

Investigating the *in vivo* functions of non-coding RNAs in *Drosophila* *melanogaster*



Jenna Schwarz
Hertford College
University of Oxford

A thesis submitted for the degree

Doctor of Philosophy

Trinity 2015

Investigating the *in vivo* functions of non-coding RNAs in *Drosophila melanogaster*

A thesis submitted for the degree of D.Phil in Physiology, Anatomy, and Genetics

Abstract

The number of annotated non-protein coding genes has increased dramatically in recent years, but the functions of very few are well understood. In this thesis, I study the functions of two classes of non-coding RNA in *Drosophila melanogaster*. Firstly, through characterisation of six putative long non-coding RNA loci containing transposon insertions, I identified a long non-coding RNA, *Inc703*, suitable for in depth characterisation. Using a GAL4/UAS fluorescent reporter I found *Inc703* to be expressed in the larval fat body and confirmed this using qRT-PCR. Transcriptome profiling identified ribosomal genes as upregulated in two independent *Inc703* mutant strains. Additionally, I have identified decreased lifespan in some *Inc703*^{gal4} strains and it will be interesting to determine whether there is a relationship between increased ribosomal protein gene expression and reduced lifespan. I also used transcriptome profiling of the musculature and CNS of miR-8 null larvae to determine tissue specific targets of miR-8. Many indirect effects of miR-8 were detected but using MRE prediction algorithms a number of potentially direct targets of miR-8 were identified. One of these, *ilp5*, which is upregulated in the CNS, was identified as a non-cell autonomous effect of loss of miR-8 in the fat body. Together this work highlights the importance of using multiple, complementary techniques to study the function of non-coding RNAs *in vivo*.

Jenna Schwarz
Hertford College

Trinity Term 2015

Acknowledgments

I would firstly like to thank my supervisors Dr Jilong Liu and Professor Chris Ponting for allowing me the opportunity to carry out this project and Dr Andrew Bassett whom I will never be able to thank enough for his help and guidance without which this work would not have been possible. I am also incredibly grateful to Dr Ana Marques for her patience and unending enthusiasm whilst teaching me to code and to the many members of the Liu and Ponting labs, past and present, who have offered support, advice and friendship over the past four years. In particular, Dr Siân Davies for her stories when I needed to laugh, Dr Gabriel Aughey for his pragmatism when things needed to be done, Dr Charlotte Tibbit for her optimism when mine failed, Dr Stuart Grice for his hyperbole when I needed perspective and Kenny Roberts for all the food.

I thank all my friends and family for their support and for understanding when flies got in the way of our plans. Especially my parents for their unquestioning support and belief in my abilities, Maisie for listening to never ending tales about the woes of lab work, Annie for keeping me connected to the real world and Kate for games nights. Finally, I thank Tris, whose calming influence made the completion of this thesis not only possible but relatively painless.

I would also like to mention Dr Nicholas Kent whose dynamic lectures first piqued my interest in the non-coding genome, the wider academic community for sharing knowledge and reagents and the MRC and ERC for funding me and the project, respectively.

Contents

List of abbreviations	9
Chapter 1 - Introduction	11
1.1 Protein-coding exons account for a small proportion of the human genome	11
1.2 There are many classes of non-coding genes	11
1.3 Prevalent transcription outside known genes uncovers many long non-coding transcripts	12
1.4 The diversity of lncRNAs confounds predictions of functionality	13
1.5 Functional lncRNAs have diverse molecular mechanisms	14
1.5.1 Dosage compensation	14
1.5.2 Chromatin associated lncRNAs	16
1.5.3 Structural lncRNAs	17
1.5.4 Competitive endogenous RNAs	17
1.5.5 RNA independent <i>cis</i> -acting lncRNAs	18
1.6 lncRNAs in disease	19
1.6.1 <i>MALAT-1</i>	19
1.7 Methods to study lncRNAs	20
1.7.1 shRNA, siRNA or dsRNA knockdown	20
1.7.2 Genetic deletion	22
1.7.3 Termination	22
1.7.5 Genome engineering	23
1.8 microRNAs	25
1.8.1 miRNA biogenesis	25
1.8.2 microRNAs control gene levels through post-transcriptional regulation	27
1.8.3 miRNAs have functions in development and disease	28
1.8.4 Methods to predict miRNA targets	28
1.8.5 Methods to confirm miRNA-target interaction	30
1.10 Project aims and thesis structure	32
Chapter 2 - Methods	34
2.1 Wet lab methods	34
2.1.1 Fly Husbandry	34
2.1.2 Isogenisation	34
2.1.3 PCR	34
2.1.4 Vector transformation	35
2.1.5 Colony PCR	35
2.1.6 Genotyping	35
2.1.7 RNA extraction	36
2.1.8 Reverse Transcription	36
2.1.9 qRT-PCR	37
2.1.10 Rapid Amplification of cDNA Ends (RACE)	37
2.1.11 Immunofluorescence	38
2.1.12 Transgenic fly stock generation	39
2.1.13 Triacylglyceride assay	42
2.1.14 Western Blotting	42

2.1.15 Longevity study	43
2.2 Computational Methods	43
2.2.1 Defining transcriptional units for differential gene expression analysis	43
2.2.2 Building a reference library	43
2.2.3 Mapping sequencing reads, transcript assembly and lncRNA discovery	44
2.2.4 Calling differential gene expression	45
2.2.5 Genome wide predictions of microRNA recognition elements	45

Chapter 3 - Characterisation of six loci containing predicted long non-coding RNA transcripts 46

Abstract	46
3.1 Introduction	47
3.1.1 <i>In vivo</i> functions of lncRNAs are largely unknown	47
3.1.2 Methods of generating lncRNA mutants	47
3.1.3 Attributes of the ideal lncRNA for functional characterisation	47
3.1.4 Aims	48
3.2 Screening methodology	50
3.2.1 Initial selection	50
3.2.2 Confirmation of presence of non-coding transcript	50
3.2.3 Analysis of mutants	51
3.3 Results	53
3.3.1 Selection of six putative long non-coding RNA loci	53
3.3.2 Initial screening of transposon insertion strains	54
3.3.3 Transcripts could be detected in all six loci at one or more developmental stages	57
3.3.4 Full transcript models	57
3.3.3 Incomplete transcript models	62
3.3.4 Coding potential	62
3.3.5 Effect of transposon insertions on lncRNA expression	65
3.3.6 Identifying putative <i>cis</i> -acting lncRNAs	66
3.4 Discussion	70
3.4.1 Methods to identify potentially functional lncRNAs	70
3.4.2 Some putative lncRNAs encode smORFs	72
3.4.3 Selection of lncRNAs candidates for further study	73

Chapter 4 - Characterisation of *Inc703*^{stop} and *Inc703*^{gal4} alleles 75

Abstract	75
4.1 Introduction	76
4.1.1 <i>Drosophila</i> innate immune response has parallels to human innate immunity	76
4.1.2 <i>Drosophila</i> immune response genes have roles in development	76
4.2 Methods	78
4.2.1 Generation of supplementary mutant strains	78
4.2.2 RNA-seq sample preparation	78
4.2.3 Enrichment analyses of differentially expressed genes	78
4.3 Results	80
4.3.1 <i>Inc703</i> is a putative long non-coding RNA with two independent transposon mediated insertion strains that reduce its expression level	80
4.3.2 A second mutation, outside the <i>Inc703</i> locus, is responsible for pupal lethal-	

ity in <i>Inc703</i> ^{stop} flies	80
4.3.3 RNA-seq identifies genes differentially expressed in <i>Inc703</i> mutant strains	83
4.3.4 Identification of dysregulated pathways in <i>Inc703</i> ^{gal4}	83
4.3.5 A significant portion of genes differentially expressed in <i>Inc703</i> ^{gal4} are also changed in <i>Inc703</i> ^{stop}	87
4.3.6 Genes differentially expressed in <i>Inc703</i> ^{gal4} overlap with genes differentially expressed in <i>atf3</i> ⁷⁶ mutants	87
4.3.7 Gene expression changes were validated by qRT-PCR	90
4.3.8 Changes in immune-related genes could not be reproduced in <i>Inc703</i> ^{gal4} mutants after isogenisation	92
4.3.9 Immune response genes are not changed in <i>Inc703</i> ^{gal4} adult females	94
4.3.10 shRNAs targeted to <i>Inc703</i> reduce transcript levels but do not reproduce gene expression changes seen in mutants	94
4.4 Discussion	98
4.4.1 The challenges in discovering lncRNA loss-of-function phenotypes	98
4.4.2 Are immune response genes innately variable?	99
4.4.3 Identification of clusters within differentially expressed genes is biased towards known genes	101
4.4.4 Summary	101

Chapter 5 - Tissue specific expression of *Inc703* **103**

Abstract	103
5.1 Introduction	104
5.1.1 Methods for determining RNA expression at the cellular level	104
5.1.2 Diverse functions of the <i>Drosophila</i> fat body	104
5.1.3 Development of the <i>Drosophila</i> larval and adult fat bodies is distinct	106
5.1.4 Ribosome levels influence cellular and organism fitness	106
5.1.5 Aim	107
5.2 Methods	108
5.2.1 Staging	108
5.2.2 GAL4/UAS reporter system	108
5.2.3 Cell fractionation	109
5.2.4 Producing mitotic clones in the larval fat body using Mosaic Analysis with a Repressible Cell Marker (MARCM)	110
5.2.5 Transcriptome profiling of larval fat bodies	110
5.3 Results	112
5.3.1 Determining expression profiles using GAL4/UAS fluorescent reporters	112
5.3.2 mCherry reporter has a dynamic expression pattern in the larval CNS	112
5.3.3 Fluorescent reporters are expressed in a subset of photoreceptors in the eye imaginal discs	114
5.3.4 Reporter expression is also observed in other larval tissues	117
5.3.5 Confirmation of reporter expression by qRT-PCR	120
5.3.6 <i>Inc703</i> is detected equally in the cytoplasmic and nuclear fractions of S2 cells	122
5.3.7 No alteration in TAG levels is observed in <i>Inc703</i> mutant larvae	124
5.3.8 Identification of genes differentially expressed in the fat body of isogenised <i>Inc703</i> ^{gal4} larvae	126
5.3.9 BAC ^{158D05} does not rescue <i>Inc703</i> expression levels in all tissues	127
5.3.10 Gene expression changes in <i>Inc703</i> ^{stop} correlate with <i>Inc703</i> ^{gal4} and	

BAC ^{158D05} , <i>Inc703</i> ^{gal4}	129
5.3.11 Gene expression changes in both <i>Inc703</i> ^{gal4} and <i>Inc703</i> ^{stop} are confirmed by qRT-PCR	130
5.3.12 Ribosomal protein genes may also be increased at the protein level	132
5.3.13 Lifespan	133
5.3.14 <i>Inc703</i> and <i>e(y)2b</i> are independently overexpressed in UAS-driven transgenic strains	135
5.3.15 Lifespan of transgenic overexpression strains	137
5.4 Discussion	139
5.4.1 Visualising RNA <i>in vivo</i>	139
5.4.2 Levels of ribosomal protein genes are increased in <i>Inc703</i> mutants	142
5.4.3 Association of ribosomal protein levels, <i>dm</i> and cell competition	143
5.4.4 <i>arc1</i> , RpS6 and starvation response	144
5.4.5 Ribosomal protein levels can influence lifespan	144
5.4.6 A putative role for <i>Inc703</i> in controlling ribosome biogenesis	145

Chapter 6 - Using transcriptome profiling to investigate tissue specific roles of miR-8 147

Abstract	147
6.1 Introduction	148
6.1.1 dme-miR-8 is a homolog of the conserved miR-200 family	148
6.1.2 The human miR-200 family controls epithelial to mesenchymal transition	148
6.1.3 Manipulation of miR-8 levels causes multiple tissue specific effects, mediated through distinct targets	149
6.1.4 Experimental aims	153
6.2 Methods	154
6.2.1 Sample preparation and sequencing	154
6.2.2 Remapping, transcript assembly and defining lncRNAs	154
6.2.3 Calling differential gene expression	155
6.2.4 miRNA Recognition Element prediction	155
6.2.5 Gene Ontology analysis	155
6.3 Results	156
6.3.1 Four distinct sample populations can be distinguished within the RNA-seq results	156
6.3.2 Differentially expressed genes include the visible marker of the miR-8 ^{Δ1} allele and represent both direct and indirect effects of loss of miR-8	158
6.3.3 Muscle samples have higher variation but differentially expressed genes are not more variable	159
6.3.4 miR-8 controls potentially overlapping pathways in different larval tissues	161
6.3.5 CNS specific effects of miR-8	163
6.3.6 Muscle specific effects of miR-8	170
6.3.7 Identification of tissue specificity of published changes	172
6.3.8 Identifying ceRNA networks	172
6.3.9 The miR-8 primary transcript is upregulated in miR-8 null larvae	174
6.4 Discussion	176
6.4.1 Levels of the <i>Drosophila</i> insulin like peptides and body size are uncoupled in <i>miR-8</i> null larvae	176
6.4.2 Comparison with published data reveals challenges in identifying miRNA	

targets	177
6.4.3 Methods of identifying true miR-8 targets	177
6.4.4 Better methods of predicting targets	179
6.4.5 Ongoing experiments	180
Chapter 7 - Discussion	181
7.1 Summary of findings	181
7.2 Challenges using current techniques to study lncRNA function	182
7.3 Future work to confirm <i>lnc703</i> function	184
7.4 Challenges in identifying true miRNA targets	185
7.5 Future work to identify true miR-8 targets	187
7.6 Concluding remarks	188
References	189
Appendices	201
Appendix A – Primers used for cloning, colony PCR and sequencing	201
Appendix B – Primers used for qRT-PCR experiments	203
Appendix C - Images of gels showing PCR products, confirming presence of transposon in lncRNA loci after isogenisation.	206
Appendix D - Sequencing results confirm introns in <i>lnc703</i> and <i>lnc921</i> aligned to BDGP6. Exon/exon boundaries are highlighted in red.	207
Appendix E – Genes changed in the same direction one day AFP in <i>lnc703</i> ^{gal4} and <i>lnc703</i> ^{stop} strains	208
Appendix F – Genes changed in the same direction in larval fat bodies of isogenised <i>lnc703</i> ^{gal4} and <i>lnc703</i> ^{stop} strains	211

List of abbreviations

APF – After puparium formation

BAC – Bacterial Artificial Chromosome

BDGP – Berkeley Drosophila Genome Project

cDNA – complementary DNA

CDS – Protein-coding sequence

ceRNA – competitive endogenous RNA

CLASH - Crosslinking, ligation, and sequencing of hybrids

CNS – Central nervous system

CRISPR – Clustered regularly interspaced short palindromic repeats

DILPs – *Drosophila* insulin-like peptides

DNA – Deoxyribonucleic acid

dsRNA – double stranded RNA

eRNA – enhancer-derived RNA

EYFP – Enhanced Yellow fluorescent protein

FC – Fold change

FDR – False discovery rate

FISH – Fluorescent *In Situ* Hybridisation

FRT – Flippase recombination target

GFP – Green fluorescent protein

GO – Gene ontology

GTF – Gene transfer format

GWAS – Genome-wide association study

lincRNA – long intergenic non-coding RNA

lncRNA – long non-coding RNA

MARCM - Mosaic Analysis with a Repressible Cell Marker

miRNA – micro RNA

MRE – miRNA recognition element

mRNA – messenger RNA

ncRNA – non-coding RNA

NMJ – Neuromuscular junction

ORF – Open reading frame

PCR – Polymerase Chain Reaction

pri-miRNA – Primary miRNA transcript

qRT-PCR – quantitative reverse transcription PCR

RACE – Rapid Amplification of cDNA Ends

RNA – Ribonucleic Acid

RNAi – RNA interference

RNAseq – high throughput cDNA sequencing

shRNA – short-hairpin RNA

smORF – small ORF

SNP – Single nucleotide polymorphism

TAG – Triacylglyceride

TF – Transcription Factor

TSS – Transcriptional start site

UAS – Upstream Activation Sequence

UTR – Untranslated region

Chapter 1 - Introduction

1.1 Protein-coding exons account for a small proportion of the human genome

The greatest surprise upon completion of the draft human genome was the discovery of just 22,000 protein-coding genes (Cunningham et al., 2015), since historical estimates predicted as many as 100,000 protein-coding genes (Gilbert, 1993). This is even more surprising in the context of the discovery that protein-coding exons account for only 1.1% of the human genome, despite evidence from sequence constraint suggesting that between 3 – 15% of the human genome is functional (Ponting and Hardison, 2011) and 62% of the genome being transcribed (The ENCODE Project Consortium (2012)). In flies, the proportion of the genome that encodes for protein-coding sequences is higher than in humans [19%, (Tweedie et al., 2009)] and the proportion that is predicted to be functional is also higher, 47% (Meader et al., 2010). It is clear that a major challenge in the post-genomic era will be to characterise the various classes of functional genetic elements encoded within complex genomes.

1.2 There are many classes of non-coding genes

Some of the transcribed non-protein coding regions are accounted for by known small RNAs including structural RNAs (such as ribosomal rRNAs and transfer tRNAs) that have roles in translation, as well as regulatory small RNAs such as microRNAs, piwiRNAs and siRNAs. However, longer RNA molecules, with no canonical signatures of protein coding sequence have also been detected using high-throughput cDNA sequencing (RNA-seq) technologies. These molecules

are variously referred to as long intergenic non-coding RNAs, long intervening non-coding RNAs (lincRNAs), or long/large non-coding RNAs (lncRNAs). They have arbitrarily been defined as transcripts of greater than 200bp in length that do not appear to encode proteins. lncRNAs are not unique to humans, and have been found in all species tested so far, including bacteria, yeast, flies, zebrafish and mice (Guttman et al., 2009; Pauli et al., 2012; Puerta-Fernandez et al., 2006; Young et al., 2012). Although definition of lncRNAs are currently based on rather arbitrary criteria, a common theme amongst many lncRNAs is that they are generally more lowly expressed and exhibit more temporally and spatially restricted expression patterns compared to protein-coding genes (Derrien et al., 2012; Kutter et al., 2012; McKinlay et al., 2011; Young et al., 2012).

1.3 Prevalent transcription outside known genes uncovers many long non-coding transcripts

Like for protein-coding genes in the pre-genomic era, the estimate for the number of lncRNA genes in humans varies greatly between 28,000 (Ulitsky and Bartel, 2013) – 58,000 (Iyer et al., 2015) and is further complicated by the varying definitions of lncRNAs. Whilst it is clear that large amounts of transcription occurs outside of protein-coding loci, its biological relevance is contentious, with some suggesting it is merely transcriptional noise. Indeed, the overlap between different catalogues of lncRNAs is often low (Jia et al., 2010; Nam and Bartel, 2012). This is likely a reflection of the low expression level and spatiotemporal restriction of lncRNA expression resulting in different lncRNAs being detected depending on the technique and tissue used. However, between 70 – 90% of novel intergenic transcripts could be detected in multiple datasets when the same sample type was used, and only ~15% have any evidence for protein products (Djebali et al., 2012). Analysis of evolutionary constraint indicates the proportion of lncRNA

sequence that is constrained in humans is similar to that for enhancer sequences and greater than unannotated sequences (Rands et al., 2014). Together this suggests that only a small proportion of lncRNAs are likely to be functional, or that the specific features important for their function are not detected by normal sequence alignment techniques.

1.4 The diversity of lncRNAs confounds predictions of functionality

Assessment of the functionality of lncRNAs as a whole is confounded by the presence of various different functional classes under the lncRNA umbrella (Ulitsky and Bartel, 2013) all of which are yet to be properly defined. These definitions largely rely on the genomic context of the lncRNA, for example intergenic, intronic, antisense to or convergent from another gene or overlapping with promoter or enhancer regions for other genes (Marques et al., 2013). Alternatively, they are based on features of the transcript, including the presence or absence of splicing or polyadenylation (Livvyatan et al., 2013; Young et al., 2012), or transcript stability (Wyers et al., 2005; Xu et al., 2009). Some analyses only include non-overlapping, spliced and polyadenylated “mRNA-like” transcripts to reduce the likelihood of inclusion of transcripts generated through spurious transcription or artefacts of cDNA synthesis (Young et al., 2012). However, as for protein-coding genes, mono-exonic transcripts can be functional (Chalei et al., 2014) and some unstable promoter-upstream transcripts (PROMPTs) are polyadenylated (Ntini et al., 2013). Other analyses include all transcripts, including unspliced transcripts and transcripts that are not products of RNA polymerase II transcription. Overlapping antisense transcripts are variously included or excluded from analyses. Estimation of the proportion of lncRNAs that are functional is complicated by these differing definitions and by the absence of defined signatures of functionality. However, efforts have been made to start classifying lncRNAs according to their features.

One distinction that can be made considers the chromatin modifications of lncRNA loci. lncRNAs whose transcription originates from enhancers or promoters can be separated by the ratio of histone H3K4 mono- and trimethylation. However, initial analyses suggest they are otherwise indistinguishable on the basis of length, exon structure or orientation relative to the neighbouring gene, although the expression of enhancer-derived RNAs (eRNA) is more likely to be correlated with expression level of the neighbouring gene (Marques et al., 2013).

1.5 Functional lncRNAs have diverse molecular mechanisms

The majority of functional lncRNAs that have been characterised act to control the level of the expression of other genes, which can be their neighbouring genes (*cis*-acting) or distal loci (*trans*-acting). Considering the molecular mechanism of action of functional lncRNAs, it is possible to separate two potential scenarios. In the first, it is the act of transcription through a region that is consequential, through remodelling of the local chromatin state or through steric hindrance of factor binding (Figure 1.1a). Alternatively, the RNA molecule generated from the lncRNA exerts a function through interaction with other RNA molecules, proteins or chromatin is *cis* (Figure 1.1b) or in *trans* (Figure 1.1c). There is reported evidence for both of these mechanisms across multiple organisms (Martens et al., 2004; Rinn et al., 2007).

1.5.1 Dosage compensation

The best studied cases of functional lncRNAs are those involved in sex dosage compensation, both in mammals and in flies. Dosage compensation of genes expressed from the X chromosome ensures XX females and XY males have equivalent levels of genes expressed from the X chromosome. In flies this is achieved by global upregulation of the X chromosome in males, which is

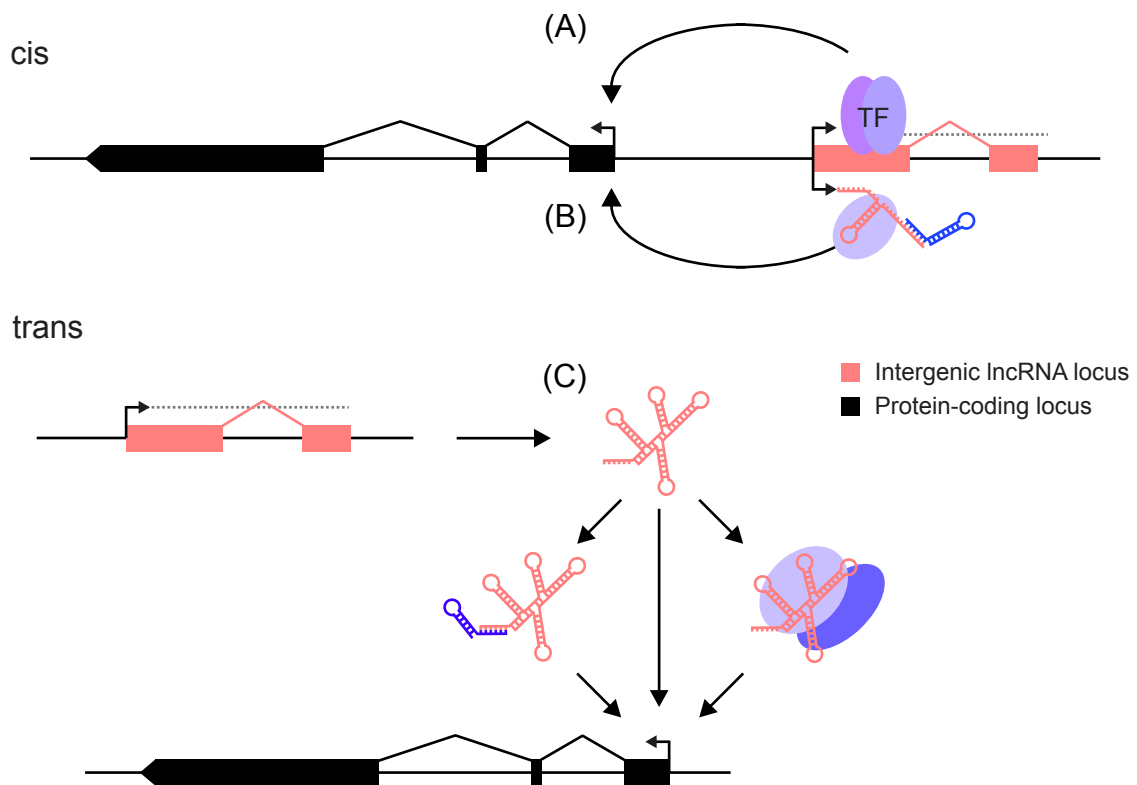


Figure 1.1 Potential molecular mechanisms of lncRNAs. lncRNAs can act in *cis* on their neighbouring genes either through (A) RNA independent mechanisms such as chromatin remodelling or transcriptional interference or (B) in RNA dependent mechanisms, through the lncRNA recruiting other factors. (C) lncRNAs can also act in *trans* to regulate distal loci through interaction of the lncRNA with DNA, or RNA or protein complexes. Adapted from (Bassett et al., 2014).

coordinated by two functionally redundant lncRNAs roX1 and roX2 that are transcribed from two distal genes on the X chromosome and which have quite divergent sequences (Franke and Baker, 1999). The roX RNAs are required for binding of the male specific lethal (MSL) protein complex to the X chromosome (Meller and Rattner, 2002) which upregulates the expression of actively transcribed genes. Recent studies have mapped the interaction of the different domains of roX1 with RNA, DNA and protein (Quinn et al., 2014) to elucidate the topology of the RNA, interaction partners and binding profile. The D3 domain has the strongest interactions with MLE and MSL protein and is able to rescue male lethality as well as the full length roX1 transgene (Quinn et al., 2014).

1.5.2 Chromatin associated lncRNAs

The role of lncRNAs controlling chromatin modification is not limited to dosage compensation. A number of novel lncRNAs have been shown to bind chromatin modifying complexes (Wang and Chang, 2011) and are thought to facilitate binding of complexes to specific loci. However, up to 25% of the entire transcriptome was found to bind to polycomb repressive complex (PRC2) in ES cells with no enrichment for lncRNAs (Zhao et al., 2010) and bacterial RNAs have been shown to associated with PRC2 as well as endogenous interactors (Davidovich et al., 2015), questioning the specificity of some of these interactions (Brockdorff, 2013). The lncRNA *HOTTIP* which is transcribed from the 5' of the *HOXA* locus in mice, acts through chromosome looping to recruit complexes that induce H3K4 trimethylation, thereby activating genes within the *HOXA* locus (Wang et al. 2011). Conversely, *HOTAIR* RNA acts as a scaffold for chromatin modifying complexes at distal loci to regulate HOX gene expression (Tsai et al., 2010). Recently, two lncRNAs located next to developmentally important transcription factors have been shown to bind to chromatin both locally and at distal regions, acting in part through interaction with the protein product of their neighbouring gene (Chalei et

al., 2014; Vance et al., 2014).

