Characterization of the role of *grainy head's* 5'UTR uORFs in regulating translation during *Drosophila melanogaster* development

by

Mathew Pattee Kirby

B.Sc., The University of Guelph, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Cell and Developmental Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis/dissertation entitled:

Characterization of the role of *grainy head's* 5'UTR uORFs in regulating translation during *Drosophila melanogaster* development

submitted by ___Mathew Pattee Kirby___ in partial fulfillment of the requirements for

the degree of ___Master of Science___

in ___Cell and Developmental Biology___

Examining Committee:

Dr. Eric Jan
Co-supervisor

Dr. Douglas Allan
Co-supervisor

Dr. Don Moerman
Supervisory Committee Member

Dr. Thibault Mayor
Additional Examiner

Additional Supervisory Committee Members:

Dr. Elizabeth Rideout
Supervisory Committee Member

Supervisory Committee Member

Abstract

Precise control of the levels and spatiotemporal domains of protein synthesis is fundamental to cellular processes. Regulation of protein synthesis largely occurs at the rate-limiting step of translation initiation in which the translation start site is selected by the scanning ribosomal pre-initiation complex (PIC) and its associated initiation factors. Upstream open reading frames (uORFs) are prevalent regulatory elements located in the 5' untranslated regions (5'UTR) of approximately 50% of mammalian transcripts. Generally, uORFs are viewed as constitutive repressors of translation of the downstream coding sequence (CDS) by sequestering ribosomes. Recent genome-wide studies have revealed that uORFs have widespread regulatory functions in different biological contexts, however our understanding of the roles played by uORFs is still in its infancy. In *Drosophila melanogaster,* the spatial and temporal expression of the transcription factor *grainy head (grh)* must be tightly controlled to ensure proper epithelial and central nervous system development. Intriguingly, *grh's* eight mRNA isoforms display uORF-containing 5'UTRs ranging from 1 to 24 uORFs. To test for a role of these uORFs in Grh function, this thesis attempts to characterize the role of *grh-RJ's* eleven uORFs in modulating the downstream CDS translation in order to fine-tune Grh's spatiotemporal expression throughout *Drosophila* development. In this study, both *in vitro* translation assays and *in vivo* genetic analyses were used to analyze the regulatory role of *grh-RJ's* uORFs on the downstream CDS translation. Our *in vitro* results showed that *grh-RJ's* eleven uORFs severely repressed translation of the downstream CDS in translation extracts. Meanwhile, our transgenic *in vivo* results showed that that *grh-RJ's* uORFs spatially restricted and repressed reporter expression in the *Drosophila* embryo. In general, we found that the role of *grh-RJ's* uORFs is to repress translation of the downstream CDS, including restricting the spatial expression of Grh during

Drosophila development. Together with the widespread prevalence of uORFs among species, this research suggests an extensive role of uORFs in regulating the level and spatiotemporal expression of proteins, which will likely contribute greatly to a fundamentally novel understanding of biological systems.

Lay Summary

One of the key questions that developmental biologists strive to answer is how a single cell can initially divide and produce different cell types that eventually form distinct tissues and organs. Throughout this developmental journey, these cells are influenced by multiple factors. These factors alter the gene expression profile of a cell based on the cell's environment and the timing of development. We are systematically analyzing the mechanism of how a single gene is controlled to produce a protein in the correct places and levels throughout fruit fly development. Determining the underlying process of how protein synthesis is regulated based on the cell's location and developmental timing will help advance our basic understanding of development and help pave the way for more effective treatments of human diseases.

Preface

All experiments were conducted in Dr. Douglas Allan and Dr. Eric Jan's Laboratory at UBC. Identification and design of this project was conducted by me in conjunction with my supervisors Dr. Eric Jan and Dr. Douglas Allan. The experiments and analysis were carried out by me with the exception listed below.

In **Chapter 1,** Figure 1.1 and 1.2 were reused with permission from applicable sources.

In **Chapter 3,** the cloning of monocistronic *grhRJ-5'UTR(WTorMut)-FLuc* reporter constructs were cloned by Mika Vivar (Jan Lab, UBC). All S2 cell transfections with *grhRJ-5'UTR(WTorMut)-FLuc* pACT and *pRLuc* vectors were conducted by Mika Vivar (Jan Lab, UBC). The *grh* HDR WT and Mut constructs were designed by Tianshun Lian (Allan Lab). RNA-sequencing performed on RNA isolated from 3rd Instar larval VNCs was conducted and analyzed by Robin Vuilleuimier and Stephane Filbotte (Allan Lab and Moerman Lab, UBC). All *Drosophila* uORFs and Cavener consensus sequence matches were mapped to the *D. melanogaster* genome by Shamsuddin Bhuiyan (Pavlidis lab, UBC). I created Table 3.1, which is modified from Supplementary Table S7 in Zhang et al. (2018). All of this work has been stated in the text.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAP	Arginine attenuator peptide
Abd-A	Abdominal-A
Ap	apterous
ATAC-seq	Transposase-accessible chromatin sequencing
ATF4	Activating transcription factor 4
ATP	Adenosine triphosphate
BiP	Binding immunoglobulin protein
C/EBPβ	CCAAT/enhancer binding protein β
Cas	Castor
cAUG	Coding AUG
cDNA	coding DNA
CDS	Coding sequence
ChIP-seq	Chromatin immunoprecipitation sequencing
CNS	central nervous system
Cora	Coracle
CP2	Connective peptide 2
CPA-1	Carboxypeptidase A1
CPE	Cytoplasmic polyadenylation elements
CRISPR	Clustered regularly interspaced short palindromic repeats
Cys	Cysteine
DA	daughterless
Ddc	Dopa decarboxylase

DE-Cad	*Drosophila* E-Cadherin
DNA	Deoxyribonucleic acid
Dpp	Decapentaplegic
eGFP	Enhanced GFP
eIF	Eukaryotic initiation factor
EMCV	Encephalomycarditis virus
ERK	Extracellular signal-regulated kinase
Fas3	Fasciclin 3
Fluc	Firefly luciferase
GADD34	Growth arrest and DNA damage-inducible protein
GCN2	General control protein GCN2
GCN4	General control protein GCN4
GDP	Guanine diphosphate
GFP	Green fluorescent protein
GMC	Ganglion mother cells
Grh	Grainy head
Grhl	Grainy head-like
gRNA	guide RNA
GTP	Guanine triphosphate
Hb	Hunchback
HDR	Homology directed repair
HRI	Hemw-regulated eIF2α
IgG	Immunoglobulin G

IGV	Integrative genomic view
IRES	Internal ribosome entry site
ITR	Inverted terminal repeats
Kr	Kruppel
LAP	Liver-activating protein
LIP	Liver inhibitory protein
m7G	7-methyl-guanosine
Met	Methionine
MM	multiple myeloma
mRNA	Messenger RNA
mTOR	Mammalian target of rapamycin
Mut	Mutant
NB	Neuroblast
nls	nuclear localization signal
NMD	Nonsense-mediated decay
nos	Nanos
nt	Nucleotides
NTF-1	Neurogenic element binding transcription factor (grh)
oORF	Out-of-frame overlapping uORF
PABP	Poly(A) binding protein
PAM	Protospacer adjacent motif
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction

Pdm	POU domain protein
PERK	PKR-like endoplasmic reticulum kinase
PFA	paraformaldehyde
PIC	Pre-initiation complex
PKR	Protein kinase R
prox	proximal
ple	Pale
pNB	Post mitotic neuroblasts
RLuc	Renilla luciferase
RNA	Ribonucleic acid
RNA-seq	RNA-sequencing
RRL	Rabbit reticulocyte lysate
RT-PCR	Reverse transcription PCR
S2	Schneider 2
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sfGFP	superfolder GFP
stit	Stitcher
SV40	Simian Virus 40
Svp	Seven up
TE	Translation efficiencies
TF	Transcription factor
tRNA	Transfer ribonucleic acid
TSS	Transcription start site

Tv	Thoracic-ventral
twi	twist
UAS	Upstream activating sequence
uAUG	Upstream AUG
uORF	Upstream open reading frame
UTR	Untranslated region
VNC	Ventral nerve cord
WT	Wild type

Acknowledgements

First and foremost, I would like to thank my two incredible supervisors Doug Allan and Eric Jan. It has been an honour to work in your labs and I am really grateful to have both of you as mentors. The past couple of years have pushed me to grow as a scientist and as a person. My graduate school experience has helped me develop critical thinking, analytical, and professional skills that will help me thrive and achieve my future goals in the biotechnology industry. Thanks again for being supportive mentors and giving me the freedom to explore unanswered questions and pursue interesting scientific endeavours.

I am grateful to have had the opportunity to work with many talented post-doctoral fellows and graduate students in both the Allan and Jan Lab. Thank you for creating a fun working environment and providing thought-provoking conversations. And most of all thank you for supporting and encouraging me these past two years, my graduate school experience would not be the same without you.

I would also like to thank the National Sciences and Engineering Council of Canada (NSERC), University of British Columbia, and Kruger Inc. for providing financial support and funding.

I would also like to thank a very supportive committee: Don Moerman and Elizabeth Rideout. Thank you for providing helpful feedback and mentorship over the last couple of years.

Finally, I would like thank my amazing friends. Thank you for believing in me and supporting me during difficult times. Spending time chatting and laughing over drinks during countless happy hours has made my time at UBC and Vancouver spectacular. I would also like to give special thanks to my Mom and Dad. Thanks for supporting me throughout my academic years, I would not have been able to do it without you. I love you guys!

Dedication

To my family and friends. Enjoy.

1. INTRODUCTION

1.1 Eukaryotic translation

1.1.1 Translation initiation

Protein synthesis is a main component in the Central Dogma of Molecular Biology and is an essential process in all forms of life. Translation includes initiation, elongation, termination, and ribosome recycling, that mediates the transfer of genetic information stored within messenger RNA (mRNA) into its nascent polypeptide. Translation initiation is the rate-limiting step in translation and is vital to translational regulation. Regulation at the initiation step allows for an immediate, reversible, and spatial response to adjust protein levels upon cellular signals or environmental stimuli.

In eukaryotes, translation of the majority of mRNAs is initiated via the cap-dependent scanning mechanism, which requires at least nine core eukaryotic initiation factors (eIFs) (reviewed in (Hinnebusch et al., 2016; Hinnebusch and Lorsch, 2012)). Eukaryotic mRNAs are comprised of two distinctive features that play intrinsic roles during initiation: the 5' cap and poly(A) tail. The modified 5' cap consists of a 7-methyl-guanosine (m^7G) that is linked to the mRNA through an inverted 5'-5' trisphosphate bridge (reviewed in (Shuman, 2002)). The eukaryotic 5' cap has many functions that include protecting the mRNA from being degraded by exonucleases, promoting pre-mRNA splicing, and allowing the pioneer round of translation for mRNA quality control (Edery and Sonenberg, 1985; Konarska et al., 1984; Lewis and Izaurflde, 1997; Maquat et al., 2010; Schwer et al., 1998).

Translation initiation begins with the formation of a 43S preinitiation complex (PIC), which is composed of the free 40S ribosomal subunit, the ternary complex (TC) eIF2-Met-tRNA$_i$-GTP, eIF1, eIF1A, eIF3, and eIF5 (reviewed in (Hinnebusch et al., 2016; Hinnebusch and

Lorsch, 2012)) (Figure 1.1). The 5' cap is first recognized by the cap-binding protein complex eiF4F, which is made up of three factors: (i) the cap binding protein, eIF4E; (ii) the RNA helicase, eIF4A; (iii) the adaptor protein, eIF4G. The adaptor protein, eIF4G, functions as a scaffold to mediate the circularization of the transcript through interaction with the poly(A) binding protein (PABP) bound to the 3' poly(A) tail. The interaction between eIF3 and eIF4G facilitates the recruitment of the PIC to the 5' cap allowing the ribosome to scan in the 5' to 3' direction along the 5' untranslated region (5'UTR) to reach an AUG initiation codon. Scanning is enabled by the helicase activity of eIF4A, which unwinds the local RNA structure, and with the combination of eIF1 and eIF1A, the ribosome adopts a scanning-competent 'open' conformation. To ensure fidelity in translation initiation and correct usage of the coding AUG (cAUG) codon, the scanning PIC discriminates against, and bypasses, unfavorable initiation codons in poor sequence context and/or near-cognate triplets (i.e. UUG, CUG, and GUG). In mammals, initiation typically occurs at an initiation codon that is directly surrounded by an optimal Kozak consensus sequence of GCC(A/G)CCAUGGG, where a purine is at the -3 position and a guanine is at the +4 position (where the A of the AUG start codon is assigned as +1) (Kozak, 1986; Kozak, 1991b). Similarly, in *Drosophila*, translation commonly initiates at a similar initiation codon that is located within an "optimal" Cavener consensus start motif of CAACAUG (Cavener, 1987). It is important to note that not all cAUGs are surrounded by optimal Kozak or Cavener sequence contexts, however, cAUGs that are located within a favorable sequence context tend to favour translation initiation.

Start codon recognition occurs through the complementary base-pairing between the anticodon of the Met-tRNA$_i$ and the AUG codon in the ribosomal P site. Upon anticodon-codon interaction, eIF1 is displaced and the 40S scanning ribosome adopts a 'closed' conformation and

remains bound to the mRNA (Passmore et al., 2007). Once the PIC is engaged, eIF5, a GTPase-activating protein, facilitates the hydrolysis of the GTP bound to eIF2, which reduces the affinity of eIF2 for Met-tRNA$_i$ and thus releasing eIF2-GDP from the 40S ribosome (Kapp and Lorsch, 2004; Paulin et al., 2001). Subsequently, eIF5B recruitment leads to the dissociation of initiation factors (eIF1, eIF1A, and eIF3) and promotes 60S ribosome subunit recruitment (Pestova et al., 2000). The newly formed 80S ribosome is now elongation-competent and is located at the start site with the Met-tRNA$_i$ base-pairing with the AUG start codon in the P site.
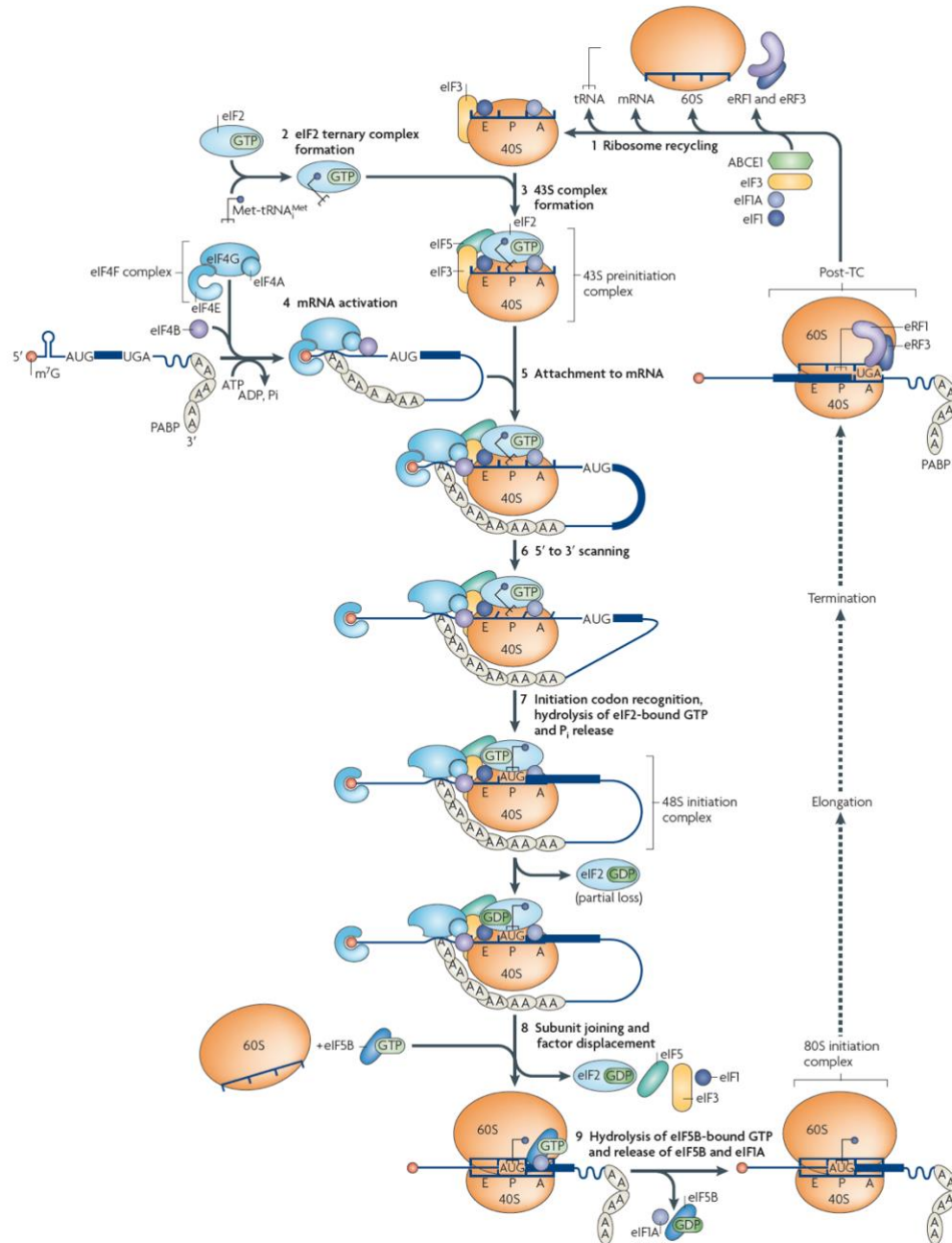
Figure 1.1 - Pathway of eukaryotic cap-dependent translation initiation. Canonical

eukaryotic translation follows a series of nine steps: (1) Ribosome recycling from the previous

translation cycle. (2) Formation of the ternary complex (TC) of the eIF2-Met-tRNA$_i$-GTP. (3)

Formation of 43S pre-initiation complex. (4) Activation of the cellular mRNA from the binding

of the 5' cap of cellular mRNA by the cap-binding complex eIF4F. (5) Recruitment of the 43S

PIC to the mRNA by the cap-binding complex. (6) ATP-dependent 5' to 3' directional scanning of the 43S PIC. (7) Start codon recognition and anticodon:codon pairing results in the hydrolysis of the eIF2-bound GTP. (8) 60S subunit joining and displacement of initiation factors. (9) Hydrolysis of the eIF2-bound GTP in a process mediated by eIF5B to form an elongation-competent 80S ribosome. Reproduced with permission from (Jackson et al., 2010).

1.1.2 5'UTR translation regulatory elements

The eukaryotic 5'UTR is a fundamental aspect for ribosome recruitment and start codon selection and consequently plays an important role in controlling translation efficiency and shaping the cellular proteome. The length of eukaryotic 5'UTRs can range from 18 to 2803 nucleotides (nt) (Mignone et al., 2002) and have an average length of 170 nt in humans (Calvo et al., 2009). Important key regulatory genes involved in proliferation, growth and differentiation tend to exhibit 5'UTR sequences with more complex secondary structure and functional features (Kozak, 1991a). Thus, this section will briefly describe a number of 5'UTR *cis* regulatory elements that regulate ribosome scanning and initiation codon selection, including stem-loops, protein binding sites, internal ribosome entry sites (IRES) and with a focus on upstream open reading frames (uORFs) (Figure 1.2).
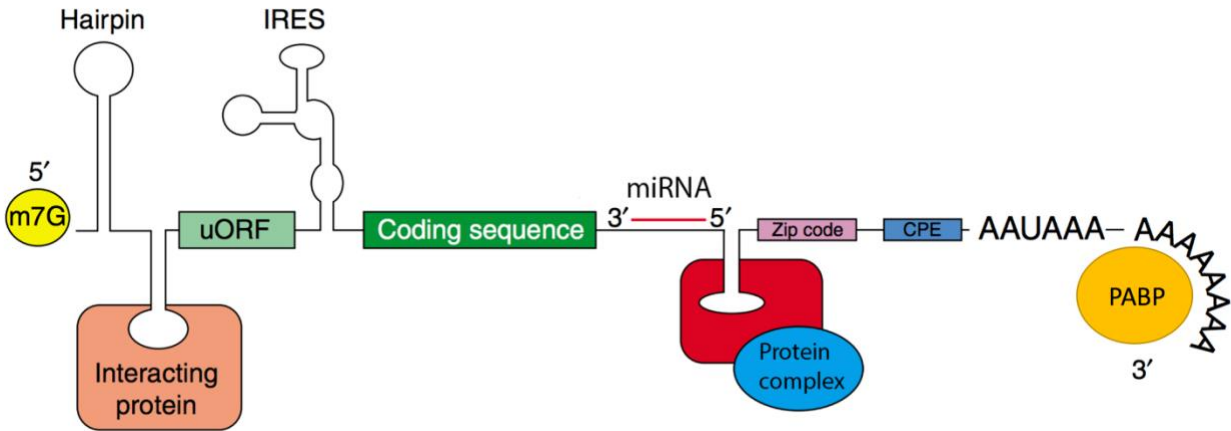
Figure 1.2 - *Cis*-regulatory elements in the cellular mRNA. The 5' end of the mRNA consists of a 7-methyl guanosine cap bound by the cap-binding complex eIF4F and may contain *cis*-regulatory elements such as, RNA hairpin stem loops and stem loops bound by interacting proteins, upstream open reading frames (uORFs), and an internal ribosome entry site (IRES). mRNA sequences also contain 3'UTR *cis*-regulatory elements such as micro RNA binding sites, stem loops for binding of protein complexes, zip codes, cytoplasmic polyadenylation elements (CPE), polyadenylation signals (AAUAAA), and a poly-A-binding proteins (PABP). Adapted with permission of (Mignone et al., 2002).

Generally, stem-loop structures can either sterically hinder the tethering of the PIC to the m^7G-cap or inhibit unwinding of the mRNA sequence, resulting in impaired ribosomal scanning (Kozak, 1989). Other stem-loops can contain structural or binding motifs for *trans*-acting regulatory proteins. For example, the translation of the iron building protein ferritin is inhibited by the interaction of iron regulatory proteins and stem-loop located in the 5'UTR (Allerson et al., 1999). Notably, some RNA 5'UTRs also contain internal ribosome entry sites (IRESs) that facilitate cap-independent translation initiation in several RNAs. First discovered in poliovirus and EMCV RNA (Jang et al., 1989; Pelletier and Sonenberg, 1988), IRESs have since been

found in 5'UTRs eukaryotic transcripts, with the first one discovered in BiP heavy-chain immunoglobulin protein (Macejak and Sarnow, 1991). Another major example of an mRNA containing an IRES, is the eukaryotic c-myc oncogene. In multiple myeloma (MM) cell lines and MM cases, a point mutation in the c-myc gene was found to drastically increase IRES activity (Chappell et al., 2000). Despite the growing evidence of IRES-mediated translation control in eukaryotic mRNAs, this cap-independent translational initiation mechanism by cellular IRESs still remains controversial (Jackson, 2013).

uORFs are short open reading frames (ORFs) in the 5'UTR of mRNA transcripts, which contain AUG start sites (uAUG) located upstream of the annotated CDS start codon (coding AUG, cAUG) (Ruiz-Orera and in Genetics, 2018). Based on the ribosomal scanning model, uORFs would be expected to repress protein synthesis due to their depletion of functional scanning ribosomes and reduction in translation initiation at the cAUG (Barbosa et al., 2013; Kozak, 2002; Wethmar, 2014; Zhang et al., 2019). Since the focus of this thesis is on uORF biology, the structure and function of uORFs will be described in detail in the following section.

1.2 5'UTR regulatory roles of upstream open reading frames (uORFs)

1.2.1 Translational control by upstream open reading frames

Approximately 50% of *Drosophila* and mammalian mRNAs contain uORFs, indicating that uORFs are prevalent genome-wide, suggesting that uORFs are key regulators of translation (Iacono et al., 2005; McGillivray et al., 2018; Zhang et al., 2018a). Intriguingly, uORFs tend to be evolutionary conserved under the assumption of neutral evolution, which suggests an evolutionary selection for advantageous uORFs (Churbanov et al., 2005; Neafsey and Galagan, 2007). Intriguingly, a substantial amount of recently introduced uORFs have been fixed in the *D.*

melanogaster population and are favoured by natural selection (Zhang et al., 2018a). With the emergence of ribosome profiling, multiple genome-wide studies have provided evidence that uORFs can sequester translation competent ribosomes and influence translation initiation of downstream coding sequences (CDS) (Chew et al., 2016; Ingolia et al., 2009; Ingolia et al., 2011; Janich et al., 2015; Johnstone et al., 2016; Zhang et al., 2018a). Specifically, an extensive study identified 35735 uORFs in *D. melanogaster* and showed that these uORFs were distributed as followed: 49.4% of genes do not contain uORFs, 35.4% of genes have one uORF, and 15.2% of genes contain at least two uORFs (Zhang et al., 2018a). It is important to note that the approximations of uORF prevalence have relied on the usage of an AUG codon to designate the uORF start site, and ribosome profiling data have illustrated that non-canonical initiation codons (e.g. UUG, CUG, and GUG) can be utilized as competent translation start sites (Na et al., 2018; Spealman et al., 2017; Zhang et al., 2018a). Consequently, the magnitude of uORF prevalence and the influential contribution of uORF translation in regulating CDS protein production have likely been underestimated.

Based on the location of the start and stop codons, uORFs can be classified into three different categories: (i) non-overlapping uORFs, of which the start and stop codons are upstream of the cAUG start site of the annotated CDS; (ii) out-of-frame overlapping uORFs (oORFs), which contain stop codons that are located downstream of cAUGs and in a different reading frame; (iii) N-terminal extensions, which are in-frame overlapping uORFs that have the same stop codon with the annotated CDS (Figure 1.3) (Johnstone et al., 2016; Torrance and Lydall, 2018; Zhang et al., 2019).
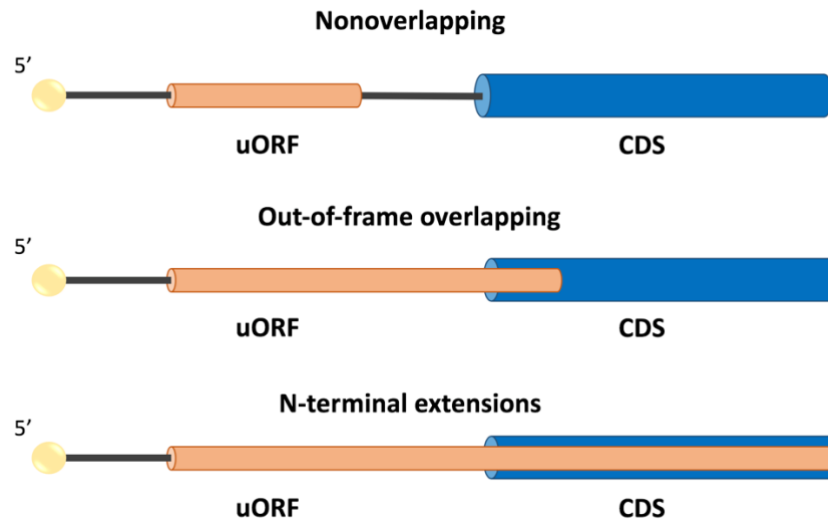
Figure 1.3 - Classification of uORFs based on the positions of their start and stop codons. A non-overlapping uORF contains a start and stop codon upstream of the annotated CDS. An out-of-frame overlapping uORF consists of a start codon in a different frame than the downstream CDS frame and terminates within the CDS. An N-terminal extension uORF has a start codon that is in frame with the downstream CDS and produces an N-terminal extension to the CDS protein.

