Post-transcriptional control of gene expression in *Drosophila melanogaster*

Stephanie Yee

Department of Biology
Faculty of Science
McGill University, Montreal
April 2019

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Table of Contents

Abstract

Post-transcriptional control is a critical determinant of gene expression that acts at the level of the messenger RNA (mRNA), which includes processes such as translational control and RNA localization, and is the focus of this thesis. This regulation is in part dictated by the characteristics of the 5' and 3' untranslated regions (UTRs) of the mRNA and the *cis*-elements they harbour.

Translational control can occur at the initiation step where the 5' cap structure of the mRNA is recognized by the eIF4E, whose activity can be modulated by the eIF4E-binding protein (4E-BP), a repressor of translation. The target of rapamycin (TOR) pathway integrates a plethora of signals and impinges on protein synthesis through its action on 4E-BPs and S6 kinases (S6Ks), two well-characterized targets. The TOR/4E-BP/eIF4E axis is known to regulate the translation of subsets of mRNAs with distinct features in their 5'UTRs. In light of recent work that demonstrated dysregulated translation of specific mRNAs in the brains of mice lacking 4E-BP2 and engendering autism spectrum disorder-like phenotypes, we endeavoured to similarly identify mRNAs regulated by d4E-BP in *Drosophila*. In Chapter 2, we performed ribosome profiling to identify specific mRNAs that are translationally regulated downstream of d4E-BP in the adult fly head, used as a proxy for the brain. Gene ontology (GO) analysis revealed that the corresponding genes of some of the upregulated mRNAs are involved in innate immunity. We determined that upregulated mRNAs possess 5'UTRs that are shorter but more complex. In our effort to validate one of the targets, dS6K, we detected elevated levels of p-RPS6, a readout of dS6K activity, in *d4E-BP*^{null} flies. We conclude there are subsets of differentially ribosome-associated mRNAs with distinct 5'UTR features in the *d4E-BP*^{null} fly head.

Subcellular localization of mRNAs in the *Drosophila* embryo establishes a molecular asymmetry of maternally-inherited determinants that is essential for its development. Of the hundreds of transcripts that localize to the primordial germ cells at the posterior of the early embryo, only 55 RNAs accumulate around posterior nuclei prior to the development of those cells, termed RNA islands. Many of the genes that encode these mRNAs have established functions in embryonic patterning and germline development. Despite their common destination to RNA

islands, a shared localization element has yet to be identified. In Chapter 3, we mapped the localization elements within the 3'UTRs of two transcripts that localize to RNA islands, *polar granule component* (*pgc*) and *germ cell-less* (*gcl*). Based on deletion mutation analysis, we report that *gcl* has redundant localization elements, while *pgc* possesses a localization element in the distal region. We show that the localization of polar granule proteins, Oskar, Tudor and Vasa, and 11 RNAs have conserved posterior localization in *Drosophilids*. Using recent findings of a sequence motif that contributes to RNA island localization, we found that this motif is enriched in the 3'UTRs of the majority of transcripts that localize in this way. Our data suggests that the RNA island type of posterior localization is an important process for directing the localization of transcripts with key roles in germline development, as highlighted by the many aspects of this process that is conserved.

Resumé

Le contrôle post-transcriptionnel, qui agit au niveau de l'ARN messager (ARNm), est un déterminant essentiel de l'expression des gènes, et comprend des processus tels que la traduction et la localisation de l'ARN et est le focus de cette thèse. Cette régulation est en partie dictée par les caractéristiques des régions non-traduites (UTRs) 5' et 3' de l'ARNm et des éléments *cis* qu'ils hébergent.

Le contrôle de la traduction peut avoir lieu à l'étape d'initiation où la structure 5 'cap de l'ARNm est reconnue par eIF4E, dont l'activité peut être modulée par la protéine de liaison eIF4E (4E-BP), un répresseur de la traduction. La voie de signalisation de *target of rapamycin* (TOR) intègre une multitude de signaux et empiète sur la synthèse des protéines par son action sur les 4E-BPs et les *S6 kinases* (S6Ks), deux cibles bien caractérisées. L'axe TOR/4E-BP/eIF4E est connu pour réguler la traduction de sous-ensembles d'ARNm présentant des caractéristiques distinctes dans leurs 5'UTRs. Étant donné certains travaux récents démontrant une traduction dérégulée d'ARNm spécifiques dans le cerveau de souris dépourvues de 4E-BP2, engendrant des troubles ressemblant au spectre autistique, nous avons tenté d'identifier de manière similaire les ARNm régulés par d4E-BP chez la *Drosophile*. Dans le chapitre 2, nous avons effectué un profilage ribosomique pour identifier des ARNm spécifiques régulés durant la traduction en aval de d4E-BP dans la tête de mouches adultes, utilisé comme proxy pour le cerveau. L'analyse des ontologies génétiques a révélé que les gènes correspondants à certains des ARNm régulés positivement sont impliqués dans l'immunité innée. Nous avons déterminé que les ARNm régulés positivement possèdent des 5'UTRs plus courts mais plus complexes. Dans un effort pour valider l'une des cibles, dS6K, nous avons détecté des niveaux élevés de p-RPS6, utilisé comme lecture de l'activité de dS6K, chez les mouches *d4E-BP^{null}*. Nous concluons qu'il existe des sous-ensembles d'ARNm associés à des ribosomes différentiels présentant des caractéristiques distinctes 5'UTR dans la tête de mouche *d4E-BP^{null}*.

La localisation subcellulaire des ARNm dans l'embryon de *Drosophile* établit une asymétrie moléculaire des déterminants hérités de la mère, essentielle à son développement. Parmi les centaines de ARN localisés dans les cellules germinales primordiales au postérieur

de l'embryon précoce, seulement 55 ARNs s'accumulent autour des noyaux postérieurs avant le développement de ces cellules appelées « *RNA islands* ». Bon nombre des gènes qui codent pour ces ARNm ont des fonctions établies dans la structuration embryonnaire et le développement de la lignée germinale. Malgré leur destination commune aux *RNA islands*, un élément de localisation partagé n'a pas encore été identifié. Dans le chapitre 3, nous avons cartographié les éléments de localisation dans les 3'UTRs de deux ARNs localisant aux *RNA islands*, *polar granule component* (*pgc*) et *germ cell-less* (*gcl*). D'après l'analyse de mutation par délétion, nous indiquons que *gcl* contient des éléments de localisation redondants, alors que *pgc* possède un élément de localisation dans la région distale. Nous montrons que la localisation des protéines granulaires polaires, Oskar, Tudor et Vasa, ainsi que 11 ARNs ont conservé leur localisation postérieure chez la *Drosophile*. À l'aide de découvertes récentes d'un motif de séquence qui contribue à la localisation aux *RNA islands*, nous avons constaté que ce motif est enrichi dans les 3'UTRs de la majorité des transcrits qui se localisent de cette façon. Nos données suggèrent que la localisation de type postérieure de *RNA islands* est un processus important pour diriger la localisation des transcrits jouant un rôle clé dans le développement de la lignée germinale, comme le montrent les nombreux aspects de ce processus qui sont conservés.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Paul Lasko, for taking a chance on me as an undergraduate student and providing me with the opportunity to do my PhD work in his lab. I am thankful for the academic freedom and the encouragement to become an independent researcher. I also want to thank the members of my supervisory committee, Dr. Eric Lécuyer, Dr. Laura Nilson and Dr. Donald van Meyel, for their guidance and advice over the years.

I would like to express my gratitude to Dr. Christos Gkogkas with whom I worked closely on the 4E-BP project. His mentorship includes the training on the polysome and ribosome profiling techniques, scientific advice, and general encouragement.

I would like to thank all the members of the Lasko lab from whom I learned a great deal, in particular Dr. Mehrnoush Dehghani, Dr. Sanjay Ghosh and Dr. Chiara Gamberi for scientific discussions, life advice, and lengthy gossip sessions. Thank you to Beili Hu for microinjecting of all of my constructs and to Mr. Li for accepting my last-minute fly food orders. I would like to thank Dr. Nahum Sonenberg and Sonenberg lab members for allowing me to conduct experiments in the lab, help with the experiments, and letting me tag along to lab events. I am grateful to Dr. Nam-Sung Moon and Moon lab members for their feedback in our joint lab meetings.

I am thankful for Alexandra Maïly Simond, Dr. Mehrnoush Dehghani and Dr. Christos Gkogkas for reviewing this thesis, and for Dr. Katherine Millette and Alexandra Maïly Simond for proofreading the French abstract. Thank you to my friends, Alexandra Maïly Simond, Jaclyn Hearnden, Dr. Mehrnoush Dehghani, Dr. Callista Yee, Minhee Kim, Mary-Rose Bradley Gill and Keemo Delos Santos, for being there.

Finally, I could not have gotten through my PhD without the unconditional love and support of my family. Thank you to my parents, Stanley and Doris, and to my brothers and sisters-in-law, Victor, Katherine, Edwin and Jenny.

Preface

This thesis presents work for two manuscripts that are in preparation.

Chapter 2:

Stephanie Yee, Christos Gkogkas, and Paul Lasko. Identification and characterization of mRNAs that are translationally regulated by d4E-BP in the *Drosophila* adult head. (In preparation)

Chapter 3:

Stephanie Yee and Paul Lasko. Analysis of mRNAs that localize to RNA islands in the *Drosophila* embryo. (In preparation)

Contribution of Authors

Chapter 2:

Paul Lasko and Christos Gkogkas designed the research. Bioinformatics analysis of ribosome profiling data was carried out by omics2view.consulting GbR, a contractor for the Gkogkas lab, with results shown in Figure 2.2A and B, Figure 2.5B, and Tables 2.1 and 2.2. I performed all experiments, analyzed all other data, and wrote the manuscript.

Chapter 3:

Paul Lasko conceived the project. I performed all experiments, analyzed the data, and wrote the manuscript.

Contribution to Original Knowledge

Chapter 2:

- Determined that global translation is unaffected by the *d4E-BP^{null}* mutation
- Identified specific subsets of mRNAs that are differentially associated with ribosomes in *d4E-BP^{null}* fly heads
- Demonstrated differences in 5'UTR features of mRNAs that are differentially associated with ribosomes where 5'UTRs of upregulated mRNAs are shorter but more complex

Chapter 3:

- Identified regions within the *gcl* and *pgc* 3'UTR that contain localization elements
- Demonstrated that 11 of the 14 mRNAs tested with RNA island type of posterior localization in the early *D. melanogaster* embryo have conserved localization in *D. simulans* and *D. virilis*.
- Demonstrated that Osk, Vas and Tud have conserved posterior localization in *D. simulans* and *D. virilis*

List of Abbreviations

Abbreviation	Definition
4E-BP	eukaryotic initiation factor 4E binding protein
AMP	antimicrobial peptide
ANOVA	analysis of variance
ARE	adenylate-uridylate-rich elements
ASD	autism spectrum disorder
AttB	*Attacin-B*
AttC	*Attacin-C*
BRD-BOX	Bearded box AGCUUUA motif
bru1	*bruno 1*
BSA	bovine serum albumin
Bsg25D	*Blastoderm-specific gene 25D*
C-to-U	cytidine to uridine
C. albicans	*Candida albicans*
cDNA	complementary DNA
CDS	coding sequence
CPE	cytoplasmic polyadenylation element
CycB	*Cyclin B*
D. ananassae	*Drosophila ananassae*
D. melanogaster	*Drosophila melanogaster*
D. pseudoobscura	*Drosophila pseudoobscura*
D. simulans	*Drosophila simulans*
D. virilis	*Drosophila virilis*
d4E-BP	Drosophila eukaryotic translation initiation factor 4E binding protein
DAPI	4',6-diamidino-2-phenylindole
DAVID	Database for Annotation, Visualization, and Integrated Discovery
DGC	Drosophila Gene Collection
DNA	deoxyribonucleic acid
DptB	*Diptericin B*
dT	deoxythymine
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced green fluorescent protein
eIF4A	eukaryotic initiation factor 4A
eIF4E	eukaryotic initiation factor 4E
eIF4F	eukaryotic initiation factor 4F
eIF4G	eukaryotic initiation factor 4G
exu	*Exuperantia*
FISH	fluorescent *in situ* hybridization
GC	guanine-cytosine
gcl	*germ cell-less*
GMP-PNP	guanosine 5′-[β,γ-imido]triphosphate trisodium salt hydrate
GO	gene ontology

grk	gurken
GU-rich	guanidine-uridine-rich
gwl	*greatwall*
GY-BOX	GUCUUCC motif
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IF	immunofluorescence
IM1	*Immune induced molecule 1*
IM3	*Immune induced molecule 3*
Imd	Immune deficiency
IRES	internal ribosomal entry site
JAK/STAT	Janus kinase/signal transducers and activators of transcription
jvl	*javelin-like*
K-BOX	cUGUGAUa motif
MBE	Musashi binding element
miRNA	micro RNA
MNase	micrococcal nuclease
modSP	*modular serine protease*
mRNA	messenger RNA
mTOR	mechanistic target of rapamycin
mTORC1	mechanistic target of rapamycin complex 1
mTORC2	mechanistic target of rapamycin complex 2
ncRNA	non-coding RNA
ND	not determined
nos	*nanos*
ns	non-significant
nt	nucleotide
Osk	Oskar
PAS	polyadenylation site
PCR	polymerase chain reaction
Pdf	Pigment-dispersing factor
pgc	*polar granule component*
PGRP-SD	*peptidoglycan recognition protein SD*
Pi3K21B	*phosphatidylinositol 3-kinase 21B*
Poly(A)	polyadenylation
PRTE	pyprimidine rich terminal element
pum	*Pumilio*
Rev	*Revertant*
RFP	ribosome footprint
RNA-Seq	deep sequencing
RNP	ribonucleoprotein
RPKM	reads per kilobase per million
RPL11	*Ribosomal protein L11*
RPS6	*Ribosomal protein S6*
rRNA	ribosomal ribonucleic acid
RSK	p90 ribosomal S6 kinase
RT-qPCR	reverse transcription quantitative PCR

S2	Schneider 2
S6K	p70 ribosomal protein S6 kinase
SEM	standard error of the mean
SXL	Sex lethal
TE	translation efficiency
Tlk	*Tousled-like kinase*
TOP	5' terminal oligopyrimidine
TOR	target of rapamycin
TORC1	target of rapamycin complex 1
tRNA	transfer RNA
TSC	tuberous sclerosis complex
Tud	Tudor
UAS	upstream activation sequence
UNR	upstream of N-ras
uORF	upstream open reading frame
UTR	untranslated region
Vas	Vasa
yw	*yellow white*
α-tub	*alpha tubulin 84B*
αTub84B	*alpha tubulin 84B*

List of Figures

List of Tables

CHAPTER 1

Literature Review

1.1 Post-transcriptional control

Post-transcriptional control of gene expression refers to the processes that impact the messenger RNA (mRNA), which include translational control and RNA localization. This literature review covers aspects related to these processes.

1.2 Overview of eukaryotic mRNA translation

1.2.1 General structure of an mRNA

Eukaryotic mRNAs typically contain a 5'UTR, open reading frame (ORF) or protein coding sequence (CDS), 3'UTR and poly(A) tail.

1.2.2 5'UTR

All eukaryotic mRNAs contain the m^7GpppN (where m^7G is 7-methylguanosine and N is any nucleotide) cap structure at their most 5' ends. The cap is indispensable for cap-dependent translation as it serves as the point of assembly of the eIF4F complex. It also has functions in protecting against 5' to 3' exonucleases, pre-mRNA splicing, and polyadenylation (Dunckley and Parker, 1999; Edery and Sonenberg, 1985; Flaherty et al., 1997; Hamm and Mattaj, 1990; Izaurralde et al., 1992; Schwer and Shuman, 1996).

Within the 5'UTR are many *cis*-acting elements with regulatory functions, these include upstream open reading frames (uORFs), internal ribosome entry sites (IRESes), terminal oligopyrimidine (TOP) motifs, and pyrimidine-rich terminal elements (PRTEs). The presence of uORFs can impart translation control on the mRNA by reducing the amount of protein synthesized through the primary ORF (Zhang et al., 2019). IRESes mediate the recruitment of the ribosome to the mRNA independent of the 5' cap structure for translation to occur (Hellen and Sarnow, 2001). TOP and PRTE are sequence elements present in the 5'UTRs of mRNAs that typically encode components involved in translation and are found to be regulated in a target of rapamycin (TOR)-eIF4E-binding protein (4E-BP) dependent manner (Hsieh et al., 2012; Thoreen et al., 2012). N^6-methyladenosine (m^6A) is an RNA modification catalysed by the methyltransferase-like 3

(METTL3) m^6A methylase that stimulates cap-independent translation where 5'UTR m^6A is directly bound by eIF3 to recruit the 43S pre-initiation complex in the absence of eIF4E, eIF4A and eIF4B (Meyer et al., 2015). The length and complexity, in terms of guanine-cytosine (GC) content and presence of secondary structures, of the 5'UTR are also important features that affect the translation of an mRNA where those with long and highly structured 5'UTRs are sensitive to levels of eIF4E and are termed "eIF4E-sensitive" mRNAs (Duncan et al., 1987; Koromilas et al., 1992).

1.2.3 Open reading frame

The open reading frame (ORF) of the mRNA contains the protein-coding sequence that dictates protein synthesis beginning with the start codon AUG and ending with the stop codon UAG, AUU or UGA. As the ribosome traverses the mRNA in a 5' to 3' direction, each codon is matched with the appropriate aminoacyl-tRNA cognate via the anticodon to allow the polypeptide chain to grow in a sequential manner. Regulatory motifs have been identified within the CDS in the mRNA, albeit not as commonly seen as with the UTRs.

1.2.4 3'UTR

The 3'UTR refers to the sequences downstream of the stop codon, and similarly to the 5'UTR, it also harbours many regulatory elements that mediate translation, localization and stability. Small noncoding RNAs called microRNAs (miRNAs) can bind to the 3'UTR to repress protein synthesis (Fabian et al., 2010). AU-rich elements (AREs) can mediate mRNA decay (Fan et al., 1997). Musashi Binding Elements (MBEs) can mediate translational repression through the recognition of the minimal element, UAG and GUAG, by Musashi proteins (Bertolin et al., 2016).

1.2.5 Poly(A) tail

The poly(A) tail is a homopolymeric sequence that is acquired by almost all nuclear-transcribed RNAs after undergoing 3' processing co-transcriptionally or post-transcriptionally, with histone transcripts being the exception (Millevoi and Vagner, 2010). Within the 3'UTR exists a hexanucleotide poly(A) signal (PAS), AAUAAA or variants of this, that lies 10-30 nucleotides

upstream of the site in the pre-mRNA where cleavage and polyadenylation occur (Proudfoot, 2011). The poly(A) tail, ranging in size of 50-250 nucleotides, is appended to the pre-mRNA by the template-independent poly(A) polymerase (PAP) (Chang et al., 2014; Choi and Hagedorn, 2003; Colgan and Manley, 1997; Meijer et al., 2007). The regulatory functions of the poly(A) tail include export into the cytoplasm, translational control and mRNA stability (Bernstein and Ross, 1989; Derry et al., 2006; Natalizio and Wente, 2013; Rutledge et al., 2014; Subtelny et al., 2014).

1.3 Cap-dependent translation

mRNA translation is a multistep process that consists of four main phases: translation initiation, translation elongation, translation termination and ribosomal recycling (Hershey et al., 2018). Translation is carried out by a macromolecular complex consisting of proteins and ribosomal RNAs (rRNAs) called the ribosome that moves along the mRNA with 5' to 3' directionality.

1.3.1 Translation initiation

Translation initiation is rate-limiting in protein synthesis and highly regulated. The 43S pre-initiation complex (43S PIC) is a ternary complex composed of the 40S ribosomal subunit, eIF2-GTP, and Met-tRNAi (Hinnebusch and Lorsch, 2012). The 43S PIC is brought to the mRNA through the interaction between eIF4G and eIF3 (Hinnebusch and Lorsch, 2012). Upon binding of eIF4E to the 5' cap structure of the mRNA, eIF4A and eIF4G are recruited to form the eIF4F complex. The 40S subunit and the eIF4F complex scan the RNA, assisted by the RNA helicase activity eIF4A, to unwind secondary structures present in the 5'UTR until the start codon is reached, this complex is now called the 48S initiation complex (Abaeva et al., 2011; Hilliker et al., 2011; Hinnebusch, 2011; Lai et al., 2008; Pisareva et al., 2008). Upon base pairing of the AUG with the Met-tRNAi, eIF2-GTP is hydrolyzed and dissociates while eIF5 promotes the recruitment of the 60S subunit to form the 80S initiation complex (Hinnebusch and Lorsch, 2012).

1.3.2 Translation elongation

Translation elongation refers to the addition of amino acids to the growing polypeptide as the ribosome ratchets along the mRNA. The eukaryotic elongation factor 1A (eEF1A), aminoacyl-tRNA and GTP are brought to the ribosome for base pairing between the codon and the aminoacyl-tRNA anticodon (Dever and Green, 2012). Base pairing leads to GTP hydrolysis by eEF1A and its release (Dever and Green, 2012). A peptide bond is formed between the aminoacyl-tRNA and the peptidyl-tRNA. Binding of the eEF2 GTPase to the ribosome and hydrolysis of the GTP allow the ribosome to translocate (Dever and Green, 2012). The deacylated tRNA leaves and elongation continues (Dever and Green, 2012).

1.3.3 Translation termination and ribosome recycling

The recognition of the stop codon by the ribosome triggers translation termination, mediated by a ternary complex of eukaryotic release factors eRF1 and eRF3-GTP (Hellen, 2018). eRF1 can recognize all three stop codons and induce the release of the polypeptide after its hydrolysis by eRF3 (Hellen, 2018). The disassembly of the ribosome and recycling is mediated by ABCE1 (Heuer et al., 2017).

1.4 eIF4F complex members

1.4.1 eIF4E

The eukaryotic initiation factor 4E (eIF4E) plays a central role in translation initiation. It was first discovered as the protein that cross-links to the 5' cap structure of the mRNA and later purified by m^7GDP affinity chromatography (Sonenberg et al., 1978; Sonenberg et al., 1979). Within a complex with eukaryotic initiation factor 4A (eIF4A) and eukaryotic initiation factor 4G (eIF4G), eIF4E has the ability to stimulate the translation of capped mRNAs (Sonenberg et al., 1980; Tahara et al., 1981). Because eIF4E is least abundant compared to other eIF4F complex members, it is considered to be rate-limiting in translation initiation (Goodfellow and Roberts, 2008).

eIF4E function is regulated post-translationally by binding proteins and by phosphorylation, processes that are themselves regulated through signaling cascades. eIF4E activity can be modulated by the family of eIF4E-binding proteins (4E-BPs) that are regulated downstream of the target of the phosphatidylinositol-3-OH kinase/target of rapamycin (PI(3)K/TOR) pathway (Pause et al., 1994). In its hypo-phosphorylated state, 4E-BP binds to eIF4E to sequester it from eIF4G; but upon phosphorylation by TOR complex 1 (TORC1), 4E-BP's binding affinity to eIF4E is reduced, allowing eIF4E to initiate translation (Gingras et al., 2001). The phosphorylation and activation of MNK1/2 by the mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/ERK) pathway leads to the phosphorylation of eIF4E at Serine 209 (Ueda et al., 2004).

Drosophila has seven eIF4E isoforms 1-7 encoded by six genes where eIF4E2 is a splicing variant of eIF4E1 (Tettweiler et al., 2012). General translation is performed by eIF4E1 which is expressed ubiquitously in all tissues (Hernandez et al., 2005). Phosphorylation on S251, which corresponds to S209 in mammalian eIF4E1 is important for growth and development (Lachance et al., 2002). eIF4E3 has a testis-specific expression and is required for spermatogenesis (Ghosh and Lasko, 2015; Hernandez et al., 2012). It was suggested that eIF4E1 and eIF4E3 in the testes regulate the translation of different mRNAs because the knockdown of one cap-binding protein did not affect the protein levels of the other, and this could be achieved through interactions with partner proteins or indirect binding with their target mRNAs (Ghosh and Lasko, 2015). The biological functions of eIF4E4, eIF4E5, eIF4E6 and eIF4E7 have yet to be established, but their transcripts are enriched in the male testes (Brown et al., 2014). All eIF4E proteins can bind to the 5' cap structure albeit with different binding affinities with the highest observed in eIF4E3 (Zuberek et al., 2016). Another yeast two-hybrid experiment showed strong interactions of eIF4G with eIF4E1, eIF4E2, and eIF4E4; weaker interactions of eIF4G with eIF4E3, eIF4E5 and eIF4E7; but no interactions between eIF4G with eIF4E6 (Hernandez et al., 2005). These results suggest that unlike the other eIF4E proteins, eIF4E6 likely functions as a translational repressor.

1.4.2 eIF4A

eIF4A is the founding member of the DEAD-box family of RNA helicases that contain the conserved DEAD motif, where D is asparagine, E is glutamine and A is alanine (Linder et al.,

1989). As a translation initiation factor, eIF4A unwinds secondary structures in the 5'UTR in an ATP-dependent manner (Rogers et al., 1999). eIF4A helicase activity is enhanced by other translation factors: eIF4E (Feoktistova et al., 2013), eIF4H (Sun et al., 2012), and eIF4G and eIF4B (Gingras et al., 1999).

1.4.3 eIF4G

eIF4G is a large scaffolding protein that can interact with multiple proteins during translation initiation. Within the eIF4F complex, eIF4G acts as a molecular bridge and interacts directly with eIF4E, through the binding motif YXXXXLΦ, where X is any amino acid, L is leucine and Φ is a hydrophobic residue, that is shared by 4E-BP, and with eIF4A (Imataka and Sonenberg, 1997; Mader et al., 1995). Through its interaction with eIF3, a translation initiation factor associated with the 40S ribosomal subunit, the 40S ribosomal subunit is brought to the 5'end of the mRNA, thus beginning ribosomal assembly (Villa et al., 2013). eIF4G can also bind to the poly(A) binding protein (PABP) and induce a circularization of the mRNA when PABP is bound to the poly(A) tail, referred to as the "closed loop model", which is proposed to enhance translation (Borman et al., 2000).

eIF4G can also bind RNA where the RNA-binding activity in *Saccharomyces cerevisiae* eIF4G1 is required for its function in translation *in vitro* (Berset et al., 2003). The RNA-binding region of mammalian eIF4G1 enhances eIF4E binding to the 5' cap by anchoring eIF4E to the mRNA (Yanagiya et al., 2009). eIF4G can mediate cap-independent translation by binding to internal ribosomal entry sites (IRES) in the 5'UTR (Pestova et al., 1996).

Drosophila has two eIF4G isoforms where eIF4G is the canonical protein and eIF4G2 is expressed in testes and is required for male fertility (Baker and Fuller, 2007; Franklin-Dumont et al., 2007; Ghosh and Lasko, 2015). Germline-specific knockdown experiments of eIF4G and eIF4G2 in the testes suggest that they act redundantly during early spermatogenesis but eIF4G2 has a distinct function in the proper meiotic divisions of early germ cells and spermatid elongation (Ghosh and Lasko, 2015). During spermatocyte maturation, specialized eIF4F complexes, eIF4E1-eIF4G2 and eIF4E3-eIF4G2, are likely formed where translation mediated by these complexes is essential for testes development (Ghosh and Lasko, 2015; Hernandez et al., 2012).

1.5 Closed loop model of translation

The closed loop model of translation is the classic model that depicts the circularization of an mRNA through the interactions of cap-eIF4E-eIF4G-PABP-poly(A) to allow 5' to 3' communication based on biochemical, genetic and structural experiments (**Fig. 1.1**) (Vicens et al., 2018). The RNA elements and proteins were shown to bind to each other (Borman et al., 2000; Imataka et al., 1998; Kahvejian et al., 2005; Tarun and Sachs, 1996). Moreover, the binding of PABP-poly(A) can increase the PABP-eIF4G and eIF4E-cap binding affinities resulting in an increase in translation initiation (Borman et al., 2000). Mutations that disrupt the binding of PABP to eIF4G leads to a decrease in levels of translation (Kahvejian et al., 2005). Electron micrographs of polysomes on the endoplasmic reticulum was one of the earliest evidences for circular mRNA that suggested the recycling of ribosomes on the polysome after translation (Christensen et al., 1987). Atomic microscopy was later used to show that the complex formed by eIF4E/eIF4G/Pab1p can circularize mRNA (Wells et al., 1998).

More recent work has raised doubts about the closed model of translation. eIF4G can interact with PABP to promote translation downstream of cap binding (Smith et al., 2017). A herpes simplex virus-1 protein, ICP27, can form a complex with PABP and eIF4G to recruit the small ribosomal subunit, independent of cap-binding, and a similar complex is formed with Deleted in Azoospermia-like (Dazl), eIF4G and PABP (Smith et al., 2017). Single molecule studies suggest that the closed-loop configuration is not formed by mRNAs when translating (Adivarahan et al., 2018; Khong and Parker, 2018). Single-molecule fluorescent *in situ* hybridization (smFISH) demonstrated that ribosome occupancy on mRNAs separates the 5' and 3' ends to adopt an open conformation during translation and the removal of ribosomes from the mRNA leads to the ends coming closer together (Adivarahan et al., 2018; Khong and Parker, 2018).

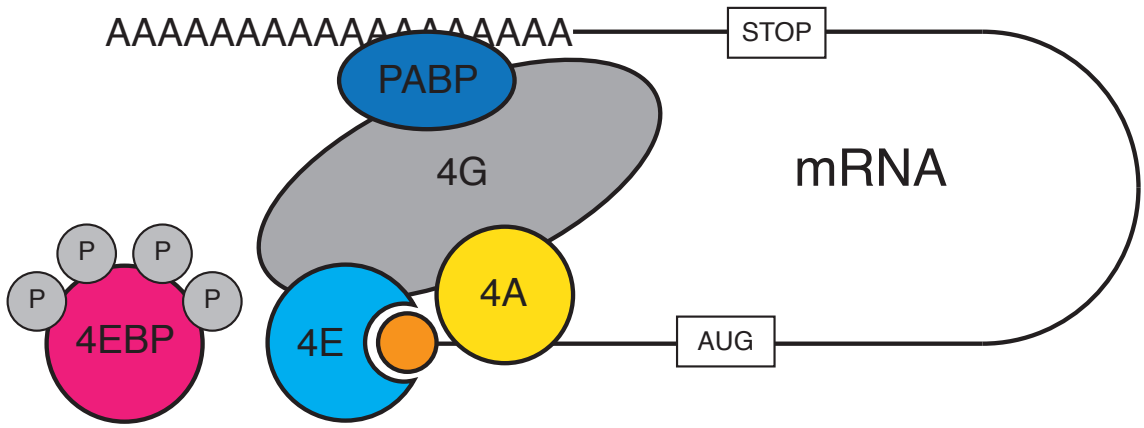


Figure 1.1 Schematic of closed loop model of translation

eIF4E is the cap-binding protein that can recognize the 5' cap structure of the mRNA. eIF4E is regulated by 4E-BP, which has reduced binding affinity when phosphorylated and release from eIF4E to allow translation initiation to take place. eIF4E recruits other eIF4F complex members to the 5' end of the mRNA. eIF4A is an RNA helicase and eIF4G is a scaffold protein. eIF4G can interact with PABP to create a closed loop structure that enables re-initiation of translation.

1.6 eIF4E-binding proteins

1.6.1 4E-BPs

4E-BP belongs to a family of small molecular weight proteins (~15-20 kDa) that bind to eIF4E to inhibit cap-dependent translation (Pause et al., 1994). There exists three 4E-BPs in mammals: 4E-BP1, 4E-BP2 and 4E-BP3. 4E-BP confers a prominent regulatory mechanism by competing for binding with the dorsal surface of eIF4E with eIF4G via the shared YXXXXLΦ motif effectively preventing eIF4F complex formation (Mader et al., 1995). Translation repression by 4E-BP is reversible and its activity is dependent on its phosphorylation status where the hypo- or unphosphorylated state allows it to bind to eIF4E but the hyperphosphorylated state has reduced binding affinity to eIF4E (Pause et al., 1994). TORC1 regulates 4E-BP activity by sequential phosphorylation (Gingras et al., 2001). The key phosphorylation sites in 4E-BP1 that lead to the release of eIF4E are T37, T46, T70 and S65, where T is threonine and S is serine (Wang et al., 2003). S83 on 4E-BP1 can be phosphorylated *in vitro* but exerts little control on 4E-BP1 (Wang et al., 2003). 4E-BP1 contains two more phosphorylation sites, S101 and S112, that are absent in 4E-BP2 and 4E-BP3 (Wang et al., 2003).

1.6.2 *Drosophila* 4E-BP

Unlike mammalian 4E-BPs, there exists a single ortholog in *Drosophila* called d4E-BP or Thor (Bernal and Kimbrell, 2000). The four key phosphorylation sites that regulate eIF4E release, T37, T46, S65 and T70, are identical in d4E-BP, but the eIF4E-binding motif diverged from the canonical YXXXXLΦ motif as YERAFMK, where Y is tyrosine, E is glutamine, R is arginine, A is alanine, F is phenylalanine, M is methionine and K is lysine (Miron et al., 2001). As with mammalian 4E-BP1, d4E-BP is also a downstream target of PI3K/TOR signaling and regulates cell growth (Miron et al., 2003; Miron et al., 2001). Yeast two-hybrid experiments demonstrated the interaction of d4E-BP with eIF4E1, eIF4E2, eIF4E4 and eIF4E5 but not with eIF4E6 or 4E-HP (Hernandez et al., 2012). d4E-BP also does not associate with eIF4E3 as shown by co-immunoprecipitation from total testis extracts (Hernandez et al., 2012).

1.6.3 Biological functions of d4E-BP

d4E-BP was identified as an immune response gene in *Drosophila* when *d4E-BP* transcript was upregulated in response to bacterial infection (Bernal and Kimbrell, 2000; Rodriguez et al., 1996). The canonical NFκB recognition sequence, known to induce immune gene expression, was identified in the *Thor* promoter (Bernal and Kimbrell, 2000). Upon bacterial infection with the non-lethal *Ecc15* pathogen, *d4E-BP* transcription is induced through GCN2-ATF4 signaling and d4E-BP inhibits cap-dependent translation to stimulate the preferential translation of immune transcripts in a cap-independent manner (Vasudevan et al., 2017). *Drosomycin* and *Attacin A*, mRNAs downstream of the Toll and IMD pathways, respectively, can be translated in a cap-independent manner, possibly via IRESes present in their 5'UTRs (Vasudevan et al., 2017). In contrast, infection by a more pathogenic bacterium *Pseudomonas entomophila*, led to a global inhibition of translation in the gut through the activation of GCN2 and inhibition of TOR signaling pathways resulted in impaired immune response and epithelial regeneration (Chakrabarti et al., 2012).

Under dietary restriction, d4E-BP protein is upregulated, and the translation of proteins involved in mitochondrial processes are also upregulated to support lifespan. The 5'UTRs of the preferentially translated mRNAs are shorter and less structured (Zid et al., 2009). d4E-BP activity in the muscle can also regulate lifespan. The forkhead box protein O (FOXO) transcription factor activates *d4E-BP* transcription and d4E-BP in turn controls proteostasis by promoting the clearance of protein aggregates through autophagy, which delays the aging process (Demontis and Perrimon, 2010). d4E-BP can be activated at a lower temperature of 18 °C to mediate a cold-induced longevity. At lower temperatures, there are higher proportions of non-phosphorylated d4E-BP that inhibit global translation but translation of mitochondrial proteins is upregulated, suggesting the switch in metabolic activity mediates longevity (Carvalho et al., 2017). Starvation and oxidative stress can lead to transcriptional activation of *d4E-BP* through FOXO to confer resistance to these stresses (Tettweiler et al., 2005). $d4E\text{-}BP^{null}$ flies had shorter lifespans when they were starved or exposed to 5% H_2O_2 (Tettweiler et al., 2005).

1.7 TOR pathway

The target of rapamycin (TOR) is a nexus for intracellular and extracellular cues to regulate cellular metabolism, growth, proliferation and survival (Laplante and Sabatini, 2012). TOR is a serine/threonine kinase that belongs to two complexes: TOR complex 1 (TORC1) and TORC2. TORC1 is composed of TOR, regulatory protein associated with TOR (Raptor) and mammalian lethal with Sec13 protein 8 (mLST8), proline-rich Akt substrate of 40 kDa (PRAS40) and DEP domain containing mTOR interacting protein (DEPTOR) (Hara et al., 2002; Kim et al., 2002; Kim et al., 2003; Peterson et al., 2009; Sancak et al., 2007). TORC1 regulates growth by promoting protein, lipid and nucleotide synthesis and inhibiting autophagy (Laplante and Sabatini, 2012). TORC2 consists of TOR, DEPTOR, rapamycin insensitive companion of mTOR (Rictor), mSin1 and Protor1/2 and regulates survival and proliferation (Laplante and Sabatini, 2012).

TORC1 regulates translation through phosphorylation of two well-characterized targets, 4E-BP and S6K. The phosphorylation of 4E-BP causes its release from eIF4E to permit cap-dependent translation to occur (Gingras et al., 1999). S6K is phosphorylated on T398 by TORC1 (Dennis et al., 1996). S6K phosphorylates several substrates that are components of the translational machinery. S6K phosphorylates RPS6 on five sites (S235, S236, S240, S244 and S247). Phosphorylated RPS6 is involved in growth and ribosome biogenesis (Chauvin et al., 2014; Ruvinsky et al., 2005). RPS6 can also be phosphorylated by p90 ribosomal S6 kinase (RSK), regulated downstream of ERK1/2, and casein kinase 1 (CK1) (Hutchinson et al., 2011; Roux et al., 2007).

1.8 4E-BP-dependent translation of specific mRNAs

Although eIF4E is regarded as a general translation factor for promoting the translation of all capped mRNAs, there is a pool of mRNAs that are “eIF4E-sensitive” that mostly encode proteins involved in proliferation and survival (Koromilas et al., 1992). These transcripts with 5’UTRs that are long and highly structured are sensitive to eIF4E levels because they are dependent on the recruitment of eIF4A by eIF4E to unwind the secondary structures (Svitkin et al., 2001). The translation of these mRNAs is in turn regulated by TOR and 4E-BP activity. The

pharmacological inhibition of translation by PP242 and metformin identified a subset of "eIF4E-sensitive" mRNAs translated in a 4E-BP-dependent manner that encode proteins that function in proliferation and tumorigenesis (Larsson et al., 2012). A study on 4E-BP2 knockout mice demonstrated that neuroligin mRNAs are translationally upregulated and these mice exhibited an increased ratio of excitatory to inhibitory synaptic inputs and Autism Spectrum Disorder (ASD)-like phenotypes (Gkogkas et al., 2013).

mRNAs with specific sequence motifs, TOP and PRTE, in the 5'UTR are also under the regulation of TOR/4E-BP/4E, and they typically encode ribosomal proteins and components of the translational machinery (Hsieh et al., 2012; Thoreen et al., 2012).

1.9 Other eIF4E-binding proteins

1.9.1 Mextli

Mextli (Mxt) is an eIF4E-binding protein found in invertebrates with roles in *Drosophila* in ovarian germline stem cell maintenance and in early embryogenesis (Hernandez et al., 2013). Through its interactions with eIF4E1 and eIF3, it promotes translation and is therefore a functional analog of eIF4G (Hernandez et al., 2013). In addition to the canonical eIF4E binding motif YXXXXLΦ, *Drosophila* Mxt also possesses a noncanonical binding motif and auxiliary binding sequences, allowing it to interact with eIF4E in a tripartite binding mode (Hernandez et al., 2013; Mader et al., 1995; Peter et al., 2015). The absence of the auxiliary binding sequences in *Caenorhabditis elegans* Mxt means that it uses a bipartite binding mode (Peter et al., 2015). The differences in binding modes confer different functional properties: the tripartite binding mode makes dMxt less competitive against eIF4G for eIF4E binding, but once it is in a complex with eIF4E, it is more difficult to be displaced by 4E-BPs or eIF4G; in contrast, the bipartite binding mode of *C. elegans* Mxt is more competitive for eIF4E binding against eIF4G, but more susceptible for displacement once in a complex with eIF4E (Peter et al., 2015).

1.9.2 Cup

Cup is an insect-specific eIF4E-binding protein that represses the translation of several maternal *Drosophila* mRNAs, *oskar*, *nanos* and *gurken*, prior to their posterior localization (Clouse et al., 2008; Nakamura et al., 2004; Nelson et al., 2004). Its inhibitory function in translation is attributed to the canonical eIF4E-binding motif YXXXXLΦ and a non-canonical motif for competing with eIF4G (Kinkelin et al., 2012; Mader et al., 1995; Nelson et al., 2004). The translation of *osk* mRNA in oogenesis is repressed through a 5'/3' interaction where Bruno binds to a Bruno response element (BRE) in the 3'UTR and to Cup that competes with eIF4G for eIF4E binding (Nakamura et al., 2004). Cup represses the translation of *nos* through its interaction with Smaug (Smg), an RNA binding protein that recognizes stem loops present in the *nos* 3'UTR called Smaug recognition elements (SREs) (Nelson et al., 2004). The translational repression of unlocalized *grk* in the oocyte is mediated by Cup, Squid (Sqd) and Bruno where Bruno binds to *grk*, Sqd and Cup, which sequesters eIF4E from eIF4G (Clouse et al., 2008).

1.9.3 Bicoid and 4EHP

4E homology protein (4EHP) is an eIF4E-related cap binding protein that does not interact with eIF4G (Hernandez et al., 2005). In the *Drosophila* embryo, d4EHP represses the translation of specific mRNAs. The *caudal* mRNA contains a Bicoid Binding Region (BBR) in its 3'UTR that is recognized by Bicoid, which then recruits d4EHP that binds the cap to inhibit translation (Cho et al., 2005). Translation of *hunchback* can also be repressed through a complex of cap-bound d4EHP, Pumilio, Brat and Nanos that recognizes the Nanos responsive element (NRE) in the *hunchback* 3'UTR (Cho et al., 2006).

1.10 RNA localization and translational control

The *Drosophila* oocyte and embryo are excellent model systems to study RNA localization and translational control where spatial restriction of maternally-inherited determinants establishes a molecular asymmetry that is essential for development. Large scale RNA *in situ* hybridization screens show that mRNA localization is widespread in the ovary and in the embryo with the

majority of mRNAs displaying patterns of localization (Jambor et al., 2015; Lecuyer et al., 2007; Tomancak et al., 2007). The correlation between the localization of some mRNAs and the proteins they encode suggests the availability of the translational machinery is also spatially restricted (Lecuyer et al., 2007). Due to the general absence of zygotic transcription in the early embryo, post-transcriptional control of maternal RNAs becomes all the more important.

1.11 Overview of Drosophila oogenesis

Each ovary consists of approximately 18 ovarioles, with each developing as an assembly line of 14 morphologically defined stages (Bastock and St Johnston, 2008). The earliest stage is called the germarium where a germline stem cells divide asymmetrically to give rise to a stem cell and daughter cell. The daughter cell divides four times with incomplete cytokinesis to generate a 16-cell cyst that remains interconnected by cytoplasmic bridges called ring canals. One of the cells will differentiate as the oocyte and complete meiosis while the remaining 15 cells become the polyploid nurse cells that synthesize and transports those products into the oocyte. RNAs that are synthesized in the nurse cells are transported into the oocyte via ring canals, some of which have key roles in embryonic patterning and germ cell specification (St Johnston, 2005). A layer of follicular epithelial cells envelopes the cysts as they move through the germarium. The development of the oocyte in the posterior and the differentiation of follicle cells establish egg chamber polarity (Assa-Kunik et al., 2007; Roth and Lynch, 2009). At stage 8, the oocyte will increase in volume from yolk protein synthesis and uptake. At stage 11, nurse cells will expel their cytoplasm into the oocyte by contraction and then undergo apoptosis (Cavaliere et al., 1998). By stage 14, the oocyte has full matured (McLaughlin and Bratu, 2015). During oocyte development, a specialized cytoplasm at the posterior end is established called the germ plasm or pole plasm, which consists of maternal proteins, including Oskar, Vasa and Tudor, and maternal RNAs, including *polar granule component* (*pgc*), *germ cell-less* (*gcl*) and *nanos* (*nos*) (Mahowald, 2001).

1.12 Overview of *Drosophila* embryogenesis

After the egg is fertilized, there are 13 rounds of nuclear division, in the absence of cytokinesis, within a shared cytoplasm of the embryo or the syncytial blastoderm. After eight

nuclear divisions, nuclei migrate to the periphery of the embryo. During cycle 9 or stage 3 of embryogenesis (Campos-Ortega and Hartenstein, 1985), nuclei at the posterior end will begin to bud to form the pole cells of the syncytial blastoderm, or stage 4 embryo. After cycle 13, cellularization occurs in the periphery forming the cellular blastoderm at stage 5. Pole cells formed at the posterior will eventually migrate inwards as embryogenesis progresses to eventually form the embryonic gonads (Dansereau and Lasko, 2008).

The development of the early embryo, within the first 2 hours of embryogenesis, is orchestrated by maternally deposited mRNAs and proteins (Zaessinger et al., 2006). During the maternal-to-zygotic transition around the 2-3 hours after fertilization, there is a shift from maternally-directed development to zygotically-directed development as maternal transcripts are degraded and zygotic transcription begins (Tadros and Lipshitz, 2009). Maternal transcripts are eliminated by two degradation pathways: the "maternal degradation pathway" is carried out by maternally encoded factors followed by the "zygotic degradation pathway" that uses zygotic factors to further clear maternal RNAs (Bashirullah et al., 1999). The RNA binding protein Smaug is a major regulator of RNA degradation in the maternal pathway (Tadros et al., 2007).

1.13 Mechanisms of posterior RNA localization

1.13.1 Active transport

osk mRNA undergoes active transport mediated by dynein-dependent movement on microtubules from the nurse cells into the oocyte (Clark et al., 2007). The mRNA cargo is coupled to dynein by Bicaudal D (BicD) and Egalitarian (Egl) (Bullock and Ish-Horowicz, 2001). Within the oocyte, RNP containing *osk* actively moves along weakly polarized microtubules using kinesin-1 in a random but posteriorly directed fashion (Parton et al., 2011; Zimyanin et al., 2008). As *osk* mRNA enriches in the posterior, its localization persists through early embryogenesis (**Fig. 1.2**).

1.13.2 Diffusion and entrapment

nos was the first RNA to be shown to undergo the diffusion and entrapment mechanism of localization to the posterior oocyte by live imaging (Forrest and Gavis, 2003). Other RNAs that also localize to the posterior oocyte in late oogenesis are *cyclin B*, *gcl* and *pgc* (Dalby and Glover, 1992; Jongens et al., 1992; Nakamura et al., 1996). Ovarian nurse cells synthesize *nos* RNA that is then transported into the oocyte upon concerted nurse cell contraction, a process referred to as nurse cell dumping, and localize to the posterior pole in late oogenesis (Forrest and Gavis, 2003). Prior to nurse cell dumping, microtubules assemble next to the oocyte cortex and rapid movements of the cytoplasm are generated in a kinesin-driver manner, called ooplasmic streaming (Glotzer et al., 1997; Lu et al., 2016). Although *nos* still localizes to the posterior when ooplasmic streaming is abrogated by colchicine, a drug that depolymerizes microtubules, posterior localization is more efficient due to bulk cytoplasmic movement (Forrest and Gavis, 2003). Once *nos* reaches the posterior of the oocyte, it associates with the germ plasm and anchors to the actin cytoskeleton (Forrest and Gavis, 2003). Transcripts localize as RNPs containing a single RNA, and once they become incorporated into the pole plasm, they act as seeds for the growth of clusters of the same RNAs, or homotypic clusters (Little et al., 2015). RNAs are incorporated into germ granules randomly, thus giving rise to heterogeneous composition of RNAs within granules (Niepielko et al., 2018).