1.5.3 Structural lncRNAs

lncRNAs can also act as molecular scaffolds, on a small scale as described for *HOTAIR* or within large cellular structures. In flies, a long non-coding RNA is transcribed from the *hsr-omega* (*hsrw*) locus. This transcript (*hsrw-n*) is retained within the nucleus upon heat shock and also upon ecdysone expression in third instar larvae. *hsrw-n* is found in clusters, named omega-speckles, which contain hnRNPs, and is required for the assembly of these structures (Prasanth et al., 2000). Similarly in mice, a lncRNA, *Nuclear Enriched Abundant Transcript 1* (*NEAT1*) is localised to nuclear paraspeckles in cultured cells and *in vivo*. *NEAT1* is required for assembly of paraspeckles (Clemson et al., 2009; Sunwoo et al., 2009). The biological function of paraspeckles is not known, but they are thought to control cellular differentiation and stress response by retention of mRNAs in the nucleus (Chen and Carmichael, 2009).

1.5.4 Competitive endogenous RNAs

lncRNAs, like mRNAs, can be post-transcriptionally regulated through regions of partial complementarity with microRNAs (miRNAs), that act to post-transcriptionally regulate RNA abundance. The abundance of one transcript can influence the levels of other transcripts which share binding sites for the same miRNAs; this can occur between coding transcripts (Poliseno et al., 2010; Tay et al., 2011) or between coding and non-coding transcripts (Cesana et al., 2011; Tan et al., 2014; Wang et al., 2013). RNAs that control the levels of other genes by this mechanism are referred to as competitive endogenous RNAs (ceRNAs). ceRNAs as with other lncRNAs are generally lowly expressed and have temporally and spatially restricted expression patterns. A recent study has suggested that lowly expressed lncRNAs cannot act as ceRNAs and that the stoichiometry is only

sufficient to affect the available pool of miRNAs for lowly expressed miRNAs (Denzler et al., 2014). However, evidence for ceRNA function *in vivo* suggests that the dynamics of interactions between lncRNA and miRNA might differ from canonical binding resulting in reduction of the miRNA pool (Cesana et al., 2011; Tan et al., 2014; Wang et al., 2013). For example, *linc-MD1* which is expressed in differentiating but not proliferating myoblasts controls differentiation timing in mouse and human myoblasts by “sponging” miR-133 and miR-135 to regulate the expression of transcription factors, *MAML1* and *MEF2C* that activate muscle specific genes (Cesana et al., 2011). Similarly *linc-RoR* controls differentiation in human embryonic stem cells by acting as a sponge for miR-145 (Wang et al., 2013). A recent genome wide study in mouse embryonic stem cells suggests that lncRNA-miRNA crosstalk is more prevalent than previously thought, with more than 50% of lncRNAs and their miRNA dependent targets being co-ordinately upregulated upon loss of Dicer (Tan et al., 2015).

1.5.5 RNA independent *cis*-acting lncRNAs

Whilst all the previously described lncRNAs function through an RNA-dependent mechanism, it is also possible for the act of transcription through a region to regulate neighbouring genes, independently of the RNA produced. Examples of RNA independent *cis*-acting lncRNAs are rarer, possibly due to the challenges associated with studying this mechanism. In yeast, there are a number of examples of genes regulated by transcriptional interference. *SER3* is repressed during growth in rich medium and deletion of the TATA box of an upstream ncRNA *SRG1* prevents its transcription and removes repression of *SER3* (Martens et al., 2004). Further experiments confirmed that the RNA molecule is not required for function and that the act of transcription of *SRG1* prevents binding of factors upstream of *SER3* inducing repression of *SER3* (Martens et al., 2004). In flies, the deletion of promoters of the *bxd* lncRNAs prevents transcription and results in

ectopic expression of neighbouring gene *Ubx*. However, knockdown of *bx**d* RNAs using dsRNA had no effect on *Ubx* expression. Repression of *Ubx* correlates with the presence of read through from the *bx**d* locus close to the *Ubx* promoter suggesting that *Ubx* is repressed through transcriptional interference (Petruk et al., 2006). However, promoter inversion experiments suggest it may in fact be the deletion of an enhancer which mediates the effects on *Ubx* expression (Pease et al., 2013), highlighting the difficulties of analysing *cis* acting mechanisms, even in such a well-studied example.

1.6 lncRNAs in disease

Disease related single nucleotide polymorphisms (SNPs) identified in genome-wide association studies (GWAS) frequently fall in intergenic regions (Maurano et al., 2012). lncRNAs that overlap these intergenic regions offer potential mechanisms by which these SNPs can be causal (Ning et al., 2014), through modulation of lncRNA expression levels (Kumar et al., 2013) or by disrupting a functional region of the lncRNA. When interpreting these data, it must be borne in mind that these mutations may also be disrupting underlying DNA elements such as transcription factor binding sites that may have roles that are independent of the lncRNA. However, there are numerous examples of lncRNAs that are dysregulated in human cancer tissues. A number of lncRNA expression changes may be useful as predictors for prognosis as they correspond with increased growth or metastatic potential of the cancer.

1.6.1 *MALAT-1*

The lncRNA, *MALAT-1* was first identified as a predictor of metastasis in human non-small cell lung cancer (Ji et al., 2003) and upregulation has since been associated with tumour growth and metastasis in a number of human cancers

(Lin et al., 2007; Wang et al., 2014; Xu et al., 2011). In a mouse model of colorectal cancer this occurs through disruption of the SFPQ/PTBP2 complex when *MALAT-1* is overexpressed (Ji et al., 2014).

1.7 Methods to study lncRNAs

Current efforts to study the function of newly identified lncRNAs have been biased by the available methods towards those in which the RNA molecule is functional (Guttman et al., 2011; Ulitsky et al., 2011) and in some cases different mutation strategies uncover conflicting results (Pease et al., 2013; Petruk et al., 2006). The availability of new genome engineering technologies will allow for more precise modifications, making it easier to study lncRNAs, particularly those acting in *cis*.

1.7.1 shRNA, siRNA or dsRNA knockdown

Knock down of lncRNAs transcripts can be achieved using siRNA, shRNA and morpholino technologies (Figure 1.2g) (Clemson et al., 2009; Guttman et al., 2011; Ulitsky et al., 2011). These methods are most effective on transcripts present in the cytoplasm. Nevertheless, there is increasing evidence of efficacious knock down of nuclear transcripts via these methods. In mammalian cell culture systems, gene expression changes and cellular phenotypes have been observed following shRNA knock down of lncRNAs that are predominantly localised to the nucleus (Chalei et al., 2014; Vance et al., 2014). Furthermore, antisense oligos can be used to induce RNase H mediated degradation in the nucleus, co-transcriptionally, but suffer from off-target effects (Figure 1.2f) (Sahu et al., 2007). Knockdown experiments can provide clear evidence for lncRNA function, particularly when multiple constructs produce the same results, reducing the likelihood of off-target effects. However, these methods will only identify changes if the RNA molecule is functional. Furthermore, it is not known what factors influence the efficacy of

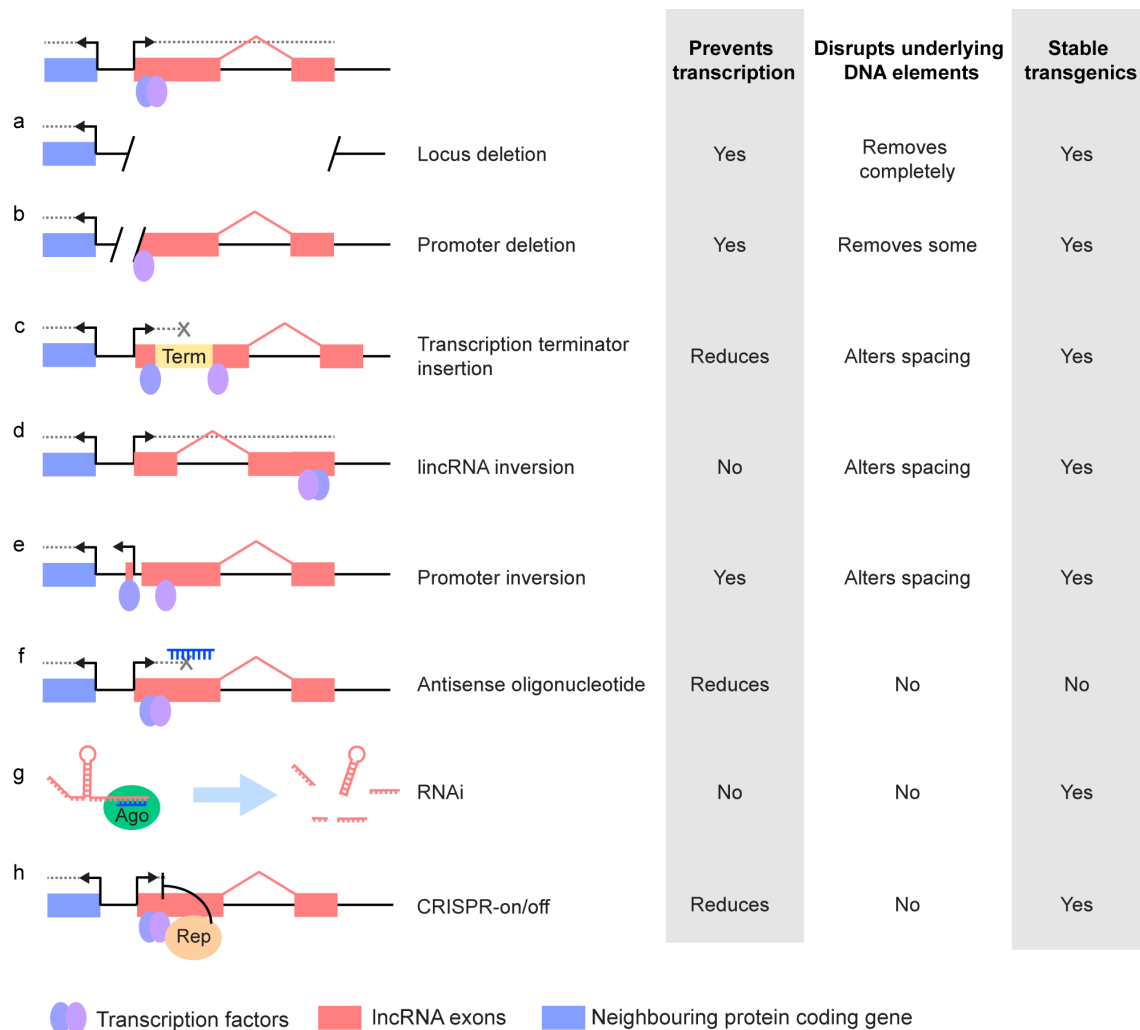


Figure 1.2 Mutation strategies available to study lncRNA function. Deletion of (A) the whole locus or (B) promoters can prevent transcription but **also** deletes underlying functional elements. (C) Insertion of terminators can also prevent transcription with minimal effect on underlying DNA. Inversion of (D) the lncRNA locus or (E) the promoter can change the orientation of transcription relative to the neighbouring gene without removing DNA elements. (F) Antisense oligos can knockdown lncRNAs co-transcriptionally reducing transcription and levels of lncRNA without affecting the DNA sequence. (G) Knockdown by RNAi reduces lncRNA level without affecting underlying DNA. (H) Inactive Cas9 can be targeted to lncRNA locus using gRNAs and coupled with activator or repressor domains to change transcription level without affecting underlying DNA sequence. Adapted from (Bassett et al., 2014a).

knockdown and in some cases the level of knockdown achieved is not sufficiently great to have any effect (Grote et al., 2013).

1.7.2 Genetic deletion

Genetic deletion of lncRNA loci or promoters offers an unbiased approach to studying lncRNA function, identifying changes for lncRNAs whether it is the act of transcription or the RNA molecule that is functional (Figure 1.2a and b). However, caution is required when interpreting results of genetic deletion as they may also remove important underlying DNA elements. Sauvageau et al. (2013) recently published preliminary results of the generation of eighteen mice strains in which lncRNA loci were replaced with lacZ reporters. This system provided both lncRNA null mutants and information on the expression pattern of the lncRNAs. Three of the twenty strains are homozygous lethal during early development, including *Fendrr*^{-/-} mutants, that die before weaning age due to multiple organ failure. However, it is not yet clear whether this is due to loss of the lncRNA or disruption of elements in the underlying DNA sequence (Bassett et al., 2014a). To confirm lncRNA-dependent phenotypes would require rescue experiments to be performed to replace lncRNA function.

1.7.3 Termination

To prevent transcription without deleting any underlying sequence, transcriptional terminators can be inserted (Figure 1.2c). For example, Grote et al. (2013) generated an independent *Fendrr* mutant in which they replaced the first intron with a strong terminator sequence, which resulted in complete loss of transcription of *Fendrr* (Grote et al., 2013). Grote et al. (2013) found that loss of *Fendrr* was lethal earlier than described by Sauvageau et al. (2013) due to defects in heart and body wall development. Importantly, this could be rescued by one copy of a bacterial artificial chromosome (BAC) containing the *Fendrr* locus with an inactive copy of

the neighbouring gene. Rescue experiments are particularly important in the case of lncRNAs to differentiate between effects caused by loss of DNA elements and loss of the lncRNA. Rescue using injected RNA or a cDNA transgene can confirm that it is loss of the RNA not disruption of the underlying DNA sequences that is functional. Rescue using a DNA fragment such as a BAC can be more effective as the lncRNA is under the control of its endogenous regulatory sequences, but it may be necessary to inactivate the neighbouring genes if they are also included in the BAC.

1.7.4 Inversion

The efficiency of termination is dependent on the genomic context and so may not always be effective (Latos et al., 2012), and still results in alteration of spacing of DNA elements. Alternatively, the promoter or whole locus can be inverted (Figure 1.2d and e). These techniques allow for retention the DNA elements, albeit with altered orientation, and change the direction of transcription. In the *Bdx/Ubx* locus, this allowed the differentiation between a transcriptional interference mechanism that had been suggested by a promoter deletion (Petruk et al., 2006) and loss of an enhancer element supported by promoter inversion (Pease et al., 2013).

1.7.5 Genome engineering

New TALEN and CRISPR/Cas9 genome engineering methods will allow for more precise genetic manipulations (Bassett et al., 2013; Jinek et al., 2012) which can be used to perturb lncRNAs. Indels generated by CRISPR/Cas9 could be used to delete promoters or splice sites to prevent production of the mature transcript, or delete putative start codons to confirm that lncRNAs are non-coding. Alternatively, combined with homologous recombination it is possible to introduce reporters or termination signals or attP sites to allow subsequent integration of multiple different constructs. Recruitment of catalytically inactive Cas9 to promoters has

been shown to block transcription (Larson et al., 2013) and can be used recruit repressor or activator domains to achieve greater repression or upregulation of the gene from its endogenous locus (Figure 1.2h) (Gilbert et al., 2013). Cas9 can also bind to single stranded RNA if a DNA oligo is supplied, resulting in cleavage and subsequent downregulation of the RNA. It is also possible to use a catalytically inactive Cas9 protein to block binding of other factors or to pull down specific mRNAs (O'Connell et al., 2014).

1.8 microRNAs

microRNAs (miRNAs) are a distinct class of endogenously encoded small RNAs that were first identified in *Caenorhabditis elegans* (Lee et al., 1993), which are found in multi-cellular animals and plants. miRNAs negatively regulate levels of a target gene by binding mRNAs via regions of sequence complementarity between the miRNA and its target mRNA.

1.8.1 miRNA biogenesis

Mature miRNAs are usually 22nt in length (Lagos-Quintana et al., 2001) and are processed from introns or exons of longer primary transcripts (Figure 1.3) which are predominantly capped and poly-adenylated RNA polymerase II transcribed genes (Lee et al., 2004a; Rodriguez et al., 2004). The first stage of processing occurs in the nucleus (Lee et al., 2002): PASHA/DGCR8 recognises and binds double stranded RNA hairpins in primary transcripts (pri-miRNAs) (Landthaler et al., 2004) and DROSHA is recruited and cleaves the transcript near the base of the hairpin via its RNase III domain (Lee et al., 2003). The resultant product (pre-miRNA) is recognised by its two nucleotide 3' overhang and exported from the nucleus (Lund et al., 2004; Yi et al., 2003). Cytoplasmic RNase III enzyme DICER cleaves the loop connecting the 3' and 5' arms to give a 22nt duplex, one arm is incorporated into the RISC complex to be used as the functional miRNA (Lee et al., 2004b; Schwarz et al., 2003). Which arm is functional varies between different miRNAs and in some cases both arms can be functional (Okamura et al., 2008), targeting different sets of target genes (Okamura et al., 2009). miRNAs with the same mature sequence can be derived from different primary transcripts.

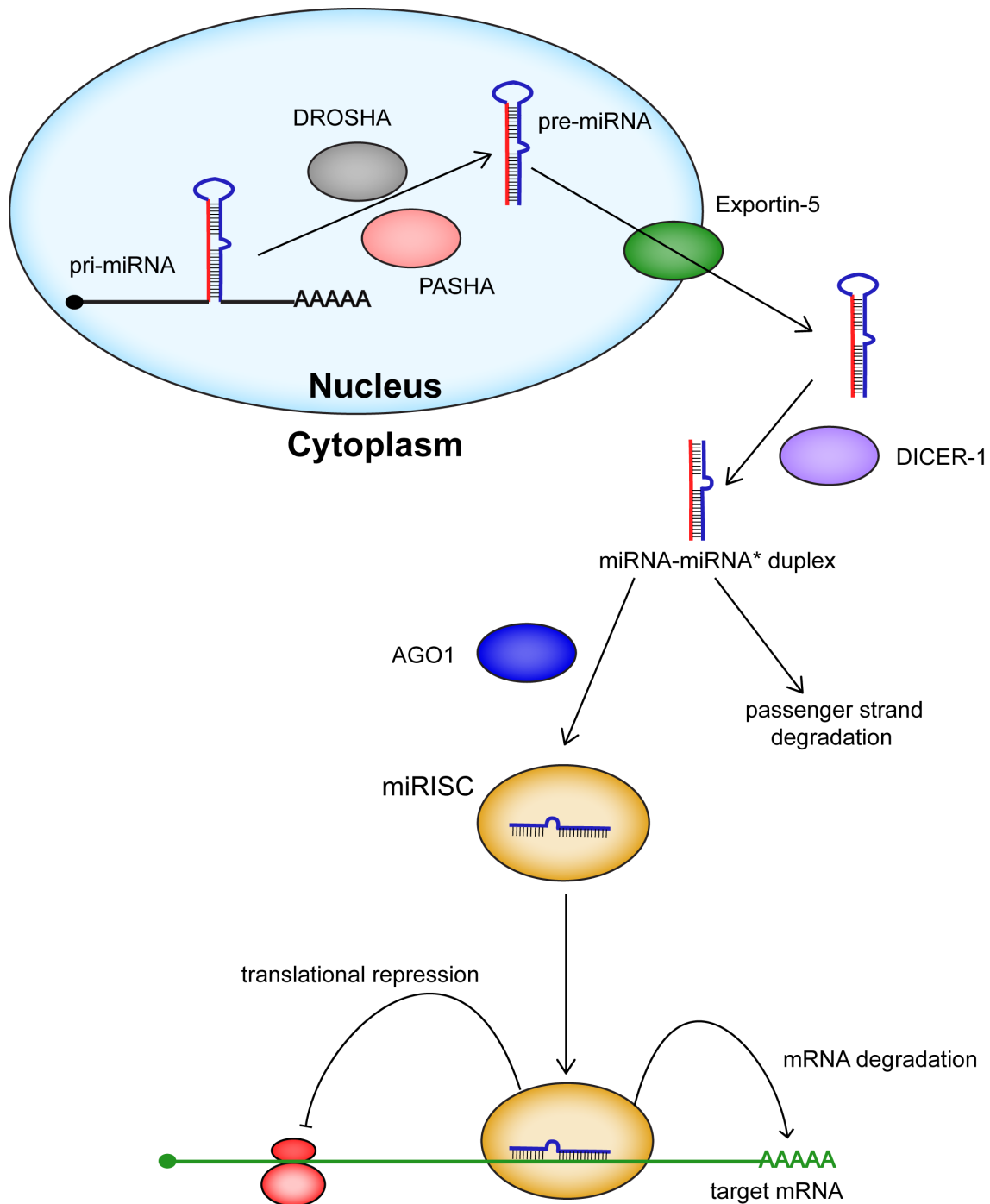


Figure 1.3 miRNA biogenesis pathway. Pre-miRNA hairpins are processed from longer transcripts in the nucleus by PASHA and DROSHA and exported to the cytoplasm by Exportin-5. Pre-miRNA is cleaved into miRNA-miRNA* duplex by DICER-1, typically one strand is incorporated into the RISC complex and the other is degraded. The miRISC complex binds mRNAs via regions of complementarity with the miRNA and causes degradation and translation repression of the target.

1.8.2 microRNAs control gene levels through post-transcriptional regulation

In animals pairing between a miRNA and its target is rarely perfect and mainly occurs through complementarity of nucleotides 2-7 of the miRNA (seed region). However, a recent genome-wide screening experiment in humans showed that only 37% of miRNA-target interactions have a perfect seed match, only 19% of interactions are mediated solely through seed matches and 16% of interactions have no seed match at all (Helwak et al., 2013). The secondary structure of the target sequence has also been shown to influence the efficacy of binding (Long et al., 2007). Target repression is not generally mediated by mRNA cleavage in animals as cleavage of transcripts requires perfect complementarity, especially around base 10 of the miRNA; instead miRNAs cause post-transcriptional repression. Through both transcript destabilisation and degradation by deadenylation and by disruption of translation.

Both mechanisms are mediated through GW182 (*gawky/dGW182*) (Behm-Ansmant et al., 2006). GW182 competes with eIF4G for binding of poly-A binding protein (PABP) to prevent circularisation of the transcript which is required for efficient translation (Zekri et al., 2013). GW182 also has a conserved ability to directly recruit deadenylases and promote deadenylation resulting in transcript degradation (Braun et al., 2011; Chekulaeva et al., 2009). High-throughput analyses indicate that repression occurs primarily as a result of transcript degradation and reduced translational efficiency is a secondary effect that is in part caused by reduced transcript levels (Guo et al., 2010). However, in some instances including pre-gastrulation in zebrafish embryos when translational efficiency is influenced by poly(A) tail length, translational repression dominates (Subtelny et al., 2014).

1.8.3 miRNAs have functions in development and disease

Current evidence suggests that miRNAs function to fine tune the levels of mRNAs, often as part of complex regulatory pathways that can feedback to control the level of the miRNA at the transcriptional level (Jin et al., 2012; Tan et al., 2014; Yuva-Aydemir et al., 2015). miRNAs have been found to have roles in a diverse range of biological processes. Using the *Drosophila* miR-8 as an example, it is thought to influence body size, body pigmentation, axon guidance, NMJ formation and function, and brain development. miRNAs are also emerging as prominent biomarkers and therapeutic targets for many disease states, including cardiovascular disease (Olson, 2014) and multiple cancer subtypes (Hayes et al., 2014).

1.8.4 Methods to predict miRNA targets

There is currently no gold standard for predicting miRNA-target interactions because multiple factors need to be considered. Firstly, not all miRNA-target interactions are mediated through seed matches (Helwak et al., 2013). Secondly, the effect of secondary structure of the target can affect binding (Long et al., 2007). Thirdly, binding requires the miRNA and target to be expressed in the same cell types and binding is also influenced by the relative levels of the miRNA and its targets, and also the levels other targets that share binding sites for the same miRNA. Finally, if binding does occur, it may not always result in repression of the target influenced by the location within the transcript and the presence of other factors bound to the target.

***In silico* predictions**

The most common method is *in silico* prediction based on identification of 7mer and 8mer seed matches. Different algorithms give different weight to influencing factors, such as conservation of the seed site (Ruby et al., 2007); non-seed

base-pairing and free energies (Enright et al., 2003); or conservation of the seed site and free energies (Grun et al., 2005). These algorithms produce many false positives, including seed matches that are not targeted *in vivo* and miRNA-target interactions that are not physiologically relevant as the two entities are never co-expressed. The rate of false positive prediction is estimated to be approximately 60% (Rajewsky, 2006). False negative results come from targets without seed matches and publically available databases often only include targets in the 3'UTR of protein-coding mRNAs which only account for 23.4% of all miRNA-mRNA interactions in one genome-wide study conducted in humans (Helwak et al., 2013).

Experimental miRNA-target interaction detection

A number of methods that identify miRNA-target interactions by capturing physical interactions between miRNAs and mRNAs using pull-down of proteins in the RISC complex with associated RNAs have been shown to work in cell culture systems: HITS-CLiP, PAR-CLiP and CLASH (Chi et al., 2009; Hafner et al., 2010; Helwak et al., 2013). Through these techniques, common rules for binding can be established that feedback into prediction algorithms. For example Chi et al. (2012) identify a G bulge at the 5-6 seed position in more than 20% of all miR-124/target interactions in the mouse brain. However, these methods will only capture the subset of interactions occurring in that particular cell line and therefore do not offer a complete set of miRNA-target interactions and not all interactions will result in repression of the target. Techniques are currently being developed to allow cell-type specific pull-down of miRNA-target interactions in *Drosophila*: for example, modified CLASH, and PAR-CLiP protocols in which a tagged-Ago1 is expressed in specific cell types, which may in the future provide a definitive approach to identify cell type specific miRNA-target interactions.

1.8.5 Methods to confirm miRNA-target interaction

Gene expression profiling

Repression of most targets by miRNAs results in reduction of transcript levels. Therefore, genome wide expression profiling can be used to detect genes that are upregulated in miRNA mutants, or downregulated upon miRNA overexpression. However, changes in expression level alone cannot distinguish between direct and indirect effects so can be combined with *in silico* predictions to identify predicted targets whose abundance is changed (Karres et al., 2007) or with *in vivo* miRNA-target interaction discovery to confirm which miRNA-target interactions result in gene expression changes (Helwak et al., 2013).

Luciferase/GFP -3'UTR fusion reporters

Putative miRNA-target interactions can also be tested through *in vitro* assays including luciferase reporters in which the 3'UTR with the intact MRE and a mutated or deleted MRE is fused to a luciferase reporter and levels are assessed in the absence or presence of the miRNA. This method has been successfully used to identify a number of targets of miR-8 in flies (Hyun et al., 2009; Morante et al., 2013). Also, GFP reporters can be generated in which 3'UTRs are fused with wildtype or mutated MREs and expressed in the cell-type of interest to assess the ability to target *in vivo*. However results of these assays need to be interpreted with caution as they do not represent the stoichiometry of the endogenous situation.

Genome engineering

Currently, *in vivo* confirmation of targeting requires either manipulation of miRNA levels, which often results in indirect effects or the use of target protector oligonucleotides which is not possible in many systems (Choi et al., 2007). Alternatively, ectopic expression of a 3'UTR fusion reporter can be used, but this might result in interactions that do not occur under physiological conditions.

