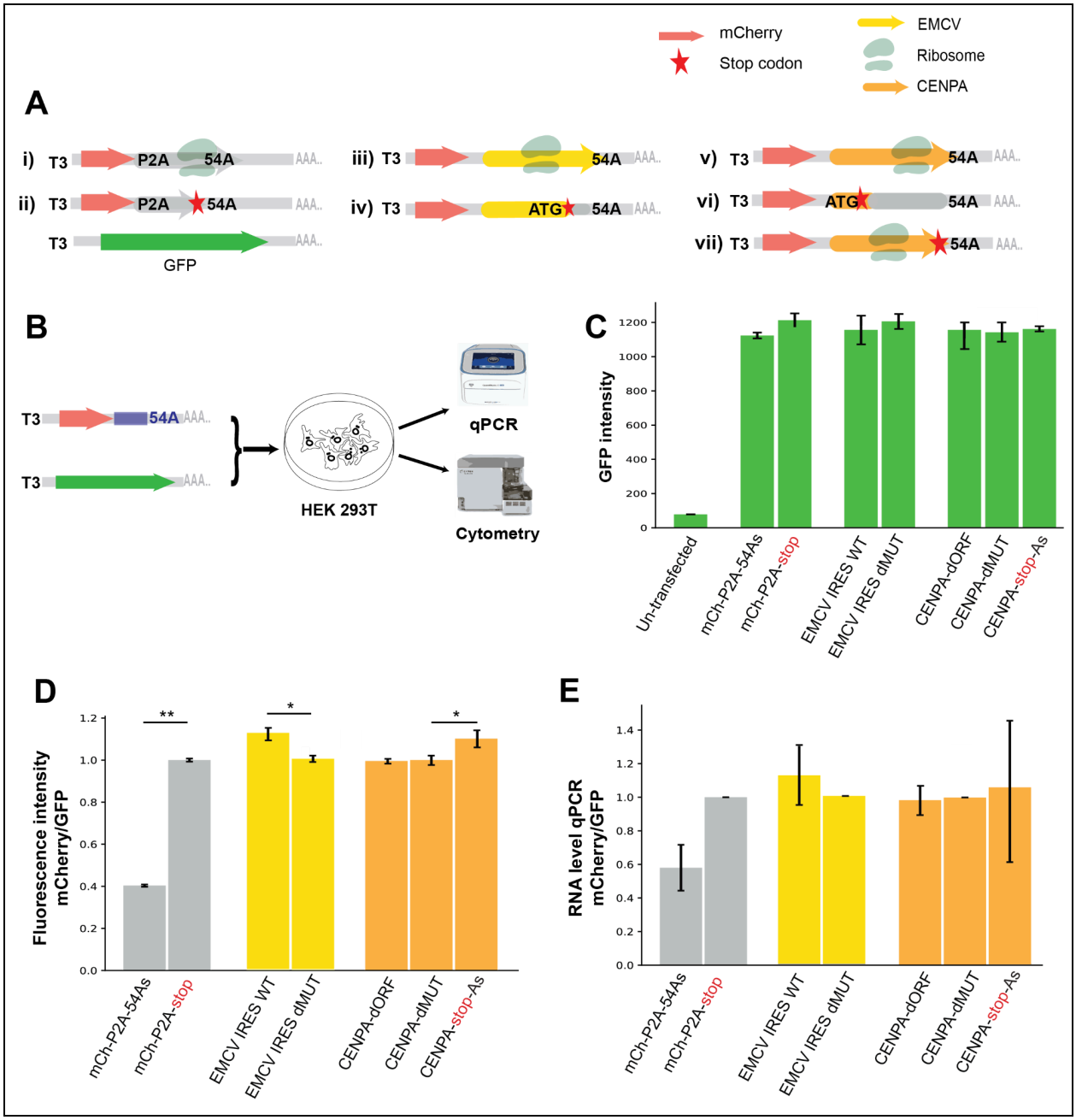

Figure 7: Experimental design to assess the effect of translation of polyA sequences within dORFs and the translation of poly-A sequences within the dORFs does not seem to cause decay of the reporters.

A. Cartoons showing the reporter constructs used for the experiment.

B. Cartoon showing the workflow of the experiment.

C. Barplot showing the transfection control (GFP) for the experiment. Error bars represent standard deviation. Differences in the compared values are not significant unless mentioned otherwise. The cytometry experiment has three biological replicates.

*D. Bar plot showing the normalized fluorescence intensity (mCherry/GFP) for the reporter constructs. The values are normalized to the relative expression levels of respective dMUTs. Data represent the mean of three biological replicates from flow cytometry experiments, with error bars representing standard deviation. Statistical significance was determined using a t-test, where ** represents P<0.01 and * P<0.05.*

E. Bar plot showing the relative RNA levels quantified using RT-qPCR. ΔΔCt values were calculated relative to the respective dMUTs. Error bars represent standard deviation. For qPCR, two biological replicates with three technical replicates were done.