Typically, uORFs have been reported to serve as constitutive repressors of CDS protein production by inhibiting downstream initiation at the annotated cAUG. For example, deletion of a single uAUG by CRISPR/Cas9 genome editing in plants was shown to enhance the translation of its downstream CDS (Zhang et al., 2018b). It has been estimated that translational efficiencies of the downstream CDS dramatically decreases when more than one uORF is present (Zhang et al., 2018b). Genes that have 5'UTRs that contain multiple uORFs generally have cAUGs in suboptimal Kozak sequence contexts, which suggests that the presence of multiple uORFs combined with a suboptimal cAUG sequence context ensures a low basal translation of the CDS (Rogozin et al., 2001). A few studies have made comparative genomic observations that suggest only a small number of uORFs are likely to encode functional conserved peptides, and that the

majority of uORFs are likely designed to hinder translation at the downstream CDS (Aspden et al., 2014; Mackowiak et al., 2015; Neafsey and Galagan, 2007; van der Horst et al., 2019). However, a recent study identified a much higher number of uORFs (1430 uORFs) that are likely to encode for functional peptides by conducting a genome-wide search for uORFs with conserved peptide sequences within a wide taxonomic range of species (human, chicken, zebrafish, and fruit fly genomes) using comparative genomics (Takahashi et al., 2019).

Several uORF-mediated regulatory mechanisms that can affect translation initiation at the downstream cAUG have been identified (Figure 1.4). In general, all of these uORF regulatory mechanisms result in a repression of translation initiation at the downstream annotated cAUG. These mechanisms include: (i) Leaky scanning. The scanning PIC fails to recognize a uAUG start site of a uORF, which tends to be surrounded by an unfavorable Kozak sequence context, and continues scanning for the next start codon (Kozak, 1981). (ii) Re-initiation. The uAUG is recognized by the scanning PIC and recruits 60S ribosomal subunit and initiates translation. Subsequent to termination at the uORF stop codon, the 40S subunit may remain bound to the RNA and re-initiate translation at a downstream cAUG after re-acquiring initiation factors (Hinnebusch, 1997; Hinnebusch, 2005; Vattem and Wek, 2004). (iii) Ribosome stalling. The translating 80S ribosome may stall at the uORF stop codon and trigger nonsense-mediated decay (Uchiyama-Kadokura et al., 2014). (4) Drop off. Both the 40S and 60S subunits dissociate from the 5'UTR.
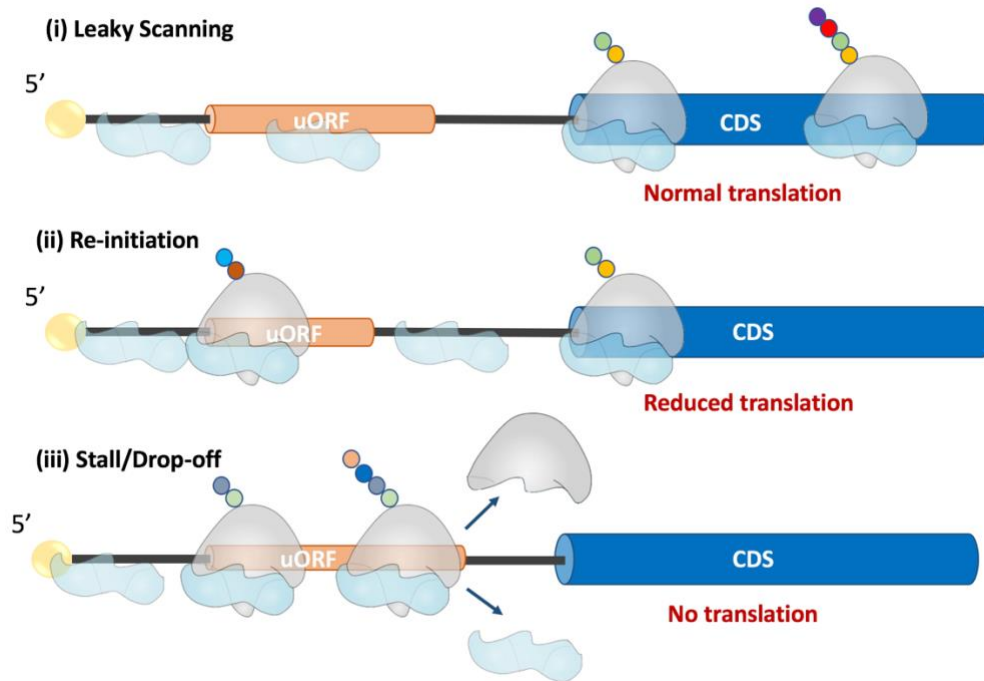
Figure 1.4 - uORF-mediated translation control of initiation of downstream CDS. Three scenarios involving the pre-initiation complex (PIC) scanning along a uORF-containing 5' untranslated region (5'UTR). (i) Leaky scanning. The scanning PIC fails to recognize the start site of the uORF and scans through in search of the next start codon. (ii) Re-initiation. The PIC recognizes the start site of the uORF, initiates translation by recruiting the 60S ribosomal subunit. After termination, the 40S subunit remains bound to the mRNA and re-initiates at the downstream CDS start side. (iii) Stall/Drop-off. Following termination at the uORF stop codon, the translating ribosome may stall and trigger nonsense-mediated decay (NMD), or dissociate from the mRNA and prevent translation of the downstream CDS.

uORFs are highly diverse in both genomic features and regulatory functions. Contextual characteristics and sequence features of the transcript's uORFs have been reported to affect the uORF's degree of repressiveness on the translation of the annotated CDS (Figure 1.5). uAUGs

11

are commonly associated with suboptimal Kozak sequence contexts for translation initiation compared to cAUGs. uORFs with high translational efficiencies (determined by ribosomal profiling) tend to have optimal Kozak sequence context surrounding their uAUG start site (Chew et al., 2016; Schleich et al., 2014; Zhang et al., 2018a). A uORF's repressive efficiency on the downstream CDS tends to increase with both a longer uORF length and with a more complex secondary structure (Kozak, 2001). A uORF also tends to be more repressive if the distance between the uAUG start site and the 5' cap increases or when a uORF's stop codon is closer to the cAUG of the downstream CDS (Calvo et al., 2009; Chew et al., 2016; Johnstone et al., 2016; Zhang et al., 2018a). Out-of-frame uORFs that are overlapping the downstream CDS (oORF) tend to be more repressive because the likelihood of translation re-initiation of the downstream CDS decreases if the oORF in the same transcript is translated (Johnstone et al., 2016; Torrance and Lydall, 2018). Furthermore, the uORF repressiveness on the downstream CDS increases with the evolutionary conservation of its uAUG start site (Calvo et al., 2009; Zhang et al., 2018a). According to the Zhang *et al.* (2018) multiple regression analysis on the translation efficiencies of genes containing uORFs in *D. melanogaster*, the most significant features that influence the degree of repressiveness of uORFs on the downstream CDS were the optimal Kozak sequence context surrounding the uAUGs, the high conservation of the uAUG codons, and the long distance between a uAUG and the 5' cap (Zhang et al., 2018a).
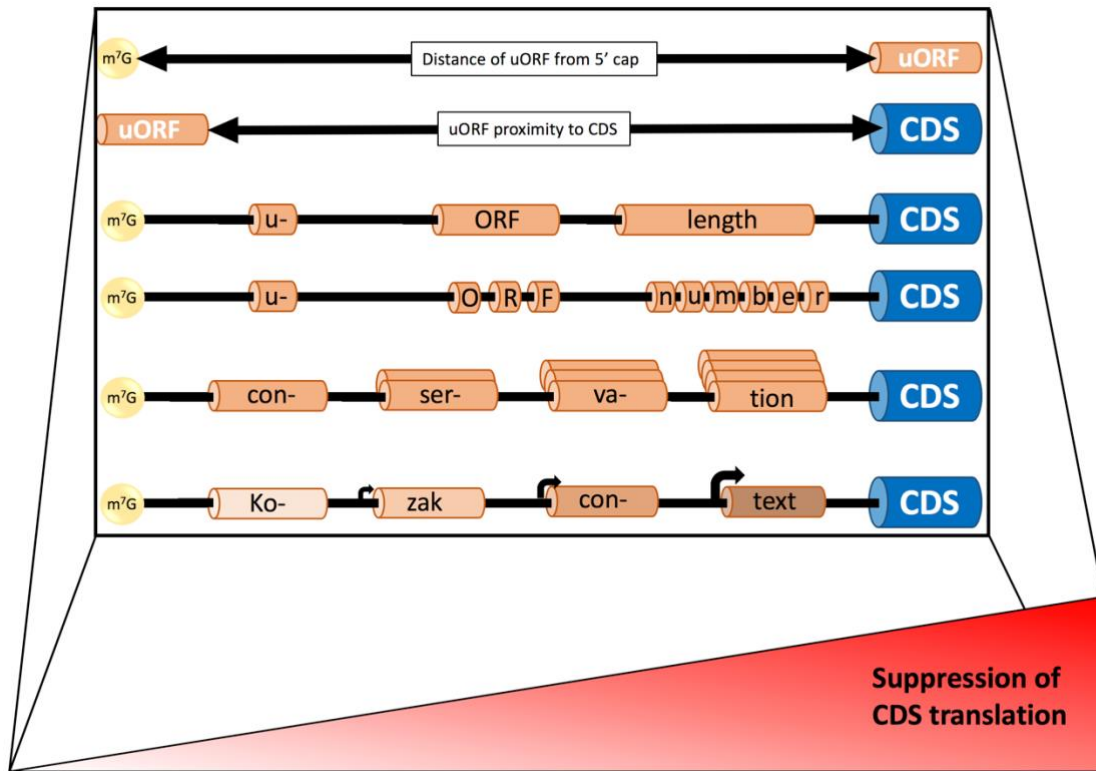
Figure 1.5 – uORF features affecting the degree of repressiveness on CDS translation.
Increase of repression on the CDS translation (bottom gradient) correlates with: (1) a longer distance between the uORF and the 5' cap, (2) a closer proximity of the uORF stop codon to the CDS start codon, (3) a longer uORF length, (4) an increase in the number of uORFs, (5) an increase in conservation of the uORF start site, and (6) a favorable Kozak start motif (reviewed in (Zhang et al., 2019)).

1.2.2 Ribosome re-initiation during the integrated stress response

uORFs may also function as rapid response elements by either dampening or promoting translation initiation at the downstream cAUG in a regulated manner. Several uORF-mediated regulatory mechanisms have been described in transcripts involved in cellular response to stress and control of cell fate (Hinnebusch et al., 2016; Morris and Geballe, 2000). This is well

characterized in transcripts involved in cellular stress adaptation and the phosphorylation status

of the α subunit of eukaryotic initiation factor 2 at serine 51 (eIF2α~P). The best studied

examples of uORF-mediated translational control involved in the integrated stress response

(ISR) are the transcription factors (TFs) GCN4 in yeast (Hinnebusch, 1997; Hinnebusch, 2005)

and ATF4 in mammals (Vattem and Wek, 2004). Both of these TFs are activated in stress

conditions and act to increase expression of genes involved in nutrient import, metabolism, and

oxidative stress management (Abastado et al., 1991; Harding et al., 2003; Hinnebusch and

Natarajan, 2002; Lu et al., 2004; Vattem and Wek, 2004). ATF4's 5'UTR contains two uORFs:

the first uORF is three codons in length, and the second out-of-frame overlapping uORF is 59

codons in length. (Figure 1.5). The 5'UTR of GCN4 contains four uORFs, with each being two

to three codons in length. During non-stressed cellular conditions, the 40S ribosome resumes

scanning after translation of the uORF1 (for both GCN4 and ATF4) and rapidly reacquires eIF2-

GTP-Met-tRNA$_i^{Met}$ ternary complex. In the case of ATF4, the newly acquired PIC re-initiates

translation at the downstream uORF2 that is out-of-frame overlapping ATF4's CDS. After

translation of uORF2, the ribosomes "drop off", resulting in a reduced translation of ATF4's

CDS (Vattem and Wek, 2004). GCN4 has a similar uORF-regulatory mechanism to ATF4,

however, GCN4's uORF4 contains a 10 nt inhibitory sequence that is suggested to promote

ribosome dissociation after uORF4's translation (Miller and Hinnebusch, 1989).

In stressed cells, eIF2α is phosphorylated by distinct eIF2α kinases (PERK, HRI, PKR,

and GCN2) and can bind with high affinity to eIF2B, which impedes the ability of eIF2B to

mediate recycling of GDP to GTP (Krishnamoorthy et al., 2001). The competitive inhibition of

eIF2B results in the reduction in eIF2α-GTP, which is required for the delivery of Met-tRNA$_i^{Met}$

for re-initiating ribosomes (Kashiwagi et al., 2016). Therefore, after translation of uORF1 there

is a deceleration of ternary complex reacquisition which allows the post-termination scanning

40S ribosomal subunit to scan past the inhibitory uORFs and re-initiate at the downstream ATF4

or GCN4 CDS (Abastado et al., 1991; Hinnebusch, 2005; Vattem and Wek, 2004). Recently, it

has been shown that the translation re-initiation of ATF4 and the changes in N^6-methyladenosine

(m^6A) modifications of its uORF2 cooperate to promote downstream CDS translation under

stress (Zhou et al., 2018). During cellular stress, eIF2α~P phosphatase GADD34 and CReP are

also highly upregulated to control proteome homeostasis (Carrara et al., 2017; Lee et al., 2009).

Similar to GCN4, the downstream CDS is translated by regulated leaky scanning through the

second uORF in the 5'UTR under stress (Lee et al., 2009; Young et al., 2015). Overall, there are

examples of genes that have uORF specific regulatory roles during stress responses to help

regulate the cellular proteome.
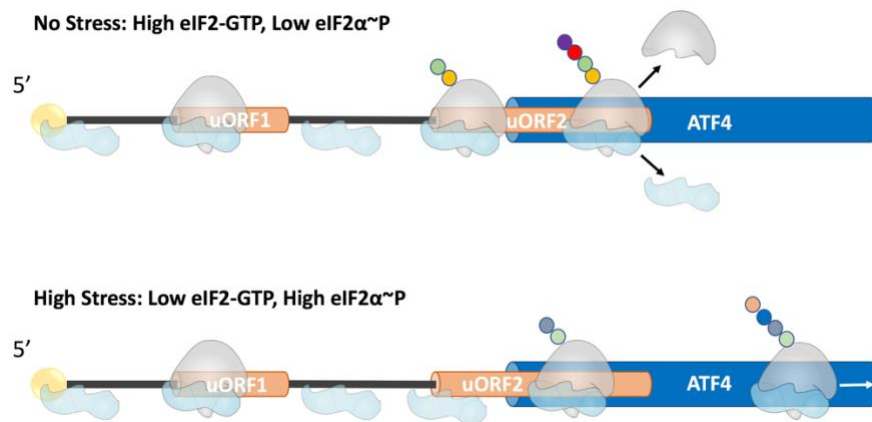
Figure 1.6 – uORF-mediated regulation of ATF4 during the integrated stress response.

During normal conditions, there are low levels of eIF2α~P and high levels of eIF2-GTP-Met-

tRNAi^{Met}. PIC scanning along the ATF4's 5'UTR recognize the start codon of the first uORF and

initiates translation. Following termination, the 40S ribosome remains bound to the mRNA and

reacquires the available ternary complex. The newly formed scanning PIC then re-initiates

translation at the second out-of-frame overlapping uORF, which results in the low ATF4

expression. During cellular stress conditions, eIF2α is phosphorylated and traps eIF2-GDP in an 'unproductive state' with eIF2B, resulting in low levels of eIF2-GTP-Met-tRNA$_i^{Met}$ (Kashiwagi et al., 2016). Following termination of translation of the first uORF, the 40S ribosome slowly reacquires a new ternary complex and 'leaky' scans passed the second uORF and re-initiates translation of the downstream start site of ATF4's CDS.

1.2.3 Ribosome stalling and nonsense-mediated mRNA decay

Not only can uORFs exert regulatory effects on the downstream CDS, but it also can have a profound impact on mRNA stability. Prolonged pausing of terminating ribosomes or premature termination of translation can trigger the nonsense-mediated decay (NMD) pathway (Isken and Maquat, 2007). This makes transcripts with uORF-containing 5'UTRs a major class of NMD substrates. For example, the stalling of terminating ribosomes on CPA-1 uORF altered the transcript's fate, making it an active substrate for NMD (Gaba et al., 2005). Intriguingly, CPA-1's uORF encodes for a fungal arginine attenuator peptide (AAP), however it is rarely translated due to the unfavorable Kozak sequence context (Werner et al., 1987). In cases where arginine levels are low, the AAP encoded uORF is skipped and the downstream CPA-1 CDS is translated. However, when arginine levels are high, translation of AAP causes terminating ribosomes to stall at the AAP uORF stop codon, which might be due to arginine inducing a conformational change or the AAP nascent peptide's interaction in the ribosome tunnel (Wu et al., 2012) triggering the NMD pathway (Gaba et al., 2005). In plants, a longer uORF (50 aa) can trigger the NMD response, whereas shorter uORFs (<30 aa) were more resistant to NMD (Nyikó et al., 2009). In some cases, uORF containing transcripts such as GCN4 are resistant to NMD. Stabilizer elements were found downstream of the GCN4's uORF that prevented the NMD

pathway. In general, the presence of uORFs does not dictate whether it will be target by NMD or not.

1.2.4 uORF-mediated translation regulation during development

Since gene expression is tightly regulated across the developmental stages of eukaryotes, it is very likely that uORF-mediated regulation is exploited to fine-tune the level and spatiotemporal control of protein synthesis throughout development (Sonenberg and Hinnebusch, 2007). A well-studied example of uORF-mediated regulation during liver development and regeneration is the mammalian transcription factor CCAAT/enhancer binding protein β (C/EBPβ), which is a key regulator in cellular processes that include cell proliferation and differentiation, innate immunity, liver development, and regeneration (Ramji and Foka, 2002). C/EBPβ consists of three different N-terminal isoforms of varying lengths: LAP* (liver-activating protein*), LAP, and LIP (liver inhibitory protein). These three isoforms have antagonistic functions and are initiated by three different consecutive AUG start codons. Since only one isoform is translated at a time, the start site selection determines the relative abundance of the antagonistic isoforms. Furthermore, C/EBPβ also contains a short uORF located in the 5'UTR upstream of the LAP and LIP start site. Following translation of the short uORF, the terminating ribosome can re-initiate downstream at the internal LIP AUG start site, producing the truncated C/EBP LIP isoform. Translation of the full length (LAP) or truncated isoform (LIP) has been shown to be modulated by the availability of initiation factors by eIF2α phosphorylation and mTOR kinase activity in cell culture experiments (Calkhoven et al., 2000; Lincoln et al., 1998). One study has developed a recombinant mouse line that contains a point mutation (ATG>TTG) at the uORF start site, and these C/EBPβ uORF-deficient mice failed to

initiate translation of the LIP isoform, which caused the mice to display phenotypes, including impaired liver regeneration after injury and defects in osteoclast differentiation (Wethmar et al., 2010). Together, these studies show that uORF-mediated translation control is vital in animal physiology.

In budding yeast, quantification of transcriptomes, translatomes, and proteomes during meiotic differentiation was performed and results showed widespread translational reprogramming by mRNA isoform switching to isoforms with varying numbers of uORFs (Cheng et al., 2018). Specifically, Cheng et al. (2018) observed in several genes that through alternate transcription start site (TSS) selection, isoform switching can occur between mRNA transcripts isoforms that differ in the content of uORFs during yeast meiosis. Recently, a genome-wide sequencing study of the ribosomal occupancy in uORFs and CDSs was conducted at specific major developmental stages of *D. melanogaster* (Zhang et al., 2018a). Zhang et al (2018) observed that switching to different mRNA isoforms with varying uORF contents are exploited to influence the translation of the downstream CDS (Zhang et al., 2018a). Specifically, Zhang et al., (2018a) observed approximately 144 isoform switching events from major transcripts with high or low amount of uORFs to other isoforms in order to regulated the translation efficiencies of the downstream CDS at all stages of *Drosophila* development (Zhang et al., 2018b). A study conducted in *Arabidopsis thaliana* showed that after exposure to blue light, alternate transcription start sites were utilized to evade uORF repression of downstream cAUG (Kurihara et al., 2018). These studies suggest that uORFs are avoided or used through isoform switching to control downstream CDS translation during stages of development. In conclusion, there are a lot of observations and correlations that uORFs modulate downstream CDS translation during development, however functional validation studies are still lacking. This

thesis focuses on the functional characterization of uORF-mediated translation control of an important developmental transcription factor called *Grainy head (grh).*

1.3 *Grainy head (grh)* gene family

The *grainy head (grh)* gene family encodes for a conserved family of TFs that are well known for their roles in epidermal barrier formation and wound healing. Grh (also known as Elf-1 and NTF-1) was first discovered in *Drosophila* with the gene name reflective of its mutant phenotype of a grainy head skeleton (Bray and Kafatos, 1991; Nüsslein-Volhard et al., 1984). Since its discovery, *grh* gene homologs have been identified in many different species including nematodes, vertebrates and even fungi. *Drosophila* and nematodes have a single *grh* gene, whereas vertebrates have multiple *grainy head-like (grhl)* gene homologues. Humans and mice have three *grhl* genes (*Grhl-1, Grhl-2* and G*rhl-3),* while zebrafish have four *grhl* genes (*grhl-1, grhl-2a*, *grhl-2b,* and *grhl-3)* (reviewed in (Wang and Samakovlis, 2012)). Since the focus of this thesis examines the uORF-mediated translational regulation of *Drosophila grh,* the following sections will focus on *Drosophila grh.*

1.3.1 *grh's* mRNA isoforms and tissue specificity

Initial studies into alternative splicing of *Drosophila grh* identified four mRNA isoforms (*grh-N, grh-N', grh-O,* and *grh-O')* (Uv et al., 1997). With the advent of whole transcriptome sequencing, another four transcripts were identified (*grh-RK, grh-RO, grh-RN* and *grh-RI*) in addition to the previous identified transcripts. Importantly, the new names of *grh* transcripts (*grh-RH, grh-RI, grh-RJ, grh-RK, grh-RL, grh-RN,* and *grh-RO)* will be used throughout this thesis (Figure 1.7). Based on the work of Uv *et al* (1997), *grh* isoforms can be classified into two

major groups. Transcripts that contain exons 4 and 5 are classified as O-isoforms (*grh-RJ, grh-RL, grh-RO, grh-RN)* and transcripts that lack exon 4 and 5 are classified as N-isoforms (*grh-RH, grh-RP, grh-RI, grh-RK*) (Figure 1.7) (Uv et al., 1997). It is suggested that the presence of these exons may affect the protein structure and function, but it is unknown how.

grh N-isoforms	*grh* O-isoforms
grh-RH = grh N'	*grh-RJ* = grh O
grh-RP = grh N	*grh-RL* = grh O'
grh-RI = grh N	*grh-RO*
grh-RK	*grh-RN*
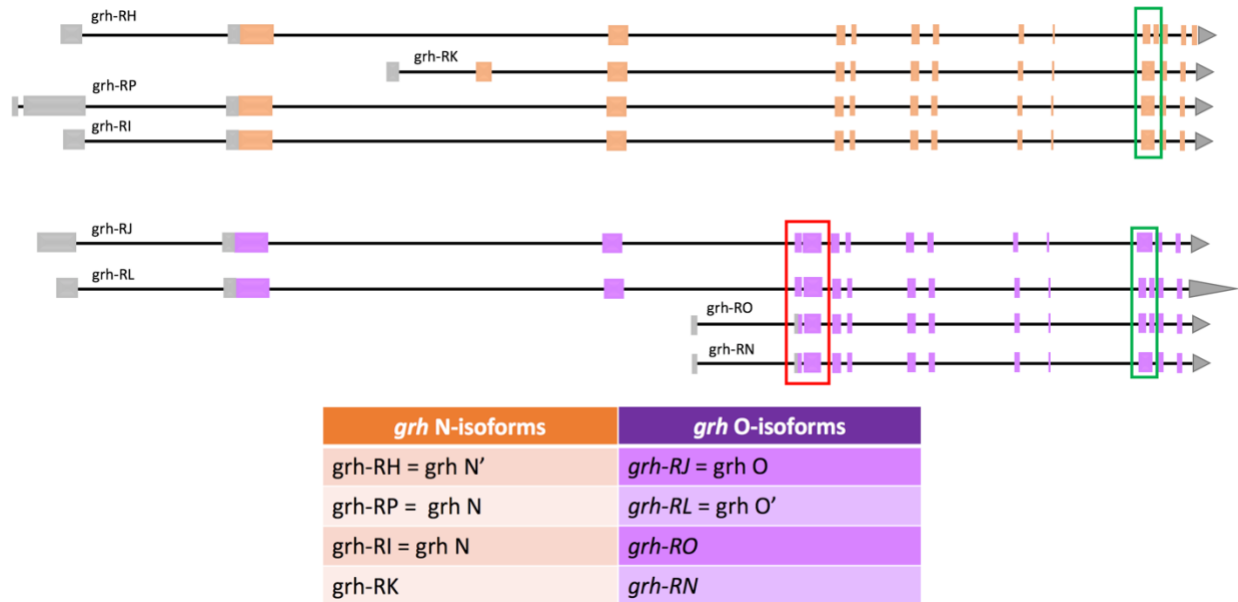
Figure 1.7 – *Drosophila's* alternatively spliced *grh* isoforms.

The single *grh* gene is alternatively spliced to produce eight mRNA isoforms, with seven different coding sequences (CDS). The eight mRNA isoforms have historically been classified into two groups, the *grh N*-isoforms (orange) that lack exon 4 and 5 (red box), and the *grh* O-isoforms (purple) which contain exon 4 and 5. The *grh* isoforms can be further subdivided by the splicing of exon 12b (green box). mRNA transcripts that have exon 12b are denoted by the apostrophe. The R nomenclature indicates that it is mRNA.