Live imaging of *nos* RNA in the early embryo showed that germ granules containing those components undergo active transport in association with nuclear division (Lerit and Gavis, 2011). *nos*-containing germ granules are initially anchored to the cortex at the posterior end of the embryo. Germ granules are released from the actin cytoskeleton, a process triggered by centrosomes associated with nuclei in the posterior embryo (Lerit and Gavis, 2011). Germ particles are then transported in a dynein-dependent along microtubules to nearby nuclei where they segregate with the dividing nuclei (Lerit and Gavis, 2011). The enrichment of germ granules around astral microtubules appears as circular structures using lower magnification microscopy, documented by the Fly-FISH screen as “RNA islands” or “perinuclear localization” (**Fig. 1.2**) (Lecuyer et al., 2007; Lerit and Gavis, 2011). This process results in these nuclei eventually forming the pole cells in the posterior, thus ensuring that the associated germ granule components are partitioned into

these cells (Lerit and Gavis, 2011). The 3'UTRs of some germ plasm RNAs are sufficient to direct this type of posterior localization (Rangan et al., 2009).

1.13.3 Protection from degradation

Maternal RNAs that are uniformly distributed throughout the early embryo can become localized through the protection from degradation mechanism. *Hsp83* is one such transcript that undergoes this type of localization in a Smaug-dependent manner where Smaug recruits the CCR4-NOT deadenylation complex to initiate degradation (Semotok et al., 2005). The open reading frame of *Hsp83* contains an Hsp83 instability element (HIE) that includes 6 SREs that are recognized by Smaug and is able to direct degradation (Semotok et al., 2008). *Hsp83* mRNA in the bulk cytoplasm of the embryo is subject to Smaug-dependent degradation, but those found in the posterior pole within the germ plasm are protected, resulting in posterior localization **(Fig. 1.2**) (Ding et al., 1993).

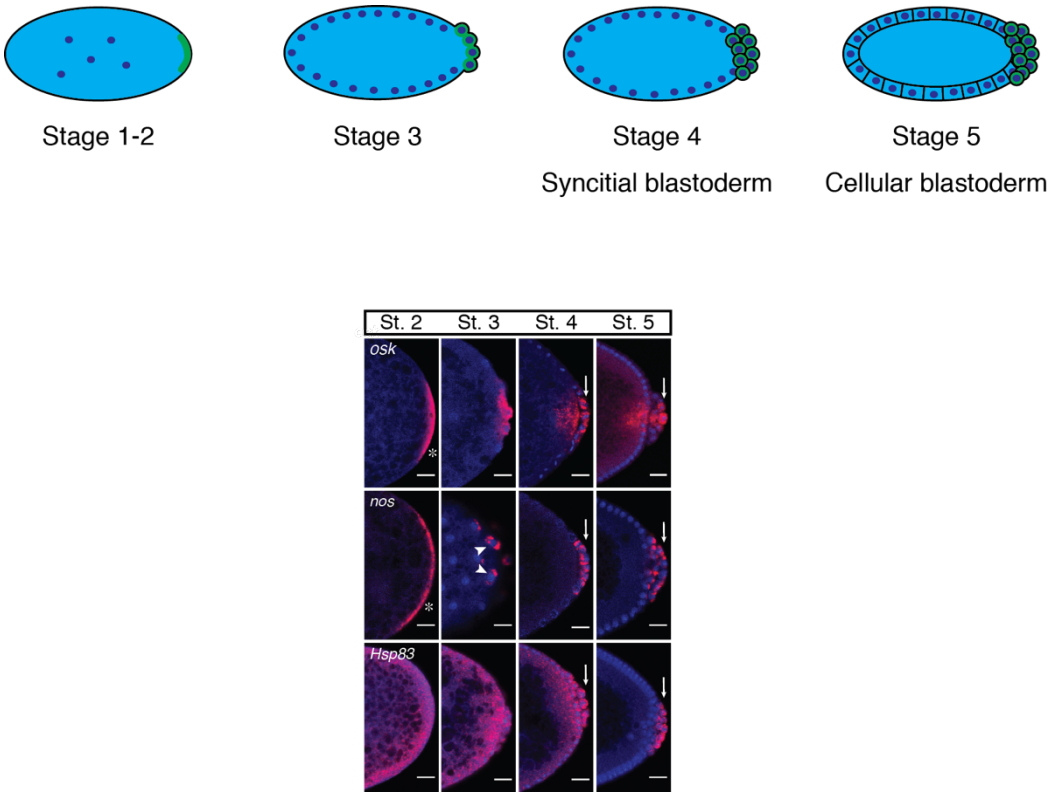


Figure 1.2 Patterns of posterior mRNA localization

Schematic of the first 5 stages of embryogenesis. Germ plasm shown in green. *osk* mRNA uses active transport to localize to the posterior of the oocyte and its localization in the posterior persists during early embryogenesis. *nos* uses the diffusion and entrapment mechanism to reach the posterior of the oocyte and it is transported along astral microtubules in the embryo to ensure proper segregation of germ granule components to pole cells. *Hsp83* mRNA is initially distributed throughout the embryo but it becomes enriched in the posterior due to protection from degradation.

1.14 Pole plasm components in *Drosophila*

1.14.1 Oskar

Oskar (*osk*) is the first mRNA to localize to the posterior pole of the oocyte, and upon translation, Osk protein has the critical role of organizing the pole plasm (Ephrussi et al., 1991). Translational repression of *osk* RNA before it reaches its destination in the posterior is essential as premature or ectopic *osk* translation results in developmental defects (Ephrussi and Lehmann, 1992; Smith et al., 1992). Anterior localization of *osk* can induce the formation of pole cells at the anterior embryo and the formation of a second posterior abdomen (Ephrussi and Lehmann, 1992).

Staufen, a double stranded RNA binding protein, is required for the posterior localization of *osk*, likely through binding to the *osk* 3'UTR (Micklem et al., 2000). Posterior localization of *osk* also requires a splicing event and the association with exon junction complex (EJC) core proteins Y14, Mago Nashi (Mago), eIF4AIII and Barentz (Ghosh et al., 2012). The active transport of *osk* RNA from the nurse cells to the oocyte is dynein-dependent, and then it switches to kinesin 1-dependent movement to move to the posterior oocyte. At the posterior, two isoforms of Osk protein are produced, Long Osk and Short Osk. Long Osk functions to anchor *osk* mRNA and Short Osk at the posterior during late oogenesis, and Short Osk directs pole plasm assembly by recruiting Vasa (Vanzo and Ephrussi, 2002). Anchoring at the posterior cortex occurs by Long Osk-induced actin remodeling coupled to endocytosis that is mediated by Mon2, a Golgi-endosomal protein (Tanaka et al., 2011).

Translation of *osk* is highly regulated by several mechanisms. Cup inhibits translation through its interaction with eIF4E and Bruno that binds to multiple BREs in the *osk* 3'UTR (Nakamura et al., 2004). Cup can also represses *osk* translation by recruiting the CCR4-NOT complex to promote deadenylation but Cup inhibits subsequent degradation with its N-terminal domain and protects against decapping with its non-canonical eIF4E-binding domain (4E-BM2) (Igreja and Izaurralde, 2011). Bruno can mediate translational repression in a Cup-independent manner via binding to BREs in the *osk* 3'UTR to mediate assembly of heavy RNPs to sequester *osk* from the translational machinery (Chekulaeva et al., 2006). Another protein involved the assembly of *osk* silencing complexes is polypyrimidine tract-binding protein (PTB) that binds to

multiple sites in the *osk* 3'UTR and mediates the formation of RNPs containing multiple *osk* transcripts (Besse et al., 2009). Hrp48 represses *osk* translation during localization by binding to the 5' and 3'UTRs of *osk* (Yano et al., 2004). Translational de-repression of *osk* occurs once it is localized to the posterior pole of the oocyte. Orb, the *Drosophila* homolog of *Xenopus* Cytoplasmic Polyadenylation Element Binding protein (CPEB) acts as an activator of *osk* translation by maintaining a long poly(A) tail (Castagnetti and Ephrussi, 2003).

1.14.2 Vasa

Vasa (Vas) is a highly conserved DEAD-box RNA helicase with important functions in germline development (Linder et al., 1989). Females homozygous for vas^{PH165}, the null allele, produce few eggs with defects in dorsal appendage formation (Dehghani and Lasko, 2015). Vas is a translational activator of *grk*, which encodes an epidermal growth factor receptor (EGFR) ligand that localizes at the dorso-anterior corner of the oocyte to activate follicle cells in the dorsal region to specify dorsal cell fates (Styhler et al., 1998; Tomancak et al., 1998). Vas does so through its interaction with eIF5B, a general translation factor involved in 60S ribosomal subunit joining (Johnstone and Lasko, 2004; Pestova et al., 2000). The Vas-eIF5B interaction is also required for activating the translation of *mei-P26* in the ovarian germline stem cells (Mader et al., 1995). Vas also regulates chromosome condensation in the germline (Pek and Kai, 2011). Vas has a role in producing Piwi-interacting RNA (piRNA), small RNAs that protect the germline genome against transposable elements (Pek et al., 2012). The C-terminal motif of Vas is important for its function in pole cell specification, posterior patterning, *grk* translation and transposon silencing, and the substitution mutation W660E in the C-terminal end abolishes its function in these processes (Dehghani and Lasko, 2015, 2016).

1.14.3 Tudor

tudor (*tud*) is essential for pole cell specification and granule formation in the embryo but otherwise dispensable for posterior patterning where pole cells are completely absent in embryos produced by *tud*-null females although posterior patterning is normal (Thomson and Lasko, 2004). In *tud*-null oocytes, Osk and Vas proteins, and *nos* mRNA localize normally to the posterior, but *pgc* and *gcl* RNAs were undetectable, demonstrating a role for Tud in pole cell specification. Polar

granules in embryos of *tud*-null mutants were severely reduced in number and size, explaining the absence of pole cells in those embryos.

Tudor (Tud) is a large protein of 285 kDa that contains 11 copies of the Tud domain (Ponting, 1997). The Tud domains can bind directly to symmetrically demethylated argnines (sDMAs) of Piwi proteins to regulate the piRNA pathway in the germline (Chen et al., 2011). Ovarian Tud complexes contain the Piwi protein Aubergine (Aub) and two glycolytic enzymes, pyruvate kinase (PyK) and glyceraldehyde-3-phosphate dehydrogenase 2 (GAPDH2). Glycolytic enzymes are recruited to germ granules where they bind Tud and protect the genome from transposable elements, thereby linking metabolism and transposon silencing in germ cell development.

1.14.4 Nanos

nanos (*nos*) has an important role in germ cell development but it is dispensable for body patterning (Gavis et al., 2008). *nos* localized to the germ plasm in the embryo leads to localized translation to create a posterior-to-anterior gradient of the Nos protein that inhibits *hunchback* (*hb*) mRNA translation for proper abdominal development to take place (Gavis and Lehmann, 1992, 1994). The ectopic expression of Nos protein in the anterior can inhibit the translation of *bicoid* (bcd) and *hb* and abrogate anterior development (Gavis and Lehmann, 1992, 1994).

nos localization to the posterior is considered inefficient with ~4% enriched in the posterior of the early embryo. Mechanisms that regulate its translation are essential for restricting the protein to the posterior (Bergsten and Gavis, 1999). *nos* uses the diffusion and entrapment mechanism to enrich in the posterior of the oocyte (Forrest and Gavis, 2003). There are multiple partially redundant localization elements within the 3'UTR of *nos* that direct posterior localization (Gavis et al., 1996). Rumpelstiltskin (Rump) binds the +2 region of the *nos* 3'UTR to regulate its localization (Jain and Gavis, 2008). Aubergine is also involved in the localization of *nos* and co-purifies with Rump (Becalska et al., 2011).

The translational control element (TCE) is a 90-nt region in the *nos* 3'UTR that regulates translational repression of unlocalized *nos* (Crucs et al., 2000; Forrest et al., 2004). Stem loop II

of the TCE contains a Smaug Recognition Element (SRE) which is recognized by Smaug to recruit the CCR4-NOT complex to lead to deadenylation (Semotok et al., 2005). The interaction between Smaug and Cup is another mechanism of translational repression done through SREs (Nelson et al., 2004). Osk can activate *nos* translation in the posterior by preventing the binding of Smaug to *nos*, leading to the stabilization and translation of *nos* (Zaessinger et al., 2006).

1.14.5 *polar granule component*

Drosophila polar granule component (*pgc*) is a maternally contributed mRNA that localizes to the germ plasm of the embryo with a critical role in germ cell specification (Nakamura et al., 1996). It was initially believed to be non-coding; however, it has since been proven to encode a small protein of 71 amino acids that is transiently expressed in blastoderm pole cells (Hanyu-Nakamura et al., 2008). Pgc mediates transcriptional quiescence in newly formed pole cells by preventing the recruitment of positive transcription elongation factor b (P-TEFb), the CTD Serine 2 kinase complex, to transcription start sites by sequestration, thereby repressing phosphorylation of Serine 2 on the carboxyl terminal domain (CTD) of RNA polymerase II, an important modification in transcription elongation (Hanyu-Nakamura et al., 2008). Most recently, Pgc has been shown to repress the zygotic transcription of multiple miRNA genes to protect germ plasm RNAs from miRNA-mediated degradation in pole cells, where misexpressed *miR-1* and *mir-10* target the *nos* 3'UTR (Hanyu-Nakamura et al., 2019).

1.14.6 *germ cell-less*

germ-cell less (*gcl*) is required for pole cells specification and formation (Jongens et al., 1992). Females with reduced *gcl* mRNA produced embryos that failed to form germ cells (Jongens et al., 1992). *gcl* translation in the oocyte outside of the pole plasm is repressed by Bruno, which binds to a sequence in the 3'UTR that is unrelated to BRE, where failure to do in the embryos from $aret^{QB72}$/+ females results in expression of Gcl in the soma and the repression of zygotic *hückebein* (*hkb*) (Moore et al., 2009).

Gcl also directs transcriptional quiescence in pole cell nuclei prior to pole cell formation in the embryo (Leatherman et al., 2002). Its function in silencing transcription is specific to a

subset of genes and does not occur globally. The localization of Gcl to the interior face of the nuclear membrane suggests that it represses transcription by anchoring chromatin to the nuclear periphery in a manner that is analogous to transcriptional silencing in yeast which is linked to the nuclear periphery (Jongens et al., 1994; Leatherman et al., 2002).

During pole cell budding in the embryo, Gcl has a role in regulating centrosome dynamics including astral microtubule organization and centrosome segregation to promote germ plasm segregation and cellularization (Lerit et al., 2017).

Gcl functions in promoting pole cell formation by inhibiting somatic signaling. Gcl is a substrate-specific adaptor of the E3 ubiquitin ligase complex $CRL3^{GCL}$ that targets the receptor tyrosine kinase (RTK) Torso, a determinant of somatic cell fate, for degradation (Pae et al., 2017).

1.15 Localization elements

Localization elements often reside in the 3'UTR of the mRNA, but they are also present in the 5'UTR and coding sequence, and they can be sequence-specific or dependent on secondary structure (Martin and Ephrussi, 2009). These *cis*-acting elements are recognized by *trans*-acting factors that function together to localize the transcript.

1.15.1 OES, SOLE and kissing-loop

The oocyte entry signal (OES) in *osk* 3'UTR is a 67-nt stem loop that mediates dynein-dependent *osk* transport from the nurse cells into the oocyte (Jambor et al., 2014). The brief localization of *osk* to the anterior of the oocyte is also driven by OES (Jambor et al., 2014). Posterior localization of *osk* to the posterior end of the oocyte in a kinesin-dependent manner requires the 28-nt stem loop called splicing *oskar* localization element (SOLE) in the 3'UTR with sequences on exons 1 and 2 (Ghosh et al., 2012). Splicing of the first intron is essential for posterior localization due to the assembly of the EJC as the *oskΔi(1)* mRNA, with a deletion of the first intron, is mislocalized (Ghosh et al., 2012). The *osk* 3'UTR contains another stem loop that promotes RNA-RNA interaction for transcripts to dimerize *in vitro* and hitchhike in the oocyte during mid-oogenesis (Jambor et al., 2011). This is accomplished by the palindromic sequence

present in the loop that allows for two loops to engage in a kissing-loop interaction (Jambor et al., 2011).

1.15.2 TLS

The transport of *K10* mRNA from the nurse cells to the oocyte's anterior cortex is mediated by a 44-nt stem loop in the 3'UTR called the Transport and Localization Sequence (TLS) (Serano and Cohen, 1995). A stem loop that is structurally similar to the TLS in the *orb* 3'UTR mediates anterior localization in the oocyte, and the localization machinery is likely shared between *K10* and *orb* (Cohen et al., 2005).

1.15.3 WLE3

The apical transport of *wingless* mRNA in the embryo relies on a stem-loop structure called the *wingless* localization element 3 (WLE3) that is necessary and sufficient for localization and is highly conserved among *Drosophila* species (dos Santos et al., 2008). The WLE3 loop sequence is not required for localization, but single-stranded sequences of secondary structures are commonly recognized by RNA binding proteins (Aviv et al., 2006; dos Santos et al., 2008). Instead, the stem sequences of WLE3 are required for localization (dos Santos et al., 2008).

1.15.4 GLS

Several localization elements found throughout the *grk* mRNA mediates its transport: the 5'UTR is sufficient for posterior localization in late oogenesis, the ORF is important for later localization, and the 3'UTR is essential for dorso-anterior localization (Saunders and Cohen, 1999; Thio et al., 2000). A *grk* localization signal (GLS) that is necessary and sufficient was identified in the coding sequence in the form of a 64-nt stem loop and can recapitulate full-length *grk* localization (Van De Bor et al., 2005). The mRNA of the *I* factor non-long-terminal-repeat (non-LTR) retrotransposon shares a similar stem loop with the GLS, called the *I* factor localization signal (ILS), and uses the same localization machinery as *grk* to colocalize in the oocyte (Van De Bor et al., 2005).

1.15.5 Homotypic clustering motif

The homotypic clustering motif mediates self-association of maternal RNAs once they are directed to the germ granules and results in enrichment in the posterior pole (Eagle et al., 2018). Although it was initially identified in the *pgc* 3'UTR because of a conserved stem loop found in the 3'UTRs of *pgc* orthologs in other *Drosophila* species, it was then determined to be a sequence specific type of localization element that is 6-nt long that is also present in the 3'UTRs of *gcl* and *nos* (Eagle et al., 2018).

1.16 Rationale and objectives

In keeping with the known function of 4E-BP as a translation inhibitor, and the upregulated translation of select mRNAs in 4E-BP2 knockout mouse brain, we sought to similarly identify mRNAs that are selectively translationally regulated by d4E-BP in the adult fly brain. We took advantage of recent advancements made in the field of translational control and applied the ribosome-profiling technique to identify mRNAs that are translationally upregulated in *d4E-BP*^{null} adult brains by using whole heads as a proxy for brains. To gain insight into the biological processes that are disrupted, we performed a gene ontology analysis. Considering 4E-BPs are known to regulate the translation of mRNAs with specific 5'UTR features, we characterized the 5'UTRs of the mRNAs that are differentially associated with ribosomes. Lastly, we assessed the levels of the proteins encoded by a few of these mRNAs to validate their preferential ribosome association. This work is presented in Chapter 2 and the Appendix.

Hundreds of RNAs were identified in recent years to localize to the pole cells of the embryo, but only a small subset of 55 transcripts localized as RNA islands. Due to their common destination, we hypothesized that they could share a localization element. We dissected the 3'UTRs of *pgc* and *gcl* in search of their posterior localization element. We also took a phylogenic approach to studying their localization considering that some localization elements have a conserved secondary structure in the other *Drosophila* species.

CHAPTER 2

Identification and characterization of mRNAs that are translationally regulated by d4E-BP in the *Drosophila* adult head

2.1 Introduction

Translation of a messenger RNA (mRNA) into its cognate protein is a fundamental step of gene expression. There are three phases of translation: initiation, elongation and termination, with initiation being most highly regulated and rate-limiting (Sonenberg and Hinnebusch, 2009). The mRNA is composed of the 5' untranslated region (UTR), a coding region and a 3'UTR. While the coding region contains the actual information that dictates protein synthesis, the untranslated regions harbor regulatory elements that regulate translation. All nuclear-transcribed mRNAs contain a m^7G- methylguanosine cap structure at the 5' end to which initiation factors bind and recruit the ribosome (Topisirovic et al., 2011). Although cap-dependent translation is the predominant form of translation in the cell, initiation and scanning of the ribosome can occur independently of the cap (Hinnebusch, 2011; Van Der Kelen et al., 2009).

The 5' cap structure is recognized by eukaryotic initiation factor 4E (eIF4E), the cap binding protein, followed by the recruitment of eIF4A, an RNA helicase, and eIF4G, a scaffolding protein that bridges the mRNA and the ribosome, to form the eIF4F complex (Gingras et al., 1999; Shahbazian et al., 2006). eIF4F unwinds secondary structures present in the 5'UTR to facilitate the binding of the 40S ribosomal subunit, together they will scan the mRNA until the complex reaches the initiation codon, AUG, then the 60S ribosomal subunit will bind to form the 80S complex that is competent to enter the elongation phase (Sonenberg, 2008). Since eIF4E is the least abundant initiation factor, eIF4F assembly is dependent on its availability, hence, it becomes a logical target for translational control (Raught and Gingras, 1999). One mechanism of regulating eIF4E is through its interaction with a family of low molecular weight phosphoproteins called eIF4E-binding proteins (4E-BPs) (Pause et al., 1994). 4E-BP competes with eIF4G for binding to eIF4E, thereby sequestering it from cap-binding complex assembly (Marcotrigiano et al., 1999).

The mechanistic target of rapamycin (mTOR) signaling pathway is essential for the regulation of 4E-BP. mTOR is a highly conserved kinase that serves as a master regulator of protein synthesis by integrating and responding to environmental and intracellular cues (Hay and Sonenberg, 2004). It exists in two distinct complexes: mTORC1, which regulates cellular growth and proliferation, and mTORC2, which regulates cell survival and actin reorganization (Dowling et al., 2010). mTORC1 phosphorylates 4E-BP and 40S ribosomal protein S6 kinases (S6Ks), two

of the best characterized direct downstream targets. Components of the TOR signaling pathway are conserved in *Drosophila*, where d4E-BP, also known as Thor, and dS6K are also downstream targets that get phosphorylated by dTOR (Miron et al., 2003). Upstream of dTOR, dAkt phosphorylates dTsc2 to prevent the dTsc1-dTsc2 interaction to form the tuberous sclerosis complex (TSC) (Potter et al., 2002), which acts to antagonize dTOR signaling (Gao et al., 2002).

The mTORC1-eIF4E pathway regulates a specific pool of mRNAs, although eIF4E is required for the translation of the majority of mRNAs (Gingras et al., 1999). mRNAs that contain extensive secondary structures in the 5'UTR are translated more efficiently in the presence of elevated levels of eIF4E, which is typically the least abundant eIF4F component (Duncan et al., 1987; Koromilas et al., 1992). Several studies have shown that 5' terminal oligopyrimidine (TOP) containing mRNAs and mRNAs containing a pyrimidine rich terminal element (PRTE) are more sensitive to eIF4F complex activity that can also be regulated by 4E-BPs (Hsieh et al., 2012; Thoreen et al., 2012). In mice where 4E-BP2, the main isoform of 4E-BP expressed in the brain, was inactivated, neuroligins 1-4, post-synaptic adhesion molecules, were translationally upregulated in the brain while other candidate genes tested were unaffected. This is potentially due to a repeated structural element present in the 5'UTRs of neuroligin mRNAs that is absent from other mRNA 5'UTRs (Gkogkas et al., 2013).

d4E-BP has been reported to regulate the translation of specific mRNAs in response to physiological changes such as defense response due to bacterial infection and dietary restriction. d4E-BP contributes to the *Drosophila* defense response whereby its expression is induced upon bacterial infection to inhibit cap-dependent translation and bias cap-independent translation of antimicrobial peptides (AMP) via the internal ribosomal entry site (IRES) in the 5'UTR of their transcripts (Vasudevan et al., 2017). *Thor* mutant flies are thus immune compromised; they are more susceptible to infection with *Staphylococcus epidermidi*s, a Gram-positive bacterium (Bernal and Kimbrell, 2000) and with the fungus *Candida albicans* (Levitin et al., 2007) compared to control flies. In response to Gram-positive bacteria and fungi, a battery of antimicrobial peptides (AMPs) are rapidly produced via the Toll pathway, while infection by Gram-negative bacteria induce the expression of different AMPs via the immune deficiency (Imd) pathway (Brennan and Anderson, 2004). Moreover, d4E-BP mRNA and protein levels increased upon infection of S2 cells by *C. albicans* (Levitin et al., 2007). Under dietary restriction, d4E-BP can extend lifespan

in flies and lead to preferential translation of mRNAs with shorter and less structured 5'UTRs that encode mitochondrial genes to enhance mitochondrial function (Zid et al., 2009).

In the present study, we provide an unbiased genome-wide characterization of mRNAs that are differentially translated within heads of *Thor* mutant flies at steady state using ribosome profiling. This analysis revealed that in *Thor* mutants, mRNAs involved in innate immunity and those with shorter but more complex 5'UTRs were preferentially associated with ribosomes. We also demonstrate that dS6K mRNA is a potential downstream target of d4E-BP-regulated translation. Together, our findings suggest that there are subsets of differentially ribosome associated mRNAs due to the d4E-BP^{null} mutation with differences in their 5'UTRs.

2.2 Results

2.2.1 Genome-wide molecular profiling of translation in *Thor* flies

To gain a comprehensive view of the role of d4E-BP in regulating mRNA translation in the brain, we assessed the translatome by using the polysome and ribosome profiling techniques. The null and revertant alleles of *d4E-BP*, hereby referred to as *Thor* and *Revertant*, were previously generated by imprecise and precise excision of a P-element within the locus (Tettweiler et al., 2005). The whole adult fly head was used as a proxy for the brain because the brain is the largest structure in the head relative to the other tissues including the eyes, trachea, antennae and proboscis (Murthy and Turner, 2013; Purice et al., 2016; Wu and Luo, 2006). Polysome profiling on head lysate from *Thor* and *Revertant* flies, a technique used to resolve mRNA on a sucrose density gradient based on the number of bound ribosomes, did not detect observable differences in translation (**Fig. 2.1A**). A puromycin incorporation assay on head lysate of flies that were fed puromycin also did not detect significant differences in *de novo* protein synthesis (**Fig. 2.1B**). Together, these results suggest that global translation is unaffected by the *d4E-BP* null mutation, consistent with other studies that demonstrate the translation of specific mRNAs is sensitive to levels of 4E-BP (Gkogkas et al., 2013; Tahmasebi et al., 2016).

To identify 4E-BP sensitive mRNAs, we carried out ribosome profiling on whole head lysate from *Thor* and *Revertant* flies (**Fig. 2.2A**). Ribosome profiling is a high-throughput technique that couples ribosome nuclease-protection assay (footprints: ~30 nucleotide (nt) regions of mRNAs protected by the ribosome) with deep sequencing (RNA-Seq) (Ingolia et al., 2009). Sequencing libraries were generated from ribosome-protected mRNA fragments to assess translation and from randomly fragmented mRNA to evaluate transcription from which the translation efficiency (TE) of a given mRNA can be measured as the ratio of ribosome footprints to mRNA fragments (**Fig. 2.2A**). Conditions to generate ribosome footprints in *Drosophila* head samples were optimized by micrococcal nuclease (MNase) digestion of polysomes where increasing concentrations of MNase resulted in an increase of the monosome peak along with a decrease in polysome peaks in the polysome profile, in agreement with a previous study that adapted the ribosome profiling technique for *Drosophila* embryos (**Fig. 2.3**) (Dunn et al., 2013).

The distribution of footprint sizes (20–30 nt) falls below the expected ~30 nt size, likely due to degradation of the RNAs during sequencing library preparation (**Fig. 2.2B**). Nonetheless, the majority of footprint reads is found in the first reading frame, confirming that ribosome footprints were indeed generated in the ribosome profiling experiment (**Fig. 2.2C**).

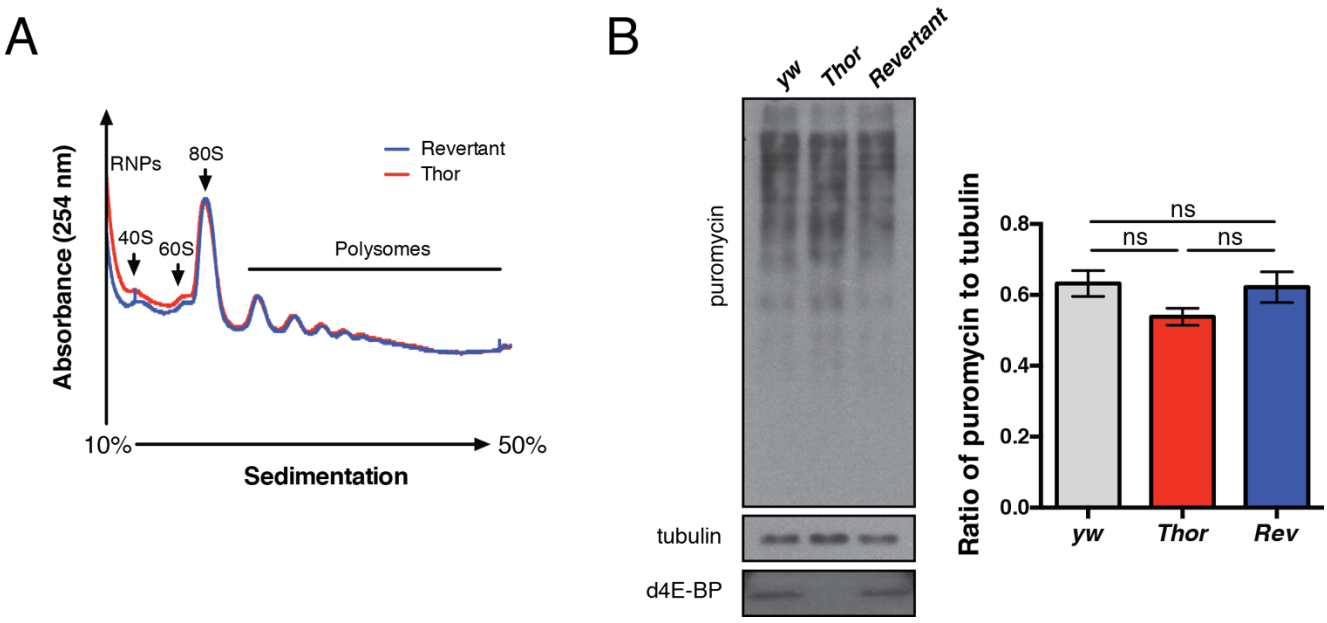


Figure 2.1 Translational control in *Revertant* and *Thor* fly heads

(A) Absorbance at 254 nm was continuously measured for polysome profiles of *Revertant* (blue) and *Thor* (red) fly head lysates resolved on a 10–50% sucrose gradient. **(B)** Immunoblot analysis of head lysate from flies fed with puromycin for 6 hours to assess levels of puromycin incorporation. *Yellow white* (*yw*) fly head sample was used as the wild type control. Quantification of puromycin-labelled proteins is shown as a ratio to the tubulin loading control. Data are presented as mean ± SEM (error bars). Significance of differences in ratios was determined by one-way ANOVA with Tukey's post-hoc. n = 3 biological replicates of 30 fly heads. ns, non-significant.

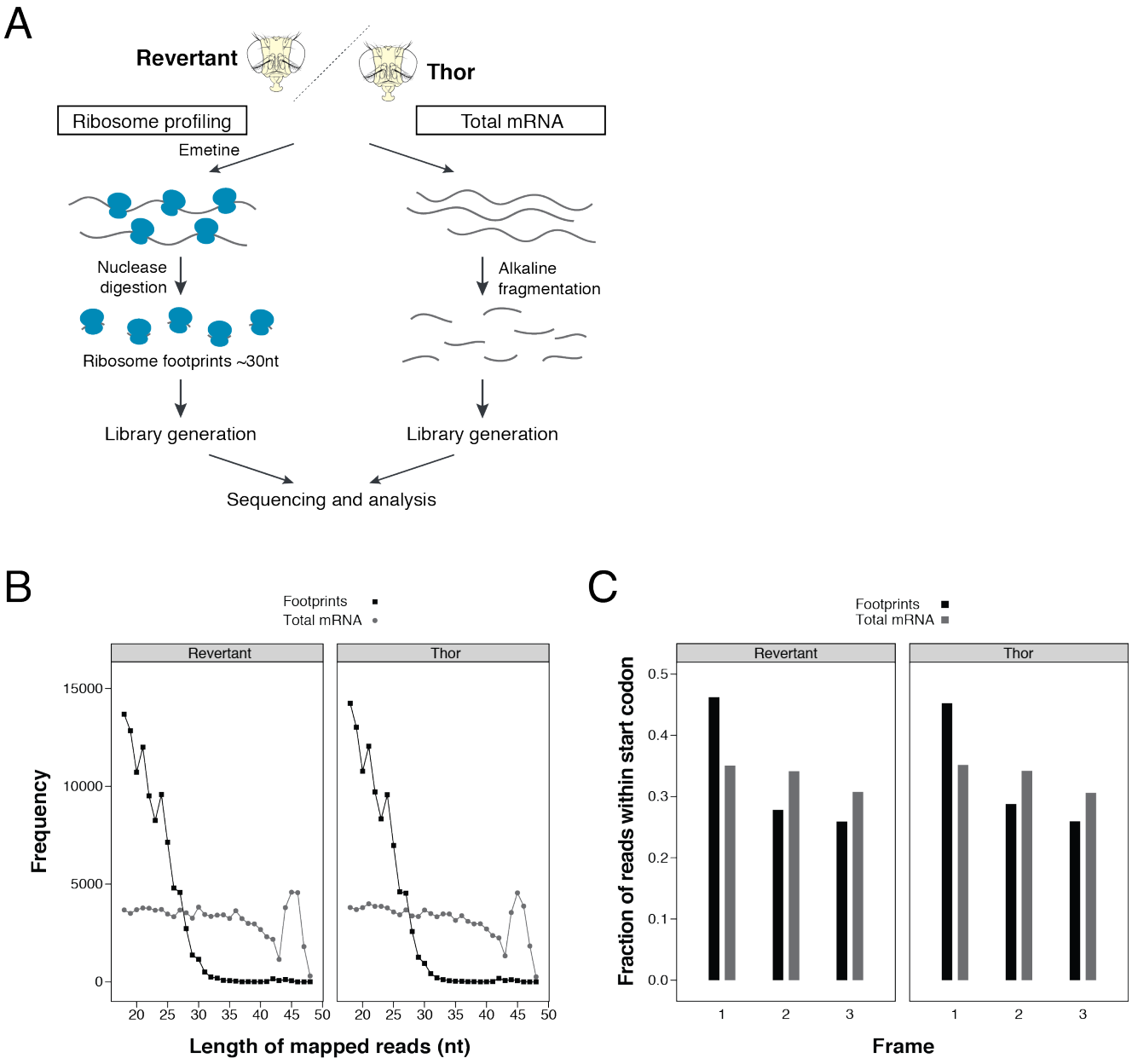


Figure 2.2 Genome-wide molecular profiling of translation in *Thor* flies

(A) Schematic illustrating ribosome profiling experiment using *Revertant* and *Thor* fly heads. **(B)** Length distribution of mapped reads of ribosome footprints (black square) and randomly fragmented mRNA (grey circle) in *Revertant* and *Thor* fly heads. **(C)** Fraction of reads of ribosome footprints (black) and randomly fragmented mRNA (grey) mapped to each 3 reading frames.

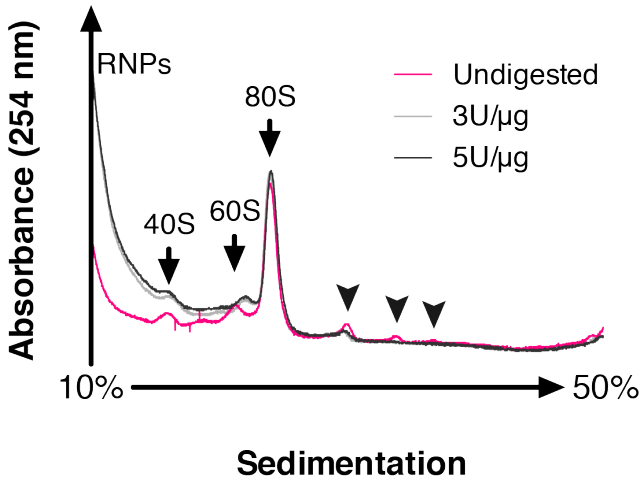


Figure 2.3 Micrococcal nuclease titration of *Revertant* fly head polysomes

Revertant fly head lysates were treated with increasing concentrations of micrococcal nuclease (MNase) to digest polysomes (0, 3 and 5 U/µg). Lysates were then resolved on a 10–50% sucrose gradient and absorbance at 254 nm was continuously measured. Arrowheads point to the polysome peaks that reduce in size with increasing concentrations of MNase.

2.2.2 Ribosome profiling reveals preferential translation of a subset of mRNAs

A total of 5413 protein-coding RNAs with a mean read cut-off of 40 from the ribosome profiling experiments were selected for further analyses, of which 5215 mRNAs had an overlap of 56% with protein coding RNAs extracted from 4 day old male fly heads in the modENCODE dataset, confirming that our technique did indeed identify RNAs expressed in 3-5 day old male fly heads (**Fig. 2.4A**) (Celniker et al., 2009). The reproducibility between replicates of ribosome footprint and mRNA libraries was high as evidenced by their correlation coefficients (r of reads per kilobase per million (RPKM) of 0.977 for footprints and 0.981 for mRNAs in *Revertant* and 0.982 for footprints and 0.983 for mRNAs in *Thor*) (**Fig. 2.4B**). Next, we sought to identify mRNAs with significant changes in translation by using the $\log_2$(TE fold change) as an index of such change. While a comparison of *Revertant* and *Thor* footprints and mRNAs also showed a strong correlation (r of RPKM of 0.988 and 0.989 respectively), the TE of *Revertant* to *Thor* had a lower correlation of r = 0.959 (**Fig. 2.5A**). The distributions of $\log_2$(TE) of *Revertant* and *Thor* showed little observable differences, further supporting that global mRNA translation is unaffected (**Fig. 2.5B**). Applying a criterion of $\log_2$(TE fold change) ± 0.697, the equivalent to a fold change of 1.5, and P-value ≤ 0.05 to identify mRNAs with significant changes in translation, we identified 60 mRNAs with upregulated translation and 128 mRNAs with downregulated translation as a result of the *d4E-BP* null mutation, hereby referred to as upregulated and downregulated mRNAs, respectively (**Fig. 2.5B**, **Table 2.1 and 2.2**). As a consequence of d4E-BP's role in inhibiting translation, it was expected that a subset of transcripts would have an increase in translation in the absence of d4E-BP. Transcripts with downregulated translation are likely due to an indirect effect due to the loss of d4E-BP activity.

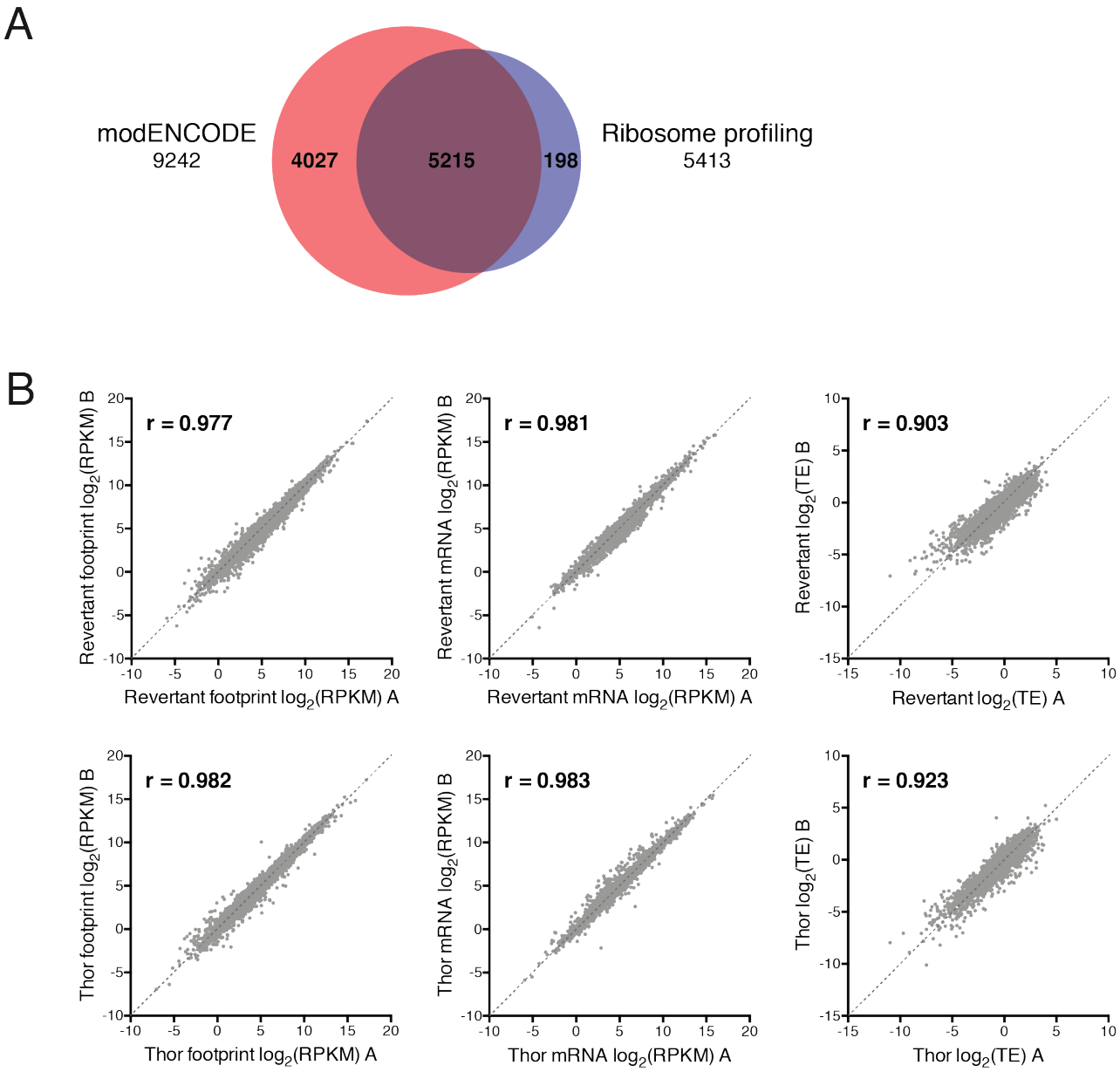


Figure 2.4 Libraries generated from fly head RNAs are reproducible

(A) Venn diagram of protein coding RNAs by modENCODE and ribosome profiling. A total of 9242 protein coding RNAs were identified in 4 day old male fly heads by modENCODE with an RPKM > 1. From the total of 5413 protein coding RNAs with a mean read count of 40 in ribosome profiling experiments, 5215 RNAs overlapped with the modENCODE dataset. **(B)** Reproducibility plots of ribosome footprints, mRNA fragments and TE from two independent ribosome profiling experiments, A and B. r indicates the Spearman's correlation coefficient of $\log_2$ comparison.

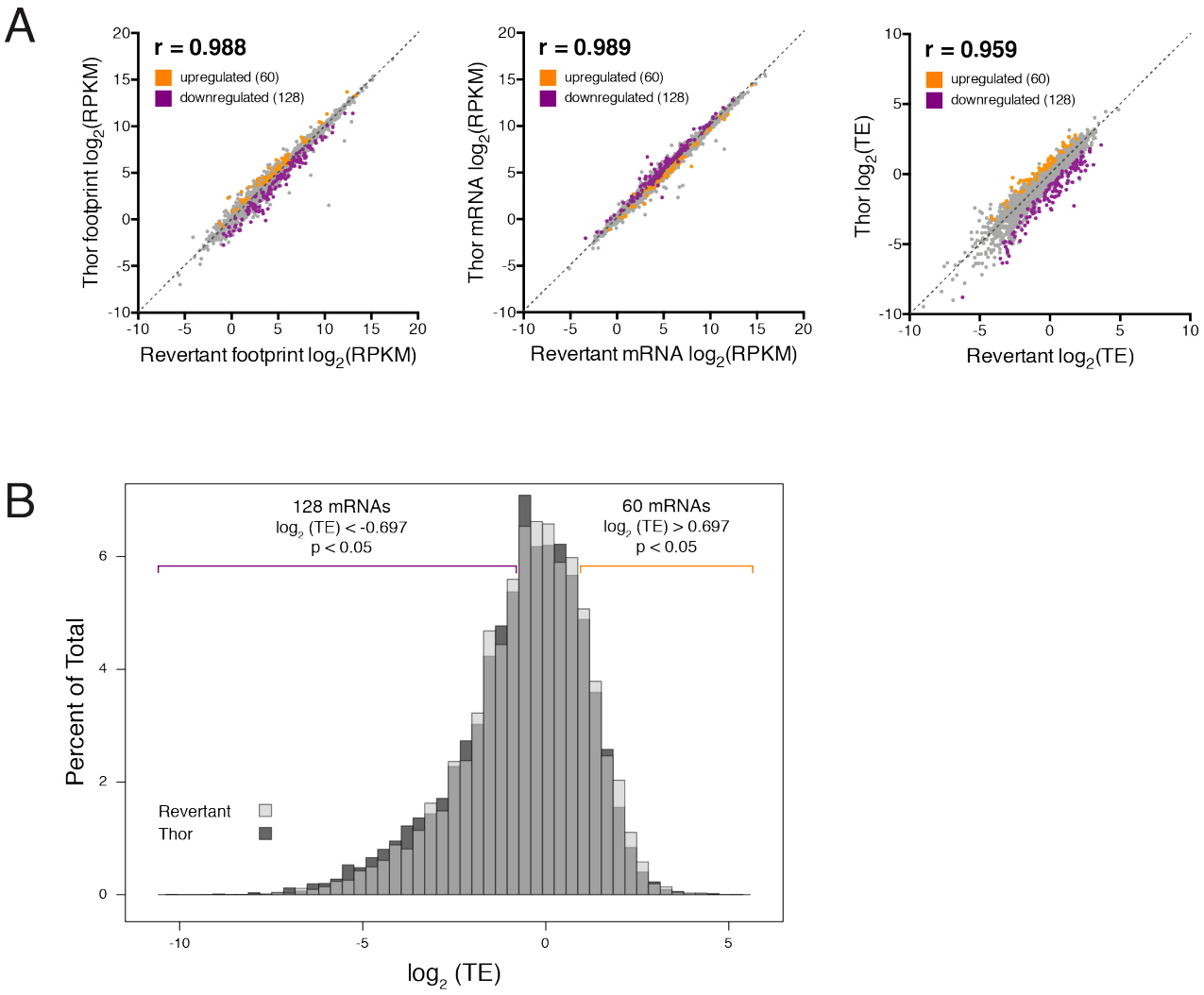


Figure 2.5 Ribosome profiling reveals preferential translation of a subset of mRNAs

(A) Correlation plots of ribosome footprints, randomly fragmented mRNA and translational efficiency (TE) of protein-coding mRNAs with a least 40 mean reads from *Revertant* and *Thor* head samples. 60 translationally upregulated (orange) and 128 translationally downregulated (purple) mRNAs were identified in the ribosome profiling experiment. **(B)** Distribution of $\log_2$ translational efficiencies (TE) of protein-coding RNAs in *Revertant* (light grey) and *Thor* (dark grey) samples. r indicates the Spearman's correlation coefficient of $\log_2$ comparison.

