Translation and degradation: The role of the polysome in exoribonuclease degradation of long non-coding RNAs in *Drosophila*

Oliver Michael Rogoyski

A thesis submitted in partial fulfilment of the requirements of the University of Brighton and the University of Sussex for the degree of Doctor of Philosophy at the Brighton and Sussex Medical School

February 2021

Table of Contents

Candidate’s declaration:

I declare that the research contained in this thesis, unless otherwise formally indicated within the text, is the original work of the author. The thesis has not been previously submitted to this or any other university for a degree and does not incorporate any material already submitted for a degree.

Signed:

[redacted]

Dated:

09/02/2021

Acknowledgements:

I would first like to thank my primary supervisor, Professor Sarah Newbury, for providing me with the initial opportunity to undertake this PhD. Additionally, the support that she has given me, both as a supervisor and a scientist, has been crucial to my completion of this project. Her interest, enthusiasm, and expertise in the field of RNA has been an inspiration and a necessity to persevere throughout this project.

Just as importantly, I would like to thank Dr. Benjamin Towler, for his continued support and determination to help me reach the end of this project, even when I struggled to see the light at the end of the doctoral tunnel. Without his advice, assistance, and friendship, I have no doubt that this project would have posed a great deal more difficulty than it already did. He has my eternal thanks for answering so many questions and proofreading so many drafts.

Similarly, Dr. Ignacio Pueyo-Marques has been absolutely crucial to both my training in the techniques required for this project, and the planning and execution of said project. He has been an inspirational example of how determined and hardworking a scientist can be. Without his knowledge (particularly regarding polysome profiling), this project would not have been possible.

I would like to thank everybody in the Newbury lab (both past and present) who has assisted me in any way during my tenure there and helped create a great environment in which to learn to be a scientist. Particularly, my thanks go to Clare Rizzo-Singh; she determinedly keeps the lab running smoothly and keeps us well supplied and organised. Additionally, Dr. Amy Pashler was incredibly helpful in keeping my spirits up. During her time here, she built up a friendly and sociable environment which pervaded even after she moved on from the lab.

I am fortunate to have had the opportunity to supervise four talented and hard-working undergraduate students during my project. I would like to thank Alexa Tataru, Harry Pink, Lauren Mulcahy, and Hannah Markham. Particularly I would like to thank Lauren and Harry for their persistence and initiative in helping me look through candidate

lncRNAs. I am proud to see that they have kept up their interest in research, and I'm sure they will both go on to be excellent scientists.

Alongside all of this help and guidance from the Newbury lab, I was fortunate enough to also have the valuable help and guidance of my supervisor Professor Juan Pablo Couso, as well as the other talented scientists working in his lab. I consider myself extremely lucky to have had continued access to their expertise, and their perspectives on the work I carried out in the Newbury lab.

On a personal note, I would like to thank all of my family for helping me through the last four years. Without their unconditional love and support, I have no doubt that I would not have made it this far, and I cannot thank them enough. My father, Andrew, for nurturing my interest in science and "inventing" from a young age. My mother, Sue, for being my staunchest supporter, and providing me with the knowledge that somebody is always fighting my corner. My older sister, Beth, for providing me with a good example of how to balance success as a scientist with pursuing other interests (and not getting boring). My younger sister, Anya, for keeping me grounded and ensuring that I always have a friend I can talk to. Crucially, Melina, who has provided me with love, support, and kindness every day; without her, I simply do not know where I would be. Finally, my faithful and wonderful dog, Alfie. He provided comfort and looked after me, all the way until the end.

Commonly used abbreviations:

Abbreviation:	Definition:
Pcm	Pacman
XRN1	Exoribonuclease 1
Dis3	Defective in sister chromatid re-joining 3
Dis3L2	Defective in sister chromatid re-joining 3 like 2
RNA	Ribonucleic acid
lncRNA	Long non-coding RNA
mRNA	Messenger RNA
pre-mRNA	Preliminary messenger RNA
miRNA	Micro RNA
RNAi	RNA interference
dsRNA	Double-stranded RNA
ssRNA	Single-stranded RNA
DNA	Deoxyribonucleic acid
cDNA	Complementary DNA
5' UTR	5' Untranslated region
3' UTR	3' Untranslated region
PCR	Polymerase chain reaction
qPCR	Quantitative PCR
RT-PCR	Reverse-transcriptase PCR
RNA-seq	RNA-sequencing
Ribo-seq	Ribosome-sequencing
Poly-ribo-seq	Poly-ribosome-sequencing
CRISPR	Clustered regular interspaced shirt palindromic repeats

Abstract:

RNA transcript abundance is largely decided by a careful balance between the transcription and degradation of any given gene's transcripts. RNA stability plays a critical role in the availability of an RNA species, and the time period over which it is able to elicit its functions and roles. The majority of the literature on regulation of RNA activity by degradation focuses on mRNAs, with the assumption that their role is to be translated into a functional protein, as described by the central dogma. Increasingly though, non-coding RNAs have been recognised as crucial to the normal function of biological organisms. The roles of RNA species such as miRNAs in controlling gene expression are now relatively well understood, as are the molecular mechanisms by which they bind to and regulate RNAs.

In contrast, long non-coding RNAs (lncRNAs) are a very poorly understood (and arguably, defined) category of RNAs. However, they clearly have their own crucial roles to play, as this relatively recently discovered RNA species regulates gene expression in diverse ways, encodes small biologically relevant peptides, and has been involved in a large variety of important biological functions. Importantly, an increasing number of lncRNAs have also been associated with a range of human diseases, including neurodegenerative pathologies and cancer. The degradation of lncRNAs requires significant study, in order to bring understanding of this key regulatory step of a crucial class of transcripts up to the level of that of the rest of the transcriptome.

The aim of this thesis is to investigate the degradation of lncRNAs by the exoribonucleases Pacman and Dis3L2, in *Drosophila melanogaster*. Within this overarching aim, several smaller goals arise. Firstly, this thesis investigates whether certain lncRNAs are specifically and significantly degraded by Pacman and Dis3L2, as is seen with canonical RNAs. By examining previous RNA sequencing data from experiments carried out on exoribonuclease deficient *Drosophila* (both in vivo, and in *Drosophila* derived cell lines,) it was possible to identify promising candidates for lncRNAs with significantly altered abundance in the absence of either Pacman or Dis3L2. This existing work was validated with qPCR, proving the principle of specific regulation of lncRNAs by Pacman and Dis3L2.

Following this, an experiment was designed and carried out to examine the role of the translating ribosome in this degradation. Existing work has shown the ribosome to be associated with XRN1 and Pacman in humans and *Drosophila* respectively, and Dis3L2 has also been shown to associate with the ribosome in humans, although whether this occurs in *Drosophila* is unclear. By using the powerful technique polyribosome sequencing (poly-ribo-seq), on exoribonuclease deficient *Drosophila* samples, this work has identified a preliminary set of lncRNAs that appear not only to be specifically regulated by Pacman and Dis3L2, but also undergoing translation, indicating the presence of small open reading frames (smORFs) within the lncRNA genes.

Ongoing work will validate not only the upregulation of these transcripts in the absence of the relevant exoribonuclease, but also the putative smORF from which a peptide is likely produced. Following this, it will investigate whether a block in transcription eliminates the differential abundance of these transcripts in the absence of Pacman or Dis3L2. This work then, identifies an initial subset of lncRNAs regulated by Pacman and Dis3L2, and shows several of them to be actively translated, identifying novel peptides, potentially of biological significance (given their active translation and specific degradation). With the completion of ongoing work, this project will also elucidate whether their translation is important to the degradation of these transcripts.

1. Introduction:

1.1 The strengths and weaknesses of Crick's central dogma of molecular biology:

In 1958, Francis Crick published a succinct, simplified framework on which to build our understanding of molecular biology, known as the central dogma (1). At its most basic level, it dictates that DNA makes RNA, which in turn makes protein. Although a powerful summary, it does not tell the whole story of gene expression, as a wide range of distinct cell types exist in any higher eukaryote, despite each of those cells possessing identical DNA.

It is the transcription of certain subsets of genes into a dynamic RNA transcriptome, and the translation of a further subset of those into the actively regulated proteome; as well as the interactions between all of these (further influenced by environment), that define each cell, tissue, and organism. As such, the mechanisms of regulation between each step of the central dogma, and roles for non-canonical RNAs must be extensively explored in order to gain a deep understanding of the complexities of gene expression in a complex living organism.

Far from a simple one-way progression to a useful endpoint, we now know there to be a much more complex network of regulatory and functional roles for both nucleic acids and proteins. In particular, many RNAs play vital roles outside of the messenger molecule to facilitate translation that it has historically been relegated to (2). Species of non-coding RNAs (ncRNAs) include not only the extensively studied ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) involved in translation, but small nuclear RNAs (snRNAs) involved in splicing, small nucleolar RNAs (snoRNAs) involved in rRNA modifications, and now many relatively novel (and less well understood) classes of RNA, including piwi-interacting RNAs (piRNAs), micro RNAs (miRNAs), and long non-coding RNAs (lncRNAs).

So although the relevant DNA is of course necessary for production of proteins through an RNA intermediate, it is the careful and ongoing regulation of gene expression, both as nucleic acids and proteins, that allows the single set of instructions present in a newly fertilised human zygote to produce the entire organism; from the complex networks of neurons, to the cardiomyocytes that cause the beating of the heart, and the pancreatic α- and β-cells responsible for the secretion of glucagon and insulin.

1.2 RNA species:

Since RNA started being explored in earnest, a broad range of different species have arisen (2). From the most canonical protein coding messenger RNAs (mRNAs), to those that elicit and regulate their functions such as microRNAs (miRNAs) and small nucleolar RNAs (snoRNAs), and the transfer RNAs (tRNAs) and ribosomal (rRNAs) required to build the proteins they encode, multiple RNA species are required in different capacities just to carry out the classical role of RNA as asserted by the central dogma. Outside of this, there are a range of alternative functions carried out by ever increasing categories of non-coding RNAs. The following sections will introduce a comprehensive range of relevant RNA species, as well as both their known and speculative functions within the cell.

1.2.1 Messenger RNA:

A messenger RNA (mRNA) is an intermediate between the DNA-encoded gene and the protein product it can produce. mRNAs are transcribed from RNA polymerase II (RNAPII), and are spliced, capped, and polyadenylated within the nucleus (as previously described). Following this, the mature mRNAs are exported to the cytoplasm, where translation initiation can occur (also previously described), allowing their production of the peptides and proteins that they encode, and subsequently their degradation through a variety of RNA decay processes (briefly discussed previously, and to be further explored in subsequent sections of this chapter).

The three main sections of an mRNA are the 5'UTR, the coding sequence, and the 3' UTR. Both the 5' and 3' UTRs are structured regulatory elements which contain binding sites for a number of trans-acting regulatory factors, such as miRNAs and RBPs (previously discussed). The coding sequence is read from a particular start codon (potentially one of several, as mentioned), and is read in-frame, as three base long combinations of adenine, guanine, cytosine, and uracil (ribonucleotide triplets). In some instances, multiple alternative reading frames may exist within the same region. These consecutive codons in the RNA transcript are complimentary to certain amino acids recruited to the growing nascent peptide by tRNAs as the ribosomes read along the RNA transcript. The 64 different combinations of ribonucleotide triplets for codons, allows several codons to correspond to each amino acid, establishing the genetic code to be degenerate, as suggested even in Crick's early work. Having been the focus of RNA biology since its inception, mRNAs are by far the best studied category of RNA.

1.2.2 Non-coding RNA:

In addition to the protein-coding mRNAs, a plethora of non-protein-coding RNAs have come to light over the past several decades. Some, like tRNAs, have been well understood for a long time, while others, such as long non-coding RNAs, are still poorly understood, and only beginning to be extensively studied. Non-coding RNAs in fact vastly outnumber mRNAs, suggesting a vast, interconnected, network of RNA species with different roles, rather than a transcriptome centered completely on protein production.

1.2.2.1 Ribosomal RNA:

The 28S, 18S, and 5.8S rRNAs are co-transcribed by RNA polymerase I (RNAPI) into a single 47S pre-rRNA precursor, which subsequently undergoes a number of regulated cleaving and trimming steps, resulting on the generation of the final three rRNA structure (28S, 18S, and 5.8S). These three structures are bound by an array of other proteins, exported to the cytoplasm, and go on to undertake their interactions with the large and small ribosomal subunits. The small 5S rRNA is separately transcribed as an immature precursor by RNA polymerase III (RNAPIII), which similarly to 28S, 18S, and 5.8S must be processed into the final 5S rRNA and exported into the cytoplasm. Due to the fundamental biological role that they have in every known species, they are highly conserved, and comparisons between them can allow better understanding of evolutionary divergence and help build phylogenetic trees.

1.2.2.2 Small nucleolar RNA:

The synthesis of the previously discussed rRNAs requires another class of RNAs to proceed. Small nucleolar RNAs (snoRNAs), small sequences of RNA localized to the nucleolus, which mostly function to undertake necessary modifications to RNA structures (by methylation and pseudouridylation, which divides them into C/D box and H/ACA box snoRNAs respectively), allowing their target RNAs to fulfil their roles properly. Additionally, some snoRNAs are known to be involved in the regulation of splicing (3, 4), chromatin structures (5), and other functions (6).

1.2.2.3 Transfer RNA:

Transfer RNAs (tRNAs) are non-coding RNAs between 76-90 nucleotides in length, which function to deliver amino acids to their corresponding codons during translation (as briefly discussed earlier). tRNAs, like the 5S rRNA, are transcribed by RNAPIII, as immature pre-tRNAs which require processing to form the processed mature tRNA, containing 3 hairpin loops and a 3’ CCA tail. The CCA tail is the site that carries the amino acid to the ribosome and is charged with said amino acid by an aminoacyl tRNA synthetase. Each different amino acid corresponds to a specific tRNA with a specific anti-codon complimentary region, able to base pair with the3 nucleotide codon sequence within the translating RNA when delivered to the T site on the translating ribosome.

1.2.2.4 Micro RNA:

Micro RNAs (miRNAs) are small, non-coding RNAs (ncRNAs) ~22 nucleotides in length that play a crucial role in post-transcriptional regulation. They are broadly transcribed by RNAPII, with miRNAs within Alu-repetitive elements (transcribed by RNAPIII) as an exception to this (7). As with other RNAPII produced transcripts, most miRNAs are processed by 5’ capping, 3’ polyadenylation (8), and in some instances, splicing (9). The final processing required to fully mature miRNAs occur due to Dicer, an RNase that processes pre-miRNA into a 22bp dsRNA. This final part of processing is often coupled with formation of the miRNA-induced silencing complex (miRISC). Mature miRNAs implement their post-transcriptional role of genes by base-pairing with the 3’ UTR of specifically targeted mRNAs, and guiding the Argonaute (AGO) proteins to adjacent target sites (10). miRNA-loaded AGO forms an important part of the previously mentioned miRISC, which promotes pre-miRNA processing, as well as translational repression and degradation of targeted mRNAs (10).

1.2.2.5 Small nuclear RNA:

Small nuclear RNAs (snRNAs) are a class of small RNAs with an average size of 150nt, a variety of which are encoded by eukaryotic genomes. As a class, they are fairly abundant, localised to the nucleus, and playing an important role in intron splicing and RNA processing as part of the large, multi-megaDalton molecular machinery of the spliceosome. As well as an array of proteins, the spliceosome contains five uridine-rich snRNAs, (U1, U2, U4, U5, and U6). These snRNAs undergo complex conformational changes to correctly recognise splice sites, allowing them to facilitate intron removal. As well as splicing, evidence suggests snRNAs presenting as ribonucleoprotein particles (snRNPs) play an important role in nuclear maturation of primary mRNA transcripts, regulation of gene expression, and 3’ end processing of histone mRNAs.

1.2.2.6 Long non-coding RNA:

lncRNAs are a complex, and poorly understood class of RNAs, despite receiving increased attention over the past several years. For the purpose of this thesis, long non-coding RNAs (lncRNAs) are defined as being RNA transcripts longer than 200 nucleotides, which lack a significant open reading frame (greater than 100 amino acids in length) (11). This definition is routinely used in the annotation of the *Drosophila* and other genomes. This definition lacks nuance and will be addressed further (with a more thorough examination of this species of RNAs in section 1.6.

1.3 Transcriptional control of gene expression:

For DNA to fulfil its downstream function beyond replication, it must be transcribed into RNA. This requires an RNA polymerase to read the DNA code and copy the message into an RNA form. An RNA polymerase (together with transcription factors) will bind a promoter region, separate the DNA strands by breaking the hydrogen bonds between them, and add complimentary RNA nucleotides. The hydrogen bonds between the DNA template and the newly synthesized RNA are then broken, freeing the nascent RNA strand (which may then be subject to further processing and to cellular transport). This crucial step, allowing the synthesis of different RNA species by specific RNA polymerases, must be tightly regulated to avoid downstream consequences of aberrant RNA levels (such as abnormal protein levels, abnormal gene expression due to deviation in transcription of RNA molecules with regulatory roles.

Several well-studied factors are known to play a role in controlling the amount of RNA synthesised from each transcriptional unit. These factors implement their regulation at several points during the process of transcription (12). RNA polymerase must gain access to the DNA in order for transcription to occur; transcription can be repressed at a given genetic locus, by the tightly packed default state of DNA (known as heterochromatin) occluding polymerase.

In order to expose the DNA, modifications must be made to DNA packing proteins called histones. Histones act as spools, around which DNA can wrap, keeping them in the transcriptionally repressive, compacted conformation. Modification of histones by the addition of chemical groups to specific amino acids causes conformational shifts (13, 14), altering their availability to RNA polymerase. For instance, the acetylation of a lysine residue neutralizes its positive charge, resulting in a reduction of electrostatic attraction between histone and negatively charged DNA backbone.

Other than the control implemented through manipulation of DNA structure, *cis*-and *trans*- acting elements can regulate transcription (12). *Cis*-acting promoters can bind and initiate transcription proximal to their binding site, while *trans*-acting enhancers regulate translation via distal intermolecular interactions.

Of the vast number of transcription factors that have been studied, several of the best known have come to prominence as oncogenes (15), further demonstrating the importance of transcriptional control in understanding metabolic processes, human pathologies, and identification of pharmacological targets.

Control of transcriptional activity produces differential expression at the RNA level, but the RNAs produced are immature pre-RNAs, requiring splicing to remove introns, and in some cases to differentiate between different isoforms.

Splicing is carried out by the multi-component spliceosome. Splice sites are identified by snRNAs U1 and U2, and the subsequent recruitment of snRNAs U4, U5, and U6 which is facilitated by the cap binding complex (CBC) to cleave the RNA at intron-exon boundaries in a two-step trans-esterification reaction and ligate exons together as necessary for the production of mature RNAs (16). Research has found that many of these processing events occur co-transcriptionally(17, 18), with co-transcriptional versus post-transcriptional timing being somewhat indicative of splicing kinetics and favourability of introns (18).

The careful balance of varied mature isoforms produced from pre-RNA is necessary not just in generating proteomic diversity from a single coding sequence (with at least 70% of human genes expressing multiple mRNAs through alternative splicing(19)), but in determining cell fate. Different isoforms of the same gene can have significantly different, and sometimes directly opposed (20), functions, meaning that misregulation of the splicing process can have severe consequences. This is demonstrated phenotypically by the role of aberrant alternative splicing as well as generation of novel isoforms in multiple cancers(19), myotonic dystrophy(21), spinal muscular atrophy (19), and Amyotropic Lateral Sclerosis (22).

Differential splicing is also an important factor in RNA stability, with as many as 33% of alternative splicing events introducing premature termination codons, leading to degradation by nonsense-mediated decay (NMD). After the splicing is completed, protective terminal additions take place, adding a 5’ methylguanosine (m7G) cap, and a 3’ chain of adenines (poly(A) tail), both serving to protect the nascent RNA from degradation, as they are subsequently exported from the nucleus to the cytoplasm. Throughout the entire transcriptional process, a tight level of control is imposed, from altering DNA conformation to allow access to RNA polymerases, right up through control of splicing while transcription is ongoing, terminal modifications are being added, and even reaching forward to exert their influence in translational control.

1.4 Post-transcriptional control of gene expression:

1.4.1 An overview of RNA stability:

Whilst most of the early attempts to shed light on the mechanisms of controlled gene expression focused on transcriptional control, it has since become clear that gene expression is controlled at several levels (23), and that post-transcriptional regulation is vital to differential gene expression (24-26). Regulation of mRNA stability and degradation is now known to play a significant role in control of expression of protein-coding genes (26, 27).

As might be expected, the longer the cytoplasmic half-life of an mRNA transcript, the more likely it is to be translated, increasing the number of protein molecules it can produce across its lifetime(25, 28). One of the main ways that mRNA stability can be regulated is by the activity of exoribonucleases (29-31), enzymes which degrade mRNA transcripts (summarized in Figure 1.1), the activity of which contribute to controlling the amount of an mRNA transcript that is allowed to accumulate.

As well as controlling protein levels, RNA degradation is of course also necessary for the regulation of RNAs that have non protein-coding biological functions. Whilst mRNAs have been the primary focus of for studies of RNA stability and degradation, increasing interest in non-canonical RNA species necessitates re-examination of existing data and paradigms, and new work is needed in this area to fully appreciate how the established regulatory pathways apply to non-canonical RNAs.

Poly-adenylated tails (poly(A) tails) on the 3' end of RNA transcripts tend to increase stability (31) by protecting from degradation from the 3' end, with RNA stability often increasing with poly(A) tail length. This in turn is often dependent on the RNA sequence in the 3' UTR, with certain sequences allowing increased poly(A) tail length compared to others. Exchanging these regions between mRNAs has been shown to accordingly alter their half-lives, as repression of translation in the context of significant RNA decay is an efficient method for RNA clearance and prevention of protein synthesis (32).

Conversely, uridylation of RNAs can specifically target RNAs for significantly increased degradation by Dis3L2 (31, 33).

Meanwhile, 5’ capping provides protection for a transcript from 5’ to 3’ acting exoribonucleases such as XRN1/Pacman. XRN1/Pacman contains a C-terminal domain able to by co-factors important to its activity, such as Dcp1/Dcp2, allowing decapping to be directly coupled to subsequent degradation (34, 35). Capped and polyadenylated mRNAs can then be circularized by interactions between poly(A) binding protein (PABP) and the cap binding protein eIF4G; producing a conformation that allows higher translational efficiency and affords greater protection from degradation.

Whilst alterations to both ends of a transcript can play a significant role in the stability of RNAs, and the efficiency and specificity with which they are targeted, these are far from the only way in which RNA stability is modulated and regulated. Complex interactions with other biological molecules (both nucleic acids and proteins) create an interwoven array of means by which RNA stability is influenced and controlled.

MicroRNAs (miRNAs) are also known to have an impact on the stability of RNA transcripts, occurring through the interaction of GW182 (part of the RNA-induced silencing complex (RISC)) and Poly(A) binding protein (PABP). GW182 is also known to recruit Ccr4-Not and Pan2-Pan3 deadenylase complexes, causing release of PABP from the poly(A) tail, disrupting circularization, and facilitating translational repression and deadenylation (36), allowing increased degradation in the 3’ to 5’ direction (36, 37). Similarly, the RISC is able to recruit factors that decap the 5’ end of a transcript, increasing its vulnerability to RNA decay machinery acting 5’ to 3’ (38).

There is also an array of RNA binding proteins (RNA-BPs) that are recruited to RNA through association with specific cis sequences within the UTR, allowing another layer of regulation to be implemented. This includes action by HuR, HuD and TTP, which bind to AU-rich Elements (AREs) in the 3’ UTR to regulate the stability of transcripts (39, 40), as well as Bruno, which also binds within the 3’ UTR, but competes for cap binding with eIF4E, in order to repress translation (41). Similarly, 4E-HP has been shown to disrupt recruitment of the translation initiation complex by binding the cap (competing with

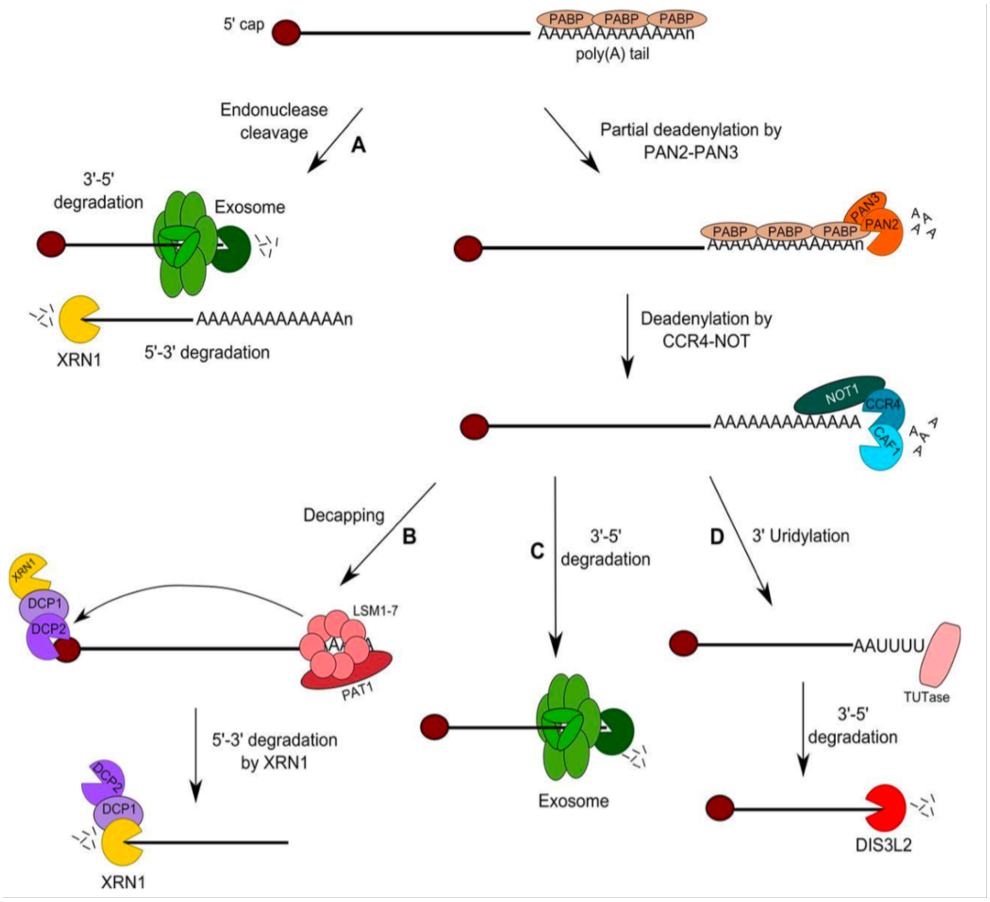


Figure 1.1 – Adapted from the Newbury lab – A cartoon depicting canonical RNA decay pathways

Pathway A – Transcripts can undergo cleavage during nonsense mediated decay or through endonucleases. This produces unprotected 3’ and 5; ends which are then degraded in either the 5’-3’ or 3’-5’ direction by XRN1 or the exosome respectively. Pathway B – Following deadenylation by complexes PAN2-PAN3 and CCR4-NOT and activating of the decapping complex by LSM1/PAT1, transcripts are degraded in a 5’-3’ direction by XRN1. [48-50] Pathway C – Alternatively, following deadenylation by PAN2-PAN3 and CCR4-NOT, transcripts can be degraded in a 3’-5’ direction by the exosome. Pathway D – Transcripts can also be degraded in a 3’-5’ direction by DIS3L2 following deadenylation and 3’uridylation by TUTases.

eIF4E, though not binding to eIF4G). The eIF4F initiation complex is a major target for regulation of translation, mostly through phosphorylation events, such as eIF4E binding proteins (4E-BPs). 4E-BPs compete with eIF4G to bind eIF4E. Once bound, eIF4E is unable to bind to eIF4G and initiate translation. This can be relieved by phosphorylation of 4E-BP by. Reduction of eIF4E-eIF4G binding thereby leads to global downregulation of translation as initiation complex formation is slowed at this crucial rate limiting step (42, 43), as well as having an important regulatory role in stress response.

In addition to managing the duration for which a protein-coding RNA persists, post-transcriptional regulation also occurs in terms of controlling the actual process of translation. A protein-coding RNA remaining intact in the cytoplasm does not guarantee it to be efficiently translated. Even if it is to be significantly translated, the extent to which it is will be dependent on several other factors separate from (but often linked with) RNA degradation (25). microRNAs (miRNAs) can also play an important role in RNA stability as well as in translation (36).

1.4.1.1 The crucial role of RNA stability in gene expression:

As previously discussed, a broad range of transcripts from different RNA species present within biological organisms are all required to carry out and regulate critical functions. These transcripts must be kept at the correct level of abundance by their careful regulation. Aberrant expression of many RNAs is indicative, and sometimes causative of, a wide range of pathologies (15, 44). RNA abundance is also time-sensitive, with the RNA profile in developing cells being crucial to determining cell fate (32).
The current level of RNA in a cell is determined by a balancing act between synthesis of RNA (by transcription) and decay of RNA (by degradation enzymes). Both of these processes are carefully maintained by an array of regulatory factors. In particular, the loss of control over RNA degradation is known to cause massive changes in RNA profile, with just as many consequences as might be expected by widespread disruption of RNA abundance (33, 34, 40, 45, 46).

1.4.2 Cis-acting factors affecting RNA degradation:

1.4.2.1 AU-rich elements:

RNA decay is a nuanced and complex mechanism, with complex means of differentially degrading different RNA transcripts, providing the variation in RNA half-lives required for the existence of dynamic, genetically complex organisms. AU-rich elements (AREs) are one of the best studied examples of specific targeting mechanisms for RNA degradation. AREs are 50-150 nucleotide regions of frequent adenine and uridine bases in the 3’ UTR of canonical RNAs. They frequently featuring multiple, overlapping “AUUUA” pentamers. The presence of at least one “UUAUUUA(U/A)(U/A)” nonamer increases turnover in chimeric RNAs, with multiple copies seeing an increased effect. These regions target the transcript for degradation and play a crucial role in gene regulation during cell growth, cell differentiation, and immune response. In addition, AREs are often found in the UTR of proto-oncogenes, transcription factors, and cytokines, further evidencing their importance as a regulatory feature.

AREs function to target transcripts for degradation by recruiting specific RNA binding proteins, which in turn are able to modulate RNA decay machinery, or in some cases by interacting directly with the RNA decay machinery. ARE-containing RNA is degraded by a broad range of cytoplasmic exoribonucleases (although the exosome is the best characterized by current literature in terms of direct interactions, with the RNase PH domain of Rrp41, Rrp43, and Rrp45 exosome subunits allowing binding to the AREs).

With the exception of direct interactions such as with these PH domains, AREs largely act through their recognition and interaction with RNA binding proteins such as tristetraprolin (TTP), AUF1, and Hu Antigen R (HuR). The precise mechanism by which this occurs is still debated, although existing research allows reasonable speculation. The effect of these proteins, however, can be easily observed. In mice, the absence of TTP leads to accumulation of *TNF-α*, and *GM-CSF*, which then leads to systemic inflammation, demonstrating the necessity of TTP (and the AREs that recruit it) in order to maintain the correct levels of certain RNAs, and avoid pathological misregulation.

Conversely to the action of AREs through TTP, the RNA-binding protein HuR enhances cell proliferation and survival by stabilising target RNAs (such as p21, c-fos, vascular

endothelial growth factor, MAPK phosphatase, tumour necrosis factor α), and modulating their translation. Interestingly, although most work on AREs as regulatory features has been carried out on mammals, AREs have also been identified and validated as being conserved in *Drosophila* 3' UTRs, with 16% of *Drosophila* genes containing the previously discusses mammalian ARE signature, according to the *Drosophila* ARE database (D-ARED).
Historically, the term ARE having been reserved for AU-rich regions containing the signature pentamer, conferring instability, or both; some regulatory regions of RNA, rich in adenosine and uracil nucleotides, that are not called AREs.

1.4.2.2 Regulation by 3' tailing:

The addition of a nucleotide tag to the 3' end of RNAs, in order to target them for degradation, is another crucial part of the complex web of regulation of degradation and has in fact been shown to be relevant to both 5' to 3' and 3' to 5' degradation pathways. While the functions of nuclear nucleotide tailing have been fairly well explored, the extent to which these mechanisms are mirrored in the cytoplasm has not been as thoroughly investigated.

Development of genome-wide techniques, such as TAIL-seq, has shown a range of 3' end modifications; with uridylation by terminal uridyl-transferases (TUTs) tending to follow shorter poly(A) tails, and guanylation tending to follow longer poly(A) tailng). As many as 80% of transcripts show some form of 3' end tag. In histone encoding mRNAs, oligo-uridylation of the 3' UTR by TUT4 stimulates binding of the Lsm1-7 complex, and subsequent decapping and degradation (from both directions). Similarly, uridylation of *pre-let7α* by TUT4 and TUT7 will decide its fate, with polyuridylation leading the premature transcript to degradation by Dis3L2, while monouridylation promotes the processing required to form the mature RNA, by Dicer2.

1.4.3 Trans-acting factors affecting RNA degradation:

1.4.3.1 Micro RNA mediated regulation:

Translation is primarily inhibited by miRISC disruption of translation initiation. The miRISC interferes with eukaryotic initiation factor 4 A-I (eIF4A-I), and eukaryotic initiation factor 4 A-II (eIF4A-II) in their interaction with target mRNAs by causing their dissociation (47), which subsequently prevents ribosome scanning and assembly of the (previously discussed) eIF4F translation initiation complex. The increase in degradation is implemented through recruitment of GW182 by AGO (48, 49), which interacts with polyadenylate-binding protein (PABPC). This subsequently promotes deadenylation of mRNAs by recruiting the complex of poly(A)-nuclease 2 and 3 (PAN2-PAN3), as well as the complex of carbon catabolite repressor protein 4 with NOT (CCR4-NOT). Deadenylation by these promotes decapping by the complex of mRNA-decapping enzymes 1 and 2 (DCP1-DCP2), thereby making the mRNA susceptible to rapid degradation (as previously discussed.

The recruitment of CCR4-NOT provides an additional means of translational repression through the recruitment of probable ATP-dependent RNA helicase (DDX6). miRNAs regulate genes in a network, with a single miRNA able to regulate hundreds of genes (50), while multiple regulatory miRNAs may be able to act on a single gene (51). Individual miRNAs and miRNA clusters are able to act on entire cellular pathways, and are able to completely shut off some genes, and fine tune the expression of others.

1.4.4 An overview of 3' to 5' ribonucleases:

1.4.4.1 The Dis3 family:

As previously stated, the removal of the poly(A) tail leaves the 3' end vulnerable to degradation by a family of 3' to 5' exoribonucleases. This family is highly conserved and is homologous to the RNaseII superfamily of bacterial ribonucleases. This superfamily is responsible for the majority of 3' to 5' RNA decay. Different members of this family are present in different species; some higher eukaryotes, including humans, feature Dis3, Dis3L1, and Dis3L2; others, such as *Drosophila* lack Dis3L1; while *S. cerevisiae* only has Dis3.

Dis3, as the only member of the family present within all eukaryotes, is responsible for providing catalytic activity to the exosome, a multi-component protein complex composed of nine (catalytically inactive) subunits forming a barrel structure, which funnels RNA towards the active site of Dis3. This structure allows for tight control over the highly processive enzyme, ensuring targeted transcripts are present for long enough to carry out their necessary function. In species lacking Dis3L1, Dis3 acts on a wider array of target transcripts, and is found to act (associated with the exosome) both within the nucleus, and on the cytoplasm. In higher eukaryotes featuring Dis3L1, the Dis3 is mostly limited to acting within the nucleus.

Dis3 features a PilT N-terminal (PIN) domain, facilitating interaction with the exosome and providing it with endoribonucleolytic cleavage activity, as well as the exoribonucleolytic activity provided by the RNB domain. This provides Dis3 with the capability to carry out degradation on a broader range of substrates via multiple mechanisms, unlike both Dis3L1 and Dis3L2. Dis3L1 maintains a PIN domain in order for it to associate with the exosome, but its PIN domain lacks endoribonucleolytic activity. While both are exosome associated, Dis3 functions mostly in the nucleus, while Dis3L1 is strictly limited to acting in the cytoplasm. Dis3L2 does not contain a PIN domain therefore, unlike Dis3 and Dis3L1, lacks both endoribonucleolytic activity and the capacity to interact with the exosome (which it functions independently of, unlike Dis3 and Dis3L1). Dis3L2 then, most closely resembles its bacterial precursor RNaseII. Of the Dis3 family, Dis3L2 is the nuclease that this thesis will focus on, and therefore will be explored further in a later chapter.

1.4.4.2 Dis3L2:

1.4.4.2.1 Structure and conservation:

Dis3L2 is a member of the highly conserved RNaseII family of 3’ to 5’ exonucleases. It is a highly processive exoribonuclease which degrades a variety of RNA transcripts with a high efficiency. As mentioned in an earlier section, Dis3L2 shares many features with the related bacterial RNaseII, lacking the N-terminal PIN domain featured in Dis3 and Dis3L1. The lack of this domain, responsible for endonuclease activity and exosome interactions, ensures Dis3L2 works only as an exosome-independent exoribonuclease, unlike Dis3 and Dis3L1 (confirmed by subsequent research). Dis3L2 does retain the exoribonuclease domain (RNB) and the three RNA binding domains.

Dis3L2 (or its relevant equivalent) has been only been well characterized in certain species, such as *S. pombe*, humans, and certain plants. The lack of Dis3L2 in the classical model organism *S. cerevisiae* (due to the previously mentioned varied conservation of Dis3 family enzymes,) has likely slowed progress on the understanding the enzyme better, as much of the founding work on RNA decay was carried out on this model, excluding Dis3L2 from being factored into this foundation of work. It functions separately from Dis3 and the exosome, as previously mentioned.

The structure of RNA-bound Dis3L2 has been resolved in mice (Figure 1.2), showing that the RNA is bound by two cold-shock domains in the N-terminal region of Dis3L2, before proceeding through the RNB domain, and into the catalytic site of the exoribonuclease. Here, the RNA is processively hydrolysed by the active site. Dis3L2 is known to degrade both coding and non-coding RNAs in this manner, with a mechanism that allows preferential targeting of transcripts uridylated by TUTs (discussed further in 1.4, 1.4.2.2, and 1.4.4.2.2).

1.4.4.2.2 Functions and phenotypes:

Although it has been the subject of significant research in the years since its discovery, its targets and roles in regulating cellular processes are only recently starting to be identified. Dis3L2 was found to be critical in cytoplasmic RNA, as a double mutation of *xrn1* and *dis3l2* in *S.pombe* was non-viable, while a double mutation of *dis3l2* and *lsm1* (a key component of the Lsm1-7 complex that leads to efficient decapping and 5' to 3' degradation) leads to an inhibition of RNA decay that is greater than that seen in either *dis3l2* or *lsm1* single mutants.

Although it clearly functions distinctly from both XRN1/Pacman and the exosome-dependent Dis3, the fact that it has been shown to degrade ARE-containing transcripts highlight it as at least having a similar role in mRNA degradation. This idea is reinforced by RNA-dependent interactions seen to occur between Pacman and Dis3L2, suggesting that although they have separate roles, there is at least some level of redundancy, and

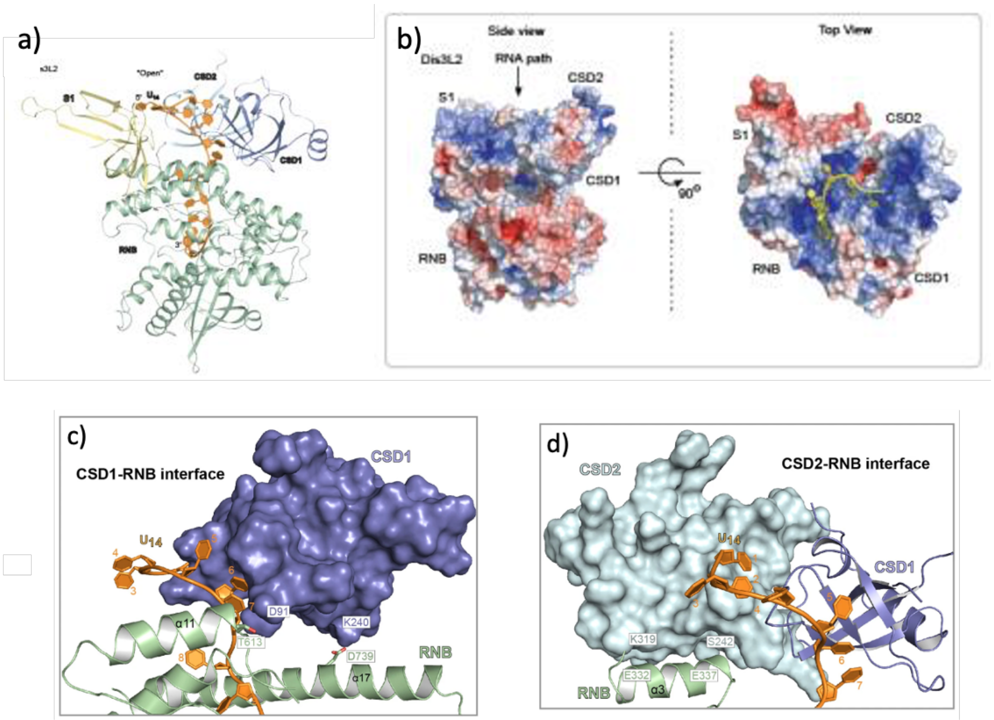


Figure 1.2 – Adapted from Faehnle et al. (2014) - Resolved structures of Dis3L2

(a) Dis3L2-U14 complex, The open funnel created by the CSD lobe and S1 allows RNA to access the “top” of the RNB.
(b) Perpendicular “side” and “top” views of the electrostatic surface potential of Dis3L2 (contoured at ±5 kT/e, white=neutral, blue=positive and red=negative). A positively charged electrostatic surface lines the wide portion of the funnel on the “top” of the RNB that can accommodate structured RNA substrates.
(c) Analysis of the CSD1–RNB interface. The conformation of CSD1 is stabilized by two protein–protein interactions with the RNB (K240 with D739 and D91 with T613) and an RNA-mediated interaction with the RNB through U6–U7 and α11.
(d) Analysis of the CSD2–RNB interface. CSD2 is intimately associated with CSD1, but also forms an interface with the RNB through α3 (S242 with E337 and K319 with E332).

that certain transcripts may be degraded from both the 5’ end and the 3’ end simultaneously. Other similarities have also been observed, such as the increase of P-bodies seen with a loss of either XRN1/Pacman or Dis3L2 (although the loss of Pacman in *Drosophila* caused an increase in P-body size, while the loss of Dis3L2 in *S. pombe* increases the number of P-bodies. This may again point to some overlap in function between the two, with migration of accumulating transcripts to P-bodies in an attempt to compensate for the loss of one nuclease by the activity of degradation in P-bodies.

Polyuridylation is now fairly well established as a conserved targeting mechanism by which certain transcripts can be marked for preferential degradation by Dis3L2. Mutation of *dis3l2* in both humans and *S. pombe* causes erroneous accumulation of transcripts featuring uridine tails. A well-documented example of this is seen in the synthesis of the miRNA precursor, *pre-let-7α*, is guided by uridylation of the transcript. Polyuridylation causes the degradation of the transcript by Dis3L2 in both humans and mice, while monouridylation promotes processing by Dicer2, and RNA maturation. This demonstrates an elegant mechanism for directing specific transcripts to certain fates, particularly degradation by Dis3L2.

Mutations in Dis3L2 cause pathological phenotypes, with a known foetal overgrowth syndrome (Perlman syndrome), resulting from *dis3l2* mutations in humans. Perlman syndrome, a congenital overgrowth syndrome causing foetal gigantism, enlargement of organs, macrocephaly, facial abnormalities, neurodevelopmental delay, and high neonatal mortality. *Dis3L2* is also associated with sporadic occurrence of a form of nephroblastoma known as Wilms’ tumour. Wilms’ tumour patients are seen to have a huge enrichment of *dis3l2* mutations (30%). Wilms’ tumour arises from uncontrolled proliferation of cells immature, un-differentiated, kidney cells. Another overgrowth condition, a Marfan-like syndrome with skeletal overgrowth is also seen associated with aberrations in the *Dis3L2* gene. These bear a striking resemblance to the overgrowth phenotypes seen in *Dis3L2* mutants in *Drosophila*.

Dis3L2 is known to be required for efficient clearance of mRNA following signals released from human cells undergoing apoptosis, and is required for proper and effective cell death, with knockdown of *dis3L2* resulting in inhibition and reduction of cell death. Of these accumulating transcripts, many do seem to be TUT4/TUT7 dependent, with uridine rich tails. In human cells, Dis3L2 has been shown to degrade miRNA *miR-27a* by target RNA-directed miRNA degradation, and associate with Ago2 in the RISC.

Depletion of Dis3L2 in *Drosophila* wing imaginal discs (WIDs) leads to increased proliferation of cells, resulting in larger WIDs (Figure 1.3), and wings (Figures 1.4 and 1.5) in the adult flies (shown to be exclusively due to an increased number of cells, with no increase in cell size. Data analysis of RNA-sequencing carried out on Dis3L2 deficient *Drosophila* WIDs determined a discrete set of transcripts, including several RNAs with known functions. Interestingly, simultaneous depletion of Pacman and Dis3L2 in *Drosophila* actually rescues the increased proliferation phenotype seen in the Dis3L2 deficient single mutant, suggesting that they work on antagonistic pathways (apoptosis and proliferation). Given how well conserved both of these enzymes are, it seems likely that these roles will be maintained in other multicellular organisms.

1.4.4.2.3 Dis3L2 expression patterns in Drosophila:

As previously mentioned, RNA sequencing data from modENCODE allows genome wide analysis of RNA expression patterns throughout *Drosophila* anatomy, developmental timeline, and *Drosophila* derived cell lines (Figure 1.6). The expression of *dis3L2* mRNA is found in by far its highest level in ovaries, followed by the imaginal discs, and the testes at progressively lower levels, with Dis3L2 is found expressed in all other tissues at low levels. Interestingly, the tissues with the highest levels of Dis3L2 and the highest levels of Pacman are very similar, suggesting these tissues to be in need of constant and thorough RNA regulation by controlled degradation.

Throughout the developmental timeline, *dis3L2* is expressed constantly at at least a low level. The first two hours of embryonic development see *dis3L2* expressed at its highest level anywhere in the *Drosophila* life cycle. From this point, it decreases steadily, being expressed at low levels for late embryo and larval stages, before increasing slightly during early pupation. Following this, *dis3L2* remains expressed at a moderate level. Again, some similarity can be seen with the expression timeline of *pacman*. *dis3l2* is also

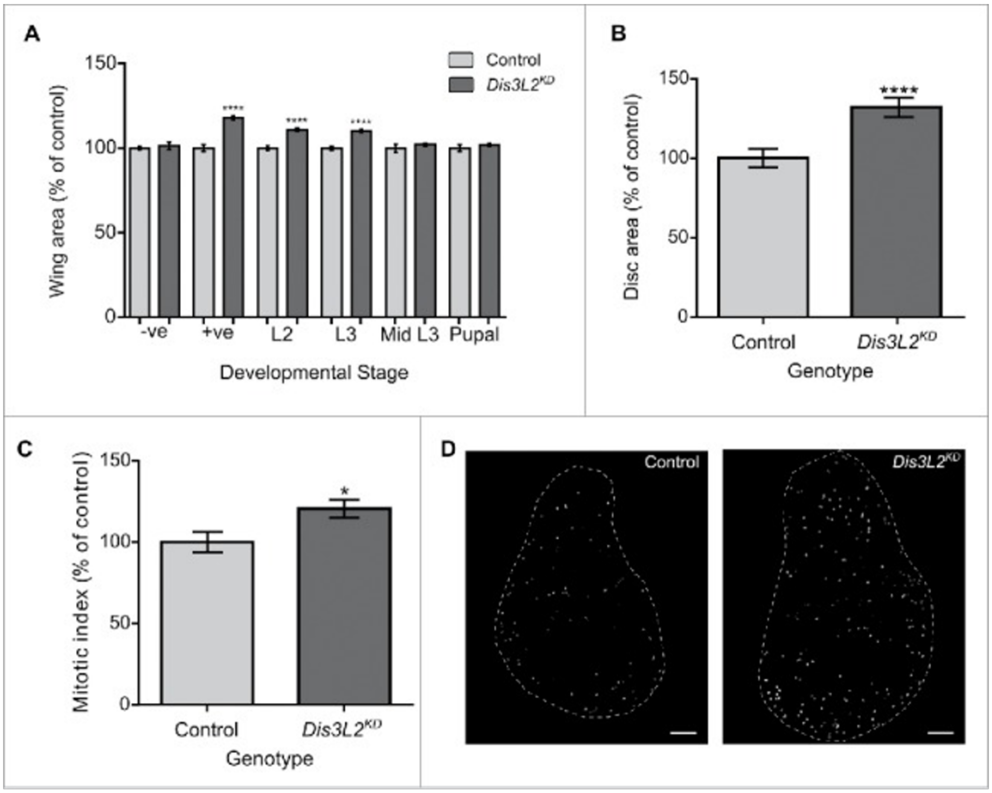


Figure 1.3 – Adapted from Towler et al. (2016) – Knockdown of dis3L2 results in increased proliferation within the wing imaginal disc leading to overgrowth of the disc

(a) Using the GAL80ts system to control the developmental timings of dis3L2 knockdown throughout the wing imaginal disc revealed that wing overgrowth is due to processes occurring during early Drosophila development. –ve represents flies that were maintained at the permissive 19°C temperature throughout their development. +ve represents flies that were maintained at the active 29°C temperature throughout their development. The positive control (+ve), start of L2 and start of L3 timings show a significant increase in adult wing area (**** = p < 0.0001) compared to parental controls. The negative control (–ve), mid L3 and start of pupal timings show no significant increase in wing area of knockdown flies. "Control" represents the mean wing area of the 2 parental controls. n ≥ 37, error bars represent 95% confidence limits.

(b) Knockdown of dis3L2 in the wing imaginal disc using the 69B-GAL4 driver results in significant overgrowth of the disc compared to controls. Dis3L2KD late L3 wing discs are 132.0% the size of parental control discs. n ≥ 21, error bars represent 95% confidence limits, **** = p < 0.0001.

(c) Dis3L2KD wing discs have a significantly higher mitotic index (20.5% increase in the percentage of cells undergoing mitosis) than control discs. n ≥ 13, * = p = 0.0241, error bars represent 95% confidence limits. For (B) and (C) "Control" shows the mean measurements of the 2 parental controls.

(d) Representative images of the DeadEasy MitoGlia output following staining of late L3 wing imaginal discs with anti-Phosphohistone H3. Scale bar = 50 µm.

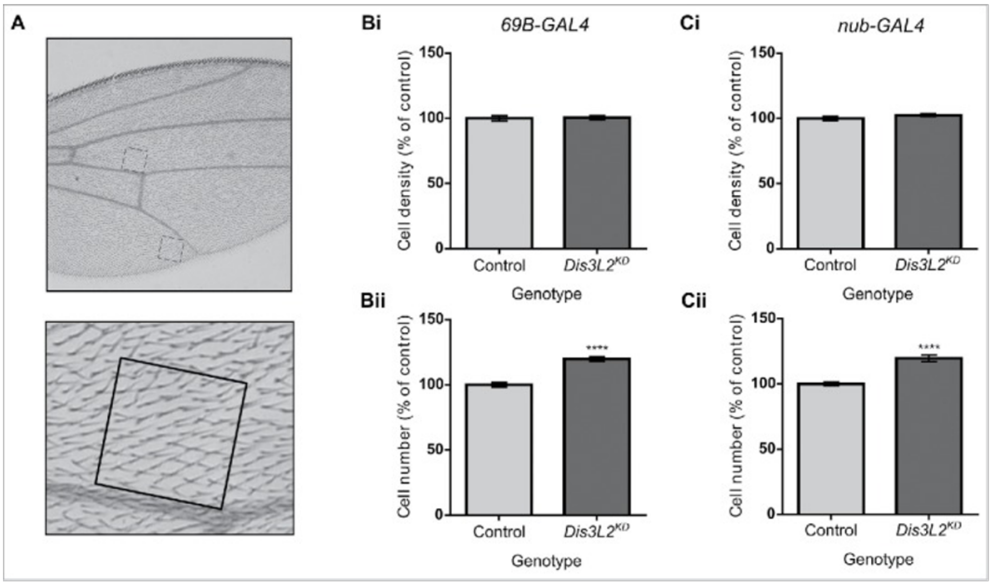


Figure 1.4 – Adapted from Towler *et al*. - Graphs and images demonstrating the overgrowth and causes of overgrowth in the wings of Drosophila Dis3L2 null mutants

Wing overgrowth is due to an increase in cell number rather than increased cell size. (a) Representation of the 0.1mm regions of the wing used to calculate the number of cells within the wing. Blown up image shows the single hairs protruding from each cell that were counted to give the number of cells. (b-c) Knockdown of dis3L2 throughout the wing imaginal disc using 69B-GAL4 (b) and nub-GAL4 (c) results in an increase in cell number (ii) rather than an increase in cell size (i) compared to controls. In each case, the “control” is the average of the parental genotypes plus a UAS-EGFPRNAi driven by the respective GAL4 driver, none of which were significantly different from each other. For (b) and (c) n ≥ 10, error bars represent SEM, **** = p < 0.0001.

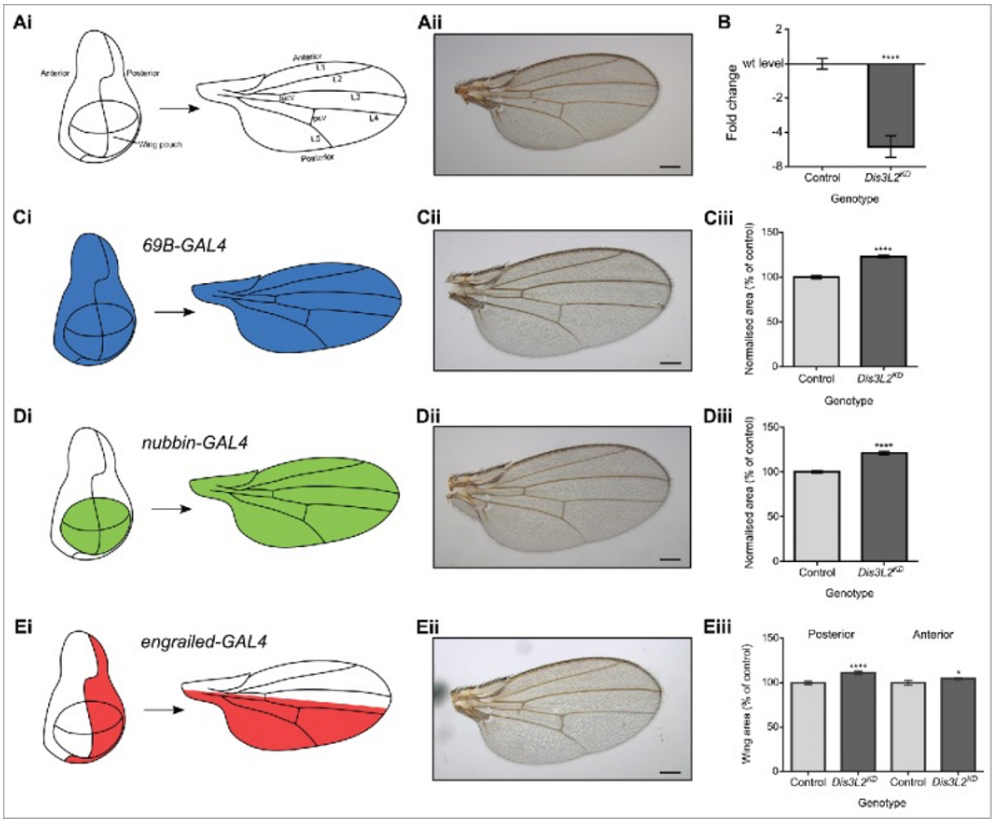


Figure 1.5 – Adapted from Towler et al. (2016) - Knockdown of dis3L2 in the wing imaginal disc results in specific overgrowth

(a) Representation of the developmental axis and regioKnockdownns of the Drosophila melanogaster wing and wing imaginal disc together with a parental control wing (UAS-dis3L2RNAi).
(b) Knockdown of dis3L2 throughout the wing imaginal disc using the 69B-GAL4 driver results in 5.2-fold knockdown (to 20% the level of controls) at the RNA level compared to the parental controls. n ≥ 10, error bars represent SEM, **** = $p < 0.0001$.
(c) Knockdown of dis3L2 throughout the wing imaginal disc using the 69B-GAL4 driver results in significant overgrowth of the wing (23%) compared to controls when wing area is normalized to fly mass.
(d) Knockdown of dis3L2 within the wing pouch of the wing imaginal disc using the nub-GAL4 driver results in significant overgrowth (21%) of the wing compared to controls when wing area is normalized to fly mass. For (c) and (d), the "control" is the mean of the parental genotypes plus a UAS-EGFPRNAi driven by the respective GAL4 driver, none of which were significantly different from each other. n ≥ 18, error bars represent 95% confidence limits, **** = $p < 0.0001$.
(e) Restricting knockdown of dis3L2 to the posterior region of the wing using the en-GAL4 driver results in specific overgrowth of the posterior compartment (11%). n ≥ 20, error bars represent 95% confidence limits, **** = $p < 0.0001$, * = $p < 0.05$. Representative images of a male wing are shown in each case, scale bar = 200 μm.

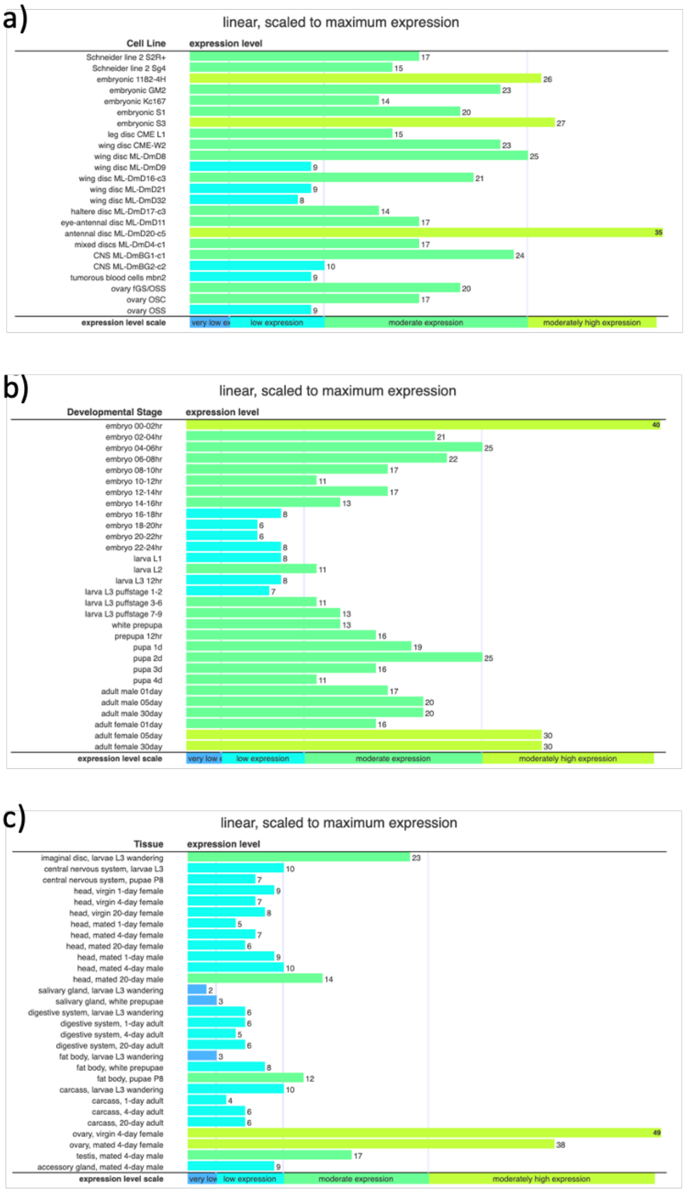


Figure 1.6 – Expression of *dis3l2* (scaled to maximum expression)

(a) *Drosophila* derived cell lines.
(b) Developmental timepoints.
(c) Specific tissues.

Graphs from modENCODE data.

Guide to modENCODE expression level colors

🟪 **no/extremely low expression (0 - 0)**
🟦 **very low expression (1 - 3)**
🟦 **low expression (4 - 10)**
🟩 **moderate expression (11 - 25)**
🟩 **moderately high expression (26 - 50)**
🟨 **high expression (51 - 100)**
🟧 **very high expression (101 - 1000)**
🟥 **extremely high expression (>1000)**

present to some level in all commonly used *Drosophila* cell lines, allowing easy exploration of *dis3l2* both in vivo, and in the closest equivalent cell lines.

1.4.5 An overview of 5' to 3' ribonucleases:

1.4.5.1 Pacman and XRN1:

1.4.5.1.1 Structure and conservation:

The XRN family of nucleases were first identified in *S. cerevisiae*, with the discovery of both XRN1 (175kDa) and XRN2 (115kDa). These exoribonucleases have been studied in great depth in the decades since their discovery, with functional orthologues of one or both of them being found pervasively through model organisms, as well as in humans. XRN1 and XRN2 show substantial conservation within the N-terminal region, active site, and mechanistically, with the sequence similarity being highest around the active site (certain conserved residues have been identified as necessary for function). Orthologs of these enzymes feature widely conserved structures and residues across many organisms (Figure 1.7).

The crystal structure of *Drosophila* XRN1, known as Pacman (184kDa) has been determined at a high resolution, showing how highly conserved it is (Figure 1.7). The structure, as well as substantial molecular research has allowed the formation of a model mechanism of action for the nuclease: Pacman features a narrow entry to the active site, thought to prevent double-stranded RNA (dsRNA) from entering, and reducing secondary structure as RNA is pulled through this entry. Once the transcript has entered, the first three nucleotides are held in place by a pair of highly conserved residues: His41 and Trp540. Once held, the first nucleotide is repositioned by a pocket of basic residues, in order to expose the phosphate bond to a pair of Mg2+ ions, allowing cleavage. The structure of the active site limits which RNAs can be degraded, as larger structures such as the m7G cap or triphosphorylated RNAs cannot fit into the basic residue pocket. An overhang of at least 4 nucleotides is required for efficient degradation, in order for the RNA strand to reach the active site and be cleaved. Once

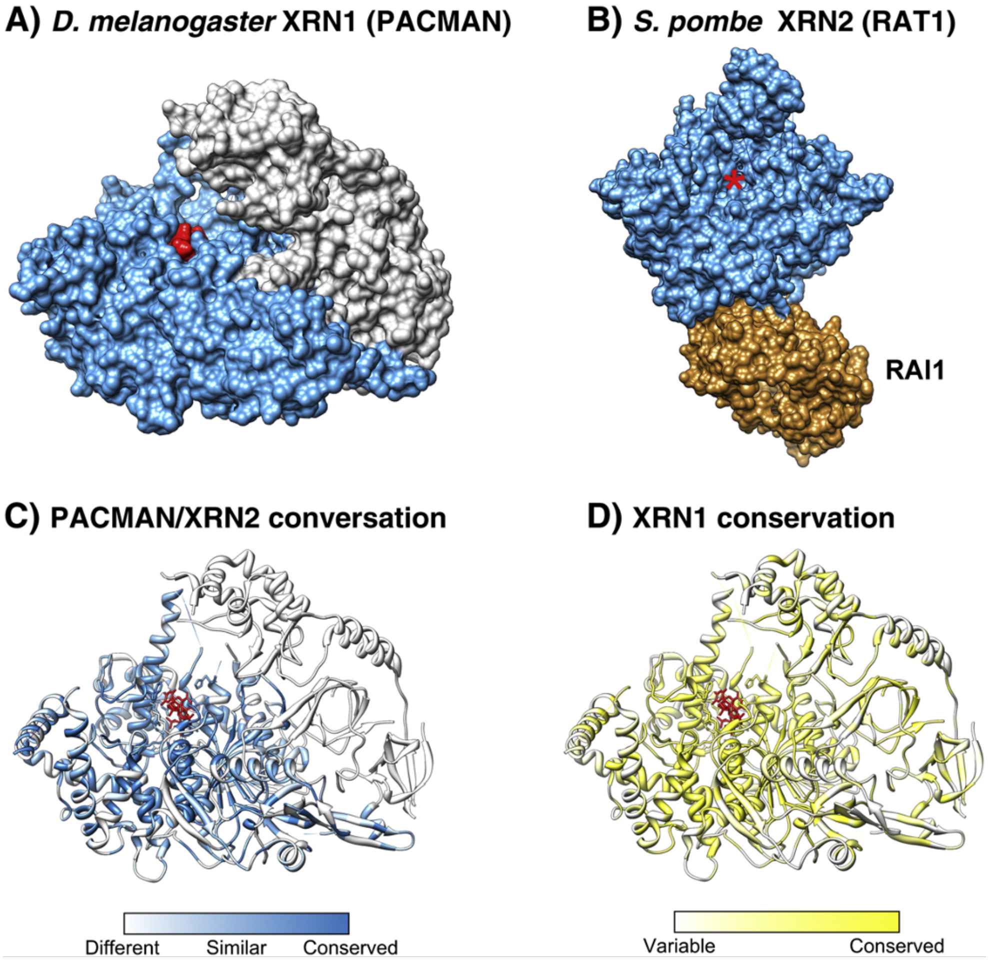


Figure 1.7 – Adapted from Nagarajan et al. (2013) – Structural comparisons between XRN1 and XRN2

a) Crystal structures of D. melanogaster XRN1 (PACMAN, PDB ID: 2Y35, residues 1–1141 of 1612) . The catalytic domain (N-terminal) of PACMAN is shown in blue with three nucleotides of decapped RNA (red) bound in the active site. The C-terminal is shown in gray.

b) S. pombe XRN2 (RAT1, PDB ID: 3FQD, residues 1–885 of 991) with its binding partner RAI1 (all of 352 residues). XRN2 is shown in blue with the active site marked by a red asterisk. RAI1 is shown in light brown.

c) Similarity of residues between XRN1 and XRN2. Residues of PACMAN are colored by similarity to residues in *D. melanogaster* XRN2. Residues are most strictly conserved around the active site and less so towards the C-terminal, much of which is not present in XRN2.

d) Conservation of residues across 195 eukaryotic XRN1s. Conservation is greatest around the active site, but there is also good conservation in parts of the C-terminal. Multiple alignment and conservation scores were calculated using ConSurf, which takes into account the phylogenetic distance between species. XRN2 alignments were excluded from the calculation. All images (and the alignment of PACMAN and XRN2 sequences) were produced in UCSF Chimera.

this cleavage has occurred, His41 and the pocket of basic residues moves in the next nucleotide by a Brownian ratchet mechanism.

The C-terminal portion of XRN1/Pacman is flexible, and low in complexity. This region is much less well conserved than other areas, but still includes some regions of conservation known as short linear motifs (SliMs), in the C-terminal interacting domain (CIR). This suggested these sequences may play an important role in XRN1/Pacman activity. Further exploration into these regions found that XRN1 interacts with CCR4-NOT via multiple short motifs embedded within the CIR. The flexible C-terminal region is also speculated to act as a scaffold for other factors that are involved in 5' to 3; degradation and has been shown to interact directly with the decapping factor Dcp1 (required for efficient mRNA decapping). A structure like this is not unheard of in this kind of enzyme, having been previously observed in another ribonuclease, RNaseE, present in *E. coli*.

1.4.5.1.2 Functions and phenotypes:

As mentioned earlier, XRN1/Pacman carries out key molecular functions including in mRNA stability, RNA interference, and miRNA-mediated gene regulation, necessary to many vital processes in the cell. XRN1/Pacman is responsible for bulk 5' to 3' RNA turnover, a function that was initially identified before the complexities of specific degradation by the enzyme were understood. In addition to this role in non-specific RNA clearance, XRN1/Pacman degrades a broad variety of cytoplasmic RNAs and is able to do so with preferential targeting of transcripts with certain features.

XRN1/Pacman is also vital in the nonsense mediated decay (NMD) process. In yeast, XRN1 is recruited to the NMD complex after 5' cap removal in order to degrade polyadenylated mRNA containing premature termination codons (PTCs), known to occur in the polysome.
Meanwhile, in *Drosophila*, NMD primarily occurs without the need for cap removal nor deadenylation, by endonucleolyitic cleavage by SMG6. In yeast, XRN1 are responsible for degrading an entire subclass of lncRNAs called XRN1-sensitive unstable transcripts

(XUTs). These XUTs are regulatory non-coding RNA transcripts, transcribed by RNAPII, and polyadenylated, as with canonical mRNAs. As many as 66% of them are antisense to open reading frames. These are not the only functional lncRNAs that XRN1/Pacman is known to degrade, with *GAL* lncRNAs (regulators of the genes encoding the GAL protein,) being another (deadenylation independent) target, as well as many more being potential lncRNA targets of XRN1/Pacman having been putatively identified by analysis of large datasets.

This thesis focuses on Pacman, the 5′ to 3′ cytoplasmic exoribonuclease, directly equivalent to XRN1 in many other model organisms, and humans, with the Drosophila Pacman protein able to complement the exonucleolytic activity of yeast Xrn1p. The *pacman* encodes the 184kDA Pacman protein, and mutations in the *pacman* gene that lead to reduced levels of Pacman have been shown to have severe phenotypic defects caused by the disruptions of the cellular functions that Pacman normally carries out. *Drosophila* with a null mutation for *pacman* are non-viable, and will not survive pupation, reinforcing the importance of the enzyme.

The previously mentioned hypomorphic mutations in *pacman* result in viable adult flies that develop at a reduced rate compared to relevant control genotypes. The adult *pacman* mutants carry a number of defects, including dull wings, ruffling of the posterior wing margin, and misshapen and erroneously placed sensory bristles. Additional reduction of the copy number of *pacman* to a single copy induces a cleft thorax phenotype, in which the wing imaginal discs fail to seal together during development, mimicking a deficiency in wound healing, which Pacman also plays a role in. An overview of these immediate phenotypes reveals Pacman to carry out substantial activity in the wing imaginal discs (WIDs), as indicated by the disproportionate number of phenotypic defects seen in this region compared to the rest of the *Drosophila* body. The WIDs are precursor tissues in *Drosophila* that go on to form the adult wings, sensory bristles, and part of the thorax. Formation of wing discs begins during embryogenesis, proceeding to grow and differentiate throughout larval and pupal development, allowing the generation of the adult tissues. The cells making up WIDs share similarities with adult stem cells, having pluripotent capabilities, able to differentiate into a number of different cell types. Aside from the importance of examining them due to how they’re impacted by *pacman* mutations, their similarities to adult stem cells make them an excellent model system for examining growth, differentiation, and specification of adult tissue.

When examined further, *pacman* hypomorphic mutant and null mutant WIDs were found to only be 82% and 45% the size of wild type WIDs respectively, with the adult wings of the hypomorphic mutants only reaching 84% of the wild type size. Wing discs from *pacman* null mutant larvae, were able to be dissected before reaching the lethal point in their disrupted development. As they die before developing into adult flies, wings do not develop, and can't be measured (Figure 1.8).

In the course of investigating the causes of the phenotypes produced by Pacman deficiency, RNA-sequencing identified 1207 genes significantly upregulated, and 1291 significantly downregulated in *pacman* null mutants compared to appropriate wild type controls. Although it might seem unusual that disruption of exoribonuclease function leads to approximately as many genes being downregulated as upregulated, the upregulated genes tended to be increased in expression by a greater magnitude than the reduction those downregulated. Those RNAs that are reduced by mutation of *pacman* are likely impacted by indirect effects.

The genes *reaper, hid, dilp8* and *Nplp2* were identified as post-transcriptionally upregulated in *pacman* null WIDs, where they would ordinarily have been degraded by Pacman. The resulting accumulation of these transcripts leads to increased translation of their apoptotic (Reaper and Hid) and signaling (Dilp8 and Nplp2) proteins, causing the increased apoptosis responsible for the undersized tissues (supported by the level of protein synthesis not significantly changing in *pacman* mutants, despite the undersized tissues). This work provided a strong example of specific regulation of biologically relevant RNAs, as well as elucidating the importance of regulation by Pacman in apoptotic pathways.

1.4.5.1.3 Pacman expression patterns in *Drosophila*:

RNA sequencing data from modENCODE allows genome wide analysis of RNA expression patterns throughout *Drosophila* anatomy, developmental timeline, and

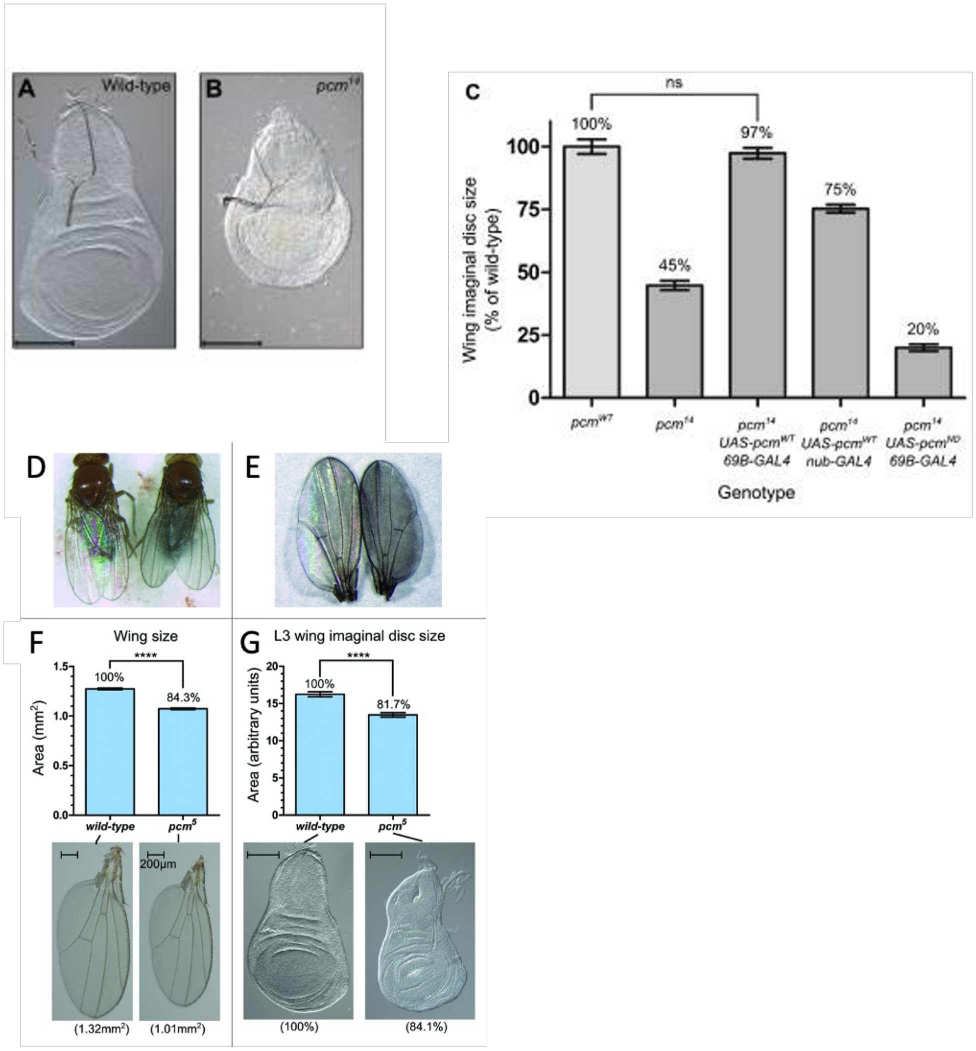


Figure 1.8 – Adapted from Jones et al. (2013), and Waldron et al. (2015) – "pcm14" genetic null *pacman* mutant larvae have significantly smaller imaginal discs than wild-type larvae.

Panels (a) and (b) show Representative wild-type and pcm14 wing imaginal discs. Scale bar represents 100 µm. Panel (c) show the mean size of pcm14 wing imaginal discs is 45% the size of wild type. This phenotype can be rescued by expressing a UAS-pcmWT construct throughout the wing imaginal disc cells using the 69B-GAL4 driver. Driving UAS-pcmWT expression with nub-GAL4 partially rescues this phenotype to 75% the size of wild type. Expressing a UAS-pcmND construct throughout the disc reduces the mean wing disc size to 20% of wild type (n≥31). "pcm5" hypomorphic pacman mutant wing and wing imaginal disc phenotypes. Panels (d) and (e) pcm5 wings (right) frequently display a dull wing phenotype where the wild-type iridescence (left) is lost. Panel (f) Shows the wings of pcm5 males are smaller than that of wild-type males. The mean wing area of pcm5 males is 1.07 mm2, 16% smaller than equivalent wild-type wings (1.27 mm2). A t-test was used to calculate significance and error bars show standard error (n = 20 for wild-type and 21 for pcm5 **** = p < 0.0001). Representative wild-type and pcm5 wings are shown below (actual size in parenthesis). Panel (g) Shows the wing imaginal discs of pcm5 L3 larvae are 18% smaller than those of wild-type discs. A t-test was used to determine significance and error bars show standard error (n = 32 for wild-type and 25 for pcm5, **** = p < 0.0001). Representative wild-type and pcm5 L3 wing discs are shown below (relative size in parenthesis).

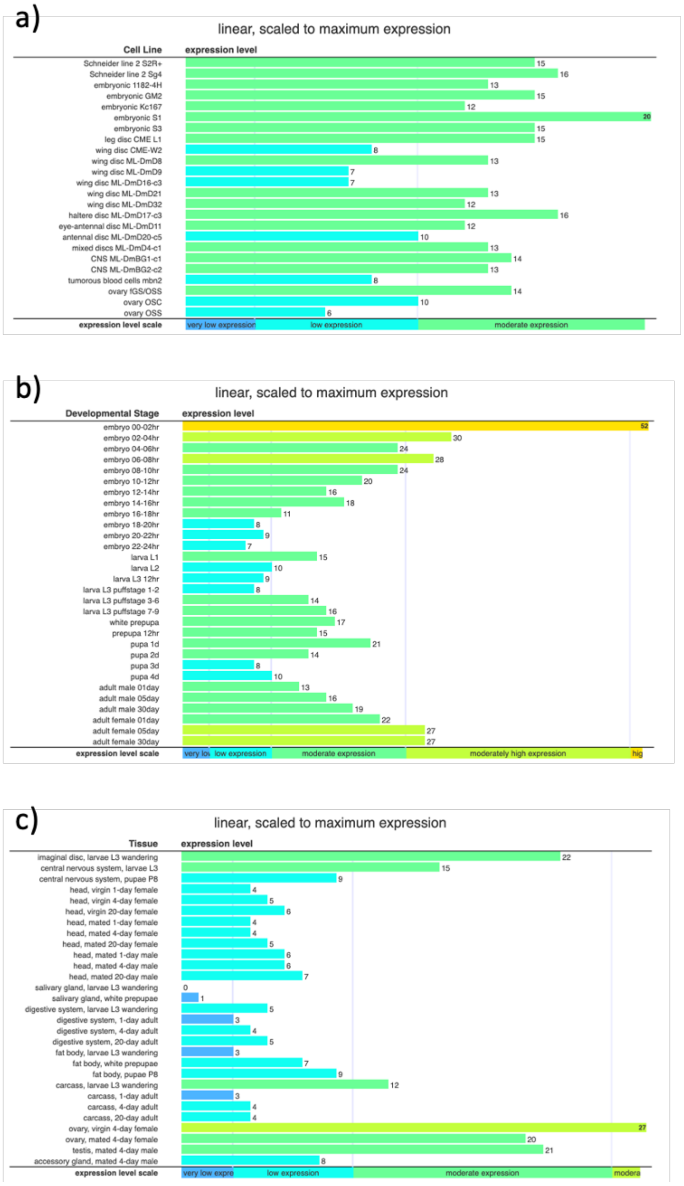


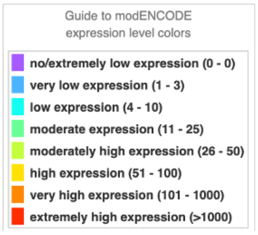
Figure 1.9 – Expression of *pacman* (scaled to maximum expression)

(a) *Drosophila* derived cell lines.

(b) Developmental timepoints.

(c) Specific tissues.

Graphs from modENCODE data.


Guide to modENCODE
expression level colors
no/extremely low expression (0 - 0)
very low expression (1 - 3)
low expression (4 - 10)
moderate expression (11 - 25)
moderately high expression (26 - 50)
high expression (51 - 100)
very high expression (101 - 1000)
extremely high expression (>1000)

Drosophila derived cell lines (Figure 1.9). The developmental timeline shows Pacman to be expressed at some level throughout all measured timepoints. During the first two hours of embryonic development, *Pacman* is expressed at its highest level anywhere in the *Drosophila* life cycle. From this point, it decreases rapidly, being expressed at low levels for late embryo and larval stages, before a temporary boost in expression appears for early pupation. Following this, *Pacman* drops back to lower levels, before a final surge in its abundance towards the later phases of the adult *Drosophila* lifespan. This timeline makes sense for an enzyme with important roles in both RNA clearance and ensuring the correct stability of transcripts needed for cellular functions, such as growth and apoptotic regulation. Anatomically, *Pacman* is present at higher levels in imaginal discs, ovaries, and testes, all tissues that require substantial regulation of transcript abundance, in order to correctly direct determination of cell fate. *Pacman* is also present to some level in all commonly used *Drosophila* cell lines, allowing easy exploration of *pacman* both in vivo, and in the closest equivalent cell lines.

1.4.6 Pathways of RNA decay:

Cytoplasmic mRNA is ordinarily protected from exoribonuclease decay by the 5’ m7G cap and the 3’ poly(A) tail; features added by post-transcriptional processing (also required to efficiently carry out subsequent translation). These protective features must be removed in order to allow efficient degradation of a transcript, starting with deadenylation. Deadenylase complexes Ccr4-Not and Pan2-Pan3 are recruited to RNA transcripts, before going about removing the poly(A) tails (reviewed in more detail by Wahler and Winkler (52)). Following this deadenylation, can be degraded in from the 3’ end by 3’ to 5’ ribonucleases in the RNase II superfamily, such as Dis3L2. Alternatively, further processing can occur to remove the 5’ m7G cap by the Dcp1-Dcp2 decapping complex. This then opens the transcripts up to degradation from the 5’ end by ribonucleases that degrade RNA in the 5’ to 3’ direction, such as XRN1/Pacman.

In addition to the conventional RNA decay pathways, RNA can also undergo internal cleavage by endoribonucleases, generating fragments with exposed ends, allowing degradation by the previously described exoribonucleases. Internal cleavage often occurs as part of a quality-control process such as nonsense mediated decay (NMD) ensuring error-containing RNA do not undergo significant translation. Internal cleavage often also occurs with the binding of high-complimentarity miRNAs or siRNAs. The canonical RNA degradation pathways tend to have been studied in the context of mRNAs, but with the muddied waters of categories that are arbitrarily defined (such as lncRNAs), these pathways could be more accurately said to be relevant to RNAs featuring the canonical mRNA structures (5’ m7G cap and 3’ poly(A) tail). miRNAs, which lack this structure, rely on alternative mechanisms to avoid unnecessary decay (53).

1.4.6.1 An overview of 3’ to 5’ RNA decay:

1.4.6.1.1 Deadenylation:

It is widely accepted that the first, and rate-limited, step in 3’ to 5’ decay of mRNAs and canonical RNAs is deadenylation, or the removal of the poly(A) tail. Deadenylase complexes are recruited to the RNAs requiring degradation, and the protective PABP must be displaced and occluded. The stability of a canonical RNA is determined (in part) by the length of the poly(A) tail (54), as determined by their maturation to precise lengths by the Pab1p-dependent poly(A) nuclease (PAN) after their initial synthesis (55). Shorter tails provide a reduction in available binding sites for protective proteins such as PABP. It is worth noting that the length of the poly(A) tail is also crucial to translocation of mature mRNAs to the cytoplasm (56) and is coupled to translational efficiency (when within a certain length range) (57).

Prior to proceeding towards degradation, mRNAs and canonical RNAs in the cytoplasm feature poly(A) tails that fluctuate in size (within an approximate range defined by the species, varying around 90 nucleotides in yeast to 200 in mammals (54, 56)). In order for degradation to occur, these tails are removed in a stepwise manner. First, the Pan2-Pan3 complex trim the tails by the deadenylase activity of Pan2, recruited and coordinated by Pan3 dimers (58), and RNA threading allowed by interactions between the Pan2-Pan3 complex and Pab1 promoters of a poly(A) RNP (59). Following this initial trimming, the highly processive Ccr4-Not complex carries out the remaining deadenylation. The Ccr4-Not complex is composed of nine-subunits and is conserved in both function and sequence throughout the entire eukaryotic kingdom (60-62). The nuclease subunits (Caf1 and CCr4 in both humans and *Drosophila (62, 63)*) are required to carry out the actual deadenylation, while the Not proteins provide a scaffold for the complex from which the deadenylases can act.

There is some level of redundancy between the two, as demonstrated by the capability of Pan2-Pan3 to partially compensate for the loss of the Ccr4-Not complex (52). Although these complexes are the only ones known to act in *Drosophila*, and are well conserved throughout eukaryotes (61), there are other factors that act at this step in other organisms, such as PARN in humans (64), while *Arabidopsis thaliana* have highly diversified deadenylases, having as many as 26 (65). Once these deadenylations have taken place, the 3' end of the transcript is exposed and vulnerable to attack and degradation by 3' to 5' acting exoribonucleases. Deadenylation also stimulates decapping, which allows 5' to 3' degradation to occur, but this will be covered in a later section.

1.4.6.1.2 Degradation by 3' to 5' ribonucleases:

The Dis3 family, although crucial, are not alone in carrying out 3' to 5' degradation. Rrp6, another exoribonuclease acting in the 3' to 5' direction, also plays a crucial role. Rrp6 is a catalytic component of the exosome complex, which participates in a variety of RNA processing and degradation events. It functions in the nucleus to allow proper maturation of snRNAs, snoRNAs, and rRNAs (including the processing of the 5.8S rRNA). It also plays a role in decay of RNA processing byproducts, pervasive non-coding transcripts in need of degradation such as cryptic unstable transcripts (CUTs), and mRNAs with processing errors. Rrp6 also associates with the barrel like exosome core, towards the top of the 9-subunit complex.

The exoribonucleases discussed are aided and regulated in their activity by important cofactors. In the nucleus, the exosome requires the targeting cofactor TRAMP. The TRAMP complex is made up of a distributive poly(A) polymerase named Trf4/5, RNA binding protein Air1/2, and Mtr4 helicase. The addition of adenine residues by Trf4/5 and unwinding of structure by Mtr4 allows degradation of targets, including CUTs, rRNAs, and snoRNA precursors. In the cytoplasm, meanwhile, the Ski complex, comprised of Ski3, Ski8, and the helicase Ski2 (Twister in *Drosophila*) is required for activity of the exosome. This complex associates with the top of the cytoplasmic exosome, together with another cofactor, called Ski7. The Ski complex then unwinds structured RNA, aiding its channeling into the central channel of the exosome, for degradation.

1.4.6.1.3 Other factors:

As well as the mechanisms regulating RNA stability already covered in this section, there are a multitude of other factors that don't easily fall into these categories. For instance, some RNA transcripts in flaviviruses generate a small structured non-coding RNA from the incomplete degradation of their 3' UTR. These subgenomic flaviviral RNAs (sfRNAs) promote stabilization of RNAs beneficial to viral pathogenesis by suppressing RNAi in the host cells (similarly to other viral RNAi suppressors such as 1A and VP3 in *Drosophila* C and X viruses respectively).

As well as nucleic acids able to regulate RNA decay, some proteins exert their influence over RNA degradation in an equivalent fashion; by blocking the RNA degradation machinery from acting upon transcripts. The previously mentioned PABP is a well-studied example, but it is far from the only RNA-binding protein able to exert an effect on RNA stability, with dozens of such proteins identified and validated in *S. pombe*.
It would be beyond the scope of this thesis to catalogue all factors that play a role in the careful fine-tuning of RNA degradation, but it is easy to see that they are many, varied in how they act, and interplay with many other such factors.

1.4.6.2 An overview of 5' to 3' RNA decay:

1.4.6.2.1 Decapping:

Following the removal of the poly(A) tail that allows 3' to 5' degradation to proceed, a transcript can also be decapped, to allow degradation in a 5' to 3' direction. This process is deandenylation dependent and requires the previously mentioned PABP being displaced following the removal of the poly(A) tail by Ccr4-Not. Once the poly(A) tail is reduced to <10 nucleotides in length, it becomes a site for the Pat1/Lsm1-7 complex to associate. Pat1/Lsm1-7 binding confers protection from 3’ to 5’ degradation, while simultaneously promoting the decapping that leads to 5’ to 3’ degradation, essentially acting to divert a transcript into that pathway of RNA decay.

The promotion of decapping is achieved by recruitment of decapping machinery to the 5’ end of the target transcript. Dcp2 provides the required catalytic activity, stimulated by a range of other factors such as Dcp1, which associates with Dcp2 to induce a conformational change in Dcp2 to a closed, active conformation. This shift is required for the decapping process to proceed efficiently. Other factors, such as Edc1-4 are known to facilitate the process, with Edc3 known to physically bind Dcp1-Dcp2 to levy its effect.

Decapping may also be stimulated by alternative means from those deadenylation-dependent mechanisms discussed above. For example, GW182 (as part of the RISC) recruits Dcp1-Dcp2 to decap transcripts in some instances of miRNA mediated silencing. In another example, addition of 3’ uridines to the poly(A) tail of a transcript by *cid1* (a nucleotide transferase) causes recruitment of the Lsm1-7 complex, which promotes decapping, and subsequent degradation. This 3’ uridylation induced decapping and 5’ to 3’ degradation was initially discovered in *S. pombe* but has since been found to be conserved in mammalian cells.

1.4.6.2.2 Degradation by 5’ to 3’ ribonucleases:

After undergoing this decapping process, a transcript is left vulnerable to attack from the 5’ end. In the cytoplasm, this degradation is carried out exclusively by the 5’ to 3’ acting exoribonuclease XRN1, known as Pacman in *Drosophila*. XRN1/Pacman has a flexible, unstructured C-terminal region that allows it to bind to decapping factors, facilitating co-recruitment, and improved efficiency of degradation. This was initially shown for Pacman in *Drosophila*, and although the binding factor varies between species, the factor binding and co-recruitment is conserved across multiple species.

The catalytic N-terminal domain of Pacman in *Drosophila* has been shown to maintain a great deal of amino acid sequence homology with human XRN1 (insert diagram from Chris' 2012 paper). A high degree of conservation is not surprising given its key role as the enzyme responsible for 5' to 3' degradation of cytoplasmic RNA transcripts. XRN1/Pacman is even able to degrade relatively structured transcripts, due to its ability to unwind secondary structure. As well as its crucial role in degrading decapped transcripts, XRN1/Pacman is vital in clearing 3' RNA fragments from endonucleolytically cleaved transcripts, such as from transcripts degraded by quality control pathways, and by RISC cleavage following miRNA targeting.

Pacman is known to be crucial in *Drosophila* development, with absence of Pacman being lethal to the developing larvae, and reduced activity in hypomorphic mutants producing delayed development and reduced size, due to increased apoptosis. In yeast, *XRN1* null mutants are viable, but show substantial accumulation of deadenylated and decapped transcripts, illustrating that even if not required for cell viability, XRN1 clearly maintains its important role in RNA degradation and clearance.

All known components for 5' to 3' degradation, from the deadenylases to XRN1/Pacman, have been found localized in processing bodies (P-bodies), and indeed deadenylation is necessary both for decapping of transcripts, and for P-body formation. P-bodies are foci of translationally repressed mRNAs, as well as proteins related to RNA decay. Initially, P-bodies were hypothesized to act as sites as mRNA decay, although it has since been demonstrated that macroscopically observable P-bodies are not required for mRNA decay to occur, and that mRNA decay can occur in mutants deficient in P-body assembly. An alternative model has emerged, which suggests that P-bodies might act as storage sites for translationally repressed RNAs and RNA decay enzymes. The stored transcripts and enzymes might then either be degraded or allowed re-entry when needed (although the means by which the activity of the stored enzymes would necessarily be inhibited is not established. XRN1/Pacman is highly enriched in P-bodies, while 3' to 5' nucleases are much more diffusely distributed throughout the cell.

1.4.6.3 Nonsense-mediated decay:

Nonsense-mediated decay (NMD) is an RNA decay pathway that specifically targets mRNAs that contain premature termination codons (PTCs) as different from normal termination codons. The mechanism by which NMDs are able to recognize PTCs is known to vary across species. In mammals, NMD is tightly coupled with the splicing of pre-mRNAs during their maturation, with a PTC 50-55 nucleotides upstream of an mRNAs final exon-exon junction allowing efficient degradation through the NMD pathway, mediated by the crucial exon-junction complex (EJC). Conversely, in *Drosophila* the components of the EJC are not necessary for NMD, and PTCs are able to be defined independently of exon boundaries. However, the *Drosophila* orthologs of UPF1-3, SMG1, SMG5, and SMG6 are required for degradation of PTC-containing mRNAs. The importance of alternative splicing, and XRN1/Pacman are also discussed in 1.3 and 1.4.5.1.2, and the possibility of NMD being linked to translation is discussed in 1.4.7.1.

1.4.7 Translation:

1.4.7.1 Linking translation and degradation:

RNA degradation and translation both have vital roles in post-translational gene expression. Although it has often been assumed that RNA undergoing translation is protected from decay, researchers have speculated on the possibility (and potential importance) of association between exoribonucleases and ribosomes since the 1990s (66, 67). The idea of co-translational degradation is not a new one, and extensive work has accrued, showing it to be a crucial nexus of different regulators of gene expression. Sequential clusters of translating ribosomes, known as polyribosomes, or polysomes, are found on regions of RNA undergoing significant levels of translation. Several existing studies have separated polysomes from lysed cells and tissues and used molecular techniques (such as ribosome pulldown and ribosome and polysome fractionation) in order to examine associated proteins and nucleic acids (68-71).

Exoribonucleases XRN1 and Dis3L2 have both been shown to be independently associated with the ribosome in humans(71). In *Drosophila*, Pacman (the XRN1 equivalent) has also been shown to be ribosome associated(69), although the details of any co-translational degradation have not been elucidated. In *S. cerevisiae*, decapping and 5’ to 3’ RNA degradation have both been shown to occur during active translation(70). It has also been noted that the three main steps of 5’ RNA decay (deadenylation, decapping, and 5’ degradation) all provide extra layers of translational regulation. These findings clash with the previous idea that sequestering RNA away from polysomes, such as by P-bodies, is not necessary for decapping and translation.

Due to ribosomes working in a 5’ to 3’ direction, the 5’ to 3’ exoribonucleases such as XRN1/Pacman seem the more obvious target for exploring co-translational degradation. Despite this, there is sufficient evidence of translation having links to 3’ to 5’ degradation and NMD (71, 72) to justify exploration of these exoribonucleases as well.

1.4.7.2 Open reading frames:

1.4.7.2.1 Canonical ORFs:

Peptides and proteins make up a vital class of molecules, essential to the existence of all biological organisms. Translation of RNA transcripts by ribosomes is responsible for the synthesis of proteins, by building a chain of amino acids corresponding to the degenerate triplet code that is read from the RNA. As previously discussed, translation is a highly regulated process, with many factors contributing to the efficiency of initiation and elongation, as well as certain structures leading to premature termination and subsequent NMD.

Canonical ORFs were described by the necessarily limited parameters included to reduce false positives in the identification of protein-coding ORFs. These parameters include the requirement for the ATG start codon, an ORF of at least 100 codons, and a predicted single ORF per transcript. The cut-off length was decided upon in an attempt to distinguish genuine protein-coding ORFs from spurious in-frame start and stop codons within genomes. As a result of this, several potentially viable ORFs have been left un-annotated. Analysis (both bioinformatic and experimental that have removed this somewhat arbitrary limit have identified the potential for a great many more potential protein-coding ORFs.

1.4.7.2.2 Small open reading frames:

The descriptor “small open reading frame” or smORF was introduced in order to identify and categorise short ORFs of less than 100 codons that are actively translated. Several small proteins and peptides have been found to be encoded by smORF, referred to as smORF-encoded peptides (SEPs), meaning a protein product of less than 100 amino acids in length, arising from a smORF. Once potential non-canonical smORFs are taken into account, many lncRNAs contain regions that may in fact be protein coding.

With recent investigation into novel bioactive peptides and small proteins from non-canonical ORFs, hundreds of thousands of previously non-annotated smORFs have been discovered across plant, animal, and bacterial genomes. In fact, according to estimates, genes with a functional smORF could make up 5-10% of genomes. Although it is time and resource intensive to prove the protein-coding potential of these genes biologically, increasingly sophisticated bioinformatic techniques along with translationally relevant high-throughput data sets, streamline the search, highlighting the most likely targets for further investigation. As previously discussed, the canonical start sequence is the methionine encoding AUG, embedded in a Kozak sequence region (of varying strength. As evidenced by these smORFs and the SEPs they encode though, the definition of what can be considered a functional ORF must be expanded beyon this narrow definition, with several variant start codons having been shown to be able to initiate translation and produce small proteins and peptides.

This area of research, although promising, is still in its infancy. Techniques such as poly-ribo-seq allow experimental identification and biological proof of whether a non-canonical ORF translation but have only emerged in the last few years. Currently, a comprehensive investigation of a smORF and corresponding smORF peptides is a laborious task, requiring use of techniques which are only now starting to see widespread use.

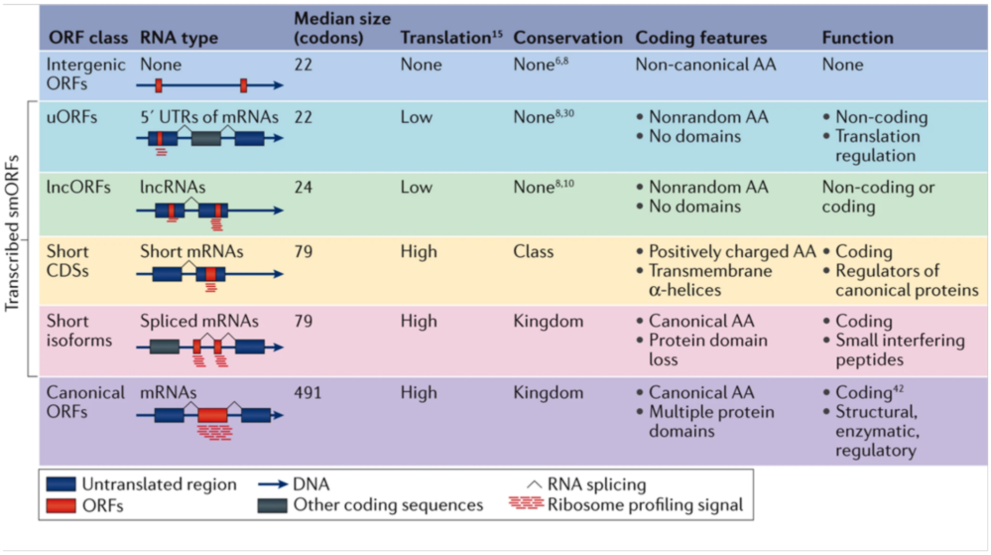
1.4.7.2.4 Upstream open reading frames:

Upstream open reading frames (uORFs) are ORFs identified within the 5’ UTR of an mRNA and are known to regulate gene expression in eukaryotes. The translation of a uORF tends to inhibit the expression of the “primary ORF” shortly downstream. A similar phenomenon is observed in bacteria, by the translation of leader peptides. uORFs are polymorphic and ubiquitous, causing widespread reduction of protein expression in humans; as many as 50% of human genes containing a uORF in their 5’ UTR.

1.4.7.2.4 Known and relevant smORF peptides:

The full catalogue of smORF peptides discovered so far would be far beyond the scope of this work to cover, for more extensive reading regarding the discovery of smORF peptides, an excellent review exists, by Saghatelian and Couso. Another excellent review by Couso and Patraquim provides a summary of smORF classifications (Figure 1.10). Here, an interesting, and particularly relevant example of smORF peptide discover (relevant to both human disease and *Drosophila*) will be discussed. The *Drosophila sarcolamban* (*scl*) gene, originally classified as a lncRNA *pncr003*, is transcribed into a 992 base-pair mRNA, which is
translated to produce two-related peptides of less than 30 amino acids. The *scl* gene is expressed in muscle cells, and *scl* null mutants show arrhythmic cardiac contractions, a phenotype produced by abnormal intracellular calcium levels in contracting muscle cells.

Interestingly, the *scl* genes were found to have homologues in humans,
namely *sarcolipin* (*sln*) and its longer paralogue, *phospholamban* (*pln*), encoding peptides of 31 and 52 amino acids, respectively (73). Phylogenetic analysis suggests that these genes belong to the same gene family, derived from a single ancestral gene, conserved for more than 550 million years. Furthermore, their function also seems to be conserved, with Sln and Pln regulating calcium transport in mammalian muscle cells, via dampening of sarco-endoplasmic reticulum Ca^{2+} adenosine triphosphate (SERCA) pump function. Scl peptides were capable of colocalising and interacting
with *Drosophila* SERCA. Exogenous expression of the human Pln and Sln peptides
in *Drosophila scl* mutant muscle cells were sufficient to rescue muscle function.



	ORF class	RNA type	Median size (codons)	Translation[15]	Conservation	Coding features	Function
	Intergenic ORFs	None	22	None	None[6,8]	Non-canonical AA	None
Transcribed smORFs	uORFs	5′ UTRs of mRNAs	22	Low	None[8,30]	• Nonrandom AA • No domains	• Non-coding • Translation regulation
	lncORFs	lncRNAs	24	Low	None[8,10]	• Nonrandom AA • No domains	Non-coding or coding
	Short CDSs	Short mRNAs	79	High	Class	• Positively charged AA • Transmembrane α-helices	• Coding • Regulators of canonical proteins
	Short isoforms	Spliced mRNAs	79	High	Kingdom	• Canonical AA • Protein domain loss	• Coding • Small interfering peptides
	Canonical ORFs	mRNAs	491	High	Kingdom	• Canonical AA • Multiple protein domains	• Coding[42] • Structural, enzymatic, regulatory

Figure 1.10 – Adapted from Couso and Patraquim (2017) – Diagram demonstratingtypes and features of ORFs

Most open reading frames (ORFs) in animal genomes are small ORFs (smORFs) in untranscribed regions (intergenic ORFs; light blue) and are considered non-functional. ORFs of 101 codons or more (canonical ORFs; purple) are generally translated and produce annotated proteins with predictable functions. In between these two extremes, our genome also encodes transcribed and putatively functional smORFs of 10–100 codons, which can be divided into different classes according to the type of their encoding transcript: upstream ORFs (uORFs; teal) in the 5′ untranslated regions (5′ UTRs) of canonical mRNAs; long non-coding ORFs (lncORFs), which are present in long non-coding RNAs (green); short coding sequences (short CDSs), which are annotated ORFs of 100 codons or fewer that are present in short mRNAs (yellow); and short isoform ORFs of 100 codons or fewer, which are generated by alternative splicing of canonical mRNAs (pink).

Importantly, aberrant levels of Sln in humans have been linked to heart arrhythmias (74). Regulation of SERCA by micropeptides (encoded by lncRNAs) has been extensively exploited in mammals, with tissue specific positive and negative regulators being found (75-77). In addition, the number of characterised lncRNA genes encoding micropeptides is rapidly increasing, with roles found in essential, conserved cellular functions, from phagocytosis (78) and cellular motility (79) to RNA degradation (80). Thus, these examples show that lncRNAs that produce biologically relevant peptides may be conserved in structure, function, and relevance to pathologies between humans and *Drosophila* (77, 80).

1.4.7.3 The process and regulation of translation:

Ribosomes are recruited to RNA, mostly by 5' cap-dependent recruitment of the 43S ribosomal complex by eIF4F initiation complex (containing eIF4G acting as a scaffold, eIF4A unwinding secondary structure, and the cap binding eIF4E), as well as poly(A) binding protein (PABP). They must gain access to a viable open reading frame (ORF) in order to begin translation of an RNA transcript. Certain miRNAs may interfere with eIF4E binding. After miRNA binding to a complimentary site in the untranslated region (UTR), these miRNAs are theorized to interact with both the 5' m7G cap and the Argonaute 2 protein of the RNA-induced silencing complex (RISC), and cause disruption to assembly of the complex (81-83).

The pre-initiation complex scans along the transcript until it reaches a suitable start codon, which triggers binding of the 60S ribosomal subunit, and dissociation of unnecessary initiation factors. This is typically the canonical methionine encoding AUG sequence, embedded in the Kozak sequence (with deviation from the prime Kozak sequence tending to reduce the "strength" of initiation (84, 85)). However, significant evidence is now available showing that the definition of what can be a functional ORF has historically been kept too narrow (86, 87), and that several variant start codons are still capable of translation initiation to a meaningful degree. As described, every step of translation initiation is regulated, ensuring a tight leash is kept on the level of proteins and peptides produced.

Further than just the capability of the RNA sequence to interact with RNA, in some instances, such as in the pre-fertilisation oocyte, mRNAs are prevented from translating by their protected storage in the cytoplasm of the oocyte. In *Drosophila*, normal cell division in the early oocyte is maintained by translation of maternal RNAs, with oocyte RNAs kept from translation until the developing embryo undergoes maternal to zygotic transition (MZT), at which point the regulatory tide is turned. Expression of maternal RNAs is repressed, and existing transcripts are degraded by context sensitive regulatory factors (32) (such as ME31B, TRAL, and Cup), allowing the necessary shift to the zygotic proteome.

Another instance of translational control, again present in *Drosophila*, is by the binding of mRNA by inhibitory proteins such as Smaug, which binds the 3' UTR of the RNA *nanos*, preventing production of the Nanos protein. Synthesis of mRNAs without their 5' methylated caps (seen in the oocyte of the tobacco hornworm moth), can ensure that they are essentially untranslatable until further modification by a methyltransferase. Alternatively, regulation of poly(A) tail length can regulate translatability as well as stability (31), as with *bicoid* mRNA.

1.4.7.4 miRNA regulation of translation:

Translation is primarily inhibited by miRISC disruption of translation initiation. The miRISC interferes with eukaryotic initiation factor 4 A-I (eIF4A-I), and eukaryotic initiation factor 4 A-II (eIF4A-II) in their interaction with target mRNAs by causing their dissociation (47), which subsequently prevents ribosome scanning and assembly of the (previously discussed) eIF4F translation initiation complex.

The increase in degradation is implemented through recruitment of GW182 by AGO (48, 49), which interacts with polyadenylate-binding protein (PABPC). This subsequently promotes deadenylation of mRNAs by recruiting the complex of poly(A)-nuclease 2 and 3 (PAN2-PAN3), as well as the complex of carbon catabolite repressor protein 4 with NOT (CCR4-NOT). Deadenylation by these promotes decapping by the complex of mRNA-decapping enzymes 1 and 2 (DCP1-DCP2), thereby making the mRNA susceptible to rapid degradation (as previously discussed. The recruitment of CCR4-NOT provides an additional means of translational repression through the recruitment of probable ATP-dependent RNA helicase (DDX6).

miRNAs regulate genes in a network, with a single miRNA able to regulate hundreds of genes (50, 88), while multiple regulatory miRNAs may be able to act on a single gene (51). Individual miRNAs and miRNA clusters are able to act on entire cellular pathways, and are able to completely shut off some genes, and fine tune the expression of others (89).

Overall, then, we can clearly see that even once transcription has occurred, a plethora of regulatory features fine-tune the final level of gene expression, whether as a nucleic acid, or a protein. The careful interplay of all these regulatory events is crucial to maintaining the functional complexity of the transcriptome and proteome necessary for any and all complex biological life.

1.5 *Drosophila* as a model organism:

1.5.1 Why use *Drosophila*:

1.5.1.1 General advantages of working on *Drosophila*:

Drosophila melanogaster, the common fruit fly, is a well-established model organism for geneticists, and one in which lncRNAs are known to be abundant. With a genome encoding just fewer than 18000 genes over four chromosomes, an estimated 75% of human disease-linked genes having a functional orthologue in *Drosophila*, and many basic molecular and biological functions conserved between species, *Drosophila* are an appealing whole animal model for understanding human disease. In addition to their genetic similarities with humans, the fly genome has been extensively studied and fully sequenced, being the first major complex organism to have its genome sequenced. Many keystone discoveries in genetics have been made in *Drosophila*; the concept of heritable traits being carried on the chromosome, the genetic control of early embryonic development, the discovery of homeotic genes, and subsequent founding of evolutionary developmental biology all being prominent examples.

Having been used as a model organism for so long, a wide range of genetic tools and gene-specific knockdown and mutant lines readily available in *Drosophila*. A very prominent example of this is the versatile and powerful UAS-GAL4 system. This system makes use of GAL4, an 881 amino acid protein, known to be a gene regulator, inducing transcription in the genes it regulates by binding to upstream activation sequences (UASs). In 1998, Fischer *et al*. showed GAL4 to be capable of stimulating transcription of a reporter gene in *Drosophila*. Following this in 1993, a bipartite system for using GAL4 and UASs to drive gene expression was devised.

In this system, the gene of interest is controlled by the presence of a UAS element, dependent on GAL4 to initiate transcription. In the absence of GAL4, the gene stays transcriptionally silent. In order to activate the target gene, the UAS flies can be mated with flies expressing GAL4 ubiquitously, in a particular pattern, or within particular tissues. The resulting progeny then express the gene of interest on a transcriptional pattern of choice, allowing tissue-specific effects to be observed, or expression of genes which might be lethal if expressed ubiquitously. It also allows stable, non-phenotypic parent lines to be easily maintained, with the target-gene expressing progeny able to be produced on relatively short notice. This system can even be combined with temperature-dependent expression of Gal4, in order to provide temporal control over expression of a target gene. Powerful tools like this (as well as a wide range of CRISPR mutants) have long made *Drosophila* a tempting choice of model organism for molecular biologists and geneticists.