3.1.2 Translation of polyA sequences within dORFs does not seem to cause significant transcript decay

As a positive control for dORF translation, we introduced the EMCV IRES (shown to initiate cap-independent translation). The EMCV IRES has been shown to drive translation in the bicistronic reporter assays performed in the Bazzini lab. The EMCV IRES was cloned downstream of the mCherry, and the poly-A track was present in the dORF **(Figure 7A iii)**. A mutated version of the EMCV dORF with a stop codon as the second codon was cloned as a negative control that would prevent translation of the poly-A track **(Figure 7A iv)**.

One dORF tested was found in the 3′ UTR of *CENP-A,* which was verified in the Bazzini lab to contain a translated dORF in a bicistronic reporter assay **(Figure 7A v)**. The iUTR and dORF from *CENP-A* were cloned in between the mCherry and poly-A track, and a truncated version (dMUT) of the iUTR and dORF was created by introducing a stop codon in the second codon of the dORF to prevent translation of the poly-A sequence **(Figure 7A vi)**. Another control includes a reporter with a stop codon immediately after the dORF but before the poly-A sequence. This ensures that while dORF is translated,

the poly-A track is not **(Figure 7A vii)**. We refer to this construct as dORFstop.As. This construct will allow us to assess the impact of dORF translation specifically while preventing ribosomes from encountering the poly-A track.

The reporter constructs containing the wild-type (WT) EMCV IRES displayed higher relative mCherry expression of about 12% (t-test, $P<0.05$) compared to the dMUT version. For the reporters containing the WT and dMUT sequence from the *CENP-A* gene, flow cytometry analysis showed comparable levels of normalized mCherry expression. However, dORFstop.As exhibited ~10% higher mCherry expression (t-test, $P<0.05$). This suggests a significant difference in expression levels based on whether the dORF is translated and if the ribosomes engage with the poly-A sequence **(Figure 7D)**.

Since the proposed mechanism of NGD is thought to act after the main ORF is translated—in this case, mCherry—we reasoned that quantifying the levels of reporter mRNA by RT-qPCR could provide us with additional insight into our hypothesis. The transfected cells were collected and lysed for RNA extraction, and RT-qPCR. Cycle threshold (Ct) values for mCherry were normalized to the Ct values of GFP, and $\Delta\Delta$Ct values of WT reporters were calculated relative to the dMUT reporters. The qPCR results indicated trends similar to the cytometry results. However, no significant differences in the mRNA abundances were observed among the reporters being compared **(Figure 7E)**. Results from the reporter experiments suggested that the translation of polyA sequences present within the dORFs can not trigger NGD.

3.2 The effect of dORF translation on miRNA targeting efficiency

3.2.1 miRNA target sites are responsive to miR transfection

To assess whether our reporter plasmid with miR-1 target sites responds to the over-expression of the miR-1. To test this, we designed a construct with mCherry as the main ORF and two tandem miR-1 target seeds in the 3′ UTR. The reporter plasmid was

co-transfected into HEK293T cells alongside plasmids that directly express miR-1 or miR-122 and a plasmid expressing GFP to use as transfection control. miR-122 was used as a control for the non-specific target effects as the target seeds of miR-122 (ACACTCCT) are similar to the miR-1 target 8-mer (ACATTCCA) but not predicted to be bound by or repressed by miR-1. The mCherry and GFP plasmids do not contain other target sites for miR-1 or miR-122. The cells were collected 36 hours post-transfection to perform cytometry and analyze the mCherry expression. The median fluorescent of mCherry in double positive cells was normalized to the median fluorescent intensity of GFP. Untransfected HEK-293T cells were used as the negative control for cytometry to define gating and optimize the flow cytometer's settings. We measured mCherry expression from cells with transfected reporter plasmids and miR plasmids (either miR-1 or miR-122) to quantify the impact of miR targeting on mCherry and detect any potential off-target effects of miR overexpression. Flow cytometry analysis of co-transfected parental plasmid with miR-1 and miR-122 showed comparable levels of GFP fluorescence across all wells, confirming similar transfection efficiencies **(Figure 8B)**. For ease of data visualization and representation, the relative mCherry/GFP expression in the miR-1 overexpression condition was set to 1, and the expression levels in the miR-122 condition were calculated relative to this baseline. The cytometry results revealed decreased mCherry expression (decrement of ~250%) in miR-1-expressing cells compared to miR-122 expressing cells **(Figure 8C)**, confirming the specificity and impact of miRNA-mediated regulation in our reporter system.

Next, we measured mCherry RNA levels normalized to GFP using qPCR to assess the impact of miRNA targeting. Cells transfected with miR-1 exhibited nearly a two-fold reduction in mCherry mRNA levels compared to those transfected with miR-122, which lacks complementary target sites in the reporter **(Figure 8D)**. The findings are consistent with the other studies reported in the literature highlighting the translational repression by miRNAs. The qPCR results were consistent with our cytometry data, showing almost five-fold less mRNA of mCherry in the case of cognate miR co-transfection as compared to the overexpression of miR-122, which did not have target sites in the plasmid.