Table 2.1 List of translationally upregulated genes identified by ribosome profiling

Ensembl Gene ID	Gene Symbol	log_2 TE Fold Change	Significance (P-value)
FBgn0259145	CG42260	2.59002266	2.29E-06
FBgn0035227	CG12090	1.83337758	4.23E-06
FBgn0261274	Ero1L	1.6997342	0.00746534
FBgn0052485	CG32485	1.69927788	0.00318425
FBgn0259716	CG42370	1.46167956	0.01307613
FBgn0040736	IM3	1.44749015	0.00321241
FBgn0041579	AttC	1.4467124	0.00021331
FBgn0034329	IM1	1.39288677	0.00107619
FBgn0040733	CG15068	1.35200032	0.00066344
FBgn0041581	AttB	1.32981308	0.00603191
FBgn0052191	CG32191	1.30257615	0.00775765
FBgn0034407	DptB	1.29084697	0.00566794
FBgn0037900	CG5276	1.28087092	0.0282789
FBgn0053512	dpr4	1.27733368	0.01182047
FBgn0030096	Zpr1	1.27550308	0.01179115
FBgn0013949	Ela	1.25880272	0.0053733
FBgn0031220	CG4822	1.22079255	0.01451847
FBgn0035806	PGRP-SD	1.20825177	0.00350807
FBgn0047135	CG32276	1.18798859	0.00354597
FBgn0032139	CG13116	1.18203851	0.01524506
FBgn0039099	GILT2	1.17001023	0.00458132
FBgn0016031	lama	1.13425839	0.03652645
FBgn0032924	Nbr	1.13390693	0.02654862
FBgn0027091	Aats-cys	1.1271119	0.02102364
FBgn0051344	CG31344	1.11911221	0.02545644
FBgn0037690	Task7	1.11428561	0.0150501
FBgn0015806	S6k	1.09064495	0.00403227
FBgn0262881	CG43236	1.06846198	0.01900248
FBgn0039690	CG1969	1.04728684	0.02898441
FBgn0039801	Npc2h	1.03505006	0.00123074
FBgn0003965	v	1.01022821	0.00336215
FBgn0039031	CG17244	1.00198251	0.03768852
FBgn0030504	CG2691	0.99185344	0.01870814
FBgn0052177	Ndfip	0.97391686	0.01378291
FBgn0261986	RASSF8	0.92519592	0.03846927
FBgn0051217	modSP	0.92181982	0.04706485
FBgn0266346	CngB	0.90853156	0.04075691
FBgn0036740	Vps60	0.89651989	0.02457442
FBgn0023178	Pdf	0.88505502	0.00381197
FBgn0011227	ox	0.87023285	0.03425808
FBgn0032518	RpL24	0.85091452	0.02477405
FBgn0003507	srp	0.84695231	0.03697099
FBgn0086367	t	0.83083639	0.0179044
FBgn0013325	RpL11	0.81916185	0.02009664
FBgn0031801	CG9498	0.80513142	0.01567926
FBgn0030838	CG5445	0.80191933	0.04329283
FBgn0015568	alpha-Est1	0.79745911	0.03242731
FBgn0038363	Acyp2	0.79431507	0.04324685
FBgn0033055	Tbce	0.79063096	0.04797066
FBgn0037874	Tctp	0.78139568	0.01276314
FBgn0024841	Pcd	0.7726446	0.04857676
FBgn0034997	CG3376	0.76602915	0.04006235
FBgn0031800	CG9497	0.75882486	0.03032572
FBgn0032836	CG10680	0.73553104	0.0085345
FBgn0003744	trc	0.72138292	0.04530375
FBgn0034215	Mtap	0.71574537	0.03889303
FBgn0001257	ImpL2	0.71300571	0.02693186
FBgn0031417	CG3597	0.70387883	0.04493478
FBgn0016013	Faa	0.69812394	0.02482385
FBgn0040074	retinin	0.69762972	0.02948397

Table 2.2 List of translationally downregulated genes identified by ribosome profiling

Ensembl Gene ID	Gene Symbol	log_2 TE Fold Change	Significance (P-value)
FBgn0038290	CG6912	-3.9299836	4.25E-12
FBgn0039190	CG5762	-3.1650799	0.00836301
FBgn0039511	CG3330	-2.9743281	0.03271262
FBgn0011669	Mst57Db	-2.6610467	2.60E-07
FBgn0011668	Mst57Da	-2.5650646	0.0426027
FBgn0004414	msopa	-2.5285609	0.00018137
FBgn0035189	CG9119	-2.487528	6.98E-10
FBgn0051029	CG31029	-2.4674693	0.03454082
FBgn0015584	Acp53Ea	-2.4672006	4.71E-06
FBgn0051709	CG31709	-2.433723	0.00703042
FBgn0040097	lectin-30A	-2.3950209	2.54E-05
FBgn0259971	CG42481	-2.3320187	2.80E-07
FBgn0032055	CG13091	-2.3264536	2.81E-06
FBgn0259973	Sfp79B	-2.2981782	0.00049968
FBgn0020509	Acp62F	-2.289326	0.00852801
FBgn0035598	CG4669	-2.2842575	0.02808024
FBgn0041102	ocn	-2.2644626	0.04207948
FBgn0264329	CG43788	-2.2396795	4.67E-07
FBgn0036973	Rbbp5	-2.1925572	0.04721885
FBgn0011670	Mst57Dc	-2.1704047	2.23E-05
FBgn0051872	CG31872	-2.1460612	0.00061243
FBgn0033286	CG2127	-2.1409159	0.01992842
FBgn0038706	CG3517	-2.1403616	0.01963089
FBgn0002863	Acp95EF	-2.1389897	7.21E-06
FBgn0263249	CG43392	-2.1156544	5.50E-06
FBgn0050376	CG30376	-2.0436577	0.04954213
FBgn0262099	CG42852	-1.9967257	0.00042251
FBgn0037763	CG16904	-1.9930745	2.63E-05
FBgn0047334	BG642312	-1.9846054	1.97E-08
FBgn0038373	CG4546	-1.960221	9.64E-05
FBgn0032360	CG14926	-1.9389796	0.00744538
FBgn0040001	CG17374	-1.9339067	0.00011825
FBgn0033774	CG12374	-1.9299854	0.00555294
FBgn0250844	CG4218	-1.8869876	0.03139268
FBgn0040747	CG12853	-1.8807617	0.04417685
FBgn0030999	Mur18B	-1.8411688	0.00042263
FBgn0034152	Acp53C14a	-1.8064924	0.00017162
FBgn0034295	CG10911	-1.8060449	0.00027048
FBgn0039321	CG10550	-1.800127	2.98E-10
FBgn0034662	CG13492	-1.7952994	0.00384231
FBgn0036415	CG7768	-1.7857284	0.00024895
FBgn0085249	CG34220	-1.7787091	4.66E-05
FBgn0036369	CG10089	-1.7654411	0.02351605
FBgn0004181	Peb	-1.7648273	0.00517388
FBgn0028987	Spn28F	-1.7291107	0.00068741
FBgn0035665	Jon65Aiii	-1.6985903	0.00695528
FBgn0033285	CG18449	-1.6899458	0.03172616
FBgn0031728	Hsp60C	-1.6892597	0.00031152
FBgn0265264	CG17097	-1.6696538	6.14E-06
FBgn0010425	epsilonTry	-1.6695403	0.00607756
FBgn0043825	CG18284	-1.6502056	0.00097813
FBgn0039755	CG15531	-1.6312454	0.00164951
FBgn0263601	mib1	-1.6275802	1.92E-05
FBgn0052081	CG32081	-1.6055791	0.02432591
FBgn0085353	CG34324	-1.5833863	0.00779449
FBgn0003356	Jon99Cii	-1.5679318	0.02454268
FBgn0011559	Acp36DE	-1.5452453	0.02731087
FBgn0267366	mil	-1.5433503	0.0014418
FBgn0010019	Cyp4g1	-1.5397362	1.15E-06
FBgn0262477	FoxP	-1.5342408	0.04950803
FBgn0004583	ex	-1.4953101	0.03960379
FBgn0259975	Sfp87B	-1.4854576	0.00945761
FBgn0069354	Porin2	-1.4850863	0.00705123

(Table continues)

Ensembl Gene ID	Gene Symbol	log_2 TE Fold Change	Significance (P-value)
FBgn0036887	CG9231	-1.4526172	0.00010415
FBgn0000047	Act88F	-1.4463416	0.01022611
FBgn0050395	CG30395	-1.4404761	0.00791566
FBgn0036969	Spn77Bb	-1.4319764	0.0064577
FBgn0051788	CG31788	-1.430962	0.04711474
FBgn0032069	CG9468	-1.4110163	0.00198322
FBgn0043530	Obp51a	-1.4098318	0.00105315
FBgn0259968	Sfp60F	-1.397235	0.01053075
FBgn0000615	exu	-1.39377	0.00059482
FBgn0037547	CG7910	-1.390906	0.00187042
FBgn0046294	CG12699	-1.3752239	0.00017199
FBgn0032669	CG15155	-1.3728548	0.01229616
FBgn0040212	Dhap-at	-1.3707666	0.00226783
FBgn0265180	CG44245	-1.3633347	0.03656046
FBgn0031176	CG1678	-1.3600729	0.03363805
FBgn0034435	CG9975	-1.3436753	0.03875985
FBgn0035915	S-Lap1	-1.3426621	0.01321627
FBgn0033027	TpnC4	-1.3239541	0.01169568
FBgn0033868	S-Lap7	-1.3181753	0.01787162
FBgn0040687	CG14645	-1.3018396	0.01046639
FBgn0051660	pog	-1.2983568	0.00010797
FBgn0038762	CG4836	-1.2826094	0.02364599
FBgn0020908	Scp1	-1.2594965	0.00611474
FBgn0052677	X11Lbeta	-1.2520047	0.00064443
FBgn0261574	kug	-1.2472871	0.01134364
FBgn0261575	tobi	-1.2272101	0.04402629
FBgn0042627	v(2)k05816	-1.1920968	0.00220353
FBgn0034144	CG5089	-1.1873485	0.04243201
FBgn0086348	se	-1.1759841	0.04561384
FBgn0002855	Acp26Aa	-1.1659147	0.00999272
FBgn0259795	loopin-1	-1.1548299	0.0294513
FBgn0083938	BG642163	-1.1415962	0.03687671
FBgn0264386	Ca-alpha1T	-1.1358188	0.01363361
FBgn0034474	Obp56g	-1.116214	0.03147322
FBgn0259952	Sfp24Bb	-1.1089616	0.04481368
FBgn0038395	CG10407	-1.1064974	0.02788996
FBgn0038200	CG9920	-1.0802117	0.04256986
FBgn0051198	CG31198	-1.0714151	0.03483637
FBgn0032481	CG16972	-1.0711976	0.03628793
FBgn0028990	Spn27A	-1.0686789	0.00465544
FBgn0000079	Amy-p	-1.0670328	0.03220947
FBgn0031746	CG9029	-1.0254752	0.01250399
FBgn0031145	Ntf-2	-1.0170174	0.01802942
FBgn0025678	CaBP1	-0.9585471	0.03953823
FBgn0023510	Rbcn-3B	-0.9553233	0.00413005
FBgn0250908	beat-VII	-0.9150052	0.03763436
FBgn0026666	l(1)G0136	-0.9120594	0.04367199
FBgn0003169	put	-0.9103395	0.03626068
FBgn0053519	Unc-89	-0.9057717	0.03145163
FBgn0039348	Npl4	-0.8999376	0.00486321
FBgn0003415	skd	-0.885163	0.04350604
FBgn0086899	tlk	-0.8827039	0.01476596
FBgn0035670	CG10472	-0.8771324	0.0444488
FBgn0036862	Gbs-76A	-0.8648149	0.03587986
FBgn0036059	nudE	-0.8437785	0.04764563
FBgn0036289	CG10657	-0.8311376	0.04927447
FBgn0261836	Msp300	-0.829341	0.0173388
FBgn0027339	jim	-0.8047053	0.04016503
FBgn0034497	CG9090	-0.7910544	0.02105792
FBgn0086906	sls	-0.7795391	0.03644236
FBgn0260003	Dys	-0.7350932	0.00651792
FBgn0002570	Mal-A1	-0.7231756	0.03939568
FBgn0043362	bchs	-0.7018823	0.02461792
FBgn0016977	spen	-0.6975415	0.03886093
FBgn0005666	bt	-0.6967558	0.03456957

2.2.3 mRNAs that are preferentially associated with ribosomes in the *Thor* mutant are linked to innate immunity

To gain further insight into the biological significance of the *d4E-BP* null mutation, we performed a gene ontology (GO) analysis for the 60 translationally upregulated and 128 translationally downregulated mRNAs and identified significantly enriched terms for biological processes, functional categories, cellular components and molecular functions ($P < 0.05$) (**Fig. 2.6**) (Dennis et al., 2003). Interestingly, terms related to the immune system were identified in biological processes and functional categories within the upregulated gene group (**Fig. 2.6A, 2.6B**). *Thor* flies were previously reported to be immune compromised upon infection (Bernal and Kimbrell, 2000; Levitin et al., 2007; Vasudevan et al., 2017). Conversely, the biological processes, functional categories and molecular functions of the downregulated gene group include multicellular organism reproduction, behavior, signal and protease activity (**Fig. 2.6A, 2.6B, 2.6D**). Genes that are downregulated produce proteins that localize in the extracellular space pr muscle (**Fig. 2.6C**). Surprisingly, three genes involved in defense response are downregulated in *Thor* fly heads (**Fig. 2.6B**). Together, these results suggest there is a potential link to innate immunity in the subset of upregulated mRNAs in the *Thor* fly heads.

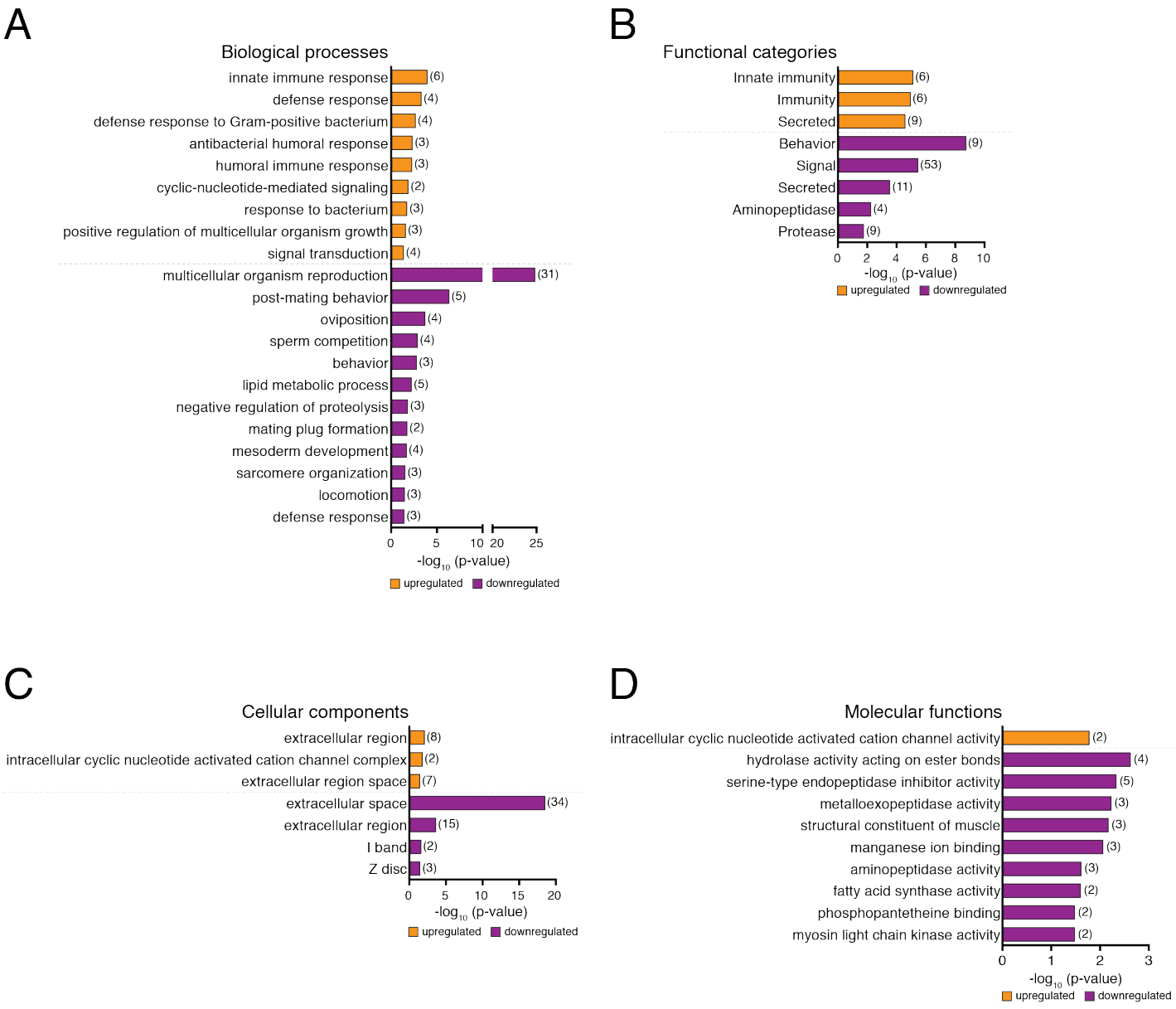


Figure 2.6 Functional analysis of upregulated and downregulated genes

Gene ontology analysis of 60 upregulated (orange) and 128 downregulated (purple) genes showing plots for biological processes (**A**), functional categories (**B**), cellular components (**C**) and molecular functions (**D**). The number of genes in each category are shown in parentheses and significance is indicated as $-\log_{10}$ (P-value). Only terms with $P < 0.05$ are presented.

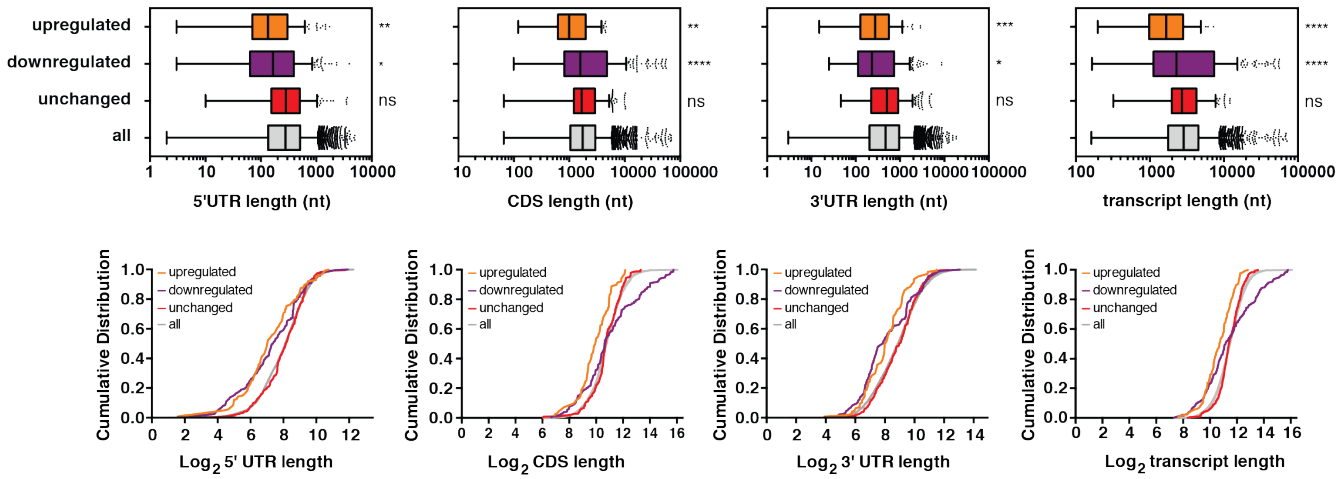
2.2.4 Upregulated and downregulated mRNAs possess UTRs with different features

We analyzed the differentially translated mRNAs to identify distinguishing features within their sequences. Two types of mRNAs with distinct 5'UTR characteristics have been identified to be translationally regulated in a TOR-4E-BP-dependent manner: (1) mRNAs containing long and structured 5'UTRs (Hinnebusch et al., 2016; Masvidal et al., 2017); and (2) mRNAs with shorter and lower complexity 5'UTRs containing TOP and/or PRTE (Hsieh et al., 2012; Thoreen et al., 2012). First, we assessed the length and guanine-cytosine content (% GC) of the upregulated and downregulated mRNAs, along with two control groups, unchanged and all mRNAs. 162 unchanged mRNAs, whose translation did not change between *Revertant* and *Thor*, were selected as those with $\log_2$(TE fold change) ± 0.01. The sequences of all known mRNA isoforms of each gene were considered: upregulated (127), downregulated (289), unchanged (610), and all (17177). This analysis revealed differences in upregulated and downregulated mRNAs compared to controls. Upregulated mRNAs were generally shorter compared to the controls in 5'UTR, CDS, 3'UTR and whole transcript length (**Fig. 2.7A, Table 2.3**). Downregulated mRNAs were longer in CDS and whole transcript length, but shorter in 5' and 3'UTRs compared to the controls (**Fig. 2.7A, Table 2.3**). This suggests that upregulated mRNAs encode smaller proteins, while downregulated mRNAs encode larger proteins. The 5'UTR length was not different between upregulated and downregulated mRNAs; however, they were significantly shorter compared to the controls (**Fig. 2.7A, Table 2.3**). Notably, 5'UTRs of upregulated mRNAs contained a higher GC content, but those of downregulated mRNAs were significantly lower (**Fig. 2.7A, 2.7B, Table 2.3**). There were significant decreases in 3'UTR length of upregulated and downregulated mRNAs and GC content of downregulated mRNAs from other control groups (**Fig. 2.7A, 2.7B, Table 2.3**). In addition to using GC content as a measure of complexity of RNA sequences, the folding $\Delta G°$ of the RNA structures is commonly used (Hsieh et al., 2012; Thoreen et al., 2012). Taking into consideration the long lengths of numerous *Drosophila* UTRs and the difficulty in accurately processing long RNA sequences and predicting structures and interactions using currently available algorithms, we only calculated the GC content and thus was unable to provide a comprehensive view of UTR complexity (Leppek et al., 2018; Lorenz et al., 2011; Zuker, 2003). Taken together, the 5'UTRs of upregulated mRNAs were shorter but more complex according to

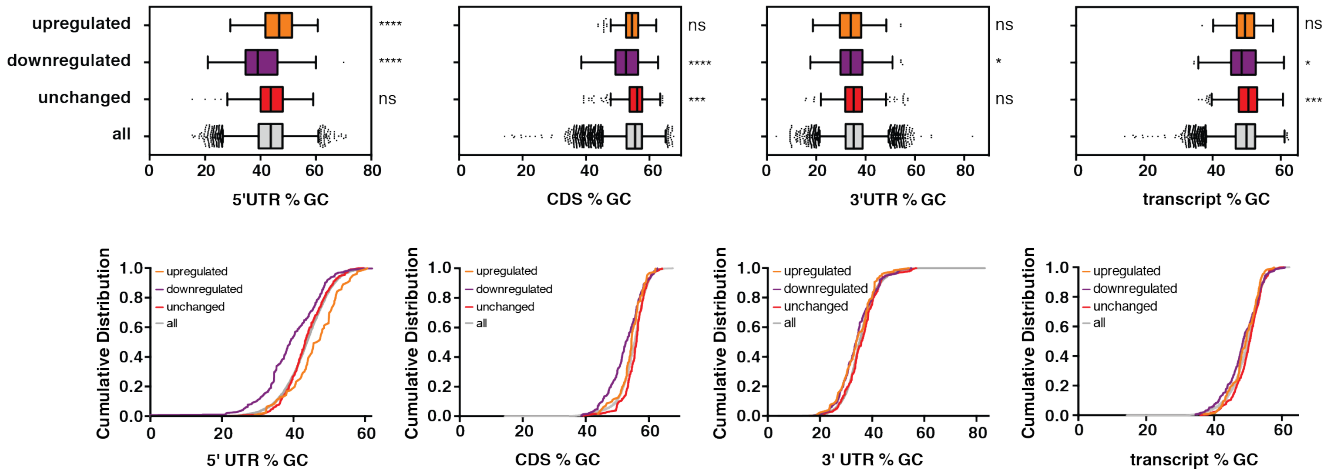
their higher % GC and their 3'UTRs were also shorter; conversely, downregulated mRNAs possessed 5' and 3'UTRs that were generally less complex based on their shorter length and lower % GC.

Next, we performed an analysis on the UTR sequences of upregulated, downregulated and unchanged mRNAs for known motifs using RegRNA and UTRscan because UTRs typically contain *cis*-regulatory elements (Grillo et al., 2010; Huang et al., 2006). The frequency of upstream open reading frames (uORFs) is reduced in 5'UTRs of upregulated and downregulated mRNAs compared to those of control mRNAs (**Fig. 2.7C, Table 2.4**). We also considered 5' TOP and TOP-like motifs, and PRTE, features contained in mRNAs that undergo mTORC1-dependent translational control via 4E-BPs (Hsieh et al., 2012; Thoreen et al., 2012). Notably, these features were not enriched in the 5'UTRs of upregulated mRNAs (**Fig. 2.7C, Table 2.4**). We next analyzed the presence of regulatory motifs in the 3'UTRs of these mRNAs. Interestingly, the incidence of Musashi binding element (MBE), a motif that mediates translational repression by Musashi (Bertolin et al., 2016; Okabe et al., 2001), was significantly reduced in the 3'UTRs of downregulated mRNAs compared to the upregulated and the control mRNAs (**Fig. 2.7C, Table 2.4**). Collectively, our data reveals distinct features present within the UTR sequences that may contribute to the upregulated translation of a subset of mRNAs, via uORFs, and downregulated translation of other mRNAs, via uORFs and MBEs.

A


upregulated
downregulated
unchanged
all
5'UTR length (nt)
CDS length (nt)
3'UTR length (nt)
transcript length (nt)
Cumulative Distribution
Log2 5' UTR length
Log2 CDS length
Log2 3' UTR length
Log2 transcript length

B


upregulated
downregulated
unchanged
all
5'UTR % GC
CDS % GC
3'UTR % GC
transcript % GC
Cumulative Distribution
5' UTR % GC
CDS % GC
3' UTR % GC
transcript % GC

C

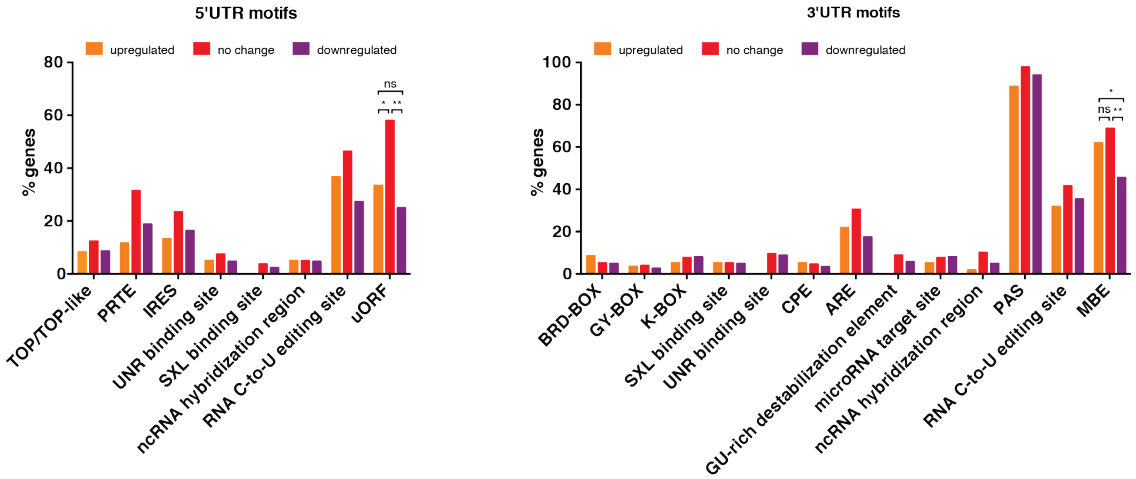

5'UTR motifs
3'UTR motifs
upregulated
no change
downregulated
% genes
TOP/TOP-like
PRTE
IRES
UNR binding site
SXL binding site
ncRNA hybridization region
RNA C-to-U editing site
uORF
BRD-BOX
GY-BOX
K-BOX
SXL binding site
UNR binding site
CPE
ARE
GU-rich destabilization element
microRNA target site
ncRNA hybridization region
PAS
RNA C-to-U editing site
MBE

Figure 2.7 Sequence analysis of upregulated and downregulated mRNAs

(A) Comparison of length of the 5'UTR, CDS, 3'UTR and whole transcript sequences of upregulated (orange), downregulated (purple), and unchanged (red) to all (grey) mRNAs identified in the ribosome profiling experiment. Sequences from all mRNA isoforms were used in the analysis. The distributions of length and % GC of each group are represented in the boxplots and cumulative distributions. The whiskers of the boxplots were drawn using the Tukey's method. Significance was determined by one-way ANOVA with Tukey's post-hoc. (**B)** Analysis of GC content of upregulated, downregulated, unchanged to all mRNAs is presented as boxplots and cumulative distributions. Boxplots whiskers were drawn using the Tukey's method. Significance was determined by one-way ANOVA with Tukey's post-hoc. (**C)** Analysis of 5' and 3'UTR sequences using RegRNA and UTRscan of upregulated (orange), downregulated (purple), and unchanged (red) mRNAs in *Thor* compared to *Revertant*. Significance was determined by two-way ANOVA with Tukey's post-hoc. Unlabelled columns are considered as non-significant proportions of genes containing motifs compared to those found in other gene groups. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; ns, non-significant.

Table 2.3 Statistical analysis of mRNA sequences using one-way ANOVA with Tukey's post-hoc

Parameter	Mean ± SEM		Significance	
5'UTR length	Up:	277 ± 30.71	Up vs. down:	$p = 0.8710$
	Down:	315 ± 24.77	Up vs. unchanged:	$p = 0.0209$
	Unchanged:	384.7 ± 18.22	Up vs. all:	$p = 0.0081$
	All:	383.3 ± 2.848	Down vs. unchanged:	$p = 0.0632$
			Down vs. all:	$p = 0.0127$
			Unchanged vs. all:	$p = 0.9998$
5'UTR % GC	Up:	46.10 ± 0.6324	Up vs. down:	$p < 0.0001$
	Down:	39.57 ± 0.5008	Up vs. unchanged:	$p = 0.0016$
	Unchanged:	43.69 ± 0.2815	Up vs. all:	$p < 0.0001$
	All:	43.58 ± 0.05054	Down vs. unchanged:	$p < 0.0001$
			Down vs. all:	$p < 0.0001$
			Unchanged vs. all:	$p = 0.8997$
CDS length	Up:	1358 ± 97.52	Up vs. down:	$p < 0.0001$
	Down:	6271 ± 661.9	Up vs. unchanged:	$p = 0.0459$
	Unchanged:	2211 ± 66.85	Up vs. all:	$p = 0.001$
	All:	2483 ± 23.84	Down vs. unchanged:	$p < 0.0001$
			Down vs. all:	$p < 0.0001$
			Unchanged vs. all:	$p = 0.2043$
CDS % GC	Up:	54.27 ± 0.3471	Up vs. down:	$p = 0.0032$
	Down:	52.6± 0.3088	Up vs. unchanged:	$p = 0.0169$
	Unchanged:	55.58 ± 0.1503	Up vs. all:	$p = 0.526$
	All:	54.82 ± 0.04697	Down vs. unchanged:	$p < 0.0001$
			Down vs. all:	$p < 0.0001$
			Unchanged vs. all:	$p = 0.0003$
3'UTR length	Up:	420.7 ± 42.47	Up vs. down:	$p = 0.2512$
	Down:	615.1 ± 57.08	Up vs. unchanged:	$p = 0.0066$
	Unchanged:	739.7 ± 36.28	Up vs. all:	$p = 0.0003$
	All:	779.8 ± 7.603	Down vs. unchanged:	$p = 0.3265$
			Down vs. all:	$p = 0.0261$
			Unchanged vs. all:	$p = 0.9058$
3'UTR % GC	Up:	34.25 ± 0.5802	Up vs. down:	$p = 0.9974$
	Down:	34.38± 0.3984	Up vs. unchanged:	$p = 0.1388$
	Unchanged:	35.54 ± 0.2494	Up vs. all:	$p = 0.1274$
	All:	35.45 ± 0.04697	Down vs. unchanged:	$p = 0.0419$
			Down vs. all:	$p = 0.0182$
			Unchanged vs. all:	$p = 0.9844$
transcript length	Up:	2053 ± 123	Up vs. down:	$p < 0.0001$
	Down:	7194 ± 676.3	Up vs. unchanged:	$p = 0.0032$
	Unchanged:	3291 ± 78.55	Up vs. all:	$p < 0.0001$
	All:	3642 ± 26.41	Down vs. unchanged:	$p < 0.0001$
			Down vs. all:	$p < 0.0001$
			Unchanged vs. all:	$p = 0.0938$
transcript % GC	Up:	49.05 ± 0.3593	Up vs. down:	$p = 0.6793$
	Down:	48.49± 0.3009	Up vs. unchanged:	$p = 0.1428$
	Unchanged:	50.03 ± 0.1649	Up vs. all:	$p = 0.9469$
	All:	49.28 ± 0.03601	Down vs. unchanged:	$p < 0.0001$
			Down vs. all:	$p = 0.0242$
			Unchanged vs. all:	$p = 0.0007$

Table 2.4 Statistical analysis of UTR motifs using two-way ANOVA with Tukey's post-hoc

UTR	Motifs	% genes containing motif		Significance	
5'	TOP/TOP-like	Up:	8.333	Up vs. down:	P = 0.9995
	(terminal oligopyrimidine)	Down:	8.594	Up vs. unchanged:	P = 0.8940
		Unchanged:	12.35	Down vs. unchanged:	P = 0.9065
5'	PRTE	Up:	11.67	Up vs. down:	P = 0.7095
	(pyrimidine-rich translational element)	Down:	18.75	Up vs. unchanged:	P = 0.0994
		Unchanged:	31.48	Down vs. unchanged:	P = 0.3497
5'	IRES	Up:	13.33	Up vs. down:	P = 0.9362
	(internal ribosomal entry site)	Down:	16.41	Up vs. unchanged:	P = 0.5050
		Unchanged:	23.46	Down vs. unchanged:	P = 0.7117
5'	UNR binding site	Up:	5.000	Up vs. down:	P = 0.9993
	(upstream of N-ras binding site)	Down:	4.688	Up vs. unchanged:	P = 0.9602
		Unchanged:	7.407	Down vs. unchanged:	P = 0.9496
5'	SXL binding site	Up:	0.000	Up vs. down:	P = 0.9623
	(Sex lethal binding site)	Down:	2.344	Up vs. unchanged:	P = 0.9088
		Unchanged:	3.704	Down vs. unchanged:	P = 0.9871
5'	ncRNA hybridization region	Up:	5.000	Up vs. down:	P = 0.9993
		Down:	4.688	Up vs. unchanged:	P > 0.9999
		Unchanged:	4.938	Down vs. unchanged:	P = 0.9996
5'	RNA C-to-U editing site	Up:	36.67	Up vs. down:	P = 0.5578
		Down:	27.34	Up vs. unchanged:	P = 0.5374
		Unchanged:	46.30	Down vs. unchanged:	P = 0.1177
5'	uORF	Up:	33.33	Up vs. down:	P = 0.6247
	(upstream open reading frame)	Down:	25.00	Up vs. unchanged:	P = 0.0364
		Unchanged:	58.02	Down vs. unchanged:	P = 0.0060
3'	BRD-BOX	Up:	8.333	Up vs. down:	P = 0.8159
	(Bearded box)	Down:	4.688	Up vs. unchanged:	P = 0.8380
		Unchanged:	4.938	Down vs. unchanged:	P = 0.9990
3'	GY-BOX	Up:	3.333	Up vs. down:	P = 0.9850
		Down:	2.344	Up vs. unchanged:	P = 0.9979
		Unchanged:	3.704	Down vs. unchanged:	P = 0.9719
3'	K-BOX	Up:	5.000	Up vs. down:	P = 0.8855
		Down:	7.813	Up vs. unchanged:	P = 0.9147
		Unchanged:	7.407	Down vs. unchanged:	P = 0.9975
3'	SXL binding site	Up:	5.000	Up vs. down:	P = 0.9985
	(Sex lethal binding site)	Down:	4.688	Up vs. unchanged:	P > 0.9999
		Unchanged:	4.938	Down vs. unchanged:	P = 0.9990
3'	UNR binding site	Up:	0.000	Up vs. down:	P = 0.3376
	(upstream of N-ras binding site)	Down:	8.594	Up vs. unchanged:	P = 0.2862
		Unchanged:	9.259	Down vs. unchanged:	P = 0.9932
3'	CPE	Up:	5.000	Up vs. down:	P = 0.9472
	(cytoplasmic polyadenylation element)	Down:	3.125	Up vs. unchanged:	P = 0.9929
		Unchanged:	4.321	Down vs. unchanged:	P = 0.9782
3'	ARE	Up:	21.67	Up vs. down:	P = 0.7366
	(AU-rich element)	Down:	17.19	Up vs. unchanged:	P = 0.3387
		Unchanged:	30.25	Down vs. unchanged:	P = 0.0940
3'	GU-rich destabilization element	Up:	0.000	Up vs. down:	P = 0.6230
		Down:	5.594	Up vs. unchanged:	P = 0.3337
		Unchanged:	8.642	Down vs. unchanged:	P = 0.8671
3'	microRNA target site	Up:	5.000	Up vs. down:	P = 0.8855
		Down:	7.813	Up vs. unchanged:	P = 0.9147
		Unchanged:	7.407	Down vs. unchanged:	P = 0.9975
3'	ncRNA hybridization region	Up:	1.667	Up vs. down:	P = 0.8693
		Down:	4.688	Up vs. unchanged:	P = 0.3697
		Unchanged:	9.877	Down vs. unchanged:	P = 0.6647
3'	PAS	Up:	88.33	Up vs. down:	P = 0.6413
	(polyadenylation site)	Down:	93.75	Up vs. unchanged:	P = 0.2908
		Unchanged:	97.53	Down vs. unchanged:	P = 0.8036
3'	RNA C-to-U editing site	Up:	31.67	Up vs. down:	P = 0.8298
		Down:	35.16	Up vs. unchanged:	P = 0.2558
		Unchanged:	41.36	Down vs. unchanged:	P = 0.5606
3'	MBE	Up:	61.67	Up vs. down:	P = 0.0297
	(Musashi binding element)	Down:	45.31	Up vs. unchanged:	P = 0.4953
		Unchanged:	68.52	Down vs. unchanged:	P = 0.0020

2.2.5 RPS6 is phosphorylated at higher levels in *Thor* fly heads

We sought to validate the upregulated target mRNAs identified by ribosome profiling and used immunoblotting to assess protein levels in *Revertant* and *Thor* heads. Gene ontology analysis identified that specific upregulated mRNAs (*Attacin-B* (*AttB*), *Attacin-C* (*AttC*), *Immune induced molecule 1* (*IM1*), *Immune induced molecule 3* (*IM3*), *modular serine protease* (*modSP*) and *peptidoglycan recognition protein SD* (*PGRP-SD*)) corresponded to terms related to the immune response (**Fig. 2.6**). Their expression in the male adult fly head is confirmed by modENCODE (Celniker et al., 2009), and they are preferentially ribosome-associated according to ribosome profiling, suggesting they are translationally upregulated. Due to the lack of available antibodies, we were only able to assess protein levels of *AttC* ($\log_2$TE = 1.45) and *PGRP-SD* ($\log_2$TE = 1.21); however, they did not differ significantly between *Revertant* and *Thor* (**Fig. 2.8**).

Another target identified by ribosome profiling that was of particular interest was *dS6K* ($\log_2$TE = 1.09) as dS6K is a downstream target of dTOR that gets phosphorylated along with d4E-BP (Miron et al., 2003). In contrast to the 2.13-fold TE increase determined by ribosome profiling, protein levels of total dS6K were not significantly different between *Revertant* and *Thor* (**Fig. 2.9A left panel and 2.9B**). Similarly, *Ribosomal protein L11* (*RPL11*), another upregulated mRNA target ($\log_2$TE = 0.82), did not have significantly elevated protein level as detected by immunoblot (**Fig. 2.9A right panel and 2.9B**). These results do not exclude the possibility that dS6K, RPL11, AttC and PGRP-SD may be synthesized at greater rates in *Thor* heads, but also turned over more rapidly, resulting in a similar steady-state level.

Next, we examined the levels of phosphorylated Ribosomal protein S6 (RPS6) as a readout of dS6K activity. Elevated levels of phosphorylated RPS6 were indeed detected in *Thor* heads while levels of RPS6 remained the same (**Fig. 2.9A right panel and 2.9B**). These results point towards increased dS6K activity; however, elevated levels of phosphorylated dS6K were not detected in *Thor* (**Fig. 2.9A and 2.9B**). We also considered changes in TOR signaling by assessing levels of dAkt, the upstream positive regulator of TOR. Neither total dAkt nor phospho-dAkt protein levels were significantly different between *Revertant* and *Thor* (**Fig. 2.9A and 2.9B**). Together, these results suggest that the higher levels of phospho-RPS6 in *Thor* may not be a consequence of increased TOR signaling, even though dS6K was determined to be translationally

upregulated in the absence of d4E-BP. This raises the possibility that increased levels of phospho-RPS6 in *Thor* is the result of kinases other than dS6K or a decrease in activity of the phosphatase that dephosphorylates phospho-RPS6.

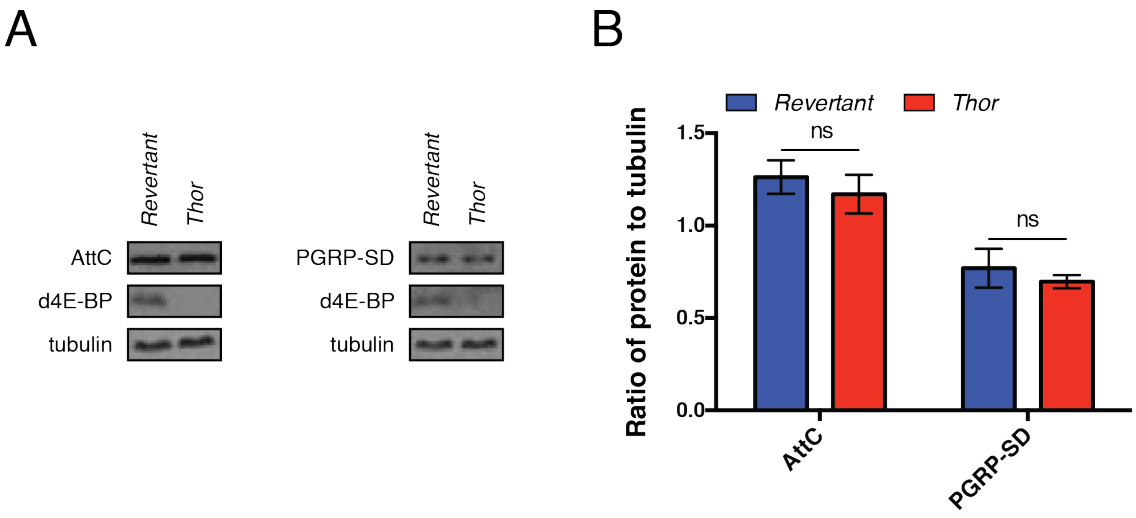


Figure 2.8 Immunoblot analysis of immune response proteins show equal levels in *Revertant* and *Thor* heads

(A) Representative immunoblots from head lysate of *Thor* and *Revertant* flies with tubulin as the loading control. The two panels represent different immunoblots using the same fly head lysates. (**B)** Quantification of immunoblots in A. Protein amounts were normalized to tubulin. Data are presented as mean ± SEM (error bars). Significance was determined by t-test with Holm-Sidak's post-hoc. n = 3 biological replicates of 30 fly heads. ns, non-significant.

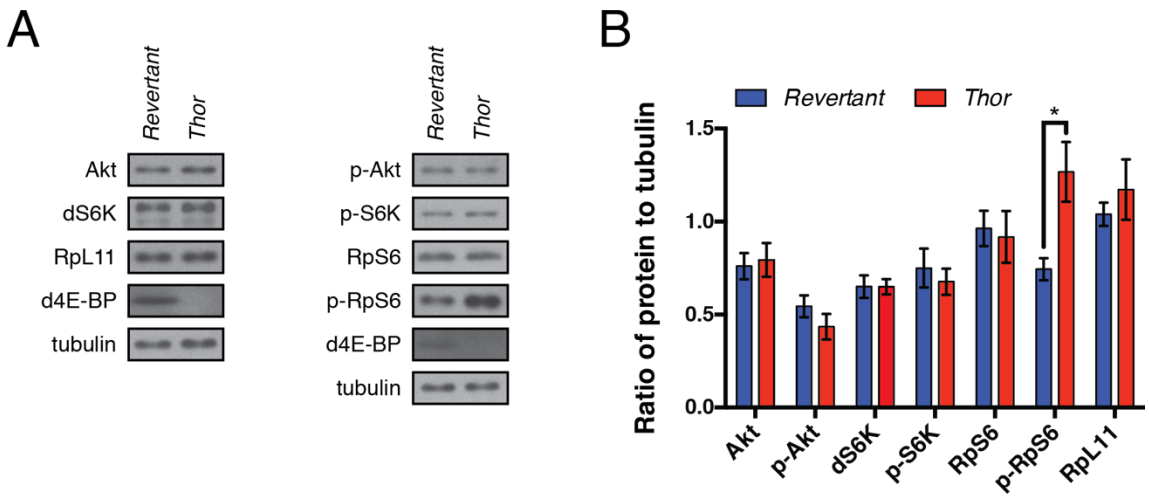


Figure 2.9 Immunoblot analysis shows higher levels of phosphorylated RPS6 in *Thor* fly head lysate

(A) Representative immunoblots from head lysate of *Thor* and *Revertant* flies. Tubulin was the loading control. The two panels represent different immunoblots using the same fly head lysates. **(B)** Quantification of immunoblots in A. Protein amounts were normalized to tubulin. Data are presented as mean ± SEM (error bars). Significance was determined by t-test with Holm-Sidak's post-hoc. n = 3 biological replicates of 30 fly heads. *P < 0.05.

2.3 Discussion

Here, we confirmed that global translation is unaffected by the *d4E-BP^{null}* mutation (**Fig. 2.1**) and identified the differential translation of a subset of mRNAs in *Thor* mutant fly heads by ribosome profiling (**Fig. 2.5**). A gene ontology analysis on target mRNAs identified by ribosome profiling indicated that some upregulated mRNAs are involved in innate immunity (**Fig. 2.6**). Furthermore, upregulated mRNAs demonstrated differences in length and complexity, as measured by GC content, in 5'UTR sequences compared to control mRNA groups (**Fig. 2.7**). Moreover, we detected elevated levels of p-RPS6 in *Thor* flies, which was used as a readout of dS6K activity, in our attempt to validate dS6K as one of the upregulated mRNAs (**Fig. 2.9**).

4E-BP2, the predominantly expressed 4E-BP in the mammalian brain, regulates the translation of neuroligins where dysregulation of this process results in autism spectrum disorders (ASD)-like phenotypes in mice (Gkogkas et al., 2013). In contrast, we did not detect significantly higher levels of *Drosophila* neuroligins in our ribosome profiling dataset in the absence of d4E-BP. Unlike the aforementioned mouse model for ASD, Neuroligin 2-deficient flies showed ASD-like behavioral symptoms, including reduced social interactions and altered acoustic communication signals (Hahn et al., 2013). The impact on behavior of enhanced cap-dependent translation of *Drosophila* neuroligins has yet to be determined. Due to d4E-BP's high degree of conservation with mammalian 4E-BP1 in sequence and in function (Miron et al., 2003) and less so to mammalian 4E-BP2 in sequence, it is therefore likely that d4E-BP is not a functional ortholog of 4E-BP2 and would not regulate the translation of orthologous mRNAs.

A link between the *Drosophila* immune system and d4E-BP has previously been established as *Thor* mutant flies are immune compromised due to their increased susceptibility to infection compared to control flies (Bernal and Kimbrell, 2000; Levitin et al., 2007). A recent study resolved the paradox of the requirement d4E-BP as a translation inhibitor in the stimulation of AMP synthesis by demonstrating the preferential translation of immune transcripts in a cap-independent manner upon d4E-BP activation by bacterial infection (Vasudevan et al., 2017). Nonetheless, it was an unanticipated finding to identify, by ribosome profiling, specific immune transcripts, *AttB*, *AttC*, *DptB*, *IM1*, *IM3*, *PGRP-SD* and *modSP*, among the mRNAs that are translationally upregulated in *Thor* fly heads (**Table 2.1**); however, AttC and PGRP-SD protein

levels were not significantly different between *Thor* and *Revertant* (**Fig. 2.8**). These transcripts encode components of the Imd and Toll pathways. AttB, AttC and DptB are AMPs that are regulated by the Imd pathway (De Gregorio et al., 2002; Dushay et al., 2000). PGRP-SD is an extracellular receptor that enhances Imd signaling activity (Iatsenko et al., 2016). IM1 and IM3 are immune induced molecules regulated by Toll signaling (Clemmons et al., 2015). modSP is a modular serine protease that mediates Toll activation upon Gram positive bacterial and fungal infection (Buchon et al., 2009).