Indels generated using CRISPR/Cas9 allow for deletion of specific MREs within the endogenous gene (Bassett et al., 2014b) and thus define the function of specific MREs *in vivo*.

1.9 *Drosophila melanogaster* as a model organism

Drosophila melanogaster was chosen as the model organism for this thesis because much of the previous work done to study the functions of lncRNAs had been carried out in mammalian cell culture system, and therefore little is currently known about the roles of lncRNAs at the organismal level. *D. melanogaster* has been favoured as a genetic organism for over 100 years because it exhibits many advantages over other systems. Firstly, flies are small, easily cultured, have a short life cycle and have large numbers of progeny, allowing for rapid expansion of flies containing new genotypes. Secondly, there is a wide range of genetic techniques available including transposon insertion mutants that cover a large proportion of the genome (Bellen et al., 2011; Bellen et al., 2004), the GAL4/UAS system that allows temporally and spatially specific expression of transgenes (Brand and Perrimon, 1993) and the ability to create mutant clonal patches (Lee and Luo, 2001).

These features make *D. melanogaster* an ideal system in which to study lncRNAs in the context of a whole organism. The availability of existing mutant strains, in particular transposon insertions strains that are randomly integrated and include many insertions within intergenic regions that may include lncRNA loci, removes the cost and risk associated with making novel lncRNA mutants in mice. Moreover, given that lncRNAs are known to be expressed in temporally and spatially restricted patterns, the ability to make manipulations with temporal and spatial specificity will allow for separation of the functions of a lncRNA in different developmental

contexts. The ability to easily generate clonal mutant tissue within a wildtype background will also be important for the study of lncRNAs, if subtle effects do not manifest in phenotypes at the whole organism level. Moreover, a large number of putative lncRNAs have recently been identified in *D. melanogaster* (Brown et al., 2014; Graveley et al., 2011; Young et al., 2012) providing a large number of putative lncRNAs from which to find appropriate loci to study. Together with the wealth of information known about *D. melanogaster* life cycle, development, life span and immune response this offers the opportunity to study the roles of lncRNAs in a wide range of biological processes.

1.10 Project aims and thesis structure

In this thesis, I sought to identify the relationship between phenotypic and gene expression changes that occur upon disruption of two classes of non-coding genes (lncRNAs and miRNAs) in *D. melanogaster* using a combination of computational, genome-wide screening and experimental approaches. The main dataset used in this thesis is the set of lncRNAs described by Young et al. (2012) that are intergenic (do not overlap known genes on either strand), multi-exonic and polyadenylated and all transcripts of the gene model were annotated as non-coding by PhyloCSF (Lin et al., 2011).

In **Chapter 3**, I describe the initial characterisation of six putative lncRNA loci which contain transposon insertions, to identify a candidate lncRNA to investigate in further detail. I find that the majority of loci are unsuitable for further study and that the effect of a transposon insertion within lncRNA loci depends on the nature of the modular elements of the insertion and its genomic context.

In **Chapter 4**, I analyse two independent transposon insertion mutant strains within the *Inc703* locus in more detail and find that one contains a closely-linked second

site mutation causing pupal lethality. I identify changes in immune response genes in both mutants using transcriptome profiling. These changes cannot be reproduced in strains with an isogenised background or when *Inc703* is knocked down using shRNAs and are therefore dependent on the genetic background.

In **Chapter 5**, I use a GAL4/UAS fluorescent reporter system to identify cells and tissues in which *Inc703* is expressed and find it to be expressed in many larval tissues including the larval fat body. Analysis of the larval fat body of *Inc703* mutants identified no gross morphological changes but transcriptome profiling found elevated levels of ribosomal protein transcripts in both *Inc703* mutant strains.

In **Chapter 6**, I describe work carried out in collaboration with Dr Markus Toegel in which I use transcriptome profiling of the CNS and musculature of miR-8 null larvae to identify tissue specific targets of miR-8. Gene expression changes in both tissues reflect both direct and indirect effects of loss of miR-8. Genes upregulated in the muscle but not in the CNS are enriched for genes whose transcripts contain miR-8 binding sites predicted by computational methods. We also find that, in the case of miR-8, deletion of the sequence encoding the hairpin results in accumulation of the primary transcript which might contribute to the mutant phenotype.

In **Chapter 7**, I discuss these findings in the context of the current understanding in the non-coding RNA field and make recommendations as to how these effects could be investigated further.

Chapter 2 - Methods

2.1 Wet lab methods

2.1.1 Fly Husbandry

Flies were maintained at 25 °C on standard *Drosophila* medium (80 g/L maize, 18 g/L dried yeast, 10 g/L soya flour, 80 g/L malt extract, 40 g/L molasses 8 g/L agar, 6.6 ml acid mix) with a 12-hour light/dark cycle, unless otherwise stated.

2.1.2 Isogenisation

All IncRNA mutant strains were isogenised by iteratively crossing virgin females containing the transposon insertion to w^{1118} males and collecting the female progeny that have retained the transposon insertion. A minimum of seven generations of back-crossing was carried out, homozygous lethal stocks were then crossed to an appropriate balancer.

2.1.3 PCR

As standard a 25 µl PCR reaction using 0.5 µl Taq Polymerase and 0.5 µl of cDNA template, 2 µl of gDNA or 5 µl of lysed bacterial template were used. The following reagents were assembled on ice to give final working concentrations shown in brackets: Taq Polymerase (0.5 Units), 10X NH₄ Reaction buffer (1X) (Bioline), 50 mM MgCl₂ (1.5 mM), 10 mM dNTPs (250 µM), 10 µM forward primer and 10 µM reverse primer (0.4 µM each), DMSO (5%). Templates were denatured by heating to 95 °C for 2 min and amplified by 30-40 cycles of denaturation (15 seconds, 95 °C), annealing (30 seconds, 55-60 °C) and elongation (60 seconds/1000 bp, 72 °C) in a G-Storm GS4 thermal cycler. PCR products were run on a 1% agarose gel in 1X TAE (40mM Tris, 20mM acetic acid, and 1mM EDTA).

2.1.4 Vector transformation

50 µl aliquots of DH5α chemically competent cells were thawed on ice, 2 µl of plasmid DNA was added, and incubated on ice for 20 min, heat-shocked at 42 °C for 30 seconds, returned to ice for 2 min then spread on LB agar plates (10 g/L NaCl, 10 g/L Tryptone, 5 g/L Yeast Extract, 15 g/L Agar) containing the appropriate antibiotic. For blue-white selection in pGEM-T Easy, 1M IPTG and 20 mg/ml X-gal were incorporated in to normal LB agar plates to give final concentrations of 0.1 mM IPTG and 40 µg/ml X-gal.

2.1.5 Colony PCR

Bacteria from individual colonies were picked using a sterile 200 µl pipet tip and resuspended in 50 µl ddH₂O. 5 µl of bacteria/water mix was removed and heated to 98 °C for 10 min to lyse the bacteria and used as PCR template. 45 µl of 2x LB was added to remaining bacteria and stored at 4 °C, positive colonies were grown overnight in liquid culture. DNA was extracted from 4 ml liquid cultures grown overnight using a QIAprep Spin Miniprep Kit (Qiagen), 5 µl of 100 ng/µl of DNA was used for Sanger sequencing (performed by Source BioScience UK) which was used to confirm the correct sequence.

2.1.6 Genotyping

Two whole adult flies were homogenised in 49µl Squishing Buffer (10 mM Tris-HCl pH 8.2, 1 mM EDTA pH 8.0, 25 mM NaCl), with 1 µl Proteinase K (200 µg/ml), incubated at 37 °C for 30 min then heated to 95 °C for 2 min to inactivate the enzyme. 2 µl of homogenate was used per 25 µl PCR reaction as previously described. PCR products were run on 1% agarose gel to confirm correct size and gel extracted using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's protocol. Purified PCR product was ligated in to pGEM-T easy vector as previously described, and white colonies were screened by colony PCR.

2.1.7 RNA extraction

Tissues were dissected in cold RNase-free PBS, whole larvae were rinsed in RNase-free PBS and pupae and adults were transferred directly to tubes. Samples were flash-frozen in liquid Nitrogen and stored at -80 °C. Total RNA was extracted using the miRNeasy Micro Kit (QIAGEN). Briefly, samples were homogenised in 70 µl of QIAzol using an electric pestle to disrupt tissues, made up to 700 µl with QIAzol and further homogenised and debris removed using a QIAshredder. After incubating at room temperature for 5 min to ensure full denaturation of proteins, 140 µl of chloroform was added and mixed by vigorous shaking. Phase separation was achieved by spinning at 12,000 g at 4 °C for 15 min. The aqueous phase containing the RNA was removed and diluted with 1.5 volumes of 100% ethanol. RNA was bound to the spin column then was washed with RWT and RPE buffers to remove remaining contaminants. An additional DNase digestion step using RNase-Free DNase Set (QIAGEN) was performed between RWT washes to remove gDNA contamination. RNA was eluted in 30-50 µl RNase-free ddH₂O depending on the expected RNA yield.

2.1.8 Reverse Transcription

100-1000 ng of RNA was reverse transcribed using the Quantitect RT Kit (QIAGEN) dependant on the RNA yield. Briefly, RNase-free ddH₂O was added to 2 µl of gDNA Wipeout buffer and RNA to make a final volume of 14 µl and heated to 42 °C for 4 min to remove any contaminating gDNA. A mix of 4 µl 5X RT Buffer, 1 µl Reverse Transcriptase Enzyme and 1 µl RT primer mix (oligo dT and random hexamers) was added to give a final volume of 20 µl. For one biological replicate of each sample an RT negative reaction in which the Reverse Transcriptase was replaced with water was carried out. Resulting cDNA was diluted and stored at -20 °C.

2.1.9 qRT-PCR

For quantitative analysis of RNA levels, qRT-PCR was performed using the SYBR Green JumpStart Taq ReadyMix (Sigma) system. cDNA was prepared as described previously. 17.5 µl of 2x ReadyMix (containing 20 mM Tris-HCl, pH 8.3; 100 mM KCl; 7 mM MgCl₂; 0.4 mM each dNTP (dATP, dCTP, dGTP, TTP); stabilizers; 0.05 Unit/µl Taq DNA Polymerase; JumpStart Taq antibody, SYBR Green I and 0.5X Internal Reference Dye), 1.75 µl DMSO and 10.75 µl of ddH₂O per sample were added to 3 µl of cDNA, mixed thoroughly and 10 µl was added to triplicate wells. All qRT-PCR experiments were performed with technical triplicates and a minimum of three biological replicates. Primers used are listed in Appendix B.

2.1.10 Rapid Amplification of cDNA Ends (RACE)

RNA was extracted as described previously. FirstChoice RLM-RACE Kit (Life Technologies) was used according to manufacturer's guidelines to perform 3' and 5' RACE. For 5' RACE the 5' phosphate of degraded and uncapped transcripts is removed by treatment with Calf Intestine Alkaline Phosphatase (CIP). The m⁷G cap is subsequently removed by Tobacco Acid Pyrophosphatase (TAP) and an RNA adaptor ligated specially to the 5' phosphate of uncapped transcripts but not dephosphorylated transcripts by T4 RNA Ligase. The reverse transcription reaction uses random primers and the first and second round PCR reactions use nested forward primers complementary to the adaptor sequence and nested reverse primers specific to the gene of interest. For 3' RACE the RNA adaptor is added by using an extended primer containing the adaptor sequence and an oligo-dT₁₂ for the reverse transcription step and primers complementary to the adaptor are used in the first and second rounds of PCR with nested gene specific primers. For samples with low RNA concentration, reaction volumes were adjusted to ensure minimum of 5 µg starting material for 5' RACE due to loss of RNA during

early processing steps. Primers were designed according to manufacturer's guidelines. Products visible after the second round of PCR with nested primers were gel extracted, transformed and Sanger sequenced as previously described. Sequences were aligned to the BDGP5.58 *D. melanogaster* genome assembly using BLAT (Kent Informatics, Inc.).

Alternatively, a circular RACE protocol was used in which the 5' and 3' ends of a transcript are ligated. After reverse transcription, PCR is performed with a forward primer designed to the 5' end and a reverse primer designed to the 3' end. This technique has the advantage of confirming the two ends are part of the same transcript. Briefly, RNA was extracted as previously described, and the phosphate groups were removed from 1.5 µg of RNA using 2U FastAP Thermosensitive Alkaline Phosphatase (Fermentas), 10x T4 Ligase reaction buffer (NEB), with 0.2 µl RNase Inhibitor, diluted to 20 µl water ddH₂O and heated to 37 °C for 30 min. Ends were ligated by adding 1 µl T4 RNA Ligase (NEB) and heat to 37 °C for 30 min. Reaction was stopped by heating to 65 °C for 10 min. 6 µl of ligation reaction was reverse transcribed using Superscript III RT (Life Technologies) according to manufacturer's guidelines, heating to 25 °C for 10 minutes, 50 °C for 30 minutes and heat inactivating at 85 °C for 5 minutes. Remaining RNA in RNA:DNA hybrids was degraded by adding 1 µl RNase H (Life Technologies) and heating to 37 °C for 20 min. Two rounds of nested PCR were performed using primers to the 5' and 3' ends concurrently. Products were gel extracted, sequenced and aligned to the genome as previously described.

2.1.11 Immunofluorescence

Tissue were dissected in PBS and fixed in 4% Paraformaldehyde (PFA) for 10 min (larval imaginal discs), 15 min (larval fat bodies) or 20 min (larval CNS or adult brains). Dissected tissues were stored at 4 °C in PBST (0.5% Horse Serum, 0.2% TritonX-100). As standard primary antibodies were incubated at 4 °C for 24

hours and secondary antibodies were used (1:1,000) incubated at 4 °C overnight.

2.1.12 Transgenic fly stock generation

All constructs were targeted to attP sites on a different chromosome arm to the endogenous gene to reduce the impact of integration site on transgene expression. For *Inc703* constructs this was the VK37 attP site on chromosome 2L. Injections were performed by Genetivision (Houston, TX). Red-eyed flies, without the appropriate balancer trait (i.e. straight-winged for injections on chr2L) were genotyped as previously described to confirm the insertion.

(i) Bacterial Artificial Chromosome (BAC) genomic rescue constructs

attB-P[acman]-CmR-BW vectors containing the BAC corresponding to the appropriate genomic region containing the lncRNA locus were obtained from Drosophila Genomics Resource Center (DGRC) (Table 2.1). Vectors were injected in to appropriate flies containing attP site on a different chromosome to the endogenous lncRNA locus by Genetivision. A full list of the BAC used for each lncRNA can be found in Table 2.1

Table 2.1 Bacterial artificial chromosomes containing lncRNA loci.

BAC	lncRNA	Fully covered protein-coding genes	Other lncRNAs	Length
CH321-73E01	<i>Inc66</i>	<i>Wg, Wnt6, Wnt10, ninaC</i>	<i>CR45304, CR44063</i>	106,451bp
CH322-134A09	<i>Inc367</i>	<i>CG17999</i>	<i>CR44667</i>	20,794bp
CH322-103H07	<i>Inc458</i>	<i>vvl</i>	<i>CR45759</i>	22,151bp
CH322-75A10	<i>Inc626</i>	<i>l(3)76BDr, Rcd7 CG9300, miR-9</i>	<i>CR44677, CR44676</i>	23,954bp
CH322-158D05	<i>Inc703</i>	<i>Ref1, Dpck, CG1091, alphaTub84B, e(y)2b</i>	<i>CR43685</i>	20,091bp

(ii) Short-hairpin RNA (shRNA) constructs

For shRNA mediated knock-down transgenic flies, hairpins were designed using the DSIR tool (Vert et al., 2006), three hairpins were selected that targeted different regions of the *Inc703* transcript. Hairpins were cloned into the pWALIUM22 vector as described by Ni and Perrimon (www.flyrnai.org/supplement/2ndGenProtocol.pdf). Briefly, ssDNA oligos containing sense and antisense sequences separated by a linker sequence and flanked by overhangs corresponding to 5' *Nhe* I and 3' *Eco* RI restriction sites were ordered from Eurofins Genomics (Ebersberg, Germany). Sense and antisense oligonucleotides were annealed by heating to 98 °C and gradually reducing the temperature to 18 °C in annealing buffer (20 mM Tris, 2 mM EDTA, 100 mM NaCl, pH 8.0). Annealed oligonucleotides were incubated at 37 °C for 30 min with T4 Polynucleotide Kinase in T4 DNA Ligase buffer containing 10 mM ATP (NEB) to phosphorylate the 5' nucleotides. Phosphorylated oligonucleotides were diluted 1:10 and 1.5 µl of oligo was ligated in to the pWALIUM22 vector that had been linearised by digestion with *Nhe* I and *Eco* RI for 2 hours at 37 °C. Vectors were transformed into DH5α *E. coli* cells, plated on to LB agar plates containing Carbenicillin (100 mg/ml) and grown overnight at 37 °C. Transformed clones were screened by colony PCR and sequences were checked by Sanger sequencing using a primer specific to pWALIUM22 (Appendix A).

(iii) Overexpression constructs

For UASp-driven overexpression constructs, the DNA sequences of both *Inc703* isoforms flanked by *Not* I and *Asc* I sites were synthesised by Life Technologies (Paisley, UK). A 323 bp region of the *Inc703* transcripts that had been identified as double stranded (Li et al. 2012) was deleted by an adapted Quickchange mutagenesis protocol using primers containing 21/24 bp of homology to the regions flanking the deleted region (Appendix A, 703del_45bp_F/R).

A UASp promoter sequence and SV40 terminator sequence were amplified from appropriate vectors using primers containing restriction sites (Table 2.3), digested and sequentially cloned in to the multiple cloning site (MCS) of the pBID vector (Wang et al. 2012). The pBID vector contains an attB phiC31 integration site, *gypsy* insulator sequences flanking the MCS and a *white* selectable marker (Wang et al. 2012). Subsequently, one of four cDNA constructs which represent the two isoforms of *Inc703* with or without a 323 bp deletion were inserted between the promoter and terminator sequences. Details of the digestions carried out at each step are shown in Table 2.2. PCR products and vectors were digested using the same conditions, with 1 µl CIP added to the vector digestion for the final 15 min to prevent self-ligation of the vector. All digestion products were run on a 2% agarose gel and gel extracted as described previously. Ligations were carried out using T4 DNA ligase (NEB), in 10X T4 Ligase Buffer (1X), using 1 µl digested vector and 7 µl of digested PCR product.

Table 2.2 Digestion conditions for generation of cDNA overexpression constructs.

Feature	Digest 1		Digest 2		Digestion conditions	Origin
	Enzyme(s)	Buffer	Enzyme	Buffer		
UASp	<i>Eco</i> RI (NEB) <i>Not</i> I (HF) (NEB)	Buffer 4 + BSA			1 hour 37°C	pWALIU22
SV40	<i>Not</i> I (Fermentas) <i>Xho</i> I (NEB)	Buffer 0			1 hour 37°C	mENTRY-SV40
<i>Inc703</i> -A	<i>Asc</i> I (NEB)	Buffer 4 + BSA	<i>Not</i> I (Fermentas)	Buffer 0 (inc. BSA)	1 hour 37°C PCR purification 1 hour 37°C	
<i>Inc703</i> -A_ del						
<i>Inc703</i> -B						
<i>Inc703</i> -B_ del						

2.1.13 Triacylglyceride assay

Larval fat bodies were dissected in cold PBS and flash frozen in liquid Nitrogen. Samples were homogenised in 200 µl of lysis buffer (PBS + 1% Tween) and heated to 65 °C for 5 min to inactivate lipases. TAG levels were determined using Infinity™ Triglycerides Liquid Stable Reagent (Thermo Scientific) according to manufacturers' instructions. Briefly, 198 µl of reagent was added to 2 µl of sample or TAG standard (0 – 600 mg/dl), incubated for 5 min at 37 °C and absorbance measured at 540 nm. TAG levels were normalised to protein content. Protein content was determined using the Pierce BCA protein assay (Life Technologies) according to manufacturers' instructions. Briefly, 25 µl of sample or protein standard (0 – 2000 µg/ml) was added to 200 µl WR buffer, mixed thoroughly and incubated for 30 min at 37 °C before cooling to room temperature and measuring absorbance at 562 nm.

2.1.14 Western Blotting

Tissues were dissected in cold PBS containing (1x) protease inhibitors; excess PBS was removed and flash frozen. 70 µl of cold RIPA buffer + 1 mM EDTA (50 mM Tris-Cl, pH 8.0, 150 mM NaCl, 0.1% SDS, 1% sodium deoxycholate, 1% TritonX-100, 1 mM EDTA) was added to samples before homogenising with an electric pestle for 1 minute. Samples were left on ice for 5 minutes to ensure complete lysis. Samples were spun at 13,000 rpm for 15 min; supernatant was removed and heated at 85 °C for 10 min with 15 µl of denaturing buffer and 50 µl of loading dye. Samples were mixed before loading into a 4-12% Bis-Tris gel (Life Technologies). Gel was run at 200 V for 40 min and transferred to a PVDF membrane at 30 V for 1 hour. Membrane was blocked for 1 hour in 5% milk, primary antibodies were incubated overnight at 4 °C, loading control mouse anti-β-tubulin (T6793, Sigma) was used 1:2,000 and mouse anti-rpS6 (Cell Signalling Technologies) was used 1:200 in TBST with 1% milk. Anti-Mouse-HRP secondary

antibody was used 1:10,000, at room temperature for 2 hours. Membranes were imaged using Image quant LAS400 (GE Healthcare).

2.1.15 Longevity study

Newly eclosed flies were allowed to mate for 24 hours before separating the sexes and placing sets of 20 flies in vials containing standard *Drosophila* medium. Dead flies were counted every two to four days and flies were transferred in to clean vials every four days.

2.2 Computational Methods

2.2.1 Defining transcriptional units for differential gene expression analysis

To reliably call differential expression of genes between different samples it is necessary to define the regions of a gene to prevent a single read being mapped multiple times to different isoforms of a gene. 'Constitutively expressed' regions, were defined as concurrent nucleotides that were present in greater than or equal to 75% of transcripts annotated to a gene. This increases the number of retained nucleotides and reduces the effect of spurious annotations which represent very lowly expressed or tissue restricted isoforms.

2.2.2 Building a reference library

All protein-coding and pseudogene exons were extracted from a gene transfer format (GTF) file of the BDGP5.68 downloaded from Ensembl, these were then compiled into a mock BED file in which the gene name replaced the chromosome and the transcript name replaced the gene name. This allowed calculation of the number of transcripts assigned to each gene. CoverageBed was run using the option `-d` giving the coverage (which is equal to the number of transcripts)

with the inputs –a being all exons and –b being a file in which all overlapping exons of the transcript are merged (generated using mergeBed). Nucleotides that were included in greater than or equal to 75% of transcripts of a given gene were called as constitutive and blocks were made containing consecutive constitutively expressed nucleotides. These were converted to a normal BED format: chromosome, start, end, gene name.

For lncRNA genes, exons were extracted from GTF files containing the 1,119 lncRNAs identified by Young et al. (2012) and the 852 lncRNAs identified by Graveley et al. (2011). Cuffcompare was run to make a non-redundant list. Constitutively expressed regions were defined as previously described. These were then intersected with the exons of the protein-coding and pseudogenes to remove any that overlap with updated annotations of genes. The two BED files were then combined to give a full list of all constitutively expressed regions of the *D. melanogaster* genome.

2.2.3 Mapping sequencing reads, transcript assembly and lncRNA discovery

Sequencing data were generated by the Wellcome Trust Centre for Human Genetics (WTCHG) and mapped to the *D. melanogaster* genome. Unmapped reads were extracted and remapped using TopHat.2.0 and Bowtie.1.0 (Trapnell et al., 2009) using the GTF file containing all protein-coding, pseudogenes and lncRNAs and added to previously mapped reads. Transcripts were assembled using Cufflinks v2.0.2 (Trapnell et al., 2010), with the three biological replicates pooled for each sample. A transcript required a minimum of six pooled reads to be assembled. Cuffcompare was run using the resultant GTF files to get a consensus list of transcripts. This GTF file was converted into a BED format and overlapping transcripts were merged using mergeBed. Options -s -nms were used to prevent transcripts on opposite strands being merged and to retain the

information about the identity of the merged transcripts. To distinguish between protein-coding and pseudogenes and putative lncRNAs, intersectBed was run using the protein-coding and pseudogenes from BDGP5.68 in a BED format. Options -u -s gave a list of regions overlapping with known protein-coding genes or pseudogenes and option -v gave intergenic regions of transcription.

The coding potential of all intergenic transcribed regions was determined using the Coding Potential Calculator (Kong et al., 2007). CPC uses 6 parameters that assess the presence, coverage and quality of the Open Reading Frames (ORFs) predicted by *framefinder* and the number, quality and distribution of hits of the output of BLASTX against UniProt Reference Clusters. A gene was classified as a lncRNA if greater than 200bp in length, multi-exonic and if all transcripts were scored as non-coding by CPC.

2.2.4 Calling differential gene expression

Read counts for genes were compiled by summing the counts of all reads that mapped within constitutively expressed regions of that gene. Read coverage across genes was calculated and only genes that had 80% coverage in all three replicates of one or more biological condition were included. These absolute count numbers were used to carry out differential gene expression analysis using edgeR (Robinson et al., 2010) in Chapter 4 or DESeq2 (Love et al., 2014) in Chapters 5 and 6.

2.2.5 Genome wide predictions of microRNA recognition elements

MicroRNA recognition elements (MREs) were predicted using the TargetScan6.0 (Lewis et al., 2005) or miRanda (Enright et al., 2003) algorithms.

Chapter 3 - Characterisation of six loci containing predicted long non-coding RNA transcripts

Abstract

The large catalogue of putative lncRNAs in *Drosophila melanogaster* provides the opportunity to study the function of lncRNAs in the context of the whole organism. This chapter describes the initial screening of six putatively non-coding loci and the corresponding mutant strains containing transposon insertions within these regions. By screening available transposon insertions I identified at least one mutation per loci that perturbed the expression level of the lncRNAs. Among the six lncRNAs, three are more highly expressed, while the other three have reduced expression in the mutant strains. Potential regulatory interactions between the lncRNA and its neighbouring gene were detected in three out of six loci. Both 5' and 3' ends of the transcripts were found in three of the six loci and all transcript models were called as non-coding by Coding Potential Calculator (CPC). Only one locus, *lnc703*, met all the criteria outlined and was therefore selected for further study.

3.1 Introduction

3.1.1 *In vivo* functions of lncRNAs are largely unknown

The functions of very few lncRNAs are known, despite having being identified in many organisms including humans, mice, zebrafish and flies. In the cases where functional characterisation has been performed it has been largely carried out using mammalian cell culture systems. The first mouse models in which lncRNAs have been deleted or disrupted show little or no overt phenotype (Nakagawa et al., 2011; Zhang et al., 2012). At present the relationship between the phenotypes seen in cell culture systems and the role of lncRNAs in their physiological setting is not yet clear. The large set of putative lncRNAs identified in *Drosophila melanogaster* (Young et al., 2012) provides the opportunity to study the effects of the loss of putative lncRNAs in the context of a whole organism.