Using *in situ* hybridization with probes that are designed against a common exon for all transcripts showed that *grh* mRNA is expressed in the foregut, epidermis, larval imaginal discs,

neuroblasts, optic lobes and trachea (Uv et al., 1997). This tissue specific expression profile corroborated Grh immunoreactivity in the embryo and larvae (Bray et al., 1989). *In situ* hybridization using probes specific to exon 4 and 5, revealed O-isoform expression in neuroblasts, however the study did not show these isoforms to be exclusive and/or unique to neuroblasts. This was further analyzed by reporter genes containing O-specific Grh DNA binding sites being displayed in wild type (WT) larval brains and the loss of this reporter expression in larval brains within the O-isoform mutant background, *grh³⁷⁰* (Uv et al., 1997). As this represents the only study to examine tissue specific isoform expression, we still have no detailed picture of *grh* mRNA isoform expression.

1.3.2 *grh's* 5'UTR and upstream open reading frames

Grh's eight isoforms consists of six different 5'UTRs (*grh-RL, grh-R*I, *and grh-RH* all have the same 5'UTR) of varying lengths, ranging from 343 nt to 2299 nt. All six 5'UTRs contain differing numbers of uORFs ranging from 1 to 24 uORFs (Table 3.1). This thesis examines the regulatory function of *grh-RJ*'s uORFs on the downstream CDS during *Drosophila* development. Specifically, *grh-RJ's* 5'UTR was chosen because it was the longest annotated *grh* 5'UTR at the time (prior to whole transcriptome sequencing) and has a high uORF content (11 uORFs). Moreover, *grh-RJ's* 5'UTR also shares part of its 5'UTR including highly repressive uORFs (uORF6-11) with *grh-RL, grh-RI, grh-RP* and *grh-RH. Grh-RJ's* 5'UTR contains eleven different uORFs of varying lengths ranging from 6 codons (uORF2) to 65 codons (uORF11) (Figure 1.8). *Grh-RJ's* 5'UTR have several uORF repressive features that can repress the downstream CDS translation, these include the close proximity of uORF9-11, the out-of-frame overlapping uORF 11 with Grh's CDS, the strong conservation of uAUG2, 8, 10 and 11, and the

optimal Cavener sequence context surrounding uAUG8. Overall, the exact regulatory role and function of *grh-RJ's* eleven uORFs on Grh's CDS is unknown, but a number of uORFs would appear to have a repressive function, based on the criteria in section 1.2.1.

Table 1.1 5'UTR properties of *grh* mRNA isoforms

RNA isoform	*grh-RJ*	*grh-RL*	*grh-RP*	*grh-RI*	*grh-RH*	*grh-RK*	*grh-RO*	*grh-RN*
5'UTR length	1458 nt	837 nt	2299 nt	837 nt	837 nt	344 nt	343 nt	337 nt
# of uORFs	11	6	24	6	6	10	1	1

Figure 1.8 – *grh-RJ's* uORF containing 5'UTR properties. *Grh-RJ's* 1458 nt 5'UTR contains eleven uORFs of varying lengths. The uAUG 2, 8, 10, and 11 are well conserved among *Drosophila* species (red). uAUG8 is surrounded by a favorable Cavener sequence context (Cavener, 1987).

1.3.3 Grh's structure and function

The Grh protein is a basic helix-loop-helix transcription factor that contains three conserved domains that consists of an isoleucine rich transactivation domain at the N-terminus, a CP2 DNA binding domain, and a dimerization domain close to the C-terminal (Attardi and Tjian, 1993; Uv et al., 1994) (Figure 1.9). The interaction between the DNA binding and dimerization domains stabilizes and promotes formation of the homo and heterodimers of Grh protein complexes (Attardi et al., 1993; Uv et al., 1994). All three Grh protein domains are well conserved among members of the *grh* gene family, which permits the family of Grh TFs to bind to the same DNA consensus sequence (CCAATT), which suggest they have conserved targets

and functions (Attardi and Tjian, 1993; Boglev et al., 2011; Ting et al., 2003; Uv et al., 1994; Venkatesan et al., 2003; Wilanowski et al., 2002).
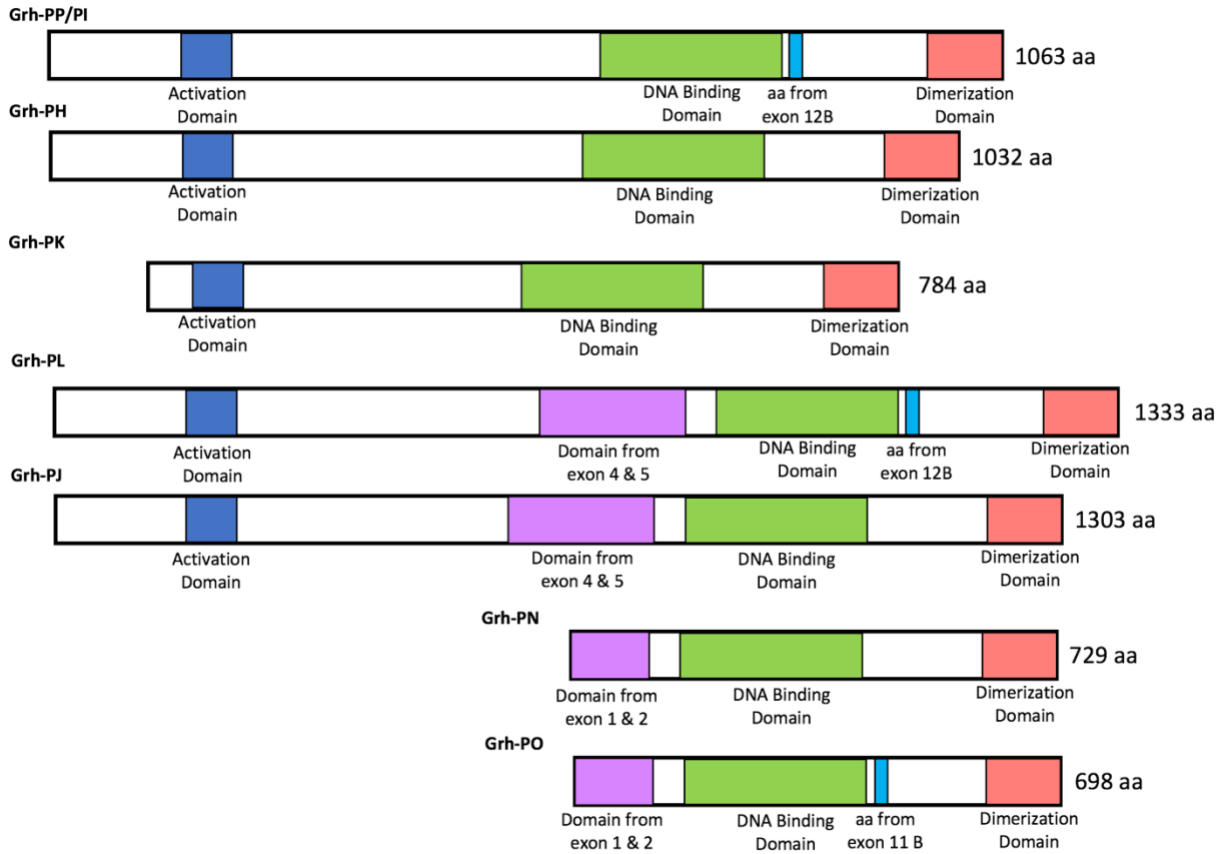
Figure 1.9 – Predicted protein domains of *Grh* isoforms. *grh* isoforms (*grh-PP, grh-PI, grh-PL,* and *grh-PJ*) contain three conserved domains: N-terminal transactivation domain (navy blue), DNA binding domain (green), and a dimerization domain (pink). The N-isoforms (*grh-PP, grh-PI,* and *grh-PH*) lack exon 4 and 5 (270 aa) and *grh-PH* has an additional exon 12b (30 aa). The O-isoforms have exon 4 and 5 (270 aa) and *grh-PL* has an additional 30 aa from exon 12b (30 aa). The P nomenclature indicates that it is protein.

The TF Grh can bind to DNA and activate or repress target genes, depending on the developmental stage and tissue. For example, in early development, Grh transcriptionally

repressing target genes such as *Dpp, zernullt,* and *tailless* to establish dorsal-ventral and terminal

patterning of the syncytial blastoderm (Huang et al., 1995; Liaw et al., 1995). Whereas in later

embryonic development (Stage 11), Grh is predominantly expressed in cuticle-producing tissues

and the central nervous system (CNS), and is known to transcriptionally activate genes such as

fushi tarazu, *Ultrabithorax,* and *Dopa decarboxylase* (Bray and Kafatos, 1991; Dynlacht et al.,

1989; Nevil et al., 2017). Currently, the mechanisms that regulate Grh's activation or repressive

functions remain unknown. ChIP-seq studies have shown that Grh exhibits stable DNA

occupancy during all stages of embryonic development (Nevil et al., 2017). Given this stable

DNA occupancy, Nevil *et al.* (2017) proposed that Grh can act as both an activator and repressor

by acting as a pioneer factor and recruiting other DNA-binding co-factors. This hypothesis was

later supported by Jacobs et al. (2018) by using ATAC-seq to demonstrate that stable Grh

binding alters chromatin accessibility (Jacobs et al., 2018). *Grh-RI* was overexpressed by a

neuron specific driver in larval neurons and showed that Grh is able to bind and open chromatin

making target genes accessible (Jacobs et al., 2018). Intriguingly, Grh is able to allow enhancer

access but is unable to activate transcription without DNA binding co-factors (Jacobs et al.,

2018). These cofactors include Atonal (a co-activator) (Jacobs et al., 2018), and Polycomb

proteins Pleiohomeotic and Ring finger domain proteins (Blastyák et al., 2006; Tuckfield et al.,

2002). This further supports the notion that the ability of Grh's to act as an activator or repressor

is mediated by the recruitment of collaborative DNA-binding co-factors.

Studies conducted in the *Drosophila* syncytial embryo illuminated Grh's ability to

compete with other TFs and regulate gene expression (Harrison et al., 2010). Grh and the

transcriptional activator, Zelda can bind to DNA elements termed TAGteam elements which are

characterized by the CAGGTAG sequence and are over-represented within promoters of genes

expressed in the syncytial blastoderm (Harrison et al., 2010). Both Grh and Zelda bind these TAGteam elements at different binding affinities located in the gene promoters of genes such as *sex lethal* and *zernullt* in the early pre-blastoderm stage. Both Grh and Zelda are unable to bind the TAGteam elements simultaneously, thus they must outcompete each other for the binding of TAGteam elements (Harrison et al., 2010). Therefore, the relative levels of Grh and Zelda determine the precise timing of target gene expression in the early embryo (Harrison et al., 2010). Similarly, Grh has also been shown to repress *tailless* and *Dpp* and it was suggested to transcriptionally repress these by outcompeting unknown co-factors during early embryonic development (Huang et al., 1995; Liaw et al., 1995). These studies also showed that overexpression of Grh in early development is associated with morphogenetic cuticle and cell division defects, which ultimately decreased embryo viability (Attardi et al., 1993; Harrison et al., 2010). Grh's protein has been shown to be phosphorylated both *in vivo* and *in vitro* (Hemphälä et al., 2003; Kim and McGinnis, 2011; Liaw et al., 1995; Wang et al., 2009). Grh is required to be phosphorylated by ERK to act as a transcriptional activator of *Ddc* and *Misshapen* wound repair genes (Kim and McGinnis, 2011). Although, it is important to note that mutations in ERK phosphorylation sites on Grh's protein did not affect its ability to form an epidermal barrier in embryos. Overall, Grh's TF activity is modulated by several mechanisms to either repress or activate target gene transcription.

1.3.4 Grh's role in *Drosophila* CNS development

During CNS development, neural stem cell like progenitor cells called neuroblasts (NB) are responsible for producing the cellular diversity of the CNS in the adult fly. Specifically, the neural diversity is generated by the sequential expression of a temporal TF cascade in the

developing NB. This cascade includes five TFs, Hunchback (Hb), Kruppel (Kr), Pdm-1/2, Castor (Cas), and Grainy head and they all progressively alter the competence of NBs to generate neurons of specific identity over time (Brody and Odenwald, 2000; Grosskortenhaus et al., 2006; Isshiki et al., 2001; Kambadur et al., 1998). (Figure 1.10) Loss- and gain-of-function studies revealed evidence for cross-regulation between the TFs in the cascade, however much of their regulation still remained unknown (reviewed in (Doe, 2017)). The best characterized lineage that Grh is involved in specifying is the NB 5-6T lineage, which produces neurons Tv1-Tv4 (Baumgardt et al., 2009). Moreover, Baumgardt et al, (2009) also found that overexpression of *grh-RL* in postmitotic Ap neurons caused ectopic neuropeptide FMRFa expression and was suggested to change Ap neuron cell fate. The exact mechanism controlling the timing of Grh's TF expression remains poorly understood. However, Grh is known to be first expressed in NBs towards the end of embryonic neurogenesis (Brody and Odenwald, 2000) and its expression is maintained in embryonic NBs that undergo quiescence and later in post-embryonic larval NBs (Bray et al., 1989).
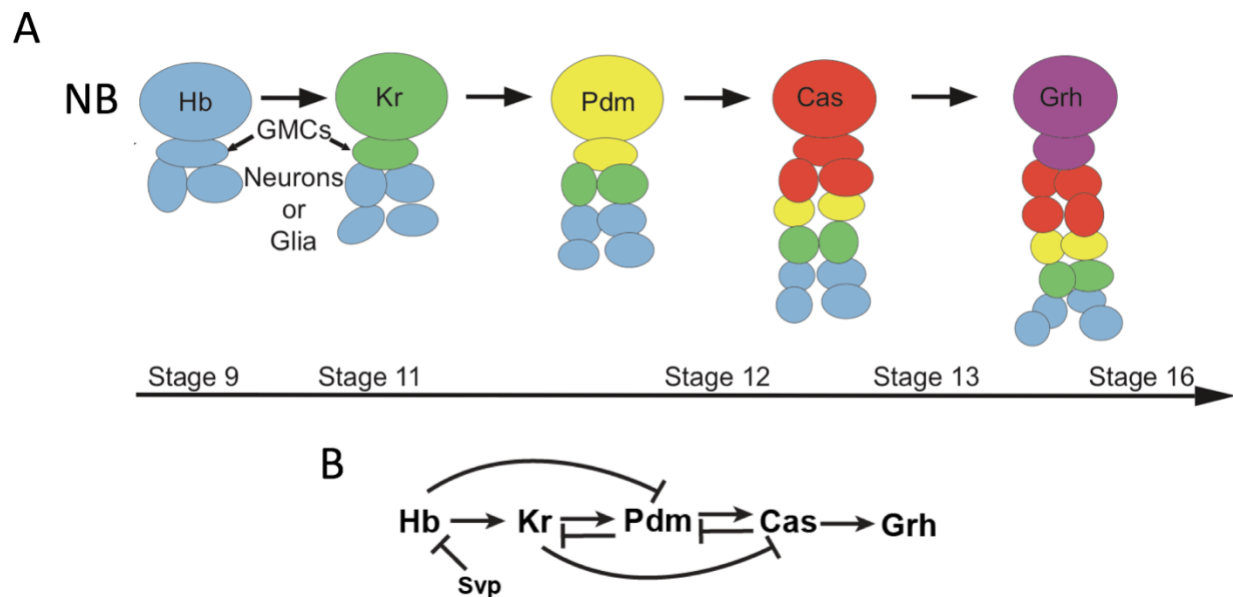
Figure 1.10 - Temporal selector expression in VNC neuroblasts. (A) Embryonic ventral nerve cord (VNC) neuroblasts (NBs) express temporal selectors sequentially at specific stages of neurogenesis producing specific ganglionic mother cells (GMCs), neurons and glia. (B) The regulatory relationship of the temporal cascade, also showing Seven up (Svp) as a switching factor.

NB proliferation is also regulated by Grh expression, but varies depending on NB identity. NBs that are mutant for *grh* located in the abdominal region of the ventral nerve cord (VNC) displayed an increase in progeny number compared to controls (Almeida and Bray, 2005; Cenci and Gould, 2005). This suggests that Grh prevents abdominal neuron production by promoting NB apoptosis. NB apoptosis is mediated by Abdominal-A (Abd-A) activity (Bello et al., 2003), and it was later found that Grh promotes apoptosis by maintaining Abd-A expression (Cenci and Gould, 2005). Meanwhile, in the thoracic region of the VNC, NBs mutant for *grh* result in a reduced progeny number, suggesting Grh's role in promoting proliferation (Almeida and Bray, 2005; Cenci and Gould, 2005). This NB proliferation is known to be partly modulated through *Drosophila* E-Cadherin (DE-Cad) expression (Dumstrei et al., 2003), and analysis of *DE-Cad* DNA locus showed Grh DNA binding sites, suggesting *DE-Cad* is a Grh target gene (Almeida and Bray, 2005). Moreover, when Grh was overexpressed in postmitotic neurons, ectopic DE-Cad levels were detected (Almeida and Bray, 2005). These two types of NB proliferation regulation by Grh suggests that Grh can modulate its downstream effector activity based on distinct cofactors and/or distinct target gene activation.

1.3.5 The *grh* gene family is an important regulator of epithelial development and maintenance of epithelial integrity

The *grh* gene family plays many important roles in development, and one of these major roles includes a conserved function in regulating epithelial structures. Interestingly, a study found that the evolution of the *grh* gene family occurred simultaneously with the evolution of the epithelia, which further suggest *grh's* crucial role in regulating epithelial development and maintenance of epithelial integrity (Traylor-Knowles et al., 2010).

1.3.5.1 Grh's role in targeting junction proteins involved in cell adhesion

Grh is responsible for activating target genes involved in cell adhesion between epithelial cells (Narasimha et al., 2008). Ectopic expression of *grh-RP/I's* CDS in the amniosera led to an overexpression of septate junction proteins, Fasciclin 3 (Fas3), Coracle (Cora) and Sinuous which also blocked the dorsal closure of the embryo (Narasimha et al., 2008). Genetic analysis of *fas3* and *cora's* DNA regulatory regions revealed a Grh protein DNA binding motif (Narasimha et al., 2008). In general, the *grh* gene family is required for epithelia structure formation to allow for a protective barrier against the external environment.

1.3.5.2 Grh's role in epidermal barrier formation and wound healing

The *grh* mutant phenotype consists of a grainy head skeleton and a flimsy and fragile cuticle, suggesting that Grh has a crucial role in cuticle formation in *Drosophila* (Bray and Kafatos, 1991). *Drosophila's* cuticle production involves the cross linking of chitin fibers and proteins by Quinone molecules which are generated through the enzymatic reactions performed by the Ddc enzyme (reviewed by (Wright, 1987)). Grh is known to transcriptionally activate the

Ddc enzyme, and therefore in *grh* mutants, the *grh* mutant cuticle phenotype is a result of a loss of Ddc expression. It has also been proposed that Grh activates other target genes involved in cell adhesion, but the expression of these genes remain unchanged in *grh* mutants (Hemphälä et al., 2003). Also, Hemphälä *et al* (2003) also found that Grh is responsible in controlling tracheal membrane growth and tube elongation and when they overexpressed *grh-RH* in all tracheal cells, the tracheal lumen growth was inhibited. Grh's activity is also crucial for the repair of the cuticle in wounded *Drosophila* embryos. Upon injury, Grh's gene targets including *Ddc*, *pale* (*ple*), and *stitcher* (*stit*) are upregulated around the wound site (Mace et al., 2005; Pearson et al., 2009; Wang et al., 2009). The *Ddc* enzyme is responsible for repairing the cuticle at the site of injury (Mace et al., 2005). Upon injury, Stit has been shown to phosphorylate Grh, which further leads to the activation of *Ddc* and *ple* creating a positive feedback loop to amplify and increase efficiency of wound healing (Kim and McGinnis, 2011; Wang et al., 2009). *Grh* mutants also exhibit an actin cable impairment phenotype, which is crucial for the flow of myosin and actin involved in site repair (Cristo et al., 2018). Overall, Grh is involved in mediating the formation of the actin cable and upregulating wound healing genes upon injury.

1.3.6 The potential repressive role of *grh's* uORFs on *grh's* CDS translation

As I previously mentioned, the *grh* gene family has many important regulatory roles throughout development. Consequently, Grh's spatial and temporal expression must be controlled to ensure proper epithelial and CNS developmental processes occur and to avoid detrimental overexpression and ectopic phenotypes. These overexpression and/or ectopic phenotypes include the following: morphogenetic cuticle defects, cell division defects, embryonic mortality, altered neuronal cell fates, NB proliferation/apoptosis defects, tracheal

lumen growth defects, and embryonic dorsal closure failure (Almeida and Bray, 2005; Attardi et al., 1993; Baumgardt et al., 2009; Harrison et al., 2010; Hemphälä et al., 2003; Narasimha et al., 2008). Therefore, *grh's* uORFs may act as a critical repressive mechanism to modulate Grh protein levels and to fine-tune Grh's spatial and temporal expression throughout *Drosophila* development.

1.4 Thesis Objectives

In summary, uORFs are important conserved regulatory elements that are prevalently found in the 5'UTR of approximately 50% of mammalian and *Drosophila* transcripts. The functional analysis of uORFs as a regulatory control mechanism to modulate the levels of spatiotemporal domains of protein synthesis to ensure proper cellular homeostasis during development has been understudied. Therefore, using *grh's* uORFs as a model to study the uORF-mediated translational control the levels of the TF Grh will serve as a functional validation of uORFs roles in tightly controlling protein levels throughout development. I aimed to test the hypothesis that *grh-RJ's* eleven uORFs (or a subset of uORFs) repress the downstream translation of *grh's* CDS in a spatiotemporal manner to regulate *Drosophila melanogaster* development by engaging ribosomes using a uORF-mediated translation control mechanism. To specifically test this hypothesis, I aimed to:

(1) Analyze the repressive role of *grh-RJ's* 5'UTR uAUGs on the downstream CDS using *in vitro* translation assays and in S2 cells.

(2) Genetically analyze *grh-RJ*'s uORFs *in vivo* function during *Drosophila* embryonic development.

2. MATERIALS AND METHODS

2.1 DNA Luciferase reporter constructs

The synthesized fragment of *grh-RJ's* wildtype or mutant (Mut) 5'UTR (uATG>uATT) including the first six codons of Grh's CDS (gBlock; IDT) was cloned in frame with Firefly luciferase between the EcoRI and NcoI sites within the p-GEM3 (Promega) (cloned by Mika Vivar). The synthesized WT and Mut 5'UTR fragments (gBlock; IDT) were also subcloned in frame between the EcoRI and NcoI sites within pAc5.1/V5-HisB (Thermo) (cloned by Mika Vivar). All eleven mutant uATT constructs were generated using Polymerase Chain Reaction (PCR)-based mutagenesis and verified by sequencing (Table 2.1).

Table 2.1 - List of primers used for site directed mutagenesis.

Site Directed Mutagenesis Primers	Primer Sequence (5' -> 3')
Prmk-ATT1-F	CTGTGTGTGTGTACGTGTGCATTTCTCTCGGAGAGCGTGAGCT
Prmk-ATT1-R	AGCTCACGCTCTCCGAGAGAAATGCACACGTACACACACACAG
Prmk-ATT2-F	GTGTGCGCTTGCGAGCTGGAATTAGACGCCATCAAAGCTGAGC
Prmk-ATT2-R	GCTCAGCTTTGATGGCGTCTAATTCCAGCTCGCAAGCGCACAC
Prmk-ATT3-F	GCTGAGCAGCGAATAATAAAATTCAAAATAAAAATAGTGATTA
Prmk-ATT3-R	TAATCACTATTTTTATTTTGAATTTTATTATTCGCTGCTCAGC
Prmk-ATT4O'-F	AACAAATTCTAGTGCGGAAAGCGCGTAA
Prmk-ATT4O'-R	GCACTAGAATTTGTTATTGCTGCTGGTT
Prmk-ATT5O'-F	CGCAAATTCCAGGCGCAACGCCCCCGAA
Prmk-ATT5O'-R	CGCCTGGAATTTGCGTTTTCAAGCCAAT
Prmk-ATT6-F	CCAGTAGCAATAGCAGCAATATTAGCAGCAGCAACATCAAATG
Prmk-ATT6-R	CATTTGATGTTGCTGCTGCTAATATTGCTGCTATTGCTACTGG
Prmk-ATT7-F	ATGAGCAGCAGCAACATCAAATTTTAGGCCAAAATGCACA
Prmk-ATT7-R	TGTGCATTTTGGCCTAAAATTTGATGTTGCTGCTGCTCAT
Prmk-ATT8-F	ACATCAAATGTTAGGCCAAAATTCACAAACCGCCAGCAACAAA
Prmk-ATT8-R	TTTGTTGCTGGCGGTTTGTGAATTTTGGCCTAACATTTGATGT
Prmk-ATT9-F	AGTGCCATAGAAAGCAGTTAATTAAACAACGACTAAGACGAAG
Prmk-ATT9-R	CTTCGTCTTAGTCGTTGTTTAATTAACTGCTTTCTATGGCACT
Prmk-ATT10-F	GTTGATAAGTGATATTTATTATTTTTATACTTGCCAGCAGCCG
Prmk-ATT10-R	CGGCTGCTGGCAAGTATAAAAATAATAAATATCACTTATCAAC
Prmk-ATT11-F	CATAGGGGATCACGGCATCGATTATCAGTCCACGACCAAGTCC
Prmk-ATT11-R	GGACTTGGTCGTGGACTGATAATCGATGCCGTGATCCCCTATG

2.2 *In vitro* transcription and translation assays

The monocistronic DNA construct containing the wild-type or Mut *grh-RJ* 5'UTR in frame with Firefly luciferase was linearized using XbaI. RNAs were *in vitro* transcribed using T7 RNA polymerase and purified afterwards using an RNeasy Mini Kit (Qiagen). The RNA purity and integrity were confirmed on a 1% denaturing formaldehyde agarose gel. For *in vitro* translation assays, uncapped RNAs (15ng/μl) were incubated in a rabbit reticulocyte lysate (RRL) translation extract (Promega) for 1 hour at 30°C in the presence of [^{35}S]-methionine/cysteine (EasyTag Express Protein Labeling Mix, Perkin Elmer). The reactions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The gels were subsequently dried and subjected to autoradiography and phosphorimager analysis (Typhoon, GE healthcare).

2.3 Cell culture and DNA transfection

Drosophila S2 line 2 (S2) cells were maintained and passaged at 25°C in Shields and Sang M3 insect media (Sigma-Aldrich) supplemented with 10% fetal bovine serum. Transfection of WT or Mut *grh-RJ* 5'UTR pAC5.1 Firefly luciferase (FLuc) plasmids and Renilla Luciferase (pRLuc) control reporter plasmid (Promega) into S2 cells was performed using X-tremeGENE™ (Sigma) as per the manufacturer's instructions. 1 μg of either WT or Mut *grh-RJ* 5'UTR DNA and 1μg of pRL DNA were used for transfection using 3.0×10^6 cells.