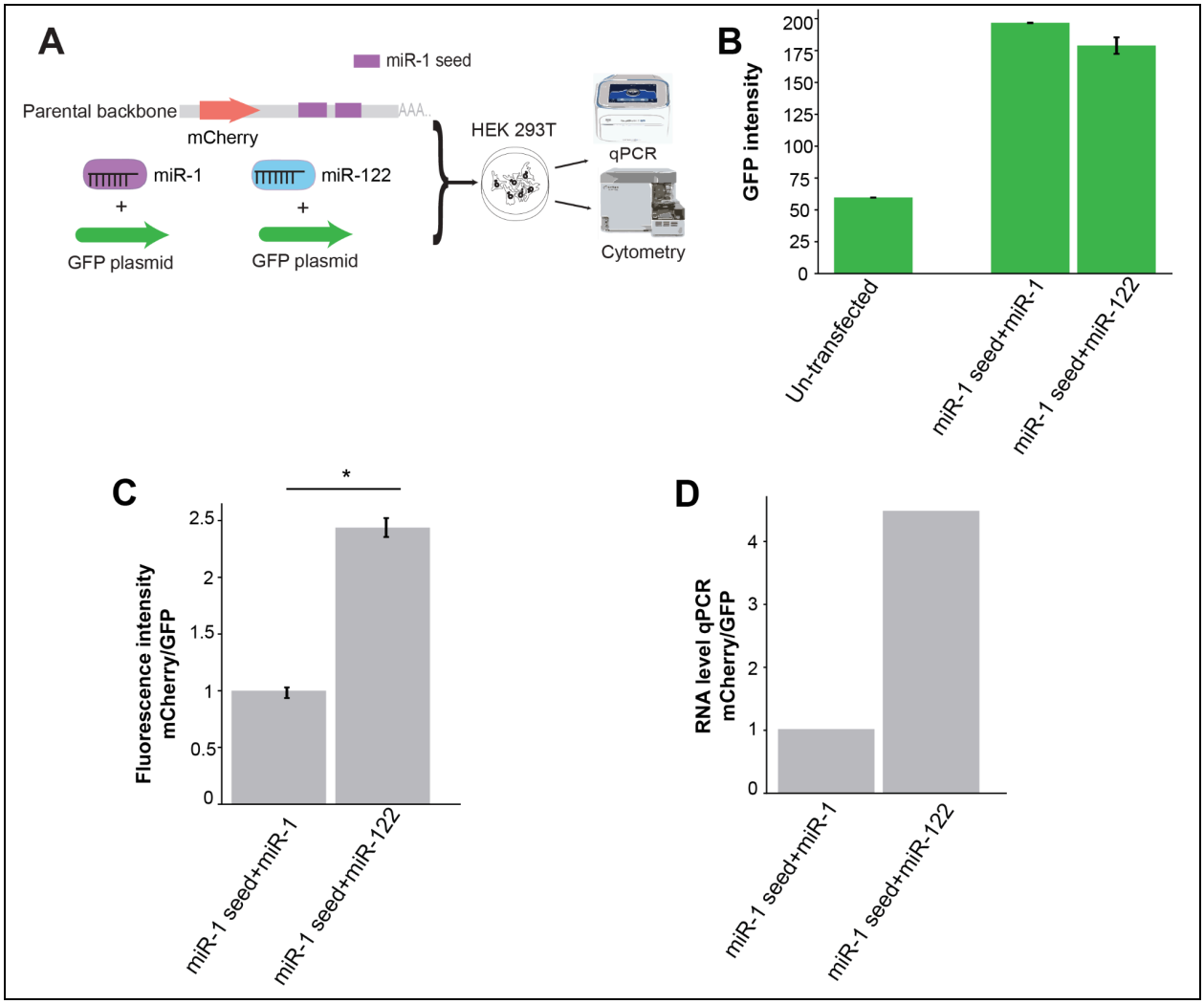

Figure 8: miR-1 target sites within 3' UTR of mCherry are responsive to the miR-1 overexpression.

A. Schematic showing the experimental design for the pilot experiment.

B. Barplot showing similar levels of GFP expression from flow cytometry when co-transfected with reporter and miR-expressing plasmids.

*C. Barplot showing the normalized fluorescence intensity (mCherry/GFP) for the parental plasmids co-transfected with either miR-1 or miR-122 expressing plasmids. The values are normalized to the relative expression levels of mCherry in the miR-1 overexpression. Data represent the mean of two biological replicates from flow cytometry experiments, with error bars representing standard deviation. Statistical Significance was determined using a t-test, where * represents P<0.05.*

D. Bar plot showing the relative RNA levels quantified using RT-qPCR. ΔΔCt values were calculated relative to the miR-1 overexpression condition. For qPCR, one biological replicates with three technical replicates were performed.