A plausible explanation is that young unchallenged flies, such as those used for this study, are able to clear high levels of immune-regulated proteins, and the gradual accumulation of these proteins contributes to the aging process. There is a growing body of evidence that indicates the overexpression of immune pathways is detrimental to the animal in which adverse effects include neurodegeneration and shortened lifespan (Badinloo et al., 2018; Cao et al., 2013; Kounatidis et al., 2017). Moreover, a role for d4E-BP in longevity has been established where the lifespan of *Thor* flies is 25% shorter in males compared to control flies (Tettweiler et al., 2005), and its expression in muscle (Demontis and Perrimon, 2010) and in the heart (Wessells et al., 2009) can extend the lifespan of adult flies. As such, it is conceivable that the preferential ribosome association to immune transcripts, and presumably the translation of these mRNAs, may be the early signs of a premature accumulation of immune proteins in *Thor* heads that contributes to an accelerated aging process in the adult flies. Given we only considered protein levels in young 3–5 day old flies, it will be important to examine immune protein expression throughout the fly's lifespan which will allow us to determine if there is indeed early onset of chronic inflammation in *Thor* flies.

Interestingly, the expression of components of the immune system in the *Drosophila* head is not necessarily indicative of an immune response; some can influence nervous system function including sleep (Dissel et al., 2015), presynaptic homeostasis (Harris et al., 2015) and non-associative learning (Bozler et al., 2017). With the majority of DptB expressed in the adult head fat body and little detected in the central brain, it is indeed the expression of DptB in the head fat body that is required for long-term memory (Barajas-Azpeleta et al., 2018). The upregulated translation of the immune transcripts identified here may impact the behavioral repertoire in the adult fly, which can be addressed by performing behavioral tests on *Thor* flies. Due to our

experimental setup of processing whole heads, our analysis presented several limitations including not being able to distinguish which head tissue or cell type expressed the transcripts and missing transcripts expressed in small populations of cells that demonstrate changes in translation as a result of the *d4E-BP*^{null} mutation. Given that the expression of an AMP outside of the brain can influence nervous system function, it will be important to assess cell type specific changes in translation. Employing a recently developed tissue-specific ribosome profiling technique would provide insight into the functional significance of the selective translation control of mRNAs in the *Thor* fly head (Chen and Dickman, 2017).

From the UTR analysis, we identified mRNAs that are translationally upregulated due to the *d4E-BP*^{null} mutation to contain 5'UTR sequences that are shorter, lower in GC content and not particularly enriched for sequence elements that confer mTORC1-eIF4E regulation, TOP and PRTE (**Fig. 2.7**). They differ from two subsets of mRNAs that are regulated by the mTORC1-eIF4E pathway: (1) mRNAs that are sensitive to levels of eIF4E and contain long and structured 5'UTRs, also termed "eIF4E-sensitive" mRNAs (Duncan et al., 1987; Koromilas et al., 1992), and (2) 5' TOP and/or PRTE containing mRNAs that are shorter and less complex (Hsieh et al., 2012). In terms of length and complexity, the 5'UTRs of upregulated mRNAs do not conform to either groups, and instead display intermediate characteristics. The differences observed may be *Drosophila*-specific. Among the upregulated mRNAs are many that do not have well-characterized mammalian orthologues, including the immune transcripts which are unique to *Drosophila.* The 5'UTRs of immune transcripts are typically short and contain IRES, allowing AMPs to be rapidly synthesized in response to infection in a cap-independent manner (Vasudevan et al., 2017). The underrepresentation of uORFs in the 5'UTRs of upregulated and downregulated mRNAs was unexpected. The lower frequency of uORFs in the 5'UTRs of upregulated mRNAs is in line with the role of an uORF in modulating translation of the downstream CDS by sequestering ribosomes (Zhang et al., 2019). Our analysis does not take into account factors that have a positive correlation with repressiveness of *D.* melanogaster uORFs including optimized Kozak context around the uORF AUG, long distance between the uORF AUG and 5'cap and high level of conservation of the uORF (Zhang et al., 2018). Moreover, crosstalk between uORFs and other *cis*-regulatory elements could also contribute to the translational regulation of mRNAs, an example of this is shown with changes in m^{6}A modification of an uORF of *ATF4* can promote the

translation of *ATF4* during integrated stress response (Zhang et al., 2019; Zhou et al., 2018). Therefore, the overall repressiveness of the predicted frequency of uORFs of each mRNA group used in our analysis will need to be considered and along with the possibility that there is crosstalk between different *cis*-regulatory elements.

In our attempt to validate the ribosome profiling results, we observed no significant changes in protein levels for some of the upregulated targets, including dS6K, RPL11, AttC and PGRP-SD (**Fig. 2.8, 2.9**). Considering d4E-BP's role as an inhibitor of translation through its function in binding eIF4E and thereby preventing the formation of the eIF4F complex, ribosomes are likely recruited to mRNAs more so in *Thor* flies. However, our results do not address the rate of translation elongation of each mRNA, which is a determining factor of the level of protein synthesized, because ribosome profiling provides a snapshot of ribosome occupancy on a given mRNA. The ribosomes may be translating at a slower rate or even stalled on the mRNA. Elongation speed can be influenced by mRNA secondary structures (Chen et al., 2013) and codon usage (Zhao et al., 2017).

dS6K protein level was also not significantly increased in *Thor* despite the increase in ribosome occupancy, nonetheless, we did observe higher levels of phosphorylated RPS6 (**Fig. 2.9**). RPS6 phosphorylation is specifically activated via TORC1-S6K in *Drosophila* wing disc, eye disc Kc167 and S2 cell lines (Kim et al., 2017; Romero-Pozuelo et al., 2017). In the adult fly brain, phosphorylated RPS6 is present in small populations of cells, namely insulin-producing cells and *Pdf*-expressing circadian neurons, and this phosphorylation is significantly decreased upon nutrient restriction, suggesting it is regulated by TORC1 activity (Kim et al., 2017). Phosphorylation of RPS6 by S6K in the *Drosophila* brain is a molecular maker for neuronal activity (Acevedo et al., 2015), consistent with phospho-RPS6 in the mouse brain (Knight et al., 2012). This suggests there may be neuronal hyperactivity in the *Thor* adult brain although further testing is required to establish if this is indeed the case. Taking into consideration that only small populations of neurons express phospho-RPS6, namely neuropeptide neurons, including insulin producing neurons and circadian neurons, as detected by an antibody that recognizes phospho-RPS6 on S233 and S235 (Kim et al., 2017), the majority of phospho-RPS6 that is detected in the head is expressed in tissues outside of the brain. It will therefore be interesting to determine the biological significance of enhanced RPS6 phosphorylation in those cell types of *Thor* fly heads.

In mammals, p90 ribosomal S6 kinase (RSK) also phosphorylates RPS6 (Anjum and Blenis, 2008), as does *Drosophila* RSK as demonstrated by a kinase assay (Kim et al., 2006). RSK is expressed in the adult fly head and has a role in modulating circadian behavior (Akten et al., 2009). Although TORC1 activity via S6K predominantly regulates RPS6 phosphorylation, there remains the possibility that RSK or other kinases can in part contribute to it in the adult head.

In conclusion, we have described mRNAs that are preferentially bound by ribosomes in the absence of d4E-BP that contain shorter but more complex 5'UTRs, some of which are linked to specific biological processes. Further studies will be required to better understand the biological significance of dysregulated translation in *Thor* flies.

2.4 Materials and Methods

2.4.1 Puromycin incorporation assay

3–4 day old male *yw*, *Revertant*, and *Thor* flies were starved for 6 hours and then placed on food supplemented with puromycin (600 µM puromycin, 1 % agarose, and 5 % sucrose) for 24 hours. Heads were collected, lysed and prepared for immunoblotting. Briefly, 30 µg total protein per sample was run on a 12 % polyacrylamide gel. Mouse anti-puromycin (Kerafast, 1:1000, 3RH11) was used as the primary antibody. *yw* flies were used as controls.

2.4.2 Polysome fractionation

Heads from 3–5 day old male *Revertant* and *Thor* flies were flash frozen and pulverized in liquid nitrogen and lysed in hypotonic buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.5 % Triton X-100, 1 mM DTT, 20 U/ml SuperasIn (Ambion by Life Technologies), 20 µg/ml emetine, and 50 µM GMP-PNP (Sigma Aldrich)). Lysates were clarified by centrifugation at 3,000 × *g* for 10 min at 4 ºC followed by second round of centrifugation at 20,000 × *g* for 10 min at 4 ºC. 300 µg total RNA was loaded onto 10–50 % (wt/vol) sucrose density gradients (50 mM Tris pH 7.5, 250 mM NaCl, 15 mM $MgCl_2$, 20 U/ml SuperasIn, and 20 µg/ml emetine) separated by centrifugation at 36 krpm for 3h at 4 ºC in an SW40 rotor (Beckman Coulter). Fractions were collected using an ISCO gradient fractionation system at 35 second intervals. The absorbance of OD 254 was continuously recorded with a Foxy JR Fractionator (Teledyne ISCO).

2.4.3 Ribosome profiling

Flash frozen heads from 3–5 day old male flies were pulverized in liquid nitrogen and lysed in hypotonic buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5mM $MgCl_2$, 0.5 % Triton X-100, 1 mM DTT, 20 U/ml Superasin (Ambion by Life Technologies), 20 µg/ml emetine, 50 µM GMP-PNP (Sigma Aldrich)). Lysate was clarified by centrifugation (3,000 × g) at 4 ºC for 10 min. Supernatant was collected and clarified again by centrifugation (20,000 × g) at 4 ºC for 10 min. The ribosome profiling assay was performed as described in (Dunn et al., 2013) with minor modifications. Briefly, 100 µg total RNA (two biological replicates) was diluted 2:1 in digestion buffer (50 mM

Tris pH 7.5, 5 mM $MgCl_2$, 0.5 % Triton X-100, 1 mM DTT, 20 U/ml Superasin, 20 µg/ml emetine, 15 mM $CaCl_2$, and 500 U micrococcal nuclease (Roche Applied Science) and digested at 25 ºC for 40 min with gentle mixing. Monosomes were sedimented by ultracentrifugation in a 34 % sucrose cushion at 70 krpm for 4 h at 4 ºC (TLA-120.2 rotor, Beckman Coulter). Resuspended monosomes were extracted twice with acid phenol and once with chloroform. RNA was precipitated with isopropanol, NaOAc and GlycoBlue (Ambion). 50 µg cytoplasmic RNA (two biological replicates) was used for mRNA-Seq analysis. Poly(A)+ mRNAs were selected on oligo-dT25 Dynabeads (Invitrogen) according to manufacturer's instructions. RNA was fragmented in alkaline fragmentation buffer (2 mM EDTA, 100 mM $NaCO_3$/$NAHCO_3$, pH 9.2) at 95 ºC for 20 min, stopped in stop/precipitation solution (300 mM NaOAc (pH 5.5) and GlycoBlue), and precipitated in isopropanol. Ribosome footprints (RFPs) and randomly fragmented mRNAs were dephosphorylated using T4 polynucleotide kinase (New England Biolabs) and size selected on a 15 % TBE-urea gel (Invitrogen). Gel slabs in the regions of 28-34 nt and 55-65 nt corresponding to RFPs and fragmented mRNAs were excised, eluted, and precipitated. Samples were analyzed using the Small RNA Bioanalyzer assay (Agilent). 10 pmol of RNA was ligated to the 3' miRNA cloning linker 1 with T4 RNA ligase 1 (New England Biolabs) and then precipitated. Ligated RNA was purified from the empty linker on a 10% TBE-urea gel (Invitrogen) and then reverse transcribed using SuperScript III (Invitrogen) and the o225-link1 primer. Two rounds of subtractive hybridization were carried out on RFP cDNA using a mixture of biotinylated complementary rRNA oligonucleotides (oJGD132, oJGD133, oJGD134, oJGD135, oJGD136, oJGD161, oJGD162, oJGD163, and oJGD164) in a ratio of 25.5:1:13:17:4:6:2:11:21. Subtracted RFP and mRNA samples were circularized using CircLigase (Epicentre). Libraries were amplified for 12 cycles by PCR with Phusion polymerase (Finnzymes) using oNTI231 and indexing primers (oCJ30, oCJ31, oCJ32 and oCJ33, oCJ35, oCJ36, oCJ37, oCJ38, oCJ39, oCJ40, oCJ41, and oCJ42) and analyzed using the High Sensitivity DNA Bioanalyzer assay (Agilent). Samples were sequenced by Illumina HiSeq 2500 SR50 (Genome Quebec). Oligonucleotides used for sequencing library generation are found in Table 2.5.

2.4.4 Bioinformatics analysis of ribosome profiling data

Raw sequencing reads were demultiplexed by Genome Quebec. The adaptor sequence was trimmed from sequencing reads with the FASTX toolkit. Bowtie was used to remove contaminant sequences including rRNAs and tRNAs. Reads were then aligned to the *Drosophila melanogaster* reference genome release 6 using STAR. Translational efficiency (TE) for each transcript was calculated as the ratio of the ribosome footprint in reads per kilobase per million (RPKM) to mRNA fragment in RPKM. The Xtail analysis pipeline was used to determine the differential translation of gene by calculating the $\log_2$ fold change and the statistical significance (Xiao et al., 2016). Genes with fewer than 40 reads were not considered in the analysis.

2.4.5 Gene ontology term analysis

The enrichment for gene ontology terms of mRNAs with $\log_2$TE ± 0.697 (TE fold change of 1.5) was determined using Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang da et al., 2009; Huang et al., 2007).

2.4.6 mRNA sequence analysis

Whole transcript, 5'UTR, CDS and 3'UTR sequences were obtained from FlyBase (Gramates et al., 2017). Length and GC content of were calculated using EMBOSS: infoseq (Rice et al., 2000). Motifs were identified by RegRNA2.0 (Huang et al., 2006) and UTRscan (Grillo et al., 2010). The TOP motif is defined as a series of 5-15 pyrimidines, beginning with a cytidine, that is located at the 5' terminus of the mRNA (Jefferies et al., 1994; Meyuhas, 2000; Thoreen et al., 2012). The TOP-like motif consists of a stretch of at least 5 pyrimidines within the first 4 nucleotides after the 5' cap structure (Thoreen et al., 2012). The PRTE is another pyrimidine rich motif consisting of an invariant uridine at position 6, but it is not located at the 5' terminus of the mRNA (Hsieh et al., 2012), and a PRTE of at least 9 nucleotides in length was considered for this analysis.

2.4.7 Immunoblotting

30 heads from 3–5 day old male flies per replicate were collected and homogenized in lysis buffer (8 M urea, 0.02 M HEPES pH 7.4, 0.1 M KCl, 2 mM EDTA, 0.5 % Triton X-100, 1 mM DTT, 2 mM PMSF, 1 X Halt™ Protease Inhibitor Cocktail (Thermo Scientific)). Samples were incubated on ice for 5 min and clarified by centrifugation at 15,000 g for 10 min at 4 °C. The supernatant was collected and protein concentration was measured by Bradford protein assay (Bio-Rad). 30 µg of protein were loaded per lane on 9–12.5 % polyacrylamide gels then transferred to 0.2 µm nitrocellulose membranes (Bio-Rad). Proteins were blocked in 5 % BSA in TBS-T (10 mM Tris pH 7.6, 150 mM NaCl, 0.1 % Tween 20) for 1 h at room temperature, incubated with primary antibodies diluted in 5 % milk TBS-T overnight at 4 °C, and finally with secondary antibodies also diluted in 5 % BSA TBS-T for 1 h at room temperature. Blots were quantified using ImageJ. The following primary antibodies were used: rabbit anti-d4E-BP (this study, 74-3, 1:500), rabbit anti-phospho-RPS6 ((Kim et al., 2017), 1:5000), mouse anti-RPS6 (Santa Cruz, C-8, 1:1000), rabbit anti-phospho-p70 S6 Kinase (Thr398) (Cell Signaling Technology, 9205, 1:1000), guinea pig anti-S6 Kinase ((Hahn et al., 2010), 1:1000), rat anti-AttC (Kim and Kim-Ha, 2006), 1:1000), rabbit anti-PGRP-SD ((Wang et al., 2008), 1:1000) and mouse anti-α-tubulin (Sigma-Aldrich, T9026, 1:5000).

Table 2.5 Oligonucleotides used for ribosome profiling sequencing library generation

Oligo	Sequence
28mer	AGUCACUUAGCGAUGUACACUGACUGUG/3Phos/
34mer	AUGUACACGGAGUCGAGCACCCGCAACGCGACUG/3Phos/
Linker-1	/5rApp/CTGTAGGCACCATCAAT/3ddC/
o225-link1	/5Phos/GATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CACTCA/iSp18/CAAGCAGA AGACGGCATACGAATTGATGGTGCCTACAG
oJGD132	/5BiotinTEG/CATTGTAATCTATTAGCATATACCAAATTT
oJGD133	/5BiotinTEG/TGATAAAGTGCTGATAGATTTATATGATTA
oJGD134	/5BiotinTEG/GCTAATTAACACAATCCCG/ideoxyI//ideoxyI/GCGTTCTAT
oJGD135	/5BiotinTEG/ACGACAATGGATGTGATGCCAATGTAATTT
oJGD136	/5BiotinTEG/GGTTGAACTCTAGATAACATGCAGATCGTA
oJGD161	/5BiotinTEG/TTTGATGCAAGCTTCTTGATCAAAGTATCACGAGT
oJGD162	/5BiotinTEG/TGCATTGTATGGCTTCTAAACCATTTAAAGTTTAT
oJGD163	/5BiotinTEG/CTTGGACTACATATGGTTGAGGGTTG
oJGD164	/5BiotinTEG/TTGGACTACATATGGTTGAGGG
oJGD165	/5BiotinTEG/GCTTGGACTACATATGGTTGAGGGT
oCJ30	AATGATACGGCGACCACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGCGACAGGTTCAGAGTTC
oCJ31	AATGATACGGCGACCACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTGACCACGACAGGTTCAGAGTTC
oCJ32	AATGATACGGCGACCACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACCAGATCCGACAGGTTCAGAGTTC
oCJ33	AATGATACGGCGACCACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATCGACAGGTTCAGAGTTC
oCJ35	AATGATACGGCGACCACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGCGACAGGTTCAGAGTTC
oCJ36	AATGATACGGCGACCACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATCGACAGGTTCAGAGTTC
oCJ37	AATGATACGGCGACCACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACCAGATCCGACAGGTTCAGAGTTC
oCJ38	AATGATACGGCGACCACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACACTTGACGACAGGTTCAGAGTTC
oCJ39	AATGATACGGCGACCACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACGATCAGCGACAGGTTCAGAGTTC
oCJ40	AATGATACGGCGACCACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACTAGCTTCGACAGGTTCAGAGTTC
oCJ41	AATGATACGGCGACCACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACGGCTACCGACAGGTTCAGAGTTC
oCJ42	AATGATACGGCGACCACCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCACCTTGTACGACAGGTTCAGAGTTC
oNTI231	CAAGCAGAAGACGGCATACGA

Code	Modification
Phos	Phosphorylation
rApp	Adenylation
ddC	Dideoxycytidine
iSp18	Internal 18-PEG spacer arm
BiotinTEG	Biotin with TEG spacer arm
ideoxyI	Internal deoxyInosine

2.5 Appendix

2.5.1 Proteomic analysis of *Revertant* and *Thor* heads

We sought to validate the results obtained in the ribosome profiling experiment of *Revertant* and *Thor* fly heads. Many of the differentially translated mRNAs identified by ribosome profiling did not have corresponding antibodies to perform immunoblotting. To identify proteins that are differentially expressed between *Revertant* and *Thor* fly heads, we thus prepared head lysate samples from 3–5 day old male flies for mass spectrometry. A total of 907 proteins were found by LC-MS/MS in our samples with an overlap of 832 proteins with another study that identified 4812 head proteins (**Fig. 2.10A**) (Aradska et al., 2015). The remaining 75 proteins in our samples that were not confirmed by the other study could have originated from the stray legs and bodies that contaminated head collections. The biological duplicates of *Revertant* and *Thor* head samples were indeed reproducible as indicated by their strong correlation (R^2 of spectral counts of 0.979 and 0.933, respectively) (**Fig. 2.10B, C**). A quantitative analysis to determine significant fold changes in protein levels identified 62 upregulated proteins and 58 downregulated proteins (**Fig. 2.10D, Table 2.6**).

Unfortunately, most of the upregulated and downregulated targets found by ribosome profiling were not validated with corresponding changes in protein levels. Moreover, d4E-BP was not identified among the 907 proteins found by mass spectrometry in the *Revertant* head lysate. To this end we selected ATPα, one of the proteins that demonstrated a dramatic increase in levels in *Thor* compared to *Rev* with a 210-fold change, for further testing to determine the validity of this experiment. Our immunoblotting results for ATPα did not detect the reported ~100 kDa band found in head samples (Jensen et al., 2013), instead bands of higher molecular weights were detected (**Fig. 2.11**). Even if the ~180 kDa band indeed corresponds to ATPα, we did not observe significant differences in protein levels between *Thor* and *Revertant* samples. We also considered AttC, a protein with a corresponding mRNA that was shown to be upregulated in *Thor* according to ribosome profiling, which was detectable by immunoblotting but not by mass spectrometry (**Fig. 2.11**). d4E-BP protein was detected by immunoblot, which was used for confirming the genotypes of the flies (**Fig. 2.11**).

Discrepancies were observed between the mass spectrometry and immunoblotting results. Although mass spectrometry analysis of our samples did not detect d4E-BP, neither did the one performed by the other study (Aradska et al., 2015). In addition to the immunoblot result that shows d4E-BP protein in the *Revertant* head, a protein-trap line where EYFP, flanked by splicing acceptor and donor sites, was inserted into the *d4E-BP* locus shows the expression of the reporter in specific adult brain structures, adult pars intercerebralis, adult brain cortex and median bundle (Knowles-Barley et al., 2010), confirming the expression of d4E-BP in the fly head. Likewise, AttC, PGRP-SD, Akt, and S6K were detectable by immunoblot in *Revertant* and *Thor* heads but not by LC-MS/MS (**Fig 2.8, 2.9**). By extension, many other proteins in these head samples could have been missed in this proteomics analysis. Taking this into account along with the disagreement in the results for ATPα using the two techniques, we conclude that the proteomics analysis did not provide an accurate identification and quantification of the proteins within our samples. We should therefore seek other methods to validate the ribosome profiling results. A possible experiment to achieve this is to express a reporter construct where a GPF cassette was inserted into a gene locus of interest, similar to the protein-trap lines, in a $d4E\text{-}BP^{null}$ background, and then determine if the protein level differs compared to the control background using immunoblot and/or immunostaining.

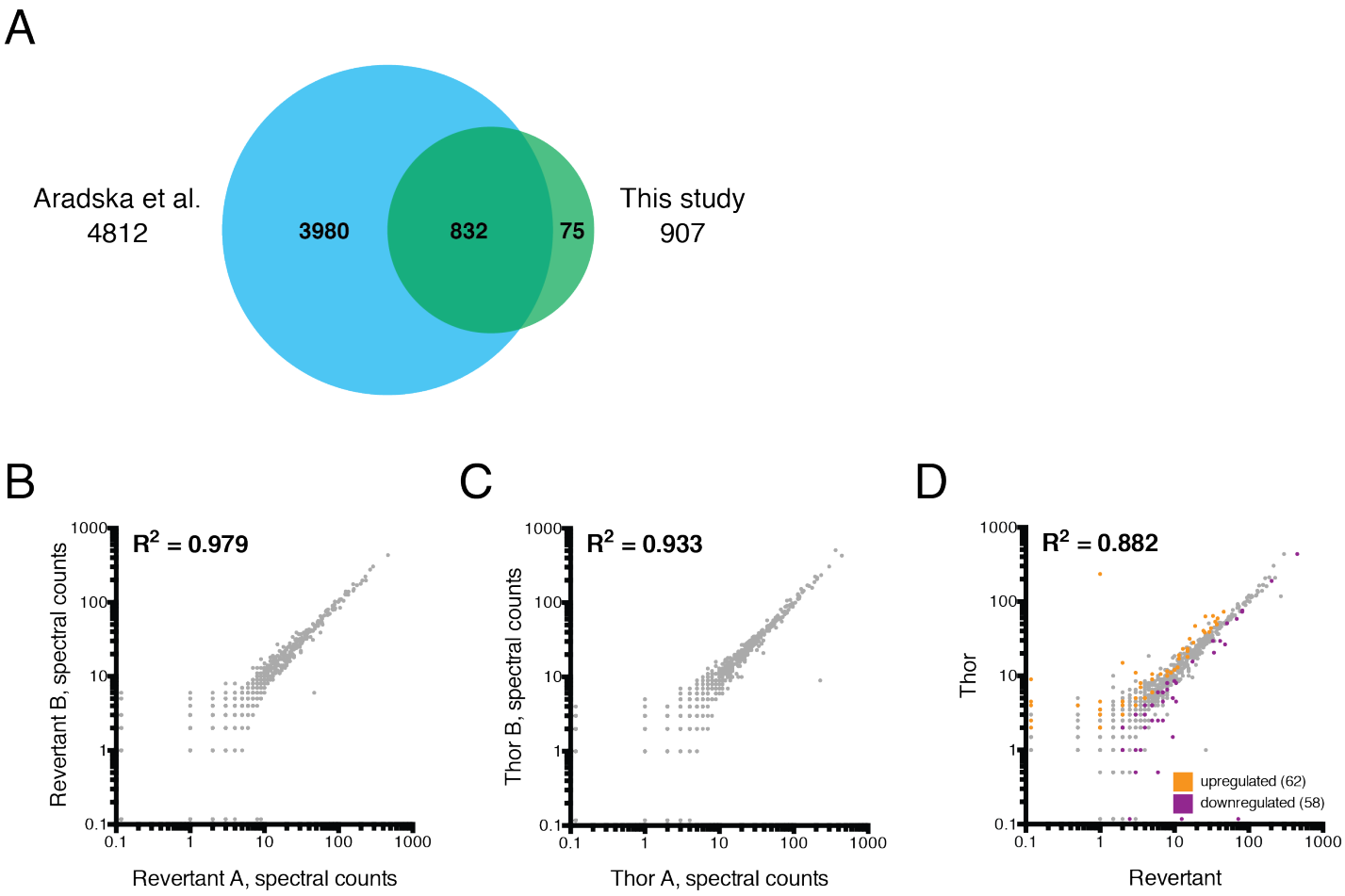


Figure 2.10 Differentially expressed proteins identified by mass spectrometry

(A) Venn diagram of proteins identified by mass spectrometry in Aradska et al. and this study. **(B and C)** Correlation plots of biological duplicates for *Revertant* and *Thor* head samples based on spectral counts. 907 proteins were identified by mass spectrometry. Quantitative analysis identified 120 proteins with $P < 0.05$ (blue) and 787 proteins with $P > 0.05$. **(D)** Scatter plot of *Revertant* and *Thor* samples in spectral counts with 62 upregulated proteins (fold change > 1.1, orange) and 58 downregulated proteins (fold change < 0.9, purple). R^2 indicates Pearson correlation.

Table 2.6 Proteins identified by mass spectrometry with significant fold changes

Identified Proteins	Gene Symbol	Significance (p-value)	Fold Change	Rev 1	Rev 2	Thor 1	Thor 2
Q961S9|Q961S9_DROME	Mtp	< 0.00010	INF	0	0	2	2
A1ZA66|A1ZA66_DROME	Strn-Mlck	< 0.00010	INF	0	0	4	4
A4UZE6|A4UZE6_DROME	Gαq	0.012	INF	0	0	8	10
Q8SZN1|Q8SZN1_DROME	CG31313	0.012	INF	0	0	4	5
Q9VSK4|Q9VSK4_DROME	CG6983	0.038	INF	0	0	3	2
P13607|ATNA_DROME	Atpα	< 0.00010	210	1	1	234	236
Q9VN44|Q9VN44_DROME	Karyβ3	0.026	7	0	1	4	4
P18053|PSA4_DROME	Prosα3	0.0081	6.6	1	3	15	15
Q95R41|Q95R41_DROME	alt	0.022	4	1	1	5	4
P12982|PP12_DROME	Pp1-87B	0.023	3.2	2	4	11	11
Q7KSQ0|Q7KSQ0_DROME	sea	0.041	3.1	1	1	3	4
Q9V3D4|IDGF2_DROME	Idgf2	0.041	3.1	1	1	3	4
Q9VW20|Q9VW20_DROME	Lon	< 0.00010	2.7	1	1	3	3
Q9V6U9|MECR_DROME	CG16935	< 0.00010	2.7	1	1	3	3
Q24388|LSP2_DROME	Lsp2	0.00079	2.2	26	26	62	64
Q9VTC3|Q9VTC3_DROME	CG6409	0.021	2.2	16	22	48	46
Q24407|ATP5J_DROME	ATPsynCF6	0.016	2	4	3	8	8
Q9VI56|Q9VI56_DROME	CG1943	0.045	2	2	2	4	5
Q9VCW6|Q9VCW6_DROME	Gclm	0.047	2	2	2	5	4
Q9V4C7|Q9V4C7_DROME	PMCA	0.0098	1.9	5	5	10	11
P91632|P91632_DROME	mub	0.00022	1.8	1	1	2	2
Q9VM19|Q9VM19_DROME	CG5171	0.00022	1.8	1	1	2	2
Q9W3Z3|Q9W3Z3_DROME	Spat	0.00022	1.8	1	1	2	2
Q9U3Z7|NH2L1_DROME	hoip	0.00022	1.8	1	1	2	2
O61231|RL10_DROME	RpL10	0.00022	1.8	2	2	4	4
P91944|P91944_DROME	eRF3	0.00022	1.8	1	1	2	2
Q0KIA8|Q0KIA8_DROME	CRAT	0.00022	1.8	1	1	2	2
P37276|DYHC_DROME	Dhc64C	0.00022	1.8	1	1	2	2
P28668|SYEP_DROME	GluProRS	0.00022	1.8	1	1	2	2
A4V0I0|A4V0I0_DROME	Pka-C1	0.027	1.8	4	3	7	7
P29845|HSP7E_DROME	Hsc70-5	0.033	1.7	14	18	32	31
Q9VA83|Q9VA83_DROME	Fer2LCH	0.042	1.7	29	37	61	67
Q9W5W8|Q9W5W8_DROME	CG9577	0.00037	1.6	5	5	9	9
Q8MKK5|Q8MKK5_DROME	kcc	0.015	1.6	13	13	24	22
Q59E58|Q59E58_DROME	zip	0.00057	1.5	3	3	5	5
P29742|CLH_DROME	Chc	0.013	1.5	11	12	19	19
Q9VW68|Q9VW68_DROME	Gabat	0.034	1.5	25	24	38	43
Q9VFF0|Q9VFF0_DROME	UQCR-C1	0.0021	1.4	38	38	60	59
Q7KRU8|Q7KRU8_DROME	Fer1HCH	0.023	1.4	16	14	23	23
Q8T477|Q8T477_DROME	kdn	0.027	1.4	48	44	76	70
Q9VNW6|Q9VNW6_DROME	CG7470	0.03	1.4	19	17	29	27
A4V4Q6|A4V4Q6_DROME	wupA	0.046	1.4	7	6	11	10
A4V439|A4V439_DROME	Tom40	0.0012	1.3	2	2	3	3
Q7KLX3|Q7KLX3_DROME	Tapδ	0.0012	1.3	2	2	3	3
P91938|TRXR1_DROME	Trxr-1	0.015	1.3	15	15	22	21
P84029|CYC2_DROME	Cyt-c-p	0.019	1.3	35	36	55	52
Q26416|CUA1_DROME	Acp1	0.027	1.3	30	32	44	44
Q9Y143|Q9Y143_DROME	CG5867	0.029	1.3	11	12	17	17
P19889|RLA0_DROME	RpLP0	0.036	1.3	8	8	11	12
Q9VWH4|IDH3A_DROME	l(1)G0156	0.038	1.3	27	24	37	40
Q2QBM1|Q2QBM1_DROME	Men	0.0012	1.2	30	29	39	39
Q4V5X9|Q4V5X9_DROME	RpL12	0.0038	1.2	3	3	4	4
P13706|GPDA_DROME	Gpdh	0.042	1.2	37	37	51	48
A0A1F4|A0A1F4_DROME	eys	0.01	1.1	8	8	10	10
A4V0I6|A4V0I6_DROME	RpL13	0.01	1.1	4	4	5	5
Q9U915|Q9U915_DROME	Adk2	0.01	1.1	4	4	5	5
P13677|KPC2_DROME	inaC	0.017	1.1	9	9	11	11
P06742|MLC1_DROME	Mlc1	0.028	1.1	15	15	18	18
A4UZB6|A4UZB6_DROME	Gαo	0.028	1.1	10	10	12	12

(Table continues)

Identified Proteins	Gene Symbol	Significance (p-value)	Fold Change	Rev 1	Rev 2	Thor 1	Thor 2
Q9VU35|Q9VU35_DROME	CG11267	0.028	1.1	5	5	6	6
Q9VKW5|Q9VKW5_DROME	CG5355	0.028	1.1	5	5	6	6
A4V2Z5|A4V2Z5_DROME	Sap47	0.048	1.1	11	11	13	13
P08255|OPS4_DROME	Rh4	0.01	0.9	2	2	2	2
Q4V3N9|Q4V3N9_DROME	Obp51a	0.01	0.9	2	2	2	2
A1Z7V8|A1Z7V8_DROME	CG1814	0.01	0.9	2	2	2	2
A4V042|A4V042_DROME	for	0.01	0.9	2	2	2	2
A4V488|A4V488_DROME	ras	0.01	0.9	2	2	2	2
Q86PE8|Q86PE8_DROME	CG5390	0.01	0.9	2	2	2	2
Q8IQG9|Q8IQG9_DROME	Adk1	0.01	0.9	2	2	2	2
Q9VLS4|Q9VLS4_DROME	CG8498	0.01	0.9	2	2	2	2
P11584|ITBX_DROME	mys	0.01	0.9	2	2	2	2
Q5BIA9|Q5BIA9_DROME	CG8613	0.01	0.9	2	2	2	2
Q9I7Q8|Q9I7Q8_DROME	Ank2	0.01	0.9	2	2	2	2
Q4V5D0|Q4V5D0_DROME	CG8329	0.01	0.9	2	2	2	2
A4V340|A4V340_DROME	CG7708	0.01	0.9	2	2	2	2
Q7K485|Q7K485_DROME	cathD	0.01	0.9	8	8	8	8
O97477|INO1_DROME	Inos	0.01	0.9	6	6	6	6
Q9VZU7|Q9VZU7_DROME	Usp5	0.01	0.9	4	4	4	4
Q0E9N2|Q0E9N2_DROME	l(2)01289	0.01	0.9	3	3	3	3
Q9V7D2|VATD1_DROME	Vha36-1	0.01	0.9	3	3	3	3
Q9VQ61|Q9VQ61_DROME	Got2	0.025	0.9	52	50	50	52
A4V0Q1|A4V0Q1_DROME	Adh	0.035	0.9	462	434	445	431
O62619|KPYK_DROME	PyK	0.0001	0.8	83	81	76	76
Q9U1K3|Q9U1K3_DROME	glob1	0.0022	0.8	7	7	6	6
A4V0H9|A4V0H9_DROME	Trx-2	0.0022	0.8	7	7	6	6
Q4QPQ0|Q4QPQ0_DROME	Aldh	0.0056	0.8	71	67	58	59
P35381|ATPA_DROME	blw	0.012	0.8	209	197	190	190
Q8IGA2|Q8IGA2_DROME	fabp	0.019	0.8	84	79	71	75
P02516|HSP23_DROME	Hsp23	0.022	0.8	18	17	16	15
Q9VZI1|Q9VZI1_DROME	Chd64	0.032	0.8	10	10	8	9
Q9VSW1|Q9VSW1_DROME	UGP	0.047	0.8	34	32	31	28
O97102|O97102_DROME	smt3	0.0011	0.7	4	4	3	3
Q8IRH0|Q8IRH0_DROME	Psa	0.0011	0.7	4	4	3	3
Q95029|CATL_DROME	Cp1	0.0015	0.7	5	5	4	4
Q9W022|Q9W022_DROME	CG8993	0.0015	0.7	5	5	4	4
Q9V3W0|Q9V3W0_DROME	UK114	0.032	0.7	10	11	8	8
Q7PL67|Q7PL67_DROME	RpL5	0.041	0.7	8	8	7	6
Q8SYA5|Q8SYA5_DROME	retinin	0.0021	0.6	41	41	30	29
Q27580|SAHH_DROME	Ahcy	0.021	0.6	7	7	4	5
A4V099|A4V099_DROME	ade2	0.014	0.5	9	10	5	5
Q9VXI1|Q9VXI1_DROME	CG9914	0.018	0.5	36	32	22	19
Q00963|SPTCB_DROME	β-Spec	0.02	0.5	45	51	26	27
A1Z7Z4|A1Z7Z4_DROME	CG1648	0.035	0.5	10	9	4	6
O17452|O17452_DROME	Peritrophin-A	0.00041	0.4	2	2	1	1
Q24150|Q24150_DROME	Nap1	0.00041	0.4	2	2	1	1
Q4V679|Q4V679_DROME	CG17242	0.00041	0.4	2	2	1	1
A1Z6R7|A1Z6R7_DROME	CG30158	0.00041	0.4	2	2	1	1
A1Z8Y2|A1Z8Y2_DROME	Cpr49Aa	0.00041	0.4	4	4	2	2
Q95SI7|Q95SI7_DROME	CG6028	0.014	0.4	6	6	3	2
A4V452|A4V452_DROME	Nrg	0.025	0.4	5	5	2	3
Q9W3L4|Q9W3L4_DROME	CG2233	0.047	0.4	12	9	4	5
Q9VW26|OAT_DROME	Oat	0.00025	0.3	3	3	1	1
Q95NS3|Q95NS3_DROME	TotA	0.0089	0.3	7	7	3	2
P10351|XDH_DROME	ry	0.03	0.3	4	3	1	1
Q9V3S0|CP4G1_DROME	Cyp4g1	0.029	0.1	3	3	0	1
Q9W0J9|Q9W0J9_DROME	CG9119	0.031	0.1	11	8	1	2
Q8IN43|Q8IN43_DROME	TotC	0.0065	0.07	6	6	0	1
P61857|TBB2_DROME	βTub85D	0.00063	0	71	73	0	0
Q86P20|Q86P20_DROME	CG34370	0.034	0	3	2	0	0
Q86NN5|Q86NN5_DROME	ND-39	0.042	0	10	15	0	0

Column 1, Accession numbers of proteins in quantitative analysis of mass spectrometry data, **column 2**, gene symbol corresponding to the protein, **column 3**, significance determined with t-test, only showing those with $P < 0.05$, **column 4**, fold change of protein levels, **column 5, 6**, spectral counts in biological duplicates of *Revertant* head samples, **column 7, 8**, spectral counts in biological duplicates of *Thor* head samples.

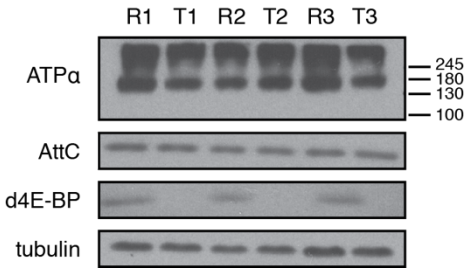


Figure 2.11 Immunoblot of ATPα does not show elevated levels in *Thor* fly head lysate

Immunoblot from *Revertant* and *Thor* fly head lysates with tubulin as the loading control. 3 biological replicates of 30 fly heads was used. R denotes *Revertant* and T denotes *Thor*.

2.5.2 Materials and Methods

2.5.2.1 Mass spectrometry

Heads from 3–5 day old male *Revertant* and *Thor* flies were collected, flash frozen and pulverized in liquid nitrogen. Head lysate was prepared using a lysis buffer (8 M urea, 50 mM Tris-HCl pH 8.1, 1 X Halt™ Protease Inhibitor Cocktail, EDTA-Free (Thermo Scientific)) and incubated on ice for 5 min. Samples were sonicated at 3 bursts of 30 seconds at max amplitude at 4 °C and placed on ice for 1 min in between sonication. Lysate was clarified by centrifugation at 15,000 g for 10 min at 4 °C. The supernatant was collected and protein concentration was measured by Bradford protein assay (Bio-Rad). Samples were submitted as biological duplicates for LC-MS/MS analysis at the Proteomics core facility of Institut de recherche en immunologie et en cancérologie (IRIC). Results were then analysed on Scaffold 4.8.4. Proteins were identified using the Drome database. The peptide threshold was set at 80% minimum and the protein thresholds at 80% minimum and 2 peptides minimum. Quantitative analysis of differences in protein level was performed by T-test with Benjamini-Hochberg multiple test correction with significance level of $P < 0.05$.

2.5.2.2 Immunoblotting

30 heads from 3–5 day old male flies per replicate were collected and homogenized in lysis buffer (8 M urea, 0.02 M HEPES pH 7.4, 0.1 M KCl, 2 mM EDTA, 0.5 % Triton X-100, 1 mM DTT, 2 mM PMSF, 1 X Halt™ Protease Inhibitor Cocktail, EDTA-Free (Thermo Scientific)). Samples were incubated on ice for 5 min and clarified by centrifugation at 15,000 g for 10 min at 4 °C. The supernatant was collected and protein concentration was measured by Bradford protein assay (Bio-Rad). 30 µg of protein were loaded per lane on 9–12.5 % polyacrylamide gels then transferred to 0.2 µm nitrocellulose membranes (Bio-Rad). Proteins were blocked in 5 % BSA in TBS-T (10 mM Tris pH 7.6, 150 mM NaCl, 0.1 % Tween 20) for 1 h at room temperature, incubated with primary antibodies diluted in 5 % milk TBS-T overnight at 4 °C, and finally with secondary antibodies also diluted in 5 % BSA TBS-T for 1 h at room temperature. Blots were quantified using ImageJ. The following primary antibodies were used: rabbit anti-d4E-BP (this study, 74-3, 1:500), rat anti-AttC ((Kim and Kim-Ha, 2006), 1:1000), mouse anti-ATPα

(Developmental Studies Hybridoma Bank, a5, 1:5000) and mouse anti-α-tubulin (Sigma-Aldrich, T9026, 1:5000)

Connecting Text

The research presented in Chapter 2 investigated mRNAs that are translationally regulated downstream of d4E-BP and features present within the 5'UTRs of transcripts that are translationally upregulated in *d4E-BP*^{null} adult heads. The work presented in Chapter 3 explores the intimate relationship shared by translational control and RNA localization and the regulation conferred by the 3'UTR of an mRNA. In this study, we wished to examine different aspects of the localization of germ plasm RNAs in the early *Drosophila* embryo. To this end, we assessed the state of translation in the early embryo of 3 RNAs with essential roles in germline biology, dissected the 3'UTRs of *pgc* and *gcl* to map *cis*-elements that function in localization, performed a phylogenetic analysis of transcripts that share a common pattern of localization, and determined the presence of a specific type of localization element, the homotypic clustering motif, within their 3'UTR sequences.

CHAPTER 3

Analysis of mRNAs that localize to RNA islands in the *Drosophila* embryo

3.1 Introduction

Subcellular localization of mRNAs is essential for establishing a spatio-temporal control of protein expression to ensure proper development of the *Drosophila* embryo. The localization of mRNAs to the pole cells is far more prevalent than previously known with hundreds identified to date by ongoing large-scale screens; however, a small subset of those RNAs follow a specific mode of localization where they accumulate around posterior nuclei prior to cellularization, forming "RNA islands" (Lecuyer et al., 2007; Tomancak et al., 2007). Some notable RNAs that localize in this manner, such as *polar granule component* (*pgc*), *germ cell-less* (*gcl*) and *nanos* (*nos*), have well defined roles in germline development (Gavis and Lehmann, 1992; Hanyu-Nakamura et al., 2008; Lerit et al., 2017). The germline-specific knockdown of mRNAs that accumulate in RNA islands resulted in defects in embryonic development, including patterning and formation of pole cells, and progression of oogenesis, suggesting roles for the genes that encode the mRNAs in those processes (Liu and Lasko, 2015).

Transcripts destined for RNA islands are initially synthesized in ovarian nurse cells to be deposited into the oocyte (Roth, 2001). Through the diffusion and entrapment mechanism, ribonucleoproteins (RNPs) containing a single mRNA diffuse during cytoplasmic streaming and are then trapped within the germ plasm to concentrate at the posterior of the oocyte (Forrest and Gavis, 2003; Little et al., 2015). Once incorporated into germ granules, the RNAs act as seeds to self-recruit additional transcripts to form homotypic clusters (Niepielko et al., 2018). During early embryogenesis, germ granules are transported to nearby posterior nuclei along astral microtubules to ensure the partitioning of germ plasm components to daughter nuclei as nuclear division progresses (Lerit and Gavis, 2011).

Germ granules are composed of a core set of proteins including Oskar (Osk), Tudor (Tud) and Vasa (Vas) (Gao and Arkov, 2013). *osk* RNA is transported to the posterior of the oocyte in a kinesin-dependent manner where it is translated to produce Osk, which then acts to nucleate germ plasm assembly by recruitment of other proteins and subsequent incorporation of additional mRNAs (Ephrussi and Lehmann, 1992). While germ plasm proteins are evenly distributed within the granules, homotypic clusters of mRNAs occupy distinct positions (Trcek et al., 2015).

The localization of mRNAs is directed by *cis*-acting regulatory elements and *trans*-acting factors (Martin and Ephrussi, 2009). *Cis*-acting elements in the mRNA have regulatory functions, including translational control and localization, and they are predominantly located in the untranslated region (UTR) (Lasko, 2009). Localization elements can be primary sequences or secondary structures that are recognized by RNA-binding proteins to form RNP complexes (Hamilton and Davis, 2007; Martin and Ephrussi, 2009). Furthermore, localization signals can function redundantly, act additively, and be conserved in a related species in sequence and in function, which is the case for *nos* (Gavis et al., 1996). The 3'UTRs of several germ plasm mRNAs that accumulate as RNA islands are sufficient to direct posterior localization in the early embryo (Rangan et al., 2009). The localization of these mRNAs is achieved by the combinatorial effects of two types of localization signals: one that targets the RNA to the germ plasm, and another for self-recruitment to allow homotypic cluster growth (Eagle et al., 2018). While numerous germ plasm RNAs localize in a similar manner, a localization element that is shared by all of their 3'UTRs has yet to be identified although a homotypic clustering motif is conserved in *pgc*, *gcl* and *nos* (Eagle et al., 2018; Lecuyer et al., 2007).

Here, we confirmed that the majority of *pgc*, *gcl* and *nos* is found in RNPs with little translation in the early embryo. We used deletion mutation analysis to delineate the localization elements of *pgc* and *gcl*. We find that the *pgc* localization signal is located in the distal region of the 3'UTR where there are sequences that are necessary and sufficient for posterior localization. The *gcl* 3'UTR possesses multiple redundant localization signals where two distinct regions are each sufficient to direct localization. We show that the protein components of the polar granules that localize in the same manner as *nos* and other RNA island transcripts, Osk, Tud and Vas, have conserved localization patterns in other *Drosophila* species. We also investigated the localization of maternal RNAs and found that most of the transcripts tested had conserved RNA island localization across *Drosophila* species. Furthermore, the homotypic clustering motif is found in the 3'UTRs of almost all RNA island transcripts, and those of *pgc*, *gcl* and *nos* appear to be conserved across *Drosophilids*.