2.4 S2 cell lysis and Firefly luciferase assay

Following transfection, cells were collected and lysed at 6 and 24 hours post transfection. Cells were pelleted and re-suspended in 50 µl of 1x passive lysis buffer (Promega Dual-Luciferase® Reporter Assay System) and subjected to a freeze-thaw cycle to completely lyse cells. Lysates were then cleared by centrifugation and protein concentration was determined using a Bradford Assay (Bio-Rad Protein Assay) as per manufacturer's instructions. For each protein sample, 5 µg of protein was plated in a 96-well plate. Upon addition of Luciferase Assay Reagent II (LAR II, Promega), luciferase activity was measured at 1000 millisecond intervals using the Tecan Spark® microplate reader.

2.5 Molecular Cloning for *in vivo* experiments

2.5.1 Transgenic reporters

The synthesized WT and Mut (uATG>uATT) *grh-RJ* 5'UTR fragments (gBlock; IDT) were inserted between AgeI and NheI sites within the pCA4B plasmid (UAS-attB vector) (Markstein et al., 2008). The superfolder GFP (*sfGFP*) with a nuclear localization signal (*NLS*) was PCR amplified from pHD-sfGFP-ScarlessDsRed (Addgene) and inserted downstream and in frame with the WT or Mut 5'UTR.

2.5.2 Construction of CRISPR homology directed repair (HDR) templates

The 1.5 kb left homology arm (chr2R:17,806,673-17,808,151) and 1.5 kb right homology arm (chr2R:17,808,148-17,809,502) were PCR amplified directly from the *D. melanogaster* genome of the *y[1]M{w[+mC]=nos-Cas9.P}ZH-2A* fly (as this is the fly which the HDR template was to be integrated). The left homology arm (H1 fragment) was PCR amplified in two fragments (H1a and H1b fragments) and combined using overlapping PCR, in order to create

through a mutational overlap. The right homology arm (H2 fragment) was PCR amplified from

the *grhRJ-(WT or Mut)-5'UTR-FLuc* reporter (H2a fragment) and from genomic DNA (H2b) and

combined by overlapping PCR (Figure 3.8B). The "scarless" dsRED marker cassette (pHD-

ScarlessDsRED) designed by the flyCRISPR group (http://flycrispr.molbio.wisc.edu/scarless)

was altered by swapping out the dsRED with the EGFP sequence. The WT and Mut HDR

template was constructed using Gibson Assembly® (New England BioLabs) by combining the

left homology fragment (H1), the EGFP reporter cassette, and the right homology fragment (H2),

and the linear blunt-end Pjet1.2 (Thermo). HDR template sequences were verified by Sanger

sequencing.

Table 2.2 List of primers used to create CRISPR/Cas9 HDR templates.

HDR PCR fragments	Primer	Primer Sequence (5' -> 3')
H1a	grh-306F	gatggctcgagttttttcagcaagatGTGCAGGGGGACTGGACGTGGA
	grh-306R	CCCCCTCATCGACCCACTGGCGATG
H1b (contains PAM site mut)	grh-307F	GTGTGTGACTGCATCGCCAGTGGGTCGA**t**GAGGGGGCGGGGGCGGGCGCCAA
	grh-307R	TTATCTTTCTAGGGTTAACTGCTTTCTATGGCACTAAGGAACTTCCTTCGTTTCGTTCT
H2a(WT)	grh-304F	TTTCTAGGGTTAATGAAACAACGACTAAGACGAAGATCGACCATCCAG
	grh-308R	CTCGTTGTGGCGGTcGATGTGGACATTATACGTGGTAGATA
H2a(Mut)	grh-304mF	TTTCTAGGGTTAAT**t**AAACAACGACTAAGACGAAGATCGACCATCCAG
	grh-308R	CTCGTTGTGGCGGTcGATGTGGACATTATACGTGGTAGATA
H2b	grh-309F	TATCTACCACGTATAATGTCCACATCgACCGCCACAACGAG
	grh-304Ra	attgtaggagatcttctagaaagatAGTATATTTCCAAACTCTTCTG
EGFP reporter cassette	grh-303F	AGAAAGCAGTTAACCCTAGAAAGATAATCATATTGTG
	grh-303R	AGTCGTTGTTTCATTAACCCTAGAAAGATAGTCTG
	grh-303mR	AGTCGTTGTTTAATTAACCCTAGAAAGATAGTCTG

2.5.3 Construction of guide RNA vectors

Guide RNA target 1 (5'- ACTGCATCGCCAGTGGGTCGAGG - 3') and target 2 (5'-CCACCGCCACAACGAGCGTTATC – 3') were designed with no predicted off-target sites using CRISPR Optimal Target Finder developed by flyCRISPR groups (O'Conner-Giles, Wildonger and Harrison labs; http://targetfinder.flycrispr.neuro.brown.edu/). Annealed gRNA primers were incorporated into the pU6-BbsI-chiRNA (Addgene plasmid #45946) at the BbsI site following the protocol provided on the flyCRISPR website (http://flycrispr.molbio.wisc.edu/protocols/gRNA). Correct insertion was confirmed by sequencing using the T3 universal primer.

2.6 Fly transgenesis and injection of CRISPR/Cas9 HDR and gRNA plasmids

All transgenesis was performed by Rainbow Transgenic Flies Inc. (Camarillo, CA, USA). One copy of each transgene was individually and unidirectionally integrated into the *attP40* locus on the second chromosome of the *D. melanogaster* genome using phiC31 integrase-based transgenesis (Bischof et al., 2007). HDR and gRNA plasmids were injected by Rainbow Transgenic Flies Inc. (Camarillo, CA, USA).

2.7 Fly Husbandry

Flies were maintained on standard cornmeal food at 70% humidity at 18ºC, or 25ºC. Strains from the Bloomington *Drosophila* Stock Center were: *P{w[+mc]=UAS-mCherry.NLS}3* (referred to as *UAS-mCherry.nls*) (BL#38424). *P{w+mW.hs=GAL4-da.G32}3a* (referred to as *da>* and *da^GAL4*) (BL#55849), *y[1]M{w[+mC]=nos-Cas9.P}ZH-2A* (BL#54591). *CyO,P{Tub-*

PBac\T}wg[Sp-1];I(3)*[*]/TM6B,Tb (BL#8285). *w^{1118}*. Balancer chromosomes for stock maintenance and construction of transgenics used were: *CyO* or *CyO-twi>GFP*.

2.8 Embryo staging and tissue processing

UAS-(WT-or-Mut)-5'UTR-sfGFP::NLS virgin females were mated with *w;;da>GAL4,UAS-mCherry::NLS* males for two to three days. Flies were transferred to an egg-laying chamber for 2 hours to clear eggs from oviduct and embryos were then collected during another 2-hour egg-laying window. Embryos were grown on grape plates supplemented with yeast paste, washed into mesh baskets, dechorionated using 50% bleach for 90 sec. Late stage 17 embryos with mouth hooks and air-filled trachea were selected for CNS dissection. VNCs were dissected in ice-cold Phosphate-Buffered Saline (PBS) and fixed in ice-cold 4% paraformaldehyde (PFA) for 20 minutes. Stage 11 whole embryos were staged according to (Campos-Ortega and Hartenstein, 1997). Whole embryos were fixed in ice-cold 4% PFA for 20 minutes and then devitellinized in methanol prior to staining (Patel, 1994).

2.9 Antibodies

Primary antibodies: rabbit anti-dsRED (1:500) (Clonetech), chicken anti-EGFP (1:1000) (ab13970, Abcam), and rat anti-Grh (1:500) (target epitope is Grh's c-terminus) (a gift from Stefan Thor). Secondary antibodies were anti-chicken, anti-rabbit, and anti-rat IgG (H+L) conjugated to Alexa 488, Cy3, or Cy5 (1:200; Jackson Immunoresearch).

2.10 Immunocytochemistry

Whole embryos were washed in 0.01% PBST (PBS containing 0.1% Triton-X-100) and then incubated with 0.01% PBST containing 5% Donkey Serum (PBSTD). Primary antibodies were incubated together overnight at 4ºC. Embryos were washed with PBST and then incubated in PBSTD. Secondary antibodies were incubated for 2.5 hours at room temperature. Tissue was washed with PBST and then mounted in Vectashield (Vector). All antibodies were diluted in PBSTD. Both *da>grhRJ-WT-5'UTR-sfGFP::NLS* and *da>grhRJ-Mut-5'UTR-sfGFP::NLS* embryos were collected, fixed, and treated in an identical manner.

2.11 Image and statistical analysis

All images for each experiment were acquired with Zeiss Axio Imager.z2 (LSM880) with a 20X lens on the same day using the same settings. Maximum intensity projections were generated using Zen blue software. Bitplane:Imaris v9.2 software (in Spots Mode) was used to identify reporter-positive nuclei in the CNS and whole embryos. Imaris settings were established independently to the internal control mCherry::NLS positive nuclei with minimal background fluorescence spot marking, in order to compare sfGFP::NLS and Grh intensities in both WT and Mut 5'UTR genotypes. We also measured the amount of sfGFP::NLS and mCherry::NLS fluorescence intensities in nuclei filtered for high and low Grh fluorescence intensities by using median intensity thresholding. More specifically, the median Grh intensity threshold was determined in all mCherry::NLS positive nuclei and all Grh intensities below the median threshold was labelled as low Grh expression and Grh intensities above the median threshold were labelled as high Grh expression. Each image was manually subtracted for spots that erroneously labelled background fluorescence. All statistical analysis and graphing were performed using Prism 6 software (GraphPad Software, San Diego, CA). A minimum of n=3

flies were used for each genotype. For comparison between two genotypes, unpaired parametric

T-tests were used (Students T-test). Statistical differences are shown if p<0.05.

3. RESULTS

The *grh* gene generates eight mRNA isoforms (*grh-RH, I, J, K, O, N, P)* by alternative splicing (Fig 3.1 A). These transcripts have varying 5'UTR lengths with differing numbers of uORFs (Table 3.1). Transcripts *grh-RL, RI,* and *RH* share the same 5'UTR with six uORFs, which are also included in the 5'UTR of *grh-RJ* (11 uORFs) and *RP* (24 uORFs) (Figure 3.1 B). Using *in situ* hybridization, Uv et al. (1997) suggested that isoforms containing exons 4 and 5 (red box, figure 3.1 A) (*grh-RJ, RL, RO,* and *RN)* are neuro-specific in larval CNS, but did not demonstrate these isoforms to be exclusive and/or unique to nervous system expression. Another study using RT-PCR on RNA extracted from early pre-cellular blastoderm embryos demonstrated that the mRNA isoforms *grh-RJ, RL, RH,* and *RI* are expressed between 0-2 hours of embryonic development, but did not analyze the other isoforms (Harrison et al., 2010). Immunostaining using Grh antibodies has shown that nuclear Grh is first detected in epidermal and tracheal cells and in the CNS at stage 11 and persists throughout embryogenesis (Bray et al., 1989; Hemphälä et al., 2003). Due to a lack of expression studies, expression patterns and timing of expression of isoforms *grh-RO, RN,* and *RK* are unknown. However, more recently, RNA-seq data for the 3rd instar larval VNC does support the notion that the major transcripts of the late L3 VNC (mostly in post mitotic neuroblasts (pNBs) and some ganglion mother cells (GMCs) (Almeida and Bray, 2005)) are *grh-RO* and *RN.* (Fig 3.1 A) (Robin Vuilleumier, Stephane Flibotte, Douglas Allan, unpublished observations). The specific spatial and temporal expression of the *grh-RJ* transcript throughout embryonic development remains unknown. This thesis explores the uORF regulatory activity of *grh-RJ's* 5'UTR on the downstream coding sequence (CDS) during *Drosophila* development.
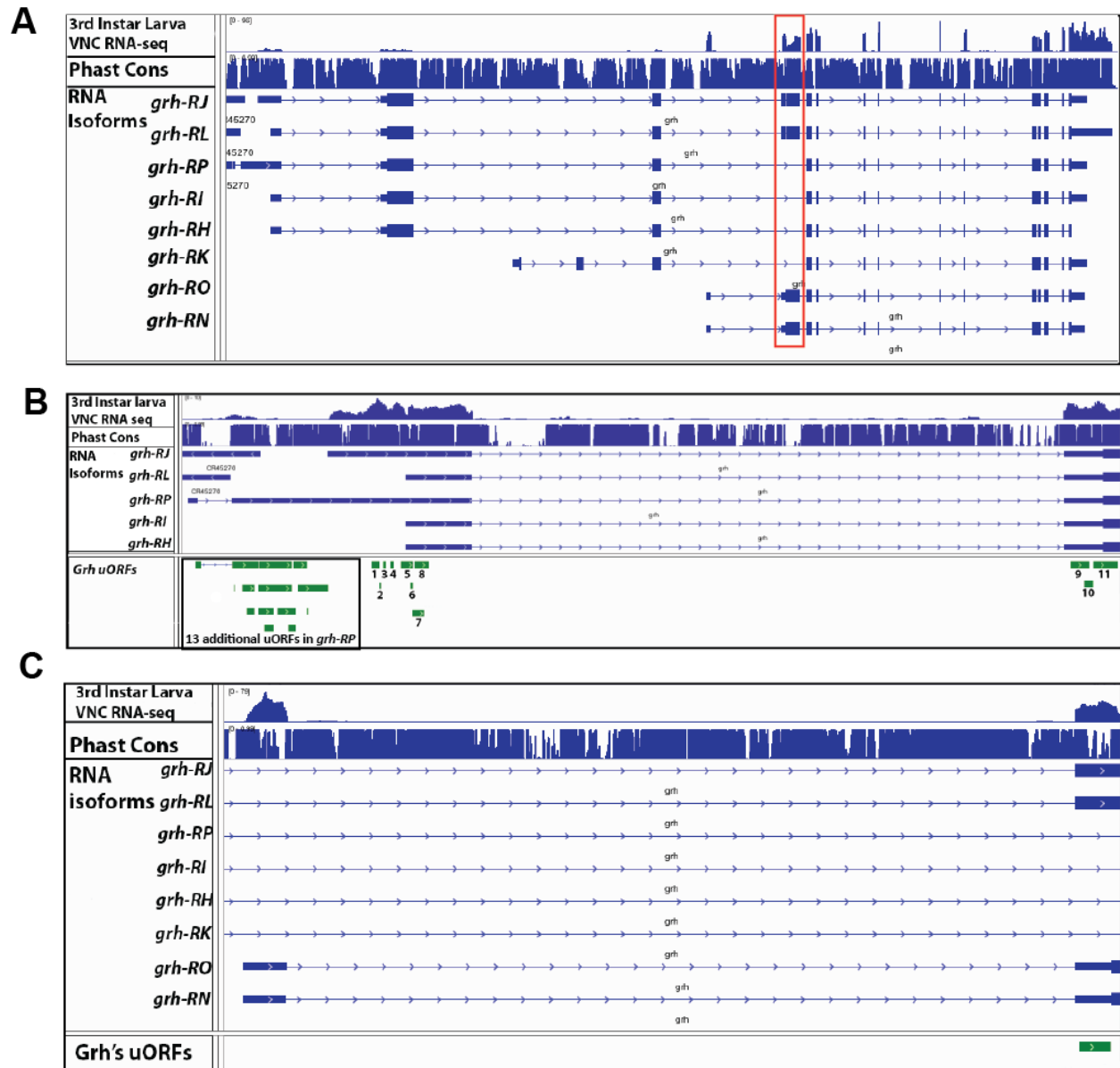
Figure 3.1 Integrative genomic view of *grh* RNA isoforms and uORFs

(A) Integrative genomics view (IGV) of the genomic regions of all eight *Grh* isoforms (grh-RJ, RL, RP, RI, RH, RK, RO, RN). RNA-sequencing was performed on RNA isolated form 3rd Instar larval VNCs (reproduced with permission of Robin Vuilleuimier and Stephane Flibotte). RNA-sequencing reads map to Grh transcript *RO* and *RN*, indicating that grh-RO isoform is predominantly expressed in the VNC at the 3rd instar larval stage. Red box outlines exon 4 and 5.

(B) Genomic view of Grh-RJ's eleven uORFs with smallest uORF containing 6 codons (uORF 2) and the longest uORF containing 65 codons (uORF 11). (C) Genomic view of Grh-RO/N's uORF of 46 codons. All uORFs and Cavener consensus matches were mapped to the *Drosophila melanogaster* genome by Shamsuddin Bhuiyan (Pavlidis lab, UBC).

A Ribo-seq study conducted by Zhang., *et al* (2018) established translation efficiencies (TE) of uORFs across all fly transcripts at various *Drosophila* developmental time points (Zhang et al., 2018a). Importantly, this study described relative translation levels for the uORFs in Grh-RJ and also showed that uORFs are widely translated in developing *Drosophila* embryos. These TEs were calculated as a ratio between translation (derived from ribosome profiling footprints) and mRNA levels (derived from mRNA sequencing reads) (Table 3.1) (calculations are shown in (Zhang et al., 2018a)). Many of *grh-RJ's* uORFs (i.e. uORF 3, 5, 6, 7, 8, 10) have a positive $\log_2$(TE) fold change at various early embryonic and larval developmental time points, which suggests that these uORFs are being translated (Table 3.1) (Supplemental Data 7 (Zhang et al., 2018a)). Whereas, *grh-RO's* uORF has $\log_2$(TE) fold changes close to the value of 0, at all developmental time points, which suggests it is not being translated or being translated at a very low efficiency (Table 3.1) (Zhang et al., 2018a).

Table 3.1 *grh-RJ* and *grh-RO/ N* uORF properties and translation efficiencies during *Drosophila melanogaster* development [a]

	Grh CDS	uORF 1 (grh-RJ)	uORF 2 (grh-RJ)	uORF 3 (grh-RJ)	uORF 4 (grh-RJ)	uORF 5 (grh-RJ)	uORF 6 (grh-RJ)	uORF 7 (grh-RJ)	uORF 8 (grh-RJ)	uORF 9 (grh-RJ)	uORF 10 (grh-RJ)	uORF 11 (grh-RJ)	uORF 12 (grh-RO/N)
uORF length (amino acids)	N/A	20	6	7	9	32	8	31	39	50	23	66	46
Cavener motif	no	no	no	no	yes	no	no	no	yes	no	no	no	no
Conservation of	17/19	6/19	11/19	6/19	6/19	8/19	5/19	7/19	12/19	5/19	16/19	10/19	15/19

uAUGs among 19 *Drosophila* species												
Translation Efficiencies[a]												
0-2hr	-1.06	-4.54	-5.8	0.72	-0.65	-1.2	0.15	-0.59	-0.52	-1.48	0.39	-2.43	-0.89
2-6hr	-0.07	-2.13	-0.1	2.36	1.43	0.87	2.18	1.3	1.2	-0.64	0.71	-0.24	0.5
6-12 hr	0.93	-2.52	-1.43	0.51	-2.19	0.67	2.36	1.42	1.45	-0.7	1.55	0.12	0.01
12-24 hr	0.62	-0.93	0.21	0.35	0.09	1.1	2.49	1.79	1.84	-0.11	1.64	0.35	-0.79
Larvae	0.32	-2.81	-0.68	3.38	N/A	1.52	3.43	1.52	0.87	0.6	1.02	1.69	0.52

-5   0   5
Log$_2$Fold-Change

[a]Translation efficiencies (S7 Data) were calculated as a ratio of translation (derived ribosome profiling footprints) and from RNA levels (from mRNA seq levels) (calculation are shown in (Zhang et al., 2018a)). Log$_2$(TE) were calculated to determine the fold change in TE for each uORF and CDS. The color scale represents the log$_2$(TE) fold change.

As I mentioned in section 1.3.3, *grh-RJ's* 5'UTR has eleven uORFs with uORF6-11 being shared with the 5'UTR of *grh-RL, grh-RI, grh-RP,* and *grh-RH*. *Grh-RJ's* uORF lengths varies from 6 codons (uORF2) to 65 codons (uORF11) (Table 3.1). *Grh-RJ's* 5'UTR have several uORF repressive features that could conceivably repress the downstream CDS translation, these include: a closer proximity of a uORF to the cAUG; a longer distance of a uORF from the 5' cap; an out-of-frame uORF that overlaps with the downstream CDS; a favourable Cavaner context surrounding uAUGs; and high conservation of the uAUG start codon. Considering these important features and our *in vitro* results, we have ranked *grh-RJ's* eleven uORFs from most to least repressive on the downstream CDS (Figure 3.2).
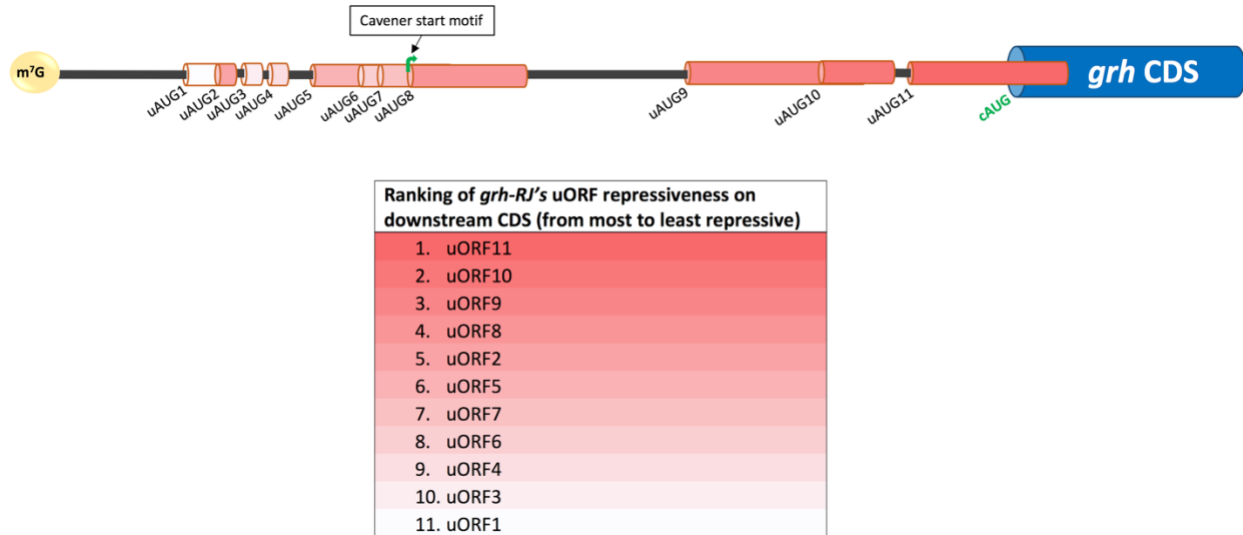
Figure 3.2 – Ranking of *grh-RJ's* uORF degree of repressiveness based on repressive uORF sequence features. Ranking is based on the following features: (1) Conservation of the uAUG (2,8,10-11) within 19 *Drosophila* species. (2) Cavener translation sequence context of uAUG8. (3) The proximity of uORF9-11 to *grh's* CDS and close proximity of uORF1-8 from the 5' cap. (4) Out-of-frame overlapping feature of uORF11 with *grh's* CDS.

3.1 *grh-RJ*'s 5'UTR uAUG start sites repress translation of the downstream CDS *in vitro* and in S2 cells

The effects of uORFs in *grh's* 5'UTR on the translation of *grh*'s CDS are unknown. To address this, we created a set of reporter constructs in which Firefly luciferase (FLuc) replaced the Grh's CDS, downstream of a wildtype (WT) or mutant (Mut) *grh-RJ* 5'UTR. These Fluc reporter constructs include: a WT *grh-RJ* 5'UTR (generated by Mika Vivar, Jan lab), 3-proximal mutant (3-prox Mut) 5'UTR (uAUG 9-11 mutated to AUU), or a Mut 5'UTR (all uAUG to AUU) (generated by Mika Vivar, Jan lab) (Figure 3.3 A). Prior to the CRISPR/Cas9 genome editing of uAUG9-11 (shown below), the 3-prox Mut construct was first generated to determine

the *in vitro* effect of the loss of uAUG9-11 on the translation of the downstream CDS. These reporter constructs were *in vitro* transcribed and then incubated in rabbit reticulocyte lysate (RRL) in the presence of [^{35}S]-Met/Cys. Protein levels of FLuc were measured by analyzing the SDS-PAGE protein bands and normalized to the WT 5'UTR FLuc protein levels. The 3-prox Mut 5'UTR and the mutant 5'UTR resulted in a 2-fold and 7-fold increase in Fluc activity, respectively (Figure 3.3 B & C). These results indicate a repressive role for the uAUGs in the WT 5'UTR on the translation of the downstream CDS in *in vitro* reactions.