A rapid life cycle (summarised in Figure 1.11) provides short generation time at a low cost. This, combined with high fecundity provides rapidly changeable and robust stocks, with a single mating pair able to produce hundreds of offspring within 10 to 12 days at 25°C. This compounds with other favourable factors, all lending themselves to ease of establishing genetic crosses. It is easy to see why *Drosophila* have emerged as one of the foremost systems for studying genetics, having already been successfully used to dissect the roles and

mechanisms of many key developmental pathways (such as *notch* and *wingless*), as well as certain lncRNAs (such as *sarcolamban*).

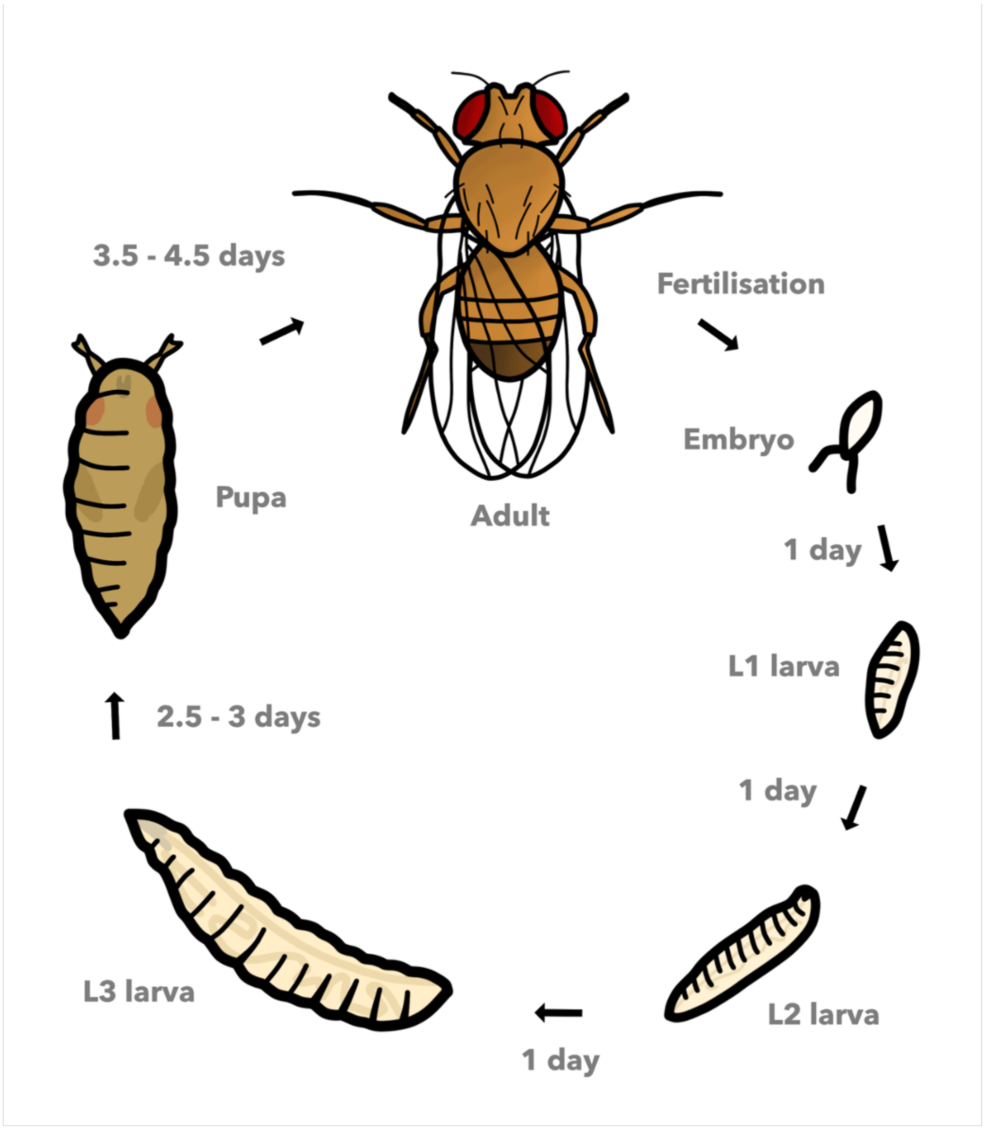


Figure 1.11 – A cartoon depicting the *Drosophila* life cycle

The Drosophila life cycle is divided into four stages: embryo, larva, pupa, and adult. The time length of the stages is approximate for stocks kept at 25°C and is shown in hours for embryos and days for larvae and pupae.

The adult fly is a complex organism with discrete tissues and structures analogous to the heart, lung, kidney, gut, reproductive system, and nervous system. The *Drosophila* brain contains over 100000 neurons and has been mapped in unbelievable detail. The entire hemibrain of an adult female fruit fly has been mapped, with the production of a 3D map of the wiring of 25000 neurons, including functional areas for learning, memory, smell, and navigation. This is particularly interesting as *Drosophila* have already been used in the discovery of several prominent drugs, and the CNS response of fruit flies to drugs is known to be similar to that seen in mammalian systems.

1.5.1.2 *Drosophila* as a model for relevant disease:

As previously discussed, 75% of human disease-causing genes have a functional homologue in *Drosophila*. The physiology of an adult fly provides tissues and structures analogous to the heart, lung, kidney, gut, reproductive system, and nervous system. There is an extensive history of the fruit fly being used in biomedical research with multiple novel therapeutic drugs having been discovered using *Drosophila* screening, and multiple Nobel prizes having been won for work using fruit flies. Particularly relevant to this project, and previously discussed throughout this introduction, *Drosophila* mirrors pathological phenotypes seen with erroneous activity of exoribonucleases, as well as phenotypes seen with mutations in certain lncRNAs.

As described in an earlier section, several lncRNA mutations are known to be either associated with, or causative of diseases in humans. Despite the the difficulties in identifying homogues of lncRNAs, and conservation being lower than in canonical RNAs, still several examples of disease-causing (or contributing) lncRNAs that have arisen that are common between humans and *Drosophila*. In summary, *Drosophila* are known as not just a strong model for biomedical and genetic research, but have evidence supporting them as directly relevant to the specific topics that this thesis aims to explore.

1.5.2 *Drosophila* life cycle and development:

The life cycle of *Drosophila melanogaster* has been extensively and thoroughly characterized, all the way from fertilized egg, through development and adult life, until death. The time taken for full development to an adult fly is heavily dependent on temperature. At 19°C, development from fertilised egg to adult fly takes around 18 days. This reduces to 10 days at 25°C, and further decreases to around 7 days at 28°C. As well as timeline shortening, the size of the adult fly decreases as temperature increases. Further discussion of *Drosophila* development is for flies cultured at 25°C.

Upon egg fertilisation, embryogenesis takes only 24 hours, with developmental axes being determined by maternal RNAs during this period. The anterior-posterior (AP) axis is governed by the genes *Bicoid* and *Hunchback* (anterior) as well as *Nanos* and *Caudal* (posterior). *Bicoid* and *Nanos*, both morphogenic genes, establish the AP concentration gradient by abundance of their transcripts, and subsequent protein products. Nanos, with its increased abundance in the posterior represses the expression of Hunchback protein, while it is expressed at higher levels (due to lack of repression) towards the anterior of the embryo. Meanwhile Bicoid blocks translation of *caudal*, so that Caudal is at lower concentration at the anterior part of the embryo, and higher towards the posterior.

These boundaries are initially set up by maternally provided transcripts, but further axis specification takes place during the mid-blastula transition. At this stage, MZT occurs, and the regulatory tide is turned from maternal transcripts to zygotisc. Maternal transcripts are degraded by context sensitive regulatory factors like ME31B, TRAL, and Cup, while the zygotic RNAs previously repressed and kept in protective cytoplasmic storage are allowed to undergo translation. Following 24 hours of embryogenesis, the 1st instar larvae (L1) hatch, remaining on the surface of the food until their first moult 24 hours later, reaching 2nd instar larval stage (L2). After a further 24 hours and another moult, they reach 3rd instar larval stage (L3). This stage is characterized by a longer (48 hour) stage spent buried in the surface of the food, until a "wandering" stage during late L3. Larvae at this stage were used for all larval RNA extractions and disc dissection. These larval stages are vital to reaching full development as an adult, as they allow for rapid growth of precursor tissues (such as WIDs), to a sufficient size to reach the pupal stage, where they enter metapmorphosis, to become the adult fly.

This pupal stage is reached 48 hours after egg fertilization and can only proceed once a controlled "checkpoint" has been cleared. This checkpoint is controlled primarily by hormones such as Ecdysone, Juvenile Hormone (JH), and insulin-like peptides. Ecdysone signaling is required for the fly to undergo pupariation and can be repressed by JH. Secretion of JH is suppressed in correctly developed larvae that are ready for metamorphosis, lifting the block on Ecdysone, and allowing the larvae to proceed into pupariation. Other hormones feed into this regulatory step, as can be seen with increased expression of *Drosophila* Insulin-like peptide 8 (Dilp8), which delays pupariation by 2-3 days.
During the metamorphosis that occurs during pupal stage, Imaginal discs change in shape, size, and location, differentiating into the tissues and physiology required for formation of the adult fly. Unnecessary larval tissues are degraded through controlled apoptosis, and after approximately 84 hours, adult flies eclose, reaching sexually capable maturity a further 12-14 hours after this.

1.5.3 L3 larvae as a model system:

As previously mentioned, the L3 stage of *Drosophila* development is a crucial one, for several reasons. The rapid growth at this stage is responsible for the development of the tissues required to metamorphise into the adult fly, so both growth and regulation are actively playing roles in determination of the adult organism. The WIDs, for example grow from 50 cells to 50000 cells during the larval stages. Additionally, this stage is the first "openly active" stage in *Drosophila* development, the larvae emerge from the food, and proceed to wander, climbing and exploring the immediate nearby environment. This not only allows for behavioural assays (not used in this thesis), but allows relatively easy collection of many L3 larvae, crucial to certain techniques (such as poly-ribo-seq).

At this stage in development, the larvae can also be dissected and separated into discrete precursor tissues of known fates. Using fine tweezers and a microscope, imaginal discs, ovaries, testes, brain, and other tissues can be gathered, allowing analysis of specific areas of interest. Of note, several studies relevant to this project have datasets available from experiments in L3 and L3 tissues. Examination of L3 larvae allows comparison with a great deal of existing work.

Wing imaginal discs were used extensively in previous work by the Newbury lab, and used in some of the preliminary stages of this project. *Drosophila* WIDs have been used in a number of important discoveries, contributing to research on growth, gene regulation, tissue regeneration, and cancer. Imaginal discs have provided such a useful tissue, primarily due to relatively easy access by dissection, and the ability to use specific drivers with tools like the UAS-GAL4 system to affect specifically the WID. In addition, WID cells share significant similarities with human epithelial cells, providing a useful platform for biomedical research.

1.5.4 S2 cells in a model system:

There is enormous and apparent value in carrying out experiments in live, complex model organisms, however complimentary work in cell culture allows experimentation and advancement that would otherwise be impossible to be carried out alongside true in-vivo work. The Schneider 2 (S2) cell line was derived from a primary culture of late stage (20-24 hour) *Drosophila melanogaster* embryos. Certain features that the cell line possesses suggests that it is derived from a macrophage-like cell lineage. S2 cells can be grown without CO_2 between 22°C and 28°C, with cell growth slowing at cooler temperatures. The standard, complete medium for S2 cell growth, is Schneider's *Drosophila* Medium, supplemented with 10% heat-inactivated Foetal Bovine Serum. This is often supplemented with Penicillin-Streptomycin in order to inhibit growth of bacterial contaminants.

S2 cells are semi-adherent, forming a monolayer in culture that can easily be dislodged by pipetting. They grow rapidly and are fairly robust, prospering at between 0.5 x 10^6 cells/mL and 5 x 10^6 cells/mL, although they can recover from being over-split or over-crowded outside of this density range, and continue to grow healthily in culture with appropriate treatment. As a cell line, they have been used extensively, with RNAi knockdown, transfections, and CRISPR-Cas9 mutation all having already been carried out in S2 cells. With such a well-used, fast-growing, and robust cell line, S2 cells are a fine model for generating large amounts of *Drosophila* genetic material and proteins, and very helpful as an exploratory tool to use alongside whole fly, or larval, work.

1.5.5 lncRNAs in *Drosophila*:

As well as the general excellence of *Drosophila* as a model organism, they stand out as particularly apt for the study of lncRNA. lncRNAs evolve rapidly and can act as flexible scaffolds tethering together one or more functional elements. *Drosophila* lncRNAs also appear to accumulate relatively few deleterious changes due to genetic drift, compared with mammalian lncRNAs, and therefore can be useful in developing strategies to identify lncRNA orthologues, as shown for *roX* (RNA on the X) lncRNA orthologues in *Drosophilid* species. Additionally, *Drosophila* is an excellent model system to functionally characterise lncRNA–protein complexes, for example by using the GAL4–UAS system to express lncRNAs in specific tissues or by characterising the localisation of RNA–proteins within cells (e.g. 7SK snRNA).

According to the Ensembl database, lncRNAs comprise 7841 of the 63 898 annotated genes in the human genome, and 2366 of the 17 559 in the *Drosophila* genome. In both species, they account for a similar and substantial proportion of the entire genome (12.4 and 13.5%, respectively). Although only a fraction of these have been investigated experimentally, information on their sequences and loci are readily available through various genomic databases, both non-specific (such as Ensembl) and dedicated non-coding RNA databases (such as LNCipedia, lncRNome, and lncRNAdb). Additionally, significant bioinformatic work has been carried out on them in terms of their expression and conservation within and across species. With so much information on lncRNAs now available, exploring this class of genes with a thorough experimental approach has become more feasible in recent years.

lncRNAs vary significantly in their distribution throughout cellular compartments, with the majority of transcripts residing predominantly in the nucleus, others in the cytoplasm, and some distributed more evenly between the two. For example, the *roX* transcripts in *Drosophila* are found in the nucleus, whereas *yar* is cytoplasmic. The localisation of lncRNAs can give clues about their function; in the case of a chromatin restructuring lncRNA such as *roX1* or *roX2*, it must be nuclear in order to access the chromatin. Localisation of particular lncRNAs can also affect their susceptibility to suppression by RNA interference and antisense oligonucleotides. An example of this is the suppression of nuclear lncRNAs *MALAT1* and *NEAT1*, which in humans is more efficient using antisense methods, whereas cytoplasmic lncRNAs *DANCR* and *OIP5-AS1* are better suppressed with RNAi methods.

However, the sub-cellular localisation of the majority of lncRNAs has not been well characterised, with the localisation of relatively few being experimentally visualised. Single molecule RNA fluorescence *in situ* hybridisation has now been used to give high resolution data for the distribution of lncRNAs in human cells, and a systematic investigation of lncRNA localisation has been suggested as an important next step in expanding our understanding of their function, as well as a useful way to shed light on the potential relevance of lncRNAs to a particular mechanism.

lncRNAs have been shown to function via a wide range of molecular mechanisms (including in *Drosophila*), falling under the broad categories of signals, molecular decoys, guide RNAs, or scaffolds (44). Some lncRNAs have convincingly been shown to be translated, with the small peptide products (smORFs) having important biological functions (90). Through these various mechanisms (Figure 1.12), they have been implicated in regulation of a diverse array of processes, such as differentiation, development, cell proliferation, nervous system function, and cardiovascular function in both *Drosophila* and humans, despite the lack of sequence conservation in lncRNAs across species. Importantly, similarities in the modes of action of lncRNAs have also been found at the molecular level between organisms, some of which have been discussed in a previous section.

Molecular functions and mechanisms of lncRNAs, such as their binding to protein complexes, definitively need to be tested *in vivo* in order to be well characterised. For example, *in vivo* experiments have shown that only the lncRNA transcribed in the reverse direction from the Polycomb/Trithorax response elements can bind the polycomb
repressive complex 2 (PRC2) component Enhancer of Zeste, which provides the critical histone methyl transferase activity required for transcriptional silencing. This level of understanding of such complex mechanisms and interactions would be extremely

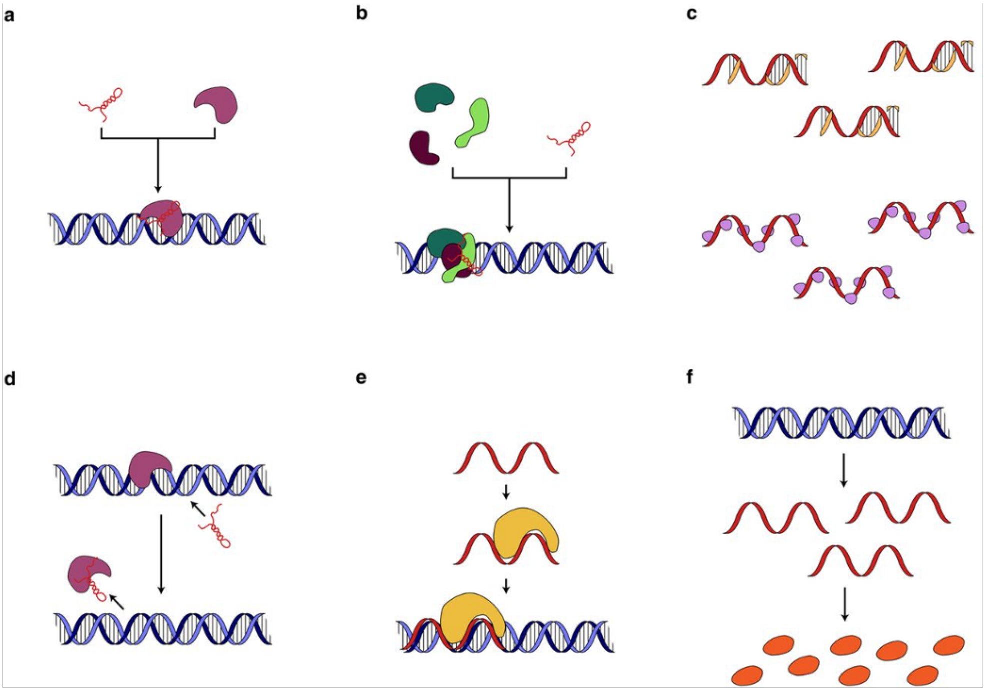


Figure 1.12 – A cartoon depicting the molecular mechanisms by which lncRNAs can function

(a) Some lncRNAs (red), such as Xist and, can act to modulate expression ofRoX1 a certain gene by binding to a transcription modifier or chromatin modifier (purple). (b) lncRNAs (red) such as HOTAIR can act as molecular scaffolds, allowing the assembly of protein complexes (teal, green, and dark purple) with genetic regulatory roles, e.g. polycomb complex PRC2. (c) lncRNAs (red) can act as molecular decoys, to sequester miRNAs (orange) or proteins (purple). (d) Alternatively, lncRNAs (red) can act as molecular decoys, occluding or removing transcription factors, proteins, or RNAs (purple) from their functional location. (e) lncRNAs (red) can act as a molecular guide, allowing formation of ribonucleoprotein complexes (yellow) to specific target sites. (f) It has also been shown that lncRNAs (blue as DNA and red as RNA) can be actively translated into functional smORF peptides (orange) such as the SclA and SclB peptides, which function in regulating calcium transport in cardiac muscle.

difficult to achieve without the use of a tractable *in vivo* system such as that provided by *Drosophila*.

It seems likely that the studies currently being carried out on lncRNA in *Drosophila* should be of interest to a far wider audience than just fly geneticists, with extensive work having shown that as a model organism, *Drosophila* is a logical choice both for better characterising this class of gene, and for precursor studies to highlight genes and mechanisms that can be carried forward into more expensive and laborious large animal and human work. The superb annotation of the *Drosophila* genome and transcriptome, coupled with constant increases in RNA-sequencing data available, will no doubt provide a candidate pool of lncRNAs, streamlining functional characterisation (with the bonus of being able to capitalise on the sophisticated genetic tools available in *Drosophila*). Therefore, lncRNA studies in *Drosophila* are likely to provide us not only with a better understanding of the basic science behind this gene class, but also promise to highlight potential biomarkers, elucidate genetic mechanisms behind a range of diseases, and perhaps in the long term, lead to therapeutic innovations.

1.6 Comprehensive overview of long non-coding RNAs:

1.6.1 Overview of lncRNAs

In the 1990s, several studies began investigating the biological purpose of longer non protein-coding RNAs, such as *Xist*, which did not fit well into the RNA classifications existing at the time. With further advances in molecular techniques suggesting that only 2% of the human genome consists of protein-coding genes, and rapidly revealing long non-coding RNA transcripts (lncRNAs) with biological functions (including in human diseases), the topic has become an extremely promising and popular avenue of investigation.

lncRNAs are highly abundant and are found in many organisms across different taxa, including humans, mice, *Xenopus tropicalis*, *Drosophila melanogaster*, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Medicago truncatula*, and *Zea mays* (91). lncRNAs have been shown to regulate gene expression transcriptionally (92-95) and post-transcriptionally

(96-100), and have a wide range of cellular and molecular functions (Figure 1.12). Despite these proven non-coding functions, there exist a handful of lncRNAs that have been shown to encode small open reading frame (smORF) peptides with proven cellular functions (73, 78, 80, 101-103). Recent work has shown that lncRNAs can simultaneously display biological function as both a coding and a non-coding RNA, for example, where primary transcripts of microRNAs encode regulatory peptides (104, 105). Additionally, ribosome profiling and bioinformatic analyses have identified the existence of thousands of lncRNAs containing putatively functional translated smORFs (73, 77, 106, 107), the extent of which may depend on developmental or tissue specific context.

lncRNAs have now been implicated as important factors linked to a range of human diseases. The broad range of biological functions of lncRNAs is reflected in the variety of different pathologies in which their aberrant expression is thought to be a contributing factor. Many lncRNAs have been shown to either be expressed at aberrant levels in cancerous cells, or their levels shown to affect the growth and behaviour of cancerous cells (Insert table from my review) (44). This has prompted speculation that if better characterised, this class of genes may present many promising biomarkers, and even novel potential therapeutic targets. This thesis cannot comprehensively cover this topic, and points the reader to a comprehensive review of the topic for more information (108), but instead demonstrate this point with two well-documented examples, below.

MALAT1, a highly conserved mammalian lncRNA, has been found to be overexpressed in human osteosarcoma cells and cell lines (109, 110). It is hypothesised to function as a molecular scaffold for ribonucleoprotein complexes, acting as a transcriptional regulator for certain genes. Higher levels of *MALAT1* have been shown to be associated with ‘aggressive’ cancer traits such as increased migration, metastasis, and clonogenic growth in non-small cell lung cancer (111-113) and pancreatic (114) and prostate cancer cells (115). Indeed, inducing a knockdown of *MALAT1* in osteosarcoma cell lines inhibited cell proliferation and invasion (109, 110).

The *HOTAIR* lncRNA, transcribed from an antisense Hox gene, plays an important role in the epigenetic regulation of genes thought to be due to its interactions with the PRC2 (116, 117), although recent work has indicated that PRC2 recruitment may be a downstream consequence of gene silencing, rather than initiating it (118). *HOTAIR* is thought to act as a molecular scaffold and is required for histone modification of particular genes across different chromosomes. Higher levels of *HOTAIR* have been found in colorectal cancer tissues and are associated with increased tumour invasion, metastasis, vascular invasion, advanced tumour stage, and a worse prognosis in patients (116, 119). *HOTAIR* has since been suggested for use as a biomarker for the progression and prognosis of certain cancers (119). A *Drosophila* homologue for *HOTAIR* has not been identified, but given the similarities in polycomb regulation between species, it is likely that a targeted search might reveal such an equivalent.

Aside from cancer, strong evidence now exists linking certain lncRNAs to certain neurological pathologies (120). lncRNAs have been shown to be relevant factors in amyotrophic lateral sclerosis, multiple sclerosis (121, 122), Alzheimer's disease (97, 123), Huntington's disease (124, 125), and Parkinson's disease, among others. For example, the *BACE1* antisense transcript (*BACE1-AS*) regulates mRNA stability of *BACE1*, a key enzyme in Alzheimer's disease pathology (97). This subsequently affects amyloid-β 1–42 abundance, the increased expression of which is a hallmark of Alzheimer's disease. One mechanism by which lncRNAs have been hypothesised to affect neurodegenerative disease is through their induction of R-loop formation (which may be triggered by trinucleotide repeat expansion). R-loops have been shown to be capable of controlling the fate of neuroprotective genes (126) and are thought to contribute to the pathogenesis of fragile X syndrome and Friedrich's Ataxia (127, 128) by their silencing of certain genes. Additionally, work in *S. pombe* and *Arabidopsis* has suggested that R-loops may regulate lncRNA expression (129, 130), although whether this is true of lncRNAs linked to neurodegenerative diseases remains unclear. Trinucleotide repeats in lncRNAs are also known to be important in the pathogenesis of SCA8, by production of toxic non-coding CUG expansion RNAs from the ataxin 8 opposite strand (*ATXN8OS*), thought to cause a toxic gain of function at both the RNA and protein level (131, 132).

Another area of disease in which lncRNAs have been proved relevant is cardiovascular disease (133, 134). Evidence now shows that lncRNAs are an important factor in susceptibility to coronary artery disease and myocardial infarction, prognosis in recovery from myocardial infarction, cardiovascular disease mortality, and heart failure (134). Once again, their correlations with prognosis and susceptibility have placed lncRNAs in the spotlight as a promising avenue of investigation in finding novel biomarkers.

1.6.2 Comparisons between *Drosophila* and human lncRNAs:

Interestingly, *Drosophila* lncRNAs have been shown to hold functional roles very relevant to the pathologies mentioned in section 1.6.1. *Hsromega* (135-139) and *bft* (140) are required for proper apoptosis process and cell differentiation, *yar* (141) and *CRG* (142) serve regulatory roles in the nervous system, and the previously mentioned *sclA* and *sclB* are required for normal calcium transients and cardiac muscle contractility (73). This is particularly promising given that these links can be made from the limited pool of *Drosophila* lncRNAs that have been experimentally characterised.

One of the most extensively studied molecular mechanisms of lncRNA modes of action, and an example of a lncRNA function with obvious parallels between species, is the role of lncRNAs in sex chromosome dosage compensation pathways. Owing to the difference in the number of X chromosome copies between males and females, there exists a compensation pathway required to maintain a similar level of expression for genes located on the X chromosome. In *Drosophila*, this is achieved by transcriptional hyperactivation of the single copy of the genes in males, allowing their expression at comparable levels to that given by the two copies of the gene found in females (143). In humans, by contrast, the genes located on the X chromosome in human females are partially transcriptionally repressed, giving a similar level of expression to that seen in males (144).

In *Drosophila*, the *RNA* on the *X* genes, *roX1* and *roX2*, are expressed in males and regulate the assembly of the male-specific lethal (MSL) complex in *Drosophila*; a chromatin modifier that functions in histone modification (145-148). The recruitment and binding of MSL proteins by high-affinity sequences on the nascent *roX* transcripts covering the X chromosome allows the assembly of the active MSL complex, which can then spread in *cis*, allowing chromatin restructuring and hyperactivation of specific regions of the chromosome.

An immediate comparison can be made between the *roX* genes in *Drosophila* and lncRNAs involved in the sex chromosome dosage compensation pathway in humans and other mammals; *X-inactive specific transcript* (*Xist*) and its antisense transcript, *Tsix*. Like the *roX* genes, *Xist* coats the X chromosome, where it regulates chromatin modifications, with consequent effects on the expression of particular target genes (149, 150). Unlike *roX*, *Xist* is expressed in females and regulates the inactivation of the X chromosome by facilitating the initiation and stabilising the X chromosome inactivation process (144). Although these lncRNA genes differ in their sequence, there are striking similarities between their role in specific regulation of the X chromosome and the molecular mechanisms by which they are thought to achieve this. Interestingly, a subset of lncRNAs involved in chromatin looping, called topological anchor point RNAs, have been identified in the human and mouse genomes, with conserved zinc-finger motifs capable of binding DNA and RNA (151). Whether these are conserved in *Drosophila* has not yet been studied, but given the involvement of lncRNAs in *Drosophila* chromatin regulation so far, this may be a promising avenue to explore and may reveal a wider conservation of this class of lncRNA chromatin regulators.

As well as these functions as nucleic acid molecules, an example of how lncRNAs have been shown to have roles in the production of small peptides is the *Drosophila sarcolamban* (*scl*) gene, previously mentioned in section 1.4.7.2.4. Originally classified as an lncRNA *pncr003* (152), it is transcribed into a 992 base-pair mRNA, which is translated to produce two-related peptides of less than 30 amino acids (73).
The *scl* gene is expressed in muscle cells, and *scl* null mutants show arrhythmic cardiac contractions, a phenotype produced by abnormal intracellular calcium levels in contracting muscle cells (73). As mentioned in 1.4.7.2.4 , *scl* genes were found to have homologues in humans encoding functionally homologous peptides.

1.7 Aims and Objectives:

1.7.1 Analyse and verify existing work to highlight lncRNA targets of Pacman and Dis3L2:

Re-analyse existing datasets as well as experimentally validate promising candidates in order to prove principles important to this project, and to take an initial look at the degradation of lncRNAs by Pacman and Dis3L2. Whilst a more comprehensive analysis of multiple datasets will follow once novel data has been gathered from the experiments later in this thesis, independent examination of the existing data aims to highlight targets and justify the rest of the project.

1.7.2 Establish a viable and useful model for exoribonuclease depleted *Drosophila* polysome work:

Having justified the principle behind the project, a viable model for carrying out poly-ribo-seq in exoribonuclease deficient *Drosophila* samples must be developed. This will draw from previous experiments, both in *Drosophila* S2 cells and whole *Drosophila* (33, 46, 69). Knockdown by dsRNA and null mutation by CRISPR-Cas9 will be tested and compared to the possibility of adapting polysome fractionation (and subsequent poly-ribo-seq) protocols to accommodate whole *Drosophila* tissues.

1.7.3 Optimise existing polysome fractionation protocols for the exoribonuclease depleted model:

Using whole *Drosophila* L3 in the delicate protocol for generating poly-ribo-seq libraries might present significant challenges, especially compared to the relative ease of protocol and clarity of results achievable when using a fast-growing, easily lysed, and low debris model such as S2 cells. In order to gain meaningful data from this protocol, significant optimisation, (particularly of the fractionation steps) must be carried out, and polysome traces compared to examples that are known to work (for example successful traces generated from S2 cells, subsequently used for poly-ribo-seq).

1.7.4 Carry out poly-ribo-seq on exoribonuclease depleted *Drosophila* model:

Once the best and most viable model has been selected, and the protocol optimised as necessary; the technically challenging and complex protocol for preparing a poly-ribo-seq library must be carried out. The process can be tested and validated at multiple steps along the way (for example by size-selection gel visualization, and bioanalyser quality control), and it must be ensured that every quality control step indicates a viable sample, before finally pooling and sending off the libraries for sequencing.

1.7.5 Use poly-ribo-seq data to identify lncRNA targets of Pacman and Dis3L2 in whole L3:

Once the sequencing has been carried out, and the data returned, it can be processed using bioinformatic pipelines, allowing comparisons of relative abundance of any given transcript, or population of transcripts, between samples. Although targets of Pacman and Dis3L2 have been analysed at the transcriptome-wide level in *Drosophila* WIDs previously (33, 46), the model eventually chosen and used here (whole L3 larvae) provides novel insight into the regulation of lncRNAs (and any other RNAs) by these exoribonucleases in whole L3 larvae. In addition, comparison to other sequencing datasets in *Drosophila* that include exoribonuclease depleted samples allows comparison and analysis of conserved and differentially regulated transcripts between different *Drosophila* models, as well as complete null mutants versus partial knockdown.

1.7.6 Evaluate the translational activity of exoribonuclease-regulated lncRNAs:

Similarly to above, and making the poly-ribo-seq technique so valuable, the novel data can be processed using bioinformatic pipelines, allowing comparisons of relative abundance of any given transcript, or population of transcripts, this time allowing comparisons between total lysate RNA and polysomal RNA from paired samples of the same genotype. This allows observation and identification of lncRNAs that are present on the polysome (and may be translating), as well as information on how the absence of these exoribonucleases may impact the association of target lncRNAs with the polysome. This could provide information on whether certain transcripts undergo co-translational degradation, and whether translation is necessary to their degradation.

1.7.7 Use combined analysis of data sets to predict whether lncRNA translation is exoribonuclease dependent:

Following the analysis of polysomal versus total lysate RNA in the different tested genotypes, further value can be extracted from this data by carrying out multi-dataset-comparisons between this novel data and other ribo-seq and poly-ribo-seq data that is available in *Drosophila*. Whilst little is available that can be used as a direct equivalent, consistency of particular candidate lncRNAs between datasets can help to increase confidence in the presence of a certain lncRNA on the polysome and help identify those more likely to have a conserved biological function. In addition, by examining the location of read pile-ups, potential ORFs can be identified for polysome associated lncRNAs that may be translated, allowing further testing of the ORF to be carried out.

Chapter 2: Methods

2.1 Polymerase Chain Reactions (PCRs):

2.1.1 Extracting DNA from whole adult *Drosophila* to test stocks:

For quick extraction and preparation of DNA from adult flies, 3 adults were selected (sex or phenotype selected when necessary) from a given stock. The flies were homogenised using a sterile miniature pestle in a 1.5mL Eppendorf tube, while suspended in “Quick Whole Fly Lysis Buffer” (as described in Table 2.1). The lysed *Drosophila* were incubated in this buffer at 37°C for 30 minutes, followed by incubation at 95°C for 5 minutes in order to denature the Proteinase K present in the buffer (added just before use). The mixture was then centrifuged for 2 minutes at 3000g to pellet the debris, and the supernatant placed in a clean Eppendorf tube. This material is then suitable to carry out PCR, using 1µL as template DNA.

2.1.2 Extracting RNA from whole *Drosophila* or L3 larvae:

RNA extractions were performed on 3 whole specimens (sex or phenotype selected when necessary) from a given stock. They were homogenized using a sterile miniature pestle in a 1.5mL Eppendorf tube, suspended in 350-700µL QIAzol lysis buffer (Qiagen, Cat No. 79306). The RNA and protein were separated by adding 0.2 volumes of chloroform, and vortexing thoroughly for 1 minute and incubated at room temperature for 3 minutes; followed by centrifugation at at 12,000g for 15 minutes. The upper, aqueous phase could then be separated into a clean Eppendorf tube.

Following this, the RNA was precipitated by the addition of 0.1 volume of 3M sodium acetate (pH 5.2), and either 2.5 volumes of >99.9% molecular grade ethanol, or 1 volume of >99.9% molecular grade isopropanol. This mixture was incubated at -80° for 1 hour, or -20° for 24 hours.

The precipitated RNA was pelleted by centrifugation (13000rpm, 30 minutes, 4°). The pellet was carefully washed and re-centrifuged twice in 75%, molecular grade

Table 2.1 - Whole fly lysis buffer:

Final concentration:	Reagent:
10mM	Tris (pH 8.2)
1mM	EDTA
25mM	NaCl
200μg/mL	Proteinase K

ethanol. After the final wash/centrifugation step, all supernatant was removed, and the pellet was left to air-dry in a fume hood. Pellets were subsequently resuspended in 30-50µL of nuclease-free water. RNA concentration was determined using a Nanodrop 1000 spectrophotometer, blanked against nuclease-free water.

2.1.3 Extracting RNA from L3 wing imaginal discs:

RNA extractions were carried out using a miRNeasy RNA extraction kit (Qiagen, Cat. No. 217084), as described in the protocol provided by the manufacturer, with an on-column DNase digestion. The RNA was eluted in 14-50µL of RNase-free water. RNA concentration was determined using a Nanodrop 1000 spectrophotometer, blanked against nuclease-free water.

2.1.4 L3 sample lysis for extracting ribosome bound RNA:

Many of the methods described in this section were subject to significant optimisation (as described in Chapter 5). The protocols listed in this section are representative of the final techniques used to gather results, whereas previous versions of the techniques will be described in the relevant results section.

2.1.1 Reverse Transcriptase Polymerase Chain Reaction (RT-PCR):

Reverse transcriptase PCR utilises the viral reverse transcriptase enzyme to generate a complementary DNA (cDNA) copy of the extracted RNA. Table 2.2 shows the reaction mixture made from the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, product code 4368814). Where necessary, an “RT negative” negative control was run alongside the other reactions, in which RNase-free water was used instead of the enzyme. Table 2.3 shows the temperatures and durations of the RT-PCR stages. Reactions were performed using the MultiScribe Reverse Transcriptase (Invitrogen, product code 4311235) and incubated in an Applied Biosystems Veriti 96-well Thermal cycler.

Table 2.2 – High-Capacity RT-PCR reaction mix:

Volume:	Reagent:
2μL	RT buffer
0.8μL	dNTP mix
2μL	RT random primers
1μL	Multiscribe RTase
4.2μL	Nuclease-free water
10μL	Sample RNA (100ng/μL)

Table 2.3 – High-Capacity RT-PCR cycling conditions:

Time (minutes):	Temperature (°C):
10	25°C
120	37°C
5	85°C
∞	4°C

2.1.2 AmpliTaq Polymerase Chain Reaction (AmpliTaq PCR):

Polymerase Chain Reaction (PCR) was used to amplify a specific sequence of DNA between two designed primers. This can be used to generate large amounts DNA within a region of interest, or to test for presence of a certain sequence in a DNA template. Reactions were performed using the AmpliTaq Gold 360 Master Mix from Applied Biosystems (product code 4398876), and heat cycles performed in an Applied Biosystems Veriti 96-well Thermal cycler. Tables 2.4 and 2.5 shows the reagents, temperatures, and durations of the PCR stages.

2.1.3 Phusion High Fidelity Polymerase Chain Reaction (HF-PCR):

Taq polymerases, including AmpliTaq, have an error rate of approximately 1 per 3,700 bases (depending on sequence). For certain experiments such as cloning, where fidelity is particularly important, various high-fidelity polymerases are available, such as Q5 and Phusion DNA polymerases. Phusion High-Fidelity DNA Polymerase (New England Biolabs, cat. no. M0530S) was used for HF-PCR in this project, with approximately 50x the fidelity of standard Taq polymerases. Reactions were carried out as per manufacturer’s instructions (as described in Table 2.6), and heat cycles performed in an Applied Biosystems Veriti 96-well Thermal cycler. Table 2.7 shows the temperatures and durations of the HF-PCR reaction stages.

2.1.4 Semi-Quantitive Polymerase Chain Reaction (sqPCR):

sqPCR is an applied use of standard PCR, followed by visualisation on a size-separation gel, that allows fluorescence of nucleic acid intercalating dyes (such as GelRed Nucleic Acid Stain, Biotin product code BT41003) to be measured as a proxy for concentration of nucleic acid concentration for the product represented by a particular band. PCR is carried out as with

standard AmpliTaq PCR (described in section 2.1.2), but at a range of different cycle numbers for the region of interest, in each sample for which abundance is to be compared. When the size-separated products are then visualised, their fluorescence

Table 2.4 – AmpliTaq PCR reaction mix:

Volume:	Reagent:
5µL	AmpliTaq 360 Master Mix
3.6µL	Nuclease-free water
0.2µL	Forward primer (10µM stock)
0.2µL	Reverse primer (10µM stock)
1µL	DNA

Table 2.5 – AmpliTaq PCR cycling conditions:

Time:	Temperature (°C):
30 seconds	98°C
10 seconds (cycled x n)	98°C
45 seconds (cycled x n)	Sequence dependent
1 minute (cycled x n)	72°C
10 minutes	72°C
∞	4°C

Table 2.6 – Phusion HF-PCR reaction mix:

Volume:	Reagent:
10µL	Phusion buffer (5x conc.)
1µL	dNTP mix
2.5µL	Forward primer (10mM stock)
2.5µL	Reverse primer (10mM stock)
0.5µL	Taq Phusion enzyme
32.5µL	Nuclease-free water
1µL	cDNA

Table 2.7 – Phusion HF-PCR cycling conditions:

Time:	Temperature (°C):
30 seconds	98°C
10 seconds (cycled x n)	98°C
45 seconds (cycled x n)	Sequence dependent
1 minute (cycled x n)	72°C
10 minutes	72°C
∞	4°C

can be measured using an Odyssey LI-COR Fc Imaging system, at a cycle number that has not saturated the reaction. By ensuring the same concentration of input cDNA and using the same cycle number between samples being compared, the fluorescence measured can then be compared to quantify the abundance of the target region in the starting DNA of each sample relative to another sample. Reactions were carried out as per manufacturer's instructions (as described in Table 2.4), and heat cycled in an Applied Biosystems Veriti 96-well Thermal cycler. Table 2.5 shows the temperatures and durations of the HF-PCR reaction stages.

2.1.5 Quantitive Real Time Polymerase Chain Reaction (qRT-PCR):

TaqMan qPCR is a variant of PCR that utilises real time detection of fluorescence emitted upon cleavage of a DNA probe (separating the fluorescent dye from its quencher) by the endogenous 5' nuclease activity of the polymerase during PCR cycling (Figure 2.1). This fluorescence, increasing with PCR cycles, can be used to calculate abundance of a certain DNA sequence, compared to another gene known not to change (a "housekeeper gene"), by how many cycles are required to reach a fluorescence threshold. Another variant of qRT-PCR, called SYBR Green uses a dsRNA dye to quantify abundance, however this will bind any dsDNA PCR products, including primer-dimer and other non-specific products. Although the SYBR Green method is cheaper, the low levels and minor differential expressions found with many lncRNAs meant that TaqMan was used in this project, for its increased specificity, and therefore accuracy. Reactions were performed using the TaqMan Universal PCR Master Mix (Applied Biosystems, product code 4324018) (specified in Table 2.8), with the fast reaction temperatures and times described in Table 2.9.

2.2 DNA product selection and purification:

2.2.1 PCR product purification:

Specific production of a desired DNA product obtained by PCR was followed by purification to remove all residual dNTPs, primer, enzyme, and ions from the buffer.

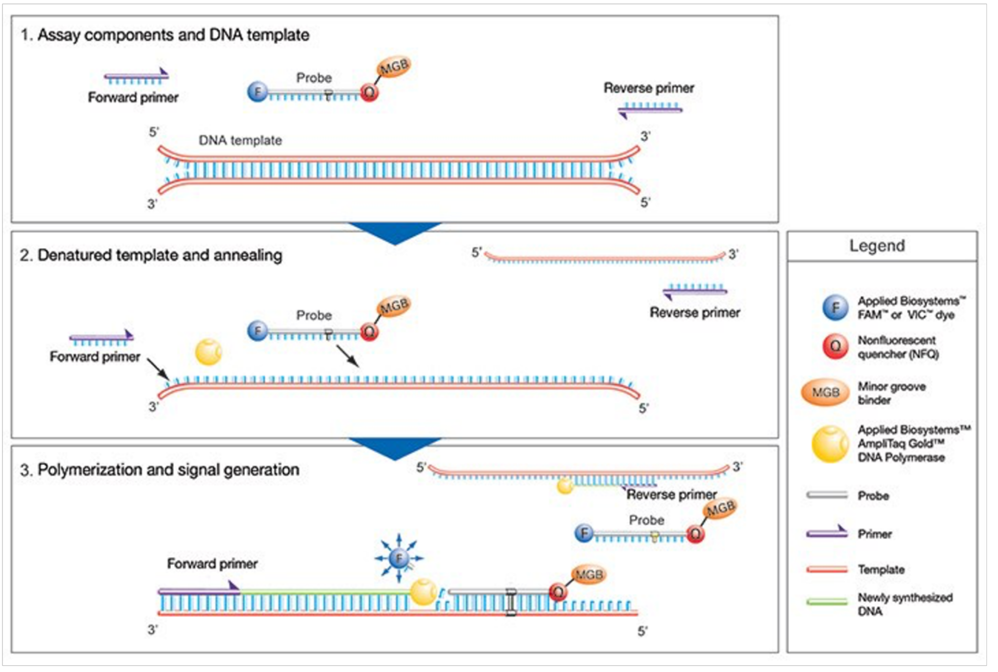

1. Assay components and DNA template
Forward primer
Probe
F
Q
MGB
Reverse primer
5'
DNA template
3'
3'
5'
2. Denatured template and annealing
5'
3'
Reverse primer
Probe
Forward primer
3'
5'
3. Polymerization and signal generation
5'
3'
Reverse primer
Probe
Forward primer
3'
5'
Legend
Applied Biosystems™ FAM™ or VIC™ dye
Nonfluorescent quencher (NFQ)
Minor groove binder
Applied Biosystems™ AmpliTaq Gold™ DNA Polymerase
Probe
Primer
Template
Newly synthesized DNA

Figure 2.1 – A graphic demonstrating the mechanism by which TaqMan qPCR can be used to calculate abundance of a target gene region (graphic from ThermoFisher).

At the start of real-time PCR, the temperature is raised to denature the double-stranded cDNA. During this step, the signal from the fluorescent dye on the 5' end of the TaqMan probe is quenched by the quencher on the 3' end. Next, the reaction temperature is lowered to allow the primers and probe to anneal to their specific target sequences. Taq DNA polymerase synthesizes new strands using the unlabelled primers and the template. When the polymerase reaches a TaqMan probe, its endogenous 5' nuclease activity cleaves the probe, separating the dye from the quencher.

Table 2.8 – TaqMan qPCR reaction mix:

Volume:	Reagent:
5µL	TaqMan Master Mix
3.7µL	Nuclease-free water
0.5µL	Assay
0.8µL	cDNA

Table 2.9 – TaqMan qPCR cycling conditions:

Time:	Temperature (°C):
2 minutes	50°C
10 minutes	95°C
15 seconds (cycled x 40-50)	95°C
1 minute (cycled x 40-50)	60°C

This was carried out using the “QIAquick PCR Purification Kit” (Qiagen, product code 28104), following the protocol provided by the manufacturer.

2.2.2 Size selection and gel purification:

In some instances, non-specific PCR products were formed, or the PCR product needed to be separated from primer dimers, either of which would complicate PCR amplification. In these instances, the gel band of the desired size was excised with a sterile, RNase-free scalpel, and the DNA within the excised gel was recovered using the MinElute Gel Extraction Kit (Qiagen, Cat. No. 28604), following the protocol provided by the manufacturer.

2.3 Western Blotting

Western blotting was used in this project to assess expression of certain proteins in S2 cells. The lysis step described is per 350µL of confluent S2 cell culture (pelleted at 3000g for 3 minutes) but can be scaled as necessary for particularly large samples. For each cell pellet, 150µL of Western Lysis Buffer (Table 2.10) with 2% volume of β-mercaptoethanol and 2% volume of 50x stock protease inhibitor cocktail was added, bringing the concentration of the protease inhibitor cocktail to 1x. The samples were then boiled for 7 minutes at 100°C. After this, the samples were centrifuged at 21,000g for 5 minutes, to separate the cellular debris. Per sample, 50µL of supernatant was recovered from the lysis mixture.

NuPAGE Tris-Acetate SDS running buffer (20X) (Invitrogen, Product code LA0041) was diluted to 1x in a volume of 750ml with UltraPure water. Pre-made NuPAGE™ 7%, Tris-Acetate, 1.5 mm, Mini Protein Gels (Invitrogen, product code EA03585BOX) were used to separate

protein products. A gel tank was prepared and set up with a 7% Tris-Acetate gel. Once the gel cassette was sealed into the tank, the central reservoir, then the rest of the tank, was filled with running buffer. The samples were each mixed with 2.5µL bromophenol-blue and loaded into the gel alongside 10µL of ColourPlus prestained protein size ladder

Table 2.10 – Western Lysis Buffer:

Volume:	Reagent:
6.0mL	Glycerol
4.8mL	Tris (pH 6.8)
12.0mL	10% SDS
7.2mL	UltraPure water

Broad Range (10-250 kDa) (New England Biolabs. product code P7719S) (after heating the ladder at 100°C for 2 minutes). The gel was run at 150V for 65 minutes.

Immobilon PVDF-FL membrane (Merck Millipore, product code IPFL00010) was cut to the size of the gel and soaked in 100% methanol for 1 minute. The membrane was subsequently washed in UltraPure water for 2 minutes. Transfer tank sponges, filter paper, and the membrane were incubated in Large Protein Transfer Buffer (Table 2.11) for 30 minutes. The transfer tank was prepared, and an ice block added, before filling it with transfer buffer. Sponges, filter paper, gel, and membrane were layered into the transfer cassette, and run at 100V for 60 minutes.

The membrane was washed in PBS for 5 minutes. After this, the membrane was blocked for 60 minutes at room temperature in 25ml of Odyssey blocking buffer (LI-COR, product code 927-40000) (PBS), on a rocking platform. The membrane was placed in a 50ml falcon tube and incubated in the desired primary antibodies diluted in 3ml of Odyssey blocking buffer (PBS) to their specified effective concentrations. To this, 0.1% volume Tween was added, and left on rolling platform overnight at 4°C.
Following this, the membrane was washed four times for 5 minute in PBS with 0.1% volume Tween. Next, the membrane was transferred to a lightproof box containing 20ml Odyssey blocking buffer (PBS) with 0.1% volume Tween, and 0.01% mass SDS. Into this, any required secondary antibodies were diluted at 1:20000. This mixture was incubated for 60 minutes
on a rocking incubator at room temperature. The fully incubated membrane was washed four times for 5 minutes in PBS with 0.1% Tween, and a further two times for 5 minutes in PBS. Imaging of the membrane on the Licor Odyssey was carried out at 600nm, 700nm, and 800nm channels.

2.4 *Drosophila* S2 cell tissue culture

2.4.1 S2 cell maintenance

Drosophila S2 cells were grown in a cell incubator at 25°C, in Complete Schneider's *Drosophila* medium, supplemented with foetal bovine serum (FBS) to a final concentration of 10% volume, and penicillin-streptomycin to a final concentration of

Table 2.11 – Large Protein Transfer Buffer:

Volume/Mass:	Reagent:
3g	Tris (pH6.8)
14.25g	Glycine
0.5g	10% SDS
200mL	Molecular grade ethanol
800mL	UltraPure water

1U/ml penicillin and 1ug/ml streptomycin. As the cells are semi-adherent, they require dislodging by pipetting and mixing before passaging. The cells were split at either a 1:2 or 1:5 dilution when they achieve a density between 6-20 x 10^6 cells/ml (at which point they are visibly confluent). When healthy, *Drosophila* S2 cells grow very quickly, with a doubling time of approximately 24 hours (temperature, density, and nutrient dependent). They usually require splitting every 2-3 days.

2.4.2 Transfection of S2 cells

Any constructs requiring transfection into S2 cells were introduced into the cells using the "FuGene HD Transfection Reagent" kit. *Drosophila* S2 cells were seeded at a density of 2 million cells/well in a 6-well plate in a total volume of 2ml of complete Schneider's S2 medium/well (scaled down if plated in smaller wells). These cells were left to settle and grow overnight, allowing the cells to be treated as a semi-adherent layer with approximately 80% confluency. After this period, all media was removed, and replaced with 950µL of serum-free Schneider's S2 medium. The transfection mixture was then made up in 50µL of complete S2 medium, per well (for a 6-well plate). A total of 2µg of each relevant construct DNA and 3µL of FuGene HD was added to the medium, with thorough vortexing between each addition. Care was taken to ensure that no undiluted FuGene touched the sides of the Eppendorf in which the mixture was made. The mixture was incubated for 15 minutes, at room temperature. After this, the entirety of the appropriate transfection mixture was added to each well containing 1mL of serum free media and mixed by gently shaking the plate. This transfection reaction was left undisturbed for 90 minutes, before 1mL of double concentration Schneider's S2 medium was added in order to provide the cells with suitable nutrients. The protocol described here is post-optimisation, which is described in Chapter 5.

2.4.3 S2 cell death assays

In order to test rate of death of S2 cells, a haemocytometer was used, with two separate counting grids. The cell population in question would be dislodged from the flask surface by gentle pipetting. A 0.4% Trypan Blue solution, buffered in pH 7.2 PBS was prepared, and mixed with an equal volume of the dislodged S2 cell culture by pipetting. This mixture was pipetted onto each counting grid of the haemocytometer, until the surface tension allowed the mixture to cover the entire area between the grid and the glass cover slip. This was immediately examined under a microscope).

2.4.4 S2 cell microscopy

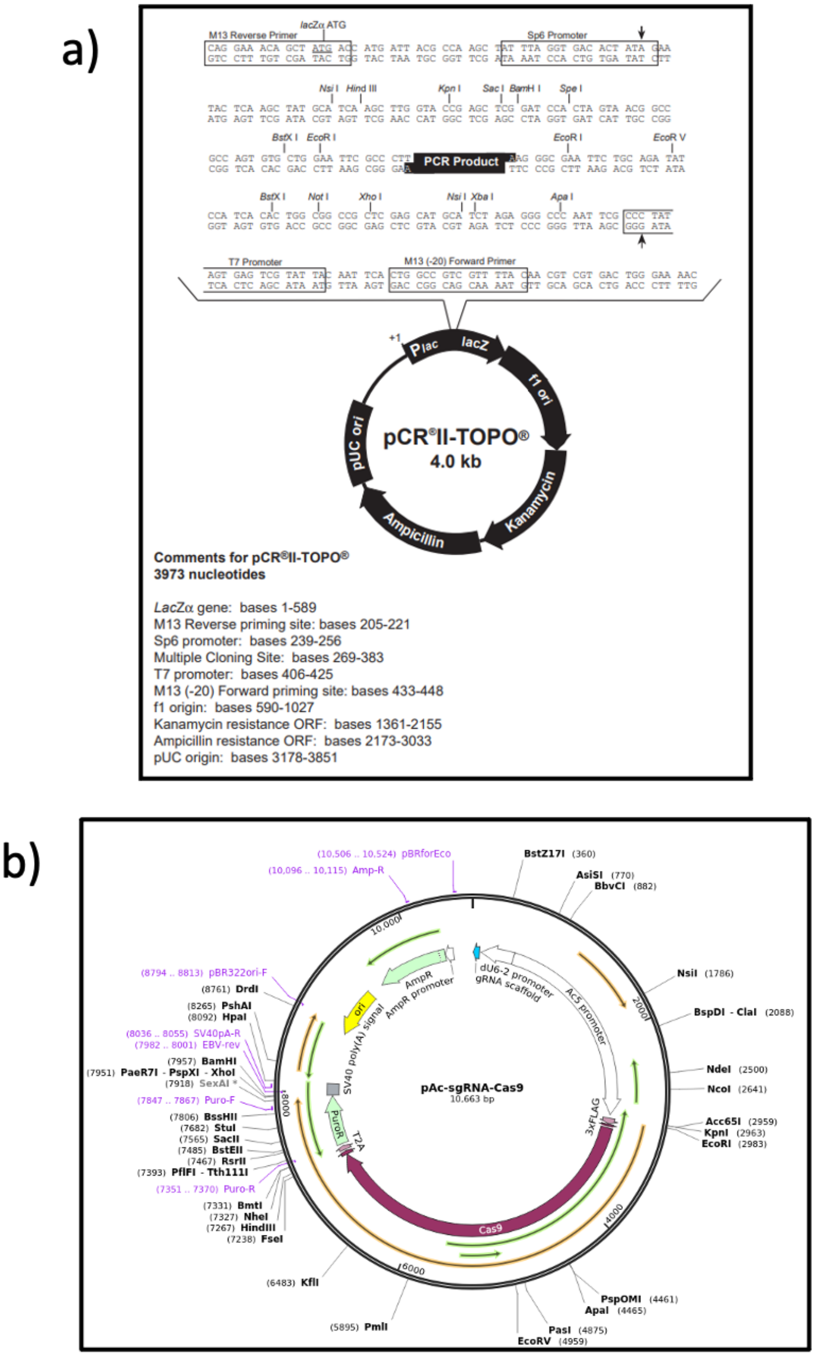
For examination of cell wellbeing, confluence, and phenotypes, cells were examined using a tissue culture microscope. When needed, LED lights were applied in order to excite fluorescence from any GFP expressing cells.

2.5 Cloning techniques:

In order to attempt a depletion of Pacman and Dis3L2 levels in *Drosophila* S2 cells, dsRNA was to be produced (complimentary to regions within the gene of interest, in order to induce a knockdown of the gene). The T7 RiboMax system was used, requiring the creation of template DNA, complimentary to the target region, with the addition of T7 promoters. Separately from that, a CRISPR-Cas9 capable plasmid was used; short guide RNAs (sgRNAs) complimentary to target gene regions were synthesised and annealed into an existing plasmid containing other components (such as Cas9) necessary to knocking out target genes with the CRISPR-Cas9 system. For both of these purposes, various cloning techniques, described subsequently, were used. Plasmid maps and catalogue numbers are shown in Figure 2.2.

2.5.1 S2 cell RNA extraction:

The region of the target genes would need to be amplified from existing total *Drosophila* cDNA, easily found in and extracted from *Drosophila* S2 cells. S2 cells were harvested from a confluent culture in a 25mL flask, after repeated pipetting to dislodge them. These S2 cells were pelleted by centrifugation (3000g, 3 minutes, room temperature). Subsequently, the supernatant was discarded, and the pellet was washed and re-pelleted in sterile phosphate-buffered saline. This wash step was carried out twice. To pelleted samples, 700µL of QIAzol lysis reagent (Qiagen cat. no.79306), and 140µL of chloroform were added. This mixture was vortexed for 1 minute, and



Comments for pCR®II-TOPO®
3973 nucleotides

LacZα gene: bases 1-589
M13 Reverse priming site: bases 205-221
Sp6 promoter: bases 239-256
Multiple Cloning Site: bases 269-383
T7 promoter: bases 406-425
M13 (-20) Forward priming site: bases 433-448
f1 origin: bases 590-1027
Kanamycin resistance ORF: bases 1361-2155
Ampicillin resistance ORF: bases 2173-3033
pUC origin: bases 3178-3851

Figure 2.2 – Plasmids used in this project.

(a) pCRII-TOPO plasmid, included as part of "TOPO™ TA Cloning™ Kit, Dual Promoter, with pCR™II-TOPO™ Vector and One Shot™ Mach1™ T1 Phage-Resistant Chemically Competent E. coli" kit (ThermoFisher, catalogue number K461020)

(b) pAc-sgRNA-Cas9 plasmid, available from Ji-Lobg Liu lab (AddGene catalogue number 49330) incubated at room temperature for a further 3 minutes, before centrifuging at 21,000g for 10 minutes. The upper, aqueous phase was transferred to a sterile, RNase-free tube. The RNA was precipitated by the addition of 0.1 volume of 3M sodium acetate (pH 5.2), and either 2.5 volumes of >99.9% molecular grade ethanol, or 1 volume of >99.9% molecular grade isopropanol. This mixture was incubated at -80° for 1 hour, or -20° for 24 hours.

The precipitated RNA was pelleted by centrifugation (13000rpm, 30 minutes, 4°). The pellet was washed and re-centrifuged in 100%, and subsequently 75%, molecular grade ethanol. After the final wash/centrifugation step, all supernatant was removed, and the pellet was left to air-dry in a fume hood. Pellets were subsequently resuspended in 30-50µL of nuclease-free water. RNA concentration was determined using a Nanodrop 1000 spectrophotometer, blanked against nuclease-free water.

2.5.2 Amplification of gene region:

Primers were designed for the target regions, with any additions such as T7 promoter regions or tags added onto the sequence. The primers were ordered from Sigma-Aldrich (now owned by Merck), and HF-PCR was used to accurately amplify the target region from cDNA produced by RT-PCR, as previously described. The product was tested by size separation on an agarose gel and subsequent imaging. Depending on the specificity of the PCR products, they were purified by either PCR product purification or further size selection and subsequent gel extraction.

2.5.3.1 TOPO cloning reaction and heat-shock transformation of TOP10 *E. coli* cells:

To make up the TOPO cloning mix, for each reaction, 2µL of purified PCR product was mixed with 0.5µL of TOPO suitable salt solution, and 0.5µL of the provided TOPO vector (ThermoFisher, catalogue number K461020; shown on Figure 2.2, panel (a)). This was subsequently incubated for 30 minutes at room temperature (increased from the 5 minutes to increase efficiency of the reaction).

A 50µL aliquot of One Shot TOP10 E. coli cells (ThermoFisher, catalogue number C404003) were defrosted from -80°C (on ice) and mixed with 3µL of the TOPO cloning mix. This was incubated on ice for 15 minutes. The tube was then mixed by gently flicking (do not vortex) and placed back on ice for a further 15 minutes. After this incubation time was completed, the tube was heat shocked at 42°C for 45 seconds and placed back on ice for 5 minutes. Following this, 950µL of SOC was added, and the sample was moved to a rocking incubator (37°C, 120rpm) for 1 hour.

2.5.3.2 Heat-shock transformation of DH5α *E. coli* cells:

Aliquots of DH5α cells (ThermoFisher, catalogue number 18265017) were defrosted from -80°C (on ice). To each 50µL aliquot of DH5α cells, 100ng of the ligated plasmid was made up (volume dependent on concentration, but no more than 3µL were used), and mixed by gently flicking the tube (do not vortex). This mixture was incubated on ice for 30 minutes, before being exposed to a heat-shock at 42°C for 30 seconds, then returned to incubate on ice for another 5 minutes. Next, 950µL of SOC was added to each tube, and the sample was moved to a rocking incubator (37°C, 120rpm) for 1 hour.

2.5.4 Making up LB-agar plates and LB-broth:

Agar plates were prepared in order to plate and grow the transformed bacteria. A volume of 500ml of LB-agar medium was made up, described in Table 2.12 A volume of 500ml of LB-broth was made up, described in Table 2.13.

The mixtures were set on a hotplate stirrer with a magnetic flea until completely dissolved, and subsequently autoclaved to ensure sterility. Immediately after autoclaving, both were kept at 55°C in a pre-warmed water bath until their temperature had dropped to the temperature maintained by the water bath. Once at 55°C, any desired antibiotics could be added at their desired concentration, and the bottles inverted repeatedly to mix. The mixed agar must be swiftly and carefully poured out onto sterile bacterial culture plates to a thickness of approximately 5mm. The plates were left to set, with lids ajar. The LB-broth was allowed to cool to room temperature

Table 2.12 – LB-agar medium:

Volume:	Reagent:
20g	LB-agar powder
500mL	UltraPure water

Table 2.13 – LB-agar broth:

Volume:	Reagent:
12.5g	LB-broth powder
500mL	UltraPure water

and kept in a sealed bottle until required. Once set, the agar plates were sealed and kept in sterile zip-lock bags at 4°C until required.

2.5.5 Growing and selecting bacterial colonies:

An hour prior to use, LB-agar plates were warmed to 37°C in a pre-heated incubator. The 1mL total volume from each TOPO reaction was split between plates as follows, to produce a high, medium, and low seeding density, described in Table 2.14.

The plates were grown up for 12-18 hours in a 37°C incubator. Subsequently, distinct, non-overlapping colonies were selected and picked with a sterile tip. The tip and colony were then placed in a sterile 15ml universal tube, containing 5ml of the LB-broth. The tubes were labelled, and corresponding labels placed on the plate they were selected from. The tubes were placed in an orbital shaking incubator overnight at 37°C and 120rpm. In cases where low growth rates were observed, these steps could be repeated whilst resuspending all of each TOPO reaction in a total of 250µL and plating all cells onto a single LB-agar plate.

2.5.6 Bacterial DNA extraction and testing for plasmid uptake:

Each tube was checked for visible bacterial growth (broth should turn cloudy) and shaken to ensure that the bacteria are suspended throughout the LB-broth. For each, a 100µL aliquot
was taken, and boiled in a sterile Eppendorf tube for 5 minutes at 100°C. The sample was then centrifuged at 3000g for 5 minutes, in order to pellet cell debris. AmpliTaq PCR was carried out with primers complimentary to the gene region of interest using 1µL of the supernatant as a template.

2.5.7 Bacterial mini-prep:

Mini-preps were carried out in order to extract plasmid DNA from bacterial colonies, using the “QIAprep Spin Miniprep Kit” (Qiagen, Cat. No. 27104) according to the

Table 2.14 – TOPO seeding densities:

Concentration:	Volume taken from TOPO reaction:	Additional SOC:
High	625µL, centrifuged at 1250rpm for 3 minutes. Supernatant discarded.	Pellet resuspended in 250µL
Medium	250µL	None
Low	125µL	125µL

protocol provided by the manufacturer. The DNA was eluted in 50µL of nuclease-free water.

2.6 dsRNA knockdown:

Having used the previous cloning techniques to produce bacterial DNA samples with the T7 promoters attached to each target gene region, the following techniques allowed production of high quantities of dsRNA. The techniques described here are post-optimisation, although the optimisation of said techniques is discussed at length in results Chapter 3.

2.6.1 RNA synthesis:

The target region was further amplified from extracted bacterial DNA using the previously described Phusion HF-PCR (section 2.1.3). The product was purified using PCR product purification or size selection and gel extraction, also as described previously (sections 2.2.1 and 2.2.2). RNA was produced from the HF-PCR product using an optimised protocol for the Promega T7 Ribomax express large-scale RNA production system (Promega, cat. no. P1320). The reagents were defrosted on ice, and added in the following order and volumes:

The reaction mixture (Table 2.15) was pipetted up and down to properly mix and incubated at 37°C for 24 hours. Once the reaction is complete, an additional 30µL of nuclease-free

water was added (to allow for more accurate nanodrop usage, as the concentration is extremely high, and the mixture ends up with a high viscosity).

2.6.2 RNA annealing reaction:

Synthesised RNA was extracted and purified using the previously described protocol for QIAzol:chloroform extraction of RNA (section 2.1.2, paragraphs 2 and 3). Once the RNA had been re-suspended in nuclease-free water, an equal volume of 2x annealing buffer (Table 2.16) was added. At this point, 1µL was taken and set aside for comparison to

Table 2.15 – T7 reaction mix:

Order added:	Reagent:	Volume:
1	T7 Ribomax 2x Buffer	10µL
2	DNA template to give 500ng total	1-8µL
3	Nuclease-free water to give 20µL total reaction volume	1-7µL
4	T7 express enzyme mix	2µL

Table 2.16 – Annealing buffer (2x concentration):

Concentration:	Reagent:
20mM	Tris (pH 7.5-8.0)
100mM	NaCl
1mM	EDTA

allow an estimation of annealing efficiency. The remaining synthesized RNA was heated to denature any secondary structure and supercoiling, and slowly cooled to room temperature (Table 2.17), to allow annealing to take place, resulting in the formation of double-stranded RNA. After the reaction, a sample from pre- and post-annealing reaction were run on an agarose gel, to observe a visible shift in running speed.

2.6.3 dsRNA treatment of *Drosophila* S2 cells:

Drosophila S2 cells (purchased from the *Drosophila* Genomics Resource Centre) were seeded on a 24-well cell culture plate (in complete Schneider's medium) at a concentration of 300000 cells/well, in a volume of 350µL per well. At least 3 replicates of each desired treatment were plated. After plating the cells, the cells were left to settle for 12 hours. The following seven-day protocol (Table 2.18) was then followed to achieve knockdown. Knockdown efficiency was observed by Western blotting, as previously described (section 2.3).

2.7 CRISPR-Cas9 system:

In order to produce a viable sgRNA-CRISPR-Cas9 plasmid, sgRNA for the target regions was synthesised, and annealed into an existing plasmid, as described below. This plasmid could then be transformed into DH5α cells, grown, selected, amplified, and mini-prepped as previously described.

2.7.1 Generating annealed and phosphorylated oligos for use as sgRNA:

Oligos to create sgRNA complimentary to a region within the XRN1 gene and Dis3L2 gene, respectively, were designed as two complimentary primers per gene. To each, an overhang allowing for annealing to the BspQ1 cut site within the plasmid was added. These primers were ordered from Sigma (primer sequences and locations described in Table 2.19 and Figure 2.3). Gene maps of protein-coding exoribonuclease transcripts are shown in Figure 2.4. Oligos were annealed and phosphorylated in their pairs. The reaction mix and reaction conditions are described in Tables 2.20 and 2.21.

Table 2.17 – Annealing conditions:

Step:	Time (minutes):	Temperature (°C):
1	10	90°C
2	1	85°C
3	1	80°C
4	1	75°C
5	1	70°C
6	1	65°C
7	3	60°C
8	3	55°C
9	3	50°C
10	3	45°C
11	3	40°C
12	60	37°C
13	∞	4°C

Table 2.18 – dsRNA treatment protocol:

Day:	Protocol:
0	Cells were plated at 300000 cells/well in 350µL of complete medium.
1	Complete medium was carefully removed from the confluent cell layer by pipetting. Immediately after, 175µL of serum-free medium (containing 15µg of the appropriate dsRNA) was added. The cells were incubated in the serum-free medium for 1 hour, after which, 175µL of 2X concentration complete medium was added. Cells were observed under the microscope, for visible phenotypic changes.
2	Cells were observed under the microscope, for visible phenotypic changes.
3	Cells were observed under the microscope, for visible phenotypic changes.
4	Complete medium was carefully removed from the confluent cell layer by pipetting. Immediately after, 175µL of serum-free medium (containing 15µg of the appropriate dsRNA) was added. The cells were incubated in the serum-free medium for 1 hour, after which, 175µL of 2X concentration complete medium was added. Cells were observed under the microscope, for visible phenotypic changes.
5	Cells were observed under the microscope, for visible phenotypic changes
6	Cells were observed under the microscope, for visible phenotypic changes
7	Cells were observed under the microscope, for visible phenotypic changes. Cells were dislodged, centrifuged at 1250 rpm for 3 minutes to pellet, washed in sterile PBS, and centrifuged again. The supernatant was removed, and the pellet was snap-frozen in liquid nitrogen.

Table 2.19 – sgRNA oligo details:

Primer name:	Position from start of gene:	Overhang:	Target sequence:
XRN1-R	300	AAC	AACCGCGCGCCGTCCGGAATCGC
XRN1-F	300	CTT	GCGCGCGGCAGGCCTTAGCGCTT
Dis3L2-R	42	AAC	AACCGTTGACGCTTGACGTTTCC
Dis3L2-F	42	CTT	GCAACTGCGAACTGCAAAGGCTT
Please note that primers were labelled Forward and Reverse based on the initial annotation of the sequences they were derived from. This doesn't align with the subsequent NGS sequencing, in which they are essentially reversed. The original labels have been retained out of convenience.			

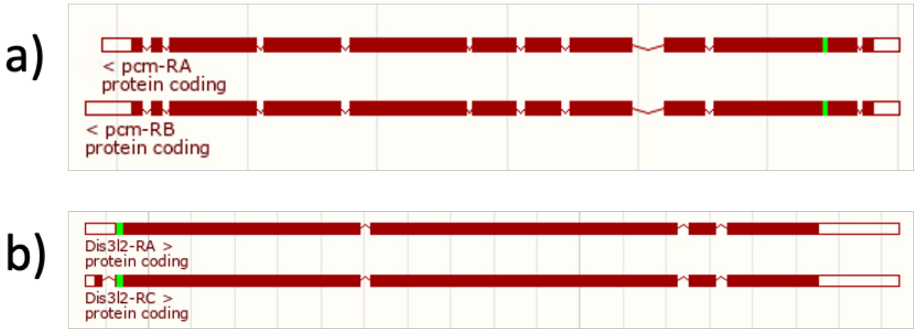


Figure 2.3 – Gene maps of protein-coding exoribonuclease transcripts.

(a) Gene maps of protein-coding *Pacman* transcript.

(b) Gene maps of protein-coding *Dis3L2* transcript.

Green highlighted region shows sgRNA binding locations.

Table 2.20 – Oligo annealing and phosphorylation mix:

Volume:	Reagent:
1µL	100µM Forward oligo (XRN1 or Dis3L2)
1µL	100µM Reverse oligo (XRN1 or Dis3L2)
1µL	10X T4 DNA ligase buffer
0.5µL	T4 polynucleotide kinase
6.5µL	Nuclease-free water

Table 2.21 – Oligo annealing and phosphorylation conditions:		
Purpose:	Time (minutes):	Temperature (°C):
Allows phosphorylation of 5’ ends.	30	37°C
Remove coiling and secondary structure, denatures enzyme	5	95°C
Allow for gradual and specific annealing of complimentary strands.	1	90°C
	1	85°C
	1	80°C
	1	75°C
	1	70°C
	1	65°C
	1	60°C
	1	55°C
	1	50°C
	1	45°C
	1	40°C
	1	35°C
	1	30°C
	1	25°C
Keep newly annealed oligos stable.	∞	4°C

2.7.2 Digesting pAc-sgRNA-Cas9 plasmid for use:

Digest mixes were made up as described in Table 2.22. The mixes were incubated for 30 minutes at 37°C, and after the addition of Calf Intestinal Phosphatase, incubated for another 30 minutes at 37°C

After the reaction had taken place, 1µL of each digest product was run on a 1.2% agarose gel, alongside a 1kb ladder, to observe whether bands were observed at the expected sizes for (single and double) digest products. Subsequently, the entire remaining products were run on a gel. The relevant bands were excised, and a gel extraction was carried out as previously described.

2.7.3 Ligation of sgRNA and plasmid

The sgRNA and pAc-sgRNA-Cas9 plasmid (Figure 2.2) were mixed with ligase and buffer, as described in Tables 2.23 and 2.24. Reaction mixtures were left overnight at room temperature to anneal. After this, 1µL of the ligated mix was then run on an agarose gel alongside some untreated plasmid to observe a shift caused by the insertion and ligation of the sgRNA.

2.7.4 Transfecting the CRISPR-Cas9 construct into S2 cells:

The constructs (described in Table 2.25) were transfected into S2 cells using the S2 transfection techniques described previously in this chapter.

After 3 days, the cells were observed under the microscope to ensure relative health of the cultures, and the UAS(GFP) + Gal4 transfected cultures were observed by fluorescence microscopy to ensure that some successful transfection had taken place. Once this was done, the media was carefully pipetted off, and replaced with 1.5ml of complete S2 media, and 0.5ml of conditioned complete S2 media (prepared by centrifuging the cells out of the media (1250rpm for 5 minutes) from healthy growing S2 cultures, leaving growth factors and signalling peptides).

Table 2.22 – pAc-sgRNA-Cas9 digest mix:

BspQI+ EcoRI- digest:	BspQI+ EcoRI+ digest:
10 units BspQI	10 units BspQI
0 units EcoRI	10 units EcoRI
1µg pAc-sgRNA-Cas9 plasmid	1µg pAc-sgRNA-Cas9 plasmid
5µL 10X NEB buffer	5µL 10X NEB buffer
Make up to 50µL with nuclease-free water	Make up to 50µL with nuclease-free water

Table 2.23 – Pacman sgRNA-plasmid ligation mix:

XRN1 sgRNA + plasmid:
50ng digested pAc-sgRNA-Cas9 plasmid
1µL annealed XRN1 sgRNA oligos (1:200 dilution)
1.5µL 10X T4 ligase buffer
1µL T4 DNA ligase
7µL nuclease-free water

Table 2.24 – Pacman sgRNA-plasmid ligation mix:

Dis3L2 sgRNA + plasmid:
50ng digested pAc-sgRNA-Cas9 plasmid
1µL annealed Dis3L2 sgRNA oligos (1:200 dilution)
1.5µL 10X T4 ligase buffer
1µL T4 DNA ligase
7µL nuclease-free water

Table 2.25 – CRISPR-Cas9 construct details:

Construct:	Purpose:
pAc-(XRN1/pacman)sgRNA-Cas9	Induce CRISPR editing of XRN1/pacman, leading to the production of several mutants, including at least one XRN1/pacman null.
pAc-(Dis3L2)sgRNA-Cas9	Induce CRISPR editing of Dis3L2, leading to the production of several mutants, including at least one Dis3L2 null.
UAS(GFP) + Gal4	Induce successfully transfected cells to produce GFP, allowing fluorescence microscopy to show that successful transfection has taken place, and provide an idea of translation efficiency (known to be low in S2 cells).

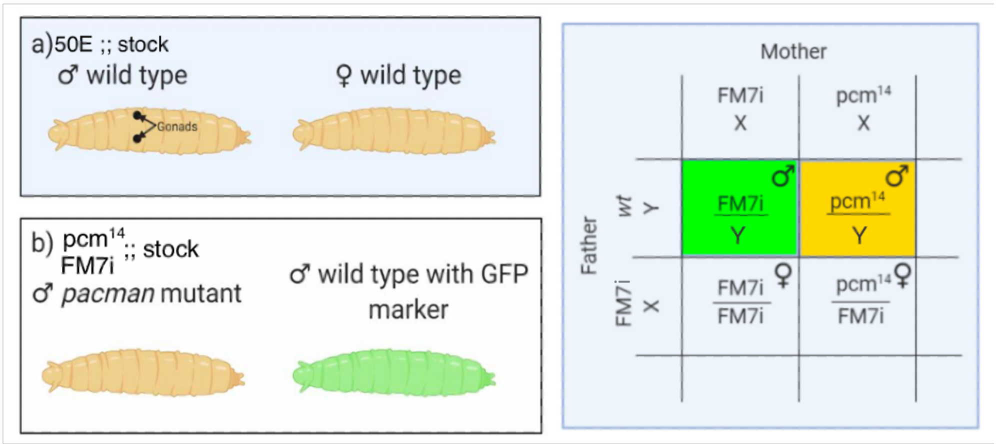
These cultures were then left to recover and grow for 3 days. Subsequently, the media was changed again for another 1.5ml complete + 0.5ml conditioned complete S2 media, this time with the addition of 10µg of Puromycin per well (in order to select for cells with the puromycin resistant pAc gene from the plasmid). After 3 days in these conditions, the media was changed again, for newly prepared 1.5ml complete + 0.5ml conditioned complete S2 media + 10µg of Puromycin per well. After this selective pressure, the cultures were observed under the microscope to ensure significant cell death. The media was replaced with 1.5ml complete + 0.5ml conditioned complete S2 media (without Puromycin, as the transfection would be transient only), and cultured in these wells, with media changed every 3 days, to allow the culture to recover back to normal growth and confluence.

2.7.5 Isolating S2 monocultures with successful CRISPR-induced mutations:

Cells were counted using a haemocytometer and diluted to extremely low concentrations, calculated based on bringing the cell count to 0.5 cells per well. These cultures were grown in 75% volume complete media + 25% conditioned complete media. The cells were grown over extended periods, and regularly observed to ensure cell presence and growth. From these, either significant cultures should be grown (sub-single cell concentration cultures), or a homogenous cell pools should be isolated (low cell density/volume cultures), and cultured.

2.8 Generating, maintaining, and using *Drosophila melanogaster* stocks to produce L3 larval samples:

Pre-existing *Drosophila melanogaster* stocks from the Newbury lab were amplified and maintained for these experiments. Their visible phenotypes and markers are described in Figure 2.4.



		Mother	
		FM7i X	pcm^{14} X
Father	*wt* Y	♂ $\frac{\text{FM7i}}{\text{Y}}$	♂ $\frac{\text{pcm}^{14}}{\text{Y}}$
	FM7i X	♀ $\frac{\text{FM7i}}{\text{FM7i}}$	♀ $\frac{\text{pcm}^{14}}{\text{FM7i}}$

Figure 2.4 – Phenotypic selection of L3 wandering larvae and the cross for *pacman* mutants and controls.

Wild type males were selected by identifying the visible presence of their gonads, when compared to the female wild type in (a). Pacman mutant L3 larvae were selected against a GFP marker expressed in wild type samples. Shown in (b), this was observed under a fluorescent microscope.

2.8.1 *Drosophila* husbandry:

Drosophila stocks were cultured on standard *Drosophila* media containing agar (81g), baker's yeast (81g), oatmeal (616g), black treacle (410g), propionic acid (40mL), nipagin (1 spatula), and distilled water (made up to 7L). Stocks were maintained at 25°C, and turned over into fresh food every 2 weeks. No crosses were required, as all genotypes used in the project had already been crossed and stabilised as stocks.

2.8.1.1 Timed egg lays:

In order to produce populations of *Drosophila* known to be at the same point in their life cycle, stocks were tipped into fresh fly food for periods of 3 hours (kept at 25°C), before being returned to the original, stock bottle. This ensured that any eggs in the bottle were laid within the same 3 hour period and could be confidently aged to the same developmental point ± 3 hours. This is important for experiments like RNA-sequencing that examine the entire RNA profile of the organism (which will of course vary significantly during development). For wandering L3 larval samples, they were aged 72 hours from egg lay, and collected as they wandered from the food.

2.8.2 L3 sample collection:

For each replicate, late L3 wandering male larvae of the desired genotype were selected for wild-type as well as *dis3l2* and *pacman* mutant samples, using "GFP" (from FM7i balancer) or "tubby" (from TM6 balancer) genetic markers to select only the desired mutants from the stocks (Figure 2.4), observed using a Leica GFP fluorescence microscope. Male samples were used due to *pacman* null mutations being lethal in females and matching this necessary sex selection in the other genotypes allowed a more direct comparison to be made, excluding differences in RNA abundance due to sex. These L3 samples were collected carefully with tweezers into an open Eppendorf tube and snap frozen within 15 minutes of collection to avoid stress, differential gene expression, or polysome collapse from lack of oxygen, starvation, or temperature change. Frozen samples were stored at -80°C until required.

2.9.1 Pestle and mortar:

An RNase-free pestle and mortar was submerged in liquid nitrogen, and left until the liquid nitrogen evaporates, at which point the pestle and mortar was placed in a bucket of dry ice, to keep the temperature extremely low. For each replicate, 0.1g of snap frozen L3 sample was transferred directly into the chilled mortar, to which 700µL of lysis buffer should be added dropwise while the sample is homogenised, using the chilled pestle. Keeping the temperature low and working quickly to integrate the lysis buffer into the homogenised sample is crucial to minimising RNA degradation and polysome collapse. Homogenised samples in lysis buffer were then transferred to RNase free 15mL falcon tubes and submerged in liquid nitrogen. The samples were stored at -80°C until required.

2.9.2 Bead beater:

A Precellys Evolution bead-beater was used to lyse some frozen whole L3 samples. Significant optimisation was carried out (as described in Chapter 3), but for final use 2mL soft tissue disruption beads were used with 0.1g of frozen larvae. To lyse the samples, 2 cycles of 7500 RPM speed disruption were used, with a 10 second pause between them. The cryolys function was used in order to keep the samples at 0°C. The Larvae Lysis Buffer for
Polysome Profiling used is described in Table 2.26. The samples were then incubated on a rolling platform at 4°C, as per section 2.10.2, and used for polysome fractionation.

2.9.3 Homogenisation:

The blades and rotor of a blade tissue homogeniser were cleaned with RNaseZap, rinsed with RNase free water, and allowed to dry. An RNase-free 15mL falcon tube containing 2100µL of lysis buffer and 0.3g of the frozen late L3 wandering male larvae was aligned with the rotary blades, in a bucket of ice, and the rotary blades were lowered into the sample, and the homogeniser turned on, and set to high speed, until the sample was visibly homogenous, and no fragments of L3 remained. This lysed sample was then incubated on a rolling platform at 4°C, as per section 2.10.2, and used for polysome fractionation.

2.10 Polysome fractionation:

Many of the methods described in this section were subject to significant optimisation (as described in Chapter 3). The protocols listed in this section are representative of the final techniques used to gather results, whereas previous versions of the techniques will be described in the relevant results section.

2.10.1 Preparing sucrose gradients:

Sucrose stocks solutions were made to a final volume of 50mL, though the initial sucrose stocks were made up in a lower volume, to allow less stable reagents to be added after the application of heat to dissolve the sucrose. Sucrose stock solutions were made up in 45mL of water at concentrations from 15% to 60% w/v, as described in Table 2.27. The RNase-free water and sucrose mixture were heated to 50°C and agitated until fully dissolved, then allowed to cool to room temperature. To these tubes, cycloheximide (CHX), dithiothreitol (DTT), and Roche cOmplete protease inhibitor cocktail (1x final concentration) were added and topped up to 50mL with more RNase-free water, to final concentrations summarised in table 2.27.

Linear gradients were produced by layering equal volumes of high and low concentration sucrose solution (15% sucrose solution upper layer, 60% sucrose solution lower layer) into SETON open-top polyallomer tubes (thin wall extra, 14mm x 95mm). The SW40Ti short cap tube marker was used to mark the point with which to fill the tube with 15% sucrose solution, using the provided sucrose syringe. Another such syringe was then filled with 60% sucrose solution, and the needle carefully inserted to the bottom of the tube, ensuring that no leakage occurs. The 60% sucrose solution was slowly ejected into the bottom of the tube, allowing it to steadily displace the 15% sucrose solution layer (with minimal mixing or disruption), and form a separate layer, with a visible meniscus between the two. The 60% sucrose solution is added until the meniscus between the two layers reaches the previously marked point on the tube,

Table 2.26 – Larvae lysis buffer for polysome profiling:

Reagent:	Concentration:
Tris HCl (pH 7.5)	10mM
NaCl	150mM
$MgCl_2$	10mM
DTT	1mM
NP40	1%
Triton X-100	60mM
Cycloheximide	14mg/mL
Turbo DNase	12U/mL
RNasein Plus	400U/replicate
Protease inhibitor	3.33µL/mL
Sodium deoxycholate	0.5%

Table 2.27 – Sucrose gradients:

Reagent:	Concentration:
Sucrose	15-60% mass/volume
NaCl	150mM
$MgCl_2$	10mM
Tris (pH7.5)	50mM
Cycloheximide	100µg/mL
DTT	1mM
Cycloheximide	33µL/mL

after which the syringe needle was slowly and carefully removed. After capping the tubes, the two layers were mixed by BioComp GradientMaster (15-60% sucrose, short cap program). After this, the gradients were left to cool overnight at 4°C before use.

Stepped gradients were made by gently pipetting layers of sucrose into the bottom of the open top polyallomer tubes, and snap freezing in liquid nitrogen, allowing a lighter layer to be pipetted over the previous layer without disruption or mixing. The volumes and sucrose concentration go from 60% sucrose solution at the lowest layer, through 50%, 47%, 42%, 34%, 26% to 18%, as previously used by Aspden *et al*., and summarised in Figure 2.5. Once completely layered, these were left to defrost and kept cool overnight at 4°C before use.

2.10.2 Sample incubation and preparation:

Lysed samples from steps covered in section 2.2 were transferred to a rolling platform at 4°C, and incubated for 30 minutes, allowing frozen samples to defrost, and thorough lysis to occur. These were then centrifuged at 3,000g at 4°C for 10 minutes to pellet the debris layer. A 200-1000 µL RNase-free pipette tip was cut to widen the opening, and used to remove as much of the visible fat layer as possible. The clear, aqueous layer was then carefully removed using a sterile needle and syringe, and transferred to a new Eppendorf tube, leaving behind the debris pellet and the majority of the remaining fat layer. This aqueous layer was then centrifuged at 21,000g at 4°c for 5 minutes, in order to pellet the nuclei. The aqueous layer was again carefully removed using a sterile needle and syringe and transferred to a new Eppendorf tube, leaving behind the nuclei pellet and the majority of the remaining fat layer. Further centrifugation at 3,000g at 4°c for 5 minutes at a time were used, in combination widened 200-1000µL RNase-free pipette tip in order to remove any fat layer that formed during centrifugation. This final fat removal step was repeated until no visible fat layer formed after centrifugation.

Once the RNA-containing layer had been depleted of lipids and nuclei, 55µL was taken from each replicate, snap frozen in liquid nitrogen, and transferred to a freezer at -80°C for storage, to be later used to extract total RNA, for comparison against polysomal RNA. Of the remaining sample, 400µL was taken for polysome fractionation. To this end,

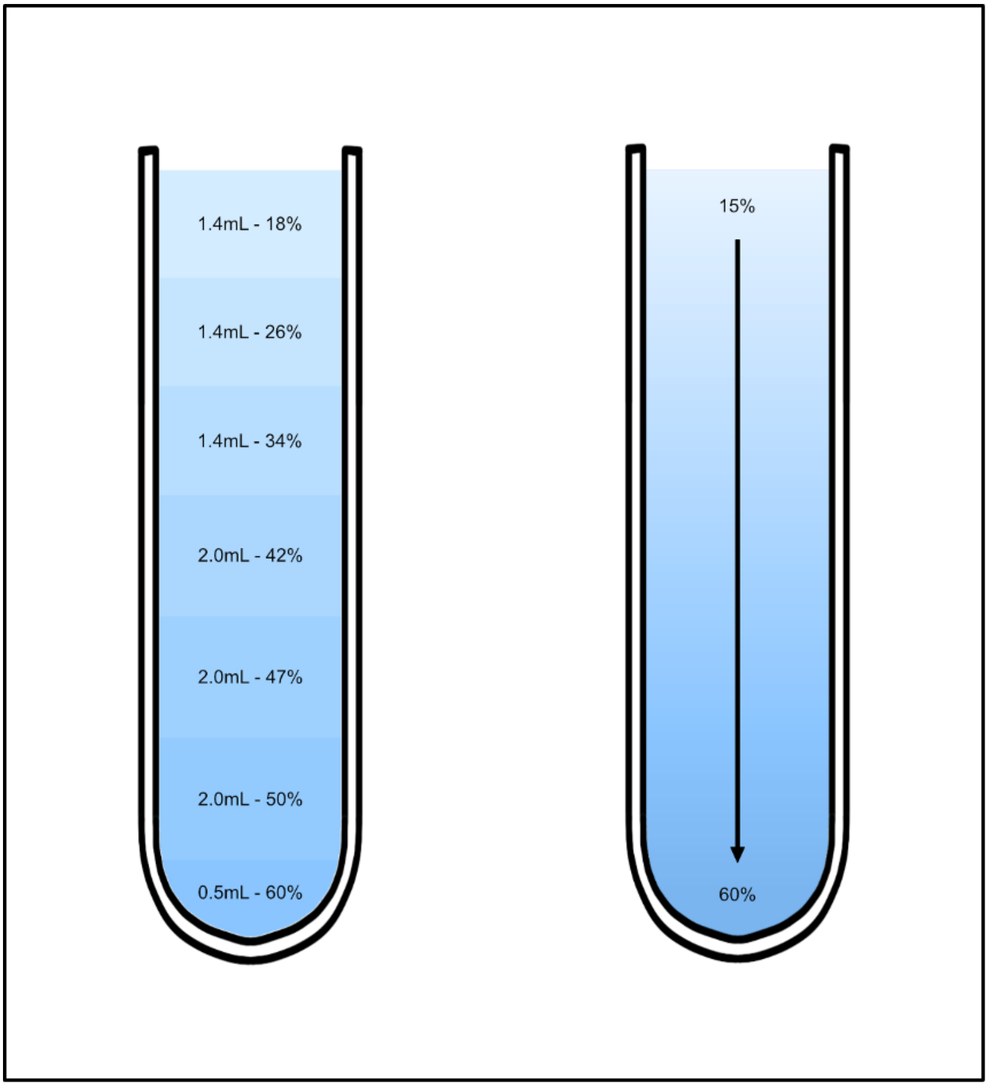


Figure 2.5 – Demonstrating the difference in composition between stepped and continuous sucrose gradients.

The updated equipment, specifically the Gradient Master Gradient Station, allowed reliable and repeatable preparation of continuous sucrose gradients. This graphic demonstrates the difference between the make up of the stepped and continuous gradients used in this project.

400µL was removed from the top of the previously prepared sucrose gradients, allowing the 400µL RNA sample to be added (slowly, dropwise running down the side of the polyallomer tube, to reduce disruption of the gradient,) in its place. These loaded gradients were labelled, and placed carefully into the Beckman SW40Ti rotor buckets, weighed, and 15% (for linear gradient) or 18% (for stepped gradient) sucrose solution added to the top of a gradient where needed to balance any opposing samples.

2.10.3 Ultracentrifugation:

The samples in the SW40Ti buckets were loaded onto the Beckman SW40Ti rotor and loaded into a Beckman ultracentrifuge. These were then spun at 121,355g (average RCF, see Table 2.28 for full ultracentrifuge conditions) at 4°C for 3 hours and 30 minutes.

2.10.4 Polysome fractionation:

Gradients were fractionated using the Triax Flow Cell system, after the system was zeroed against RNase-free water. For poly-ribo-seq or total translational RNA, a manual advance was used following the 80s ribosomal peak to ensure fractions contained polysomal RNA only. For individual polysome peaks, a manual advance was used to separate every resolvable polysome peak. Fractions were collected in 1.5mL Eppendorf tubes, and snap frozen in liquid nitrogen.

2.10.5 RNA extraction from sucrose fractions for use as PCR template:

2.10.5.1 Pooling and dilution of fraction:

Fractions containing the desired polysome and associated RNA were defrosted and pooled, and the sucrose percentage calculated across these fractions. Addition of 'Sucrose Fraction Dilution Buffer' (contents shown in Table 2.29), was added in a sufficient volume to bring the sucrose percentage to 10%, to facilitate efficient RNA extraction.

Table 2.28 – SW40Ti centrifugation speeds:

Measurement:	Force:
RCF (average)	121355
RCF (max)	170920
RPM	31000

Table 2.29 – Sucrose Fraction Dilution Buffer:

Reagent:	Concentration:
Tris HCl (pH7.5)	50mM
NaCl	150mM
$MgCl_2$	10mM

2.10.5.2 Concentration of RNA in pooled sucrose fractions:

Samples were concentrated using a MWCO Ultrafiltration concentrator and spun at 4000g at 4°C in 10-minute bursts, and flow through was discarded. This was repeated until there was 1mL of material per sample.

2.10.6 RNA extraction from concentrated samples:

To the concentrated samples, 4000µL of QIAzol lysis reagent was added and mixed with 1000µL of sample, split into 4 x 1250µL in 2mL Eppendorf tubes, and incubated for 1 hour at room temperature. Following this, 250µL of RNase-free chloroform was then added to each Eppendorf, and mixed thoroughly, before being left to stand for 3 minutes at room temperature.

These samples were then spun at 12,000g at 4°C for 15 minutes, the aqueous layer was then carefully pipetted off, and 750µL of RNase free isopropanol and 75µL of RNase-free 3M sodium acetate were added and mixed thoroughly. This was incubated at -20°C overnight to precipitate. Precipitated samples were then centrifuged at 12,000g at 4°C for 30 minutes and the supernatant was carefully removed, leaving the pellet. If no pellet could be seen, 1µL of GlycoBlue was added, the tubes vortexed, and the previous centrifugation was repeated. To the isolated pellet, 750µL of 100% RNase-free ethanol was added, vortexed, and centrifuged at 7,500g at 4°C for 5 minutes. This was then repeated using 75% RNase-free ethanol and the supernatant discarded before the pellet was left to air dry. Following this, the pellet was then resuspended in 50µL of RNase-free water.

RNA concentration was measured using the Thermo Scientific NanoDrop One Microvolume UV-Vis Spectrophotometer, through the sample's absorbance peak at 260nm. For each sample, 1µL was measured and calibrated with a blank of 1µL of RNase-free water.

2.11 Poly-Ribo-Seq:

To carry out poly-ribo-seq, samples were lysed, 55µL held back and frozen, and the remainder fractionated, as described in 2.10.1 - 2.10.5.1. Once fractionated, pooled and diluted, samples went through a series of further steps, until ready to undergo library prep.

2.11.1 RNase digest of fractions:

RNase I was added to pooled and diluted fractions at a concentration of 24U/mL. This was incubated overnight at 4°C on a rolling platform, to provide constant agitation. Following this, the RNase I was inactivated by the addition of SuperaseIN, which was thoroughly mixed with the sample, and incubated for 5 minutes at room temperature on a rolling platform, to provide constant agitation. Subsequent concentration of samples was carried out as described in 2.10.5.2.

2.11.2 Sucrose cushioning:

In order to pellet ribosome-bound RNA, ultracentrifugation through a sucrose cushion was carried out, exerting sufficient force over time to separate cellular components approximately equal to the density of a ribosome.

Open-Top 3.5mL Thickwall Polycarbonate Tube (Beckman-Coulter, catalogue number 349622) ultracentrifuge tubes were loaded with 2mL of “34% sucrose cushion solution” (components and concentrations described in Table 2.30). Once this had settled, the 1mL sample of concentrated RNA was loaded carefully on top, by pipetting gently against the internal wall of the tube, in order to avoid disrupting the separated layers. Tubes were weighed against one another and balanced using additional “34% sucrose cushion solution” where necessary. These tubes were then loaded into a TLA-100.3 tabletop ultracentrifuge rotor, and spun at 70,000rpm (264360g) for 4 hours at 4°C (average RCF, see Table 2.31 for full ultracentrifuge conditions).

Table 2.30 – 34% Sucrose Cushion Solution:

Reagent:	Concentration:
Tris HCl (pH7.5)	50mM
NaCl	150mM
$MgCl_2$	10mM
Sucrose	34%
UltraPure water	Make up to 50mL

Table 2.31 – TLA-100.3 centrifugation speeds:

Measurement:	Force:
RCF (average)	207188
RCF (max)	264360
RPM	70000

Following this, the tubes were carefully removed and examined for the visible presence of a significant pellet. The sucrose layer was carefully removed by pipetting, leaving only the pellet. To each tube, 1mL of QIAzol was added, and pipetted up and down until the pellet was entirely resuspended. This served the purpose of not only resuspending the pellet, but also stripping the ribosome (and any other bound proteins) from the RNA. The contents of each tube was transferred into an RNase-free Eppendorf, and left to incubate for 1 hour at room temperature. Following this, 250µL of RNase-free chloroform was added to each tube, and all tubes were vortexted for 15 seconds, before being left at room temperature for 3 minutes. Subsequently, the tubes were centrifuged at 12,000g for 15 minutes at 4°C to separate the aqueous and phenol phases completely. The aqueous phases were carefully removed by pipetting, and transferred to new RNase-free Eppendorf tubes. To each of these, 750µL of RNase-free isopropanol and 75µL of RNase-free 3M Sodium Acetate were added and mixed by pipetting. The mixtures were then left to precipitate at -20°C overnight. Once this step was complete, 1µL of GlycoBlue was added to each tube, and mixed thoroughly. The samples were centrifuged at 21,000g for 30 minutes at 4°C, and checked for a visible RNA pellet. If not visible, the centrifugation step was repeated.

2.11.3 Total RNA extraction from lysed L3 larvae:

At this point, the 55µL aliquots of lysed sample from 2.10.2 were defrosted, and an RNA extraction carried out as follows: First, 1mL of QIAzol was added to each sample, mixed by pipetting up and down, and transferred to clean RNase-free Eppendorf tubes. To each tube, 750µL of RNase-free isopropanol, 75µL of RNase-free 3M Sodium Acetate, and 0.5µL GlycoBlue were added, and incubated for 1 hour at -80°C. After this, the samples were centrifuged at 12000g for 15 minutes, the pellets were then washed with 750µL of 100% RNase-free ethanol and centrifuged at 7500g for 5 minutes. This step was repeated with 75% RNase-free ethanol to wash a second time. The ethanol was removed, and the pellet left to air-dry at room temperature for 5-10 minutes.

The pellet was re-suspended in "1x DNase buffer with Turbo DNase" (ThermoFisher, catalogue number AM2238), and incubated at 37°C for 30 minutes. Immediately after this, all of the prior RNA extraction steps up to drying the RNA pellet were repeated in order to re-isolate the RNA. The pellets were then re-suspended in 100µL of RNase-free water and quantified using a nanodrop spectrophotometer.

2.11.4 Bead extraction of poly-A RNA transcripts from total RNA:

First, “Bead Binding Buffer” and “RNA Fragmentation Buffer” were prepared as specified in Tables 2.32 to 2.34. The total RNA samples from Section 2.11.3 were heated to 65°C for 2 minutes, and subsequently mixed with 200µL of oligo-dT binding beads per 75ug of RNA. This mixture was incubated for 10 minutes at room temperature with constant agitation. The beads were then washed in 1 volume of “Bead Binding Buffer” and placed on a magnetic separation strip until the liquid is clear, and the beads have formed a distinct and separate pellet. The liquid was pipetted off and discarded, taking care not to disrupt the beads. The beads were resuspended in 100µL of “Bead Binding Buffer”. The amount of oligo-dT binding beads was calculated and pipetted out into Eppendorf tubes (200µL of oligo-dT beads per 75ug of RNA). The beads were placed on a magnetic separation rack until the liquid is clear, and the beads formed a distinct and separate pellet.

The liquid was pipetted off and discarded, taking care not to disrupt the beads. The beads were resuspended in 100µL of “Bead Binding Buffer” and mixed thoroughly. This mixture was incubated for 10 minutes at room temperature with constant agitation (in order to wash the beads) and separated again by use of the magnetic separation rack. The total RNA samples from 2.11.3 were heated to 65°C for 2 minutes. The “Bead Binding Buffer” used to wash the beads was carefully removed, as before, taking care not to disrupt the beads. The beads were then resuspended in 150µL of the RNA (per sample) and 100µL of “Bead Binding Buffer”. This mixture was incubated for 10 minutes at room temperature with constant agitation (in order to allow binding of polyA-RNAs to the oligo-dT beads) and separated again by use of the magnetic separation rack (for 2 minutes, or until clear). The previous wash step was repeated twice more, and after the second wash, the “Bead Binding Buffer” was carefully removed, the remainder centrifuged at 2000g for 10 seconds, and the last of the liquid then removed again. In order to elute the RNA, 50µL of 10mM Tris was added to each sample, and heated at 75°C for 3 minutes to release the selected polyA-RNA.

Table 2.32 – Bead Binding Buffer (2x conc.):

Reagent:	Concentration:
Tris HCl (pH 7.5)	10mM
EDTA	1mM
NaCl	1mM

Table 2.33 – RNA Fragmentation Buffer:

Reagent:	Volume
EDTA	2µL
Na_2CO_3	10µL
$NaHCO_3$	110µL
UltraPure Water	378µL

Table 2.34 – Urea-Acrylamide denaturing gel mix:

Reagent:	Volume/Mass:
10x TBE	5mL
40% Acrylamide	12.5mL
UltraPure Water	10mL
40% APS	125µL
TEMED	32µL
Urea	25g

The samples were once more placed on a magnet until the liquid was clear and the beads completely separated. The liquid (containing polyA selected RNA) was pipetted off, into new RNase-free Eppendorf tubes. An alcohol precipitation was carried out as described in 2.11.2. After the pelleting step, the RNA was resuspended in 10µL of "RNA chemical fragmentation buffer" (as described in Table 2.33). The RNA and buffer were thoroughly mixed by pipetting and heated to 95°C for 20 minutes. Immediately after this, the samples are cooled on ice, quick-centrifuged for 5 seconds to ensure all sample was at the bottom of the tube, and then placed back on ice to cool. Another alcohol precipitation was carried out as described in 2.11.2.

2.11.5 Denaturing gel purification of RNA fragments:

A "Urea-Acrylamide denaturing gel mix" (Table 2.34) was made up and poured into an RNase-cleaned (IMS, followed by RNase-zap, and rinsed with Millipore water) large vertical Bio-Rad gel tank, before carefully adding the comb. This was left to set for half an hour, then loaded into its running tank and filled with 1xTBE. The tank and gel were pre-run at 300V for 30 minutes, before rinsing out the wells with 1xTBE to displace excess acrylamide polymers. All suspended samples of RNA from 2.11.2 (RNase-digested polysomal RNA after sucrose cushioning) and 2.11.4 (fragmented, polyA-RNA from total RNA) were pelleted and resuspended in 30µL of formamide dyes. These were mixed well, and 15µL of each were loaded into a well of the denaturing Urea-Acrylamide gels, alongside Low Molecular Weight Gel Ladder, and custom markers for 28 and 34 base pair products. The gel was run for 2 hours and 45 minutes at 300V. Once run, the gel was carefully transferred to a foil-wrapped flat tray (in order to keep out light) and soaked in 200mL of 1xTBE and 20µL of SybrGold. This was placed on a rocking platform for 30 minutes to allow for constant agitation to mix and soak the gel in the SybrGold-mixed buffer. The gel was viewed on a DarkReader, and the desired bands based on the expected products of RNase-digest and chemical fragmentation (polysome footprint = 28-34nt, mRNA fragments = 50-80nt) were excised with a sterile, RNase-free scalpel. Excised gel slices were placed into pre-prepared shredder tubes (0.5mL RNase-free Eppendorf tubes, perforated 5 times at the base with a sterile and RNase-free syringe needle). The tubes were placed into larger, 2mL tubes, and centrifuged at 3000g for 2 minutes at a time until all gel had passed through the shredder tubes. To each of these, 750µL of “Acrylamide Gel Elution Buffer” (as described in Table 2.35) was added and mixed thoroughly by vortexing. These were placed on a rotating Eppendorf rack, and left at 4°C overnight.

2.11.6 Recovering RNA from Urea-Acrylamide gel elution:

Wide-opening p1000 pipette tips were prepared by trimming RNase-free p1000 pipette tips 5mm from the narrow opening with a sterile, RNase-free scalpel. These tips were used to transfer the entire gel and elution buffer mix from 2.11.5 into Spin-X filter tubes (Corning 8161). These tubes were centrifuged at 12,000g in 2 minute bursts until all liquid had passed through the filters. Another alcohol precipitation was carried out on this buffer, as described in 2.11.2. The washed and dried RNA pellets were resuspended in 20µL of 10mM Tris pH7.5 and mixed well.

2.11.7 T4 Phosphonucleotide Kinase treatment of RNA samples:

To each sample, 67µL of RNase-free water was added, and heated to 80°C for 2 minutes. To these, 10µL of 10x T4 PNK buffer, 2µL of T4 PNK enzyme, and 1µL of SuperRNasin were added, and mixed by gentle pipetting. This mixture was incubated at 37°C for 1 hour, then transferred to 70°C for a further 10 minutes to inactivate the T4 PNK enzyme. Another alcohol precipitation was carried out on the mixture, as described in 2.11.2, and the pellet resuspended in 6µL of RNase-free water.

2.11.8 Modifications to NEBNext Small RNA Library Prep Kit from Illumina’s provided protocol:

Steps 1-4, “Ligate the 3’ SR Adaptor” were carried out exactly as described according to the “NEBNext Multiplex Small RNA Library Prep Set for Illumina” protocol. From here, the samples from total RNA were placed on ice, while rRNA depletion was carried out on the

polysomal RNA samples (as they had not gone through polyA selection. For each sample, 25µL of “500bp rRNA bead mix” (available as a custom reagent from the Couso lab, Table 2.36), 25 µL of 1kbp of rRNA bead mix (available as a pre-made custom

Table 2.35 – Acrylamide Gel Elution Buffer

Reagent:	Concentration:
NaAOC	300mM
EDTA	1mM
SDS	0.25%
Nuclease-free water	Dilute to correct concentration

reagent from the Couso lab, Table 2.36), and 12.5µL of Streptavidin magnetic beads, were pipetted into a clean RNase-free Eppendorf, and mixed well. This mixture was placed on a magnetic separation strip until the liquid is clear, and the beads have formed a distinct and separate pellet. The liquid was discarded, and the bead pellet was resuspended in 150µL of "1x Bind and Wash Buffer" (as described in Table 2.37) and mixed well. This mixture was separated again by use of the magnetic separation rack.

To each polysomal RNA sample, 1µL of "rRNA Oligo Depletion Mix" (available as a custom reagent from the Couso lab, Table 2.38) was added, and mixed well by pipetting. The "1x Bind and Wash Buffer" was removed from the magnetically separated beads, and the beads were resuspended in the polysomal RNA and "rRNA Oligo Depletion Mix" mixture. These were heated to 70°C for 2 minutes (to unbind the RNA from the beads), then incubated with constant agitation for 20 minutes at room temperature.

During this incubation, a second mixture of 25µL of "500bp rRNA bead mix", 25µL of 1kbp of rRNA bead mix, and 12.5µL of Streptavidin magnetic beads was prepared for each sample, and the beads washed in 150µL of "1x Bind and Wash Buffer", as previously described.