3.2 Results

3.2.1 *pgc*, *gcl* and *nos* mRNAs are mostly found in subpolysomal fractions in the early embryo

To investigate the translational activity of maternally-inherited germ plasm mRNAs, we performed polysome profiling on early (0–2 hour) *yw* embryos, a technique that sediments mRNAs on a sucrose density based on the number of bound ribosomes (**Fig. 3.1A**). We examined the distributions of *pgc*, *gcl*, and *nos* across the sucrose fractions using RT-qPCR. Only a small proportion of those transcripts were present in polysome fractions (8–14) at 25.7%, 32.6% and 24.9% for *pgc*, *gcl* and *nos* respectively (**Fig. 3.1B**). Conversely, the majority of those germ plasm transcripts was found in subpolysomal fractions with more than 40% of *pgc*, *gcl* and *nos* mRNAs in fractions 1–3 at 43.4%, 43.7% and 40.8%, respectively (**Fig. 3.1B**). These findings support the translational regulation of these RNAs where their onset of translation once they reach the germ plasm is varied (Rangan et al., 2009). Notably, the *nos* mRNA distribution is consistent with previous findings that showed only a small proportion of *nos* was found in the polysome fractions (Qin et al., 2007). Although another study found that 53% of *nos* mRNA sedimented with polysomes in the heavier fractions, twice as much as shown here, they also demonstrate that only about 4% of posterior localized *nos* is translated in the preblastoderm embryo, suggesting there are additional mechanisms of translational control of *nos* (Clark et al., 2000). Collectively, these results confirm that the majority of those transcripts are found in RNP complexes, sequestered from translational machinery, and there is little translation of the germ plasm mRNAs, *pgc*, *gcl* and *nos*, in the early embryos.

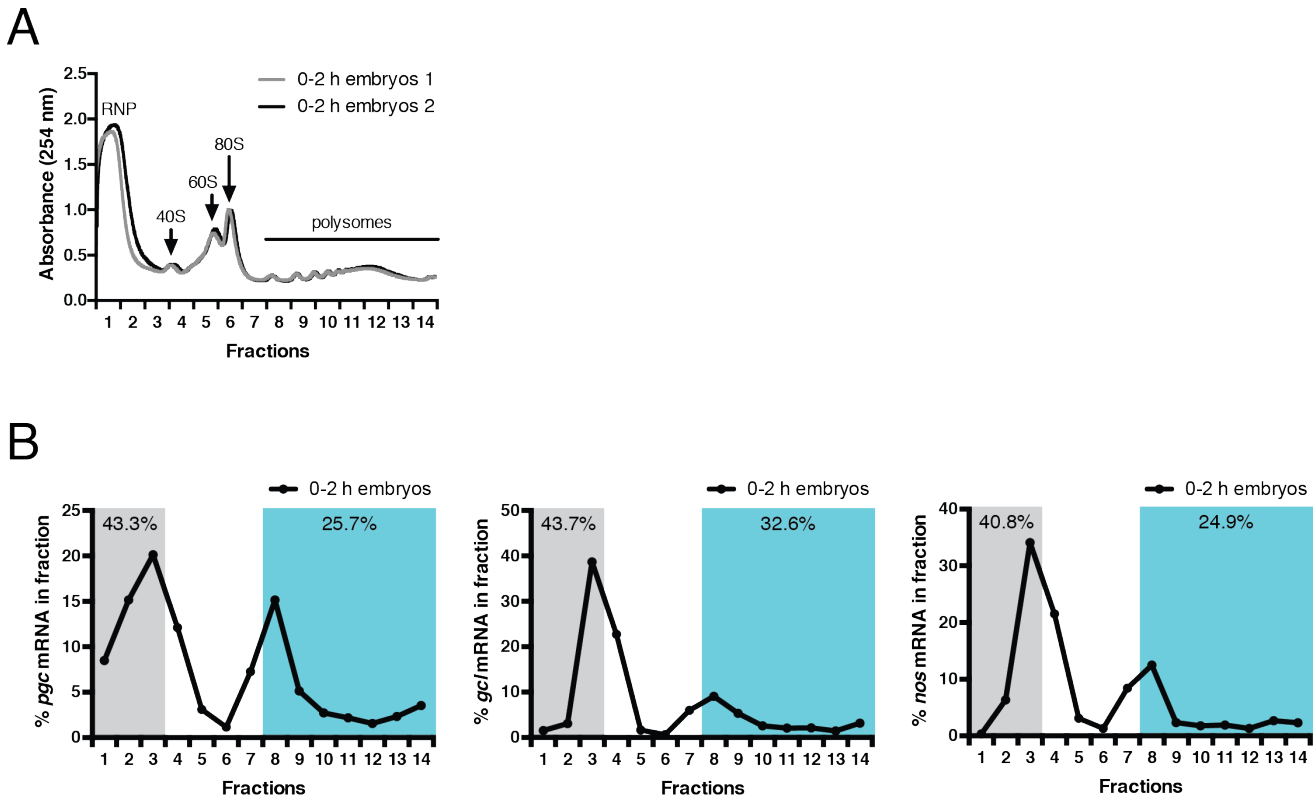


Figure 3.1 *pgc*, *gcl* and *nos* mRNA are found predominantly in subpolysomal fractions

(A) Absorbance at 254 nm was continuously measured for polysome profiles of two independent biological replicates of 0–2 hour (h) embryo lysates resolved on a 10–50% sucrose gradient. **(B)** RT-qPCR on RNA extracted from 0–2 h embryo polysome fractions. The mRNA distributions of *pgc*, *gcl*, and *nos* are shown as percentage of total mRNA, averaged from two independent fractionations, detected in all 14 fractions as determined using the standard curve method. Grey boxes indicate the fractions that correspond to RNPs (1–3) and the blue boxes indicate the polysome fractions (8–14). The percentages shown are the sum of the percentages of mRNA within the fractions highlighted by the boxes.

3.2.2 The distal region of the *pgc* 3'UTR contains a localization signal that is necessary and sufficient for posterior localization

To identity the localization signal within the *pgc* 3'UTR, we generated UAS-eGFP fly lines expressing the coding sequence of eGFP fused to the *pgc* 3'UTR containing various deletions. Expression of these mRNAs was controlled by the germline-specific driver, *nos*-Gal4-VP16 (**Fig. 3.2A, B**). Transgenes where *pgc* 3'UTR sequences were fused to *alpha tubulin 84B* (*α-tub*) 3'UTR were also generated to produce *eGFP-pgc-α-tub-3'UTR* or *eGFP-α-tub-pgc3'UTR* (**Fig. 3.2A, B**). *α-tub* is not an mRNA that is enriched in the posterior and its 3'UTR was used to provide additional length to *pgc* 3'UTR fragments shorter than 150 nucleotides (nt) in length in order to test their sufficiency in directing localization in a similar approach to map the *nos* localization element (Gavis et al., 1996). The full length *pgc* 3'UTR transgenic RNA, *eGFP-pgc*, localized to the pole plasm, RNA islands and pole cells in the posterior embryo, as expected of *pgc* mRNA (**Fig. 3.2C**). Interestingly, *eGFP-pgcΔ168-401* failed to localize to the posterior of the embryo (**Fig. 3.2C**). Furthermore, *eGFP-pgc(+238-401)-α-tub* does indeed localize to the posterior (**Fig. 3.2C**). These results suggest that the sequence between 238-401 is necessary and sufficient for RNA localization to the pole plasm and pole cells. To further examine the sequence of 238-401, finer mapping was performed with additional transgenic constructs. *eGFP-pgc(Δ238-321)* and *eGFP-pgc(Δ291-360)* both resulted in posterior localization (**Fig. 3.2C, bottom 2 panels on left**). The reciprocal experiment to test the sufficiency of the 238-321 and 291-360 sequences showed that those transgenic RNAs did not localize to the posterior (**Fig. 3.2C, bottom 2 panels on right**). The sequence from 361-401 is therefore a likely candidate for the minimal localization element that is necessary and sufficient for pole plasm and pole cell localization. While it is possible that there is a single localization element in the *pgc* 3'UTR within the 238-401 sequence, these observations do not rule out the possibility of a multipartite localization signal where several elements found throughout the 238-401 sequence are individually insufficient to localize the transcript but instead must all work in concert.

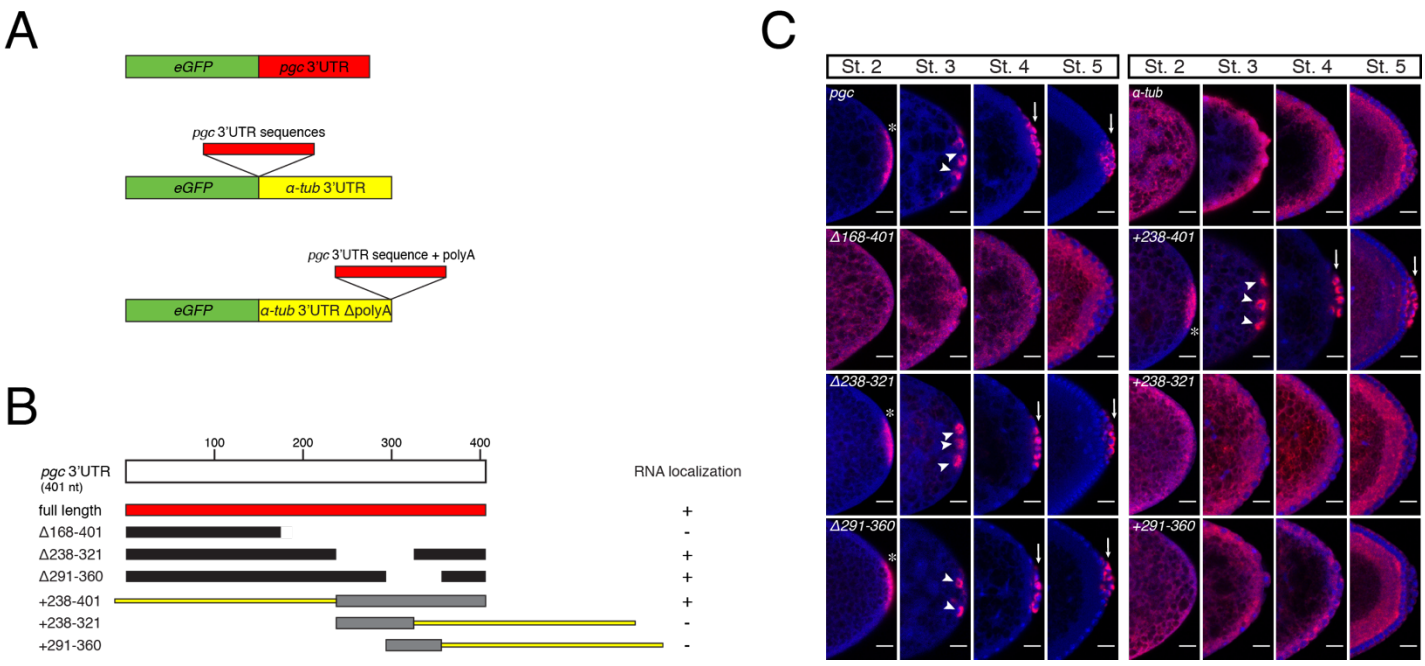


Figure 3.2 Region 238-401 of the *pgc* 3'UTR is necessary and sufficient for posterior localization

(A) Schematic diagram of *pgc* transgenic constructs where the *pgc* 3'UTR (red), including full length and deletion mutants, were cloned downstream of the *eGFP* coding region (green). Short sequences less than 100 nt in length were cloned upstream of the α-tubulin 3 'UTR (yellow). The *pgc* 3'UTR sequence containing the polyA signal (+238-401) was cloned downstream of the α-tubulin 3'UTR. **(B)** Schematic diagram of deletion mutation analysis of *pgc* 3'UTR. Full length *pgc* 3'UTR is shown as the red bar. Black bars correspond to *pgc* 3'UTR deletion mutants. Short 3'UTR sequences, in grey, that were fused to *α-tubulin* 3'UTR, in yellow. The constructs were tested for RNA localization in embryos using FISH where (+) denotes posterior localization and (-) denotes a lack of enrichment in the posterior. **(C)** Localization of the *eGFP*-reporter mRNA (red) was detected by FISH using an antisense *eGFP* riboprobe as indicated by asterisks in stage 2, arrowheads in stage 3 and arrows in stage 4–5 embryos. Transgenes were expressed by *nos*-Gal4-VP16 in a wild type genetic background. DNA was stained with DAPI (blue). All images are oriented with posterior to the right. The scale bars represent 20 μm.

3.2.3 The *gcl* 3'UTR contains functionally redundant localization signals

To identify the localization signal in the *gcl* 3'UTR, we used a similar deletion mutation analysis to the one used on the *pgc* 3'UTR (**Fig. 3.3A, B**). The full length *eGFP-gcl-3'UTR* reporter transcript localized to the pole plasm in stage 2 embryos, RNA islands in stage 3 embryos and pole cells in stage 4 and 5 embryos, as expected of *gcl* mRNA (**Fig. 3.3C, asterisk, arrowheads and arrows**). Deletions in increments of 100 nt were made along the *gcl* 3'UTR with the exception of the final 124 nt at the 3' end (**Fig. 3.3B**) and all of them localized as wild type (**Fig. 3.3C, left panels**), suggesting these individual regions are not necessary for posterior localization. While most *eGFP-gcl3'UTR* transgenic RNAs are unlocalized in the embryo (**Fig. 3.3C**), *eGFP-gcl(+101-200)* and *eGFP-gcl(+201-300)* recapitulated the localization of the full length *eGFP-gcl* transgenic RNA (**Fig. 3.3C**). Transgenes that overlap with the 101-200 region were generated: Δ70-169 and +70-169 (**Fig. 3.3B**). *eGFP-gcl(Δ70-169)* RNA localized to the posterior, but *eGFP-gcl(+70-169)-α-tub* RNA remained unlocalized in the embryo (**Fig. 3.3C**). This suggests that the 101-169 sequence is dispensable for posterior localization and further maps the essential and possibly sufficient localization sequence to 170-200 region. Additionally, transgenes that overlap with the 201-300 region were also generated: Δ251-350 and +251-350 (**Fig. 3.3B**). *eGFP-gcl(Δ251-350)* localized to the posterior while *eGFP-gcl(+251-350)* did not (**Fig. 3.3C**). These results indicate that the region 201-250 contains a sequence that is important for posterior localization; however, the sequence of 251-300 likely does not contribute to the localization element. Taken together, our data demonstrates that there is not a single localization element found within the *gcl* 3'UTR, rather the presence of two redundant localization signals that are each sufficient but not required for posterior localization in the embryo.

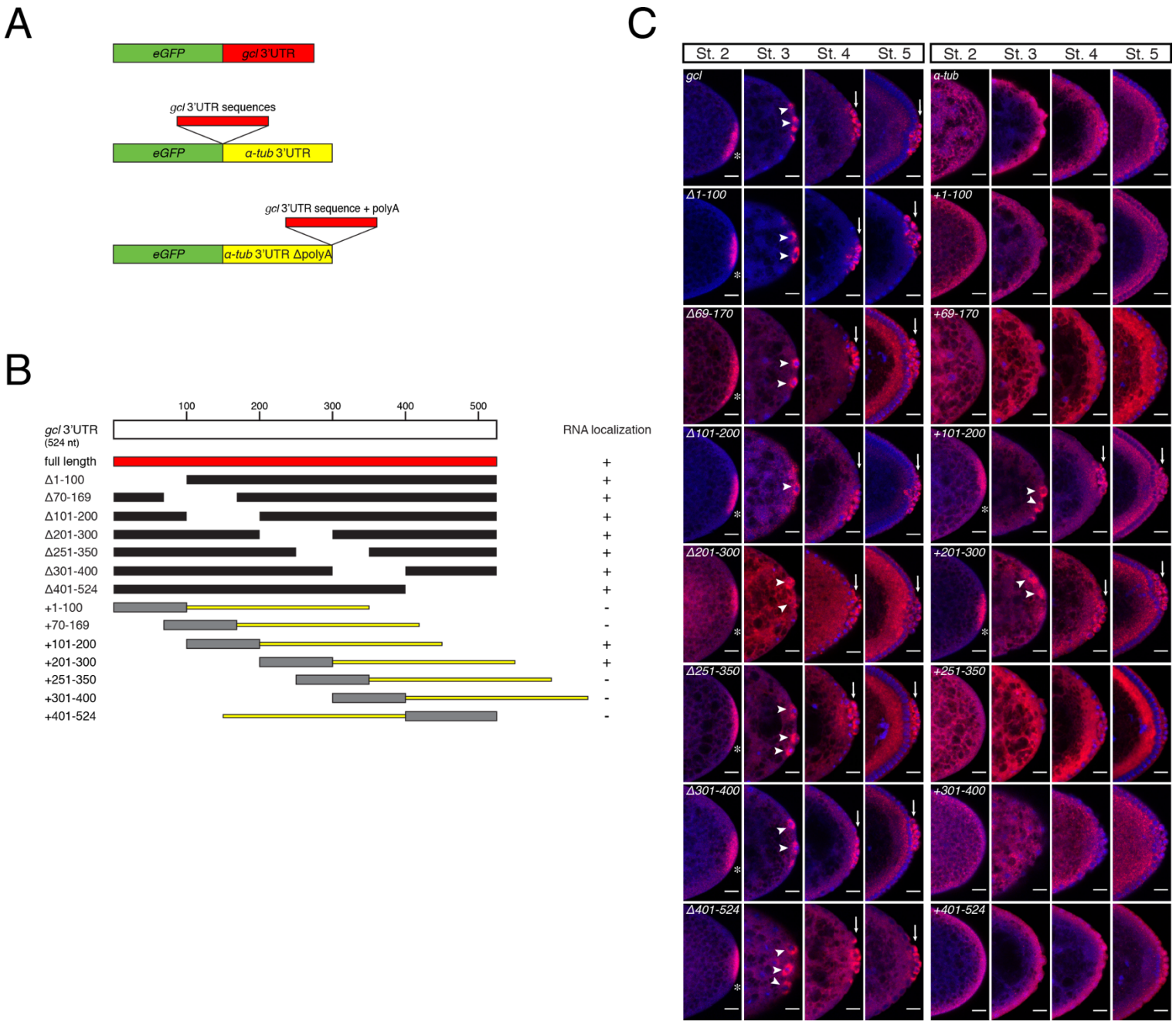

A
eGFP
gcl 3'UTR
gcl 3'UTR sequences
eGFP
α-tub 3'UTR
gcl 3'UTR sequence + polyA
eGFP
α-tub 3'UTR ΔpolyA
B
100
200
300
400
500
gcl 3'UTR
(524 nt)
RNA localization
full length +
Δ1-100 +
Δ70-169 +
Δ101-200 +
Δ201-300 +
Δ251-350 +
Δ301-400 +
Δ401-524 +
+1-100 -
+70-169 -
+101-200 +
+201-300 +
+251-350 -
+301-400 -
+401-524 -
C
St. 2
St. 3
St. 4
St. 5
St. 2
St. 3
St. 4
St. 5
gcl
α-tub
Δ1-100
+1-100
Δ69-170
+69-170
Δ101-200
+101-200
Δ201-300
+201-300
Δ251-350
+251-350
Δ301-400
+301-400
Δ401-524
+401-524

Figure 3.3 The *gcl* 3'UTR contains multiple localization signals

(A) Schematic diagram of *gcl* transgenic constructs where the full-length sequence and deletion mutants of *gcl* 3'UTR (red) were cloned downstream of the cloned downstream of the *eGFP* coding region (green). Short sequences of the *gcl* 3'UTR were cloned upstream of the *α-tubulin* 3'UTR (yellow), with the exception of the terminal fragment containing the polyA signal (+401-524), which was cloned downstream. **(B)** Schematic diagram of deletion mutation analysis of *gcl* 3'UTR. Full length *gcl* 3'UTR is shown as the red bar. Deletion mutants of *gcl* 3'UTR are shown as black bars. The corresponding 3'UTR fragments are shown in grey and those were fused to the *α-tubulin* 3'UTR shown in yellow. The constructs depicted were tested for RNA localization in the posterior region of the embryos using FISH as indicated by (+) for localization and (-) for a lack of enrichment in the posterior. **(C)** Localization of the *eGFP* mRNA (red) in embryos of transgenic flies is indicated by asterisks in stage 2, arrowheads in stage 3 and arrows in stage 4–5 embryos. Transgenes were expressed by *nos*-Gal4-VP16 in a wild type genetic background. DNA was stained with DAPI (blue). All images are oriented with posterior to the right. The scale bars represent 20 μm.

3.2.4 *nos* RNA co-localization with polar granule components, Osk, Tud and Vas is conserved in *Drosophila*

Next, we considered the distribution of core polar granule components, Vas, Osk, and Tud proteins, alongside *nos* mRNA, one of the transcripts that accumulate in RNA islands, in the early embryo using a combined immunostaining and FISH approach. A previous study showed that polar granules containing Vas or Osk aggregate around posterior nuclei prior to pole bud formation in a similar manner as *nos* mRNA (Lerit and Gavis, 2011). As expected, *nos*, Vas and Osk co-localize to the pole plasm in the form of a crescent, then as circular structures surrounding posterior nuclei followed by incorporation into pole cells (**Fig. 3.4A**). Tud also follows the same localization pattern (**Fig. 3.4B**). We wanted to establish if the composition of polar granules is conserved in *Drosophilids* by extending the analysis of the subcellular localization of these polar granule components in embryos to two other *Drosophila* species, *D. simulans* and *D. virilis*, which have diverged from *D. melanogaster* 5.4 and 62.9 million years ago, respectively (Tamura et al., 2004). The localization pattern of *nos* mRNA, Vas, Osk and Tud observed in the early stages of embryogenesis of *D. simulans* and *D. virilis* show remarkable similarities to that observed in *D. melanogaster* embryos (**Fig. 3.4A, B**). These results suggest that the coordinated active transport of polar granules with posterior nuclei division is an evolutionary conserved mechanism, highlighting the importance of their proper segregation into pole cells for germline development. While the proteins is homogenously distributed in polar granules, as shown by the high degree of co-localization between Vas with Osk and Tud, there is a heterogeneous distribution of germ plasm mRNAs within polar granules (Trcek et al., 2015). To determine the spatial organization of the core proteins and RNAs within the granules of *D. simulans* and *D. virilis*, a high-resolution view of the granules using single-molecule FISH (smFISH) and structured illumination microscopy (SIM) will be needed as it was done in *D. melanogaster* embryos (Trcek et al., 2015).

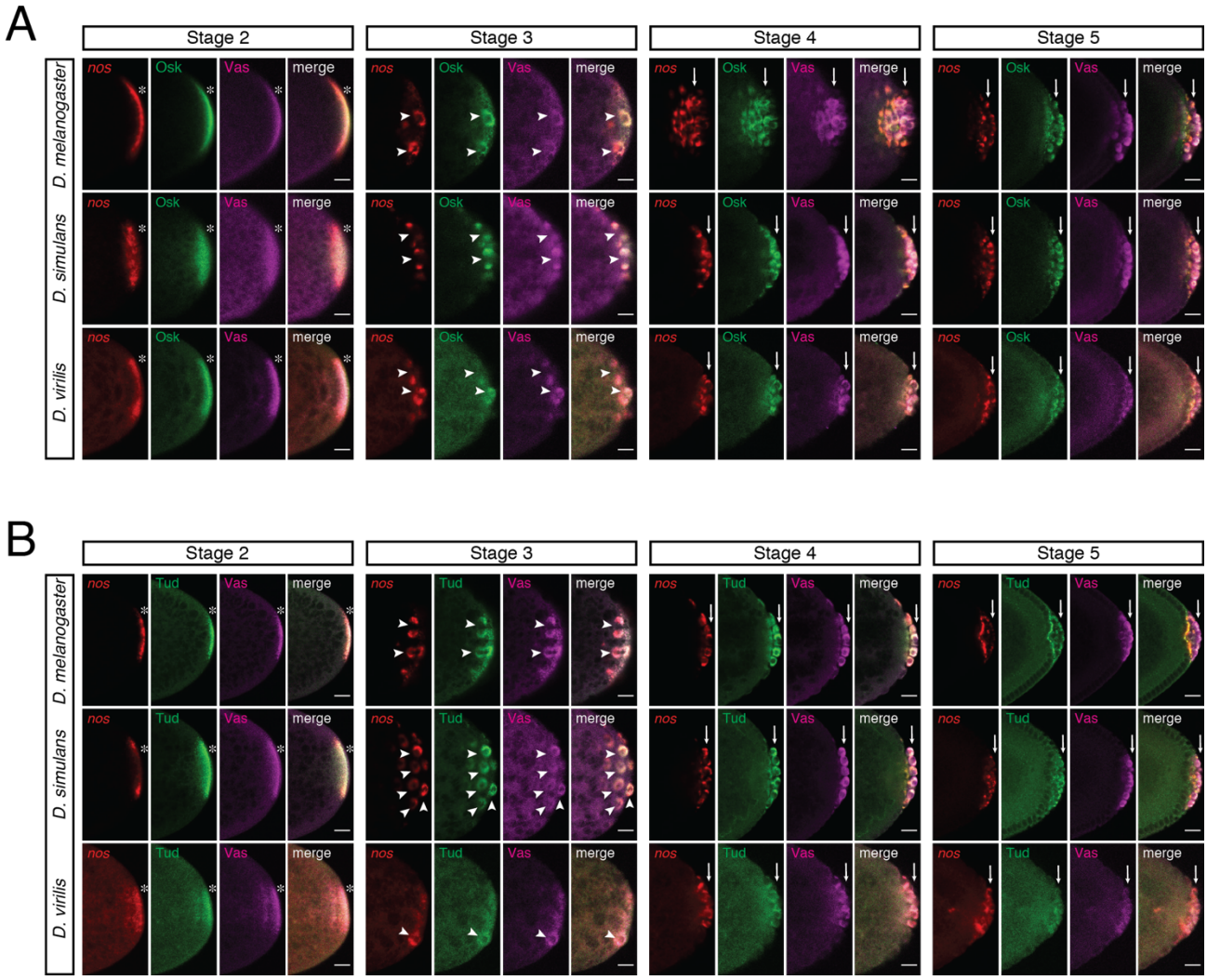


Figure 3.4 Co-localization of *nos* RNA with polar granule components, Osk, Tud and Vas is conserved in *Drosophila*

(A and B) FISH for *nos* (red) and immunofluorescence for Osk or Tud (green) and Vas (magenta) in stage 2–5 embryos of wild type *D. melanogaster*, *D. simulans* and *D. virilis*. *nos* mRNA and Osk, Tud and Vas localize to the posterior cortex in stage 2 (asterisks), surrounding posterior nuclei in stage 3 (arrowheads) and to pole cells in stage 4 and 5 (arrows). In all images, posterior is to the right. The scale bars represent 20 μm.

3.2.5 Posterior localization of mRNAs is conserved in *Drosophila*

Numerous RNAs that localize to the pole cells of posterior embryo have been identified by two large-scale ongoing FISH screens (Lecuyer et al., 2007; Tomancak et al., 2007; Wilk et al., 2016). Prior to incorporation into pole cells, a small subset of posterior localizing mRNAs enriches around posterior nuclei as transient circular formations, also known as RNA islands, in the stage 3 embryo. To date, 55 mRNAs that localize as RNA islands have been identified by Fly-FISH (Lecuyer et al., 2007; Wilk et al., 2016). Given that some RNA island localizing mRNAs have well characterized roles in germline biology, we sought to gain insight into their localization by determining if it is conserved across other *Drosophila*, which would be indicative of their importance in germ cell development and posterior patterning. We randomly selected 14 candidate mRNAs that localize as RNA islands for further analysis in *D. simulans* and *D. virilis* embryos using the FISH technique. Not surprisingly, all 14 mRNAs tested in *D. simulans* embryos have comparable localization patterns to that observed in *D. melanogaster* as they are closely related species (**Fig. 3.5A and 3.5B**). A high proportion of those RNAs also have conserved posterior localization in *D. virilis* embryos (**Fig. 3.5A**). The 11 mRNAs with conserved localization in *D. simulans* and *D. virilis* are *pgc*, *gcl*, *nos*, *CycB*, *exu*, *Tao*, *eIF5*, *Pi3K21B*, *jvl*, *pum* and *Bsg25D* (**Fig. 3.5A and 3.5C**). The localization of *Tlk*, *bru1* and *gwl* is ubiquitous in the early *D. virilis* embryo with no particular enrichment in the posterior that was consistently observed across embryos in those samples (**Fig. 3.5B**). We mainly used the localization of the transcript to RNA islands in the stage 3 embryo because it is an obvious and striking feature of this particular type of posterior localization. In negative controls, including the *α-tub* FISH (**Fig. 3.2C and 3.3C**) and no probe control (data not shown), fluorescence is slightly detectable in the pole plasm of stage 2 embryos and in the pole cells of stage 4 and 5 embryos. Taking into consideration that posterior localization of the RNA was scored by eye, it is possible that the posterior localization of *Tlk*, *bru1* and *gwl* in *D. virilis* embryos was missed using this qualitative approach and a more quantitative one, such as smFISH, would be more appropriate. Our results raise the possibility that many more mRNAs that localize as RNA islands in *D. melanogaster* have conserved localization patterns in other *Drosophila* embryos and implies that their function, which is dependent on their localization, is likely also conserved.

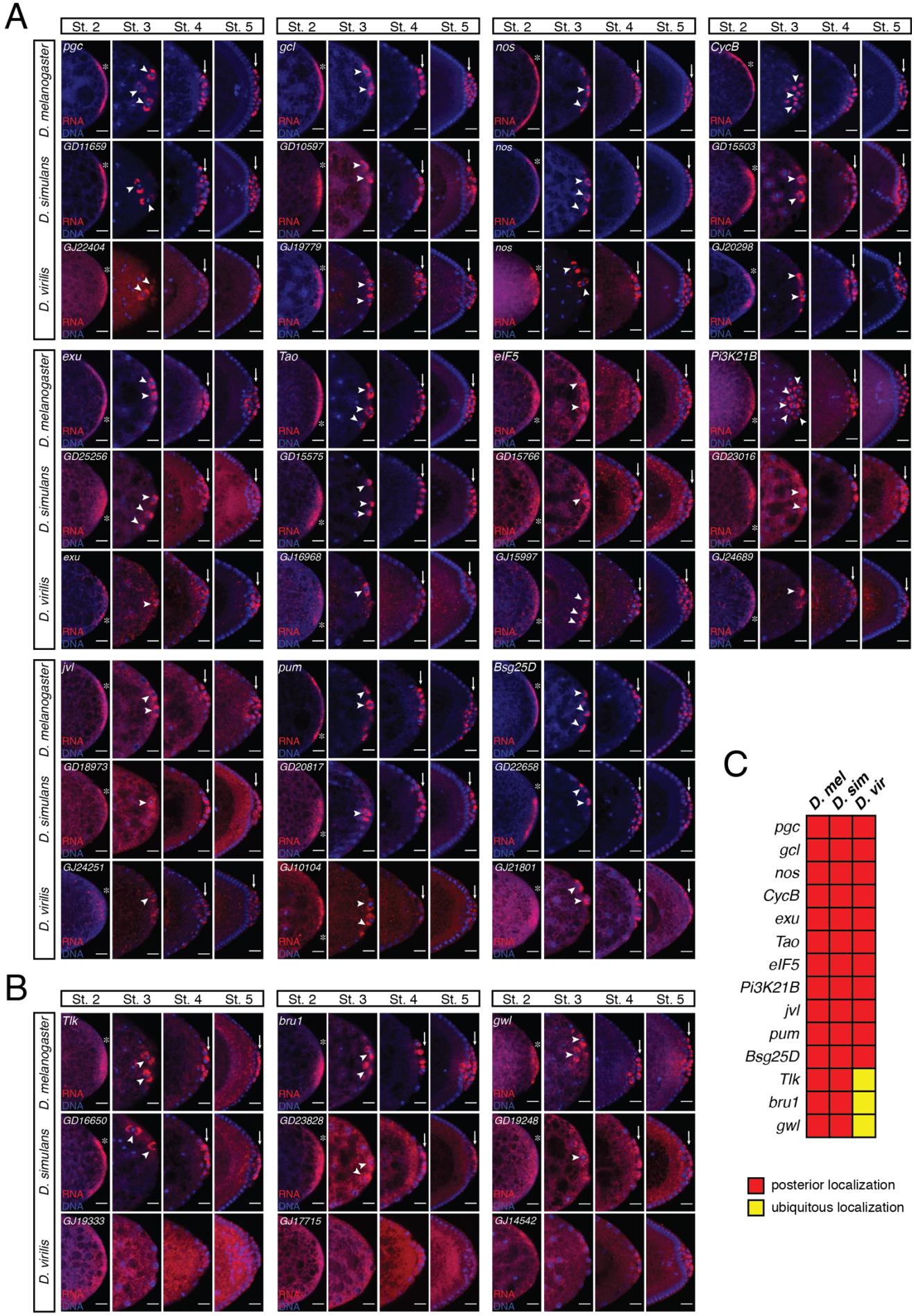

A
St. 2
St. 3
St. 4
St. 5
D. melanogaster
D. simulans
D. virilis
pgc
gcl
nos
CycB
GD11659
GD10597
nos
GD15503
GJ22404
GJ19779
nos
GJ20298
exu
Tao
eIF5
Pi3K21B
GD25256
GD15575
GD15766
GD23016
exu
GJ16968
GJ15997
GJ24689
jvl
pum
Bsg25D
GD18973
GD20817
GD22658
GJ24251
GJ10104
GJ21801
RNA
DNA
B
Tlk
bru1
gwl
GD16650
GD23828
GD19248
GJ19333
GJ17715
GJ14542
C
D. mel
D. sim
D. vir
pgc
gcl
nos
CycB
exu
Tao
eIF5
Pi3K21B
jvl
pum
Bsg25D
Tlk
bru1
gwl
posterior localization
ubiquitous localization

Figure 3.5 Conserved RNA localization to pole plasm and pole cells in *Drosophila*

(A) RNA localization was detected by FISH in pole plasm and pole cells of stage 2–5 embryos of *D. melanogaster*, *D. simulans* and *D. virilis* (RNA in red and DNA in blue). Localization is denoted by asterisks in stage 2, arrowheads in stage 3 and arrows in stage 4–5 embryos. RNAs with conserved posterior localization in *D. simulans* and *D. virilis* are *pgc*, *gcl*, *nos*, *CycB*, *exu*, *Tao*, *eIF5*, *Pi3K21B*, *jvl*, *pum*, and *Bsg25D*. **(B)** Localization of *Tlk*, *bru1* and *gwl* in pole plasm and pole cells is conserved in *D. simulans* but not in *D. virilis* (mRNA in red and DNA in blue). RNAs with conserved posterior localization in *D. simulans* but not in *D. virilis* are *Tlk*, *bru1* and *gwl*. **(C)** Summary of RNA localization in *Drosophila* embryos where posterior localization is indicated in red and ubiquitous localization is indicated in yellow. All images are oriented with the posterior to the right. The scale bars represent 20 μm.

3.2.6 The homotypic clustering motif is present in the majority of RNA island localizing transcripts

Due to the distinct pattern of localization that this group of mRNAs undergo, we considered the presence of a consensus *cis*-acting localization element amongst these mRNAs which can exist as a secondary structure or primary sequence that is shared. A recent study demonstrated that the localization of germ plasm transcripts requires targeting of the mRNAs to germ plasm RNPs followed by their homotypic clustering, and a 6-nt motif that regulates the self-assembly of mRNAs into homotypic clusters was identified within the 3'UTRs of *pgc*, *gcl* and *nos* (Eagle et al., 2018). Based on the sequences identified in *pgc*, *gcl* and *nos,* the homotypic clustering element has an invariant core sequence (AAGU) with flexible nucleotides on either ends, appearing as (C/U)AAGU(A/C/U). A search within the 3'UTR of the candidate RNAs revealed that all but two, *eIF5* and *exu*, contained matches to the homotypic clustering motif in at least one of their transcript isoforms (**Table 3.1**).

This analysis was extended to the remaining 41 transcripts that also localize to RNA islands identified in Fly-FISH (Lecuyer et al., 2007; Wilk et al., 2016), and similarly, the majority contained sequences that resembled the motif with the exception of *Ack*, *AhcyL1*, *CG3295* and *eIF4G1* (**Table 3.2**). Notably, RNAs whose 3'UTR is sufficient for localization to germ plasm were amongst the transcripts that had matches, some of which were identical to the sequences identified in *pgc*, *gcl* and *nos* (**Table 3.1, 3.2**) (Eagle et al., 2018). While these sequences are present in the 3'UTRs of 49 out of 55 transcripts that localize to RNA islands, their functional significance in mediating homotypic clustering will need to be determined. For the 6 RNAs that did not have predicted motifs in their 3'UTRs, it will be important to determine if their 3'UTRs are sufficient to confer posterior localization and also consider if localization elements are present in the 5'UTRs or coding regions.

Given the short length of the motif, we considered the possibility that it may be randomly occurring in any 3'UTR. We thus selected 50 RNAs that localize differently and are excluded from the pole cells of the syncytial blastoderm to search their 3'UTRs for the clustering element (**Table 3.3**). The motif is indeed present in the 3'UTR sequences albeit at a lower occurrence with 32 out of 50 RNAs possessing the homotypic clustering motif. This indicates that the element is not

specific to germ plasm transcripts; however, it alone is not sufficient for RNAs to self-recruit as they need to be targeted to the germ plasm in order for them to grow as clusters.

We next considered if this motif was conserved in the *Drosophila* orthologs of these RNAs given that 11 of the 14 RNAs tested had conserved posterior localization in the early embryo. The expression of a transgene carrying *D. virilis nos* in *D. melanogaster* resulted in the posterior localization of the RNA in the embryo, suggesting that there is evolutionary conservation of the recognition of localization elements by the localization machinery (Gavis et al., 1996). Since the homotypic clustering element is primary sequence in nature, we searched for similar sequences within multiple alignments of the 3'UTRs for *pgc*, *gcl*, and *nos* and their orthologs in four *Drosophila* species: *D. simulans*, *D. ananassae*, *D. pseudoobscura* and *D. virilis*.

The *pgc* (230-235) sequence, CAAGUA, in *D. melanogaster* is present as an identical match in *D. simulans pgc* 3'UTR (230-235) (**Fig. 3.6A, Table 3.4**). Versions of the (C/U)AAGU(A/C/U) motif are present in *D. ananassae pgc* 3'UTR, as UAAGUC, at 306-311, and in *D. virilis pgc* 3'UTR, as CAAGUU at 269-274 (**Fig. 3.6A, Table 3.4**). While a second identical CAAGUA at position 327-332 is found in the *D. melanogaster pgc* 3'UTR, it was not tested for mediating self-assembly into homotypic clusters (Eagle et al., 2018). While the *gcl* (363-368) sequence, UAAGUA, is conserved in *D. simulans*, *D. ananassae* and *D. pseudoobscura* There is an identical match to *D. melanogaster gcl* (420-425) sequence, CAAGUU, in *D. simulans* only (**Fig. 3.6B, Table 3.4**). Versions of the *D. melanogaster* homotypic clustering motif is also present in 3'UTR of *D. ananassae*, *D. pseudoobscura* and *D. virilis* orthologs of *gcl* (**Fig. 3.6B, Table 3.4**). Two homotypic clustering motifs were identified in *D. melanogaster nos*; however, only the *nos* (218-233) sequence, CAAGUC, is conserved in *D. simulans*. (**Fig. 3.6C, Table 3.4**). Because the lengths of *nos* 3'UTRs in other *Drosopholids*, with lengths less than 300 nucleotides, were noticeably shorter than the *D. melanogaster nos* 3'UTR of 880 nucleotides, we considered that they might not be fully annotated. To address this, we amplified the *nos* 3'UTR from *D. virilis* unfertilized eggs and obtained an 806-nt 3'UTR of comparable length to *D. melanogaster*. Within the extended *D. virilis nos* 3'UTR lies the (500-505) sequence, CAAGUU, a version of the *D. melanogaster* clustering element (**Fig. 3.7**). While the clustering motif is identical in sequence and location within the 3'UTR of *D. simulans* because *D. simulans* is a closely related species to *D. melanogaster*, it may have evolved rapidly and the sequence is not necessarily the same in other

species despite the conservation in localization in the most distantly related species used in this analysis, *D. virilis.* We also uncovered that related species may have incomplete 3'UTR annotations which should be taken into consideration before performing a phylogenetic comparison of localization signals. Further testing is required to determine if sequences in the other species that match the *D. melanogaster* clustering element are functional in contributing to cluster growth or are simply randomly occurring sequences.

Table 3.1 RNAs, used in this study, that localize to RNA islands and sequences within their 3'UTRs that correspond to the homotypic clustering motif

RNA	Transcript ID	3'UTR Length	Sequence(s)	3'UTR Sufficiency
pgc	FBtr0112519	401	(230-235) **CAAGUA**	Rangan et al., 2009
	FBtr0112520	401	(230-235) **CAAGUA**	
	FBtr0342987	401	(230-235) **CAAGUA**	
gcl	FBtr0088710	524	(363-368) **UAAGUA**; (420-425) **CAAGUU**	Rangan et al., 2009
	FBtr0339337	524	(363-368) **UAAGUA**; (420-425) **CAAGUU**	
nos	FBtr0083732	880	(218-223) **CAAGUC**; (645-650) **CAAGUA**	Gavis and Lehmann, 1992
	FBtr0335019	880	(218-223) **CAAGUC**; (645-650) **CAAGUA**	
CycB	FBtr0071911	773	(254-259) **CAAGUU**	Dalby and Glover, 1992
	FBtr0071913	773	(254-259) **CAAGUU**	
	FBtr0071914	773	(254-259) **CAAGUU**	
	FBtr0309858	773	(254-259) **CAAGUU**	
bru1	FBtr0080345	503		Rangan et al., 2009
	FBtr0080344	503		
	FBtr0080347	253		
	FBtr0300567	154		
	FBtr0331644	1146	(589-594) UAAGUU; (799-804) **CAAGUA**; (1112-1117) **UAAGUA**	
	FBtr0331645	1361	(532-537) UAAGUU; (602-607) **CAAGUC**; (669-674) **CAAGUU**; (830-835) **UAAGUA**	
	FBtr0331646	1361	(532-537) UAAGUU; (602-607) **CAAGUC**; (669-674) **CAAGUU**; (830-835) **UAAGUA**	
	FBtr0331647	1146	(589-594) UAAGUU; (799-804) **CAAGUA**; (1112-1117) **UAAGUA**	
	FBtr0331648	253		
	FBtr0331998	1361	(532-537) UAAGUU; (602-607) **CAAGUC**; (669-674) **CAAGUU**; (830-835) **UAAGUA**	
Bsg25D	FBtr0079084	950	(151-156) UAAGUU; (715-720) UAAGUU	Kowanda et al., 2016
	FBtr0079085	950	(151-156) UAAGUU; (715-720) UAAGUU	
	FBtr0302567	188		
	FBtr0302568	950	(151-156) UAAGUU; (715-720) UAAGUU	
	FBtr0310624	950	(151-156) UAAGUU; (715-720) UAAGUU	
	FBtr0310625	950	(151-156) UAAGUU; (715-720) UAAGUU	
	FBtr0342939	188		
Tao	FBtr0074774	412	(295-300) **CAAGUA**	ND
	FBtr0074773	412	(295-300) **CAAGUA**	
	FBtr0074772	412	(295-300) **CAAGUA**	
	FBtr0074771	412	(295-300) **CAAGUA**	
	FBtr0303999	1860	(295-300) **CAAGUA**	
	FBtr0304000	934	(295-300) **CAAGUA**	
	FBtr0309237	441		
	FBtr0336691	412	(295-300) **CAAGUA**	
	FBtr0336692	393		
jvl	FBtr0305690	2378	(400-405) UAAGUC; (433-438) UAAGUU; (675-680) **CAAGUA**; (1972-1977) UAAGUU; (2041-2046) **UAAGUA**; (2138-2143) UAAGUC	ND
	FBtr0305689	117		
	FBtr0305691	117		
	FBtr0305692	806	(400-405) UAAGUU; (469-474) **UAAGUA**; (566-571) UAAGUC	
	FBtr0305693	1738	(1332-1337) UAAGUU; (1401-1406) **UAAGUA**; (1498-1503) UAAGUC	
	FBtr0305694	2441	(170-175) **CAAGUA**	
	FBtr0305695	806		
pum	FBtr0081990	1188	(93-98) **CAAGUA**; (516-521) **CAAGUC**; (638-643) UAAGUU; (980-985) UAAGUC; (987-992) **CAAGUA**; (1121-1126) **CAAGUA**	ND
	FBtr0081993	3367	(93-98) **CAAGUA**; (516-521) **CAAGUC**; (638-643) UAAGUU; (980-985) UAAGUC; (987-992) **CAAGUA**; (1121-1126) **CAAGUA**; (2138-2143) **CAAGUA**; (2555-2560) **CAAGUC**; (2851-2856) **CAAGUA**; (3033-3038) **CAAGUA**; (3222-3227) **CAAGUU**; (3351-3356) **CAAGUU**	
	FBtr0081992	1188	(93-98) **CAAGUA**; (516-521) **CAAGUC**; (638-643) UAAGUU; (980-985) UAAGUC; (987-992) **CAAGUA**; (1121-1126) **CAAGUA**	

(Table continues)

RNA	Transcript ID	3'UTR Length	Sequence(s)	3'UTR Sufficiency
pum	FBtr0081991	1188	(93-98) **CAAGUA**; (516-521) **CAAGUC**; (638-643) UAAGUU; (980-985) UAAGUC; (987-992) **CAAGUA**; (1121-1126) **CAAGUA**	ND
	FBtr0081994	1188	(93-98) **CAAGUA**; (516-521) **CAAGUC**; (638-643) UAAGUU; (980-985) UAAGUC; (987-992) **CAAGUA**; (1121-1126) **CAAGUA**	
	FBtr0305197	1188	(93-98) **CAAGUA**; (516-521) **CAAGUC**; (638-643) UAAGUU; (980-985) UAAGUC; (987-992) **CAAGUA**; (1121-1126) **CAAGUA**	
	FBtr0333667	4602	(93-98) **CAAGUA**; (516-521) **CAAGUC**; (638-643) UAAGUU; (980-985) UAAGUC; (987-992) **CAAGUA**; (1121-1126) **CAAGUA**; (2138-2143) **CAAGUA**; (2555-2560) **CAAGUC**; (2851-2856) **CAAGUA**; (3033-3038) **CAAGUA**; (3222-3227) **CAAGUU**; (3351-3356) CAAGUU; (3643-3648) **CAAGUU**; (3743-3748) **UAAGUA**; (4026-4031) **CAAGUC**; (4289-4294) UAAGUC	
	FBtr0333668	4602	(93-98) **CAAGUA**; (516-521) **CAAGUC**; (638-643) UAAGUU; (980-985) UAAGUC; (987-992) **CAAGUA**; (1121-1126) **CAAGUA**; (2138-2143) **CAAGUA**; (2555-2560) **CAAGUC**; (2851-2856) **CAAGUA**; (3033-3038) **CAAGUA**; (3222-3227) **CAAGUU**; (3351-3356) **CAAGUU**; (3643-3648) **CAAGUU**; (3743-3748) **UAAGUA**; (4026-4031) **CAAGUC**; (4289-4294) UAAGUC	
Tlk	FBtr0112680	232		ND
	FBtr0070596	1573	(1268-1273) **UAAGUA**	
	FBtr0299579	1573	(1268-1273) **UAAGUA**	
	FBtr0299580	2414	(1268-1273) **UAAGUA**; (1746-1751) UAAGUC; (1802-1807) **UAAGUA**	
	FBtr0299581	1573	(1268-1273) **UAAGUA**	
	FBtr0299582	1573	(1268-1273) **UAAGUA**	
	FBtr0299583	565		
	FBtr0301659	565		
	FBtr0333751	3366	(1268-1273) **UAAGUA**; (1746-1751) UAAGUC; (1802-1807) **UAAGUA**; (3235-3240) **UAAGUA**	
	FBtr0333752	565		
	FBtr0333753	3366	(1268-1273) **UAAGUA**; (1746-1751) UAAGUC; (1802-1807) **UAAGUA**; (3235-3240) **UAAGUA**	
	FBtr0345319	232		
gwl	FBtr0083674	356	(302-307) **CAAGUA**	ND
	FBtr0331352	356	(302-307) **CAAGUA**	
Pi3K21B	FBtr0331205	1275	(765-770) UAAGUC	ND
	FBtr0331206	2574	(765-770) UAAGUC; (1933-1938) UAAGUU	
	FBtr0331207	2574	(765-770) UAAGUC; (1933-1938) UAAGUU	
	FBtr0331208	2574	(765-770) UAAGUC; (1933-1938) UAAGUU	
eIF5	FBtr0074144	119		ND
	FBtr0074145	119		
	FBtr0074146	119		
	FBtr0074147	277		
	FBtr0074148	119		
	FBtr0074149	119		
	FBtr0074150	119		
exu	FBtr0086242	248		ND
	FBtr0086243	248		
	FBtr0086244	881		
	FBtr0302285	881		

Column 1, RNAs used in this study; **column 2**, all known transcript isoforms; **column 3**, 3'UTR length; **column 4**, regions within 3'UTR, in parentheses, that contain the sequences corresponding to the homotypic clustering motif (Eagle et al., 2018), and blank, no sequence identified; **column 5**, sufficiency of 3'UTR to localize to pole plasm and the references, ND, not determined. Sequences in bold are identical to the motifs found in *pgc*, *gcl* and *nos*. eIF5 and *exu* 3'UTRs did not contain sequences that matched the motif.