3.1.2 Methods of generating lncRNA mutants

Traditional methods of mutagenesis in *D. melanogaster* include EMS induced point mutations, transposon insertion mediated gene disruptions or genomic deletions by homologous recombination or excision of transposon insertions. Transposon insertions account for the largest proportion of mutant strains available from stock centres. At the time of writing, more than 14,000 transposon insertion strains had been generated, covering greater than 9,400 genes and more than 1,700 intergenic regions (Bellen et al., 2011), which could be expected to include putative lncRNAs.

3.1.3 Attributes of the ideal lncRNA for functional characterisation

The 1,119 putative lncRNAs described by Young et al. (2012) were identified

as regions of transcription outside of known protein-coding genes, which were defined from the transcriptomes of 30 *Drosophila melanogaster* developmental stages (Graveley et al., 2011). Initial characterisation of these putative lncRNAs indicated enrichments for expression in the adult head or in males, and that they are enriched near to transcription factors (Young et al., 2012). The presence of a transcript in the locus was confirmed by qRT-PCR for a subset of the putative lncRNAs and recapitulated the expression patterns expected based on RNA-seq data (Young et al., 2012). However, these predictions represent incomplete gene models and therefore only act as a guide for regions of non-coding transcription and further information is required prior to in depth study of these regions. For a candidate lncRNA to be suitable for detailed functional characterisation it will be necessary to know the following information:

1. Can it reliably be detected by standard methods i.e. qRT-PCR?
2. Does it represent a characteristic of lncRNAs, e.g. head or male expressed, or proximal to transcription factor?
3. Do available fly strains perturb expression of the lncRNA?
4. What is the full gene model, including 5' and 3' ends, introns and alternative splice isoforms?
5. Do all transcripts within the full gene models have no coding potential as determined by computational prediction?

3.1.4 Aims

I aimed to identify a number of suitable candidate lncRNAs with available mutants that could be investigated in greater detail. This chapter will describe the characterisation of a small subset of putatively non-coding transcripts identified

in Young et al. (2012) that were initially selected was based on the availability of fly strains with transposon insertions within the region of non-coding transcription. Through expression profiling to confirm transcription, defining complete transcript models to confirm the structure of the transcripts in that region and confirming the ability of the transposon insertion to specifically affect these transcripts.

3.2 Screening methodology

3.2.1 Initial selection

A screening process was designed to address each of the criteria detailed above. We collated a list of all transposon insertions that lie within putative lncRNAs and that have a corresponding fly stock available from the Bloomington or Kyoto stock centres. Candidate lncRNAs that had detectable expression in qPCR expression profiling experiments and represented common features identified by Young et al. (2012), including those expressed in males or heads, or those proximal to transcription factors were selected for further study.

3.2.2 Confirmation of presence of non-coding transcript

Expression profiles were determined by qRT-PCR of RNA extracted from *yw* flies at different developmental stages (in collaboration with Dr Charlotte Tibbit). To determine the full-length gene model of the transcripts within these loci, Rapid Amplification of cDNA Ends (RACE) was performed to determine the transcriptional start sites and transcriptional termination sites. The predicted gene models were generated from poly-A selected sequencing libraries (Graveley et al., 2011) and are therefore expected to be specifically RNAPII transcribed transcripts. Therefore, the techniques used were specific for m⁷G capped and polyadenylated transcripts. Putative introns were identified using the exon-exon junction track on flybase GBrowse (Daines et al., 2011), or by manual inspection for gaps in the read tracks from the modENCODE data. PCR primers were designed to the flanking regions; amplified regions were selected on size and Sanger sequenced to confirm exon-exon junctions. Where no 5' RACE results were obtained and the lncRNA was transcribed on the same strand as the neighbouring gene, primers were designed to amplify any continuous transcripts. Where full gene models

were determined, CPC was run to test the coding potential of all transcripts of a gene. CPC assigns scores that reflect coding potential, scores greater than 1 are protein-coding, scores between 0 and 1 are weakly coding, scores between 0 and -1 are weakly non-coding and scores less than -1 are non-coding. All transcripts with scores less than 0 were considered non-coding.

3.2.3 Analysis of mutants

Mutant strains were assessed for lethality by counting of homozygous adults, where the number of homozygous flies was reduced or absent mutant lines were crossed to deficiency strains to map the cause of lethality. To remove any background mutations in the transposon insertion mutant lines, strains were isogenised by nine generations of backcrossing to wildtype (w^{1118}), flies in duplicate. In parallel, mutant strains were crossed to flies expressing the appropriate transposase to achieve excision of the transposon. *Piggybac* transposons are precisely excised (Hacker et al., 2003) and therefore excision should cause reversion of any phenotypes or changes in gene expression. *P-element* excisions are also usually precise, but a low frequency of excisions are imprecise, resulting in deletions at the site of insertion (Hummel and Klambt, 2008) so may not result in reversion of phenotypes but excision could generate novel mutant alleles.

The effect of the insertion on the expression level of the lncRNA and the neighbouring protein-coding genes was assessed using qRT-PCR comparing isogenised mutant strains to w^{1118} controls. To identify lncRNAs that might act in *cis* on their neighbouring protein-coding gene, the expression profiles of genes next to the lncRNAs were determined using the same developmental profile samples used for expression profiling of lncRNAs. Pearson's correlation coefficients were used to calculate the linear dependence of the expression levels of the

two genes. Data for S2 cells were excluded from this analysis as they do not represent a distinct development stage. Positive correlation coefficients (r) show a positive relationship between the two variables with +1 being complete positive correlation; conversely negative coefficients show negative correlation with -1 being total negative correlation. A coefficient of 0 indicates no correlation.

3.3 Results

3.3.1 Selection of six putative long non-coding RNA loci

A total of 234 transposon insertion strains that represented 115 of the 1,119 putatively non-coding loci identified in Young et al. (2012) were available from either the Bloomington or Kyoto stock centres. Of these, six regions were selected for screening because transcription in the region had previously been confirmed by qRT-PCR (Young et al., 2012) and the locus represented a common feature identified in Young et al. (2012). Genomic coordinates and associated fly stock numbers are listed in Table 3.1.

Table 3.1 Genomic coordinates of six loci containing putative long non-coding RNAs that contain at least one transposon insertion.

	Genomic coordinates	Stock number	Attribute	Transposon specific primer	Gene specific primer
<i>Inc66</i>	2L:7318462..7320229	f03265 (Pbac) 207305 (P)		5R 5pIn/5pOut	<i>Inc66</i> insR <i>Inc66</i> F2/R2
<i>Inc367</i>	2R:16724969..16726852	205574 (P)	Male specific	5pIn/5pOut	<i>Inc367</i> insR
<i>Inc458</i>	3L:6775314..6785313	f05223 (PBac)	Next to TF	3F/3R	<i>Inc458</i> insR
<i>Inc626</i>	3L:19547917..19552119	16556 (P)	Head expressed	Plac1/Pwht1	<i>Inc626</i> insR
<i>Inc703</i>	3R:2917681..2919527	19598 (PBac gal4) 141392 (PBac stop)	Next to TF	PLF/PLR 5F0/5R2	<i>Inc703</i> insR
<i>Inc921</i>	4:475388-479955	f04237 (PBac)	Next to TF Head specific	5F/5R	<i>Inc921</i> insR

TF = Transcription Factor.

3.3.2 Initial screening of transposon insertion strains

Many of the transposon insertion strains available from stock centres have not been well characterised further than the mapping of the site of insertion. It is therefore necessary to use a number of methods to ensure the specificity of these lines. All strains were isogenised through nine generations of crossing transposon containing females to *w¹¹¹⁸* males.

3.3.2.1 Confirming insertion sites

The genomic location of the transposon insertion was confirmed by PCR in which one primer is specific for the transposon and the other specific for the flanking region. All transposons could be detected by this method before (data not shown) and after nine generations of isogenisation (Appendix C).

3.3.2.2 Phenotypes before and after isogenisation

The transposon insertion strains were screened for overt phenotypes before and after isogenisation. The *Inc921^{PBac}* strains had no overt phenotype. The *Inc66^{PBac}* strain but not *Inc66^P* strain is homozygous lethal during larval development both before and after isogenisation. The *Inc367^P* and *Inc458^{PBac}* are also homozygous lethal before and after isogenisation in early larval and late embryonic stages, respectively. The *Inc626^P* strain was initially found to be homozygous lethal during late pupal development with homozygotes failing to eclose. After isogenisation, the *Inc626^P* strain is homozygous viable with only a small proportion of homozygotes failing to eclose. Before isogenisation, the *Inc703^{gal4}* strain was partially homozygous lethal with a large proportion of flies failing to eclose and *Inc703^{stop}* was fully lethal with all homozygotes failing to progress past day two after puparium formation. After isogenisation the *Inc703^{gal4}* was fully homozygous viable however homozygotes of the *Inc703^{stop}* strain still failed to progress past day two after puparium formation.

3.3.2.3 Excision of transposon insertions

Excision of the transposon by crossing mutants to flies containing the appropriate transposase was attempted for all strains. Successful excision, as judged by the loss of the visible marker (red eyes in most cases and EYFP or dsRED expression in *Inc703* mutants) was only achieved for *Inc458*. Despite the insertion within *Inc458* being a *Piggybac* element that should always excise precisely, excision flies were still homozygous lethal during embryonic development. Sanger sequencing of four independent excision lines identified sequences that map to the backbone of the *Drosophila* transposon mutagenesis vector pMiET (Metaxakis et al., 2005) and transposable P vector conferring G418 resistance in *Drosophila* (Steller and Pirrotta, 1985) suggesting the presence of residual transposon fragments.

3.3.2.4 Deficiency crosses

The specificity of a mutant can also be determined by generating hemizygotes, which carry one copy of the mutant and lack the corresponding genomic region. Hemizygotes were generated for the six homozygous lethal mutant strains prior to isogenisation, using two different deficiency lines where possible. No hemizygous flies were viable for *Inc66*, *Inc367*, *Inc458*, *Inc626* or *Inc703* strains (Figure 3.1), showing that the lethality in these strains maps to the genomic region surrounding the insertion.

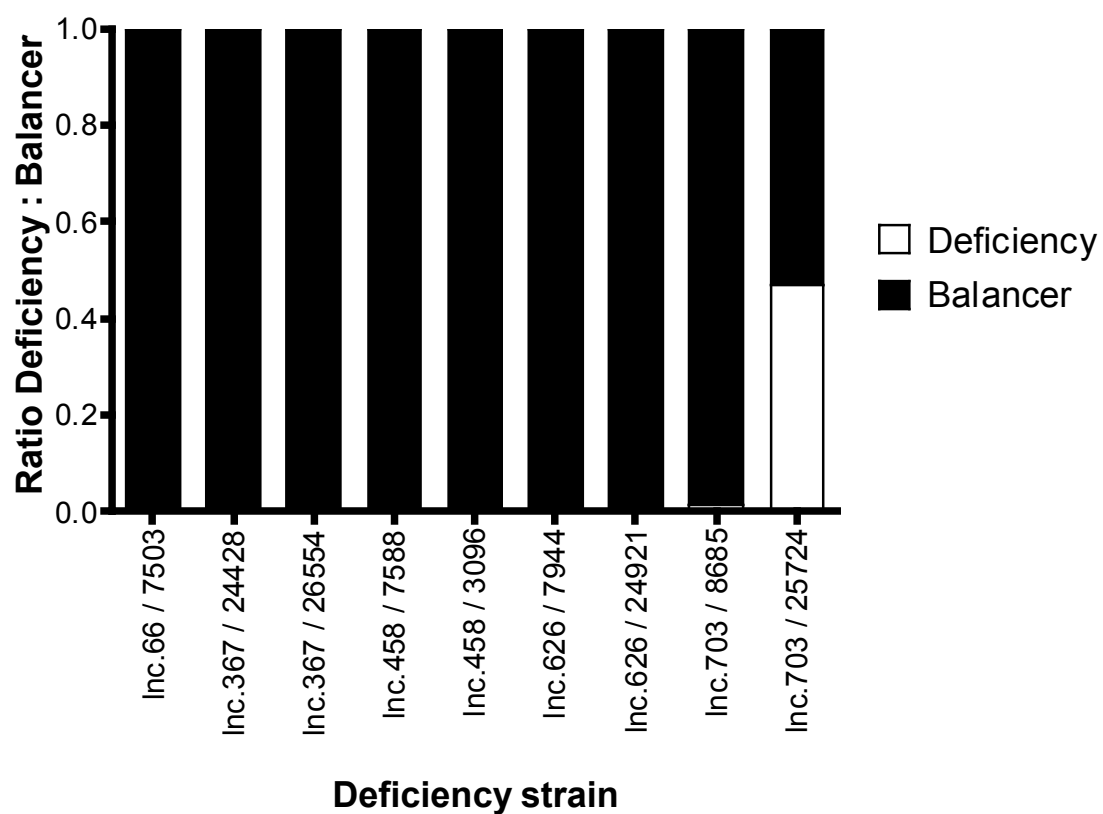


Figure 3.1 Adult survival ratio. Transposon mutant strains with observed lethal phenotypes were crossed to corresponding deficiency lines. No hemizygous flies were observed for any of the strains. N = 4 – 84.

3.3.3 Transcripts could be detected in all six loci at one or more developmental stages

In general, lncRNAs are more lowly expressed and have more tissue and temporally restricted expression patterns than protein-coding genes. qRT-PCR was performed on samples from different developmental stages to identify the tissues in which each lncRNAs was most highly expressed. Consistent with being selected as male-specific, *lnc367* showed highest expression in adult males and was also detectable in all larval and pupal stages (Figure 3.2). qRT-PCR analysis further confirmed that both *lnc626* and *lnc921* were highly expressed in the head (Figure 3.2). Of the lncRNAs selected due to their proximity to a transcription factor, *lnc458* was highly expressed in the adult head compared to other developmental stages assayed, whereas *lnc703* was expressed broadly throughout development, with lowest expression in first instar larvae and *lnc66* had the lowest levels of expression the lncRNAs tested with highest expression in 4 - 8 hour embryos (Figure 3.2).

3.3.4 Full transcript models

The 1,119 putative lncRNAs used as the initial dataset for this work were not annotated in the genome assembly (BDGP5.38) when this work started in 2011. It was therefore necessary to experimentally determine the transcript models to confirm the transcripts are all distinct from their neighbouring genes and get full length transcripts to assess their coding potential. 5' and 3' RACE were carried out on all six loci. In 2012, BDGP5.43 was released which included many more annotated lncRNAs based in EST data. In 2014, BDGP6.0 was released which contained a greater number of lncRNAs collated from various sources and includes many of the lncRNAs identified by Young et al. (2012).

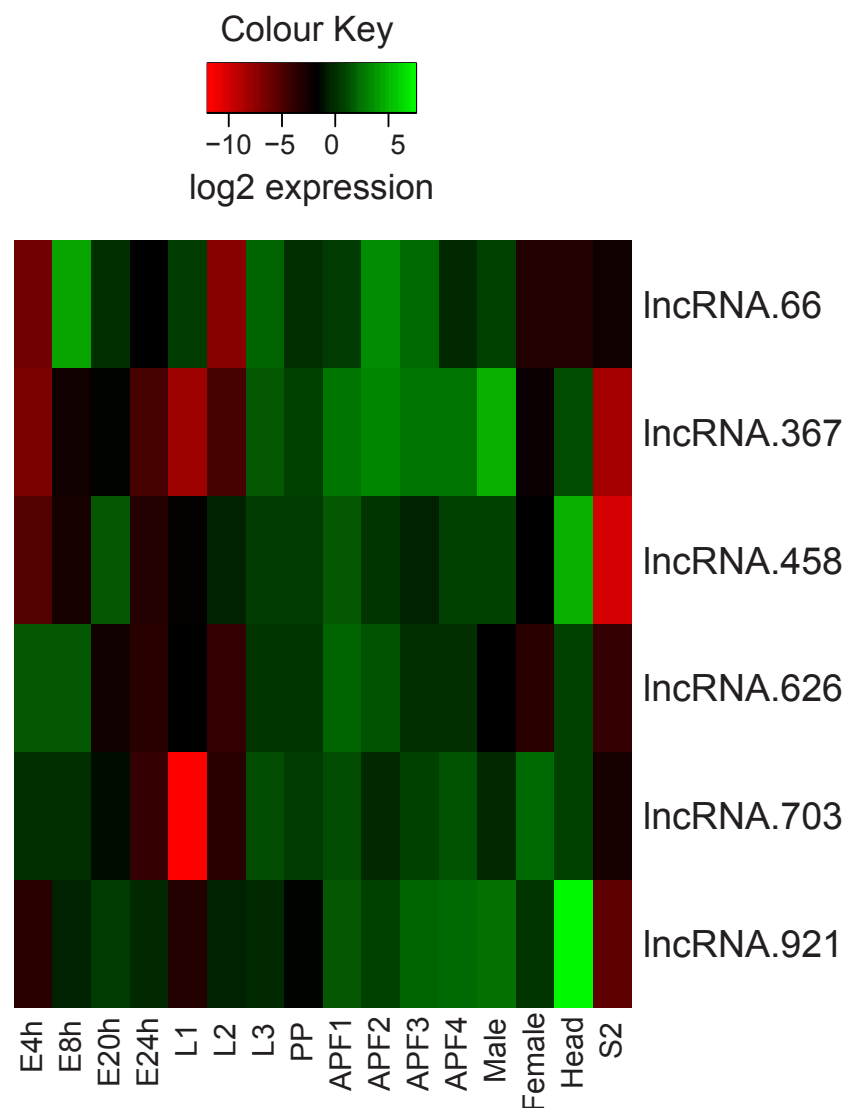


Figure 3.2 Developmental expression profiles of six lncRNAs in yw flies determined by qRT-PCR. Expression values are normalised to the median expression value for each lncRNA.

Full transcript models could be assembled for three of six loci tested and revised transcript annotations for each gene are shown in Figure 3.3. In BDGP5.38 no transcripts were annotated in the *Inc703* locus, although EST data indicated the presence of a transcript. A non-coding transcript, CR43685 (Figure 3.3a) was annotated in BDGP5.43. 5' RACE revealed that the gene's transcriptional start site (TSS) is 28bp closer to downstream gene *e(y)2b* than annotated. Both annotated introns were confirmed by PCR and Sanger sequencing and two isoforms, differing in the size of the first intron were identified (Figure 3.3a and Appendix D). The location of the primers used for 5' RACE, downstream of the first exon, confirmed that both isoforms share the same TSS.

No transcripts had previously been annotated in the region containing *Inc921*, however stranded RNA-seq from adult heads (Graveley et al., 2011) suggested the transcript is transcribed from the positive strand and a 3' end was identified by RLM-RACE 493bp upstream of the 3' of neighbouring *bip2* (Figure 3.3b). No 5' results were obtained using the RLM-RACE technique, so a circular RACE method was used. This identified a TSS, 564bp downstream of the 3' end of neighbouring *CAMKI*, and an alternative 3' end, 346bp upstream of *bip2* (Figure 3.3b). Visual inspection of the modENCODE RNA-seq tracks on flybase GBrowser (Graveley et al., 2011) indicated an intron in *Inc921*, which was confirmed by PCR and Sanger sequencing (Sequence in Appendix D).

No transcripts had previously been annotated in the region containing *Inc626* in release 5.38 of the *D. melanogaster* genome assembly (BDGP5.38) or BDGP.43. 5' RACE identified a single TSS 88bp downstream of the 3' end of neighbouring CG9300. Two alternative 3' ends were identified 2416bp and 4173bp downstream of the TSS respectively (Figure 3.3c). Exon junctions predicted in Flybase (Daines et al., 2011) could not be validated by PCR (data not shown). In more recent annotations of the genome (BDGP6.0), a non-coding transcript, CR44677, has

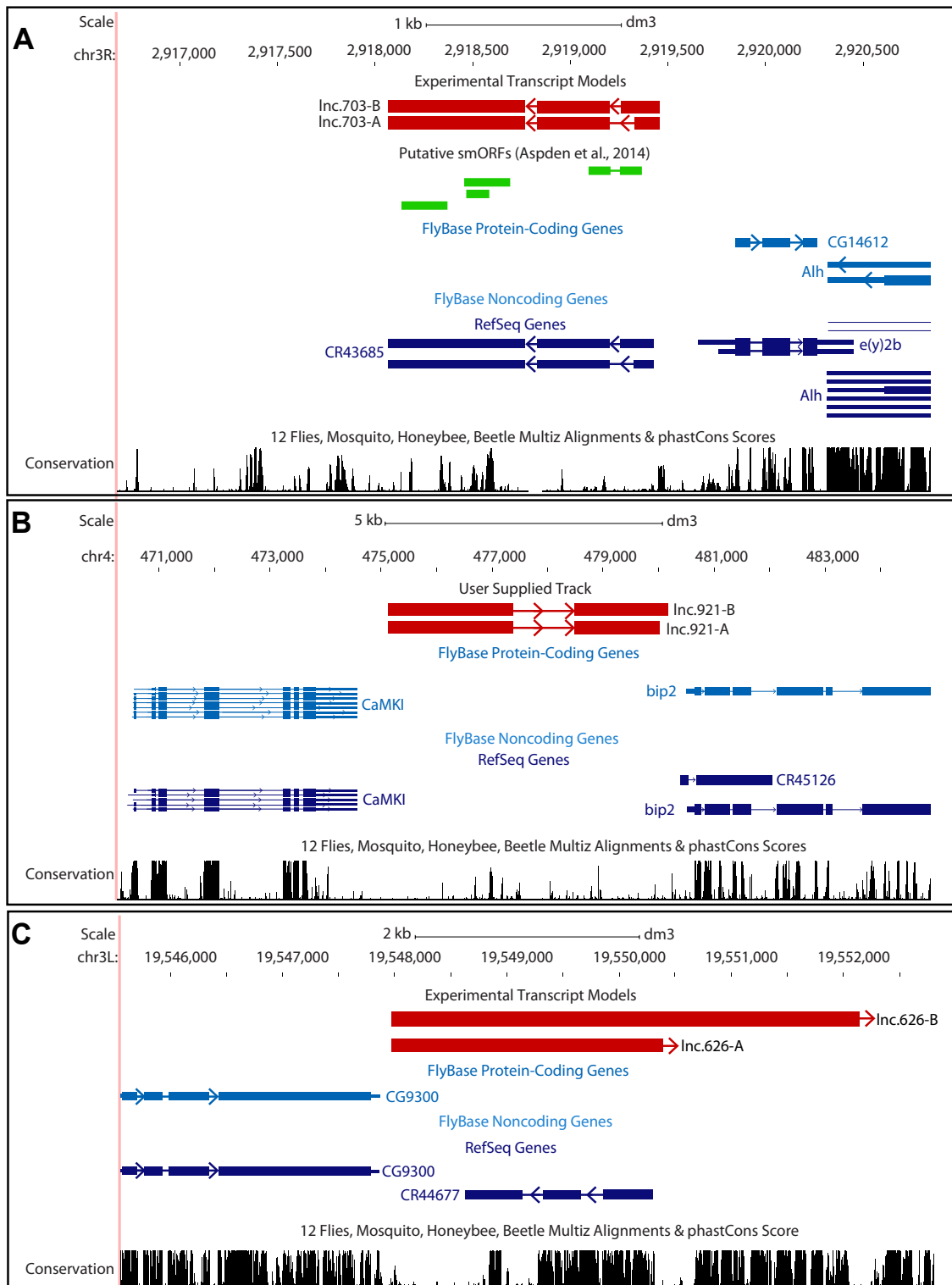


Figure 3.3 Experimentally determined transcript models. lncRNAs are shown in red, surrounding genes are shown in light and dark blue. Direction of transcription is indicated by arrows. (A) Two isoforms of *Inc703* were previously annotated in the locus (shown in dark blue), the 5' end of both isoforms was found to be 28bp upstream. In the *Inc921* locus (B), a single TSS, two 3' ends and a single intron were found downstream of *CaMKI* (shown in blue), another non-coding transcript, *CR45126* (shown in blue) has

been annotated downstream of *Inc921* overlapping the *bip2* 3' end (shown in blue) but not *Inc921*. The *Inc626* locus (C) two alternative 3' ends and a single TSS were detected. Neighbouring protein-coding gene, *CG9300* and newly annotated non-coding transcript, *CR44677* are shown in blue.

been annotated and is transcribed from the opposite strand to *Inc626* (Figure 3.3c).

3.3.3 Incomplete transcript models

Young et al. (2012) identified three putative transcripts (*Inc66*, *Inc67* and *Inc68*) in the region between *wg* and *Wnt6*. No further information about the strandedness or connectedness of these three transcripts could be determined using more recently available stranded RNA-seq data (Graveley et al., 2011). I identified two 3' ends by 3' RACE, which indicated two overlapping transcripts on opposite strands (Figure 3.4). No 5' ends could be identified using any tested combination of primers spanning a 12.9kb region between *wg* and *Wnt6*. Areas of transcription detected by tiling PCR across this region are shown in Figure 3.4. It was not possible to determine the relationship between the areas of transcription and the two 3' ends detected. For the rest of the chapter I will distinguish between the two putative transcripts in this locus using identifiers the *Inc66a* and *Inc66b*. No 5' or 3' ends were detected for *Inc458*, which is predicted to lie upstream of transcription factor *vvl* or *Inc367*, which lies downstream of the protein coding genes *insc* and *skt1* (data not shown).

3.3.4 Coding potential

The CPC algorithm was rerun on the full length transcripts for *Inc626*, *Inc703* and *Inc921*. The algorithm had previously been run on the annotated *Inc703* transcripts (CR43685) and the largest ORF detected for both splice isoforms was 63 amino acids (aa), accounting for approximately 15% of the transcript. No homology to known protein domains was found for this peptide and both transcripts were determined to be non-coding (Score = -1.15). No new ORFs were found for the

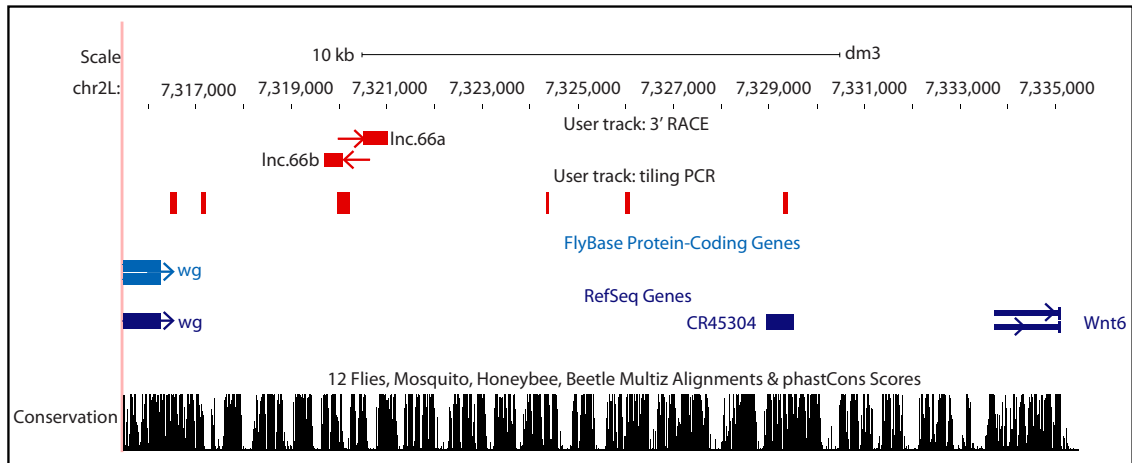


Figure 3.4 Experimentally determined regions of transcription in the *Inc66* locus. Two 3' ends were found on opposite strands (shown in red, labelled *Inc66a* and *Inc66b*). Regions of transcription between *wg* and *Wnt6* (shown in blue) identified by tiling PCR are shown in red. In the most recent genome assembly (BDGP.6.01) a non-coding RNA *CR45304* (shown in dark blue) is annotated, overlapping the most downstream region transcription detected.