3.2.2 Reporters with downstream open reading frames are not responsive to the miR-mediated regulation

To check if dORF translation can reduce the efficiency of the miR-mediated repression, we cloned the iUTR and dORF from *CENP-A* was cloned downstream to the mCherry. ensuring that the miR-1 target sites were in the dORF **(Figure 9A i)**. A truncated version of the dORF was created by introducing a stop codon right after the first AUG to prevent dORF translation, called dMUT **(Figure 9A ii)**. As a positive control for dORF translation, we used the sequence from the EMCV IRES sequence and cloned downstream of the mCherry, ensuring that the miR-1 target seeds were present in the dORF **(Figure 9A iii)**. A mutated version was created to assess the effect of ribosomes not being present near the miR-1 target seeds; a stop codon was introduced right after the first start codon of the EMCV IRES and dORF. We refer to the construct as EMCV dMUT **(Figure 9A iv)**. Untransfected HEK-293T cells were used as the negative control. Flow cytometry analysis showed the normalized mCherry expression remained unchanged between the cells transfected with miR-1 target reporters, regardless of whether the cells expressed miR-1 or miR-122 **(Figure 9C)**. The findings suggest that the miRNA sites in our reporters are not responding to the miRNA-mediated repression of mCherry expression.

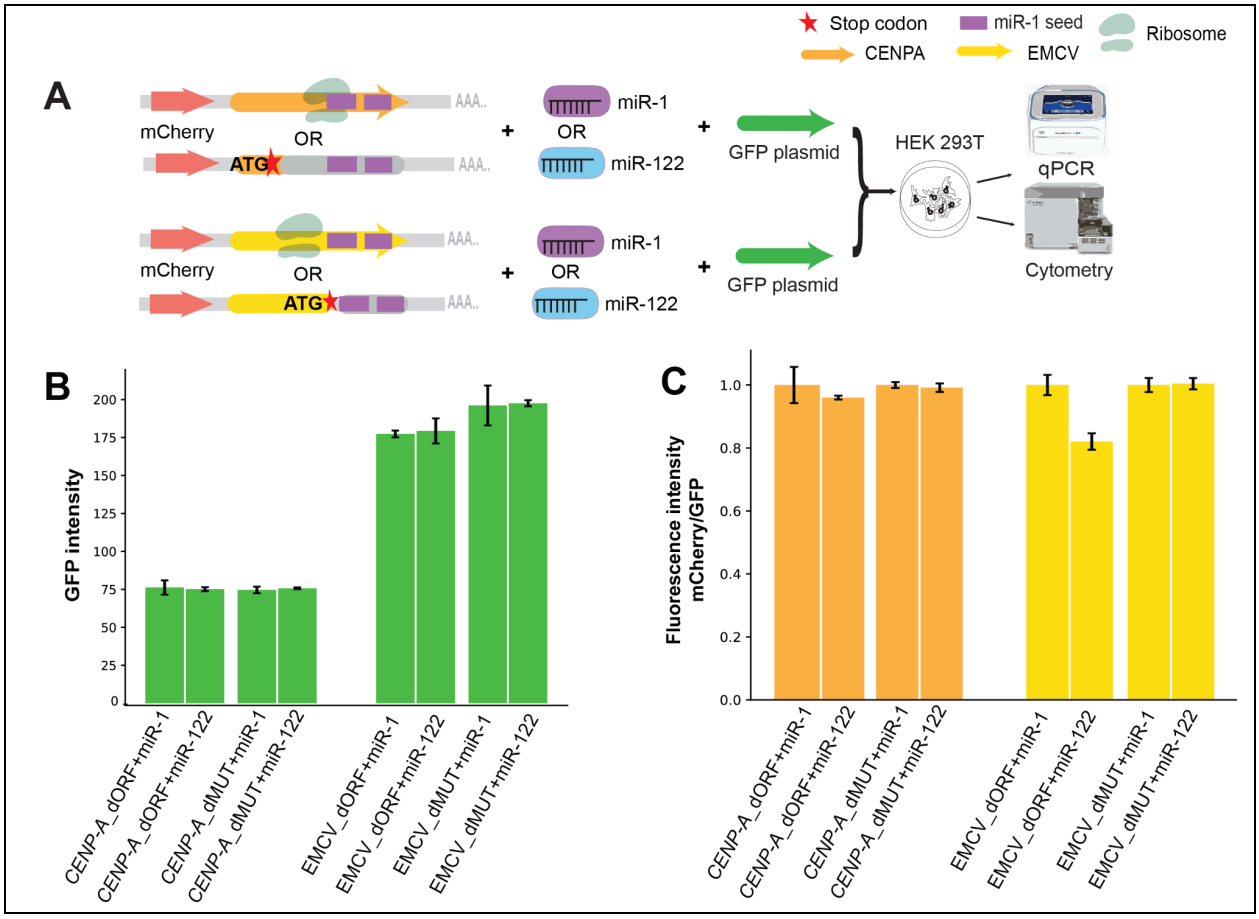

Figure 9: Reporters with downstream open reading frames are not responsive to the miR transfection.

A. Cartoon showing the reporter constructs and the experimental design.

B. Barplot showing similar levels of GFP expression from flow cytometry when co-transfected with reporter and miR-expressing plasmids. Three biological replicates were performed.