To determine whether the *grh-RJ* 5'UTR uAUGs have a similar repressive effect *in vivo,* S2 cells were co-transfected with the WT or Mut *grh-RJ* 5'UTR FLuc DNA reporter constructs and the control Renilla luciferase (RLuc) SV40 plasmid for 6 and 24 hours, and the ratio of FLuc/RLuc activities were measured and normalized to the WT 5'UTR FLuc/RLuc ratio (Figure 3.3 D) (conducted by Mika Vivar, Jan Lab). At both 6 and 24 h post-transfection, the Mut 5'UTR reporter construct led to a ~2.5-fold increase in the ratio of FLuc/RLuc (Figure 3.3 D). Similar to *in vitro* results, the mutant *grh-RJ* 5'UTR resulted in higher FLuc activity, indicating that the WT 5'UTR uAUGs repress the translation of the downstream CDS.
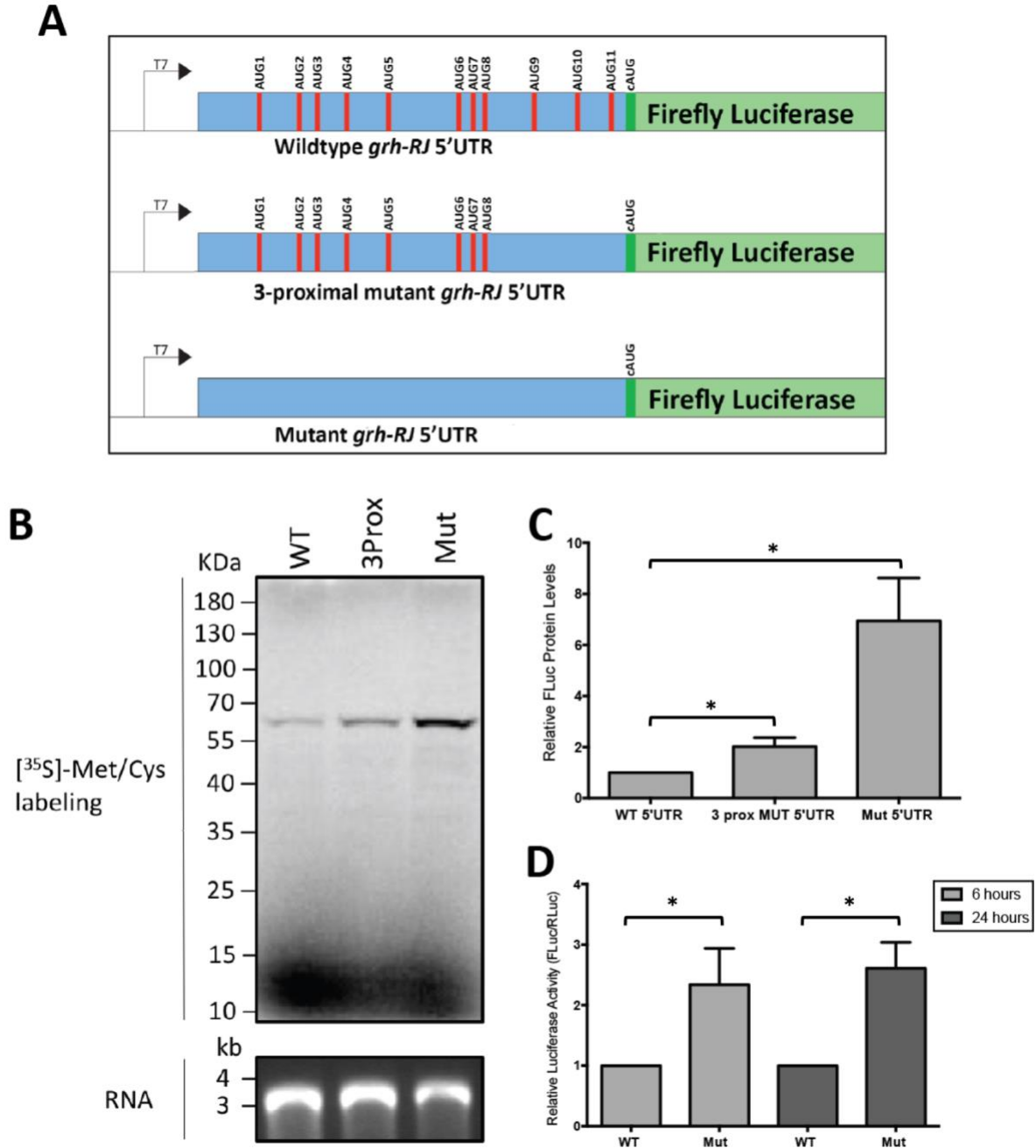
Figure 3.3 *grh-RJ* wildtype 5'UTR repress luciferase activity *in vitro*

(A) Schematics of reporter constructs. Wildtype *grh-RJ* 5'UTR RNA (top) contain all eleven

upstream AUG start sites, the 3-proximal mutant *grh-RJ* 5'UTR RNA (middle) contain nine

upstream AUG start sites with the 3 proximal AUG start sites (AUG9-11) mutated to AUU, and

the mutant *grh-RJ* 5'UTR (bottom) has all AUG start sites mutated to AUU. (B) *In vitro*

synthesized RNA from all three constructs were incubated in RRL translation extracts for 1 h at 30°C in the presence of [35S]-methionine/cysteine. Reactions were resolved be SDS-PAGE and radioactive proteins were visualized by phosphorimager analysis. Shown is a representative gel from three independent experiments. (C) Quantification of firefly luciferase protein bands of mutant *grh-RJ* 5'UTR constructs (3-prox and mutant) relative to wildtype *grh-RJ* 5'UTR. Band intensities were quantified using ImageQuant. Error bars represent the standard deviation. (D) Translational activity in S2 cells. S2 cells are co-transfected with the wildtype *grh-RJ* 5'UTR DNA construct (or mutant *grh-RJ* 5'UTR) and Renilla SV40 DNA constructs for 6 h or 24 h at 25°C. Firefly luciferase activity is normalized to Renilla luciferase activity to obtain a ratio of FLuc/RLuc (conducted by Mika Vivar, Jan Lab). The Fluc/RLuc ratio of the mutant *grh-RJ* 5'UTR construct is normalized to the wildtype *grh-RJ* ratio. Shown are mean ±S.D from three independent replicates. Statistical comparisons were independently conducted between Mut 5'UTR and WT 5'UTR or 3-prox Mut 5'UTR and WT 5'UTR using unpaired parametric Student T-tests. Statistical differences are shown if p<0.05.

3.2 No single *grh-RJ* 5'UTR uAUG has a significant repressive effect on downstream CDS translation

We next investigated the regulatory role of each individual uAUG on the translation activity of the downstream CDS. To this end, we created eleven reporter constructs in which we mutated one of the uAUG's to AUU in the wildtype 5'UTR reporter construct (Figure 3.4). *In vitro* transcribed uAUG mutant and wildtype reporter constructs were incubated in RRL in the presence of [35S]-Met/Cys and the protein levels of FLuc were measured and normalized to the wildtype FLuc protein levels. No *grh-RJ* 5'UTR with a single mutated uAUG>uAUU showed a

significant change in FLuc protein levels compared to the WT *grh-RJ* 5'UTR (Figure 3.4).

These results suggest that no single uAUG, but rather the combined action of multiple uAUGs,

repress downstream CDS translation *in vitro*. Currently, it is unknown if these uAUGs would act

additively or synergistically.
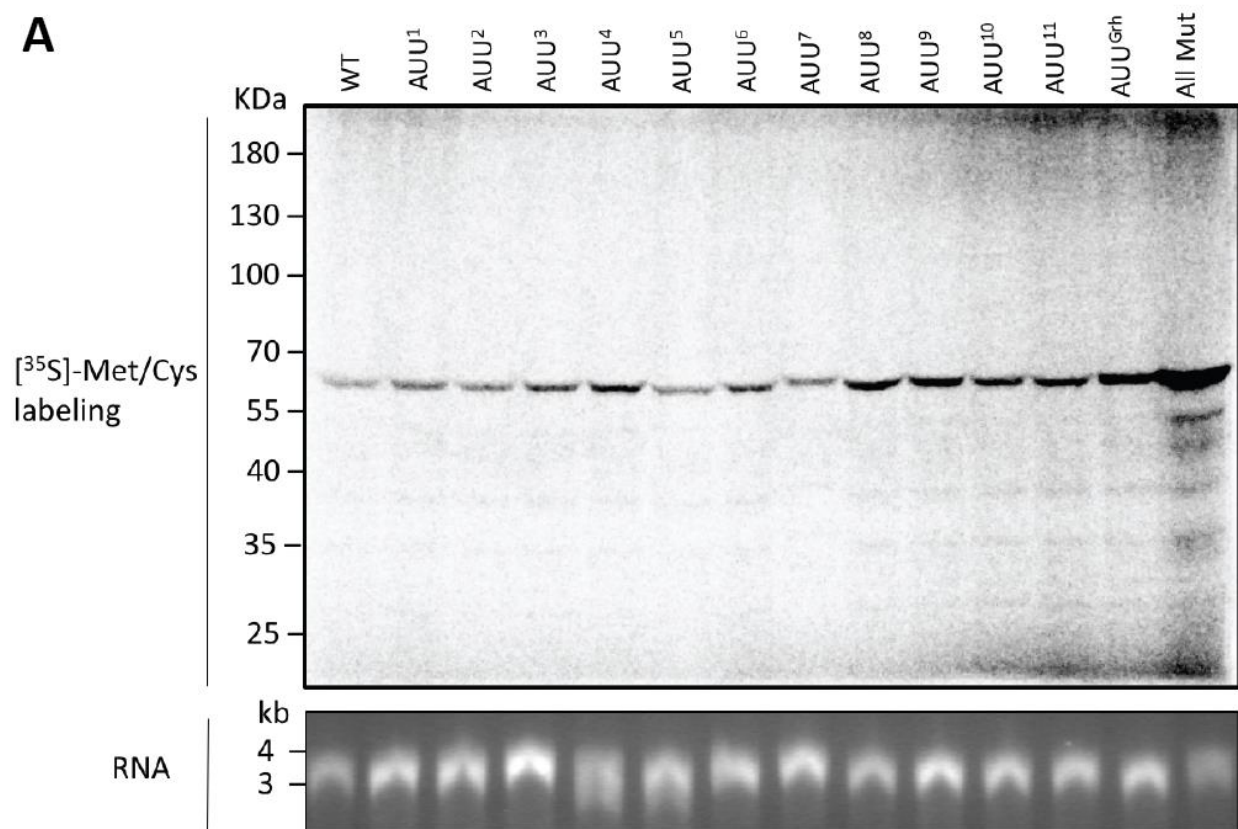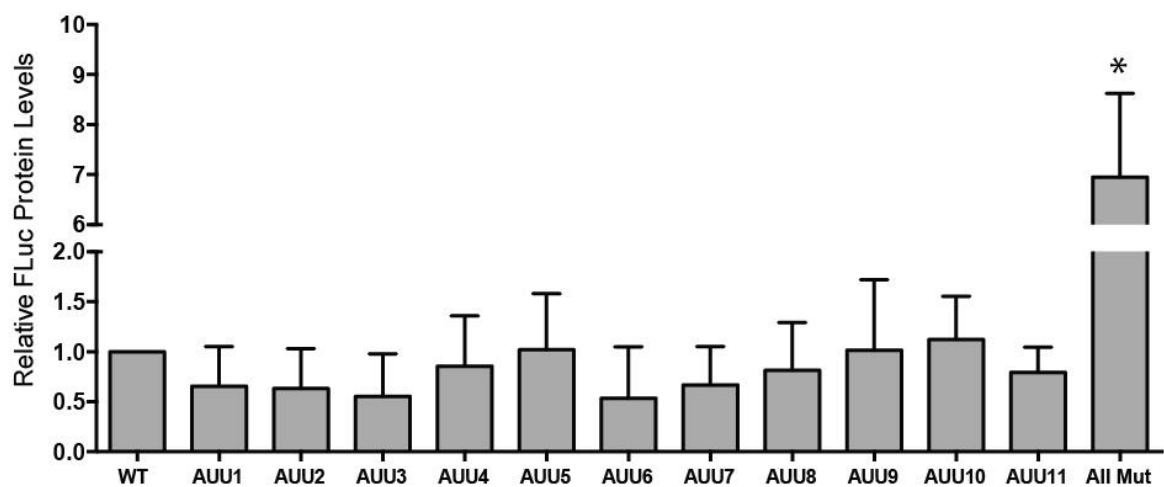
Figure 3.4 No single upstream AUG start site in *grh-RJ's* 5'UTR repress luciferase activity

(A) Each upstream AUG start site was individually mutated to AUU in the wildtype *grh-RJ* 5'UTR backbone. *In vitro* synthesized RNA of all constructs were incubated in RRL translation extracts for 1 h at 30°C in the presence of [^{35}S]-methionine/cysteine. Reactions were resolved by SDS-PAGE and radioactive proteins were analyzed by phosphorimager analysis. Shown is a representative gel from three independent experiments. (B) Quantification of firefly luciferase protein bands of mutant *grh-RJ* 5'UTR constructs (3-prox and mutant) relative to wildtype *grh-RJ* 5'UTR. Band intensities were quantified using ImageQuant. Error bars represent the standard deviation. Shown are mean ±S.D from at least three independent replicates. Data from each mutant 5'UTR construct was compared to the WT 5'UTR construct using unpaired parametric Student T-tests. Statistical differences are shown if p<0.05.

### 3.3 *In vivo* analysis of *grh-RJ's* 5'UTR uORF function on downstream CDS translation.

Using a transgenic approach, we investigated the *in vivo* function of *grh-RJ's* uORFs on the downstream CDS translation of a fluorescent reporter, using the *GAL4/UAS* system in *Drosophila*. We made inducible constructs in which 10 concatemerized upstream activating sequences (10X UAS) and a minimal promoter was placed upstream of the wild type (WT) or Mutant (Mut) *grh-RJ 5'UTR* sequences followed by a *sfGFP*::*NLS* reporter sequence. These constructs contain an *attB* transgenic site for germline integration into the *Drosophila* genome at the *attP40* locus (Figure 3.5 B) (see Materials and Methods) (Bischof et al., 2007). This site-specific genomic integration ensures equal transcript levels for all constructs across different fly

strains, to permit direct comparison of translation rates of the downstream CDS (coding for

sfGFP) downstream of WT and mutant *grh-RJ* 5'UTRs.
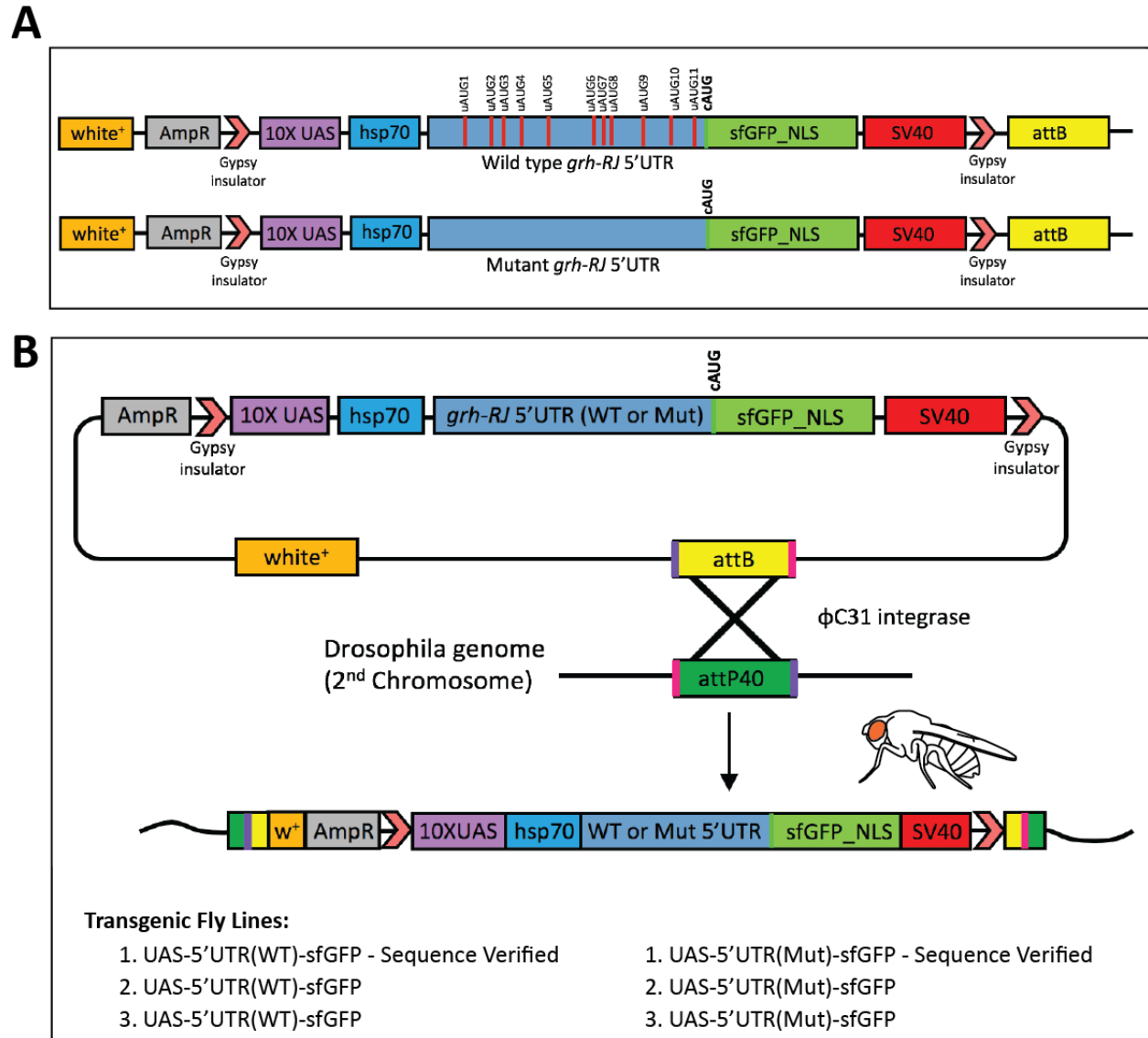
Figure 3.5 Generation *grh-RJ* 5'UTR (WT or Mut) *sfGFP::NLS* reporter in *D.*

melanogaster

(A) WT or uAUG mutated *grh-RJ* 5'UTR was cloned upstream of a superfolder

GFP::*NLS* (*sfGFP*::*NLS*) reporter in frame with the coding AUG start site (cAUG). The

transgenic constructs contain the following: (i) 10 concatemerized upstream activating sequences (10xUAS) that efficiently binds yeast GAL4. (ii) hsp70Bb promoter. (ii) Simian virus 40 (SV40) poly(A) signal. (iv) Gypsy chromatin insulators. (B) Transgenesis was performed by Rainbow Transgenic Flies Inc. (Camarillo, CA, USA). Both WT and Mut reporter constructs were integrated into the attP40 site on the second chromosome of the *D. melanogaster* genome by phiC31 integrase-based transgenesis (Bischof et al., 2007). The resulting progeny were screened for mini-white[+] reporter and crossed to flies carrying the CyO balancer to establish balanced stocks, which were then sequence verified. Three independent homozygous transgenic fly lines were established for WT and mutant 5'UTR sfGFP reporters, following loss of the balancer chromosome.

To test a role for *grh-RJ's* uORFs on CDS translation globally, we used a ubiquitous *daughterless-GAL4 (daGAL4)* driver recombined with *UAS-mCherry::NLS* (an internal control) to drive expression of wildtype or mutant *5'UTR-sfGFP::NLS i*n the whole embryo. As a preliminary test, we analyzed the ubiquitous CDS (*sfGFP::NLS*) expression downstream of the *grh-RJ*'s WT or Mut 5'UTR by comparing spatiotemporal localization and quantitative levels of sfGFP:NLS (normalized to mCherry::NLS) in the CNS at late embryonic stage 17 (Figure 3.6). Using the imaging analysis software, Imaris 9.2.0, mCherry::NLS nuclei were detected in the CNS of both *WT-5'UTR sfGFP::NLS* and *Mut-5'UTR-sfGFP::NLS* genotypes and the relative sfGFP::NLS and mCherry:NLS intensities were measured within the detected nuclei (see Materials and Methods). Using Imaris 9.2.0, the relative sfGFP::NLS intensities were normalized to the mCherry::NLS intensities per cell and compared between *WT-5'UTR sfGFP::NLS* and *Mut-5'UTR-sfGFP::NLS* genotypes (see Materials and Methods).

We found that the WT *grh-RJ* 5'UTR constrained sfGFP::NLS reporter expression in a subset of cells in the VNC (Figure 3.6 A''-A'''), with a general low uniform level of sfGFP::NLS expression within the CNS (Figure 3.6 A – A'''). In contrast, we found that the mutant *grh-RJ* 5'UTR permitted ubiquitous sfGFP::NLS expression at an elevated level in the CNS and showed a loss of spatial expression pattern throughout the VNC (Figure 3.6 B – B'''). These results suggest that *grh-RJ's* WT uORFs have a role in Grh's cell type specific CDS translation (Figure 3.6 A-A''') and the loss of uORFs leads to ectopic expression and an overall elevated level of expression in both ectopic regions and in regions where the sfGFP::NLS downstream of the WT 5'UTR is expressed. The quantified expression level of sfGFP::NLS downstream of Mut 5'UTR was greater than sfGFP::NLS downstream of the WT 5'UTR, suggesting that *grh-RJ's* uORFs have a repressive role on the downstream CDS (Figure 3.6 C).
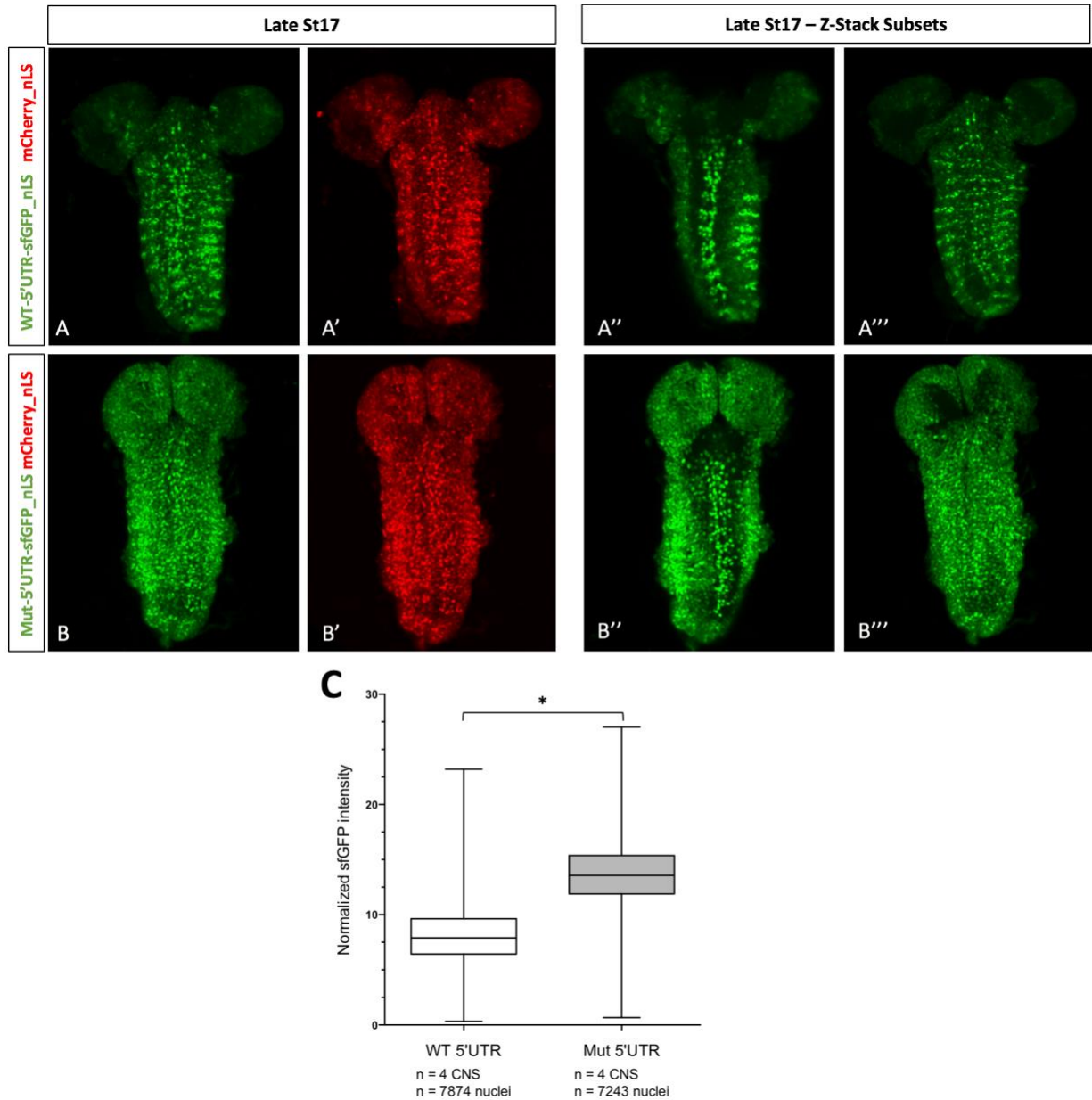
Figure 3.6 uORFs in *grh-RJ's* 5'UTR regulate spatial CDS translation in *Drosophila* CNS

We ubiquitously expressed WT (A-A''') and Mutant (B-B''') *grh-RJ 5'UTR sfGFP::NLS*

reporters using the *daughtlerless-GAL4* driver. In both genotypes, we also co-expressed *UAS-*

mCherry::NLS to serve as an internal control. sfGFP::NLS expression was examined in the

whole intact CNS of late stage 17 embryos imaged by confocal microscopy (A-B'''). A, A', B,

and B' are full projection images showing all optical sections (z-stacks) overlayed. A'', A''', B'',

and B''' are single optical sections (10 z-stacks) taken from two planes of the CNS (A'' and B'' are taken from the same plane, as are A''' and B'''). The mutant *grh-RJ* 5'UTR permitted ubiquitous sfGFP::NLS expression (B''-B'''), but the WT *grh-RJ* 5'UTR exhibited more constrained expression in a subset of cells seen in optical sections (A''-A'''). (C) Using the image analysis software, Imaris 9.2.0 (see Materials and Methods), we quantified the intensity of sfGFP::NLS intensity in each nucleus, and normalized it to the intensity of mCherry::NLS in the same nucleus. In this way, by comparing the expression of sfGFP::NLS and mCherry::NLS driven from the same GAL4 in the same nucleus, we can directly compare the WT and Mut *grh-RJ* 5'UTR sfGFP::NLS intensities to a constant. The boxplots indicate that the WT 5'UTR has a median relative intensity of 7.85 with 50% of the sfGFP nuclei relative intensities lying between 6.35 and 9.60. Whereas, the mutant 5'UTR had a median relative intensity of 13.55 with 50% of the sfGFP nuclei relative intensities lying between 11.87 and 15.35. Relative mean intensities of WT 5'UTR and mutant 5'UTR *grh-RJ* were compared using unpaired parametric Student T-tests. Statistical differences are shown if $p < 0.05$.