The incubating samples and beads were placed on a magnetic separation strip until the liquid is clear, and the beads have formed a distinct and separate pellet. The liquid was removed and discarded from the previously prepared second set of bead mixture, and these beads were re-suspended in the liquid taken from the heated and incubated beads along with another 11µL of "rRNA Oligo Depletion Mix", in order to carry out a second round of depletion. This was washed, separated, heated, and incubated as previously described, and the liquid phase separated by use of a magnetic separation rack. This liquid phase for each sample (now containing polysomal RNA not depleted in the common rRNA regions that would bind the oligos and beads) was taken aside, and another alcohol precipitation carried out, as described in 2.11.2. The pellet was resuspended in "Replacement Adaptor Ligation Buffer" (as described in Table 2.39), in order to bring these rRNA depleted samples back to the conditions they were in following step 4 of the library prep protocol. All subsequent steps (5+) for library prep

Table 2.36 – rRNA depletion primers:

Primer:	Sequence (5’ to 3’):
2S Reverse	biotin-TACAACCCTCAACCATATGTAGTCCAAGCA
5S Forward	GCCAACGACCATACCACGCT
5S Reverse	biotin-AAAAAGTTGTGGACGAGGCC
5.8S Forward	AACTCTAAGCGGTGGATCAC
5.8S Reverse	biotin-CAGCATGGACTGCGATATGCG

Table 2.36 (cont.) – rRNA depletion primers (Set 1):

18S Forward (A)	ATTCTGGTTGATCCTGCCAG
18S Reverse (A)	biotin-CAAGAATTTCACCTCTCGCGT
18S Forward (B)	GACCGTCGTAAGACTAACTT
18S Reverse (B)	biotin-TAATGATCCTTCCGCAGGTTC
28S Forward (A)	TTATATACAACCTCAACTCAT
28S Reverse (A)	biotin-AAGTATAGTTCACCATCTTTC
28S Forward (B)	GATCAGGTTGAAGTCAGGGG
28S Reverse (B)	biotin-CATGCTCTTCTAGCCCATCTA
28S Forward (C)	ACATATACTGTTGTGTCGATA
28S Reverse (C)	biotin-AAATACATAAATGCATCGTTT
28S Forward (D)	TTGATTTGAAAATTTGGTATA
28S Reverse (D)	biotin-TCGAATCATCAAGCAAAGGAT

Table 2.36 (cont.) – rRNA depletion primers (Set 2):	
18S Forward (A)	CCGAGGCCCTGTAATTGGAAT
18S Reverse (A)	biotin-ATATGAGTCCTGTATTGTTATTTT
18S Forward (B)	ATTGTGTTTGAATGTGTTTATGTAAG
18S Reverse (B)	biotin-AAGCATTTTACTGCCAACATGAAT
28S Forward (A)	ATATAAGGACATTGTAATCTATTAGC
28S Reverse (A)	biotin-GGAAAAAATGCACACTATTCTCAT
28S Forward (B)	GCGCTTAAGTTGTATACCTATAC
28S Reverse (B)	biotin-CATCCATTTTAAGGGCTAGTTG
28S Forward (C)	GCGGGTGTTGACACAATGTGA
28S Reverse (C)	biotin-TAGGGCCATCACAATGCTTTGT
28S Forward (D)	CAAAACGTTGTTGCGACAGCA
28S Reverse (D)	biotin-TCATTAGTAGGGTAAAACTAACC

Set 1 was used to generate 1 Kb fragments and Set 1 and Set 2 were used in combination to generate 500 bp fragments

Table 2.37 – 1x Bind and Wash Buffer

Reagent:	Concentration:
Tris HCl (pH 7.5)	5mM
EDTA	0.5mM
NaCl	2mM
Nuclease-free water	Dilute to correct concentration

Table 2.38 – rRNA Oligo Depletion Mix:

28S	
1	GGGTAGTCCCATATGAGTTGAGGTTG
2	ATTGTGGAACTTTCTTGCTAAAATTTTTAAGA
3	TATAAACTTTAAATGGTTTAGAAGCCATACAATGC
18S	
4	CGCTTGGTTTTAGCCTAATAAAAGCACAC
5	ATACGATCTGCATGTTATCTAGAGTTCAACCAATA
6	GGGACAAACCAACAGGTACGGCTCCACTTAC

Oligos 1-6 were resuspended to 100µM and mixed in the ratio 2:3:2:2:3.25:1.75 (determined by the number of reads in a previous sucrose cushion run) to make 14µL of oligo mixture, subsequently siluted 1:1 with nuclease-free water to make a 50µM mix.

were carried out exactly as specified in the “NEBNext Multiplex Small RNA Library Prep Set for Illumina” (catalogue number E7300S) protocol.

2.11.9 Gel Size Selection:

The “Acrylamide non-denaturing gel mix” was made up (as described in Table 2.40), and poured into an RNase-cleaned (IMS, followed by RNase-zap, and rinsed with Millipore water) large vertical Bio-Rad gel tank, before carefully adding the comb. This was left to set for half an hour, then loaded into its running tank, and filled with 1xTBE. The tank and gel were pre-run at 300V for 30 minutes, before rinsing out the wells with 1xTBE to displace excess acrylamide polymers. To all library-prepped samples, 30µL of 1x TBE was added, and samples were loaded into wells, alongside Low Molecular Weight Ladder. The gel was run for 2 hours and 45 minutes at 200V. Once run, the gel was carefully transferred to a foil-wrapped flat tray (in order to keep out light) and soaked in 200mL of 1xTBE and 20µL of SybrGold. This was placed on a rocking platform for 30 minutes to allow for constant agitation to mix and soak the gel in the SybrGold-mixed buffer. The gel was viewed on a DarkReader, and the desired bands based on the expected products of RNase-digest and chemical fragmentation with 3’ and 5’ adaptors added (polysome footprint = 155-161nt, mRNA fragments = 177-227nt) were excised with a sterile, RNase-free scalpel.

Excised gel slices were placed into pre-prepared shredder tubes (0.5mL RNase-free Eppendorfs, perforated 5 times at the base with a sterile and RNase-free syringe needle). The tubes were placed into larger, 2mL tubes, and centrifuged at 3000g for 2 minutes at a time until all gel had passed through the shredder tubes. To each of these, 750µL of “Acrylamide Gel Elution Buffer” (Table 2.35) was added and mixed thoroughly by vortexing. These were placed on a rotating Eppendorf rack, and left at 4°C overnight. The next day, another alcohol precipitation was carried out, as described in 2.11.2, and each pellet was resuspended in 13µL of TE buffer.

Table 2.39 – Replacement Adator Ligation Buffer

Reagent:	Volume:
10x NEB T4 RNA Ligase truncated buffer	2µL
PEG 8000	4µL
Rnase Inhibitor	1µL
Nuclease-free water	13µL

Table 2.40 – Acrylamide non-denaturing gel mix:

Reagent:	Volume:
10x TBE	5mL
40% Acrylamide	10mL
UltraPure Water	35mL
40% APS	125µL
TEMED	32µL

2.11.10 Quality control using Bioanalyser:

In order to ensure that RNA fragments were present at the right size, and at the right concentration, 1µL of each sample was tested using a Bioanalyser High-Sensitivity DNA Chip, as per the manufacturer provided by the manufacturer. Based on the RNA sizes and concentrations given, the Gel Size Selection step (2.11.9) was repeated in some instances, and once all samples were satisfactory, they were pooled as advised by Leeds University RNA-sequencing facility.

2.12 External services:

Samples were sent for sequencing at external facilities that the Newbury lab has previously had success in using.

2.12.1 Eurofins DNA sequencing:

Basic DNA sequencing was carried out by Eurofins TubeSeq Service, with samples sent at concentrations and volumes advised by the Eurofins TubeSeq submission guide.

2.12.2 Illumina NextSeq sequencing:

Illumina NextSeq was carried out by the University of Leeds in house RNA-sequencing facility. Samples were submitted in pooled mixtures at concentrations and volumes as advised by staff from the facility.

2.13 Primers

All primers used in the project are listed in Table 2.41.

Table 2.41 – Primer catalogue:

Primer:	Sequence (5’ to 3’):
HsrOmega-Cloning F	CACCCTCTCGAAAACTGAACATTA
HsrOmega-Cloning R	ATCTTTCAAAATCCGCAGGT
CR40469-Cloning F	CACCCACGTTCCTCACTAATTGTG
CR40469-Cloning R	ATCAATTTTCATCAATTCAC
Pacman CRISPR sequencing primer F	ACGGTTCCTTTGCAGATACCC
Pacman CRISPR sequencing primer R	AGGCAGTGTCGTGTATTGGG
Rp49 F	CCAGTCGGATCGATATGCTAA
Rp49 R	TCTGCATGAGCAGGACCTC
BglII sequencing primer (for pAc-sgRNA-Cas9) F	CGTATTTCAGGCTGCAAGTCGAAC
BglII sequencing primer (for pAc-sgRNA-Cas9) R	AAAAAAGCACCGACTCGGTGC
CRISPR-42F	GTCGGAAACGTCAAGCGTCAACG
CRISPR-42R	AAACCGTTGACGCTTGACGTTTCC
sgRNA XRN1 primer F	AACCGCGCGCCGTCCGGAATCGC
sgRNA XRN1 primer R	GCGCGCGGCAGGCCTTAGCGCTT
sgRNA Dis3L2 primer F	AACCGTTGACGCTTGACGTTTCC
sgRNA Dis3L2 primer R	GCAACTGCGAACTGCAAAGGCTT
CR42719 F	GCAAAGCGACAAAGAGCGAG
CR42719 R	ATCTCCAAAACTCGGCCTCC

Table 2.41 (cont.) – Primer catalogue:	
Primer:	Sequence (5’ to 3’):
CR6900 F	GGTGACGAGATGTCCAAGCA
CR6900 R	AGTAACTTGGAGGTTCAGTGC
CR45177 F	CGGTTATGTGGGGTGATGGT
CR45177 R	TTGGGGTGCCAAGATTGTCT
CR43635 F	GCATATGTGCACATCTCTCC
CR 43635 R	GTCCCACAATTGCTGTTGCT
CR44677 F	TGGCTTTGGTCATGAGTCCC
CR44677 R	GTGGCACTAGAACTGTGGCT
CR43466 F	GAATTCCTTCGGACCTGGCA
CR43466 R	ACCAAAATCCCGCGTCTTCT
CR43260 F	TCATTGCGAGCAGGTTATTCC
CR43260 R	GACACGCGGCTTATAGGTGA
CR44587 F	TTGCCGAGTGCTCTCCATTT
CR44587 R	CGATCGACGATGGAGCTTGA
Uhg8 F	AAAGCTGACCGTATGGGCTC
Uhg8 R	TGTTAAAAGATTGCGATTTAG CACG
CR44367 F	ACATACGCTTGGGGTGCTAA
CR44367 R	ATGAGCTGGAGTCCCTTTGC
CR34006 F	GTCAGTCAGCCTCCGATTCC
CR34006 R	GGCCATATACTCGACTGGGC
CR18217 F	TCGATCGAAGCGACGTG
CR18217 R	GAATTTCGGCCATCGACAGG

Table 2.41 (cont.) – Primer catalogue:

Primer:	Sequence (5’ to 3’):
CR42719 TaqMan qPCR F	GCACGTCGACCACATATTCCA
CR42719 TaqMan qPCR R	GCTGCAGTCGACGTCTTCA
CR42719 TaqMan qPCR Probe	CCGGCGTCACTCTTT
CR45177 TaqMan qPCR F	CGGACAACGGACCTCATATGTG
CR45177 TaqMan qPCR R	GCTGGATTATTAGCGGTCCGAATTT
CR45177 TaqMan qPCR Probe	ACGATTATGGTAGATGATCCTC
CR6900 TaqMan qPCR F	AGGCGCTGATCAAGGATGTC
CR6900 TaqMan qPCR R	CCACTCCCGCCTGAAGTTC
CR6900 TaqMan qPCR Probe	CACGGCAAAATTGTTG
Dis3L2 dsRNA primer 1 F	TAATACGACTCACTATAGGCGAACCCAACCAAACTCTGT
Dis3L2 dsRNA primer 1 R	TAATACGACTCACTATAGGACGCAGATCCTCTTGGCTTA
Dis3L2 dsRNA primer 2 F	TAATACGACTCACTATAGGTTGCTCGCAATTTGCTTATG
Dis3L2 dsRNA primer 2 R	TAATACGACTCACTATAGGGCTGACTTAGGCAGGCATTC
Pacman dsRNA primer F	TAATACGACTCACTATAGGGAGATGAACTGATCGAGGAACTGTGCC
Pacman dsRNA primer R	TAATACGACTCACTATAGGGAGACCAGCTGGCGCTTGCG
HsrOmega-Flag F	CACGCGGCCGCCTCTCGAAAACTGAACATTA
HsrOmega-Flag R	CGCACTAGTCTACTTGTCGTCATCGTCTTTGTAGTCATCTTTCAAAATCCGCAG
CR40469-Flag F	CACGCGGCCGCCACGTTCCTCACTAATTGTG
CR40469-Flag R	CGCACTAGTCTACTTGTCGTCATCGTCTTTGTAGTCATCAATTTTCATCAATTXCAC

Chapter 3: Results – Exploring depletion of Pacman and Dis3L2 levels in *Drosophila* cell culture and tissues to examine the degradation and translation of lncRNAs in *Drosophila*

3.1 Introduction

With most existing work on RNA decay focusing on the canonical degradation of highly abundant RNAs; the regulation of lowly expressed and poorly annotated genes like lncRNAs is lacking by comparison. Despite this, many useful resources do already exist that can be re-examined and re-applied in a different context, in order to shed light upon the mechanisms of decay for lncRNAs.

Any dataset produced by next generation, high-throughput sequencing techniques (including RNA-seq, ribo-seq, and poly-ribo-seq, among others) contains a wealth of information. In order to meaningfully interpret this data to answer specific questions, and provide relevant information to be examined, bioinformatic filters must be carefully designed and applied. Although this allows us to make use of almost unfathomably huge datasets, the majority of the raw information is not used in any single comparison or analysis. However, by trawling existing literature and datasets, datasets with the potential to help with a current question can be identified, downloaded, and (with substantial work) re-interpreted using a different framework of relevance.

This project aimed to explore the degradation and translation of lncRNA in *Drosophila*, and the roles of exoribonucleases Pacman and Dis3L2 within this. As such, to begin the project, a thorough search of high-throughput sequencing was carried out to allow for proof of principle, identification of trends, and to help with planning and carrying out a novel experiment.

In order to examine the regulation of lncRNA by Pacman and Dis3L2, three main papers were identified as being of the most use. Two previous papers by the Newbury lab (Jones *et al*. (33) and Towler *et al*. (46)) had used RNA-sequencing to categorise the decay targets of these enzymes in *Drosophila* wing imaginal discs (WIDs), with a focus on mRNAs and RNAs with known roles within biological pathways and processes. This had been carried out using comparison between the abundance of RNAs in the WID in wild-type control larvae, Pacman null mutant larvae, and Dis3L2 null mutant larvae. These provided a great source of data to re-examine in the context of lncRNA decay in these tissues.

Alongside this, previous work by Antic *et al*. (69) had carried out RNA-sequencing on *Drosophila* S2 cells under normal conditions, and with dsRNA knockdown of Pacman. Likewise, this provided another promising investigation of lncRNA decay by Pacman, in a cell line, using an incomplete knockdown rather than the complete depletion that a null mutant provides.

Along with reanalysis of these data to begin the examination of degradation, other poly-ribo-seq and ribo-seq data from *Drosophila* cells and tissues were taken to examine the translational activity of lncRNAs in *Drosophila* (90, 153). This would not allow a full examination of both translation and degradation together, as no existing Pacman or Dis3L2 depleted models had undergone polysome-sensitive sequencing. Despite this, individual candidates may be examined using different data sets, thus providing a strong starting point for experimental validation. Reanalysis of these data sets could also provide proof of principal that certain lncRNAs may be specifically degraded by Pacman and Dis3L2, and also that this degradation could be carried out on the polysome.

3.2 Project background and aims

This chapter aims to use re-analysis of existing datasets as well as experimental validation in order to prove principles important to this project, and to take an initial look at the degradation of lncRNAs by Pacman and Dis3L2. Whilst a more comprehensive analysis of multiple datasets will follow in Chapter 5, and allow more meaningful conclusions to be drawn, independent examination of the existing data can highlight the targets necessary to justify the rest of the project. As such, this chapter will not only construct an initial overview of the role of these exoribonucleases; but will be crucial in framing and designing future experiments.

The main stages of this work are as follows:

1) Carry out the initial, necessary steps to process the data in a way that allows differential abundance analysis of lncRNAs.
2) Use Antic *et al.* data (69) to examine and identify lncRNA targets of Pacman in *Drosophila* S2 cells.
3) Use Newbury lab data (33, 46) to examine and identify candidate lncRNA targets of Pacman and Dis3L2 in *Drosophila* wing imaginal discs.
4) Carry out preliminary overview of lncRNAs present on the ribosome or polysome in *Drosophila* models using existing data from the Couso lab (90, 153) and Zhang *et al*. (154) data.
5) Validate promising examples using molecular techniques.

3.3 Analysis of previous data by Antic et al. identified potential lncRNA targets of Pacman in *Drosophila* S2 cells

A previous paper, "General and microRNA-mediated mRNA degradation occurs on ribosome complexes in Drosophila cells" by Antic et al. (69) uses several molecular methods, including ribosome affinity purification, luciferase assays, and RNA-seq to explore the hypothesis that mRNA degradation takes place on the ribosome in *Drosophila*. As part of the investigation, an RNA-seq dataset was produced from Pacman-depleted *Drosophila* S2 cells, with libraries prepared both from total cell lysate, and ribosome pulldown.

The dataset provided by Antic et al. used ribo-seq to examine ribosome association with Pacman degraded transcripts in *Drosophila* S2 cells. The authors isolated decapped mRNA degradation intermediates from ribosome complexes and performed high-throughput sequencing analysis on them. The sequencing was carried out RNA derived from S2 cells; both in untreated and dsRNA induced Pacman knockdown conditions. In both conditions, RNA from ribosome complexes and total cell lysate were extracted, processed, and sequenced. Although the authors carried out these experiments for different purposes, and with no interest in lncRNAs, this work provided valuable information for evaluating the degradation of certain lncRNAs (in S2 cells) by Pacman.

This work also provided data for ribosome associated RNAs (similar to that gained by fractionation followed by ribo-seq), though the ribosome bound samples being decapped Pacman degradation fragments. This was done as Pacman is the enzyme responsible for degradation of decapped RNA, as well as actively encouraging the decapping process; only decapped fragments will bind the 5' linker, and only fragments with both the 3' and 5' linkers will be amplified by the library preparation PCR steps. This chapter instead focuses on how Pacman can degrade lncRNAs in S2 cells by examining sequencing from untreated control and Pacman knockdown S2 cell total cell lysate. By examining those lncRNA that increase in abundance with the depletion of Pacman, candidates for direct degradation by Pacman can be identified.

There are some limitations to using this dataset for the purposes of this thesis. The data from this paper explores the link between translation and RNA degradation by Pacman only, and in cell culture rather than whole organism or tissue. In addition, due to using ribosome pulldown (using a tagged version of the RpL10Ab component of the 60S ribosomal subunit,) rather than polysome fractionation, some RNAs that are not actively translated may be protected by putatively bound ribosomes or ribosomal subunits, essentially generating false positives. Despite this, the value of the dataset was identified, and the sequencing data from total lysate of untreated and Pacman knockdown S2 cells was acquired and re-analysed in order to provide a preliminary insight into lncRNAs in *Drosophila* that are differentially abundant or differentially translated in the absence of Pacman, as well as providing a proof of principle to validate and justify a larger and more complex poly-ribo-seq experiment.

The unaligned .bam files from the supplemental data of Antic et al. were downloaded and converted into .fastq files. A .fasta file of the *Drosophila melanogaster* genome (Flybase release 6.18) was used to build an index to align the reads to with the HiSat2 algorithm. All 4 *Drosophila* chromosomes were used. This index file was used to align the reads, using HiSat2. CuffLinks, CuffMerge, CuffQuant, and CuffDiff were used to produce a spreadsheet of normalised read counts for each condition (Summarised in Figure 3.1). Read counts were normalised by total reads per sample and by gene length to give FPKM, and these were extracted in Pacman knockdown and untreated control. From these, the abundance in Pacman knockdown cells were compared to the abundance in the untreated cells. Those with a fold change equal than or greater than

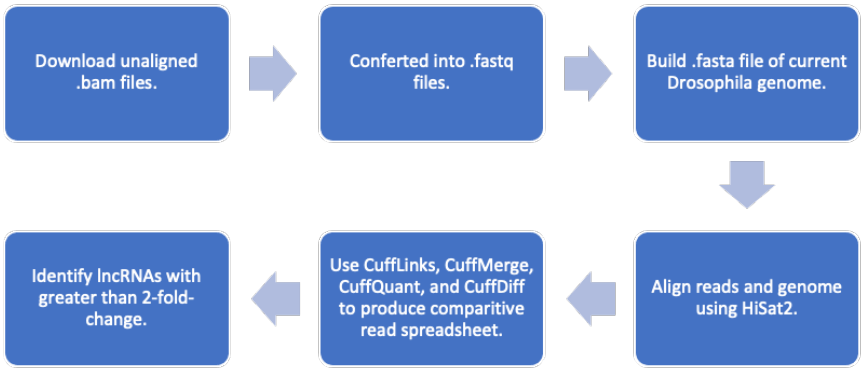

Download unaligned .bam files.
Conferted into .fastq files.
Build .fasta file of current Drosophila genome.
Align reads and genome using HiSat2.
Use CuffLinks, CuffMerge, CuffQuant, and CuffDiff to produce comparitive read spreadsheet.
Identify lncRNAs with greater than 2-fold-change.

Figure 3.1 – Summary of workflow for re-analysing Antic et al. sequencing data:

The flow diagram summarises the necessary steps to download data, map to a genome, and make comparative analysis.

two-fold (an arbitrary cut off, chosen to increase confidence in the increased abundance in data with low replicate counts) were extracted and plotted (Figure 3.2).

This process produced a shortlist of 14 down-regulated lncRNAs, and 8 up-regulated lncRNAs (Figure 3.3), all of these differentially abundant by two-fold or more in *Drosophila* S2 cell total lysate. Those that increase in abundance are of greater interest in this study as they are potential direct RNA targets of Pacman that may be stabilised in its absence. These candidates were profiled according to available FlyBase data. Very little data were available for these lncRNA (as non-annotated lncRNAs, this is not unusual,) with only one target, *CR44371*, having been suggested by experimental evidence to have a biological role. The knockout of *CR44371* has been suggested to cause defects in spermatogenesis, a process in which Pacman itself is involved in (155). From the available modENCODE sequencing data, the candidates were lowly expressed to not detected at all throughout tested timepoints and tissues (although the age and depth of some of the sequencing data means that lowly but significantly expressed transcripts may not be detected at all).

Interestingly, the majority of those passing the fold-change cutoff implemented were downregulated in Pacman mutants, the opposite of what might be initially expected from the depletion of an RNA decay enzyme. This could be due to several reasons: firstly, this experiment used dsRNA depletion of Pacman, rather than a null mutation or other means of guaranteeing a complete knockout of Pacman. As shown in Figure 3.4, panel (a), the knockdown achieved by Antic *et al*. (69) is clearly strong (though not quantified), but is not a complete knockdown, as some Pacman still shows up on the Western blot. Therefore, some Pacman-regulated RNAs may not see substantial change in abundance whilst there is still functional Pacman to carry out their degradation.

Secondly, indirect regulation by Pacman is a strong possibility; this can occur when transcripts that are directly targeted and regulated by Pacman degradation are themselves involved in downstream pathways that can subsequently feed back into other mechanisms of control for RNA abundance and gene expression. For example, the knowledge that Pacman activity is needed to maintain and balance normal apoptotic signalling through *reaper* and *hid*, informs us that misregulated apoptosis (along with

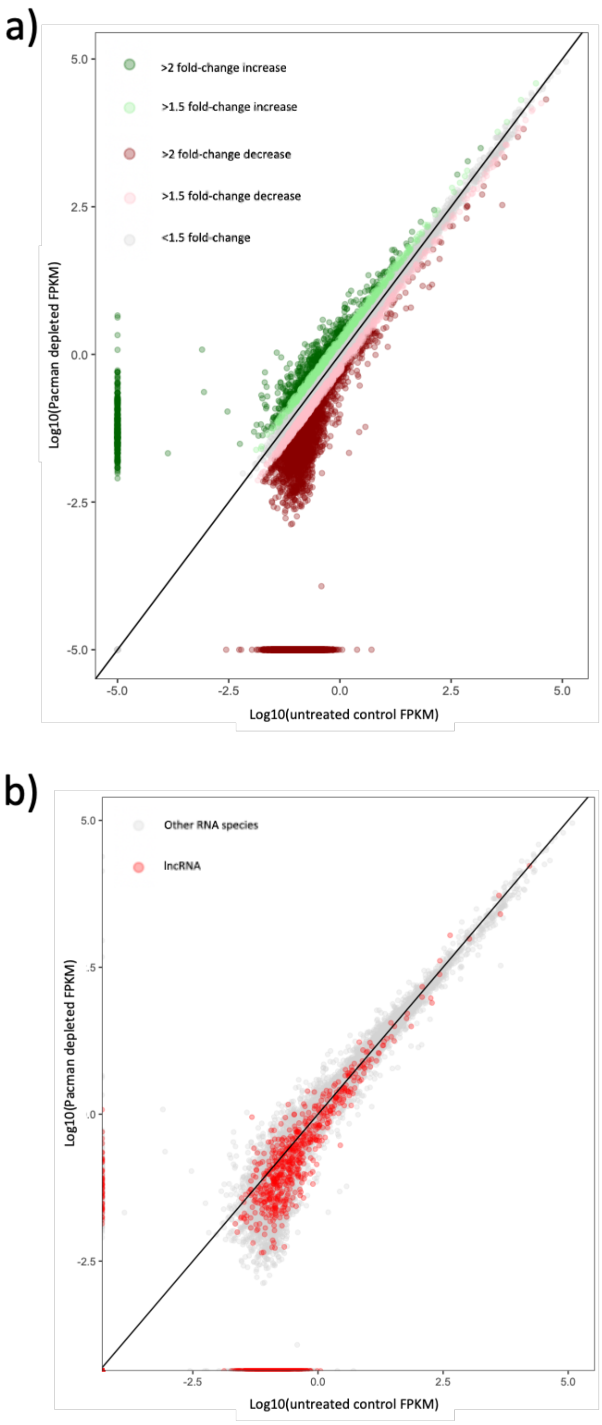
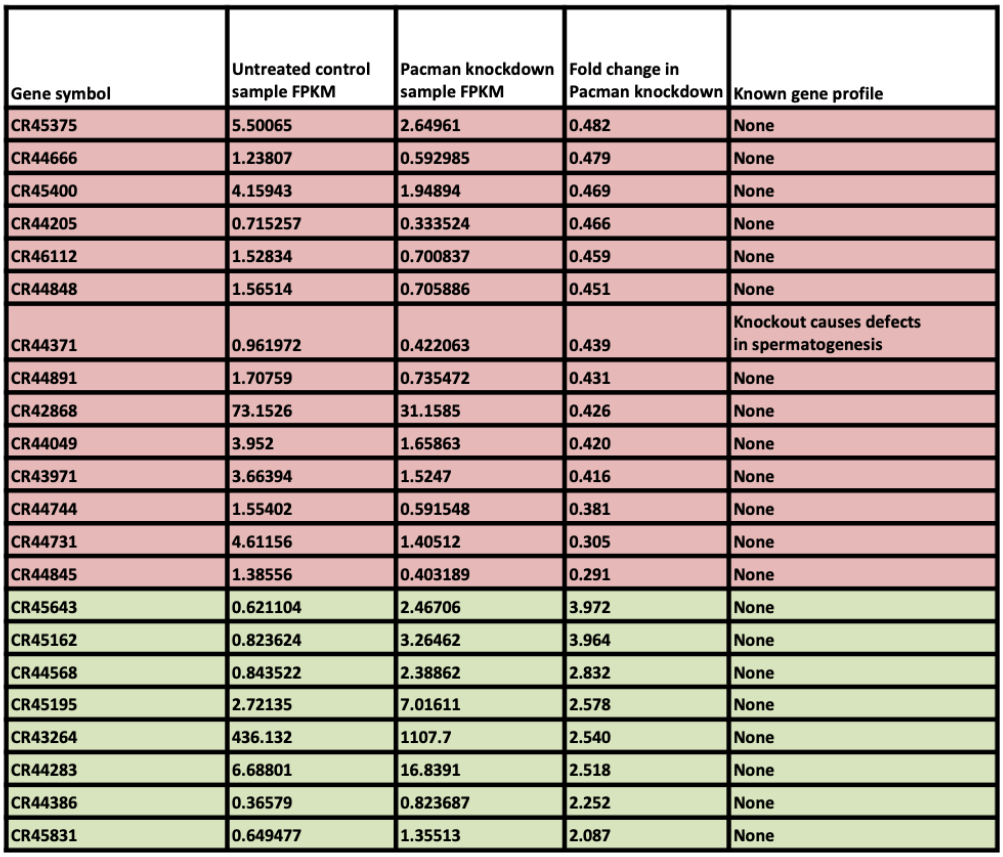


Figure 3.2 – RNAs shown by new analysis of the data of Antic et al. to be more than two-fold differentially abundant in Pacman depleted cells

(a) The log10 FPKM of Pacman depleted cells is plotted against log10 FPKM of wild untreated S2 cells. Upregulated RNAs are highlighted in green, and downregulated RNAs in red, with darker and lighter shades used to highlight >2 or >1.5 fold-change, respectively. Any smaller changes were left in light grey. Values of "0" were replaced with 0.0001 (lower than any other non-zero value) to prevent infinite values, and a cut off for the plot of log10(FPKM)<-4 was used to plot the graph at a meaningful scale.

(b) The log10 FPKM of Pacman depleted cells is plotted against log10 FPKM of wild untreated S2 cells again, now with lncRNA species highlighted in red, while other RNA species remain plotted in light grey. Values of "0" were replaced with 0.0001 (lower than any other non-zero value) to prevent infinite values, and a cut off for the plot of log10(FPKM)<-4 was used to plot the graph at a meaningful scale.



Gene symbol	Untreated control sample FPKM	Pacman knockdown sample FPKM	Fold change in Pacman knockdown	Known gene profile
CR45375	5.50065	2.64961	0.482	None
CR44666	1.23807	0.592985	0.479	None
CR45400	4.15943	1.94894	0.469	None
CR44205	0.715257	0.333524	0.466	None
CR46112	1.52834	0.700837	0.459	None
CR44848	1.56514	0.705886	0.451	None
CR44371	0.961972	0.422063	0.439	Knockout causes defects in spermatogenesis
CR44891	1.70759	0.735472	0.431	None
CR42868	73.1526	31.1585	0.426	None
CR44049	3.952	1.65863	0.420	None
CR43971	3.66394	1.5247	0.416	None
CR44744	1.55402	0.591548	0.381	None
CR44731	4.61156	1.40512	0.305	None
CR44845	1.38556	0.403189	0.291	None
CR45643	0.621104	2.46706	3.972	None
CR45162	0.823624	3.26462	3.964	None
CR44568	0.843522	2.38862	2.832	None
CR45195	2.72135	7.01611	2.578	None
CR43264	436.132	1107.7	2.540	None
CR44283	6.68801	16.8391	2.518	None
CR44386	0.36579	0.823687	2.252	None
CR45831	0.649477	1.35513	2.087	None

Figure 3.3 – Re-analysis of data by Antic et al. provides a novel shortlist of candidate lncRNAs that may be specifically regulated by Pacman:

The list shows 14 down-regulated lncRNAs, and 8 up-regulated lncRNAs, in Pacman depleted S2 cells vs control S2 cells. Mean sample values are listed, along with p-values from statistical analysis, and any known biological profile. Statistical analyses should not be read as truly significant, as only one replicate was analysed for each condition.

Interestingly a majority of those passing the fold-change cutoff implemented were downregulated in Pacman, the opposite of what might be initially expected from the depletion of an RNA decay enzyme. Almost no biological information was available on the candidate genes, as is common for poorly annotated genes such as lncRNAs.

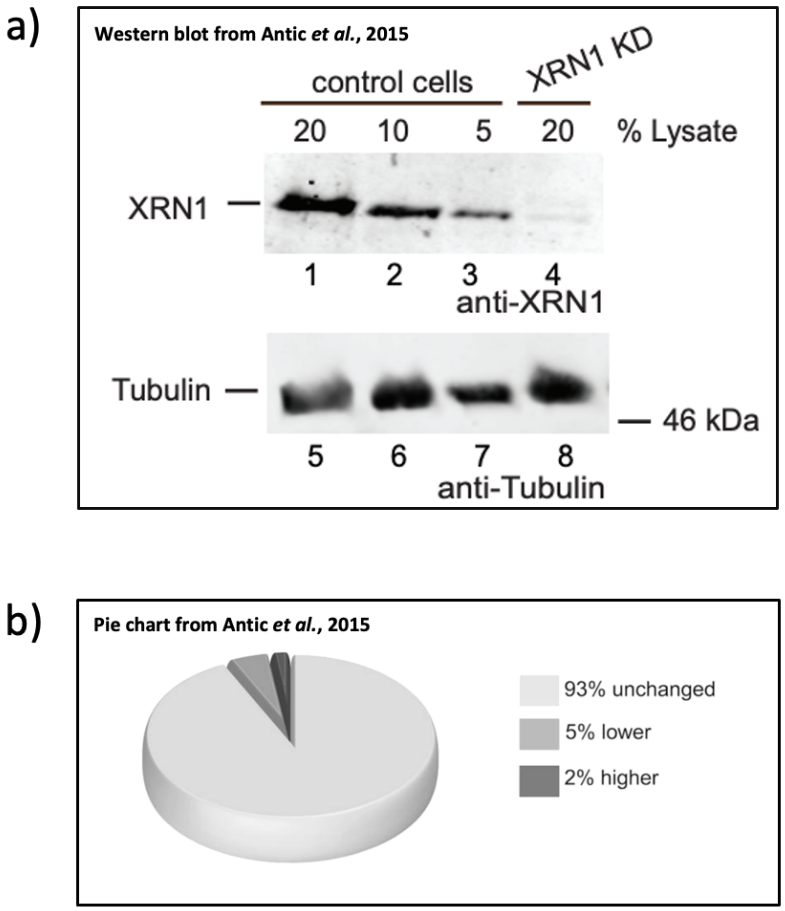


Figure 3.4 – Useful supplemental figures from Antic *et al*. paper

(a)) A Western blot testing Pacman expression in control cells versus Pacman dsRNA treated knockdown cells. A strong (though unquantified) knockdown can clearly be observed with treatment of Pacman complimentary dsRNA.

(b) This pie chart (adapted from Antic et al., 2015) shows the differential abundance of the decapped decay intermediates found associated with the ribosome compared to those in the total cell lysates. The relative abundance was unchanged for most RNA transcripts (93%, pale grey), with a minority being lower on the ribosome than the cell lysate (5%, mid grey), and a smaller minority being higher on the ribosome than the cell lysate (2%, dark grey).

overgrowth, metabolic shifts dependent on altered proliferation, etc.) will all follow on from the initial impact of lack of Pacman degradation of one of its direct targets.

The alteration of core biological processes like this necessarily feed back into the real-time genetic profile of the affected cells, tissues, and organism. The knock-on results of Pacman depletion may then lead to a wide array of regulatory changes, which may manifest as both up- and down-regulation, despite both of them being dependent on Pacman degradation of certain transcripts. It is also worth noting that lncRNAs are often expressed at very low levels, and therefore may not have been detected in some datasets, and replicates, depending on depth of sequencing and variability. In the Antic data, for example, the single replicate analysis will limit the ability to pick up some RNAs that may be differentially abundant in a way that would be apparent across a larger scale experiment but is undetectable due to the potential for variability in a single replicate.

As previously mentioned, lncRNAs directly targeted by Pacman (those that would be most easily highlighted as a lncRNA target of Pacman) would be expected to be up-regulated in its absence or depletion, due to their reduced degradation; and although indirect targets of Pacman may end up up-regulated by its depletion, it was surprising to find so few substantially up-regulated lncRNAs, compared to down-regulated lncRNAs. Despite this, the fact that some lncRNAs were differentially abundant in untreated and Pacman depleted S2 cells provides support to the idea that certain lncRNAs may be specifically targeted and degraded by exoribonucleases such as Pacman, although further investigation in whole organism or tissue sample would provide a more accurate portrayal of normal transcript regulation.

The finding that almost none of the lncRNAs had predicted biological roles, and all were poorly annotated, is not indicative of very much when examining lncRNAs. Due to the somewhat arbitrary criteria by which RNAs and ORFs have historically been defined and identified, this is to be expected when working on lncRNAs; only a certain few have been the subject of any scientific investigation, and most bioinformatic approaches exclude them by default.

Although, as mentioned, the ribosome bound samples being decapped degradation fragments led to the decision for the purpose of this chapter to focus on how Pacman can degrade lncRNAs in S2 cells, without factoring the ribosome-pulldown RNA sequencing data, it is worth noting that in the original paper by Antic *et al*., a large majority of decay intermediates were found at the same relative abundance in both total cell lysate and in ribosome pulldown RNA. This is encouraging going forward, as it shows on a global level that RNA decay is carried out to a significant extent on the translation machinery, and that substantial and interesting results can likely be found in the intersection of degradation and translation.

3.3.2 Analysis of previous data from the Newbury lab identified potential lncRNA targets of Pacman and Dis3L2 in *Drosophila* WIDs

Previous work in the Newbury lab has produced RNA-seq datasets in *Drosophila* WIDs, comparing wild type to Pacman null mutants (33), and wild type to Dis3L2 null mutant (46). The Pacman null mutant was generated using a P-element insertion. One Pacman null mutant (*pcm*^{14}) allele is a 3,501 bp deletion extending from the P-element insertion site towards Pacman, deleting 3,068 bp into the 3′ of the gene, completely removing exons 7–11 and part of exon 6. The 5′ of the neighbouring non-coding RNA *CR43260* is also deleted, although these deletions were shown not to contribute to the Pacman mutant phenotypes (156). The other Pacman null mutant (*pcm*^{15}) allele was created by imprecise P-element excision of *P{EP}pcm*^{G1726} from stock 33263 (*w* P{EP}pcm*^{G1726}) (156). For the isogenic wild-type controls, a line from which the P-element was excised without causing a deletion was selected. The Dis3L2 null mutant was generated using the CRISPR-Cas9 system, an efficient way to generate the advantages that can be gained from a null mutant versus a knockdown. A guide RNA (gRNA) targeting the first common exon of the two dis3L2 isoforms was used generating a line with an 8bp mutation. The successful knockout of Dis3L2 was tested using an antibody to the first 198 amino acids which was specific to the *Drosophila* Dis3L2 protein. The control for these experiments consisted of a line which went through the CRISPR process but was either unedited or repaired correctly.

Although this work was focused primarily on mRNAs, the data provided an excellent opportunity, and re-analysis of this existing data would allow the identification of differentially abundant lncRNAs in both mutants, and in a whole *Drosophila* tissue. As well as the obvious advantages in identifying biologically relevant targets in an actual living organism, the shared use of WIDs also provides the opportunity (at a later point in this thesis) for more meaningful cross-comparisons between the two datasets.
The .bam files from these RNA-seq experiments were processed with the CuffLinks pipeline to produce normalised expression counts and allow differential abundance analysis. From this, RNAs annotated as lncRNAs were extracted and taken forward. These were then filtered to remove any genes with a normalised read count of zero in the Pacman or Dis3L2 mutant samples (as the genes of interest are those degraded by these exoribonucleases, and therefore should be stabilised in the exoribonuclease mutants); as well as removing any duplicate gene annotations where the software had failed to reliably map reads. These were then sorted by fold-change increase in the mutant samples. It is worth noting that due to the experiments being carried out at different points, with advances in technologies available, the RNA-seq on the Pacman mutant was polyA selected, while the RNA used for the Dis3L2 mutant sequencing was rRNA depleted.

This provided a list of 480 lncRNAs, 156 of which were listed as "infinite upregulation" in the absence of Pacman. This can be due to either an "on-off switch" gene with a Boolean expression profile in the absence of Pacman, or by low enough expression levels that no reads are detected in the wild-type tissues. In order to filter out the latter, those with a calculated infinite upregulation were filtered out if they had a normalised read count of less than 1 FPKM in the Pacman mutant. This left 325 lncRNAs, of which 161 had a higher read count in the Pacman mutant than in the isogenic control. A cut off of 1.5-fold increase or more was introduced, producing a list of 105 potential Pacman degraded lncRNAs. In the Dis3L2 dataset, this provided a list of 195 lncRNAs, 71 of which were listed as "infinite upregulation". In order to filter out the false infinite upregulation calculations due to very low expression levels, those with a calculated infinite upregulation were filtered out if they had a normalised read count of less than 1 FPKM in the Dis3L2 mutant. This left 129 lncRNAs, of which 56 had a higher read count in the Dis3L2 mutant than in the isogenic control. A cut off of 1.5-fold increase or more was introduced, producing a list of 40 potential Dis3L2 degraded lncRNAs. These steps are summarised in Figure 3.5.

As the aim of this re-analysis was primarily to identify candidate lncRNAs degraded by Pacman and Dis3L2 to be validated experimentally, these quality controlled, mapped, and counted datasets were then taken and used to create further shortlists of the most likely upregulated lncRNAs, according to degree of differential abundance (greater than 2-fold increase in the absence of Pacman in order to highlight only the strongest candidates), and variability between replicates and statistical significance according to CuffDiff (this should not be taken as confident or definitive statistical testing, beyond its use for filtering for likely candidates, due to the low replicate numbers, with 3 for each genotype, except Pacman null mutants, and corresponding controls which had 6 replicates available (3 *pcm14* mutants, 3 *pcm15* mutants, 3 *50E* isogenic control (to *pcm14*), and 3 166V isogenic control (to *pcm15*), allowing higher confidence for that genotype). The newly compiled shortlist identified 16 substantially and significantly upregulated lncRNAs in the Pacman null mutant (Figure 3.6, panel (a)), and an additional 15 in the Dis3L2 null mutant (Figure 3.6, panel (b)).

3.4 Preliminary exploration of candidate lncRNAs

Having identified some likely candidates, experimental validation was used in order to experimentally distinguish those lncRNA definitively regulated at a post-transcriptional level by Pacman or Dis3L2. Of the candidates identified from the Newbury lab work, an initial batch of lncRNAs were selected (by choosing lncRNAs with only one known transcript, and with a suggested exon-exon boundary suitable for primer design (Figure 3.7, panel (a)). All lncRNA primers are designed to amplify over the exon-exon boundary, therefore product length can be used to determine whether primers are binding to the spliced mature-lncRNA transcript. Primers meeting these criteria were designed and ordered from Sigma (Figure 3.7, panel (b)). Initially, semi-quantitive PCR was used for this purpose, as it allows screening of more candidates to be undertaken for a lower cost, reducing the investment required from systems like TaqMan qPCR by allowing those more costly probes to only be ordered to further validate proven candidates.

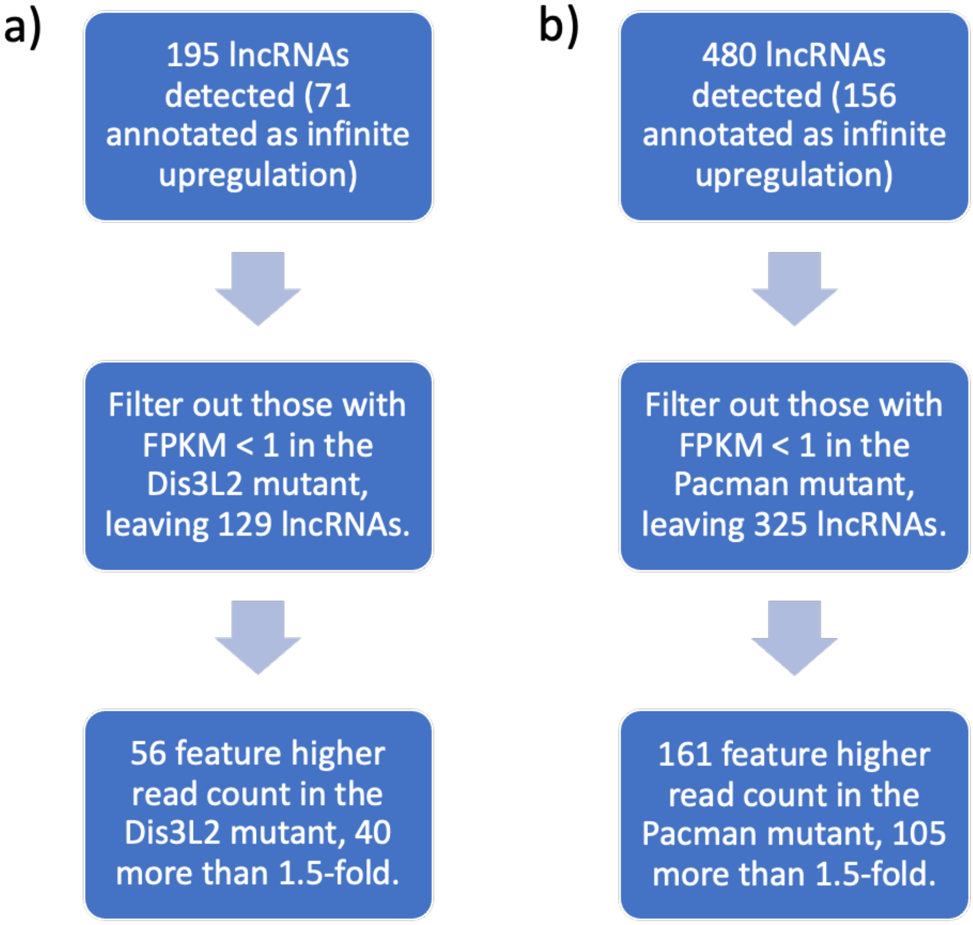
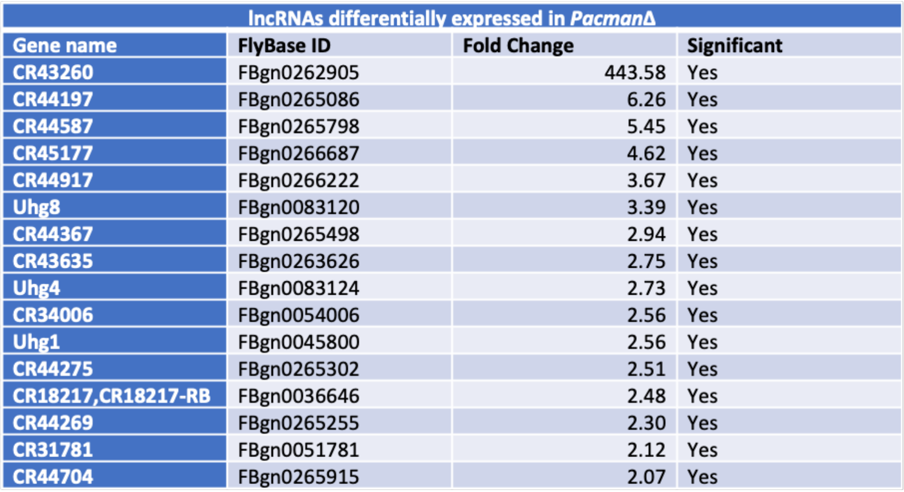

a)
195 lncRNAs detected (71 annotated as infinite upregulation)
Filter out those with FPKM < 1 in the Dis3L2 mutant, leaving 129 lncRNAs.
56 feature higher read count in the Dis3L2 mutant, 40 more than 1.5-fold.
b)
480 lncRNAs detected (156 annotated as infinite upregulation)
Filter out those with FPKM < 1 in the Pacman mutant, leaving 325 lncRNAs.
161 feature higher read count in the Pacman mutant, 105 more than 1.5-fold.

Figure 3.5 – Summary workflow of selection steps to analyse previous data from the Newbury lab:

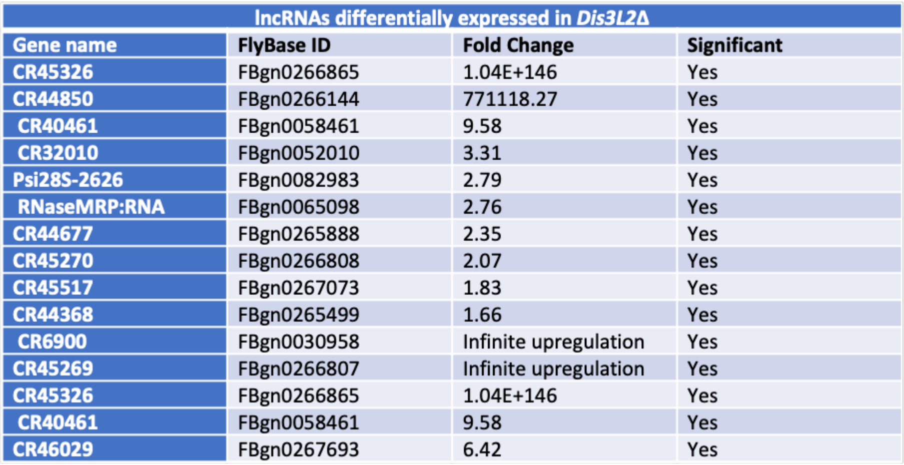
Flowcharts summarising the refinement steps to select candidate exoribonuclease sensitive lncRNAs (Pacman-sensitive in panel (a), Dis3L2-sensitive in panel (b)) from pre-existing Newbury lab data.

a)



lncRNAs differentially expressed in *PacmanΔ*			
Gene name	FlyBase ID	Fold Change	Significant
CR43260	FBgn0262905	443.58	Yes
CR44197	FBgn0265086	6.26	Yes
CR44587	FBgn0265798	5.45	Yes
CR45177	FBgn0266687	4.62	Yes
CR44917	FBgn0266222	3.67	Yes
Uhg8	FBgn0083120	3.39	Yes
CR44367	FBgn0265498	2.94	Yes
CR43635	FBgn0263626	2.75	Yes
Uhg4	FBgn0083124	2.73	Yes
CR34006	FBgn0054006	2.56	Yes
Uhg1	FBgn0045800	2.56	Yes
CR44275	FBgn0265302	2.51	Yes
CR18217,CR18217-RB	FBgn0036646	2.48	Yes
CR44269	FBgn0265255	2.30	Yes
CR31781	FBgn0051781	2.12	Yes
CR44704	FBgn0265915	2.07	Yes

b)



lncRNAs differentially expressed in *Dis3L2Δ*			
Gene name	FlyBase ID	Fold Change	Significant
CR45326	FBgn0266865	1.04E+146	Yes
CR44850	FBgn0266144	771118.27	Yes
CR40461	FBgn0058461	9.58	Yes
CR32010	FBgn0052010	3.31	Yes
Psi28S-2626	FBgn0082983	2.79	Yes
RNaseMRP:RNA	FBgn0065098	2.76	Yes
CR44677	FBgn0265888	2.35	Yes
CR45270	FBgn0266808	2.07	Yes
CR45517	FBgn0267073	1.83	Yes
CR44368	FBgn0265499	1.66	Yes
CR6900	FBgn0030958	Infinite upregulation	Yes
CR45269	FBgn0266807	Infinite upregulation	Yes
CR45326	FBgn0266865	1.04E+146	Yes
CR40461	FBgn0058461	9.58	Yes
CR46029	FBgn0267693	6.42	Yes

Figure 3.6 – Re-analysis of Newbury lab data provides a novel shortlist of lncRNAs upregulated in Pacman and Dis3L2 null mutant wing imaginal discs:

(a) A shortlist of 16 lncRNAs up-regulated in Pacman null mutants. All candidate lncRNAs were present at significantly different levels between the wild type control and the Pacman null mutants, and mean fold change is listed.

(b) A shortlist of 15 lncRNAs up-regulated in Dis3L2 null mutants WIDs in previous Newbury lab data. All candidate lncRNAs were present at significantly different levels between the wild type control and the Pacman null mutants, and mean fold change is listed

a) The example of *CR45177* is shown below;

TTGCAAATACCCTCCTGCTAGCTGTAGGAACTTGTAACGATTCATAAGGCTAGGGTGGTGTTATGAGATATTCCAGGGC
GCACGCGCGAATGTCCAGCTGCTCCTTGTCGGTTATGTGGGGTGATGGTCTTGTCCCGGAATCTCGATGTGACGACAAT
TAAAAAAGGGCGGCAAATCGCCATCCACGACACGGACAACGGACCTCATATGTGGTGGAGGATCATgtaatattaatataaa
tattaatattaataataacacaactaaaatttcttacagCTACCATAATCGTAACTTAGTAAAATTCGGACCGCTAATAATCCAGCAATT
TGCAGTTTTCGATTCCCACTCGAATATTTTATAATCAGCTACCCGCGCTTAATGGATATGGAAAGTTGATCCCTTAAAAAA
TAATTAACTTAGATGAAATAATTTTAGAGTTGAGAGACAATCTTGGCACCCCAAGAATTAATTAAAAATAAACATACATT
GCTG

b)

Gene	Forward Primer	Reverse Primer	Predicted Product Length (bp)	Upregulated in which mutant	Band seen at expected size
CR44677	TGGCTTTGGTCATGAGTCCC	GTGGCACTAGAACTGTGGCT	522	dis3l2	No
CR6900	GGTGACGAGATGTCCAAGCA	AGTAACTTGGAGGTTCAGTGC	207	dis3l2	Yes
CR42719	GCAAAGCGACAAAGAGCGAG	ATCTCCAAAACTCGGCCTCC	143	dis3l2	Yes
CR43466	GAATTCCTTCGGACCTGGCA	ACCAAAATCCCGCGTCTTCT	257	dis3l2	No
CR43260	TCATTGCGAGCAGGTTATTCC	GACACGCGGCTTATAGGTGA	391	pacman	Yes
CR44587	TTGCCGAGTGCTCTCCATTT	CGATCGACGATGGAGCTTGA	304	pacman	No
CR45177	CGGTTATGTGGGGTGATGGT	TTGGGGTGCCAAGATTGTCT	302	pacman	Yes
uhg8	AAAGCTGACCGTATGGGCTC	TGTTAAAAGATTGCGATTTAGCACG	378	pacman	No
CR44367	ACATACGCTTGGGGTGCTAA	ATGAGCTGGAGTCCCTTTGC	359	pacman	Yes
CR43635	GCATATGTGCACATCTCTCC	GTCCCACAATTGCTGTTGCT	301	pacman	Yes
CR34006	GTCAGTCAGCCTCCGATTCC	GGCCATATACTCGACTGGGC	441	pacman	No
CR18217	TCGATCGAAGCGACGTG	GAATTTCGGCCATCGACAGG	533	pacman	No
rp49	CCAGTCGGATCGATATGCTAA	TCTGCATGAGCAGGACCTC	206	housekeeper	Yes

Figure 3.7 – Designing primers to pick up only the mature transcript:

(a) A visual example of how target lncRNAs were selected, and primers designed. Yellow shows the binding location of the forward primer, green shows the binding location of the reverse primer, and the intronic sequence is shown in blue and lowercase. Hence a mature-lncRNA product will be able to be differentiated between gDNA or pre-lncRNA by the presence of the intronic sequence.

(b) A reference table of specificity and binding for primers designed for candidate lncRNAs, the mutant in which they were seen to be upregulated, their predicted product length, and whether a band of this size was seen.

It was decided that in order to test primer binding, whole L3 larvae could be used for the template RNA, as it is significantly less time consuming to cultivate and select enough L3 to extract sufficient RNA from, without the additional step of WID dissection (which also vastly reduces the amount of RNA extracted and requires use of more expensive column extraction reagents). In addition, RNA from whole organism or tissues tends to be more representative of biological realities than established cell lines. Whole *Drosophila* L3 larvae were available for both *pacman* and *dis3L2* null mutants, as well as isogenic controls, allowing straightforward harvesting of the required genetic material. A cDNA copy of the RNA present in L3 larvae was produced via L3 homogenisation and lysis (by liquid-nitrogen cooled pestle and mortar homogenisation of L3 larvae with a lysis buffer), RNA extraction (including a DNase step) and RT-PCR, as described in methods. Although the high debris and fat content in whole larvae can complicate delicate extractions, this did not pose a problem for simple extraction of total RNA. The cDNA from this was then used to test whether primers for potential lncRNAs upregulated in *pacman* and *dis3l2* mutants were producing the expected product, by running the product on an agarose gel (alongside a molecular ladder,) and imaging it.

Out of the 8 lncRNA candidates upregulated in the absence of Pacman, 7 primer pairs produced a product, and of these 4 were of the expected size (*CR43260, CR45177, CR44367, CR43635*). Of the 4 lncRNA candidates upregulated in the absence of Dis3L2, 2 of the expected products were seen (*CR6900, CR42719*), out of the 4 tested lncRNA genes. These are summarised in panel (b) of Figure 3.7. The primers that failed to produce a band of the right size likely either do not feature strong and specific enough binding or may bind to transcripts only present at very low levels in whole L3 larvae. An example of the gels used to verify the primer product can be seen in Figure 3.8.

Given that these RNAs were identified in L3 WIDs (a tissue present within the whole L3), we do know that they are expressed at some level. It is possible that they are present only in the WID (without expression throughout the rest of the larvae, or at very low levels throughout the rest of the larvae), vastly reducing the relative abundance of the transcript when examined from the RNA extracted from total (non-tissue-specific) L3 larvae. Due to this experiment aiming only to offer proof of principle of Pacman and Dis3L2 regulated lncRNA in L3, and offer a preliminary look into differential regulation, it was decided that the pursuing the successful primers and PCRs should be prioritised

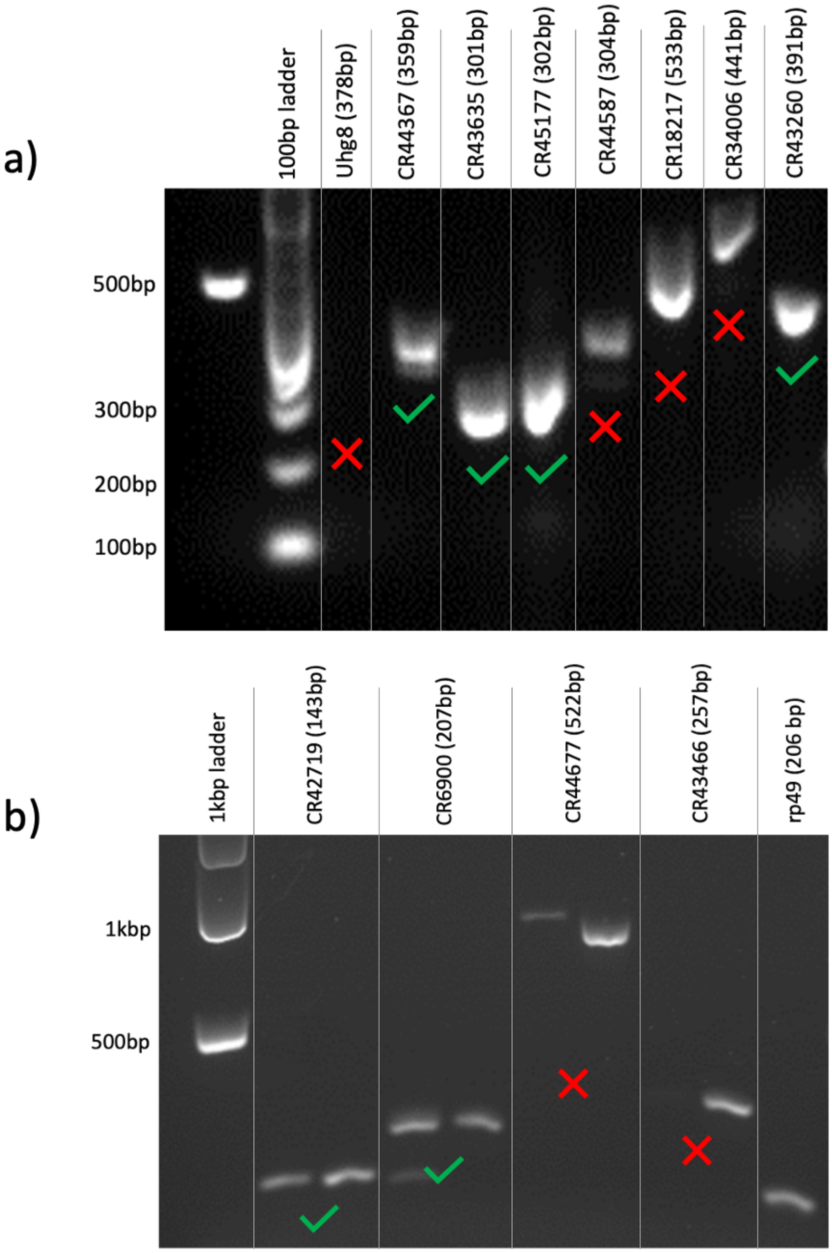


Figure 3.8 – Testing sqPCR primers ensured that correct, and specific products were identified and carried forward:

An example of the primers tested, to be used for sqPCR. Ticks indicate the presence of a band of the right size. Where a non-specific band was seen (as seen in CR44367), this was noted, so the PCR products could be run out over a higher density gel, or for a longer period, to ensure that only the band of interest would be quantified. Panel (a) shows the primers designed to test the *pacman*-regulated lncRNAs, while Panel (b) shows those designed to test *dis3l2*-regulated lncRNAs.

(Some data presented in this figure is from work carried out by Harry Pink, supervised by Oliver Rogoyski)

over optimisation of PCR for the entire shortlist (especially given that successful proof of principle would lead to use of high-throughput techniques). From this point onwards, the template cDNA used was a cDNA copy of the RNA present in pre-dissected frozen stocks of WIDs was produced via column RNA extraction (including a DNase step) and RT-PCR, as described in methods. This allowed a more meaningful comparison to be made with the Newbury lab RNA-seq data (also from WIDs).

sqPCR was performed for four of the genes for which the primers bound specifically and produced a product of the correct size (*CR42719, CR6900, CR43635, CR45177*, summarized in Figure 3.9). Primers for the gene *rp49* were used alongside the genes of interest, as *rp49* is known to have constant expression in *Drosophila* L3 larvae, in both mutants and the isogenic control. This means it can be used as a housekeeper gene, allowing normalization of gene expression across samples.

3.4.1 Validating *CR42719* upregulation in *dis3l2* mutant wing imaginal discs

RNA-seq data from the Newbury lab showed *CR42719* to be "infinitely" upregulated in *dis3l2* mutant WIDs, indicating either a binary "on-off" mechanism for its presence, or such a strong upregulation that it was beyond the resolving power of that sequencing experiment to meaningfully quantify it. Figures 3.6, panel (b) and 3.7 show *CR42719* to be one of the lncRNA genes upregulated in *dis3l2* mutants for which the primers bound and amplified the correct region. This allowed it to be carried forward for sqPCR validation of upregulation. sqPCR was performed using *CR42719*-specific primers at cycles 24, 28, 30, 35, and 40, with template cDNA from either *dis3L2* mutant or isogenic control WIDs.

Having ran the products on a 1% agarose gel, the *CR42719* band of interest is seen at 143bp, along with a non-specific ~210bp band. Initially, the band of interest could not be quantified without including some of the non-specific band, and the reaction required some optimisation. Cycle 30 seemed to be the optimum point to visualise the *CR42719* specific band, showing a clear 143bp band in *dis3l2* mutants, but not in isogenic controls (Figure 3.10). The sqPCR was repeated at cycle 30 and was run out further than previously on a 1.5% gel to further separate the specific and non-specific

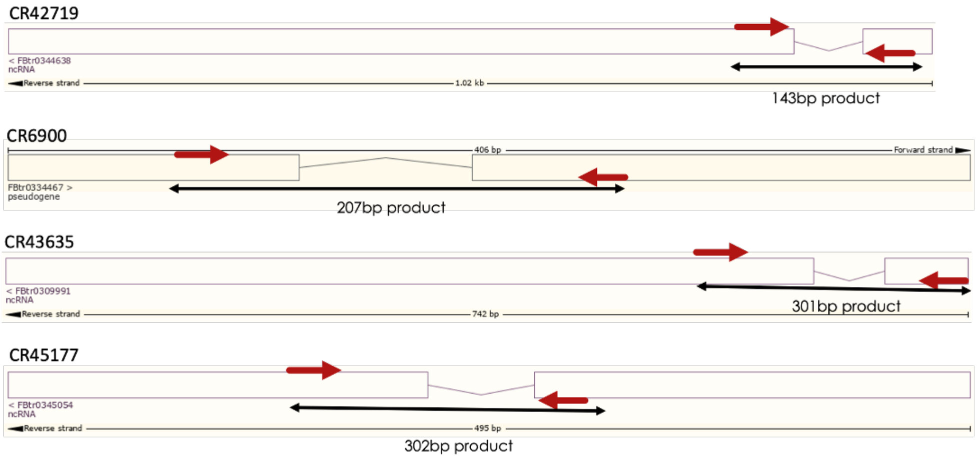


Figure 3.9 – Primer binding sites:

A visual representation of the tested lncRNAs, and where the PCR primers bind across exon-exon boundaries. Product size is listed below each .

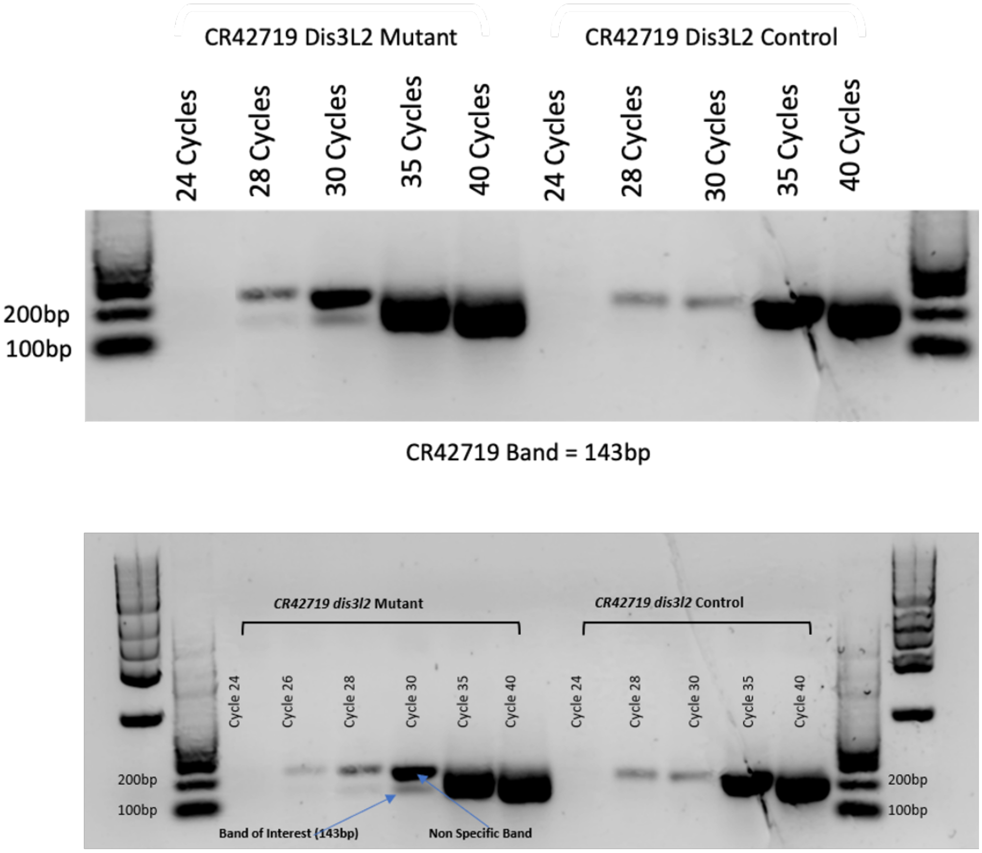


Figure 3.10:

Initial gels optimising cell cycles and comparing CR42719 abundance in *dis3l2* mutant and wild type control. The CR42719 band of interest is seen at 143bp, along with a non-specific ~210bp band. Cycle 30 seemed to be the optimum point to visualise the CR42719 specific band, showing a clear ~143bp band in dis3l2 mutants, but not in isogenic controls.

(Some data presented in this figure is from work carried out by Harry Pink, supervised by Oliver Rogoyski)

bands. This gel was then imaged on the Li-COR, allowing gel band quantification (Figure 3.11, panel (a)). When normalised to *rp49* expression, the gel band quantification showed a statistically significant 3.9-fold upregulation in the *dis3l2* mutant (Figure 3.11, panel (b)), supporting the RNA-Seq data.
Following confirmation of a statistically significant sqPCR upregulation in *dis3l2* mutants, custom TaqMan qPCR probes were designed for *CR42719*, across its exon-exon boundary (to ensure only mature, spliced transcript is detected). Following qPCR and $\Delta\Delta C_t$ analysis, a statistically significant 2.6-fold upregulation was shown in the *dis3l2* mutant, shown in Figure 3.11, panel (c). $\Delta\Delta C_t$ analysis calculates the fold change in expression from isogenic controls to mutants, normalised to *rp*49, confirming the RNA-Seq data. TaqMan qPCR relies on 3 sequence-specific binding events, meaning it is very specific, allowing high confidence in this upregulation.

3.4.2 *CR6900* shows no significant upregulation in *dis3l2* mutant wing imaginal discs

RNA-seq data from the Newbury lab showed *CR6900* to be “infinitely” upregulated in *dis3l2* mutants. Figure 3.6 also showed that the *CR6900* primers produced a band of the correct size. Upon following up with sqPCR on *dis3l2* mutant and isogenic control wing disc cDNA (Figure 3.12, panel (a)) non-significant upregulation (3.9-fold by sqPCR, and 1.4-fold by qPCR) of *CR6900* was seen in the *dis3l2* mutant wing discs after normalisation to *rp49* (Figure 3.12, panel (b)). Also seen in Figure 3.12, panel (a) is a non-specific band of ~300bp that also looks to be more abundant in the *dis3l2* mutant, this is likely to be the unspliced pre-lncRNA. With the inclusion of the one predicted intron, a product of 277bp would be produced, and this may be responsible for the band (precise size is hard to assess without running on a higher density gel for a longer time). It was ensured that this band was separated sufficiently to not influence band intensity quantification.

Despite the lack of statistical significance (Figure 3.12, panel (b)) (potentially due to the variability of technique), sqPCR had shown a 3.9-fold upregulation in the *dis3l2* mutant. Following this, a TaqMan probe was designed for *CR6900*. TaqMan qPCR carried out on *dis3l2* mutant and isogenic control wing disc cDNA showed a 1.4-fold up regulation with

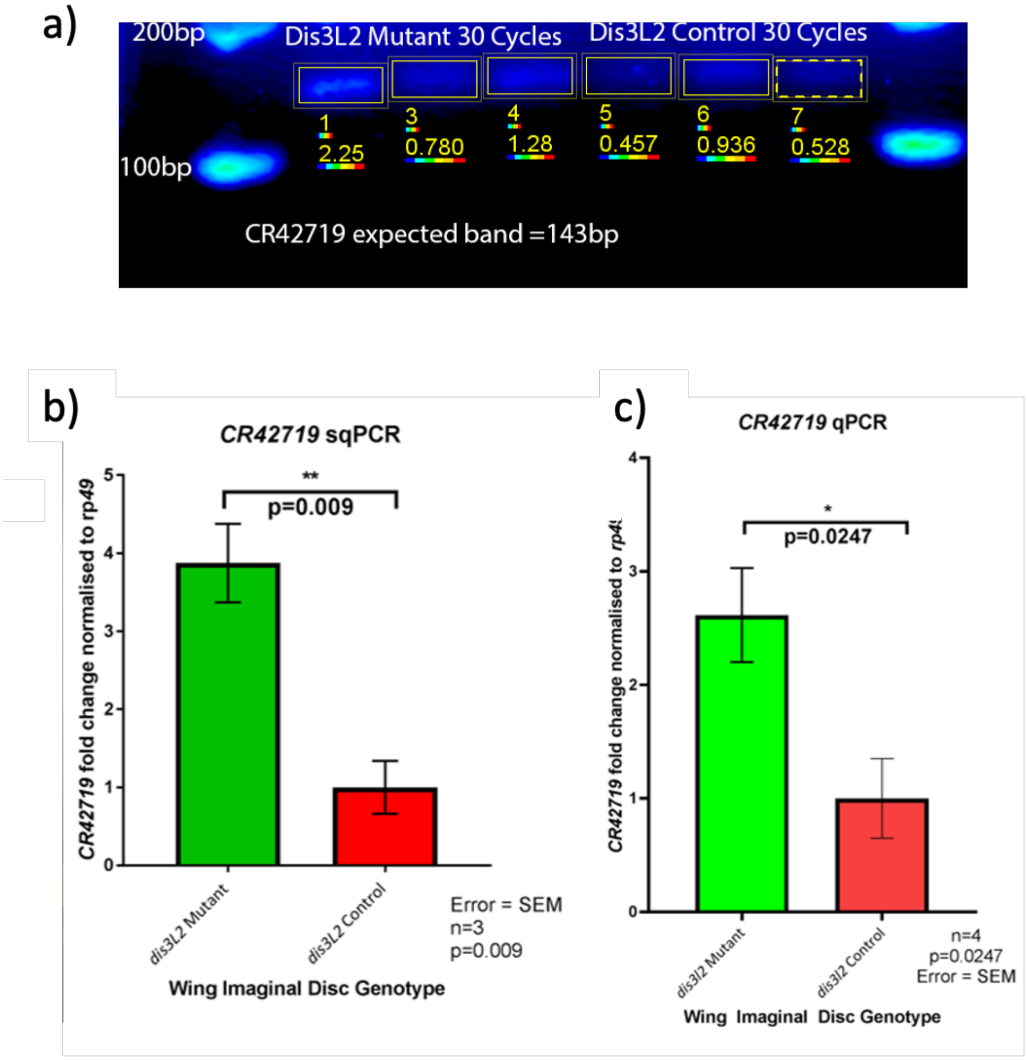


Figure 3.11 - CR42719 sqPCR showed upregulation in dis3l2 mutant, and was statistically significant:

Panel (a) Shows visualisation and imaging of sqPCR shows upregulation of *CR42719* in the Pacman deficient mutant: The sqPCR was repeated at cycle 30 (with 3 replicates per genotype,) and was run out further than previously on a 1.5% gel to further separate the specific and non-specific bands. This gel was then imaged on the Li-COR, allowing gel band quantification

Panel (b) shows that after gel band quantification at cycle 30 and normalisation to *rp49* expression, a 3.9-fold upregulation was shown. An unpaired t-test showed this was statistically significant, with a p-value of 0.009.

Panel (c) using qPCR shows a statistically significant upregulation in the Dis3l2 mutant. A 2.6-fold upregulation is seen, and a t-test reveals a p-value of 0.0247. This is greater than the 0.05 confidence level, showing statistical significance

(Some data presented in this figure is from work carried out by Harry Pink, supervised by Oliver Rogoyski)

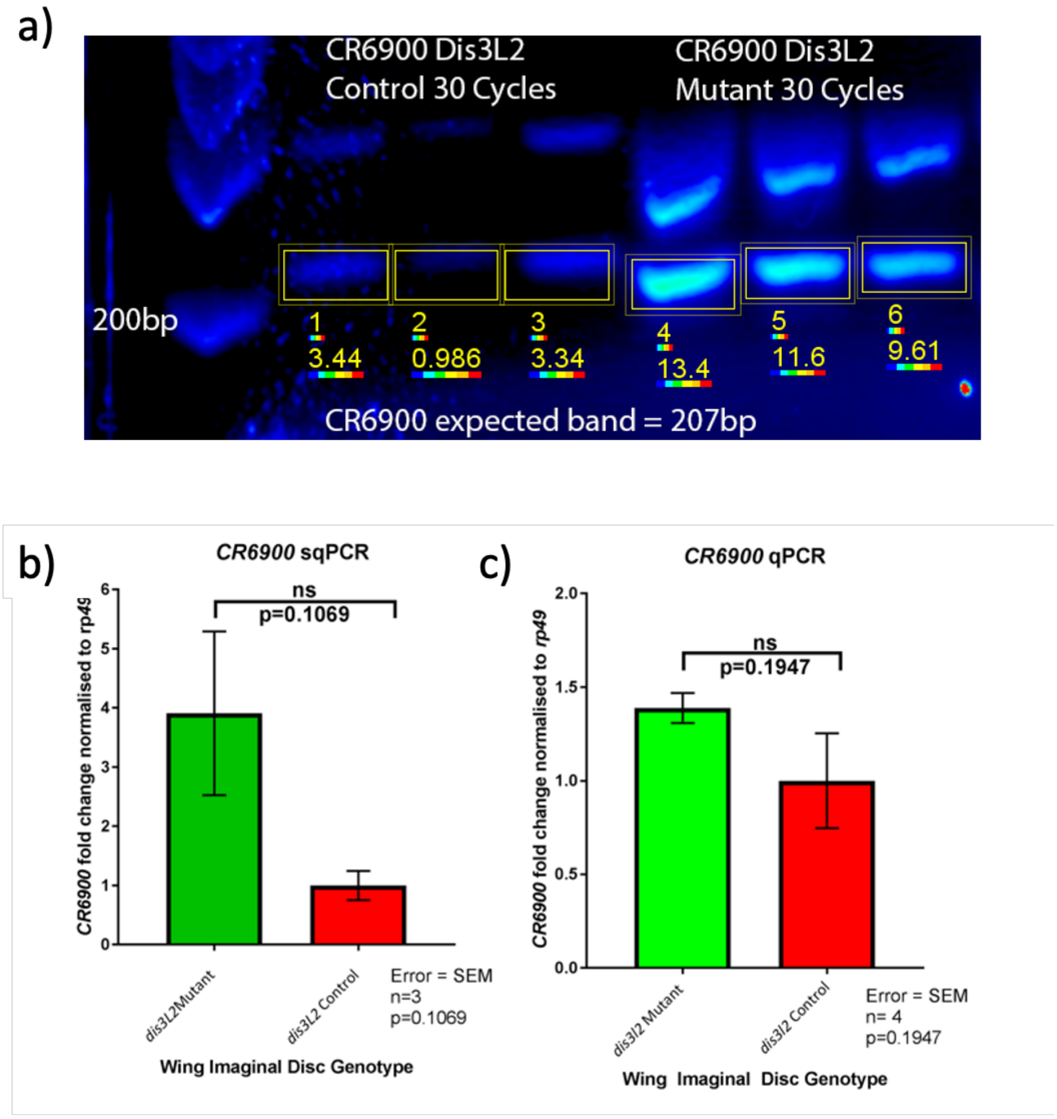


Figure 3.12 - CR6900 sqPCR showed upregulation in dis3l2 mutant, however this was not statistically significant:

Panel (a) shows visualisation and imaging of sqPCR shows upregulation of *CR6900* in the Pacman deficient mutant: The sqPCR was repeated at cycle 30 (with 3 replicates per genotype,) and was run on a 1.5% agarose gel. This gel was then imaged on the Li-COR, allowing gel band quantification.

Panel (b) shows that after gel band quantification at cycle 30 and normalisation to *rp49* expression, a 3.9-fold upregulation was shown. However, an unpaired t-test showed this was statistically insignificant.

Panel (c) qPCR shows no statistically significant upregulation in the Dis3l2 mutant. A 1.4-fold upregulation is seen; however, a t-test reveals a p-value of 0.1947. This is greater than the 0.05 confidence level, hence no statistical significance.

(Some data presented in this figure is from work carried out by Harry Pink, supervised by Oliver Rogoyski)

no statistically significant difference between *dis3l2* mutant and isogenic control (p=0.1947), shown in Figure 3.12, panel (b). This is insufficient to confirm the RNA-seq data's preliminary findings. However, this repeated lack of statistical significance shows the importance of performing RNA-seq data with multiple molecular techniques, in order to attempt to differentiate between false positives and minor but significant changes.

3.4.3 *CR43635* sqPCR confirms upregulation in *pacman* mutant wing discs, validating the RNA-seq data

RNA-Seq data showed a 2.75-fold upregulation of *CR43635* in *pacman* mutant wing discs (compared to isogenic controls). Figure 3.7 showed that the *CR43635* primers were producing a band of the expected size. Following this, the lncRNA was then selected as a worthwhile candidate for verification of upregulation.

Figure 3.13 shows the sqPCR data for *CR43635* on *pacman* mutant versus isogenic control wing disc cDNA. In Figure 3.13, panel (a), the Li-COR imaged sqPCR gel and gel band quantification data at cycle 28 can be seen. When normalised to *rp49*, this lncRNA showed a statistically significant 9.3-fold upregulation in the *pacman* mutant (Figure 3.13, panel (b)). A TaqMan qPCR probe for this gene was not available at the time of writing, so the conclusion of significant upregulation is supported by fewer methods of testing, and with less confidence, than *CR42719*, *CR6900*, and *CR45177*.

3.4.4 *CR45177* is upregulated in *pacman* mutants and quantified differential expression shows statistical significance in qPCR

The initial investigation from re-analysis of RNA-Seq data showed a 4.62-fold regulation of *CR45177* in *pacman* mutant wing discs (compared to isogenic controls), and Figure 3.7 showed the *CR45177* primers to bind and amplify a product of the right size. Following this, Figure 3.14, panel (a) shows the Li-COR imaged gel with gel band quantification data of the *CR45177* sqPCR on the Pacman mutant and isogenic control wing disc cDNA after 28 cycles Analysis of this suggested greater abundance in the mutant, and after normalisation to *rp49*, a 14.4-fold upregulation was shown, although

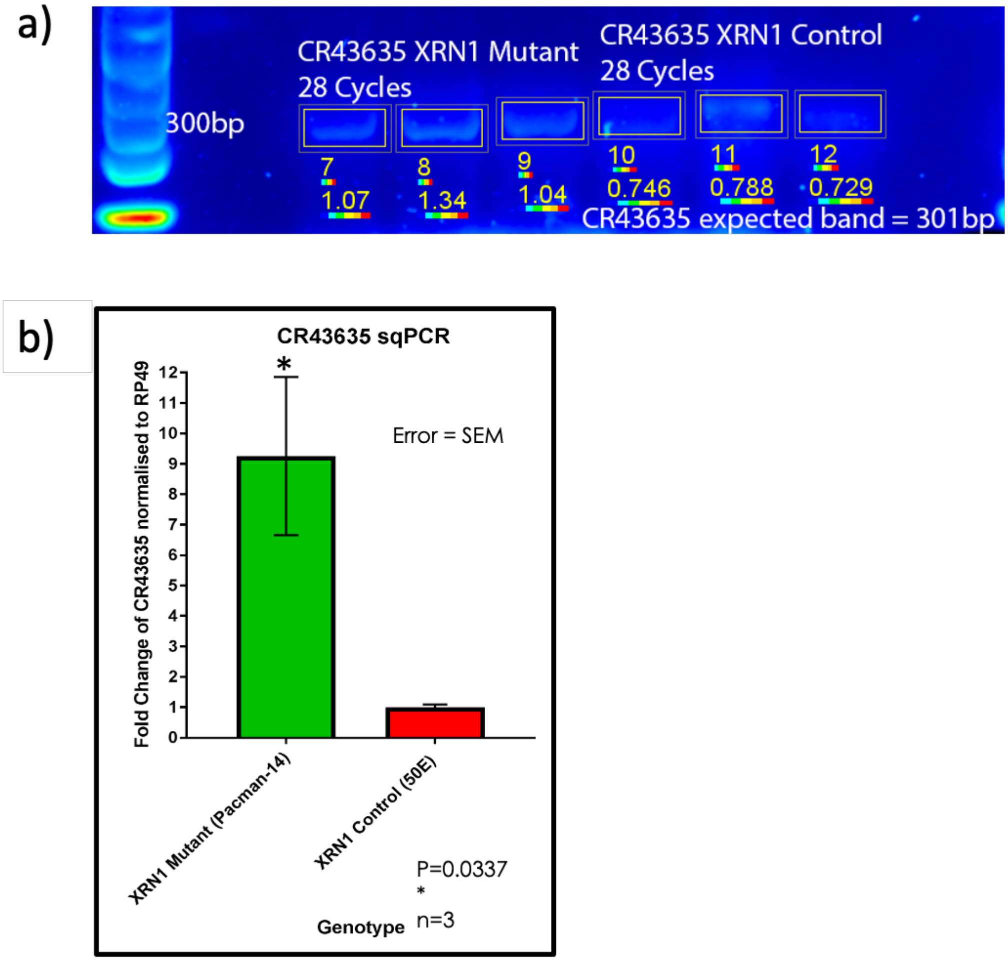


Figure 3.13 - CR43635 sqPCR showed statistically significant upregulation in Pacman functional null mutant:

Panel (a) shows visualisation and imaging of sqPCR shows upregulation of *CR43635* in the Pacman deficient mutant: The sqPCR was repeated at cycle 28 (with 3 replicates per genotype,) and was run on a 1.5% agarose gel. This gel was then imaged on the Li-COR, allowing gel band quantification.

Panel (b) shows after gel band quantification at cycle 28, and normalisation to *rp49* expression the a statistically significant 9-fold upregulation was seen in the Dis3l2 mutant, (compared to the expression seen in the wild type control,) supporting the RNA-Seq data

(Some data presented in this figure is from work carried out by Harry Pink, supervised by Oliver Rogoyski)

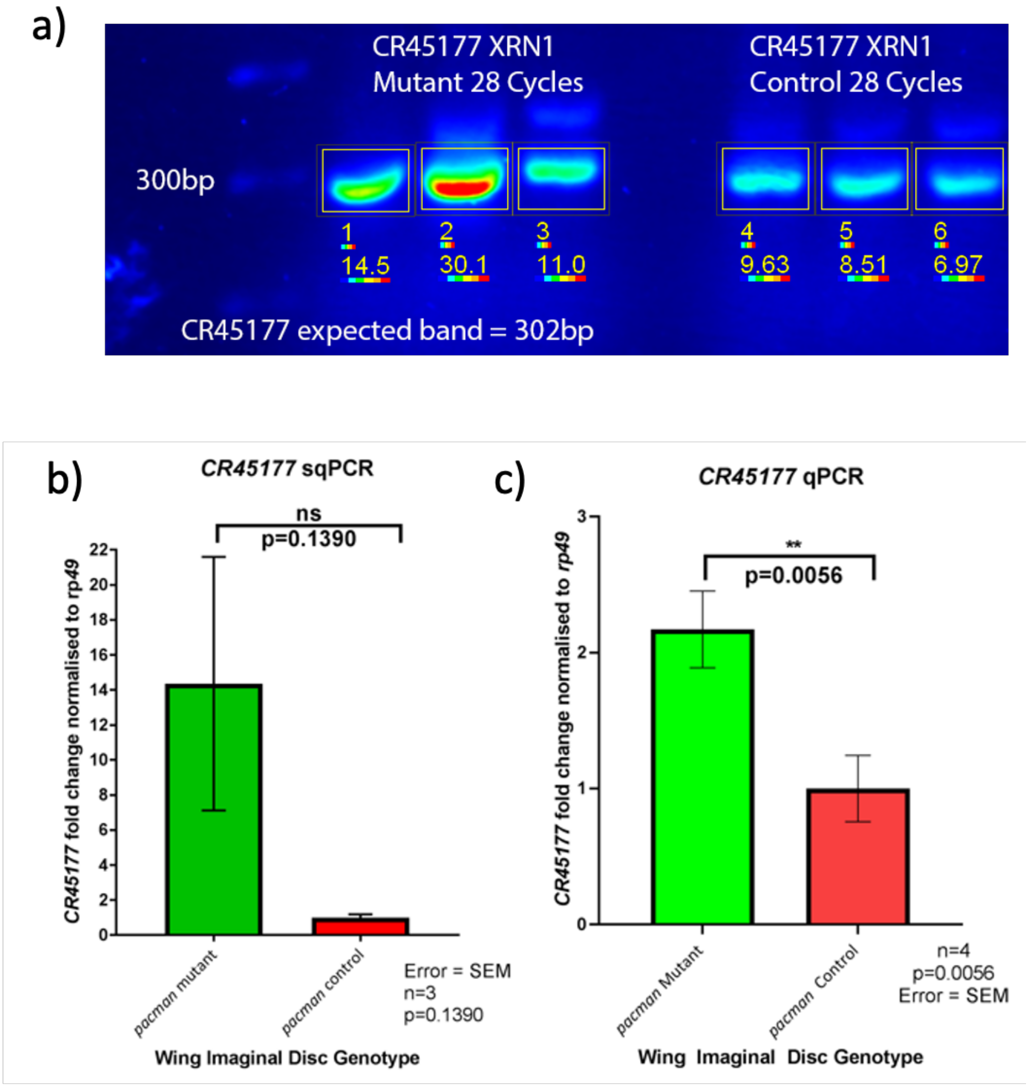


Figure 3.14 - *CR45177* sqPCR showed upregulation in Dis3l2 mutant, shown by qPCR to be statistically significant:

Panel (a) shows visualisation and imaging of sqPCR shows upregulation of *CR45177* in the Pacman deficient mutant: The sqPCR was repeated at cycle 28 (with 3 replicates per genotype,) and was run on a 1.5% agarose gel. This gel was then imaged on the Li-COR, allowing gel band quantification.

Panel (b) shows that after gel band quantification at cycle 328 and normalisation to *rp49* expression, a 14-fold upregulation was shown. However, an unpaired t-test showed this was statistically insignificant.

Panel (c) qPCR shows statistically significant upregulation in the Dis3l2 mutant. A 2-fold upregulation is seen, and t-test reveals a p-value of 0.0056. This is less than than the 0.05 confidence level, showing statistical significance.

(Some data presented in this figure is from work carried out by Harry Pink, supervised by Oliver Rogoyski)

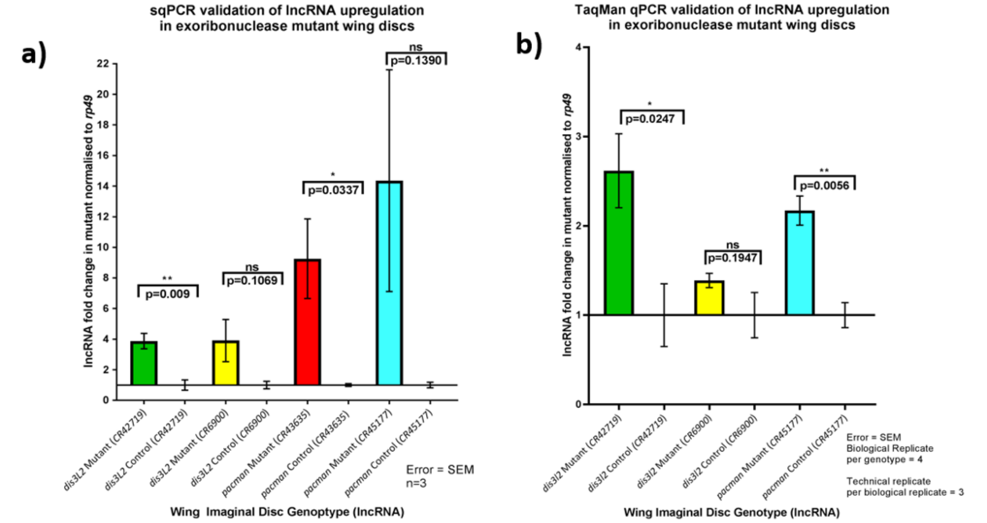
without statistical significance (Figure 3.14, panel (b)). Due to the potential magnitude of the upregulation, with uncertain significance, qPCR TaqMan probes were designed for *CR45177*. qPCR showed a statistically significant (p=0.0056) 2.2-fold upregulation in the *pacman* mutant (Figure 3.14, panel (c)). This result supported the RNA-Seq data and suggests the original statistical insignificance might be due to the high variability, and less accurate quantification of sqPCR.

3.4.5 Summary of candidate lncRNA exploration

sqPCR data shows 2 examples of lncRNAs significantly upregulated in each exoribonuclease mutant. *CR42719* was significantly upregulated in the *dis3l2* mutant and *CR43635* was significantly upregulated in the *pacman* mutant. There also looked to be upregulation in *CR6900* and *CR45177*, despite this not being statistically significant. TaqMan qPCR confirms the statistically significant upregulation of *CR42719* in the dis3l2 mutant. Additionally, *CR45177* was also shown to be significantly upregulated by TaqMan qPCR. *CR6900* showed no significant upregulation, which highlights the importance of validating RNA-Seq data, as false positives can occur in RNA-seq data due to mapping errors, especially with poorly annotated lncRNAs. Figure 3.15 summarises the sqPCR and qPCR data for all 4 tested lncRNAs with their unpaired t-test p-values. Although some of these lncRNAs could benefit from further verification to clarify the extent to which they are differentially upregulated in exoribonuclease deficient mutants, the findings from these experiments, combined with the multiple RNA-seq datasets, is enough to prove the principle of specific lncRNA regulation by Pacman and Dis3L2. The link between translation and degradation is a compelling area for research, and this data provides a worthwhile reason to examine it further in lncRNAs.

3.5 Exploring the polysomal presence of lncRNAs in *Drosophila* samples:

As well as the existing RNA-seq data that can be collected and re-analysed to identify potential candidate lncRNAs regulated by Pacman and Dis3L2; some ribo-seq and poly-ribo-seq data can be analysed in order to identify examples of lncRNAs that are present on the polysome (and by extension, likely candidates for lncRNA translation). This thesis takes advantage of work by the Couso lab (Patraquim *et al*. (90), Aspden *et al*. (153)),



c)

lncRNA	upregulation candidate in;	sqPCR mutant upregulation	qPCR mutant upregulation
CR42719	*dis3l2* mutant	3.9-fold (p=0.009/**)	2.6-fold (p=0.0247/*)
CR6900	*dis3l2* mutant	3.9-fold (p=0.1069/ns)	1.4-fold (p=0.1947/ns)
CR45177	*pacman* mutant	14.4-fold (p=0.1390/ns)	2.2-fold (p=0.0056/**)
CR43635	*pacman* mutant	9.3-fold (p=0.0337/*)	Unable to bind probe over exon-exon boundary

Figure 3.15 - lncRNAs have been shown to be significantly upregulated in an exoribonuclease mutant:

(a) sqPCR data shows 2 lncRNAs to be significantly upregulated in the exoribonuclease mutant. *CR42719* was significantly upregulated in the Dis3l2 mutant and *CR43635* was significantly upregulated in the Pacman mutant. There also looked to be upregulation in *CR6900* and *CR45177*, despite this not being statistically significant. (b) TaqMan qPCR confirms the statistically significant upregulation of *CR42719* in the Dis3l2 mutant. Additionally, *CR45177* was also shown to be significantly upregulation by TaqMan qPCR. *CR6900* showed no significant upregulation, which highlights the importance of validating RNA-Seq data, as false positives can occur in RNA-seq data due to mapping errors, especially with poorly annotated lncRNAs. (c) Table summarising the sqPCR and qPCR data for all 4 tested lncRNAs with their unpaired t-test p-values.

(Some data presented in this figure is from work carried out by Harry Pink, supervised by Oliver Rogoyski)

and Zhang *et al*. (154) in order to take a preliminary look at polysome-associated lncRNAs, and then aims to follow this work with further high-throughput techniques, as well as molecular validation.

3.5.1.1 Global overview and summary of Zhang et al. RNA sequencing of harringtonine treated S2 cells

The paper “Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during *Drosophila* development” by Zhang et al. explores the roles of upstream open reading frames in ribosome occupancy and translational efficiency in *Drosophila* S2 cells. Their use of harringtonine treatment before ribosomal profiling and subsequent sequencing provides a dataset that shows not only the overall RNA profile of S2 cells, but also the profile of ribosome-bound RNAs and where the ribosome binds to transcripts. This is because harringtonine freezes ribosomes after initiation, preventing extension, providing a snapshot of ribosome initiation sites. Although the use of a wild-type genotype prevents this data from being used to answer the questions posed in this thesis regarding lncRNA degradation by Pacman or Dis3L2; or the interplay between translation and degradation by these enzymes, it still offers a valuable resource to assess the likelihood of lncRNA translation, as well as elucidating potential specific initiation sites for novel ORFs.

3.5.1.2 Initial examination of elongation-inhibited lncRNA reads in harringtonine-treated S2 cell ribo-seq data

The data of interest from the Zhang et al. dataset features a single read count, genome wide distribution, for *Drosophila* S2 cells treated with Harringtonine (no comparison to untreated available in the dataset). As there is only one replicate in one condition being examined here, statistical testing was impossible, and normalisation of their data was unnecessary for the purposes of this project. From the initial data set, 252 lncRNA were detected with at least 1 mapped read. Of these, 113 detected at least 3 reads. Although this cut off is used in later points when examining the novel dataset, the data here does not have the benefit of multiple replicates, so a higher threshold was set in order to supply a more promising list of candidates. Of all lncRNAs, 51 were detected as having

20 or more reads detected in the ribosome-bound RNA sample; though this is a higher cut off than for the novel data set, the increased read depth of this data allows for more stringent cutoffs to be used (Figure 3.16).

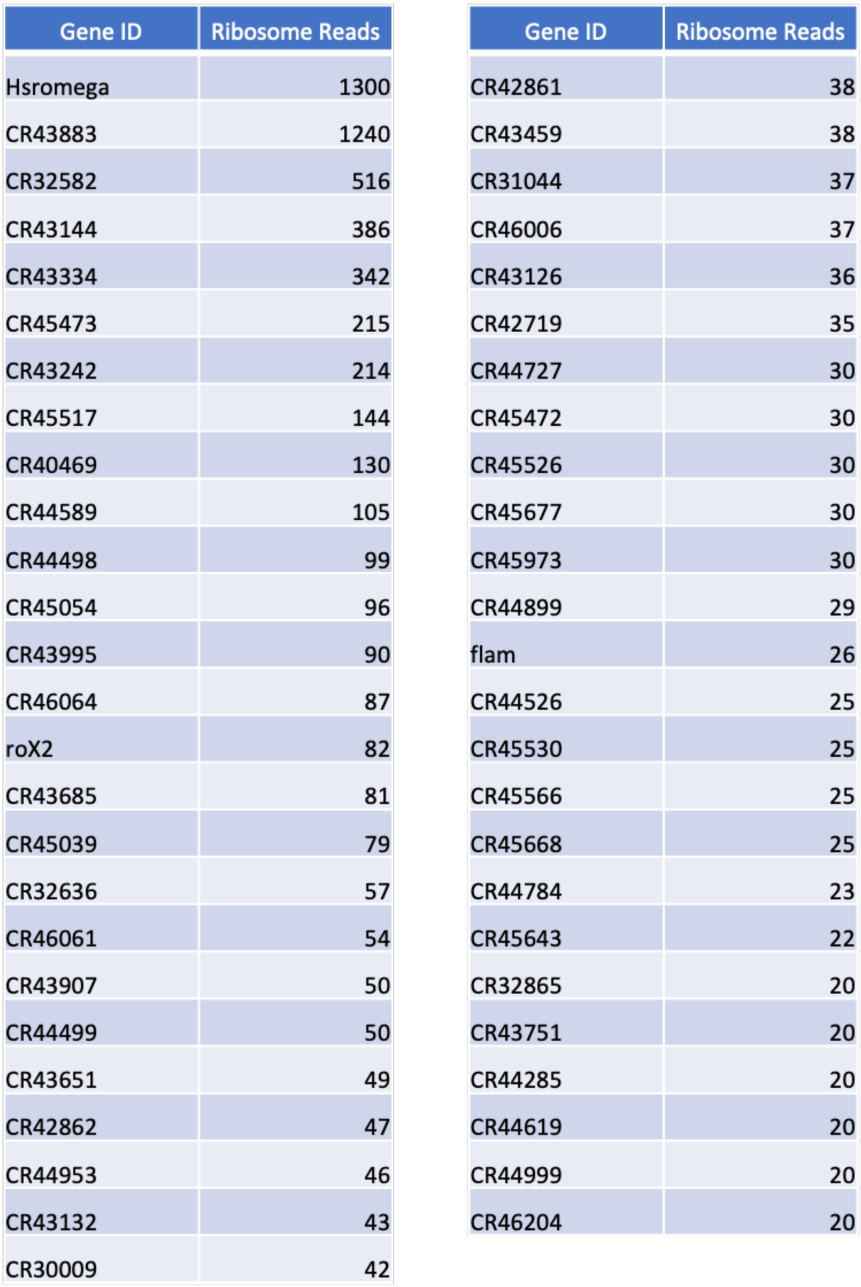
These ribosome-bound lncRNAs include *HsrOmega, CR43883, CR32582, CR43144*, and *CR43334*. Although this is simply normalised relative to total counts, and not by expression ratios or gene length, it will primarily be used to compare to reads found on the same gene in different datasets, making that stage of normalisation unnecessary. Although not meaningful enough to draw significant conclusions from without further comparison, this does provide examples of lncRNAs on the polysome (and given the blocking of translation elongation, likely examples of lncRNAs undergoing translation initiation).

3.5.2.1 Global overview and summary of Couso lab datasets

Extensive work from the Couso lab has developed and used poly-ribo-seq to identify small open reading frames (smORFs) in *Drosophila*, allowing follow up experiments to identify and characterise novel smORF peptides. To this end they carried out poly-ribo-seq on both *Drosophila* S2 cells and multiple developmental points in *Drosophila* embryos. They produced 2 biological replicates for Embryonic Stages 1, 2, and 3 (E1, E2, and E3), which are respectively 0-8, 8-16, and 16-24 hours after egg lay; from which both total RNA and polysome-associated fractions were processed for poly-ribo-seq. The S2 cell dataset had a single replicate, again split into total RNA and polysome associated RNA, which were then processed for poly-ribo-seq. Their extensive analysis used both polysome association and analysis of ribosome phasing to identify promising novel smORFs throughout the genome.

3.5.2.2 Initial examination of elongation inhibited lncRNA reads in *Drosophila* embryonic developmental timepoints and S2 cell poly-ribo-seq data

To start with, the read count and counts per million (CPM) for all lncRNAs were extracted from the Patraquim *et al*. (90) poly-ribo-seq data carried out on samples from *Drosophila* embryos at E1, E2, and E3. These were sorted in order of average number of



Gene ID	Ribosome Reads
Hsromega	1300
CR43883	1240
CR32582	516
CR43144	386
CR43334	342
CR45473	215
CR43242	214
CR45517	144
CR40469	130
CR44589	105
CR44498	99
CR45054	96
CR43995	90
CR46064	87
roX2	82
CR43685	81
CR45039	79
CR32636	57
CR46061	54
CR43907	50
CR44499	50
CR43651	49
CR42862	47
CR44953	46
CR43132	43
CR30009	42

Gene ID	Ribosome Reads
CR42861	38
CR43459	38
CR31044	37
CR46006	37
CR43126	36
CR42719	35
CR44727	30
CR45472	30
CR45526	30
CR45677	30
CR45973	30
CR44899	29
flam	26
CR44526	25
CR45530	25
CR45566	25
CR45668	25
CR44784	23
CR45643	22
CR32865	20
CR43751	20
CR44285	20
CR44619	20
CR44999	20
CR46204	20

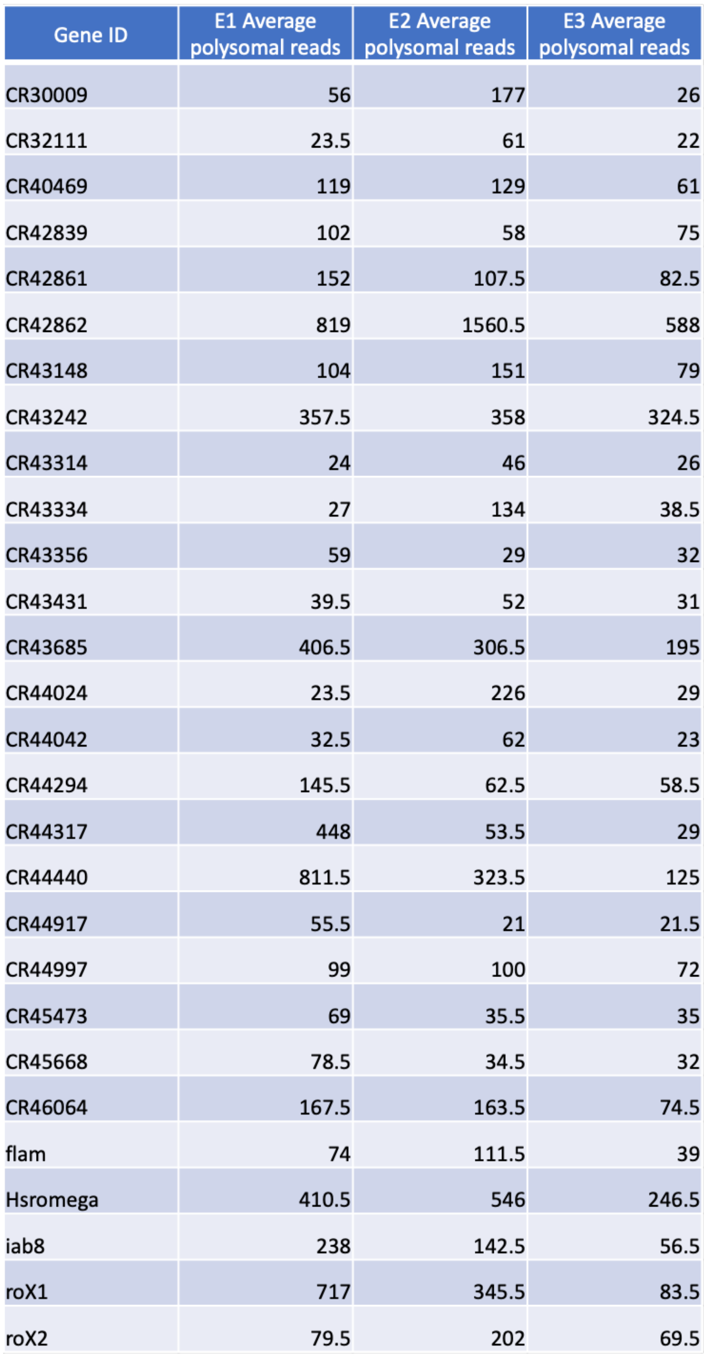
Figure 3.16 – Tables showing lncRNAs present with 20 or more reads in ribosomal RNA from harringtonine treated S2 cells

The 51 lncRNAs listed here were all detected with at least 20 reads in the polysomal RNA for Drosophila S2 cells treated with harringtonine. Data from Zhang et al.

reads detected in polysomal RNA at each developmental timepoint. In order to increase confidence in detection of polysome-association, only lncRNAs with an average of at least 20 reads were kept. Of these (90 in E1, 87 in E2, 47 in E3), those lncRNAs detected with more than 20 reads in all developmental stages were taken, in order to identify candidate lncRNAs likely to be translated consistently and stably throughout *Drosophila* embryogenesis, allowing an interesting comparison to L3 larvae in the novel poly-ribo-seq dataset to be produced in this thesis. This produced a list of 28 lncRNAs detected with at least 20 polysomal reads in *Drosophila* embryos at E1, E2, and E3 developmental timepoints (Figure 3.17). These lncRNAs include *CR30009, CR32111, CR40469, CR42839*, and *CR42861*. Similarly, to 3.5.1.2, this initial work is not enough to draw substantial conclusions, but instead provides further examples of polysome-associated lncRNAs, particularly managing to highlight some that are consistently polysome-associated at different developmental timepoints in *Drosophila* embryo. The persistent association across different developmental timepoints may suggest a functional and pervasive polysomal (and possibly protein-coding) role for some of these lncRNAs.

3.5.3 Summary of initial examination of polysome-associated lncRNAs

The initial examination of polysome associated lncRNAs in *Drosophila* cells and tissues is sufficient only to provide evidence of lncRNAs that do associate with the polysome. Due to the lack of exoribonuclease depleted samples, and the inconsistency in *Drosophila* samples, a unifying dataset (that examines both the translational and degradation aspects simultaneously) is necessary to make meaningful comparisons that would allow stronger identification of candidate lncRNAs. Even so, the evidence from all of these datasets that multiple lncRNAs are present on the polysome is encouraging; and justifies experimental validation of the principle. Given that a small number of candidate Pacman and Dis3L2 regulated lncRNAs have already been identified and validated in L3 WIDs, a follow-up experiment to establish the presence or absence of these candidates in polysomal fractions could then be carried out in order to definitively show the existence of polysomally present lncRNA candidates of both Pacman and Dis3L2.



Gene ID	E1 Average polysomal reads	E2 Average polysomal reads	E3 Average polysomal reads
CR30009	56	177	26
CR32111	23.5	61	22
CR40469	119	129	61
CR42839	102	58	75
CR42861	152	107.5	82.5
CR42862	819	1560.5	588
CR43148	104	151	79
CR43242	357.5	358	324.5
CR43314	24	46	26
CR43334	27	134	38.5
CR43356	59	29	32
CR43431	39.5	52	31
CR43685	406.5	306.5	195
CR44024	23.5	226	29
CR44042	32.5	62	23
CR44294	145.5	62.5	58.5
CR44317	448	53.5	29
CR44440	811.5	323.5	125
CR44917	55.5	21	21.5
CR44997	99	100	72
CR45473	69	35.5	35
CR45668	78.5	34.5	32
CR46064	167.5	163.5	74.5
flam	74	111.5	39
Hsromega	410.5	546	246.5
iab8	238	142.5	56.5
roX1	717	345.5	83.5
roX2	79.5	202	69.5

Figure 3.17 – Table showing polysome associated lncRNAs common to all embryo developmental stages

The 28 lncRNAs listed here are present with 20 or more reads in polysomal RNA from *Drosophila* embryo in stages E1, E2, and E3.

3.6 Testing polysomal presence of lncRNAs of interest:

As previously mentioned, some lncRNAs have been shown to contain small open reading frames (smORFs) which can be translated into short peptides. Although the vast majority of lncRNAs have not been investigated in this manner, and require significant lab work to fully elucidate, a preliminary investigation can be made with readily available bioinformatic tools, based on the candidate Pacman and Dis3L2 degraded lncRNAs identified in 3.6.

In order to identify whether the targets were present on the polysome, a few straightforward molecular and bioinformatic tools can be used. Polysome fractionation followed by RNA extraction and PCR were used to experimentally determine the presesnce of the lncRNAs on the polysome. Subsequently ORFfinder was used, and the sequence of each lncRNA candidate was searched for potential ORFs and any viable candidates were noted.

Polysome fractionation can be used in combination with PCR testing, in order to detect the presence or absence of the lncRNA in paired samples, one from total lysate, and another only containing polysomal RNA. Isogenic control L3 larvae were lysed, separated from tissue debris, and run through a density gradient in order to allow fractionation of only polysomally bound RNAs (described further in methods). An RNA extraction was then carried out on the polysome fractions, alongside total RNA from the same isogenic control replicate taken before the fractionation.

3.7.1 RNA extracted from polysomal fractions retains sufficient integrity for examination:

Following lysis, sample preparation, polysome fractionation, fraction concentration, and RNA extraction (as described in methods), the RNA from each sample underwent RT-PCR in order to provide cDNA. In order to test that the RNA that was used to generate the cDNA had not degraded substantially before this point, the polysomal RNA would be collected as per the described methods, and PCR used to amplify a gene region from a positive control gene, known to be present in both total lysate and polysome fractionated RNA. If the RNA is intact and has retained sufficient integrity, the expected product should be amplified in every sample (discernible by gel electrophoresis size separation). This testing allows us to discount the possibility of RNA degradation being the cause of a lncRNA not showing up. This step is therefore crucial to any conclusions being drawn from the PCR results, and to validating suitability of the techniques used.