C. Barplot showing the normalized fluorescence intensity (mCherry/GFP) for the parental plasmids co-transfected with either miR-1 or miR-122 expressing plasmids. The values are normalized to the relative expression levels of mCherry in the miR-1 overexpression. Data represent the mean of three biological replicates from flow cytometry experiments, with error bars representing standard deviation.

The result from the previous experiment necessitated further investigation to address the potential explanation for the observations. Our hypothesis was that introducing dORF sequences between mCherry and the miR target seeds led to an unexpected splicing event. To address this, we extracted RNA from the cells that were transfected with our reporters containing the *CENP-A* iUTR and dORF/dMUT, followed by a reverse transcription reaction to make cDNA **(Figure 10A)**. We then performed the PCR using primers that can amplify the mCherry and the inserted sequences, which include dORF and miR target seeds. RNA from untransfected cells and no-template control were included as negative controls to rule out any non-specific amplification. RT-PCR result revealed a single amplicon of size ~1350 base pairs. The RT-PCR did not detect any aberrant splicing events, as no unexpected amplicons were observed, ruling out abnormal splicing as a cause **(Figure 10B)**.

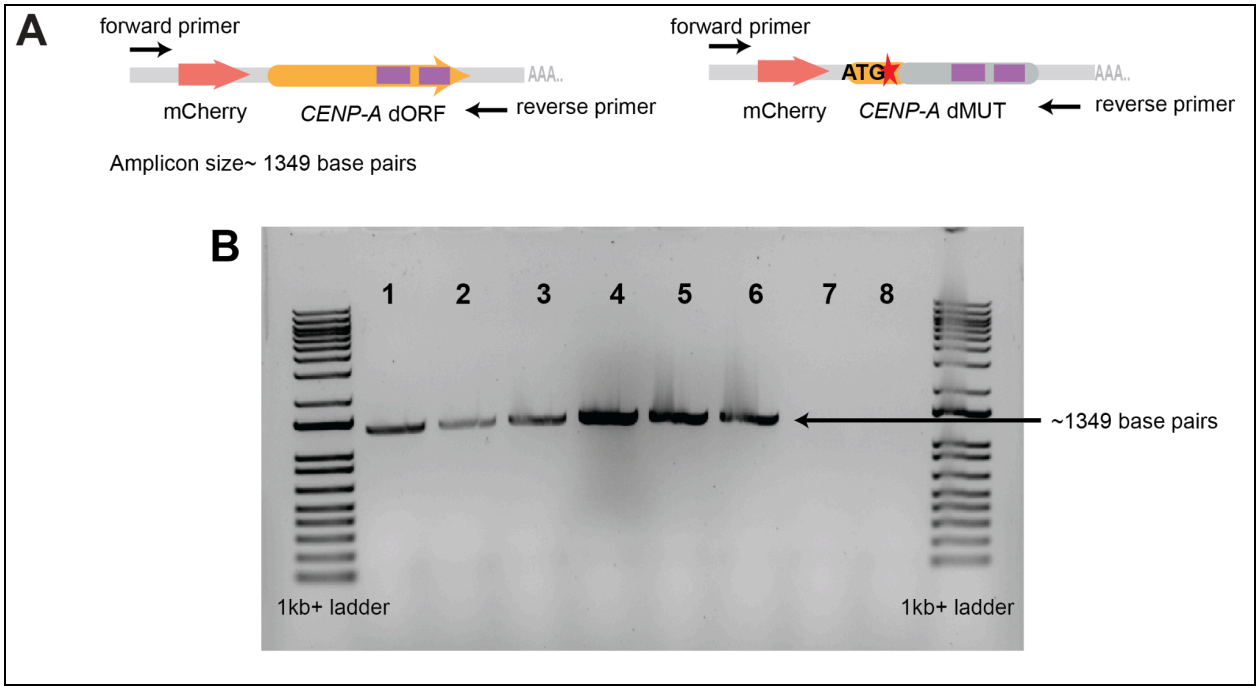

Figure 10: Figure showing RT-PCR analysis to assess potential splicing events in dORF-containing reporters.

A. Schematic representation of the experimental workflow.

B. PCR amplification of the mCherry and inserted sequences revealed a single amplicon of ~1350 bp. No unexpected bands were observed. Wells represents the RNA extracted from the transfections. 1: *CENP-A dORF*+GFP+miR-1; 2: *CENP-A*

dORF+GFP+miR-122; 3: *CENP-A dORF*+GFP; 4: *CENP-A dMUT*+GFP+miR-1; 5: *CENP-A dMUT*+GFP+miR-122; 6: *CENP-A dMUT*+GFP; 7: Untransfected; 8: No template.

Chapter 4

Discussion

4.1 dORFs containing poly-A sequences

The results from our reporter experiments assessing the effect of translating dORFs containing poly-A sequences indicate that ribosomes translating these sequences do not significantly impact mRNA stability or protein output. No significant differences were observed in either mCherry fluorescence or mRNA abundance between the reporters being compared. The findings contrast with the previous studies, which demonstrated that the translation of poly-A sequences in the main ORF can trigger NGD and cause transcript decay (Arthur *et al., 2015*). One possible explanation for this discrepancy could be the efficiency of ribosome recruitment by iUTRs for dORF translation. It is possible that the density of ribosomes translating the dORFs is lower, thus reducing the potential ribosome stalling and subsequent NGD.