Table 3.2 RNAs identified in Fly-FISH that localize to RNA islands and sequences within their 3'UTRs that correspond to the homotypic clustering motif

RNA	Transcript ID	3'UTR Length	Sequence(s)	3'UTR Sufficiency
CG18446	FBtr0088376	371	(324-329) UAAGUU	Rangan et al., 2009
CG5292	FBtr0083410	595	(453-458) **UAAGUA**	Rangan et al., 2009
	FBtr0337042	1259	(453-458) **UAAGUA**; (795-800) UAAGUU	
orb	FBtr0084384	1206	(411-416) **CAAGUU**; (1074-1079) UAAGUU; (1112-1117) UAAGUC	Lantz and Schel, 1994
	FBtr0084385	1206	(411-416) **CAAGUU**; (1074-1079) UAAGUU; (1112-1117) UAAGUC	
	FBtr0084386	1206	(411-416) **CAAGUU**; (1074-1079) UAAGUU; (1112-1117) UAAGUC	
	FBtr0100227	213		
	FBtr0100228	213		
	FBtr0301491	199		
	FBtr0334498	5243	(411-416) **CAAGUU**; (1074-1079) UAAGUU; (1112-1117) UAAGUC; (1608-1613) **CAAGUA**; (1887-1892) **UAAGUA**; (2148-2153) **UAAGUA**; (2237-2242) **CAAGUU**; (3252-3257) **CAAGUC**; (4964-4969) **CAAGUA**; (5124-5129) UAAGUU	
sra	FBtr0083255	669	(271-276) **CAAGUU**	Rangan et al., 2009
	FBtr0344473	669	(271-276) **CAAGUU**	
Ank	FBtr0089173	425	(51-56) UAAGUU; (283-288) UAAGUU	ND
	FBtr0089171	425	(51-56) UAAGUU; (283-288) UAAGUU	
	FBtr0089172	425	(51-56) UAAGUU; (283-288) UAAGUU	
	FBtr0300497	1059	(51-56) UAAGUU; (724-729) **CAAGUC**	
aqz	FBtr0081932	201		ND
	FBtr0081931	3178	(413-418) **CAAGUA**; (799-804) **UAAGUA**; (1594-1599) UAAGUU; (2119-2124) **CAAGUA**; (2143-2148) **CAAGUU**; (2182-2187) **CAAGUU**; (2194-2199) **CAAGUA**; (2786-2791) **CAAGUU**; (3156-3161) **CAAGUC**	
	FBtr0334659	4373	(413-418) **CAAGUA**; (799-804) **UAAGUA**; (1594-1599) UAAGUU; (2119-2124) **CAAGUA**; (2143-2148) **CAAGUU**; (2182-2187) **CAAGUU**; (2194-2199) **CAAGUA**; (2786-2791) **CAAGUU**; (3156-3161) **CAAGUC**; (4111-4116) **CAAGUU**	
	FBtr0334660	3178	(413-418) **CAAGUA**; (799-804) **UAAGUA**; (1594-1599) UAAGUU; (2119-2124) **CAAGUA**; (2143-2148) **CAAGUU**; (2182-2187) **CAAGUU**; (2194-2199) **CAAGUA**; (2786-2791) **CAAGUU**; (3156-3161) **CAAGUC**	
Atf6	FBtr0086079	726	(317-322) **CAAGUU**; (616-621) UAAGUU	ND
	FBtr0086080	726	(317-322) **CAAGUU**; (616-621) UAAGUU	
	FBtr0086081	164		
bip2	FBtr0089110	134	(103-108) **CAAGUU**	ND
CG10077	FBtr0076940	839	(114-119) **CAAGUC**; (127-132) **CAAGUC**; (790-795) **UAAGUA**	ND
	FBtr0300937	1955	(723-728) UAAGUC; (1087-1092) UAAGUU; (1166-1171) **CAAGUA**; (1415-1420) **CAAGUU**; (1434-1439) UAAGUU; (1563-1568) **UAAGUA**	
	FBtr0333083	839	(114-119) **CAAGUC**; (127-132) **CAAGUC**; (790-795) **UAAGUA**	
CG11597	FBtr0076080	141	(26-31) **CAAGUA**	ND
	FBtr0110780	141	(26-31) **CAAGUA**	
	FBtr0302137	141	(26-31) **CAAGUA**	
CG14322	FBtr0083476	90		ND
	FBtr0331346	90		
	FBtr0331347	120		
	FBtr0344347	858	(700-705) **CAAGUC**	
CG2865	FBtr0070408	52		ND
	FBtr0332962	3547	(322-327) UAAGUU; (901-906) **CAAGUU**; (1187-1192) UAAGUU; (1521-1526) UAAGUC; (1741-1746) UAAGUC; (1867-1872) UAAGUU; (3182-3187) **CAAGUU**; (3388-3393) **CAAGUA**	
CG31998	FBtr0089145	186	(104-109) **UAAGUA**	ND
	FBtr0308249	1203	(104-109) **UAAGUA**; (726-731) **UAAGUA**; (766-771) UAAGUU; (855-860) UAAGUU; (1102-1107) **UAAGUA**	
chrb	FBtr0076093	3960	(1065-1070) **CAAGUC**; (1877-1882) **CAAGUA**; (2237-2242) **CAAGUA**; (2562-2567) **CAAGUA**; (2938-2943) **CAAGUA**; (3452-3457) UAAGUU; (3719-3724) **CAAGUC**	ND

(Table continues)

RNA	Transcript ID	3'UTR Length	Sequence(s)	3'UTR Sufficiency
chrb	FBtr0344910	3960	(1065-1070) **CAAGUC**; (1877-1882) **CAAGUA**; (2237-2242) **CAAGUA**; (2562-2567) **CAAGUA**; (2938-2943) **CAAGUA**; (3452-3457) UAAGUU; (3719-3724) **CAAGUC**	ND
cta	FBtr0111126	400		ND
	FBtr0334157	1008	(607-612) **UAAGUA**; (957-962) UAAGUC	
Dlg5	FBtr0080247	901	(425-430) **UAAGUA**	ND
	FBtr0080246	526	(425-430) **UAAGUA**	
	FBtr0336630	2281	(144-149) **CAAGUC**; (360-365) UAAGUU; (1011-1016) **CAAGUU**; (1152-1157) UAAGUC; (1805-1810) **UAAGUA**	
	FBtr0336631	901	(425-430) **UAAGUA**	
Hip14	FBtr0075518	783		ND
	FBtr0330161	729		
	FBtr0333577	1607	(915-920) **UAAGUA**; (1490-1495) UAAGUU	
mei-P26	FBtr0071303	844	(272-277) **UAAGUA**	ND
	FBtr0333473	18494	(272-277) **UAAGUA**; (1394-1399) **UAAGUA**; (1982-1987) UAAGUU; (4113-4118) UAAGUU; (4297-4302) UAAGUU; (5824-5829) **UAAGUA**; (5982-5987) **UAAGUA**; (6045-6050) **CAAGUA**; (6724-6729) **CAAGUU**; (6895-6900) **CAAGUU**; (7063-7068) **CAAGUA**; (7160-7165) **CAAGUU**; (7574-7579) UAAGUC; (8287-8292) **CAAGUU**; (8707-8712) **CAAGUA**; (9004-9009) **UAAGUA**; (10955-10960) **UAAGUA**; (11158-11163) UAAGUU; (12589-12594) UAAGUU; (15764-15769) **UAAGUA**; (15913-15918) **UAAGUA**; (18194-18199) **CAAGUU**	
	FBtr0333474	18494	(272-277) **UAAGUA**; (1394-1399) **UAAGUA**; (1982-1987) UAAGUU; (4113-4118) UAAGUU; (4297-4302) UAAGUU; (5824-5829) **UAAGUA**; (5982-5987) **UAAGUA**; (6045-6050) **CAAGUA**; (6724-6729) **CAAGUU**; (6895-6900) **CAAGUU**; (7063-7068) **CAAGUA**; (7160-7165) **CAAGUU**; (7574-7579) UAAGUC; (8287-8292) **CAAGUU**; (8707-8712) **CAAGUA**; (9004-9009) **UAAGUA**; (10955-10960) **UAAGUA**; (11158-11163) UAAGUU; (12589-12594) UAAGUU; (15764-15769) **UAAGUA**; (15913-15918) **UAAGUA**; (18194-18199) **CAAGUU**	
	FBtr0333475	4375	(272-277) **UAAGUA**; (1394-1399) **UAAGUA**; (1982-1987) UAAGUU; (4113-4118) UAAGUU	
	FBtr0340341	844	(272-277) **UAAGUA**	
	FBtr0474151	481		
milt	FBtr0306228	291	(127-132) **CAAGUU**	ND
	FBtr0306230	291	(127-132) **CAAGUU**	
	FBtr0306229	1454	(127-132) **CAAGUU**; (614-619) **UAAGUA**	
	FBtr0306231	1454	(127-132) **CAAGUU**; (614-619) **UAAGUA**	
	FBtr0306232	291	(127-132) **CAAGUU**	
	FBtr0306233	1454	(127-132) **CAAGUU**; (614-619) **UAAGUA**	
Nost	FBtr0302905	42		ND
	FBtr0302906	42		
	FBtr0300000	42		
	FBtr0300001	42		
	FBtr0302907	1298	(700-705) **CAAGUU**	
	FBtr0343357	1298	(700-705) **CAAGUU**	
nrv1	FBtr0079384	818	(635-640) **UAAGUA**	ND
	FBtr0332370	549		
pAbp	FBtr0086738	331	(244-249) **CAAGUA**	ND
	FBtr0086740	1904	(244-249) **CAAGUA**; (482-487) **CAAGUA**; (818-823) **CAAGUU**; (1203-1208) UAAGUC; (1598-103) **CAAGUC**	
	FBtr0086736	948	(244-249) **CAAGUA**; (482-487) **CAAGUA**; (818-823) **CAAGUU**	
Patr-1	FBtr0083392	539		ND
	FBtr0339155	1292	(593-598) **UAAGUA**	
RapGAP1	FBtr0334846	1220	(262-267) **CAAGUU**; (375-380) UAAGUC; (468-473) **UAAGUA**; (771-776) UAAGUU; (777-782) UAAGUU; (828-833) UAAGUU; (1120-1125) **UAAGUA**	ND
	FBtr0334847	1220	(262-267) **CAAGUU**; (375-380) UAAGUC; (468-473) **UAAGUA**; (771-776) UAAGUU; (777-782) UAAGUU; (828-833) UAAGUU; (1120-1125) **UAAGUA**	

(Table continues)

RNA	Transcript ID	3'UTR Length	Sequence(s)	3'UTR Sufficiency
RapGAP1	FBtr0334848	1220	(262-267) **CAAGUU**; (375-380) UAAGUC; (468-473) **UAAGUA**; (771-776) UAAGUU; (777-782) UAAGUU; (828-833) UAAGUU; (1120-1125) **UAAGUA**	ND
	FBtr0334849	1220	(262-267) **CAAGUU**; (375-380) UAAGUC; (468-473) **UAAGUA**; (771-776) UAAGUU; (777-782) UAAGUU; (828-833) UAAGUU; (1120-1125) **UAAGUA**	
	FBtr0334850	1220	(262-267) **CAAGUU**; (375-380) UAAGUC; (468-473) **UAAGUA**; (771-776) UAAGUU; (777-782) UAAGUU; (828-833) UAAGUU; (1120-1125) **UAAGUA**	
RasGAP1	FBtr0076367	876	(206-211) UAAGUU; (505-510) UAAGUU; (510-515) **UAAGUA**	ND
	FBtr0076368	1469	(206-211) UAAGUU; (505-510) UAAGUU; (510-515) **UAAGUA**; (974-979) **CAAGUU**; (1388-1393) UAAGUU	
	FBtr0330168	822	(206-211) UAAGUU; (505-510) UAAGUU; (510-515) **UAAGUA**	
Sep4	FBtr0074365	935	(405-410) **CAAGUA**; (516-521) **CAAGUA**; (594-599) **CAAGUU**; (718-723) **CAAGUU**	ND
	FBtr0074367	935	(405-410) **CAAGUA**; (516-521) **CAAGUA**; (594-599) **CAAGUU**; (718-723) **CAAGUU**	
	FBtr0074370	935	(405-410) **CAAGUA**; (516-521) **CAAGUA**; (594-599) **CAAGUU**; (718-723) **CAAGUU**	
	FBtr0074368	935	(405-410) **CAAGUA**; (516-521) **CAAGUA**; (594-599) **CAAGUU**; (718-723) **CAAGUU**	
	FBtr0074369	935	(405-410) **CAAGUA**; (516-521) **CAAGUA**; (594-599) **CAAGUU**; (718-723) **CAAGUU**	
	FBtr0074366	935	(405-410) **CAAGUA**; (516-521) **CAAGUA**; (594-599) **CAAGUU**; (718-723) **CAAGUU**	
	FBtr0074371	935	(405-410) **CAAGUA**; (516-521) **CAAGUA**; (594-599) **CAAGUU**; (718-723) **CAAGUU**	
sl	FBtr0074230	1696	(312-317) **CAAGUU**; (364-369) **CAAGUC**; (544-549) **UAAGUA**; (553-558) UAAGUU; (653-658) **UAAGUA**; (1212-1217) UAAGUU	ND
spir	FBtr0081352	1217	(549-554) **UAAGUA**; (932-937) UAAGUU; (1085-1090) **UAAGUA**; (1116-1121) UAAGUU	ND
	FBtr0081353	1217	(549-554) **UAAGUA**; (932-937) UAAGUU; (1085-1090) **UAAGUA**; (1116-1121) UAAGUU	
	FBtr0081355	1217	(549-554) **UAAGUA**; (932-937) UAAGUU; (1085-1090) **UAAGUA**; (1116-1121) UAAGUU	
	FBtr0081354	307		
	FBtr0301884	1217	(549-554) **UAAGUA**; (932-937) UAAGUU; (1085-1090) **UAAGUA**; (1116-1121) UAAGUU	
	FBtr0306556	1217	(549-554) **UAAGUA**; (932-937) UAAGUU; (1085-1090) **UAAGUA**; (1116-1121) UAAGUU	
	FBtr0307893	307		
	FBtr0307894	114		
	FBtr0339233	114		
	FBtr0339234	3236	(572-577) **CAAGUU**; (1020-1025) **CAAGUC**; (1239-1244) **CAAGUU**; (1440-1445) **CAAGUC**; (2568-2573) **UAAGUA**; (2951-2956) UAAGUU; (3104-3109) **UAAGUA**; (3135-3140) UAAGUU	
Tm1	FBtr0089960	806		ND
	FBtr0089962	740	(413-418) **CAAGUA**; (617-622) UAAGUU	
	FBtr0089965	305		
	FBtr0089959	70		
	FBtr0089964	1104		
	FBtr0089963	740	(413-418) **CAAGUA**; (617-622) UAAGUU	
	FBtr0089961	740	(413-418) **CAAGUA**; (617-622) UAAGUU	
	FBtr0089967	305		
	FBtr0089968	1249	(835-840) UAAGUU	
	FBtr0301959	305		
	FBtr0301960	305		
	FBtr0305654	305		
	FBtr0305655	158	(4-9) **CAAGUU**	
	FBtr0305656	129		
	FBtr0333917	672		
	FBtr0333918	305		

(Table continues)

RNA	Transcript ID	3'UTR Length	Sequence(s)	3'UTR Sufficiency
ttk	FBtr0085829	896	(243-248) UAAGUU; (506-511) **CAAGUU**	ND
	FBtr0085827	896	(243-248) UAAGUU; (506-511) **CAAGUU**	
	FBtr0085828	2070	(415-420) **CAAGUA**; (1265-1270) **CAAGUA**; (1281-1286) **CAAGUA**	
	FBtr0085830	2070	(415-420) **CAAGUA**; (1265-1270) **CAAGUA**; (1281-1286) **CAAGUA**	
	FBtr0085825	896	(243-248) UAAGUU; (506-511) **CAAGUU**	
	FBtr0085826	2070	(415-420) **CAAGUA**; (1265-1270) **CAAGUA**; (1281-1286) **CAAGUA**	
	FBtr0303227	896	(243-248) UAAGUU; (506-511) **CAAGUU**	
	FBtr0303228	2070	(415-420) **CAAGUA**; (1265-1270) **CAAGUA**; (1281-1286) **CAAGUA**	
Unr	FBtr0076661	421		ND
	FBtr0114592	1237	(769-774) C**AAGUU**; (857-862) **CAAGUA**	
	FBtr0308954	2147	(769-774) **CAAGUU**; (857-862) **CAAGUA**; (1863-1868) **CAAGUA**; (2004-2009) UAAGUC	
	FBtr0332984	1237	(769-774) C**AAGUU**; (857-862) **CAAGUA**	
	FBtr0345595	2147	(769-774) **CAAGUU**; (857-862) **CAAGUA**; (1863-1868) **CAAGUA**; (2004-2009) UAAGUC	
CAH2	FBtr0076001	600	(553-558) UAAGUU	ND
	FBtr0333381	600	(553-558) UAAGUU	
del	FBtr0081502	162	(87-92) UAAGUU	ND
	FBtr0333084	162	(87-92) UAAGUU	
	FBtr0333085	162	(87-92) UAAGUU	
dock	FBtr0078013	1318	(1020-1025) UAAGUU	ND
	FBtr0078014	1318	(1020-1025) UAAGUU	
	FBtr0078015	1318	(1020-1025) UAAGUU	
	FBtr0332973	408		
pbl	FBtr0076771	615	(207-212) UAAGUU; (538-543) UAAGUC; (571-576) UAAGUU	ND
	FBtr0076772	615	(207-212) UAAGUU; (538-543) UAAGUC; (571-576) UAAGUU	
	FBtr0076773	615	(207-212) UAAGUU; (538-543) UAAGUC; (571-576) UAAGUU	
	FBtr0076774	615	(207-212) UAAGUU; (538-543) UAAGUC; (571-576) UAAGUU	
	FBtr0112817	615	(207-212) UAAGUU; (538-543) UAAGUC; (571-576) UAAGUU	
	FBtr0306647	615	(207-212) UAAGUU; (538-543) UAAGUC; (571-576) UAAGUU	
Pino	FBtr0077988	1080	(335-340) UAAGUU	ND
	FBtr0077987	1080	(335-340) UAAGUU	
	FBtr0332272	1080	(335-340) UAAGUU	
Ack	FBtr0073211	370		ND
AhcyL1	FBtr0073016	918		ND
CG3295	FBtr0071547	308		ND
eIF4G1	FBtr0089243	195		ND
	FBtr0112904	195		
	FBtr0289951	195		

Column 1, RNAs that localize to RNA islands identified by Fly-FISH excluding those used in this study (Lecuyer et al., 2007; Wilk et al., 2016); **column 2**, all known transcript isoforms; **column 3**, 3'UTR length; **column 4**, regions within 3'UTR, in parentheses, that contain the sequences corresponding to the homotypic clustering motif (Eagle et al., 2018), and blank, no sequence identified; **column 5**, sufficiency of 3'UTR to localize to pole plasm and the references, ND, not determined. Sequences in bold are identical to the motifs in *pgc*, *gcl* and *nos*. *Ack*, *AhcyL1*, *CG3295* and *eIF4G1* 3'UTRs did not contain any sequences that matched the motif.

Table 3.3 RNAs identified in Fly-FISH that are pole cell excluded and sequences within their 3'UTRs that correspond to the homotypic clustering motif

RNA	Transcript ID	3'UTR Length	Sequence(s)
AGO1	FBtr0087612	4225	(150-155) UAAGUU; (359-364) **CAAGUU**; (564-569) **CAAGU**U; (989-994) UAAGUU; (1712-1717) **CAAGUU**; (1849-1854) **CAAGUU**; (3070-3075) **UAAGUA**; (3085-3090) **CAAGUU**; (4127-4132) **CAAGUU**
	FBtr0087613	435	(150-155) UAAGUU; (359-364) **CAAGUU**
	FBtr0087614	2563	(150-155) UAAGUU; (359-364) **CAAGUU**; (564-569) **CAAGUU**; (915-920); UAAGUU; (989-994) UAAGUU; (1712-1717) **CAAGUU**; (1849-1854) **CAAGUU**
	FBtr0087612	4585	(150-155) UAAGUU; (359-364) **CAAGUU**; (564-569) **CAAGUU**; (915-920); (989-994) UAAGUU; (1712-1717) **CAAGUU**; (1849-1854) **CAAGUU**; (3070-3075) **UAAGUA**; (3085-3090) **CAAGUU**; (4127-4132) **CAAGUU**
alien	FBtr0079780	57	
	FBtr0079781	870	(416-421) **CAAGUU**
	FBtr0339550	1721	(416-421) **CAAGUU**
	FBtr0343095	123	
Ance-5	FBtr0072420	117	(6-11) **CAAGUA**; (86-91) **CAAGUA**
Baldspot	FBtr0075355	1633	(88-93) **CAAGUU**; (773-778) UAAGUC; (861-866) UAAGUU; (998-1003) UAAGUC
	FBtr0075356	1234	(88-93) **CAAGUU**; (773-778) UAAGUC; (861-866) UAAGUU; (998-1003) UAAGUC
Best1	FBtr0082183	281	(39-44) UAAGUC; (233-238) **UAAGUA**; (263-268) UAAGUU
	FBtr0114590	521	(39-44) UAAGUC; (233-238) **UAAGUA**; (263-268) UAAGUU
	FBtr0306023	281	(39-44) UAAGUC; (233-238) **UAAGUA**; (263-268) UAAGUU
	FBtr0337020	281	(39-44) UAAGUC; (233-238) **UAAGUA**; (263-268) UAAGUU
bib	FBtr0079929	880	(175-180) **CAAGUC**; (424-429) UAAGUU; (866-871) **CAAGUU**
	FBtr0329984	757	(52-57) **CAAGUC**; (301-306) UAAGUU; (743-748) **CAAGUU**
blot	FBtr0075232	523	(252-257) UAAGUC; (413-418) UAAGUC; (442-447) **UAAGUA**
	FBtr0075233	523	(252-257) UAAGUC; (413-418) UAAGUC; (442-447) **UAAGUA**
	FBtr0075234	523	(252-257) UAAGUC; (413-418) UAAGUC; (442-447) **UAAGUA**
	FBtr0075235	523	(252-257) UAAGUC; (413-418) UAAGUC; (442-447) **UAAGUA**
	FBtr0339898	523	(252-257) UAAGUC; (413-418) UAAGUC; (442-447) **UAAGUA**
brat	FBtr0081158	593	(82-87) UAAGUU; (465-470) **CAAGUC**
	FBtr0081159	1308	(82-87) UAAGUU; (465-470) **CAAGUC**
	FBtr0081160	8499	(82-87) UAAGUU; (465-470) **CAAGUC**; (1339-1344) **CAAGUU**; (1833-1838) **UAAGUA**; (4034-4039) **CAAGUU**; (4390-4395) **UAAGUA**; (4632-4637) **UAAGUA**; (4876-4881) **UAAGUA**; (4953-4958) **CAAGUA**; (5194-5154) **CAAGUU**; (5926-5931) **CAAGUU**; (6398-6403) **CAAGUU**; (6801-6806) UAAGUU; (6920-6925) **CAAGUU**; (7518-7523) UAAGUU
	FBtr0336627	2235	(134-139) **CAAGUU**; (537-542) UAAGUU; (656-661) **CAAGUU**; (1254-1259) UAAGUU
	FBtr0344846	1308	(82-87) UAAGUU; (465-470) **CAAGUC**
bs	FBtr0072271	483	(171-176) UAAGUC; (177-182) UAAGUC; (294-299) UAAGUU; (312-317) **CAAGUU**
	FBtr0290089	1495	(301-306) **CAAGUC**; (783-788) **CAAGUA**; (849-854) **CAAGUU**; (1212-1217) **CAAGUU**
	FBtr0343314	483	(171-176) UAAGUC; (177-182) UAAGUC; (294-299) UAAGUU; (312-317) **CAAGUU**
clumsy	FBtr0081476	933	(25-30) **CAAGUU**; (378-383) **CAAGUU**; (442-447) **CAAGUC**; (598-603) **CAAGUC**
	FBtr0110895	164	
CycB3	FBtr0084728	445	(159-164) **CAAGUC**
	FBtr0334578	876	(159-164) **CAAGUC**; (632-637) UAAGUC
dikar	FBtr0303748	340	
	FBtr0303749	309	(218-223) **UAAGUA**
	FBtr0303750	1671	(218-223) **UAAGUA**; (600-605) **CAAGUU**; (976-981) **CAAGUA**; (1090-1095) **UAAGUA**; (1255-1260) **CAAGUA**; (1531-1536) UAAGUU
	FBtr0303751	1734	(982-987) UAAGUC; (1060-1065) **CAAGUA**; (1199-1204) **CAAGUU**; (1463-1468) **CAAGUC**
	FBtr0330145	192	(101-106) **UAAGUA**
	FBtr0333047	1671	(218-223) **UAAGUA**; (600-605) **CAAGUU**; (976-981) **CAAGUA**; (1090-1095) **UAAGUA**; (1255-1260) **CAAGUA**; (1531-1536) UAAGUU
gammaTub23C	FBtr0077641	254	(93-98) **UAAGUA**
Gas8	FBtr0070589	79	(27-32) **CAAGUC**
	FBtr0343540	502	(27-32) **CAAGUC**
	FBtr0344969	502	(27-32) **CAAGUC**
halo	FBtr0114554	199	(127-132) **CAAGUU**; (164-169) UAAGUC

(Table continues)

Him	FBtr0074549	234	(213-218) **UAAGUA**

RNA	Transcript ID	3'UTR Length	Sequence(s)
lbm	FBtr0089033	217	(178-183) **UAAGUA**
	FBtr0336873	217	(178-183) **UAAGUA**
lt	FBtr0111123	81	
	FBtr0111125	177	
	FBtr0333519	1777	(255-260) **CAAGUA**; (353-358) **UAAGUA**; (596-601) UAAGUU; (1537-1542) UAAGUC
	FBtr0344082	2689	(255-260) **CAAGUA**; (353-358) **UAAGUA**; (596-601) UAAGUU; (1537-1542) UAAGUC; (2006-2011) **UAAGUA**; (2063-2068) **UAAGUA**; (2538-2543) **CAAGUU**
mal	FBtr0077306	284	(95-100) **UAAGUA**; (145-150) **UAAGUA**; (202-207) UAAGUU
	FBtr0344987	284	(95-100) **UAAGUA**; (145-150) **UAAGUA**; (202-207) UAAGUU
Mtl	FBtr0085191	395	
	FBtr0085192	1171	(820-825) **UAAGUA**; (1052-1057) **CAAGUC**
	FBtr0085193	1171	(820-825) **UAAGUA**; (1052-1057) **CAAGUC**
Nfl	FBtr0089141	2136	(126-131) UAAGUU; (595-600) UAAGUU; (693-698) UAAGUU; (1012-1017) **CAAGUA**; (1621-1626) **CAAGUU**; (2059-2064) UAAGUU
	FBtr0333678	2275	(126-131) UAAGUU; (595-600) UAAGUU; (693-698) UAAGUU; (1012-1017) **CAAGUA**; (1621-1626) **CAAGUU**; (2059-2064) UAAGUU
pain	FBtr0333814	259	(158-163) **UAAGUA**; (162-167) **UAAGUA**
Pep	FBtr0075198	1909	(175-180) **CAAGUC**; (286-291) **CAAGUC**; (781-786) **UAAGUA**; (856-861) **CAAGUU**; (1520-1525) **UAAGUA**; (1777-1782) **CAAGUU**
	FBtr0075199	402	(255-260) **CAAGUU**
	FBtr0075200	402	(255-260) **CAAGUU**
	FBtr0304977	402	(255-260) **CAAGUU**
sax	FBtr0088832	648	(158-163) **CAAGUA**; (288-293) UAAGUU; (603-608) UAAGUU
	FBtr0088833	648	(158-163) **CAAGUA**; (288-293) UAAGUU; (603-608) UAAGUU
	FBtr0308825	648	(158-163) **CAAGUA**; (288-293) UAAGUU; (603-608) UAAGUU
skd	FBtr0078328	1573	(161-166) **CAAGUA**; (322-337) UAAGUC; (595-600) **UAAGUA**; (987-992) **UAAGUA**; (1329-1334) UAAGUU; (1350-1355) **UAAGUA**; (1513-1518) **UAAGUA**
	FBtr0078329	2890	(161-166) **CAAGUA**; (322-337) UAAGUC; (595-600) **UAAGUA**; (987-992) **UAAGUA**; (1329-1334) UAAGUU; (1350-1355) **UAAGUA**; (1513-1518) **UAAGUA**; (2182-2187) **CAAGUU**; (2218-2223) **CAAGUU**; (2359-2364) UAAGUU; (2392-2397) UAAGUC; (2641-2646) UAAGUU; (2659-2664) **CAAGUC**
	FBtr0112830	1573	(161-166) **CAAGUA**; (322-337) UAAGUC; (595-600) **UAAGUA**; (987-992) **UAAGUA**; (1329-1334) UAAGUU; (1350-1355) **UAAGUA**; (1513-1518) **UAAGUA**
	FBtr0333284	1573	(161-166) **CAAGUA**; (322-337) UAAGUC; (595-600) **UAAGUA**; (987-992) **UAAGUA**; (1329-1334) UAAGUU; (1350-1355) **UAAGUA**; (1513-1518) **UAAGUA**
Skeletor	FBtr0305663	290	
	FBtr0305664	2251	(550-555) UAAGUC; (1631-1636) **CAAGUC**
	FBtr0305665	224	
	FBtr0305666	224	
Swim	FBtr0071784	543	(177-182) UAAGUU; (195-200) **CAAGUU**; (464-469) UAAGUU
	FBtr0071785	543	(177-182) UAAGUU; (195-200) **CAAGUU**; (464-469) UAAGUU
tinc	FBtr0089771	1736	(88-93) UAAGUU; (193-198) **CAAGUC**; (218-223) UAAGUU; (542-547) **UAAGUA**; (1484-1489) **UAAGUA**
	FBtr0089772	3244	(79-84) UAAGUC; (1596-1601) UAAGUU; (1701-1706) **CAAGUC**; (1726-1731) UAAGUU; (2050-2055) **UAAGUA**; (2992-2997) **UAAGUA**
	FBtr0304905	1736	(88-93) UAAGUU; (193-198) **CAAGUC**; (218-223) UAAGUU; (542-547) **UAAGUA**; (1484-1489) **UAAGUA**
	FBtr0337038	1736	(88-93) UAAGUU; (193-198) **CAAGUC**; (218-223) UAAGUU; (542-547) **UAAGUA**; (1484-1489) **UAAGUA**
	FBtr0337039	1736	(88-93) UAAGUU; (193-198) **CAAGUC**; (218-223) UAAGUU; (542-547) **UAAGUA**; (1484-1489) **UAAGUA**
alpha-Man-IIb	FBtr0083179	1181	(782-787) UAAGUU
	FBtr0334593	1181	(782-787) UAAGUU
	FBtr0345238	1181	(782-787) UAAGUU
Bsg25A	FBtr0077423	172	(117-122) UAAGUC
Surf4	FBtr0083117	611	(440-445) UAAGUU
Tom	FBtr0075620	403	(266-271) UAAGUU
alphaTub67C	FBtr0076393	82	
amx	FBtr0071331	292	
	FBtr0307281	292	
betaTub85D	FBtr0082046	270	

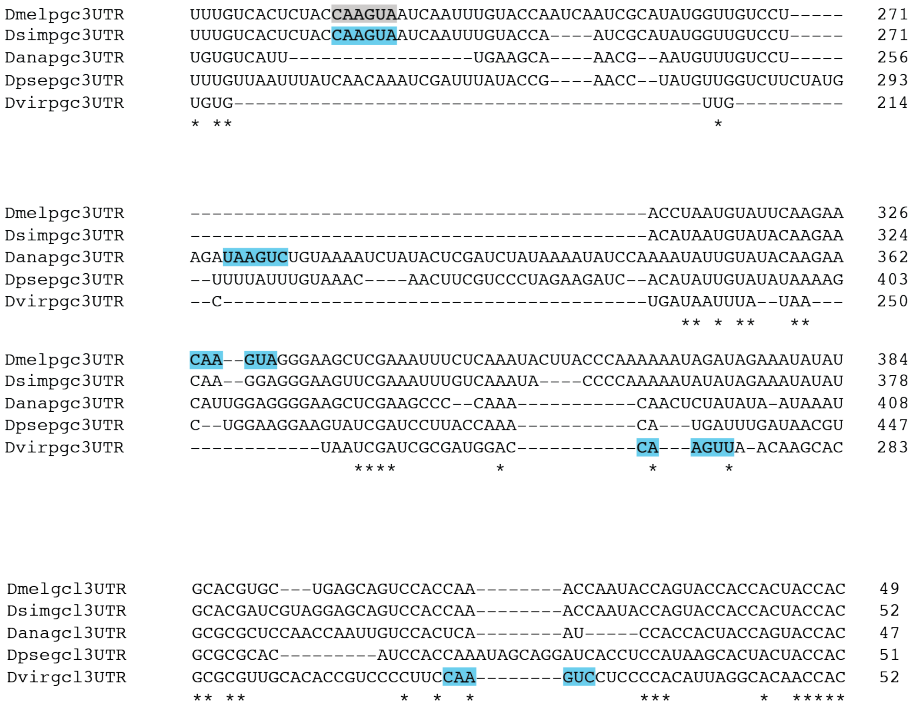
(Table continues)

Catsup	FBtr0081143	264	

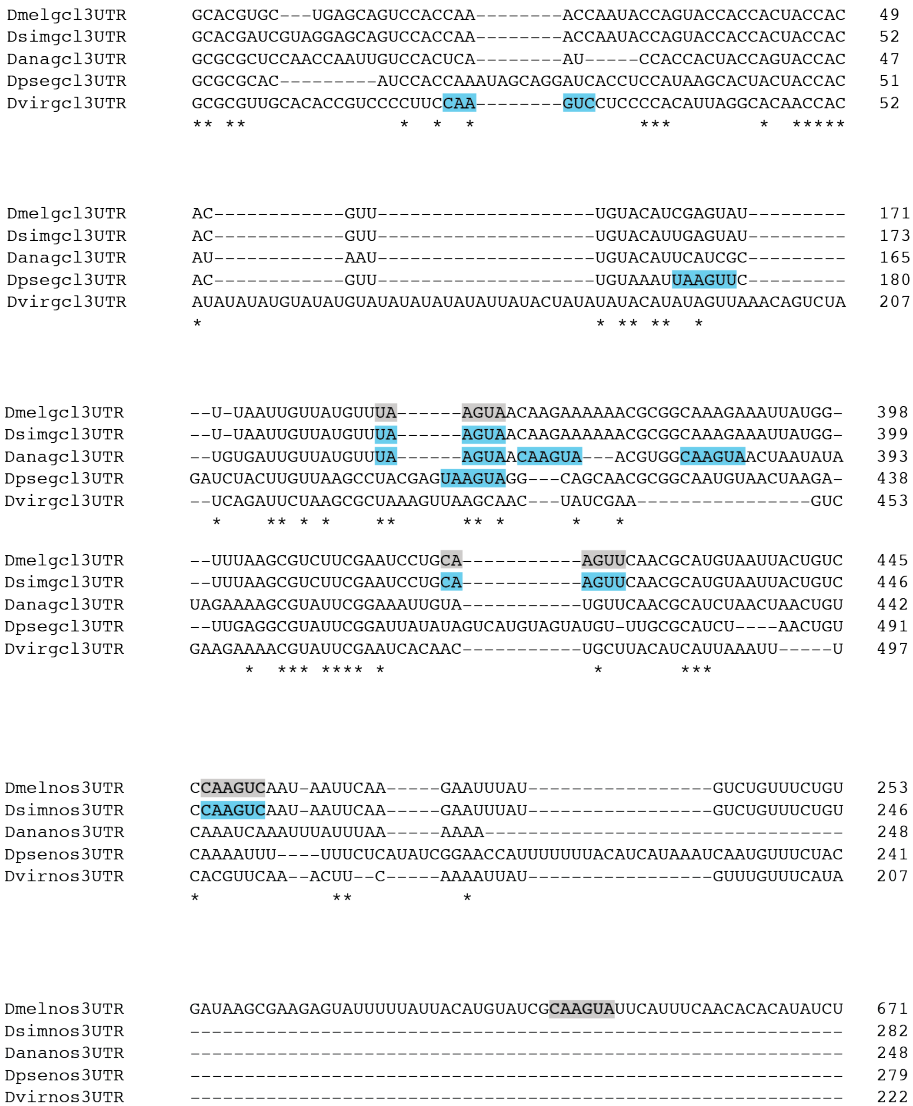
RNA	Transcript ID	3'UTR Length	Sequence(s)
CHORD	FBtr0084554	55	
fau	FBtr0344444	501	
key	FBtr0072428	41	
	FBtr0305672	41	
Lsd-1	FBtr0084455	139	
	FBtr0084456	139	
	FBtr0084457	139	
	FBtr0335002	139	
NitFhit	FBtr0072483	67	
odd	FBtr0077557	770	
Orc6	FBtr0088482	107	
ppl	FBtr0078412	198	
pug	FBtr0082264	168	
	FBtr0082265	168	
	FBtr0082266	168	
	FBtr0100144	168	
	FBtr0305662	168	
sktl	FBtr0071563	710	
	FBtr0342801	178	
	FBtr0344196	178	
miple2	FBtr0072506	129	
	FBtr0332836	129	
	FBtr0332837	129	
	FBtr0332838	129	
	FBtr0332839	129	
slp1	FBtr0077499	347	
yellow-h	FBtr0089115	29	
zetaTry	FBtr0088164	71	

Column 1, 50 RNAs that are pole cell excluded identified by Fly-FISH (Lecuyer et al., 2007; Wilk et al., 2016); **column 2**, all known transcript isoforms; **column 3**, 3'UTR length; **column 4**, regions within 3'UTR, in parentheses, that contain the sequences corresponding to the homotypic clustering motif (Eagle et al., 2018), and blank, no sequence identified. Sequences in bold are identical to the motifs in *pgc*, *gcl* and *nos*. *alphaTub67C*, *amx*, *betaTub85D*, *Catsup*, *CHORD*, *fau*, *key*, *Lsd-1*, *NitFhit*, *odd*, *Orc6*, *ppl*, *pug*, *sktl*, *miple2*, *slp1*, *yellow-h* and *zetaTry* 3'UTRs did not contain any sequences that matched the motif.

A


Dmelpgc3UTR UUUGUCACUCUACCAAGUAAUCAAUUUGUACCAAUCAAUCGCAUAUGGUUGUCCU----- 271
Dsimpgc3UTR UUUGUCACUCUACCAAGUAAUCAAUUUGUACCA----AUCGCAUAUGGUUGUCCU----- 271
Danapgc3UTR UGUGUCAUU----------UGAAGCA----AACG--AAUGUUUGUCCU----- 256
Dpsepgc3UTR UUUGUUAAUUUAUCAACAAAUCGAUUUAUACCG----AACC--UAUGUUGGUCUUCUAUG 293
Dvirpgc3UTR UGUG----------UUG---------- 214
Dmelpgc3UTR ----------ACCUAAUGUAUUCAAGAA 326
Dsimpgc3UTR ----------ACAUAAUGUAUACAAGAA 324
Danapgc3UTR AGAUAAGUCUGUAAAAUCUAUACUCGAUCUAUAAAAUAUCCAAAAUAUUGUAUACAAGAA 362
Dpsepgc3UTR --UUUUAUUUGUAAAC----AACUUCGUCCCUAGAAGAUC--ACAUAUUGUAUAUAAAAG 403
Dvirpgc3UTR --C----------UGAUAAUUUA--UAA--- 250
Dmelpgc3UTR CAA--GUAGGGAAGCUCGAAAUUUCUCAAAUACUUACCCAAAAAAUAGAUAGAAAUAUAU 384
Dsimpgc3UTR CAA--GGAGGGAAGUUCGAAAUUUGUCAAAUA----CCCCAAAAAUAUAUAGAAAUAUAU 378
Danapgc3UTR CAUUGGAGGGGAAGCUCGAAGCCC--CAAA----------CAACUCUAUAUA-AUAAAU 408
Dpsepgc3UTR C--UGGAAGGAAGUAUCGAUCCUUACCAAA----------CA---UGAUUUGAUAACGU 447
Dvirpgc3UTR ----------UAAUCGAUCGCGAUGGAC----------CA---AGUUA-ACAAGCAC 283

B


Dmelgcl3UTR GCACGUGC---UGAGCAGUCCACCAA--------ACCAAUACCAGUACCACCACUACCAC 49
Dsimgcl3UTR GCACGAUCGUAGGAGCAGUCCACCAA--------ACCAAUACCAGUACCACCACUACCAC 52
Danagcl3UTR GCGCGCUCCAACCAAUUGUCCACUCA--------AU-----CCACCACUACCAGUACCAC 47
Dpsegcl3UTR GCGCGCAC---------AUCCACCAAAUAGCAGGAUCACCUCCAUAAGCACUACUACCAC 51
Dvirgcl3UTR GCGCGUUGCACACCGUCCCCUUCCAA--------GUCCUCCCCACAUUAGGCACAACCAC 52
Dmelgcl3UTR AC----------GUU----------UGUACAUCGAGUAU--------- 171
Dsimgcl3UTR AC----------GUU----------UGUACAUUGAGUAU--------- 173
Danagcl3UTR AU----------AAU----------UGUACAUUCAUCGC--------- 165
Dpsegcl3UTR AC----------GUU----------UGUAAAUUAAGUUC--------- 180
Dvirgcl3UTR AUAUAUAUGUAUAUGUAUAUAUAUAUUAUACUAUAUAUACAUAUAGUUAAACAGUCUA 207
Dmelgcl3UTR --U-UAAUUGUUAUGUUUA------AGUAACAAGAAAAAACGCGGCAAAGAAAUUAUGG- 398
Dsimgcl3UTR --U-UAAUUGUUAUGUUUA------AGUAACAAGAAAAAACGCGGCAAAGAAAUUAUGG- 399
Danagcl3UTR --UGUGAUUGUUAUGUUUA------AGUAACAAGUA---ACGUGGCAAGUAACUAAUAUA 393
Dpsegcl3UTR GAUCUACUUGUUAAGCCUACGAGUAAGUAGG---CAGCAACGCGGCAAUGUAACUAAGA- 438
Dvirgcl3UTR --UCAGAUUCUAAGCGCUAAAGUUAAGCAAC---UAUCGAA----------GUC 453
Dmelgcl3UTR --UUUAAGCGUCUUCGAAUCCUGCA----------AGUUCAACGCAUGUAAUUACUGUC 445
Dsimgcl3UTR --UUUAAGCGUCUUCGAAUCCUGCA----------AGUUCAACGCAUGUAAUUACUGUC 446
Danagcl3UTR UAGAAAAGCGUAUUCGGAAAUUGUA----------UGUUCAACGCAUCUAACUAACUGU 442
Dpsegcl3UTR --UUGAGGCGUAUUCGGAUUAUAUAGUCAUGUAGUAUGU-UUGCGCAUCU----AACUGU 491
Dvirgcl3UTR GAAGAAAACGUAUUCGAAUCACAAC----------UGCUUACAUCAUUAAAUU-----U 497

C

Dmelnos3UTR CCAAGUCAAU-AAUUCAA-----GAAUUUAU----------GUCUGUUUCUGU 253
Dsimnos3UTR CCAAGUCAAU-AAUUCAA-----GAAUUUAU----------GUCUGUUUCUGU 246
Dananos3UTR CAAAUCAAAUUUAUUUAA-----AAAA---------- 248
Dpsenos3UTR CAAAAUUU----UUUCUCAUAUCGGAACCAUUUUUUUACAUCAUAAAUCAAUGUUUCUAC 241
Dvirnos3UTR CACGUUCAA--ACUU--C-----AAAAUUAU----------GUUUGUUUCAUA 207
Dmelnos3UTR GAUAAGCGAAGAGUAUUUUUAUUACAUGUAUCGCAAGUAUUCAUUUCAACACACAUAUCU 671
Dsimnos3UTR ---------- 282
Dananos3UTR ---------- 248
Dpsenos3UTR ---------- 279
Dvirnos3UTR ---------- 222

Figure 3.6 3'UTR sequence alignments of *pgc*, *gcl* and *nos*

(A) Sequence alignment of *pgc* 3'UTR from FBtr0112519 (*D. melanogaster*), FBtr0353986 (*D. simulans*), FBtr0117079 (*D. ananassae*), FBtr0279950 (*D. pseudoobscura*), and FBtr0235704 (*D. virilis*). **(B)** Sequence alignment of *gcl* 3'UTR from FBtr0088710 (*D. melanogaster*), FBtr0357586 (*D. simulans*), FBtr0387874 (*D. ananassae*), FBtr0277627 (*D. pseudoobscura*), and FBtr0435510 (*D. virilis*). **(C)** Sequence alignment of *nos* 3'UTR from FBtr0083732 (*D. melanogaster*), FBtr0219184 (*D. simulans*), FBtr0120753 (*D. ananassae*), FBtr0380189 (*D. pseudoobscura*), and FBtr0437357 (*D. virilis*). The homotypic clustering motifs (grey boxes) in *D. melanogaster* and versions of the *D. melanogaster* homotypic clustering element (light blue boxes) in *D. simulans*, *D. ananassae*, *D. pseudoobscura* and *D. virilis* are shown. Only portions of the 3'UTR alignments that contain the relevant sequences are shown. Asterisks denote conserved bases.

Figure 3.7 Sequence alignment of *D. melanogaster* and *D. virilis nos* 3'UTRs

Dvirnos3UTRe is the 806-nt sequence obtained from amplifying the *nos* 3'UTR from *D. virilis* unfertilized egg samples where "e" denotes extended. In blue is the sequence that extends beyond the longest *D. virilis* nos 3'UTR sequence (FBtr0437357). The homotypic clustering motifs (grey boxes) in *D. melanogaster nos* and one version of the *D. melanogaster* homotypic clustering element (light blue boxes) in *D. virilis nos* are shown. Asterisks denote conserved bases.