PCR was performed on both the total and polysomal wild type L3 samples following the RT-PCR, to test if the cDNA was intact and that the precursor RNA had not been significantly degraded. *rp49*, a transcript that encodes for a ribosomal protein, was used as a positive translational control as it is known to be actively translated in third instar larval tissue. Figure 3.18 shows an example of an imaged gel following PCR. Clear bands were seen at the correct size for all samples, indicating that *rp49* can be amplified from the samples by PCR, and the RNA used to make cDNA for the PCR has not been degraded.

3.7.2 PCR shows the presence of *CR42719* transcripts on the polysome and multiple putative ORFs are predicted

Having successfully extracted total and polysome associated RNA, the experiment could then progress to testing for presence of the Pacman and Dis3L2 sensitive lncRNAs by PCR amplification; with primers for *CR42719* was performed on wild type L3 cDNA for both the total lysate and polysomal samples, using primers for the lncRNA *CR42719* following RT-PCR. As a transcript expressed at a very low level, the input concentration of template cDNA required optimisation to reach a level that allowed visualisation of the lncRNA, at least in the total lysate PCR, in which *CR42719* is known to be expressed. Having increased starting concentration to 100ng/µl, a band of the expected size (143bp) could be seen in both the total lysate and polysome fraction reactions (Figure 3.19, panel (a) and (b)), showing *CR42719* to be present on the polysome, and thus, likely translated. Several additional bands were seen, likely due to the high concentration of starting material and high cycle number required to visualise a lncRNA expressed at such a low level. The non-specific bands were not seen in the (reverse transcriptase negative) negative control reaction. By inputting the known sequence of *CR42719* into ORFfinder, the gene was searched for any ORFs that could potentially be

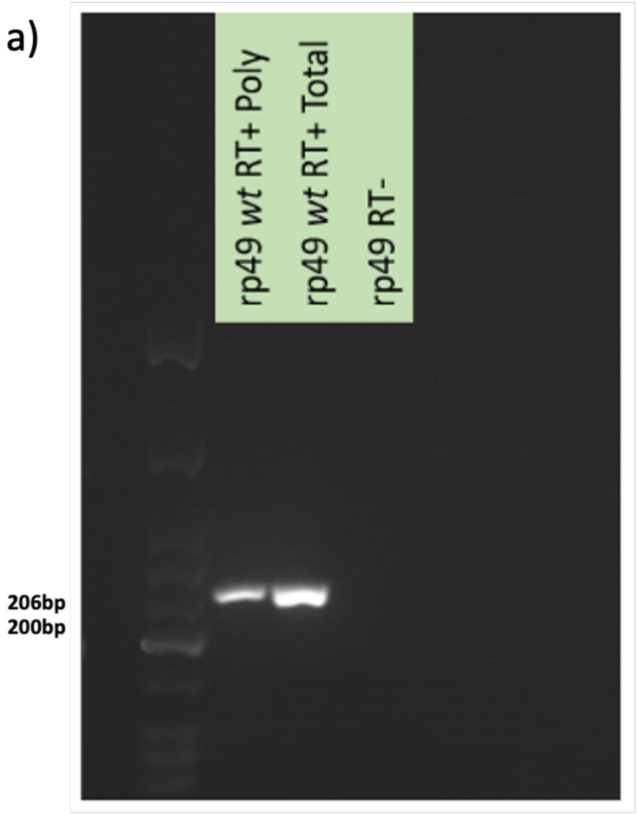
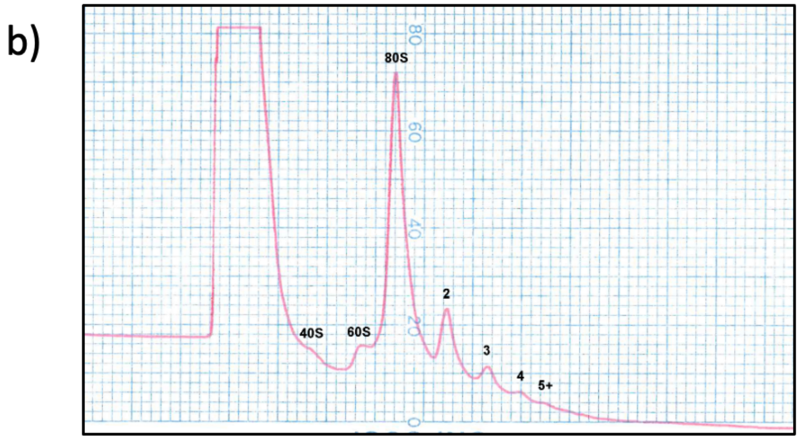


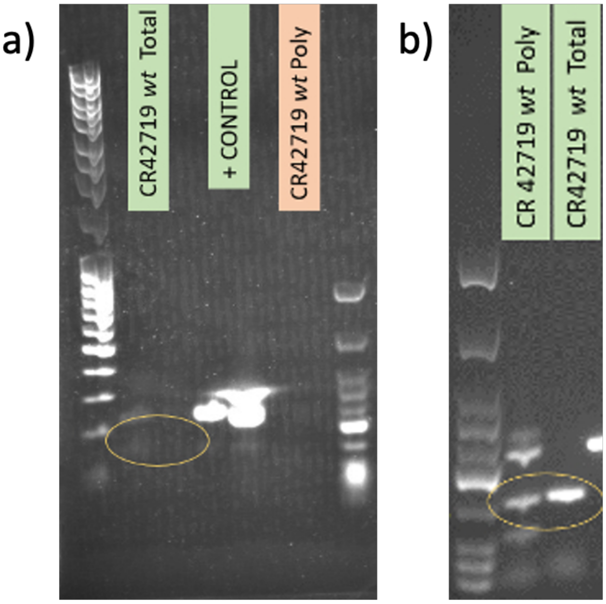
Figure 3.18 - Testing RNA integrity in L3 samples using *rp49*:

a) PCR products (40 cycles, samples at 200ng/µl)) amplified from cDNA (from a reverse-transcription reaction on the extracted RNA) from wild type whole larvae that had gone through polysome fractionation and sucrose extraction. Both the RNA from pooled polysome fractionss (2+ ribosomes per transcript), as well as total RNA extracted from a paired sample before fractionation, were shown to maintain RNA integrity, by the presence of a single rp49 band of the expected size (206bp).

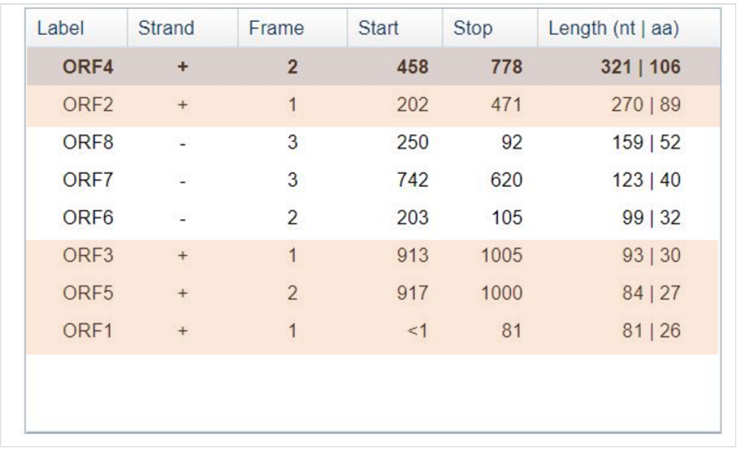
b) Representative profile from polysome fractionated RNA, tested for RNA integrity by PCR.



(Some data presented in this figure is from work carried out by Lauren Mulcahy, supervised by Oliver Rogoyski)



c)



Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF4	**+**	**2**	**458**	**778**	**321 | 106**
ORF2	+	1	202	471	270 | 89
ORF8	-	3	250	92	159 | 52
ORF7	-	3	742	620	123 | 40
ORF6	-	2	203	105	99 | 32
ORF3	+	1	913	1005	93 | 30
ORF5	+	2	917	1000	84 | 27
ORF1	+	1	<1	81	81 | 26

Figure 3.19 – Observed presence of *cr42719* on the polysome

(a) For the wild type total lysed RNA sample, a band was seen at the expected size of 143bp, shown circled, but this was very faint. This was only visible in one of the two replicates. The positive control *Drosophila* RNA sample (established in previous experiments to contain intact RNA that can amplify an *rp49* product) showed two strong bands of the expected size, meaning that the PCR had been successful.

(b) The PCR was repeated at a higher concentration of template cDNA. Two bands of the expected size were seen in both the polysomal and total samples, suggesting this lncRNA is present on the polysome. A non-specific band, under 100bp, was seen in the polysome sample alongside the expected band.

(c) Potential ORFs for cr42719. Highlighted are those found on the positive strand.

(Some data presented in this figure is from work carried out by Lauren Mulcahy, supervised by Oliver Rogoyski)

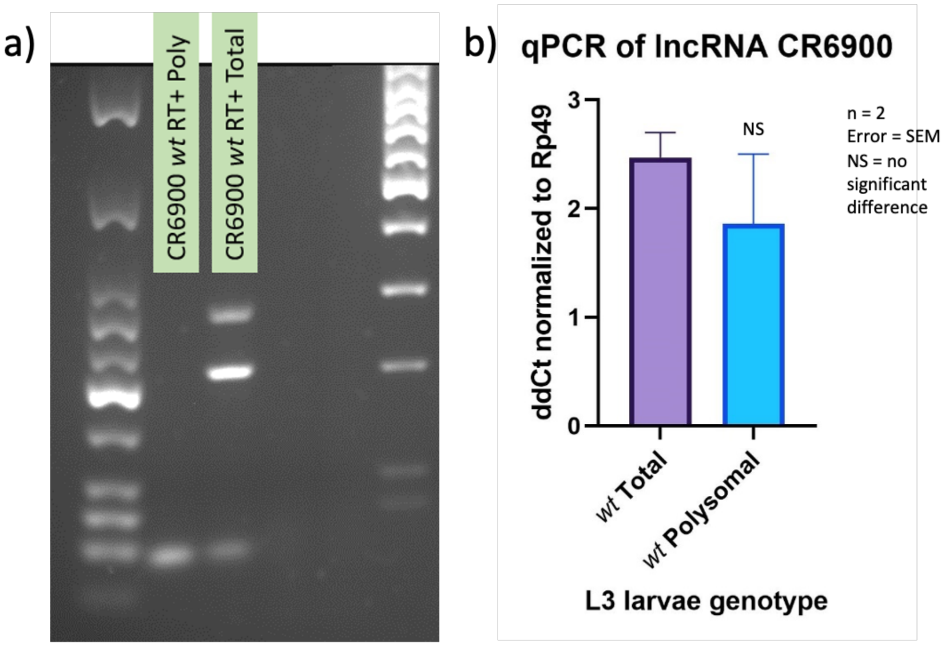
undergoing translation. For *CR72719*, 5 potential ORFs were found, ranging from 26-231 nucleotides in length (Figure 3.19, panel (c)).

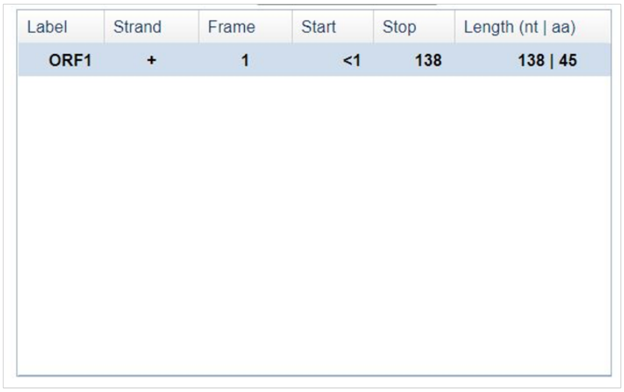
It should be noted that no reads for *CR42719* were found in the novel poly-ribo-seq data in whole L3 larvae (described in Chapter 6), however 35 reads were found in the harringtonine-treated S2 cell ribo-seq dataset from Zhang et al. (154). Given that *CR42719* has showed up in both L3 larval RNA (by PCR) and in the harringtonine-treated S2 cell ribo-seq (in sequencing data), this lends additional support to the polysome-association and possible translation of the gene. Having been detected by PCR in the same type of sample, we would expect to see reads in the L3 larvae poly-ribo-seq data. The low read depth from this sequencing data may explain the lack of reads, as PCR is able to amplify a product from extremely low concentration, and the higher read depth of the dataset from Zhang *et al*. would be more sensitive to lowly expressed lncRNAs (and expression may be higher in S2 cells).

3.7.3 PCR shows *CR6900* not to be present on the polysome and only a single putative ORF is predicted

PCR amplification with primers for *CR6900* was performed on wild type L3 cDNA for both the total lysate and polysomal samples, using primers for the lncRNA *CR6900* following RT-PCR. Figure 3.20, panel (a) shows the imaged gel. The expected band was seen in the total samples, but not in the polysomal samples, which would suggest lncRNA *CR6900* is not present on the polysome.

Further testing of *CR6900* was done using qPCR, as this technique is much more specific than standard PCR and uses a different set of primers which may illuminate any artefacts caused by issues with initial primer binding. The qPCR amplification plot is shown in Figure 3.20, panel (b) and shows that the *CR6900* transcript was being amplified in both the total and polysomal samples after approximately 26 cycles. This does not support the previous data which suggested *CR6900* is not present on the polysome. This could be due to the increased sensitivity of TaqMan qPCR allowing detection of lowly expressed lncRNA.





Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF1	+	1	<1	138	138 | 45

Figure 3.20 – Inconsistent observed presence of *cr6900* on the polysome.

(a) One band of the expected size was seen in the total sample; non-specific bands were also seen alongside this. No relevant bands were seen in the polysomal sample, suggesting that the lncRNA is not present on the polysome, or is present only at very low levels.

(b) Observed presence of cr6900 on the polysome. Normalised to rp49, showing the relative amount of lncRNA transcripts in the total and polysomal samples of wildtype L3 larvae. qPCR data shows that lncRNA cr6900 is present on the polysome in similar abundance to that seen in the total sample, contradicting the sqPCR.

(c) Figure 3.13 – Potential ORF for *cr6900*: Found on the positive strand.

(Some data presented in this figure is from work carried out by Lauren Mulcahy, supervised by Oliver Rogoyski)

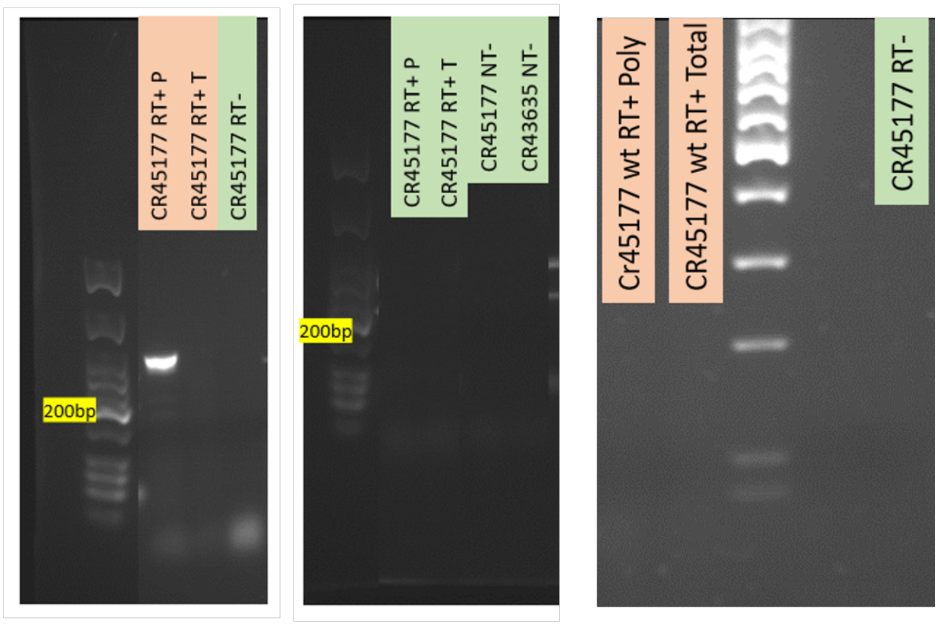
Although it could also be a result of non-specific amplification, TaqMan qPCR primers are highly specific, reducing the chance that this is the cause. A 138nt ORF is predicted by ORFfinder, as shown in Figure 3.20, panel (c).

It should be noted that extremely low levels of reads for *CR6900* were found in the novel poly-ribo-seq data in whole L3 larvae, in both isogenic control (an average of 1 read) and Dis3L2 mutant (an average of 0.5 reads) genotypes. Similarly, an extremely low number of reads (2 reads) were found in the harringtonine-treated S2 cell ribo-seq dataset from Zhang *et al*. (154). Given that *CR6900* has showed up in both L3 larval RNA (by PCR and poly-ribo-seq data) and in the harringtonine-treated S2 cell ribo-seq (in sequencing data), this lends additional support to the polysome-association and possible translation of the gene. The read counts are low enough to not provide a high level of confidence, and further sequencing at a higher depth would help to elucidate the presence or absence of *CR6900* on the polysome. The low read depth from this sequencing data may explain the lack of reads, as PCR is able to amplify a product from extremely low concentration, and the higher read depth of the dataset from Zhang *et al*. would be more sensitive to lowly expressed lncRNAs (and expression may be higher in S2 cells).

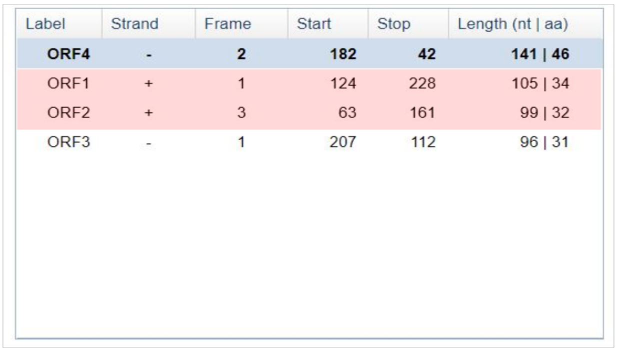
3.7.4 PCR is inconclusive for the presence of *CR45177* transcripts on the polysome and multiple putative ORFs are predicted

PCR amplification with primers for *CR45177* was performed on wild type L3 cDNA for both the total lysate and polysomal samples. In all tested replicates, no relevant bands were seen in either the polysomal or total samples (Figure 3.21, panel (a)). This could suggest that this lncRNA is expressed at very low levels. The primers have previously been tested and shown to bind and amplify a product of the correct size. The template cDNA used for this experiment is the same as that used to test for the presence of other candidate lncRNAs, suggesting this not to be an issue with either the template. It is possible that the expression of *CR45177* is low enough that that it can’t be picked up by PCR at even this relatively high concentration of RNA, in either total or polysomal RNA pools.

a)



b)



Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF4	**-**	**2**	**182**	**42**	**141 | 46**
ORF1	+	1	124	228	105 | 34
ORF2	+	3	63	161	99 | 32
ORF3	-	1	207	112	96 | 31

Figure 3.21 – Inconclusive observed polysomal presence of lncRNA *cr45177*:

(a) In all but one polysomal sample, no bands were seen in either the total or polysomal samples. The band seen in the polysomal sample is not of the expected size. This suggests an issue with the primers as we would expect to see *cr45177* in the total sample regardless of its presence on the polysome, however, these primers have previously been successfully used to amplify a product of the correct size, indicating that the expression of *cr45177* may just be too low to pick up in these sample.

(b) Potential ORFs for *cr45177*. Highlighted are those found on the positive strand

(Some data presented in this figure is from work carried out by Lauren Mulcahy, supervised by Oliver Rogoyski)

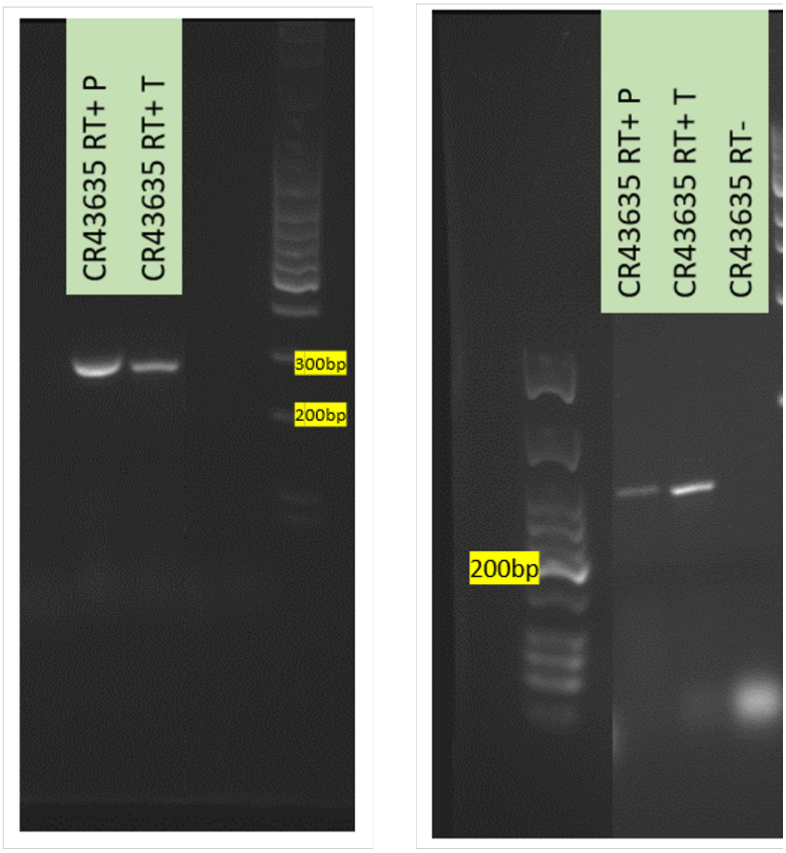
By inputting the known sequence of *CR45177* into ORFfinder, the gene was searched for any ORFs that could potentially be undergoing translation. For *CR45177*, 4 potential ORFs were found, with two on the positive strand. These ORFs were of 99 and 105 nucleotides in length, respectively (Figure 3.20, panel (b)).

It should be noted that extremely low levels of reads for *C45177* were found in the novel poly-ribo-seq data in whole L3 larvae, in the Dis3L2 mutant (an average of 2 reads) genotypes. However, no reads were found in the harringtonine-treated S2 cell ribo-seq dataset from Zhang et al. (154). Given that *CR45177* has showed up in L3 larval RNA (by PCR and inpoly-ribo-seq data), this lends additional support to the polysome-association and possible translation of the gene. The read counts are low enough to not provide a high level of confidence, and further sequencing at a higher depth would help to elucidate the presence or absence of *CR45177* on the polysome. Although the data from Zhang *et al*. has a higher read depth and is generally more sensitive to detecting the expression of a given gene, it must be remembered that tissue-specific expression of genes is to be expected in many instances, and it may be the case that S2 cells do not express *C45177*.

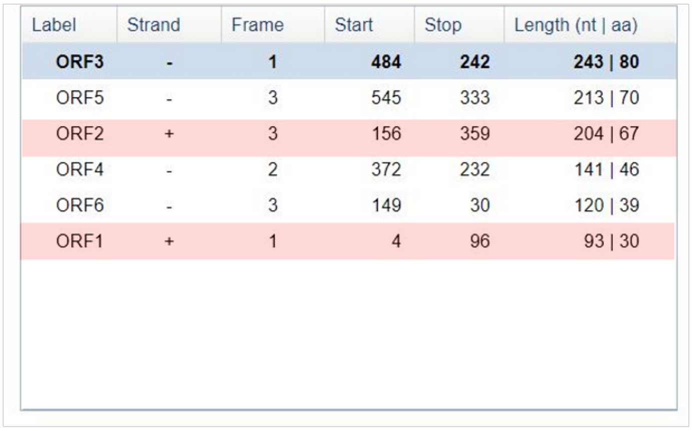
3.7.5 PCR shows the presence of *CR43635* transcripts on the polysome and multiple putative ORFs are predicted

PCR amplification with primers for *CR43635* was performed on wild type L3 cDNA for both the total lysate and polysomal samples. Figure 3.22, panel (a) shows the imaged gels of two replicates. A band of the expected size was seen in both the total and polysomal samples which suggests this lncRNA is present on the polysome. Further validation using qPCR would be useful but suitable primers for *CR43635* could not be produced by the manufacturer. TaqMan qPCR probes for *Drosophila* are not readily available for every gene, and especially for poorly annotated lncRNAs. While ordinary primers require only a complimentary region, and so can be easily designed and ordered to fit any sequence, the more complex fluorescence-quencher system of TaqMan probes must be ordered from the manufacturer, who will not readily give out a protocol to custom design probes for this system.

a)



b)



Label	Strand	Frame	Start	Stop	Length (nt | aa)
ORF3	**-**	**1**	**484**	**242**	**243 | 80**
ORF5	-	3	545	333	213 | 70
ORF2	+	3	156	359	204 | 67
ORF4	-	2	372	232	141 | 46
ORF6	-	3	149	30	120 | 39
ORF1	+	1	4	96	93 | 30

Figure 3.22- Observed presence of lncRNA *cr43635* on the polysome:

(a) Shown for two replicates. In both gels, a band of the expected size was seen in the polysomal and total cDNA samples. No bands were seen in the no-RT template or the no template control., This suggests that *cr43635* is present on the polysome.

(b) Potential ORFs for cr43635. Highlighted are those found on the positive strand.

(Some data presented in this figure is from work carried out by Lauren Mulcahy, supervised by Oliver Rogoyski)

By inputting the known sequence of *CR43635* into ORFfinder, the gene was searched for any ORFs that could potentially be undergoing translation. For *CR43635*, 6 potential ORFs were found, with two on the positive strand (Figure 3.22, panel (b)). These ORFs were of 93 and 204 nucleotides in length, respectively. Reads were not found on this gene in the data from Zhang *et al* (154).

It should be noted that no reads for *CR43635* were detected in the novel poly-ribo-seq data in whole L3 larvae, or in the harringtonine-treated S2 cell ribo-seq dataset from Zhang et al. (154). Given that *CR43635* has showed up in L3 larval RNA (by PCR), but not in either set of sequencing data, this only lends limited support to the polysome-association and possible translation of the gene. It is possible that further sequencing at a higher depth would help to elucidate the presence or absence of *CR43635* on the polysome, as PCR amplification is able to produce detectable levels of a product from an extremely low concentration of the target in template cDNA. It is also highly possible that the primers used for PCR testing may be binding off-target and producing a product of a similar size. Without the TaqMan qPCR probes available, it is hard to put as much confidence in the PCR results. Use of a second set of primers complimentary to the target region, or sequencing of the product, could also help to increase confidence as to whether the detection by PCR is a real result.

3.7.6 Molecular testing identifies multiple candidates for exoribonuclease degraded lncRNAs present in polysomal RNA

Throughout the previous sections (3.4 to 3.4.5 and 3.7.1 to 3.7.6), variants of PCR were conclusively used in order to prove that previously identified candidate lncRNAs (identified as degraded by Pacman or Dis3L2 by previous sequencing data), are verifiably both degraded by Pacman and Dis3L2, but also present on the polysome. This proves the principle that certain lncRNA transcripts are likely to be specifically degraded by these exoribonucleases, and also play some role (possibly translated) on the polysome. Although this is only a preliminary explanation, it justifies and informs a follow-up experiment that requires a genome-wide, high-throughput approach to look at total lysate and polysomal RNA in both Pacman mutant, Dis3L2 mutant, and wild-type control *Drosophila* samples.

3.8 Overview of results

3.8.1 Re-analysing RNA-seq data from Antic et al:

Thorough and critical analysis of the Antic et al. data on a similar experiment uncovered some points of weakness in the initial analysis. Only two control replicates were used, severely limiting how meaningful the statistical analysis carried out on this data can be. It was also noted that XRN1/Pacman was not shown to be significantly downregulated at the RNA level in their XRN1/Pacman dsRNA treated S2 cells, although they did provide a Western blot showing strong, although unquantified knockdown at a protein level.

By mapping their raw data to a newer genome build and changing selection criteria so as to include (and indeed focus on) lncRNAs, several interesting candidate lncRNAs were found. These include *CR45643, CR45162, CR44568, CR45195*, and *CR43264*. These data will allow better comparison to other existing (and upcoming) data sets, allowing variability between different labs, cell lines, tissues, etc. to be seen.

3.8.2 Analysing previous Newbury Lab RNA-seq experiments uncovered candidate lncRNAs, differentially expressed in XRN1/Pacman and Dis3L2 null mutants:

The existing Newbury lab data from *Drosophila* wing discs provided a snapshot of regulation in a different tissue and developmental context. Although the depth of sequencing, and mapped genome (among other things) will also contribute to any differences that are detected, given the notably different lists of candidates produced by each, it seems likely that Pacman and Dis3L2 might have a broad range of specific and context-dependent targets. This work has then found specific candidate lists of at least some of the lncRNAs sensitive to Pacman and Dis3L2. Now that lists of candidates sensitive to degradation by Pacman, Dis3L2, or both have been identified, further examination can always be carried out to maximise the data extracted from them. For example, looking at these data sets while sorting targets by translational efficiency might elucidate how translational activity impacts rate and specificity of degradation.

3.8.3 Verification with semi-quantitive PCR (sqPCR), and subsequent qPCR supported the exoribonuclease-dependent differential abundance of several candidate lncRNAs:

The variation between data sets, between replicates, and low fold-changes give significant reason to validate any interesting findings with additional methods and using a sensitive and time-proven technique like qPCR allowed validation with a comparatively simple workflow, and allows higher replicate values to be achieved than is financially viable in RNA-seq experiments. Whilst the workflow is relatively simple, TaqMan probes are costly enough that it can be prohibitively expensive to qPCR screen all interesting candidates. The cheaper SYBR-green could have been used, but this would have significantly reduced specificity. As such, the cheaper but less accurate method of sqPCR was used to start, with promising results followed up by TaqMan qPCR.

As shown in this initial screen, lncRNA candidates identified by RNA-seq could be verified by these techniques. With a moderate number of qPCR replicates, each internally normalized to a housekeeper gene like *rp49*, statistical significances could be established to a high degree of confidence. The example lncRNAs identified here act both as potentially interesting targets to look at in the context of future experiments, and as proof of principle of lncRNA degradation by Pacman and Dis3L2. It also demonstrated that a validation step is useful and necessary, as not all candidates saw a significant difference according to the RNA-seq data.

3.8.4 Translational activity of exoribonuclease regulated lncRNAs:

Two lncRNAs, *CR42719* and *CR43635*, were successfully shown to be present within the polysomal RNA samples obtained from wild-type whole L3 larvae. *CR6900* showed conflicting results from PCR and qPCR results, and therefore no definitive conclusions can be drawn regarding its presence on the polysome without further investigation. The contradictory results shown by PCR and qPCR for lncRNA *CR6900* are unexpected but higher confidence can be placed in the qPCR data due to its higher specificity and sensitivity. *CR45177* produced no conclusive results for polysomal association within this project but could possibly be expressed at a very low level. All tested lncRNAs were found to have at least one potential ORF (although many suggested ORFs are not biologically usable), as identified by ORFfinder, and summarised in Figures 3.19 – 3.22. When examined in the novel poly-ribo-seq data (in L3 larvae) and the existing ribo-seq data (in harringtonine-treated S2 cells, by *Zhang et al*. (154)), some reads were found associated with the polsyomal and monosomal samples (respectively). The lncRNAs *CR6900* and *C45177* were detected (albeit at extremely low levels) in the polysomal fractions of at least one replicate of at least one genotype. The lncRNAs *CR42719* and *CR6900* were detected (albeit at an extremely low level for *CR6900*) in the polysomal fractions of at least one replicate of at least one genotype. It is interesting to examine these data, and to see how it lends some small support to the presence of the aforementioned lncRNAs on the polysome, but it must be acknowledged that further replicates at a higher read depth would be required to be confident in the assessment of these data.

This work shows then, that at least 2 of the lncRNAs subject to specific degradation by exoribonuclease mutants are present on the polysome of wild-type L3 whole larvae. This might indicate that these 2 lncRNAs are actively translated into smORF peptides; as their presence in polysome fractions removes the possibility of spurious and transient association with individual ribosomes. This is further supported by the fact they have 4 and 2 predicted ORFs respectively, also providing specific regions to further explore for novel smORF encoding genes. The fact these lncRNA transcripts were shown to be upregulated in *dis3L2* or *pacman* mutants suggests they are degradation targets of these exoribonucleases, and their presence on the polysome could be initial evidence that these for co-translational degradation of lncRNAs and justifies further exploration of the idea.

It should be noted that several of the sqPCR and qPCR experiments, quantifying the levels of candidate lncRNAs, and their presence on the polysome, were carried out by undergraduate students (Harry Pink and Lauren Mulcahy) under the candidate’s supervision.

Chapter 4: Results – Manipulating levels of Pacman and Dis3L2 in *Drosophila* S2 cells in order to develop a model for exoribonuclease depleted poly-ribo-seq

4.1 Introduction

In Chapter 3, re-analysis of high-throughput sequencing datasets was used in order to identify multiple shortlists of lncRNAs that were differentially expressed in the absence of Pacman and Dis3L2 (implying the roles of these exoribonuclease in their regulation and degradation), as well as showing some of them to be present on the polysome, and even identifying read pileups near potential ORFs in *Drosophila* S2 cells treated with harringtonine. By following some of these candidate lncRNAs with molecular techniques (polysome fractionation, RNA extraction, sqPCR, and qPCR,) concrete examples of exoribonuclease-degraded lncRNAs (some of which were present on the polysome, possibly even undergoing translation) were discovered, such as *CR42719* and *CR43635*.

With some preliminary data showing at least some lncRNAs to be both differentially abundant in the absence of either Pacman or Dis3L2, this project aimed to pursue the use of a high throughput method (poly-ribo-seq) that would allow genome-wide screening of both these, and other candidates. In order to do this, a large amount of *Drosophila* genetic material must be generated, for wild type, Pacman depleted, and Dis3L2 depleted samples. This was pursued with three different approaches.

Previous work by Antic *et al*. (69) has used dsRNAi to induce a transient knockdown of *Pacman* in *Drosophila* S2 cells by use of dsRNA treatment, particularly promising in *Drosophila* cells as they lack the interferon response. Although not as revealing as a total knockout or null mutation of a gene, replicating this in S2 cells (as well as attempting to knockdown *dis3L2* in the same way), was attempted, due to the ease of growing large amounts of S2 cells. *Drosophila* S2 cells are an embryonically derived cell line from dissociated embryos, near to hatching used extensively in *Drosophila* work, including to understand RNA decay in work by Izzauralde *et al*. (157-159)

CRISPR-Cas9 has been used extensively in *Drosophila* in recent years (including in S2 cells), in order to generate a wide range of specific mutations. Although growing a stable stock from a single suitably mutated cell can be both time-consuming and fraught with difficulty, the promise of a null mutant stock that can be rapidly amplified at will justified attempting to generate exoribonuclease null mutants in S2 cells by use of CRISPR.

Existing null mutants for both *pacman* and *dis3L2* exist in whole *Drosophila* (see section 3.3.2) Pacman null mutants are not viable, dying during the pupal stage of development. As direct comparison between the genotypes would offer the greatest value as a dataset, this made adult *Drosophila* tissues a poor choice. Several of the existing RNA-seq datasets in whole *Drosophila*, including those that acted as foundations for this work, was carried out on WIDs from L3 larvae, this developmental timepoint was deemed a good option. Whole larvae were used, rather than WIDs, due to the large quantity of genetic material required by poly-ribo-seq, and the time consuming nature of dissecting and isolating WIDs, which yields minimal RNA in comparison to the whole L3. The downside of this option, and why it was not initially undertaken in this chapter, is that a protocol for polysome fractionation of late-stage (wandering) L3, the desired developmental point, had never been attempted, although techniques were available for polysome fractionation of *Drosophila* embryonic material. This meant that although the exoribonuclease mutants were readily available, a refined protocol to produce high resolution polysome traces and fractions needed to be developed to work with this kind of sample. The potential for comparison of large-scale data sets, as well as enabling future polysomal work on L3, provided reason enough to attempt optimization of this process (further discussed in Chapter 5).

4.2 Project background and aims

Although some functional methods of generating exoribonuclease depleted *Drosophila* genetic material have been established; for this project, a method of generating large quantities of genetic material of this kind is needed. Similar and relevant existing data exists in *Drosophila* S2 cells, L3 larval wing imaginal discs, and whole embryos, which will all allow for comparisons. Exoribonuclease depleted RNA-sequencing data is available in L3 WIDs, although not poly-ribo-seq data. This model, however, is difficult to gather large amounts of. The use of S2 cells, on the other hand, seems viable (as Pacman knockdown has previously been achieved using dsRNA (69). In addition, with the powerful CRISPR-Cas9 system being used increasingly for specific generation of mutant lines (including in the generation of the whole *Drosophila* Dis3L2 null mutant used here), this was also explored as an efficient way of generating exoribonuclease deficient mutants (in S2 cell lines). All of these models seem worth exploring but will require significant work in order to overcome the potential issues that might come with either.

The main stages to this work are as follows:

1) Attempt to recreate a substantial and significant knockdown of Pacman in *Drosophila* S2 cells, using dsRNA treatment.
2) Attempt to expand the use of the aforementioned dsRNA protocol to knockdown of Dis3L2 in the same S2 cell model.
3) Explore use of the CRISPR-Cas9 system to generate null mutants S2 cell lines for Pacman and Dis3L2.
4) Evaluate cell line exoribonuclease depletion versus the potential use of whole *Drosophila* tissues.

4.3 Exploration of the use of dsRNA to induce exoribonuclease knockdown in *Drosophila* S2 cells

Drosophila S2 cells are an easy to grow, commonly used cell line, with multiple RNA-seq datasets available in a range of conditions. Their easy and rapid growth allows for quick amplification of stocks, and generation of genetic material. The presence of so many datasets in S2 cells provides invaluable resources for cross-comparisons, for example with Harringtonine treated S2 cells (154), further detailed in Chapter 6. In addition, a protocol for dsRNA knockdown (specifically of Pacman, was already available (69), providing a straightforward, pre-optimised method to achieve the desired knockdown, and a promising starting point for knocking down Dis3L2.

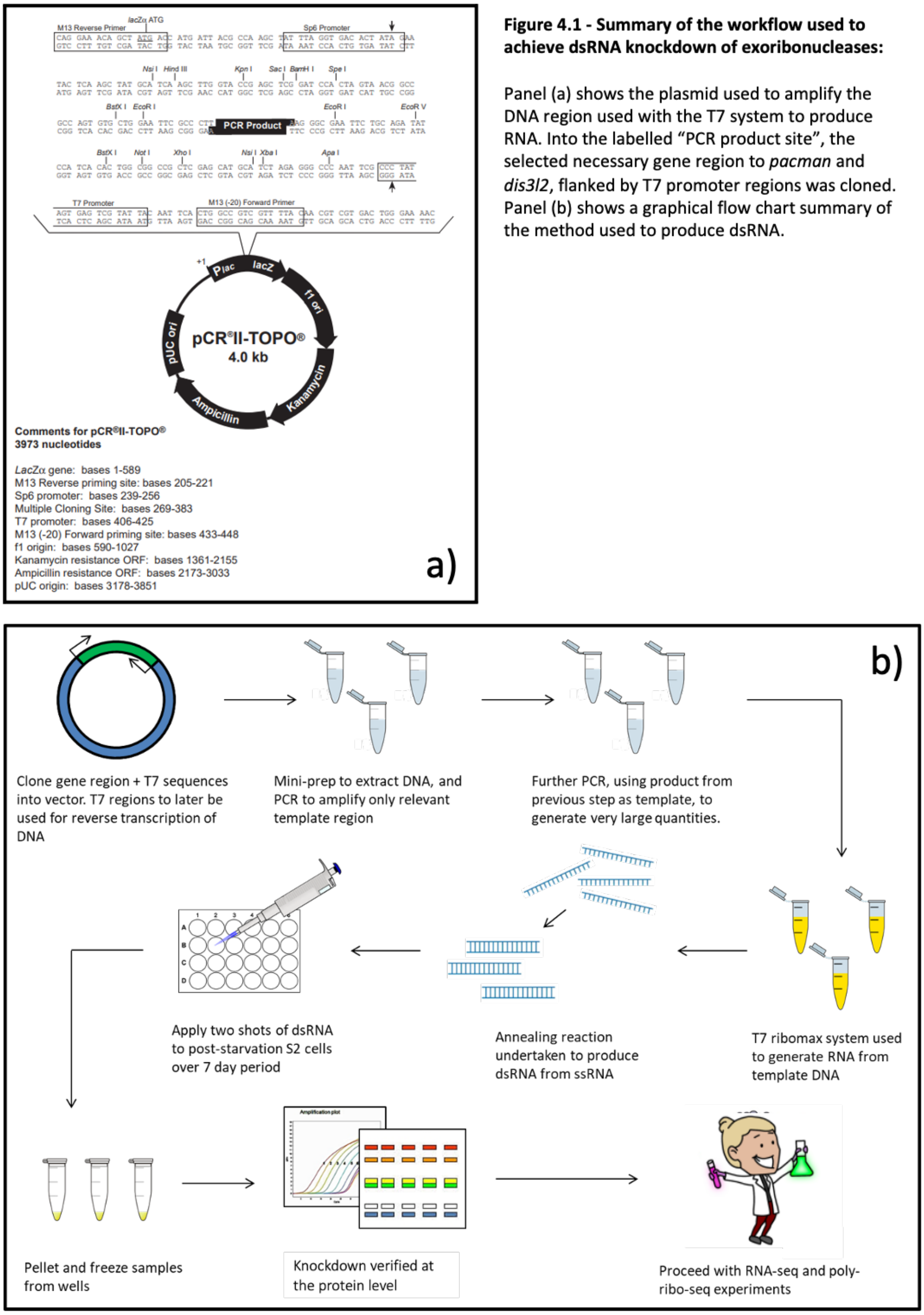
4.3.1 Pacman and dis3L2 were successfully cloned into competent TOP10 *E. coli* cells

The first aim of this part of the project was to attempt the amplification and cloning of a portion of the *pacman* and *dis3L2* genes into TOP10 *E. coli* cells, in order to allow amplification of *pacman* and *dis3l2* gene regions flanked by T7 promoters, which (combined with the use of the T7 RiboMax system) would allow synthesis of RNA complimentary to these regions (summarised in Figure 4.1). The successful completion of this step, verified by DNA sequencing, would allow the project to proceed onto attempting the actual knockdown. This was initially attempted in DH5α cells, but after seeing virtually no growth on selective plates (implying low transformation efficiency, and therefore a lack of Ampicillin resistance), the experiment was carried out in TOP10 cells (which have a higher transformation efficiency).

Following this, several colonies from each transformation reaction were selected and amplified separately. After growing these colonies up in Ampicillin-containing broth (in order to select for only transformed bacteria, which would contain the Ampicillin resistance gene), they were pelleted, and underwent DNA extraction, followed by a PCR amplification of the target region of *pacman* or *dis3L2*, as appropriate. Gel electrophoresis of the PCR products then allowed identification of colonies that contained the target region (Figure 4.2, panel (a)). These rest of the broth, containing stocks from individual colonies, then underwent a mini-prep, and were sent for DNA sequencing (Figure 4.2, panel (b)). After analysis, one colony for *pacman* and one for *dis3L2* were selected based on having complete sequence similarity between the plasmid insertion site and the known target region from *pacman* or *dis3l2* (Figure 4.2, panel (c)).

4.3.2 Successful amplification of the relevant section of the *pacman* and *dis3L2* genes

High-Fidelity Polymerase Chain Reaction (HF-PCR) was used to generate template DNA from the region of target DNA cloned into the pCRII-TOPO plasmid (Figure 4.1 panel (a)). In some instances, non-specific bands were seen, so gel extraction was used to excise


lacZα ATG
M13 Reverse Primer
Sp6 Promoter
CAG GAA ACA GCT ATG AGC ATG ATT ACG CCA AGC TAT TTA GGT GAC ACT ATA GAA
GTC CTT TGT CGA TAC TCG TAC TAA TGC GGT TCG ATA AAT CCA CTG TGA TAT CTT
Nsi I Hind III Kpn I Sac I BamH I Spe I
TAC TCA AGC TAT GCA TCA AGC TTG GTA CCG AGC TCG GAT CCA CTA GTA ACG GCC
ATG AGT TCG ATA CGT AGT TCG AAC CAT GGC TCG AGC CTA GGT GAT CAT TGC CGG
BstX I EcoR I EcoR I EcoR V
GCC AGT GTG CTG GAA TTC GCC CTT PCR Product AAG GGC GAA TTC TGC AGA TAT
CGG TCA CAC GAC CTT AAG CGG GAA TTC CCG CTT AAG ACG TCT ATA
BstX I Not I Xho I Nsi I Xba I Apa I
CCA TCA CAC TGG CGG CCG CTC GAG CAT GCA TCT AGA GGG CCC AAT TCG CCC TAT
GGT AGT GTG ACC GCC GGC GAG CTC GTA CGT AGA TCT CCC GGG TTA AGC GGG ATA
T7 Promoter
M13 (-20) Forward Primer
AGT GAG TCG TAT TAC AAT TCA CTG GCC GTC GTT TTA CAA CGT CGT GAC TGG GAA AAC
TCA CTC AGC ATA ATG TTA AGT GAC CGG CAG CAA AAT GTT GCA GCA CTG ACC CTT TTG
+1
Plac
lacZ
f1 ori
Kanamycin
Ampicillin
pUC ori
pCR®II-TOPO®
4.0 kb
a)
b)
Amplification plot
1 2 3 4 5 6
A
B
C
D

Comments for pCR®II-TOPO®
3973 nucleotides

LacZα gene: bases 1-589
M13 Reverse priming site: bases 205-221
Sp6 promoter: bases 239-256
Multiple Cloning Site: bases 269-383
T7 promoter: bases 406-425
M13 (-20) Forward priming site: bases 433-448
f1 origin: bases 590-1027
Kanamycin resistance ORF: bases 1361-2155
Ampicillin resistance ORF: bases 2173-3033
pUC origin: bases 3178-3851

Clone gene region + T7 sequences into vector. T7 regions to later be used for reverse transcription of DNA

Mini-prep to extract DNA, and PCR to amplify only relevant template region

Further PCR, using product from previous step as template, to generate very large quantities.

T7 ribomax system used to generate RNA from template DNA

Annealing reaction undertaken to produce dsRNA from ssRNA

Apply two shots of dsRNA to post-starvation S2 cells over 7 day period

Pellet and freeze samples from wells

Knockdown verified at the protein level

Proceed with RNA-seq and poly-ribo-seq experiments

Figure 4.1 - Summary of the workflow used to achieve dsRNA knockdown of exoribonucleases:

Panel (a) shows the plasmid used to amplify the DNA region used with the T7 system to produce RNA. Into the labelled “PCR product site”, the selected necessary gene region to *pacman* and *dis3l2*, flanked by T7 promoter regions was cloned. Panel (b) shows a graphical flow chart summary of the method used to produce dsRNA.

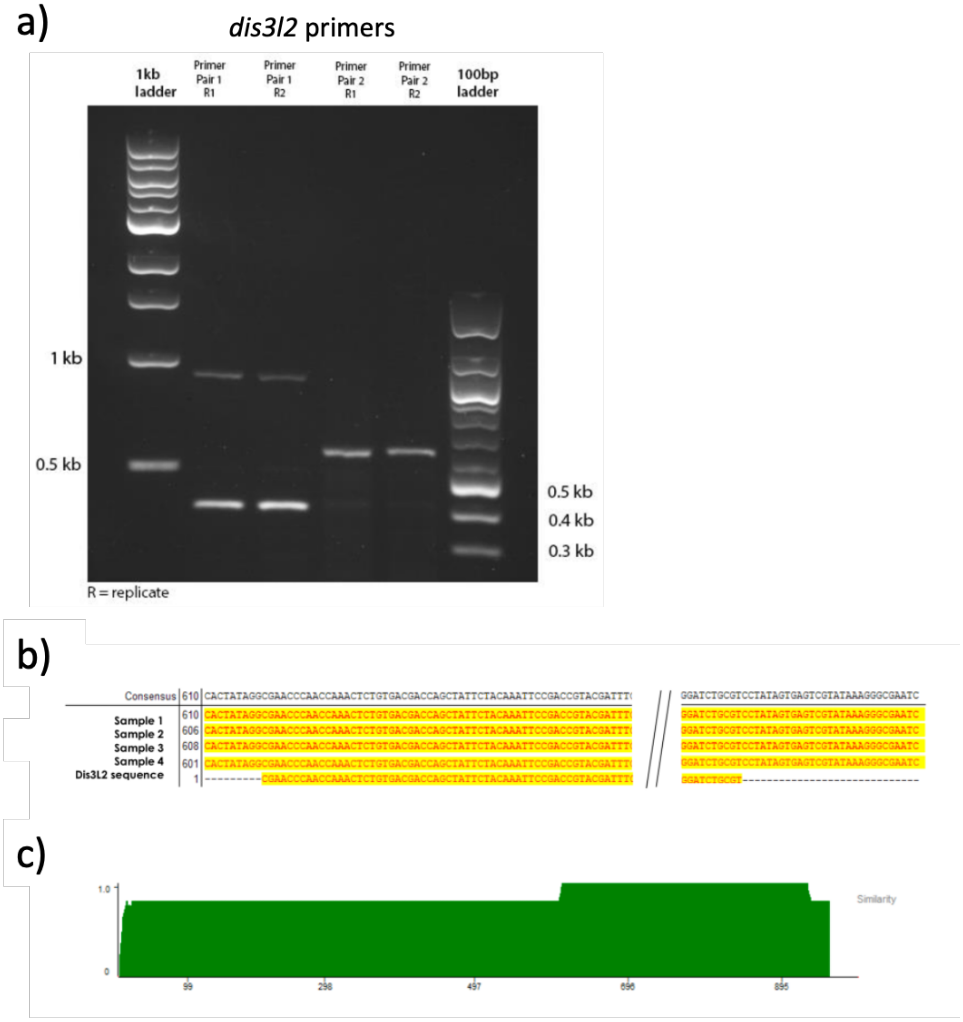

a)
dis3l2 primers
1kb ladder
Primer Pair 1 R1
Primer Pair 1 R2
Primer Pair 2 R1
Primer Pair 2 R2
100bp ladder
1 kb
0.5 kb
0.5 kb
0.4 kb
0.3 kb
R = replicate
b)
Consensus 610 CACTATAGGCGAACCCAACCAAACTCTGTGACGACCAGCTATTCTACAAATTCCGACCGTACGATTT GGATCTGCGTCCTATAGTGAGTCGTATAAAGGGCGAATC
Sample 1 610 CACTATAGGCGAACCCAACCAAACTCTGTGACGACCAGCTATTCTACAAATTCCGACCGTACGATTT GGATCTGCGTCCTATAGTGAGTCGTATAAAGGGCGAATC
Sample 2 606 CACTATAGGCGAACCCAACCAAACTCTGTGACGACCAGCTATTCTACAAATTCCGACCGTACGATTT GGATCTGCGTCCTATAGTGAGTCGTATAAAGGGCGAATC
Sample 3 608 CACTATAGGCGAACCCAACCAAACTCTGTGACGACCAGCTATTCTACAAATTCCGACCGTACGATTT GGATCTGCGTCCTATAGTGAGTCGTATAAAGGGCGAATC
Sample 4 601 CACTATAGGCGAACCCAACCAAACTCTGTGACGACCAGCTATTCTACAAATTCCGACCGTACGATTT GGATCTGCGTCCTATAGTGAGTCGTATAAAGGGCGAATC
Dis3L2 sequence 1 ---------CGAACCCAACCAAACTCTGTGACGACCAGCTATTCTACAAATTCCGACCGTACGATTT GGATCTGCGT-----------------------------
c)
1.0
0
99
298
497
696
895
Similarity

Figure 4.2 – Example of successful amplification of the relevant section for vector ligation:

The PCR product that was used in the final vector ligation reaction for *dis3l2* was primer pair 1. The PCR product obtained with these primers was the expected size (382bp long).

Initially, the experiment was carried out using primer pair 2, as these primers produced no non-specific products that could be seen by size separation (using gel electrophoresis). Sequencing of this product, however, revealed that the product was actually a fragment of histone RNA.

The primary product from primer pair 1 seemed to be of the right size, but a non-specific band was seen at approximately 900bp (unlike the PCR amplification for primer pair 2). Following this, a gel extraction of the correct sized band for primer pair 1 was carried out, used as template cDNA for another PCR (eliminating the non-specific product), and sequencing of this showed it to be the desired *dis3l2* fragment (panel (b)).

Sequencing results (panel (b)) of the 4 samples and the known *dis3L2* sequence are aligned, showing a perfect match for all samples. Panel (c) shows sequence similarity, in this case is 100%. The region of reduced similarity corresponds to sections of the plasmid itself.

the band of interest, which could then be used as a template for a second round of HF-PCR (summarised in a graphical flow chart in Figure 4.1, panel (b)), ensuring a pure and high-fidelity template to be used for the T7 reaction. The generation of large amounts of a pure and high-fidelity template for the region of interest, flanked by T7 promoter regions, allows easy generation of the desired region of RNA.

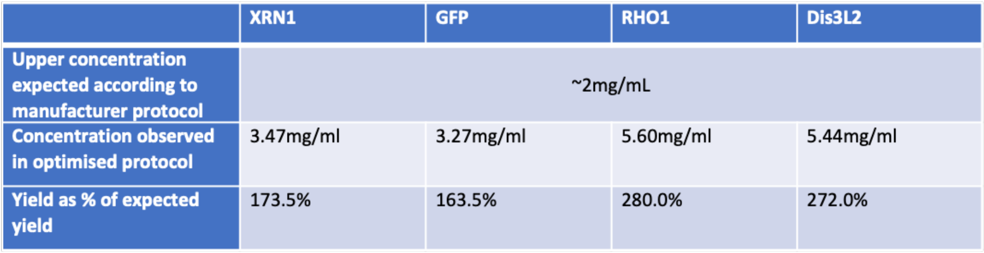
4.3.3 Optimised T7 express protocol generates increased yield of RNA

As previously mentioned in section 4.2 and 4.3, treatment of S2 cells with dsRNA has repeatedly been shown to be capable of inducing genetic knockdown; and work by Antic *et al*., appeared to establish it to be capable of knocking down *Drosophila pacman* specifically. As previously mentioned, Figure 4.1 panel (b) summarises the experimental plan and workflow required to do so.

As mentioned, poly-ribo-seq requires enormous amounts of RNA in order to generate a library with high enough concentration for sequencing. The protocol used by Aspden *et al.* uses as many as 1 billion cells per replicate. As a result, the ordinarily moderate quantity of dsRNA required to knockdown expression in a small population of S2 cells (as described by Antic *et al*.,) needed to be increased by orders of magnitude for treating this quantity of cells. In order to resolve this problem, the manufacturer's protocol for generation of RNA by the T7 express system was examined and altered, in an attempt to optimize the output of RNA.

According to the manufacturer's provided protocol, the T7 RiboMax system was intended to produce RNA in quantities of up to 2mg per ml of reaction (standard reaction volume of 20µl). Given that subsequent cleanup reaction reduces the yield, and each 350µl well in a 24-well plate of S2 cells to be treated requires up to 20µg of RNA, the protocol was adjusted to get as many reactions as possible with the highest yield.

Over multiple attempts, it was found that increasing incubation time for the T7 reaction (from 30 mins to 36 hours) and increasing time for the precipitation step during RNA extraction increased the final yield significantly (from ~2mg/mL to between 3.27-5.60mg/ml (Figure 4.3)). In addition, it was found that a significantly reduced T7 enzyme



	XRN1	GFP	RHO1	Dis3L2
Upper concentration expected according to manufacturer protocol	~2mg/mL			
Concentration observed in optimised protocol	3.47mg/ml	3.27mg/ml	5.60mg/ml	5.44mg/ml
Yield as % of expected yield	173.5%	163.5%	280.0%	272.0%

Figure 4.3: Table summarizing showing the output of the T7 ribomax reaction pre- and post-optimization

input concentration (80% of recommended volume) did not significantly decrease RNA output, allowing a greater number of reactions, and greater total RNA generated, per volume of T7 enzyme (the limiting reagent per kit bought).

4.3.4 Time and concentration optimised annealing reaction generates an efficient shift from single-stranded to double-stranded RNA

Exposure to high temperature, followed by gradual cooling, allows any secondary structure or partial annealing occurring within a sample of ssRNA to be disrupted, and the linearized strands to anneal into dsRNA. This can often be visualized by running paired samples of before and after annealing reaction on an agarose gel. dsRNA has an increased charge/mass ratio, and so runs slightly further in the same conditions. In addition, commonly used dyes like GelRed and Ethidium Bromide are brighter in double stranded nucleic acids, due to the structure of double-stranded molecules allowing full intercalation of the dyes. This increased stacking of fluorescent dye molecules generates a more intense band when imaged.

Across the attempts to efficiently generate dsRNA, the cooling rate, and concentration of RNA were both optimised, in order to determine whether a particular set of conditions might provide significantly increased efficiency in the shift from ssRNA to dsRNA (as inevitably, some molecules will not anneal into the desired double-stranded structure. This was tested by visual comparison of imaged gels. Both the concentration and cooling rate seem to be important to the efficiency of annealing (Figure 4.4), with more gradual cooling steps, and higher concentration lending themselves to more efficient annealing. Annealing in a salt buffer (described in methods) also provided a better annealing efficiency than annealing in pure water. Optimization of these allowed for improved efficiency of the annealing reaction (Figure 4.4 and 4.5).

4.3.5 dsRNA treatment achieved possible knockdown of Pacman, whilst effects on Dis3L2 expression are not known

According to the work of Antic *et al*., knockdown of Pacman can be carried out on *Drosophila* S2 cells using a protocol from earlier work by Barisic-Jager *et al*. (160). This

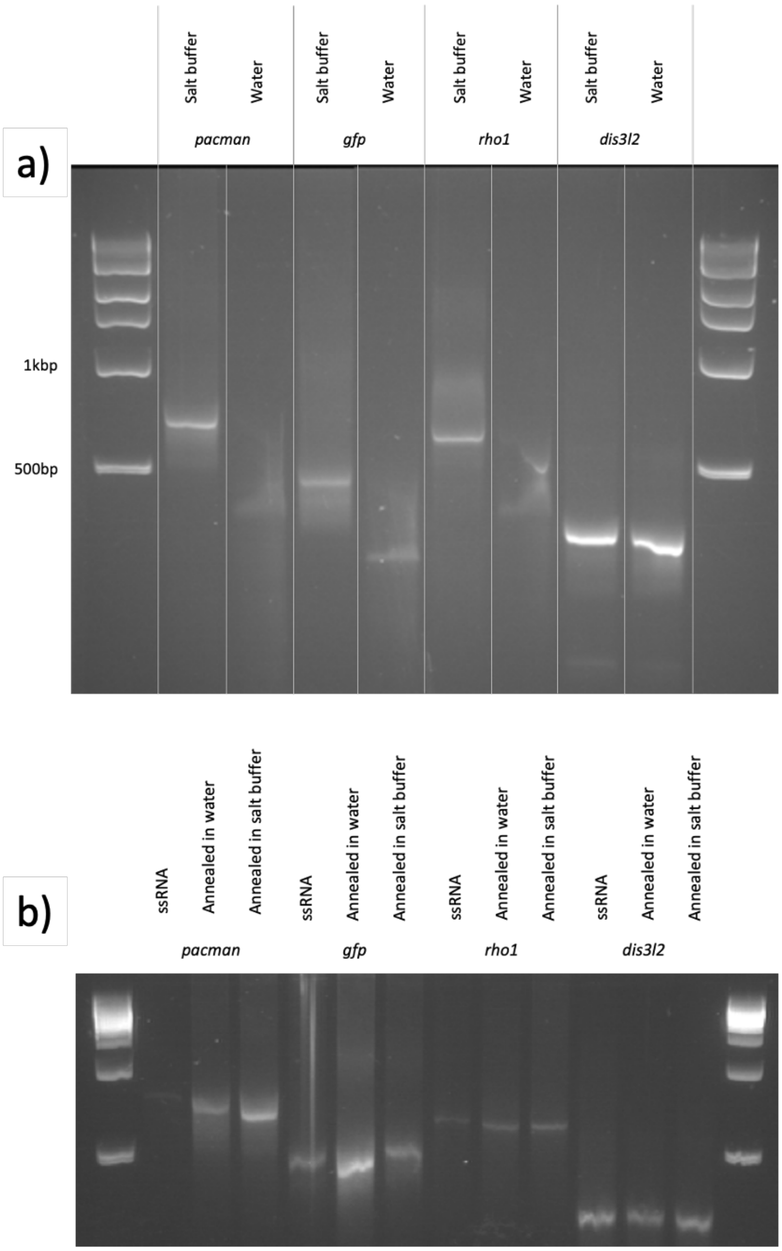
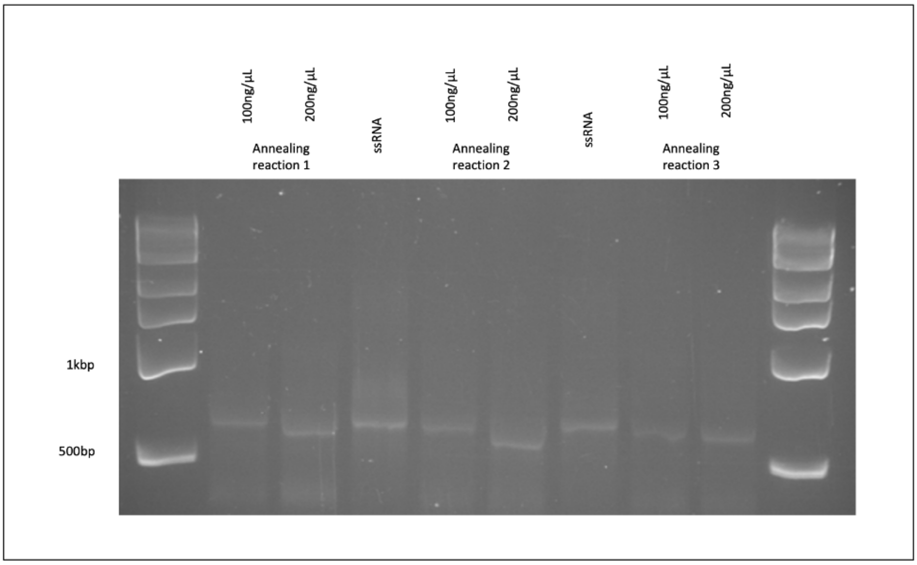


Figure 4.4 - Optimisation conditions for liquid conditions of dsRNA annealing reaction:

Panel (a) shows the annealing reaction carried out in water versus a salt buffer. The reactions carried out in water are prone to smear, as a wider range of conformations form during the re-annealing step. Even when this does not occur (panel (b)), a slightly stronger shift to double-strand conformation can be observed by the bend running further, and appearing brighter, for samples of the same volume and concentration. The composition of the salt buffer is specified in the Methods chapter.

a)



b)

Annealing reaction 1	Annealing reaction 2	Annealing reaction 3
90°C – 10 mins	90°C – 10 mins	90°C – 10 mins
	85°C – 1 min	80°C – 1 min
	80°C – 1 min	
	75°C – 1 min	70°C – 1 min
	70°C – 1 min	
	65°C – 1 min	60°C – 1 min
Leave at room temperature to cool	60°C – 1 min	
	55°C – 1 min	50°C – 1 min
	50°C – 1 min	
	45°C – 1 min	40°C – 1 min
	40°C – 1 min	
	37°C – 60 min	37°C – 60 min

Figure 4.5 - Optimisation of concentrations and temperatures for dsRNA annealing reaction:

Panel (a) shows a concentration dependent element to RNA annealing, with reactions at the higher concentration (200ng/µL) showing a greater shift to dsRNA than the same reaction at the lower concentration (100ng/µL). Annealing reaction 2, with the more gradual cooling step, appears to have a slightly increased annealing effect compared to annealing reactions 1 and 2. Annealing reaction temperatures are summarised in Panel (b).

protocol describes applying dsRNA in two shots at a concentration of 15µg/mL, although in this protocol it was targeting a range of different genes (*AGO1*, *NOT1*, *EDC4, DCP1* and *Pacman*). Treatment of S2 cells with dsRNA for these genes were carried out alongside S2 cells treated with dsRNA for *GFP*, and *rho1*, and untreated S2 cells. *GFP* is a gene not present in *Drosophila*, allowing control replicates that undergo dsRNA treatment, without risk of knockdown. *rho1* is a *Drosophila* gene with a known and visible phenotype when knocked down, allowing some visible indicator of whether the dsRNA treatment is having any effect. The protocol was also followed at double and triple the prescribed concentration of dsRNA, in order to maximise chance of a successful knockdown.

Once the knockdown protocol was followed, and the cells harvested, Western blotting was carried out to observe whether the knockdown was successful at the protein level. A known antibody for Pacman was available immediately, so the full protocol could be carried out and verified as soon as the dsRNA was synthesized. Western blotting showed that at a concentration of 45ug/mL dsRNA complimentary to *pacman* to be causing a (non-significant) 71% knockdown of Pacman at the protein level, while the *gfp* and *rho1* dsRNA treated control had the same expression of Pacman as the untreated cells (Figure 4.6). Although non-significant, Western blot quantification is unreliable, and the knockdown can be visibly observed. At 30ug/mL and 15ug/mL no knockdown was seen (Figure 4.7). Of note, the double-nucleus phenotype expected in *rho1* knockdown cells was not seen consistently, but did appear in some cells, at some points.

When the antibody for Dis3L2 became available, the same verification of knockdown experiment was due to be carried out, at the same range of concentrations. Alongside this, a scaled-up version of the Pacman knockdown was under way, with the intention of producing the number of Pacman depleted cells required to proceed with polysome fractionation, and subsequently RNA sequencing. This was prevented due to technical complications in the lab leading to death of S2 cell stocks, discussed in section 4.6. The phenotyping of cells treated with each dsRNA was also being undertaken, and was stopped short by these complications, and before knockdown could be confirmed.

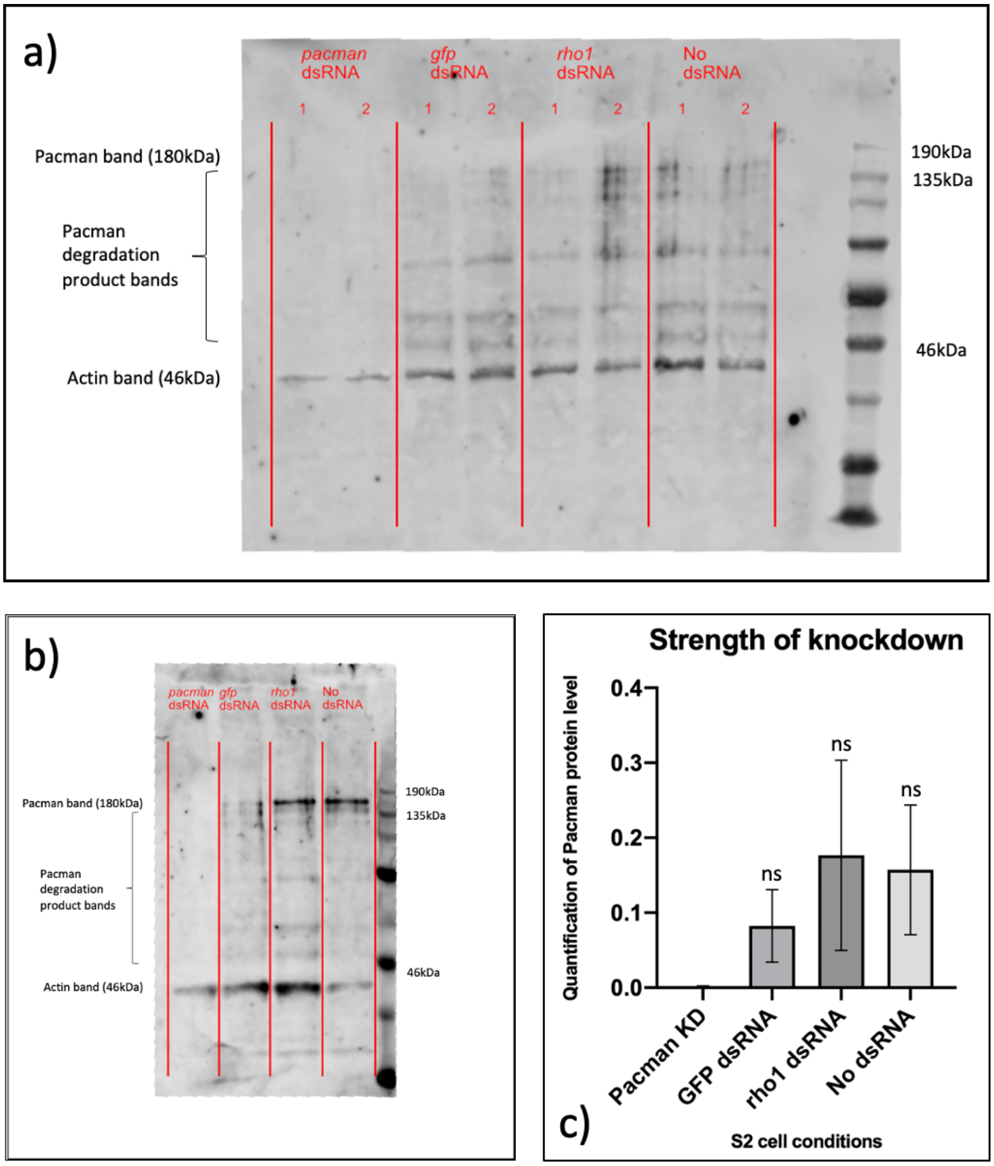


Figure 4.6 – Western blot shows successful Pacman knockdown:

Successful knockdown of Pacman at the protein level, by dsRNA, at a concentration of 45µg/mL. Pacman is unaffected by the dsRNA complimentary to *gfp*, *rho1*, or *dis3l2*. This is shown with a total of 3 replicates of each condition, across 2 different Western blots panel (a) and (b). Panel (c) shows quantification of the knockdown. Although the depletion is not statistically significant, this is likely due to the variability seen in the non-knockdown conditions. Error bars are SEM, and n = 3.

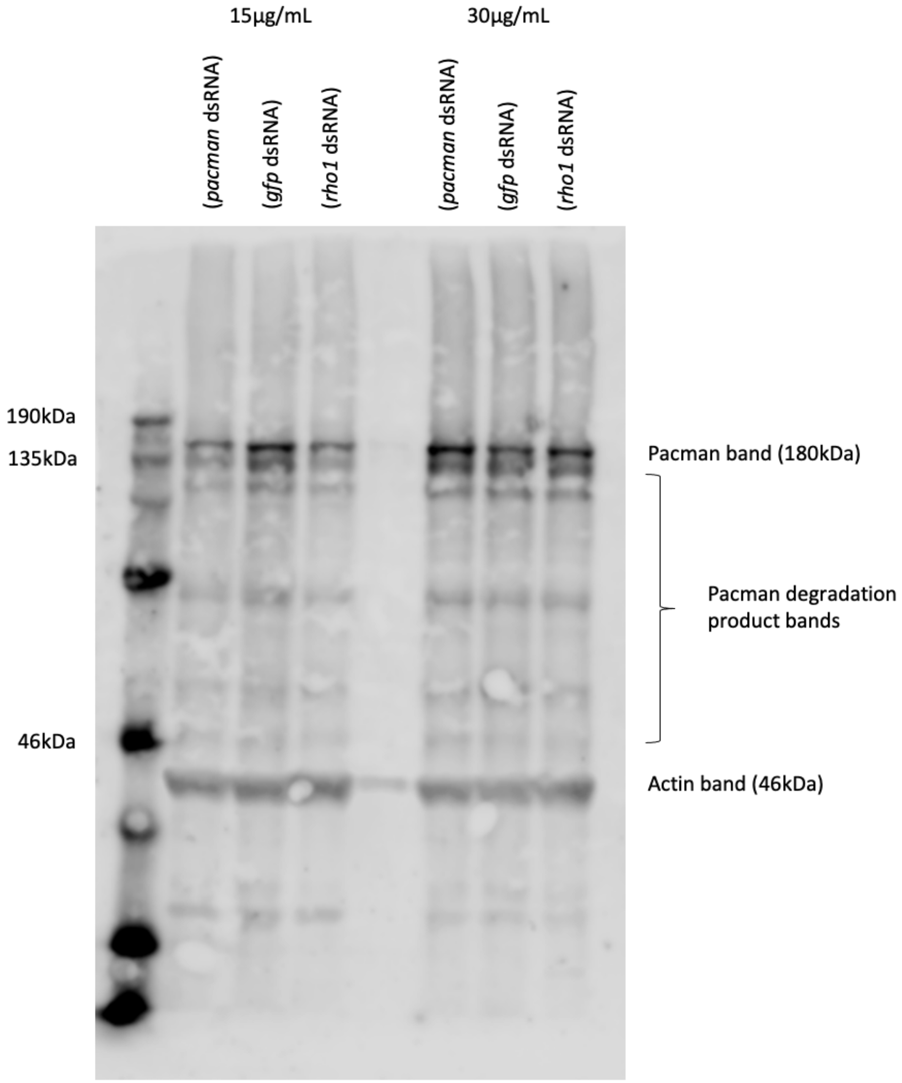


Figure 4.7 – Knockdown is not seen at dsRNA concentration suggested by Antic et al.:

Knockdown of Pacman at the protein level is not achieved by dsRNA at a concentration of 15µg/mL or 30µg/mL. Pacman is also unaffected by the dsRNA complimentary to *gfp or rho1*, as would be expected.

4.3.6 Possible phenotypic effects of exoribonuclease knockdown observed in S2 cells

Although the technical complications discussed in section 4.6 throw the validity of these results into question, as well as preventing further replicates in the same conditions, the possible phenotypic effects and trends that were observed will be discussed here.

As previously mentioned, GFP does not exist in the *Drosophila* genome, and therefore it is not expected to lead to an RNAi response due to the lack of known targets in the organism. Therefore, it helps distinguish between the sequence-specific gene silencing induced by the particular dsRNA used, and any non-specific effects. Further, the untreated control is also used to ascertain the baseline phenotype. The experiment was set up using two biological replicates for each of these conditions, and cell growth was assessed by way of daily cell counts. These were obtained using a haemocytometer with two separate counting grids. The cells were counted for two biological replicates in each of these grids, and the average of the four counts was recorded, as described in methods.

As seen in Figure 4.8, cell counts for the *pacman* dsRNA treated cells significantly decreased on day four, when the medium was removed and replaced with serum-free medium, likely as a result of the additional dose of dsRNA that was added on this particular day. The procedure requires the removal of the medium, followed by the addition of a serum-free medium and a second dose of dsRNA (as described in methods). The semi-adherent S2 cells grew floating within the medium, making it very difficult to remove without disturbing them. Consequently, some S2 cells were unavoidably removed during the medium exchange.

However, after this point the growth rates did not recover, in contrast to the situation seen for the GFP and untreated control conditions. This could be suggestive of a phenotype induced by the knockdown of Pacman. Previous studies have reported that knockdown is achieved in *Drosophila* S2 cells after three days of dsRNA treatment. Thus, by day four it might be expected for the cells to display a phenotype due to a successful knockdown of Pacman. Therefore, the rates of cell growth for Pacman might have

a)

	XRN1		*Dis3L2*		*GFP*		*Untreated control*	
Day	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2
1	470000	350000	600000	545000	565000	385000	565000	550000
2	360000	570000	380000	490000	670000	660000	480000	560000
3	830000	600000	560000	430000	810000	710000	870000	710000
4	400000	600000	810000	820000	600000	570000	350000	620000
5	530000	470000	660000	1070000	810000	430000	410000	540000
6	420000	550000	880000	940000	700000	780000	580000	630000

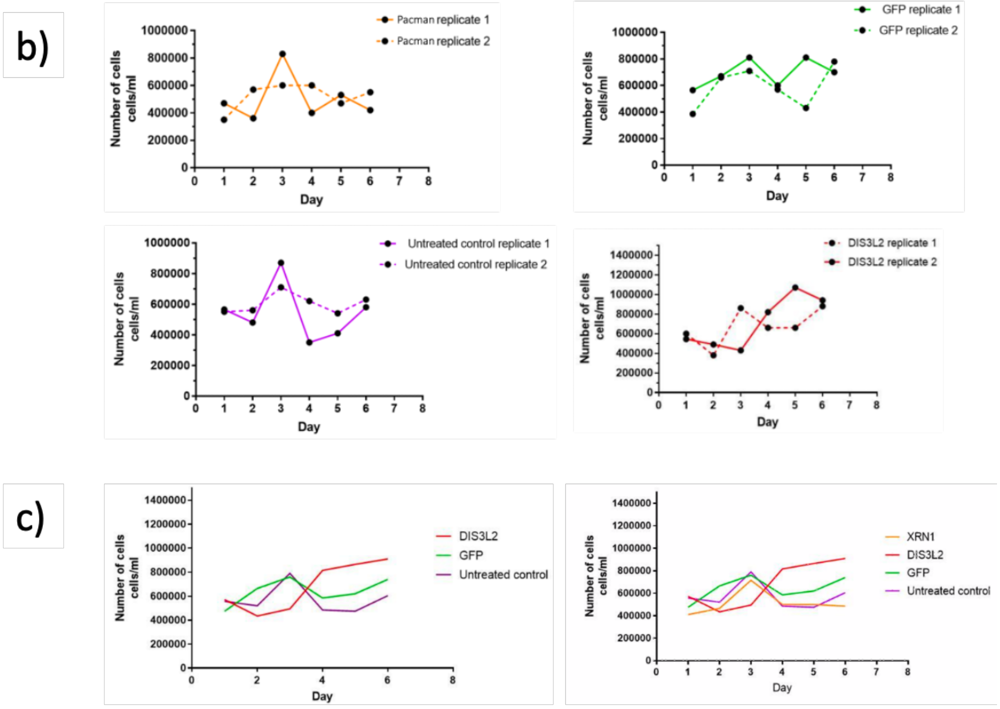


Figure 4.8 – Potential growth phenotypes in dsRNA treated cells:

(a) Daily cell count (cells/mL) in the days following dsRNA treatment.

(b) The cell count replicates plotted against time, showing the significant variability between replicates, making meaningful statistical analysis impossible. This is possibly accounted for by the technical problems later discovered and discussed in section 4.6.

(c) The treatment of the S2 cells by the *dis3l2* dsRNA does however seem to increase growth rate compared to *pacman* treated, untreated and GFP controls. This cannot be taken to be significant without repeating these experiments in the absence of technical issues.

remained low as a result of a successful knockdown. However, while the knockdown for these particular cells itself is not confirmed by means of qRT-PCR or Western Blot, the lack of proliferative recovery could also be a result of the stress generated by the medium-exchange procedure, or arguably any other external factors, including technical complications.

Assuming that the dsRNA treatment in these conditions was successful, the knockdown of Pacman appears to have decreased the proliferation rates of S2 *Drosophila* cells. However, this reduction did not appear to be a result of increased apoptosis rates. This was assessed via cell count with Trypan Blue staining. Surprisingly, according to Trypan Blue staining, no dead cells were observed on any day of the experiment, which may be an unusual consequence of the technical issues discussed in section 4.6 loosening the dead cells enough that all dead cells were lost with changes of media, as at least some cell death would be occurring over this timeframe, and trypan blue should allow at least some of it to be visualised.

As seen in Figure 4.8, the expected drop in number of cells/ml on day four, in the *dis3l2* dsRNA treated cells, was less drastic (in the mean) compared to the other conditions. It seems likely that this difference is a result of varying numbers of cells being removed during the medium exchange between conditions. This seems especially likely when the variability of replicates is considered (Figure 4.8, panel (b)). An alternative (and optimistic) interpretation of the cell counts could be that the *dis3L2* dsRNA treated cells had higher proliferation rates (as seen in previous work by Towler *et al*. (46)) due to a successful Dis3L2 depletion, allowing them to double by day four. Therefore, even if some were lost during the medium removal, the loss was not as drastic as that seen for the controls. Once again, no dead cells were observed on any day of the experiment. The time course of the experiment could not proceed beyond day 6, due to sudden and total death of S2 cells, caused by technical issues discussed in section 4.6. Due to the time limitations and technical difficulties, it was decided that attempted use of CRISPR-Cas9 would be a promising technique to try and produce exoribonuclease depleted cells.

4.4 Producing S2 cell XRN1/pacman and Dis3L2 mutants using CRISPR:

CRISPR-Cas9 is a powerful genetic tool with rapidly expanding uses throughout several biological fields (Figure 4.9). By introducing Cas9 (an RNA-guided DNA endonuclease), and a guide RNA construct complimentary to a region within the gene of interest (in this case *pacman* and *dis3l2*), a complex is formed between Cas9, the guide RNA, and the target DNA. The DNA is cut by Cas9, resulting in a double-strand break (DSB). Should this be repaired by the error-prone non-homologous end joining pathway, an indel may be caused, resulting in a non-functional gene.

4.4.1 Successful cloning of the sgRNA into a suitable plasmid

The pAc-sgRNA-Cas9 plasmid (Figure 4.10) used for this experiment was grown up as per manufacturer instructions, and the sgRNA regions for *pacman* and *dis3l2* were cloned in, as described in methods. The sgRNA regions were designed to be complimentary to necessary 5' regions in the CDS of the *pacman* (a 300bp region) and *dis3L2* (a 42bp region) genes, with overhangs for the BspQ1 cut site, in order to allow specific cutting of the plasmid. The *dis3l2* sgRNA was the same as used to make CRISPR-Cas9 null mutants in flies, in order to make the most direct comparisons with existing mutants and data from them. Subsequent DNA sequencing identified the region of interest within the re-ligated pAc-sgRNA-Cas9 plasmid for *pacman* and *dis3L2*. The insertions were identified in the correct location across multiple tested colonies, which successfully grew through selection on LB-Agar plates containing Ampicillin (which the plasmid contains a resistance gene to), and further selection of these colonies being transferred into LB-Broth containing Ampicillin.

4.4.2 Transfection of a test construct expressing GFP into S2 cells

By carrying out a transfection into S2 cells with a two-part GFP expressing construct (a construct with a gene driver, and a corresponding construct with GFP that can be driven by the first construct, using the UAS-GAL4 system to introduce Actin-GAL4 and UAS-GFP, allowing their combination to produce GFP expression in the Actin expression pattern), a lower end (as the CRISPR-Cas9 transfections only require one construct to be

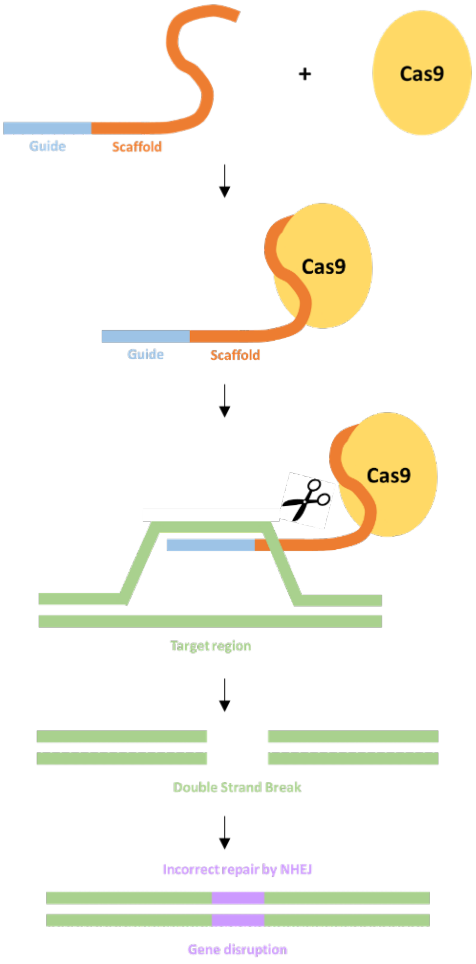


Figure 4.9 - A cartoon depicting the basic principles of CRISPR-induced mutation:

A complex is formed between Cas9, the guide RNA, and the target DNA. Cas9 causes a double-strand break, allowing an error-prone repair pathway to potentially cause an indel that disrupts gene function.

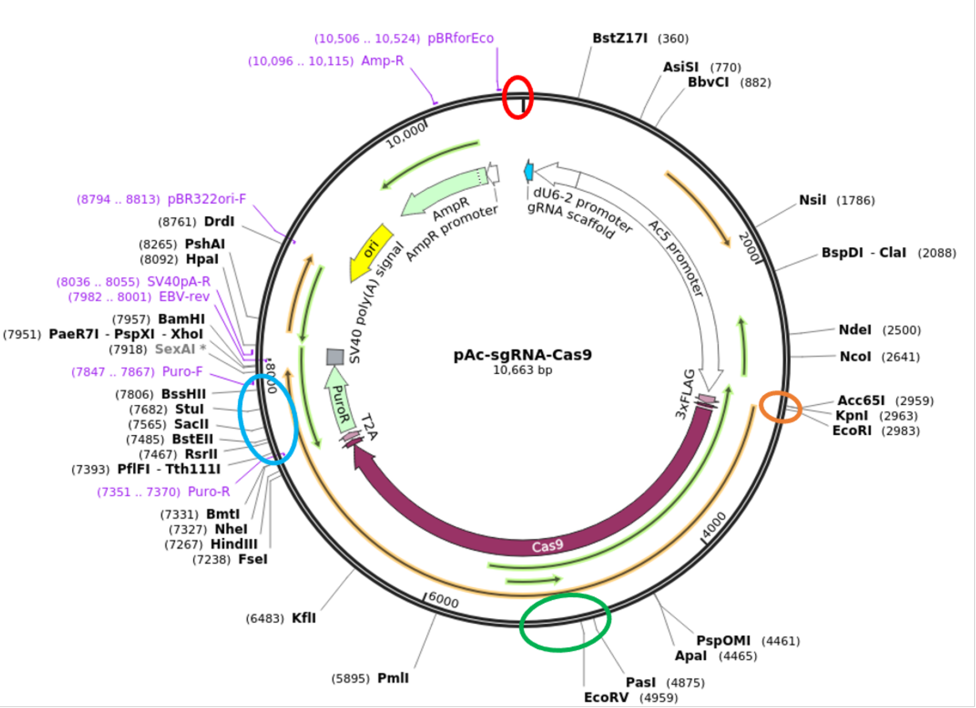


Figure 4.10 - Plasmid map for the pAc-sgRNA-Cas9 plasmid used for the CRISPR experiment:

Crucial components include Cas9 (allowing the cut necessary for the CRISPR process to take place, region tagged in green), the BspQI digest site (not labelled, site of guide insertion, immediately downstream of U6 promoter and guide RNA scaffold, region tagged in red), EcoRI (used for double digest testing, as described in methods, region tagged in orange), and the pAc (puromycin resistance) gene (region tagged in blue).

transfected) estimate for the efficiency of transfection was estimated at 0.41% by counting total cells vs. GFP expressing cells with fluorescent microscopy (Figure 4.11).

By starving the S2 cells in serum-free media for 3 hours prior to transfection (whereupon the media was supplemented with FBS to bring it to full concentration of serum), the efficiency of the transfection was vastly improved, increasing to as much as 4.05% (Figure 4.12).

4.4.3 Attempted transfection of pAc-sgRNA-cas9 construct into S2 cells

Following confirmation that the transfection protocol had a reasonable efficiency, the same procedure was carried out on S2 cells in order to transfect the pAc-(*pacman*)sgRNA-cas9 and pAc-(*dis3l2*)sgRNA-cas9 plasmids into S2 cells. Puromycin was used in order to select for cells containing the plasmid, which provides resistance to Puromycin (normally lethal to S2 cells), via the Puromycin N-acetyltransferase activity conferred by the pAc gene.

Cells were incubated in Puromycin containing media for 4 days, at a concentration of 5µg/mL (shown to be sufficient for complete lethality in this timeframe (161)), to allow for the drug to kill off as many non-transfected S2 cells as possible, within the timeframe of a transient transfection. After this, the cells were centrifuged to pellet them, washed in fresh media, and plated in fresh media, in order to prevent residual Puromycin from harming the cells after the plasmid was ejected from the cells.

In order to generate a population of a homogenous mutant genotype, all treated cells were split to sub-single cell concentrations and plated en masse in 96-well plates. These were plated with 50µL of conditioned media mixed with 50µL of fresh media and left to proliferate. Many wells inevitably did not contain even a single cell, or contained a cell that died, or failed to proliferate. Of those that did begin to proliferate, they were left in the initial media until they were seen to be semi-adherent when examined under the microscope. At this point, the media was replaced with new conditioned media, which was replaced once a week from this point onwards.

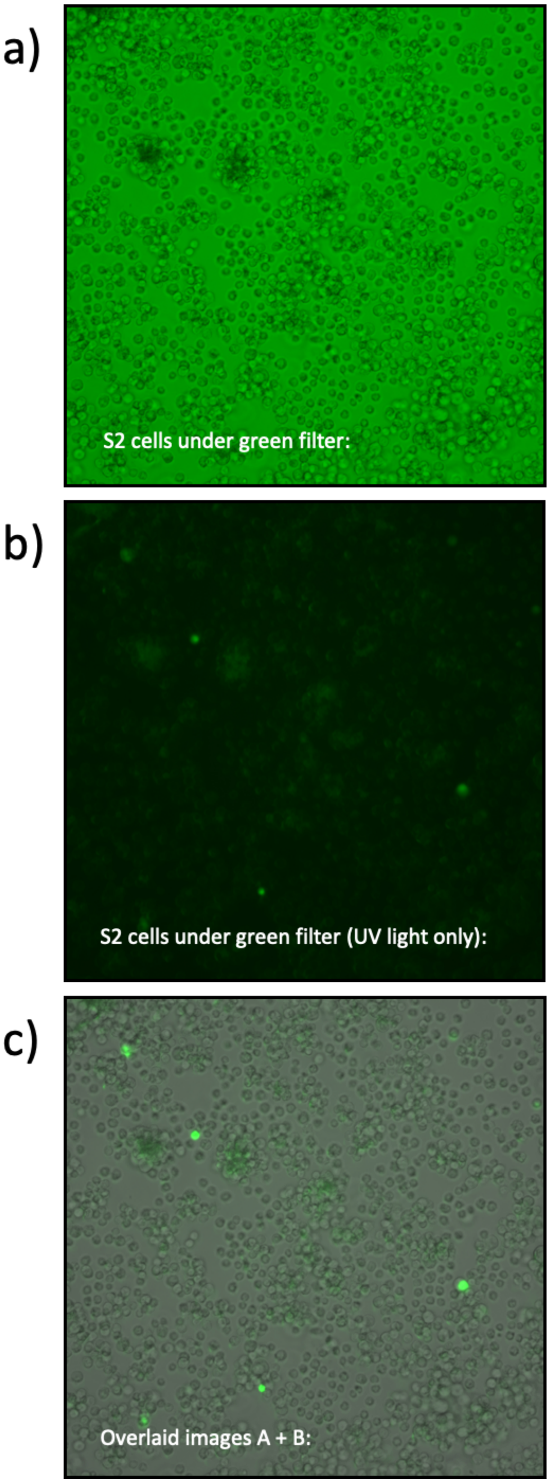

a)
S2 cells under green filter:
b)
S2 cells under green filter (UV light only):
c)
Overlaid images A + B:

Figure 4.11 – GFP microscopy images showing transfected cells expressing GFP

Efficiency of transfection in S2 cells in these conditions was calculated as 0.42%. Panel (a) shows S2 cells all under a green filter with background white light. Panel (b) shows S2 cells all under a green filter with UV light only. Panel (c) shows the overlaid S2 cells with background light and UV light only, to highlight transfected cells versus all cells.

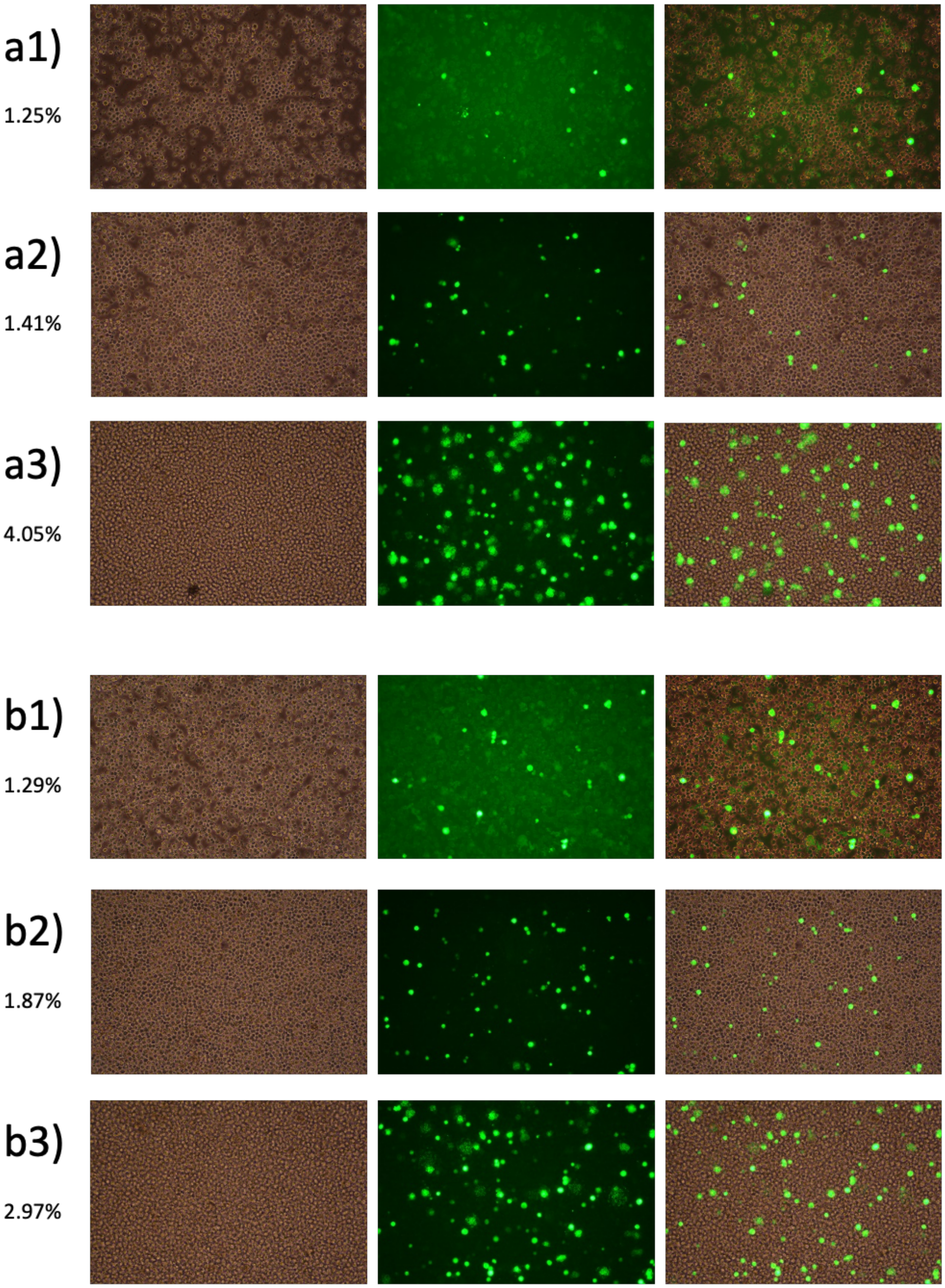


Figure 4.12 – GFP microscopy images showing improved efficiency of transfection

From prior efficiency of 0.42%, optimization by starvation of cells and observation over time was able to increase efficiency substantially. Rows (a) 1-3 show transfection of Pacman + Actin + GFP at 24, 48, and 72 hours respectively; rows (b) 1-3 show transfection of Pacman + Actin + GFP at 24, 48, and 72 hours respectively. Efficiency of transfection is listed below each row number.

Many cell lines (including S2 cells) struggle to proliferate and thrive beyond a certain minimum concentration of cells. As a result of this, the process of growing up cell populations from single cells is time-consuming, and prone to failure. Many cell populations died before amplifying to a level suitable to use. Further to this, the primary recovering stocks were killed due to high temperatures and dehydration due to technical issues, as discussed in section 4.6.

4.5 Overview of results:

4.5.1 Optimising protocols to produce large quantities of dsRNA:

Following the dsRNA knockdown of Pacman achieved in S2 cells by Antic *et al*., this project achieved significant optimization in the production of dsRNA using the T7 RiboMax system, and knockdown of *pacman* was achieved (although it required significant alteration to the existing protocol by Antic *et al*.). Both the efficiency of RNA synthesis and the formation of dsRNA were optimized, in order to facilitate continuous testing of the attempted knockdown, with the ultimate aim of carrying out the knockdown of both exoribonucleases for multiple replicates, at a large scale.

4.5.2 Attempted dsRNA knockdown of Pacman and Dis3L2 in S2 cells:

Western blotting showed successful knockdown of *XRN1/pacman* in both 24-well plates, and in full 100mm tissue culture dishes, but the required dose of dsRNA was higher than expected, requiring hundreds of micrograms of dsRNA to induce a knockdown in 100mm dishes. Given that between 6 and 10 plates may be required per condition to carry out poly-ribo-seq, this raises significant problem with this technique, especially as the success of the knockdown can be impacted by the health and growth rate of the cells, whether the annealing has been completely efficient (which can be observed by gel electrophoresis of annealed versus non-annealed RNA, but is difficult to accurately measure).

Throughout all conditions and concentrations trialed, dsRNA treatment with Dis3L2 targeting dsRNA was not produced in viable enough cells to test to show knockdown. A significant number of the optimization assay attempts for the Dis3L2 dsRNA treatment were carried out S2 cells that later proved to be unhealthy (see section 4.6). Some further testing was later carried out on viable S2 cells, and the knockdown was still unsuccessful in these attempts. The dsRNA Dis3L2 knockdown attempt was less thoroughly tested on healthy cells than the Pacman as a result of these circumstances; however, combined with the aforementioned difficulties in the Pacman knockdown procedure and the stage in the project at which these issues arose, it was deemed more promising to switch to an alternative method.

4.5.3 Phenotyping potential Pacman and Dis3L2 knockdown S2 cells:

Cell growth assays, with trypan blue staining, were carried out on the cells over the time-course of treatment with dsRNA. A possible increase in cell proliferation was seen in the S2 cells treated with dsRNA complimentary to *dis3L2*, although this could not be statistically verified across only 2 replicates, and the death of the cells due to technical issues (section 4.6) makes it hard to draw any meaningful conclusion from this phenotyping. No cell death was seen by Trypan blue staining, although small specks of stained debris were seen, later thought to be dead cells that had been rapidly dehydrated due to the technical issues.

4.5.4 Attempted CRISPR-Cas9 editing of *Drosophila* S2 cells:

Although it is useful to know that dsRNA can be used for this purpose, and undoubtedly with further refinement the technique might be made more reliable, ultimately it was decided that with another technique (better suited to large scale cell culture) might be better for this project.

Despite being a simple, efficient, and powerful tool for genetic manipulation, there are still difficulties in utilisation of the CRISPR-Cas9 system. Within this project, the necessity to isolate homogenous populations of a single mutant added the challenge of growing up a stable stock of S2 cells from single cells. Although possible in S2 cells, this is a challenge, requiring significant time and resources. With the added issue of having to repeat this experiment due to aforementioned technical issues (section 4.6), by the time individual populations had been isolated and were ready to test, an alternative model to examine polysome-associated lncRNAs regulated by Pacman and Dis3L2 had already been optimised, and was ready to use. Perhaps it should be noted that if similar work along these lines were to be repeated, a better match might be *Drosophila* clone 8 cells or DMD21 cells (both derived from 3rd instar larvae wing imaginal discs), in order to more closely match existing in-vivo experiments.

Despite this, some positive outcomes can be taken from this experiment. The sgRNA necessary to generate null mutants was designed and successfully cloned into constructs that can easily be transfected into S2 cells. The efficiency of S2 cell transfection was substantially improved throughout this work, improving chances of a successful transfection of the CRISPR construct in any future attempts. Several cell pellets were harvested post-transfection, but pre-dilution to sub-single cell concentrations. Overall, the work carried out may be useful in future work to generate and use a stable stock of exoribonuclease null mutant S2 cell lines but was put aside in favour of whole tissue work in *Drosophila* L3 larvae.

4.6 Technical difficulties with the culturing of *Drosophila* S2 cells:

As mentioned throughout this chapter, the work in this project using *Drosophila* S2 cells was plagued by technical difficulties not scientifically relevant to the focus of the project. Although relatively straightforward to grow, S2 cells do require a stable temperature of anywhere between 23-27°C, and sufficient humidity to keep their media at a high enough liquid volume to adequately cover the cultures. Growth rates for the cells in this project were found to be wildly inconsistent, and survivability varied between 2 and 20 passages, with some cultures dying as soon as they were re-suspended from stock pellets. At various points (mentioned through the chapter), recently passaged or treated cells were found to rapidly dehydrate overnight, leaving dead or highly stressed cells in crystallised media, with the cells either partially or completely dry.

Temperature was monitored, and technical assistance from faculty managers was requested repeatedly, but it was only after significant investment in the S2 cell work that the incubator was identified as being incapable of supporting a humidity suitable for S2 cell culture. Instead, rapid airflow in the incubator used had been sporadically dehydrating cell stocks, causing sudden death of cells. Additionally, while during certain experiments cells survived these stresses, the variability of these factors cannot be known, and how this stress affected the cells and any results gained from them cannot be meaningfully analysed.

It should be noted that one replicate of dsRNA treatment of S2 cells with dsRNA, and subsequent cell counts and phenotyping, were carried out by an undergraduate student (Alexa Tataru) under the candidate’s supervision.

Chapter 5: Results – Optimisation of Poly-Ribo-Seq protocol for whole *Drosophila* L3 larvae

5.1 Introduction

As previously discussed, post-transcriptional control of gene expression is one of many regulatory layers acting within the central dogma of molecular biology. RNA stability and degradation, as part of this, play crucial roles in determining the abundance and temporal availability of RNA species in any given environment. In turn, this affects how readily the RNA can exert its function, whether it be as a protein-coding mRNA, the silencing function of a miRNA, the crucial structural roles of rRNA, amongst many other roles. The decay of RNA is dependent upon degradation by enzymes from either the 5' end (such as degradation by Pacman, in *Drosophila*), or from the 3' end (by the Dis3 family). The Newbury lab has produced extensive work in the area of RNA decay and degradation and has made great use of RNA-sequencing experiments in order to produce large-scale datasets that allow examination of RNA abundance at a genome wide level.

Whilst RNA sequencing has been a powerful tool for this area of RNA biology (among others), the information it can provide is not sufficient to adequately explore the interplay between degradation and translation of RNA. Techniques building on standard RNA sequencing have emerged, improved, and gained popularity over the last decade. Ribosome profiling (also known as Ribo-Seq or ribosome footprinting,) is a technique that built on the work of Marilyn Kozak and Joan Steitz (84). By carrying out digestion of unprotected RNA, genome-wide sequencing can target only RNA that is actively protected by the ribosome, during translation. Isolation of the polysome and associated molecules by fractionation has been of great use in the field of molecular genetics, helping us understand the control of translation and factors associated with the polysome.

Kozak and Steitz had established that treatment with endoribonucleases is capable of digesting translating RNA, converting it to short fragments protected by individual ribosomes. This discovery was at the time successfully used in order to identify and sequence RNA initiation sites, providing a wealth of new knowledge to those within the field, and shedding light on the workings of internal ribosome entry sites (IRESs). The dawn of modern high-throughput genome wide sequencing technologies opened up the possibility of combining these techniques in order to capture a global snapshot of ribosome bound RNA. This technique has repeatedly demonstrated its usefulness, and in recent years has added extra levels at which it can be used, having been used to detect numerous small translating ORFs, and providing a look at non-canonical ORFs.

Whilst the power of ribosome profiling as a technique is not to be understated, it does have its limits; notably a relatively high false positive rate, due to putative binding of RNAs to ribosomes without functional translation taking place. Building on the strengths of the technique, Ingolia et al. and Aspden *et al*. (153, 162), developed a variant of this technique, poly-ribo-seq, that uses the isolation of transcripts containing multiple ribosomes (called polysomes or poly-ribosomes) followed by endoribonuclease digest and sequencing. This leaves pools of RNA that were separated by density to select only those transcripts bound by multiple ribosomes simultaneously (with a higher certainty of active translation compared to association with a single ribosome). The subsequent digestion by nucleases then leaves only the fragments of RNA that were bound and shielded by the ribosome, in order to identify and sequence actively translating RNA with a much lower rate of false positives.

Initially applied to *Drosophila* S2 cells, the technique proved useful, allowing Aspden *et al*. to identify two separate types of short ORF (153). The technique has since been applied to other models and tissues, including *Drosophila* embryos. Although this has provided a wealth of data on translation of *Drosophila* RNA, an opportunity exists to use this technique in combination with established mutant lines lacking in crucial enzymes of the RNA degradation pathways, in order to better understand the interplay between the translation and degradation. With the previous chapter discussing attempts to use different *Drosophila* based models, and concluding with the decision to use wandering L3 larvae, a new variant of the protocol developed by Aspden *et al.* (153) was necessary. This chapter will discuss the process required to optimise use of whole *Drosophila* L3 larvae for poly-ribo-seq.

5.2 Project background and aims

The Newbury lab has previously explored the impact of Pacman and Dis3L2 depletion upon L3 larval tissues, and the Couso and Aspden labs have explored non-canonical translation in *Drosophila* cells and tissues (90, 153). The overlap between these two areas (exploring non-canonical translation, and the role of translation in facilitating degradation) is promising but requires significant work to open it up for exploration. The work in this chapter aimed to create a protocol capable of generating high-quality samples for poly-ribo-seq from whole *Drosophila* L3 larvae.

The main stages to this work are as follows:

1) Establish a baseline by carrying out polysome fractionation on a *Drosophila* model that has a working protocol already, namely S2 cells.
2) Carry out this protocol on whole *Drosophila* L3 larval samples, and evaluate viability compared to the previously fractionated samples from S2 cells.
3) Optimise steps throughout the protocol for higher quality fractionation and polysome trace.
4) Use this optimised protocol to provide samples to go through the remaining stages of poly-ribo-seq.

5.3 Whole *Drosophila* L3 larvae as a model organism for poly-ribo-seq

Extensive previous work in the Newbury lab has been carried out on *Drosophila*, and of particular interest, has RNA-seq datasets from *pacman* and *dis3L2* null mutant L3 wing imaginal discs (WIDs). Although poly-ribo-seq in S2 cells would still produce interesting and valuable data, whole biological model organisms provide a more representative look at RNA behaviour in living organisms. Use of *Drosophila* WIDs would provide a valuable dataset, but the time taken for dissection, combined with the small amount of RNA that can be extracted from wing discs made this option non-viable. By using whole L3, not only will the data produced have these advantages, but any interesting lncRNAs identified by analysis of the existing Newbury lab data will be known to be expressed in the L3 (by virtue of their presence in at least one tissue, the WIDs,) at that developmental timepoint.

At the time of the experiment, this protocol had not yet been carried out in wandering L3 larvae, although some polysome fractionation work had been published in earlier stages of *Drosophila* development (embryonic, L2, and early L3), as well as cell lines. Although testing and optimising the protocol for this model organism would be challenging, the advantages of the data that could result from the experiment, (as well as the help available from the protocols in similar work on other *Drosophila* developmental stages and cells,) led to the decision to pursue the experiment in whole L3 larvae.

5.4.1 Direct comparison between S2 cell and whole L3 polysome traces shows the necessity of an alternative protocol

Multiple previous works have provided protocols for, and results from, polysome fractionation (in some instances followed by poly-ribo-seq,) in *Drosophila* S2 cells. Examples of these (Figure 5.1) can largely be seen to be of a high resolution, with a significant and active polysome, indicating significant translation.

Following the protocol provided by the Couso lab (summarised in Figure 5.2 and 5.3), a test run of polysome fractionation of *Drosophila* S2 cells was carried out, in order to test samples, reagents, and technical equipment were sufficient to produce similar results. Following a successful demonstration of the existing protocol in S2 cells (Figure 5.4, panel (a)), along with examination of previous work in both S2 cells and *Drosophila* embryos (Figure 5.1), the protocol from polysome fractionation of *Drosophila* embryos was trialled and applied to whole *Drosophila* L3 larvae (Figure 5.4, panel (b)).