The expression pattern observed for the EMCV dORF and dMUT reporters may be attributed to the characteristics of group 3 IRESs, which can initiate cap-independent translation using a subset of the initiation factors required for canonical ORF translation (Kieft, 2008). A similar mechanism may also be true for dORF reporters, as iUTRs are also believed to recruit ribosomes for dORF translation (Wu *et al.*, 2020). It is possible that the full complement of translation factors is likely required for ribosomes to engage in quality control mechanisms, such as no-go decay. The differences in translation factor usage between ribosomes recruited by EMCV IRES/dORF iUTRs and those initiating at the main ORF may explain the reduced mCherry decay in our reporters.

4.2 miRNA target seed within the dORFs

The results from the pilot experiment, in which miR-1 target seeds were present in the 3′ UTR of the mCherry, suggest that the miR-1 seeds are specifically responsive to the

miR-1 overexpression. The findings are consistent with the previous studies showing that miRNA-mediated repression is more efficient when the target seeds are present in the 3' UTR (Gu *et al.*, 2009). However, the results from the reporters containing the miR-1 target sites within the dORF do not show repression of the target reporters.

Our hypothesis was that the dORF translation would sterically hinder the miRNA-AGO complex from effectively repressing the target mRNA output due to the presence of a bulky group, such as the ribosome. However, the results from our experiments indicate that the levels of normalized mCherry remain the same regardless of miR-1 or miR-122 overexpression.

Several explanations could account for this observation. One possibility is the occurrence of any out-of-frame translation events that allow the ribosomes to traverse through the sequences and interfere with miR-mediated repression. To test this, additional cloning is required to introduce stop codons in all three frames, ensuring the termination of any out-of-frame translation events. Alternatively, It has been shown previously that target site accessibility is an important factor in determining the efficiency of miRNA-mediated translational repression (Kertesz *et al.*, 2007; Long *et al.*, 2007). It is possible that the introduction of additional sequences in our reporters might have increased the propensity of the mRNA to form stable secondary structures. These stable secondary structures have the potential to prevent the miR from accessing the target sites efficiently, thus saving the mRNAs from being targeted.

References

Arthur, LL, Pavlovic-Djuranovic, S, Koutmou, KS, Green, R, Szczesny, P, and Djuranovic, S (2015). Translational control by lysine-encoding A-rich sequences. Sci Adv 1, e1500154.

Barrett, LW, Fletcher, S, and Wilton, SD (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cell Mol Life Sci 69, 3613–3634.

Bazzini, AA, Johnstone, TG, Christiano, R, Mackowiak, SD, Obermayer, B, Fleming, ES, Vejnar, CE, Lee, MT, Rajewsky, N, Walther, TC, *et al.* (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J 33, 981–993.

Crick (1970). Central Dogma of Molecular Biology | Nature. Available at: https://www.nature.com/articles/227561a0. Accessed March 7, 2025.

Cobb, M (2017). 60 years ago, Francis Crick changed the logic of biology. PLoS Biol 15, e2003243.

Dodbele, S, and Wilusz, JE (2020). Ending on a high note: Downstream ORFs enhance mRNA translational output. EMBO J 39, e105959.

Dueck, A, Ziegler, C, Eichner, A, Berezikov, E, and Meister, G (2012). microRNAs associated with the different human Argonaute proteins. Nucleic Acids Res 40, 9850–9862.

Gavis, ER, and Lehmann, R (1992). Localization of nanos RNA controls embryonic polarity. Cell 71, 301–313.

Gu, S, Jin, L, Zhang, F, Sarnow, P, and Kay, MA (2009). Biological basis for restriction of microRNA targets to the 3’ untranslated region in mammalian mRNAs. Nat Struct Mol Biol 16, 144–150.

Hanson, G, Alhusaini, N, Morris, N, Sweet, T, and Coller, J (2018). Translation elongation and mRNA stability are coupled through the ribosomal A-site. RNA 24, 1377–1389.

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

Ingolia, NT, Brar, GA, Stern-Ginossar, N, Harris, MS, Talhouarne, GJS, Jackson, SE, Wills, MR, and Weissman, JS (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. Cell Rep 8, 1365–1379.

Ingolia, NT, Ghaemmaghami, S, Newman, JRS, and Weissman, JS (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science 324, 218–223.

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. Nature 431, 931–945.

Johnstone, TG, Bazzini, AA, and Giraldez, AJ (2016). Upstream ORFs are prevalent translational repressors in vertebrates. EMBO J 35, 706–723.

Kertesz, M, Iovino, N, Unnerstall, U, Gaul, U, and Segal, E (2007). The role of site accessibility in microRNA target recognition. Nat Genet 39, 1278–1284.

Kieft, JS (2008). Viral IRES RNA structures and ribosome interactions. Trends Biochem Sci 33, 274–283.