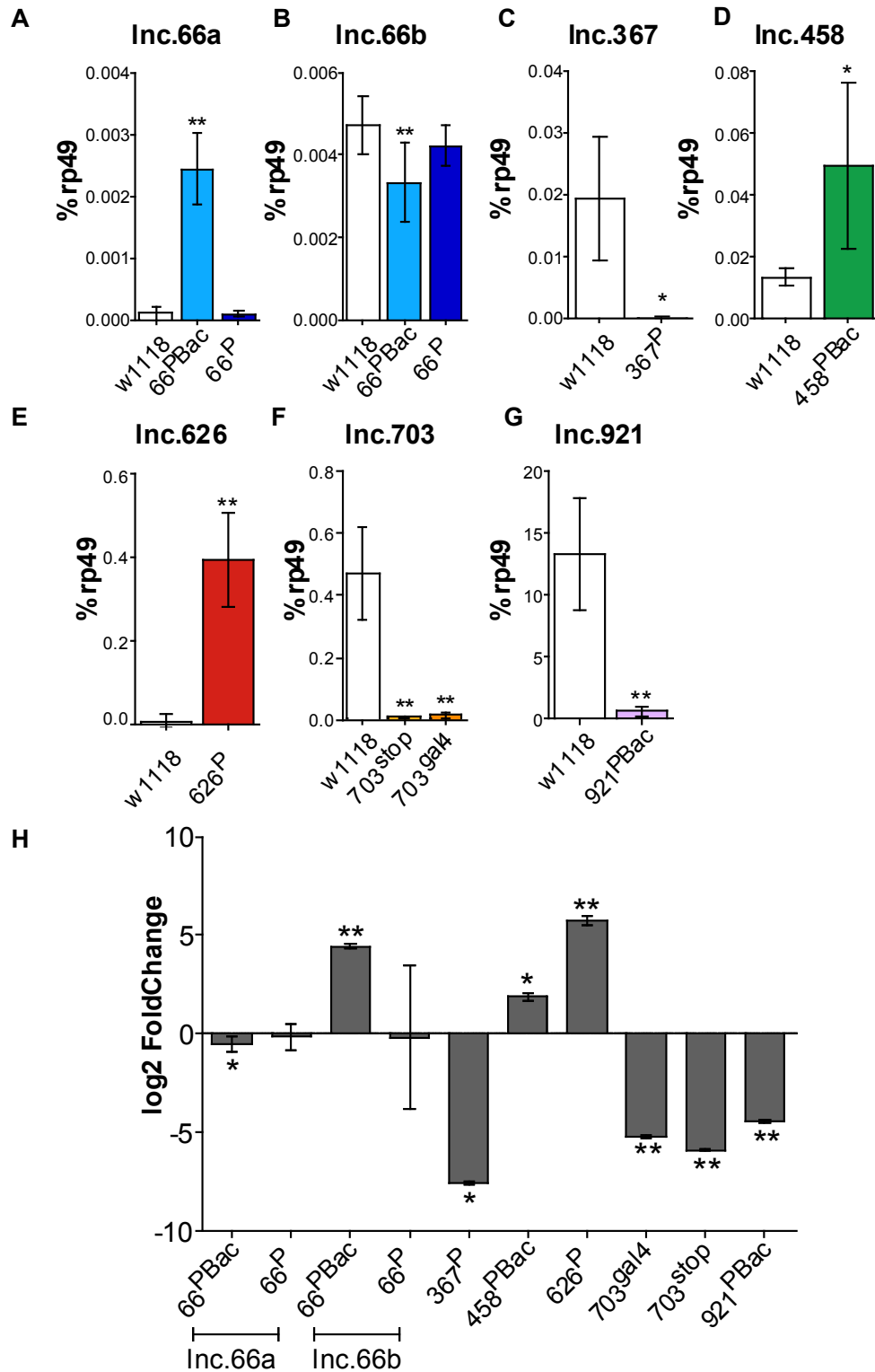


Figure 3.5 Transposon insertions affect the expression levels of lncRNAs in the tissues in which they are most highly expressed. (a-g) level of lncRNAs reported as a percentage of *rp49* expression. (h) Data is represented as log₂ fold change expression level compared to *w*¹¹¹⁸. Significance compared to *w*¹¹¹⁸ levels is indicated by * $p < 0.05$ or ** $p < 0.01$, two-tailed Student t-test.

extended *Inc703* transcripts defined by RACE and both isoforms were classified as non-coding (Score = -1.15). Furthermore, the start codon of the longest ORF in *Inc703* is not conserved in other *Drosophila* species. Both the long and short experimentally predicted isoforms of *Inc626* were called as non-coding (Scores = -1.19 and -1.39, respectively). Both *Inc921-A* and *Inc921-B* were also determined to be non-coding (Scores = -0.37) with the longest ORF detected in both being 20aa with no homology to any known protein domains.

3.3.5 Effect of transposon insertions on lncRNA expression

The effect of a transposon insertion within a region depends on the modular elements within the transposon, the genomic context, and the orientation and location of the transposon relative to the gene. qRT-PCR analysis was used to compare the expression level of the lncRNAs in wildtype and isogenised mutant strains in the tissue with highest expression or at the stage prior to lethality for homozygous lethal strains. A significant change in expression level was detected in eight out of nine strains (Figure 3.5h). Consistent with the expected variability of transposon insertions, strain *Inc66^{PBac}* has opposing effects on the two putative transcripts *Inc66a* and *Inc66b*. *Inc66a* is more highly expressed (Figure 3.5a, 3.5h, log₂ fold change = 4.4, $p < 0.01$), whereas *Inc66b* is lower expressed (Figure 3.5b, 3.5h, log₂ fold change = -0.5, $p < 0.05$). Despite containing similar elements (UAS sites for GAL4 driven misexpression) and being close in location to the *Inc66^{PBac}* insertion site, the *Inc66^P* strain has no effect on either transcript (Figure 3.5a and 3.5b).

Inc367 (Figure 3.5c), *Inc703* (Figure 3.5f) and *Inc921* (Figure 3.5g) are significantly reduced in expression in the corresponding mutant strains compared to wildtype (log₂ fold change = -7.5, -5.2, -5.9 and -4.4; $p < 0.05$, 0.01, 0.01 and 0.01,

respectively). Conversely, *Inc458* (Figure 3.5d) and *Inc626* (Figure 3.5g) are more highly expressed in mutant strains ($\log_2$ fold change = 1.9 and 5.7, respectively, $p < 0.05$).

3.3.6 Identifying putative *cis*-acting lncRNAs

3.3.6.1 Correlated expression

To investigate if there was a relationship between the expression of each lncRNA and its nearest neighbour(s), expression profiles across development were determined by qRT-PCR for each of the neighbouring protein-coding genes. These expression profiles were compared to the profiles I had previously determined by qRT-PCR for the lncRNAs (Figure 3.2) using Pearson's correlation. Table 3.2 shows the correlation coefficients for pairwise comparisons between the six lncRNAs and their nearest neighbours. For comparison, *Inc927* and its neighbour *toy* are included because *Inc927* is an extended 3'UTR of *toy* (Personal communication, A. Bassett) and as expected a strong positive correlation is observed ($r = 0.907$, $df = 13$, $p = 3.06 \times 10^{-06}$). A significant positive relationship was detected between *CG9300* and *Inc626* ($r = 0.540$, $df = 13$, $p = 0.029$) indicating the two genes are expressed in a similar developmental profile. No other significant correlations were observed.

Whilst correlation of expression patterns could indicate an interaction between a lncRNA and its protein-coding neighbour, the method described above does not inform about the dependency of the expression of the lncRNA and its neighbour. Moreover, correlated expression will not detect interactions that occur specifically in a single tissue if the lncRNA is expressed in a single tissue and its neighbour is more broadly expressed. lncRNAs have been shown to have more tissue restricted expression profiles than protein-coding genes, therefore it is possible

that a lncRNA will have an effect on a neighbouring gene in a tissue specific manner even if the neighbour is more broadly expressed.

Table 3.2. Pearson's correlation of developmental expression profiles of lncRNAs and their nearest protein-coding genes.

lncRNA	Protein-coding neighbour	Pearson's correlation (<i>r</i>)	Df	<i>p</i>-value
<i>Inc367</i>	<i>Insc</i>	0.261	13	0.348
	<i>Skittles</i>	-0.393	13	0.147
<i>Inc458</i>	<i>Vvl</i>	-0.044	13	0.874
<i>Inc626</i>	<i>CG93000</i>	0.562	13	0.029
<i>Inc703</i>	<i>α-Tub84B</i>	-0.172	13	0.540
	<i>e(y)2b</i>	0.248	13	0.374
<i>Inc921</i>	<i>bip2</i>	-0.089	13	0.751
	<i>CaMKI</i>	0.195	13	0.487
<i>Inc927</i>	<i>Toy</i>	0.907	13	3.058 x 10⁻⁰⁶

3.3.6.2 Affect in mutant lines

To investigate the potential for a tissue specific effect of cis-acting lncRNAs and the dependency of expression levels, the expression level of the nearest protein-coding neighbour(s) was compared in wild-type and isogenised transposon insertion strains in the tissue in which highest expression of the lncRNA was observed (Figure 3.6m). In two loci a significant change in the neighbour was observed in the same direction as the change in the lncRNA. A significant decrease was observed in both *Inc367* and its neighbour *insc* (Figure 3.5c and Figure 3.6c, respectively) and a significant increase was observed in *Inc626* and its neighbour *CG9300* (Figure 3.5e and Figure 3.6f, respectively). Due to the positive correlation in expression profiles, coordinated change in expression in the mutant line and the same direction of transcription of the two regions, it is

possible that *Inc626* is an extended 3'UTR of *CG9300*. *Inc703* neighbour *e(y)2b* was significantly more highly expressed in third instar larvae of *Inc703*^{gal4} but not *Inc703*^{stop} strains compared with wild-type (Figure 3.6g and 3.6i). The cause of the difference between these two strains is unclear as they have been isogenised, have similar residual levels of *Inc703* expression (Figure 3.5f) and the insertion sites are only 1bp different. It is likely that this results from the specific sequences of the integrated transposon, which differ between the two mutant lines. The relationship between expression levels of *Inc703* and *e(y)2b* is investigated further in Chapter 4.

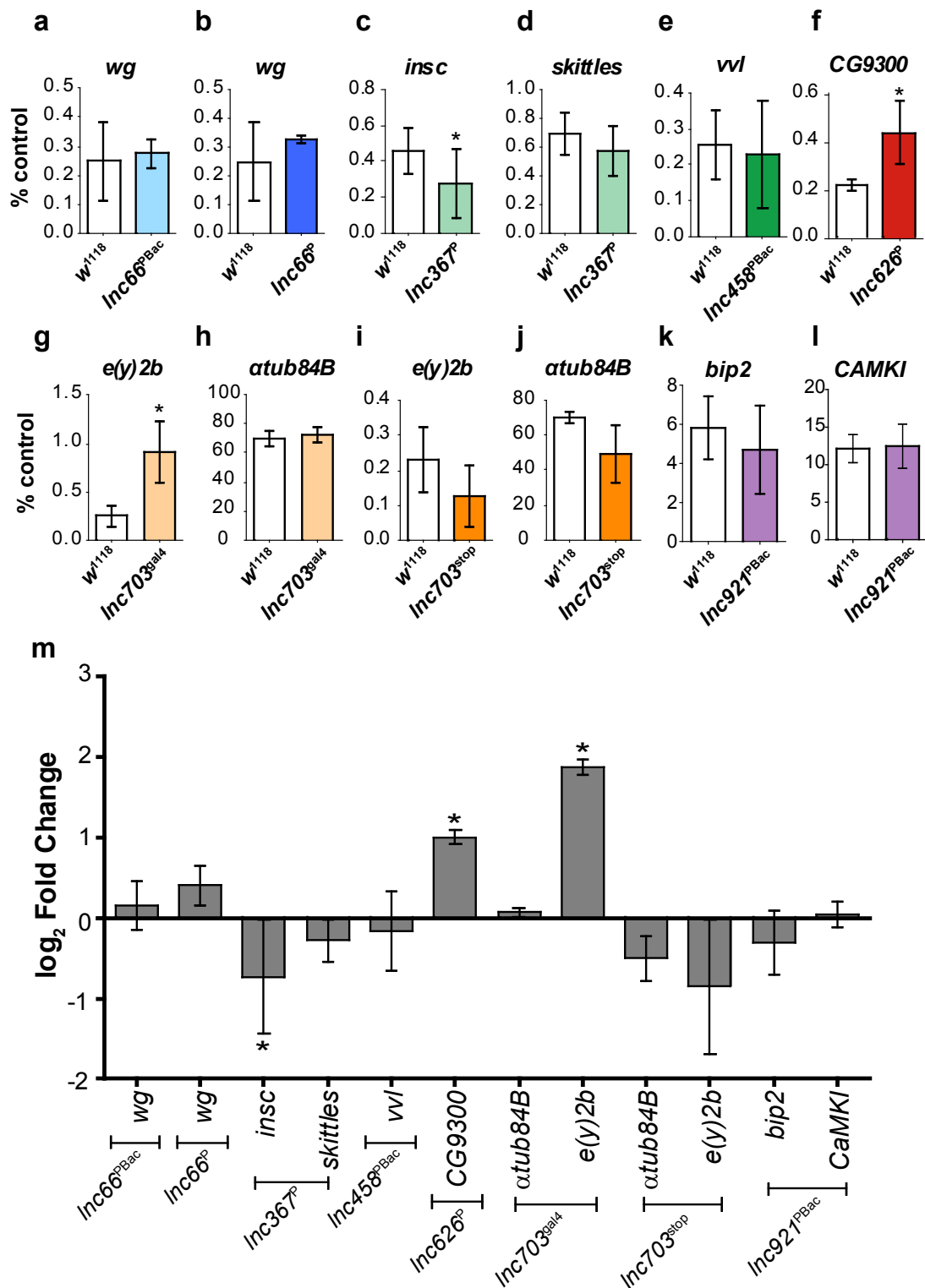


Figure 3.6 Changes in mRNA levels of neighbouring protein-coding genes can be detected in isogenised transposon insertion strains. (a-l) Level of neighbouring gene reported as a percentage of *rp49* expression. (m) Levels are reported as $\log_2$ fold change compared to *w*¹¹¹⁸ in the tissue in which the lncRNA is expressed. Significance compared to *w*¹¹¹⁸ levels is indicated by * $p < 0.05$, two-tailed Student t-test.

3.4 Discussion

3.4.1 Methods to identify potentially functional lncRNAs

It is not yet clear what proportion of the large number of lncRNAs that have been identified from transcriptome profiling experiments is functional or how best to identify lncRNAs most likely to be functional. This study used the availability of mutants to perturb the expression of the lncRNA *in vivo* as the initial selection criteria. Other studies have used different criteria to identify putatively function lncRNAs including conservation, synteny or functional class.

Through intensive *in silico* analysis, 29 lncRNAs were found in zebrafish that shared sequence homology with human lncRNAs (Ulitsky et al., 2011). However, these homologous regions only represent very small regions of the whole transcript making them difficult to identify. Remarkably, in the case of *cyrano* the loss of function phenotype of the zebrafish could be rescued using the cDNA of the mouse or human homologous lncRNA (Ulitsky et al., 2011) suggesting conservation may be a good indicator of function. In this case, the conserved function is due to a miRNA target site present in a small segment of the lncRNA (Ulitsky et al., 2011). However, to date no lncRNAs have been found to share sequence similarity or functionality between mammals and *Drosophila*. Although, lncRNAs in *Drosophilids* are more highly conserved than lncRNAs in mammals (Haerty and Ponting, 2013). Even in the case of dosage compensation of the sex chromosomes between males and females, which involves lncRNAs in both flies and mammals, the mechanisms are not the same. In mammals, the lncRNA *Xist* is transcribed from, coats and inactivates one copy of the X chromosome in females. However, in flies, the lncRNAs *roX1* and *roX2* are transcribed from and upregulate genes on the X chromosome exclusively in males.

Using the less stringent measure of synteny between species, it is possible to find

113 analogous lncRNAs between zebrafish and mouse or human. Synteny can also be used to identify analogous lncRNAs between fly and mouse (Young et al., 2012), and would only be expected to be conserved across long evolutionary distances if the lncRNA act in *cis* to regulate neighbouring genes. Of the six loci characterised in this chapter, synteny could be found for only one, *lnc458*, located upstream of *vvf*. The mouse orthologue of *vvf*, *Pou3f3* has a putative lncRNA upstream of its promoter.

Others have tried to characterise different functional classes of lncRNAs, for example by the chromatin marks in cells or tissues that express the lncRNA. One class that has been identified by this method are promoter lncRNAs (plncRNAs) that are marked by the presence of H3K4me3 at the start of the gene and H3K36me3 in the body of the gene. Enhancer RNAs (eRNAs or elncRNAs) are transcribed from active enhancers, which are marked by H3K4me1. It is not yet clear if transcription of elncRNAs is a consequence of the recruitment of activating complexes to the region or a functional requirement of enhancer activity, or whether plncRNAs or elncRNAs are more likely to be functional. However, elncRNAs are more likely to be correlated in expression with their neighbouring gene than plncRNAs during mouse erythropoiesis (Marques et al., 2013).

A combination of these methods was used in a recent paper that describes the deletion of 18 lncRNA loci in mice. Loci were preferentially selected if a lncRNA was annotated in the syntenic region of the human genome and a number of loci with chromatin marks indicative of elncRNAs were chosen (Sauvageau et al., 2013). Of these, three mutants had reduced viability, with 50 – 100% of pups failing to reach weaning age. The authors claim this proves that some lncRNAs are required for viability, suggesting conservation may be a good indicator of functionality. However, in the absence of rescue experiments it is not possible to confirm if it is the loss of the lncRNAs or loss of underlying DNA sequences in the

deleted regions that confer lethality.

3.4.2 Some putative lncRNAs encode smORFs

Methods for predicting the coding potential of genes are continually developing. Initially, genes for which the longest ORF was less than 100aa, were classified as non-coding, however studies have shown that peptides as small as 11aa can have conserved function from flies to mammals (Pueyo and Couso, 2008). The algorithm used in this study to predict coding potential (CPC), calculates the presence, coverage and quality of ORFs and assesses number, distribution and quality of BLASTX hits for putative peptides (Kong et al., 2007). More recently the association of ribosomes to RNAs has been used to determine the likelihood of translation. However, association between an RNA moiety and a ribosome does not necessarily confirm translation. Ribosomal profiling experiments have shown occupancy of ribosomes of 5'UTRs, and indicate that occupancy of lncRNAs is similar to that of 5'UTRs (Guttman et al., 2013).

A recent study, published during the preparation of this thesis, separated polyribosomes from S2 cells based on size before sequencing the associated RNA, based on the assumption that larger polyribosomes assemble on more highly translated transcripts (Aspden et al., 2014). Consistent with this the median translational efficiency of annotated smORFs was lower in small polyribosomes (0.265) than in large polyribosomes (1.131). Although there was some evidence for smORFs originating from annotated lncRNAs, the median translation efficiencies for 918 smORFs initiating with an AUG codon originating from ncRNAs, were much lower than for annotated smORFs in both small polyribosomes (0.000) and large polyribosomes (0.384). *Inc626* and *Inc921* were not annotated in the assembly of the genome used by Aspden et al. (2014), so the occupancy for these transcripts is unknown. Both isoforms of *Inc703* can be found associated with small and large polyribosomes in S2 cells, although considering the small

and large polyribosome data separately no smORFs reach the thresholds set for translation described in the paper. Combining small and large polyribosome data together, four putative smORFs have sufficient evidence of translation (Figure 3.3a). Therefore, it is possible that *Inc703* may encode for a small peptide. Three of these smORFs would be disrupted in the two transposon insertion lines within the *Inc703* locus.

3.4.3 Selection of lncRNAs candidates for further study

Three of the six candidate lncRNAs, *Inc66*, *Inc367* and *Inc458*, were dismissed for further study because the transcript structure could not be experimentally determined. Failure to detect the 5' and 3' ends using RACE could be due to a number of factors including expression level, lack of information about transcript TSS and termination sites or that transcripts are not capped or polyadenylated. Both RACE techniques require high levels of starting material, particularly 5' RACE as RNA is lost during each of the processing steps. Although absolute levels of expression have not been determined for the lncRNAs, using the percentage of *rp49* levels in qRT-PCR experiments as a proxy for expression level, the three transcripts that could not be detected by RACE are 1 – 3 magnitudes more lowly expressed than the three transcripts that were successfully detected by RACE (Figure 3.5). Lower expression also affects the coverage of the region in the RNA-seq data sets that were used to guide the design of primers. It is therefore possible that the primers used were not sufficiently close to the true ends to achieve amplification.

Inc626 was excluded from further study because the insertion in the locus did not cause a loss of the lncRNA, conversely upregulation of *Inc626* and its neighbour *CG9300* were detected in the mutant strain. The transposon insertion in this locus contains two visible markers and UAS-binding sites that can cause activation when crossed to flies containing the transcriptional activator GAL4. In

another locus it has been shown that regulatory elements associated with these markers can cause upregulation of surrounding genes (A. Bassett, Personal communication). Other transposon insertions which cause upregulation in the locus (*Inc458*^{PBac} and *Inc66*^{PBac}) also contain UAS sites and the mini-white visible marker suggesting this may be the cause of the upregulation of the surrounding genes. However, *Inc458*^{PBac} and *Inc921*^{PBac} contain the same type of transposon, *PBac*[WH], and the insertions have opposing effects at the two loci, suggesting that the effect of an transposon is highly dependent on the genomic context of the insertion. *Inc703* was preferentially chosen over *Inc921* for further study because two independent insertions strains were available and the associated phenotypes of *Inc703* were more severe.

Chapter 4 - Characterisation of *Inc703*^{stop} and *Inc703*^{gal4} alleles

Abstract

Previously I have described two insertions that disrupt the transcription of a putative long non-coding RNA, *Inc703*. Further analysis showed that the lethal phenotype of the *Inc703*^{stop} strain was not caused by loss of *Inc703* but a tightly linked second mutation located downstream of the *Inc703* locus. This was not separated from the transposon insertion by nine generations of backcrossing. Transcriptome profiling of *Inc703*^{stop} and *Inc703*^{stop} pupae identified numerous differentially expressed genes including 88 genes that are changed in the same direction in both mutants. Gene ontology enrichment analysis indicated these genes were involved in immune response and expression changes in a subset of genes were confirmed by qRT-PCR. However, results could not be recapitulated using isogenised *Inc703*^{gal4} strains or shRNA-mediated knockdown of *Inc703*. Moreover, the levels of these genes were found to be highly variable between the two isogenised lines suggesting these genes are incredibly sensitive to genetic background.

4.1 Introduction

4.1.1 *Drosophila* innate immune response has parallels to human innate immunity

In *Drosophila* the innate immune defense in response to systemic infection by bacteria and fungus is coordinated by the fat body which releases antimicrobial peptides (AMPs) into the hemolymph. Challenge by Gram-positive bacteria or fungus activates the Toll pathway resulting in expression of different classes of AMPs: Drosomycins and Metchnikowin, or Defensin, respectively. Gram-negative bacteria primarily activate the Imd pathway which results in expression of Attacin, Cecropin, Drosocin and Diptericin AMPs (Lemaitre and Hoffmann, 2007). Both pathways activate gene expression through Rel/NF- κ B transcription factors: Relish (rel) is activated by the Imd pathway and Dorsal (DI) and Dorsal-related immunity factor (Dif) by the Toll pathway.

4.1.2 *Drosophila* immune response genes have roles in development

It is becoming increasingly apparent that genes involved in immune response genes are co-opted multiple times during development (Govind, 1999). *DI* was first identified through its role in dorsal/ventral patterning of the embryo (Steward et al., 1984) and NF- κ B factors also function in muscle development (Halfon and Keshishian, 1998), hematopoiesis (Qiu et al., 1998), and have unclear roles in the larval central nervous system (Cantera et al., 1999) and body shape (Letsou et al., 1991). There is also an association between genes involved in immunity and metabolism. Under normal conditions MEF2 is phosphorylated by p70 S6 kinase (S6K), and controls expression of genes involved in lipogenesis and glycogenesis. Upon infection, MEF2 is not phosphorylated and binds to the core promoters of AMP genes associated with the TATA-binding protein. Consequently,

the phosphorylation state of MEF2 controls the expression of either anabolic or immune response genes (Clark et al., 2013).

4.2 Methods

4.2.1 Generation of supplementary mutant strains

Transgenic BAC genomic rescue and shRNA expressing flies were generated by injection of the attB-P[acman]-CmR-BW and pWALIU22 vectors containing the Bacterial Artificial Chromosome CH322-158D05 (BAC^{158D05}) or *Inc703* targeting shRNA sequences respectively into *w¹¹¹⁸* embryos containing an attP integration site on chromosome 2L by Genetivision (Houston). Homozygous flies were selected for by eye colour and absence of curly wings. BAC^{158D05} homozygotes were crossed to *Inc703^{stop}* and *Inc703^{gal4}* strains. shRNA containing flies were crossed to *actin5C-GAL4/CyO* to achieve ubiquitous expression.

4.2.2 RNA-seq sample preparation

Age matched larvae were sexed and separated into triplicate vials containing 20 males and 20 females. Equal numbers of pupae were collected 24 hours after white puparium formation and flash frozen in liquid nitrogen. RNA extractions were performed as previously described. RNA integrity and concentration were measured using the 2100 Bioanalyser (Agilent) and male and female RNA was mixed to give 6 µg per sample. cDNA libraries were prepared and sequencing was performed by the Wellcome Trust Centre for Human Genetics (Oxford).

Reads were mapped to a reference GTF file containing all annotated genes from the *Drosophila* genome release 5.38 and additional putative lncRNAs from Graveley et al. (2011) and Young et al. (2012). edgeR (Robinson et al., 2010) was used to call differential gene expression in pairwise comparisons between control and *Inc703^{stop}* and control and *Inc703^{gal4}*.

4.2.3 Enrichment analyses of differentially expressed genes

Significance of overlap between gene expression changes in different mutant

strains was calculated using Chi-square test with Yate's correction. Enrichment for genes changing the same direction within the overlap was calculated using Fisher's exact test. Gene Ontology enrichment was performed using The Database for Annotation, Visualization and Integrated Discovery (DAVID). A background of all *D. melanogaster* genes was used as no difference was found using a background of either all *D. melanogaster* genes or only genes expressed in pupae.