Although this protocol was capable of separating RNA transcripts by RNA density (Figure 5.4, panel (b)), to some degree, the resolution was poor compared to previous results, and the polysome profile has collapsed, likely due to ribosome dissociation or RNA degradation. Given the aims of the project to explore lncRNAs that are associated with the polysome, and

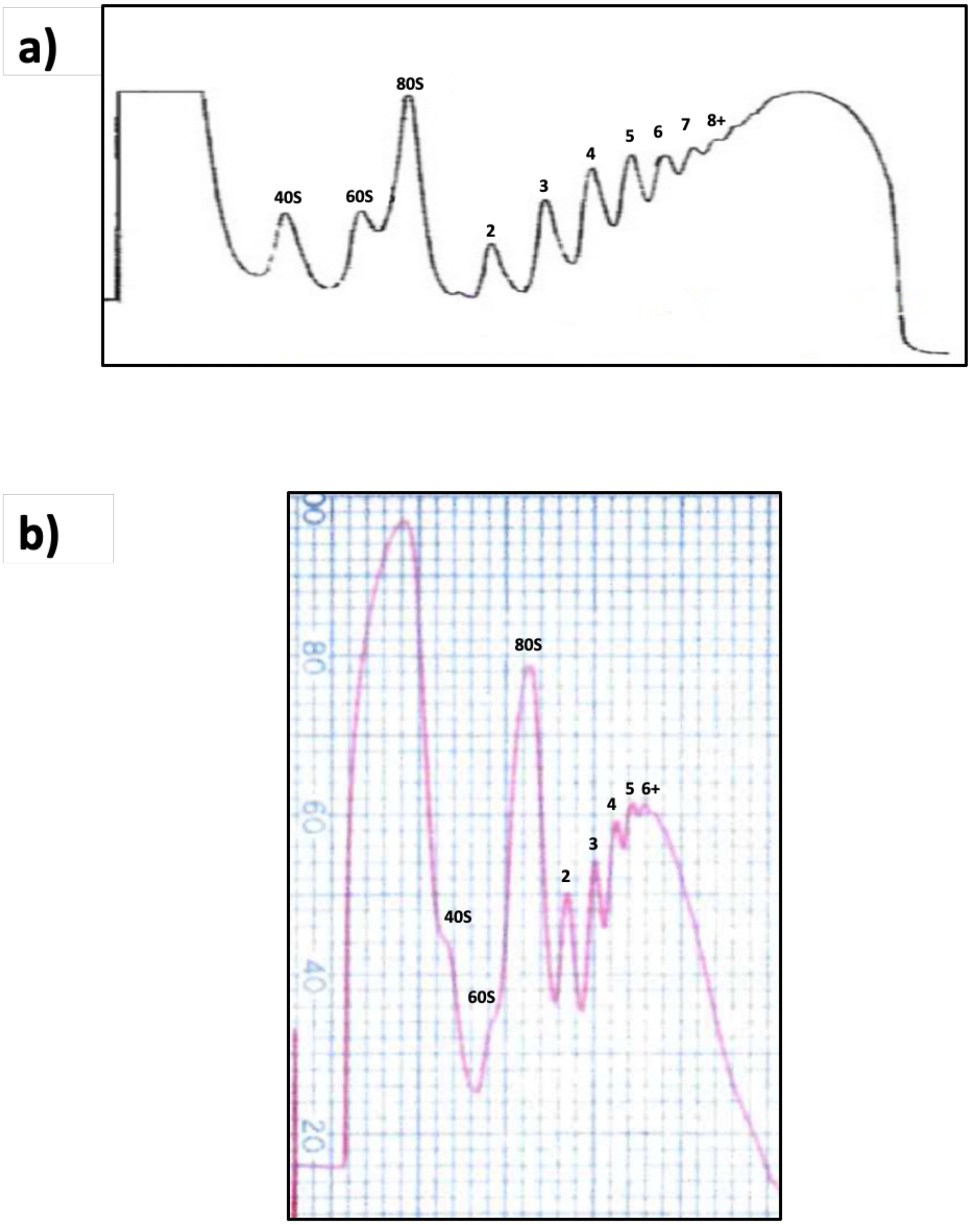


Figure 5.1 – Example of polysome profiles in S2 cells produced by and adapted from the work of Aspden et al. in her 2014 paper (panel a), and *Drosophila* embryo by A. Mumtaz in his thesis (panel b) .

Previous work has successfully produced high resolution polysome profiles from both S2 cell and *Drosophila* embryo samples.

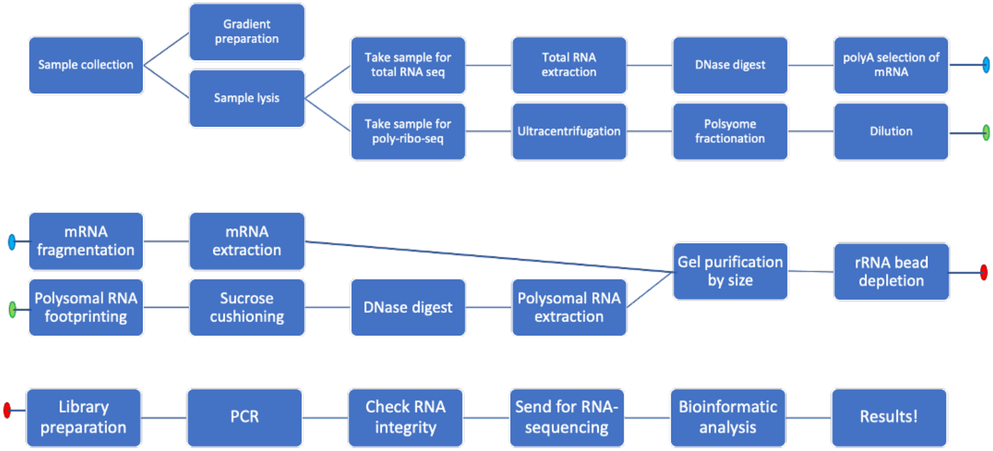


Figure 5.2 – A flow chart summarising the main steps involved in poly-ribo-seq:

The chart includes parallel workflows for total lysate RNA and polysomal RNA. Individual workflows can be followed between lines according to the colour at the start and end of the line.

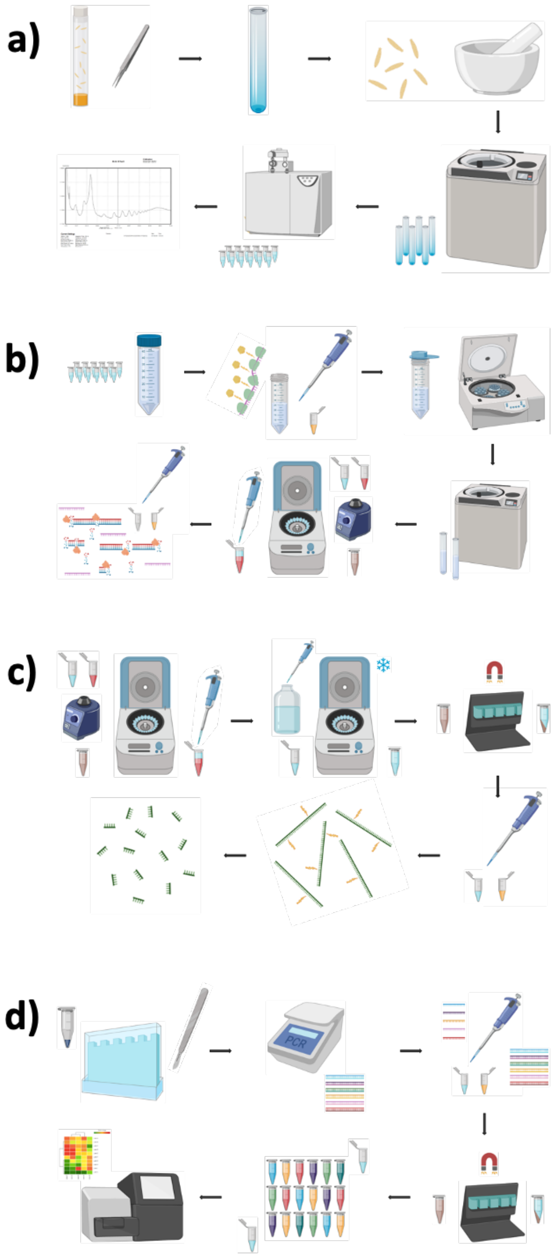


Figure 5.3 – Visual summary of the steps of poly-ribo-seq

(a) Polysome fractionation: Lysis at extremely low temperature, directly into cycloheximide and Rnase inhibitor containing lysis buffer stabilises the polysome as much as possible, reducing polysomal collapse and RNA degradation. Different methods of lysis are available, and can be chosen depending on sample type. A sucrose gradient allows high-speed centrifugation to separate by density, so RNA transcripts can be fractionated by the number of associated ribosomes. Absorbance at 260nm is measured while the gradient is fractionated, allowing separation by polysome peak. Can be split to different resolutions, depending on needs.

(b) Polysomal RNA footprinting: Dilution allows effective treatment and digestion of RNA. This can be calculated and adjusted based on sucrose gradient used. Digestion with a non-specific RNase, such as RNase I, allows near-complete digestion of any unbound RNA. This leaves only the ribosome and polysome associated transcripts to proceed. The large volume must then be concentrated to a usable size, ultrafiltration concentrator tubes are available with membranes suitable for this purpose. Ultracentrifugation through a 34% sucrose cushion allows enrichment of RNA bound to proteins of the correct density to be a ribosome. Extraction and DNase treatment gives the usable RNA footprints.

(c) Total lysate RNA selection and fragmentation: A portion of the lysed sample is held back to extract and compare total mRNA. This will allow comparison between levels of translationally active and total RNA. mRNA is selected from total RNA by magnetic polyA bead binding. This will prevent the majority of reads from being taken up by ubiquitous rRNAs and similar. Chemical fragmentation brings RNA to sizes suitable for library prep and sequencing alongside the polysomal RNA.

(d) Purification and preparation of final samples: Running the samples on a denaturing urea-acrylamide gel allows size selection of RNA fragments. For polysome footprinting, a band is extracted between 28 and 34 bp. For total mRNA, a broader smear is extracted from 50-100bp (variable dependant on chemical fragmentation protocol). Biotinylated oligos complimentary to rRNAs are bound to magnetic beads, allowing depletion of rRNA, freeing up reads in the sequencing for the RNA of interest. From this point normally library preparation can be carried out (as per manufacturer instructions). Send off for sequencing!

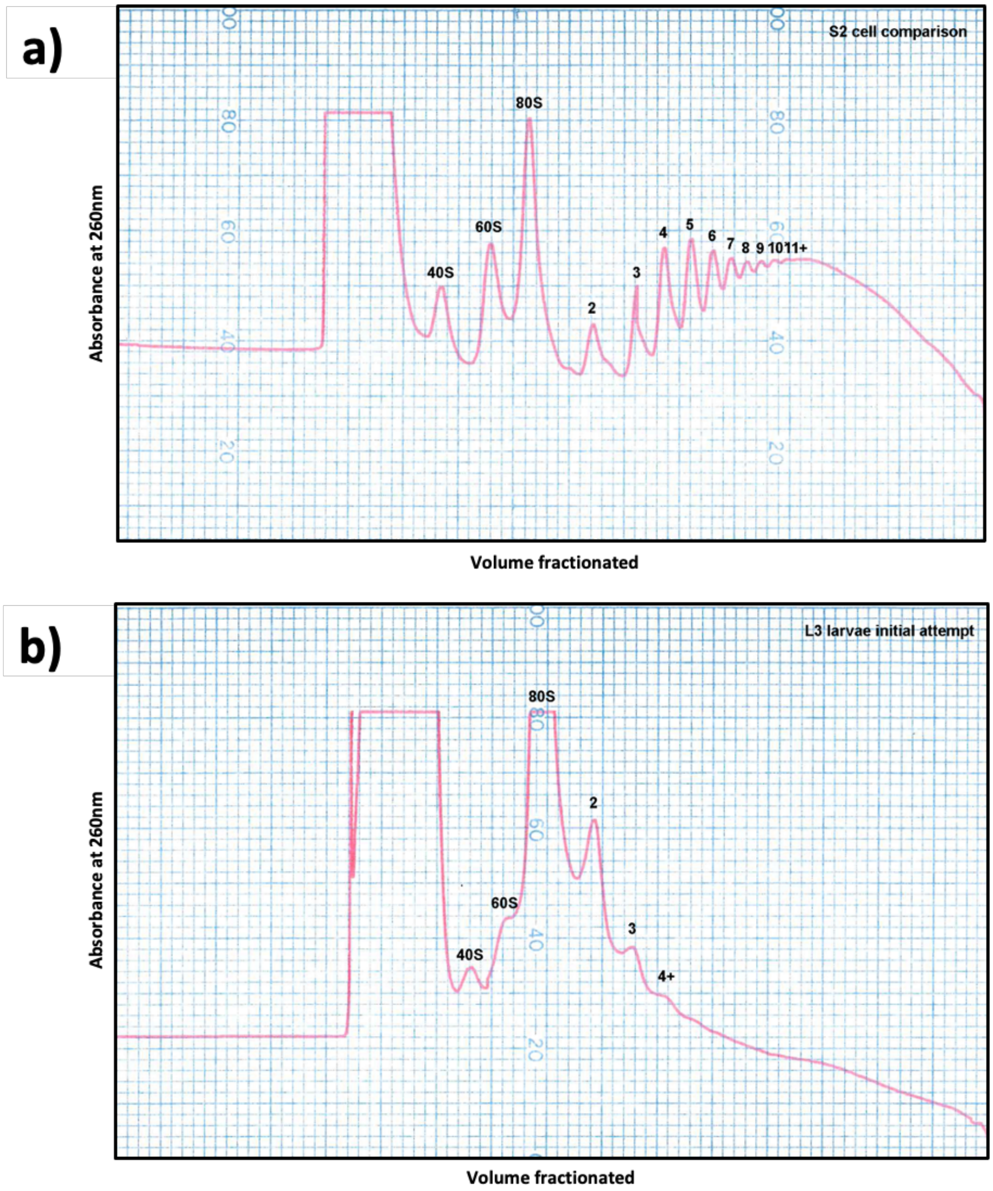


Figure 5.4 – Polysome fractionation protocol based on previous work by Aspden et al. and Patraquim et al. is not suitable for producing high quality polysome profiles from whole *Drosophila* L3 larval samples

Using the existing protocol for polysome fractionation of *Drosophila* embryo samples, detailed in the thesis of A. Mumtaz, does produce a polysome trace with visible separation of the 40S ribosomal subunit, the 60S ribosomal subunit, and the 80S ribosome; as well as polysome peaks corresponding to 2, 3, and 4 ribosomes. However, when this is compared to a sample produced by following the very similar protocol specified by Aspden et al. for polysome fractionation from S2 cell samples, the resolution is vastly superior, showing resolution of polysome peaks up to 11 ribosomes. In addition, the shape of the trace in S2 cells implies active translation and intact polysomes, whilst the "collapse" of the translating polysome in L3 larvae implies that ribosomes may have dissociated from RNA transcripts.

may be actively translating, as well as the need for substantial quantities of polysome-associated RNA, an alternative and optimised protocol was clearly needed in order to progress further. The reduced resolution as initially seen (4+ polysomes in L3 versus 11+ in S2 cells) provides an issue; given that only the 2+ polysome will be harvested for these experiments, the best possible resolution, and highest translational activity, is needed in order to capture as much active translation as possible, as well as to increase confidence in what is part of the true polysome.

5.4.2 Increased concentration of cycloheximide is important in maintaining high resolution polysome traces

Cycloheximide is a naturally occurring fungicide identified in the bacterium *Streptomyces griseus*, known to interfere with translocation of the ribosome during protein synthesis, and blocking translational elongation in eukaryotes. Cycloheximide has been extensively used in biological research to inhibit the process of translation in eukaryotic biological samples (although significant toxicity means that the use of the drug in *in vivo* usage has limitations). In polysome profiling, the use of cycloheximide can stabilise the association between ribosomes and the RNA, by preventing them from running off the strand of RNA, providing a “snapshot” of where the ribosomes are bound on the RNA when the cycloheximide exerts its effects.

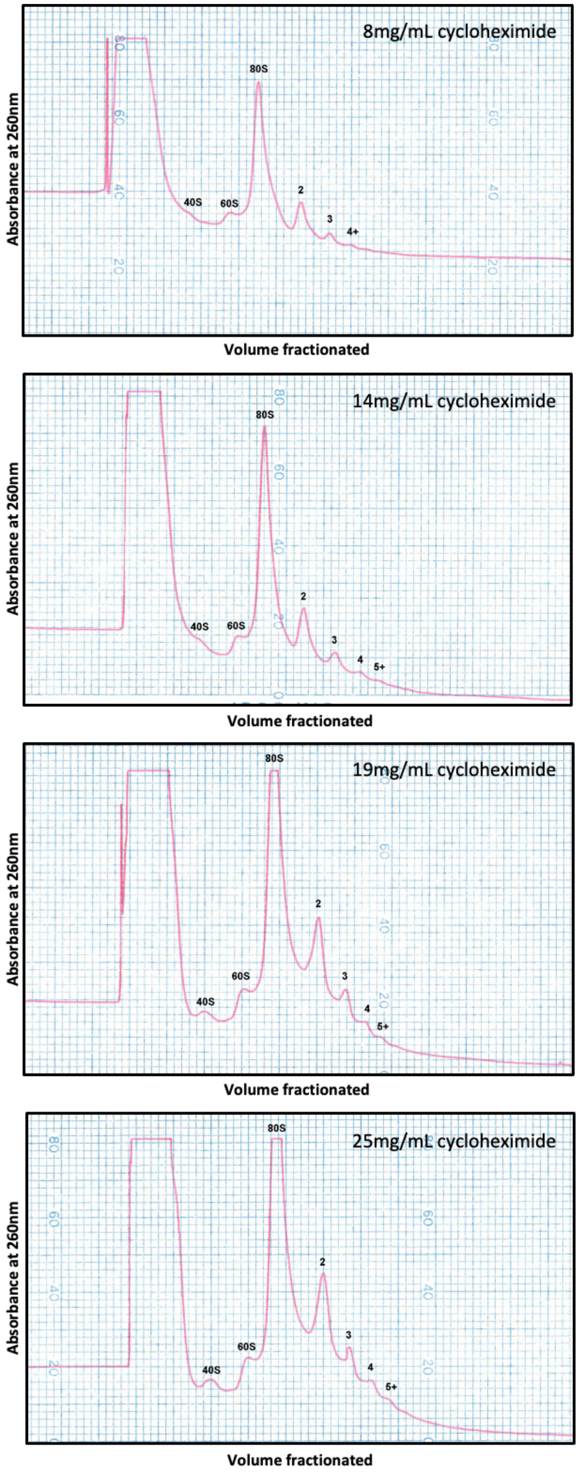
Previous work by the Couso and Aspden labs (whose protocols were formative in developing the protocol used in this thesis) have used cycloheximide for this purpose previously, and work by Antic *et al.* (69) have shown that cycloheximide treatment of *Drosophila* S2 cells while harvesting and lysing does not cause a significant difference in RNA abundances, compared to untreated control. Although the same has not been shown in L3, given that the samples will be snap frozen, and lysed directly into the cycloheximide, a similar lack of impact on RNA profile can reasonably be expected.

Given that the initial attempts on L3 larvae showed a collapsed polysome profile, and no more than 4 ribosomes seen clearly on RNA, cycloheximide concentrations were optimised, in order to see if an increased concentration of cycloheximide might stabilise the association of ribosomes on the RNA, allowing maximum preservation of the polysome. This could be due to a range of factors; the uptake into frozen cells may be less efficient, the excess debris may absorb some of the drug, or the S2 cell type may allow quicker uptake than the WID cells.

As shown in Figure 5.5, polysome fractionation was performed as per the previous protocol, but with varied concentrations of cycloheximide, increasing to a point of maximum solubility in water (8mg/mL, 14mg/mL, 19mg/mL, and 25mg/mL). As seen in Figure 5.5, increasing concentration of cycloheximide to 14mg/mL seems to significantly improve the resolution of the polysome trace, although has much less effect on preventing the collapse of the polysomal profile. Given that the RNA profile shouldn't be significantly impacted by the cycloheximide treatment, and that the highest concentration tested gave the best polysome trace, future experimentation used cyclohexomide at this higher concentration (14mg/mL), it should be noted that the concentration tested by Antic *et al.*, at which they saw no significant difference in the transcript levels between the two expression libraries, was at the much lower concentration of 1mg/mL. It was decided overall though, that the higher concentration was required, and that it is likely that the principle of the cells already being frozen, dead, and lysed preventing differential transcription and translation (for example of stress genes), likely holds true even at a very high concentration.

5.4.3 Increased Triton-X concentration increases quality of the polysome profile

In an attempt to solve the small quantity of RNA detected in the heavy polysome, seen in the previous traces, an increased concentration of Triton-X was tested (60mM compared to the previous 50mM). Triton-X acts as a non-ionic surfactant, and a detergent, able to lyse cells and permeabilise membranes by disrupting their polar structure. At the increased level of Triton-X, the polysome profile can be seen with slightly improved resolution, a slowed tailing-off, and a reduced collapse of the active polysome peaks, allowing the emergence of a 6+ ribosome peak (Figure 5.6). This is likely due to the increased strength of the detergent lysing cells in the more robust L3 tissue (compared to embryonic tissue or S2 cells), allowing access to more RNA and attached ribosomes, as well as providing better uptake of cycloheximide.



8mg/mL cycloheximide

14mg/mL cycloheximide

19mg/mL cycloheximide

25mg/mL cycloheximide

Figure 5.5 – Increasing the concentration of cycloheximide used in the protocol improves resolution, and stabilises the heavier polysome

Using the existing protocol for polysome fractionation of *Drosophila* embryo samples previously failed to give polysome profiles with sufficient resolution, and none maintained significant polysome presence beyond 4 ribosomes.

By incrementally increasing the concentration of cycloheximide through otherwise identical replicates (using the protocol provided by Patraquim et al.) to a concentration that approaches the maximum solubility of cycloheximide in water, the resolution can be seen to slightly increase, and produce a less "collapses" profile, with RNA absorbance showing some polysome bound RNA with 5 or more ribosomes attached.

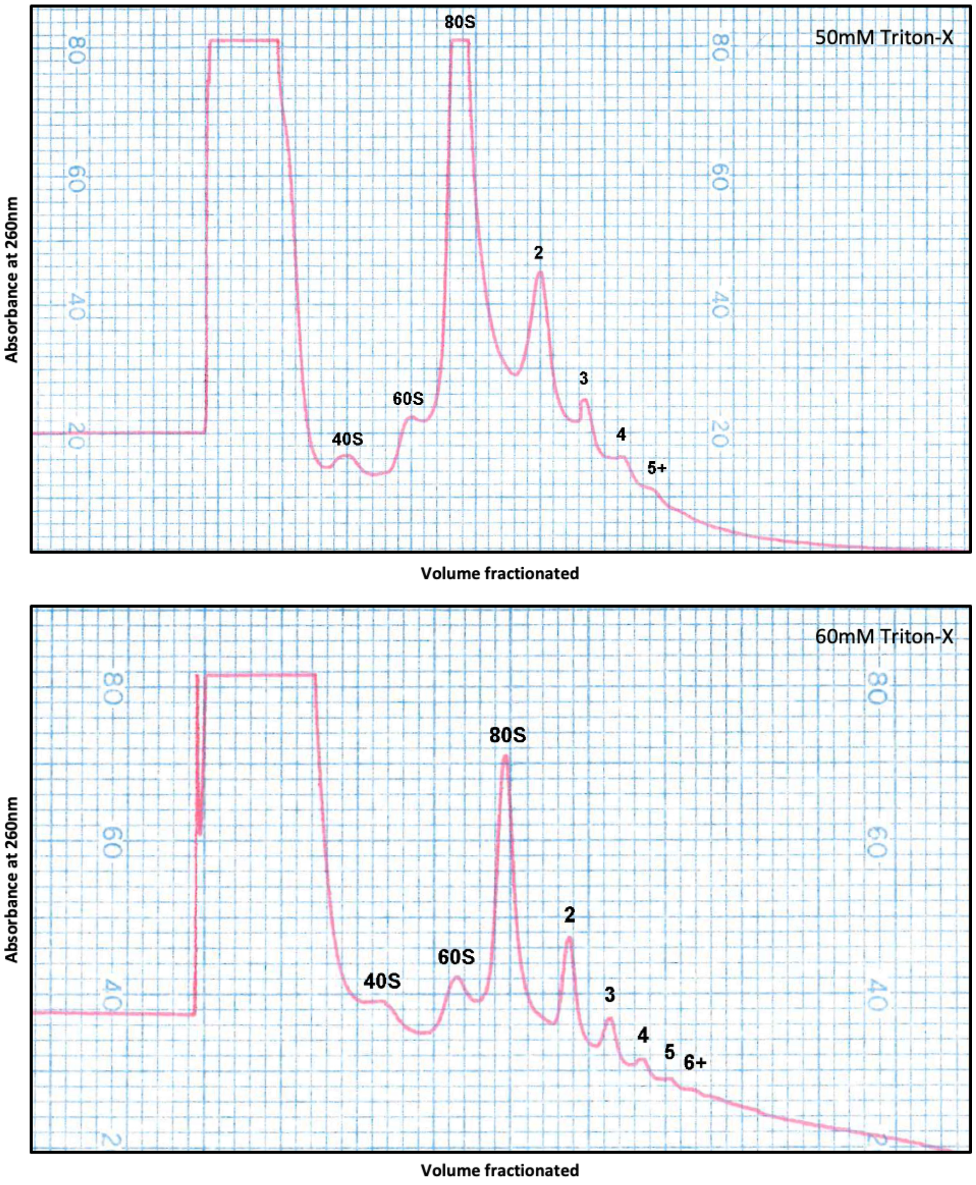


Figure 5.6 – Increasing the concentration of Triton-X used in the protocol improves resolution, and stabilises the heavier polysome

Using the existing protocol for polysome fractionation of *Drosophila* embryo samples previously failed to give polysome profiles with sufficient resolution, and none maintained significant polysome presence beyond 5 ribosomes. Increasing the concentration to 60mM from 50mM provided slightly improved polysome profile, both in terms of resolution and reducing polysome collapse.

5.4.4 Polysome traces see insignificant differences between stepped and continuous sucrose gradient or different fractionation equipment

An update in polysome fractionation equipment (from the Gilson MINIPULS2 with UA-6 UV/Vis ISCO) to the Biocomp Fractionator and Biocomp TRIAX Flow Cell, with Gradient Master Gradient Station) allowed the opportunity to create continuous sucrose gradients, with a higher reproducibility than can be achieved by snap-frozen stepped gradients (previously used by Aspden *et al.* (153)). Both the stepped gradients (used in 5.1 and 5.2.1), and continuous gradients with a minimum sucrose density of 15%, and a maximum density of 60% were prepared (Figure 5.7), and identically treated samples were run on both kinds of gradient, in the same conditions (Figure 5.8). Due to a difference in the internal diameter of the tubes used to create stepped versus continuous gradients, the stepped gradients were fractionated using the Gilson Fractionator, while the continuous gradients were fractionated using the Biocomp Fractionator.

As can be seen in Figure 5.8, there is very little difference seen between the two different approaches. Given the ease and reproducibility of the continuous gradients produced by the Gradient Master Gradient Station, and the advantages of a digitalised readout, and more sensitive measurements available from the Biocomp TRIAX Fractionator and FlowCell, all further experiments used these methods. The BioComp Fractionator is pictured in Figure 5.9.

5.4.5 Increased RNase inhibitor concentration provides some stabilisation of the polysome profile

In an attempt to solve the polysomal collapse seen in the previous traces, an increased concentration of RNase inhibitor was tested (from 200U/replicate to 400U/replicate). There is no significant difference seen between the two conditions. At the increased level of RNase inhibitor, the polysome profile initially looks to have a very slightly improved resolution, and a slightly higher yield (Figure 5.10), but the difference was not consistent across all replicates, and was determined to be too small to justify the high cost of the reagent. The lower concentration of RNase inhibitor was used for all future

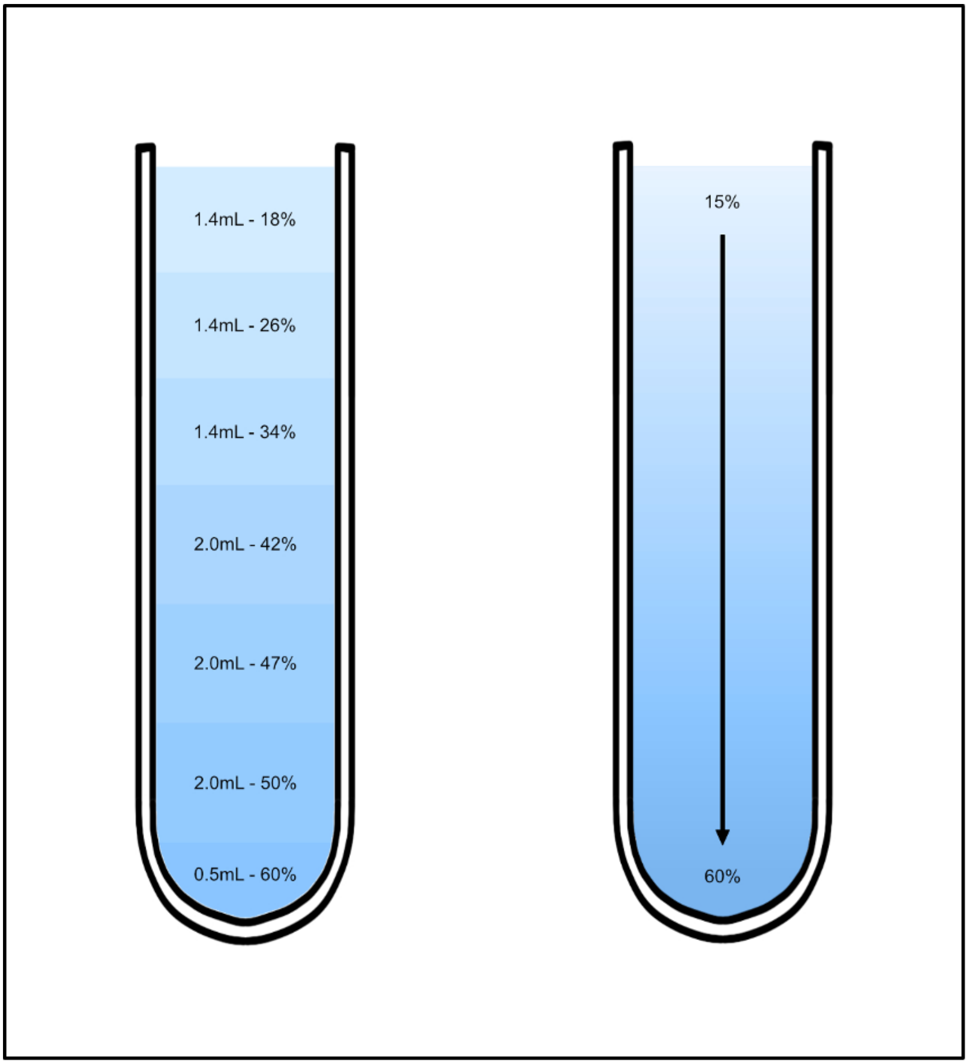


Figure 5.7 – Demonstrating the difference in composition between stepped and continuous sucrose gradients.

The updated equipment, specifically the Gradient Master Gradient Station, allowed reliable and repeatable preparation of continuous sucrose gradients. This graphic demonstrates the difference between the make up of the stepped and continuous gradients used in this project.

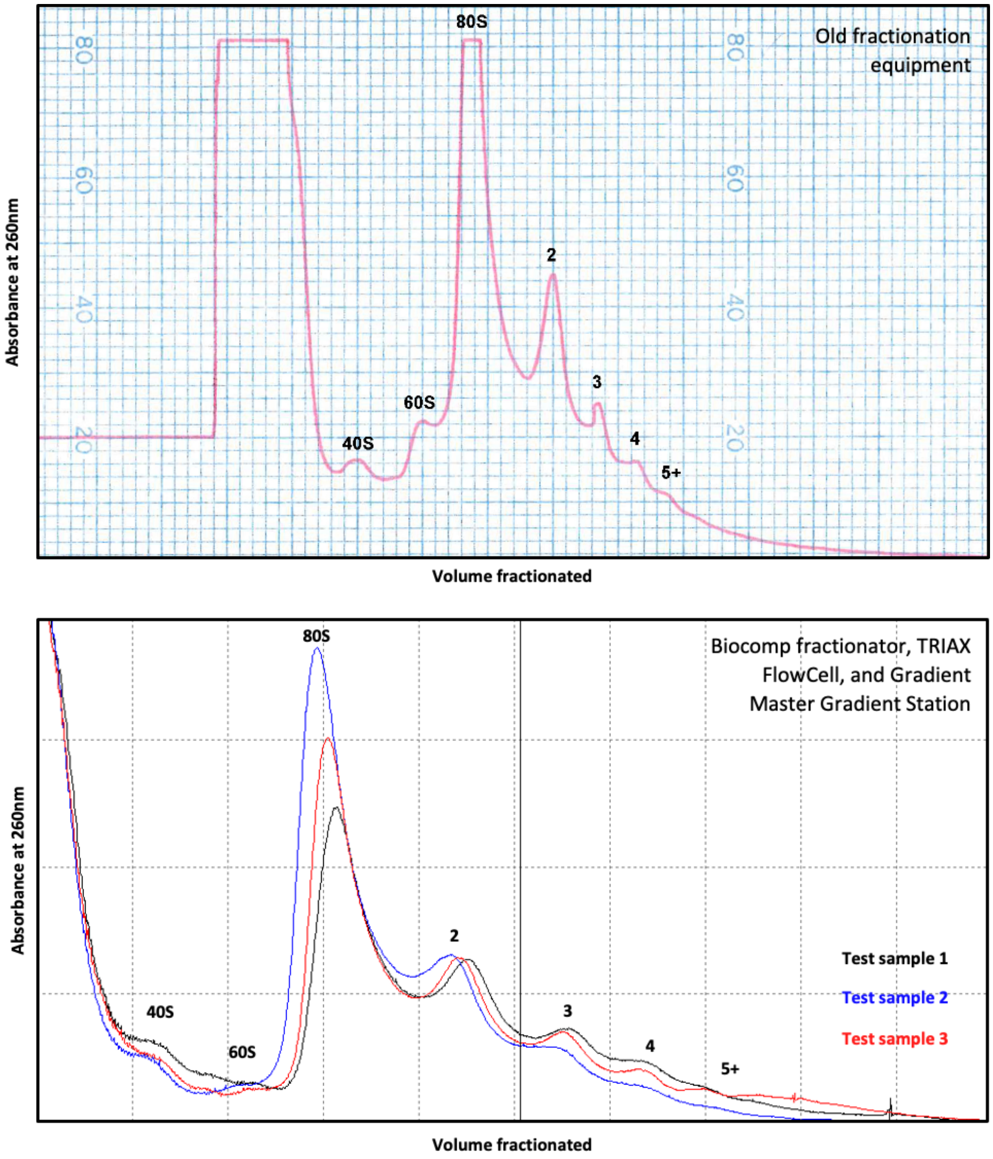


Figure 5.8 – Polysome traces see insignificant differences between stepped and continuous sucrose gradient or different fractionation equipment

The update in polysome fractionation and profiling equipment brought an update in the pump, flowcell, and gradients used. In order to continue the protocol that was being optimised previously, both pieces of kit were tested, in order to show that they were comparable in their output. The Biocomp TRIAX equipment allows for multiple samples to be overlaid on a single plot, which has been done here in order to show the trace to be representative of more than a single run. Although some differences can be seen, this is partly due to the scaling differences between the output on either machine. On both traces, all ribosomal components can be seen, along with polysome peaks up to 5-6 ribosomes. Given the increased sensitivity of the TRIAX FlowCell, and the improved reproducibility of the Gradient Master Gradient Station, this new equipment was used for all subsequent experiments.

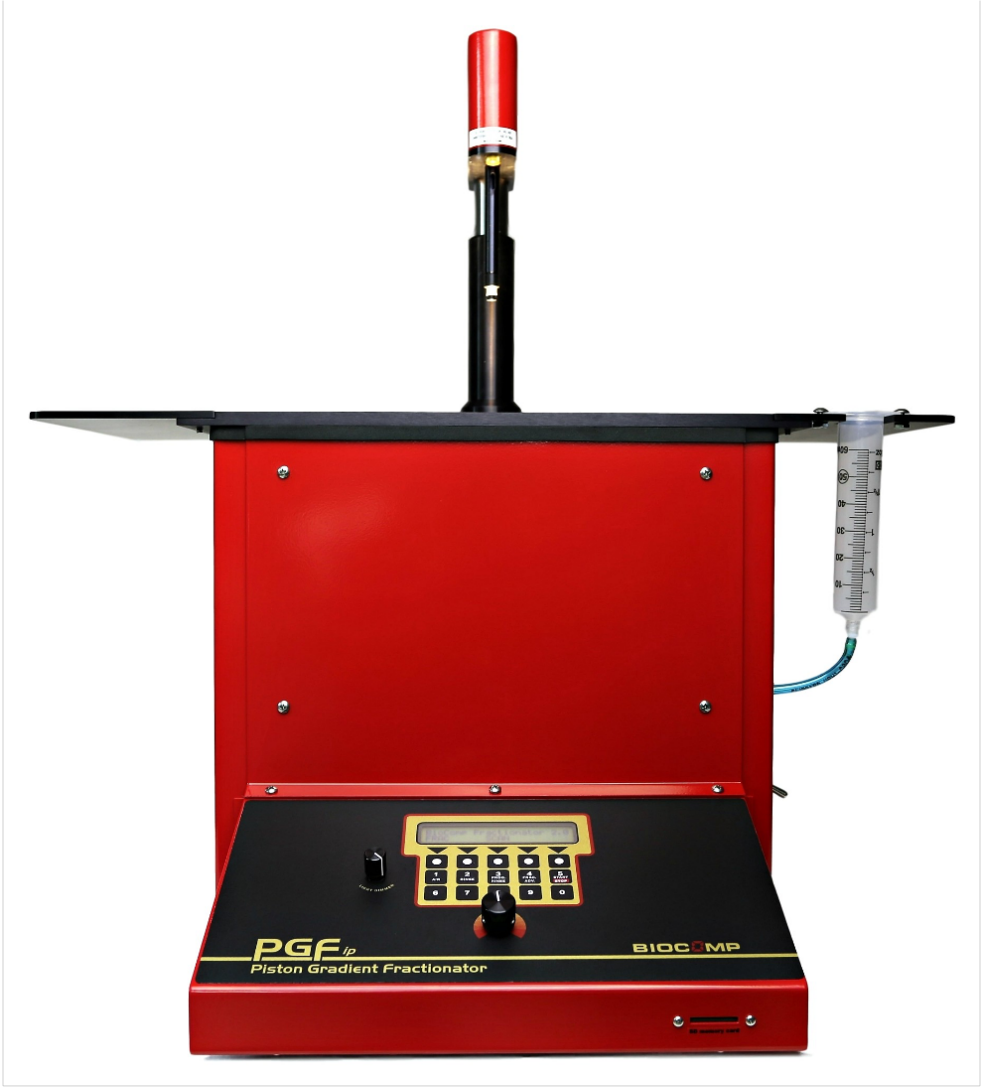


Figure 5.9 – Image of the BioComp Fractionation System used in this work

experiments. Interestingly, this. Does suggest that RNA degradation is not the (main)reason for the low quality of ribosome traces.

5.4.6 Requirement for high yield must be balanced with the reduced resolution seen at higher sample concentration

Poly-ribo-seq, as a technique, requires a large amount of input RNA if it is to be used successfully and meaningfully. This is in part due to several steps in the protocol where significant amounts of RNA will be lost. Ultracentrifugation through sucrose, selection of fractions, RNA extraction from sucrose, RNase digest, subsequent RNA extractions, rRNA depletion/polyA selection, and multiple gel extractions all deplete the concentration of RNA
from the initial input. Due to the necessity of PCR steps during the library prep, reduced concentration of samples requires extra rounds of PCR during library preparation, which can cause increased bias towards simple transcripts which may be preferentially amplified.

Another point to consider is the separation of RNA transcripts by density which relies on sucrose density gradients. These cause polysome bound RNAs to migrate differentially through the gradient during ultracentrifugation, with a migration rate determined by number of bound ribosomes. The “bands” of RNAs bound by a certain number of ribosomes are diffuse, as shown by the absorbance at 260nm against volume fractionated appearing as a peak, rather than a thin spike. By overloading a gradient with overly high concentrations of RNA, the diffuse nature of the peaks may be exacerbated, broadening the individual peaks, and reducing peak resolution.

In addition to this, the fatty nature of L3 sample material poses a challenge not present in cells. Despite efforts to separate the fat layer before centrifugation (further detailed in 5.2.7.2), some residual fat layer tends to persist (although it can and should be minimised).

This layer is seen to interfere with the migration of RNA through the rest of the gradients, which could result in the collapsed, low-resolution trace (as can be seen in

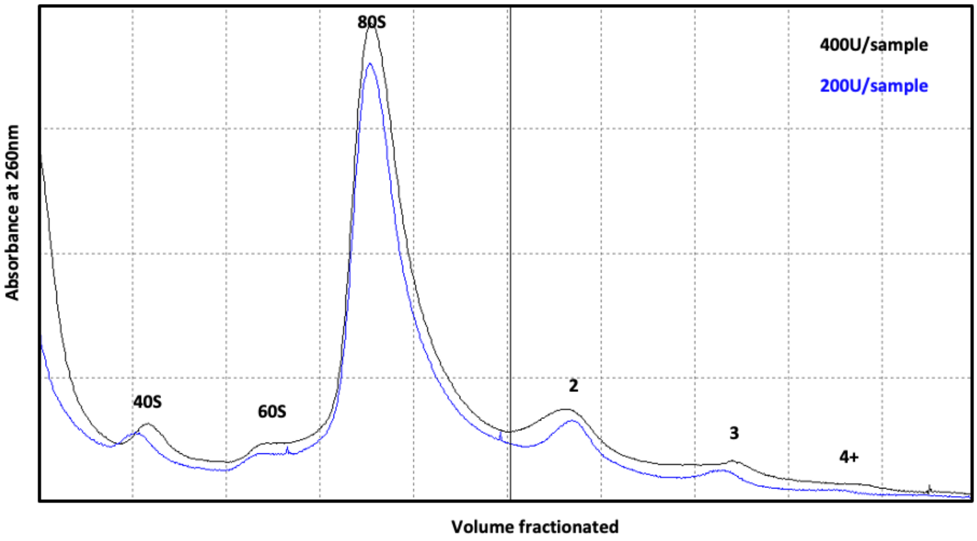


Figure 5.10 – Polysome traces see insignificant difference with increased RNase inhibitor concentration

There is no significant difference seen between the polysome profiles produced by varying RNase inhibitor concentration.

Figures 5.4 - 5.10). With increased input material of L3, the fat layer seems to be both larger, and more pervasive, resulting in a reduced resolution on the polysome profile (Figure 5.10 panel (a)).

In order to optimise the input amount of L3 larvae, different masses of sample were used: In polysome fractionation for whole L3 samples, the yield could consistently be seen to be higher than in the S2 cell samples previously used in poly-ribo-seq experiments (Figure 5.11, panel (b)) with an input of 0.1g of L3 larvae per sample. While an input of 0.3g per sample produced an extremely high concentration of RNA (compared to the previous 0.1g input), the resolution was reduced to the point of indistinguishable peaks. Given that the optimisation running parallel to the yield testing was showing promising resolution with input samples of 0.1g, and similar input (0.08g) in *Drosophila* embryos (albeit with less of the sample mass lost to fat layer removal) had yielded high quality poly-ribo-seq data in previous work; all future samples were carried out with an input of 0.1g of L3 larvae.

5.4.7 Lysis method of L3 larvae has significant impact on the quality of polysome profile

As discussed, the method for lysing and fractionating samples of L3 larvae had not been optimised, and although a method for lysing the relatively similar L2 stage larvae has been established, the differences between the stages (cuticle thickness, fat body prevalence, rate of growth and development, etc.) made it worth evaluating whether the same method was appropriate, or whether other avenues should be investigated. After reference to the available methods, three possible options were tested; manual pestle and mortar lysis, blender tissue homogenisation, and bead-beater. The pestle and mortar method followed the technique used by the Couso lab in their previous work (ground into a fine powder in a liquid Nitrogen cooled pestle and mortar placed in dry ice (90, 153)), while the blender tissue homogenisation was straightforward and relatively inflexible (adding the L3 larvae (stored at -80°C to ice-cooled lysis buffer (still on ice), and straight away homogenising with the blender blades until visibly homogenous, with no visible debris). Contrasting these, several different programs and bead sizes were used to explore bead-beater homogenisation. For the comparison with

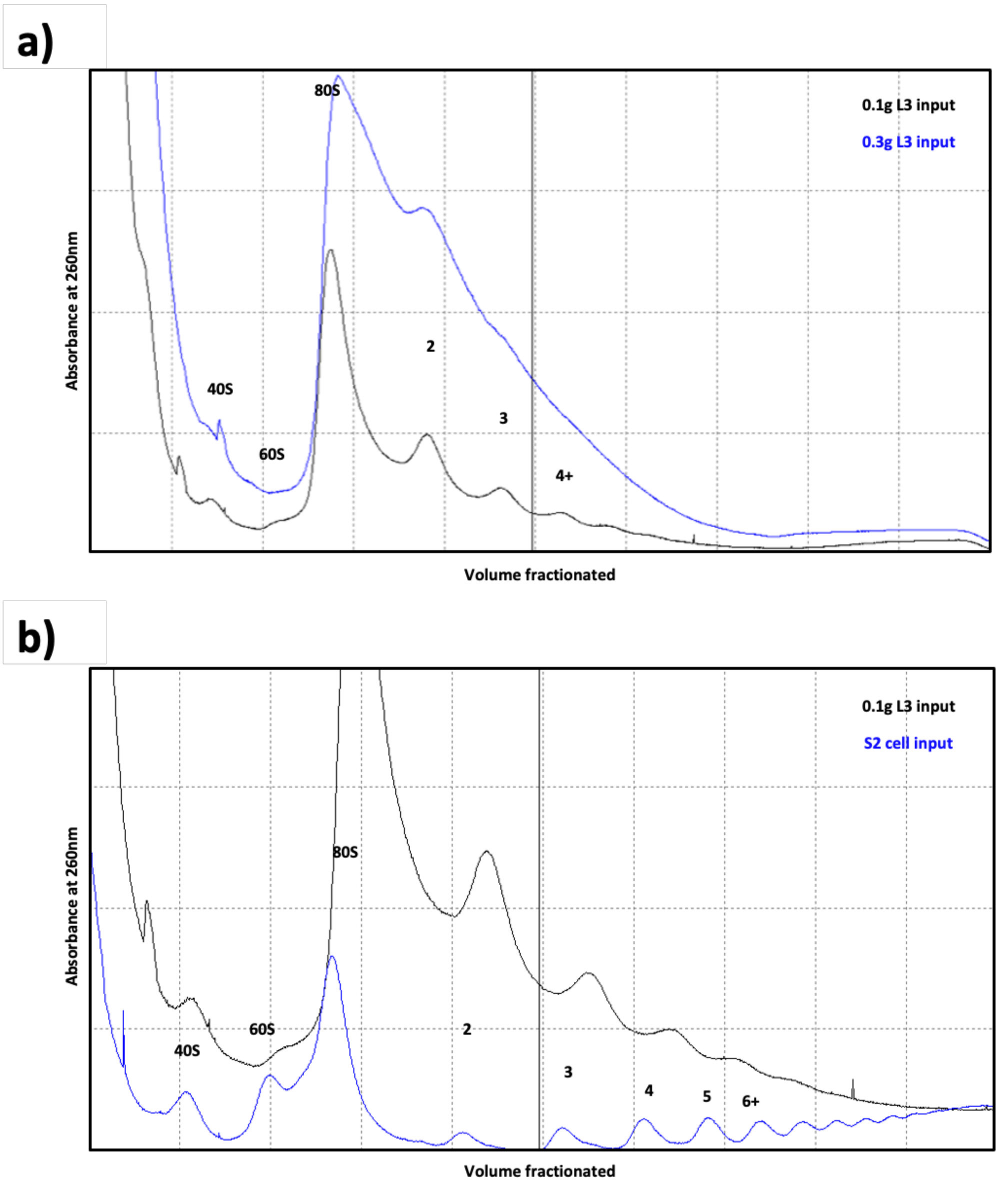


Figure 5.11 – Input mass of sample requires balance between yield and resolution

While an input of 0.3g per sample produced an extremely high concentration of RNA, the resolution was reduced to the point of indistinguishable peaks (panel a). An input mass of 0.1g produced sufficient resolution (panel a), while still maintaining a higher total yield of RNA than the input from the previously successful S2 cell poly-ribo-seq protocol.

other methods, the final optimised bead-beater program and conditions is used. The same quantity of L3 larvae (0.1g) was input for all three methods, and the same lysis buffer was used for all. All methods are further described below.

For pestle and mortar homogenisation, An RNase-free pestle and mortar was submerged in liquid nitrogen, and left until the liquid nitrogen evaporates, at which point the pestle and mortar was placed in a bucket of dry ice, to keep the temperature extremely low. For each replicate, 0.1g of snap frozen L3 sample was transferred directly into the chilled mortar, to which 700µL of lysis buffer should be added dropwise while the sample is homogenised, using the chilled pestle. Keeping the temperature low and working quickly to integrate the lysis buffer into the homogenised sample is crucial to minimising RNA degradation and polysome collapse. Homogenised samples in lysis buffer were then transferred to RNase free 15mL falcon tubes and submerged in liquid nitrogen. The samples were stored at -80°C until required

For bead-beater homogenisation, a custom program was developed and used on the Precellys Evolution cooled bead-beater. The settings were as follows: 2mL tube volume, cryolys on, temperature at 0°C, speed at 5500RPM, 4x 25 second cycle, 30 second pause. Prior to this, two unsuccessful programs were trialled as follows. Failed program 1 (based on Precellys Soft tissue disruption program): 2mL tube volume, cryolys on, temperature at 0°C, speed at 5800RPM, 2x 15 second cycle, 30 second pause. Failed program 2: 2mL volume, cryolys on, temperature at 0°C, speed at 7500RPM, 2x 20 second cycle, 10 second pause.

The initial attempt at bead-beater homogenisation was unsuccessful, failing to lyse the samples sufficiently to run on a sucrose gradient. Following this, different types of disruption beads were tested, along with higher and lower sample:lysis buffer ratios. Figure 5.12 shows that an increased amount of sample per volume of liquid resulted not just in an increase in the material lysed (as would be expected), but an increase of thoroughness of the lysis, with whole L3 larvae still visible in the homogenisation attempts with lower sample:lysis buffer ratio. Soft tissue disruption beads and soil disruption beads were selected to be tested, as there was no recommended kit available for L3 or L3-like samples.

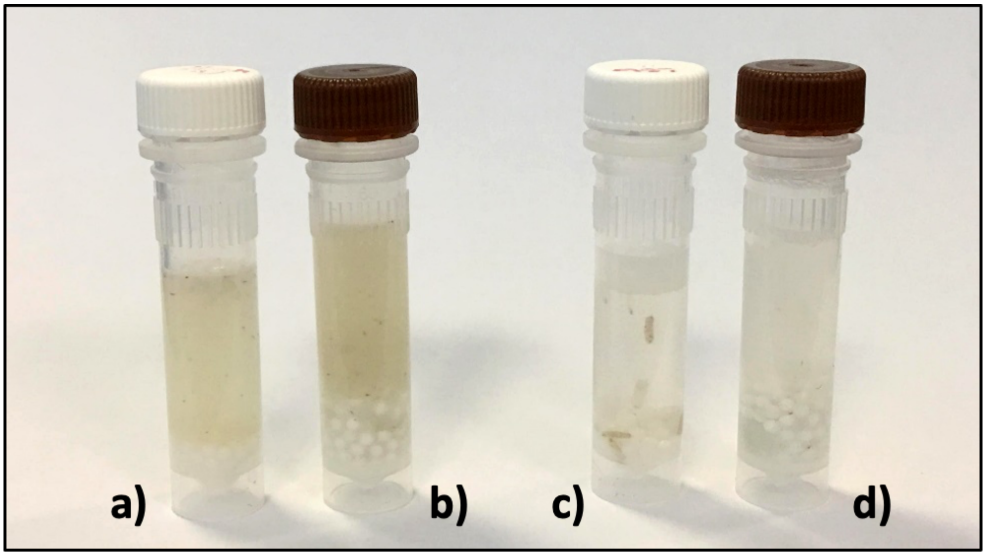


Figure 5.12 – Input mass of sample requires balance between yield and resolution

Left to right: (a) High sample:lysis buffer ratio with soft tissue disruption beads, (b) high sample:lysis buffer ratio with soil disruption beads, (c) low sample:lysis buffer ratio with soft tissue disruption beads, (d) low sample:lysis buffer ratio with soil disruption beads. Full lysis of the L3 larvae can be seen to only take place thoroughly with the higher sample:lysis buffer ratio. The variant of beads used appears to have minimal difference to the efficiency of the lysis.

Soft tissue beads use 1.4mm ceramic (zirconium oxide) beads and are designed for more fragile tissue homogenization such as brain, liver, kidney, skin, plant leaves, and mammalian cells. Soil disruption beads a mix of glass and ceramic (zirconium oxide) beads between 0.1mm and 4.4mm in diameter, in order to ensure thorough lysis of smaller and larger sample fragments of variable resilience, and are recommended for unspecified human or animal tissues, soil, roots, and faeces. These two products (described further in methods) were selected as the most suitable for lysing L3 larvae completely and thoroughly. As can be seen in Figure 5.12, little difference can be seen between the bead types, although with a lower sample:lysis buffer ratio, the soil disruption beads appear less likely to leave intact larvae. As the higher sample:lysis buffer homogenised better than samples with a lower ratio (and were therefore used for all subsequent experiments), both kits were deemed suitable for this lysis. As such, the larger beads of the soft tissue disruption kit were used in all subsequent experiments, as separation of sample and beads by pipetting is significantly easier. All bead beater lysis was carried out at 0-1°C. For rotor-blade homogenisation, the blades and rotor of a (brand) tissue homogeniser were cleaned with RNaseZap, rinsed with RNase free water, and allowed to dry. An RNase free 15mL falcon tube containing 2100µL of lysis buffer and 0.3g of the frozen late L3 wandering male larvae was aligned with the rotary blades, in a bucket of ice, and the rotary blades were lowered into the sample, and the homogeniser turned on, and set to high speed, until the sample was visibly homogenous, and no fragments of L3 remained. This lysed sample was then incubated on a rolling platform at 4°C for 1 hour and used for polysome fractionation.

An obvious and significant fat layer was readily observed in bead-beater homogenised samples, which was pervasive, and had a strong negative impact on polysome profile resolution. This prompted optimisation of fat layer removal (from the initial step used in L2 sample preparation in previous work), both in bead-beater, and pestle and mortar, homogenised samples. This is further discussed in 5.2.7.2.

As can be seen in Figure 5.13, bead-beater homogenisation preserves the profile of the heavy polysome (higher translational activity) region better than either of the other methods. The bead-beater homogenisation and the pestle and mortar method appear to have approximately the same resolution to distinguish between peaks, while the blender tissue homogenisation method provided the lowest peak resolution. Pestle and

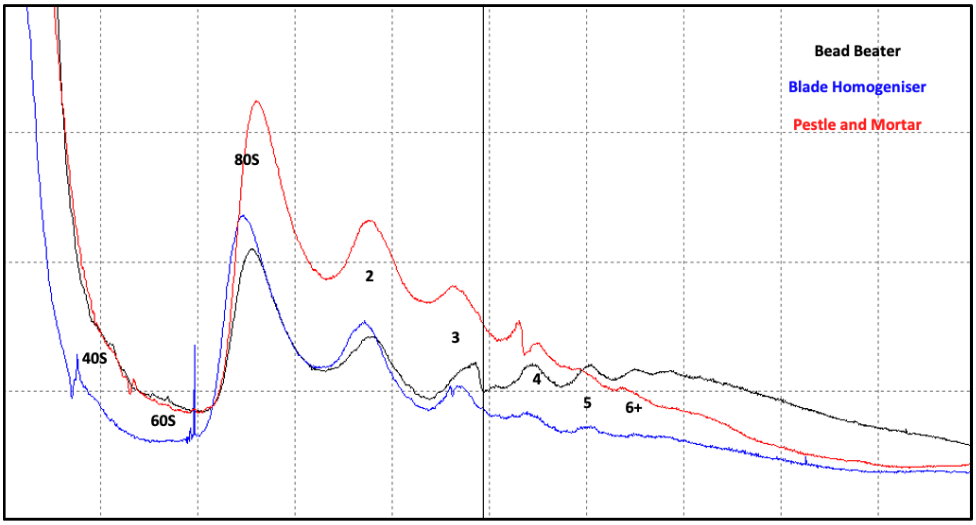


Figure 5.13 – Comparison of lysis methods shows noticeable differences between methods:

Bead-beater homogenisation better preserves the profile of the heavy polysome (higher translational activity) region better than either of the other methods. The bead-beater homogenisation and the pestle and mortar method appear to have approximately the same resolution to distinguish between peaks, while the blender tissue homogenisation method provided the lowest peak resolution. Pestle and mortar homogenisation appears to provide the highest yield, although with lower RNA concentration in the heavier polysome compared to bead-beater homogenisation, the difference between the two is not substantial. Blender tissue homogenisation provided the lowest yield of the three methods.

mortar homogenisation appears to provide the highest yield, although with lower RNA concentration in the heavier polysome compared to bead-beater homogenisation, the difference between the two is not substantial, with the same number of peaks (plus or minus 1) seen in both, and a similar yield and separation between peaks. Blender tissue homogenisation provided the lowest yield of the three methods. Notably, the 40S and 60S peaks are either very difficult to see in many of the traces seen with these lysis methods (Figure 5.13). As such, the traces were compared to an S2 cell profile produced under the same conditions (post-homogenisation with pestle and mortar), in order to ensure that the subsequent peaks aligned with the equivalent peaks in the S2 samples, and are multi-ribosome peaks as expected. Figure 5.14 confirms that the polysome peaks do in fact align as expected, and the lack of an obvious 40S or 60S peak does not skew the subsequent peaks.

Comparing the methods, bead-beater lysis was deemed to produce the highest quality polysomes profiles, followed by the slightly inferior pestle and mortar lysis. Blender tissue homogenisation was the worst method by yield and resolution, so was not further considered. Although the bead-beater method was able to produce the higher quality polysome profiles, it also was less reproducible than the manual pestle and mortar technique, sometimes failing to lyse individual larvae, which were then left floating, whole, in the lysing sample. In addition, this method could not be carried out at less than 0°C, potentially reducing RNA stability and preservation of ribosome association. As a result, manual pestle and mortar homogenisation was used in experiments going forward. Having seen the large and obvious fat layer produced by the bead-beater homogenisation, and the importance of removing this layer in producing a usable polysome profile, increased fat layer separation and removal was prioritised for pestle and mortar homogenisation as well.

5.4.8 Effective removal of fat layer is crucial to high-quality polysome fractionation

As discussed in the previous section, the presence of a fat layer in the samples significantly disrupts their running through the sucrose gradient. Although a fat and debris layer was present in the embryo poly-ribo-seq previously used, this was easily removed by a single centrifugation step. The fat layer produced by homogenisation of

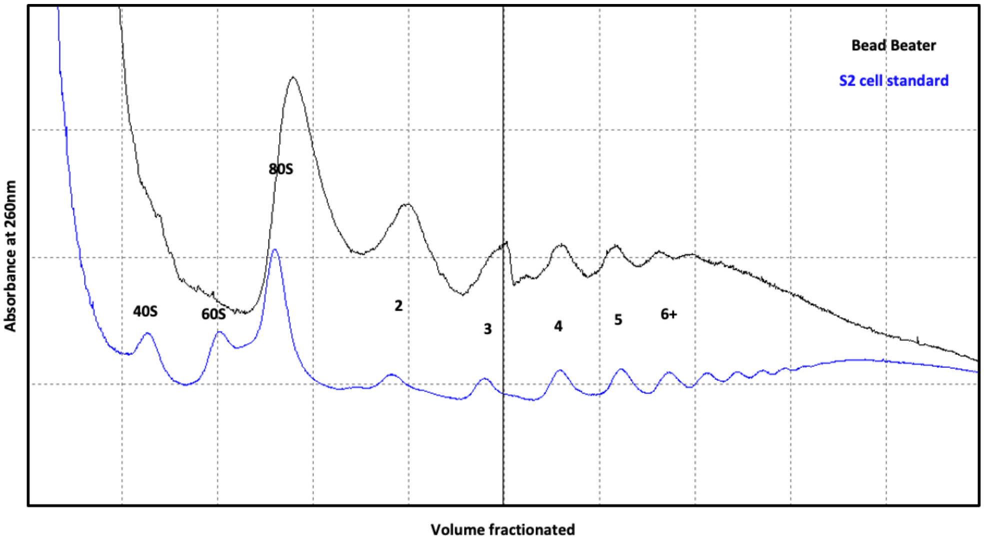


Figure 5.14 – Comparison of Bead Beater lysis of L3 with S2 cell standard:

The total volume fractionated at each peak aligns between the L3 and S2 samples, showing that the polysome peaks are where they would be expected, despite the apparent absence of ribosome subunit peaks.

L3 larvae was larger, and more pervasive; still present sufficiently after this step to disrupt the polysome profile.

After testing, an improved process was found. Following the initial centrifugation at 3000g) at 4°C for 10 minutes., the majority of the fat layer was removed by widened-tip pipette (created by use of a sterile scalpel to cut a p1000 pipette tip 4mm above the standard aperture. This was followed by additional centrifugation steps of the same speed and duration, after which the lysed sample (beneath the remaining fat layer) was removed by a fine-tipped syringe (21G 1 ½" – Nr. 2 – 0.8 x 40mm), taking liquid from the bottom of the tube. The centrifugation and syringe steps were repeated twice more, before proceeding with the sample. The result of these extra steps, along with the prior optimisation steps already described, provided a much improved polysome profile (Figure 5.15).

5.4.9 Application of relative centrifugal force over a greater time negatively impacts polysome resolution versus a shorter centrifugation

The relative centrifugal force needed to achieve sufficient separation of ribosome bound transcripts has been established in the previously mentioned work. Due to the decreased resolution seen in the whole L3 samples, exerting this relative centrifugal force at reduced speed spins (over a greater time,) was carried out for comparison, and to see whether the resolution might be improved. The initial "high speed" run was carried out using a swing out SW40Ti rotor at 31000rpm, providing average relative centrifugal force of 121355g, and maximum relative centrifugal force of 170920g. With these settings, sufficient separation could be achieved after 210 minutes.

To compare to this, samples were also centrifuged on the same rotor at 25000rpm (average rcf 121355, maximum rcf 170920g) for 325 minutes, and 18000rpm (average rcf 40915g, maximum rcf 57625g) for 623 minutes; in order to produce the equivalent force over time. Fractionation of these samples provided traces showing that the shorter centrifugation timeat a higher speed produces improved resolution of polysome peaks (Figure 5.16). Interestingly, the lower speed centrifugation produces a polysome trace very similar to the attempts prior to the optimisation of lysis methods and fat

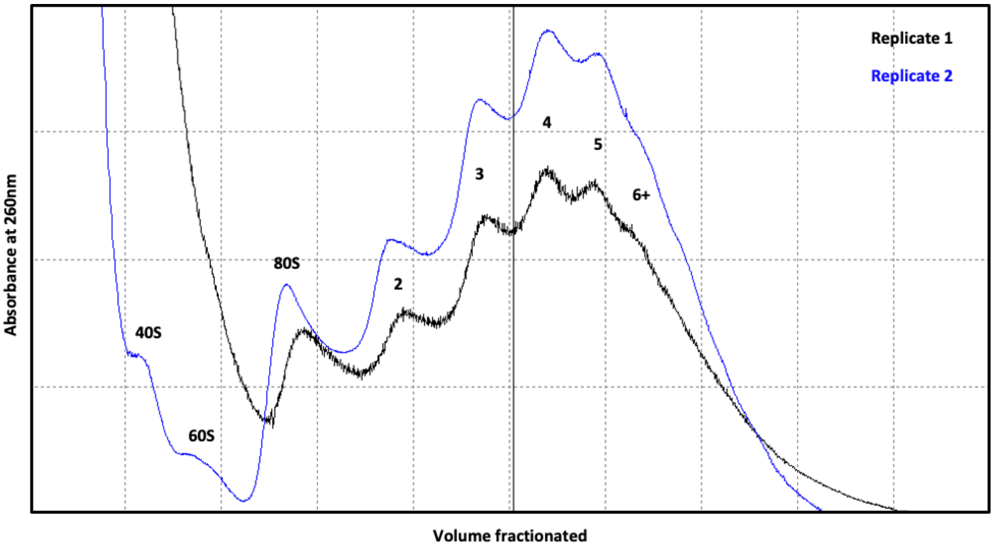


Figure 5.15 – Multi-stage removal of fat layer by centrifugation is crucial in maintaining an actively translating polysome profile:

The additional centrifugation steps, combined with the use of both wide-tipped pipette and narrow gauge syringe (to stop disruption of separated fluid layers) produces a high resolution, high yield, actively translating polysome.

layer separation and removal. This suggests that the lower speeds may be insufficient to pass the polysome-associated RNA through the residual fat present in the lysed sample.

5.4.10 Centrifugation of the samples is a time-critical step in producing high quality polysome profiles

Given the large number of polysome fractions required to guarantee high-yield, high-quality polysome profiles for multiple samples (some of which may require pooling multiple fractionated samples), working in batches was a more efficient way of generating samples. In order to test how many lysed samples ready to centrifuge could effectively be produced at once (as ultracentrifugation with a rotor that only holds 6 samples is the limiting step for how many samples can be processed at once), the effect of leaving lysed samples, loaded onto a sucrose gradient, overnight was tested. Mass production of samples ready to centrifuge, followed by back-to-back ultracentrifugation, would allow for samples to be processed with a much higher efficiency. To this end, two samples were prepared from the same lysis of L3 larvae and loaded onto sucrose gradients. One was immediately put through ultracentrifugation, whilst the other was left at 4°C for 24 hours before ultracentrifugation.

As seen in Figure 5.17, the sample left overnight has lower RNA presence recorded from 80S ribosome onwards (through the active polysome). The 40S and 60S subunits are better resolved on the immediately centrifuged sample, and the RNA detected is higher. It seems likely that the ribosomes are more likely to detach from RNA over a longer timeline, despite the relative stability provided by the protective reagents (cycloheximide, etc.) in the lysis buffer. Both polysome profiles are of a high enough quality to confidently use, however. Ultimately all future samples were centrifuged immediately after lysis to maximise quality and yield, but it is useful to know that samples can feasibly be stored for a short while before use.

5.5 Summary using whole *Drosophila* L3 larvae

In order to produce the L3 reuquired for this optimised protocol, adult breeding stocks of *Drosophila* capable of producing each desired genotype (Pacman mutant, Dis3L2

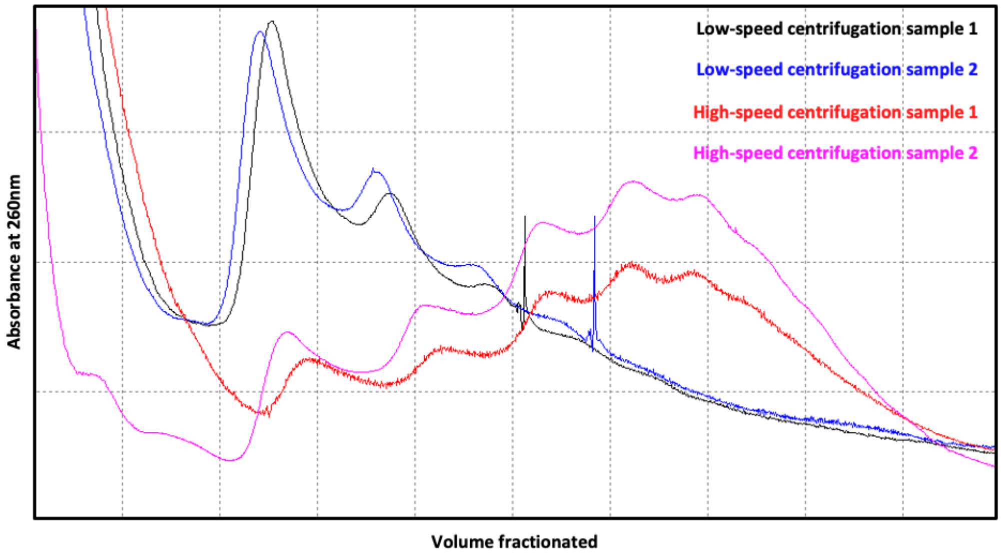


Figure 5.16 – Reduced speed centrifugation over a longer time does not produce sufficient resolution of polysome peaks:

The reduced speed centrifugation over an extended time produces a polysome trace very similar to the attempts prior to the optimisation of lysis methods and fat layer separation and removal. This suggests that the lower speeds may be insufficient to pass the polysome-associated RNA through the residual fat present in the lysed sample.

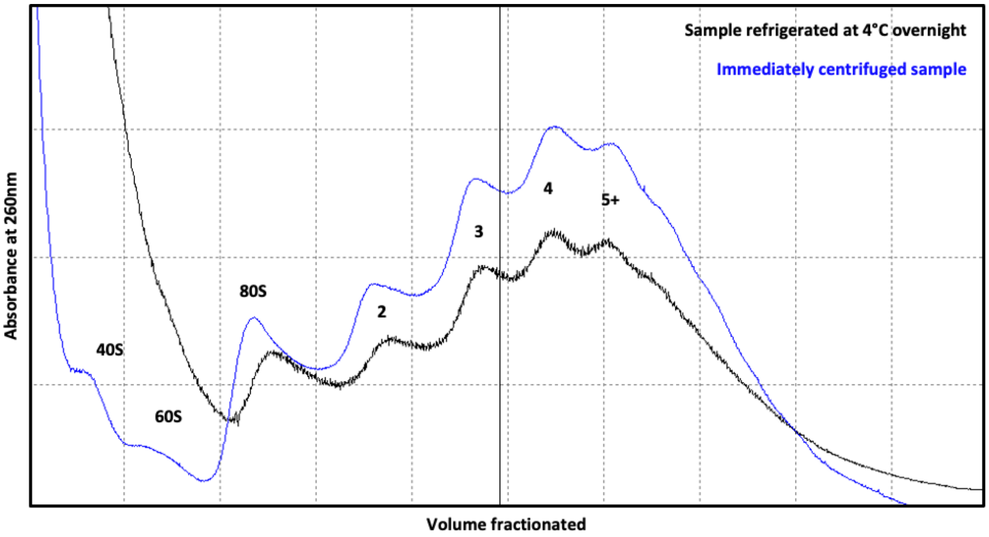


Figure 5.17 – Centrifugation of the samples is a time-critical step in producing high quality polysome profiles with immediate centrifugation improving polysome stability:

The sample left overnight has lower RNA presence recorded from 80S ribosome onwards (through the active polysome). The 40S and 60S subunits are better resolved on the immediately centrifuged sample, but the RNA detected is higher.

mutant, isogenic wild-type control) were tipped into new bottles, and left in these bottles for 3 hours, to ensure that all eggs present in the new stock were born within the same 3-hour window. All bottles were kept at 25°C, to ensure that all stocks matured at the same rate. In order to collect wandering L3 larval samples, they were aged 72 hours (at 25°C) from egg lay and examined under the microscope in order to visualise markers for the desired genotype and identify the male larvae (as only male larvae were used for this experiment due to Pacman null mutations being lethal in females), these were collected carefully with tweezers as they wandered from the food. Once collected carefully with tweezers into an open Eppendorf, they were snap frozen within 15 minutes of collection to avoid stress, differential gene expression, or polysome collapse from lack of oxygen, starvation, or temperature change. Frozen samples were stored at -80°C until required.

Having gone through significant optimisation (of cycloheximide concentration, Triton-X concentration, RNase inhibitor concentration, lysis methods, fat and debris removal methods, ultracentrifugation speeds, fractionation equipment, input quantity, gradient type), although L3 samples were now producing polysome traces of a much higher quality, the newly optimised protocol was required in order to produce anything of a usable quality. Comparisons between the different traces were made in order to evaluate the viability of different protocols. Once the viability was evaluated, the rest of the experiment would be able to continue: Polysome fractionation, RNA footprinting and fragmentation, polyA selection and rRNA depletion, purification of desired RNA fragments, library preparation, and sequencing.

5.5.1 Comparison between optimised L3 larvae polysome profile and S2 cell polysome profile show the optimised L3 protocol to be of comparable quality

To ensure that the optimised polysome fractionation protocol was suitable and worthwhile for continuing with into the subsequent stages, a comparison was run between the initial “ideal” of polysome fractionation of *Drosophila* S2 cells, and *Drosophila* L3 larvae using the new protocol. As can be seen in Figure 5.18, the resolution of polysome profile of the samples carried forward are nearing the quality of the S2 cell ideal. Although the quality does not quite reach the same high standard reliably, the presence of an active polysome in the sample, along with clear resolution of peaks, is sufficient to confidently separate the monosome from the polysome, which is all that the poly-ribo-seq part of this project requires. All samples were prepared with a small fraction of the total lysed sample held back for a “total RNA” comparison to the “polysomal RNA”, as described previously in Chapter 2 - Methods. It should be noted that biological differences between cells (such as the very hight translation in S2 cells versus the slowing translation of L3 tissues as they enter pupation) likely is a strong contributing factor.

5.5.2 High quality polysome fractions achieved for all replicates across all genotypes

An excess of replicates, beyond those planned to carry on through the rest of the poly-ribo-seq process (at least twice the 2 required replicate for each condition) were fractionated, and their polysome profiles compared. As shown in Figure 5.18, those with the best resolution and polysome stability were selected to be carried forward for the rest of the experiment. At least two replicates of sufficiently good quality were available for each genotype. These traces show that post-optimisation, the protocol to carry out polysome fractionation on whole *Drosophila* larvae is capable of reliably producing high-quality polysome profiles, suitable for separation into fractions, and with a high enough yield throughout the profile to explore polysomally associated, and actively translated RNA transcripts through methods like poly-ribo-seq. As discussed in section 5.4.10, the quality does not reliably reach the same high standard as the S2 cell fractions, which are much easier to reproduce at a high resolution. This is likely due to factors that had to be dealt with during the optimisation steps. For instance, although numerous steps were taken to ensure the minimisation of the fat layer, some residual layer tends to be pervasive, and likely still provides some disruption to the fractionation; however, the traces produced were deemed of sufficient quality to continue with the experiment.

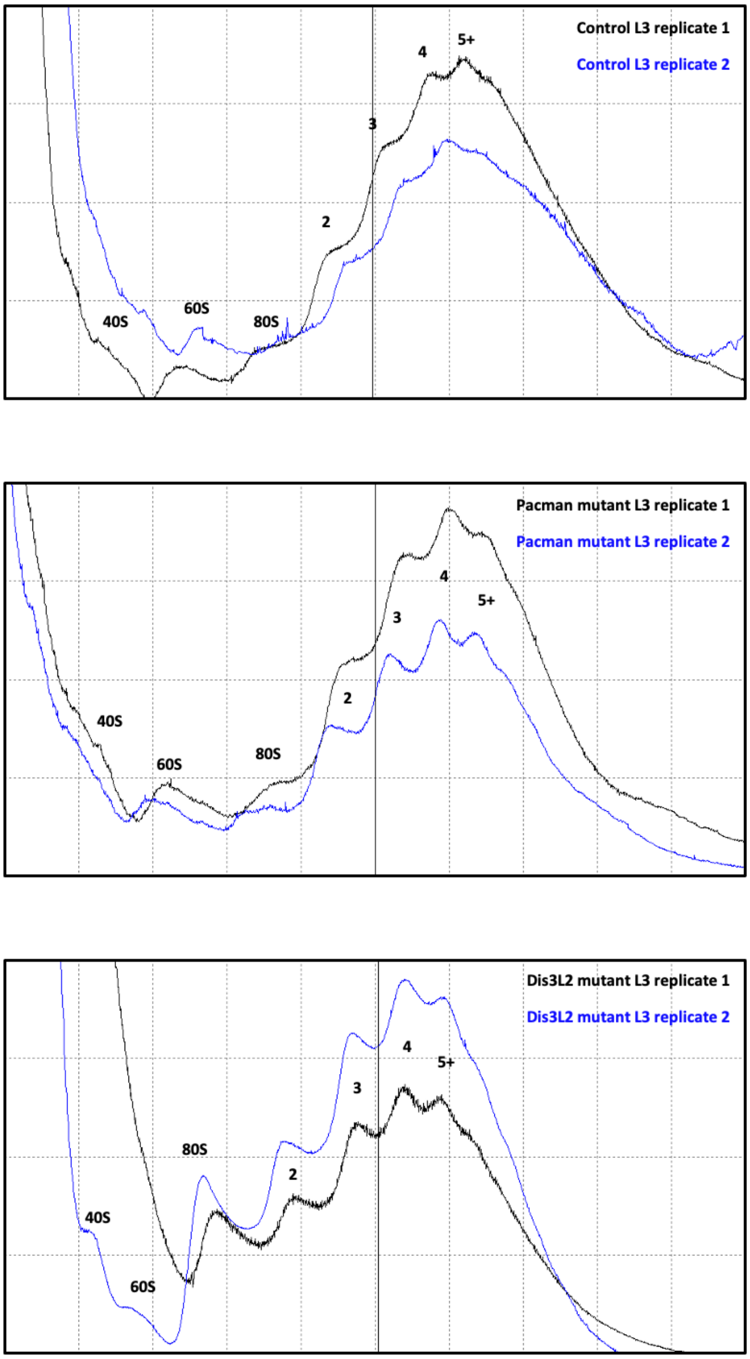


Figure 5.18 – High quality polysome fractionation achieved across all genotypes:

Sufficient high quality polysome fractions were generated for poly-ribo-seq of Pacman mutant, Dis3L2 mutant, and isogenic control *Drosophila* L3 larvae. These samples were carried forward for sequencing and analysis.

5.6 Size selection shows expected distribution of digested and fragmented RNA fragments across all samples

To reach the point of separating RNA library fragments by size, samples were lysed, centrifuged at high speed through sucrose density gradients, and polysome fractions (everything upwards of 2 ribosomes) collected and snap frozen in liquid Nitrogen. These were then pooled and diluted to a maximum of 10% sucrose in a ribosome and RNA protecting buffer. An RNA digest and fragmentation were then carried out, using RNase for the polysomal fractions, and chemical fragmentation for total lysate RNA (taken prior to ultracentrifugation). Following this, the samples were concentrated using centrifugation at 4000g at at 4°C through MWCO Ultrafiltration concentrator filters. The concentrated samples were then centrifuged at high speed in a tabletop ultracentrifuge, through a 34% sucrose cushion, allowing ribosome-bound RNA fragments to pellet. These were subsequently resuspended in QIAzol, and a standard RNA extraction carried out. Oligo-dT selection and rRNA depletion were carried out to produce libraries of polyA RNAs, with depletion of overly-abundant rRNAs. After running the digested and fragmented RNA samples on an acrylamide gel (as described in Chapter 2 – Methods), the visualised gel (Figures 5.19 and 5.20) shows the expected distribution of RNA molecule lengths across all samples. The ribosome footprinted samples show a strong, distinct band at 28-32 base pairs (the length of RNA that a ribosome occupies and protects when bound), while the chemically fragmented total RNA samples show a broad smear, as would be expected from a random fragmentation, centred around approximately 50-80 base pairs (as per the aim of the protocol designed by the Couso and Aspden labs (90, 153)).

5.7 Size selection shows expected distribution of adaptor-bound RNA fragments across all samples

After polyA selection of RNA from total lysate, and rRNA depletion of polysomal RNA, library preparation (using NEBNext Multiplex Small RNA Library Prep Set for Illumina) had to be carried out. Input concentrations and pooling details are described in section 5.8.3. Following the adaptor ligation step of the library prep, the adaptor-bound RNA

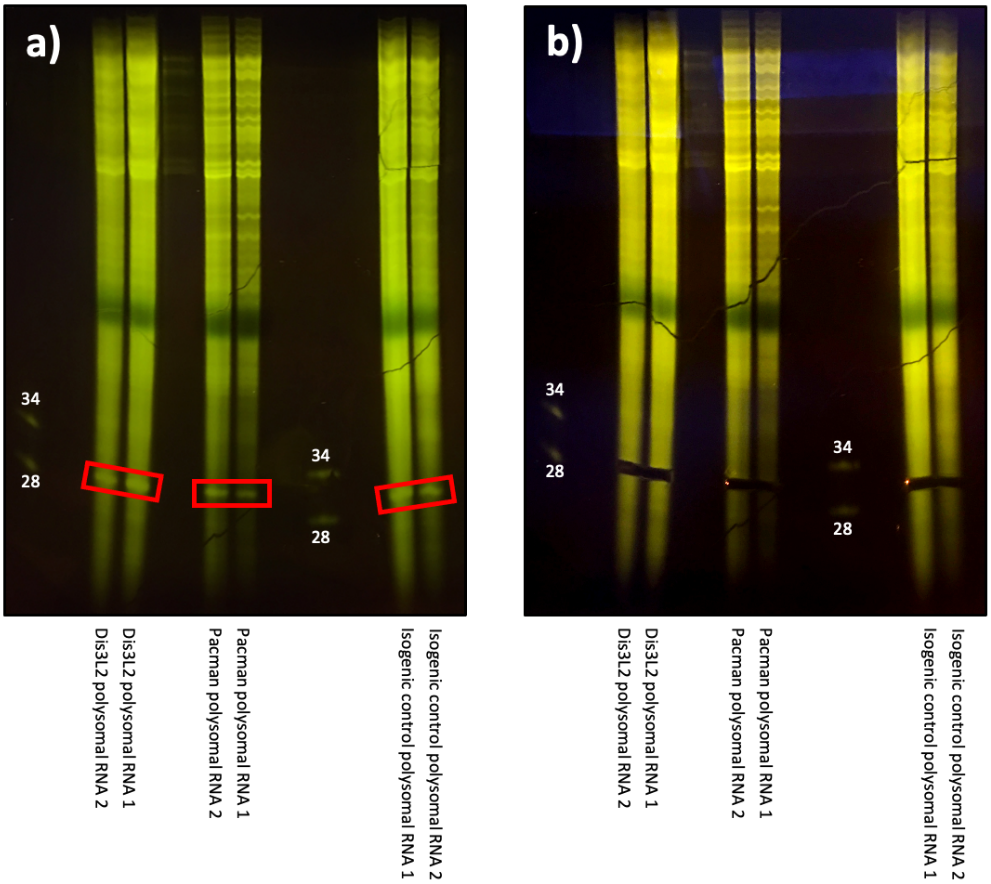


Figure 5.19 – Size selection shows expected distribution of digested RNA fragments across all samples

The visualised gel shows the expected distribution of RNA molecule lengths across all samples. The ribosome footprinted polysomal RNA samples show a strong, distinct band (panel a) at 28-32 base pairs (the length of RNA that a ribosome occupies and protects when bound). This is cleanly excised with a scalpel (panel b).

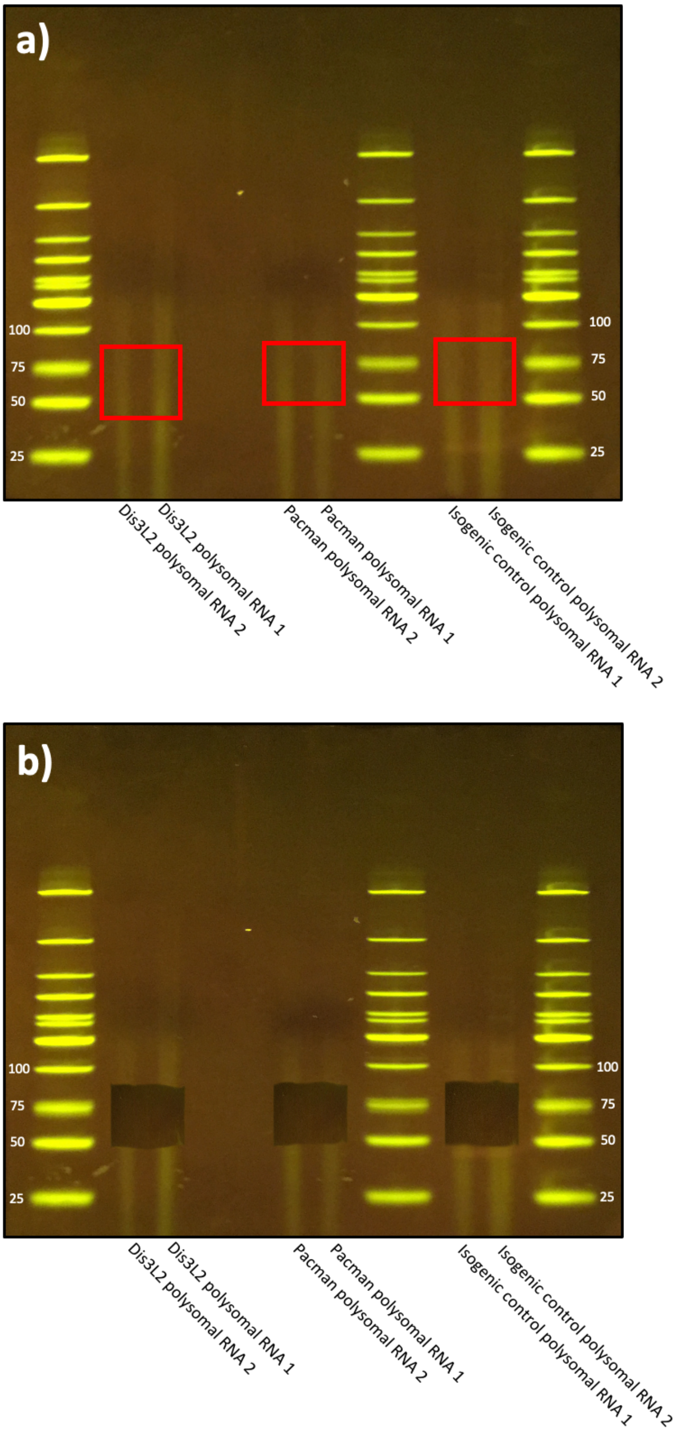


Figure 5.20 – Size selection shows expected distribution of chemically fragmented RNA fragments across all samples

The visualised gel shows the expected distribution of RNA molecule lengths across all samples. The chemically fragmented total RNA samples show a broad smear (panel a), as would be expected from a random fragmentation, centred around approximately 50-80 base pairs (as per the the aim of the protocol designed by the Couso and Aspden labs). This is cleanly excised with a scalpel (panel b).

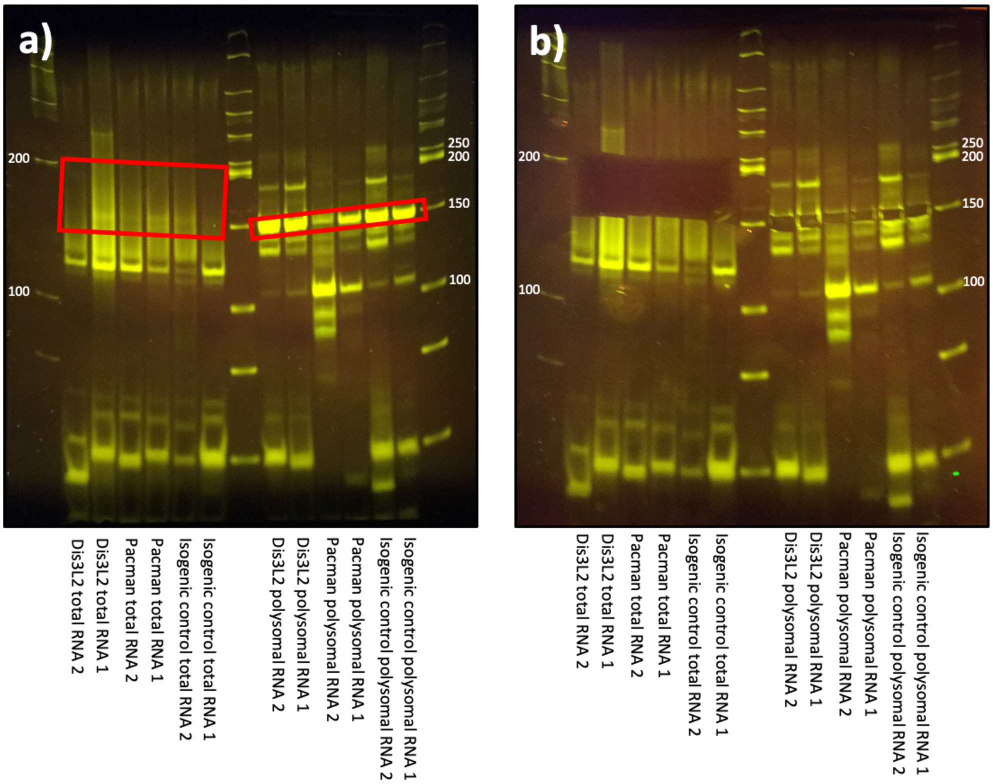


Figure 5.21 – Size selection shows expected distribution of adaptor-bound RNA fragments across all samples

The visualised gel shows the expected distribution of RNA molecule lengths across all samples (panel a). The ribosome footprinted samples feature a band at 155-161 base pairs (expected size of individual ribosome binding plus length of annealed adaptors), while the fragmented total RNA features a smear between 177-227 base pairs (expected size of fragmented RNA plus length of annealed adaptors). This is cleanly excised with a scalpel (panel b).

samples were run on an acrylamide gel (as described in Chapter 2 – Methods), the visualised gel (Figure 5.21) shows the expected distribution of RNA molecule lengths across all samples. The samples that had undergone ribosome footprinting feature a band at 155-161 base pairs (expected size of individual ribosome binding plus length of annealed adaptors), while the fragmented total RNA features a smear between 177-227 base pairs, now with a sharp cut off at the upper and lower limits of this size range, due to the previous size selection step. There is reduced fluorescence outside of these regions, as expected, although some adaptor-adaptor dimer is seen at ~120 base pairs.

5.8.1 Initial Bioanalyser traces show insufficient enrichment of desired peak in some instances

The Bioanalyser measurements of the samples should show sufficient enrichment of RNA library fragments of the expected size (155-161 and 177-227 for polysomal and total RNA, respectively), and severe depletion of RNA fragments outside of these ranges. This was the case in several of the samples, though with others, there are peaks showing the presence of RNA outside of the target range (as seen in Figure 5.22). A peak around 120 base pairs likely indicates that some residual adaptor-adaptor binding was not removed by the previous gel size-selection step. Future attempts can have adaptor-adaptor dimer prevalence reduced (at the cost of library output concentration) by reduced adaptor availability in the adaptor ligation reaction; but with the samples already prepared, these traces show the necessity of a second size selection step for adaptor-bound sample RNA.

5.8.2 Repeated size selection shows further enrichment of the desired RNA fragments

The adaptor-bound RNA samples found to contain substantial adaptor-adaptor dimer were run on an acrylamide gel (as described in Chapter 2 – Methods), now at a lower total RNA concentration compared to the earlier adaptor-RNA size selection step. Figure 5.23, compared to Figure 5.21, shows further separation between the adaptor-adaptor dimer and the desired adaptor-RNA fragments, with a reduction in fluorescence outside

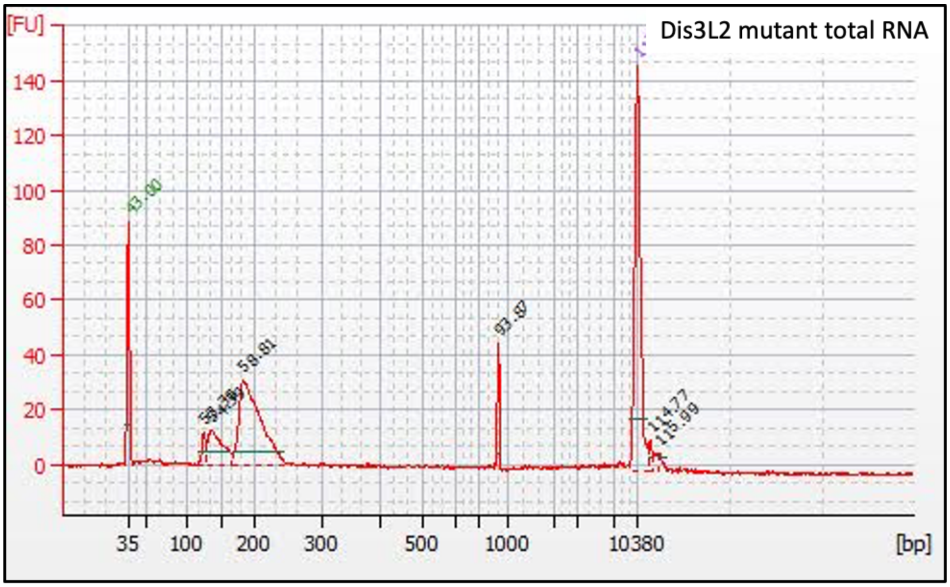


Figure 5.22 – Representative Bioanalyser traces from failed quality control of samples for poly-ribo-seq

In addition to the expected and desired RNA presence, indicated by either a slim peak at 160bp (polysomal RNA fragments with adaptors), or a broader peak at 200bp (total RNA fragments with adaptors), as appropriate; secondary peaks were seen (here at 110-130bp and 950bp. Concentrations were also too low in some instances. Markers, present as part of the kit, can be seen at 35bp and 10380bp.

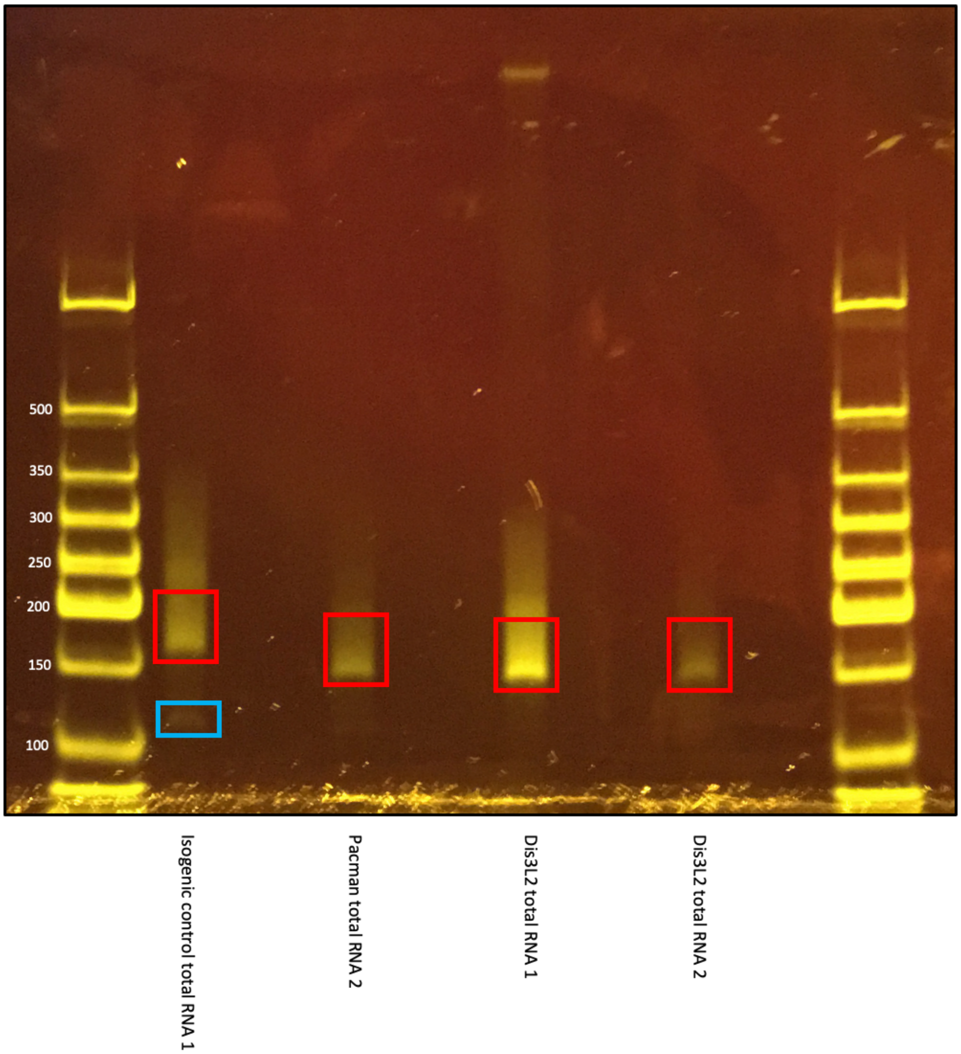


Figure 5.23 – Size selection shows enriched distribution of adaptor-bound RNA fragments across all samples

The visualised gel shows the expected distribution of RNA molecule lengths across all samples. All re-run samples were fragmented total RNA, featuring a smear between 177-227 base pairs (expected size of fragmented RNA plus length of annealed adaptors,) highlighted in red. Adaptor-adaptor dimer was detected in all re-run instances, particularly in “Isogenic control total RNA 1”, where it has been highlighted in blue.

of these size ranges. This allowed further size selection by careful incision and recovery of the gel regions desired (all further described in Chapter 2 - Methods.

5.8.3 Bioanalyser traces show sufficient enrichment and concentration of all desired peaks

These final Bioanalyser measurements of the samples (Figure 5.24) did indeed show sufficient enrichment of RNA fragments of the expected size (155-161 and 177-227 for polysomal and total RNA, respectively), and severe depletion of RNA fragments outside of these ranges, in all samples. Some samples such were at a significantly reduced concentration after all size-selection steps, but all were calculated to be of a concentration that can be sequenced by the NextSeq 500 platform to be used in the experiment. The samples, with their concentrations recorded (Figure 5.25, panel (a)) were pooled to run on two NextSeq lanes (Figure 5.25, panel (b)), and sent to Leeds University RNA-sequencing facility. These traces then, show 2 suitable replicates for Pacman mutant, Dis3L2 mutant, and isogenic control genotypes, with paired total RNA and polysomal RNA samples for each replicate.

5.9 Summary of the importance of optimisation of the protocol

With new and complex molecular techniques, (especially those with long, multi-step protocols,) it is crucial to have every step from start to finish carefully optimised and carried out, as with each inefficiency or failure at any step the final data produced can be significantly lowered in quality. With poly-ribo-seq (a technique that loses a lot of RNA through multiple selection steps), substantial work has been carried out to produce optimised protocols for use in *Drosophila* (90, 153). The required sample type in this project though (whole *Drosophila* L3 larvae), has not previously been used, and although the other *Drosophila* poly-ribo-seq optimisation was extremely useful in informing this work, these protocols could only provide a starting point. Without the substantial optimisation carried out in this chapter, no suitable protocol would exist for examining whole L3 larvae with poly-ribo-seq. Without the enormous amount of work that goes into optimisation, new techniques would not develop, and existing techniques would not find new and improved uses.

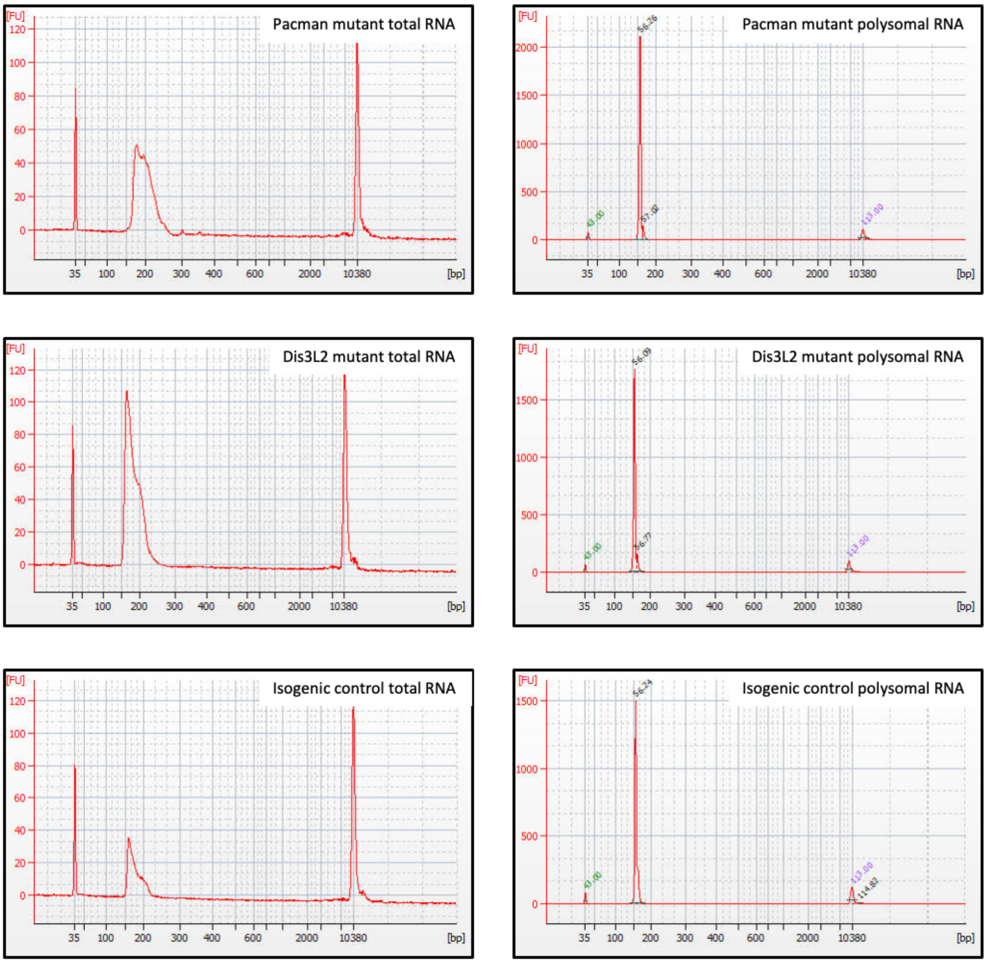
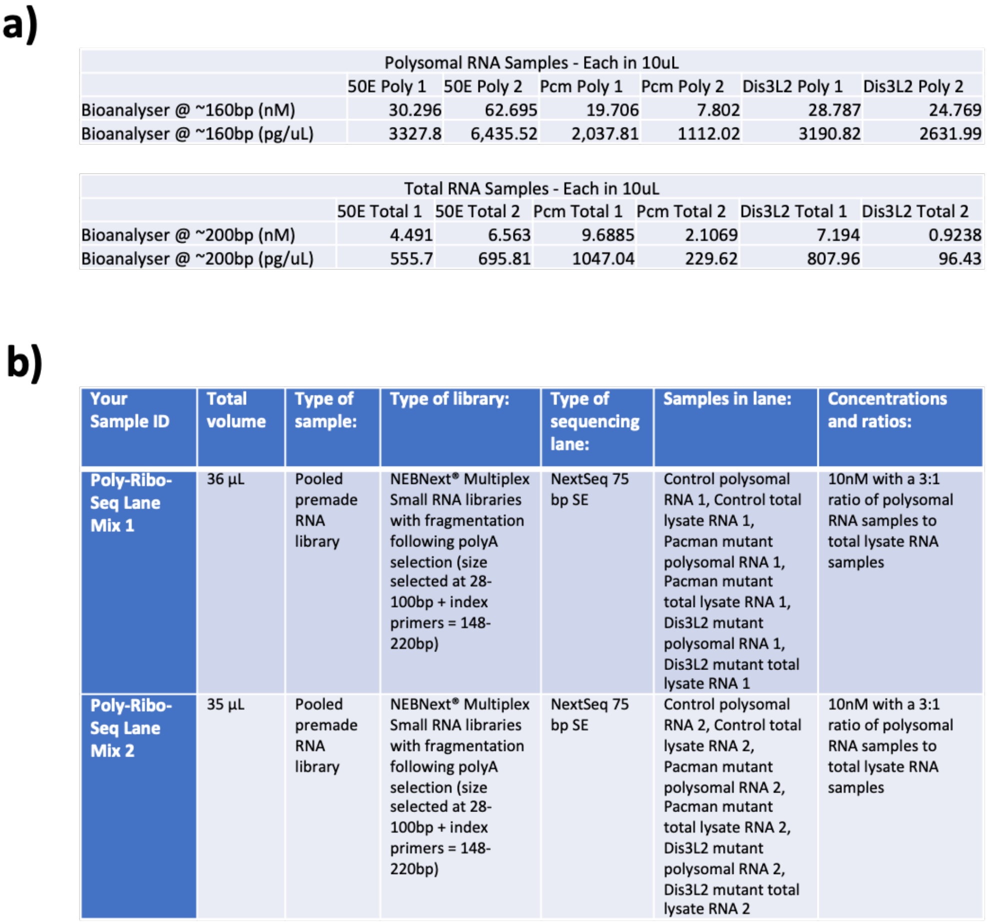


Figure 5.24 – Representative Bioanalyser traces from final quality control of samples for poly-ribo-seq

All samples can be seen to have either a slim peak at 160bp (polysomal RNA fragments with adaptors), or a broader peak at 200bp (total RNA fragments with adaptors), as appropriate. No other peaks were detected. Concentrations vary, but all were enough to send for sequencing. Markers, present as part of the kit, can be seen at 35bp and 10380bp.



Polysomal RNA Samples - Each in 10uL						
	50E Poly 1	50E Poly 2	Pcm Poly 1	Pcm Poly 2	Dis3L2 Poly 1	Dis3L2 Poly 2
Bioanalyser @ ~160bp (nM)	30.296	62.695	19.706	7.802	28.787	24.769
Bioanalyser @ ~160bp (pg/uL)	3327.8	6,435.52	2,037.81	1112.02	3190.82	2631.99

Total RNA Samples - Each in 10uL						
	50E Total 1	50E Total 2	Pcm Total 1	Pcm Total 2	Dis3L2 Total 1	Dis3L2 Total 2
Bioanalyser @ ~200bp (nM)	4.491	6.563	9.6885	2.1069	7.194	0.9238
Bioanalyser @ ~200bp (pg/uL)	555.7	695.81	1047.04	229.62	807.96	96.43

Your Sample ID	Total volume	Type of sample:	Type of library:	Type of sequencing lane:	Samples in lane:	Concentrations and ratios:
Poly-Ribo-Seq Lane Mix 1	36 µL	Pooled premade RNA library	NEBNext® Multiplex Small RNA libraries with fragmentation following polyA selection (size selected at 28-100bp + index primers = 148-220bp)	NextSeq 75 bp SE	Control polysomal RNA 1, Control total lysate RNA 1, Pacman mutant polysomal RNA 1, Pacman mutant total lysate RNA 1, Dis3L2 mutant polysomal RNA 1, Dis3L2 mutant total lysate RNA 1	10nM with a 3:1 ratio of polysomal RNA samples to total lysate RNA samples
Poly-Ribo-Seq Lane Mix 2	35 µL	Pooled premade RNA library	NEBNext® Multiplex Small RNA libraries with fragmentation following polyA selection (size selected at 28-100bp + index primers = 148-220bp)	NextSeq 75 bp SE	Control polysomal RNA 2, Control total lysate RNA 2, Pacman mutant polysomal RNA 2, Pacman mutant total lysate RNA 2, Dis3L2 mutant polysomal RNA 2, Dis3L2 mutant total lysate RNA 2	10nM with a 3:1 ratio of polysomal RNA samples to total lysate RNA samples

Figure 5.25 – Summary of final RNA-sequencing libraries

Panel (a) shows the concentrations of each sample determined by Bioanalyser trace, with Panel (b) showing the pooling used for each lane.

Chapter 6: Results – Multi-dataset analysis of translation and degradation of RNAs (including lncRNAs) in *Drosophila*

6.1 Introduction

The previous chapter has outlined not only the value and importance of poly-ribo-seq as a technique, but also walked through the steps required to optimise a protocol to the point where it can be meaningfully used to gather the transcriptional and translational profile from *Drosophila* L3 larvae. Following this then, it remains to extract as much purpose and useful information as reasonably possible from the gathered data.

The advantage of all RNA sequencing based techniques, including poly-ribo-seq, is the ability to extensively screen the entire transcriptome, allowing identification of differential expression between developmental conditions. Provided sufficient sequencing depth has been achieved, the (largely) unbiased nature of the technique (sequencing all transcripts, rather than an array of pre-selected targets, as would be the case for techniques such as microarrays, PCR arrays, and nanostring), RNA sequencing techniques are able to identify novel and non-canonical RNA transcripts. In combination with the translational context provided by the polysome fractionation in poly-ribo-seq, this presents an ideal technique for identifying not only poorly annotated RNA transcripts, and those with non-canonical behaviours (such as lncRNAs on the polysome), but instances where non-canonical peptides might be produced.

In order to increase confidence in the validity of the data (especially in cases of lowly transcribed and translated, novel or non-canonical ORFs) specific bioinformatic approaches must be applied in order to not only extract meaning from this novel data set, but to compare it with other relevant RNA-sequencing datasets (of which there are many). This has the additional benefit of facilitating comparison and speculation regarding the expression, and non-canonical translation, of certain genes in different cell lines, tissues, and developmental time points.

As a starting point, this meant identifying and downloading the most informative datasets regarding to the project, which will be summarised here. The paper “General and MicroRNA-Mediated mRNA Degradation Occurs on Ribosome Complexes in Drosophila Cells" by Antic *et al.* has been covered already in Chapter 3. To recap, the study investigated the possibility of ribosomal degradation of mRNAs by Pacman in *Drosophila* S2 cells. Amongst other experiments the authors isolated decapped mRNA degradation intermediates from ribosome complexes and performed high-throughput sequencing analysis on them, alongside RNA from total cell lysates. This was performed in control and Pacman knockdown S2 cells, depleted using RNA interference. They found that a large majority (93%) of these transcripts could be detected at the same relative abundance on ribosome complexes as in the polyA selected cell lysate (Figure 6.1); supporting that many mRNAs will often associate with the ribosome during 5' to 3' decay by Pacman. In summary, this dataset provides the transcriptional profile for S2 cells with and without Pacman depletion, in total and from ribosome association. Likewise, earlier mentioned were the RNA sequencing datasets already available in the Newbury lab. RNA sequencing had been carried out by Jones *et al*. on *Drosophila* L3 wing imaginal discs, both for null mutants for Pacman and for isogenic wild type control samples. Also available was the RNA sequencing data from Towler *et al*., which was carried out on *Drosophila* L3 wing imaginal discs, both for null mutants for Dis3L2 and for isogenic wild type control samples, with a preliminary analysis in Chapter 3. To summarise, these data provided a transcriptome wide snapshot of the RNA profile in L3 wing imaginal discs (notably the same developmental stage as that used in generating the novel dataset generated in this work) in wild type *Drosophila*, the absence of Pacman, and the absence of Dis3L2.