Kute, PM, Soukarieh, O, Tjeldnes, H, Trégouët, D-A, and Valen, E (2022). Small Open Reading Frames, How to Find Them and Determine Their Function. Front Genet 12.

Kowalski, MH, Wessels, H-H, Linder, J, Dalgarno, C, Mascio, I, Choudhary, S, Hartman, A, Hao, Y, Kundaje, A, and Satija, R (2024). Multiplexed single-cell characterization of alternative polyadenylation regulators. Cell 187, 4408-4425.e23.

Lambert, SA, Jolma, A, Campitelli, LF, Das, PK, Yin, Y, Albu, M, Chen, X, Taipale, J, Hughes, TR, and Weirauch, MT (2018). The Human Transcription Factors. Cell 172, 650–665.

Lareau, LF, Hite, DH, Hogan, GJ, and Brown, PO (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. eLife 3, e01257.

Lee, TI, and Young, RA (2013). Transcriptional Regulation and Its Misregulation in Disease. Cell 152, 1237–1251.

Lefèvre, C, Cook, GM, Dinan, AM, Torii, S, Stewart, H, Gibbons, G, Nicholson, AS, Echavarría-Consuegra, L, Meredith, LW, Lulla, V, *et al.* (2024). Zika viruses encode 5′ upstream open reading frames affecting infection of human brain cells. Nat Commun 15, 8822.

Liu, Z, Chen, O, Wall, JBJ, Zheng, M, Zhou, Y, Wang, L, Ruth Vaseghi, H, Qian, L, and Liu, J (2017a). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci Rep 7, 2193.

Long, D, Lee, R, Williams, P, Chan, CY, Ambros, V, and Ding, Y (2007). Potent effect of target structure on microRNA function. Nat Struct Mol Biol 14, 287–294.

Matoulkova, E, Michalova, E, Vojtesek, B, and Hrstka, R (2012). The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biology 9, 563–576.

McGeary, SE, Lin, KS, Shi, CY, Pham, TM, Bisaria, N, Kelley, GM, and Bartel, DP (2019). The biochemical basis of microRNA targeting efficacy. Science 366, eaav1741.

Mignone, F, Gissi, C, Liuni, S, and Pesole, G (2002). Untranslated regions of mRNAs. Genome Biol 3, reviews0004.1.

Moore, MJ, and Proudfoot, NJ (2009). Pre-mRNA Processing Reaches Back toTranscription and Ahead to Translation. Cell 136, 688–700.

Park, EG, Ha, H, Lee, DH, Kim, WR, Lee, YJ, Bae, WH, and Kim, H-S (2022). Genomic Analyses of Non-Coding RNAs Overlapping Transposable Elements and Its Implication to Human Diseases. Int J Mol Sci 23, 8950.

Peng, SS, Chen, CY, Xu, N, and Shyu, AB (1998). RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. EMBO J 17, 3461–3470.

Pesole, G, Mignone, F, Gissi, C, Grillo, G, Licciulli, F, and Liuni, S (2001). Structural and functional features of eukaryotic mRNA untranslated regions. Gene 276, 73–81.

Pray, L. (2008) Discovery of DNA structure and function: Watson and Crick. Nature Education 1(1):100

Ringnér, M, and Krogh, M (2005). Folding Free Energies of 5′-UTRs Impact Post-Transcriptional Regulation on a Genomic Scale in Yeast. PLOS Comput Biol 1, e72.

Shah, AN, Leesch, F, Lorenzo-Orts, L, Grundmann, L, Novatchkova, M, Haselbach, D, Calo, E, and Pauli, A (2024). A dual ribosomal system in the zebrafish soma and germline. 2024.08.29.610041.

Simms, CL, Kim, KQ, Yan, LL, Qiu, J, and Zaher, HS (2018). Interactions between the mRNA and Rps3/uS3 at the entry tunnel of the ribosomal small subunit are important for

no-go decay. PLOS Genetics 14, e1007818.

Somers, J, Pöyry, T, and Willis, AE (2013). A perspective on mammalian upstream open reading frame function. Int J Biochem Cell Biol 45, 1690–1700.

Takahashi, K, and Yamanaka, S (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 126, 663–676.

Venter, JC, Adams, MD, Myers, EW, Li, PW, Mural, RJ, Sutton, GG, Smith, HO, Yandell, M, Evans, CA, Holt, RA, *et al.* (2001). The sequence of the human genome. Science 291, 1304–1351.

Walsh, C Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications - Walsh - 2005 - Angewandte Chemie International Edition - Wiley Online Library. Available at: https://onlinelibrary.wiley.com/doi/10.1002/anie.200501023. Accessed March 7, 2025.

Wang, C, and Lehmann, R (1991). Nanos is the localized posterior determinant in Drosophila. Cell 66, 637–647.

Wieder, N, D'Souza, EN, Martin-Geary, AC, Lassen, FH, Talbot-Martin, J, Fernandes, M, Chothani, SP, Rackham, OJL, Schafer, S, Aspden, JL, *et al.* (2024). Differences in 5'untranslated regions highlight the importance of translational regulation of dosage sensitive genes. Genome Biol 25, 111.