We investigated the impact of *grh-RJ's* uORFs at embryonic stage 11 in whole embryos. This time was selected because this is when zygotic Grh protein is first detected in the epidermis (Bray et al., 1989). WT or mutant *grh-RJ 5'UTR-sfGFP::NLS* were driven by the ubiquitous *daughterless-GAL4* driver (together with *UAS-mCherry::NLS)*. Epidermal sfGFP::NLS expression was compared at stage 11 in both high and low expressing Grh nuclei determined by Grh immunohistochemistry and median intensity thresholding (Figure 3.7) (see Materials and Methods). The global level of normalized sfGFP::NLS relative intensities in the mutant *grh-RJ* 5'UTR transgenic embryos was higher compared to the WT *grh-RJ* 5'UTR genotype (Figure 3.7

A – D). However, the spatial pattern of expression of sfGFP::NLS remained unchanged when comparing the WT and mutant 5'UTR (Figure 3.7 A-B). Using median intensity thresholding, we examined the effect of *grh-RJ's* uORFs in cells that express high Grh expression and cells that express no or low Grh (see Materials and Methods). We found that the sfGFP::NLS in WT-5'UTR embryos is expressed at the same level regardless of whether the cells express high or low Grh protein (Figure 3.7 E). Notably, the sfGFP::NLS in Mut-5'UTR embryos was expressed at higher levels compared to the WT-5'UTR embryos in both high and low Grh protein expressing cells (Figure 3.7 E). Overall, this suggests that WT uORFs repress downstream CDS translation in all cell types, regardless of whether the cell expresses high or low amount of Grh protein.
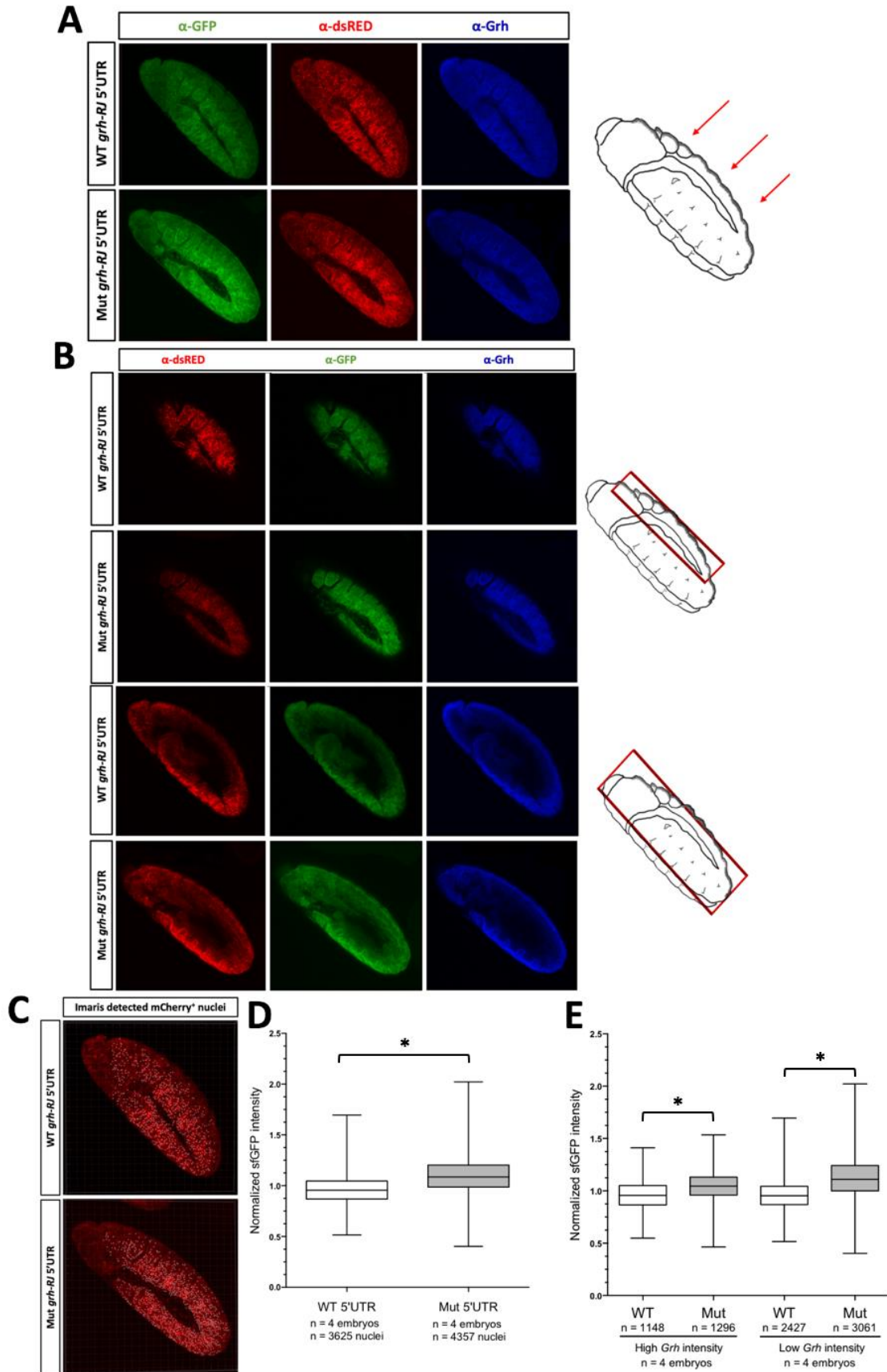
Figure 3.7 *grh-RJ* 5'UTR uORFs repress downstream CDS translation in embryonic epidermal cells

(A-B) *WT* or mutant *grh-RJ 5'UTR sfGFP::NLS* (together with *UAS-mCherry::NLS*) was overexpressed in embryos using *daughterless-GAL4*. Tissues were stained with anti-GFP, anti-dsRed, and anti-Grh. The sfGFP::NLS expression was compared between *WT-5'UTR-sfGFP:NLS* and *Mut-5'UTR-sfGFP:NLS* at stage 11 of embryogenesis. The mutant 5'UTR sfGFP::NLS is expressed at a higher intensity, whereas the WT 5'UTR sfGFP::NLS is expressed at a lower intensity. The mCherry::NLS relative intensities were used as a normalizing control when quantifying sfGFP::NLS relative intensities per nucleus. (C) Using Imaris v9.2, mCherry positive nuclei were detected in both *WT-5'UTR* and *Mut-5'UTR* embryos (see Materials and Methods). (D) Using boxplots, we compared the normalized sfGFP::NLS intensity in all mCherry::NLS positive nuclei of both WT and Mut *grh-RJ* 5'UTR sfGFP::NLS reporters. (E) Using median intensity thresholding, normalized sfGFP:NLS intensities were quantified in cells with high Grh intensity and in cells with low Grh intensity (see Materials and Methods). There is no difference in median sfGFP::NLS relative intensity in the WT 5'UTR nuclei that have high Grh or low Grh intensity. Relative mean intensities of WT 5'UTR and mutant 5'UTR *grh-RJ* were compared using unpaired parametric Student T-tests. Statistical differences are shown if $p<0.05$.

3.4 *grh-RJ's* 5'UTR CRISPR/Cas9 mutant uORF genome editing and phenotypic analysis

The previous experiments suggest that *grh-RJ's* uORFs play an important regulatory role in controlling translation of its downstream CDS in both *in vitro* and *in vivo* transgenic models.

Next, we wanted to address whether *grh's* uORF-mediated translational control is an active control mechanism in transcripts derived from the endogenous *grh* gene locus. Using CRISPR/Cas9 genome editing, I initially wanted to determine the phenotypic consequences of mutating all uAUG start codons of *grh-RJ*. Since *grh-RJ's* 5'UTR is divided into two exons and is separated by a large intron (4.7 kb), CRISPR/Cas9 genome editing of *grh-RJ's* 5'UTR had to be conducted by two separate homology directed repair (HDR) events. We hypothesized that the three proximal uORFs (uORF9-11) would have the high degree of repressiveness on the downstream primarily due to their close proximity to the cAUG, and partly due to the high conservation of uAUG10 and 11 among *Drosophila* species (Table 3.1). Furthermore, uORF11 is likely to be the most repressive due to the low probability of translation re-initiation at the cAUG once uORF11 is translated (Johnstone et al., 2016; Torrance and Lydall, 2018). However, none of the proximal uAUGs 9-11 have a strong Cavener translation start motif. Therefore, I decided to first edit *grh-RJ, RH, RP, RI* and *RL's* three proximal uAUGs (uAUG 9-11) by HDR. This approach was successful and I report the homozygous viability phenotypes below. I also initiated editing the first eight uAUGs (uAUG 1-8), but this is not completed at this time. Using the CRISPR/Cas9-based genome editing approach, we generated fly strains that have point mutated *grh-RJ, RP, RI, RH, and RL's* proximal 5'UTR uAUGs (uAUG9-11) to AUU (Figure 3.8 A). Using guide RNAs (designed by CRISPR Optimal Target Finder, see Materials and Methods) flanking the *grh's* proximal 5'UTR, we constructed an HDR template which incorporates PiggyBac inverted terminal repeats (ITR) flanking an EGFP selection marker for scarless gene editing (Gratz et al., 2015) (see Materials and Methods) (Figure 3.8 B). In parallel, we also generated HDR templates for knocking in *grh-RJ's* WT 5'UTR sequence to control for the PAM site mutations (Figure 3.8 B) (see Materials and Methods).
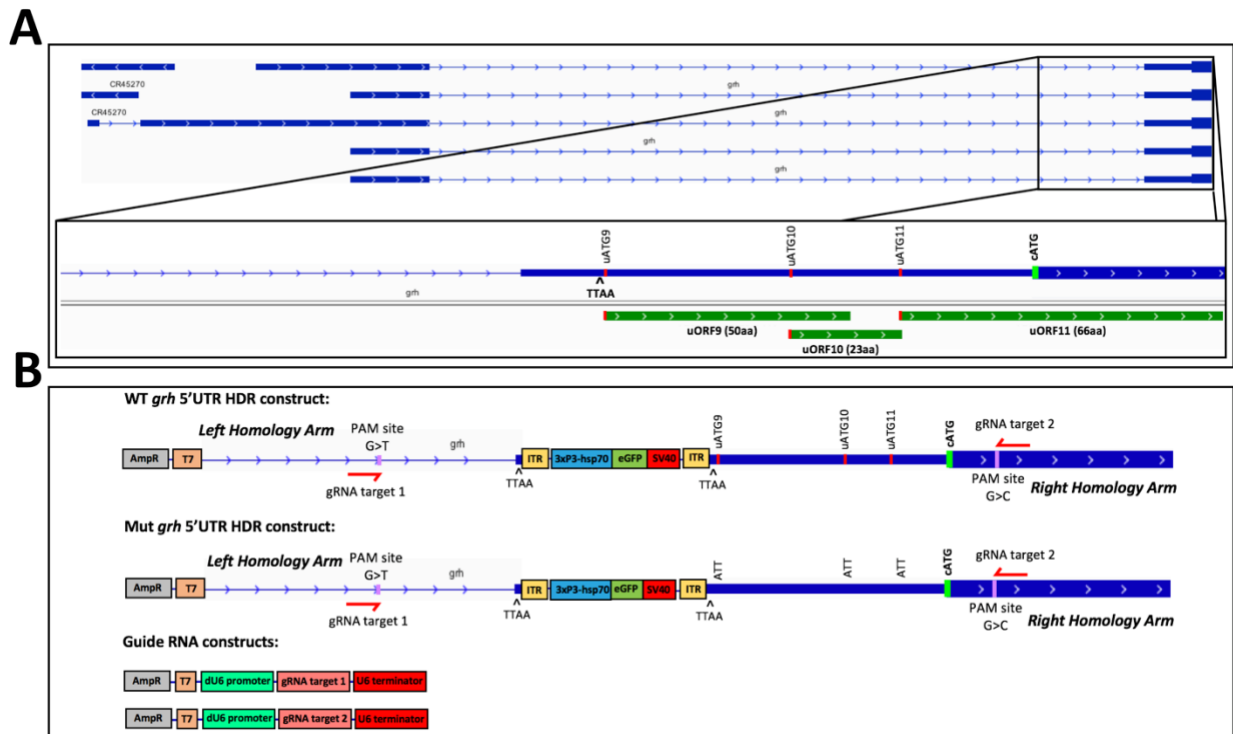
Figure 3.8 Generating WT and 3-prox-Mut HDR plasmids for CRISPR/Cas9 genome editing

(A) *grh-RJ, RH, RP, RI,* and *RL's* proximal 5'UTR contains three uORFs: uORF9 (50 aa), uORF10 (23 aa), and uORF11 (65 aa). None of the proximal uAUGs have a strong Cavener translation start motif and uAUG 10 and 11 are well conserved among *Drosophila* species. (B) We generated two homology directed repair (HDR) templates, the WT and the Mutant *grh-RJ* proximal (*3-prox Mut*) 5'UTR. The left and right homology arms of both WT and 3-prox mutant were PCR amplified from the *Drosophila* genome of the flies with the *nos-Cas9* transgene (injected flies). Protospacer adjacent motif (PAM) site mutations (G>T and G>C) were introduced (and selected based on lack of conservation for this nucleotide – or to create a silent mutation) to prevent further CRISPR/Cas9 gene editing after HDR. An eye-expressing EGFP selection marker cassette was included within a PiggyBac transposon for precise excision that

contains two flanking PiggyBac transposon inverted terminal repeats (ITR), an artificial Pax6 promoter (3xP3) that drives EGFP expression in the eye and a SV40 terminator sequence was inserted at an endogenous "TTAA" site. The three uAUGs in the 3-prox-mut 5'UTR HDR construct were mutated to AUU. We also generated two guide RNA (gRNA) constructs (gRNA target 1 and 2). gRNA scaffolds were inserted upstream of an RNA polymerase III dU6-2 promoter in their own respective pU6 vector (see Materials and Methods). All plasmid sequences were verified by PCR assay and sequencing.

Both guide RNA target plasmids and the WT or 3-prox Mut HDR template were injected into fly embryos that express the Cas9 protein during oogenesis under control of nanos regulatory sequences by Rainbow Transgenic Flies Inc. (Port et al., 2014) (Figure 3.9 A). Injected embryos grew to adulthood and were crossed to w^{1118} flies. F$_1$ progeny were then screened for eyes expressing EGFP and then outcrossed to w^{1118} flies 6 times to remove possible off-target mutations (Figure 3.9 A). The EGFP selection cassette was then removed by crossing to flies that express PiggyBac transposase. After the crossing scheme, the resulting WT and 3-prox Mut homozygote fly progeny were then sequenced to ensure that the proper PAM site mutations and/or 5'UTR mutations were present (Figure 3.9 B).
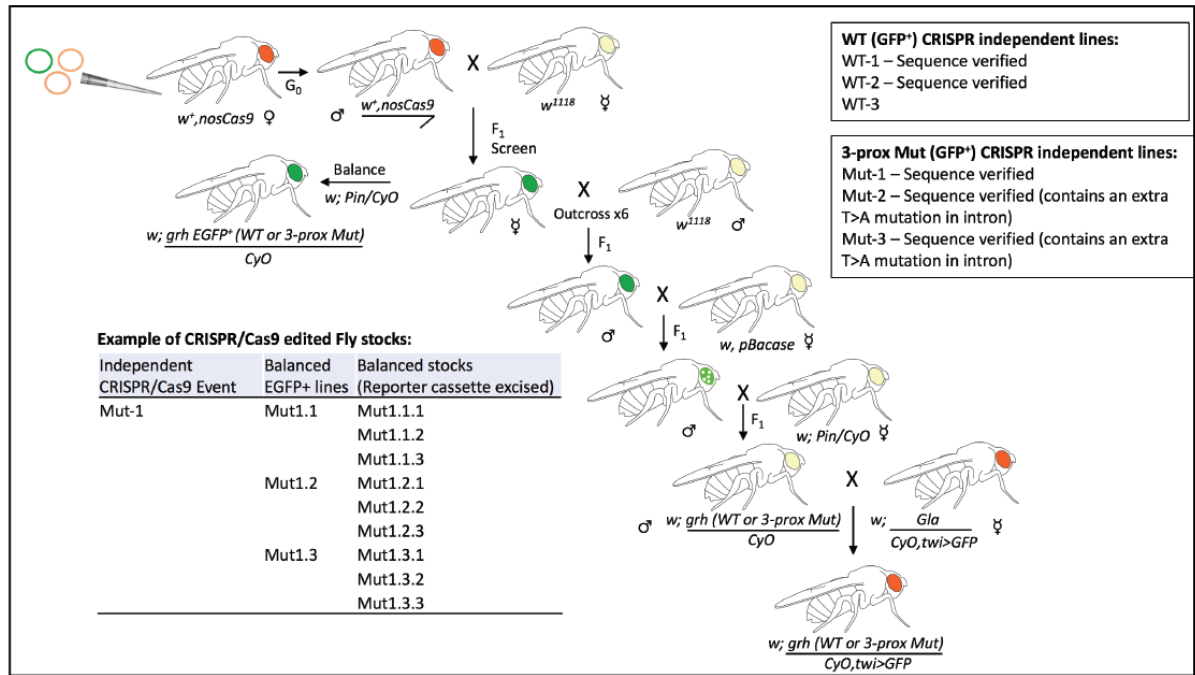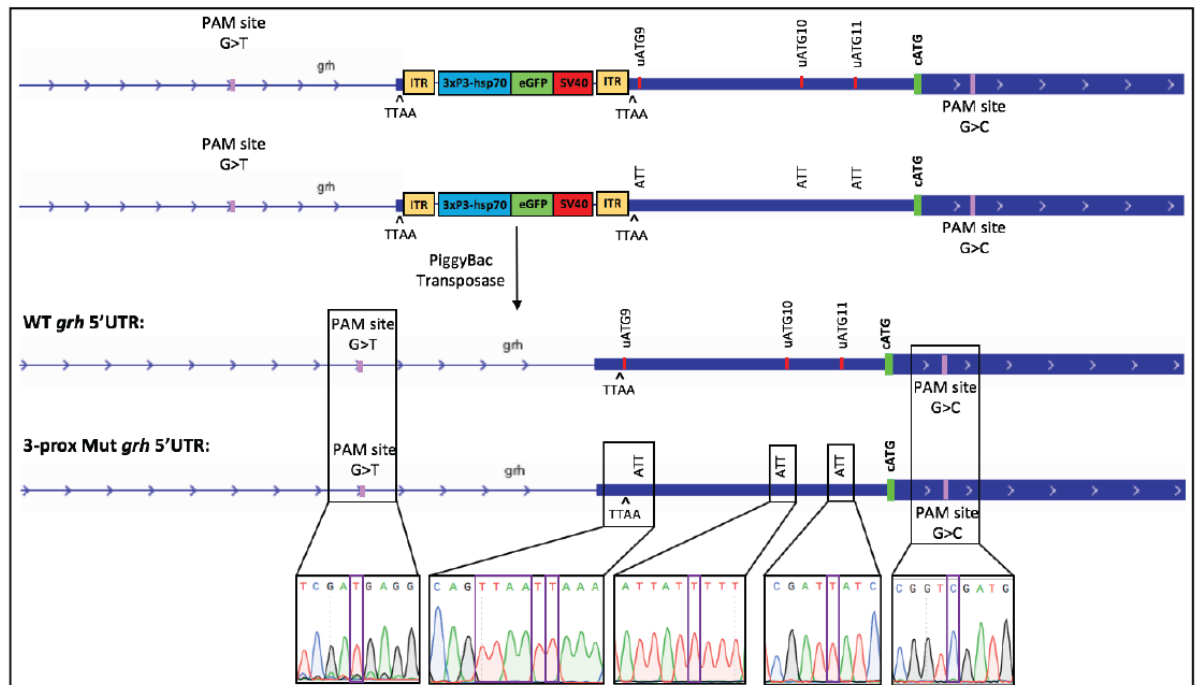
Figure 3.9 CRISPR/Cas9 gene editing of *grh-RJ's* proximal 5'UTR uORFs in *Drosophila*

(A) Crossing scheme for generating CRISPR/Cas9 *grh-RJ* 5'UTR edited lines. Both gRNA

plasmids and the HDR plasmid (WT or 3-prox Mut) were injected into flies with *nos-Cas9*

transgene. Injected embryos were grown to adulthood and the male progeny carrying the *nos-Cas9* transgene on the X-chromosome were crossed to female w^{1118} virgin flies and F_1 progeny were screened for EGFP expressing eyes. Three independent 3-prox Mut *nos-Cas9* and five independent WT *nos-Cas9* male progeny displayed EGFP-positive eyes, indicating that CRISPR/Cas9 homology directed repair had likely occurred. Three males with EGFP positive eyes from each independent CRISPR/Cas9 edited line were crossed to flies carrying the CyO balancer to establish balanced stocks. F_1 GFP-positive female progeny from each independent CRISPR event were then outcrossed six times to w^{1118} males to remove possible off-target CRISPR/Cas9 events. The EGFP reporter cassette was then removed from the outcrossed WT and 3-prox Mut CRISPR/Cas9 edited flies in a scarless manner. Outcrossed GFP$^+$ male progeny were crossed to females carrying PiggyBac transposase. Male progeny with mosaic EGFP-positive eyes (indicative of active PiggyBac transposase activity) were then crossed to white eyed female flies carrying the CyO balancer. Three balanced male progeny without GFP expressing eyes from each line were then crossed to flies carrying the *CyO,twi>GFP* balancer to establish stocks. (B) Schematic showing the WT and Mut *grh-RJ* 2nd exons before and after removal of the PiggyBac selection marker cassette. Final homozygous *grh* WT and 3-prox Mut progeny were then sequenced. Both contain PAM site mutations and the 3-prox Mut sequenced flies contain the appropriate ATT mutation in the proximal 5'UTR.

Using CRISPR/Cas9 genome editing to mutate the three proximal uAUGs in *grh-RJ, RH, RP, RI,* and *RL,* we expected a similar phenotype to a *grh* overexpression phenotype which could include morphogenetic cuticle defects, cell division defects, embryonic mortality, altered neuronal cell fates, NB proliferation/apoptosis defects, tracheal lumen growth defects, and

embryonic dorsal closure failure (Almeida and Bray, 2005; Attardi et al., 1993; Baumgardt et al., 2009; Harrison et al., 2010; Hemphälä et al., 2003; Narasimha et al., 2008). As a preliminary phenotype analysis, we determined if *grh 3-prox-mut* or *WT grh* (with PAM site mutations) flies are homozygous viable. 30 virgin males were crossed with 30 virgin females that are heterozygous for the *grh 3-prox-mut (or WT)* (combined with *CyO,twi>GFP* balance). Flies were allowed to mate for 48 hours at 25°C on fresh medium and F_1 adult progeny were counted. If the flies are homozygous viable, we'd expect a F_1 progeny ratio of 1 homozygous fly to 2 heterozygous flies. F_1 adult progeny that contain *grh 3-prox-mut* mutations produced a 1:2 ratio of homozygous to heterozygous F_1 flies, which indicates that the *grh 3-prox-mut* mutant flies are homozygous viable (Table 3.2). Three out of the four WT *grh* CRISPR lines (contains PAM site mutations) did not give rise to homozygous WT F_1 adult progeny, however one WT *grh* CRISPR line (1.1.3) did give rise to homozygous and heterozygous progeny at the 1:2 expected ratio (Table 3.2). The lack of F_1 WT *grh* homozygous progeny (1.4.3, 1.1.3, 2.1.3) suggests that there are other off-target mutations that causes a lethal homozygous phenotype. Overall, these viability experiments indicate that the *grh-3-prox-Mut* mutations in the *grh-RJ, RH, RP, RI,* and *RL* transcripts are not lethal. However, given the low translation efficiencies of uORF9-11 from Zhang et al. (2018) (Table 3.1), mutating the uAUG9-11 may not cause an impactful increase in translation of the downstream CDS, which may explain the lack of a lethal phenotype.

Table 3.2 - F_1 progeny counts from the heterozygous *grh 3-prox-Mut* and *WT* CRISPR lines

Fly lines	F_1 Heterozygotes (contains CyO,twi>GFP balancer)	F_1 Homozygotes	Total F_1 Progeny
EXPECTED COUNTS	66	34	100
1.1.1 – *3-prox-Mut*	101	36	137
1.2.1 – *3-prox-Mut*	87	53	140
2.2.1 – *3-prox-Mut*	66	28	94
2.3.1 – *3-prox-Mut*	62	48	110

1.1.3 – *WT*	92	32	124
1.4.3 – *WT*	62	0	62
1.1.3 – *WT*	70	0	70
2.1.3 – *WT*	78	0	78

Given the non-lethal phenotype, we decided to analyze Grh's immunoreactivity in the 3-proximal *grh-RJ* 5'UTR mutant homozygotes and w^{1118} flies in stage 11 embryos to determine if there is a detectable increase in Grh protein levels. We expected that if these three uORFs repress CDS translation, then the 3-proximal uORF mutations in *grh-RJ's* 5'UTR would exhibit an increase in Grh immunoreactivity compared to w^{1118} flies. Using Imaris v9.2, we quantified and compared Grh fluorescence intensities of both w^{1118} and 3-prox Mut embryos (Figure 3.10). This experiment should also be repeated with the CRISPR/Cas9 edited WT *grh* flies as they contain the PAM site mutation as a control. Quantified Grh relative intensities did not show an increase in the *3*-prox mutant embryos (Figure 3.10). The mutations of *grh-RJ's* proximal uORFs did not upregulate Grh's expression, thus suggesting that these three uAUG mutations do not have a detectable impact on Grh CDS translation in stage 11 embryos.
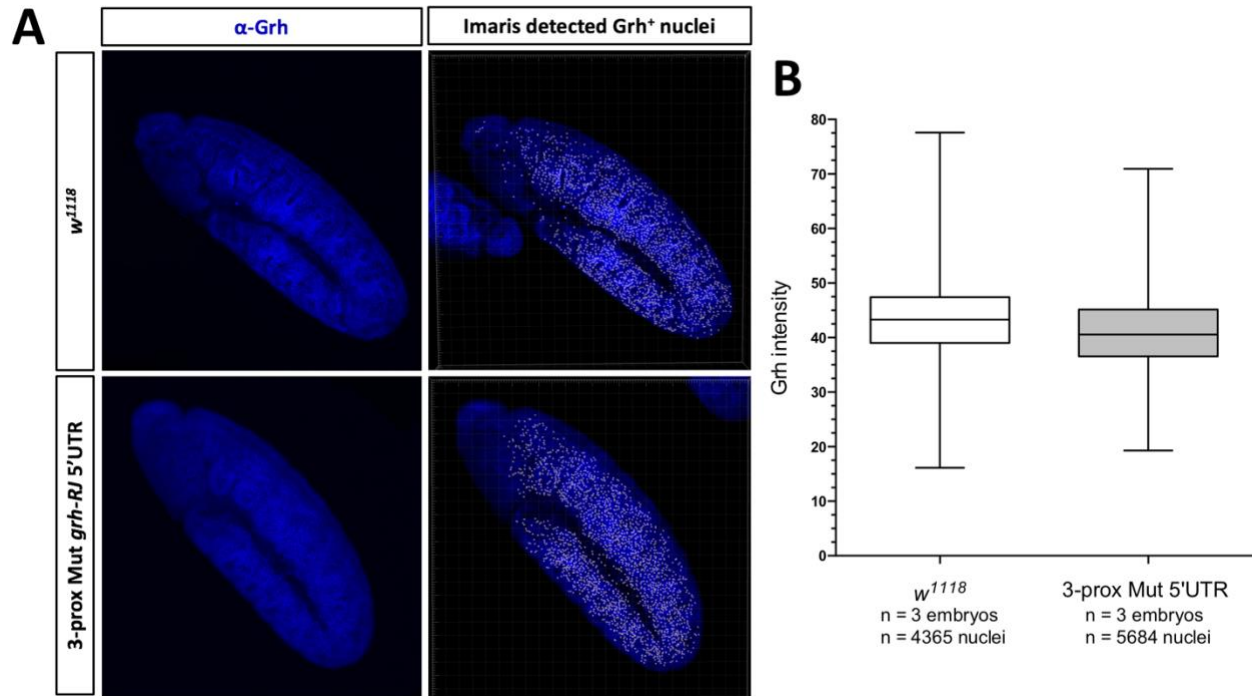
Figure 3.10 Epidermal Grh expression remains unchanged in 3-prox Mut Grh embryos

(A) w^{1118} and 3-prox Mut *grh-RJ* 5'UTR Stage 11 embryos were stained with Grh antibody.