Table 3.4 Homotypic clustering motif sequences identified in *pgc*, *gcl* and *nos* 3'UTR

RNA	Species	Transcript ID	3'UTR Length	Sequence(s)
pgc	*D. melanogaster*	FBtr0112519	401	(230-235) **CAAGUA**; *(327-332) CAAGUA
		FBtr0112520	401	(230-235) **CAAGUA**; *(327-332) CAAGUA
		FBtr0342987	401	(230-235) **CAAGUA**; *(327-332) CAAGUA
	D. simulans	FBtr0353986	392	(234-239) CAAGUA
	D. ananassae	FBtr0387874	442	(306-311) UAAGUC
	D. pseudoobscura	FBtr0277627	477	
	D. virilis	FBtr0435510	312	(269-274) CAAGUU
gcl	*D. melanogaster*	FBtr0088710	524	(363-368) **UAAGUA**; (420-425) **CAAGUU**
		FBtr0339337	524	(363-368) **UAAGUA**; (420-425) **CAAGUU**
	D. simulans	FBtr0357586	535	(364-369) UAAGUA; (421-426) CAAGUU
	D. ananassae	FBtr0117079	523	(360-365) UAAGUA; (367-372) CAAGUA; (379-384) CAAGUA
	D. pseudoobscura	FBtr0279950	587	(406-411) UAAGUA
	D. virilis	FBtr0235704	516	(24-29) CAAGUC
nos	*D. melanogaster*	FBtr0083732	880	(218-223) **CAAGUC**; (645-650) **CAAGUA**
		FBtr0335019	880	(218-223) **CAAGUC**; (645-650) **CAAGUA**
	D. simulans	FBtr0219184	282	
		FBtr0351812	282	
	D. ananassae	FBtr0120753	248	
		FBtr0383542	248	
	D. pseudoobscura	FBtr0380189	279	
	D. virilis	FBtr0239461	219	
		FBtr0437357	222	
	*extended 3'UTR	N/A	806	(500-505) CAAGUU

Column 1, *pgc*, *gcl* and *nos* RNAs; **column 2**, *Drosophila* species with annotated transcripts; **column 3**, IDs of all known transcripts; **column 4**, length of 3'UTR sequences; **column 5**, regions within 3'UTR, in parentheses, that contain the sequences corresponding to the homotypic clustering motif (Eagle et al., 2018) in *D. melanogaster* and similar sequences in other *Drosophila* species. Asterisks denote a sequence found in *D. melanogaster pgc* 3'UTR that does not contribute to homotypic clustering. Sequences in black and bold contribute to homotypic clustering in *D. melanogaster* (Eagle et al., 2018). Versions of the *D. melanogaster* homotypic clustering element, (C/U)AAGU(A/C/U), that are present in the 3'UTRs of other *Drosophila* species are shown in blue.

3.3 Discussion

In this study, we showed that the majority of *pgc*, *gcl* and *nos* is translationally repressed in the early embryo (**Fig 3.1**). Using a deletion mutation analysis, we determined that the *gcl* 3'UTR possesses multiple functionally redundant localization elements, and the distal region of the *pgc* 3'UTR is necessary and sufficient to direct localization (**Fig. 3.2, 3.3**). Furthermore, the localization of the protein components of germ plasm, Osk, Tud and Vas, and most of the mRNAs tested is conserved in related species where they also accumulate around posterior nuclei in early embryos (**Fig. 3.4, 3.5**). Sequences that resemble the homotypic clustering is found in almost all 3'UTRs of transcripts that localize to RNA islands in *D. melanogaster* and in the 3'UTRs of *pgc*, *gcl* and *nos* in *D. simulans* (**Table 3.1, 3.2, 3.3**).

While this thesis chapter was in preparation, a different research group also characterized the localization elements of *pgc* and *gcl* (Eagle et al., 2018). We used a binary assessment of posterior localization, where the RNA was either enriched or not, in the characterization of sequences containing localization signals. In contrast, they were able to distinguish between two classes of localization signals that function jointly to ensure the localization of pole plasm RNAs: the targeting element that directs RNAs to the pole plasm and the clustering element for RNAs to self-recruit and accumulate as clusters (Eagle et al., 2018). Based on their results, we were able to further delineate the targeting and clustering elements within the *pgc* and *gcl* 3'UTR sequences (**Fig. 3.8**). Localization elements in the *pgc* 3'UTR are located at nucleotides 1-150, 150-249, including the homotypic clustering motif at 230-235 and 255-392 (Eagle et al., 2018). A notable difference between the constructs generated in this study compared to that of Eagle et al. is the use of the *α-tub* 3'UTR to test short 3'UTR fragments. While *α-tub* is not enriched in the posterior, its 3'UTR contains two putative homotypic clustering motifs, (187-192) CAAGUU and (203-208) UAAGUU, one of which is identical to the *gcl* (420-425) CAAGUU clustering element. The *pgc(Δ168-401)* transgene failed to localize to the posterior because it was lacking the clustering element located at 230-235 despite having a targeting element in 1-150, consistent with the localization result of the *pgc(1-150)* transgene (**Fig. 3.8A**). Although there is a sequence that resembles a homotypic clustering element at 327-332, it likely has weak clustering ability; *pgc(255-292)* did not accumulate in the posterior. In contrast, *α-tub+pgc(238-401)* localized to the

posterior, likely due to the additional clustering elements that are contributed by the *α-tub* 3'UTR (**Fig. 3.8A**). Because *pgc(238-321)+α-tub* and *pgc(291-360)+α-tub* did not localize, the targeting element in 255-392 either spans a large region and each construct contains a truncated form of it or the sequence in 361-392 is the minimal targeting element (**Fig. 3.8A**).

The *gcl* 3'UTR contains multiple weak localization elements that are spread across 4 regions, 1-149, 150-247, 276-399 and 400-524, with two homotypic clustering motifs located at 364-369 and 419-424 (Eagle et al., 2018). Based on our deletions mutation constructs, *gcl(Δ1-100)*, *gcl(Δ101-200)*, *gcl(Δ201-300)*, *gcl(Δ301-401)*, and *gcl(Δ401-524)*, we also concluded that there are multiple functionally redundant localization elements in the *gcl* 3'UTR. Each of those constructs contain at least one targeting element and one clustering element (**Fig. 3.8B**). The *gcl(150-399)* transgenic RNA did not localize to the posterior even though it has the 364-369 clustering element and the 150-249 targeting element but *gcl(150-399)x2* did, suggesting that there are weak localization elements in 150-399 and additional targeting elements are required to achieve posterior enrichment (Eagle et al., 2018). Given that *gcl(1-399)* and *gcl(Δ401-524)* can localize with a single 364-369 clustering element each, this further supports that the targeting elements in *gcl(150-399)* are collectively too weak to result in accumulation (**Fig. 3.8B**). Our constructs, *gcl(101-200)+α-tub* and *gcl(201-300)+α-tub*, resulted in posterior accumulation, indicating that each *gcl* 3'UTR fragment must possess at least one targeting element while the *α-tub* 3'UTR provided two clustering elements. Because 101-200 and 201-300 overlap with 150-399, sequences that did not result in localization, we considered that there are two targeting elements present in 150-249, 150-200 and 201-249 (**Fig. 3.8B**). Therefore, *gcl(101-200)+α-tub* has targeting elements in 101-149 and 150-200 and *gcl(201-300)+α-tub* has targeting elements in 201-249 and another in 250-300 (**Fig. 3.8B**). Our result for *gcl(201-300)+α-tub* contradicts theirs for *gcl(150-399)*, but we propose that the two putative clustering elements in *α-tub* 3'UTR can partially compensate for a weak targeting element (**Fig. 3.8B**). Collectively, we conclude that a single targeting element is insufficient for localization of *gcl*, and multiple targeting elements in addition to homotypic clustering elements are required for localization. Further work will be required to validate the targeting ability of the *gcl* 3'UTR sequences at 101-149, 150-200, 201-249 and 250-300, and the clustering ability of the *α-tub* 3'UTR. We can generate constructs to test the targeting abilities of the proposed regions by placing several copies in tandem. Chimeric RNAs

that fuse the minimal region containing clustering elements predicted in the *α-tub* 3'UTR to known targeting elements will determine if sequences in *α-tub* 3'UTR are competent for cluster growth.

While the targeting elements in *pgc* and *gcl* have been delineated and the homotypic clustering motif has been identified (Eagle et al., 2018), there remains unresolved issues. It has yet to be determined if the localization signal to target transcripts to polar granules exists as a primary sequence, similarly to the clustering element, or as a secondary structure, and if it is shared amongst all the RNAs with similar patterns of posterior localization. The homotypic clustering motif is conserved amongst *pgc*, *gcl* and *nos*, with similar 6-nt sequences in the 3'UTRs of other pole plasm transcripts (Eagle et al., 2018). Consequently, this raises the question of how self-association to drive cluster growth is accomplished if the same motif is present in many different pole plasm RNAs. These localization signals are likely recognized by *trans*-acting factors with the possibility of different sets of proteins orchestrating the growth of specific RNA clusters. Although the clustering motif in *pgc* was deemed sequence-specific, it was predicted to contribute to the stem of a stem loop that is conserved across *Drosopholids* (Eagle et al., 2018). Furthermore, only RNAs that are targeted to polar granules are capable of seeding clusters, but not those that are found in the bulk cytoplasm of the embryo (Niepielko et al., 2018). Therefore, there is likely remodeling of the RNA architecture by RNA helicases to expose the homotypic clustering motif to enable self-recruitment once it reaches the polar granules. Candidate RNA helicases include Vasa, a core polar granule protein, Belle, both of which enrich in the posterior of the early embryo (Hay et al., 1988; Johnstone et al., 2005; Trcek et al., 2015).

Many aspects of the localization of germ plasm RNAs are conserved, highlighting the importance of this process to ensure proper germline development. While the majority of the RNAs tested localized to the posterior in related species, there were three exceptions: *Tlk*, *bru1* and *gwl*. Tlk is a serine/threonine kinase that maintains follicle cell morphology and its expression in polar cells regulates border cell migration by activating JAK/STAT (Xiang et al., 2016; Yeh et al., 2015). Bruno, encoded by *bru1*, is an RNA-binding protein that represses the translation of numerous RNAs, including *osk*, *gurken* (*grk*), *gcl* and *pgc*, during oogenesis (Chekulaeva et al., 2006; Filardo and Ephrussi, 2003; Flora et al., 2018; Moore et al., 2009; Nakamura et al., 2004). Gwl is a kinase that is required for female meiosis in the oocyte (Archambault et al., 2007). Given the important roles that these proteins encoded by germ plasm RNAs play in oogenesis, we consider the

possibility that these RNAs are localized to the posterior in *D. virilis* embryos but they do not possess sequences that are complementary to the RNA probes used in this study or a more sensitive and quantitative technique to detect enrichment in the posterior pole plasm is required. For RNAs with conserved localization, we can begin to decipher their targeting elements, which is also likely to be conserved in sequence or structure, using a similar approach that led to the identification of the homotypic clustering motif.

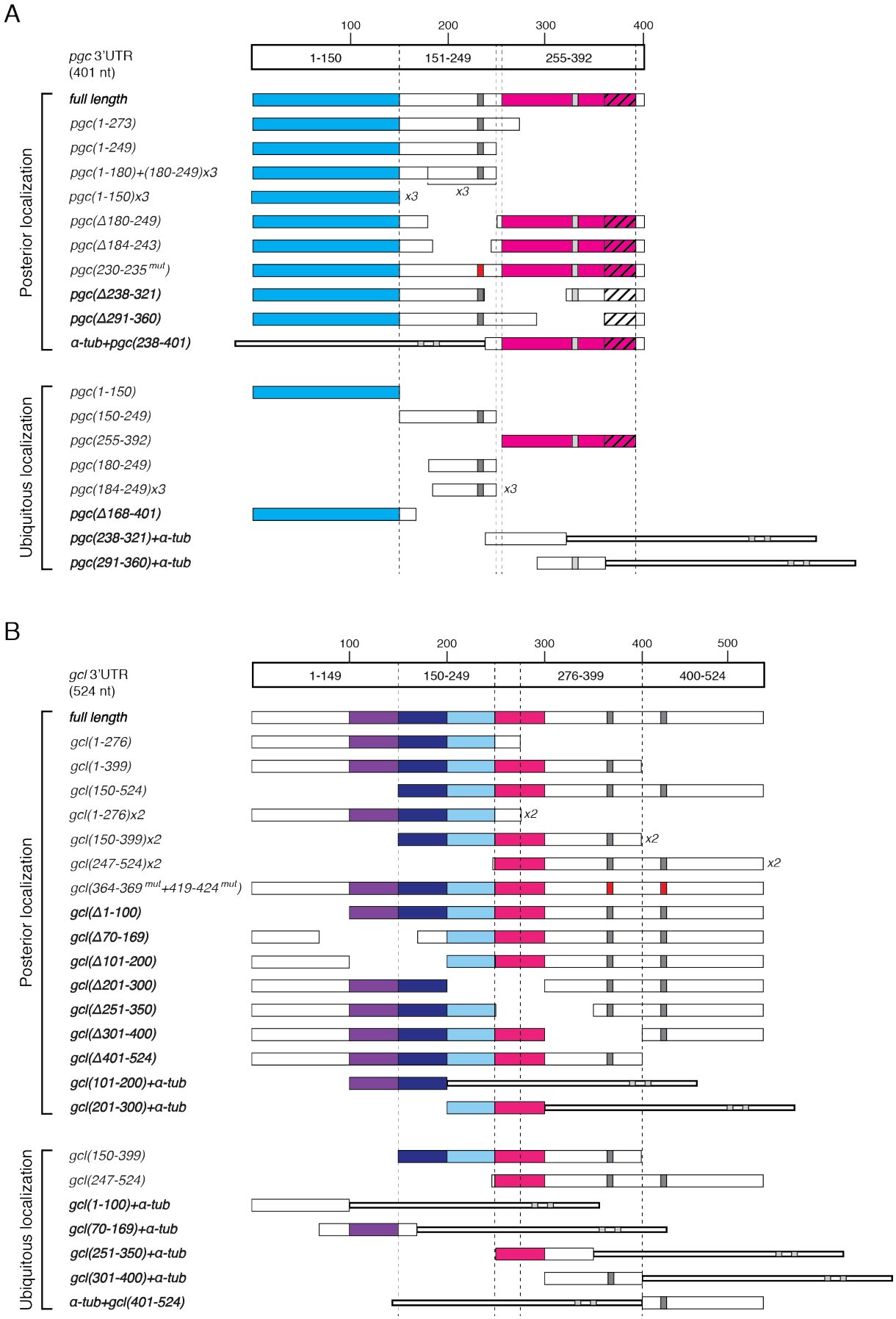

A
100
200
300
400
pgc 3'UTR
(401 nt)
1-150
151-249
255-392
Posterior localization
full length
pgc(1-273)
pgc(1-249)
pgc(1-180)+(180-249)x3
pgc(1-150)x3
pgc(Δ180-249)
pgc(Δ184-243)
pgc(230-235 mut)
pgc(Δ238-321)
pgc(Δ291-360)
α-tub+pgc(238-401)
x3
x3
Ubiquitous localization
pgc(1-150)
pgc(150-249)
pgc(255-392)
pgc(180-249)
pgc(184-249)x3
pgc(Δ168-401)
pgc(238-321)+α-tub
pgc(291-360)+α-tub
x3
B
100
200
300
400
500
gcl 3'UTR
(524 nt)
1-149
150-249
276-399
400-524
Posterior localization
full length
gcl(1-276)
gcl(1-399)
gcl(150-524)
gcl(1-276)x2
gcl(150-399)x2
gcl(247-524)x2
gcl(364-369 mut+419-424 mut)
gcl(Δ1-100)
gcl(Δ70-169)
gcl(Δ101-200)
gcl(Δ201-300)
gcl(Δ251-350)
gcl(Δ301-400)
gcl(Δ401-524)
gcl(101-200)+α-tub
gcl(201-300)+α-tub
x2
x2
x2
Ubiquitous localization
gcl(150-399)
gcl(247-524)
gcl(1-100)+α-tub
gcl(70-169)+α-tub
gcl(251-350)+α-tub
gcl(301-400)+α-tub
α-tub+gcl(401-524)

Figure 3.8 Summary of posterior localization results of *pgc* and *gcl* reporter constructs used in this study and Eagle et al.

Schematic diagrams of **(A)** *pgc* 3'UTR and **(B)** *gcl* 3'UTR transgenic constructs used to delineate posterior localization elements in (Eagle et al., 2018) and in this study (names in bold). **(A)** Targeting elements are found at 1-150 (cyan) and 255-392 (magenta). The putative minimal localization element at 361-392 is indicated by the hash box. The clustering element (grey box) and its mutated sequence (red box) is located at 230-235. The second putative homotypic clustering motif in *pgc* is located at 327-332 (light grey box). **(B)** Regions in the *gcl* 3'UTR containing targeting elements are 101-149 (purple), 150-200 (blue), 201-249 (light blue) and 250-300 (pink). Clustering elements (grey boxes) and their mutated sequences (red boxes) are located at 364-369 and 419-424. The dashed lines delineate the sequences containing localization elements identified in Eagle et al. *αTub84B* 3'UTR is shown as the narrow bar with the two putative clustering elements (light grey boxes) at 187-192 and 203-208.

3.4 Materials and Methods

3.4.1 Polysome fractionation and reverse transcription quantitative PCR (RT-qPCR)

0–2-hour old embryos from *yw* flies were pulverized in liquid nitrogen and lysed in hypotonic buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.5% Triton X-100, 1 mM DTT, 20U/ml SuperasIn (Ambion by Life Technologies), 20 µg/ml emetine, and 50 µM GMP-PNP (Sigma Aldrich)). Lysates were cleared by centrifugation at 3,000 × *g* for 10 min at 4 ºC followed by second round of centrifugation at 20,000 × *g* for 10 min at 4 ºC. 300 µg total RNA was loaded onto 10–50% (wt/vol) sucrose density gradients (50 mM Tris pH 7.5, 250 mM NaCl, 15 mM $MgCl_2$, 20 U/ml SuperasIn, and 20 µg/ml emetine) separated by centrifugation at 36 krpm for 3h at 4 ºC in an SW40 rotor (Beckman Coulter). Fractions were collected at 35 second intervals using an ISCO gradient fractionation system. The absorbance of OD254 was continuously recorded with a Foxy JR Fractionator (Teledyne ISCO). Total RNA was extracted from the fractions using TRIzol® Reagent (Thermo Fisher Scientific), treated with Turbo DNase (Ambion), then the RNA was precipitated with acid-phenol:chloroform (Ambion). Reverse transcription of mRNA was performed from 500 mg RNA using Maxima H Minus First Strand cDNA synthesis kit (Thermo Fisher Scientific) using a 1:1 mixture of oligo dT and random hexamers. RT-qPCR was carried out using DyNAmo Flash SYBR Green (Thermo Fisher Scientific) in a CFX96 Real-Time machine (Bio-Rad). mRNA was quantified using the standard curve method and the percentage of total mRNA for each fraction was calculated. The following primers were used for RT-qPCR:

pgc F 5'-CCGGTCATCGCGGATAGATGG-3'
pgc R 5'-TCGCCTCTGAAGGCTGGTAGT-3'
gcl F 5'-TGCCCTCGCACAATATGTAGGC-3'
gcl R 5'-AACTTGCAGGATTCGAAGACGCT-3'
nos F 5'-GGTTTGCAGGCCCAAACAGC-3'
nos R 5'-GCAGTGGCGGCTGATCTCTT-3'

3.4.2 Transgenic fly generation and 3'UTR deletion analysis

The 3'UTR sequences of *gcl* and *pgc* were amplified from cDNA clones LD23660 and RE14873, respectively, with NotI and PstI restriction sites added to the 5' and 3' ends, respectively, and then cloned into a pENTR vector (Life Technologies). Deletions were generated by PCR-based mutagenesis. The 3'UTR of *αTub84B* was amplified from genomic DNA extracted from whole *D. melanogaster* flies with NotI/BamHI and PstI restriction sites added to the 5' and 3' ends, respectively. BamHI restriction sites were added to the 3' ends of *gcl* and *pgc* 3'UTR sequences that were subsequently cloned upstream of the *αTub84B* 3'UTR. The 3'UTR of *αTub84B* was cloned upstream of *gcl* and *pgc* 3'UTR fragments containing their respective polyA signals, in which case, the XhoI restriction site was added to the 5' ends of *gcl* and *pgc* 3'UTR sequences, the polyA signal of *αTub84B* was removed and XhoI/PstI restriction sites were added to the 3' end of the *αTub84B* 3'UTR sequence. NotI and PstI digested 3'UTR sequences were then cloned into an pUASp-eGFP plasmid with a deletion of the K10 terminator sequence. All primers used for cloning are found in Table 3.5. Transgenic flies were generated by P element-mediated germline transformation (Rubin and Spradling, 1982). Transgenes were expressed with *nos*-Gal4-VP16 in a wild type background. 0–3 hour old embryos were collected for fluorescent *in situ* hybridization using an antisense *eGFP* probe to determine the localization of the transgene. The following primers were used for probe synthesis:

eGFP T3 F 5'-AATTAACCCTCACTAAAGGGAGAGTGAGCAAGGGCGAGGAGC-3'
eGFP T7 R 5'-TAATACGACTCACTATAGGGAGACTTGTACAGCTGCTCCATGCCG-3'

3.4.3 RNA fluorescent *in situ* hybridization (RNA-FISH) and immunofluorescence (IF)

0–3 hour old embryos were collected from *D. melanogaster* (*OregonR* (*OrR*)) and *D. simulans* flies and 0–5-hour old embryos were collected from *D. virilis* flies on yeasted grape juice plates. Embryos were processed for RNA-FISH and IF as previously described (Lecuyer et al., 2008). Antisense RNA probes for FISH on *D. melanogaster* embryos were synthesized with T3 or T7 RNA polymerase on amplified cDNA sequences from the *Drosophila* gene collection 1 and 2 (DGC1 and 2) bacterial cDNA libraries. cDNA clones used for probe synthesis are as follows: *pgc*

(RE14873), *gcl* (LD23660), *nos* (LD32741), *CycB* (LD23613), *exu* (LD26657), *Tao* (LD21241), *eIF5* (GM02147), *Pi3K21B* (LD42724), *jvl* (LD41490), *pum* (RE63138), *Bsg25D* (LD21844), *Tlk* (GH07910), *bru1* (LD29068), *gwl* (LD35132). Genomic DNA was extracted from whole *D. simulans* and *D. virilis* female flies to generate species specific RNA probes. *D. simulans* and *D. virilis* primers are found in Tables 3.6 and 3.7, respectively. *D. melanogaster* RNA probes against *Tao*, *eIF5*, *Pi3K21B*, *jvl*, *pum*, *Bsg25D*, *Tlk*, *bru1*, and *gwl* were used for FISH on *D. simulans* embryos. Primary antibodies used for IF are rabbit anti-Osk #3842 (1:500), affinity purified rabbit anti-Tud (1:500) and rat anti-Vas #4 (1:500).

3.4.4 Confocal microscopy

Images were acquired on a Leica SP8 confocal microscope using a 40× dry objective. The posterior end of embryos was imaged in single focal planes at 1.2× magnification with the pinhole set to 1.5 airy units at a resolution of 1024 × 1024 pixels.

3.4.5 3'UTR analysis

3'UTR sequences were obtained from FlyBase for *D. melanogaster* (Release 6.25), *D. simulans* (Release 2.02), *D. ananassae* (Release 1.06), *D. pseudoobscura* (Release 3.04) and *D. virilis* (Release 1.07). Sequence alignment was performed with Clustal Omega (McWilliam et al., 2013).

Unfertilized eggs were collected from virgin *D. virilis* females on grape juice plates. Total RNA was extracted from the eggs/embryos using TRIzol® Reagent (Thermo Fisher Scientific), treated with Turbo DNase (Ambion), then the RNA was precipitated with acid-phenol:chloroform (Ambion). Reverse transcription of mRNA was performed from 3 µg RNA using Maxima H Minus First Strand cDNA synthesis kit (Thermo Fisher Scientific) using an oligo dT primer containing an anchor sequence. The *D. virilis nos* 3'UTRs was amplified using gene-specific primer and the anchor primer, and then submitted for sequencing.

anchor dT 5'- CGTATGATCGGTTAACCGTCTTTTTTTTTTT -3'
anchor R 5'- CGTATGATCGGTTAACCGTC -3'
D.vir nos F 5'- TGTTACAATGGAGGATGCC -3

Table 3.5 Cloning primers used for transgene generation

Construct	Primer	Sequence
gcl 3'UTR full length	*gcl* full length NotI F	CACCGCGGCCGCGCACGTGCTGAGCAGTCC
	gcl full length PstI R	ATATCTGCAGTTTTTATTAAGTGTAATCTTTAATAAATGATCACTCAAG
gcl 3'UTR Δ1-100	*gcl* Δ1-100 NotI F	CACCGCGGCCGCGTAAATTCGTTTATCGCCATCG
	gcl full length PstI R	ATATCTGCAGTTTTTATTAAGTGTAATCTTTAATAAATGATCACTCAAG
gcl 3'UTR Δ20-219	*gcl* Δ20-219 F	ATCGTTTACATCACTAGTATCATGTAATGACGC
	gcl Δ20-219 R	TGGACTGCTCAGCACGTG
gcl 3'UTR Δ101-200	*gcl* Δ101-200 F	CAGATCTATATTTTTGTATATCGTTTACATCACTAGTATCATGTAATGAC
	gcl Δ101-200 R	AAAACAACTACAACGAGCCCGAACTG
gcl 3'UTR Δ201-300	*gcl* Δ201-300 F	CACAATATGTAGGCGTTGCAAATAGTTTG
	gcl Δ201-300 R	CGCATTCCAAAACGATTGCGG
gcl 3'UTR Δ251-350	*gcl* Δ251-350 F	AATTGTTATGTTTAAGTAACAAGAAAAAACGCGGC
	gcl Δ251-350 R	GTCATTACATGATACTAGTGATGTAAACGATATACAAAAATATAGATCTGC
gcl 3'UTR Δ301-400	*gcl* Δ301-400 F	TAAGCGTCTTCGAATCCTGCAAGTTC
	gcl Δ301-400 R	CGAGGGCACTAGAGTGTGCGTC
gcl 3'UTR Δ401-524	*gcl* full length NotI F	CACCGCGGCCGCGCACGTGCTGAGCAGTCC
	gcl Δ401-524 polyA PstI R	TATACTGCAGTTTAATAACCATAATTTCTTTGCCGCG
gcl 3'UTR +1-100	*gcl* full length NotI F	CACCGCGGCCGCGCACGTGCTGAGCAGTCC
	gcl 1-100 BamHI R	TATAGGATCCAAAACAACTACAACGAGCCCG
gcl 3'UTR +101-200	*gcl* 101-200 NotI F	CACCGCGGCCGCGTAAATTCGTTTATCGCCATCGCGATTAG
	gcl 101-200 BamHI R	TATAGGATCCCGCATTCCAAAACGATTGCGG
gcl 3'UTR +201-300	*gcl* 201-300 NotI F	CACCGCGGCCGCCAGATCTATATTTTTGTATATCGTTTACATCACTAGTATCATGTAATG
	gcl 201-300 BamHI R	TATAGGATCCCGAGGGCACTAGAGTGTGCGTC
gcl 3'UTR +251-350	*gcl* 251-350 NotI F	CACCGCGGCCGCGCTTTACGTCATCTACTTAAATGGAGTAGACGCACAC
	gcl 251-350 BamHI R	ATATGGATCCAATCTTTAAATTCCAAATTTACAAACTATTTGCAACGC
gcl 3'UTR +301-400	*gcl* 301-400 NotI F	CACCGCGGCCGCCACAATATGTAGGCGTTGCAAATAGTTTG
	gcl 301-400 BamHI R	TATAGGATCCAACCATAATTTCTTTGCCGCGTTTTTTC
gcl 3'UTR +401-524	*gcl* 401-524 XhoI F	CACCCTCGAGTAAGCGTCTTCGAATCCTGCAAGTTC
	gcl full length PstI R	ATATCTGCAGTTTTTATTAAGTGTAATCTTTAATAAATGATCACTCAAG
pgc 3'UTR full length	*pgc* full length NotI F	CACCGCGGCCGCCTGGACCTCCCAAAAGCCAAC
	pgc full length PstI R	TATACTGCAGGAACGATTGCGAATCGAAAATATATTTCTATC
pgc 3'UTR Δ168-401	*pgc* full length NotI F	CACCGCGGCCGCCTGGACCTCCCAAAAGCCAAC
	pgc Δ168-401 polyA PstI R	TATACTGCAGTATATTACTTCAAGAAACTATGCATACGATCG
pgc 3'UTR Δ238-321	*pgc* Δ238-321 F	AAGAACAAGTAGGGAAGCTCGAAATTTCTC
	pgc Δ238-321 R	ATTACTTGGTAGAGTGACAAAACAATGCGAG
pgc 3'UTR Δ291-360	*pgc* Δ291-360 F	CCCAAAAAATAGATAGAAATATATTTTCGATTCGCAATC
	pgc Δ291-360 R	TATTGCCATTTTTAGATCTAGGACAACCATATGC
pgc 3'UTR +238-401	*pgc* 238-401 XhoI F	CACCCTCGAGCAATTTGTACCAATCAATCGCATATGG
	pgc full length PstI R	TATACTGCAGGAACGATTGCGAATCGAAAATATATTTCTATC
pgc 3'UTR +238-321	*pgc* 238-321 NotI F	CACCGCGGCCCAATTTGTCCGTGTATTCAAATGTTTGC
	pgc 238-321 BamHI R	ATATGGATCCGAATACATTAGGTGCAATACCGAAGGCA
pgc 3'UTR +291-360	*pgc* 291-360 NotI F	CACCGCGGCCGCATTTGCCTTCGGTATTGCACCTAATGTATTC
	pgc 291-360 BamHI R	ATATGGATCCTAAGTATTTGAGAAATTTCGAGCTTCCCTACTTGTTC
α-tub 3'UTR full length	*α-tub* NotI/BamHI F	CACCGCGGCCGCGGATCCGCGTCACGCCACTTCAAC
	α-tub PstI R	TATACTGCAGCTTATTTCTGACAACACTGAATCTGGCC
	α-tub ΔpolyA XhoI PstI R	CTGCAGCTCGAGCTTGTGTACACAACTTATCGCCGAGTTA

Table 3.6 Primers for *D. simulans* RNA probe synthesis

Target	Primer	Sequence
GD11659 (pgc)	T3 F	ATTAACCCTCACTAAAGGAACTACAAGACCCGAAAATGTGCG
	T7 R	TAATACGACTCACTATAGGGTATTTGACAAATTTCGAACTTCCCTC
GD10597 (gcl)	T3 F	GAAATTAACCCTCACTAAAGGGCCTTCTTACTACCACCCAGTAC
	T7 R	TGTAATACGACTCACTATAGGGCTCCTGCCCATGATCGATCC
GD19274 (nos)	T3 F	GAAATTAACCCTCACTAAAGGGCCTTCATCTGTTGCTTGTAGTAG
	T7 R	GTAATACGACTCACTATAGGGCCAGCAACTTGGAAGGCAGTG
GD15503 (CycB)	T7 F	AATTAACCCTCACTAAAGGGAGAGTGGGCACAACACTGAAAATGCGTAG
	T3 R	TAATACGACTCACTATAGGGTATTTCCTCTGGCTCTGGCCCAC
GD25256 (exu)	T3 F	ATTAACCCTCACTAAAGGATCTGATATAGTGTCCGATTCGGTATC
	T7 R	TAATACGACTCACTATAGGGTGAATACGCGGCTTGTTTG

Table 3.7 Primers for *D. virilis* RNA probe synthesis

Target	Primer	Sequence
GJ22404 (pgc)	T3 F	ATTAACCCTCACTAAAGGTGTGACTACGAGCACGAATACTCC
	T7 R	TAATACGACTCACTATAGGGTATTTGACAAATTTCGAACTTCCCTC
GJ19779 (gcl)	T3 F	GAAATTAACCCTCACTAAAGGGCGACGACCACACAGTATATATAC
	T7 R	TGTAATACGACTCACTATAGGGCATCCCAGTGGCTGTGTATG
GJ23536 (nos)	T7 F	GTAATACGACTCACTATAGGGCGACGCGATGCAGGATTTTG
	T3 R	GAAATTAACCCTCACTAAAGGGCGAACTTGCCACTGTTGTACC
GJ20298 (CycB)	T3 F	AATTAACCCTCACTAAAGGGAGAACTGAAAGACGTTAAATTGACCAAGGCG
	T7 R	TAATACGACTCACTATAGGGCTCGGAGGGCTTGTAGCTGGC
GJ18765 (exu)	T3 F	AATTAACCCTCACTAAAGGGAGACATCAAATCTAAATCAGAGGTCGCTGCG
	T7 R	TAATACGACTCACTATAGGGGTTGGTGGCCGCAATGGGCAC
GJ16968 (Tao)	T3 F	AATTAACCCTCACTAAAGGGAGAGCCAAGCATCAGCAGGATTCGG
	T7 R	TAATACGACTCACTATAGGGCGCGCAACAGCTGATCCTCCAG
GJ15997 (eIF5)	T3 F	AATTAACCCTCACTAAAGGGAGAGAAGCGATCCGTGCACGTTTGC
	T7 R	TAATACGACTCACTATAGGGAGATCTCATCGATATCTATCTCGTCGCC
GJ24689 (Pi3K21B)	T7 F	TGTAATACGACTCACTATAGGGCGACCCTGATGAAGGATGGC
	T3 R	GAAATTAACCCTCACTAAAGGGCTCCTTGCGTAACTGCAATTCC
GJ24251 (jvl)	T3 F	AATTAACCCTCACTAAAGGGAGACGCCAGCCAGTCAGAGCTTTC
	T7 R	TAATACGACTCACTATAGGGGCTGCTGCTGCTCGGGCG
GJ10104 (pum)	T3 F	ATTAACCCTCACTAAAGCAACTTTTTCGCTCACAGAATCCGG
	T7 R	TAATACGACTCACTATAGGGCTGATACGGCTGATTCTCGGCG
GJ21801 (Bsg25D)	T7 F	TAATACGACTCACTATAGGGCTGGAGGCTGTTAGCGTCGG
	T3 R	AATTAACCCTCACTAAAGGTGCCACAGGTAGCTGATG
GJ19333 (Tlk)	T3 F	ATTAACCCTCACTAAAGGTCCGATGAGCAAAAGTGCC
	T7 R	TAATACGACTCACTATAGGGCTGATCTTCGTTGAGTATGCGC
GJ17715 (bru1)	T7 F	TAATACGACTCACTATAGGGACCTGAAGGCTGCAATTTG
	T3 R	AATTAACCCTCACTAAAGTAAACATTTCTTGTTGCTCTTTTAGC
GJ4542 (gwl)	T3 F	GCGCGAAATTAACCCTCACTAAAGACATTGAATGGCCTGAAGG
	T7 R	TAATACGACTCACTATAGGGTTTAAATGAAGTATGTGGATAACTTTAATG

CHAPTER 4

Conclusion and future directions

4.1 Conclusion

The research presented in this thesis addressed the role of d4E-BP-regulated translational control in the fly head and RNA localization of germ plasm mRNAs in the embryo. In Chapter 2, we found that global translation is unaffected by the *d4E-BP*^{null} mutation, identified specific mRNAs that are preferentially ribosome bound, identified the immune response as the one of the biological processes that is linked to this group of mRNAs and determined that their 5'UTRs of were shorter but more complex. While attempts were made to validate the differential translation of mRNAs in terms of protein levels, we did not detect corresponding changes by immunoblot or by mass spectrometry. In Chapter 3, we found that the distal region of the *pgc* 3'UTR is necessary and sufficient for localization, and the *gcl* 3'UTR possesses multiple redundant localization elements with two regions, at 101-200 and 201-300, that are each sufficient to direct localization. We determined that the protein components of germ granules, Osk, Tud and Vas, how conserved localization in other *Drosophila* species. Similarly, 11 out of the 14 mRNAs used in this study also demonstrated conserved posterior localization in *Drosophila* embryos. We conclude that this mechanism of posterior localization is likely reserved for mRNAs that have important functions in germline development. The following sections will discuss some shortcomings of this research and future experiments.

4.2 Limitations of the ribosome profiling technique

Ribosome profiling is a powerful technique when coupled to RNAseq and it is increasingly being used to study translational control; nevertheless, it does present several limitations (Masvidal et al., 2017). The ribosome-protected fragments do not allow a straightforward way of determining which mRNA isoform is translated or the number of ribosomes that are bound to the mRNA and actively translating it. Due to the first limitation mentioned, we included all mRNA isoforms in the UTR analysis which does not provide the most accurate depiction of the UTRs and their regulation on translation. The ideal setup would allow us to distinguish which isoforms of each mRNA were differentially translated within our ribosome profiling dataset. To address the second limitation mentioned, we can complement the ribosome profiling results with polysome profiling which sediments mRNAs based on the number of bound ribosomes. A comparison of RNAs

extracted from fractions corresponding to “light” polysomes possessing 2-4 ribosomes to “heavy” polysomes possessing 5 or more ribosomes in *Thor* and *Revertant* head samples would provide additional insight into which RNAs are differentially translated. A caveat to this approach is that the monosome peak can contain translating RNAs.

4.3 Validation of preferentially ribosome bound mRNA targets

Although attempts were made to validate the mRNAs that were preferentially ribosome associated in *Thor*, corresponding increases in protein levels were not detected. To rule out protein turnover as one of the reason for this observation, we could treat the samples with a proteasome inhibitor, MG132, then reassess protein levels (Kisselev and Goldberg, 2001).

One of the more interesting targets identified by ribosome profiling is S6K as we were able to detect significantly higher levels of phospho-RPS6 in *Thor* heads. Determining if the phosphorylation of RPS6 is a direct consequence of higher levels of S6K would help validate this finding. We would need to assess if the activity of other kinases that phosphorylate RPS6 is elevated in *Thor* such as RSK.

4.4 Biological significance of upregulated immune transcripts in *d4EBP^{null}* flies

We proposed several possible explanations for the translational upregulation of immune transcripts in *Thor*. It was speculated that the shortened lifespan of *Thor* flies compared to control flies could be attributed to the premature accumulation of immune proteins. Our results obtained from ribosome profiling provides with a snapshot in time of the state of preferential ribosome association with mRNAs. A time course of immune proteins levels using flies at different ages could provide some insight into this.

4.5 Revisiting the *pgc* and *gcl* 3'UTRs in posterior RNA localization

As addressed in Chapter 3, a recent publication mapped the localization elements in the 3'UTRs of *pgc* and *gcl* and made the distinction between sequences involved in targeting the RNA to polar granules and those required for self-recruitment within the granules (Eagle et al., 2018). Using their findings, we were able to confirm our results, propose a finer mapping of the localization signals in *gcl*, and deduce that the *α-tub* 3'UTR possesses putative homotypic clustering motifs (**Fig. 3.8**).

To test our hypothesis, additional constructs would need to be generated to demonstrate the targeting abilities of 101-149, 150-200, 201-249 and 250-300 in the *gcl* 3'UTR. The multimerization of the putative targeting signals with two or three copies in tandem in the presence of a clustering motif would confirm if they are competent for targeting. We also proposed that some elements are weaker than others, and multimerization would address this and determine if there is an additive effect. To assess the targeting of these transcripts to germ granules, we would need to quantify the co-localization of the transgenes with a native germ plasm RNA such as *nos* from which we can infer that the RNAs are incorporated into the same granules (Eagle et al., 2018; Little et al., 2015; Niepielko et al., 2018).

The comparison of our deletion mutation analysis results with those of Eagle et al. led us to propose that there are two putative clustering elements at 187-192 and 203-208 in the *α-tub* 3'UTR. To test the clustering abilities of these sequences, we would need to generate two types of constructs: those containing the minimal region possessing the putative homotypic clustering motifs in *α-tub* and known targeting elements; and constructs where the *α-tub* are abrogated by deletions or mutations in the presence of targeting elements. If these *α-tub* sequences demonstrate clustering abilities, there would be implications on how these chimeric RNAs segregate within polar granules. Homotypic clusters of RNAs are spatially segregated within polar granules (Little et al., 2015; Trcek et al., 2015). We would need to assess if there are differences in the spatial distribution between clusters of the endogenous RNA and clusters of chimeric RNAs. If their spatial distribution is mediated by the homotypic clustering motif, then studying the localization of chimeric RNAs would address that.

The shortcomings of our experimental setup are due to the qualitative approach in assessing posterior localization of the transgenic RNAs and the use of comparatively low-resolution microscopy. To gain a comprehensive understanding of the complexities of these localization elements that lead to posterior localization, we would need to refine our techniques to be more quantitative by using smFISH to distinguish between targeting and clustering elements, and SIM to assess spatial distribution of RNA clusters within germ granules.

References

Abaeva, I.S., Marintchev, A., Pisareva, V.P., Hellen, C.U., and Pestova, T.V. (2011). Bypassing of stems versus linear base-by-base inspection of mammalian mRNAs during ribosomal scanning. EMBO J *30*, 115-129.

Acevedo, S.F., Peru y Colon de Portugal, R.L., Gonzalez, D.A., Rodan, A.R., and Rothenfluh, A. (2015). S6 Kinase Reflects and Regulates Ethanol-Induced Sedation. J Neurosci *35*, 15396-15402.

Adivarahan, S., Livingston, N., Nicholson, B., Rahman, S., Wu, B., Rissland, O.S., and Zenklusen, D. (2018). Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. Mol Cell *72*, 727-738 e725.

Akten, B., Tangredi, M.M., Jauch, E., Roberts, M.A., Ng, F., Raabe, T., and Jackson, F.R. (2009). Ribosomal s6 kinase cooperates with casein kinase 2 to modulate the Drosophila circadian molecular oscillator. J Neurosci *29*, 466-475.

Anjum, R., and Blenis, J. (2008). The RSK family of kinases: emerging roles in cellular signalling. Nat Rev Mol Cell Biol *9*, 747-758.

Aradska, J., Bulat, T., Sialana, F.J., Birner-Gruenberger, R., Erich, B., and Lubec, G. (2015). Gel-free mass spectrometry analysis of Drosophila melanogaster heads. Proteomics *15*, 3356-3360.

Archambault, V., Zhao, X., White-Cooper, H., Carpenter, A.T., and Glover, D.M. (2007). Mutations in Drosophila Greatwall/Scant reveal its roles in mitosis and meiosis and interdependence with Polo kinase. PLoS Genet *3*, e200.

Assa-Kunik, E., Torres, I.L., Schejter, E.D., Johnston, D.S., and Shilo, B.Z. (2007). Drosophila follicle cells are patterned by multiple levels of Notch signaling and antagonism between the Notch and JAK/STAT pathways. Development *134*, 1161-1169.

Aviv, T., Lin, Z., Ben-Ari, G., Smibert, C.A., and Sicheri, F. (2006). Sequence-specific recognition of RNA hairpins by the SAM domain of Vts1p. Nat Struct Mol Biol *13*, 168-176.

Badinloo, M., Nguyen, E., Suh, W., Alzahrani, F., Castellanos, J., Klichko, V.I., Orr, W.C., and Radyuk, S.N. (2018). Overexpression of antimicrobial peptides contributes to aging through cytotoxic effects in Drosophila tissues. Arch Insect Biochem Physiol *98*, e21464.

Baker, C.C., and Fuller, M.T. (2007). Translational control of meiotic cell cycle progression and spermatid differentiation in male germ cells by a novel eIF4G homolog. Development *134*, 2863-2869.

Barajas-Azpeleta, R., Wu, J., Gill, J., Welte, R., Seidel, C., McKinney, S., Dissel, S., and Si, K. (2018). Antimicrobial peptides modulate long-term memory. PLoS Genet *14*, e1007440.

Bashirullah, A., Halsell, S.R., Cooperstock, R.L., Kloc, M., Karaiskakis, A., Fisher, W.W., Fu, W., Hamilton, J.K., Etkin, L.D., and Lipshitz, H.D. (1999). Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in Drosophila melanogaster. EMBO J *18*, 2610-2620.

Bastock, R., and St Johnston, D. (2008). Drosophila oogenesis. Curr Biol *18*, R1082-1087.

Becalska, A.N., Kim, Y.R., Belletier, N.G., Lerit, D.A., Sinsimer, K.S., and Gavis, E.R. (2011). Aubergine is a component of a nanos mRNA localization complex. Dev Biol *349*, 46-52.

Bergsten, S.E., and Gavis, E.R. (1999). Role for mRNA localization in translational activation but not spatial restriction of nanos RNA. Development *126*, 659-669.

Bernal, A., and Kimbrell, D.A. (2000). Drosophila Thor participates in host immune defense and connects a translational regulator with innate immunity. Proc Natl Acad Sci U S A *97*, 6019-6024.

Bernstein, P., and Ross, J. (1989). Poly(A), poly(A) binding protein and the regulation of mRNA stability. Trends Biochem Sci *14*, 373-377.

Berset, C., Zurbriggen, A., Djafarzadeh, S., Altmann, M., and Trachsel, H. (2003). RNA-binding activity of translation initiation factor eIF4G1 from Saccharomyces cerevisiae. RNA *9*, 871-880.

Bertolin, A.P., Katz, M.J., Yano, M., Pozzi, B., Acevedo, J.M., Blanco-Obregon, D., Gandara, L., Sorianello, E., Kanda, H., Okano, H.*, et al.* (2016). Musashi mediates translational repression of the Drosophila hypoxia inducible factor. Nucleic Acids Res *44*, 7555-7567.

Besse, F., Lopez de Quinto, S., Marchand, V., Trucco, A., and Ephrussi, A. (2009). Drosophila PTB promotes formation of high-order RNP particles and represses oskar translation. Genes Dev *23*, 195-207.

Borman, A.M., Michel, Y.M., and Kean, K.M. (2000). Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end. Nucleic Acids Res *28*, 4068-4075.

Bozler, J., Kacsoh, B.Z., Chen, H., Theurkauf, W.E., Weng, Z., and Bosco, G. (2017). A systems level approach to temporal expression dynamics in Drosophila reveals clusters of long term memory genes. PLoS Genet *13*, e1007054.

Brennan, C.A., and Anderson, K.V. (2004). Drosophila: the genetics of innate immune recognition and response. Annu Rev Immunol *22*, 457-483.

Brown, J.B., Boley, N., Eisman, R., May, G.E., Stoiber, M.H., Duff, M.O., Booth, B.W., Wen, J., Park, S., Suzuki, A.M.*, et al.* (2014). Diversity and dynamics of the Drosophila transcriptome. Nature *512*, 393-399.

Buchon, N., Poidevin, M., Kwon, H.M., Guillou, A., Sottas, V., Lee, B.L., and Lemaitre, B. (2009). A single modular serine protease integrates signals from pattern-recognition receptors upstream of the Drosophila Toll pathway. Proc Natl Acad Sci U S A *106*, 12442-12447.

Bullock, S.L., and Ish-Horowicz, D. (2001). Conserved signals and machinery for RNA transport in Drosophila oogenesis and embryogenesis. Nature *414*, 611-616.

Campos-Ortega, J.A., and Hartenstein, V. (1985). The embryonic development of Drosophila melanogaster (Berlin ; New York: Springer-Verlag).

Cao, Y., Chtarbanova, S., Petersen, A.J., and Ganetzky, B. (2013). Dnr1 mutations cause neurodegeneration in Drosophila by activating the innate immune response in the brain. Proc Natl Acad Sci U S A *110*, E1752-1760.

Carvalho, G.B., Drago, I., Hoxha, S., Yamada, R., Mahneva, O., Bruce, K.D., Soto Obando, A., Conti, B., and Ja, W.W. (2017). The 4E-BP growth pathway regulates the effect of ambient temperature on Drosophila metabolism and lifespan. Proc Natl Acad Sci U S A *114*, 9737-9742.

Castagnetti, S., and Ephrussi, A. (2003). Orb and a long poly(A) tail are required for efficient oskar translation at the posterior pole of the Drosophila oocyte. Development *130*, 835-843.

Cavaliere, V., Taddei, C., and Gargiulo, G. (1998). Apoptosis of nurse cells at the late stages of oogenesis of Drosophila melanogaster. Dev Genes Evol *208*, 106-112.

Celniker, S.E., Dillon, L.A., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M.*, et al.* (2009). Unlocking the secrets of the genome. Nature *459*, 927-930.

Chakrabarti, S., Liehl, P., Buchon, N., and Lemaitre, B. (2012). Infection-induced host translational blockage inhibits immune responses and epithelial renewal in the Drosophila gut. Cell Host Microbe *12*, 60-70.

Chang, H., Lim, J., Ha, M., and Kim, V.N. (2014). TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. Mol Cell *53*, 1044-1052.

Chauvin, C., Koka, V., Nouschi, A., Mieulet, V., Hoareau-Aveilla, C., Dreazen, A., Cagnard, N., Carpentier, W., Kiss, T., Meyuhas, O.*, et al.* (2014). Ribosomal protein S6 kinase activity controls the ribosome biogenesis transcriptional program. Oncogene *33*, 474-483.