Wu, Q, and Bazzini, AA (2018). Systems to study codon effect on post-transcriptional regulation of gene expression. Methods 137, 82–89.

Wu, Q, Medina, SG, Kushawah, G, DeVore, ML, Castellano, LA, Hand, JM, Wright, M, and Bazzini, AA (2019). Translation affects mRNA stability in a codon-dependent manner in human cells. eLife 8, e45396.

Wu, Q, Wright, M, Gogol, MM, Bradford, WD, Zhang, N, and Bazzini, AA (2020). Translation of small downstream ORFs enhances translation of canonical main open reading frames. EMBO J 39, e104763.

Zhong, Q, Xiao, X, Qiu, Y, Xu, Z, Chen, C, Chong, B, Zhao, X, Hai, S, Li, S, An, Z, *et al.* (2023). Protein posttranslational modifications in health and diseases: Functions, regulatory mechanisms, and therapeutic implications. MedComm 4, e261.

Zubiaga, AM, Belasco, JG, and Greenberg, ME (1995). The Nonamer UUAUUUAUU Is the Key AU-Rich Sequence Motif That Mediates mRNA Degradation. Molecular and Cellular Biology 15, 2219–2230.

Primers

Serial no.	Name	Sequence
1	5861_miRNA_targetsite_fw	gcatctgcagTCGACGAATTC
2	5862_miRNA_targetsite_rv	atgcttacgtaatacgactcact
3	91_mir1_targetsite_coro1c	gcatctgcagTCGACGAATTCGGCCTGTACAatttggagacattccatttcagatttgtcaaccagcgataggccacattccagtaagaacACCGGTTGAaTAGcTAATTAATTAAGAAGAGCACACGTCTGAACTCCAGTCACActagaactatagtgagtcgtattacgta
4	5547_Rv_seq_main_ORFome	gcagcgtatccacatagcgt
5	3835_seq_TRE_bicis_2_Fw_CB016	gtcatcagaaaggatgtaccc
6	5386_Fw_CENPA_STARR_dORF	TAGAGCATGCACCGGCCTGCACCCAGTGTTTCTGT
7	5387_Rv_CENPA_STARR_dORF	GGCCGAATTCGTCGAAAGCAACACACACATACTGT
8	5661_Fw_CENPA_complete_STARR	TAGAGCATGCACCGGGCTCCTGCACCCAGTGTTTCTGT
9	5663_Rv_CENPA_complete_enhanc	GGCCGAATTCGTCGATTAAAGCAACACACACATACTGTTC
10	3836_seq_TRE_bicis_3_Rv_CB017	CCTTGGAAAAGGCGCAACC
11	5536_Fw_gBlock_main	catttcaggtgtcgtgaggc
12	5537_Rv_gBlock_main	agaggttgattgtcgactta
13	6317_t3_mCh_orfeome	aattaaccctcactaaagggacatttcaggtgtcgtgaggc
14	6318_60xT_orfeome	ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttagaggttgattgtcgactta
15	6121_seq_newBB_mCherry2_fw_CB268	aatacgaaagggcagaaggga
16	3959_mCherry_Seq_Rv_CB138	actgacccttccatatgaacttt
17	6374_starrseq_fw	TAGAGCATGCACCGG
18	6375_starrseq_rv	GGCCGAATTCGTCGA
19	6376_ORFeome_mcherry_wo_stop_fw	TAGAGCATGCACCGGCGTAC
20	6377_ORFeome_mcherry_wo_stop_rv	CATTTGTAAAGTTCATCCATCCCTCC
21	6383_stop codons_polyA_cenpa	GGCCGAATTCGTCtAAAGCAACtCACACA
22	6403_remove_mCh_stop1_fw	gaactttacaaaTATAGAGCATGCACCGG

23	6404_remove_mCh_stop1_rv	gagacccgtttatgtatcgagctaggcact
24	6463_remove_stop_mCh_orf_3fw	tgtatttaagtgcctagctcgatacataaacgggtctctc
25	6464_remove_stop_mCh_orf_3rv	CCGGTGCATGCTCTATAtttgtaaagttcatccatcc ctcctg
26	6383_stop codons_polyA_cenpa	GGCCGAATTCGTCtAAAGCAACtCACACA
27	6530_mch_qpcr_a_fw	tcagatgggccagttatgca
28	6531_mch_qpcr_a_rv	atgcccctggaagttgaact
29	704_GFP FRWD	TGGGTCAGTTCAACTTGCAG
30	705_GFP_RVRSE	GAAAGCGCTGATTGTGTTGA
31	261-ACTB_F	ACCTTCTACAATGAGCTGCG
32	262-ACTB_R	CCTGGATAGCAACGTACATGG
33	3848_Gib_pCW57_Qbi_Rv_1_CB029	AGGCGCAACCCCAACCCCGGATCacctgaaa cataaaatgaatgc
34	3851_seq_pCW57_Fw_CB032	CCTATATAAGCAGAGCTCGT