4.3 Results

4.3.1 *Inc703* is a putative long non-coding RNA with two independent transposon mediated insertion strains that reduce its expression level

The initial characterisation of the *Inc703* locus (described in detail in Chapter 3), revealed *Inc703* to be a putatively non-coding gene with three exons and two alternative splice isoforms, both of which were confirmed by RACE. qRT-PCR analysis corroborated existing RNA-seq data showing that *Inc703* is expressed broadly across development, with lowest expression in late embryonic and early larval stages and highest expression in pupae and adults. Two independent *Piggybac* transposon insertions located in the first intron of *Inc703* were shown to significantly reduce *Inc703* expression level. One allele, *Inc703*^{stop}, was homozygous lethal during pupal development before and after isogenisation to *w*¹¹¹⁸. The second allele, *Inc703*^{gal4}, had a reduced viability phenotype before isogenisation that was not observed after isogenisation to *w*¹¹¹⁸. Successful excision of either insertion was not achieved and in this chapter, I will describe more detailed analysis of the phenotypes and gene expression changes associated with the two mutant strains.

4.3.2 A second mutation, outside the *Inc703* locus, is responsible for pupal lethality in *Inc703*^{stop} flies

In the absence of the reversion of the lethal phenotype of the *Inc703*^{stop} strain by excision of the transposon, a transgenic line containing a Bacterial Artificial Chromosome (BAC) containing the *Inc703* locus and the surrounding genes was inserted into an attP site on chr2L. When crossed to the *Inc703*^{stop} strain, neither one nor two copies of the BAC^{158D05} were able to rescue the lethality, suggesting 1) the BAC^{158D05} does not contain all the necessary regulatory sequences to express *Inc703* in the endogenous pattern 2) *Inc703* is only functional in its endogenous

genomic context or 3) the cause of lethality lies outside of the region covered by the BAC^{158D05}. Although it is not possible to confirm the correct expression pattern of *Inc703* expressed from the BAC^{158D05}, levels of *Inc703* in pupae were restored to a comparable level to those observed in *Inc703*^{+/stop} by one copy of the BAC^{158D05} (Figure 4.1a). Neighbouring genes α -tub84B and *e(y)2b* both exhibit expression levels comparable to *Inc703*^{+/stop} heterozygotes in mutants containing one copy of the BAC (Figure 4.1a). It is not easy to test the theory that *Inc703* is only functional in the endogenous location, so further deficiency crosses were performed to more precisely map the cause of lethality.

As previously observed, the large (~1Mb) region covered by strain 8685 (Figure 4.1b) was lethal in combination with the *Inc703*^{stop} allele (Figure 4.1c). However, strain 25724 which covers a smaller (~100kb) region that contains only *Inc703* and four neighbouring protein-coding genes (Figure 4.1b) was not lethal in combination with the *Inc703*^{stop} allele (Figure 4.1c), suggesting that the transposon insertion within *Inc703* is not the cause of lethality. Further crosses using deficiencies spanning the region covered by 8685 identified a region covered by deficiencies 8685, 9201 and 9076 that was fully or partially lethal in combination with the *Inc703*^{stop} allele (Figure 4.1c). The deleted region in 9076 covers a region >300kb downstream of *Inc703* (Figure 4.1b), suggesting that this region contains the cause of the lethality of *Inc703*^{stop}.

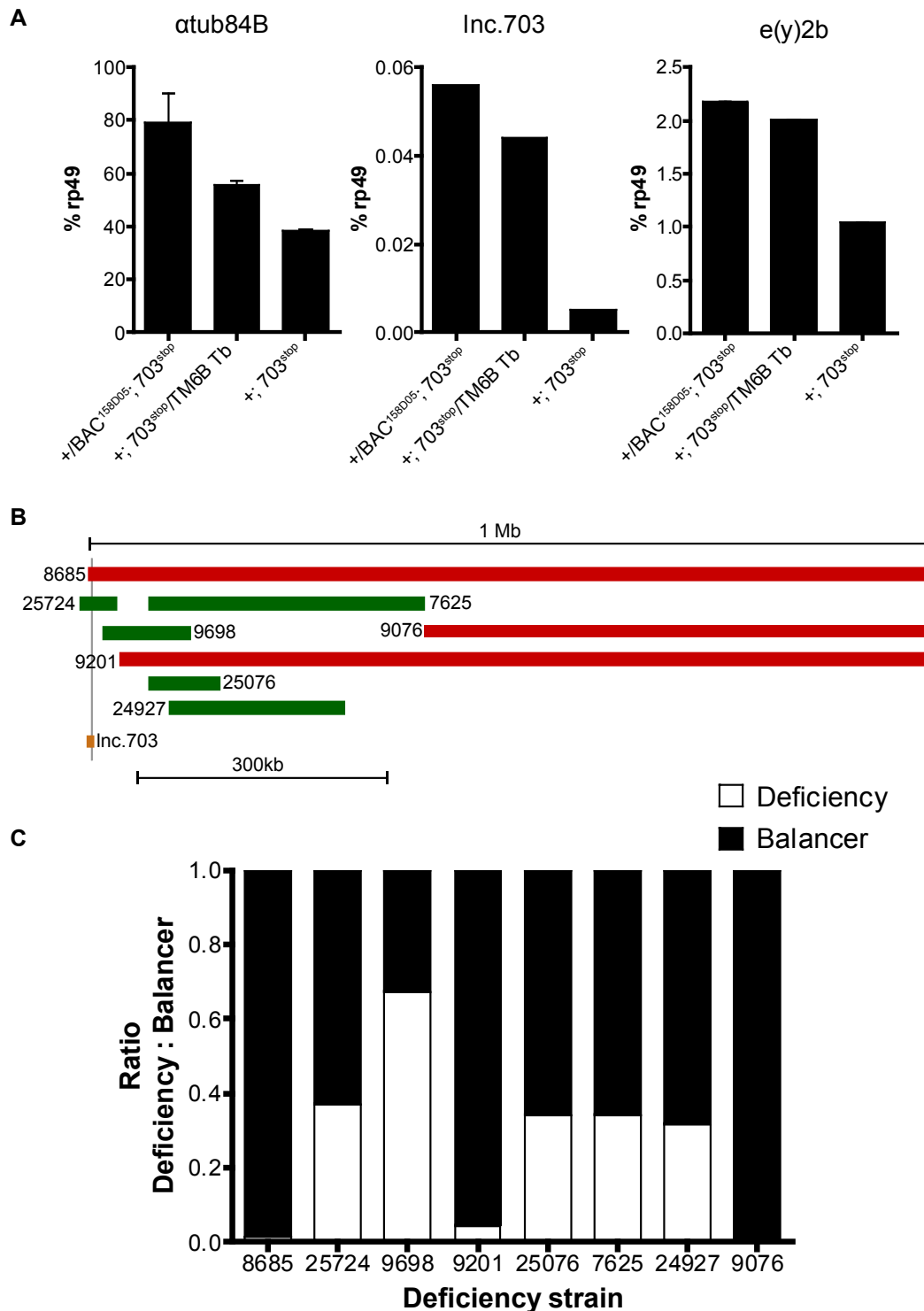


Figure 4.1 Lethality of *Inc703*^{stop} allele maps to region downstream of the *Inc703* locus. (A) Expression of *Inc703* is restored by one copy of BAC^{158D05} and levels of the nearest genes are comparable to *Inc703*^{stop} heterozygotes. (B) Schematic of the deficiencies covering the *Inc703* locus (orange), those in green are viable in combination with *Inc703*^{stop}, those in red are lethal in combination. (C) Quantification of hemizygous viability for *Inc703*^{stop} in combination with each deficiency strain.

4.3.3 RNA-seq identifies genes differentially expressed in *Inc703* mutant strains

In parallel to the genetic analysis of the mutant strains, transcriptome profiling was used to investigate the cause of the pupal lethality of *Inc703*^{stop} and identify common changes between the two mutant strains. RNA was extracted from *Inc703*^{stop}, *Inc703*^{gal4} and control pupae one day after puparium formation (APF) - approximately one day before lethality is observed in *Inc703*^{stop}. Samples were first tested by qRT-PCR to confirm reduction in *Inc703* levels (Figure 4.2a) which were comparable to the reduction seen in third instar larvae (Figure 3.4f) in both mutants.

Initial analysis of the read counts mapping to the *Inc703* locus confirmed that *Inc703* levels were reduced in both mutants, however it also revealed that expression of the first exon, upstream of the insertion site, was slightly increased in *Inc703*^{gal4} mutants (Figure 4.2b). Pairwise comparisons between the control and *Inc703*^{gal4} and *Inc703*^{stop} identified 252 and 1,956 differentially expressed genes, respectively. In both strains more genes were downregulated than upregulated (Figure 4.3a). Full length *Inc703* was detected as significantly downregulated in both mutants ($\log_2$ FC = -2.8 and -4.1, $\text{adj}P = 1.29 \times 10^{-04}$ and 3.96×10^{-09}), but the first exon that is upstream of the insertion sites is unchanged ($\log_2$ FC = 0.87 and -0.15, $\text{adj}P = 1$ and 0.97).

4.3.4 Identification of dysregulated pathways in *Inc703*^{gal4}

The initial focus was on genes differentially expressed in the *Inc703*^{gal4} mutant, given the inability to distinguish between changes caused by loss of *Inc703* and changes caused by the secondary mutation in *Inc703*^{stop} mutants. Of the 252 genes changed, 224 were downregulated and 28 were upregulated (Figure

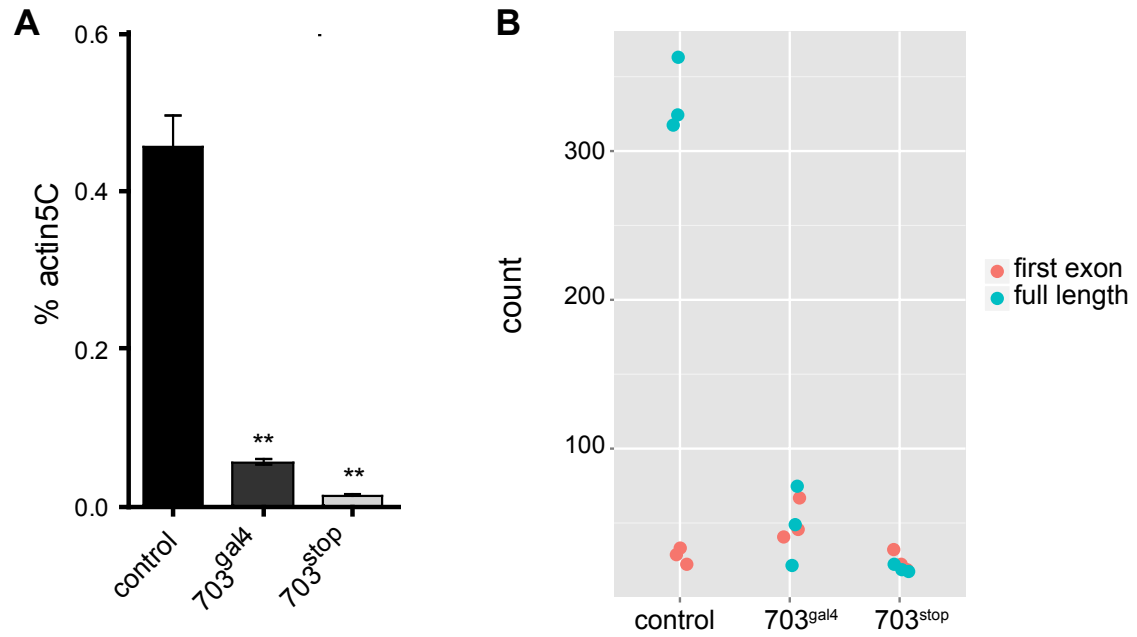


Figure 4.2 RNA sequencing of confirms reduction of *Inc703* levels in *Inc703^{gal4}* and *Inc703^{stop}* mutants. (A) qRT-PCR results show expression of *Inc703* is significantly reduced in pupae one day APF in both mutants (B) Read counts mapping to *Inc703* upstream and downstream of insertion sites confirm both transposon insertions terminate transcription. ** $p < 0.01$ compared to control, two-tailed student T-test.

4.3a). Gene ontology analysis was performed on upregulated and downregulated genes separately and in combination using the DAVID and Flymine online resources, to identify common pathways in the genes differentially expressed in the mutants. Only 14 of the upregulated genes had associated gene ontologies. Within these, were enrichments for genes involved in the cell wall (cell wall organisation or biogenesis [GO:0071554]; cell wall macromolecule metabolic process [GO:0044036]; cell wall macromolecule catabolic process [GO:0016998]) ($\text{adj}P = 2.9 \times 10^{-06}$) and immune response genes (antimicrobial humoral response [GO:0019730], $\text{adj}P = 0.002$; humoral immune response [GO:0006959], $\text{adj}P = 0.004$). Of the 215 down regulated genes, 101 have associated gene ontology terms, significant enrichments were found for genes involved in metabolic processes (aminoglycan metabolic process [GO:0006022], $\text{adj}P = 9.99 \times 10^{-04}$; chitin metabolic process [GO:0006030], $\text{adj}P = 0.002$; glucosamine-containing compound metabolic process [GO:1901071], $\text{adj}P = 0.003$; amino sugar metabolic process [GO:0006040], $\text{adj}P = 0.004$) and genes with molecular functions in the cuticle (chitin binding [GO:0008061], $\text{adj}P = 2.18 \times 10^{-04}$; structural constituent of chitin-based cuticle [GO:0005214], $\text{adj}P = 3.01 \times 10^{-04}$; structural constituent of cuticle [GO:0042302], $\text{adj}P = 4.29 \times 10^{-04}$).

Considering all differentially expressed genes together, no new parent terms for biological processes or molecular functions were found to be enriched (Figure 4.3c). However, larger numbers of genes contribute to the signal, indicating genes involved in both metabolism and immune response pathways are both upregulated and downregulated. The largest number of genes supporting a gene ontology is for the parent term 'defense response' (GO:0006952) which contains 15 genes, only three of which were upregulated (Figure 4.3c).

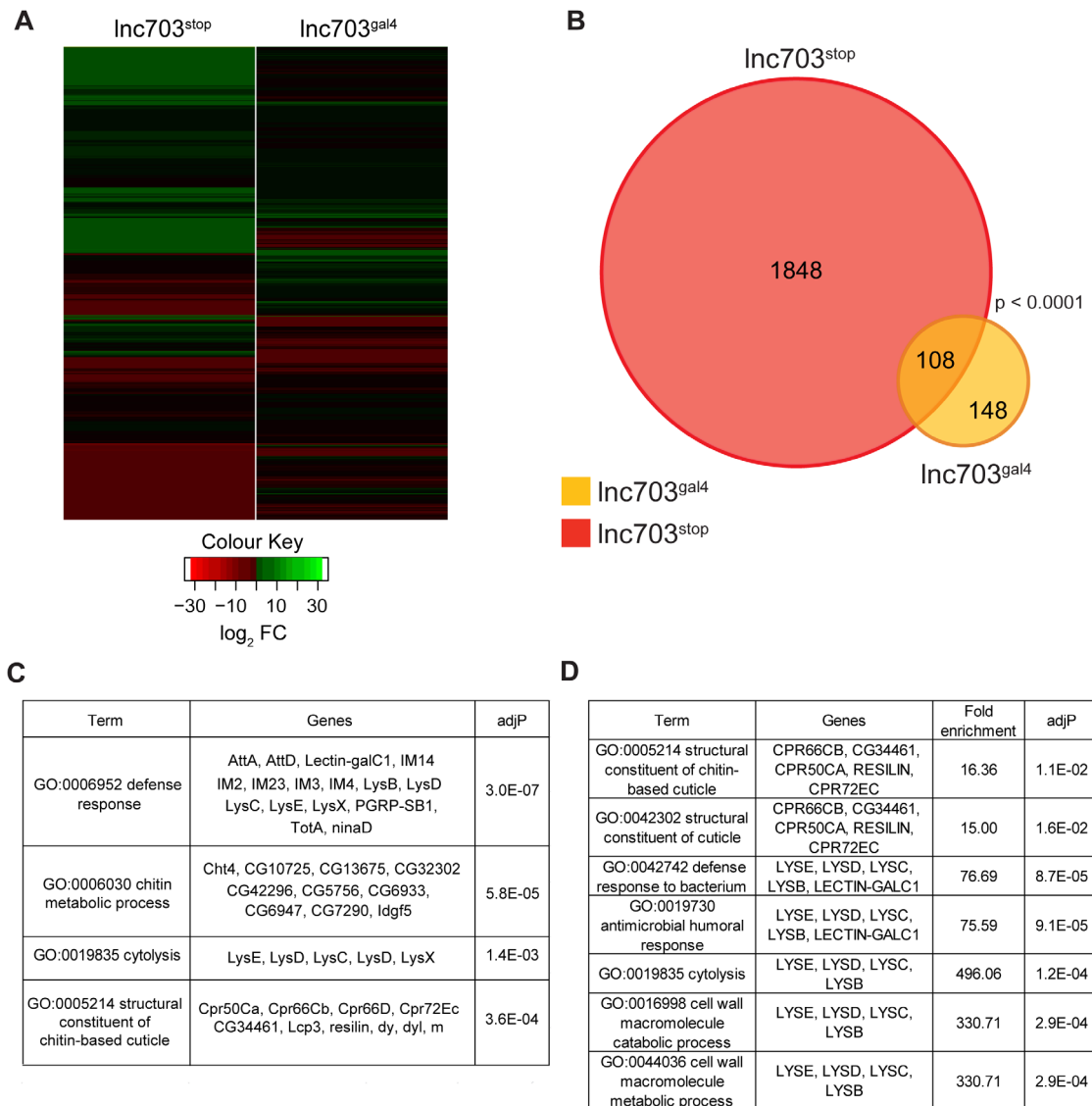


Figure 4.3 Genes differentially expressed in *Inc703* mutants include components of the cuticle and genes involved in defense response. (A) Heatmap of differentially expressed genes in mutant strains vs. wildtype. (B) Overlap between genes significantly changed *Inc703*^{gal4} and *Inc703*^{stop} mutants ($p < 0.0001$, Chi-square). (C and D) Gene ontology terms for genes changed in (C) *Inc703*^{gal4} or (D) both *Inc703*^{gal4} and *Inc703*^{stop} are enriched for structural components of the cuticle and defence response genes.

4.3.5 A significant portion of genes differentially expressed in *Inc703*^{gal4} are also changed in *Inc703*^{stop}

It was expected that genes changed as a consequence of loss of *Inc703* would behave similarly in both *Inc703*^{gal4} and *Inc703*^{stop} mutants so I analysed those genes that were differentially expressed in both mutants. 108 genes, including *Inc703*, were differentially expressed in both *Inc703*^{gal4} and *Inc703*^{stop} alleles (Figure 4.3b). Within these a significant enrichment is seen for genes changing in the same direction in both conditions (88 genes, $p < 0.0001$, Fisher's Exact) (Appendix E). Genes downregulated in both mutants were enriched for the gene ontologies involved in cuticle development (cuticle development [GO:0042335], $\text{adj}P = 0.01$; chitin-based cuticle development [GO:0040003], $\text{adj}P = 0.049$), upregulated genes were enriched for genes involved in the cell wall (cell wall organisation or biogenesis [GO:0071554]; cell wall macromolecule metabolic process [GO:0044036]; cell wall macromolecule catabolic process [GO:0016998]) ($\text{adj}P = 9.5 \times 10^{-07}$) and immune response (antimicrobial humoral response [GO:0019730], $\text{adj}P = 4.01 \times 10^{-04}$; humoral immune response [GO:0006959], $\text{adj}P = 0.001$; immune response [GO:0006955], $\text{adj}P = 0.04$). No further enrichments were found when the upregulated and downregulated genes were considered together (Figure 4.3d).

4.3.6 Genes differentially expressed in *Inc703*^{gal4} overlap with genes differentially expressed in *atf3*⁷⁶ mutants

To better define the potentially dysregulated pathways in the mutants, the differentially expressed genes were compared to those of other mutants in which metabolic and immune-related genes were changed. *atf3* is a transcription factor that controls the expression of metabolic and immune-related genes during fly

Table 4.1 Genes significantly changed in both *Inc703^{gal4}* and *atf3⁷⁶* mutants.

	703^{gal4} log₂FC	adjP	atf3⁷⁶ FC	GO term
AttA**	-2.84	7.51 x 10 ⁻⁰⁵	4.36	D/ir
AttD**	-4.95	8.28 x 10 ⁻¹²	2.14	D/ir
Brf	-1.76	3.45 x 10 ⁻⁰²	1.50	
CG10799	-1.96	2.01 x 10 ⁻⁰²	1.87	
CG10911	-3.74	2.17 x 10 ⁻⁰³	3.89	
CG13177	-2.76	1.85 x 10 ⁻⁰²	2.66	
CG14598	2.28	3.60 x 10 ⁻⁰³	-1.55	
CG15043	-2.93	3.13 x 10 ⁻⁰³	4.31	
CG15282	-2.53	2.41 x 10 ⁻⁰²	1.99	
CG15553	-2.77	2.71 x 10 ⁻⁰³	-2.42	
CG17352	3.98	1.20 x 10 ⁻⁰⁹	-2.53	
CG18641	-2.61	2.64 x 10 ⁻⁰⁴	2.14	
CG30090	-2.08	8.21 x 10 ⁻⁰³	1.95	
CG31373	1.97	2.42 x 10 ⁻⁰²	-1.65	
CG31606	1.94	2.79 x 10 ⁻⁰²	1.71	
CG32302	-2.87	3.87 x 10 ⁻⁰²	8.13	P/cm
CG33128	-4.22	1.58 x 10 ⁻⁰³	1.88	
CG5011	1.96	3.89 x 10 ⁻⁰²	2.26	
CG6836	-2.08	3.87 x 10 ⁻⁰²	1.56	
CG6933	-2.77	3.00 x 10 ⁻⁰²	1.66	P/cm
CG8665	-2.76	1.33 x 10 ⁻⁰⁴	1.62	
Cht4**	-2.92	3.47 x 10 ⁻⁰⁴	2.21	P/cm
Cpr66D	-3.29	4.20 x 10 ⁻⁰⁵	2.03	
Cyp301a1	-1.84	4.74 x 10 ⁻⁰²	2.00	
Cyp6a17	-8.60	1.35 x 10 ⁻²⁰	1.54	
Ect	-3.30	3.68 x 10 ⁻⁰⁵	1.52	
Fst	-3.27	4.42 x 10 ⁻⁰⁴	2.48	
IM2**	-2.29	9.71 x 10 ⁻⁰³	27.85	D/ir
IM23**	-2.39	1.73 x 10 ⁻⁰²	6.94	D/ir
IM3	-2.17	1.94 x 10 ⁻⁰²	38.12	D/ir
IM4	-2.07	2.49 x 10 ⁻⁰²	15.29	D/ir
LysB**	6.85	8.95 x 10 ⁻¹⁸	-35.31	D/ir
LysC**	6.16	1.18 x 10 ⁻¹⁶	-18.68	D/ir
LysD**	5.82	3.48 x 10 ⁻¹⁵	-38.61	D/ir
LysE**	4.21	4.50 x 10 ⁻⁰⁹	-32.54	D/ir
mthl8**	-9.44	2.12 x 10 ⁻²⁷	2.18	
Osi14	-3.35	2.62 x 10 ⁻⁰³	3.24	
Ptth	2.83	6.29 x 10 ⁻⁰⁵	1.50	
sisA	-3.50	1.08 x 10 ⁻⁰³	8.56	
TotA**	-2.10	1.27 x 10 ⁻⁰²	2.87	Sr/wh

D/Ir = Defense/immune response; P/Cm = Polysaccharide/chitin metabolism; Sr/Wh = Stress response/wound healing. Blue = genes validated by qRT-PCR, red = genes changed in the same direction in both mutants.

development. Mutation of *atf3* results in chronic inflammation, starvation response, lipid accumulation in the fat body and death via activation of pathways downstream of NF-KB/Relish, JNK and FOXO (Rynes et al., 2012). A significant overlap was observed between differentially expressed genes in *Inc703^{gal4}* mutants, and *atf3⁷⁶* null flies (40 genes, Fisher's exact, $p = 0.038$) (Figure 4.4a), with 36 of the 40 genes changing in opposing directions in the two mutants. Genes changed in both mutants (Table 4.1) are enriched for immune response genes (Figure 4.4b), the four genes that change in the same direction in both mutants are not annotated with GO terms relating to immune response or metabolism, possibly suggesting that *atf3* and *Inc703* may have opposing functions in these pathways. *Atf3* mRNA levels are not significantly changed in *Inc703^{gal4}* pupae ($\log_2$ FC = -0.443, $p = 1$), suggesting that the changes in the *Inc703^{gal4}* strain are independent of *atf3* levels. *Inc703* was not annotated in the version of the genome assembly used in the *atf3* study, so it is not known if *Inc703* levels are changed in *atf3⁷⁶* mutants.

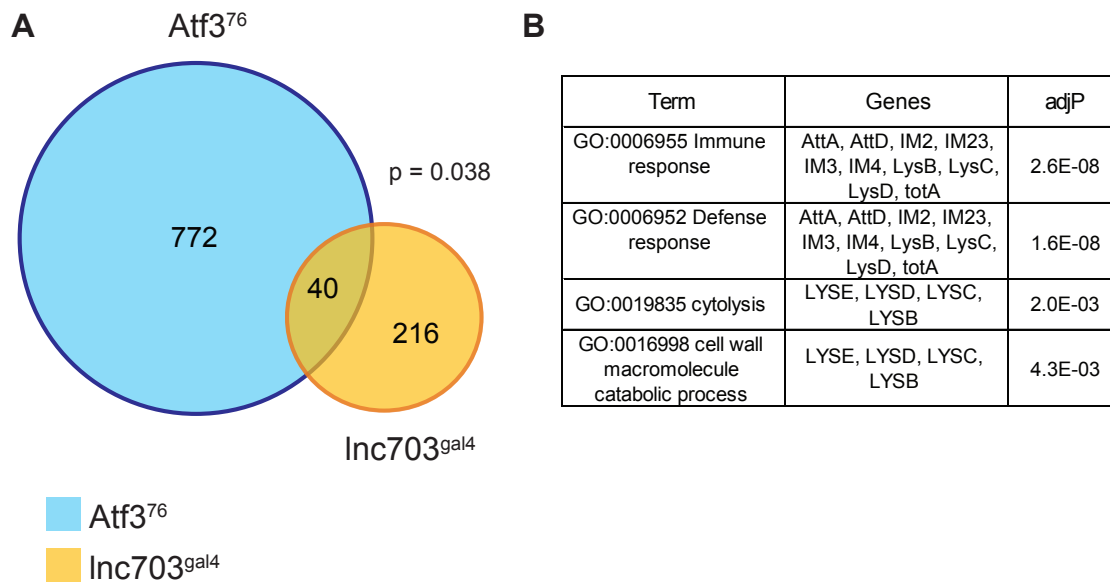


Figure 4.4 Genes differentially expressed in *Inc703^{gal4}* significantly overlap with genes differentially expressed in both *atf3⁷⁶*. (A) Overlap between *Inc703^{gal4}* mutants and *atf3⁷⁶* mutants ($p = 0.038$, Fisher's Exact). (B) Gene ontology terms for genes changed both *Inc703^{gal4}* and *atf3⁷⁶* mutants are enriched for immune response genes.