Another paper, "Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during Drosophila development" from Zhang *et al.* examined ribosome occupancy and translational efficiency (at the codon level) throughout the developmental stages of the *Drosophila* life cycle (154). The stages at which this was carried out were embryos at 0–2 hours, 2–6 hours, 6–12 hours, and 12–24 hours old; L3 instar larvae; stage P7–8 pupae; female heads; male heads; adult female bodies (heads removed); male bodies; and *Drosophila* S2 cells (only S2 cells were treated with harringtonine). The protocol they used for L3 larvae was found to not be suitable for reliably producing ribosome profiles of a high enough translational activity and resolution for poly-ribo-seq, which was not a problem for their collection of monosomes. In order to carry this out, the authors used mRNA sequencing alongside

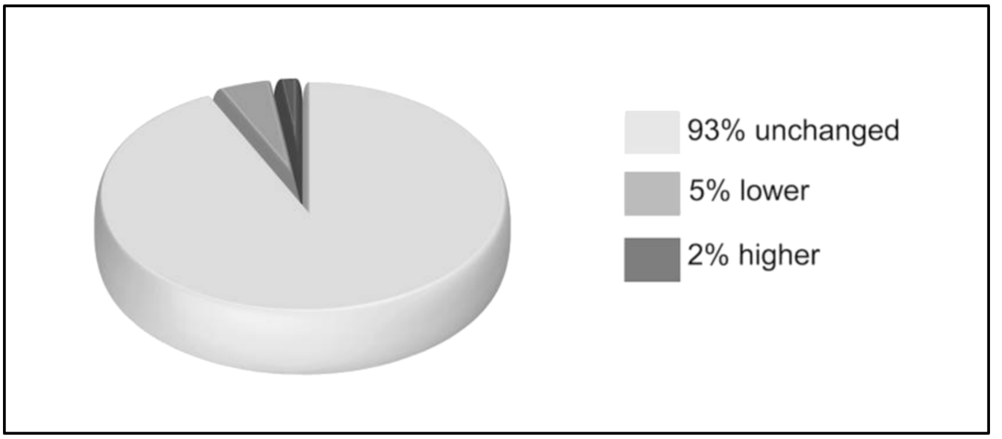


Figure 6.1 – Summary of the difference in ribosomally associated 5' to 3' decay intermediates compared to those found in total cell lysate

This pie chart (adapted from Antic et al., 2015) shows the differential abundance of the decapped decay intermediates found associated with the ribosome compared to those in the total cell lysates. The relative abundance was unchanged for most RNA transcripts (93%, pale grey), with a minority being lower on the ribosome than the cell lysate (5%, mid grey), and a smaller minority being higher on the ribosome than the cell lysate (2%, dark grey).

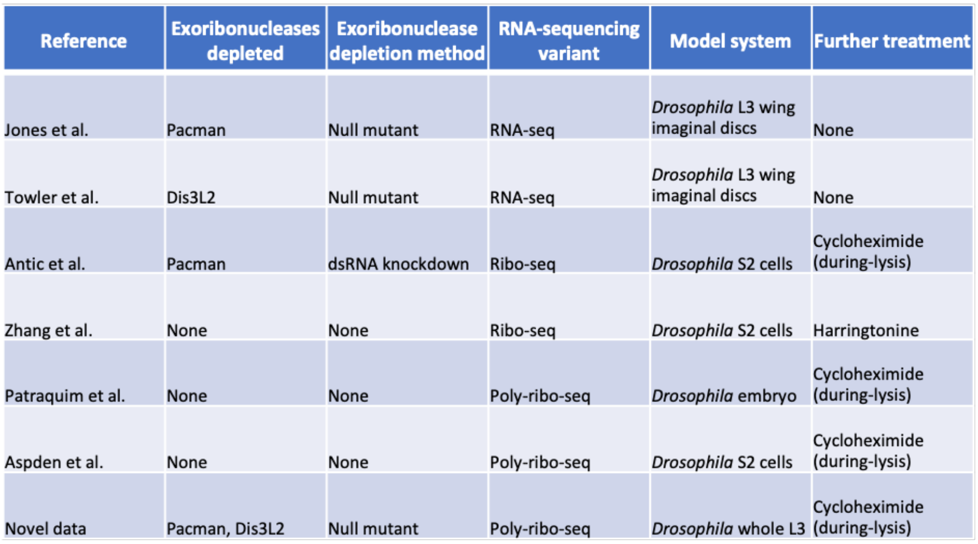
ribosomal profiling. Notably, and most relevant for this line of work, they used ribosome profiling on *Drosophila* S2 cells following harringtonine treatment in order to characterise genome-wide translation initiation events at upstream open reading frames (as harringtonine treatment is able to block the elongation stage of translation). In summary, this dataset provided a transcriptome wide look at RNA levels in *Drosophila* S2 cells, with the availability of the translational data from ribosome profiling, on cells where the initiation site could be clearly identified thanks to the harringtonine treatment.

Data already available from the Couso lab provided the last piece of the puzzle for this examination of the interplay between translation and degradation(90). Their work provided full poly-ribo-seq data from both *Drosophila* embryos and from *Drosophila* S2 cells. This meant that the novel dataset could be compared to data from similarly styled poly-ribo-seq experiments, which examined only polysome-associated RNA, rather than all ribosome bound RNA. It also provided a comparison to both a popularly used embryonic cell line (S2 cells) and an earlier point in *Drosophila* development (embryo).

By making select comparisons between these data (summarised in Figure 6.2), the novel data set can be used to a greater efficacy than would be possible if examined in isolation. Additionally, this chapter highlights not only the wealth of deep sequencing data now available, but how they can be compared and combined to more thoroughly answer any given question.

6.2 Project background and aims

The previous chapter's optimisation has allowed for sequencing of L3 RNA in the context of both total RNA profile and RNA associated with the polysome. Both of these conditions are examined in the presence and absence of Pacman (known to be involved in co-translational degradation in *Drosophila* and mammalian cells) and in the presence and absence of Dis3L2 (known to be associated with the ribosome in human cells) (69, 71). The work in this chapter aimed to gather together all useful, relevant data, and examine it from multiple angles.



Reference	Exoribonucleases depleted	Exoribonuclease depletion method	RNA-sequencing variant	Model system	Further treatment
Jones et al.	Pacman	Null mutant	RNA-seq	*Drosophila* L3 wing imaginal discs	None
Towler et al.	Dis3L2	Null mutant	RNA-seq	*Drosophila* L3 wing imaginal discs	None
Antic et al.	Pacman	dsRNA knockdown	Ribo-seq	*Drosophila* S2 cells	Cycloheximide (during-lysis)
Zhang et al.	None	None	Ribo-seq	*Drosophila* S2 cells	Harringtonine
Patraquim et al.	None	None	Poly-ribo-seq	*Drosophila* embryo	Cycloheximide (during-lysis)
Aspden et al.	None	None	Poly-ribo-seq	*Drosophila* S2 cells	Cycloheximide (during-lysis)
Novel data	Pacman, Dis3L2	Null mutant	Poly-ribo-seq	*Drosophila* whole L3	Cycloheximide (during-lysis)

Figure 6.2 – Table summarising datasets for comparison in this chapter

Basic details of all datasets used for comparison and analysis in this chapter.

The main stages to this work are as follows:

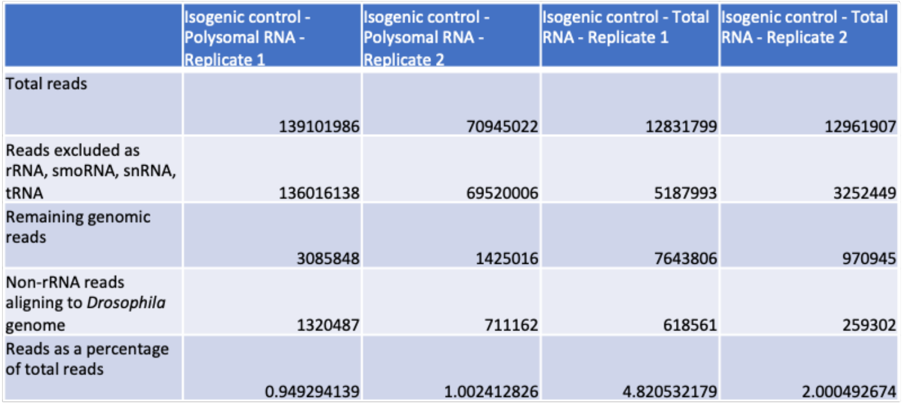
1) Carry out the initial, necessary steps to process the data, and then evaluate the depth and distribution of available reads.
2) Examine the novel data alongside data from the Couso lab and from Zhang *et al*., identify candidate lncRNAs that may be polysome associated and potentially undergoing translation.
3) Examine the novel data alongside data from the Newbury lab and from Antic *et al.*, identify lncRNA targets of Pacman degradation, and which of those may be polysome associated and potentially undergoing translation.
4) Examine the novel data alongside data from the Newbury lab, identify lncRNA targets of Dis3L2 degradation, and which of those may be polysome associated and potentially undergoing translation.
5) Create a shortlist of the most promising candidate lncRNAs regulated by Pacman or Dis3L2, also found on the polysome, and use molecular techniques to assess their presence on the polysome and potential translational activity.
6) Examine the novel data s*et al*ongside data from the Newbury lab, identify target lncRNAs that are consistently differentially abundant in the absence of both Pacman and Dis3L2, and use the data from Zhang *et al*. and the Couso lab to see whether any of these may be polysome associated and potentially undergoing translation.

5.3.1 Tuxedo suite processing of poly-ribo-seq data shows total number of *Drosophila* genome aligned reads

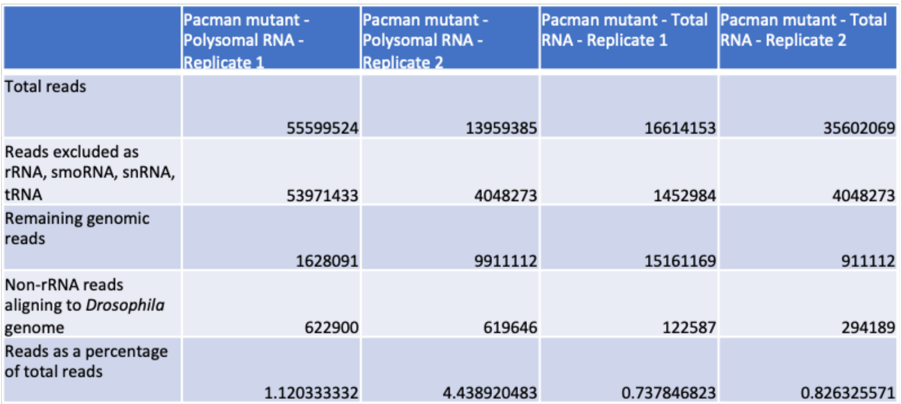
This dataset provides the first genome wide examination of RNA levels with the additional context of polysomal association in *Drosophila* L3 larvae. In order to make use of this, the answers to specific questions needed to be retrieved from huge, high-throughput data output. Upon receiving the raw reads as unaligned .fastq files from the sequencing facility, the following pipeline was used in order to extract valuable information and candidate genes from the datasets. An initial quality assessment was carried out using FastQC, a program that assesses overall sequence quality and identifies over-represented sequences within a .fastq file, along with the frequency of its occurrence. This allows the identification of the adaptor sequences for each sample and provides a measure of each of their abundance. The adaptors were analysed (and found to be as expected) for each sample were subsequently removed using CutAdapt and reads shorter than 15 base pairs were discarded. This was achieved using Sickle. Sickle is a "sliding window" algorithm, able to remove low quality bases.

Following the above quality assessment and trimming steps, Bowtie was used to align the trimmed reads, using a custom .fasta reference index file containing all known tRNA, rRNA, snoRNA, and snRNA sequences in the *Drosophila melanogaster* genome. All sequences that mapped to this index file were discarded, and the remaining reads carried forward for further analysis. This was carried out as the length and abundance of these RNAs causes selection for them throughout the ribosome footprinting procedure. Although there were steps taken to deplete these RNA species, it is almost impossible to completely remove them, and most eukaryotic sequencing data still shows their presence. A current .fasta version of the *Drosophila melanogaster* genome was built using Flybase release (6.29) with the HiSat2 algorithm. All 4 *Drosophila* chromosomes were used. This index file was used to align the reads, using the HiSat2 algorithm.

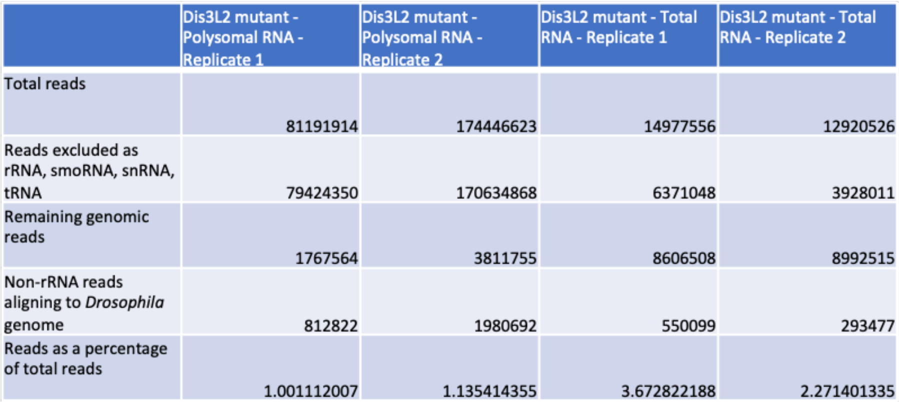
This step produced unexpected, and disappointing, results. The quantity of reads mapping to tRNA, rRNA, snoRNA, and snRNA was significantly higher than expected, causing a disproportionate number of reads to be discarded (Figure 6.3). Following this, a significant proportion of the remaining reads did not map to the *Drosophila* genome, further reducing the number of useful reads. Together, these issues reduced the usable data dramatically, essentially reducing the resolution of the snapshot of the RNA profile (Figure 6.3). Although the data can, and was, taken forward to still unearth interesting and novel findings, it must be acknowledged that this reduced sequencing depth provides limitations on what can be detected and statistically verified. From further investigation, it seems that the cause of the high level of rRNA left was due to the rRNA depletion step (discussed in Chapter 5) not adequately removing these transcripts. The non-*Drosophila* RNA (primarily in the total lysate RNA samples) was identified as mapping to several strains of bacteria (such as Mycobacterium abcessus, a mycobecteria species common in soil and water contaminents), likely present due to the intact larval digestive system in the samples. Given that the polyA bead selection



	Isogenic control - Polysomal RNA - Replicate 1	Isogenic control - Polysomal RNA - Replicate 2	Isogenic control - Total RNA - Replicate 1	Isogenic control - Total RNA - Replicate 2
Total reads	139101986	70945022	12831799	12961907
Reads excluded as rRNA, smoRNA, snRNA, tRNA	136016138	69520006	5187993	3252449
Remaining genomic reads	3085848	1425016	7643806	970945
Non-rRNA reads aligning to *Drosophila* genome	1320487	711162	618561	259302
Reads as a percentage of total reads	0.949294139	1.002412826	4.820532179	2.000492674



	Pacman mutant - Polysomal RNA - Replicate 1	Pacman mutant - Polysomal RNA - Replicate 2	Pacman mutant - Total RNA - Replicate 1	Pacman mutant - Total RNA - Replicate 2
Total reads	55599524	13959385	16614153	35602069
Reads excluded as rRNA, smoRNA, snRNA, tRNA	53971433	4048273	1452984	4048273
Remaining genomic reads	1628091	9911112	15161169	911112
Non-rRNA reads aligning to *Drosophila* genome	622900	619646	122587	294189
Reads as a percentage of total reads	1.120333332	4.438920483	0.737846823	0.826325571



	Dis3L2 mutant - Polysomal RNA - Replicate 1	Dis3L2 mutant - Polysomal RNA - Replicate 2	Dis3L2 mutant - Total RNA - Replicate 1	Dis3L2 mutant - Total RNA - Replicate 2
Total reads	81191914	174446623	14977556	12920526
Reads excluded as rRNA, smoRNA, snRNA, tRNA	79424350	170634868	6371048	3928011
Remaining genomic reads	1767564	3811755	8606508	8992515
Non-rRNA reads aligning to *Drosophila* genome	812822	1980692	550099	293477
Reads as a percentage of total reads	1.001112007	1.135414355	3.672822188	2.271401335

Figure 6.3 – Tables showing read counts before and after index file alignments

The majority of reads can be seen to be lost to non useful (i.e. rRNA, smoRNA, snRNA, and tRNA) RNA species, as well as non *Drosophila* RNA.

steps performed on the total RNA samples should have depleted both bacterial RNA and rRNA, but high quantities of both are seen, it is likely that the influence of *Drosophila* gut flora might be responsible for over-saturating the beads during the depletion steps, resulting in neither bacterial RNA or rRNA being sufficiently removed. A summary of the workflow for processing the data is visually represented in Figure 6.4

By inputting a reference genome and the aligned reads, the CuffLinks program was able to assemble the remaining transcripts using the .gtf annotation file for the same Flybase genome release (6.29) and outputs a single .gtf file containing the reads assembled into transcripts. From the assembled transcripts for each sample, the CuffMerge program was used to create a reference assembly containing all processed sample data. In order to quantify abundance of hits for each sequence, the CuffQuant program was used. CuffQuant is able to take the merged .GTF file from the CuffMerge output, along with a .fasta file reference genome and the accepted processed hits from the earlier HiSat2 step, and produce a .cxb file containing the abundances of each transcript.

Following this, CuffDiff was used to process the information from the CuffQuant step, the .fasta reference *Drosophila* genome, the merged .gtf file, and a text input file of which conditions or datasets to compare. The CuffDiff algorithm was able to process these data, and output a .diff file of normalised read counts and differential expression analysis, that can be used in most spreadsheet or workbook capable software (such as Microsoft Excel), to be read. The TuxedoSuite package CummeRbund was used to extract replicate values for use and analysis. The R package ggPlot2 was used alongside this in order to visualise data in plots and graphs. In a parallel pipeline, in order to more thoroughly identify statistical significances, FeatureCounts was used, after HiSat2 alignment, to quantify, before using the EdgeR package for R in order to quantify read counts separately, enforcing more stringent statistical and replicate value filters (replicate values plotted in Figure 6.5)., and normalising read count (allowing an additional level of stringent filtering when desired). It is worth noting that CuffLinks includes a transcript length normalisation, whilst EdgeR provides normalisation by the expected distribution of differentially and non-differentially expressed genes.

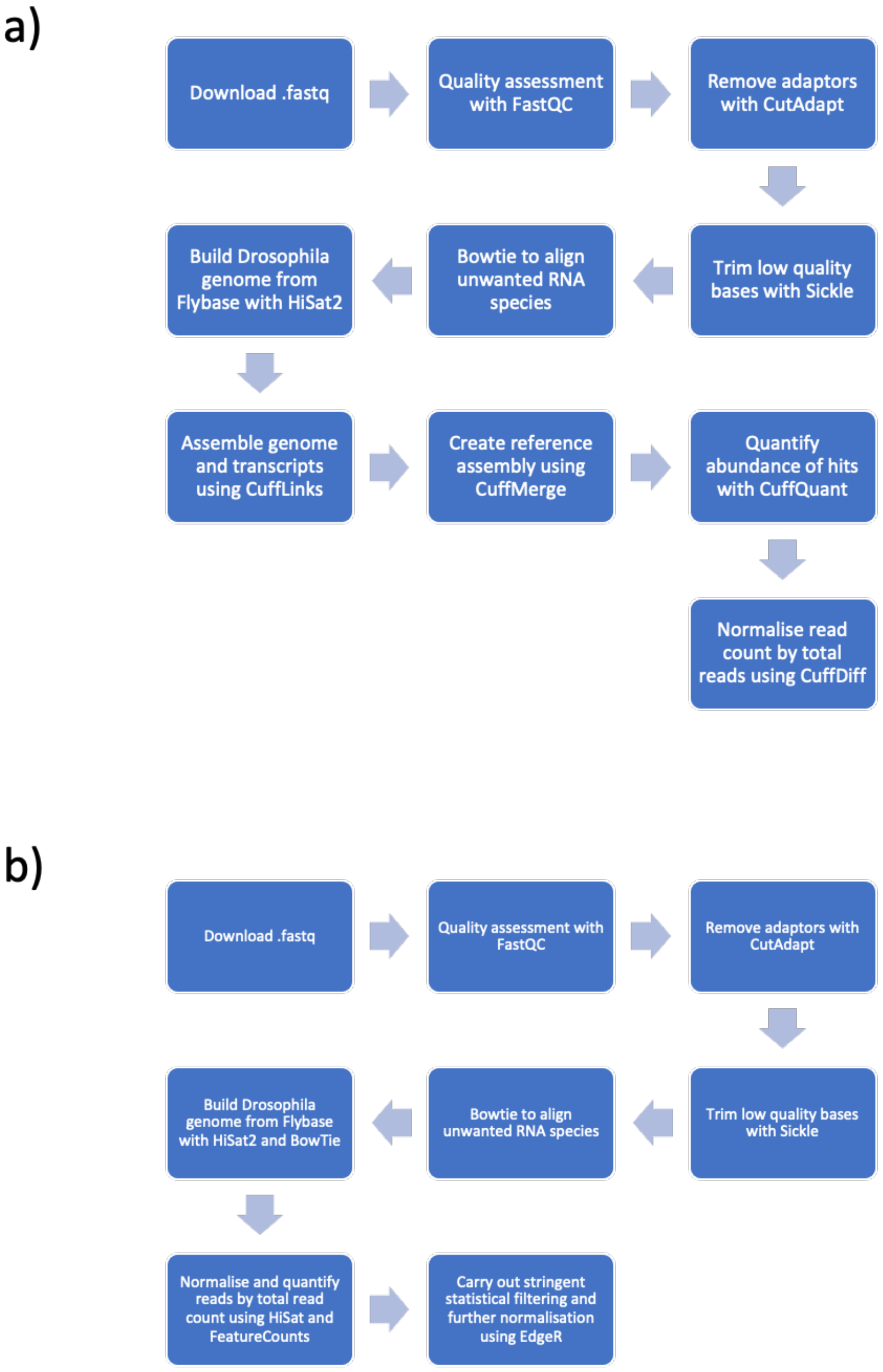

a)
Download .fastq
Quality assessment with FastQC
Remove adaptors with CutAdapt
Trim low quality bases with Sickle
Bowtie to align unwanted RNA species
Build Drosophila genome from Flybase with HiSat2
Assemble genome and transcripts using CuffLinks
Create reference assembly using CuffMerge
Quantify abundance of hits with CuffQuant
Normalise read count by total reads using CuffDiff
b)
Download .fastq
Quality assessment with FastQC
Remove adaptors with CutAdapt
Trim low quality bases with Sickle
Bowtie to align unwanted RNA species
Build Drosophila genome from Flybase with HiSat2 and BowTie
Normalise and quantify reads by total read count using HiSat and FeatureCounts
Carry out stringent statistical filtering and further normalisation using EdgeR

Figure 6.4 – A summary of the data processing workflow

Panel (a) shows the data analysis pipeline without further stringency filtering by EdgeR, while Panel (b) shows the pipeline used in order to implement these.

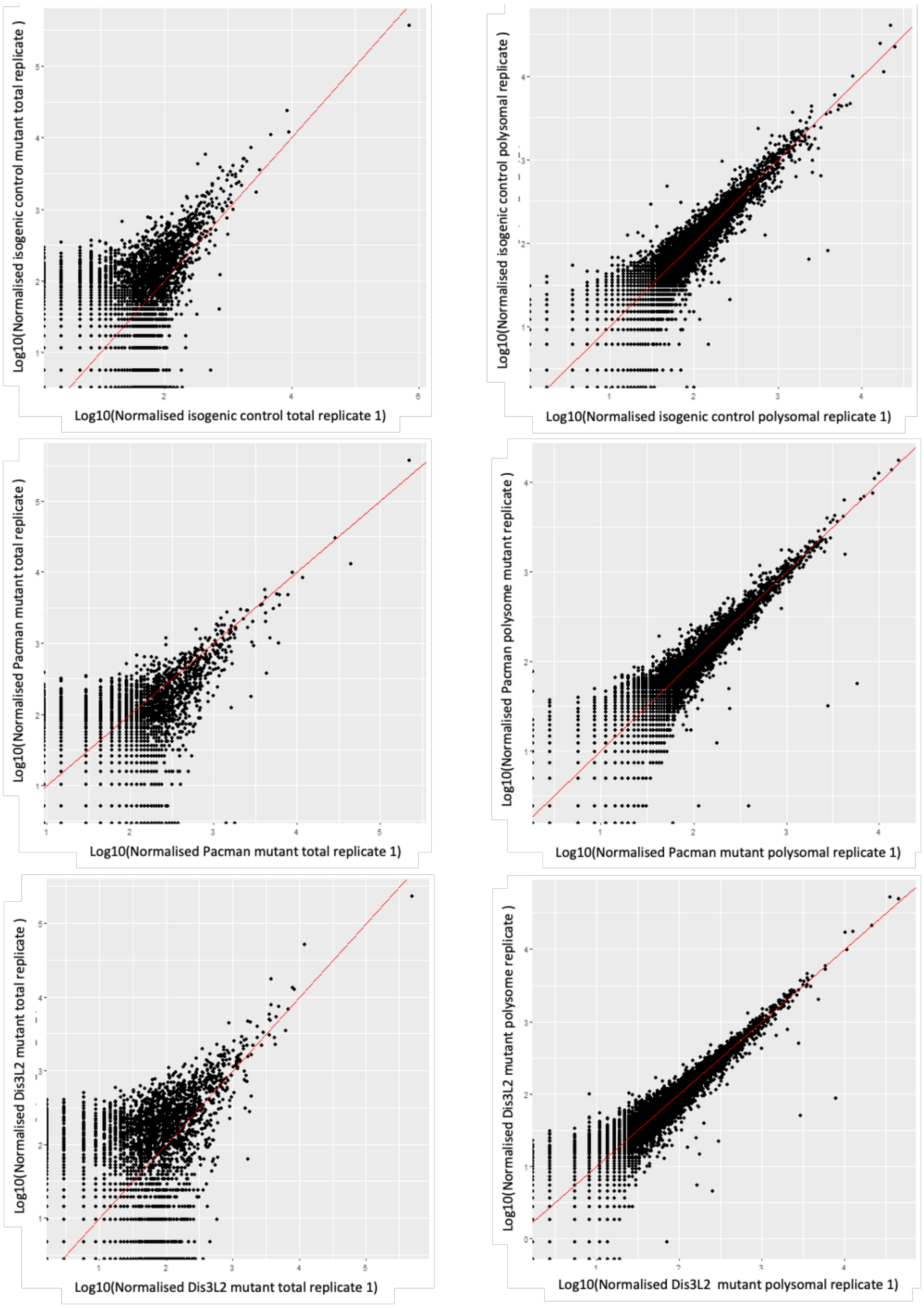


Figure 6.5 – Plots showing correlation of sequencing replicate values

Normalised read counts for each sample was plotted against the normalised read count for the second replicate of each sample. Polysomal RNA refers to RNA prepared from polysome fractions with two or more ribosomes attached per transcript. Total RNA refers to RNA prepared from the total lysate of the L3 sample before fractionation.

6.3.2 Overall dataset comparisons of differential abundance between polysomal and total lysate in exoribonuclease mutant genotypes

Due to the low read depth of the dataset, the strongest use of the data is the use of multiple comparisons to validate trends and defined candidates, to compare and contrast differences, and to provide an informative overview, rather than relying on the depth from a single dataset to provide comprehensive analysis. Work in a previous chapter (Chapter 3) has already covered a similar (although less comprehensive) independent overview of the other datasets used in this chapter. With the aim of generating the greatest value from the novel dataset, the majority of this chapter (6.4 onwards) will focus on the composite use of datasets in order to pursue any course of inquiry. This initial section will, however, provide an initial analysis of the data, independently of other datasets.

6.3.2.1 Observing proportions of RNA populations differentially abundant in the presence and absence of each exoribonuclease in total L3 lysate

Previous work by Jones *et al*. and Towler *et al*. have already explored the global role of Pacman and Dis3L2 in degradation of RNA, using *Drosophila* larval wing imaginal discs. Although not the aim of this project, a short evaluation of the role of Pacman and Dis3L2 in the global degradation of RNA in whole L3 larvae is a simple and incidental analysis to carry out while sorting data for more specific and informative comparisons. The novel data (in addition to the information on polysome-associated RNAs specific to carrying out poly-ribo-seq) contains two replicates of total lysate RNA (comparable to straightforward RNA-seq) in Pacman null mutant, Dis3L2 null mutant, and isogenic control genotypes. The two replicates for each genotype (Pacman mutant, Dis3L2 mutant, isogenic control,) in each RNA pool (total lysate RNA, polysomal RNA,) were averaged (total lysate RNA replicate correlation plotted in Figure 6.6, polysomal RNA replicate correlations plotted in Figure 6.7) for simple comparison to each other.

From these average CPM values, each mutant genotype was plotted against the isogenic control from the RNA extracted from total L3 lysate, revealing the detected populations of all up- and down-regulated RNAs for each exoribonuclease. Given the

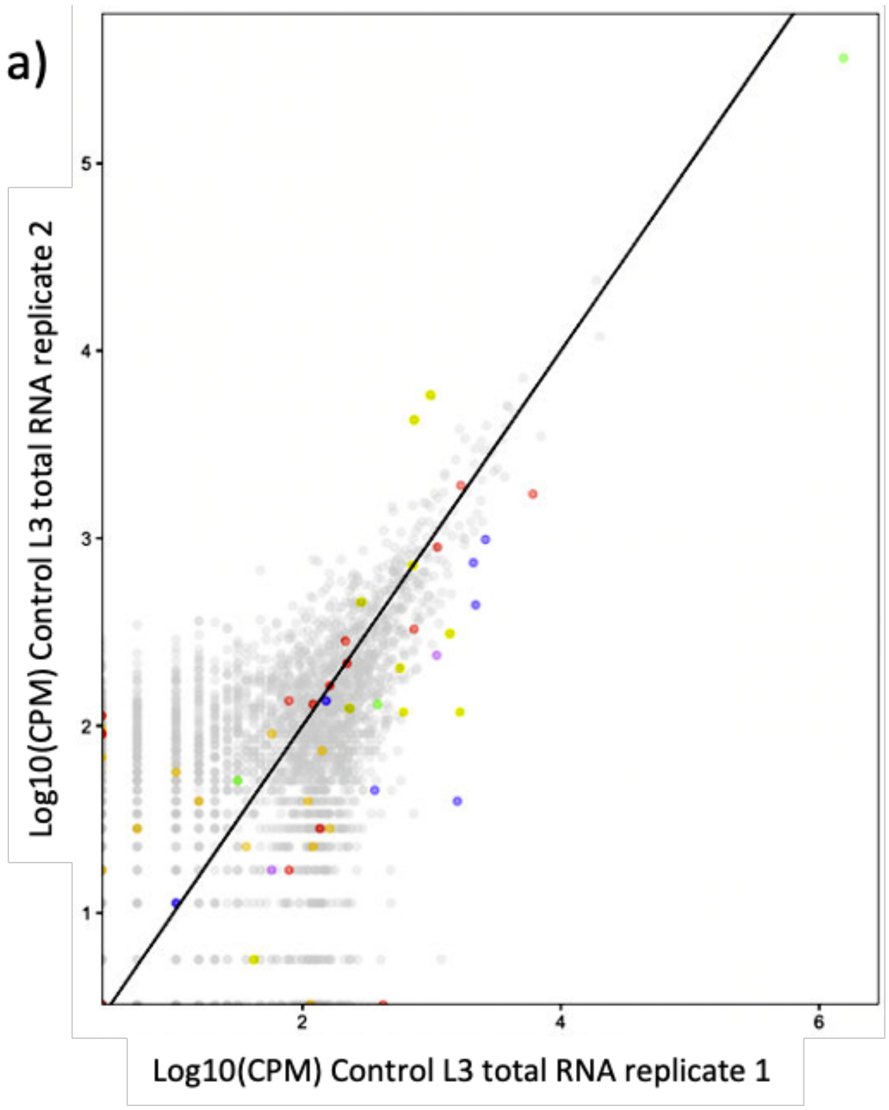


Figure 6.6 (a) – Correlation between CPM of RNA transcripts in total lysates from L3 larvae replicates for each genotype

A plot of a log10 of the CPM of RNAs between replicates of each genotype of total RNA showing strong correlation (isogneic control (a), Pacman mutant (b), Dis3L2 mutant (c)). Points that align perfectly with the x=y line would have perfect correlation between replicates, while deviation from the line shows increased variability between replicates.

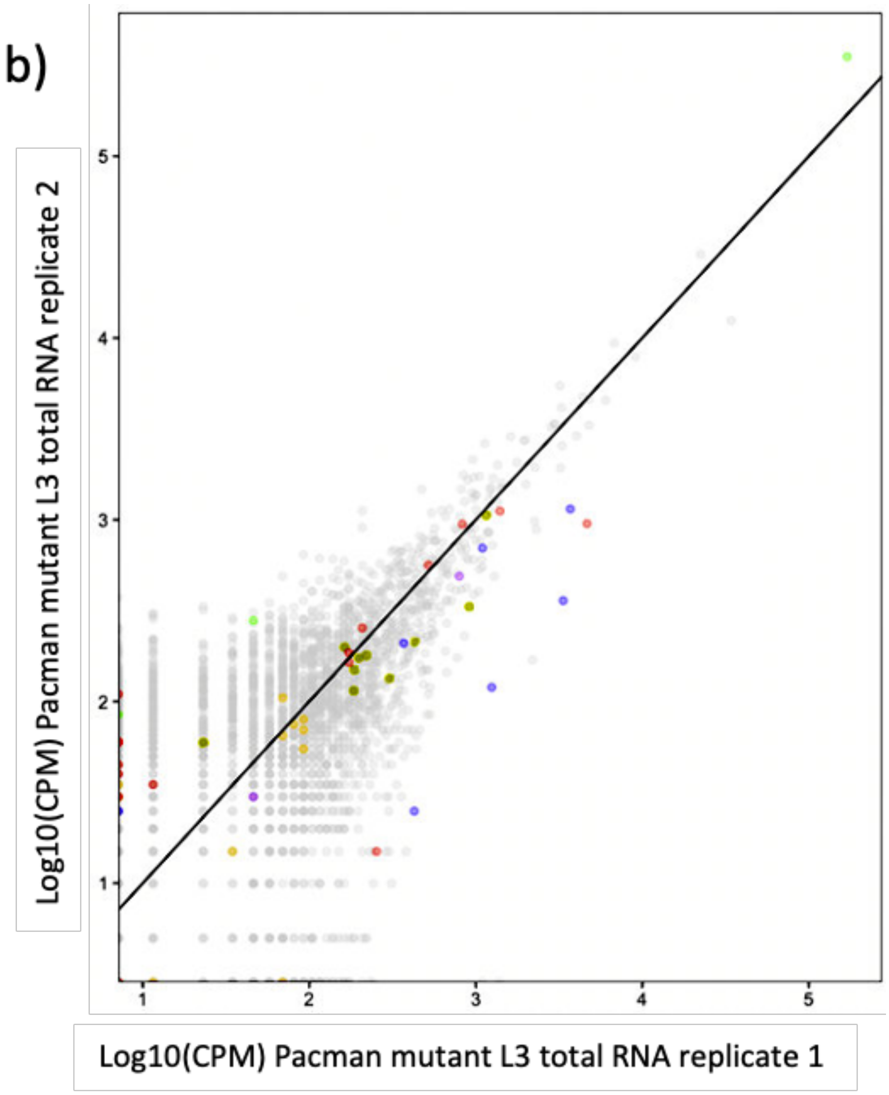


Figure 6.6 (b) – Correlation between CPM of RNA transcripts in total lysates from L3 larvae replicates for each genotype

A plot of a log10 of the CPM of RNAs between replicates of each genotype of total RNA showing strong correlation (isogneic control (a), Pacman mutant (b), Dis3L2 mutant (c)). Points that align perfectly with the x=y line would have perfect correlation between replicates, while deviation from the line shows increased variability between replicates.

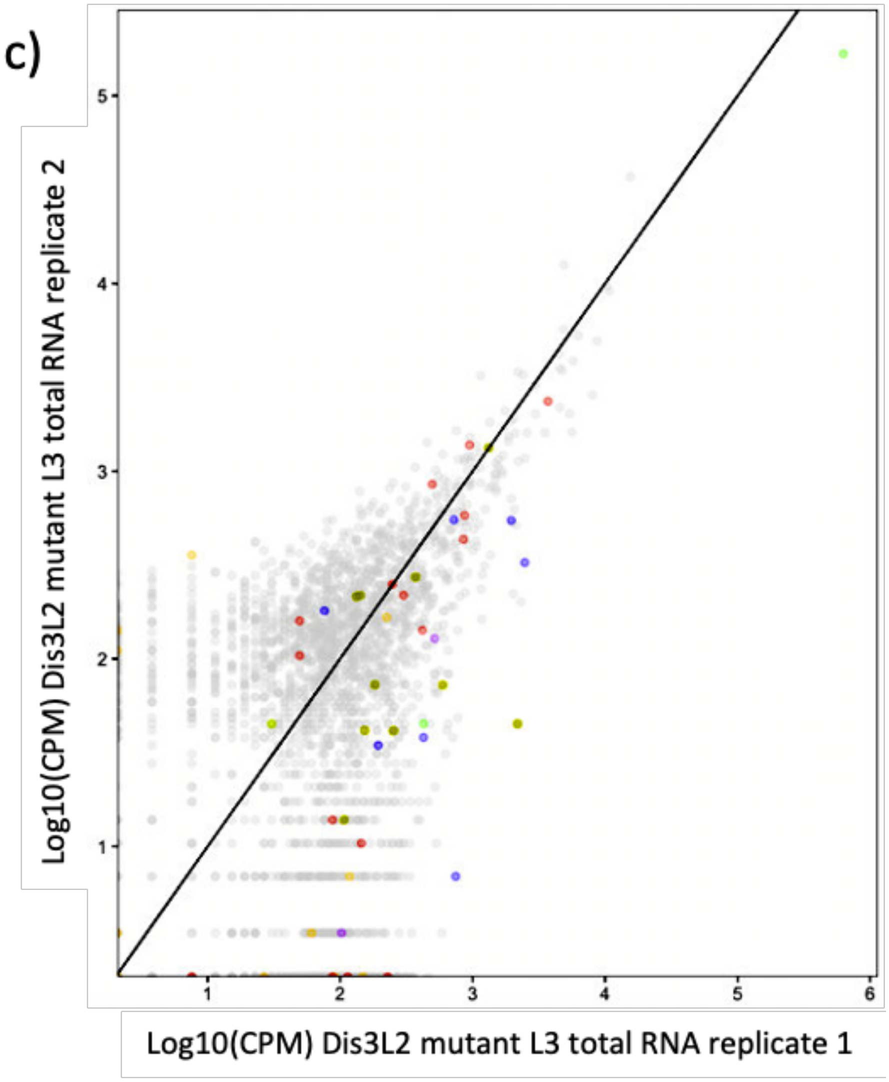


Figure 6.6 (c) – Correlation between CPM of RNA transcripts in total lysates from L3 larvae replicates for each genotype

A plot of a log10 of the CPM of RNAs between replicates of each genotype of total RNA showing strong correlation (isogneic control (a), Pacman mutant (b), Dis3L2 mutant (c)). Points that align perfectly with the x=y line would have perfect correlation between replicates, while deviation from the line shows increased variability between replicates.

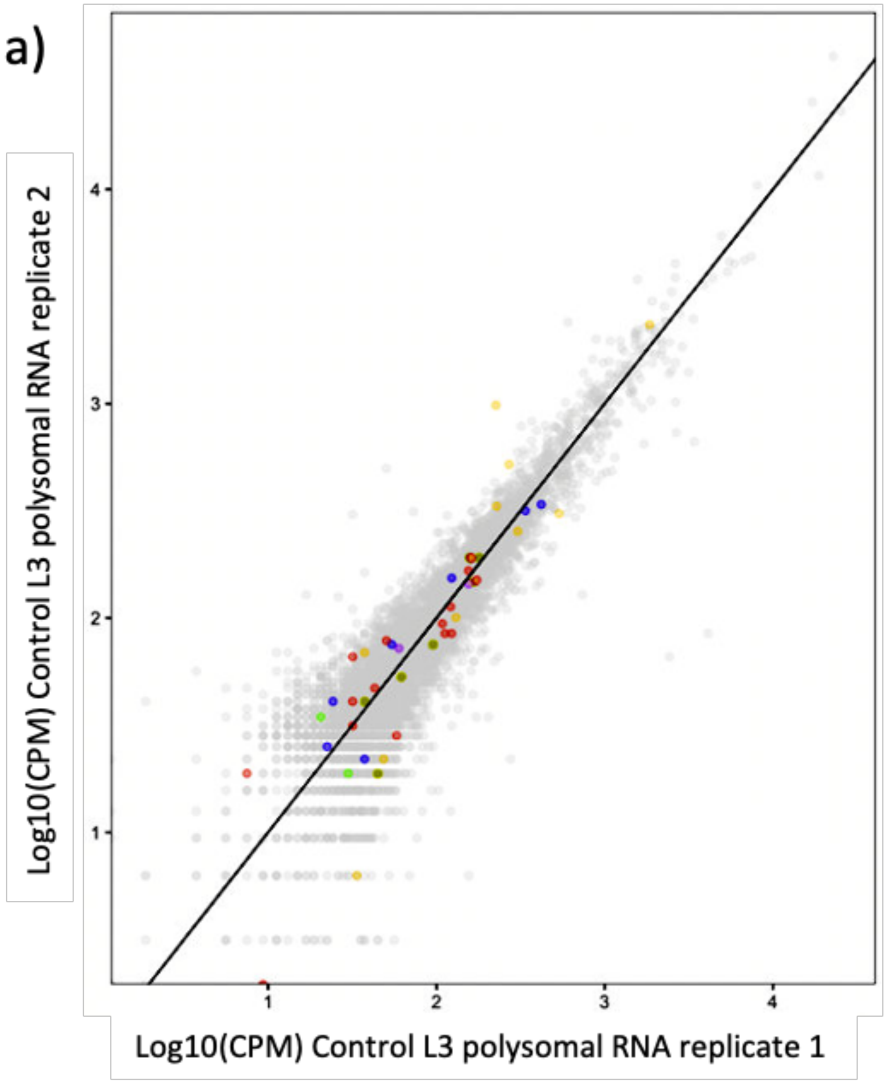


Figure 6.7 (a) – Correlation between CPM of RNA transcripts in polysomal fractions from L3 larvae replicates for each genotype

A plot of a log10 of the CPM of RNAs between replicates of each genotype of polysomal RNA showing strong correlation (isogneic control (a), Pacman mutant (b), Dis3L2 mutant (c)). Points that align perfectly with the x=y line would have perfect correlation between replicates, while deviation from the line shows increased variability between replicates. Cut-offs on the axis are imposed at the lowest value for both x-axis and y-axis.

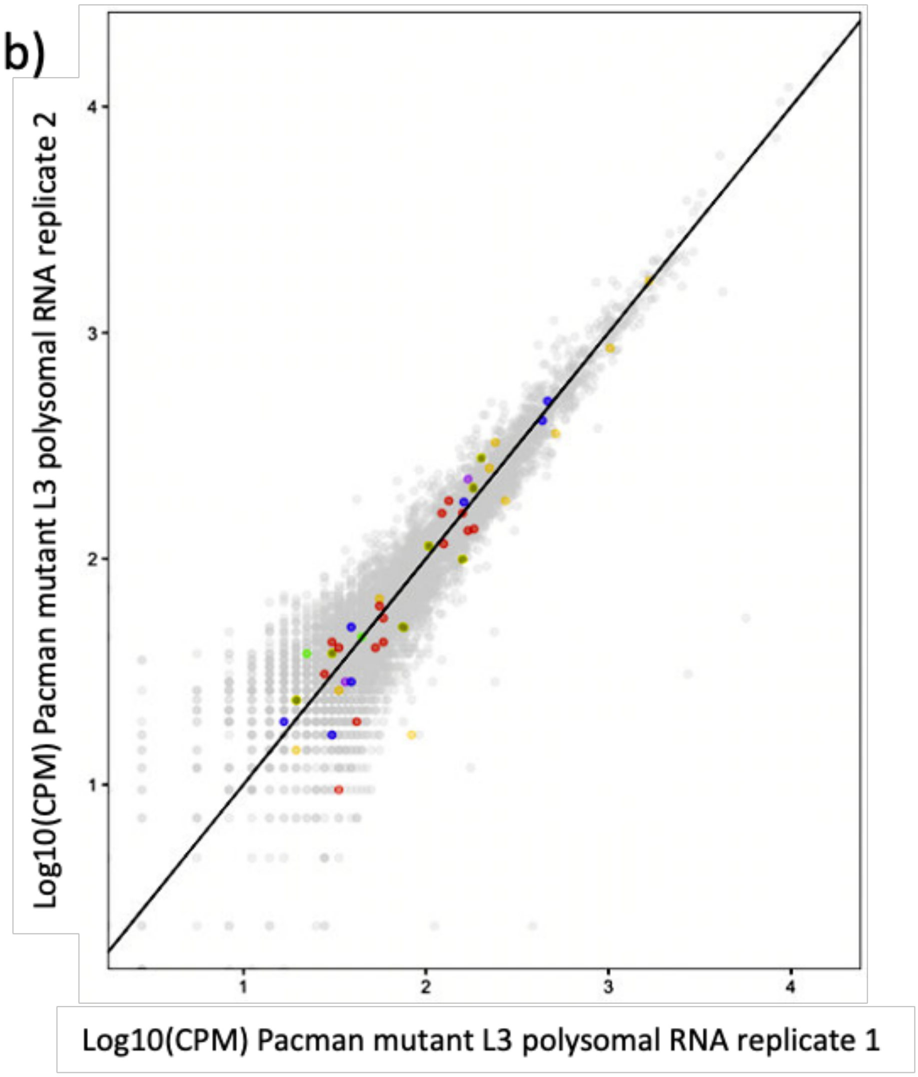


Figure 6.7 (b) – Correlation between CPM of RNA transcripts in polysomal fractions from L3 larvae replicates for each genotype

A plot of a log10 of the CPM of RNAs between replicates of each genotype of polysomal RNA showing strong correlation (isogneic control (a), Pacman mutant (b), Dis3L2 mutant (c)). Points that align perfectly with the x=y line would have perfect correlation between replicates, while deviation from the line shows increased variability between replicates. Cut-offs on the axis are imposed at the lowest value for both x-axis and y-axis.

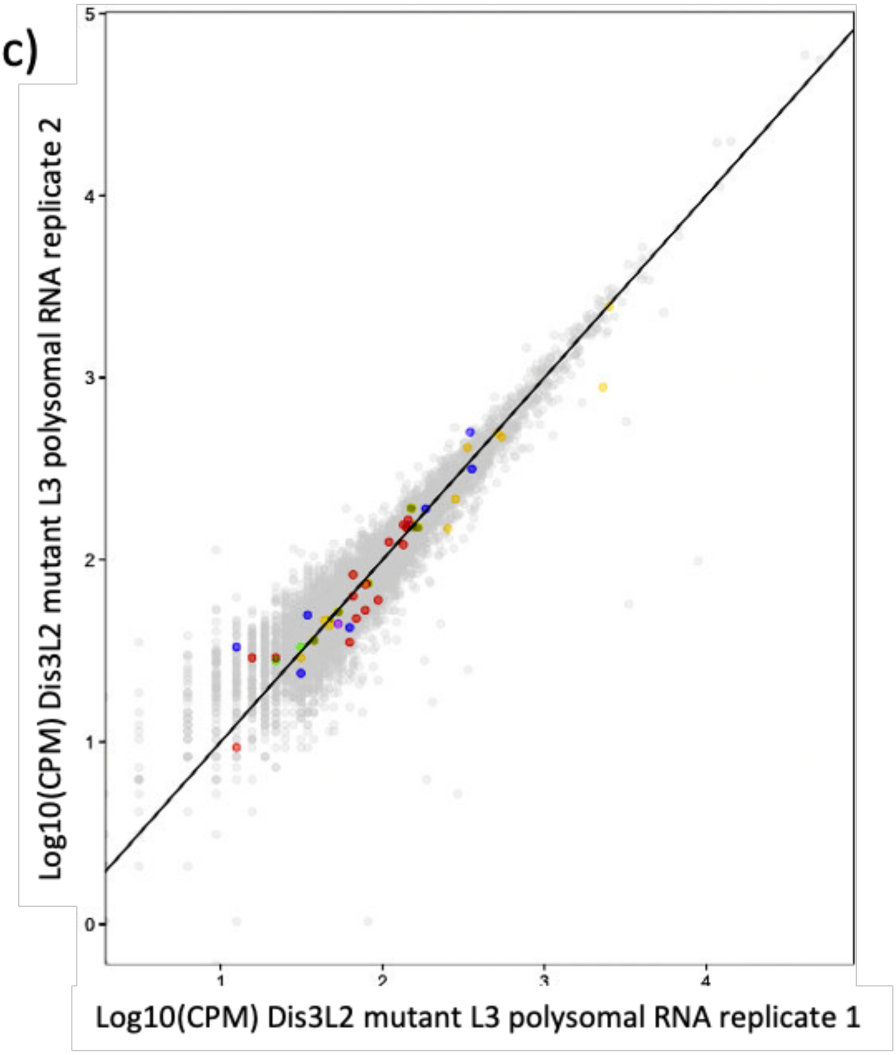


Figure 6.7 (c) – Correlation between CPM of RNA transcripts in polysomal fractions from L3 larvae replicates for each genotype

A plot of a log10 of the CPM of RNAs between replicates of each genotype of polysomal RNA showing strong correlation (isogneic control (a), Pacman mutant (b), Dis3L2 mutant (c)). Points that align perfectly with the x=y line would have perfect correlation between replicates, while deviation from the line shows increased variability between replicates. Cut-offs on the axis are imposed at the lowest value for both x-axis and y-axis.

lack of other datasets in this analysis to improve confidence, EdgeR statistical stringency filters were used to allow meaningful comparison even at this low read depth. With the filters, there were 2999 genes that passed all tests in all total RNA replicates and could be meaningfully analysed in this context by direct comparison. Of these 2999, 610 RNAs are upregulated more than 2-fold in the absence of Pacman, and another 615 downregulated more than 2-fold (Figure 6.8, panel (a)). When comparing to Dis3L2, 638 RNAs were upregulated more than 2-fold, and 767 downregulated more than 2-fold (Figure 6.9, panel (a)). This is a substantial proportion of genes that appear to be changing in expression levels; although some of these changes are in some instances of relatively low confidence (partially due to having only 2 replicates available per genotype), such a high number really emphasises the importance of Pacman and Dis3L2 in the regulation of many RNAs in whole *Drosophila* models. Of these genes analysed, the vast majority are protein coding, as would be expected (2948), with 16 non-coding RNAs (13 of which are long non-coding RNAs), and a remaining minority of other RNA types (13 tRNAs, 10 pseudogenes, 7 snoRNAs, 3 rRNAs, and 2 srRNAs). These are plotted on panel (b) and (c) of Figures 6.8 and 6.9). Of interest, even in this limited form of analysis, one lncRNA is more than 2-fold upregulated in the Pacman mutant (CR43334), with three lncRNA more than 2-fold upregulated in the Dis3L2 mutant larvae (*CR43334*, *cherub*, and *CR40469*).

6.3.2.2 Observing proportions of RNA populations differentially abundant in the presence and absence of each exoribonuclease in polysomal L3 extract

Following this, the read count of each mutant genotype was plotted against the isogenic control from the polysomal RNA extracted from L3 lysate, revealing the detected populations of all up- and down-regulated RNAs for each exoribonuclease. Given the lack of other datasets in this analysis to improve confidence, EdgeR statistical stringency filters were used to allow meaningful comparison even at this low read depth. With the filters, there were 5519 genes that passed all tests in all replicates and could be meaningfully analysed in this context by direct comparison. Of these 5519, 335 RNAs are upregulated more than 2-fold in the absence of Pacman, and another 362 downregulated more than 2-fold (Figure 6.10, panel (a)). When comparing to Dis3L2, 168 RNAs were upregulated more than 2-fold, and 177 downregulated more than 2-fold

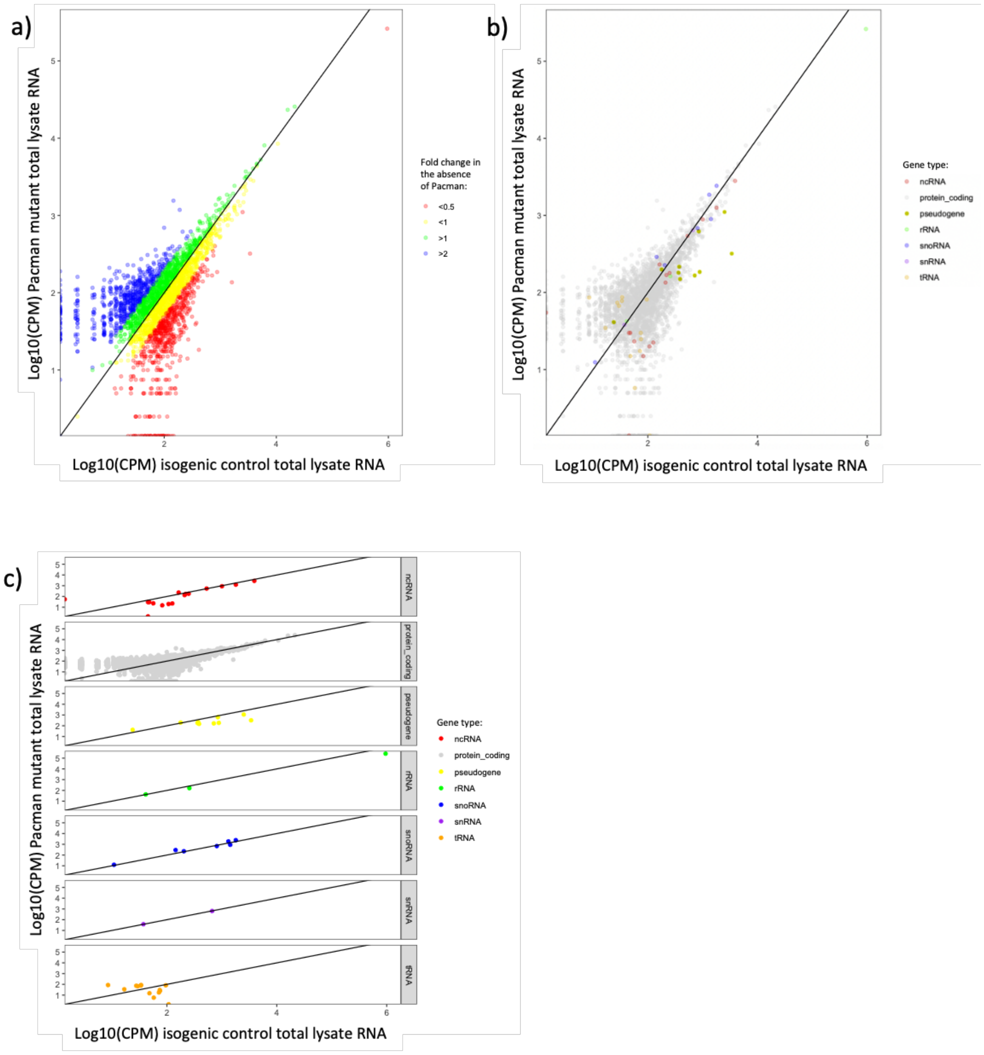


Figure 6.8 – Plots of log10 average CPM in Pacman mutant genotype plotted against isogenic control from the RNA extracted from total L3 lysate

A plot of a log10 of the average CPM of RNAs between Pacman mutant genotype and isogenic control total L3 lysate RNA. Panel (a) highlights the genes with greater than a 2 fold-change either up or down. Panels (b) shows the distribution and abundance of different RNA species in the data, also represented in separated plots in panel (c).

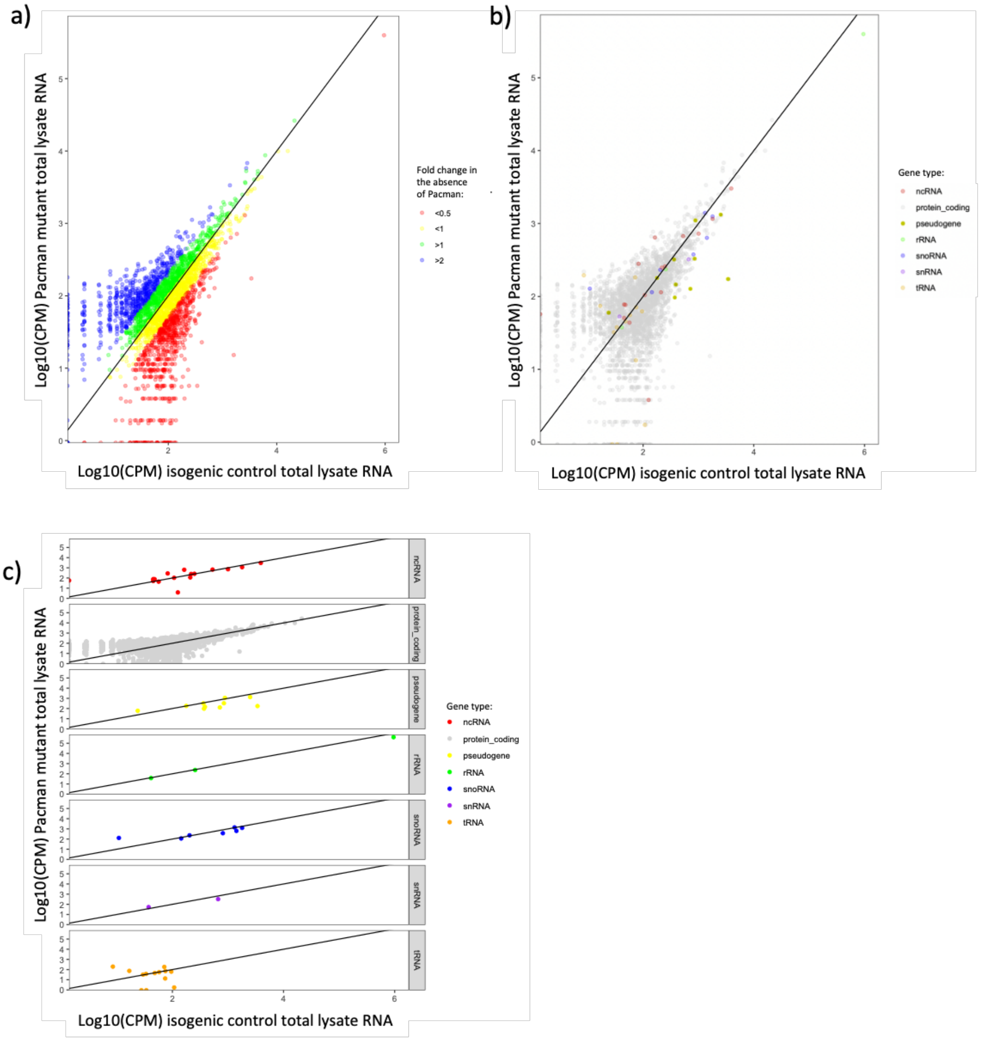


Figure 6.9 – Plots of log10 average CPM in Dis3L2 mutant genotype plotted against isogenic control from the RNA extracted from total L3 lysate

A plot of a log10 of the average CPM of RNAs between Dis3L2 mutant genotype and isogenic control total L3 lysate RNA. Panel (a) highlights the genes with greater than a 2 fold-change either up or down. Panels (b) shows the distribution and abundance of different RNA species in the data, also represented in separated plots in panel (c).

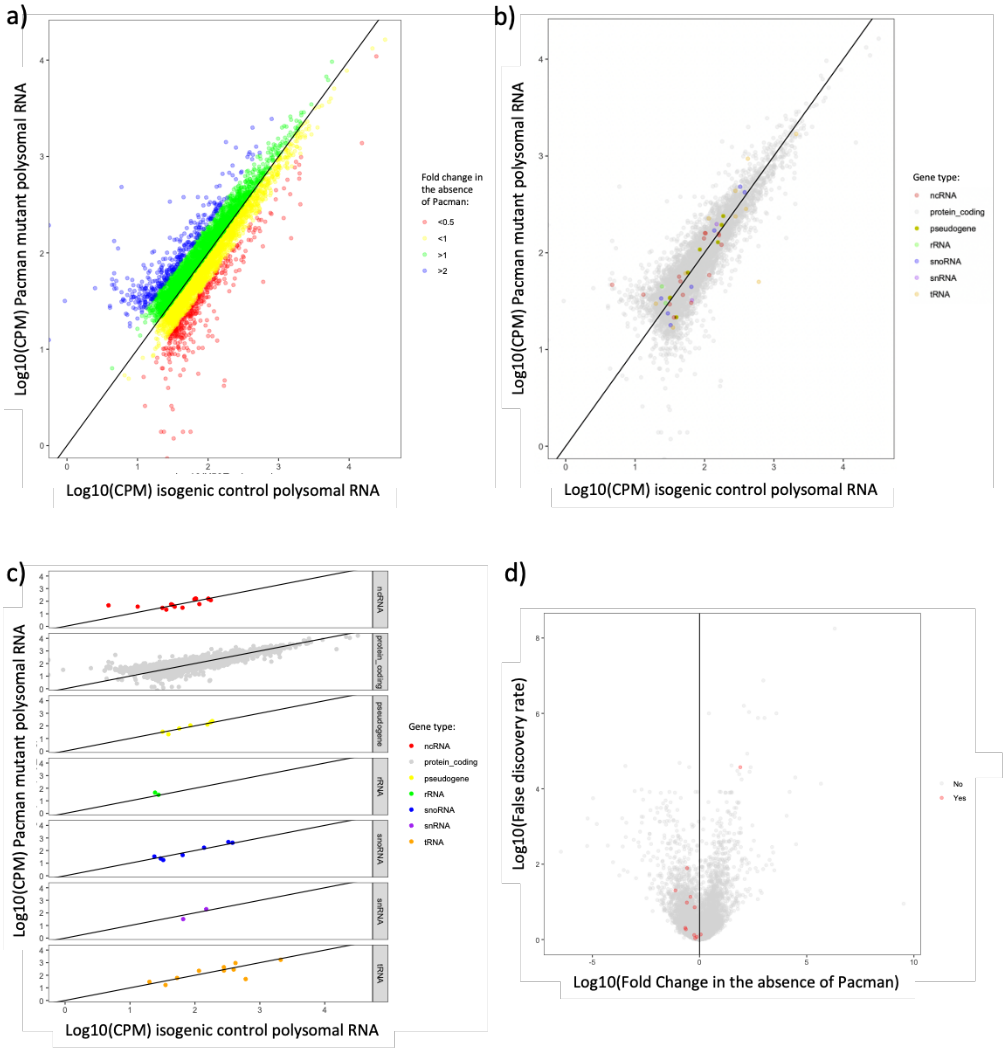


Figure 6.10 – Plots of log10 average CPM in Pacman mutant genotype plotted against isogenic control from the polysomal RNA extracted from L3 larvae

A plot of a log10 of the average CPM of RNAs between Pacman mutant genotype and isogenic control polysomal RNA. Panel (a) highlights the genes with greater than a 2 fold-change either up or down. Panels (b) shows the distribution and abundance of different RNA species in the data, also represented in separated plots in panel (c). Panel (d) compares these fold-changes against false discovery rate, providing a graphical representation of confidence in these data.

(Figure 6.11, panel (a)). Of these genes analysed, the vast majority are protein coding, as would be expected (5476), with 15 non-coding RNAs (13 of which are long non-coding RNAs), and a remaining minority of other RNA types (10 tRNAs, 7 pseudogenes, 7 snoRNAs, 2 rRNAs, and 2 srRNAs). It is unexpected that tRNAs were still present after the BowTie alignment and selection step; especially as the same FlyBase genome was used for both BowTie alignment and analysis of RNA type. These are plotted on panel (b) and (c) of Figures 6.10 and 6.11). The fold change of the lncRNAs against false discovery rate are plotted in panel (d) of Figures 6.10 and 6.11

6.3.2.3 Observing proportions of RNA populations differentially abundant in the presence and absence of each exoribonuclease in total L3 lysate compared to polysomal L3 extract

In order to fully appreciate the novel insights that this dataset can offer, comparisons between control and mutant genotypes must be analysed in both total L3 lysate and polysomal L3 extract, and the overlap and difference between the exoribonuclease targets explored. This will be more fully accomplished in further cross-dataset comparisons
throughout the rest of the chapter, but an initial look here can identify targets that may be specifically degraded by the enzymes in polysomal RNA only, or less so on the polysome.

In Pacman mutant samples, of the 2859 genes able to pass the EdgeR filters in all polysomal and total lysate samples, 140 are more than 1.5-fold upregulated on both the polysome and in total lysate; 126 are upregulated more than 1.5-fold upregulated on the polysome but more than 1.5-fold downregulated in total lysate; 84 are upregulated more than 1.5-fold in total lysate but more than 1.5-fold downregulated on the polysome, with a final 133 more than 1.5-fold downregulated on both the polysome and in total lysate (Figure 6.12).

In Dis3L2 mutant samples, of the 2859 genes able to pass the EdgeR filters in all polysomal and total lysate samples, 34 are more than 1.5-fold upregulated on both the polysome and in total lysate; 94 are upregulated more than 1.5-fold upregulated on the

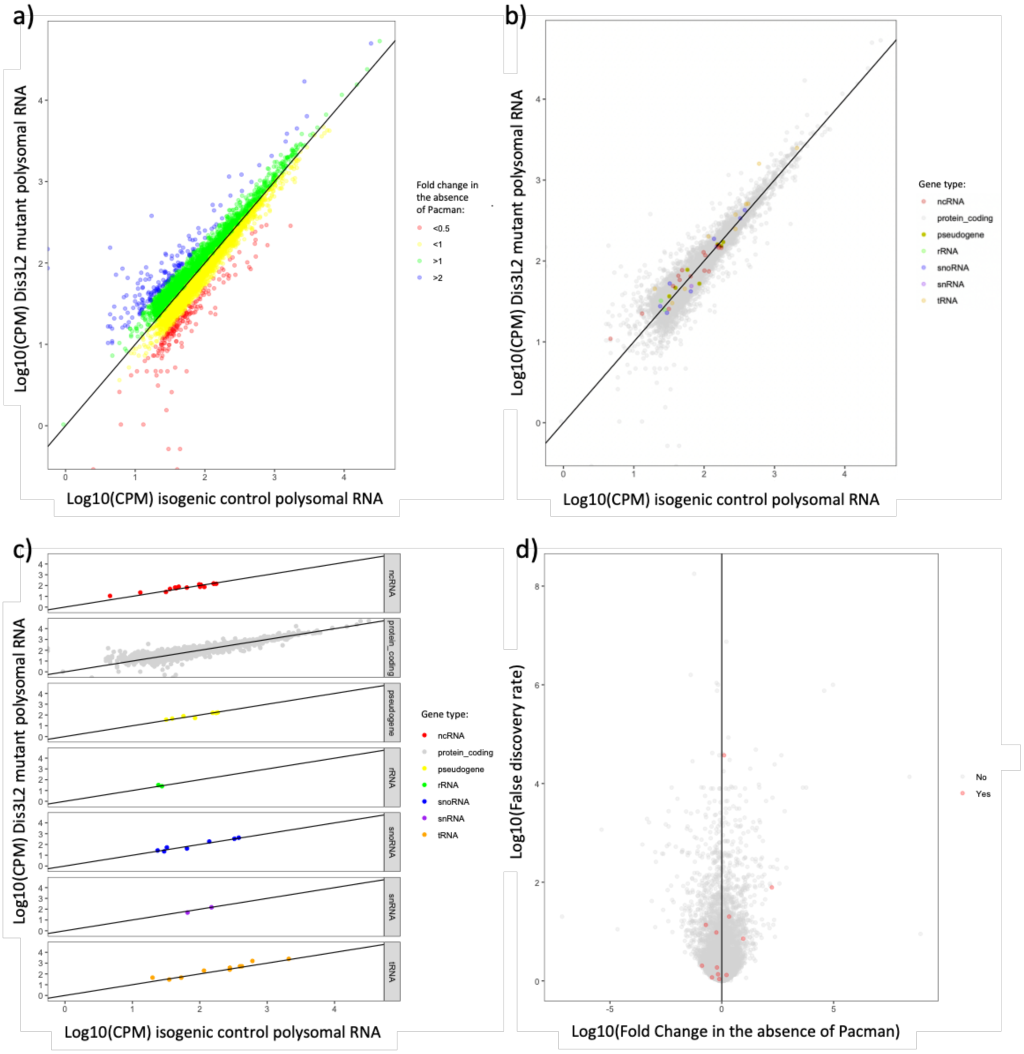


Figure 6.11 – Plots of log10 average CPM in Dis3L2 mutant genotype plotted against isogenic control from the polysomal RNA extracted from L3 larvae

A plot of a log10 of the average CPM of RNAs between Dis3L2 mutant genotype and isogenic control polysomal RNA. Panel (a) highlights the genes with greater than a 2 fold-change either up or down. Panels (b) shows the distribution and abundance of different RNA species in the data, also represented in separated plots in panel (c). Panel (d) compares these fold-changes against false discovery rate, providing a graphical representation of confidence in these data.

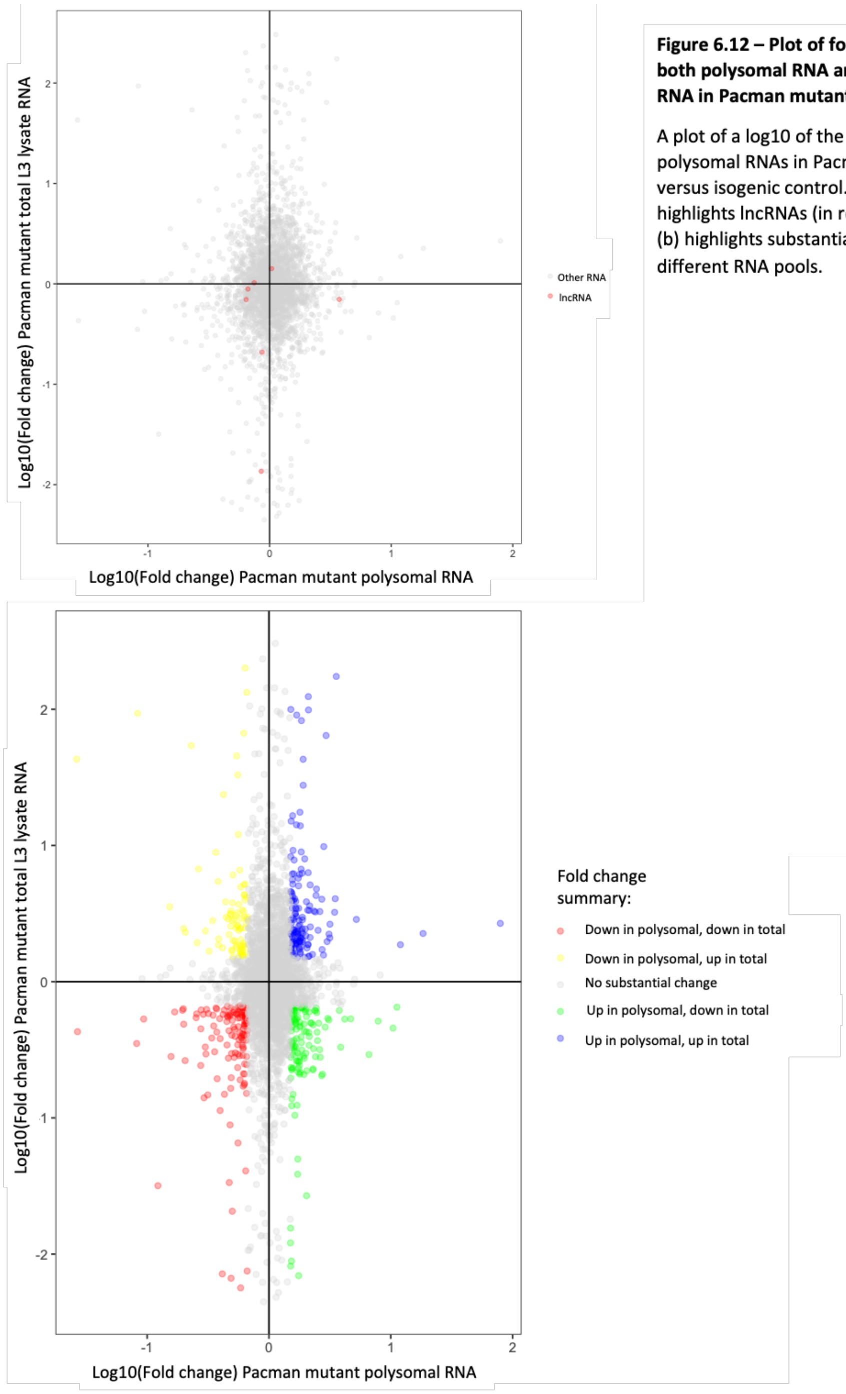


Figure 6.12 – Plot of fold change in both polysomal RNA and total L3 lysate RNA in Pacman mutant L3

A plot of a log10 of the fold change of polysomal RNAs in Pacman mutant versus isogenic control. Panel (a) highlights lncRNAs (in red), while panel (b) highlights substantial fold change in different RNA pools.

polysome but more than 1.5-fold downregulated in total lysate; 20 are upregulated more than 1.5-fold in total lysate but more than 1.5-fold downregulated on the polysome, with a final 64 more than 1.5-fold downregulated on both the polysome and in total lysate (Figure 6.13).

Differential abundance has significantly different implications for both different genotypes, and for polysomal versus total lysate RNAs. Some interpretations are included below:

1) Increase in polysomal RNA + Increase in total lysate RNA:

   An increase in both polysomal and total lysate RNA for a given lncRNA (in the absence of either Pacman or Dis3L2) would suggest that not only is the given exoribonuclease degrading the lncRNA, but this increased abundance in its absence is responsible for the increase in polysomal abundance.

2) Increase in polysomal lysate RNA + No change in total lysate RNA:

   An increase in polysomal abundance with no change in total lysate abundance would suggest that the depletion of functional Pacman or Dis3L2 is playing a role in co-translational regulation, and specifically targeting polysome associated lncRNAs (which could possibly be due to a co-translational degradation mechanism).

3) No change in polysomal lysate RNA + Increase in total lysate RNA:

   An increase in total lysate abundance, with no substantial change in polysomal abundance would imply that the transcript is degraded by whichever exoribonuclease, but is protected from this degradation on the polysome, or simply present at low enough levels that a substantial difference is simply detected (although the EdgeR filters requiring reads in each replicate should remove those that fall into this category by virtue of not being present on the polysome).

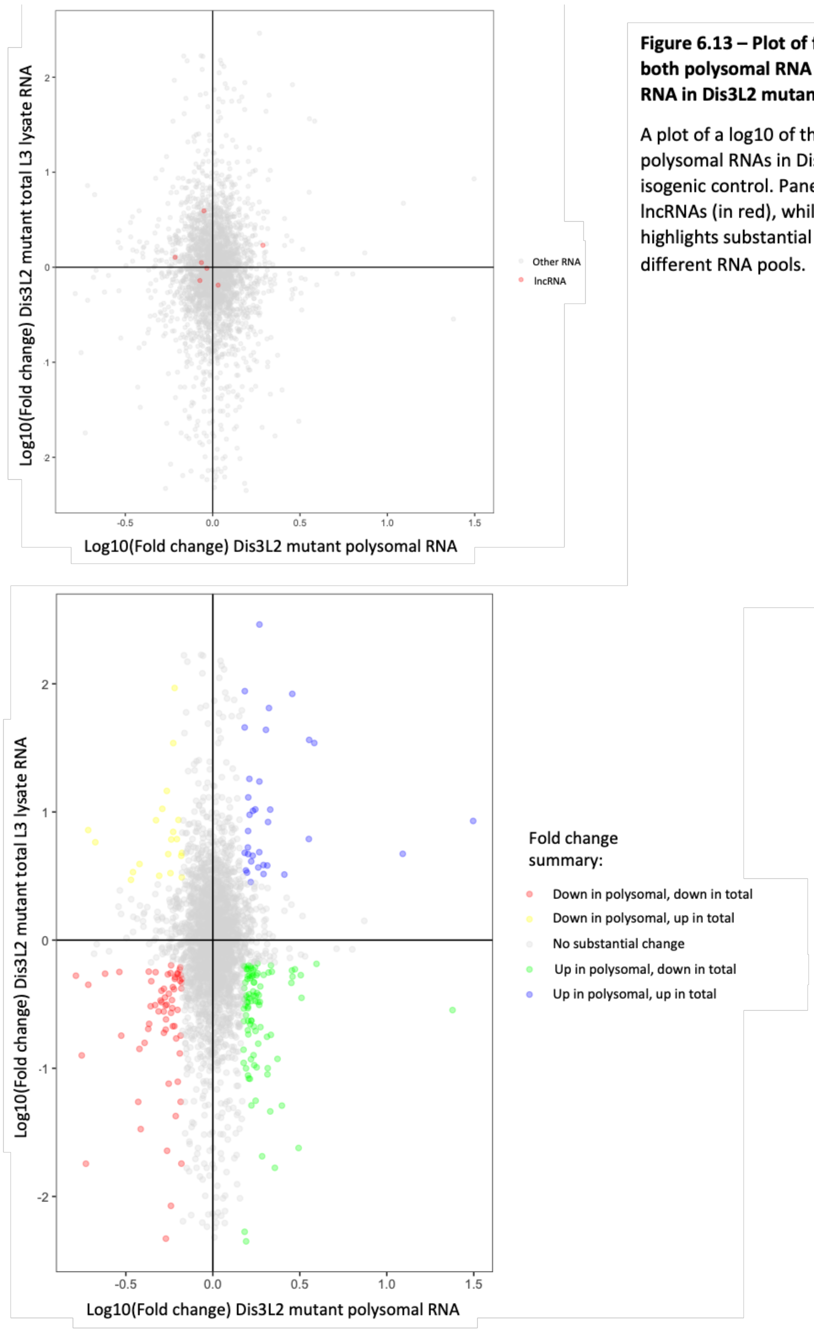


Figure 6.13 – Plot of fold change in both polysomal RNA and total L3 lysate RNA in Dis3L2 mutant L3

A plot of a log10 of the fold change of polysomal RNAs in Dis3L2 mutant versus isogenic control. Panel (a) highlights lncRNAs (in red), while panel (b) highlights substantial fold change in different RNA pools.

4) Increase in polysomal lysate RNA + Decrease in total lysate RNA:

   An increase in polysomal abundance in the absence of either exoribonuclease, accompanied by a decrease of the lncRNA in total lysate RNA, would be particularly interesting, as it may suggest some kind of feedback inhibition between the two.

5) Decrease in polysomal lysate RNA + Increase in total lysate RNA:

   A decrease in polysomal abundance, accompanied by an increase in total lysate abundance would suggest that whilst the transcript does seem to undergo degradation, there is a factor that inhibits this on the polysome, and that the presence of the exoribonuclease in fact leads to an increased presence of the transcript on the polysome.

6) Decrease in polysomal lysate RNA + No change in total lysate RNA:

   A decrease in polysomal abundance with no change in total lysate abundance would suggest that the absence of the exoribonuclease (whilst seemingly not effecting the overall stability of the transcript) is indirectly decreasing its association with the polysome. This could be an indirect action through another of the RNA targets of Pacman or Dis3L2, a target transcript which may inhibit the association of the lncRNA with the polysome.

7) No change in polysomal lysate RNA + Decrease in total lysate RNA:

   A decrease in total lysate abundance with no change in polysomal abundance would suggest that the transcript is not directly degraded by either exoribonuclease, although appears to be a target of indirect action, possibly a transcript directly targeted by Pacman or Dis3L2 may decrease the transcription or stability of the lncRNA of interest.

8) Decrease in polysomal lysate RNA + Decrease in total lysate RNA:

   If both polysomal and total lysate abundance decrease in the absence of either exoribonuclease this likely rules out direct action of the exoribonuclease on the degradation of that transcript. However, given the change, the absence of the exoribonuclease clearly has some impact. This can be caused by one of the transcripts that the exoribonuclease does degrade having a knock on effect on the transcription or stability of another transcript, allowing indirect regulation by Pacman or Dis3L2.

9) No changes:

   If no substantial changes in abundance (in either polysomal or total lysate RNA) is seen in the absence of Pacman or Dis3L2, then the exoribonuclease likely plays no role in the regulation of that transcript, or a role that cannot be detected with this number of replicates and limited read depth.

With this stringent filtering, only 7 lncRNAs in both the Pacman and Dis3L2 mutant (out of 2859 RNAs passed all filtering criteria. Of these, in the Pacman mutant, two had no substantial changes in abundance in any of the samples (*CR42682, cherub*), two had a greater than 1.5-fold decrease in the polysomal RNA, with no substantial change in the total lysate RNA (*CR32652, roX1*), two had a greater than 1.5-fold decrease in the total lysate RNA, with no substantial change in the polysomal RNA (*CR33938, roX2*), and one had a greater than 1.5-fold increase in polysomal RNA, with no substantial change in total lysate RNA (*HsrOmega*). In the Dis3L2 mutant, three had no substantial changes in abundance in any of the samples (*CR32652, roX1, roX2*), one had a greater than 1.5-fold decrease in the polysomal RNA, with no substantial change in the total lysate RNA (*CR42862*), one had a greater than 1.5-fold decrease in the total lysate RNA, with no substantial change in the polysomal RNA (*HsrOmega*), and one had a greater than 1.5-fold increase in polysomal RNA, with no substantial change in total lysate RNA (*CR33938*). It should of course be noted that (due to the nature of these stringent filters) there are likely many more lncRNAs that could be analysed and assigned informative profiles like these with either additional replicates, or greater read depth.

6.3.3 Comparison between total and polysomal RNA identifies population of lncRNAs that are associated with the polysome and likely to be actively translated in *Drosophila* L3 larvae

Due to the low coverage, the first approach used here to identify potentially translated lncRNAs used all six replicates, regardless of genotypes. This approach is not without its own problems, but it was decided on balance to be the most promising avenue to identify translated lncRNAs. By initially extracting all RNAs annotated as lncRNAs (using Flybase 6.29 annotation), those with substantial detectable polysomal association could be identified. All lncRNAs present on the genome file were sorted by average polysomal reads (across all six samples; 2 in wild type isogenic control, 2 in Pacman mutant, 2 in Dis3L2 mutant *Drosophila* L3 larvae). Of 1918 lncRNAs mapped to the *Drosophila* genome used, 101 showed greater than or equal to an average of 1 read per sample. Given the nature of sequencing experiments, low numbers of detected reads may be present as a technical artifact, so those with more consistent and higher numbers of detected reads provide higher confidence examples of polysome associated lncRNAs. Conversely, given the limited depth of this sequencing data, any sample that shows up consistently is likely underrepresented compared to its prevalence in higher resolution datasets. From all lncRNAs, those with an average of 3 or more reads were extracted, providing a total of 33 candidate polysome-associated lncRNAs to be further examined (Figure 6.14).

Although all of the candidate RNAs here have evidence showing them to be present to some extent on the (translationally active) polysome, a second more stringent set of criteria was applied to try to further highlight prime examples of lncRNAs that might be translating. All reads were normalised to the total number of reads counted in that particular sample, expressed as counts per million (CPM). Although ordinarily, some form of normalisation for gene length is useful for quantification, this is not easily achieved in the context of ribosome footprinting, as this would depend on ORF length, in situations where the ORF is not even known. For comparisons between the same gene, and to observe its presence on the polysome in a non-quantitive manner, CPM with no further normalisation is sufficient. Following normalisation, the data was passed through expression filters using the statistical analysis package EdgeR. The filters

a)

Gene ID	Average Polysome Reads
CR42862	304.33
roX1	257
cherub	142.67
Hsromega	111
CR32652	26.67
roX2	24.17
CR9284	23.33
CR33938	20.5
CR42850	14.67
CR31781	13.5
CR33948	13.17
CR42767	9.83
CR46006	9.67
CR32835	9.33
CR31451	7.33
CR43626	6.17
CR45102	6.17
CR45388	6.17
CR43334	6
CR31044	5.67
CR46003	5.67
CR43314	5.33
CR40469	5
CR43432	4.83
CR43459	4.5
CR42657	4.33
CR45234	4.17
CR43278	4
CR44455	3.67
flam	3.67
CR44206	3.33
CR43685	3.17
CR44957	3

b)

Gene ID	Average Total Reads
Hsromega	275
roX1	171.50
CR42862	121.83
cherub	79.67
CR40469	31.33
CR44684	27.50
roX2	17.67
CR43626	12.17
CR33938	9.50
CR45102	9.50
CR32652	9.17
CR44042	9.17
CR43334	8.67
CR46006	8.50
CR43314	6
CR43174	5.50
CR45668	4.83
CR43253	4.67
CR42767	4.17
iab8	4.17
CR9284	3.33
let7C	3.33

Figure 6.14 – Tables showing all polysome associated lncRNAs identified by poly-ribo-seq with an average of 3 reads mapped in either polysomal or total RNA

Table (a) shows those with an average of 3 or more reads in polysomal RNA, while table (b) shows those with an average of 3 or more reads in total L3 lysate RNA. Of these, 15 lncRNAs were common to both datasets (marked in bold), demonstrating significant overlap, and likely showing these to be generally abundant transcripts, which also happen to be found on the polysome.

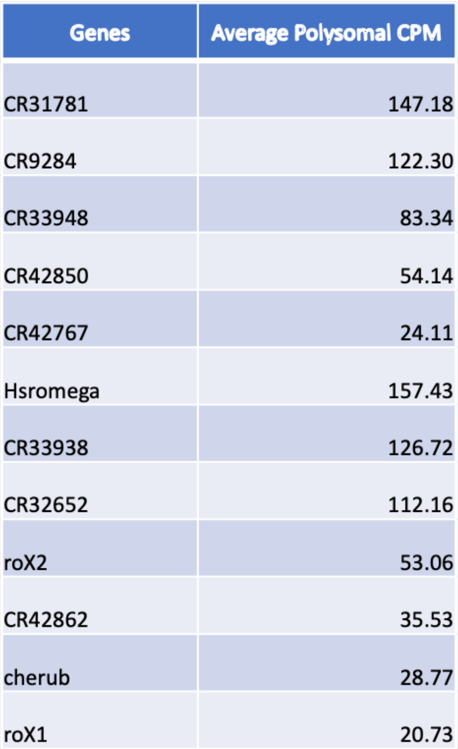
selected only the genes with a sufficient number of reads to be used in statistical analysis. CPM were also calculated for the same genes in the total RNA (rather than polysomal) samples. The same statistical filters were applied, and genes that had passed the filters for both total and polysomal RNA samples (and can confidently asserted to have been detected in both polysome bound and total RNA) were identified. In order to ascertain whether any of the candidates were specifically enriched in polysomal RNA, the ratio of CPM polysomal:CPM total was calculated. This further analysis is summarised in Figures 6.15 and 6.16. Using this filtering methods, *CR33938* and *CR32652* showed an obvious enrichment on the polysome (by average normalised polysomal RNA reads versus average normalised total rRNA depleted RNA reads), suggesting that these lncRNA candidates are in fact predominantly found associated with the polysome.

6.4 Global overview and summary of further uses and comparisons available from Zhang *et al*. RNA sequencing of harringtonine treated S2 cells

In Chapter 3, an initial examination of the Harringtonine treated S2 cell sequencing data by *Zhang et al.* was carried out in order to spot preliminary trends and help prove the principle of ribosome-associated lncRNAs (154). This was limited in scope when taken in isolation, whereas this novel work, along with other datasets previously explored, allows for a more thorough and informative examination of that same data now. Despite this being carried out in S2 cells (compared to the whole L3 of the novel dataset), the principles of ribosome pile-up, searching for ORFs, and relative abundance of lncRNAs in *Drosophila* can still be explored.

6.4.1 Comparison of elongation inhibited reads with reads from novel poly-ribo-seq dataset

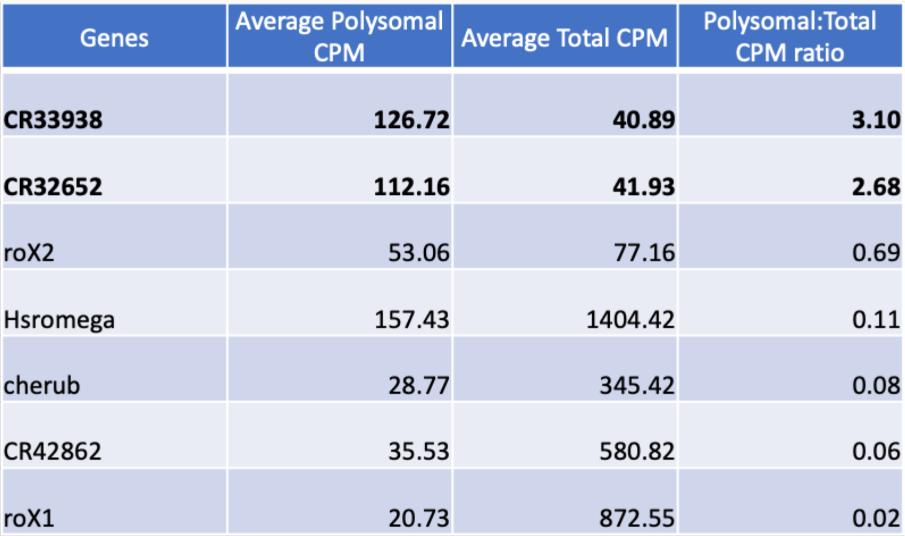
Previously, of all ribosome-associated lncRNAs found in the data of Zhang *et al.*, 51 were detected as having 20 or more reads detected in the ribosome-bound RNA sample. By comparing the 51 lncRNAs selected in this shortlist (detected with at least 20 reads in the ribosomal RNA in Chapter 3, Figure 3.14) and the 33 candidate lncRNAs selected from the novel dataset (detected with an average at least 3 polysomal reads), a



Genes	Average Polysomal CPM
CR31781	147.18
CR9284	122.30
CR33948	83.34
CR42850	54.14
CR42767	24.11
Hsromega	157.43
CR33938	126.72
CR32652	112.16
roX2	53.06
CR42862	35.53
cherub	28.77
roX1	20.73

Figure 6.15 – Tables showing all polysome associated lncRNAs identified by poly-ribo-seq with sufficient reads for EdgeR statistical testing

The 12 lncRNAs listed here were all detected in the polysomal RNA with sufficient reads in all samples to pass through the stringent filtering for statistical analysis, as defined by EdgeR.



Genes	Average Polysomal CPM	Average Total CPM	Polysomal:Total CPM ratio
CR33938	**126.72**	**40.89**	**3.10**
CR32652	**112.16**	**41.93**	**2.68**
roX2	53.06	77.16	0.69
Hsromega	157.43	1404.42	0.11
cherub	28.77	345.42	0.08
CR42862	35.53	580.82	0.06
roX1	20.73	872.55	0.02

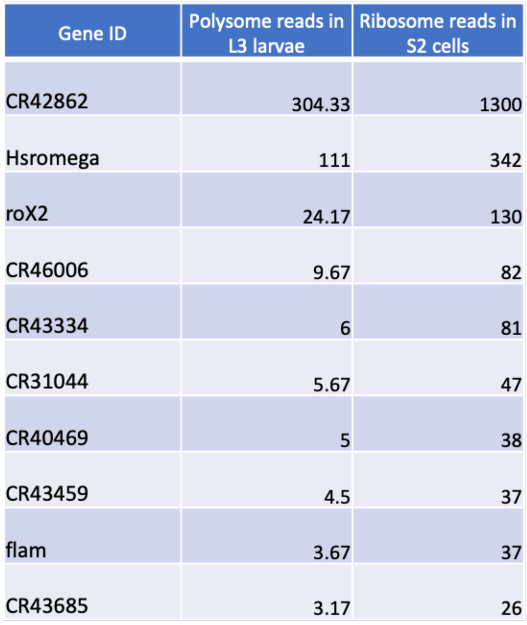
Figure 6.16 – Tables showing lncRNAs significantly present in both polysomal and total RNA samples

The 7 lncRNAs listed here were all detected in both the polysomal and total RNA with sufficient reads in all samples to pass through the stringent filtering for statistically significant analysis, as set by EdgeR. The ratio between Polysomal CPM and Total CPM indicates whether the lncRNAs are specifically enriched on the polysome, with a higher ratio showing a higher prevalence in the polysomal fractions. Two lncRNAs (highlighted in bold) identified here show an obvious enrichment on the polysome, showing that these lncRNA candidates are in fact predominantly found associated with the polysome.

new shortlist of lncRNAs present on ribosomes in both L3 larvae and S2 cells was produced. As mentioned in Chapter 3, a higher cut off was used in the interpretation of the data from Zhang *et al*. than for the novel data set; as the increased read depth of this data allows for more stringent cutoffs to be used. A total of 10 lncRNAs were common to both of these datasets and criteria, listed in Figure 6.17. It is worth noting that the order of abundance is relatively well conserved between datasets; this is encouraging, as although the novel data set has limitations (as previously discussed), this provides evidence that the available comparisons on transcripts that have been significantly detected can still be used. The correlation between normalised reads from novel L3 poly-ribo-seq and Zhang *et al*. RNA-seq are plotted in Figure 6.18. The implications from this shortlist would be twofold. First, the consistent association of these lncRNAs with ribosomes and polysomes reinforces a prediction that they are likely to be undergoing translation. Secondly, the fact that the association occurs in both L3 larvae and S2 cells shows that these lncRNA likely have a role at both the L3 developmental timepoint, and in the healthy growth and proliferation of S2 cells. The polysomally enriched candidate lncRNAs identified in 6.3.3 (*CR33938* and *CR32652*) were not present in this analysis, however this could be due to specific expression in L3, or S2 cells not expressing these genes.

6.5 Global overview and summary of further uses and comparisons available from Zhang *et al*. RNA sequencing of poly-ribo-seq *Drosophila* developmental timepoints and cells

Following the initial examination of the data from Zhang *et al.* in Chapter 3, an initial examination of the data by Zhang *et al.* was carried out in order to spot preliminary trends and help prove the principle of polysome-associated lncRNAs (154). This was limited in scope when taken in isolation, whereas this novel work, along with other datasets previously explored, allows for a more thorough and informative examination of that same data now.



Gene ID	Polysome reads in L3 larvae	Ribosome reads in S2 cells
CR42862	304.33	1300
Hsromega	111	342
roX2	24.17	130
CR46006	9.67	82
CR43334	6	81
CR31044	5.67	47
CR40469	5	38
CR43459	4.5	37
flam	3.67	37
CR43685	3.17	26

Figure 6.17 – Table showing ribosome associated lncRNAs common to both S2 cell ribosomal RNA and L3 polysomal RNA

The 10 lncRNAs listed here are present with 20 or more reads in ribosomal RNA from harringtonine treated S2 cells and an average of 3 or more reads in L3 larvae.

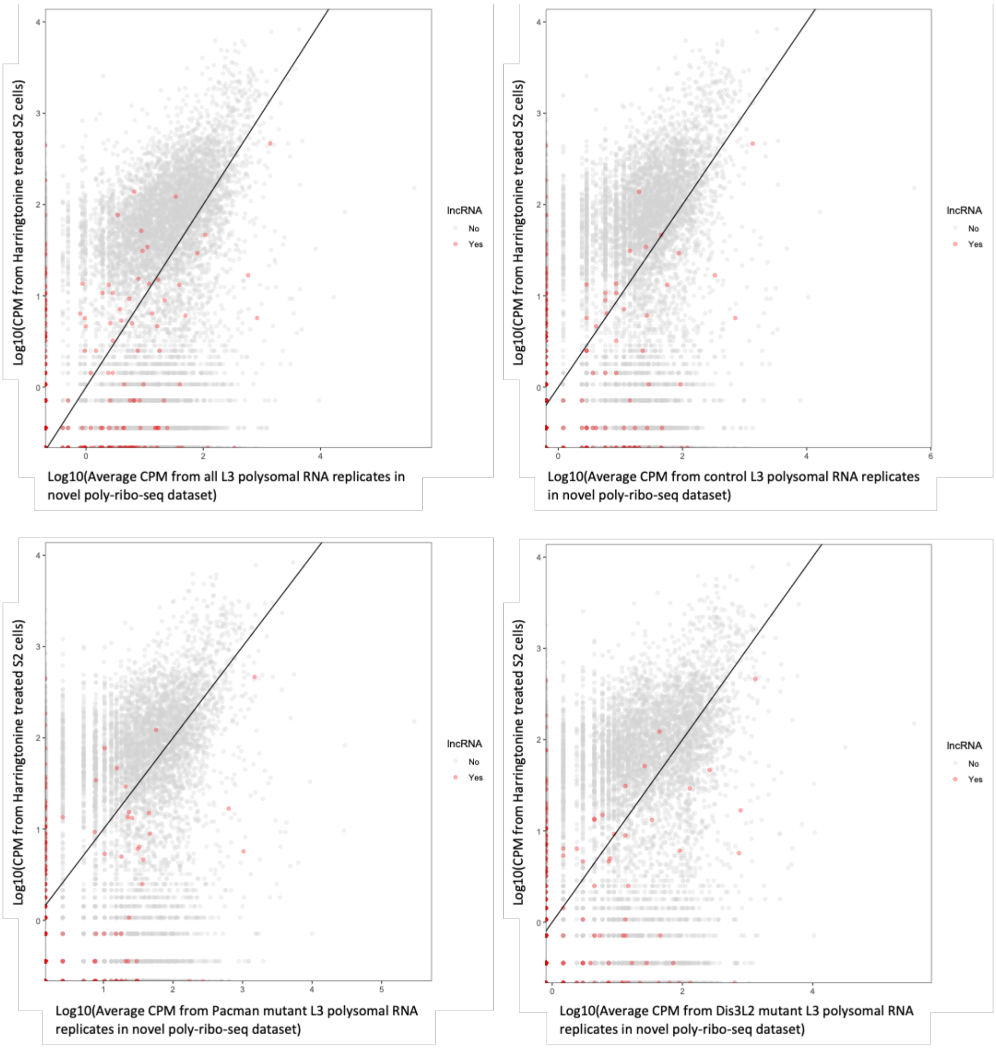


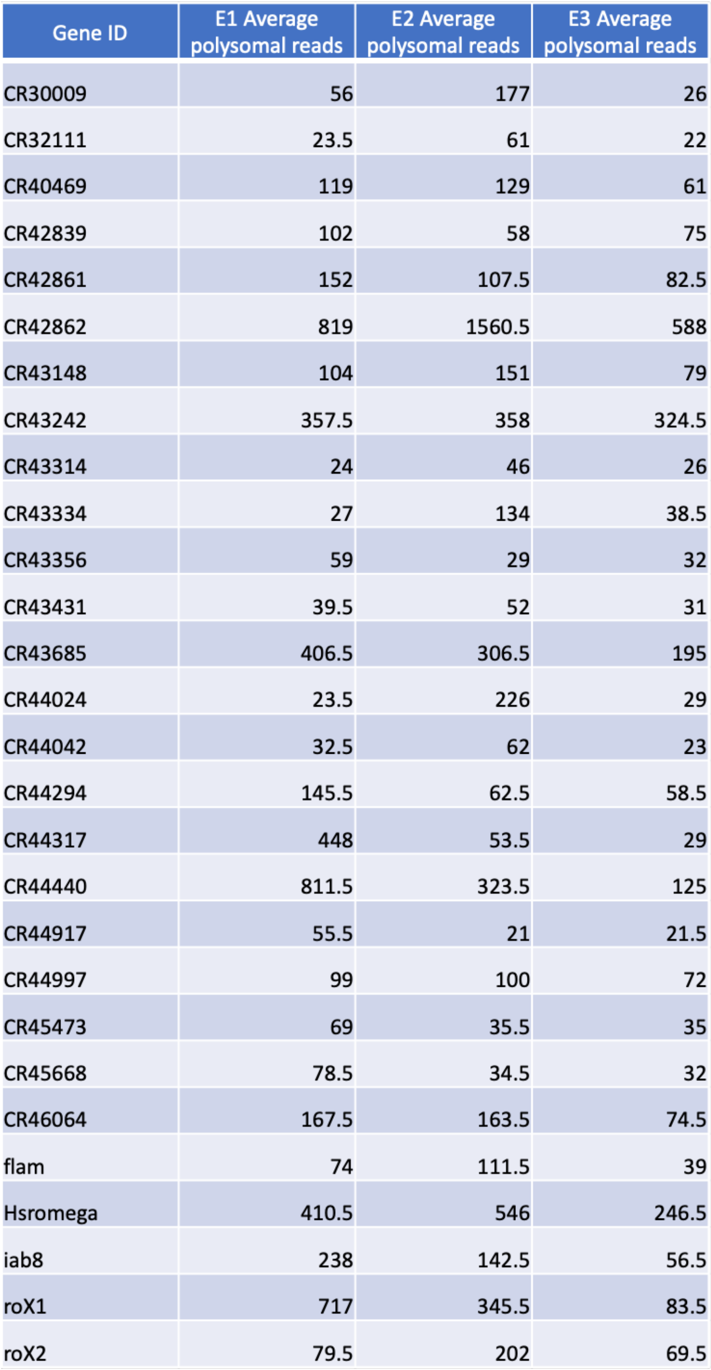
Figure 6.18 – Scatter plots showing correlation between normalised reads from novel L3 poly-ribo-seq and normalised reads from Zhang et al. RNA-seq data

Novel data is plotted as an average from all replicates (a), control replicates (b), Pacman mutants (c), and Dis3L2 mutants (d). lncRNAs are highlighted in red.

6.5.1 Comparison in poly-ribo-seq data from *Drosophila* S2 cells, embryo, and L3 larvae

Previously, a list was produced of 28 lncRNAs detected with at least 20 polysomal reads in *Drosophila* embryos at E1, E2, and E3 developmental timepoints (Figure 6.19). This shortlist was then compared to the polysomal shortlist in from the comparative analysis of the data developed in this thesis and that from Zhang *et al.* in Chapter 3 (Figure 3.15), (lncRNAs detected with an average of at least 3 polysomal reads in L3 larvae). This produced a new shortlist (Figure 6.20, panel (a)), of 9 of the lncRNAs with the highest number of reads consistently detected on the polysome through all tested developmental timepoints (E1, E2, E3, and L3). This list therefore provides strong candidates for lncRNAs consistently expressed and translated throughout the *Drosophila* life cycle.

Although the comparison between cell lines and living tissues or whole organisms is more difficult to draw meaningful conclusions from (differences would always be expected between the two), the novel poly-ribo-seq data was also compared to the S2 cell poly-ribo-seq data from the Couso lab. As with the embryo data, the read count and counts per million (CPM) for all lncRNAs were extracted from the Couso lab's poly-ribo-seq data and sorted in order of average number of reads detected in polysomal RNA at each developmental timepoint. Only lncRNAs with an average of at least 20 reads were kept, in order to increase confidence in detection of polysome-association. The 46 lncRNA matching these criteria were compared to the polysomal shortlist in Figure 6.19 (lncRNAs detected with an average of at least 3 polysomal reads in L3 larvae). This produced a new shortlist (Figure 6.20, panel (b)) of the 13 lncRNAs with the highest number of reads consistently detected on the polysome in both S2 cells and L3 larvae. Due to the nature of fast-proliferating stable cell lines, and their differences from the controlled and specific growth of cells within the tissues of a model organism, it is hard to speculate meaningfully what the role of these lncRNAs may be as a result of their consistent polysome association between both models (which further increases confidence). That said, it does provide further supporting evidence for these lncRNAs to be meaningfully translated. To further demonstrate this point, the 50 lncRNAs with highest counts per million in polysomal RNA for each sample were listed and compared



Gene ID	E1 Average polysomal reads	E2 Average polysomal reads	E3 Average polysomal reads
CR30009	56	177	26
CR32111	23.5	61	22
CR40469	119	129	61
CR42839	102	58	75
CR42861	152	107.5	82.5
CR42862	819	1560.5	588
CR43148	104	151	79
CR43242	357.5	358	324.5
CR43314	24	46	26
CR43334	27	134	38.5
CR43356	59	29	32
CR43431	39.5	52	31
CR43685	406.5	306.5	195
CR44024	23.5	226	29
CR44042	32.5	62	23
CR44294	145.5	62.5	58.5
CR44317	448	53.5	29
CR44440	811.5	323.5	125
CR44917	55.5	21	21.5
CR44997	99	100	72
CR45473	69	35.5	35
CR45668	78.5	34.5	32
CR46064	167.5	163.5	74.5
flam	74	111.5	39
Hsromega	410.5	546	246.5
iab8	238	142.5	56.5
roX1	717	345.5	83.5
roX2	79.5	202	69.5

Figure 6.19 – Table showing polysome associated lncRNAs common to all embryo developmental stages

The 28 lncRNAs listed here are present with 20 or more reads in polysomal RNA from *Drosophila* embryo in stages E1, E2, and E3.

a)

Gene ID	L3 Average Polysome Reads	E1 Average Polysome Reads	E2 Average Polysome Reads	E3 Average Polysome Reads
CR40469	5	119	129	61
CR42862	304.33	819	1560.5	588
CR43314	5.33	24	46	26
CR43334	6	27	134	38.5
CR43685	3.17	406.5	306.5	195
flam	3.67	74	111.5	39
Hsromega	111	410.5	546	246.5
roX1	257	717	345.5	83.5
roX2	24.17	79.5	202	69.5

b)

Gene ID	S2 polysomal reads	Average L3 polysomal reads
CR31044	170	5.67
CR31451	137	7.33
CR42862	19074	304.33
CR43334	1053	6
CR43459	1667	4.5
CR43685	159	3.17
CR44206	534	3.33
CR45102	82	6.17
CR46006	37	9.67
flam	65	3.67
Hsromega	2189	111
roX1	51	257
roX2	300	24.17

Figure 6.20 – Tables showing polysome associated lncRNAs common to E1, E2, E3, and L3 developmental stages in *Drosophila*

(a) The 9 lncRNAs listed here were all detected with an average of at least 20 reads in all embryonic stages, and an average of at least 3 reads in L3 larvae.

(b) The 13 lncRNAs listed here were all detected with at least 20 reads S2 cells, and an average of at least 3 reads in L3 larvae.

to the 50 lncRNAs with highest average counts per million in polysomal RNA in the L3 larvae dataset. The percentage of said top 50 common between both L3 and each other sample are plotted in Figure 6.21.

The comparisons between these datasets provide mostly information to narrow down the search for promising candidates of lncRNAs present on the polysome in the novel L3 data, by highlighting those that retain this polysome-association at other developmental stages (E1, E2, and E3), as well as in a commonly used *Drosophila* cell line (S2). We can take from these comparisons that a significant proportion of lncRNAs associated with the polysome are common between the developmental timepoints and cells tested by the Couso lab, and the novel poly-ribo-seq data on L3 larvae. Of the most abundant lncRNAs found associated with the polysome in each sample (according to average CPM across replicates), between 32% and 40% were conserved from each sample type tested by the Couso lab versus the new data in L3 larvae. This substantial proportion of lncRNAs common to different samples not only shows that important targets are still being detected at a high enough read count to identify those that are of a notably increased abundance; but also demonstrates that a substantial proportion of polysome associated lncRNAs in L3 are found on the polysome at other developmental stages. This is a promising indication of the potential importance of these lncRNAs, and more specifically that their translation may play an important role.

5.6 The effect of Pacman depletion on the transcriptional and translational landscape in *Drosophila* model systems: Global overview and summary of Antic *et al.* dataset

Following the initial examination of the data from Antic *et al.* in Chapter 3, an initial examination of the data by Antic *et al.* was carried out here in order to spot preliminary trends and help prove the principle of polysome-associated lncRNAs (69). This was limited in scope when taken in isolation, whereas this novel work, along with other datasets previously explored, allows for a more thorough and informative examination of that same data now.

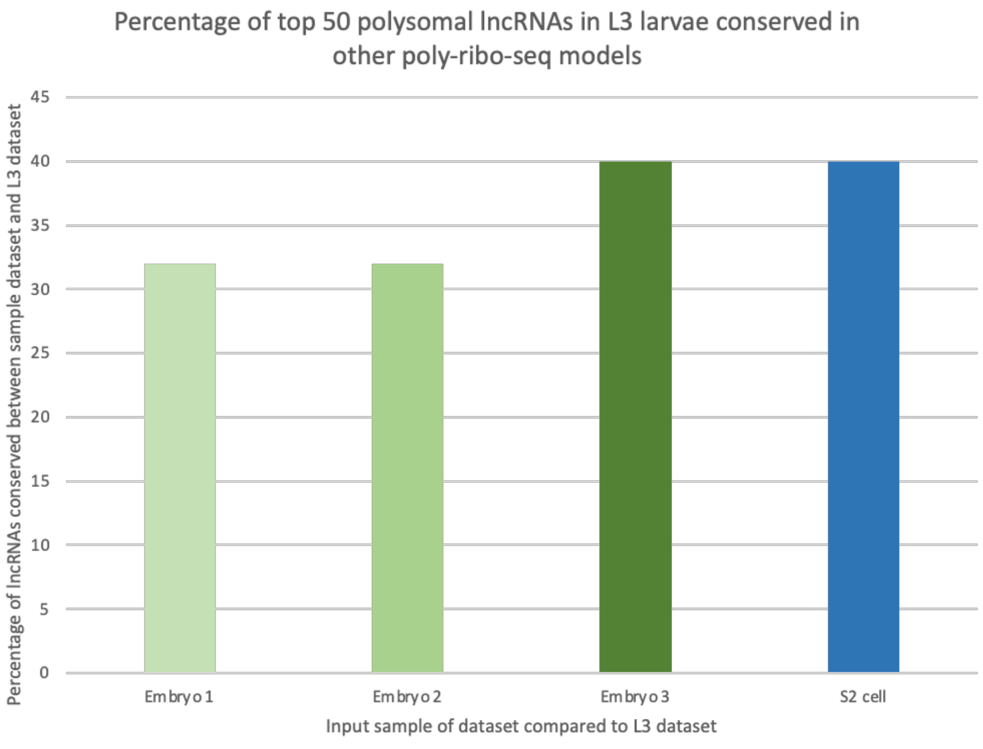


Figure 6.21 – Tables showing the percentage of top polysome associated lncRNAs common between each of E1, E2, E3, and S2 samples, and the L3 sample

The 50 lncRNAs with highest counts per million in polysomal RNA for each sample was listed, and compared to the 50 lncRNAs with highest average counts per million in polysomal RNA in the L3 larvae dataset. The percentage of said top 50 common between both L3 and each other sample is plotted above.