Using Imaris v9.2, approximately 1500 Grh positive nuclei were detected per embryo using

appropriate settings (see Materials and Methods). (B) w^{1118} nuclei have a median relative

intensity level of 42.73 and the *grh* 3-prox Mut 5'UTR nuclei have a median relative intensity of

40.46. Relative mean intensities of WT 5'UTR and mutant 5'UTR *grh-RJ* were compared using

unpaired parametric Student T-tests. Statistical differences are shown if $p<0.05$.

4. DISCUSSION

4.1 Summary of major findings

This thesis aimed to characterize the role of *grh's* 5'UTR uORF-mediated translational control on the downstream CDS as an active and physiological mechanism of protein expression regulation to limit Grh's spatiotemporal activity and protein levels during *Drosophila* development. Specifically, I have used both *in vitro* and *in vivo* assays to evaluate the level of *grh's* uORFs repressiveness on downstream CDS reporters. The major findings of this thesis work include:

In vitro translation assay major findings:

(1) We demonstrated that *grh-RJ's* uORFs repress downstream CDS translation of the Firefly luciferase reporter (Figure 3.2).

(2) Using mutational analysis and *in vitro* translation assays, we demonstrated that there is no single *grh-RJ's* uORF that has a dominant role in repressing translation of the downstream CDS, suggesting that there is a combinatorial role of several uORFs that are responsible in repressing the downstream CDS expression (Figure 3.3).

In vivo major findings:

(1) Using a transgenic approach, we demonstrated that *grh-RJ* uORFs repress the overall downstream CDS translation of the sfGFP::NLS reporter and spatially regulate translation in a subset of cells in the late stage embryonic CNS. We also showed that *grh-RJ's* uORFs repress the downstream sfGFP::NLS reporter expression in Grh expressing epidermal cells.

(2) We found that mutating the three proximal uORFs of *grh* isoforms *grh-RH, RJ, RI, RP,* and *RL* using CRISPR/Cas9 did not result in an increase of Grh expression in stage 11

embryos (from Grh immunostaining) and these *grh 3-prox-Mut* flies are homozygous viable.

4.2 Biological relevance of our findings

Our general understanding of the regulatory functions of uORFs is broadening. The majority of uORFs in eukaryotic 5'UTRs are known to act as repressor elements on the downstream CDS by depleting functional scanning ribosomes (Johnstone et al., 2016). Recent genomic studies involving ribosome profiling and RNA-sequencing enables us to identify sequence features of uORFs that affect the degree of repressiveness on the translation of the downstream CDS. Nevertheless, beyond mere observation that uORFs are translated which implies that they may impact downstream CDS translation, there has been little systematic attempt to determine if they have any role during development. We have chosen to examine this in transcription factors (TFs) that are critical for developmental processes. Developmental TFs have precise spatiotemporal patterns of expression that drive *Drosophila* embryonic patterning and organogenesis, and the loss of TF regulation can result in severe embryonic phenotypes. This study examining the function of *grh's* uORFs in *Drosophila* is one of the first to experimentally show *in vitro* and *in vivo* that *grh-RJ's* uORFs control CDS expression during development.

4.3 uORF regulatory features of *Grh's* 5'UTR on the downstream CDS translation

At the genomic scale, a common feature amongst uORF-containing eukaryotic 5'UTRs is the repressive nature that uORFs may impose on the downstream CDS (Johnstone et al., 2016). From our *in vitro* translation assays and *in vivo* transgenic analysis, *grh-RJ's* uORFs are also repressive in nature by repressing translation of the downstream reporter CDS. Our *in vitro*

results indicate that the removal of the three proximal uAUGs (uAUG 9-11) resulted in a mild increase in downstream CDS translation of reporter, therefore suggesting the combination of these uAUGs have a repressive effect on Grh's CDS translation. Although, this mild increase in downstream CDS translation of reporter due the removal of the three proximal uAUGs is minimal compared to the drastic increase in downstream CDS translation by the removal of all *grh-RJ* uAUGs. Also, recent Ribo-seq studies showed that uORFs 9-11 have low translation efficiencies (Zhang et al., 2018a), which suggest they are not translated and consequently may explain why uORF9-11 have little or no repressive effect on the Grh's CDS. Overall, our results don't exactly elucidate the uORF-mediated translation control mechanism of how ribosomes can bypass the eleven uAUGs in order to initiate translation at *grh's* cAUG. However, given the long 5'UTR of *grh-RJ* (1458 nt), the presence of RNA structural elements such as three-dimensional stem loop structures and/or IRES-like structures could potentially be present in *grh-RJ's* 5'UTR and could be implicated in the regulation of the *grh's* CDS translation. Enhanced global translation conditions are associated with the mTOR signaling pathway and the levels of important eukaryotic initiation factors (eIFs), including eIF4E (Corradetti and Guan, 2006). As seen in C/EBP, overexpression of eIF4E favored the translation initiation at the uORF start site and the inhibition of the mTOR pathway resulted in a decrease of uORF translation initiation (Calkhoven et al., 2000). Perhaps, *grh-RJ's* uORFs are being translated by a similar mechanism which involve an activation of mTOR signaling and/or increase in eIF4E levels. However further studies which include RNA structural analysis and more in-depth *in vitro* translation analysis will be required to help elucidate the *grh's* uORF-mediated translation control mechanism.

4.4 uORF-mediated translation repression and spatial restriction of CDS reporter expression during *Drosophila* development

Mutating *grh-RJ, RH, RP, RI* and *RL's* endogenous proximal uAUGs (uAUG9-11) did not increase Grh immunoreactivity in stage 11 embryonic epidermal cells or lead to lethality. It is likely that the removal of the proximal uAUGs was not able to cause a visible increase in Grh expression due to the presence of the other eight uAUGs in *grh-RJ's* 5'UTR and/or due to the mild loss of the potential repressive effect of uAUG9-11 on Grh's CDS. Nonetheless, further detailed expression analysis could potentially identify phenotypes in the *grh 3-prox-Mut* embryos. But, before we make final conclusions of the repressive role of *grh's* endogenous uORFs on *Grh's* CDS translation, we will have to remove the other eight uAUGs in *grh-RJ's* 5'UTR by CRISPR/Cas9 genome editing. This may lead to a visible Grh overexpression phenotype, and will help us determine the impact and the role of *grh's* uORFs on regulating *grh's* spatiotemporal protein levels.

In spite of the lack of obvious phenotype in the *grh 3-prox-Mut* flies, this thesis has shown that translation of the reporter CDS is repressed by the presence of uORFs in *grh-RJ's* 5'UTR in both transgenic and *in vitro* analysis. This suggests that there is an importance for low basal expression of Grh in order to have adequate Grh-mediated transcriptional gene regulation during *Drosophila* development. As mentioned in section (1.3.4), the onset of Grh protein expression is important for specifying neuroblasts and any deviation in the timing of Grh's protein expression would be deleterious. This deleterious effect was shown by Baumgardt et al (2009) through the overexpression of *grh-RL* in Ap neurons, which resulted in the conversion of all four Ap neruons into an Ap4/FMRFa cell fate (Baumgardt et al., 2009). The uORFs in *grh's* 5'UTR could play an important role in ensuring low levels of Grh within NBs in order to specify

the correct neuronal cell fates. At the syncytial blastoderm stage, the maternally loaded Grh competes with Zelda for TAGteam DNA elements in the promoters of important developmental genes such as *sex lethal* and *zernullt* (Harrison et al., 2010). Overexpression of *grh-RH* caused Grh to outcompete Zelda for DNA binding whiched resulted in an increase in transcriptional repression of Grh's target genes resulting in cell division defects (Harrison et al., 2010). Conceivably, the repressive uORFs in *grh's* 5'UTR may be responsible in maintaining low Grh protein levels in order to prevent Grh from outcompeting Zelda for DNA binding (Figure 4.2). Another example where Grh's level of activity is crucial in maintaining homeostasis is Grh's ability to activate septate junction proteins. Removal of *grh's* uORFs could lead to a Grh overexpression phenotype in the embryo, which involves embryonic dorsal closure defects and embryonic lethality (Attardi et al., 1993; Narasimha et al., 2008).
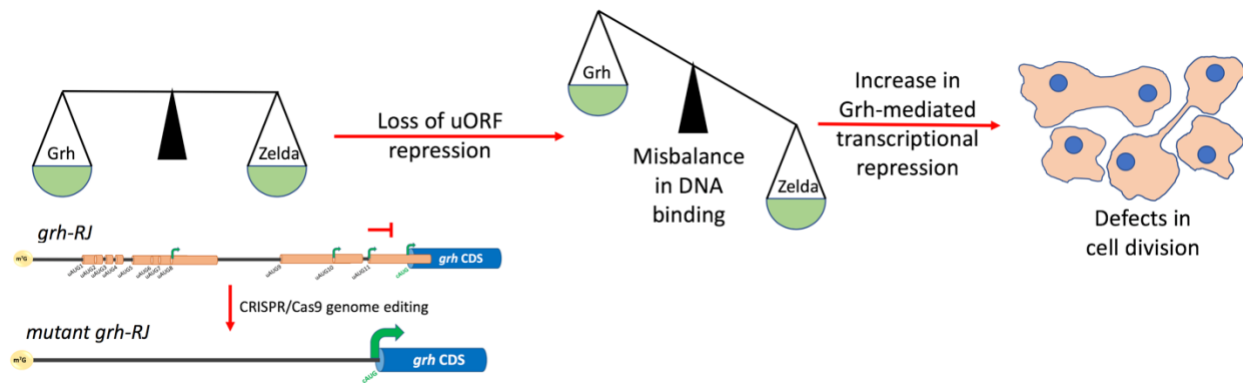
Figure 4.1 - Proposed phenotype of *grh-RJ's* mutant 5'UTR in the syncytial blastoderm. Grh proteins have been shown to outcompete Zelda for binding TAGteam elements in promoters of several important genes (Harrison et al., 2010). We propose that the mutant 5'UTR in *grh-RJ* generated by CRISPR/Cas9 will result in a loss of uORF repression which will lead to an increase in translation of the downstream CDS. This will then cause a misbalance in TAGteam DNA binding activity between Grh and Zelda. This misbalance will likely phenocopy the Grh

overexpression phenotype or Zelda depletion phenotype and result in cell division defects and lethality (Harrison et al., 2010).

Intriguingly, the overexpression experiments of *grh-RJ-WT-5'UTR-sfGFP::NLS* in the CNS showed a restriction of spatial sfGFP::NLS expression in a subset of cells in the VNC, whereas the overexpression of the *grh-RJ-Mut-5'UTR-sfGFP::NLS* showed a loss of this spatial expression pattern and showed overall high levels and ectopic expression of sfGFP::NLS. These results suggest that *grh's* uORFs modulate the downstream CDS translation in a cell specific manner. This novel finding suggests there may be cell-specific mechanisms that function to prevent *grh* uORF repression and promote CDS translation initiation. Grh's onset of expression is tightly controlled based on the cell context and time, which suggests that mechanisms that prevent aberrant Grh expression is crucial for proper development. Our transgenic results suggest that *grh's* uORFs control the spatial expression of Grh, however an analysis of any mechanisms that restrict Grh expression is beyond the scope of this thesis. Complete removal of all *grh* uORFs may help elucidate the uORF mechanism and uORF functional role in controlling the spatial and temporal expression of Grh. By using Grh immunoreactivity, we could determine if there is a loss of spatial restriction of Grh protein expression in the CNS and throughout the whole embryo during *Drosophila* development.

4.5 The neuro tissue specificity of *grh* isoforms (O vs N–isoform)

The *grh* O-isoforms (*grh-RJ, RL, RO, and RN*) are suggested to be neuro-specific based on the previous *in situ* hybridization studies using probes specific to exon 4 and 5 and O-specific Grh reporter expression studies (Uv et al., 1997). Interestingly, based on the RNA-seq results

(performed by Robin Vuilleuimier and Stephane Flibotte, Allan Lab), *grh-RO/N* is the most highly expressed isoform in late stage 3rd instar larval VNC. *Grh-RO/N*'s 5'UTRs contains a single conserved uORF with a uAUG that is surrounded by a poor Cavener sequence context. Our *in vitro* results suggest that *grh-RO/RN's* uORF does not repress translation of the reporter CDS (Appendix Figure A). Previous studies in yeast have shown pervasive translational reprogramming through the usage of alternate transcription start sites (TSS) in order to switch between mRNA isoforms that differ in the content of uORFs (Chen et al., 2017). Switching to an isoform that has a low uORF-containing 5'UTR will allow for a higher translation efficiency of the downstream CDS by avoiding uORF-repressive transcripts (Chen et al., 2017; Cheng et al., 2018). Perhaps *D. melanogaster* has adopted a similar mechanism by using an alternate TSS to switch to the *grh-RO/N* isoforms (which contains a single non-repressive uORF) in the larval VNC to prevent uORF-mediated translation repression present in the other *grh* isoforms (Figure 4.3).
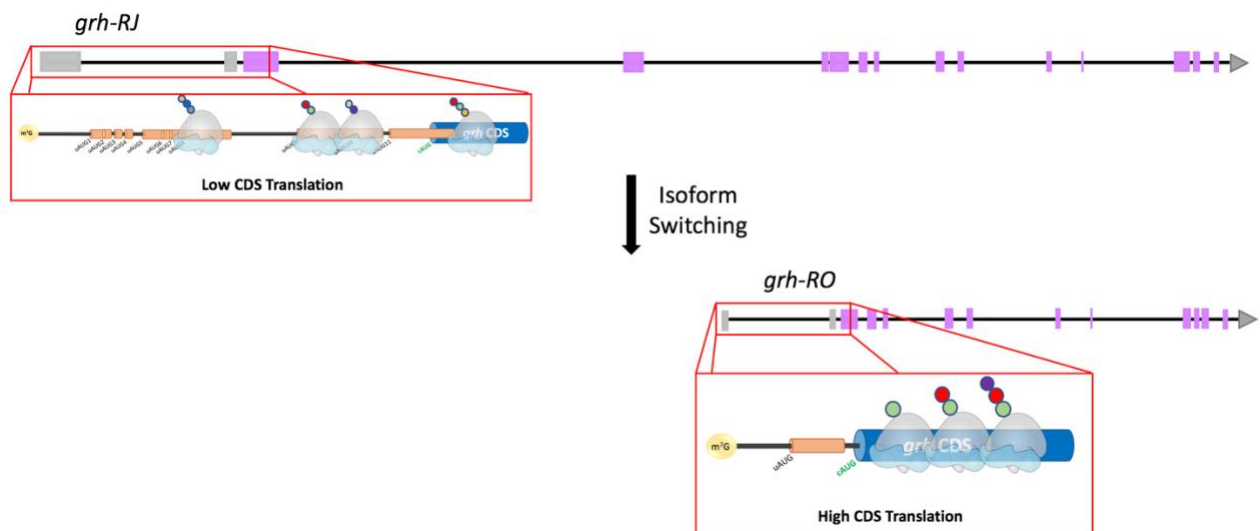
Figure 4.2 – Proposed mechanisms of translational reprogramming of the CDS by switching between mRNA transcript isoform *RJ* to *RO* to prevent uORF repression. In

order to have a higher translation efficiency of *grh's* main CDS, NBs in the larval brain could perhaps switch from *grh-RJ* to *grh-RO* isoform by using alternate transcription start sites (TSS) to prevent uORF-mediated translation repression. Switching to *grh-RO* will allow for a shorter 5'UTR the contains only one uORF with a weak Cavener sequence context, although well conserved among *Drosophila* species. Functional scanning ribosomes can now efficiently scan along the 5'UTR and initiate translation at *grh-RO's* cAUG.

4.6 Considerations and limitations of this study

The repressive function of *grh-RJ*'s uORFs on downstream CDS translation both *in vitro* and *in vivo* is an important finding. However, we must keep in mind the possible limitations and considerations of our study. These include:

(1) Using mutational analysis followed by translation assays to determine which uAUG(s) are primarily responsible for downstream repression of the FLuc reporter presents a major consideration. Mutating a single uAUG in the WT 5'UTR did not have a significant repressive effect. This is perhaps due to the combination effect of the other remaining uAUGs. To circumvent this, we propose to add back each individual uAUG into the mutant 5'UTR and then measure the luciferase activity. This well help determine if that particular uAUG is sufficient for repressing downstream CDS translation.

(2) Using RRL as an *in vitro* translation system has many limitations. First of all, the pre-existing cellular machinery in reticulocytes is highly specialized and how well RRL can recapitulate downstream events such as translation arrest and regulation of translation initiation is unclear. Rabbit reticulocytes lack a nucleus but protein synthesis (primarily of hemoglobin) continues robustly. RRL is treated with a calcium-dependent nuclease

degrades endogenous mRNAs, making it an optimal translation system for exogenous mRNA. This high efficiency of protein synthesis does not recapitulate the endogenous regulatory system in cells, which typically displays lower translation efficiency. To avoid these limitations, using translation extracts from early *Drosophila* embryos may be a better *in vitro* translation system to recapitulate endogenous translation machinery.

(3) Overexpression of *grh-RJ's 5'UTR-sfGFP*::*NLS* using the GAL4/UAS system may have potentially overwhelmed the endogenous regulatory mechanisms. Saturating the endogenous regulatory mechanisms with high levels of mRNA transcripts may cause the uORFs in the *grh-RJ 5'UTR-sfGFP::NLS* to deplete the ribosomal machinery and suppress CDS translation at higher efficiency. To resolve this issue, we plan to completely knock out the other eight distal uAUGs using CRISPR/Cas9 genome editing. This will help us determine the combined function of all uORFs *in vivo* at normal transcript levels.

(4) Using CRISPR/Cas9 genome editing, mutating the three proximal uAUGs (uAUG9-11) of isoforms *grh-RH, RJ, RI, RP,* and *RL* did not lead to a detectable increase in Grh immunoreactivity in stage 11 embryos. Perhaps an increase in Grh expression by altered uORF activity was not detected using immunohistochemistry at stage 11 embryos. To resolve this, knocking in a *sfGFP*::*NLS* protein sequence, followed by a T2A peptide sequence (to prevent the alteration Grh protein structure) at the 5' end of Grh's CDS using CRISPR/Cas9 editing in both WT and mutant 5'UTR. This will allow us to quantify and visualize the endogenous Grh's spatial and temporal expression by directly analyzing sfGFP expression and determine there is a physiological consequence of the loss of *grh* uAUGs.

4.7 Future directions

We have determined that *grh-RJ*'s uORFs repress ribosome initiation at the downstream CDS in both *in vitro* and *in vivo* translation reporter assays. But we have not yet characterized the exact mechanism of translation repression and the endogenous role of *grh-RJ's* uORFs during *Drosophila* development.

4.7.1 Immediate Goals

To help characterize the translation control mechanism of *grh-RJ's* uORFs, the ribosome positioning along *grh's* 5'UTR can be analyzed using an *in vitro* toeprinting assay. Toeprinting, commonly known as the primer extension inhibition assay will provide direct evidence for specific usage of uAUG as translation initiation sites. The toeprinting assay involves a primer extension carried out by a reverse transcriptase that halts once it encounters cycloheximide-stalled initiating 80S ribosomal complexes along the 5'UTR (Hartz et al., 1991; Pestova et al., 1998). Full-length primer extension cDNA products can then be analyzed on a sequencing gel, and strong bands are observed at translation initiating 80S ribosomes (Au et al., 2015). Also, to further help characterize the *grh-RJ's* 5'UTR translation control mechanism, each uAUG(s) will be added back in the mutant 5'UTR one-by-one or as a combination of uAUGs, which will help determine which uAUG(s) is sufficient to repress the downstream CDS translation.

To determine *grh-RJ*'s uORF function in regulating Grh translation during epidermal and CNS development, one could use established transgenic 5'UTR translation assays. This study only examined *sfGFP*::*NLS* reporter expression driven by a ubiquitous GAL4 driver (*daughterless)* at two distinct stages of embryonic development, stage 11 in epidermal cells and

late stage 17 in the CNS. Therefore, examining *grh's 5'UTR-sfGFP::NLS* reporter expression at

other important stages where Grh is known to be expressed and in cellular contexts where

overexpression of Grh resulted in a severe phenotype. Also, using a *grh* specific GAL4 driver

that drives the *5'UTR-sfGFP::NLS* expression in the endogenous pattern of Grh will allow one to

compare *sfGFP::NLS* expression with the expression of Grh itself by examining of the

immunoreactivity of Grh protein. However, it is important to take into account the lag-time in

GAL4/UAS-mediated gene expression and the differences in protein perdurance, maturation

times and detection sensitivity that will exist between sfGFP::NLS and native protein

immunoreactivity. We will generate a *UAS-(Mut or WT)-5'UTR-Grh-RJ* transgene in which the

grh CDS is overexpressed downstream of a uORF-mutated 5'UTR or a WT 5'UTR. This would

allow us to determine if the phenotypic consequences of Grh overexpression is more severe with

the *UAS-Mut-5'UTR-grh-RJ* transgene compared to the *WT-5'UTR-grh-RJ* transgene.

Furthermore, using the same cell-specific GAL4 drivers used in previous Grh overexpression

studies, such as *ap>GAL4,* will allow one to directly compare to the observed Grh

overexpression phenotypes.

4.7.2 Long-term goals

We have only explored the function of *Grh's* uORFs on the translation control of Grh's

downstream CDS during *Drosophila* development. As mentioned before, there are 33224 uORFs

that are occupied by 80S ribosomes. Stephane Filbotte (from the Allan and Moerman lab) has

filtered through Zhang *et al.,* (2018) ribosome profiling data and found 305 developmentally

expressed TFs with a high uORF:CDS ratio of translating ribosome occupancy, during 0-12

hours of *Drosophila* development. Out of these 305 TFs we have prioritized 100 of these TFs for

a critical role in patterning of the embryo and/or specifying cell/tissue fates. Using the tools established in this thesis, it is now possible to rapidly screen 5'UTRs of prioritized developmental TFs using *in vitro* luciferase translation assays. The TFs with uORF-containing 5'UTRs that have a high repressive index on the downstream CDS determined through *in vitro* translation assays can then be selected for the *in vivo* genetic analysis of their 5'uORF function in developmental activity and patterning.

5. CONCLUSIONS

We have made progress in understanding how the uORFs in *grh's* 5'UTR limit translation on the downstream CDS during *Drosophila* development. Using both *in vitro* translation assays and *in vivo* genetic analysis, we have developed methods and a streamlined experimental workflow for analyzing uORF function in important developmental TFs. This approach will allow us further understand a fundamental level of regulation of important TFs and help illuminate the potential regulatory role of uORFs and mechanisms of uORF-mediated translation control during crucial developmental stages.

BIBLIOGRAPHY

Abastado, J.P., P.F. Miller, B.M. Jackson, and A.G. Hinnebusch. 1991. Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for GCN4 translational control. *Molecular and cellular biology*. 11:486-496.

Allerson, C.R., M. Cazzola, and T.A. Rouault. 1999. Clinical severity and thermodynamic effects of iron-responsive element mutations in hereditary hyperferritinemia-cataract syndrome. *The Journal of biological chemistry*. 274:26439-26447.

Almeida, M.S., and S.J. Bray. 2005. Regulation of post-embryonic neuroblasts by Drosophila Grainyhead. *Mechanisms of development*. 122:1282-1293.

Aspden, J.L., Y. Eyre-Walker, R.J. Phillips, U. Amin, M.S. Mumtaz, M. Brocard, and J.-P. Couso. 2014. Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *eLife*. 3.

Attardi, L.D., V.D. Seggern, and R. Tjian. 1993. Ectopic expression of wild-type or a dominant-negative mutant of transcription factor NTF-1 disrupts normal Drosophila development. *Proceedings of the National Academy of Sciences*. 90:10563-10567.

Attardi, L.D., and R. Tjian. 1993. Drosophila tissue-specific transcription factor NTF-1 contains a novel isoleucine-rich activation motif. *Genes & development*. 7:1341-1353.

Au, H.H., G. Cornilescu, K.D. Mouzakis, Q. Ren, J.E. Burke, S. Lee, S.E. Butcher, and E. Jan. 2015. Global shape mimicry of tRNA within a viral internal ribosome entry site mediates translational reading frame selection. *Proceedings of the National Academy of Sciences of the United States of America*. 112:55.

Barbosa, C., I. Peixeiro, and L. Romão. 2013. Gene expression regulation by upstream open reading frames and human disease. *PLoS genetics*. 9.

Baumgardt, M., D. Karlsson, J. Terriente, F.J. Díaz-Benjumea, and S. Thor. 2009. Neuronal

    subtype specification within a lineage by opposing temporal feed-forward loops. *Cell*.

    139:969-982.

Bello, B.C., F. Hirth, and A.P. Gould. 2003. A Pulse of the Drosophila Hox Protein Abdominal-

    A Schedules the End of Neural Proliferation via Neuroblast Apoptosis. *Neuron*. 37:209-

    219.

Bischof, J., R.K. Maeda, M. Hediger, F. Karch, and K. Basler. 2007. An optimized transgenesis

    system for Drosophila using germ-line-specific phiC31 integrases. *Proceedings of the*

    National Academy of Sciences of the United States of America. 104:3312-3317.

Blastyák, A., R.K. Mishra, F. Karch, and H. Gyurkovics. 2006. Efficient and Specific Targeting

    of Polycomb Group Proteins Requires Cooperative Interaction between Grainyhead and

    Pleiohomeotic. *Molecular and cellular biology*. 26:1434-1444.

Boglev, Y., T. Wilanowski, J. Caddy, V. Parekh, A. Auden, C. Darido, N.R. Hislop, M.

    Cangkrama, S.B. Ting, and S.M. Jane. 2011. The unique and cooperative roles of the

    Grainy head-like transcription factors in epidermal development reflect unexpected target

    gene specificity. *Developmental Biology*. 349:512-522.

Bray, S.J., B. Burke, N.H. Brown, and J. Hirsh. 1989. Embryonic expression pattern of a family

    of Drosophila proteins that interact with a central nervous system regulatory element.

    Genes & development. 3:1130-1145.

Bray, S.J., and F.C. Kafatos. 1991. Developmental function of Elf-1: an essential transcription

    factor during embryogenesis in Drosophila. *Genes & development*. 5:1672-1683.

Brody, T., and W.F. Odenwald. 2000. Programmed transformations in neuroblast gene expression during Drosophila CNS lineage development. *Developmental biology*. 226:34-44.

Calkhoven, C.F., C. Müller, and A. Leutz. 2000. Translational control of C/EBPalpha and C/EBPbeta isoform expression. *Genes & development*. 14:1920-1932.