4.3.7 Gene expression changes were validated by qRT-PCR

Changes in expression were first confirmed by qRT-PCR using cDNA generated from the same RNA that the sequencing libraries were generated from to confirm similar results could be obtained by qRT-PCR and RNA-seq. Genes for validation were selected to include genes from the overlap between *Inc703^{gal4}* and *atf3⁷⁶* mutants with associated gene ontology terms (Table 4.1) and genes that were changed in *atf3⁷⁶* mutants but not in *Inc703* mutants that are master regulators of the immune response (*drs* and *relish*). Four further genes: *mlth8* which had the largest negative fold change of all genes changed in *Inc703^{gal4}* and was also changed in *atf3⁷⁶* flies; *lcp3* which had the largest positive fold change in *Inc703^{gal4}*; *CG7249* which was changed in *atf3⁷⁶* and not *Inc703^{gal4}* flies; and *e(y)2b* which had previously been shown to change in *Inc703^{gal4}* flies but was not significantly changed in the RNA-seq analysis and was not changed in *atf3⁷⁶* flies.

Excluding genes for which the expression level did not reach the required threshold (C_T value of less than 30 in wildtype samples), all 15 genes tested had the same direction change as calculated by edgeR (Figure 4.5c) in both mutant strains. Generally the magnitude of the change was comparable (adj $R^2 = 0.425$ and 0.238 , $p = 0.005$ and 0.037 , *Inc703^{gal4}* and *Inc703^{stop}*, respectively) (Figure 4.5a), although for *mlth8* and *lcp3* - the genes with the largest downregulation and upregulation respectively - the magnitude determined by qRT-PCR analysis was much lower. Excluding *mlth8* and *lcp3*, positive correlation between the two methods was much stronger for both *Inc703^{gal4}* (adj $R^2 = 0.956$, $p = 5.5 \times 10^{-09}$) and *Inc703^{stop}* (adj $R^2 = 0.916$, $p = 1.8 \times 10^{-07}$) (Figure 4.5b). Despite the correlation between the two methods, only six of the 11 genes that were significantly changed as determined by edgeR were significant in qRT-PCR experiments ($p < 0.05$, two-tailed student t-test), and two genes (*rel* and *e(y)2b*) that were not significant by edgeR showed equivalent fold change differences

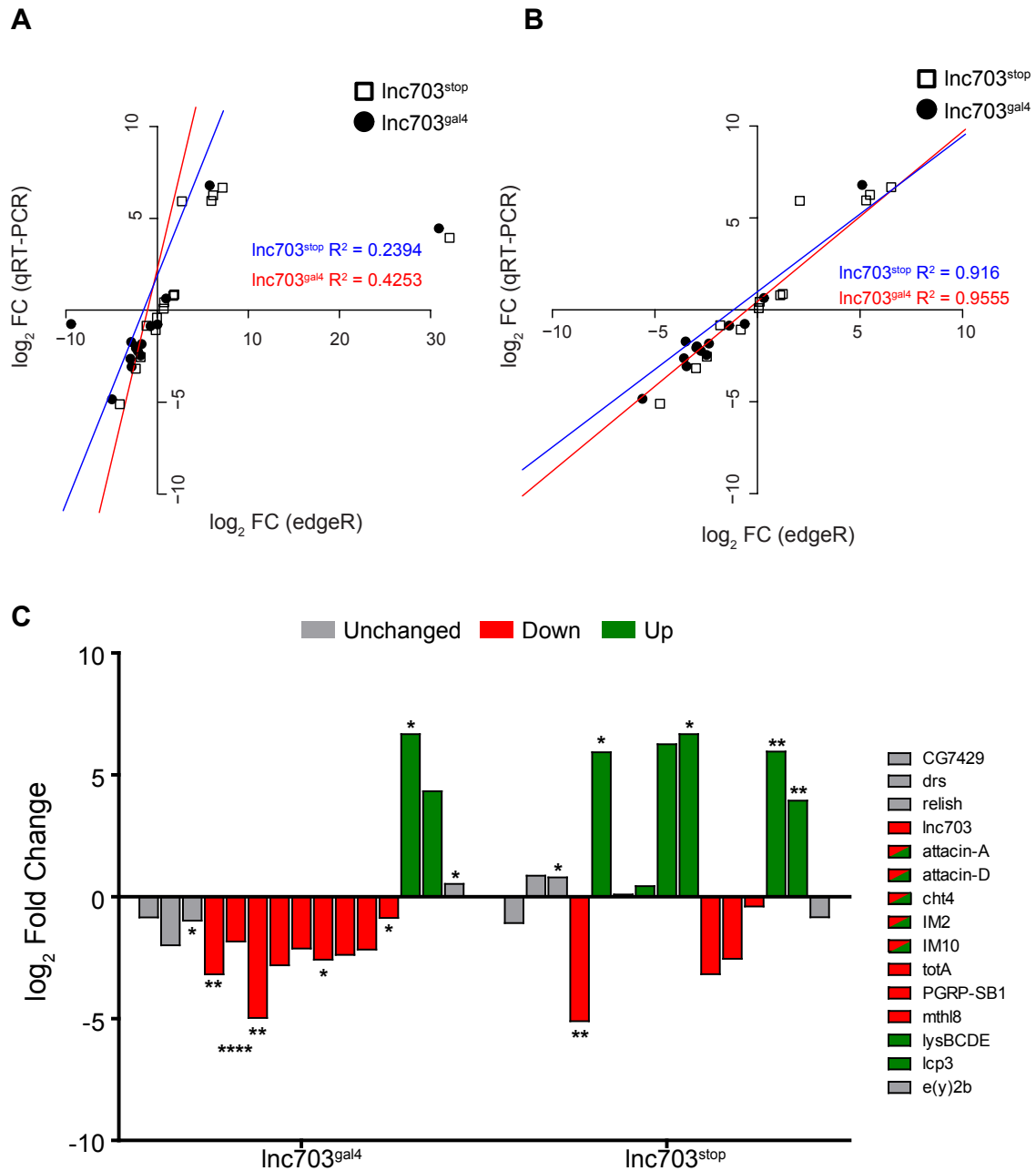


Figure 4.5 Gene expression changes calculated from RNA-seq data can be reproduced by qRT-PCR. (A) Correlation of expression changes measured by qRT-PCR and RNA-seq. (B) Correlation excluding *mth18* and *lcp3*. (A and B) Regression lines show strong positive correlation (red = *Inc703*^{stop}; blue = *Inc703*^{gal4}). (C) Log₂ fold change of subset of genes, grey = not significantly changed, red = significantly downregulated and green = significantly upregulated. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$, two-tailed Student t-test.

but were considered significant by qRT-PCR (Figure 4.5c). These differences may be due to the different normalisation techniques, qRT-PCR experiments are normalised to a single control gene *actin5C*, whereas edgeR uses scaling factors that minimise the fold change for most genes.

4.3.8 Changes in immune-related genes could not be reproduced in *Inc703*^{gal4} mutants after isogenisation

To determine if the same gene expression changes could be seen in *Inc703*^{gal4} mutants after isogenisation, levels of the same subset of genes that had been validated by qRT-PCR were measured in two isogenised *Inc703*^{gal4} strains compared to *w*¹¹¹⁸ controls.

Levels of *Inc703* are reduced in pupae compared to *w*¹¹¹⁸ controls in both isogenised strains (Figure 4.6a, log₂ FC = -3.81 and -5.21, *p* < 0.0001, One-way ANOVA, Bonferroni Multiple Comparison Test). None of the other genes tested changed significantly in either isogenised strain compared to controls (Figure 4.6a). However, not all changes reached significance by qRT-PCR before isogenisation (Figure 4.5c) so I calculated the Pearson's correlation coefficient for the log₂ fold change determined by qRT-PCR of the two isogenised strains and the log₂ fold change before isogenisation to test if similar changes that did not reach significance were found. No significant correlation was observed for either isogenised strain (Figure 4.6b), suggesting that the changes detected in immune-related genes might not be caused by loss of *Inc703* but represent changes caused either by changes in developmental timing of the *Inc703*^{gal4} strain before isogenisation, an interaction between loss of *Inc703* and the genetic background or an unidentified second mutation.

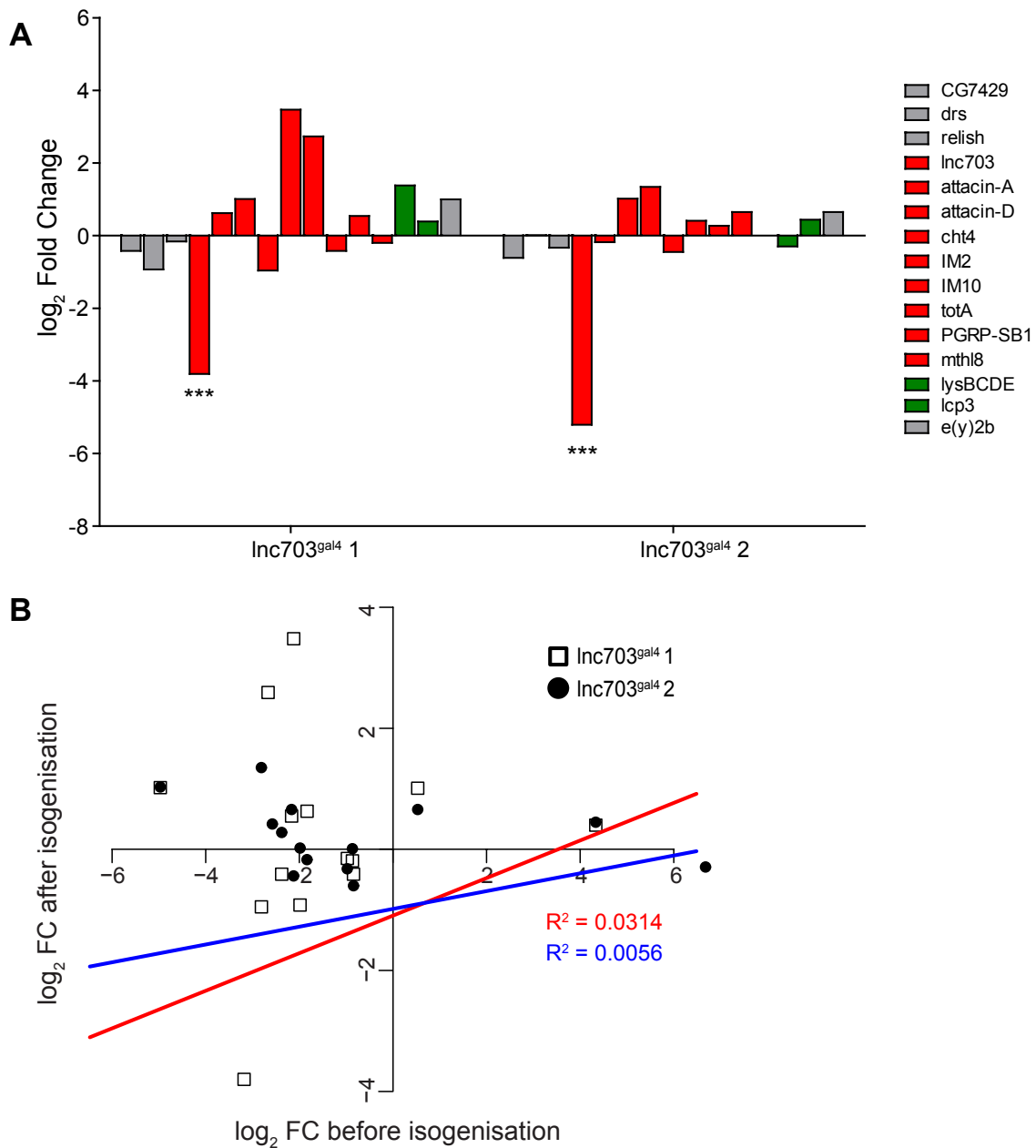


Figure 4.6 Reduction in *Inc703* levels but not of other genes are replicated in pupae one day APF in two independent isogenised *Inc703*^{gal4} strains. (A) qRT-PCR results showing log₂ fold change of validated genes in two isogenised *Inc703*^{gal4} strains compared to *w*¹¹¹⁸ controls. Grey = no change, red = downregulated, green = upregulated. *** $p < 0.001$ one-way ANOVA compared to *w*¹¹¹⁸. (B) No significant correlation is observed between the log₂ FC of either isogenised strain with the original *Inc703*^{gal4} strain ($p = 0.528$ and 0.791).

4.3.9 Immune response genes are not changed in *Inc703*^{gal4} adult females

The relationship between the role of immune related genes during development and in response to infection is unclear and the response to infection in *Drosophila* is very dependent on developmental timing, sex and mating status. Therefore most experiments studying immune related genes are carried out on mated adult females between two and four days after eclosion. Furthermore, *Inc703* is expressed highly in adult females (Figure 3.2). Therefore, the levels of the same subset of immune response genes identified by RNA-seq were measured in *Inc703*^{gal4} adult females.

The changes in *Inc703* ($\log_2$ FC = -5.33 and -5.70, $p < 0.0001$, 1-way ANOVA) and neighbouring gene *e(y)2b* ($\log_2$ FC = 1.61 and 1.44, $p = 0.0007$, 1-way ANOVA) that were observed in the *Inc703*^{gal4} strain before isogenisation were also observed in adult females in both isogenised strains (Figure 4.7a). However, none of the other genes tested showed a significant change in either isogenised strain. Moreover, no significant correlation was found between the $\log_2$ fold change relative to *w*¹¹¹⁸ controls for the two different isogenised strains ($R^2 = 0.247$, $p = 0.071$, Pearson's correlation), indicating that these genes are highly variable in adult females.

4.3.10 shRNAs targeted to *Inc703* reduce transcript levels but do not reproduce gene expression changes seen in mutants

Given the effect of genetic background on the expression of immune related genes it is possible that the failure to reproduce the same gene expression changes in *Inc703*^{gal4} after isogenisation might be caused by the different genetic background. In an attempt to reproduce gene expression changes observed in the *Inc703*^{gal4} mutants using an independent method, expression levels were measured in pupae expressing shRNAs targeting *Inc703* driven by *actin5C*-GAL4.

Levels of *Inc703* were reduced between 2.0 – 2.8 fold in pupae one day APF in three lines expressing independent shRNAs targeting *Inc703*, compared to *actin5C*-GAL4 lines expressing RFP (Figure 4.8a). The levels of immune related genes were measured in *actin5C*-GAL4; *UAS-shRNA-11.3* pupae which had the greatest reduction in *Inc703* levels. Whilst genes were variable between strains, the gene expression changes observed in the original *Inc703*^{gal4} strain were not reproduced (Figure 4.8b).

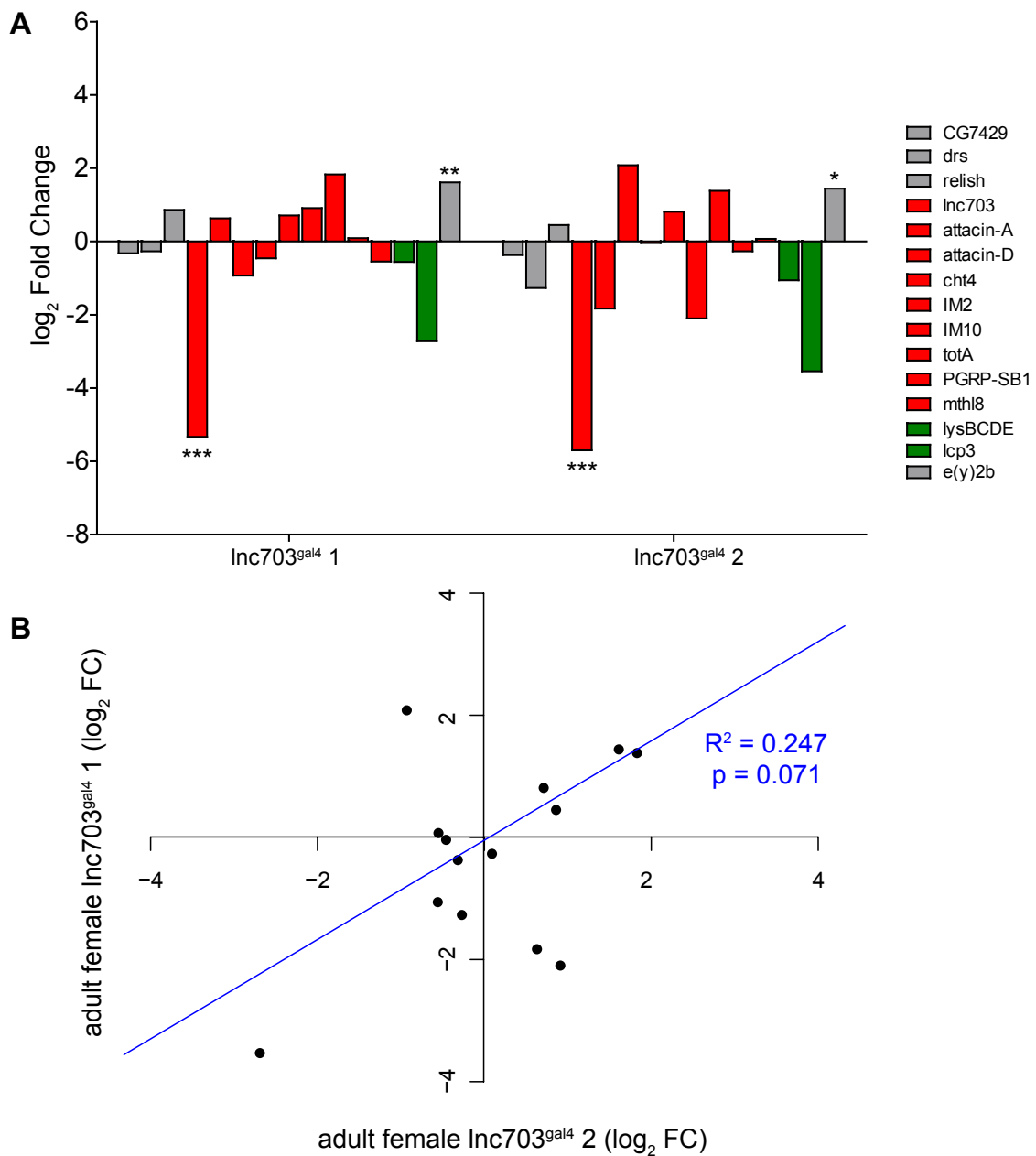


Figure 4.7 Immune response genes are not significantly changed in *Inc703*^{gal4} adult females. (A) log₂ fold change calculated by qRT-PCR of immune related genes in two independent isogenised *Inc703*^{gal4} strains compared to *w*¹¹¹⁸ controls. Colour corresponds to change in RNA-seq. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, two-tailed student T-test. (B) No significant correlation between log₂ fold change of immune response genes in the two different isogenised *Inc703*^{gal4} strains (excluding *Inc703*), Pearson's correlation.

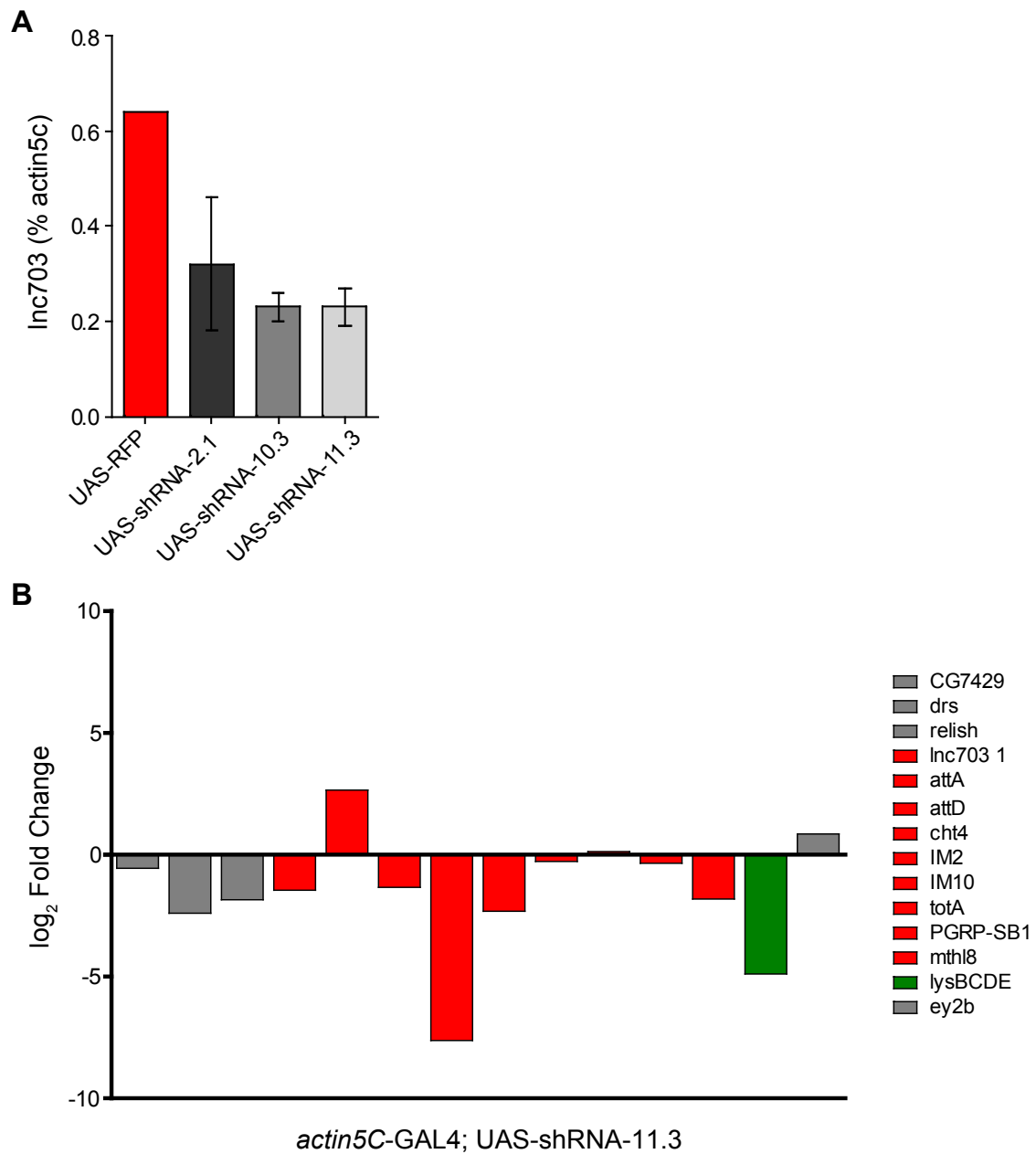


Figure 4.8 Immune response genes are not changed in pupae with shRNA- mediated *Inc703* knockdown. (A) Levels of *Inc703* in pupae expressing three different shRNA targeted to *Inc703* or expressing RFP under the control of *actin5C-GAL4*. (B) Gene expression changes in immune response genes in *actin5C-GAL4; UAS-shRNA-11.3* pupae.

4.4 Discussion

4.4.1 The challenges in discovering lncRNA loss-of-function phenotypes

The majority of lncRNA loss-of-function studies have been carried out in cell culture systems where lncRNAs have been proposed to have diverse functions as molecular scaffolds (*NEAT1*) (Clemson et al., 2009), controllers of differentiation fate (*DALI* and *linc-RoR*) (Chalei et al., 2014; Wang et al., 2013) and regulators of metabolic rate (*CEROX*, Personal Communication, Dr Tamara Sirey). However, how these cellular effects manifest as phenotypic changes on a tissue or organismal level is unclear. For example, loss of *NEAT1* in mice results in complete loss of paraspeckle formation in cell lines and *in vivo*, but initial characterisation found the mice to be indistinguishable from littermate controls in terms of development, viability and behaviour (Nakagawa et al., 2011). However, further analysis found *Neat1*^{-/-} mothers were unable to provide sufficient milk to wean their offspring due to reduction in the number of milk producing cells (Standaert et al., 2014) although a causal link to lack of paraspeckle formation has not been established. In recent study in which 18 putative lncRNA locus were replaced with lacZ reporters in mice, three mutant strains were lethal (Sauvageau et al., 2013). However, it is not yet clear whether these phenotypes are due to loss of the lncRNA or a result of loss of the underlying DNA sequence effecting proper function of nearby developmentally important genes. The latter seems particularly likely in the case of *Mdgt* in which the 12.4 kb that was deleted is only 84 bp from the 5' end of *HoxD1*. The deletion includes many transcription factor binding sites conserved in human and mice whose deletion is likely to result in misregulation of *HoxD1*.

Therefore, it is perhaps unsurprising that the lethality of the *Inc703*^{stop} strain is caused by a second mutation outside of the *Inc703* locus. That lncRNAs may only have subtle phenotypes, or phenotypes that manifest only under certain

conditions (Bassett et al., 2014a) makes reverse genetic approaches challenging to study lncRNAs. Instead, it may be more profitable to identify lncRNAs involved in a particular process or cell-type transition using high-throughput cDNA sequencing or custom microarrays from a more defined cell type at different time points or under different conditions. For example, *lincRNA-Cox2* was identified as upregulated upon LPS stimulation and found to control immune response genes in a NFkB dependent manner (Carpenter et al., 2013). Alternatively, forward genetic approaches could be employed to identify lncRNAs involved in particular processes, although very few lncRNAs have so far been identified from such phenotypic screens. There are a number of possible reasons why lncRNAs have not been identified. Firstly, the mutation methods used may not be suited to disruption of lncRNAs. Transposon insertions have inconsistent effects on lncRNA expression (discussed in Chapter 3), and EMS screens introduce point mutations that are unlikely to disrupt the function of lncRNAs, as in many cases only a small part of the lncRNA sequence is required. It is also likely that RNA secondary structures are more tolerant of mutations than protein-coding sequences. Secondly, mutation of lncRNAs may result in more subtle phenotypes and therefore may not be selected over alleles with stronger effects. Finally, there may be increased redundancy between lncRNAs, particularly if only a small region of the RNA is functional.

4.4.2 Are immune response genes innately variable?

Comparison of changes in levels of immune genes between multiple strains in this chapter showed a large variability in expression level not consistent with the presence or absence of expression of *Inc703*. It is therefore not currently possible to separate any effect of *Inc703* from the variability of immune related genes. Genes involved in the immune response pathway are frequently identified as

differentially expressed in genome wide studies (Teixeira, 2012). Whether this is due to innate variability of expression levels between strains or whether immune response genes are particularly sensitive to homeostasis is unknown. Recent evidence suggests a role of nutritional state in immune response, which is in turn, influenced by genetic variation (Unckless et al., 2015). The relationship between *Inc703* and the genes defined as immune-related is further complicated by the apparent dual role of many genes initially identified for their role in the immune response during development and in maintaining homeostasis. It is therefore possible that the changes observed are secondary effects resulting from altered metabolism or disturbed homeostasis caused by loss of *Inc703* that are dependent on genetic background.