Chekulaeva, M., Hentze, M.W., and Ephrussi, A. (2006). Bruno acts as a dual repressor of oskar translation, promoting mRNA oligomerization and formation of silencing particles. Cell *124*, 521-533.

Chen, C., Nott, T.J., Jin, J., and Pawson, T. (2011). Deciphering arginine methylation: Tudor tells the tale. Nat Rev Mol Cell Biol *12*, 629-642.

Chen, C., Zhang, H., Broitman, S.L., Reiche, M., Farrell, I., Cooperman, B.S., and Goldman, Y.E. (2013). Dynamics of translation by single ribosomes through mRNA secondary structures. Nat Struct Mol Biol *20*, 582-588.

Chen, X., and Dickman, D. (2017). Development of a tissue-specific ribosome profiling approach in Drosophila enables genome-wide evaluation of translational adaptations. PLoS Genet *13*, e1007117.

Cho, P.F., Gamberi, C., Cho-Park, Y.A., Cho-Park, I.B., Lasko, P., and Sonenberg, N. (2006). Cap-dependent translational inhibition establishes two opposing morphogen gradients in Drosophila embryos. Curr Biol *16*, 2035-2041.

Cho, P.F., Poulin, F., Cho-Park, Y.A., Cho-Park, I.B., Chicoine, J.D., Lasko, P., and Sonenberg, N. (2005). A new paradigm for translational control: inhibition via 5'-3' mRNA tethering by Bicoid and the eIF4E cognate 4EHP. Cell *121*, 411-423.

Choi, Y.H., and Hagedorn, C.H. (2003). Purifying mRNAs with a high-affinity eIF4E mutant identifies the short 3' poly(A) end phenotype. Proc Natl Acad Sci U S A *100*, 7033-7038.

Christensen, A.K., Kahn, L.E., and Bourne, C.M. (1987). Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary. Am J Anat *178*, 1-10.

Clark, A., Meignin, C., and Davis, I. (2007). A Dynein-dependent shortcut rapidly delivers axis determination transcripts into the Drosophila oocyte. Development *134*, 1955-1965.

Clark, I.E., Wyckoff, D., and Gavis, E.R. (2000). Synthesis of the posterior determinant Nanos is spatially restricted by a novel cotranslational regulatory mechanism. Curr Biol *10*, 1311-1314.

Clemmons, A.W., Lindsay, S.A., and Wasserman, S.A. (2015). An effector Peptide family required for Drosophila toll-mediated immunity. PLoS Pathog *11*, e1004876.

Clouse, K.N., Ferguson, S.B., and Schupbach, T. (2008). Squid, Cup, and PABP55B function together to regulate gurken translation in Drosophila. Dev Biol *313*, 713-724.

Cohen, R.S., Zhang, S., and Dollar, G.L. (2005). The positional, structural, and sequence requirements of the Drosophila TLS RNA localization element. RNA *11*, 1017-1029.

Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes Dev *11*, 2755-2766.

Crucs, S., Chatterjee, S., and Gavis, E.R. (2000). Overlapping but distinct RNA elements control repression and activation of nanos translation. Mol Cell *5*, 457-467.

Dalby, B., and Glover, D.M. (1992). 3' non-translated sequences in Drosophila cyclin B transcripts direct posterior pole accumulation late in oogenesis and peri-nuclear association in syncytial embryos. Development *115*, 989-997.

Dansereau, D.A., and Lasko, P. (2008). The development of germline stem cells in Drosophila. Methods Mol Biol *450*, 3-26.

De Gregorio, E., Spellman, P.T., Tzou, P., Rubin, G.M., and Lemaitre, B. (2002). The Toll and Imd pathways are the major regulators of the immune response in Drosophila. EMBO J *21*, 2568-2579.

Dehghani, M., and Lasko, P. (2015). In vivo mapping of the functional regions of the DEAD-box helicase Vasa. Biol Open *4*, 450-462.

Dehghani, M., and Lasko, P. (2016). C-terminal residues specific to Vasa among DEAD-box helicases are required for its functions in piRNA biogenesis and embryonic patterning. Dev Genes Evol *226*, 401-412.

Demontis, F., and Perrimon, N. (2010). FOXO/4E-BP signaling in Drosophila muscles regulates organism-wide proteostasis during aging. Cell *143*, 813-825.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.

Dennis, P.B., Pullen, N., Kozma, S.C., and Thomas, G. (1996). The principal rapamycin-sensitive p70(s6k) phosphorylation sites, T-229 and T-389, are differentially regulated by rapamycin-insensitive kinase kinases. Mol Cell Biol *16*, 6242-6251.

Derry, M.C., Yanagiya, A., Martineau, Y., and Sonenberg, N. (2006). Regulation of poly(A)-binding protein through PABP-interacting proteins. Cold Spring Harb Symp Quant Biol *71*, 537-543.

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

Ding, D., Parkhurst, S.M., Halsell, S.R., and Lipshitz, H.D. (1993). Dynamic Hsp83 RNA localization during Drosophila oogenesis and embryogenesis. Mol Cell Biol *13*, 3773-3781.

Dissel, S., Seugnet, L., Thimgan, M.S., Silverman, N., Angadi, V., Thacher, P.V., Burnham, M.M., and Shaw, P.J. (2015). Differential activation of immune factors in neurons and glia contribute to individual differences in resilience/vulnerability to sleep disruption. Brain Behav Immun *47*, 75-85.

dos Santos, G., Simmonds, A.J., and Krause, H.M. (2008). A stem-loop structure in the wingless transcript defines a consensus motif for apical RNA transport. Development *135*, 133-143.

Dowling, R.J., Topisirovic, I., Fonseca, B.D., and Sonenberg, N. (2010). Dissecting the role of mTOR: lessons from mTOR inhibitors. Biochim Biophys Acta *1804*, 433-439.

Duncan, R., Milburn, S.C., and Hershey, J.W. (1987). Regulated phosphorylation and low abundance of HeLa cell initiation factor eIF-4F suggest a role in translational control. Heat shock effects on eIF-4F. J Biol Chem *262*, 380-388.

Dunckley, T., and Parker, R. (1999). The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. EMBO J *18*, 5411-5422.

Dunn, J.G., Foo, C.K., Belletier, N.G., Gavis, E.R., and Weissman, J.S. (2013). Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. Elife *2*, e01179.

Dushay, M.S., Roethele, J.B., Chaverri, J.M., Dulek, D.E., Syed, S.K., Kitami, T., and Eldon, E.D. (2000). Two attacin antibacterial genes of Drosophila melanogaster. Gene *246*, 49-57.

Eagle, W.V.I., Yeboah-Kordieh, D.K., Niepielko, M.G., and Gavis, E.R. (2018). Distinct cis-acting elements mediate targeting and clustering of Drosophila polar granule mRNAs. Development *145*.

Edery, I., and Sonenberg, N. (1985). Cap-dependent RNA splicing in a HeLa nuclear extract. Proc Natl Acad Sci U S A *82*, 7590-7594.

Ephrussi, A., Dickinson, L.K., and Lehmann, R. (1991). Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. Cell *66*, 37-50.

Ephrussi, A., and Lehmann, R. (1992). Induction of germ cell formation by oskar. Nature *358*, 387-392.

Fabian, M.R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. Annu Rev Biochem *79*, 351-379.

Fan, X.C., Myer, V.E., and Steitz, J.A. (1997). AU-rich elements target small nuclear RNAs as well as mRNAs for rapid degradation. Genes Dev *11*, 2557-2568.

Feoktistova, K., Tuvshintogs, E., Do, A., and Fraser, C.S. (2013). Human eIF4E promotes mRNA restructuring by stimulating eIF4A helicase activity. Proc Natl Acad Sci U S A *110*, 13339-13344.

Filardo, P., and Ephrussi, A. (2003). Bruno regulates gurken during Drosophila oogenesis. Mech Dev *120*, 289-297.

Flaherty, S.M., Fortes, P., Izaurralde, E., Mattaj, I.W., and Gilmartin, G.M. (1997). Participation of the nuclear cap binding complex in pre-mRNA 3' processing. Proc Natl Acad Sci U S A *94*, 11893-11898.

Flora, P., Wong-Deyrup, S.W., Martin, E.T., Palumbo, R.J., Nasrallah, M., Oligney, A., Blatt, P., Patel, D., Fuchs, G., and Rangan, P. (2018). Sequential Regulation of Maternal mRNAs through a Conserved cis-Acting Element in Their 3' UTRs. Cell Rep *25*, 3828-3843 e3829.

Forrest, K.M., Clark, I.E., Jain, R.A., and Gavis, E.R. (2004). Temporal complexity within a translational control element in the nanos mRNA. Development *131*, 5849-5857.

Forrest, K.M., and Gavis, E.R. (2003). Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila. Curr Biol *13*, 1159-1168.

Franklin-Dumont, T.M., Chatterjee, C., Wasserman, S.A., and Dinardo, S. (2007). A novel eIF4G homolog, Off-schedule, couples translational control to meiosis and differentiation in Drosophila spermatocytes. Development *134*, 2851-2861.

Gao, M., and Arkov, A.L. (2013). Next generation organelles: structure and role of germ granules in the germline. Mol Reprod Dev *80*, 610-623.

Gao, X., Zhang, Y., Arrazola, P., Hino, O., Kobayashi, T., Yeung, R.S., Ru, B., and Pan, D. (2002). Tsc tumour suppressor proteins antagonize amino-acid-TOR signalling. Nat Cell Biol *4*, 699-704.

Gavis, E.R., Chatterjee, S., Ford, N.R., and Wolff, L.J. (2008). Dispensability of nanos mRNA localization for abdominal patterning but not for germ cell development. Mech Dev *125*, 81-90.

Gavis, E.R., Curtis, D., and Lehmann, R. (1996). Identification of cis-acting sequences that control nanos RNA localization. Dev Biol *176*, 36-50.

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

Gavis, E.R., and Lehmann, R. (1994). Translational regulation of nanos by RNA localization. Nature *369*, 315-318.

Ghosh, S., and Lasko, P. (2015). Loss-of-function analysis reveals distinct requirements of the translation initiation factors eIF4E, eIF4E-3, eIF4G and eIF4G2 in Drosophila spermatogenesis. PLoS One *10*, e0122519.

Ghosh, S., Marchand, V., Gaspar, I., and Ephrussi, A. (2012). Control of RNP motility and localization by a splicing-dependent structure in oskar mRNA. Nat Struct Mol Biol *19*, 441-449.

Gingras, A.C., Raught, B., and Sonenberg, N. (1999). eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. Annu Rev Biochem *68*, 913-963.

Gingras, A.C., Raught, B., and Sonenberg, N. (2001). Regulation of translation initiation by FRAP/mTOR. Genes Dev *15*, 807-826.

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

Glotzer, J.B., Saffrich, R., Glotzer, M., and Ephrussi, A. (1997). Cytoplasmic flows localize injected oskar RNA in Drosophila oocytes. Curr Biol *7*, 326-337.

Goodfellow, I.G., and Roberts, L.O. (2008). Eukaryotic initiation factor 4E. Int J Biochem Cell Biol *40*, 2675-2680.

Gramates, L.S., Marygold, S.J., Santos, G.D., Urbano, J.M., Antonazzo, G., Matthews, B.B., Rey, A.J., Tabone, C.J., Crosby, M.A., Emmert, D.B.*, et al.* (2017). FlyBase at 25: looking to the future. Nucleic Acids Res *45*, D663-D671.

Grillo, G., Turi, A., Licciulli, F., Mignone, F., Liuni, S., Banfi, S., Gennarino, V.A., Horner, D.S., Pavesi, G., Picardi, E.*, et al.* (2010). UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Res *38*, D75-80.

Hahn, K., Miranda, M., Francis, V.A., Vendrell, J., Zorzano, A., and Teleman, A.A. (2010). PP2A regulatory subunit PP2A-B' counteracts S6K phosphorylation. Cell Metab *11*, 438-444.

Hahn, N., Geurten, B., Gurvich, A., Piepenbrock, D., Kastner, A., Zanini, D., Xing, G., Xie, W., Gopfert, M.C., Ehrenreich, H.*, et al.* (2013). Monogenic heritable autism gene neuroligin impacts Drosophila social behaviour. Behav Brain Res *252*, 450-457.

Hamilton, R.S., and Davis, I. (2007). RNA localization signals: deciphering the message with bioinformatics. Semin Cell Dev Biol *18*, 178-185.

Hamm, J., and Mattaj, I.W. (1990). Monomethylated cap structures facilitate RNA export from the nucleus. Cell *63*, 109-118.

Hanyu-Nakamura, K., Matsuda, K., Cohen, S.M., and Nakamura, A. (2019). Pgc suppresses the zygotically acting RNA decay pathway to protect germ plasm RNAs in the Drosophila embryo. Development *146*.

Hanyu-Nakamura, K., Sonobe-Nojima, H., Tanigawa, A., Lasko, P., and Nakamura, A. (2008). Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. Nature *451*, 730-733.

Hara, K., Maruki, Y., Long, X., Yoshino, K., Oshiro, N., Hidayat, S., Tokunaga, C., Avruch, J., and Yonezawa, K. (2002). Raptor, a binding partner of target of rapamycin (TOR), mediates TOR action. Cell *110*, 177-189.

Harris, N., Braiser, D.J., Dickman, D.K., Fetter, R.D., Tong, A., and Davis, G.W. (2015). The Innate Immune Receptor PGRP-LC Controls Presynaptic Homeostatic Plasticity. Neuron *88*, 1157-1164.

Hay, B., Jan, L.Y., and Jan, Y.N. (1988). A protein component of Drosophila polar granules is encoded by vasa and has extensive sequence similarity to ATP-dependent helicases. Cell *55*, 577-587.

Hay, N., and Sonenberg, N. (2004). Upstream and downstream of mTOR. Genes Dev *18*, 1926-1945.

Hellen, C.U., and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. Genes Dev *15*, 1593-1612.

Hellen, C.U.T. (2018). Translation Termination and Ribosome Recycling in Eukaryotes. Cold Spring Harb Perspect Biol *10*.

Hernandez, G., Altmann, M., Sierra, J.M., Urlaub, H., Diez del Corral, R., Schwartz, P., and Rivera-Pomar, R. (2005). Functional analysis of seven genes encoding eight translation initiation factor 4E (eIF4E) isoforms in Drosophila. Mech Dev *122*, 529-543.

Hernandez, G., Han, H., Gandin, V., Fabian, L., Ferreira, T., Zuberek, J., Sonenberg, N., Brill, J.A., and Lasko, P. (2012). Eukaryotic initiation factor 4E-3 is essential for meiotic chromosome segregation, cytokinesis and male fertility in Drosophila. Development *139*, 3211-3220.

Hernandez, G., Miron, M., Han, H., Liu, N., Magescas, J., Tettweiler, G., Frank, F., Siddiqui, N., Sonenberg, N., and Lasko, P. (2013). Mextli is a novel eukaryotic translation initiation factor 4E-binding protein that promotes translation in Drosophila melanogaster. Mol Cell Biol *33*, 2854-2864.

Hershey, J.W.B., Sonenberg, N., and Mathews, M.B. (2018). Principles of Translational Control. Cold Spring Harb Perspect Biol.

Heuer, A., Gerovac, M., Schmidt, C., Trowitzsch, S., Preis, A., Kotter, P., Berninghausen, O., Becker, T., Beckmann, R., and Tampe, R. (2017). Structure of the 40S-ABCE1 post-splitting complex in ribosome recycling and translation initiation. Nat Struct Mol Biol *24*, 453-460.

Hilliker, A., Gao, Z., Jankowsky, E., and Parker, R. (2011). The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex. Mol Cell *43*, 962-972.

Hinnebusch, A.G. (2011). Molecular mechanism of scanning and start codon selection in eukaryotes. Microbiol Mol Biol Rev *75*, 434-467, first page of table of contents.

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

Hinnebusch, A.G., and Lorsch, J.R. (2012). The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harb Perspect Biol *4*.

Hsieh, A.C., Liu, Y., Edlind, M.P., Ingolia, N.T., Janes, M.R., Sher, A., Shi, E.Y., Stumpf, C.R., Christensen, C., Bonham, M.J.*, et al.* (2012). The translational landscape of mTOR signalling steers cancer initiation and metastasis. Nature *485*, 55-61.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C.*, et al.* (2007). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res *35*, W169-175.

Huang, H.Y., Chien, C.H., Jen, K.H., and Huang, H.D. (2006). RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. Nucleic Acids Res *34*, W429-434.

Hutchinson, J.A., Shanware, N.P., Chang, H., and Tibbetts, R.S. (2011). Regulation of ribosomal protein S6 phosphorylation by casein kinase 1 and protein phosphatase 1. J Biol Chem *286*, 8688-8696.

Iatsenko, I., Kondo, S., Mengin-Lecreulx, D., and Lemaitre, B. (2016). PGRP-SD, an Extracellular Pattern-Recognition Receptor, Enhances Peptidoglycan-Mediated Activation of the Drosophila Imd Pathway. Immunity *45*, 1013-1023.

Igreja, C., and Izaurralde, E. (2011). CUP promotes deadenylation and inhibits decapping of mRNA targets. Genes Dev *25*, 1955-1967.

Imataka, H., Gradi, A., and Sonenberg, N. (1998). A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. EMBO J *17*, 7480-7489.

Imataka, H., and Sonenberg, N. (1997). Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. Mol Cell Biol *17*, 6940-6947.

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

Izaurralde, E., Stepinski, J., Darzynkiewicz, E., and Mattaj, I.W. (1992). A cap binding protein that may mediate nuclear export of RNA polymerase II-transcribed RNAs. J Cell Biol *118*, 1287-1295.

Jain, R.A., and Gavis, E.R. (2008). The Drosophila hnRNP M homolog Rumpelstiltskin regulates nanos mRNA localization. Development *135*, 973-982.

Jambor, H., Brunel, C., and Ephrussi, A. (2011). Dimerization of oskar 3' UTRs promotes hitchhiking for RNA localization in the Drosophila oocyte. RNA *17*, 2049-2057.

Jambor, H., Mueller, S., Bullock, S.L., and Ephrussi, A. (2014). A stem-loop structure directs oskar mRNA to microtubule minus ends. RNA *20*, 429-439.

Jambor, H., Surendranath, V., Kalinka, A.T., Mejstrik, P., Saalfeld, S., and Tomancak, P. (2015). Systematic imaging reveals features and changing localization of mRNAs in Drosophila development. Elife *4*.

Jefferies, H.B., Reinhard, C., Kozma, S.C., and Thomas, G. (1994). Rapamycin selectively represses translation of the "polypyrimidine tract" mRNA family. Proc Natl Acad Sci U S A *91*, 4441-4445.

Jensen, L., Farook, M.F., and Reiter, L.T. (2013). Proteomic profiling in Drosophila reveals potential Dube3a regulation of the actin cytoskeleton and neuronal homeostasis. PLoS One *8*, e61952.

Johnstone, O., Deuring, R., Bock, R., Linder, P., Fuller, M.T., and Lasko, P. (2005). Belle is a Drosophila DEAD-box protein required for viability and in the germ line. Dev Biol *277*, 92-101.

Johnstone, O., and Lasko, P. (2004). Interaction with eIF5B is essential for Vasa function during development. Development *131*, 4167-4178.

Jongens, T.A., Ackerman, L.D., Swedlow, J.R., Jan, L.Y., and Jan, Y.N. (1994). Germ cell-less encodes a cell type-specific nuclear pore-associated protein and functions early in the germ-cell specification pathway of Drosophila. Genes Dev *8*, 2123-2136.

Jongens, T.A., Hay, B., Jan, L.Y., and Jan, Y.N. (1992). The germ cell-less gene product: a posteriorly localized component necessary for germ cell development in Drosophila. Cell *70*, 569-584.

Kahvejian, A., Svitkin, Y.V., Sukarieh, R., M'Boutchou, M.N., and Sonenberg, N. (2005). Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. Genes Dev *19*, 104-113.

Khong, A., and Parker, R. (2018). mRNP architecture in translating and stress conditions reveals an ordered pathway of mRNP compaction. J Cell Biol *217*, 4124-4140.

Kim, D.H., Sarbassov, D.D., Ali, S.M., King, J.E., Latek, R.R., Erdjument-Bromage, H., Tempst, P., and Sabatini, D.M. (2002). mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery. Cell *110*, 163-175.

Kim, D.H., Sarbassov, D.D., Ali, S.M., Latek, R.R., Guntur, K.V., Erdjument-Bromage, H., Tempst, P., and Sabatini, D.M. (2003). GbetaL, a positive regulator of the rapamycin-sensitive pathway required for the nutrient-sensitive interaction between raptor and mTOR. Mol Cell *11*, 895-904.

Kim, J., and Kim-Ha, J. (2006). Ovarian tumors in Rbp9 mutants of Drosophila induce an immune response. Mol Cells *22*, 228-232.

Kim, M., Lee, J.H., Koh, H., Lee, S.Y., Jang, C., Chung, C.J., Sung, J.H., Blenis, J., and Chung, J. (2006). Inhibition of ERK-MAP kinase signaling by RSK during Drosophila development. EMBO J *25*, 3056-3067.

Kim, W., Jang, Y.G., Yang, J., and Chung, J. (2017). Spatial Activation of TORC1 Is Regulated by Hedgehog and E2F1 Signaling in the Drosophila Eye. Dev Cell *42*, 363-375 e364.

Kinkelin, K., Veith, K., Grunwald, M., and Bono, F. (2012). Crystal structure of a minimal eIF4E-Cup complex reveals a general mechanism of eIF4E regulation in translational repression. RNA *18*, 1624-1634.

Kisselev, A.F., and Goldberg, A.L. (2001). Proteasome inhibitors: from research tools to drug candidates. Chem Biol *8*, 739-758.

Knight, Z.A., Tan, K., Birsoy, K., Schmidt, S., Garrison, J.L., Wysocki, R.W., Emiliano, A., Ekstrand, M.I., and Friedman, J.M. (2012). Molecular profiling of activated neurons by phosphorylated ribosome capture. Cell *151*, 1126-1137.

Knowles-Barley, S., Longair, M., and Armstrong, J.D. (2010). BrainTrap: a database of 3D protein expression patterns in the Drosophila brain. Database (Oxford) *2010*, baq005.

Koromilas, A.E., Lazaris-Karatzas, A., and Sonenberg, N. (1992). mRNAs containing extensive secondary structure in their 5' non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. EMBO J *11*, 4153-4158.

Kounatidis, I., Chtarbanova, S., Cao, Y., Hayne, M., Jayanth, D., Ganetzky, B., and Ligoxygakis, P. (2017). NF-kappaB Immunity in the Brain Determines Fly Lifespan in Healthy Aging and Age-Related Neurodegeneration. Cell Rep *19*, 836-848.

Kowanda, M., Bergalet, J., Wieczorek, M., Brouhard, G., Lecuyer, E., and Lasko, P. (2016). Loss of function of the Drosophila Ninein-related centrosomal protein Bsg25D causes mitotic defects and impairs embryonic development. Biol Open *5*, 1040-1051.

Lachance, P.E., Miron, M., Raught, B., Sonenberg, N., and Lasko, P. (2002). Phosphorylation of eukaryotic translation initiation factor 4E is critical for growth. Mol Cell Biol *22*, 1656-1663.

Lai, M.C., Lee, Y.H., and Tarn, W.Y. (2008). The DEAD-box RNA helicase DDX3 associates with export messenger ribonucleoproteins as well as tip-associated protein and participates in translational control. Mol Biol Cell *19*, 3847-3858.

Lantz, V., and Schedl, P. (1994). Multiple cis-acting targeting sequences are required for orb mRNA localization during Drosophila oogenesis. Mol Cell Biol *14*, 2235-2242.

Laplante, M., and Sabatini, D.M. (2012). mTOR signaling in growth control and disease. Cell *149*, 274-293.

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

Lasko, P. (2009). Translational control during early development. Prog Mol Biol Transl Sci *90*, 211-254.

Leatherman, J.L., Levin, L., Boero, J., and Jongens, T.A. (2002). germ cell-less acts to repress transcription during the establishment of the Drosophila germ cell lineage. Curr Biol *12*, 1681-1685.

Lecuyer, E., Necakov, A.S., Caceres, L., and Krause, H.M. (2008). High-resolution fluorescent in situ hybridization of Drosophila embryos and tissues. CSH Protoc *2008*, pdb prot5019.

Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell *131*, 174-187.

Leppek, K., Das, R., and Barna, M. (2018). Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat Rev Mol Cell Biol *19*, 158-174.

Lerit, D.A., and Gavis, E.R. (2011). Transport of germ plasm on astral microtubules directs germ cell development in Drosophila. Curr Biol *21*, 439-448.

Lerit, D.A., Shebelut, C.W., Lawlor, K.J., Rusan, N.M., Gavis, E.R., Schedl, P., and Deshpande, G. (2017). Germ Cell-less Promotes Centrosome Segregation to Induce Germ Cell Formation. Cell Rep *18*, 831-839.

Levitin, A., Marcil, A., Tettweiler, G., Laforest, M.J., Oberholzer, U., Alarco, A.M., Thomas, D.Y., Lasko, P., and Whiteway, M. (2007). Drosophila melanogaster Thor and response to Candida albicans infection. Eukaryot Cell *6*, 658-663.

Linder, P., Lasko, P.F., Ashburner, M., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J., and Slonimski, P.P. (1989). Birth of the D-E-A-D box. Nature *337*, 121-122.

Little, S.C., Sinsimer, K.S., Lee, J.J., Wieschaus, E.F., and Gavis, E.R. (2015). Independent and coordinate trafficking of single Drosophila germ plasm mRNAs. Nat Cell Biol *17*, 558-568.

Liu, N., and Lasko, P. (2015). Analysis of RNA Interference Lines Identifies New Functions of Maternally-Expressed Genes Involved in Embryonic Patterning in Drosophila melanogaster. G3 (Bethesda) *5*, 1025-1034.

Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol Biol *6*, 26.

Lu, W., Winding, M., Lakonishok, M., Wildonger, J., and Gelfand, V.I. (2016). Microtubule-microtubule sliding by kinesin-1 is essential for normal cytoplasmic streaming in Drosophila oocytes. Proc Natl Acad Sci U S A *113*, E4995-5004.

Mader, S., Lee, H., Pause, A., and Sonenberg, N. (1995). The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. Mol Cell Biol *15*, 4990-4997.

Mahowald, A.P. (2001). Assembly of the Drosophila germ plasm. Int Rev Cytol *203*, 187-213.

Marcotrigiano, J., Gingras, A.C., Sonenberg, N., and Burley, S.K. (1999). Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. Mol Cell *3*, 707-716.

Martin, K.C., and Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. Cell *136*, 719-730.

Masvidal, L., Hulea, L., Furic, L., Topisirovic, I., and Larsson, O. (2017). mTOR-sensitive translation: Cleared fog reveals more trees. RNA Biol *14*, 1299-1305.

McLaughlin, J.M., and Bratu, D.P. (2015). Drosophila melanogaster Oogenesis: An Overview. Methods Mol Biol *1328*, 1-20.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., and Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res *41*, W597-600.

Meijer, H.A., Bushell, M., Hill, K., Gant, T.W., Willis, A.E., Jones, P., and de Moor, C.H. (2007). A novel method for poly(A) fractionation reveals a large population of mRNAs with a short poly(A) tail in mammalian cells. Nucleic Acids Res *35*, e132.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.B., and Jaffrey, S.R. (2015). 5' UTR m(6)A Promotes Cap-Independent Translation. Cell *163*, 999-1010.

Meyuhas, O. (2000). Synthesis of the translational apparatus is regulated at the translational level. Eur J Biochem *267*, 6321-6330.

Micklem, D.R., Adams, J., Grunert, S., and St Johnston, D. (2000). Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation. EMBO J *19*, 1366-1377.

Millevoi, S., and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. Nucleic Acids Res *38*, 2757-2774.

Miron, M., Lasko, P., and Sonenberg, N. (2003). Signaling from Akt to FRAP/TOR targets both 4E-BP and S6K in Drosophila melanogaster. Mol Cell Biol *23*, 9117-9126.

Miron, M., Verdu, J., Lachance, P.E., Birnbaum, M.J., Lasko, P.F., and Sonenberg, N. (2001). The translational inhibitor 4E-BP is an effector of PI(3)K/Akt signalling and cell growth in Drosophila. Nat Cell Biol *3*, 596-601.

Moore, J., Han, H., and Lasko, P. (2009). Bruno negatively regulates germ cell-less expression in a BRE-independent manner. Mech Dev *126*, 503-516.

Murthy, M., and Turner, G. (2013). Dissection of the head cuticle and sheath of living flies for whole-cell patch-clamp recordings in the brain. Cold Spring Harb Protoc *2013*, 134-139.

Nakamura, A., Amikura, R., Mukai, M., Kobayashi, S., and Lasko, P.F. (1996). Requirement for a noncoding RNA in Drosophila polar granules for germ cell establishment. Science *274*, 2075-2079.

Nakamura, A., Sato, K., and Hanyu-Nakamura, K. (2004). Drosophila cup is an eIF4E binding protein that associates with Bruno and regulates oskar mRNA translation in oogenesis. Dev Cell *6*, 69-78.

Natalizio, B.J., and Wente, S.R. (2013). Postage for the messenger: designating routes for nuclear mRNA export. Trends Cell Biol *23*, 365-373.

Nelson, M.R., Leidal, A.M., and Smibert, C.A. (2004). Drosophila Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression. EMBO J *23*, 150-159.

Niepielko, M.G., Eagle, W.V.I., and Gavis, E.R. (2018). Stochastic Seeding Coupled with mRNA Self-Recruitment Generates Heterogeneous Drosophila Germ Granules. Curr Biol *28*, 1872-1881 e1873.

Okabe, M., Imai, T., Kurusu, M., Hiromi, Y., and Okano, H. (2001). Translational repression determines a neuronal potential in Drosophila asymmetric cell division. Nature *411*, 94-98.

Pae, J., Cinalli, R.M., Marzio, A., Pagano, M., and Lehmann, R. (2017). GCL and CUL3 Control the Switch between Cell Lineages by Mediating Localized Degradation of an RTK. Dev Cell *42*, 130-142 e137.

Parton, R.M., Hamilton, R.S., Ball, G., Yang, L., Cullen, C.F., Lu, W., Ohkura, H., and Davis, I. (2011). A PAR-1-dependent orientation gradient of dynamic microtubules directs posterior cargo transport in the Drosophila oocyte. J Cell Biol *194*, 121-135.

Pause, A., Belsham, G.J., Gingras, A.C., Donze, O., Lin, T.A., Lawrence, J.C., Jr., and Sonenberg, N. (1994). Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function. Nature *371*, 762-767.

Pek, J.W., and Kai, T. (2011). A role for vasa in regulating mitotic chromosome condensation in Drosophila. Curr Biol *21*, 39-44.

Pek, J.W., Patil, V.S., and Kai, T. (2012). piRNA pathway and the potential processing site, the nuage, in the Drosophila germline. Dev Growth Differ *54*, 66-77.

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

Pestova, T.V., Shatsky, I.N., and Hellen, C.U. (1996). Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. Mol Cell Biol *16*, 6870-6878.

Peter, D., Weber, R., Kone, C., Chung, M.Y., Ebertsch, L., Truffault, V., Weichenrieder, O., Igreja, C., and Izaurralde, E. (2015). Mextli proteins use both canonical bipartite and novel tripartite binding modes to form eIF4E complexes that display differential sensitivity to 4E-BP regulation. Genes Dev *29*, 1835-1849.

Peterson, T.R., Laplante, M., Thoreen, C.C., Sancak, Y., Kang, S.A., Kuehl, W.M., Gray, N.S., and Sabatini, D.M. (2009). DEPTOR is an mTOR inhibitor frequently overexpressed in multiple myeloma cells and required for their survival. Cell *137*, 873-886.

Pisareva, V.P., Pisarev, A.V., Komar, A.A., Hellen, C.U., and Pestova, T.V. (2008). Translation initiation on mammalian mRNAs with structured 5'UTRs requires DExH-box protein DHX29. Cell *135*, 1237-1250.

Ponting, C.P. (1997). Tudor domains in proteins that interact with RNA. Trends Biochem Sci *22*, 51-52.

Potter, C.J., Pedraza, L.G., and Xu, T. (2002). Akt regulates growth by directly phosphorylating Tsc2. Nat Cell Biol *4*, 658-665.

Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev *25*, 1770-1782.

Purice, M.D., Speese, S.D., and Logan, M.A. (2016). Delayed glial clearance of degenerating axons in aged Drosophila is due to reduced PI3K/Draper activity. Nat Commun *7*, 12871.

Qin, X., Ahn, S., Speed, T.P., and Rubin, G.M. (2007). Global analyses of mRNA translational control during early Drosophila embryogenesis. Genome Biol *8*, R63.

Rangan, P., DeGennaro, M., Jaime-Bustamante, K., Coux, R.X., Martinho, R.G., and Lehmann, R. (2009). Temporal and spatial control of germ-plasm RNAs. Curr Biol *19*, 72-77.

Raught, B., and Gingras, A.C. (1999). eIF4E activity is regulated at multiple levels. Int J Biochem Cell Biol *31*, 43-57.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet *16*, 276-277.

Rodriguez, A., Zhou, Z., Tang, M.L., Meller, S., Chen, J., Bellen, H., and Kimbrell, D.A. (1996). Identification of immune system and response genes, and novel mutations causing melanotic tumor formation in Drosophila melanogaster. Genetics *143*, 929-940.

Rogers, G.W., Jr., Richter, N.J., and Merrick, W.C. (1999). Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. J Biol Chem *274*, 12236-12244.

Romero-Pozuelo, J., Demetriades, C., Schroeder, P., and Teleman, A.A. (2017). CycD/Cdk4 and Discontinuities in Dpp Signaling Activate TORC1 in the Drosophila Wing Disc. Dev Cell *42*, 376-387 e375.

Roth, S. (2001). Drosophila oogenesis: coordinating germ line and soma. Curr Biol *11*, R779-781.

Roth, S., and Lynch, J.A. (2009). Symmetry breaking during Drosophila oogenesis. Cold Spring Harb Perspect Biol *1*, a001891.

Roux, P.P., Shahbazian, D., Vu, H., Holz, M.K., Cohen, M.S., Taunton, J., Sonenberg, N., and Blenis, J. (2007). RAS/ERK signaling promotes site-specific ribosomal protein S6 phosphorylation via RSK and stimulates cap-dependent translation. J Biol Chem *282*, 14056-14064.

Rubin, G.M., and Spradling, A.C. (1982). Genetic transformation of Drosophila with transposable element vectors. Science *218*, 348-353.

Rutledge, C.E., Lau, H.T., Mangan, H., Hardy, L.L., Sunnotel, O., Guo, F., MacNicol, A.M., Walsh, C.P., and Lees-Murdock, D.J. (2014). Efficient translation of Dnmt1 requires cytoplasmic polyadenylation and Musashi binding elements. PLoS One *9*, e88385.

Ruvinsky, I., Sharon, N., Lerer, T., Cohen, H., Stolovich-Rain, M., Nir, T., Dor, Y., Zisman, P., and Meyuhas, O. (2005). Ribosomal protein S6 phosphorylation is a determinant of cell size and glucose homeostasis. Genes Dev *19*, 2199-2211.

Sancak, Y., Thoreen, C.C., Peterson, T.R., Lindquist, R.A., Kang, S.A., Spooner, E., Carr, S.A., and Sabatini, D.M. (2007). PRAS40 is an insulin-regulated inhibitor of the mTORC1 protein kinase. Mol Cell *25*, 903-915.

Saunders, C., and Cohen, R.S. (1999). The role of oocyte transcription, the 5'UTR, and translation repression and derepression in Drosophila gurken mRNA and protein localization. Mol Cell *3*, 43-54.

Schwer, B., and Shuman, S. (1996). Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing in vivo. RNA *2*, 574-583.

Semotok, J.L., Cooperstock, R.L., Pinder, B.D., Vari, H.K., Lipshitz, H.D., and Smibert, C.A. (2005). Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early Drosophila embryo. Curr Biol *15*, 284-294.

Semotok, J.L., Luo, H., Cooperstock, R.L., Karaiskakis, A., Vari, H.K., Smibert, C.A., and Lipshitz, H.D. (2008). Drosophila maternal Hsp83 mRNA destabilization is directed by multiple SMAUG recognition elements in the open reading frame. Mol Cell Biol *28*, 6757-6772.

Serano, T.L., and Cohen, R.S. (1995). A small predicted stem-loop structure mediates oocyte localization of Drosophila K10 mRNA. Development *121*, 3809-3818.

Shahbazian, D., Roux, P.P., Mieulet, V., Cohen, M.S., Raught, B., Taunton, J., Hershey, J.W., Blenis, J., Pende, M., and Sonenberg, N. (2006). The mTOR/PI3K and MAPK pathways converge on eIF4B to control its phosphorylation and activity. EMBO J *25*, 2781-2791.

Smith, J.L., Wilson, J.E., and Macdonald, P.M. (1992). Overexpression of oskar directs ectopic activation of nanos and presumptive pole cell formation in Drosophila embryos. Cell *70*, 849-859.

Smith, R.W.P., Anderson, R.C., Larralde, O., Smith, J.W.S., Gorgoni, B., Richardson, W.A., Malik, P., Graham, S.V., and Gray, N.K. (2017). Viral and cellular mRNA-specific activators harness PABP and eIF4G to promote translation initiation downstream of cap binding. Proc Natl Acad Sci U S A *114*, 6310-6315.

Sonenberg, N. (2008). eIF4E, the mRNA cap-binding protein: from basic discovery to translational research. Biochem Cell Biol *86*, 178-183.

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

Sonenberg, N., Morgan, M.A., Merrick, W.C., and Shatkin, A.J. (1978). A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. Proc Natl Acad Sci U S A *75*, 4843-4847.

Sonenberg, N., Rupprecht, K.M., Hecht, S.M., and Shatkin, A.J. (1979). Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. Proc Natl Acad Sci U S A *76*, 4345-4349.

Sonenberg, N., Trachsel, H., Hecht, S., and Shatkin, A.J. (1980). Differential stimulation of capped mRNA translation in vitro by cap binding protein. Nature *285*, 331-333.

St Johnston, D. (2005). Moving messages: the intracellular localization of mRNAs. Nat Rev Mol Cell Biol *6*, 363-375.

Styhler, S., Nakamura, A., Swan, A., Suter, B., and Lasko, P. (1998). vasa is required for GURKEN accumulation in the oocyte, and is involved in oocyte differentiation and germline cyst development. Development *125*, 1569-1578.

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature *508*, 66-71.

Sun, Y., Atas, E., Lindqvist, L., Sonenberg, N., Pelletier, J., and Meller, A. (2012). The eukaryotic initiation factor eIF4H facilitates loop-binding, repetitive RNA unwinding by the eIF4A DEAD-box helicase. Nucleic Acids Res *40*, 6199-6207.

Svitkin, Y.V., Pause, A., Haghighat, A., Pyronnet, S., Witherell, G., Belsham, G.J., and Sonenberg, N. (2001). The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. RNA *7*, 382-394.

Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007). SMAUG is a major regulator of maternal mRNA destabilization in Drosophila and its translation is activated by the PAN GU kinase. Dev Cell *12*, 143-155.

Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. Development *136*, 3033-3042.

Tahara, S.M., Morgan, M.A., and Shatkin, A.J. (1981). Two forms of purified m7G-cap binding protein with different effects on capped mRNA translation in extracts of uninfected and poliovirus-infected HeLa cells. J Biol Chem *256*, 7691-7694.

Tahmasebi, S., Jafarnejad, S.M., Tam, I.S., Gonatopoulos-Pournatzis, T., Matta-Camacho, E., Tsukumo, Y., Yanagiya, A., Li, W., Atlasi, Y., Caron, M.*, et al.* (2016). Control of embryonic stem cell self-renewal and differentiation via coordinated alternative splicing and translation of YY2. Proc Natl Acad Sci U S A *113*, 12360-12367.

Tamura, K., Subramanian, S., and Kumar, S. (2004). Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol Biol Evol *21*, 36-44.

Tanaka, T., Kato, Y., Matsuda, K., Hanyu-Nakamura, K., and Nakamura, A. (2011). Drosophila Mon2 couples Oskar-induced endocytosis with actin remodeling for cortical anchorage of the germ plasm. Development *138*, 2523-2532.

Tarun, S.Z., Jr., and Sachs, A.B. (1996). Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. EMBO J *15*, 7168-7177.

Tettweiler, G., Kowanda, M., Lasko, P., Sonenberg, N., and Hernandez, G. (2012). The Distribution of eIF4E-Family Members across Insecta. Comp Funct Genomics *2012*, 960420.

Tettweiler, G., Miron, M., Jenkins, M., Sonenberg, N., and Lasko, P.F. (2005). Starvation and oxidative stress resistance in Drosophila are mediated through the eIF4E-binding protein, d4E-BP. Genes Dev *19*, 1840-1843.

Thio, G.L., Ray, R.P., Barcelo, G., and Schupbach, T. (2000). Localization of gurken RNA in Drosophila oogenesis requires elements in the 5' and 3' regions of the transcript. Dev Biol *221*, 435-446.

Thomson, T., and Lasko, P. (2004). Drosophila tudor is essential for polar granule assembly and pole cell specification, but not for posterior patterning. Genesis *40*, 164-170.

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., and Sabatini, D.M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. Nature *485*, 109-113.

Tomancak, P., Berman, B.P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S.E., and Rubin, G.M. (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome Biol *8*, R145.

Tomancak, P., Guichet, A., Zavorszky, P., and Ephrussi, A. (1998). Oocyte polarity depends on regulation of gurken by Vasa. Development *125*, 1723-1732.

Topisirovic, I., Svitkin, Y.V., Sonenberg, N., and Shatkin, A.J. (2011). Cap and cap-binding proteins in the control of gene expression. Wiley Interdiscip Rev RNA *2*, 277-298.

Trcek, T., Grosch, M., York, A., Shroff, H., Lionnet, T., and Lehmann, R. (2015). Drosophila germ granules are structured and contain homotypic mRNA clusters. Nat Commun *6*, 7962.

Ueda, T., Watanabe-Fukunaga, R., Fukuyama, H., Nagata, S., and Fukunaga, R. (2004). Mnk2 and Mnk1 are essential for constitutive and inducible phosphorylation of eukaryotic initiation factor 4E but not for cell growth or development. Mol Cell Biol *24*, 6539-6549.

Van De Bor, V., Hartswood, E., Jones, C., Finnegan, D., and Davis, I. (2005). gurken and the I factor retrotransposon RNAs share common localization signals and machinery. Dev Cell *9*, 51-62.

Van Der Kelen, K., Beyaert, R., Inze, D., and De Veylder, L. (2009). Translational control of eukaryotic gene expression. Crit Rev Biochem Mol Biol *44*, 143-168.

Vanzo, N.F., and Ephrussi, A. (2002). Oskar anchoring restricts pole plasm formation to the posterior of the Drosophila oocyte. Development *129*, 3705-3714.

Vasudevan, D., Clark, N.K., Sam, J., Cotham, V.C., Ueberheide, B., Marr, M.T., 2nd, and Ryoo, H.D. (2017). The GCN2-ATF4 Signaling Pathway Induces 4E-BP to Bias Translation and Boost Antimicrobial Peptide Synthesis in Response to Bacterial Infection. Cell Rep *21*, 2039-2047.

Vicens, Q., Kieft, J.S., and Rissland, O.S. (2018). Revisiting the Closed-Loop Model and the Nature of mRNA 5'-3' Communication. Mol Cell *72*, 805-812.

Villa, N., Do, A., Hershey, J.W., and Fraser, C.S. (2013). Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome. J Biol Chem *288*, 32932-32940.

Wang, L., Gilbert, R.J., Atilano, M.L., Filipe, S.R., Gay, N.J., and Ligoxygakis, P. (2008). Peptidoglycan recognition protein-SD provides versatility of receptor formation in Drosophila immunity. Proc Natl Acad Sci U S A *105*, 11881-11886.

Wang, X., Li, W., Parra, J.L., Beugnet, A., and Proud, C.G. (2003). The C terminus of initiation factor 4E-binding protein 1 contains multiple regulatory features that influence its function and phosphorylation. Mol Cell Biol *23*, 1546-1557.

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell *2*, 135-140.

Wessells, R., Fitzgerald, E., Piazza, N., Ocorr, K., Morley, S., Davies, C., Lim, H.Y., Elmen, L., Hayes, M., Oldham, S.*, et al.* (2009). d4eBP acts downstream of both dTOR and dFoxo to modulate cardiac functional aging in Drosophila. Aging Cell *8*, 542-552.

Wilk, R., Hu, J., Blotsky, D., and Krause, H.M. (2016). Diverse and pervasive subcellular distributions for both coding and long noncoding RNAs. Genes Dev *30*, 594-609.

Wu, J.S., and Luo, L. (2006). A protocol for dissecting Drosophila melanogaster brains for live imaging or immunostaining. Nat Protoc *1*, 2110-2115.

Xiang, W., Zhang, D., and Montell, D.J. (2016). Tousled-like kinase regulates cytokine-mediated communication between cooperating cell types during collective border cell migration. Mol Biol Cell *27*, 12-19.

Xiao, Z., Zou, Q., Liu, Y., and Yang, X. (2016). Genome-wide assessment of differential translations with ribosome profiling data. Nat Commun *7*, 11194.

Yanagiya, A., Svitkin, Y.V., Shibata, S., Mikami, S., Imataka, H., and Sonenberg, N. (2009). Requirement of RNA binding of mammalian eukaryotic translation initiation factor 4GI (eIF4GI) for efficient interaction of eIF4E with the mRNA cap. Mol Cell Biol *29*, 1661-1669.

Yano, T., Lopez de Quinto, S., Matsui, Y., Shevchenko, A., Shevchenko, A., and Ephrussi, A. (2004). Hrp48, a Drosophila hnRNPA/B homolog, binds and regulates translation of oskar mRNA. Dev Cell *6*, 637-648.

Yeh, T.H., Huang, S.Y., Lan, W.Y., Liaw, G.J., and Yu, J.Y. (2015). Modulation of cell morphogenesis by tousled-like kinase in the Drosophila follicle cell. Dev Dyn *244*, 852-865.

Zaessinger, S., Busseau, I., and Simonelig, M. (2006). Oskar allows nanos mRNA translation in Drosophila embryos by preventing its deadenylation by Smaug/CCR4. Development *133*, 4573-4583.

Zhang, H., Dou, S., He, F., Luo, J., Wei, L., and Lu, J. (2018). Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during Drosophila development. PLoS Biol *16*, e2003903.

Zhang, H., Wang, Y., and Lu, J. (2019). Function and Evolution of Upstream ORFs in Eukaryotes. Trends Biochem Sci.

Zhao, F., Yu, C.H., and Liu, Y. (2017). Codon usage regulates protein structure and function by affecting translation elongation speed in Drosophila cells. Nucleic Acids Res *45*, 8484-8492.

Zhou, J., Wan, J., Shu, X.E., Mao, Y., Liu, X.M., Yuan, X., Zhang, X., Hess, M.E., Bruning, J.C., and Qian, S.B. (2018). N(6)-Methyladenosine Guides mRNA Alternative Translation during Integrated Stress Response. Mol Cell *69*, 636-647 e637.

Zid, B.M., Rogers, A.N., Katewa, S.D., Vargas, M.A., Kolipinski, M.C., Lu, T.A., Benzer, S., and Kapahi, P. (2009). 4E-BP extends lifespan upon dietary restriction by enhancing mitochondrial activity in Drosophila. Cell *139*, 149-160.

Zimyanin, V.L., Belaya, K., Pecreaux, J., Gilchrist, M.J., Clark, A., Davis, I., and St Johnston, D. (2008). In vivo imaging of oskar mRNA transport reveals the mechanism of posterior localization. Cell *134*, 843-853.

Zuberek, J., Kuchta, K., Hernandez, G., Sonenberg, N., and Ginalski, K. (2016). Diverse cap-binding properties of Drosophila eIF4E isoforms. Biochim Biophys Acta *1864*, 1292-1303.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res *31*, 3406-3415.