Calvo, S.E., D.J. Pagliarini, and V.K. Mootha. 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proceedings of the National Academy of Sciences*. 106:7507-7512.

Campos-Ortega, J.A., and V. Hartenstein. 1997. The Embryonic Development of Drosophila melanogaster. *springer*:9-102.

Carrara, M., A. Sigurdardottir, and A. Bertolotti. 2017. Decoding the selectivity of eIF2α holophosphatases and PPP1R15A inhibitors. *Nature Structural & Molecular Biology*. 24:708-716.

Cavener, D.R. 1987. Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. *Nucleic Acids Research*. 15:1353-1361.

Cenci, C., and A.P. Gould. 2005. Drosophila Grainyhead specifies late programmes of neural proliferation by regulating the mitotic activity and Hox-dependent apoptosis of neuroblasts. *Development (Cambridge, England)*. 132:3835-3845.

Chappell, S.A., J.P. LeQuesne, F.E. Paulin, M.L. deSchoolmeester, M. Stoneley, R.L. Soutar, S.H. Ralston, M.H. Helfrich, and A.E. Willis. 2000. A mutation in the c-myc-IRES leads to enhanced internal ribosome entry in multiple myeloma: a novel mechanism of oncogene de-regulation. *Oncogene*. 19:4437-4440.

Chen, J., A. Tresenrider, M. Chia, D.T. McSwiggen, G. Spedale, V. Jorgensen, H. Liao, F. van Werven, and E. Unal. 2017. Kinetochore inactivation by expression of a repressive mRNA. *eLife*. 6.

Cheng, Z., G. Otto, E. Powers, A. Keskin, P. Mertins, S. Carr, M. Jovanovic, and G. Brar. 2018. Pervasive, Coordinated Protein-Level Changes Driven by Transcript Isoform Switching during Meiosis. *Cell*. 172:910.

Chew, G.-L., A. Pauli, and A.F. Schier. 2016. Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nature Communications*. 7:11663.

Churbanov, A., I.B. Rogozin, V.N. Babenko, H. Ali, and E.V. Koonin. 2005. Evolutionary conservation suggests a regulatory function of AUG triplets in 5′-UTRs of eukaryotic genes. *Nucleic Acids Research*. 33:5512-5520.

Corradetti, M.N., and K.L.L. Guan. 2006. Upstream of the mammalian target of rapamycin: do all roads pass through mTOR? *Oncogene*. 25:6347-6360.

Cristo, I., L. Carvalho, S. Ponte, and A. Jacinto. 2018. Novel role for Grainy head in the regulation of cytoskeletal and junctional dynamics during epithelial repair. *Journal of cell science*. 131.

Doe, C.Q. 2017. Temporal Patterning in the Drosophila CNS. *Annual review of cell and developmental biology*. 33:219-240.

Dumstrei, K., F. Wang, and V. Hartenstein. 2003. Role of DE-Cadherin in Neuroblast Proliferation, Neural Morphogenesis, and Axon Tract Formation in DrosophilaLarval Brain Development. *Journal of Neuroscience*. 23:3325-3335.

Dynlacht, B.D., L.D. Attardi, A. Admon, M. Freeman, and R. Tjian. 1989. Functional analysis of NTF-1, a developmentally regulated Drosophila transcription factor that binds neuronal cis elements. *Genes & development*. 3:1677-1688.

Edery, I., and N. Sonenberg. 1985. Cap-dependent RNA splicing in a HeLa nuclear extract. *Proceedings of the National Academy of Sciences*. 82:7590-7594.

Gaba, A., A. Jacobson, and M.S. Sachs. 2005. Ribosome occupancy of the yeast CPA1 upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. *Molecular cell*. 20:449-460.

Gratz, S.J., C.D. Rubinstein, M.M. Harrison, J. Wildonger, and K.M. O'Connor-Giles. 2015. CRISPR-Cas9 Genome Editing in Drosophila. *Curr Protoc Mol Biol*. 111:31 32 31-20.

Grosskortenhaus, R., K.J. Robinson, C.Q. Doe, R. Grosskortenhaus, K.J. Robinson, and C.Q. Doe. 2006. Pdm and Castor specify late-born motor neuron identity in the NB7-1 lineage. *Genes & development*.

Harding, H.P., Y. Zhang, H. Zeng, I. Novoa, P.D. Lu, M. Calfon, N. Sadri, C. Yun, B. Popko, R. Paules, D.F. Stojdl, J.C. Bell, T. Hettmann, J.M. Leiden, and D. Ron. 2003. An Integrated Stress Response Regulates Amino Acid Metabolism and Resistance to Oxidative Stress. *Molecular Cell*. 11:619-633.

Harrison, M.M., M.R. Botchan, and T.W. Cline. 2010. Grainyhead and Zelda compete for binding to the promoters of the earliest-expressed Drosophila genes. *Developmental Biology*. 345:248-255.

Hartz, D., D.S. McPheeters, L. Green, and L. Gold. 1991. Detection of Escherichia coli ribosome binding at translation initiation sites in the absence of tRNA. *Journal of Molecular Biology*. 218:99-105.

Hemphälä, J., A. Uv, R. Cantera, and B.-S. Development. 2003. Grainy head controls apical membrane growth and tube elongation in response to Branchless/FGF signalling. *Development*.

Hinnebusch, A.G. 1997. Translational Regulation of Yeast GCN4 A WINDOW ON FACTORS THAT CONTROL INITIATOR-tRNA BINDING TO THE RIBOSOME. *Journal of Biological Chemistry*. 272:21661-21664.

Hinnebusch, A.G. 2005. Translational regulation of GCN4 and the general amino acid control of yeast. *Annual review of microbiology*. 59:407-450.

Hinnebusch, A.G., I.P. Ivanov, and N. Sonenberg. 2016. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science (New York, N.Y.)*. 352:1413-1416.

Hinnebusch, A.G., and J.R. Lorsch. 2012. The Mechanism of Eukaryotic Translation Initiation: New Insights and Challenges. *Cold Spring Harbor perspectives in biology*. 4.

Hinnebusch, A.G., and K. Natarajan. 2002. Gcn4p, a Master Regulator of Gene Expression, Is Controlled at Multiple Levels by Diverse Signals of Starvation and Stress. *Eukaryotic Cell*. 1:22-32.

Huang, J.D., T. Dubnicoff, G.J. Liaw, Y. Bai, S.A. Valentine, J.M. Shirokawa, J.A. Lengyel, and A.J. Courey. 1995. Binding sites for transcription factor NTF-1/Elf-1 contribute to the ventral repression of decapentaplegic. *Genes & development*. 9:3177-3189.

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

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

Ingolia, N.T., L.F. Lareau, and J.S. Weissman. 2011. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell*. 147:789-802.

Isken, O., and L.E. Maquat. 2007. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes & development*. 21:1833-3856.

Isshiki, T., B. Pearson, S. Holbrook, and C.Q. Doe. 2001. Drosophila neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell*. 106:511-521.

Jackson, R.J. 2013. The Current Status of Vertebrate Cellular mRNA IRESs. *Cold Spring Harbor perspectives in biology*. 5.

Jackson, R.J., C.U.T. Hellen, and T.V. Pestova. 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*. 11:113.

Jacobs, J., M. Atkins, K. Davie, H. Imrichova, L. Romanelli, V. Christiaens, G. Hulselmans, D. Potier, J. Wouters, I.I. Taskiran, G. Paciello, C.B. González-Blas, D. Koldere, S. Aibar, G. Halder, and S. Aerts. 2018. The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nature Genetics*. 50:1011-1020.

Jang, S.K., M.V. Davies, R.J. Kaufman, and E. Wimmer. 1989. Initiation of protein synthesis by internal entry of ribosomes into the 5' nontranslated region of encephalomyocarditis virus RNA in vivo. *Journal of virology*. 63:1651-1660.

Janich, P., A.B. Arpat, V. Castelo-Szekely, M. Lopes, and D. Gatfield. 2015. Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames. *Genome research*. 25:1848-1859.

Johnstone, T.G., A.A. Bazzini, and A.J. Giraldez. 2016. Upstream ORFs are prevalent translational repressors in vertebrates. *The EMBO journal*. 35:706-723.

Kambadur, R., K. Koizumi, C. Stivers, J. Nagle, S.J. Poole, and W.F. Odenwald. 1998. Regulation of POU genes by castor and hunchback establishes layered compartments in the Drosophila CNS. *Genes & development*. 12:246-260.

Kapp, L.D., and J.R. Lorsch. 2004. GTP-dependent Recognition of the Methionine Moiety on Initiator tRNA by Translation Factor eIF2. *Journal of Molecular Biology*. 335:923-936.

Kashiwagi, K., M. Takahashi, M. Nishimoto, T.B. Hiyama, T. Higo, T. Umehara, K. Sakamoto, T. Ito, and S. Yokoyama. 2016. Crystal structure of eukaryotic translation initiation factor 2B. *Nature*. 531:122.

Kim, M., and W. McGinnis. 2011. Phosphorylation of Grainy head by ERK is essential for wound-dependent regeneration but not for development of an epidermal barrier. *Proceedings of the National Academy of Sciences*. 108:650-655.

Konarska, M.M., R.A. Padgett, and P.A. Sharp. 1984. Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell*. 38:731-736.

Kozak, M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Research*. 9:5233-5252.

Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*. 44:283-292.

Kozak, M. 1989. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Molecular and cellular biology*. 9:5134-5142.

Kozak, M. 1991a. An analysis of vertebrate mRNA sequences: intimations of translational control. *The Journal of cell biology*. 115:887-903.

Kozak, M. 1991b. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *The Journal of biological chemistry*. 266:19867-19870.

Kozak, M. 2001. Constraints on reinitiation of translation in mammals. *Nucleic Acids Research*. 29:5226-5232.

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

Krishnamoorthy, T., G.D. Pavitt, F. Zhang, T.E. Dever, and A.G. Hinnebusch. 2001. Tight binding of the phosphorylated alpha subunit of initiation factor 2 (eIF2alpha) to the regulatory subunits of guanine nucleotide exchange factor eIF2B is required for inhibition of translation initiation. *Molecular and cellular biology*. 21:5018-5030.

Kurihara, Y., Y. Makita, M. Kawashima, T. Fujita, S. Iwasaki, and M. Matsui. 2018. Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in Arabidopsis. *Proceedings of the National Academy of Sciences*. 115:201804971.

Lee, Y.-Y.Y., R.C. Cevallos, and E. Jan. 2009. An upstream open reading frame regulates translation of GADD34 during cellular stresses that induce eIF2alpha phosphorylation. *The Journal of biological chemistry*. 284:6661-6673.

Lewis, J.D., and E. Izaurflde. 1997. The Role of the Cap Structure in RNA Processing and Nuclear Export. *European Journal of Biochemistry*. 247:461-469.

Liaw, G.J., K.M. Rudolph, J.D. Huang, T. Dubnicoff, A.J. Courey, and J.A. Lengyel. 1995. The torso response element binds GAGA and NTF-1/Elf-1, and regulates tailless by relief of repression. *Genes & development*. 9:3163-3176.

Lincoln, J.A., Y. Monczak, S.C. Williams, and P.F. Johnson. 1998. Inhibition of CCAAT/Enhancer-binding Protein α and β Translation by Upstream Open Reading Frames. *Journal of Biological Chemistry*. 273:9552-9560.

Lu, P.D., H.P. Harding, and D. Ron. 2004. Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *The Journal of cell biology*. 167:27-33.

Mace, K.A., J.C. Pearson, and W. McGinnis. 2005. An Epidermal Barrier Wound Repair Pathway in Drosophila Is Mediated by grainy head. *Science*. 308:381-385.

Macejak, D.G., and P. Sarnow. 1991. Internal initiation of translation mediated by the 5' leader of a cellular mRNA. *Nature*. 353:90-94.

Mackowiak, S.D., H. Zauber, C. Bielow, D. Thiel, K. Kutz, L. Calviello, G. Mastrobuoni, N. Rajewsky, S. Kempa, M. Selbach, and B. Obermayer. 2015. Extensive identification and analysis of conserved small ORFs in animals. *Genome biology*. 16:179.

Maquat, L.E., W.-Y. Tarn, and O. Isken. 2010. The Pioneer Round of Translation: Features and Functions. *Cell*. 142:368-374.

Markstein, M., C. Pitsouli, C. Villalta, S.E. Celniker, and N. Perrimon. 2008. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nature Genetics*. 40.

McGillivray, P., R. Ault, M. Pawashe, R. Kitchen, S. Balasubramanian, and M. Gerstein. 2018. A comprehensive catalog of predicted functional upstream open reading frames in humans. *Nucleic Acids Research*.

Mignone, F., C. Gissi, S. Liuni, and G. Pesole. 2002. Untranslated regions of mRNAs. *Genome biology*. 3.

Miller, P.F., and A.G. Hinnebusch. 1989. Sequences that surround the stop codons of upstream open reading frames in GCN4 mRNA determine their distinct functions in translational control. *Genes & development*. 3:1217-1225.

Morris, D.R., and A.P. Geballe. 2000. Upstream Open Reading Frames as Regulators of mRNA Translation. *Molecular and cellular biology*. 20:8635-8642.

Na, C., M.A. Barbhuiya, M.-S. Kim, S. Verbruggen, S.M. Eacker, O. Pletnikova, J.C. Troncoso, M.K. Halushka, G. Menschaert, C.M. Overall, and A. Pandey. 2018. Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. *Genome Research*. 28:25-36.

Narasimha, M., A. Uv, A. Krejci, N.H. Brown, and S.J. Bray. 2008. Grainy head promotes expression of septate junction proteins and influences epithelial morphogenesis. *Journal of cell science*. 121:747-752.

Neafsey, D.E., and J.E. Galagan. 2007. Dual Modes of Natural Selection on Upstream Open Reading Frames. *Molecular Biology and Evolution*. 24:1744-1751.

Nevil, M., E.R. Bondra, K.N. Schulz, T. Kaplan, and M.M. Harrison. 2017. Stable Binding of the Conserved Transcription Factor Grainy Head to its Target Genes Throughout Drosophila melanogaster Development. *Genetics*. 205:605-620.

Nüsslein-Volhard, C., E. Wieschaus, and H. Kluding. 1984. Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster. *Wilhelm Roux's archives of developmental biology*. 193:267-282.

Nyikó, T., B. Sonkoly, Z. Mérai, A. Benkovics, and D. Silhavy. 2009. Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Molecular Biology*:367.

Passmore, L.A., M.T. Schmeing, D. Maag, D.J. Applefield, M.G. Acker, M.A. Algire, J.R. Lorsch, and V. Ramakrishnan. 2007. The Eukaryotic Translation Initiation Factors eIF1 and eIF1A Induce an Open Conformation of the 40S Ribosome. *Molecular Cell*. 26:41-50.

Patel, N.H. 1994. Imaging neuronal subsets and other cell types in whole-mount Drosophila embryos and larvae using antibody probes. *Methods Cell Biol*. 44:445-487.

Paulin, F., L.E. Campbell, K. O'Brien, J. Loughlin, and C.G. Proud. 2001. Eukaryotic translation initiation factor 5 (eIF5) acts as a classical GTPase-activator protein. *Current Biology*. 11:55-59.

Pearson, J.C., M.T. Juarez, M. Kim, Ø. Drivenes, and W. McGinnis. 2009. Multiple transcription factor codes activate epidermal wound–response genes in Drosophila. *Proceedings of the National Academy of Sciences*. 106:2224-2229.

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

Pestova, T.V., I.B. Lomakin, J.H. Lee, S.K. Choi, T.E. Dever, and C.U. Hellen. 2000. The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature*. 403:332-335.

Pestova, T.V., I.N. Shatsky, S.P. Fletcher, R.J. Jackson, and C. Hellen. 1998. A prokaryotic-like

 mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal

 translation initation of hepatitis C and classical swine fever virus RNAs. *Genes &*

 development. 12:67-83.

Port, F., H.-M. Chen, T. Lee, and S.L. Bullock. 2014. Optimized CRISPR/Cas tools for efficient

 germline and somatic genome engineering in Drosophila. *Proceedings of the National*

 Academy of Sciences. 111.

Ramji, D.P., and P. Foka. 2002. CCAAT/enhancer-binding proteins: structure, function and

 regulation. *Biochem J*. 365:561-575.

Rogozin, I.B., A.V. Kochetov, F.A. Kondrashov, E.V. Koonin, and L. Milanesi. 2001. Presence

 of ATG triplets in 5′ untranslated regions of eukaryotic cDNAs correlates with a 'weak'

 context of the start codon. *Bioinformatics*. 17:890-900.

Ruiz-Orera, J., and A.-M.M. in Genetics. 2018. Translation of small open reading frames: roles

 in regulation and evolutionary innovation. *Trends in Genetics*.

Schleich, S., K. Strassburger, P.C. Janiesch, T. Koledachkina, K.K. Miller, K. Haneke, Y.-S.S.

 Cheng, K. Küchler, G. Stoecklin, K.E. Duncan, and A.A. Teleman. 2014. DENR-MCT-1

 promotes translation re-initiation downstream of uORFs to control tissue growth. *Nature*.

 512:208-212.

Schwer, B., X. Mao, and S. Shuman. 1998. Accelerated mRNA decay in conditional mutants of

 yeast mRNA capping enzyme. *Nucleic Acids Research*. 26:2050-2057.

Shuman, S. 2002. What messenger RNA capping tells us about eukaryotic evolution. *Nature*

 reviews. Molecular cell biology. 3:619-625.

Sonenberg, N., and A.G. Hinnebusch. 2007. New modes of translational control in development, behavior, and disease. *Molecular cell*. 28:721-729.

Spealman, P., A. Naik, G. May, S. Kuersten, L. Freebert, R. Murphy, and J. McManus. 2017. Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. *Genome research*.

Takahashi, H., S. Miyaki, H. Onouchi, T. Motomura, N. Idesako, A. Takahashi, S. Fukuyoshi, T. Endo, K. Satou, S. Naito, and M. Itoh. 2019. Exhaustive identification of conserved upstream open reading frames with potential translational regulatory functions from animal genomes. *bioRxiv*:672840.

Ting, S.B., T. Wilanowski, L. Cerruti, L.-L. Zhao, J.M. Cunningham, and S.M. Jane. 2003. The identification and characterization of human Sister-of-Mammalian Grainyhead (SOM) expands the grainyhead-like family of developmental transcription factors. *Biochemical Journal*. 370:953-962.

Torrance, V., and D. Lydall. 2018. Overlapping open reading frames strongly reduce human and yeast STN1 gene expression and affect telomere function. *PLOS Genetics*. 14.

Traylor-Knowles, N., U. Hansen, T.Q. Dubuc, M.Q. Martindale, L. Kaufman, and J.R. Finnerty. 2010. The evolutionary diversification of LSF and Grainyhead transcription factors preceded the radiation of basal animal lineages. *BMC evolutionary biology*. 10:101.

Tuckfield, A., D.R. Clouston, T.M. Wilanowski, L.-L. Zhao, J.M. Cunningham, and S.M. Jane. 2002. Binding of the RING Polycomb Proteins to Specific Target Genes in Complex with the grainyhead-Like Family of Developmental Transcription Factors. *Molecular and cellular biology*. 22:1936-1946.

Uchiyama-Kadokura, N., K. Murakami, M. Takemoto, N. Koyanagi, K. Murota, S. Naito, and H. Onouchi. 2014. Polyamine-Responsive Ribosomal Arrest at the Stop Codon of an Upstream Open Reading Frame of the AdoMetDC1 Gene Triggers Nonsense-Mediated mRNA Decay in Arabidopsis thaliana. *Plant and Cell Physiology*. 55:1556-1567.

Uv, A.E., E.J. Harrison, and S.J. Bray. 1997. Tissue-specific splicing and functions of the Drosophila transcription factor Grainyhead. *Molecular and cellular biology*. 17:6727-6735.

Uv, A.E., C.R. Thompson, and S.J. Bray. 1994. The Drosophila tissue-specific factor Grainyhead contains novel DNA-binding and dimerization domains which are conserved in the human protein CP2. *Molecular and cellular biology*. 14:4020-4031.

van der Horst, S., B. Snel, J. Hanson, and S. Smeekens. 2019. Novel pipeline identifies new upstream ORFs and non-AUG initiating main ORFs with conserved amino acid sequences in the 5′ leader of mRNAs in Arabidopsis thaliana. *RNA*. 25:292-304.

Vattem, K.M., and R.C. Wek. 2004. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci U S A*. 101:11269-11274.

Venkatesan, K., H.R. McManus, C.C. Mello, T.F. Smith, and U. Hansen. 2003. Functional conservation between members of an ancient duplicated transcription factor family, LSF/Grainyhead. *Nucleic Acids Research*. 31:4304-4316.

Wang, S., and C. Samakovlis. 2012. Chapter two Grainy Head and Its Target Genes in Epithelial Morphogenesis and Wound Healing. *Current Topics in Developmental Biology*. 98:35-63.

Wang, S., V. Tsarouhas, N. Xylourgidis, N. Sabri, K. Tiklová, N. Nautiyal, M. Gallio, and C. Samakovlis. 2009. The tyrosine kinase Stitcher activates Grainy head and epidermal wound healing in Drosophila. *Nature cell biology*. 11.

Werner, M., A. Feller, F. Messenguy, and A. Piérard. 1987. The leader peptide of yeast gene CPA1 is essential for the translational repression of its expression. *Cell*. 49:805-813.

Wethmar, K. 2014. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdisciplinary Reviews: RNA*. 5:765-768.

Wethmar, K., V. Bégay, J.J. Smink, K. Zaragoza, V. Wiesenthal, B. Dörken, C.F. Calkhoven, and A. Leutz. 2010. C/EBPbetaDeltauORF mice--a genetic model for uORF-mediated translational control in mammals. *Genes & development*. 24:15-20.

Wilanowski, T., A. Tuckfield, L. Cerruti, S. O'Connell, R. Saint, V. Parekh, J. Tao, J.M. Cunningham, and S.M. Jane. 2002. A highly conserved novel family of mammalian developmental transcription factors related to Drosophila grainyhead. *Mechanisms of Development*. 114:37-50.

Wright, T.R. 1987. The genetic and molecular organization of the dense cluster of functionally related, vital genes in the DOPA decarboxylase region of the Drosophila melanogaster genome. *Results Probl Cell Differ*. 14:95-120.

Wu, C., J. Wei, P.-J. Lin, L. Tu, C. Deutsch, A.E. Johnson, and M.S. Sachs. 2012. Arginine Changes the Conformation of the Arginine Attenuator Peptide Relative to the Ribosome Tunnel. *Journal of Molecular Biology*. 416:518-533.

Young, S.K., J.A. Willy, C. Wu, M.S. Sachs, and R.C. Wek. 2015. Ribosome Reinitiation Directs Gene-specific Translation and Regulates the Integrated Stress Response. *The Journal of biological chemistry*. 290:28257-28271.

Zhang, H., S. Dou, F. He, J. Luo, L. Wei, and L.-J. biology. 2018a. Genome-wide maps of
ribosomal occupancy provide insights into adaptive evolution and regulatory roles of
uORFs during Drosophila development. *PLoS biology*.

Zhang, H., X. Si, X. Ji, R. Fan, J. Liu, K. Chen, D. Wang, and C. Gao. 2018b. Genome editing of
upstream open reading frames enables translational control in plants. *Nature
biotechnology*.

Zhang, H., Y. Wang, and J. Lu. 2019. Function and Evolution of Upstream ORFs in Eukaryotes.
Trends in Biochemical Sciences.

Zhou, J., J. Wan, X.E. Shu, Y. Mao, X.-M.M. Liu, X. Yuan, X. Zhang, M.E. Hess, J.C. Brüning,
and S.-B.B. Qian. 2018. N6-Methyladenosine Guides mRNA Alternative Translation
during Integrated Stress Response. *Molecular cell*.

APPENDIX

1.1 The uORF in *grh-RO/N's* 5'UTR has no repressive effect on the translation of downstream CDS *in vitro*

To determine the effect of the uORF in *grh-RO* 5'UTR, we constructed a *grh-RO* 5'UTR mutant FLuc reporter construct that consists of the single uAUG mutated to AUU (Figure 3.4). *In vitro* transcribed *grh-RO* WT 5'UTR FLuc and *grh-RO* Mut 5'UTR FLuc constructs were incubated in RRL translation extract in the presence of [^{35}S]-Met/Cys and the translation activity of FLuc was measured and normalized to the wildtype FLuc activity. FLuc activity of the mutant 5'UTR construct was not greater than the wildtype 5'UTR construct (Figure 3.4). To conclude, based on a preliminary test of one biological replicate, *grh-RO's* uORF does not have a suppressive effect on the translation of the downstream CDS, however more biological replicates will need to be conducted to confirm these results.
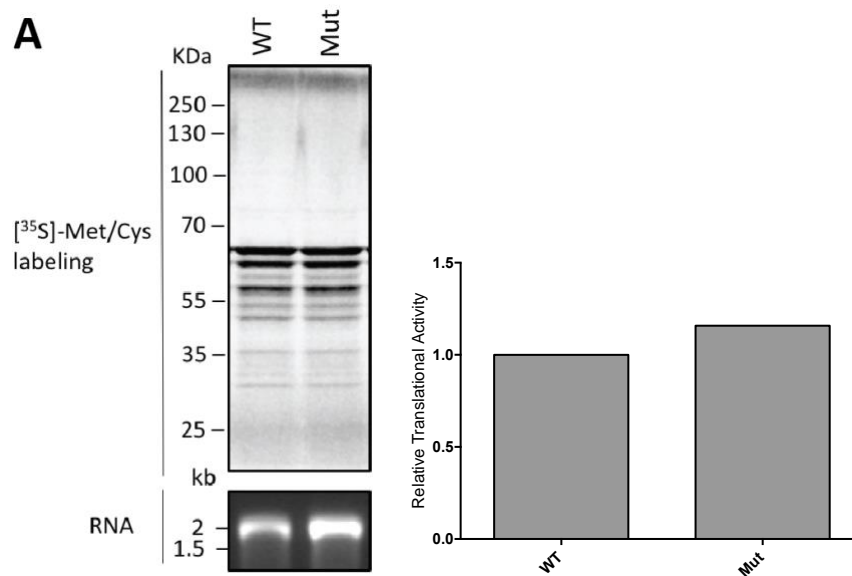
Figure A. *grh-RO* **5' UTR uORF does not inhibit luciferase activity** *in vitro*

(A) *In vitro* synthesized RNA from WT or Mutant (AUU) constructs were incubated in RRL translation extracts for 1 h at 30°C in the presence of [^{35}S]-methionine/cysteine. Reactions were resolved by SDS-PAGE and radioactive proteins were visualized by phosphorimager analysis. The top band below 70 kDa is the Firefly luciferase protein. It is unclear why there are several bands present underneath the Firefly luciferase protein band. Shown is a representative gel from one experiment. Only one biological replicate was conducted.