6.6.1 Comparison of Pacman degraded lncRNAs between Antic *et al*. RNA-seq in Pacman depleted S2 cells and total RNA-seq from novel dataset

The unaligned .bam files from the supplemental data of Antic *et al*. were downloaded and converted into .fastq files. A .fasta file of the *Drosophila melanogaster* genome was built using Flybase release 6.18 with the HiSat2 algorithm. All 4 *Drosophila* chromosomes were used. This index file was used to align the reads, using the Bowtie algorithm. CuffLinks, CuffMerge, CuffQuant, and CuffDiff were used to produce a spreadsheet of read counts for each condition. Read counts were normalised by total reads per sample, to give CPM and by gene length to give FPKM, and these were extracted in Pacman knockdown and untreated control. From these, the abundance in Pacman knockdown cells were compared to the abundance in the untreated cells. Those with an increased abundance in the Pacman knockdown (implying the direct targeting of these transcripts by Pacman) with a fold change equal than or greater than two-fold (an arbitrary cut off, chosen to increase confidence in the increased abundance in data with low replicate counts) were extracted and plotted (Figure 3.2).

In order to observe Pacman regulation of lncRNAs in L3 larvae versus S2 cells, the normalised read count of RNA from the total L3 lysate (in isogenic control and Pacman null mutant) and the normalised read count of S2 cell RNA (in untreated wild type and Pacman dsRNA knockdown) were compared (Figure 6.22). Of all recorded RNAs, only lncRNAs were extracted and kept. From these, any genes with an average of 0 reads in Pacman mutant or knockdown samples were discarded. Of the remaining lncRNAs, different fold-change criteria were assessed in order to observe conservation of degradation targets. Of lncRNAs with Pacman mutant reads detected, any lncRNA with increased levels in the absence of Pacman was taken, giving 49 in L3 larvae, and 205 in S2 cells, with only 6 common to both (CR41257, CR44948, CR46234, CR30009, roX1, and HsrOmega). To narrow the scope of these very loose criteria, only those lncRNA with a 1.5-fold or greater increase in normalised reads in the absence of Pacman were selected and taken forward, leaving 114 in the S2 cell data, and 45 in the L3 larval data, with only 1 lncRNA common to both (CR46234). These common targets are listed in Figure 6.23, panels (a) and (b).

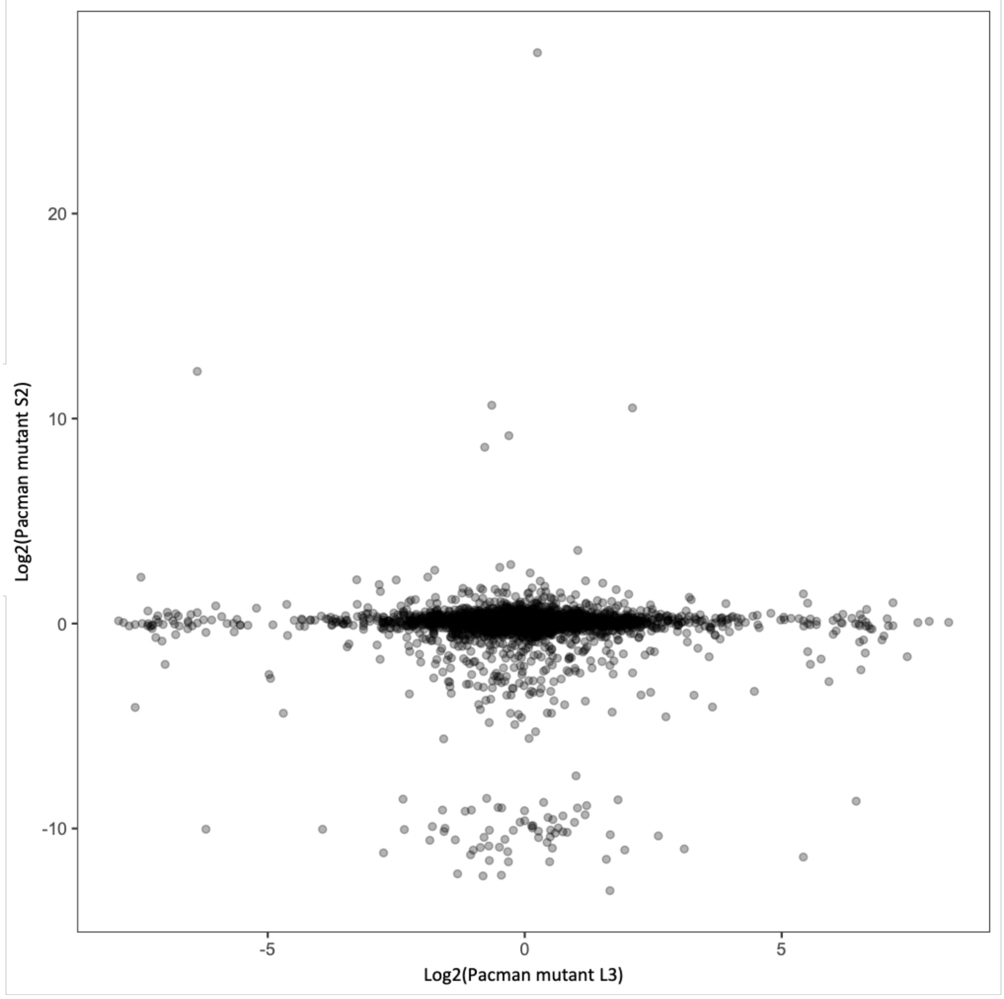


Figure 6.22 – Correlation between fold-change of RNA transcripts in the absence of Pacman in *Drosophila* L3 larvae vs S2 cells

A plot of a log2 of the fold change of RNAs between control and Pacman depleted samples (L3 on the x axis, S2 on the y axis). A strong correlation would be shown by data points with equal values on the x-axis and y-axis (with many of them forming a line on x=y). This data shows virtually no correlation between data-sets, and a trend of very little fold-change in Pacman depleted S2 cells (shown by a tight plotting of data points along the x-axis. There is still a trend for RNAs to not change substantially in Pacman mutant L3 larve either, as shown by some data point clustering along the y-axis. The lack of x=y correlation will be due to tissue specific expression of many RNAs.

a)

Gene name	Fold change (L3)	Fold change (S2)
CR30009	Infinite	1.13
CR41257	Infinite	1.11
CR44948	Infinite	1.25
CR46234	Infinite	Infinite
HsrOmega	1.11	1.50
roX1	1.46	1.08

b)

Gene name	Fold change (L3)	Fold change (S2)
CR46234	Infinite	Infinite

c)

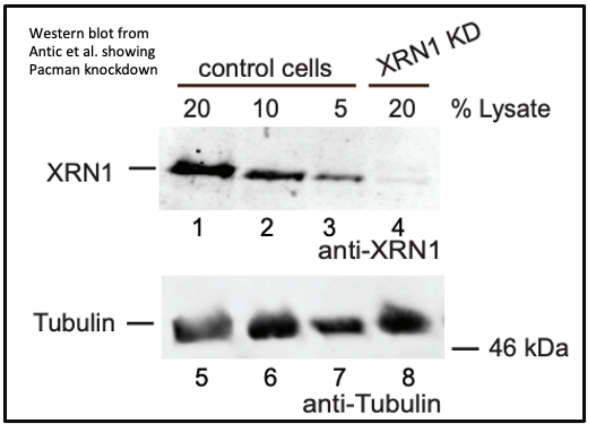


Figure 6.23 – Conserved lncRNA targets of Pacman in S2 cells and L3 larvae

Of all lncRNAs detected as upregulated in RNA sequencing of total lysate from both S2 cells and L3 larvae, those common to both were taken and listed. A total of 50 lncRNAs with increased normalised read count in the absence of Pacman were detected in L3 larvae, and 205 in S2 cells. Of these, the 6 listed in table (a) are common to both. With a cut off of 1.5-fold increase or more, there were 114 in the S2 cell data, and 45 in the L3 larval data, with only one common to both, listed in table (b). The S2 cells treated with dsRNA complimentary to Pacman were tested for knockdown at the protein level by Western blot (c). A strong (though unquantified) knockdown can clearly be observed.

The lack of conservation in targets between the two datasets is perhaps surprising, although several factors must be considered when looking at this data. Firstly, as previously mentioned, the low read depth on the novel poly-ribo-seq dataset, especially in the total L3 lysate extracted RNA, prevents lowly expressed RNAs from being detected whatsoever, and further, prevents small but real differential abundances from being detected during analysis. Secondly, the difference between whole *Drosophila* at the L3 developmental timepoint (shortly before pupation) is very different from the perpetual growth with no further differentiation of S2 cells (derived from late stage embryos). Finally, the difference between a strong but incomplete dsRNA knockdown of Pacman by Antic *et al*. (Figure 6.23, panel (c)) and a complete functional null mutant of Pacman from the Newbury lab would be expected to impact which RNAs are differentially abundant, and to what extent, as some degradation targets may utilise other compensatory degradation pathways, redundancies, or require only low levels of Pacman to degrade it.

6.6 Global overview and summary of further uses and comparisons available from Jones *et al*. wing imaginal disc RNA-seq data

The Newbury lab has worked extensively in characterising the function of Pacman in *Drosophila*, and the phenotypes caused by the loss of Pacman. As part of this, *Drosophila* wing tissues were used as a localised model to observe the impact of hypomorphic mutations. After observing the reduced size, increased apoptosis was observed in the wing imaginal discs (the developmental precursor tissue that goes on to form the wing in adult flies) of Pacman null mutant *Drosophila* larvae which were then used as input samples for an RNA-seq dataset observing the transcriptome-wide impacts of Pacman depletion. For the purposes of this project, this provided a dataset for the RNA profile of a *Drosophila* tissue from the same developmental stage as the novel dataset (L3), in both Pacman mutant and isogenic control (Figure 6.6).

6.6.1 Comparison of Pacman degraded RNAs between Jones *et al*. RNA-seq in Pacman mutant L3 wing imaginal discs and total RNA-seq from novel dataset

Previously, in Chapter 3, the lncRNAs detected in the data from Jones *et al.* were sorted by fold-change increase in the Pacman mutant samples. This provided a list of 480 lncRNAs, 156 of which were listed as “infinite upregulation”, meaning that they were detected in the mutant samples, but not in the control. Those with a calculated infinite upregulation were filtered out if they had a normalised read count of less than 1 CPM in the Pacman mutant. This left 325 lncRNAs, of which 161 had a higher read count in the Pacman mutant. A cut off of 1.5-fold increase or more was introduced, producing a list of 105 lncRNAs that showed Pacman sensitivity.

The novel sequencing data from whole L3 larval lysate in both Pacman mutant and isogenic control were treated with the same set of filters (correlation between datasets plotted in Figure 6.24), and compared (Figure 6.25), producing 49 lncRNAs with higher expression in the Pacman mutant, and 45 of these passing the 1.5-fold increase threshold. Of the lncRNAs with increased normalised read count in the absence of Pacman (161 from wing imaginal disc sequencing, 49 from novel L3 sequencing data,) 8 were common to both lists (*CR44506, CR31781, CR44347, CR42646, CR45312, CR32111, cherub*, and *CR43626*). Of the lncRNAs which increased by 1.5-fold or more (105 from wing imaginal disc sequencing, 45 from novel L3 sequencing data,) only 3 were common to both (*CR44506, CR32781*, and *CR32111*) (Figure 6.26). This is an informative approach to identify candidate lncRNAs conserved in their degradation by Pacman between the wing imaginal disc tissue and the whole L3 organism and should not be taken as either a definitive or exhaustive list, and it should be remembered that tissue-specific expression of genes is likely.

6.7 Comparison between genotypes identifies RNA species that are differentially abundant in the absence of Pacman in *Drosophila* L3 larvae

The novel RNA-seq data from total lysate of L3 larvae, both for Pacman mutant and isogenic control, allowed identification of lncRNAs are differentially abundant, and therefore likely to be degraded by Pacman, in whole L3 larvae. In order to identify the transcripts sensitive to Pacman in this data set, , the normalised read counts for both genotypes were compared. From all reads, RNAs annotated as lncRNAs were extracted and taken forward. These were then filtered to remove any genes with a normalised

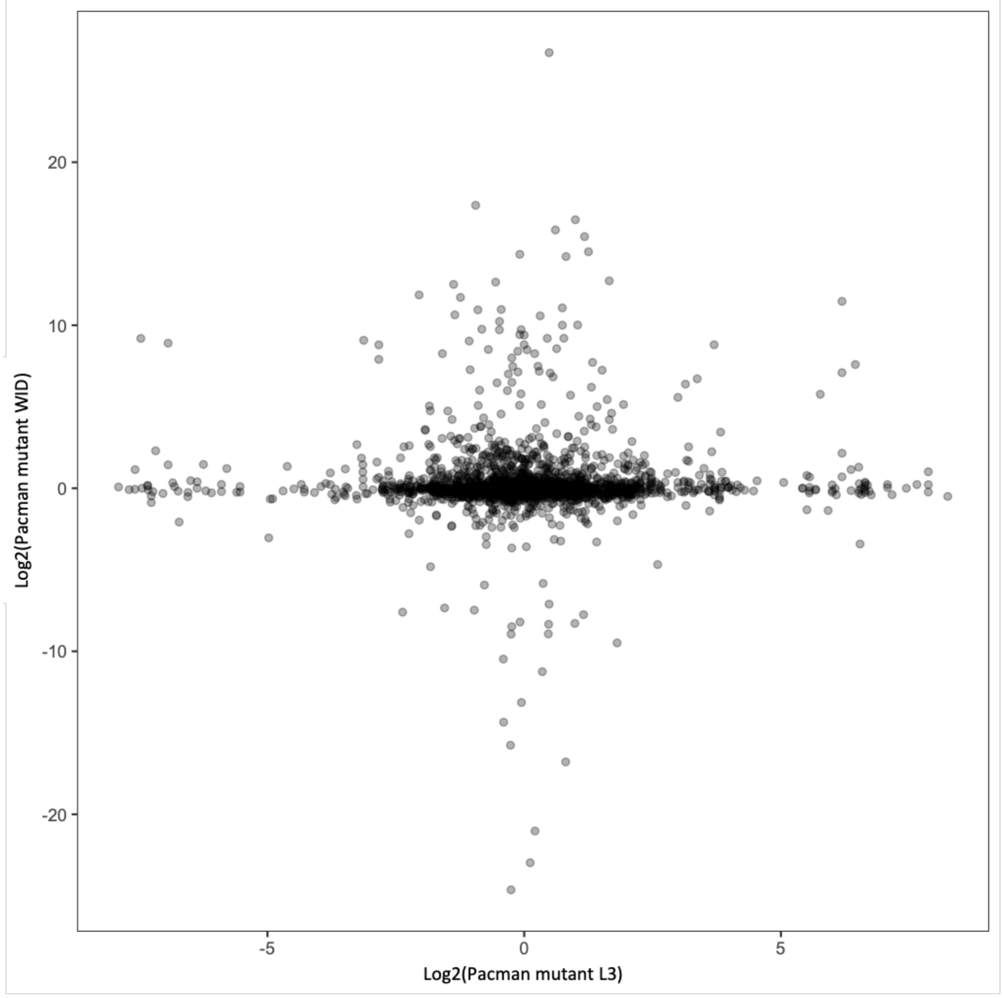


Figure 6.24 – Correlation between fold-change of RNA transcripts in the absence of Pacman in *Drosophila* L3 larvae vs L3 wing imaginal discs

A plot of a log2 of the fold change of RNAs between control and Pacman depleted samples (L3 on the x axis, WIDs on the y axis). A strong correlation would be shown by data points with equal values on the x-axis and y-axis (with many of them forming a line on x=y). This data shows virtually no correlation between data-sets, and a trend of very little fold-change in Pacman depleted WIDs (shown by a tight plotting of data points along the x-axis), although more variation than is seen previously in S2 cells. The majority of RNA targets do not change substantially in Pacman mutant L3 larvae either, as shown by data point clustering along the y-axis. The lack of x=y correlation will be due to tissue specific expression of many RNAs.

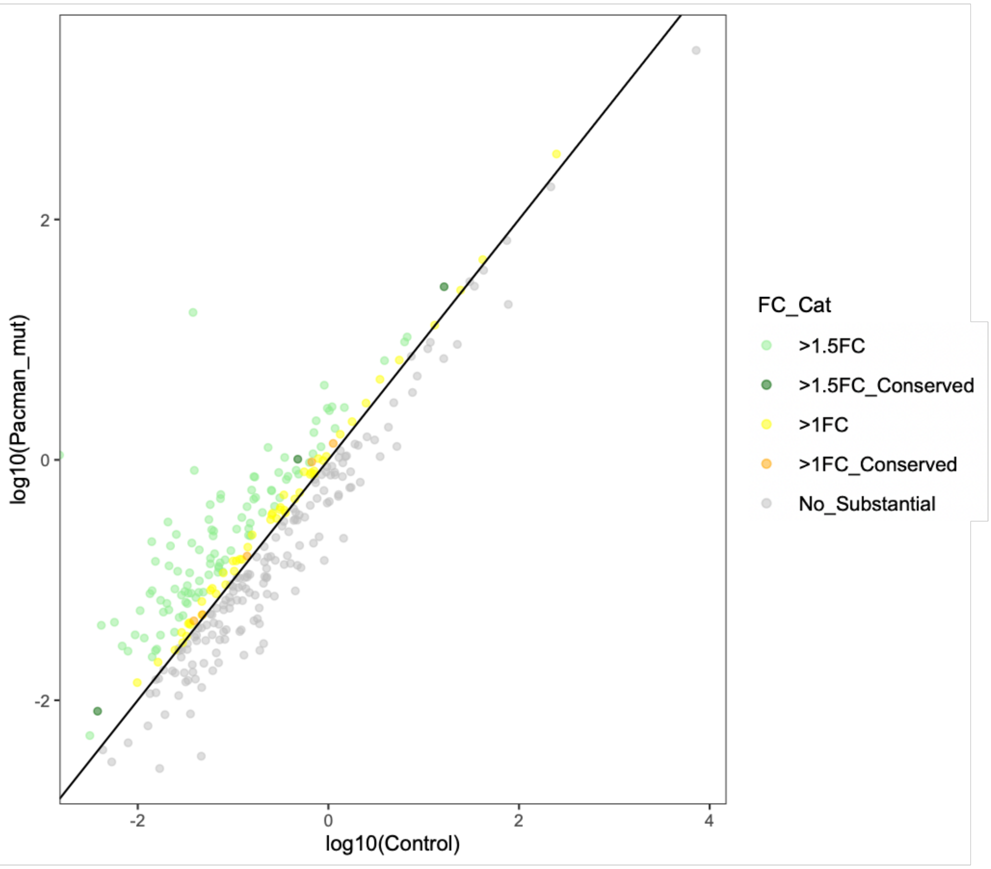
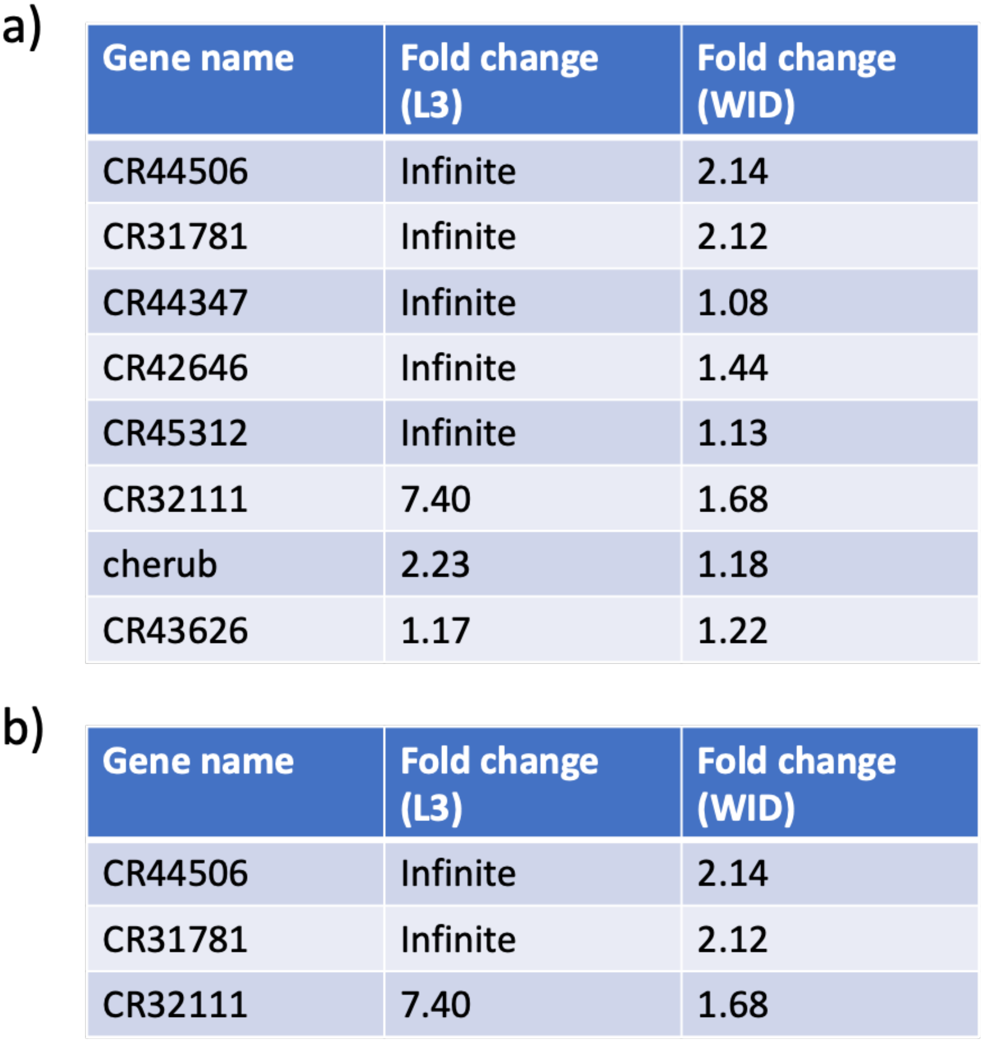


Figure 6.25 – Plot of normalised read counts of polysomally present and polysomally enriched lncRNAs in both genotypes

All 45 lncRNAs with reads detected in all L3 polysomal samples are plotted by normalised read count in control and Pacman mutant polysomal RNA. Those which were upregulated in polysomal RNA more than 1.5-fold in the absence of Pacman are highlighted in green.



a)

Gene name	Fold change (L3)	Fold change (WID)
CR44506	Infinite	2.14
CR31781	Infinite	2.12
CR44347	Infinite	1.08
CR42646	Infinite	1.44
CR45312	Infinite	1.13
CR32111	7.40	1.68
cherub	2.23	1.18
CR43626	1.17	1.22

b)

Gene name	Fold change (L3)	Fold change (WID)
CR44506	Infinite	2.14
CR31781	Infinite	2.12
CR32111	7.40	1.68

Figure 6.26 – Conserved lncRNA targets of Pacman in L3 wing imaginal discs (WIDs) and L3 larvae

Of all lncRNAs detected in RNA sequencing of total lysate from both L3 WIDs and L3 larvae, those common to both were taken and listed. After quality control steps discussed in 5.7.1, a total of 161 lncRNAs with increased normalised read count in the absence of Pacman were detected in L3 WIDs, and 49 in S2 whole L3 larvae. Of these, the 8 listed in table (a) are common to both. With a cut off of 1.5-fold increase or more, there were 161 lncRNAs with increased normalised read count in the absence of Pacman were detected in L3 WIDs, and 49 in S2 whole L3 larvae, with only 3 common to both, listed in table (b).

read count of zero in the Pacman mutant samples (as the genes of interest are those degraded by Pacman); as well as removing any duplicate gene annotations where the software had failed to reliably map reads. These were then sorted by fold-change increase in the Pacman mutant samples. This provided a shortlist of 49 lncRNAs that had an increased normalised read count. When a 1.5-fold cut off was implemented, this left 45 lncRNAs (plotted in Figure 6.27). The confidence in these candidates was too low for anything other than an informative overview, and in order to identify the more promising candidates, a more thorough filtering process must be used in future work that may follow that this thesis.

6.7.1 Implementation of stringent criteria produces a shortlist of the most interesting candidate RNAs to examine translationally active lncRNA targets of Pacman

As this thesis aims to explore the links between translation and degradation, one way in which this novel dataset can be used is in highlighting the best candidates to further understand the mechanisms of how the two processes are linked. In order to identify potential target lncRNAs, associated with the polysome and also degraded by Pacman, a more stringent approach to compiling a shortlist can be helpful. Although the depth of coverage provided by this poly-ribo-seq is not as substantial as would be needed to provide the most comprehensive overview of Pacman degraded lncRNAs and their association with the polysome, it does limit the number of candidates from the start to only those which were abundant enough to be detected at this limited depth. The FeatureCounts normalised CPM were taken for all replicates of the isogenic control L3, as well as the Pacman null mutant L3. From these, only the lncRNA were selected and taken forward. In order to ensure that the lncRNA can be found on the polysome, individual replicates were examined, and only those with reads in every polysomal RNA replicate were taken forward, leaving 45 lncRNAs (plotted in Figure 6.28). In order to identify those that are targets of Pacman degradation (potentially even occurring on the polysome,) it was decided that those which increased in polysomal abundance in the absence of Pacman would be taken forward. The lncRNAs were, to this end, sorted by fold-change between Pacman mutant and isogenic control. Those with an increase of greater than 1.5-fold were selected and taken forward, leaving 7 lncRNAs of interest

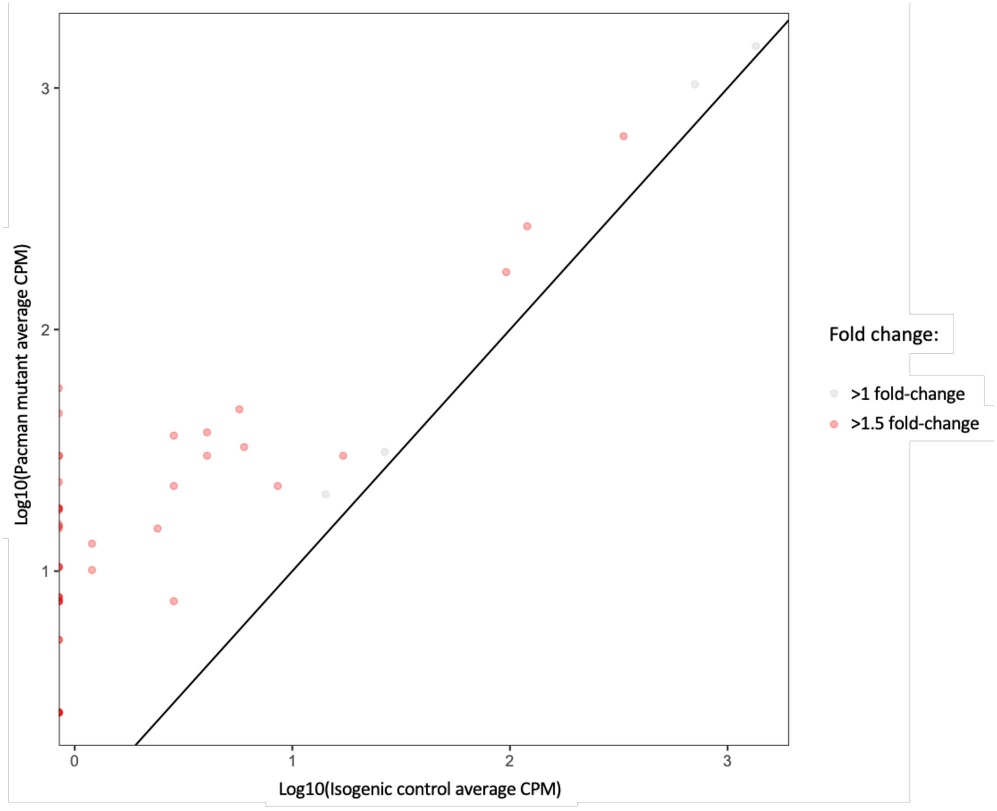


Figure 6.27 – Plot of normalised read counts of lncRNAs from total L3 lysate in Pacman mutant versus control genotype

All 49 lncRNAs passing the selection criteria described in 5.8 in Pacman mutant total L3 lysate are plotted by normalised read count, in control and Pacman mutant genotypes. The 45 which were upregulated in the absence of Pacman more than 1.5-fold are highlighted in red.

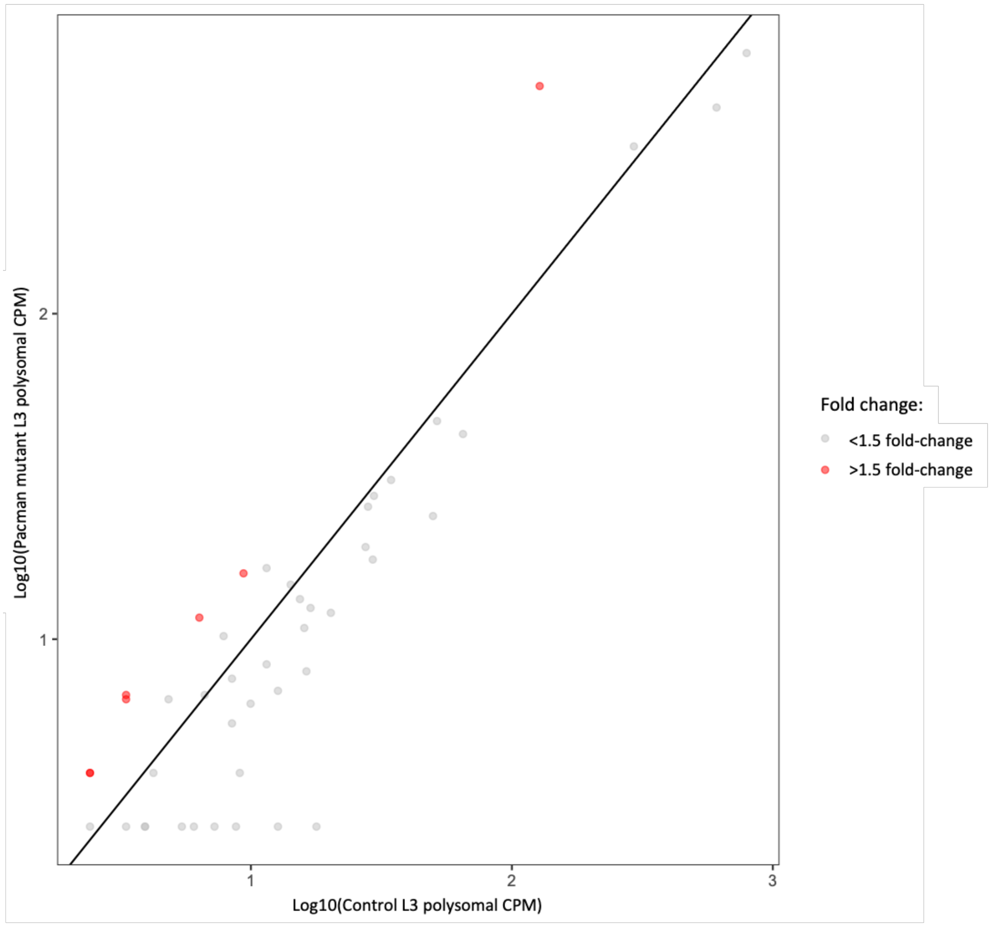


Figure 6.28 – Plot of normalised read counts of polysomally present and polysomally enriched lncRNAs in both genotypes

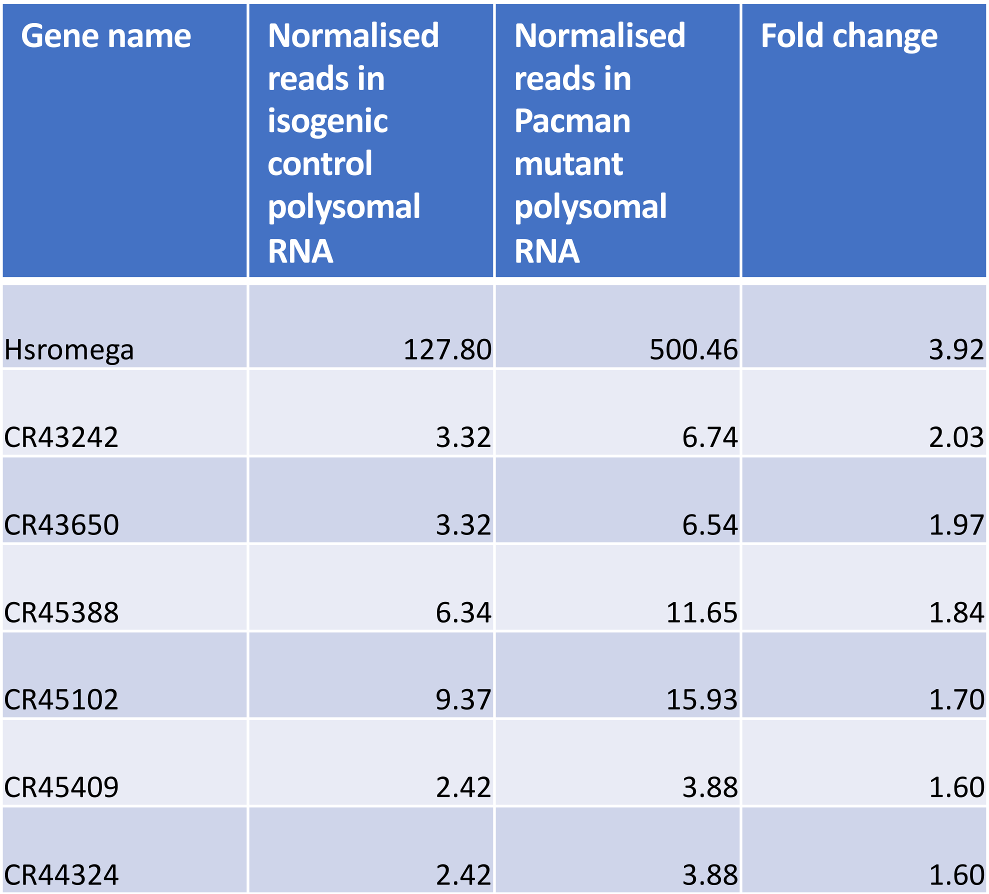
All 45 lncRNAs with reads detected in all L3 polysomal samples are plotted by normalised read count in control and Pacman mutant polysomal RNA. Those which were upregulated in polysomal RNA more than 1.5-fold in the absence of Pacman are highlighted in red.

(*HsrOmega, CR43242, CR43650, CR45388, CR45102, CR45409,* and *CR44324* Figures 6.28 and 6.29).

6.7.2 GFP reporter fails to show likely translation of a novel ORF in lncRNA HsrOmega

Whilst the informative candidate tables and comparisons produced by the multiple tables are valid ways of observing trends, and highlighting targets for future work, a biologically tested example allows the work in this thesis to be proven effective in a more practical sense. From the shortlist of targets to investigate the interplay between translation and Pacman degradation, the top candidate, *HsrOmega* (or *Hsrω*) was taken for further experimental testing. *HsrOmega* was determined as the top candidate, as it underwent the largest increase in abundance of polysomal RNA in the absence of Pacman (with a 3.92-fold increase), in addition to being the most abundant on lncRNA on the polysome in either condition. Although we can be confident of polysomal association from the multiple poly-ribo-seq and ribo-seq datasets in which it appears, whether or not it is actively translating is harder to be sure of. The fact that *HsrOmega* is present as translated in data from S2 cells (including with Harringtonine treatment), and from embryonic stages E1, E2, and E3 increases confidence beyond what we could achieve from this data alone. In order to examine the translational activity of *HsrOmega* more thoroughly, a biological profile was needed. Known information on the gene was gathered from FlyBase. *HsrOmega* is annotated as a long non-coding RNA, a gene on chromosome 3, with a length of 21,710 base pairs.

The expression profile (Figure 6.30) from catalogued sequencing data shows very low to low expression in all tissues (very low in L3 WID, and low in late L3), other than ovaries and testes in 4-day old *Drosophila*; and present in the organism at all timepoints after 2 hours. The lncRNA is also expressed to some extent in all commonly used cell lines. Speculative biological processes, supported by some experimental evidence are also recorded (Figure 6.31), alongside preliminary work on the cellular localisation of the lncRNA, with presence on the chromatin and the omega speckle inferred from direct assay.



Gene name	Normalised reads in isogenic control polysomal RNA	Normalised reads in Pacman mutant polysomal RNA	Fold change
Hsromega	127.80	500.46	3.92
CR43242	3.32	6.74	2.03
CR43650	3.32	6.54	1.97
CR45388	6.34	11.65	1.84
CR45102	9.37	15.93	1.70
CR45409	2.42	3.88	1.60
CR44324	2.42	3.88	1.60

Figure 6.29 – A table listing polysome associated lncRNAs that increase in polysomal abundance in the absence of Pacman

Of all lncRNAs detected in RNA sequencing of polysome fractions from both isogenic control and Pacman null mutant L3 larvae, those with reads detected in every replicate (implying a detectable background level polysomal association and possible translation) were taken, and filtered to select only those with an increase of 1.5-fold or more in the absence of Pacman. All 7 detected candidates are listed above.

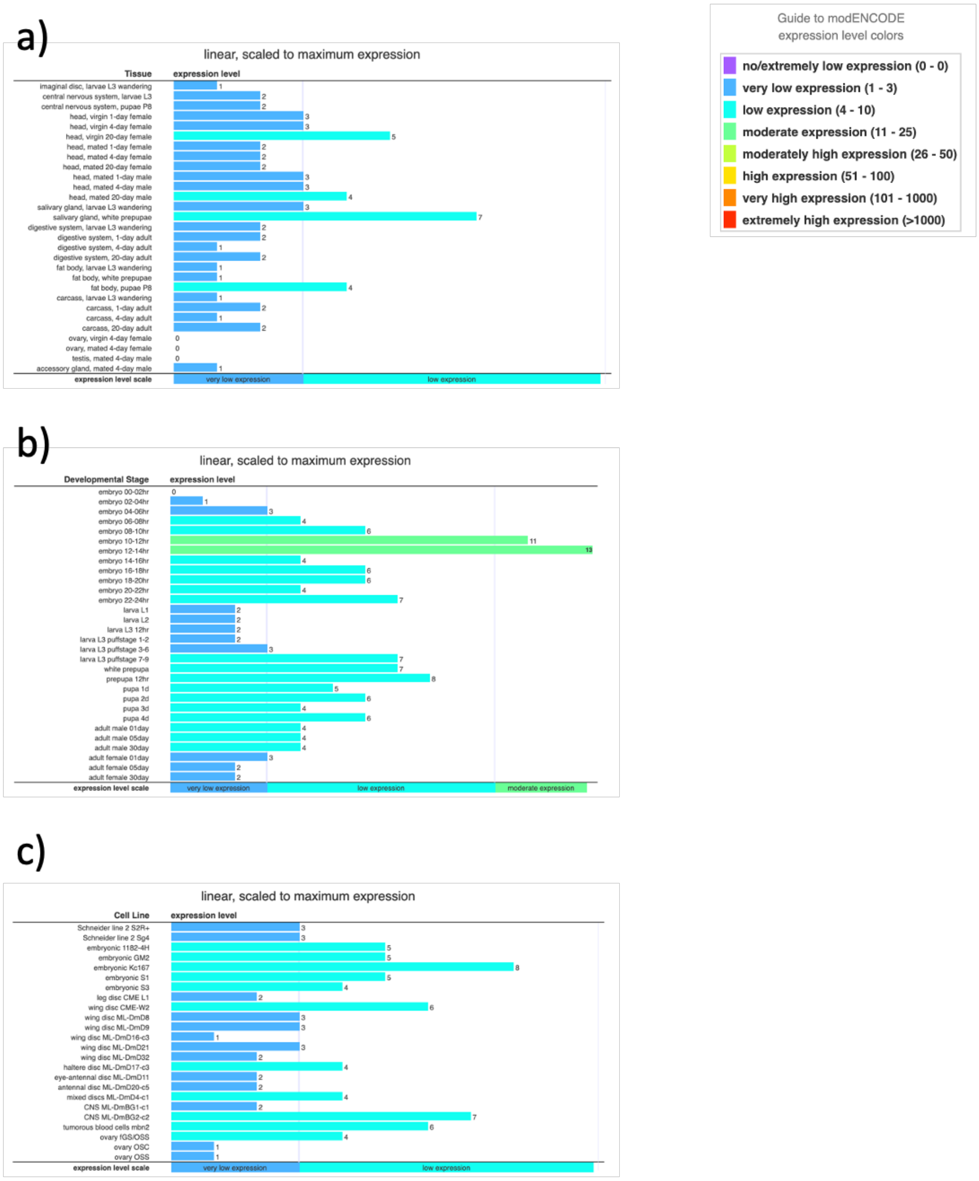
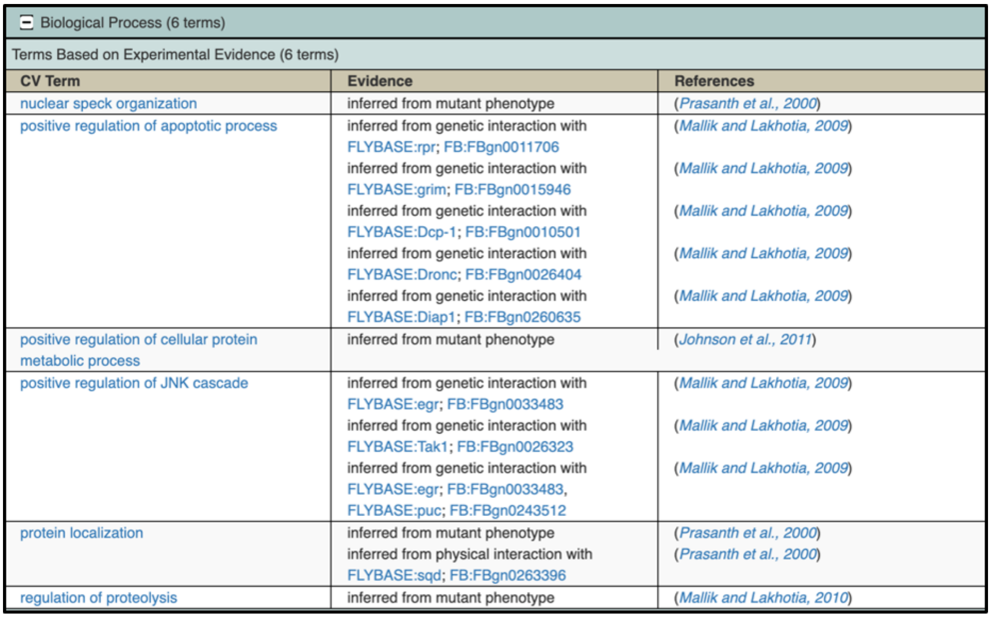


Figure 6.30 – Expression profiles from catalogued RNA-seq data for the lncRNA HsrOmega

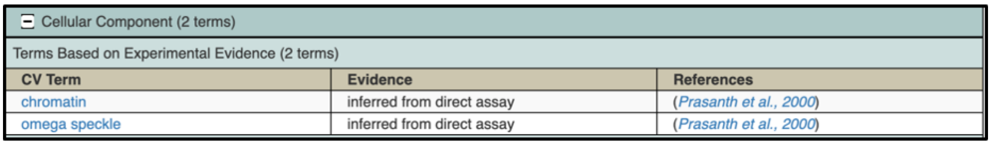
Expression level is shown through developing and developed tissues (a), developmental timepoints (b), and commonly used and available *Drosophila* cell lines (c). Diagram adapted from FlyBase expression data.



⊟ Biological Process (6 terms)

Terms Based on Experimental Evidence (6 terms)

CV Term	Evidence	References
nuclear speck organization	inferred from mutant phenotype	(*Prasanth et al., 2000*)
positive regulation of apoptotic process	inferred from genetic interaction with FLYBASE:rpr; FB:FBgn0011706 inferred from genetic interaction with FLYBASE:grim; FB:FBgn0015946 inferred from genetic interaction with FLYBASE:Dcp-1; FB:FBgn0010501 inferred from genetic interaction with FLYBASE:Dronc; FB:FBgn0026404 inferred from genetic interaction with FLYBASE:Diap1; FB:FBgn0260635	(*Mallik and Lakhotia, 2009*) (*Mallik and Lakhotia, 2009*) (*Mallik and Lakhotia, 2009*) (*Mallik and Lakhotia, 2009*) (*Mallik and Lakhotia, 2009*)
positive regulation of cellular protein metabolic process	inferred from mutant phenotype	(*Johnson et al., 2011*)
positive regulation of JNK cascade	inferred from genetic interaction with FLYBASE:egr; FB:FBgn0033483 inferred from genetic interaction with FLYBASE:Tak1; FB:FBgn0026323 inferred from genetic interaction with FLYBASE:egr; FB:FBgn0033483, FLYBASE:puc; FB:FBgn0243512	(*Mallik and Lakhotia, 2009*) (*Mallik and Lakhotia, 2009*) (*Mallik and Lakhotia, 2009*)
protein localization	inferred from mutant phenotype inferred from physical interaction with FLYBASE:sqd; FB:FBgn0263396	(*Prasanth et al., 2000*) (*Prasanth et al., 2000*)
regulation of proteolysis	inferred from mutant phenotype	(*Mallik and Lakhotia, 2010*)



⊟ Cellular Component (2 terms)

Terms Based on Experimental Evidence (2 terms)

CV Term	Evidence	References
chromatin	inferred from direct assay	(*Prasanth et al., 2000*)
omega speckle	inferred from direct assay	(*Prasanth et al., 2000*)

Figure 6.31 – Preliminary and speculated biological roles and localisation of the lncRNA HsrOmega

Speculative roles and cellular compartmentalisation of the candidate lncRNA HsrOmega. Adapted from FlyBase Gene Ontology Annotations.

IGV was used to view localisation of read mapping in the novel poly-ribo-seq data, and visually detect open reading frames. Due to the nature of the polysome footprinting step of poly-ribo-seq, the sequenced RNA fragments are from regions where ribosomes bound, protecting them from RNase digest. Examining the read mapping allows visualisation of read pile up locations (Figure 6.32), illuminating likely gene regions to identify any potential open reading frames (as the ribosome would have been bound on the regions the reads map to). Promising regions can then be examined by amino acid sequence, in order to identify likely start and stop codons that fit the ribosome bound region. A promising ORF was identified (Figure 6.32) and using Gateway cloning (summarised in Figure 6.33), a construct was made with a GFP encoding region downstream of and in-frame of the potential ORF. This was then transfected into S2 cells and observed by microscopy to check for GFP fluorescence. Neither conventional optical microscopy (using Zeiss Axio Imager 2) nor confocal microscopy using a Lecia SP8 Microscope observed GFP fluorescence (Figure 6.34). Further examination of the gene region and comparison with the elongation-inhibited data from Zhang *et al*. identified multiple other potential ORFs which might be translated within the gene region (Figure 6.35). These must be followed up in order to definitively show whether the gene has translational activity; although its conservation in polysomal RNA throughout multiple datasets, as well as identification of potential ORFs, support the translation of the gene.

6.8 Global overview and summary of Towler *et al*. wing imaginal disc RNA-seq data

In addition to exploring the molecular functions of Pacman, the Newbury lab, have also worked extensively in characterising the functions of Dis3L2 in *Drosophila*, the pathways it regulates, and the phenotypes caused by the loss of Dis3L2 function. As part of these studies, *Drosophila* wing tissues were used as a localised model to observe the impact of mutations. After observing an increased growth phenotype in this tissue (contrasting the reduced size seen in Pacman mutants), the wing imaginal discs of Dis3L2 null mutant *Drosophila* larvae were used as input samples for a novel RNA-seq dataset observing the transcriptome-wide impacts of Dis3L2 null mutation. For the purposes of this project, this provided a dataset for the RNA profile of a *Drosophila* tissue from the same developmental stage as the novel dataset (L3), in both Dis3L2 mutant and isogenic control.

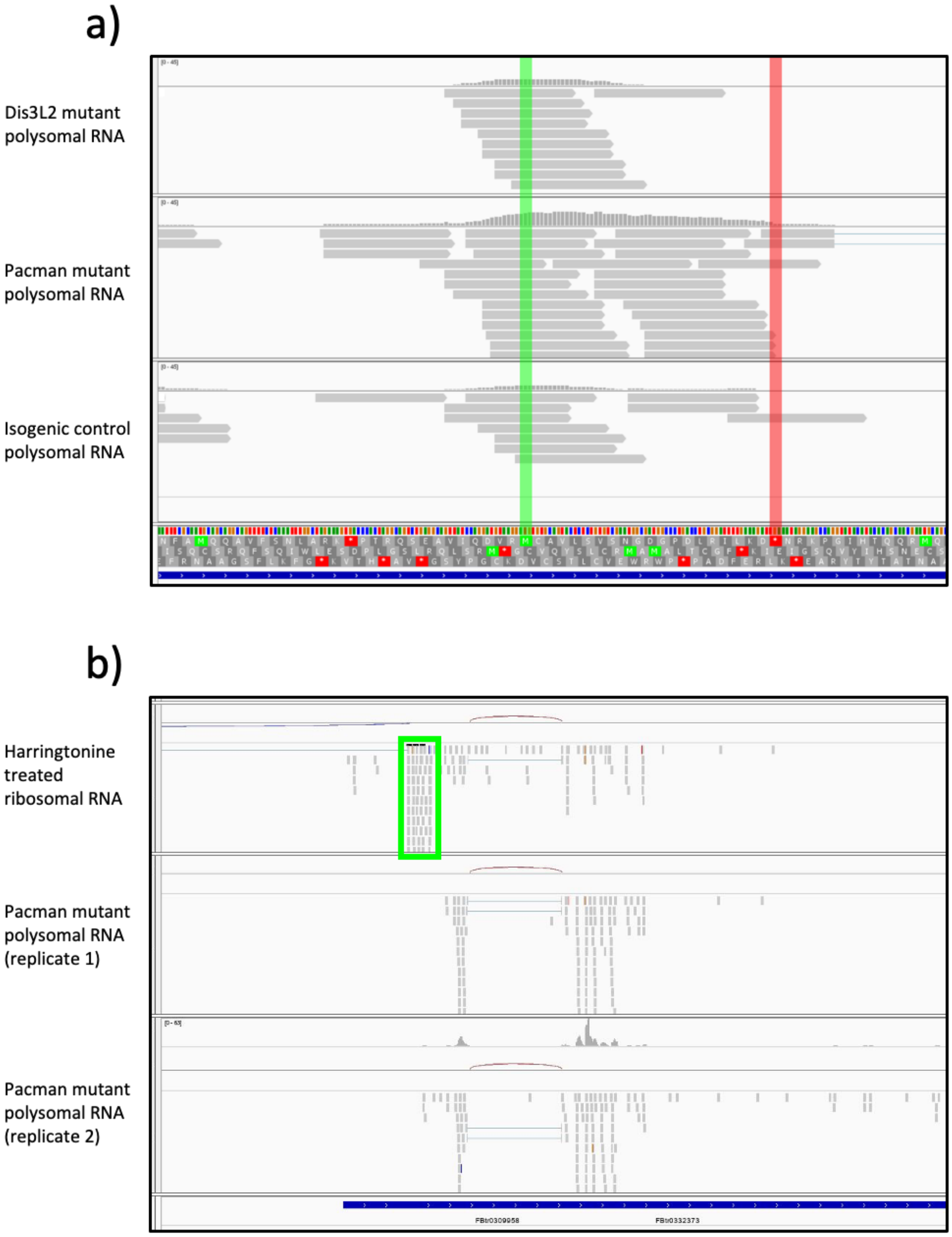
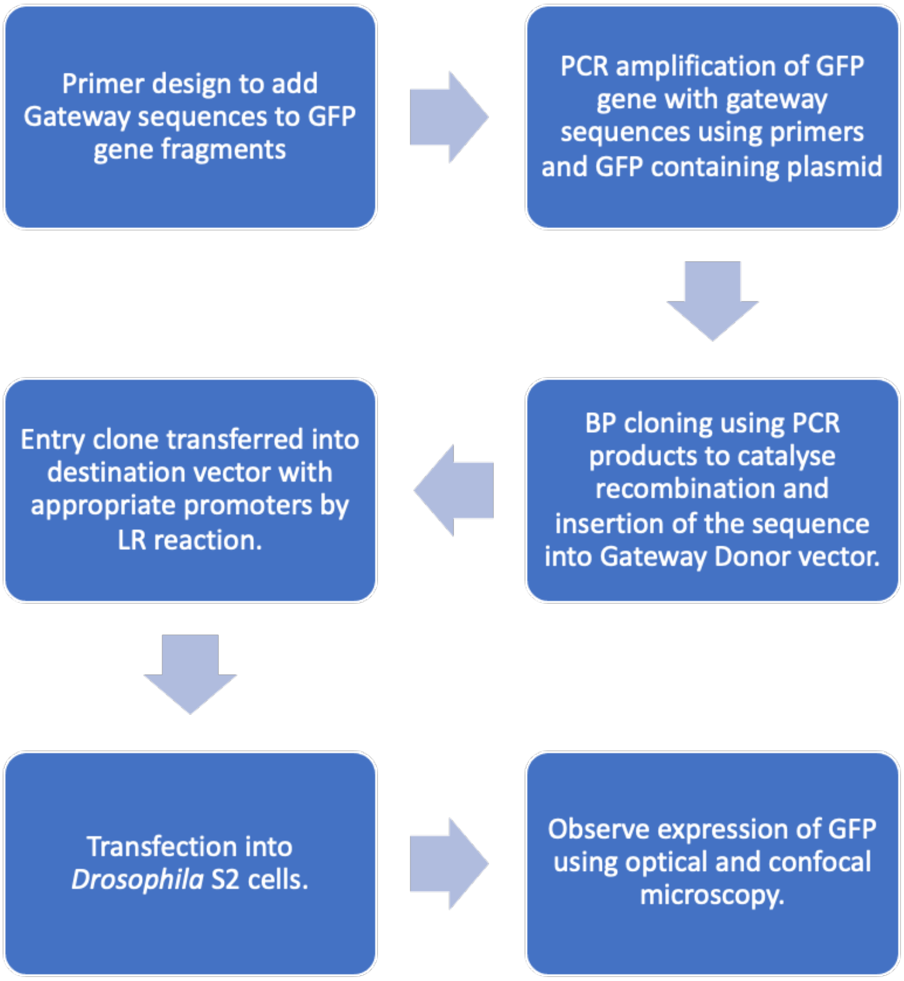

a)
Dis3L2 mutant polysomal RNA
Pacman mutant polysomal RNA
Isogenic control polysomal RNA
[0 - 45]
b)
Harringtonine treated ribosomal RNA
Pacman mutant polysomal RNA (replicate 1)
Pacman mutant polysomal RNA (replicate 2)
[0 - 63]
FBtr0309958
FBtr0332373

Figure 6.32 – Promising potential ORFs identified in HsrOmega viewed in IGV

From the novel poly-ribo-seq data, a promising potential ORF was identified (panel (a)). The start codon is highlighted in green, and the stop codon in red. Subsequent analysis of the data from Zhang *et al*. (panel (b)) identified an alternative, and likely more promising potential location of read pile-ups (outlined in green). However, by this point, the follow-up cloning experiments were already in progress for the ORF highlighted in panel (a).



Primer design to add Gateway sequences to GFP gene fragments

PCR amplification of GFP gene with gateway sequences using primers and GFP containing plasmid

BP cloning using PCR products to catalyse recombination and insertion of the sequence into Gateway Donor vector.

Entry clone transferred into destination vector with appropriate promoters by LR reaction.

Transfection into *Drosophila* S2 cells.

Observe expression of GFP using optical and confocal microscopy.

Figure 6.33 – Workflow of gateway cloning

Gateway cloning is a fast and straightforward way to test a potential ORF, the workflow is summarised here.

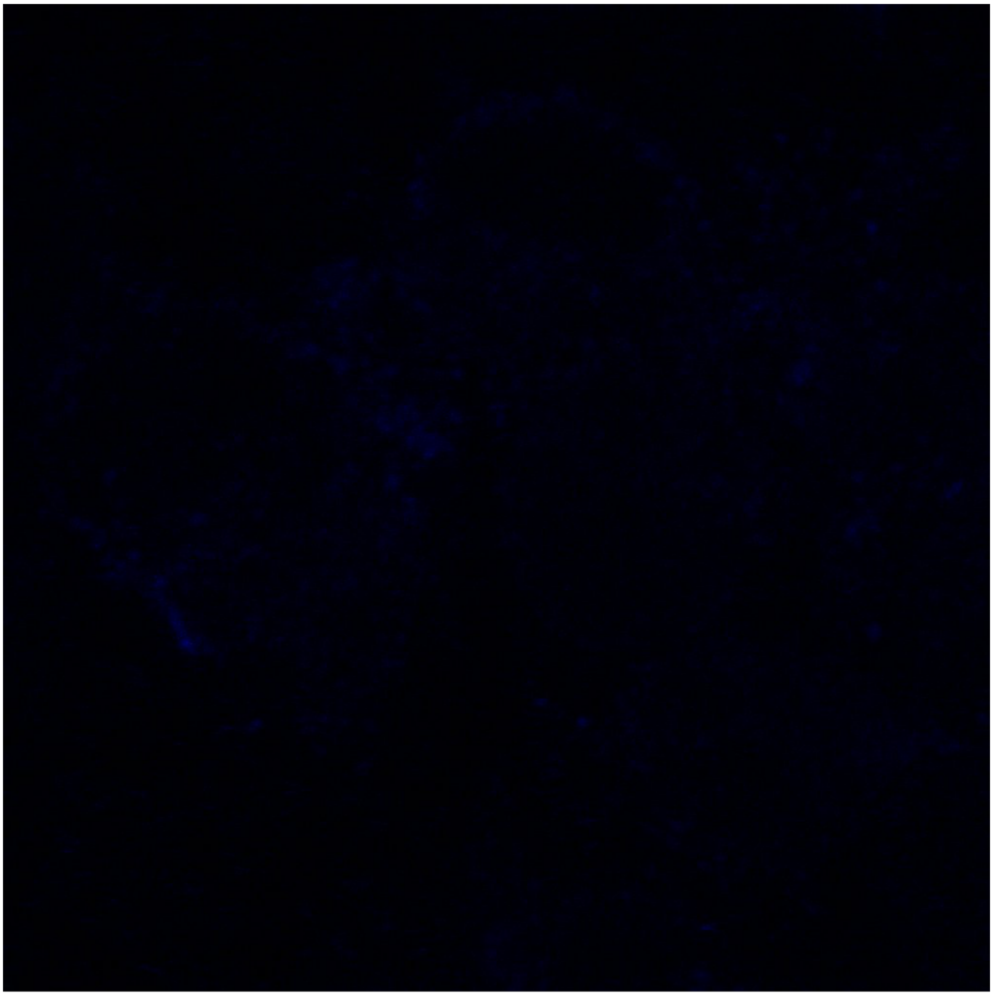


Figure 6.34 – Confocal microscopy image showing a complete lack of GFP signal.

This field of view, showing no GFP signal, was representative of all areas imaged using the Leica SP-8 confocal microscope. The blue fluorescence shows DAPI signal, indicating the presence of S2 cells, despite the lack of GFP signal to indicate transfected cells.

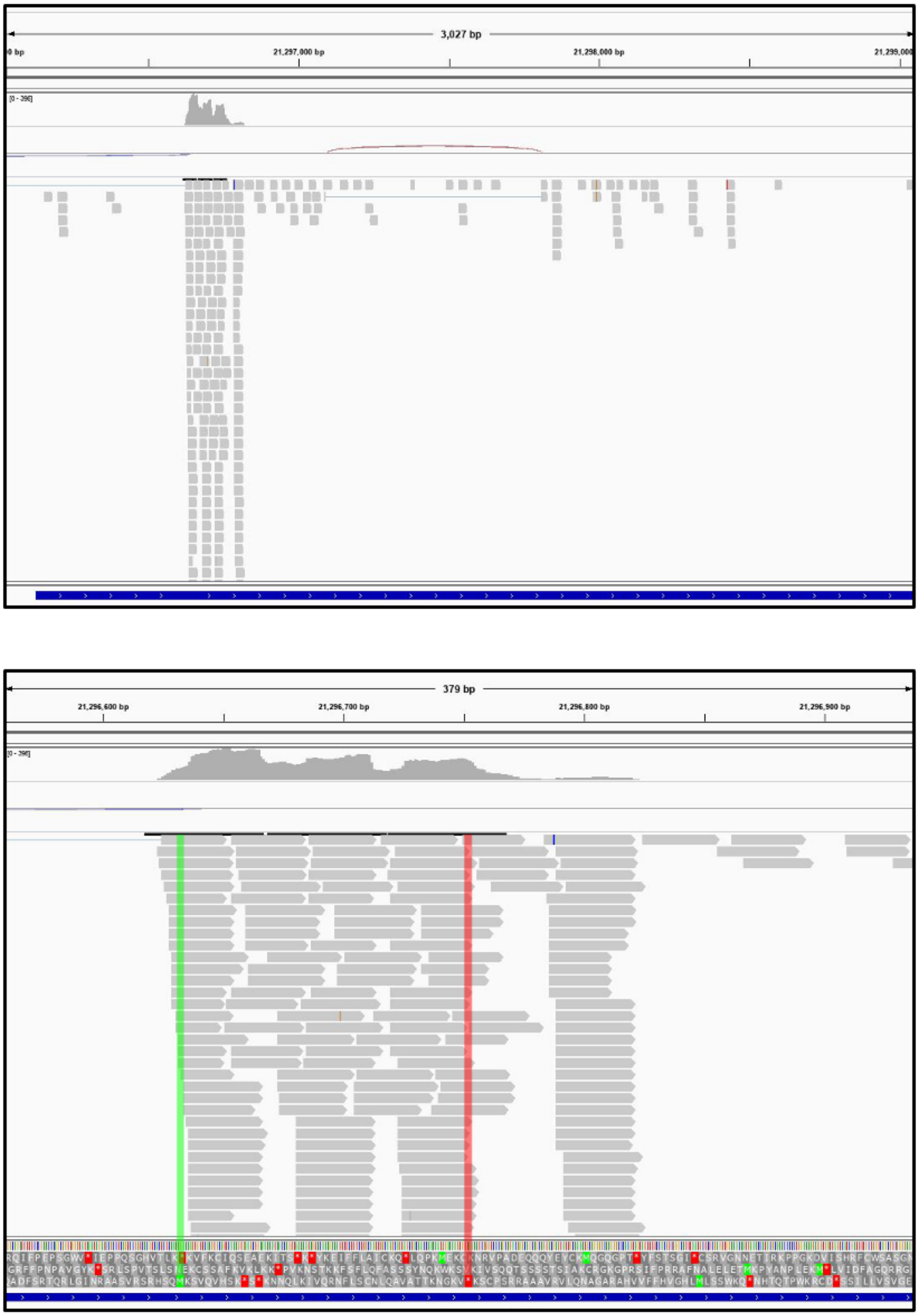


Figure 6.35 – Alternative ORFs identified in the data from Zhang *et al*. using IGV

The point with the most pile-up (most protected RNA frangements mapping to that region) can be clearly seen in panel (a). In panel (b), the viewer has been zoomed in, and a potential ORF has been highlighted (start codon in green, and stop codon in red). Although there are stop codons between those that are highlighted, the may be read through, or a splicing event may prevent them from ending translation.

6.8.1 Comparison of Dis3L2 degraded RNAs between Towler *et al*. RNA-seq in Dis3L2 mutant L3 wing imaginal discs and total RNA-seq from novel dataset

As with the previous analysis from the Newbury lab RNA-seq data, the .bam files from this RNA-seq experiment were run through FeatureCounts to produce read count and differential abundance data. From this, RNAs annotated as lncRNAs were extracted and taken forward. These were then filtered to remove any genes with a normalised read count of zero in the Dis3L2 mutant samples (as the genes of interest are those degraded by Dis3L2); as well as removing any duplicate gene annotations where the software had failed to reliably map reads. These were then sorted by fold-change increase in the Dis3L2 mutant samples. This provided a list of 195 lncRNAs, 71 of which were listed as “infinite upregulation”. In order to filter out the false infinite upregulation calculations due to very low expression levels, those with a calculated infinite upregulation were filtered out if they had a normalised read count of less than 1 CPM in the Dis3L2 mutant. This left 129 lncRNAs, of which 56 had a higher read count in the Dis3L2 mutant than in the isogenic control. A cut off of 1.5-fold increase or more was introduced, producing a list of 40 potential Dis3L2 degraded lncRNAs. The novel sequencing data from whole L3 larval lysate in both Dis3L2 mutant and isogenic control were treated with the same set of filters (data correlation plotted in Figure 6.36), producing 55 lncRNAs with higher expression in the Dis3L2 mutant, and 48 of these passing the 1.5-fold increase threshold. Of the lncRNAs with increased normalised read count in the absence of Dis3L2 (56 from wing imaginal disc sequencing, 55 from novel L3 sequencing data,) 3 were common to both lists (CR42646, CR45517, CR33938). Of the lncRNAs which increased by 1.5-fold or more (40 from wing imaginal disc sequencing, 48 from novel L3 sequencing data,) only 1 was common to both (CR45517) (Figure 6.36). This is an informative approach to identify candidate lncRNAs conserved in their degradation by Dis3L2 between the wing imaginal disc tissue and the whole L3 organism (Figure 6.37) but should not be taken as either a definitive or exhaustive list, and it should be remembered that tissue-specific expression of genes is likely.

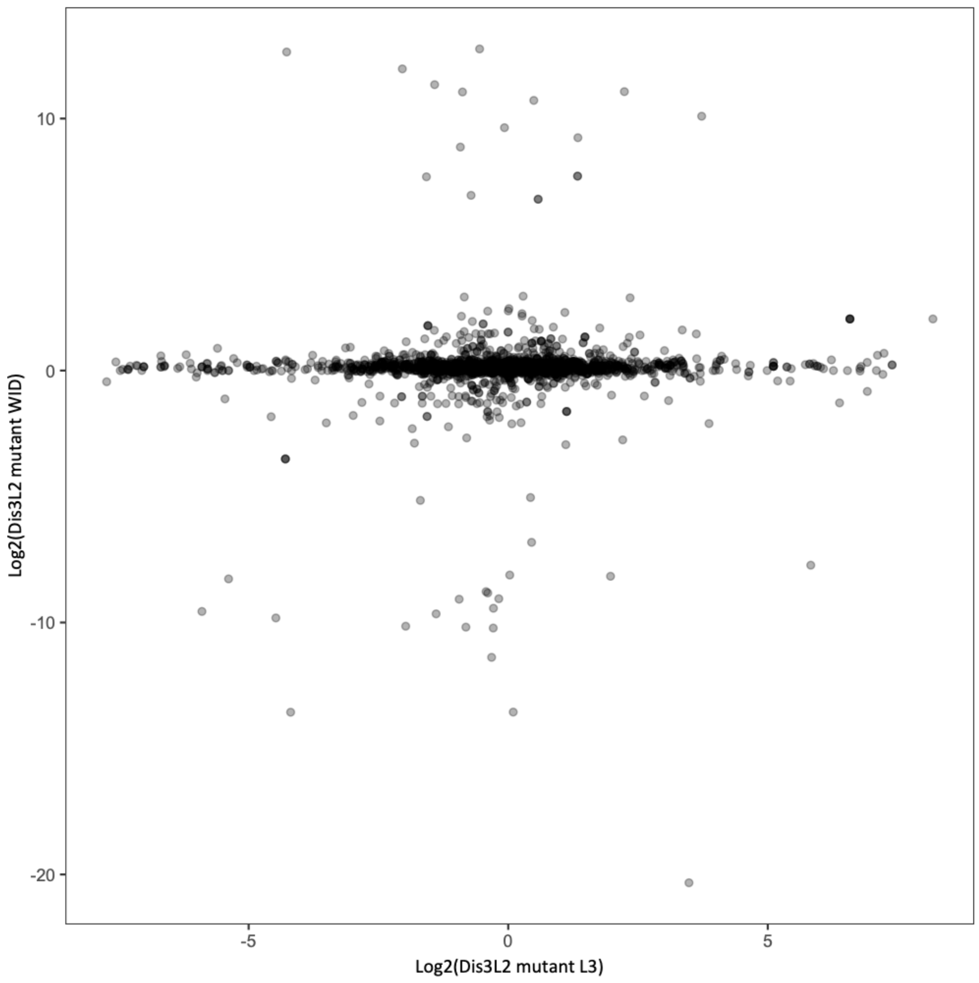
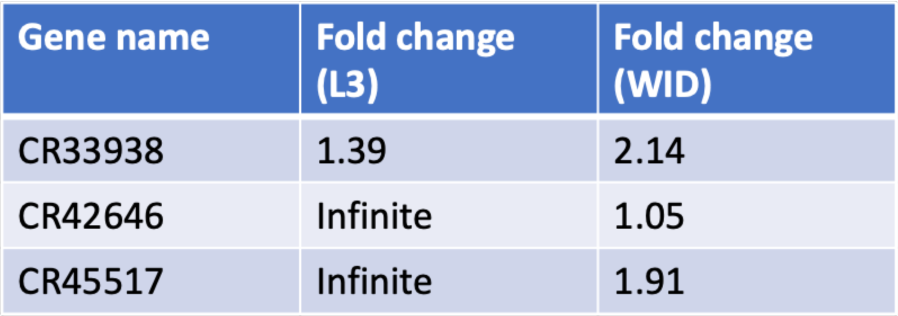


Figure 6.36 – Correlation between fold-change of RNA transcripts in the absence of Dis3L2 in *Drosophila* L3 larvae vs L3 wing imaginal discs

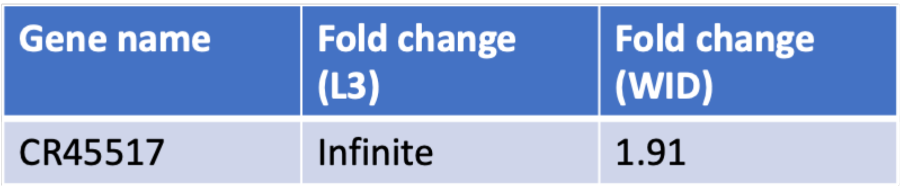
A plot of a log2 of the fold change of RNAs between control and Dis3L2 depleted samples (L3 on the x axis, WIDs on the y axis). A strong correlation would be shown by data points with equal values on the x-axis and y-axis (with many of them forming a line on x=y). This data shows virtually no correlation between data-sets, and a trend of very little fold-change in Dis3L2 mutant WIDs (shown by a tight plotting of data points along the x-axis. There is still a trend for RNAs to not change substantially in Dis3L2 mutant L3 larve either, as shown by some data point clustering along the y-axis. The lack of x=y correlation will be due to tissue specific expression of many RNAs.

a)



Gene name	Fold change (L3)	Fold change (WID)
CR33938	1.39	2.14
CR42646	Infinite	1.05
CR45517	Infinite	1.91

b)



Gene name	Fold change (L3)	Fold change (WID)
CR45517	Infinite	1.91

Figure 6.37 – Conserved potential lncRNA targets of Dis3L2 in L3 wing imaginal discs (WIDs) and L3 larvae

Of all lncRNAs detected in RNA sequencing of total lysate from both L3 WIDs and L3 larvae, those common to both were taken and listed. After quality control steps discussed in 5.9.1, a total of 56 lncRNAs with increased normalised read count in the absence of Dis3L2 were detected in L3 WIDs, and 55 in S2 whole L3 larvae. Of these, the 3 listed in table (a) are common to both. With a cut off of 1.5-fold increase or more, there were 40 lncRNAs with increased normalised read count in the absence of Dis3L2 were detected in L3 WIDs, and 48 in S2 whole L3 larvae, with only 1 common to both, listed in table (b).

6.9 Comparison between genotypes identifies RNA species that are differentially abundant in the absence of Dis3L2 in *Drosophila* L3 larvae

The novel RNA-seq data from total lysate of L3 larvae, both for Dis3L2 mutant and isogenic control, allowed analysis of the lncRNAs which are differentially abundant, and therefore likely to be degraded by Dis3L2, in whole L3 larvae. In order to identify the transcripts sensitive to Dis3L2 in this data set, the normalised read counts for both genotypes were compared. From all reads, RNAs annotated as lncRNAs were extracted and taken forward. These were then filtered to remove any genes with a normalised read count of zero in the Dis3L2 mutant samples (as the genes of interest are those degraded by Dis3L2); as well as removing any duplicate gene annotations where the software had failed to reliably map reads. These were then sorted by fold-change increase in the Dis3L2 mutant samples. This provided a shortlist of 55 lncRNAs that had an increased normalised read count. When a 1.5-fold cut off was implemented, this left 48 lncRNAs (Figure 6.38). The confidence in these candidates was too low for anything other than an informative overview, and in order to identify the more promising candidates, a more thorough filtering process must be used.

6.9.1 Implementation of stringent criteria produces a shortlist of the most interesting candidate RNAs to examine translationally active lncRNA targets of Dis3L2

Similar to the strategy discussed in 5.7.4, in order to identify potential target lncRNAs associated with the polysome and also degraded by Dis3L2, a more stringent approach to compiling a shortlist can be helpful. Although the depth of coverage provided by this poly-ribo-seq is not as substantial as would be needed to provide the most comprehensive overview of Dis3L2 degraded lncRNAs and their association with the polysome, it does limit the number of candidates from the start to only those which were abundant enough to be detected at this limited depth.

The FeatureCounts normalised CPM were taken for all replicates of the isogenic control L3, as well as the Dis3L2 functional mutant L3. From these, only the lncRNA were selected and taken forward. In order to ensure that the lncRNA can be found on the

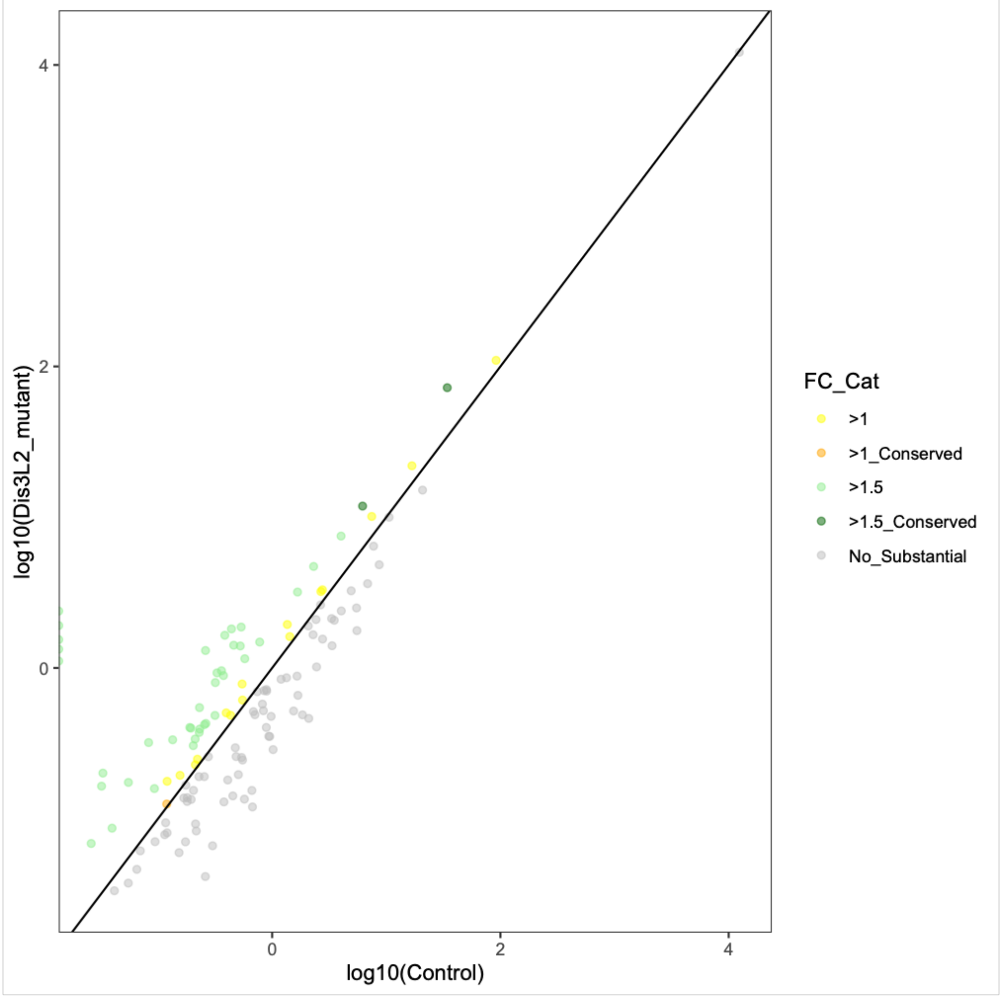


Figure 6.38 – Plot of normalised read counts of polysomally present and polysomally enriched lncRNAs in both genotypes

All 45 lncRNAs with reads detected in all L3 polysomal samples are plotted by normalised read count in control and Pacman mutant polysomal RNA. Those which were upregulated in polysomal RNA more than 1.5-fold in the absence of Pacman are highlighted in green.

polysome, individual replicates were examined, and only those with reads in every polysomal RNA replicate were taken forward, leaving 41 lncRNAs (Figure 6.39). In order to identify those that are targets of Dis3L2 degradation (potentially even occurring on the polysome,) it was decided that those which increased in polysomal abundance in the absence of Dis3L2 would be taken forward. The lncRNAs were, to this end, sorted by fold-change between Dis3L2 mutant and isogenic control. Those with an increase of greater than 1.5-fold were selected and taken forward, leaving 13 lncRNAs (Figure 6.40 and 6.41). Interestingly, of these 13 Dis3L2-dependent lncRNAs found in the polysomal RNA of whole L3 larvae, 3 lncRNAs (*CR40469*, *CR42839*, and *flam*) are also found in the polysomal RNA of all embryonic stages (E1, E2, and E3); while 1 (*flam*) is also found in the polysomal RNA of S2 cells (with *flam* being common to embryonic polysomal, S2 polysomal, and L3 polysomal RNA, as well as being more than 1.5-fold upregulated in the absence of Dis3L2).

6.9.2 GFP reporter indicates likely translation of a novel ORF in lncRNA CR40469

Whilst the informative candidate tables and comparisons produced by the multiple tables are valid ways of observing trends, and highlighting targets for future work, a biologically tested example allows the work in this thesis to be proven effective in a more practical sense. From the shortlist of targets to investigate the interplay between translation and Dis3L2 degradation, the top candidate, CR40469 was taken for further experimental analysis. CR40469 was determined as the top candidate, as it underwent the largest increase in abundance of polysomal RNA in the absence of Dis3L2 (with a 4.81-fold increase). Although we can be confident of polysomal association from the multiple poly-ribo-seq and ribo-seq datasets in which it appears, whether or not it is actively translating is harder to be sure of. In order to examine the translational activity of CR40469 more thoroughly, a biological profile was needed. Known information on the gene was gathered from FlyBase. CR40469 is annotated as a long non-coding RNA, a gene on the X chromosome, with a length of 214 base pairs. Subsequent work, building on the experiments of this thesis may also wish to examine *CR33948*, as the most abundant of those differentially expressed in the absence of Dis3L2.

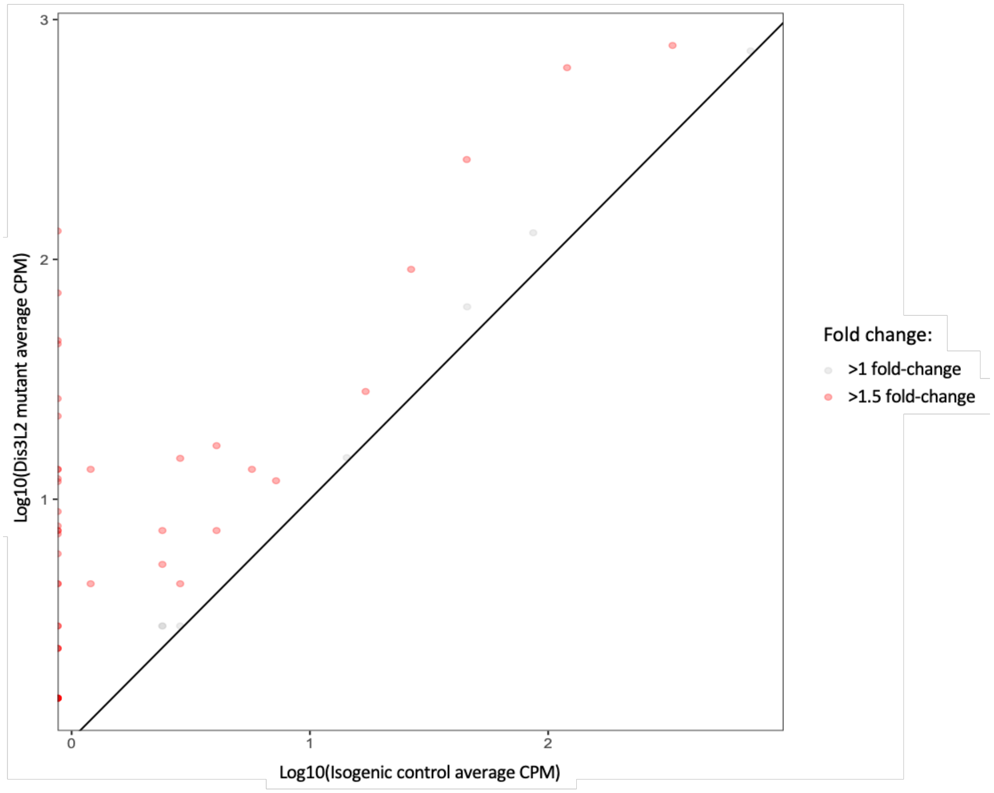


Figure 6.39 – Plot of normalised read counts of lncRNAs from total L3 lysate in Dis3L2 mutant versus control genotype

All 55 lncRNAs passing the selection criteria described in 5.10 in Dis3L2 mutant total L3 lysate are plotted by normalised read count, in control and Dis3L2 mutant genotypes. The 48 which were upregulated in the absence of Dis3L2 more than 1.5-fold are highlighted in red.

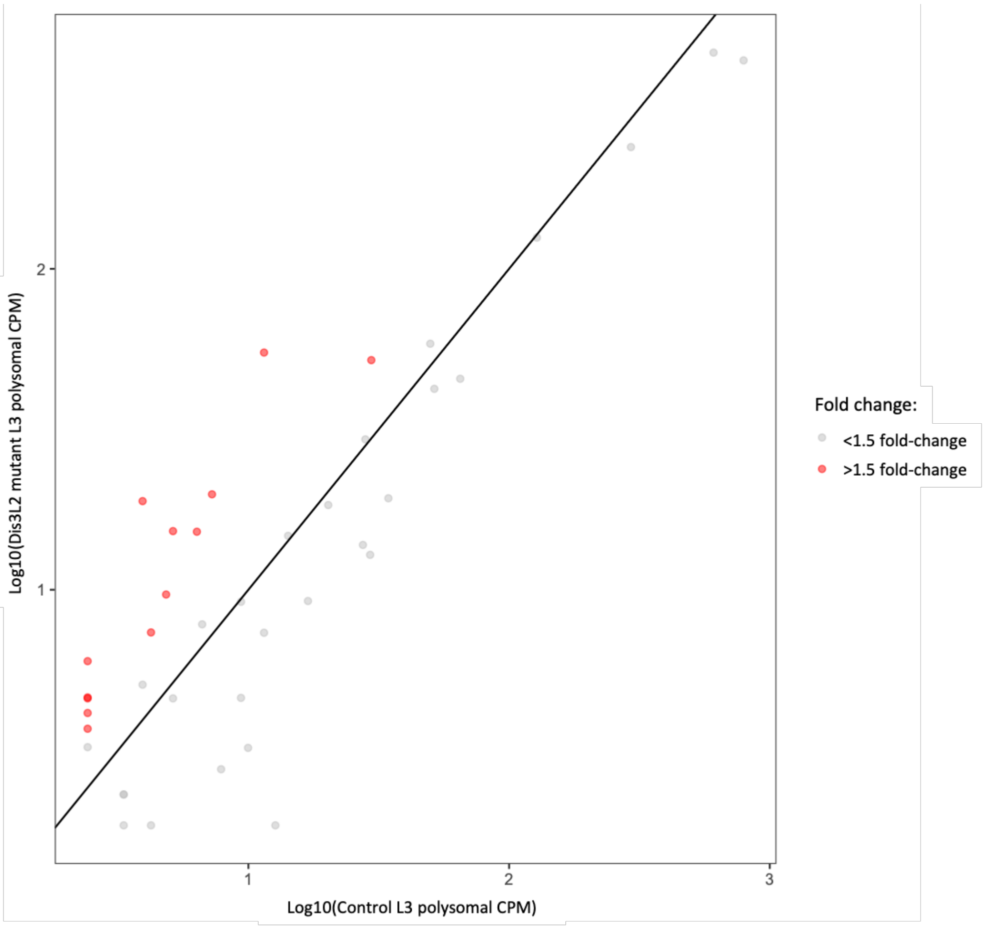
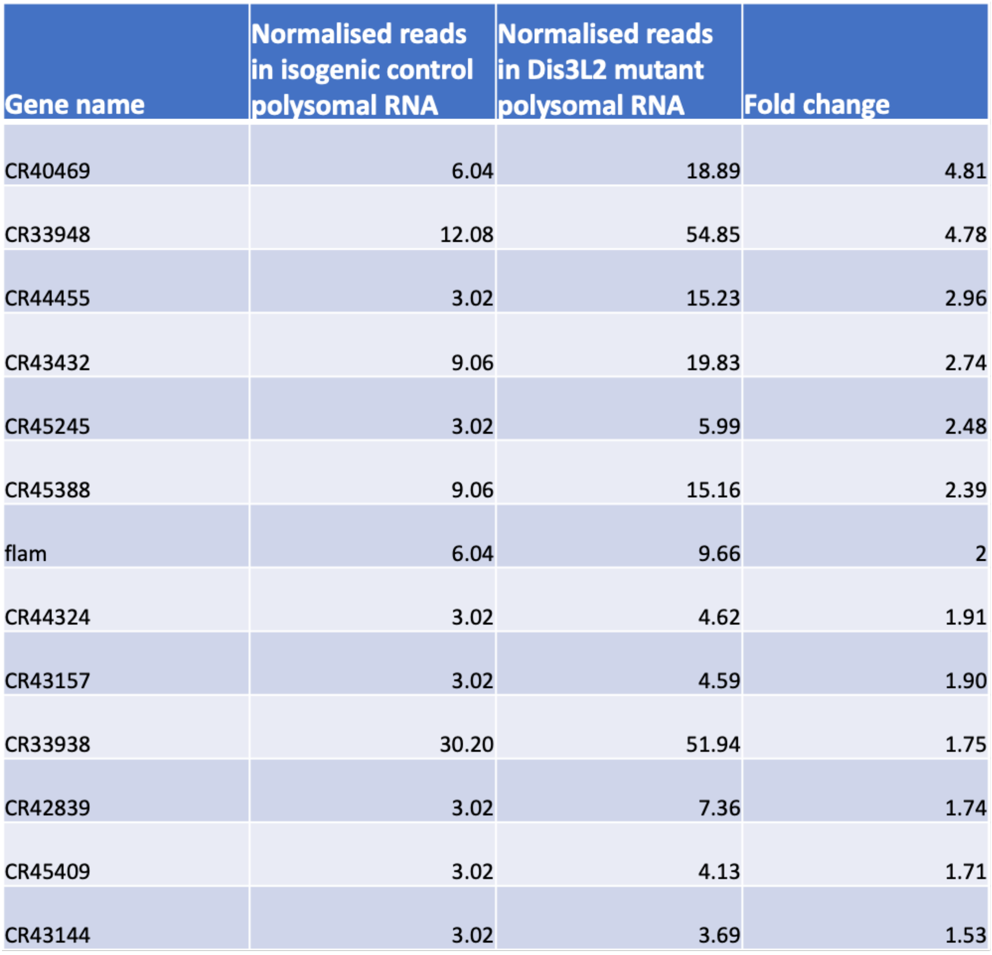


Figure 6.40 – Plot of normalised read counts of polysomally present and polysomally enriched lncRNAs in both genotypes

All 41 lncRNAs with reads detected in all L3 polysomal samples are plotted by normalised read count in control and Dis3L2 mutant polysomal RNA. The 13 which were upregulated in polysomal RNA more than 1.5-fold in the absence of Dis3L2 are highlighted in red.



Gene name	Normalised reads in isogenic control polysomal RNA	Normalised reads in Dis3L2 mutant polysomal RNA	Fold change
CR40469	6.04	18.89	4.81
CR33948	12.08	54.85	4.78
CR44455	3.02	15.23	2.96
CR43432	9.06	19.83	2.74
CR45245	3.02	5.99	2.48
CR45388	9.06	15.16	2.39
flam	6.04	9.66	2
CR44324	3.02	4.62	1.91
CR43157	3.02	4.59	1.90
CR33938	30.20	51.94	1.75
CR42839	3.02	7.36	1.74
CR45409	3.02	4.13	1.71
CR43144	3.02	3.69	1.53

Figure 6.41 – A table listing polysome associated lncRNAs that increase in polysomal abundance in the absence of Dis3L2

Of all lncRNAs detected in RNA sequencing of polysome fractions from both isogenic control and Dis3L2 null mutant L3 larvae, those with reads detected in every replicate (implying a detectable background level polysomal association and possible translation) were taken, and filtered to select only those with an increase of 1.5-fold or more in the absence of Dis3L2. All 13 detected candidates are listed above.

The expression profile (Figure 6.42) from catalogued sequencing data shows varied expression through all tissues, with some presence in all tested tissues, and very high expression in L3 central nervous system, heads of 4-day virgin female *Drosophila,* and the head and accessory gland of mated 4-day male *Drosophila*; and extremely high in the fat body of P8 pupae. Its overall expression through development is much lower, showing no to very low expression throughout development, peaking with low expression in 2 day-old pupae. This contrast between high expression in some tissues and low expression overall is likely explained by extremely low expression in parts of the whole body (for example, the cuticle), diluting the total abundance of the transcript. Interestingly, the lncRNA is also expressed to some extent in all commonly used cell lines, and extremely highly in most of them, including S2 cells. Together this might suggest its important in localised tissues rather than pervasively throughout the organism, and as having a role that results in its increased expression in the specific conditions that cell lines share with each other that is significantly different from what may be found in whole organisms. No speculative biological processes have been noted.

IGV was used to view localisation of read mapping in the novel poly-ribo-seq data. Due to the nature of the polysome footprinting step of poly-ribo-seq, the sequenced RNA fragments are from regions where ribosomes bound, protecting them from RNase digest. Examining the read mapping allows visualisation of read pile up locations (Figure 6.43), illuminating likely gene regions to identify any potential open reading frames (as the ribosome would have been bound on the regions the reads map to). Promising regions can then be examined by amino acid sequence, in order to identify likely start and stop codons that fit the ribosome bound region. A promising ORF was identified (Figure 6.43), and using Gateway cloning, a construct was made with a GFP encoding region downstream of and in-frame of the potential ORF. This was then transfected into S2 cells and observed by microscopy to check for GFP fluorescence. GFP fluorescence was seen clearly using both conventional optical microscopy (using Zeiss Axio Imager 2) and when subsequently imaged using confocal microscopy using (Leica SP8 Microscope), throughout the cell (Figure 6.44). Although further experiments would strengthen the evidence and ensure that the GFP wasn’t due to read through from the previous ORF, this combined with the conservation in polysomal RNA throughout multiple datasets, as well as identification of potential ORFs, strongly support the translation of the gene. The same GFP tag was used for HsrOmega (which did not show

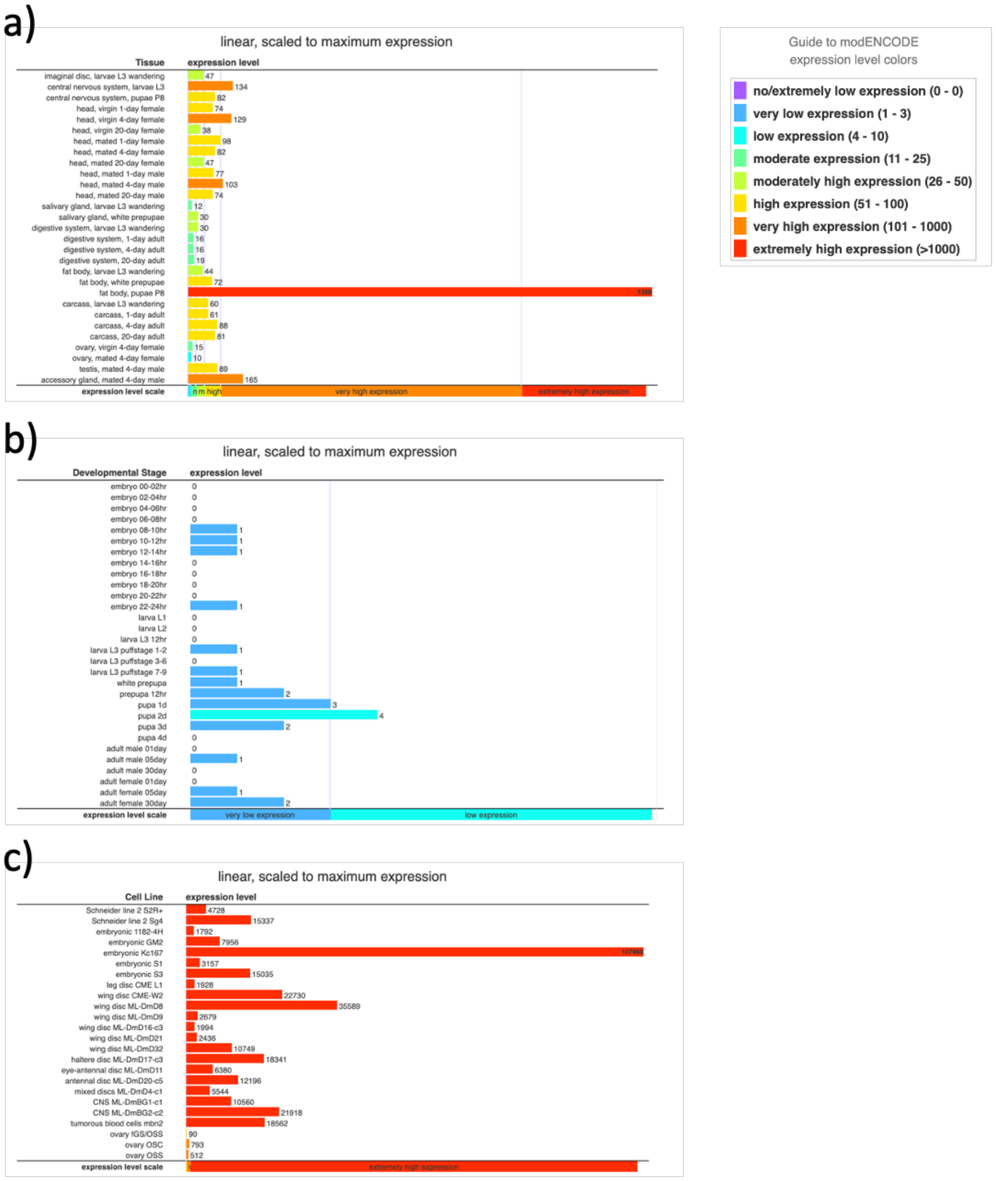


Figure 6.42 – Expression profiles from catalogued RNA-seq data for the lncRNA CR40469

Expression level is shown through developing and developed tissues (a), developmental timepoints (b), and commonly used and available *Drosophila* cell lines (c). Diagram adapted from FlyBase expression data.

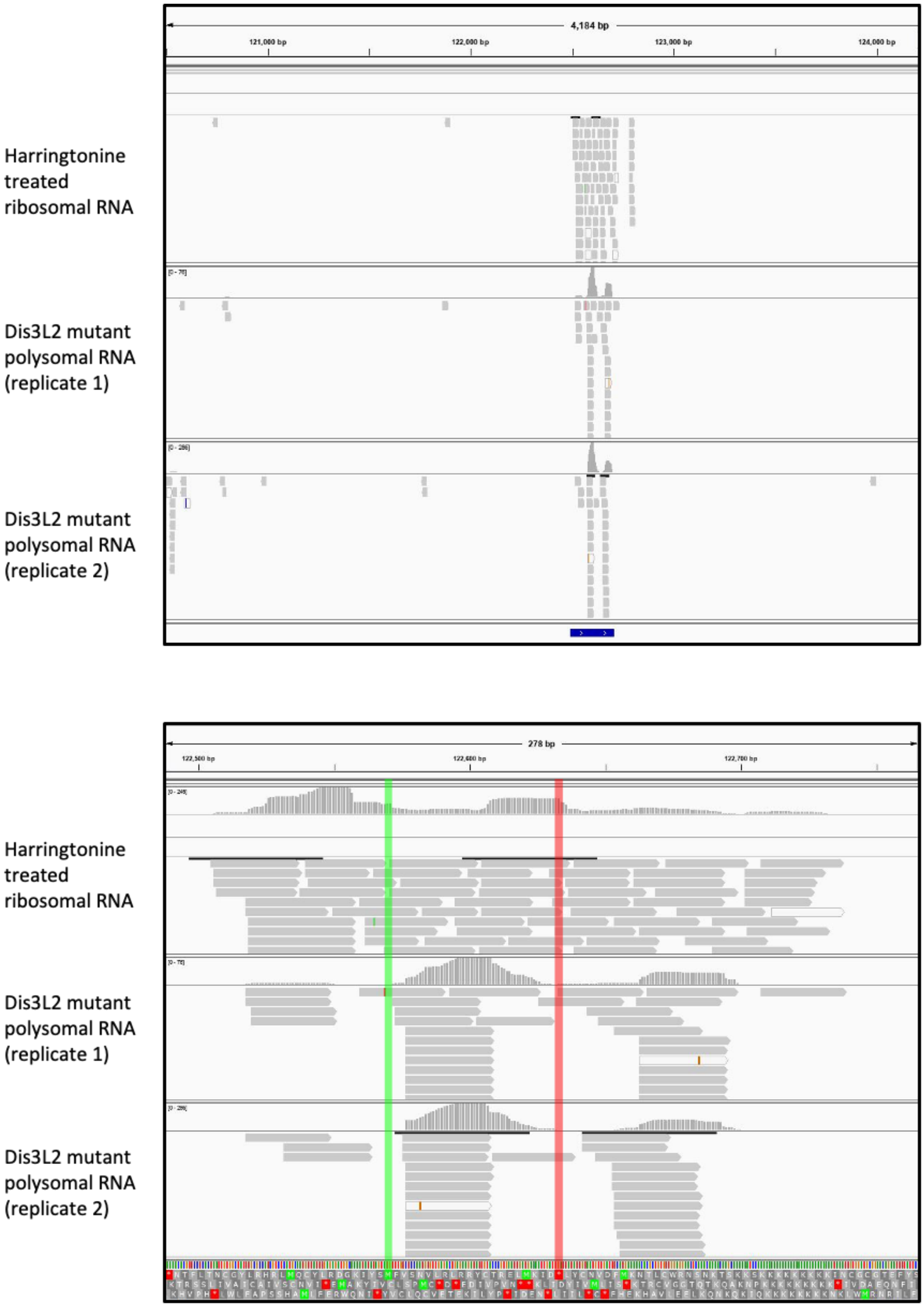

Harringtonine treated ribosomal RNA
Dis3L2 mutant polysomal RNA (replicate 1)
Dis3L2 mutant polysomal RNA (replicate 2)
4,184 bp
121,000 bp
122,000 bp
123,000 bp
124,000 bp
[0 - 78]
[0 - 286]
Harringtonine treated ribosomal RNA
Dis3L2 mutant polysomal RNA (replicate 1)
Dis3L2 mutant polysomal RNA (replicate 2)
278 bp
122,500 bp
122,600 bp
122,700 bp
[0 - 246]
[0 - 78]
[0 - 286]

Figure 6.43 – CR40469 ORFs identified in the data from Zhang *et al*. using IGV

The point with the most pile-up (most protected RNA frangements mapping to that region) can be clearly seen in panel (a). In panel (b), the viewer has been zoomed in, and a potential ORF has been highlighted (start codon in green, and stop codon in red).

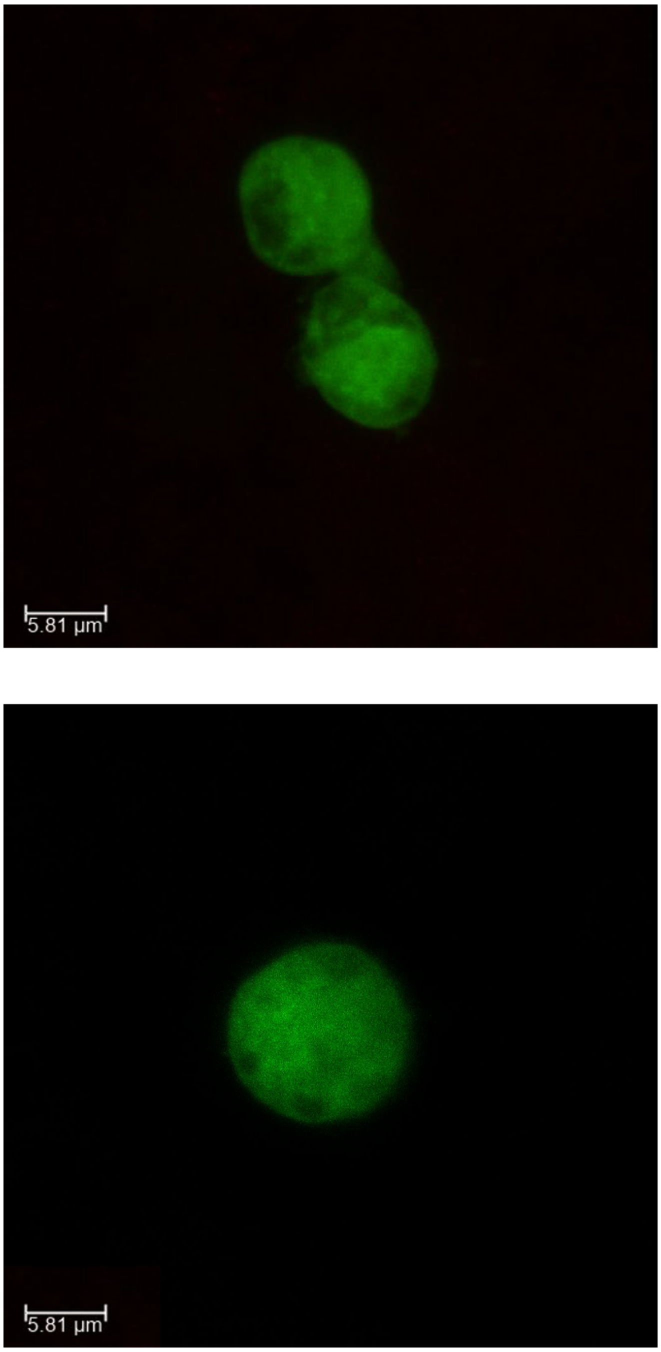


Figure 6.44 – Confocal microscopy shows translation of GFP in frame of the potential CR40469 ORF.

In both panels, GFP can be seen to be expressed throughout the cell, showing the translational viability of the potential ORF.

expression), making it unlikely to merely be read through. In future work, building on that of this thesis, Western blotting could be carried out in order to confirm correct size of peptide, and look at differential abundance at the protein level.

6.10 Comparison between genotypes identifies subset of RNAs affected by the absence of both Pacman and Dis3L2 in *Drosophila* L3 larvae

Although both Pacman and Dis3L2 are known to have specific targets, their important roles in degradation have some overlap, and there are partial and complete redundancies between degradation pathways. Adding to the growing picture of lncRNA degradation by these exoribonucleases, the targets of both Pacman and Dis3L2 were compared, in order to establish whether certain lncRNAs might be targeted for degradation by both Pacman and Dis3L2.

First, the 161 lncRNAs upregulated in Pacman mutant L3 WIDs (as established Chapter 3 and covered again in 6.6.1) were compared with the 56 lncRNAs upregulated in Dis3L2 WIDs (as established Chapter 3 and covered again in 6.8.1). From the comparison, 17 lncRNAs were seen to be upregulated in both (Figure 6.46, panel (a)), with 7 of these being upregulated more than 1.5-fold in both mutant genotypes (Figure 6.45, panel (b)). This is visually represented in Figure 6.46.

Similarly, the 49 lncRNAs upregulated in whole Pacman mutant L3 (as established in 5.8) were compared with the 55 lncRNAs upregulated in whole Dis3L2 (as established in 5.10). From the comparison, 21 lncRNAs were seen to be upregulated in both (Figure 6.47, panel (a)), with 17 of these being upregulated more than 1.5-fold in both mutant genotypes (Figure 6.47, panel (b)). This is visually represented in Figure 6.48.

6.11 Summarising discussion of the information gathered from data comparison and analysis

This chapter has made a large number of cross-comparisons and unearthed extensive information regarding differential abundances between different genotypes and polysomal versus total lysate RNA. There is a substantial amount of information to

a)

Gene name	Fold change in the absence of Pacman in WIDs	Fold change in the absence of Dis3L2 in WIDs
CR42646	1.44	1.05
CR42719	1.78	Infinite
CR43635	2.75	2.49
CR43643	2.40	2.05
CR43685	1.06	1.36
CR43724	1.63	1.27
CR43751	1.45	2.64
CR44537	4.51	3.87
CR44566	3.16	1.47
CR44786	2.01	6.10
CR45359	1.20	1.50
CR45380	2.91	1.43
CR45437	1.85	2.66
CR45517	1.22	1.91
CR45700	2.68	2.19
CR45721	2.50	1.12
CR46090	1.01	1.87

b)

Gene name	Fold change in the absence of Pacman in WIDs	Fold change in the absence of Dis3L2 in WIDs
CR42719	1.78	Infinite
CR43635	2.75	2.49
CR43643	2.40	2.05
CR44537	4.51	3.87
CR44786	2.01	6.10
CR45437	1.85	2.66
CR45700	2.68	2.19

Figure 6.45 – Tables showing potential lncRNA targets of both Pacman and Dis3L2 in *Drosophila* L3 wing imaginal discs (WIDs)

Tables showing lncRNAs upregulated in the absence of both Pacman and Dis3L2 in L3 WIDs, according to data from Towler et al. and Jones et al., processed and selected as described in 5.7.1 and 5.9.1. Panel (a) shows all lncRNAs shown to be upregulated in both genotypes compared to their isogenic control, while panel (b) shows only those that increase more than 1.5-fold.

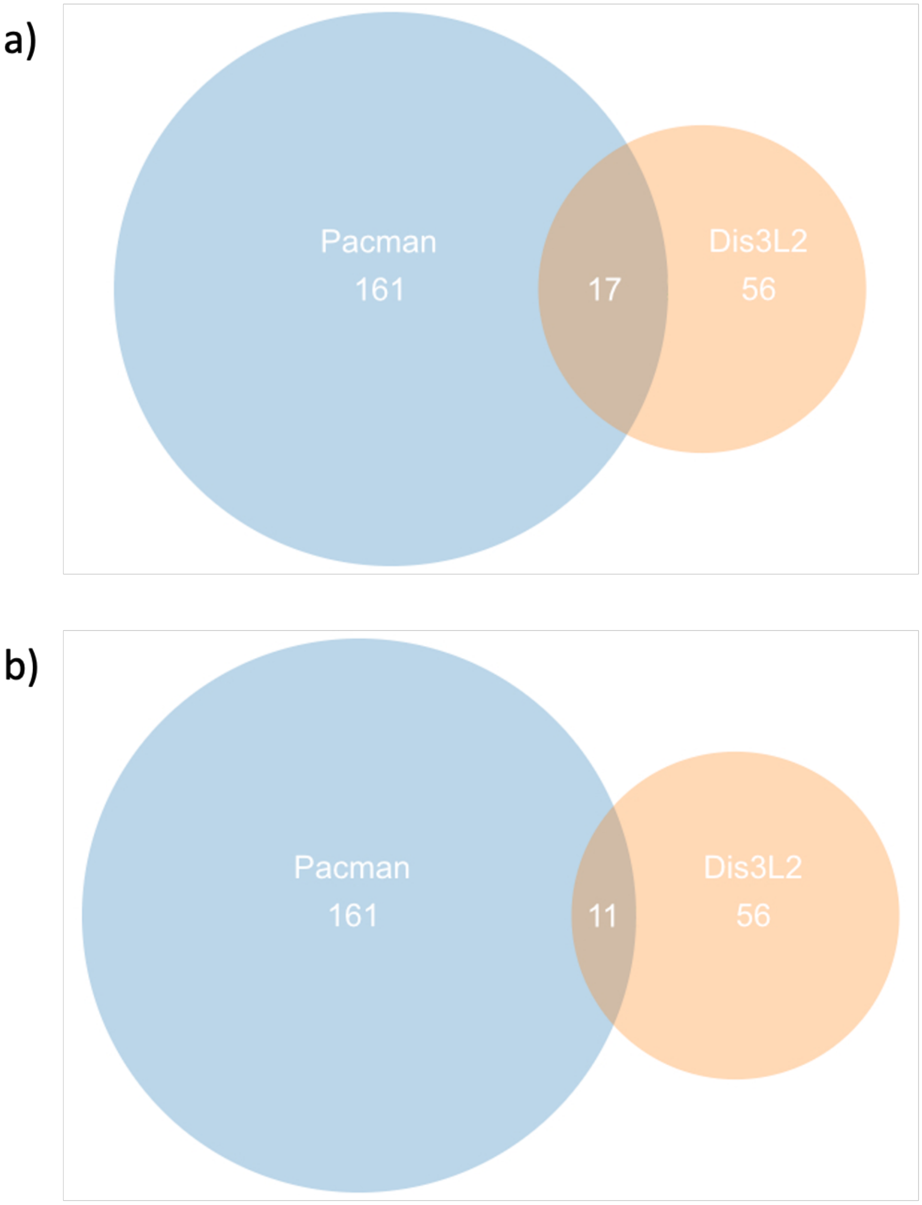
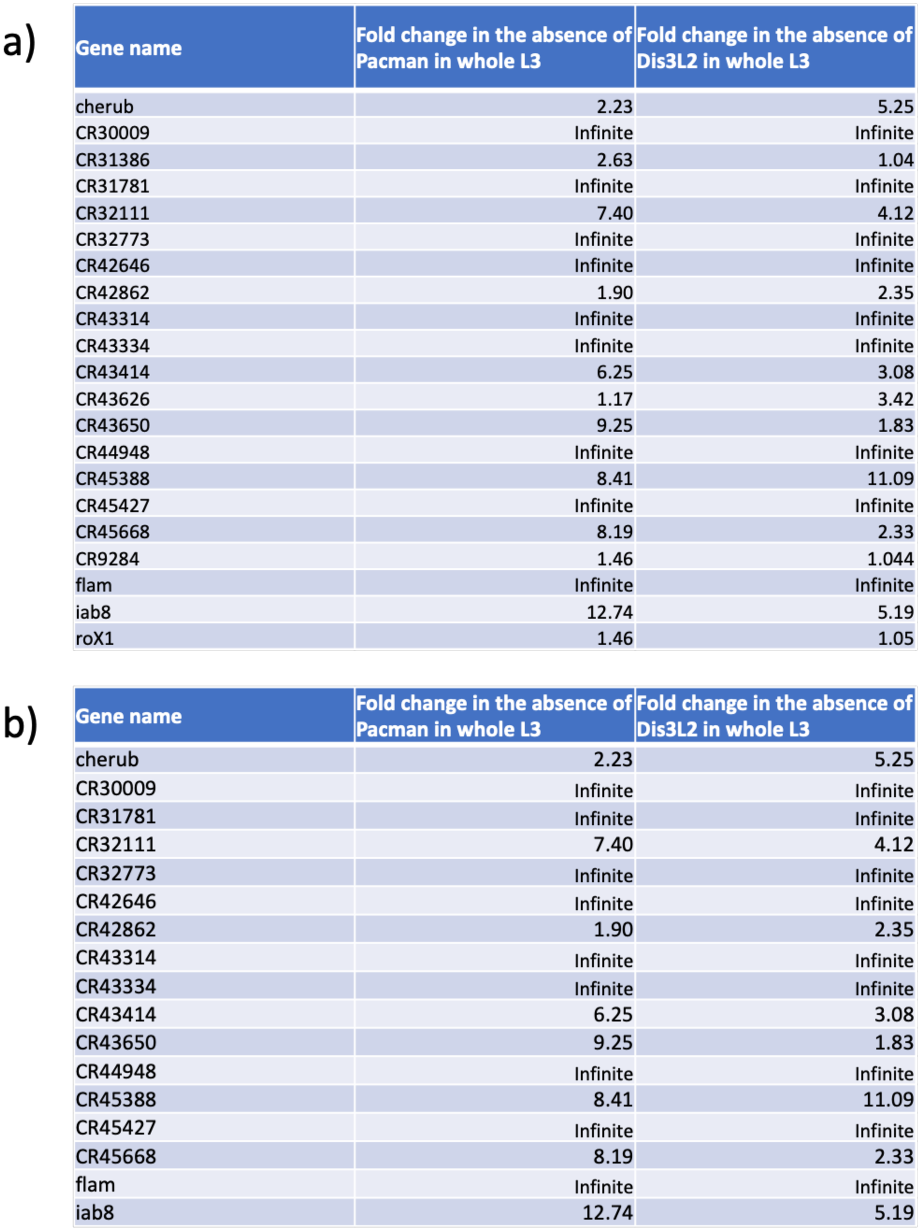


Figure 6.46 – A visual representation of the overlap in lncRNAs regulated by both Pacman and Dis3L2 in L3 WIDs

Venn diagrams showing lncRNAs upregulated in the absence of both Pacman and Dis3L2 in L3 WIDs, according to data from Towler et al. and Jones et al., processed and selected as described in 5.7.1 and 5.9.1. Panel (a) shows all lncRNAs shown to be upregulated in both genotypes compared to their isogenic control, while panel (b) shows only the overlap of those that increase more than 1.5-fold.



a)

Gene name	Fold change in the absence of Pacman in whole L3	Fold change in the absence of Dis3L2 in whole L3
cherub	2.23	5.25
CR30009	Infinite	Infinite
CR31386	2.63	1.04
CR31781	Infinite	Infinite
CR32111	7.40	4.12
CR32773	Infinite	Infinite
CR42646	Infinite	Infinite
CR42862	1.90	2.35
CR43314	Infinite	Infinite
CR43334	Infinite	Infinite
CR43414	6.25	3.08
CR43626	1.17	3.42
CR43650	9.25	1.83
CR44948	Infinite	Infinite
CR45388	8.41	11.09
CR45427	Infinite	Infinite
CR45668	8.19	2.33
CR9284	1.46	1.044
flam	Infinite	Infinite
iab8	12.74	5.19
roX1	1.46	1.05

b)

Gene name	Fold change in the absence of Pacman in whole L3	Fold change in the absence of Dis3L2 in whole L3
cherub	2.23	5.25
CR30009	Infinite	Infinite
CR31781	Infinite	Infinite
CR32111	7.40	4.12
CR32773	Infinite	Infinite
CR42646	Infinite	Infinite
CR42862	1.90	2.35
CR43314	Infinite	Infinite
CR43334	Infinite	Infinite
CR43414	6.25	3.08
CR43650	9.25	1.83
CR44948	Infinite	Infinite
CR45388	8.41	11.09
CR45427	Infinite	Infinite
CR45668	8.19	2.33
flam	Infinite	Infinite
iab8	12.74	5.19

Figure 6.47 – Tables showing potential lncRNA targets of both Pacman and Dis3L2 in whole *Drosophila* L3 larvae

Tables showing lncRNAs upregulated in the absence of both Pacman and Dis3L2 in L3 larvae, according to data from the total L3 lysate carried out as part of the novel poly-ribo-seq experiment, processed and selected as described in 5.7.1 and 5.9.1. Panel (a) shows all lncRNAs shown to be upregulated in both genotypes compared to the isogenic control, while panel (b) shows only those that increase more than 1.5-fold.

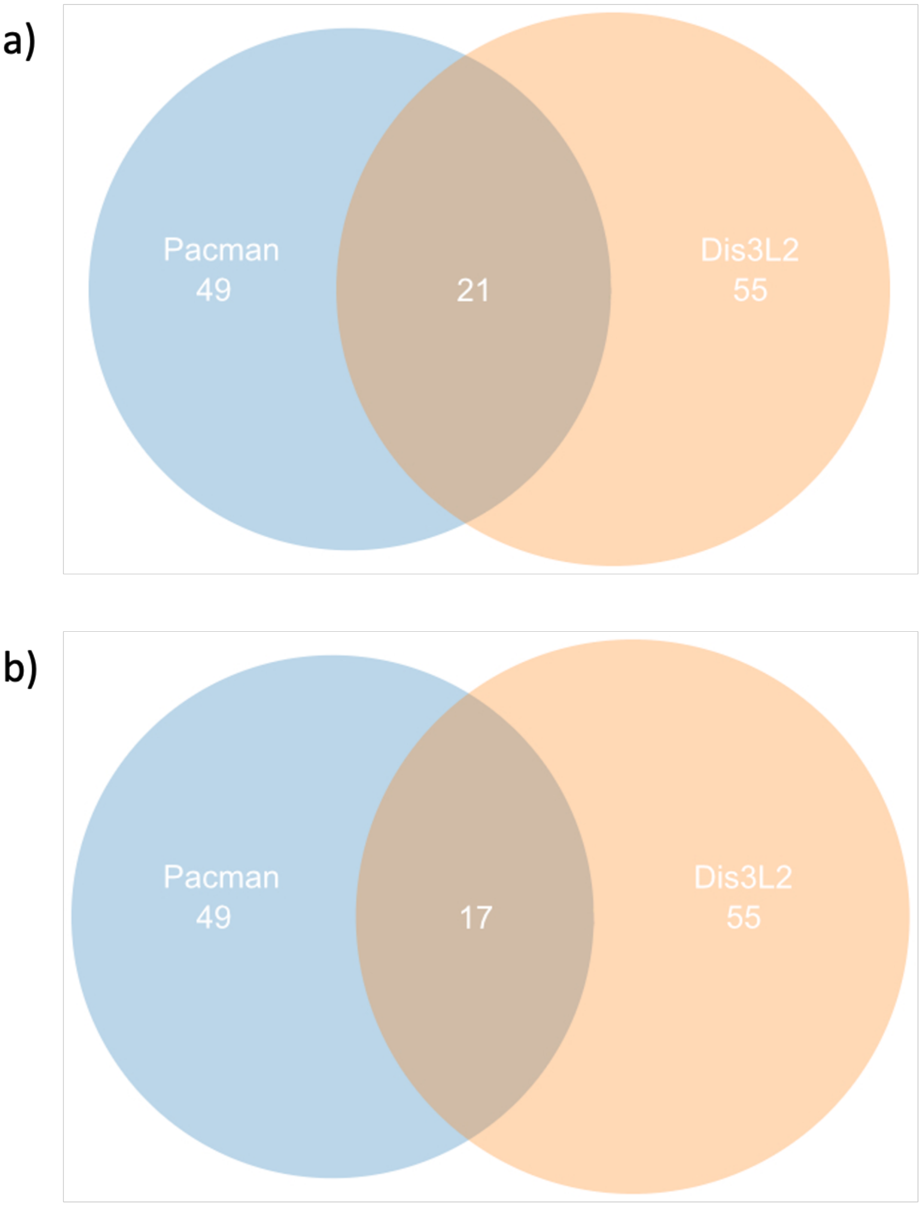


Figure 6.48 – A visual representation of the overlap in lncRNAs regulated by both Pacman and Dis3L2 in whole L3 larvae

Venn diagrams showing lncRNAs upregulated in the absence of both Pacman and Dis3L2 in L3 WIDs, according to data from the total L3 lysate carried out as part of the novel poly-ribo-seq experiment., processed and selected as described in 5.7.1 and 5.9.1. Panel (a) shows all lncRNAs shown to be upregulated in both genotypes compared to their isogenic control, while panel (b) shows only the overlap of those that increase more than 1.5-fold.

assess, and a summarising figure has been added in order to help clarify the overall conclusions and message that can be taken from this data. These summarising “master tables” can be found in Figure 6.49.

A total of 7 data sets have been compared and analysed in this chapter (Figure 6.49, panel (a)): Pacman and Dis3L2 mutant WID RNA-seq data from the Newbury lab (33, 46); Pacman knockdown S2 cell ribo-seq by Antic *et al.* (69); Harringtonine treated S2 cell ribo-seq by Zhang *et al.* (154); poly-ribo-seq on *Drosophila* embryonic stages 1-3, and S2 cells, by the Couso lab (90, 153), and the novel data generated for this thesis (poly-ribo-seq on Pacman mutant, Dis3L2 mutant, and isogenic wild type control whole *Drosophila* L3 larvae).

From this novel data, polysomally associated lncRNAs, with an average of 3 or more reads in polysomal RNA (from all genotypes), of which there were 26, were extracted. These were further sorted by then highlighting those with an average of 3 or more reads in total lysate RNA (from all genotypes), of which there were 15. These 15 would be more likely to be generally abundant transcripts (relatively speaking, for lncRNAs), compared to those from the initial 26 not included in this second list, which may be enriched on the polysome. These are listed in Figure 6.49, panel (b). From these, EdgeR was used to identify those 12 polysome-associated lncRNAs with sufficient reads in all replicates to pass the stringent filtering steps imposed by EdgeR to maximise usefulness of statistical testing (Figure 6.49, panel (c)). Those that passed these filters in both polysomal and total lysate samples (of which there were 7) were compared, providing a ratio of their relative abundance between polysomal RNA and total lysate RNA (Figure 6.49, panel (d)).

With the most promising lncRNAs (both relatively highly expressed, and polysome associated) identified, the lncRNAs that passed the expression cutoff in the polysome were then cross-compared against the RNA-seq data from harringtonine-treated S2 cells carried out by *Zhang et al.* (154). A higher cutoff of an average of 20 reads or more was imposed for the second dataset, as their depth of coverage was higher. The 10 lncRNAs common to these two analyses were extracted (Figure 6.49, panel (e)), likely identifying candidates that are likely to be both on the polysome, and in the Zhang data,

Figure 6.49 – Master table part (a)

Reference	Exoribonucleases depleted	Exoribonuclease depletion method	RNA-sequencing variant	Model system	Further treatment
Jones et al.	Pacman	Null mutant	RNA-seq	*Drosophila* L3 wing imaginal discs	None
Towler et al.	Dis3L2	Null mutant	RNA-seq	*Drosophila* L3 wing imaginal discs	None
Antic et al.	Pacman	dsRNA knockdown	Ribo-seq	*Drosophila* S2 cells	Cycloheximide (post-lysis)
Zhang et al.	None	None	Ribo-seq	*Drosophila* S2 cells	Harringtonine
Patraquim et al.	None	None	Poly-ribo-seq	*Drosophila* embryo	Cycloheximide (post-lysis)
Aspden et al.	None	None	Poly-ribo-seq	*Drosophila* S2 cells	Cycloheximide (post-lysis)
Novel data	Pacman, Dis3L2	Null mutant	Poly-ribo-seq	*Drosophila* whole L3	Cycloheximide (post-lysis)

Panel (a), above – Table summarising datasets for comparison in this chapter

Basic details of all datasets used for comparison and analysis in this chapter.

Gene ID	Average Polysome Reads
CR42862	304.33
roX1	257
cherub	142.67
Hsromega	111
CR32652	26.67
roX2	24.17
CR9284	23.33
CR33938	20.50
CR42850	14.67
CR31781	13.50
CR33948	13.17
CR42767	9.83
CR46006	9.67
CR32835	9.33
CR31451	7.33
CR43626	6.17
CR45102	6.17
CR45388	6.17
CR43334	6
CR31044	5.67
CR46003	5.67
CR43314	5.34
CR40469	5
CR43432	4.83
CR43459	4.5
CR42657	4.33
CR45234	4.17
CR43278	4
CR44455	3.67
flam	3.67
CR44206	3.33
CR43685	3.17
CR44957	3

Gene ID	Average Total Reads
Hsromega	275
roX1	171.50
CR42862	121.83
cherub	79.67
CR40469	31.33
CR44684	27.50
roX2	17.67
CR43626	12.17
CR33938	9.50
CR45102	9.50
CR32652	9.17
CR44042	9.17
CR43334	8.67
CR46006	8.50
CR43314	6
CR43174	5.50
CR45668	4.83
CR43253	4.67
CR42767	4.17
iab8	4.17
CR9284	3.33
let7C	3.33

Panel (b), left – Tables showing all polysome associated lncRNAs identified by poly-ribo-seq with an average of 3 reads mapped in either polysomal or total RNA

The left table shows those with an average of 3 or more reads in polysomal RNA, while the right table shows those with an average of 3 or more reads in total L3 lysate RNA. Of these, 15 lncRNAs were common to both datasets (marked in bold), demonstrating significant overlap, and likely showing these to be generally abundant transcripts, which also happen to be found on the polysome.

Gene ID	Average Polysomal Reads
CR31781	147.18
CR9284	122.30
CR33948	83.34
CR42850	54.14
CR42767	24.11
Hsromega	157.43
CR33938	126.72
CR32652	112.16
roX2	53.06
CR42862	35.53
cherub	28.77
roX1	20.73

Panel (c), above – Tables showing all polysome associated lncRNAs identified by poly-ribo-seq with sufficient reads for EdgeR statistical testing

The 12 lncRNAs listed here were all detected in the polysomal RNA with sufficient reads in all samples to pass through the stringent filtering for statistical analysis, as defined by EdgeR.

Genes	Average Polysomal CPM	Average Total CPM	Polysomal:Total CPM ratio
CR33938	**126.72**	**40.89**	**3.10**
CR32652	**112.16**	**41.93**	**2.68**
roX2	53.06	77.15	0.69
Hsromega	157.43	1404.42	0.11
cherub	28.77	345.42	0.08
CR42862	35.53	580.82	0.06
roX1	20.73	872.55	0.02

Panel (d), above – Tables showing lncRNAs significantly present in both polysomal and total RNA samples

The 7 lncRNAs listed here were all detected in both the polysomal and total RNA with sufficient reads in all samples to pass through the stringent filtering for statistically significant analysis, as set by EdgeR. The ratio between Polysomal CPM and Total CPM indicates whether the lncRNAs are specifically enriched on the polysome, with a higher ratio showing a higher prevalence in the polysomal fractions. Two lncRNAs (highlighted in bold) identified here show an obvious enrichment on the polysome, showing that these lncRNA candidates are in fact predominantly found associated with the polysome.

Gene ID	Polysome reads in L3 larvae	Ribosome reads in S2 cells
CR42862	304.33	1300
Hsromega	111	342
roX2	24.17	130
CR46006	9.67	82
CR43334	6	81
CR31044	5.67	47
CR40469	5	38
CR43459	4.50	37
flam	3.67	37
CR43685	3.17	26

Panel (e), above – Table showing ribosome associated lncRNAs common to both S2 cell ribosomal RNA and L3 polysomal RNA

The 10 lncRNAs listed here are present with 20 or more reads in ribosomal RNA from harringtonine treated S2 cells and an average of 3 or more reads in L3 larvae.

Figure 6.49 – Master table part (b)

Gene ID	E1 Average polysomal reads	E2 Average polysomal reads	E3 Average polysomal reads
CR30009	56	177	26
CR32111	23.5	61	22
CR40469	119	129	61
CR42839	102	58	75
CR42861	152	107.5	82.5
CR42862	819	1560.5	588
CR43148	104	151	79
CR43242	357.5	358	324.5
CR43314	24	46	26
CR43334	27	134	38.5
CR43356	59	29	32
CR43431	39.5	52	31
CR43685	406.5	306.5	195
CR44024	23.5	226	29
CR44042	32.5	62	23
CR44294	145.5	62.5	58.5
CR44317	448	53.5	29
CR44440	811.5	323.5	125
CR44917	55.5	21	21.5
CR44997	99	100	72
CR45473	69	35.5	35
CR45668	78.5	34.5	32
CR46064	167.5	163.5	74.5
flam	74	111.5	39
Hsromega	410.5	546	246.5
iab8	238	142.5	56.5
roX1	717	345.5	83.5
roX2	79.5	202	69.5

Panel (f), above – Table showing polysome associated lncRNAs common to all embryo developmental stages

The 28 lncRNAs listed here are present with 20 or more reads in polysomal RNA from *Drosophila* embryo in stages E1, E2, and E3.

Gene ID	L3 Average Polysome Reads	E1 Average Polysome Reads	E2 Average Polysome Reads	E3 Average Polysome Reads
CR40469	5	119	129	61
CR42862	304.33	819	1560.50	588
CR43314	5.33	24	46	26
CR43334	6	27	134	38.50
CR43685	3.17	406.50	306.50	195
flam	3.67	74	111.50	39
Hsromega	111	410.5	546	246.5
roX1	257	717	345.50	83.50
roX2	24.17	79.50	202	69.50

Gene ID	S2 polysomal reads	Average L3 polysomal reads
CR31044	170	5.67
CR31451	137	7.33
CR42862	19074	304.33
CR43334	1053	6
CR43459	1667	4.50
CR43685	159	3.17
CR44206	534	3.33
CR45102	82	6.17
CR46006	37	9.67
flam	65	3.67
Hsromega	2189	111
roX1	51	257
roX2	300	24.17

Panel (g), above – Tables showing polysome associated lncRNAs common to E1, E2, E3, and L3 developmental stages in *Drosophila*

The upper table shows 9 lncRNAs listed here were all detected with an average of at least 20 reads in all embryonic stages, and an average of at least 3 reads in L3 larvae. The lower table shows 13 lncRNAs listed here were all detected with at least 20 reads S2 cells, and an average of at least 3 reads in L3 larvae.

Gene name	Fold change (L3)	Fold change (S2)
CR30009	Infinite	1.13
CR41257	Infinite	1.11
CR44948	Infinite	1.25
CR46234	Infinite	Infinite
HsrOmega	1.11	1.50
roX1	1.46	1.08

Gene name	Fold change (L3)	Fold change (S2)
CR46234	Infinite	Infinite

Figure (h), left – Conserved lncRNA targets of Pacman in S2 cells and L3 larvae

Of all lncRNAs detected in RNA sequencing of total lysate from both S2 cells and L3 larvae, those common to both were taken and listed. A total of 50 lncRNAs with increased normalised read count in the absence of Pacman were detected in L3 larvae, and 205 in S2 cells. Of these, the 6 listed in the upper table are common to both. With a cut off of 1.5-fold increase or more, there were 114 in the S2 cell data, and 45 in the L3 larval data, with only one common to both, listed in the lower table.

Gene name	Fold change (L3)	Fold change (WID)
CR44506	Infinite	2.14
CR31781	Infinite	2.12
CR44347	Infinite	1.08
CR42646	Infinite	1.44
CR45312	Infinite	1.13
CR32111	7.404	1.68
cherub	2.23	1.18
CR43626	1.17	1.22

Gene name	Fold change (L3)	Fold change (WID)
CR44506	Infinite	2.14
CR31781	Infinite	2.12
CR32111	7.40	1.68

Panel (i), left – Conserved lncRNA targets of Pacman in L3 wing imaginal discs (WIDs) and L3 larvae

Of all lncRNAs detected in RNA sequencing of total lysate from both L3 WIDs and L3 larvae, those common to both were taken and listed. After quality control steps discussed in 5.7.1, a total of 161 lncRNAs with increased normalised read count in the absence of Pacman were detected in L3 WIDs, and 49 in S2 whole L3 larvae. Of these, the 8 listed in the upper table are common to both. With a cut off of 1.5-fold increase or more, there were 161 lncRNAs with increased normalised read count in the absence of Pacman were detected in L3 WIDs, and 49 in S2 whole L3 larvae, with only 3 common to both, listed in the lower table.

Gene name	Normalised reads in isogenic control polysomal RNA	Normalised reads in Pacman mutant polysomal RNA	Fold change
Hsromega	127.80	500.46	3.92
CR43242	3.32	6.74	2.03
CR43650	3.32	6.54	1.97
CR45388	6.34	11.65	1.84
CR45102	9.37	15.93	1.70
CR45409	2.42	3.88	1.61
CR44324	2.42	3.88	1.61

Panel (j), above – A table listing polysome associated lncRNAs that increase in polysomal abundance in the absence of Pacman

Of all lncRNAs detected in RNA sequencing of polysome fractions from both isogenic control and Pacman null mutant L3 larvae, those with reads detected in every replicate (implying a detectable background level polysomal association and possible translation) were taken, and filtered to select only those with an increase of 1.5-fold or more in the absence of Pacman. All 7 detected candidates are listed above.

Figure 6.49 – Master table part (c)

Gene name	Fold change (L3)	Fold change (WID)
CR33938	1.39	2.14
CR42646	Infinite	1.05
CR45517	Infinite	1.91

Gene name	Fold change (L3)	Fold change (WID)
CR45517	Infinite	1.91

Panel (k) , above – Conserved lncRNA targets of Dis3L2 in L3 wing imaginal discs (WIDs) and L3 larvae

Of all lncRNAs detected in RNA sequencing of total lysate from both L3 WIDs and L3 larvae, those common to both were taken and listed. After quality control steps discussed in 5.9.1, a total of 56 lncRNAs with increased normalised read count in the absence of Dis3L2 were detected in L3 WIDs, and 55 in S2 whole L3 larvae. Of these, the 3 listed in the upper table are common to both. With a cut off of 1.5-fold increase or more, there were 40 lncRNAs with increased normalised read count in the absence of Dis3L2 detected in L3 WIDs, and 48 in S2 whole L3 larvae, with only 1 common to both, listed in the lower table.

Gene name	Normalised reads in isogenic control polysomal RNA	Normalised reads in Dis3L2 mutant polysomal RNA	Fold change
CR40469	6.04	18.89	4.81
CR33948	12.08	54.85	4.78
CR44455	3.02	15.23	2.96
CR43432	9.06	19.83	2.74
CR45245	3.02	5.99	2.48
CR45388	9.06	15.16	2.39
flam	6.04	9.66	2
CR44324	3.02	4.62	1.91
CR43157	3.02	4.59	1.90
CR33938	30.20	51.94	1.75
CR42839	3.02	7.36	1.74
CR45409	3.02	4.13	1.71
CR43144	3.02	3.69	1.53

Panel (l), above – A table listing polysome associated lncRNAs that increase in polysomal abundance in the absence of Dis3L2

Of all lncRNAs detected in RNA sequencing of polysome fractions from both isogenic control and Dis3L2 null mutant L3 larvae, those with reads detected in every replicate (implying a detectable background level polysomal association and possible translation) were taken, and filtered to select only those with an increase of 1.5-fold or more in the absence of Dis3L2. All 13 detected candidates are listed in this table.

Gene name	Fold change in the absence of Pacman in WIDs	Fold change in the absence of Dis3L2 in WIDs
CR42646	1.437179239	1.05028692
CR42719	1.782570332	Infinite
CR43635	2.753869128	2.48569623
CR43643	2.397365703	2.05243993
CR43685	1.057677062	1.35700685
CR43724	1.630339024	1.27210522
CR43751	1.450394139	2.63769911
CR44537	4.512415762	3.86872988
CR44566	3.164439532	1.47340802
CR44786	2.010563637	6.10406362
CR45359	1.195984755	1.50097505
CR45380	2.912726013	1.43443839
CR45437	1.853660453	2.66157459
CR45517	1.221012526	1.91329986
CR45700	2.683247252	2.18796552
CR45721	2.504860049	1.1212096
CR46090	1.013536679	1.86723822

Gene name	Fold change in the absence of Pacman in WIDs	Fold change in the absence of Dis3L2 in WIDs
CR42719	1.782570332	Infinite
CR43635	2.753869128	2.48569623
CR43643	2.397365703	2.05243993
CR44537	4.512415762	3.86872988
CR44786	2.010563637	6.10406362
CR45437	1.853660453	2.66157459
CR45700	2.683247252	2.18796552

Panel (m), above – Tables showing potential lncRNA targets of both Pacman and Dis3L2 in *Drosophila* L3 wing imaginal discs (WIDs)

Tables showing lncRNAs upregulated in the absence of both Pacman and Dis3L2 in L3 WIDs, according to data from Towler et al. and Jones et al., processed and selected as described in 5.7.1 and 5.9.1. The upper table shows all lncRNAs shown to be upregulated in both genotypes compared to their isogenic control, while the lower table shows only those that increase more than 1.5-fold.

Gene name	Fold change in the absence of Pacman in whole L3	Fold change in the absence of Dis3L2 in whole L3
cherub	2.23	5.25
CR30009	Infinite	Infinite
CR31386	2.63	1.04
CR31781	Infinite	Infinite
CR32111	7.40	4.12
CR32773	Infinite	Infinite
CR42646	Infinite	Infinite
CR42862	1.90	2.35
CR43314	Infinite	Infinite
CR43334	Infinite	Infinite
CR43414	6.25	3.08
CR43626	1.17	3.42
CR43650	9.25	1.83
CR44948	Infinite	Infinite
CR45388	8.41	11.09
CR45427	Infinite	Infinite
CR45668	8.20	2.33
CR9284	1.46	1.04
flam	Infinite	Infinite
iab8	12.74	5.19
roX1	1.46	1.05

Gene name	Fold change in the absence of Pacman in whole L3	Fold change in the absence of Dis3L2 in whole L3
cherub	2.23	5.25
CR30009	Infinite	Infinite
CR31781	Infinite	Infinite
CR32111	7.40	4.12
CR32773	Infinite	Infinite
CR42646	Infinite	Infinite
CR42862	1.90	2.35
CR43314	Infinite	Infinite
CR43334	Infinite	Infinite
CR43414	6.25	3.08
CR43650	9.25	1.83
CR44948	Infinite	Infinite
CR45388	8.41	11.09
CR45427	Infinite	Infinite
CR45668	8.19	2.33
flam	Infinite	Infinite
iab8	12.74	5.19

Panel (n) – Tables showing potential lncRNA targets of both Pacman and Dis3L2 in whole *Drosophila* L3 larvae

Tables showing lncRNAs upregulated in the absence of both Pacman and Dis3L2 in L3 larvae, according to data from the total L3 lysate carried out as part of the novel poly-ribo-seq experiment, processed and selected as described in 5.7.1 and 5.9.1. The left table shows all lncRNAs shown to be upregulated in both genotypes compared to the isogenic control, while the right table shows only those that increase more than 1.5-fold.

frozen at their potential initiation site. This identifies not only the lncRNAs, but sites to begin looking for potential ORFs.

In a similar vein, the lncRNAs that passed the expression cutoff in the polysome (in the novel data) were then cross-compared against the poly-ribo-seq data from the Couso lab (90, 153), again with a higher cutoff of an average of 20 reads or more was imposed for these dataset, as their depth of coverage was higher. There were 28 lncRNAs identified as passing these criteria in all three tested embryo stages (E1-3), shown in Figure 6.49, panel (f). These were then compared to the novel L3 poly-ribo-seq, and 9 were also found to pass the previously described criteria in the L3 polysomal samples. Further comparison between the novel L3 data and the polysomal samples from the Couso lab's S2 cell poly-ribo-seq data identified 13 that were common between the L3 and S2 datasets, again with a higher cutoff of an average of 20 reads or more was imposed for the S2 cell dataset, as their depth of coverage was higher. These are shown in Figure 6.49, panel (g).

Next, by comparing the total lysate samples from the novel data with that of the Pacman depleted S2 cell data from Antic *et al.* (69), the conservation of Pacman targets between the two models could be analysed. A total of 50 lncRNAs with increased normalised read count in the absence of Pacman were detected in L3 larvae, and 205 in S2 cells. Of these, 6 were common to both, with only 1 (*CR46234*) exceeding a 1.5-fold increase in both models; showing that although significant variation between models is to be expected, and was found, there are still conserved targets to be found (Figure 6.49, panel (h)).

By taking a similar approach to comparing the total lysate samples from the novel data with that of the Pacman mutant WID data from the Newbury lab (33), those common to both were taken and listed. A total of 161 lncRNAs with increased normalised read count in the absence of Pacman were detected in L3 WIDs, and 49 in S2 whole L3 larvae. Of these, 8 were identified as common to both. Of those with a fold-change increase of 1.5 or greater in the absence of Pacman, 3 were identified (Figure 6.49, panel (i)). Next, looking in the polysomal RNA in both isogenic control and Pacman mutant samples from the novel L3 poly-ribo-seq, those with reads detected in every polysomal replicate were taken, and filtered to identify those with a 1.5-fold or greater increase in the absence of Pacman, of which 7 were identified (Figure 6.49, panel (j)).

The equivalent comparisons were also carried out between the Dis3L2 mutant samples and the isogenic controls, this time identifying in the total lysate a total of 56 lncRNAs with increased normalised read count in the absence of Dis3L2 in L3 WIDs, and 55 in S2 whole L3 larvae. Of these, 3 were common to both (Figure 6.49, panel (k)). With a cut off of 1.5-fold increase or more, there were 40 lncRNAs with increased normalised read count in the absence of Dis3L2 detected in L3 WIDs, and 48 in S2 whole L3 larvae, with only 1 common to both. Next, looking in the polysomal RNA in both isogenic control and Dis3L2 mutant samples from the novel L3 poly-ribo-seq, those with reads detected in every polysomal replicate were taken, and filtered to identify those with a 1.5-fold or greater increase in the absence of Pacman, of which 13 were identified (Figure 6.49, panel (l)).

In order to get the most from these previous comparisons in the previous Newbury lab RNA-seq data, and the novel poly-ribo-seq data, the lncRNAs upregulated more than 1.5-fold in Pacman and Dis3L2 mutant WIDs were compared, and those 7 common to both (likely to be targets of non-specific, or less-specific, exoribonuclease degradation) were extracted (Figure 6.49, panel (m)). The same was done in the total lysate samples from the novel poly-ribo-seq data, identifying 17 lncRNAs upregulated more than lncRNAs in the absence of both Pacman, and the absence of Dis3L2 (Figure 6.49, panel (n)).

In summary, these comparisons have been able to identify lncRNAs found at relatively high levels on the polysome in *Drosophila* L3 larvae, and of those, which are also relatively abundant in the total lysate RNA. This provides candidates for future work that may wish to investigate generally polysomally abundant lncRNAs (possibly undergoing translation) and separate those which may be abundant throughout versus those enriched on the polysome. The analysis with EdgeR allows different targets to be included or excluded based on the stringency that one wishes to apply (and how many targets one wishes to investigate). Further, the comparison to harringtonine treated S2 cells allows us to identify not only those lncRNAs with conserved polysomal presence between two very different *Drosophila* models (more likely to have a conserved, biologically relevant role), but also use the pile-up location to examine the potential ORF of the most interesting candidates. Further conservation comparisons could be made with embryonic and untreated S2 cell poly-ribo-seq data, providing more comprehensive lists of which lncRNAs are conserved between which models and conditions.

Following these comparisons that allowed useful polysomal and translational information to be discovered, the impact of the exoribonucleases Pacman and Dis3L2 was examined by comparisons with the previous work on dsRNA-treated Pacman knockdown S2 cells, as well as Pacman and Dis3L2 null mutant L3 WIDs. This allowed identification of lncRNAs that were specifically targeted by Pacman or Dis3L2 across multiple tissues and models, providing novel insight into lncRNA degradation, especially in whole L3 larvae. This was followed by examination of lncRNA degradation by these enzymes of lncRNAs present on the polysome, by identifying polysome-associated lncRNAs that increased in abundance in the absence of either Pacman or Dis3L2. By examining the changes at both the total RNA level and the polysome-associated lncRNA level, the targeting of polysome-bound (possibly translating) lncRNAs versus RNAs present elsewhere by Pacman and Dis3L2 can be better understood: By comparing those lncRNAs that increase and decrease (in the absence of either exoribonuclease) in polysomal and total lysate RNA with different profiles (summarised in depth in 6.3.2.3), we also are able to identify targets that may be require translation in order to undergo their canonical degradation (among other examinations of the mechanics of the degradation of these lncRNAs. Finally, by comparing the genotypes examined in this novel data, specificity and redundancy of lncRNA degradation can be explored, and this work identifies lncRNAs likely to undergo degradation by multiple exoribonucleases, with less potential for specific targeting by either.

This chapter provides a substantial amount of novel information on both the degradation and potential translation of lncRNAs in *Drosophila*, from a new dataset (using a model that hasn’t been examined in this context before), as well as a re-imagining and new analysis of existing data. Although further work would be required to pursue, validate, and better understand specific candidates, this chapter shows that a strong, informative overview can be gained by analysing the data within.

Chapter 7: Discussion

7.1 Pacman and Dis3L2 are required for specific degradation of certain lncRNAs in *Drosophila melanogaster*

7.1.1 Existing genome-wide data sets can identify candidate lncRNAs for degradation by exoribonucleases

The work presented in Chapter 3 used previously existing datasets in order to identify pools of potential candidate lncRNAs upregulated in the absence of either Pacman or Dis3L2. The data from Antic *et al*. provided an initial pool of 8 lncRNAs upregulated in the absence of Pacman in *Drosophila* S2 cells.

By then carrying out similar analysis on work Jones *et al*. and Towler *et al*. the comparison produced a further 16 lncRNAs upregulated in Pacman null mutant *Drosophila* L3 wing imaginal discs, and 15 lncRNAs upregulated in Dis3L2 null mutant *Drosophila* L3 wing imaginal discs. This initial use of existing data provides initial support for the degradation of certain lncRNAs by the canonical Pacman and Dis3L2 RNA decay pathways, and highlighted potential targets that could be followed up experimentally. With the same stringency and filtering between the two datasets, a similar proportion of RNAs were up- and down-regulated in he absence of both Pacman (610 RNAs more than 2-fold upregulated, 615 RNAs more than 2-fold downregulated) and Dis3L2 (638 RNAs more than 2-fold upregulated, 767 RNAs more than 2-fold downregulated).

Although at first we might imagine the depletion of a crucial exoribonuclease to lead to a greater proportion of RNA transcripts being increased in abundance (due to the regulating enzyme being unable to degrade its targets), this fairly equal spread can be explained in other ways. The complex layers of RNA degradation mean that RNA stability can be degraded by indirect means. For example, a transcript that is directly and specifically degraded by an exoribonuclease (and increases in abundance when said exoribonuclease is depleted) may play its own role in regulating the stability of other transcripts, by affecting protective feature (5’ capping and 3’ tailing), degradation-specific structure formation, or recruitment of other parts of the decay machinery.

In a yet more indirect fashion, many transcripts will (as part of their standard function) cause downstream effects in the cell (for example the role of Pacman regulated *reaper* and *hid* in apoptotic pathways) which will pass on other effects downstream leading to a shift in the equilibrium of various relevant RNAs (by modulating either synthesis or decay rates). With such complex and interwoven levels of genetic regulation, high-throughput methods of examining exoribonuclease depleted samples alone can not identify all targets of decay by the tested enzyme. This once again shows the importance of using multiple related datasets, as well as further dissecting the informative data gained from high-throughput datasets using follow up molecular investigation (with specific experimental design depending on what question is being answered).

When examining only lncRNAs, the numbers are greatly reduced. Only 1 lncRNA is more than 2-fold upregulated in the Pacman mutant (CR43334), with three lncRNA more than 2-fold upregulated in the Dis3L2 mutant (CR43334, cherub, and CR40469). Interestingly, despite the low number of identified targets there is still some overlap in the effected lncRNAs, with CR43334 increasing in abundance more than 2-fold in the absence of either Pacman or Dis3L2. It is apparent from the higher number of targets identified by an earlier, lower stringency approach (used in Chapter 3 to identify preliminary proof of principle) we can also gather that while high stringency approaches can be useful to identify the strongest candidates and largest changes for differential abundance, the drawback is that more subtle (but still scientifically valid) targets can be missed. Given that some of these less stringent candidates were followed up individually and the differential abundance proven with further molecular techniques, it shows the validity of implementing varied cut-offs for different purposes; from carrying out early, informative overviews of data, and proof-of-principle, to pointed identification of the most promising candidates.

The use and re-use of existing high-throughput datasets is a valuable approach in recent years. The very nature of genome-wide, high-throughput techniques (like RNA-seq) provide vast amounts of information, which requires dedicated work to be interpreted in the context of a specific question or set of questions. Undoubtedly, most datasets of this kind still hold potentially valuable information, and by perusing the relevant literature a substantial amount of previous work can shed light upon a novel question.

Having found candidates for both Pacman and Dis3L2 degraded lncRNAs, their differential abundance in the absence of these enzymes were verified using both semi-quantitive and quantitive PCR, providing specific examples of lncRNAs definitively degraded by Pacman and Dis3L2. By experimentally proving several candidates to be at an increased abundance in the absence of either Pacman or Dis3L2, definitive proof of principle was provided that both Pacman and Dis3L2 are involved in the degradation of certain (but not all) candidate lncRNAs. Use of semi-quantitive PCR showed a candidate lncRNA significantly upregulated in the absence of each exoribonuclease (CR42179 upregulated in the absence of Dis3L2, and CR43635 upregulated in the absence of Pacman). Additionally, CR45177 was shown by qPCR to be significantly increased in abundance in the absence of Dis3L2 function. No known function was available for any of these candidates, as is common with lncRNAs; although informatic tools like ORFfinder allow identification of potential ORFs regardless of their other cataloguing, providing a useful tool for further investigation of completely novel and unknown lncRNAs. The variability between differential abundance in sequencing data and PCR techniques also showed the importance of validating low-confidence targets, showing that some lncRNAs, for example *CR6900*, did not show significant difference in the absence of Dis3L2, despite having been identified in the high-throughput data from Towler *et al*. (46)

7.1.2 Specific exoribonuclease sensitive lncRNAs can be identified on the polysome

After having established that certain lncRNAs are specifically degraded by Pacman and Dis3L2, these existing, proven candidates could then be examined for their presence on the polysome. Use of polysome fractionation, RNA extraction, and PCR techniques identified examples of both Pacman and Dis3L2 degraded lncRNAs that were present on the polysome. CR43635 was shown by sqPCR to be upregulated in the absence of Pacman, and CR42719 was shown by sqPCR and qPCR showed to be upregulated in the absence of Dis3L2; and both were identified by PCR to be present on the polysome. This successfully proved the principle that not only are lncRNAs degraded by the canonical Pacman and Dis3L2 pathways (presumably along with their regulating steps), but some of those Pacman and Dis3L2 regulated lncRNAs are present on the polysome, with the possibility of translation of smORFs within these lncRNAs.

7.2 Generating Pacman and Dis3L2 deficient samples suitable for poly-ribo-seq is a difficult undertaking

7.2.1 dsRNA treatment requires unfeasibly high amounts of exoribonuclease complimentary double strand RNA

In order to provide a novel dataset that would allow a genome wide look at lncRNAs in the absence of Pacman and Dis3L2, as well as in both total sample lysate versus polysomal RNA only, a usable model had to be developed. Although previous poly-ribo-seq experiments have used *Drosophila* models (as both S2 cells and multiple embryonic stages), this experiment required this to be done in a way that allowed the role of Pacman and Dis3L2 to be explored, and also allowed useful comparison to existing data.

Multiple approaches to this were attempted during this project. The previous work by Antic *et al*. showed a promising approach to knockdown Pacman in S2 cells; one that was relatively straightforward. This project hoped to apply it to Pacman cells once again, as well as apply it to Dis3L2 cells. S2 cells have the advantage of being very simple and straightforward for poly-ribo-seq, producing very high quality polysome traces, as well as being very simple to amplify up to create large amounts of input RNA. In addition, several relevant ribo-seq datasets, as well as RNA-seq datasets already exist in S2 cells. Unfortunately, the low reliability of the knockdown along with the vast quantity of dsRNA required led to the decision that this supposedly straightforward method was not feasible for this project.

7.2.2 Technical issue inhibited CRISPR-Cas9 exoribonuclease depletion in *Drosophila* S2 cells

Due to the aforementioned ease of polysome fractionation in S2 cells, another approach was attempted in this cell line. With the boom of popularity in recent years of the CRISPR-Cas9 system of gene editing, an attempt was made to create null mutants for both Pacman and Dis3L2 in S2 cells, which would allow the generation of stable S2 cell stocks, easily amplified up for input samples to any experiment needed. In addition, this may have provided a new phenotyping opportunity for complete Pacman and Dis3L2 depletion in S2 cells.

Unfortunately, two significant factors inhibited this experiment from going forward further. First, the recovery phase from sub-single cell concentration in S2 cells is slow, with a high rate of death for cell populations. The rate at which the cells recovered, along with relatively low transfection efficiency (even post-optimisation) meant a lot of time was invested in each well, which had a low chance of payoff. The second problem was the technical issues leading to the death of all recovering cell stocks. As discussed in Chapter 4, S2 cells do require a stable temperature of anywhere between 23-27°C, and sufficient humidity to keep their media at a high enough liquid volume to adequately cover the cultures. Growth rates for the cells in this project were found to be wildly inconsistent, and survivability varied between 2 and 20 passages, with some cultures dying as soon as they were re-suspended from stock pellets. After substantial investigation, the incubator was identified as being incapable of supporting a humidity suitable for S2 cell culture. Instead, rapid airflow in the incubator used had been sporadically dehydrating cell stocks, causing sudden death of cells. Due to the previously mentioned time required for the cell growth and recovery, this put an end to the feasibility of this experiment within the scope of this project.

7.2.3 Successful development of an exoribonuclease depleted *Drosophila* in vivo model suitable for poly-ribo-seq

Following the unsuccessful S2 cell attempts to produce a viable source for exoribonuclease depleted samples for poly-ribo-seq, the subsequent attempt to generate the sample was undertaken in whole *Drosophila* L3 larvae. As previously existing Pacman and Dis3L2 mutant larvae were available, the task here was optimizing polysome fractionation protocols to produce high quality fractions from whole L3 larvae. The high content of fat and debris, as well as the high concentration of RNases, was problematic. However, after substantial optimization, the work in this project showed that it is possible to produce high quality polysome profiles even from difficult whole organism samples. The steps used to get this result would produce valuable points to work from for any future attempts with difficult tissue samples.

7.2.4 Limitations and future adjustments to novel protocol for poly-ribo-seq in exoribonuclease depleted *Drosophila* samples

Although the exploration of steps required to optimise the polysome fractionation procedure revealed several crucial steps between sample lysis and producing a high quality profile, and ultimately gave samples that could be taken through the rest of the preparation steps for poly-ribo-seq, the final sequencing from these samples was left with some limitations due to the novel protocol.

Despite the oligo-dT depletion and polyA selection steps, depletion of rRNA and bacterial RNA steps was not complete therefore affected the sequencing stage. With the amount of bacterial RNA present, a hypothesis needs to be put forward to explain why these problems may have arisen. Given that the entire *Drosophila* digestive tract was present in the sample used, the gut flora would be pervasive throughout the input material, likely providing a huge source of non-target RNA. Although the polyA selection step would be expected to select against bacterial RNA, this as well as the oligo-dT would be expected to massively deplete rRNA. It is possible (based on the proportion of desired RNA versus rRNA and bacterial RNA) that the sheer amount of bacterial RNA essentially saturated the selection beads, preventing efficient selection and depletion.

Although RNA-sequencing has been carried out successfully on *Drosophila* L3 wing imaginal discs (without the same issue of bacterial contamination), the time required for the dissection process for isolating WIDs, along with the low input RNA per disc makes this an unfeasible approach. A compromise between these two may, however, be possible. *Drosophila* L3 heads can easily be dissected by separating the top third of the larvae from the bottom two thirds. This third contains a substantial amount of RNA (including the aforementioned WIDs,) and would exclude the gut fauna (hopefully eliminating the problem). This would provide a straightforward way to separate the major source of contamination whilst retaining a lot of input sample, with only a small amount of time added for the simple dissection step.

7.3 Successful generation of a novel dataset examining regulation of RNAs and lncRNAs in the absence of Pacman and Dis3L2 in both polysomal and total lysate *Drosophila* samples

7.3.1 An overview of RNA regulation by Pacman and Dis3L2 in *Drosophila* whole L3 larvae

Although the dataset comes with its limitations, it still allowed an overview of the degradation patterns of the exoribonucleases of interest, even without comparison to other datasets. Even with the stringent filters introduced by EdgeR, it gave 610 RNAs upregulated more than 2-fold in the absence of Pacman, with 638 RNAs upregulated more than 2-fold in the absence of Dis3L2. Although this data is of limited use in answering the questions of lncRNA degradation and lncRNA non-canonical translation, it still provides a novel dataset for examining exoribonuclease degradation in whole *Drosophila* L3 larvae (as other RNA-seq datasets examining this have been in L3 WID null mutants, and in Pacman dsRNA knockdown S2 cells). Only one lncRNA that passes these EdgeR filters is more than 2-fold upregulated in the Pacman mutant (*CR43334*), with another three lncRNA more than 2-fold upregulated in the Dis3L2 (*CR43334, cherub, and CR40469*). Although the strength of RNA-sequencing based techniques is their genome-wide approach, this initial overview showed that with the low read depth of this dataset, highly stringent filters cannot be used for this dataset without risking cutoff of most potential candidates. This strengthened the case for using this dataset as an informative dataset, rather than a comprehensive one, and encouraged the later use of cross-comparisons with other existing datasets.

7.3.2 Comparison with other datasets allows a meaningful and informative overview of lncRNA degradation by Pacman in *Drosophila* whole L3 larvae

Pacman depleted *Drosophila* tissues have undergone RNA-sequencing in previous studies. For Pacman, the S2 cell Pacman dsRNA knockdown by Antic *et al*., as well as the L3 WIDs from Pacman null mutants from Jones *et al*. RNA-sequencing datasets were available to download and re-analyse (now in the context of lncRNAs rather than their original context). Both of these datasets featured a higher read depth than in the novel data, allowing analysis of these datasets separately, as well as subsequent comparison to the low-confidence novel poly-ribo-seq datasets, which can allow them to be meaningfully examined with less stringent statistical filters than EdgeR filters used in previous comparisons.

Of lncRNAs with Pacman mutant reads detected, 49 were found in L3 larvae, and 205 in S2 cells, with only 6 common to both; reducing to 114 in the S2 cell data, and 45 in the L3 larval data, with only 1 lncRNA common to both when a cut-off of 1.5-fold increase was implemented. In L3 WIDs, 161 lncRNAs had a higher read count in the Pacman mutant than in the isogenic control (with 8 common to both whole L3 and WIDs), with only 105 potential Pacman degraded lncRNAs passing the 1.5-fold change threshold (with only 3 lncRNA common to both the whole L3 and the WIDs).

Whilst stable cell culture models and whole tissue or whole organism models are generally considered and expected to be wildly different (due to the nature of constant tissue culture growth, lack of biological systems and structure, homogenous culture, etc.) it is interesting to observe the similarities and differences that can be seen between whole L3 samples and the WID samples dissected from L3. With only the 3 lncRNAs (*CR44506, CR31781*, and *CR32111*) common to being increased more than 1.5-fold between the two Pacman mutant models, we can see that although there do appear to be conserved examples of polysome-associated Pacman targets across multiple tissues at this developmental timepoint (and we should not assume that all of them are necessarily identified here,) spatial and tissue specific expression and control are important to this kind of transcript (as they are, to many steps, all throughout control of gene expression and function).

As can be seen in Figure 7.1, *CR44506* and *CR32111* expression are both very low throughout, with low enough read count and variation to prevent meaningful speculation as to tissue specificity throughout L3 (which presents one of the barriers to examining many lowly expressed and poorly annotated lncRNAs). The third lncRNA, CR31781, is more highly expressed, and can be seen to be much more variably abundant throughout the L3 tissues in which it is annotated in the modENCODe data (from lowest to highest; L3 fat body, L3 imaginal disc, L3 salivary gland, L3 digestive system, L3 central nervous system, L3 carcass). Although this information is not sufficient to speculate meaningfully on localised roles, it does demonstrate localised regulation of these transcripts' abundance as a factor that must be considered and would impact the weighting of RNA detected from a whole L3 sample.

Although only a minority of Pacman degraded lncRNAs were common between the different models, this comparison allows identification of more lncRNAs that may be degraded by Pacman (in the higher read depth WID dataset); and with the conservation comparison, some idea of lncRNAs that might have a role pervasively through the organism, with consistent degradation by the Pacman enzyme, can be gathered (and potentially followed in later work).

7.3.3 Comparison with other datasets allows a meaningful and informative overview of lncRNA degradation by Dis3L2 in *Drosophila* whole L3 larvae

Dis3L2 depleted *Drosophila* tissues have undergone RNA-sequencing before. In Dis3L2, the L3 WIDs from Pacman null mutants from Towler *et al.* RNA-sequencing datasets were available to download and re-analyse (now in the context of lncRNAs rather than their original context). Similarly to the Pacman comparisons, the pre-existing dataset features a higher read depth than in the novel data, allowing analysis of these datasets separately, as well as subsequent comparison to the low-confidence novel poly-ribo-seq datasets, which can allow them to be meaningfully examined with less stringent statistical filters than EdgeR filters used in previous comparisons.

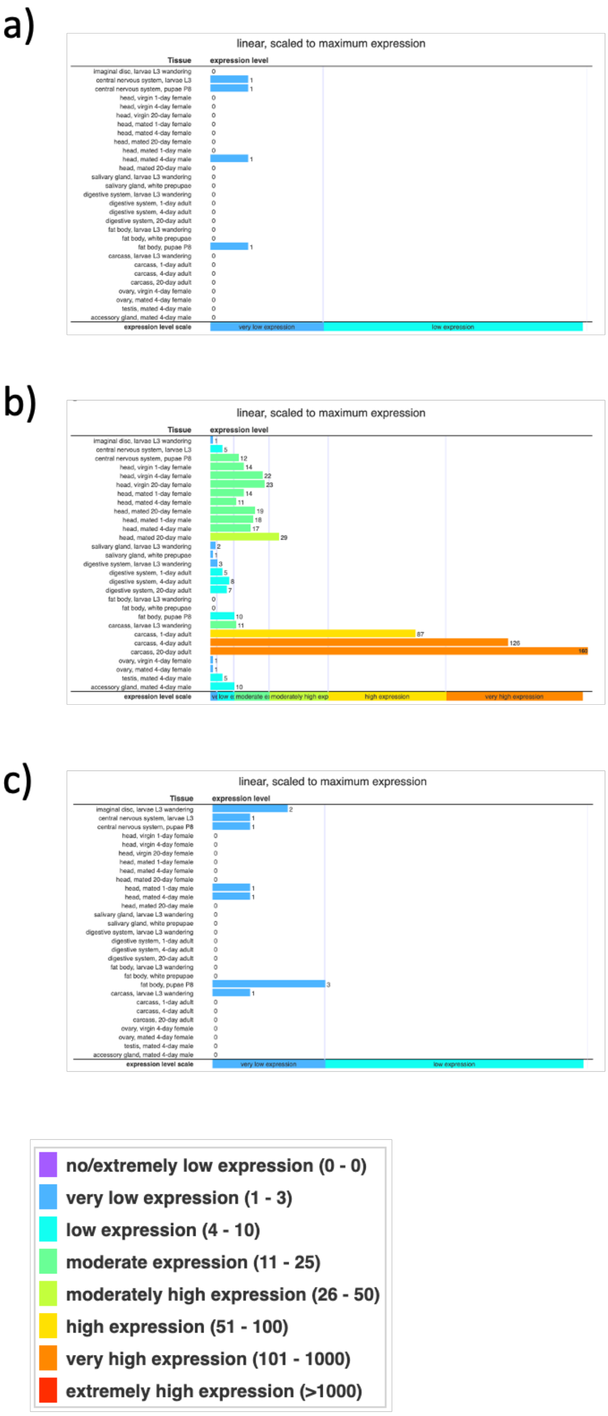


Figure 7.1 – modENCODE expression data for whole L3 and WID conserved polysome-associated Pacman targets

Although mostly lowly expressed, and with very little other information available, an initial overview of tissue specific abundance of the candidate lncRNAs can be gathered from available modENCODe data. Panel (a) ahows *CR44506* expression, panel (b) shows *CR31781* expression, and panel (c) shows *CR32111* expression throughout recorded RNA-seq data in available *Drosophila* tissues.

Of lncRNAs with Dis3L2 mutant reads detected 129 were found in L3 WIDs, with 56 having a higher read count in the absence of Dis3L2, and only 40 of them passing a 1.5-fold cut-off. From the novel sequencing data, 55 lncRNAs were identified with higher expression in the Dis3L2 mutant, and 48 of these passing the 1.5-fold increase threshold. Of the lncRNAs with increased normalised read count in the absence of Dis3L2 (56 from wing imaginal disc sequencing, 55 from novel L3 sequencing data,) 3 were common to both lists. Of the lncRNAs which increased by 1.5-fold or more (40 from wing imaginal disc sequencing, 48 from novel L3 sequencing data,) only 1 was common to both (*CR45517*).

As discussed previously (in 7.3.2,) the differences between stable cell culture models and whole tissue or whole organism models are unsurprising; but comparing the differences between localised L3 WID and total L3 can provide some small insight into the pervasiveness and localised control of targets. Again, only a minority are seen to be conserved between the two models (despite the WID model being present within the whole L3, the WID makes up so little of the overall L3 that any subtle differences that are specific to that tissue might not be noticeable in the total L3). As an informative approach to identify candidate lncRNAs conserved in their degradation by Dis3L2 between the wing imaginal disc tissue and the whole L3 organism, this still provides a useful resource. Once again, some idea of lncRNAs that might have a role pervasively through the organism, with consistent degradation by the Dis3L2 enzyme, can be gathered. Figure 7.2 shows the available modENCODE data for localized tissue expression of *CR45517*. As with *CR44506* and *CR32111*, expression is very low throughout, with low enough read count and variation to prevent meaningful speculation as to tissue specificity throughout L3 larvae.

7.3.4 Comparison of lncRNA degradation datasets gives preliminary profile of lncRNA decay by Pacman and Dis3L2 in *Drosophila melanogaster*

In summary, these comparisons provide a preliminary, global, profile for the degradation of lncRNAs in *Drosophila*, along with a limited look at the conservation of their degradation between whole L3 and L3 WIDs. This could be a useful resource for any RNA decay work in *Drosophila* in the future, and the comparisons specifically with

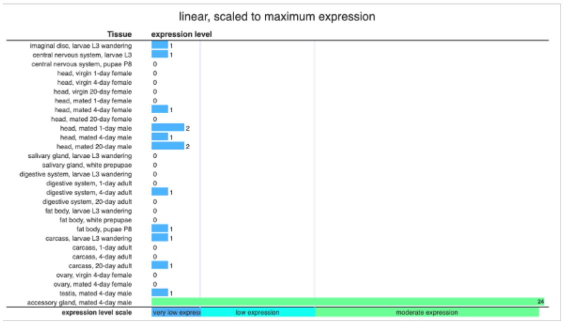
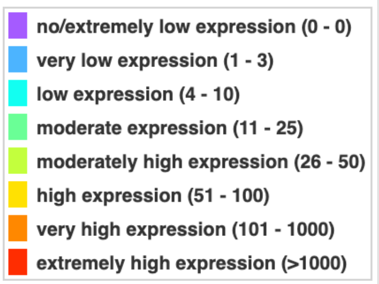


Figure 7.2 – modENCODE expression data for whole L3 and WID conserved polysome-associated Dis3L2 target

Although mostly lowly expressed, and with very little other information available, an initial overview of tissue specific abundance of the candidate lncRNA can be gathered from available modENCODe data. The figure ahows *CR45517* expression throughout recorded RNA-seq data in available *Drosophila* tissues.



🟪 no/extremely low expression (0 - 0)

🟦 very low expression (1 - 3)

🟦 low expression (4 - 10)

🟩 moderate expression (11 - 25)

🟩 moderately high expression (26 - 50)

🟨 high expression (51 - 100)

🟧 very high expression (101 - 1000)

🟥 extremely high expression (>1000)

lncRNAs provides the further opportunity to identify functional roles for unknown lncRNA, and a dataset that could subsequently be examined in regard to their distribution between total lysate and polysomal RNA. Almost none of the lncRNAs that arise have significant annotation, or a biological profile, making it difficult to draw conclusions without more specific questions, or further comparisons.

7.4 An overview of polysome associated lncRNAs in *Drosophila* whole L3 larvae

7.4.1 Examination of polysomal fractions allows identification of potential smORF bearing lncRNAs in *Drosophila* whole L3 larvae

Due to the low coverage achieved in this poly-ribo-seq experiment, the first approach used here was to identify potentially translated lncRNAs used all replicates from all three tested genotypes (isogenic control, Pacman null mutant, and Dis3L2 null mutant). By observing all lncRNAs with presence in the polysomal fractions, a total of 33 candidate polysome-associated lncRNAs with an average of at least 3 reads were found (even with this limited read depth). With further stringent filters applied, and comparison between polysomal and total lysate samples, two lncRNAs showed enrichment on the polysome compared to total lysate. These examples show that in some lncRNAs, not only are they present on the polysome, but specifically enriched, implying active translation (or any other polysome associated role) likely as a primary function. Further specific exploration of the mechanism for their degradation could potentially reveal the details of how highly polysomally associated lncRNAs undergo decay, and whether there’s any link between the potential translation and the degradation.

7.5 Analysis of polysomally present Pacman and Dis3L2 regulated lncRNAs identifies candidates for exploration of co-translational degradation

7.5.1 By identifying exoribonuclease targeted lncRNAs and comparing with data describing polysome association initial candidates for co-translational degradation can be identified

Using stringent criteria for polysomally present lncRNA also substantially upregulated in the absence of Pacman and Dis3L2 respectively, the most promising candidates for lncRNAs degraded by the exoribonucleases of interest to the project (Pacman and Dis3L2), that are also likely to be translated (as gauged by their polysome association) can be identified. Although the genome wide dataset produced by this work is only informative (rather than a comprehensive testing of the relationships between translation and degradation), it does help identify those lncRNAs that may be worth subsequent full mechanistic investigation. This work did in fact achieve this; and although the limitations on this as a thesis project prevents all of these from being followed up, and overarching questions being comprehensively answered, a working base for further work has been established by this work.

Two of these potential targets (*HsrOmega* and *CR40469*) were the subject of follow-up experiments (described below) in order to ascertain more information about their potential translation. Relatively little is known about the role and function of these genes (as with most lncRNAs); *HsrOmega* has been inferred to have roles in nuclear speck organization, regulation of apoptosis, regulation of cellular protein metabolism, regulation of JNK cascade, protein localization, and regulation of proteolysis, although these roles have not been sufficiently validated using direct molecular techniques. *CR40469* has no known or speculated biological or molecular function.

7.5.2 Further analysis and testing identifies at least one likely protein-coding ORF in a Dis3L2 degraded lncRNA in *Drosophila*

Following up on the two specific targets (*HsrOmega* and *CR40469*) identified as regulated by Pacman and Dis3L2 respectively while also present on the polysome, relatively simple experiments were able to be used to test predicted ORFs within these genes. By cloning a GFP reporter in-frame of the start codon of the predicted ORF, *CR40469* was shown to translate the GFP reporter, providing strong evidence of the translational activity of the tested ORF within the *CR40469* gene. The *HsrOmega* ORF tested did not show any GFP activity, although it was only one of multiple potential ORFs within the gene region. This simple approach for testing translation is not sufficient to prove with certainty either way, nonetheless the work does strongly suggest that *CR40469* at least does undergo translation, and the bioinformatic data from the poly-ribo-seq backs up the translation of both genes. This support for translation, and differential abundance in the absence of Pacman and Dis3L2, means that either of these may have co-translational aspects to their degradation, and beg the question of how important the translation of candidates like this may be to their degradation.

7.6 Comparison with other datasets allows identification of promising smORF bearing lncRNAs in *Drosophila* whole L3 larvae

Although the sequencing depth on the polysomal data from the novel data is better than the total lysate RNA, it still suffers from the low depth obscuring potential candidates. Once again, in order to get the most from the data, multiple comparisons between re-analysed pre-existing datasets allows for an improved chance of identifying translated lncRNAs, as well as potentially their ORFs, due to pile-up locations in different datasets.

A paper by Zhang *et al*. provided ribo-seq data from *Drosophila* S2 cells treated by a translation elongation inhibiting drug called Harringtonine. Not only does this allow an examination of which lncRNAs are found on the polysome in S2 cells, but also the profile of where ribosomes bind to RNA transcripts, as the blocking of elongation leaves ribosome pile up at the start codon of any potential open reading frame. As the ribosome protects RNA from digestion, the sequenced transcripts are then strong indicators of these sites and allow precise examination of these regions. With 252 lncRNA detected in these ribosome-protected transcripts (with the list shortened by the application of filters), this dataset could be compared to the novel poly-ribo-seq data.

Similarly, pre-existing data from the Couso lab (Aspden *et al*., Patraquim *et al*.) could be re-analysed to provide further examples, and comparisons with the novel data. These samples (from three embryonic stages, and from S2 cells) were treated only with cycloheximide (during and post-lysis), as with the whole L3 samples in the novel datasets. Having such high similarity in lysis conditions to the novel dataset, these were ideal for conservation comparisons.

Although the nature of the RNAs being examined (often lowly expressed and poorly annotated) provides a substantial barrier to identifying and highlighting specific profiles, points of interest can still be found. The comparison between different developmental stages and cell lines allowed highlighting of promising lncRNAs present on the polysome, particularly those present consistently enough to likely be conserved on the polysome by purpose rather than chance. They may provide promising indicators of transcripts that may have important roles, spanning multiple developmental timepoints.

For example, the percentage of the top 50 polysomal lncRNAs conserved between poly-ribo-seq datasets from different *Drosophila* developmental timepoints was reliably seen to remain between 32-40%, a proportion that even remained consistent in S2 cells. This finding is encouraging, showing that despite the limited depth of the novel sequencing data, the findings are real, and comparable to similar datasets. This provides an increased confidence in the novel dataset, as well as demonstrating the likely presence of a key cohort of biologically important, polysome associated lncRNAs. To demonstrate this, a few examples of consistently polysome-associated lncRNAs with some known biological profile can be identified (despite the shortcomings of lncRNA annotations).

The lncRNA *flam*, or *flamenco* was detected with at least 20 polysomal reads in all tested embryonic stages and S2 cells, and at least 3 polysomal reads in the novel L3 sequencing data. Although not detectable at any stage, in any tissue, in the modENCODE sequencing data available on FlyBase, there is still some information available. The *flamenco* gene is projected to have some role related to male courtship behaviour, oogenesis, and ovarian follicle cell development (all inferred from mutant phenotypes (163, 164)). The gene is understood to control mobilisation of the endogenous retrovirus, gypsy, through the repeat-associated small interfering RNA silencing pathway and is required somatically for morphogenesis of the follicular epithelium. Gene ontology clustering links *flamenco* to development, reproduction, and behaviour.

Both *roX1* and *roX2* are also detected with at least 20 polysomal reads in all tested embryonic stages and S2 cells, and at least 3 polysomal reads in the novel L3 sequencing data. These two genes are some of the best examples of lncRNAs with a known and studied role. In Drosophila, the RNA on the X genes, *roX1* and *roX2*, are expressed in males, and regulate the assembly of the Male Specific Lethal (MSL) complex in Drosophila; a chromatin modifier that functions in histone modification (44). The recruitment and binding of MSL proteins by high affinity sequences on the nascent *roX* transcripts covering the X chromosome allows the assembly of the active MSL complex, which can then spread in cis, allowing chromatin restructuring and hyperactivation of specific regions of the chromosome. Compared to many lncRNAs, these are highly expressed as well as well studied, providing examples of biologically important polysome-associated lncRNAs that may still have new light shed on them by their association with the polysome.

Finally, the lncRNA *HsrOmega* also detected with at least 20 polysomal reads in all tested embryonic stages and S2 cells, and at least 3 polysomal reads in the novel L3 sequencing data. It is also one of the lncRNAs seen to increase in its abundance in the absence of Pacman. *HsrOmega* is known to associate with a variety of heterogeneous nuclear RNA-binding proteins and other RNA-binding proteins to assemble the nucleoplasmic omega speckles (136). It has also been shown that RNAi depletion of *HsrOmega* is able to dominantly suppress apoptosis, in eye and other imaginal discs, triggered by induced expression of *reaper*, *grim*, or caspases (138). Given that Pacman is known to act through the pro-apoptotic *reaper* pathway to exert its effects, and that depletion of Pacman causes significant increase in the expression of the pro-apoptotic mRNAs, *hid* and *reaper*, with this increase mostly occurring at the post-transcriptional level, this candidate presents an interesting target for investigating the roles of lncRNAs in the workings of the decay machinery that regulates them. The implication that this lncRNA is not only degraded by Pacman, but that its depletion is able to suppress some of the same pro-apoptotic signals that arise in the absence of normal degradation by Pacman, whilst also having a role on the polysome begs further investigation. It could be the case that translational activity of *HsrOmega* is necessary for the ordinary degradation of the lncRNA by Pacman, and that this co-translational degradation is important to the action of Pacman in degrading the pro-apoptotic *reaper* and *hid*; perhaps even producing a small functional peptide that has a role in controlling this pathway.

Overall, this part of the work allows the compilation of shortlists of potentially translated lncRNAs, with the additional context of when in *Drosophila* development they may have a (polysomal) role, and how well conserved this role is. The exciting thing about producing this dataset with the additional layer of degradation to examine with paired samples (paired total lysate and polysomal RNA samples in each genotype), is the ability to examine the degradation activity of potentially translating lncRNA, and how one might affect the other.

7.7 Potential interplay between the translation and exoribonuclease degradation of lncRNAs in *Drosophila*

As mentioned, the limited scope of this project is not sufficient to prove definitively the relationship between translation and degradation of lncRNA. That said, it makes promising steps in the direction of fully exploring these topics. The multiple datasets allow not only an overview of all of the RNA degradation in whole L3 larvae (including of lncRNAs) but also allows observation of how this degradation occurs on the polysome and polysomal RNA.

RNA degradation is already known to be a targeted and specific process, for example with 3’ polyuridylation targeting specific transcripts for preferential degradation by Dis3L2(165), or by specific RNA binding proteins or miRNAs binding to the 3’ UTRs of target RNAs, as speculated for degradation of certain transcripts by Pacman(34). Either of these mechanisms would be viable explanations for why certain lncRNAs may be sensitive to Pacman or Dis3L2 degradation. What this project provides, in terms of examining specific degradation, is a glimpse into how co-translational degradation may be facilitating specific degradation of lncRNAs by Pacman, and potentially Dis3L2 as well.

The candidates identified and followed up can be identified as strong starting points for further experimental work into exploring the interplay between translation and exoribonuclease degradation. As transcripts that are not only differentially abundant in the absence of Pacman and Dis3L2, but regulated by those exoribonucleases while on the polysome, use of further techniques (such as protein-pulldown techniques, qPCR on polysomal fractions, and genetic manipulation of target genes) and conditions affecting translation and degradation (further use of compounds that block decay and translation at various stages, and observing the effects on RNA abundance and association with the polysome) will be required to further dissect the links between them.

7.8 Limitations of this project

Although a large amount of work went into the project, its limitations must be acknowledged. Firstly, the generation of input material and an overall model for examining degradation of lncRNAs in both total lysate and polysomal RNA came across multiple issues. The initial attempts to recreate the dsRNA Pacman knockdown by Antic *et al.*, and to produce it in Dis3L2, were not hugely successful. The concentration of Pacman dsRNA required to induce the knockdown was substantially higher than that used in the work of Antic *et al.*, for reasons unknown. The reliability of the knockdown was also low, as it was not produced over all attempts, even with the same protocol. Attempts were made to phenotype these treated cells, but the consistency of the knockdown was not able to be reliably measured. With the eventual reveal that the conditions that the cells were kept in were problematic (and ultimately lethal) for cells and cell growth, all work on the dsRNA knockdown and any attempted phenotype must (unfortunately) be acknowledged as unreliable. The incubator, (as mentioned in 4.5) was identified as being incapable of supporting a humidity suitable for S2 cell culture.

Instead, rapid airflow in the incubator used had been sporadically dehydrating cell stocks, causing sudden death of cells Any leads from the dsRNA knockdown protocol or phenotyping must be repeated if any conclusion were to be drawn from it.

Following this, substantial time and effort was invested in generating a stable mutant for Pacman and Dis3L2 in *Drosophila* S2 cells. This could have provided a useful resource, allowing new phenotyping in S2 cells, easily amplified stable stocks, able to be used across the vast array of techniques that S2 cells are well adapted to. Compared to the work by Antic *et al*. this would also provide a total null mutant (and therefore total exoribonuclease depletion) rather than a partial knockdown. The unfortunate death of these cells and time limitations called an end to this experiment within the scope of this project. Unfortunately, not even whether the gene editing was successful could be established, and from the work as it stands, we cannot even be sure that a total null mutation of either exoribonuclease is viable in S2 cells, (Pacman null mutants are lethal in whole *Drosophila* at pupation, and the existing Pacman depletion in S2 cells was incomplete).

The extensive optimization of polysome fractionation in whole L3 larvae will be a useful resource for future work aiming to produce high quality polysome fractionation in other difficult sample types (robust cells, whole organism, robust tissues). However, although high quality polysome profiles were produced, by the end of the sequencing, it was clear that the over-abundance of bacterial RNA from the *Drosophila* gut dramatically reduced the read depth of the sequencing. This limited how many definitive conclusions could be gained from the entire project and prevented the high throughput datas*et al*one from being able to identify trends between degradation of lncRNAs and how polysomal localization might affect that. This was somewhat overcome by the use of multiple datasets, and the comparisons that could then be made between them. Nonetheless, a repeat of the experiment, or a similar variant, would allow a great resource. Looking at how much information can still be gained from the data set as is (especially if looking at RNAs with higher abundancies than lncRNAs tend to), a similar dataset without bacterial RNA and high levels of rRNA contamination could reveal substantially more than the novel poly-ribo-seq as it stands now.

Finally, although an early investigation into some of the prime candidates has been begun within this project, with the current work, not much can be asserted as certain. Examples of lncRNAs not only degraded by Pacman or Dis3L2 but degraded on the polysome were found. Of these, the strongest candidates for both Pacman and Dis3L2 were taken forward to be tested by cloning in a GFP reporter to a potential ORF. One of the two showed translation of the GFP, providing further evidence of translation, but further experimentation would be required to ensure that translation is from the ORF of interest, and that it isn't read-through from a previous start codon. Similarly, only one potential ORF was tested for each gene, meaning that the overall translational activity of each gene cannot be confidently stated at this point.

7.9 Future work

In future work, carrying out a similar experiment (likely with *Drosophila* larval heads as the input sample, to avoid bacterial contamination) treated with a translation blocking drug such as harringtonine or puromycin would provide interesting information for how necessary the translational activity is to the degradation of targets identified by the genome-wide approach. A drug like puromycin, (which works by causing premature chain termination during ribosomal translation, due to it entering the A-site and transferring to the growing chain, causing the formation of a puromycylated nascent chain and premature chain release) would allow observation of the abundance of the transcripts on the polysome, and their degradation in conditions where translation terminates prematurely.

Similarly, treating with an elongation blocking drug like harringtonine, which immobilises ribosomes immediately after initiation, would allow a detailed observation of initiation sites, and could explore whether initiation is sufficient for a potential handover mechanism between translation and degradation machinery (as might be the case for ribosomal machinery and Dis3L2, which work in opposite directions).

By the thorough examination of the impact of inhibiting translation on the stability of lncRNAs like those identified in this work, examples of definitive co-translational and translation-dependent degradation in lncRNAs (in *Drosophila*) might be identified, providing targets to further study the mechanism (if it is as such) using protein-based and protein-RNA techniques.

To better understand the links and associations between Dis3L2 and the ribosome, carrying out a similar poly-ribo-seq experiment on (both control and Dis3L2 knockout) human cell culture (in which association between Dis3L2 and the ribosome has already been definitively shown(71)) might allow comparisons to be made to a model in which the relevance of the interaction and direct association between Dis3L2 and the ribosome is already established.

In terms of gaining a deeper understanding of lncRNA regulation and degradation in general, similar experiments to gain RNA-sequencing datasets could be carried out on equivalent *Drosophila* tissues in which key components of other RNA-degradation pathways have been knocked out (such as endonuclease cleavage, or nonsense mediated decay).

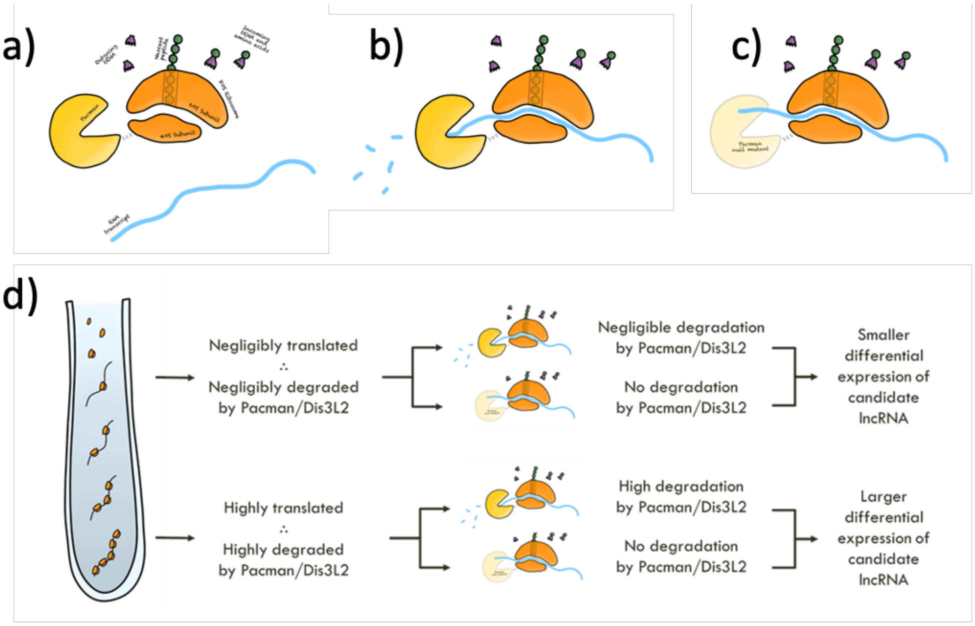
Extrapolating out from this, a better understanding of these topics could reveal the workings of yet another layer in regulating gene expression. From the earliest concept of the Central Dogma to now, many layers of regulation have since been uncovered, and understood to various extents, and the translation, degradation, and the interplay between the two in lncRNAs is another layer that must be examined carefully in order to further our ever-growing understanding of gene regulation.

Finally, an in-depth characterization and phenotyping of candidate lncRNAs (particularly *HsrOmega* and *CR40469*,) would not only be the natural and necessary next step in a comprehensive examination of these genes (elucidating their biological roles and relevance) but would allow better informed speculation on the regulation of translation and degradation of these genes and their products. Establishing a biological role, whether they produce functional peptides, and the circumstances impacting their translation and degradation would provide a much more comprehensive of what is currently an exploratory approach to these relatively unknown genes and their regulation.

7.10 Concluding remarks

The work here provides a stepping-stone towards a true understanding of the interplay between translation and degradation in the context of lncRNAs. By examining the genome-wide profile of lncRNA degradation, we can add to existing datasets, increasing our ability to produce in depth maps of degradation pathways in *Drosophila*, and compare the degradation of lncRNAs (by Pacman and Dis3L2) to the degradation of canonical RNAs. Alongside this, the use of poly-ribo-seq, rather than conventional sequencing (or even ribo-seq) allows confident examination of polysome associated lncRNAs, likely the most promising source of lncRNAs that produce hitherto unknown and novel small peptides. Due to the samples being paired (for polysomal and total lysate RNA) for each replicate of each genotype, we can therefore examine the degradation of lncRNAs for those genes that have potential translational activity. Even with the limited depth available from this dataset, examples of polysomally present lncRNAs being degraded by Pacman and Dis3L2 on the polysome could be identified, and an experimental model has been developed that allows insight and meaningful speculation into a potential model for the sensitivity of certain lncRNAs to degradation by Pacman and Dis3L2 (Figure 7.3).

This data was provided the chance to further inform the understanding of these topics by way of comparison to other similar datasets and follow up experimental work. By large scale sorting and plotting of degradation targets, multiple lists of lncRNAs degraded by Pacman and Dis3L2 across different *Drosophila* developmental stages, cell lines, and tissues could be compiled, once again providing a strong starting point for any further investigation into lncRNA degradation. Similarly, the comparison with multiple poly-ribo-seq and ribo-seq datasets allowed the compilation of lists and plots of lncRNA with suspected translational activity, and even identification of promising novel ORFs. With previous examples of biologically relevant ORFs having been discovered in a similar fashion, this provides an excellent resource for further work into better cataloguing non-canonical peptides in *Drosophila*, and large-scale knockdown screens or cloning in reporters could be used to pick out multiple actively translated lncRNA-encoded peptides, which can then be phenotyped and characterized more fully, beginning to fill the substantial gaps that exist in the understanding of the roles of



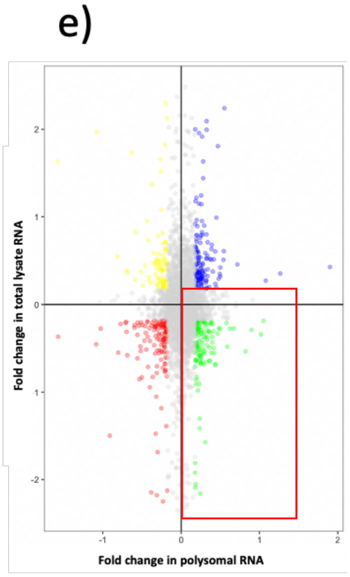


Figure 7.3 – A diagram summarising the hypothetical model for co-translational degradation of lncRNAs as well as how this can be explored using poly-ribo-seq

(a) This panel shows the components of a Pacman-associated ribosome, along with the components necessary for translation (tRNAs, amino acids, and RNA transcript).

(b) This panel shows a simple visual representation of Pacman following a translating ribosome, and degrading the RNA transcript following its translation.

(c) This panel shows a simple visual representation of a non-functional mutant Pacman following a translating ribosome, and failing to degrade the RNA transcript following its translation.

(d) This panel shows a model for how polysome fractionation (and subsequent digest and sequencing) can be used to produce results that may be interpreted in a way that sheds light on the relationship between translation and degradation.
If translation is necessary for the degradation of these transcripts, then the exoribonuclease sensitive lncRNA transcripts found, that are polysome-associated (potentially actively translated), will be differentially abundant to a significantly higher degree in the absence of the exoribonuclease than the same transcript in the total RNA of the same sample (lower potential translational activity), due to the degradation only taking place in the presence of both a translating ribosome and a functional exoribonuclease.

(e) A visual highlighting the combination of increased abundance in polysomal RNA and no substantial change (or a decrease, indicating a possible feedback signal if not degraded) that indicates co-translational degradation in an exoribonuclease mutant sample.

lncRNAs. Again, this can be combined with the degradation data, not only to observe how degradation for any peptide-encoding lncRNAs may occur, but to observe possible links between the translation and degradation.

1. Crick F. Central dogma of molecular biology. Nature. 1970;227(5258):561-3.
2. Dai X, Zhang S, Zaleta-Rivera K. RNA: interactions drive functionalities. Mol Biol Rep. 2020;47(2):1413-34.
3. Lu Z, Matera AG. Developmental analysis of spliceosomal snRNA isoform expression. G3 (Bethesda). 2014;5(1):103-10.
4. Luhrmann R, Kastner B, Bach M. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. Biochim Biophys Acta. 1990;1087(3):265-92.
5. Yin Y, Lu JY, Zhang X, Shao W, Xu Y, Li P, et al. U1 snRNP regulates chromatin retention of noncoding RNAs. Nature. 2020;580(7801):147-50.
6. Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, Chen W, et al. A human snoRNA with microRNA-like functions. Mol Cell. 2008;32(4):519-28.
7. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004;23(20):4051-60.
8. Cai X, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA. 2004;10(12):1957-66.
9. Bracht J, Hunter S, Eachus R, Weeks P, Pasquinelli AE. Trans-splicing and polyadenylation of let-7 microRNA primary transcripts. RNA. 2004;10(10):1586-94.
10. Jonas S, Izaurralde E. Towards a molecular understanding of microRNA-mediated gene silencing. Nat Rev Genet. 2015;16(7):421-33.
11. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22(9):1775-89.
12. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. Cell. 2013;152(6):1237-51.
13. Annunziato AT, Hansen JC. Role of histone acetylation in the assembly and modulation of chromatin structures. Gene Expr. 2000;9(1-2):37-61.
14. Jaskiewicz M, Conrath U, Peterhansel C. Chromatin modification acts as a memory for systemic acquired resistance in the plant stress response. EMBO Rep. 2011;12(1):50-5.
15. Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013;153(2):320-34.
16. Gornemann J, Kotovic KM, Hujer K, Neugebauer KM. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol Cell. 2005;19(1):53-63.
17. Zhang G, Taneja KL, Singer RH, Green MR. Localization of pre-mRNA splicing in mammalian nuclei. Nature. 1994;372(6508):809-12.
18. Neugebauer KM. On the importance of being co-transcriptional. J Cell Sci. 2002;115(Pt 20):3865-71.
19. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet. 2007;8(10):749-61.
20. Paronetto MP, Bernardis I, Volpe E, Bechara E, Sebestyen E, Eyras E, et al. Regulation of FAS exon definition and apoptosis by the Ewing sarcoma protein. Cell Rep. 2014;7(4):1211-26.
21. Mankodi A, Takahashi MP, Jiang H, Beck CL, Bowers WJ, Moxley RT, et al. Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. Mol Cell. 2002;10(1):35-44.
22. Cieply B, Carstens RP. Functional roles of alternative splicing factors in human disease. Wiley Interdiscip Rev RNA. 2015;6(3):311-26.
23. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature. 2011;473(7347):337-42.

24. Amrani N, Dong S, He F, Ganesan R, Ghosh S, Kervestin S, et al. Aberrant termination triggers nonsense-mediated mRNA decay. Biochem Soc Trans. 2006;34(Pt 1):39-42.
25. Garneau NL, Wilusz J, Wilusz CJ. The highways and byways of mRNA decay. Nat Rev Mol Cell Biol. 2007;8(2):113-26.
26. Parker R, Song H. The enzymes and control of eukaryotic mRNA turnover. Nat Struct Mol Biol. 2004;11(2):121-7.
27. Newbury SF. Control of mRNA stability in eukaryotes. Biochem Soc Trans. 2006;34(Pt 1):30-4.
28. Amrani N, Sachs MS, Jacobson A. Early nonsense: mRNA decay solves a translational problem. Nat Rev Mol Cell Biol. 2006;7(6):415-25.
29. Fadda A, Farber V, Droll D, Clayton C. The roles of 3'-exoribonucleases and the exosome in trypanosome mRNA degradation. RNA. 2013;19(7):937-47.
30. Manful T, Fadda A, Clayton C. The role of the 5'-3' exoribonuclease XRNA in transcriptome-wide mRNA degradation. RNA. 2011;17(11):2039-47.
31. Malecki M, Viegas SC, Carneiro T, Golik P, Dressaire C, Ferreira MG, et al. The exoribonuclease Dis3L2 defines a novel eukaryotic RNA degradation pathway. EMBO J. 2013;32(13):1842-54.
32. Wang M, Ly M, Lugowski A, Laver JD, Lipshitz HD, Smibert CA, et al. ME31B globally represses maternal mRNAs by two distinct mechanisms during the Drosophila maternal-to-zygotic transition. Elife. 2017;6.
33. Jones CI, Pashler AL, Towler BP, Robinson SR, Newbury SF. RNA-seq reveals post-transcriptional regulation of Drosophila insulin-like peptide dilp8 and the neuropeptide-like precursor Nplp2 by the exoribonuclease Pacman/XRN1. Nucleic Acids Res. 2016;44(1):267-80.
34. Waldron JA, Jones CI, Towler BP, Pashler AL, Grima DP, Hebbes S, et al. Xrn1/Pacman affects apoptosis and regulates expression of hid and reaper. Biol Open. 2015;4(5):649-60.
35. Jones CI, Zabolotskaya MV, Newbury SF. The 5' → 3' exoribonuclease XRN1/Pacman and its functions in cellular processes and development. Wiley Interdiscip Rev RNA. 2012;3(4):455-68.
36. Zekri L, Kuzuoglu-Ozturk D, Izaurralde E. GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. EMBO J. 2013;32(7):1052-65.
37. Braun JE, Huntzinger E, Fauser M, Izaurralde E. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol Cell. 2011;44(1):120-33.
38. Nishihara T, Zekri L, Braun JE, Izaurralde E. miRISC recruits decapping factors to miRNA targets to enhance their degradation. Nucleic Acids Res. 2013;41(18):8692-705.
39. Carballo E, Blackshear PJ. Roles of tumor necrosis factor-alpha receptor subtypes in the pathogenesis of the tristetraprolin-deficiency syndrome. Blood. 2001;98(8):2389-95.
40. Akamatsu W, Fujihara H, Mitsuhashi T, Yano M, Shibata S, Hayakawa Y, et al. The RNA-binding protein HuD regulates neuronal cell identity and maturation. Proc Natl Acad Sci U S A. 2005;102(12):4625-30.
41. Richter JD, Sonenberg N. Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature. 2005;433(7025):477-80.
42. Gingras AC, Raught B, Sonenberg N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. Annu Rev Biochem. 1999;68:913-63.

43. Clemens MJ. Initiation factor eIF2 alpha phosphorylation in stress responses and apoptosis. Prog Mol Subcell Biol. 2001;27:57-89.

44. Rogoyski OM, Pueyo JI, Couso JP, Newbury SF. Functions of long non-coding RNAs in human disease and their conservation in Drosophila development. Biochem Soc Trans. 2017;45(4):895-904.

45. Pashler AL, Towler BP, Jones CI, Newbury SF. The roles of the exoribonucleases DIS3L2 and XRN1 in human disease. Biochem Soc Trans. 2016;44(5):1377-84.

46. Towler BP, Jones CI, Harper KL, Waldron JA, Newbury SF. A novel role for the 3'-5' exoribonuclease Dis3L2 in controlling cell proliferation and tissue growth. RNA Biol. 2016;13(12):1286-99.

47. Meijer HA, Kong YW, Lu WT, Wilczynska A, Spriggs RV, Robinson SW, et al. Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. Science. 2013;340(6128):82-5.

48. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature. 2010;466(7308):835-40.

49. Liu J, Rivas FV, Wohlschlegel J, Yates JR, 3rd, Parker R, Hannon GJ. A role for the P-body component GW182 in microRNA function. Nat Cell Biol. 2005;7(12):1261-6.

50. Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. Nature. 2008;455(7209):58-63.

51. Mestdagh P, Bostrom AK, Impens F, Fredlund E, Van Peer G, De Antonellis P, et al. The miR-17-92 microRNA cluster regulates multiple components of the TGF-beta pathway in neuroblastoma. Mol Cell. 2010;40(5):762-73.

52. Wahle E, Winkler GS. RNA decay machines: deadenylation by the Ccr4-not and Pan2-Pan3 complexes. Biochim Biophys Acta. 2013;1829(6-7):561-70.

53. Ameres SL, Horwich MD, Hung JH, Xu J, Ghildiyal M, Weng Z, et al. Target RNA-directed trimming and tailing of small silencing RNAs. Science. 2010;328(5985):1534-9.

54. Eckmann CR, Rammelt C, Wahle E. Control of poly(A) tail length. Wiley Interdiscip Rev RNA. 2011;2(3):348-61.

55. Brown CE, Sachs AB. Poly(A) tail length control in Saccharomyces cerevisiae occurs by message-specific deadenylation. Mol Cell Biol. 1998;18(11):6548-59.

56. Jalkanen AL, Coleman SJ, Wilusz J. Determinants and implications of mRNA poly(A) tail size--does this protein make my tail look big? Semin Cell Dev Biol. 2014;34:24-32.

57. Park JE, Yi H, Kim Y, Chang H, Kim VN. Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. Mol Cell. 2016;62(3):462-71.

58. Jonas S, Christie M, Peter D, Bhandari D, Loh B, Huntzinger E, et al. An asymmetric PAN3 dimer recruits a single PAN2 exonuclease to mediate mRNA deadenylation and decay. Nat Struct Mol Biol. 2014;21(7):599-608.

59. Schafer IB, Yamashita M, Schuller JM, Schussler S, Reichelt P, Strauss M, et al. Molecular Basis for poly(A) RNP Architecture and Recognition by the Pan2-Pan3 Deadenylase. Cell. 2019;177(6):1619-31 e21.

60. Miller JE, Reese JC. Ccr4-Not complex: the control freak of eukaryotic cells. Crit Rev Biochem Mol Biol. 2012;47(4):315-33.

61. Albert TK, Lemaire M, van Berkum NL, Gentz R, Collart MA, Timmers HT. Isolation and characterization of human orthologs of yeast CCR4-NOT complex subunits. Nucleic Acids Res. 2000;28(3):809-17.

62. Temme C, Zaessinger S, Meyer S, Simonelig M, Wahle E. A complex containing the CCR4 and CAF1 proteins is involved in mRNA deadenylation in Drosophila. EMBO J. 2004;23(14):2862-71.

63. Maryati M, Airhihen B, Winkler GS. The enzyme activities of Caf1 and Ccr4 are both required for deadenylation by the human Ccr4-Not nuclease module. Biochem J. 2015;469(1):169-76.
64. Lee JE, Lee JY, Trembly J, Wilusz J, Tian B, Wilusz CJ. The PARN deadenylase targets a discrete set of mRNAs for decay and regulates cell motility in mouse myoblasts. PLoS Genet. 2012;8(8):e1002901.
65. Goldstrohm AC, Wickens M. Multifunctional deadenylase complexes diversify mRNA control. Nat Rev Mol Cell Biol. 2008;9(4):337-44.
66. Beelman CA, Parker R. Differential effects of translational inhibition in cis and in trans on the decay of the unstable yeast MFA2 mRNA. J Biol Chem. 1994;269(13):9687-92.
67. Mangus DA, Jacobson A. Linking mRNA turnover and translation: assessing the polyribosomal association of mRNA decay factors and degradative intermediates. Methods. 1999;17(1):28-37.
68. Pelechano V, Wei W, Steinmetz LM. Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. Cell. 2015;161(6):1400-12.
69. Antic S, Wolfinger MT, Skucha A, Hosiner S, Dorner S. General and MicroRNA-Mediated mRNA Degradation Occurs on Ribosome Complexes in Drosophila Cells. Mol Cell Biol. 2015;35(13):2309-20.
70. Hu W, Sweet TJ, Chamnongpol S, Baker KE, Coller J. Co-translational mRNA decay in Saccharomyces cerevisiae. Nature. 2009;461(7261):225-9.
71. Lubas M, Damgaard CK, Tomecki R, Cysewski D, Jensen TH, Dziembowski A. Exonuclease hDIS3L2 specifies an exosome-independent 3'-5' degradation pathway of human cytoplasmic mRNA. EMBO J. 2013;32(13):1855-68.
72. Tuck AC, Rankova A, Arpat AB, Liechti LA, Hess D, Iesmantavicius V, et al. Mammalian RNA Decay Pathways Are Highly Specialized and Widely Linked to Translation. Mol Cell. 2020;77(6):1222-36 e13.
73. Magny EG, Pueyo JI, Pearl FM, Cespedes MA, Niven JE, Bishop SA, et al. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. Science. 2013;341(6150):1116-20.
74. Shanmugam M, Molina CE, Gao S, Severac-Bastide R, Fischmeister R, Babu GJ. Decreased sarcolipin protein expression and enhanced sarco(endo)plasmic reticulum Ca2+ uptake in human atrial fibrillation. Biochem Biophys Res Commun. 2011;410(1):97-101.
75. Anderson DM, Makarewich CA, Anderson KM, Shelton JM, Bezprozvannaya S, Bassel-Duby R, et al. Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. Sci Signal. 2016;9(457):ra119.
76. Nelson BR, Makarewich CA, Anderson DM, Winders BR, Troupes CD, Wu F, et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science. 2016;351(6270):271-5.
77. Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell. 2015;160(4):595-606.
78. Pueyo JI, Magny EG, Sampson CJ, Amin U, Evans IR, Bishop SA, et al. Hemotin, a Regulator of Phagocytosis Encoded by a Small ORF and Conserved across Metazoans. PLoS Biol. 2016;14(3):e1002395.
79. Pauli A, Norris ML, Valen E, Chew GL, Gagnon JA, Zimmerman S, et al. Toddler: an embryonic signal that promotes cell movement via Apelin receptors. Science. 2014;343(6172):1248636.

80. D'Lima NG, Ma J, Winkler L, Chu Q, Loh KH, Corpuz EO, et al. A human microprotein that interacts with the mRNA decapping complex. Nat Chem Biol. 2017;13(2):174-80.
81. Eulalio A, Huntzinger E, Izaurralde E. Getting to the root of miRNA-mediated gene silencing. Cell. 2008;132(1):9-14.
82. Kiriakidou M, Tan GS, Lamprinaki S, De Planell-Saguer M, Nelson PT, Mourelatos Z. An mRNA m7G cap binding-like motif within human Ago2 represses translation. Cell. 2007;129(6):1141-51.
83. Chendrimada TP, Finn KJ, Ji X, Baillat D, Gregory RI, Liebhaber SA, et al. MicroRNA silencing through RISC recruitment of eIF6. Nature. 2007;447(7146):823-8.
84. Kozak M. The scanning model for translation: an update. J Cell Biol. 1989;108(2):229-41.
85. Hellen CU, Sarnow P. Internal ribosome entry sites in eukaryotic mRNA molecules. Genes Dev. 2001;15(13):1593-612.
86. Saghatelian A, Couso JP. Discovery and characterization of smORF-encoded bioactive polypeptides. Nat Chem Biol. 2015;11(12):909-16.
87. Vanderperre B, Lucier JF, Bissonnette C, Motard J, Tremblay G, Vanderperre S, et al. Direct detection of alternative open reading frames translation products in human significantly expands the proteome. PLoS One. 2013;8(8):e70698.
88. Cannell IG, Kong YW, Bushell M. How do microRNAs regulate gene expression? Biochem Soc Trans. 2008;36(Pt 6):1224-31.
89. Kong YW, Cannell IG, de Moor CH, Hill K, Garside PG, Hamilton TL, et al. The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene. Proc Natl Acad Sci U S A. 2008;105(26):8866-71.
90. Patraquim P, Mumtaz MAS, Pueyo JI, Aspden JL, Couso JP. Developmental regulation of canonical and small ORF translation from mRNAs. Genome Biol. 2020;21(1):128.
91. Au PC, Zhu QH, Dennis ES, Wang MB. Long non-coding RNA-mediated mechanisms independent of the RNAi pathway in animals and plants. RNA Biol. 2011;8(3):404-14.
92. Hirota K, Miyoshi T, Kugou K, Hoffman CS, Shibata T, Ohta K. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature. 2008;456(7218):130-4.
93. Tian D, Sun S, Lee JT. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell. 2010;143(3):390-403.
94. Yoo EJ, Cooke NE, Liebhaber SA. An RNA-independent linkage of noncoding transcription to long-range enhancer function. Mol Cell Biol. 2012;32(10):2020-9.
95. Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature. 2013;494(7438):497-501.
96. Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, et al. LincRNA-p21 suppresses target mRNA translation. Mol Cell. 2012;47(4):648-55.
97. Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, et al. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. Nat Med. 2008;14(7):723-30.
98. Wang H, Iacoangeli A, Lin D, Williams K, Denman RB, Hellen CU, et al. Dendritic BC1 RNA in translational control mechanisms. J Cell Biol. 2005;171(5):811-21.
99. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature. 2011;470(7333):284-8.

100. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell. 2010;39(6):925-38.
101. Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. PLoS Biol. 2007;5(5):e106.
102. Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S, Kageyama Y. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. Nat Cell Biol. 2007;9(6):660-5.
103. Pueyo JI, Couso JP. The 11-aminoacid long Tarsal-less peptides trigger a cell signal in Drosophila leg development. Dev Biol. 2008;324(2):192-201.
104. Lauressergues D, Couzigou JM, Clemente HS, Martinez Y, Dunand C, Bécard G, et al. Primary transcripts of microRNAs encode regulatory peptides. Nature. 2015;520(7545):90-3.
105. Williamson L, Saponaro M, Boeing S, East P, Mitter R, Kantidakis T, et al. UV Irradiation Induces a Non-coding RNA that Functionally Opposes the Protein Encoded by the Same Gene. Cell. 2017;168(5):843-55.e13.
106. Mackowiak SD, Zauber H, Bielow C, Thiel D, Kutz K, Calviello L, et al. Extensive identification and analysis of conserved small ORFs in animals. Genome Biol. 2015;16:179.
107. Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM. Long non-coding RNAs as a source of new peptides. Elife. 2014;3:e03523.
108. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12(12):861-74.
109. Dong Y, Liang G, Yuan B, Yang C, Gao R, Zhou X. MALAT1 promotes the proliferation and metastasis of osteosarcoma cells by activating the PI3K/Akt pathway. Tumour Biol. 2015;36(3):1477-86.
110. Cai X, Liu Y, Yang W, Xia Y, Yang C, Yang S, et al. Long noncoding RNA MALAT1 as a potential therapeutic target in osteosarcoma. J Orthop Res. 2016;34(6):932-41.
111. Ji P, Diederichs S, Wang W, Böing S, Metzger R, Schneider PM, et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene. 2003;22(39):8031-41.
112. Schmidt LH, Spieker T, Koschmieder S, Schäffers S, Humberg J, Jungen D, et al. The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. J Thorac Oncol. 2011;6(12):1984-92.
113. Tano K, Mizuno R, Okada T, Rakwal R, Shibato J, Masuo Y, et al. MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. FEBS Lett. 2010;584(22):4575-80.
114. Li L, Chen H, Gao Y, Wang YW, Zhang GQ, Pan SH, et al. Long Noncoding RNA MALAT1 Promotes Aggressive Pancreatic Cancer Proliferation and Metastasis via the Stimulation of Autophagy. Mol Cancer Ther. 2016;15(9):2232-43.
115. Ren S, Liu Y, Xu W, Sun Y, Lu J, Wang F, et al. Long noncoding RNA MALAT-1 is a new potential therapeutic target for castration resistant prostate cancer. J Urol. 2013;190(6):2278-87.
116. Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. Cancer Res. 2011;71(20):6320-6.
117. Meller VH, Joshi SS, Deshpande N. Modulation of Chromatin by Noncoding RNA. Annu Rev Genet. 2015;49:673-95.

118. Portoso M, Ragazzini R, Brenčič Ž, Moiani A, Michaud A, Vassilev I, et al. PRC2 is dispensable for HOTAIR-mediated transcriptional repression. EMBO J. 2017;36(8):981-94.
119. Wu ZH, Wang XL, Tang HM, Jiang T, Chen J, Lu S, et al. Long non-coding RNA HOTAIR is a powerful predictor of metastasis and poor prognosis and is associated with epithelial-mesenchymal transition in colon cancer. Oncol Rep. 2014;32(1):395-402.
120. Roberts TC, Morris KV, Wood MJ. The role of long non-coding RNAs in neurodevelopment, brain function and neurological disease. Philos Trans R Soc Lond B Biol Sci. 2014;369(1652).
121. Mori K, Arzberger T, Grässer FA, Gijselinck I, May S, Rentzsch K, et al. Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. Acta Neuropathol. 2013;126(6):881-93.
122. Zu T, Liu Y, Bañez-Coronel M, Reid T, Pletnikova O, Lewis J, et al. RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc Natl Acad Sci U S A. 2013;110(51):E4968-77.
123. Lee DY, Moon J, Lee ST, Jung KH, Park DK, Yoo JS, et al. Distinct Expression of Long Non-Coding RNAs in an Alzheimer's Disease Model. J Alzheimers Dis. 2015;45(3):837-49.
124. Johnson R, Teh CH, Jia H, Vanisri RR, Pandey T, Lu ZH, et al. Regulation of neural macroRNAs by the transcriptional repressor REST. RNA. 2009;15(1):85-96.
125. Johnson R. Long non-coding RNAs in Huntington's disease neurodegeneration. Neurobiol Dis. 2012;46(2):245-54.
126. Salvi JS, Mekhail K. R-loops highlight the nucleus in ALS. Nucleus. 2015;6(1):23-9.
127. Groh M, Lufino MM, Wade-Martins R, Gromak N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. PLoS Genet. 2014;10(5):e1004318.
128. Colak D, Zaninovic N, Cohen MS, Rosenwaks Z, Yang WY, Gerhardt J, et al. Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. Science. 2014;343(6174):1002-5.
129. Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science. 2013;340(6132):619-21.
130. Nakama M, Kawakami K, Kajitani T, Urano T, Murakami Y. DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. Genes Cells. 2012;17(3):218-33.
131. Tan H, Xu Z, Jin P. Role of noncoding RNAs in trinucleotide repeat neurodegenerative disorders. Exp Neurol. 2012;235(2):469-75.
132. Mutsuddi M, Marshall CM, Benzow KA, Koob MD, Rebay I. The spinocerebellar ataxia 8 noncoding RNA causes neurodegeneration and associates with staufen in Drosophila. Curr Biol. 2004;14(4):302-8.
133. Uchida S, Dimmeler S. Long noncoding RNAs in cardiovascular diseases. Circ Res. 2015;116(4):737-50.
134. Archer K, Broskova Z, Bayoumi AS, Teoh JP, Davila A, Tang Y, et al. Long Non-Coding RNAs as Master Regulators in Cardiovascular Diseases. Int J Mol Sci. 2015;16(10):23651-67.
135. Lakhotia SC, Mallik M, Singh AK, Ray M. The large noncoding hsrω-n transcripts are essential for thermotolerance and remobilization of hnRNPs, HP1 and RNA polymerase II during recovery from heat shock in Drosophila. Chromosoma. 2012;121(1):49-70.

136. Prasanth KV, Rajendra TK, Lal AK, Lakhotia SC. Omega speckles - a novel class of nuclear speckles containing hnRNPs associated with noncoding hsr-omega RNA in Drosophila. J Cell Sci. 2000;113 Pt 19:3485-97.
137. Perrimon N, Lanjuin A, Arnold C, Noll E. Zygotic lethal mutations with maternal effect phenotypes in Drosophila melanogaster. II. Loci on the second and third chromosomes identified by P-element-induced mutations. Genetics. 1996;144(4):1681-92.
138. Mallik M, Lakhotia SC. The developmentally active and stress-inducible noncoding hsromega gene is a novel regulator of apoptosis in Drosophila. Genetics. 2009;183(3):831-52.
139. Johnson TK, Cockerell FE, McKechnie SW. Transcripts from the Drosophila heat-shock gene hsr-omega influence rates of protein synthesis but hardly affect resistance to heat knockdown. Mol Genet Genomics. 2011;285(4):313-23.
140. Hardiman KE, Brewster R, Khan SM, Deo M, Bodmer R. The bereft gene, a potential target of the neural selector gene cut, contributes to bristle morphogenesis. Genetics. 2002;161(1):231-47.
141. Soshnev AA, Ishimoto H, McAllister BF, Li X, Wehling MD, Kitamoto T, et al. A conserved long noncoding RNA affects sleep behavior in Drosophila. Genetics. 2011;189(2):455-68.
142. Li M, Wen S, Guo X, Bai B, Gong Z, Liu X, et al. The novel long non-coding RNA CRG regulates Drosophila locomotor behavior. Nucleic Acids Res. 2012;40(22):11714-27.
143. Deng X, Meller VH. roX RNAs are required for increased expression of X-linked genes in Drosophila melanogaster males. Genetics. 2006;174(4):1859-66.
144. Pontier DB, Gribnau J. Xist regulation and function explored. Hum Genet. 2011;130(2):223-36.
145. Park Y, Kelley RL, Oh H, Kuroda MI, Meller VH. Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. Science. 2002;298(5598):1620-3.
146. Kelley RL, Lee OK, Shim YK. Transcription rate of noncoding roX1 RNA controls local spreading of the Drosophila MSL chromatin remodeling complex. Mech Dev. 2008;125(11-12):1009-19.
147. Oh H, Park Y, Kuroda MI. Local spreading of MSL complexes from roX genes on the Drosophila X chromosome. Genes Dev. 2003;17(11):1334-9.
148. Kelley RL, Kuroda MI. The Drosophila roX1 RNA gene can overcome silent chromatin by recruiting the male-specific lethal dosage compensation complex. Genetics. 2003;164(2):565-74.
149. Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B. Xist RNA and the mechanism of X chromosome inactivation. Annu Rev Genet. 2002;36:233-78.
150. McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, et al. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature. 2015;521(7551):232-6.
151. Amaral PP, Leonardi T, Han N, Vire E, Gascoigne DK, Arias-Carrasco R, et al. Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. Genome Biol. 2018;19(1):32.
152. Tupy JL, Bailey AM, Dailey G, Evans-Holm M, Siebel CW, Misra S, et al. Identification of putative noncoding polyadenylated transcripts in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2005;102(15):5495-500.
153. Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MA, Brocard M, et al. Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. Elife. 2014;3:e03528.

154. Zhang H, Dou S, He F, Luo J, Wei L, Lu J. Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during Drosophila development. PLoS Biol. 2018;16(7):e2003903.

155. Zabolotskaya MV, Grima DP, Lin MD, Chou TB, Newbury SF. The 5'-3' exoribonuclease Pacman is required for normal male fertility and is dynamically localized in cytoplasmic particles in Drosophila testis cells. Biochem J. 2008;416(3):327-35.

156. Jones CI. Post-transcriptional gene regulation by the exoribonuclease Pacman [PhD]. Brighton and Sussex Medical School: Brighton and Sussex Medical School; 2011.

157. Gatfield D, Izaurralde E. Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila. Nature. 2004;429(6991):575-8.

158. Gatfield D, Unterholzner L, Ciccarelli FD, Bork P, Izaurralde E. Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. EMBO J. 2003;22(15):3960-70.

159. Rehwinkel J, Herold A, Gari K, Kocher T, Rode M, Ciccarelli FL, et al. Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. Nat Struct Mol Biol. 2004;11(6):558-66.

160. Barisic-Jager E, Krecioch I, Hosiner S, Antic S, Dorner S. HPat a decapping activator interacting with the miRNA effector complex. PLoS One. 2013;8(8):e71860.

161. Iwaki T, Figuera M, Ploplis VA, Castellino FJ. Rapid selection of Drosophila S2 cells with the puromycin resistance gene. Biotechniques. 2003;35(3):482-4, 6.

162. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009;324(5924):218-23.

163. Mevel-Ninio M, Pelisson A, Kinder J, Campos AR, Bucheton A. The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. Genetics. 2007;175(4):1615-24.

164. Subocheva EA, Romanova NI, Karpova NN, Iuneva AO, Kim AI. [Male reproductive behavior in Drosophila melanogaster strains with different alleles of the flamenco gene]. Genetika. 2003;39(5):675-81.

165. Ustianenko D, Hrossova D, Potesil D, Chalupnikova K, Hrazdilova K, Pachernik J, et al. Mammalian DIS3L2 exoribonuclease targets the uridylated precursors of let-7 miRNAs. RNA. 2013;19(12):1632-8.