The inability to reproduce the same gene expression changes using shRNA mediated knockdown could be due to the extent of knockdown being insufficient, or in the wrong tissues, or that the act of transcription is critical for *Inc703* function rather than the levels of mature transcript. The pWALIUM22 vector used for shRNA expression primarily expresses the the female germline (Ni et al. 2010). Therefore, *Inc703* levels may not be reduced in the somatic cells of pupae resulting in the moderate reduction observed (Figure 4.8a). It is therefore possible that if the changes in immune response genes are driven by changes in *Inc703* levels in the soma they cannot be reproduced by germline shRNA knockdown using the pWALIUM22 system. The pWALIUM20 vector which expresses less efficiently in the germline but more strongly in the soma could be used to assess if somatic reduction of *Inc703* levels results in altered expression of immune response genes. However, taken together with the inconsistency between the mutant strains before and after isogenisation, it is most likely that the gene expression changes detected by RNA-seq are the result of differences between the genetic backgrounds or developmental timing of the strains.

4.4.3 Identification of clusters within differentially expressed genes is biased towards known genes

Although the changes in immune response genes appear to be independent of *Inc703*, these genes represent only a proportion of the genes differentially expressed in *Inc703^{gal4}* and a very small proportion of genes differentially expressed in *Inc703^{stop}*. It is therefore possible that other gene expression changes are caused by *Inc703* but did not contribute to the analysis, as not all the genes had gene ontology annotations. Whilst pathway and gene ontology analysis can be useful to identify common features in large sets of genes, they rely on prior annotation of the function of genes. Consequently, these methods are biased towards previously studied genes and processes. For example, 42 out of 81 of protein-coding genes changed in the same direction in both *Inc703^{gal4}* and *Inc703^{stop}* mutants were excluded from the gene ontology analysis and 75 out of the same 81 were excluded from pathway enrichment analysis. Although genome wide screens, particularly genome-wide RNAi screens, are increasing the amount of information known about previously unstudied genes, many genes are still not covered by these analyses, particularly non-coding genes.

4.4.4 Summary

A second mutation 300kb downstream of the *Inc703* locus, covered by the deficiency stain 9076 is responsible for lethality of the *Inc703^{stop}* strain. 108 genes were found to be differentially expressed in pupae of both *Inc703^{gal4}* and *Inc703^{stop}* mutants, including 88 that are changed in the same direction in both. A small but significant enrichment for immune response genes was found, but changes in these genes could not be reproduced in isogenised or shRNA-mediated knockdown. Although these genes represent only a small proportion of the genes

changed in both mutants, no other common pathways were identified using GO analysis, making it difficult to identify other phenotypes resulting from loss of *Inc703*.

Chapter 5 - Tissue specific expression of *Inc703*

Abstract

Broad level temporal and tissue specific expression information for *Inc703* was previously established using both RNA-seq (Graveley et al., 2011) and qRT-PCR (Chapter 3, Figure 3.2) data. In this chapter, the cell type specific expression was determined using fluorescent reporters driven by the expression of GAL4 present within the transposon insertion of the *Inc703^{gal4}* strain. Fluorescent reporters indicated expression in a subset of cells of the larval cuticle, larval CNS, imaginal eye discs and cells of the larval fat body but not the embedded gonads. Analysis of *Inc703^{gal4}* larval fat bodies did not reveal any gross changes in triacylglyceride (TAG) levels or lipid droplet size. Upregulation of genes involved in translation were detected in the larval fat body of both *Inc703^{gal4}* and *Inc703^{stop}* mutant strains, and initial results indicate protein levels of RpS6 are also increased. These changes were not rescued by two copies a BAC covering the *Inc703* locus inserted at a second site. The lack of rescue in *trans* suggests that this effect may be dependent on the genomic context of *Inc703* at its endogenous locus.

5.1 Introduction

5.1.1 Methods for determining RNA expression at the cellular level

As part of the modENOCDE project, high throughput cDNA sequencing was performed on 30 different *D. melanogaster* tissues and developmental stages (Graveley et al., 2011). The developmental expression profile for *Inc703* (Figure 5.1) is in good agreement with the profile determined by qRT-PCR in Chapter 3 (Figure 3.2). Expression is lowest in late embryos and first instar larvae, and peaks in late third instar larvae and early pupae and is highest in adult females. Whilst some additional RNA-seq data is available about expression in different tissues, it is limited and does not distinguish between low level expression in all cells or high expression in a small number of cells. Direct visualisation of RNA can be achieved using Fluorescent *In Situ* Hybridisation (FISH); however this may be problematic for lowly expressed lncRNAs. A common technique for discovering expression patterns of unknown genes is through “enhancer-trapping” (Bellen et al., 1989). The insertion of a transposon containing the transcriptional activator GAL4 under the control of a minimal promoter, allows for expression of GAL4 to be dependent on surrounding enhancer sequences (Kaiser, 1993). By crossing enhancer trap flies to UAS-driven lacZ or fluorescent reporters the expression pattern can be visualised. The activation of reporter expression by GAL4 acts as an amplification step meaning it may be suitable for lowly expressed transcripts.

5.1.2 Diverse functions of the *Drosophila* fat body

The primary role of the larval fat body is to provide a store of energy that can be mobilised to fuel metamorphosis and protect against starvation in the early adult. More than that, it has the endocrine functions analogous to the vertebrate liver and coordinates organismal growth based on nutritional status (Colombani et

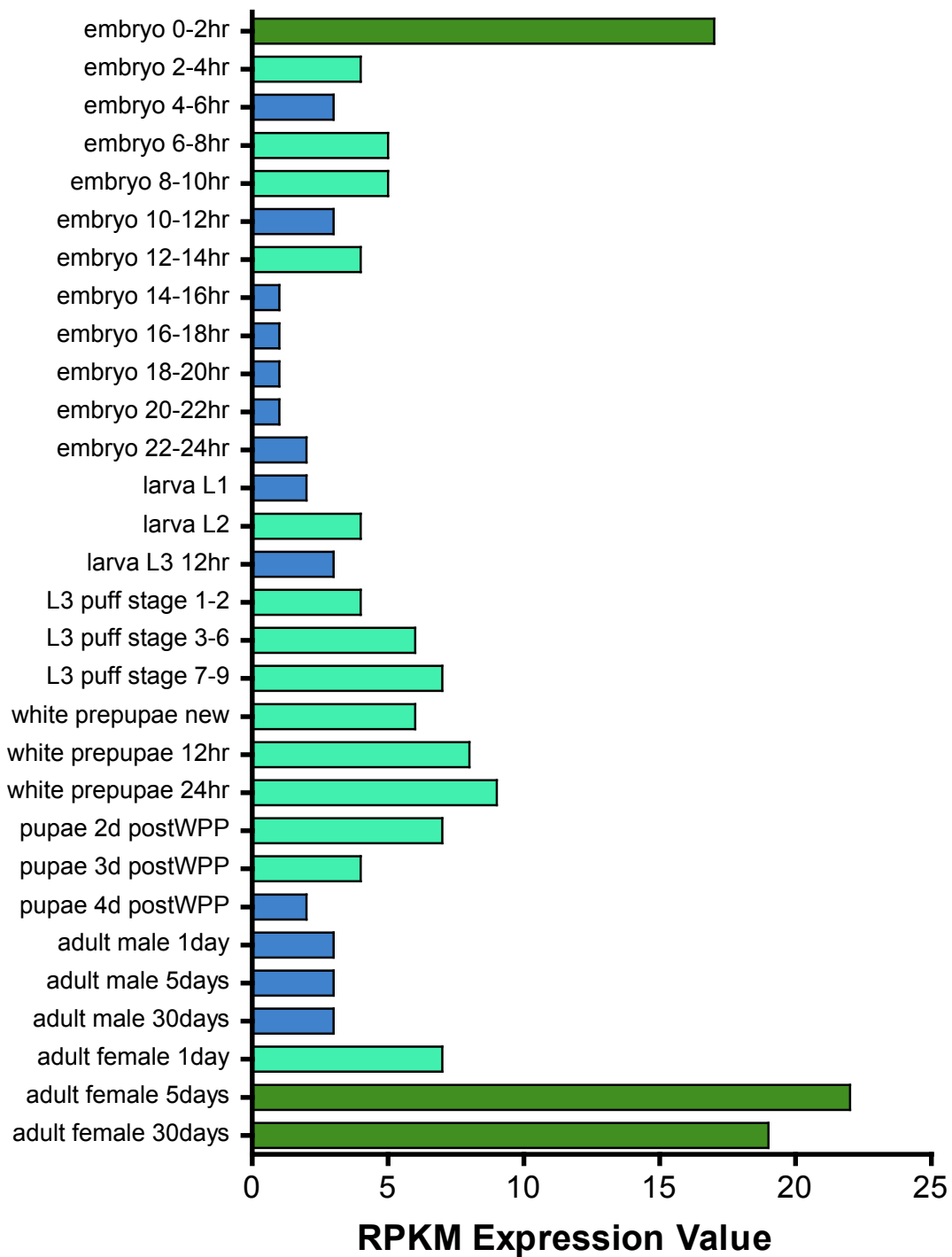


Figure 5.1 Developmental expression profile of *Inc703* (CR43685) from RNA-seq in OregonR. Blue = very low expression, turquoise = low expression, green = moderate expression. Data from Graveley et al. (2011.4.13).

al., 2003; Slaidina et al., 2009). Both the larval and adult fat bodies also secrete AMPs in response to infection (Bulet et al., 1999).

5.1.3 Development of the *Drosophila* larval and adult fat bodies is distinct

The cells of the larval fat body are specified early in development and grow by mitotic division between six and eight hours after fertilisation, after which they switch to endoreplicative growth throughout the rest of embryogenesis and larval development (Neufeld, 2008). Fat stores are mobilised from dissociated fat body cells which are protected from autophagic cell death during pupal development and although some larval fat body cells persist to protect newly eclosed adults from starvation, these are removed by the caspase cascade once the adult fat body has been formed (Aguila et al., 2007). The adult fat body is not descended from the larval fat body but from mesodermal progenitors (Hoshizaki et al., 1995).

5.1.4 Ribosome levels influence cellular and organism fitness

A series of heterozygous mutations that result in small adults map to ribosomal proteins and mutations in the *Drosophila* Myc homologue *diminutive* (*dm*) which results in decreased ribosome levels decreases body size (Johnston et al., 1999). Together this suggests that levels of ribosomal proteins influence final body size. Moreover, *dm* levels in the fat body can have a non-cell autonomous effect on whole animal size. Increased *dm* levels in the larval fat body increase adult body size (Parisi et al., 2013) and reduced *dm* levels in the larval fat body reduce adult body size (Delanoue et al., 2010). On a cellular level, mutant cells with reduced ribosome levels are less fit than surrounding wildtype cells (Morata and Ripoll, 1975) and are outcompeted by them and lost by apoptosis (Moreno et al., 2002). Conversely, mutant cells with increased *dm* levels are “super-competitors” which

can induce apoptosis of surrounding wildtype cells (Moreno and Basler, 2004).

5.1.5 Aim

In this chapter I identify the tissue specific expression pattern of *Inc703* using a GAL4/UAS reporter system driven by the GAL4 transcriptional activator within the *Inc703^{gal4}* allele. The expression pattern will then be used to determine appropriate tissues for phenotypic analysis assessing for both gross morphological differences and identifying dysregulated pathways using high-throughput cDNA sequencing.

5.2 Methods

5.2.1 Staging

To ensure correct developmental timing of samples, all samples were collected from food bottles which had been populated by the same number of adults for the same time and kept under the same light/dark, temperature and humidity conditions. Third instar larvae were selected when unresponsive to mechanical stimulus and sex was determined by identification of the gonads in males.

5.2.2 GAL4/UAS reporter system

Two independent strains containing UAS-driven fluorescent reporters (*UAS-mCD8::GFP* and *UAS-mCD8::GFP*, *UAS-H2B::mCherry*) were driven by the GAL4 within the *Inc703^{gal4}* allele (*PBac[GAL4D,EYFP]PL00061*, Hacker et al. 2003). For the *UAS-mCD8::GFP*, *UAS-H2B::mCherry* reporter it was not possible to generate a stable stock as the alleles are on the same chromosome as the *Inc703^{gal4}* allele. Therefore, flies homozygous for each allele were crossed so all progeny contained one copy of each allele and a wildtype copy of *Inc703*. A stable stock of *UAS-mCD8::GFP/CyO*; *Inc703^{gal4}/TM6B Tb* flies was generated and analysis was performed on non-CyO, non-*Tubby* larvae.

mCD8::GFP signal was distinguished from *PAX6*-driven EYFP using mouse anti-mCD8a::APC (1:200, Miltenyi Biotec). Known markers were used to identify specific cell types: rat anti-Ci (1:100, a gift from Robert Holmgren), mouse anti-Repo (1:100, DSHB Hybridoma Product 8D12), rat anti-Elav (1:1,000, DSHB Hybridoma Product 7E8A10), rabbit anti-Pros (1:100, a gift from Denan Wang), guinea pig anti-Dpn (1:10,000, a gift from James B. Skeath). Secondary antibodies donkey anti-rat-Cy5, donkey anti-rabbit-Cy5, donkey anti-mouse Cy5, (Jackson ImmunoResearch Laboratories) goat anti-guinea pig Alexa633 (Molecular Probes) were used 1:1,000.

5.2.3 Cell fractionation

RNA from the cytoplasm, nucleoplasm, loosely associated with DNA (salt extract) and tightly bound to the chromatin was separated sequentially from a set number of S2 cells. All steps were performed at 4 °C. Briefly, cells were pelleted, washed in PBS, counted and 4×10^7 cells were resuspended in 2 ml of cold lysis buffer (15 mM HEPES pH 7.5, 10 mM KCl, 5 mM MgCl_2 , 0.1 mM EDTA, 0.5 mM EGTA, 250 mM Sucrose, 0.4% NP-40; with 1 mM DTT, 0.5 mM PMSF, 1x Protease inhibitors and 100 U/ml RNasin). After agitating for 20 min, 10% was taken as input. Nuclei were pelleted at 2000 *g* for 10 min and the supernatant was taken as the cytoplasmic fraction. Pellet was resuspended in an equal volume of cold nuclei lysis buffer (10 mM HEPES pH 7.5, 0.1 mM EDTA, 0.1 mM EGTA; with 1 mM DTT, 0.5 mM PMSF, 1x Protease inhibitors and 100 U/ml RNasin) and passed through a needle to break open the nuclei. The supernatant (nucleoplasm fraction) was removed after spinning at 17000 *g* for 5 min. The pellet was resuspended in an equal volume of cold salt extraction buffer (25 mM HEPES pH 7.5, 10% Glycerol, 420mM NaCl, 5 mM MgCl_2 , 0.1 mM EDTA; with 1 mM DTT, 0.5 mM PMSF, 1x Protease inhibitors and 100 U/ml RNasin. After agitating at for 30 min, chromatin was pelleted at 17 000 *g* for 20 min and the supernatant was taken as the salt extract. The pellet containing chromatin bound RNA was resuspended in an equal volume of cold salt extraction buffer. SDS, EDTA and Proteinase K were added to 400 μl of each sample to give final concentration of 1% SDS, 10 mM EDTA and 200 $\mu\text{g/ml}$ Proteinase K. RNA was extracted by phenol chloroform extraction. The above steps were performed by Drs Andrew Bassett and Charlotte Tibbit. Reverse transcription and qRT-PCR were performed as previously described using 1 μl of sample per reverse transcription reaction.

5.2.4 Producing mitotic clones in the larval fat body using Mosaic Analysis with a Repressible Cell Marker (MARCM)

GFP labelled patches of cells, homozygous for *Inc703^{gal4}* were generated by MARCM (Lee and Luo, 2001) (Figure 5.2). 0 – 12 hour embryos were placed in a water bath at 37°C for 1 hour, to induce *hsFLP* expression resulting in recombination between FRT82B sites on the third chromosome of *Inc703^{gal4}* and MARCM of cells in G2 phase. After mitosis daughter cells possess either two copies of the *Inc703^{gal4}* allele or two copies of the *GAL80* allele. In the absence of *GAL80*, cells mutant for *Inc703* will express *GAL4* and subsequently GFP, whereas in cells wildtype for *Inc703*, *GAL4* will be continued to be repressed by *GAL80* and they will be GFP negative.

5.2.5 Transcriptome profiling of larval fat bodies

The fat bodies of six male and six female third instar larvae were dissected in cold RNase free PBS and flash frozen in liquid nitrogen in triplicate for each genotype. Samples were prepared and sequenced as described in Chapter 4. Reads were mapped as described in Chapter 4 and pairwise comparisons between each sample and the *w¹¹¹⁸* control were performed using DESeq2, using an FDR of 5% (Love et al., 2014).

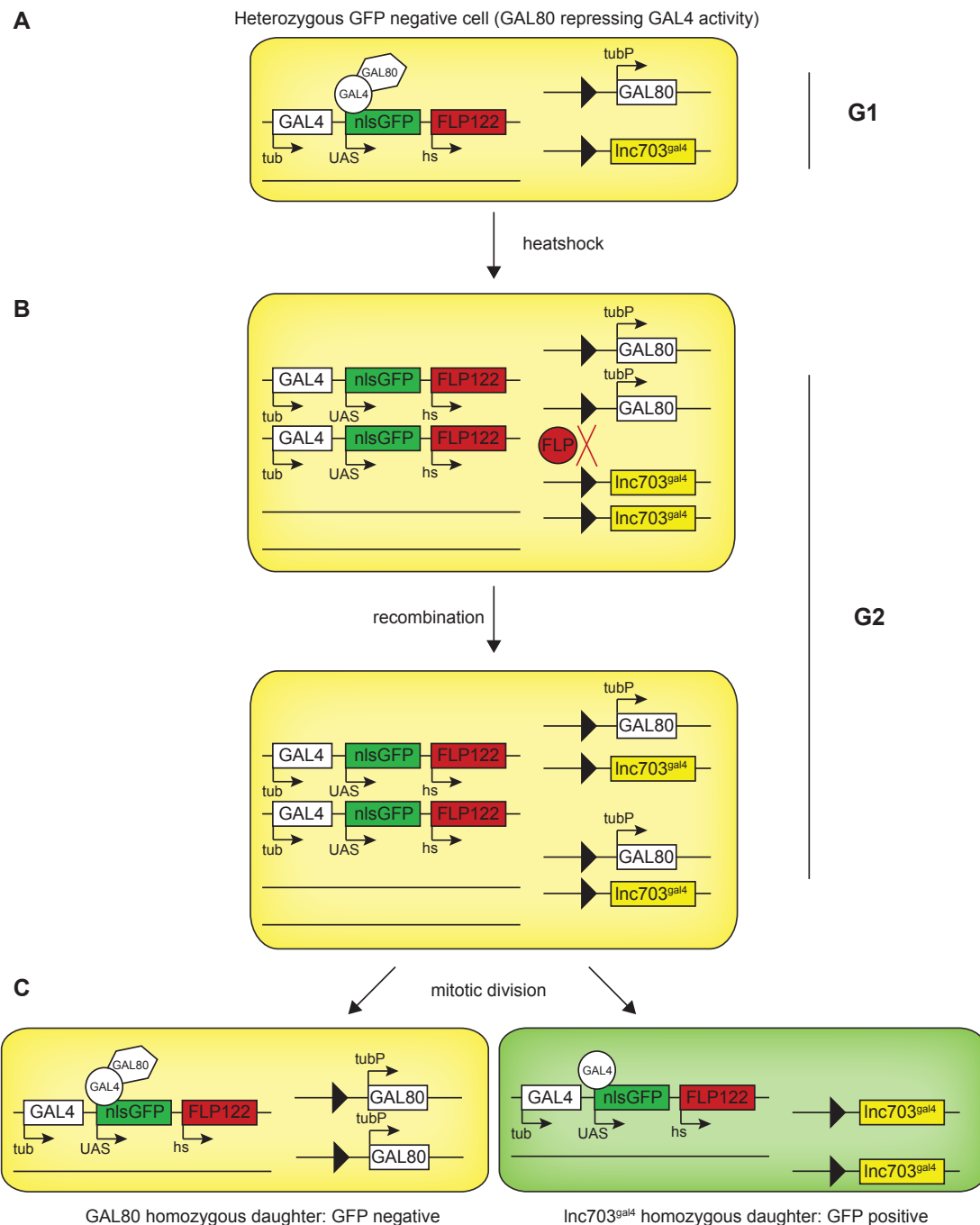


Figure 5.2 Schematic showing generation of GFP labelled *Inc703^{gal4}* homozygous clones using MARCM. (A) Before heatshock all cells are GFP negative as expression of GAL80 prevents GAL4 activating UAS-GFP expression. (B) Induction of FLP using heatshock causes recombination between FRT82B sites in cells at G2 phase. (C) After mitotic division cells homozygous for *Inc703^{gal4}* are GFP positive as UAS-GFP is driven by GAL4, whereas GAL4 and subsequently GFP expression continues to be repressed in cells homozygous for GAL80. Yellow cells are GFP negative. Black triangle = FRT82B site, red circle = FLP enzyme, white circle = GAL4 transcriptional activator, white polygon = GAL80, repressor of GAL4.

5.3 Results

5.3.1 Determining expression profiles using GAL4/UAS fluorescent reporters

Previous experiments had focussed on determining the role of *Inc703* during pupal development, driven by the pupal lethality of *Inc703*^{stop} mutants. Given the cause of lethality does not map to the *Inc703* locus and the absence of consistent changes to immune-related genes in *Inc703*^{gal4} pupae or adult females, it was necessary to determine which developmental stage and tissue would be most appropriate for further analysis. The *Piggybac*[GAL4D, EYFP] transposon in the *Inc703*^{gal4} strain was generated to allow mutagenesis as well as acting as an enhancer trap to determine the expression patterns of unknown genes (Hacker et al., 2003). The transposon is located within the first intron of *Inc703* and the GAL4 sequence is correctly orientated to act as an enhancer trap for sequences upstream of *Inc703*. I took advantage of this feature of the transposon to catalogue the tissue and cell type specificity of *Inc703* expression during larval development by crossing *Inc703*^{gal4} flies to flies containing either *UAS-mCD8::GFP* or *UAS-mCD8::GFP*, *UAS-H2B::mCherry* fluorescent reporters. Larval expression of *Inc703* was chosen as the focus because the *Inc703*^{stop} strain could still be used to validate discoveries made in the *Inc703*^{gal4} strain as the second site mutation is not lethal at this earlier stage. Given that *Inc703* levels are high in adult females (Figure 5.1), I also used a *UASp-RFP* reporter, which is active in the germline as well as somatic tissues to detect expression in the ovaries but no expression was observed.

5.3.2 mCherry reporter has a dynamic expression pattern in the larval CNS

Analysis of the larval CNS revealed a subset of mCherry positive cells; however repeat experiments showed the pattern was very variable during development making it difficult to conclude which cell type this corresponds to. However, partial

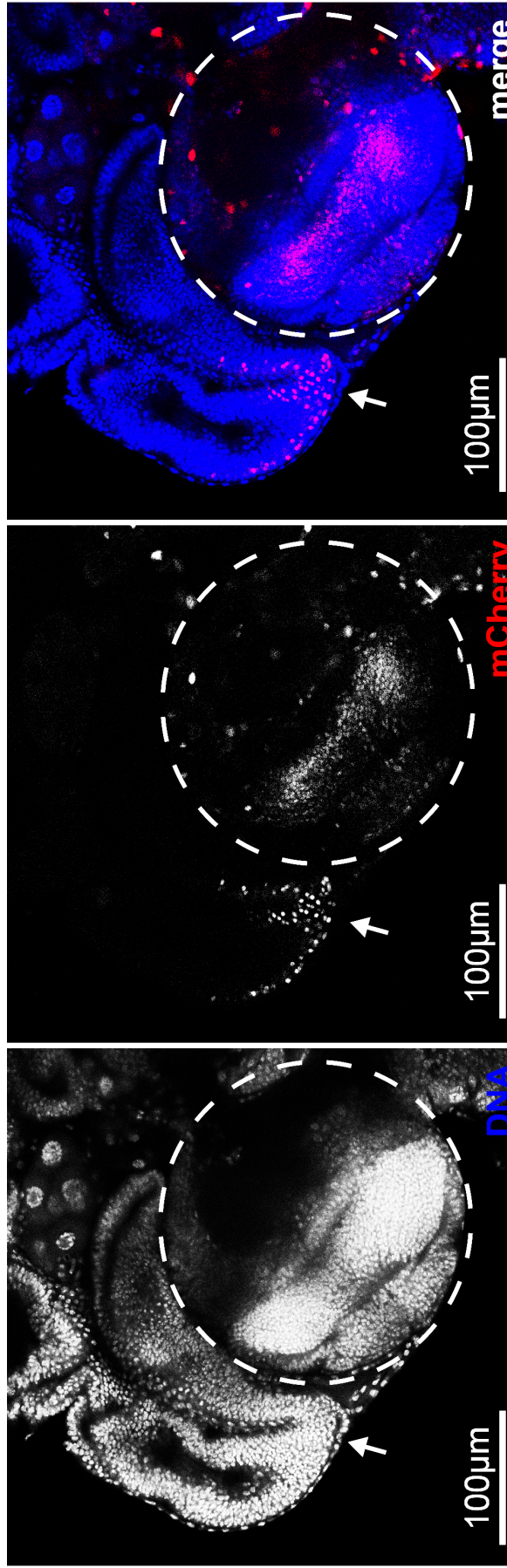


Figure 5.3 Larval CNS preparations with eye imaginal discs attached show mCherry expression. mCherry expressing third instar larval CNS with eye imaginal discs attached, stained with Hoechst (blue). Dashed line shows the boundary of the optic lobe. Arrow indicates area of eye imaginal disc in which mCherry is observed.

overlap was observed with Repo, a marker for glial cells, at some developmental stages. mCherry expression was also observed in the eye imaginal discs of CNS preparations from which the eye discs had not been removed (Figure 5.3). The eye and antennal imaginal discs are connected to the optic lobes of the larval brain via the optic stalk (Figure 5.4a). A wave of differentiation occurs from posterior to anterior of the disc which is visible as the morphogenetic furrow (MF) (Figure 5.4a). Posterior to the MF there is a basal layer of glia and an apical layer of neurons in clusters of five cells which will become the photoreceptors the adult eye (Figure 5.4b). A second round of mitosis occurs to generate the other photoreceptor and cone cells (Figure 5.4b).

5.3.3 Fluorescent reporters are expressed in a subset of photoreceptors in the eye imaginal discs

To ensure the signal observed in the eye discs was not due to EYFP fluorescence driven by the P3 promoter present as a marker for the transposon *Piggybac*[GAL4D, EYFP] which expresses in the larval light sensory organ (Hacker et al., 2003). Reporter expression in the eye discs was confirmed using constructs containing both *H2B::mCherry* and *mCD8::GFP* (Figure 5.4c) or *mCD8::GFP* alone (Figure 5.4d). Reporters were consistently observed in a subset of cells in the eye disc but not antennal discs. To further investigate the subpopulation of cells in which reporter expression was observed, eye discs were first co-stained for *cubitus interruptus* (Ci), a marker of the MF. mCherry was only observed in the differentiated cells, posterior to the MF (Figure 5.4e). Since, mCherry had previously been shown to colocalise with the glial cell marker Repo in the CNS, eye discs were stained using the Repo antibody. No cells in the eye disc were found to coexpress repo and mCherry (Figure 5.5a). mCherry expression was restricted to the apical layer of cells (Figure 5.5a) and Repo expression was confined to the basal layer of cells of the eye disc (Figure 5.5a). However, single

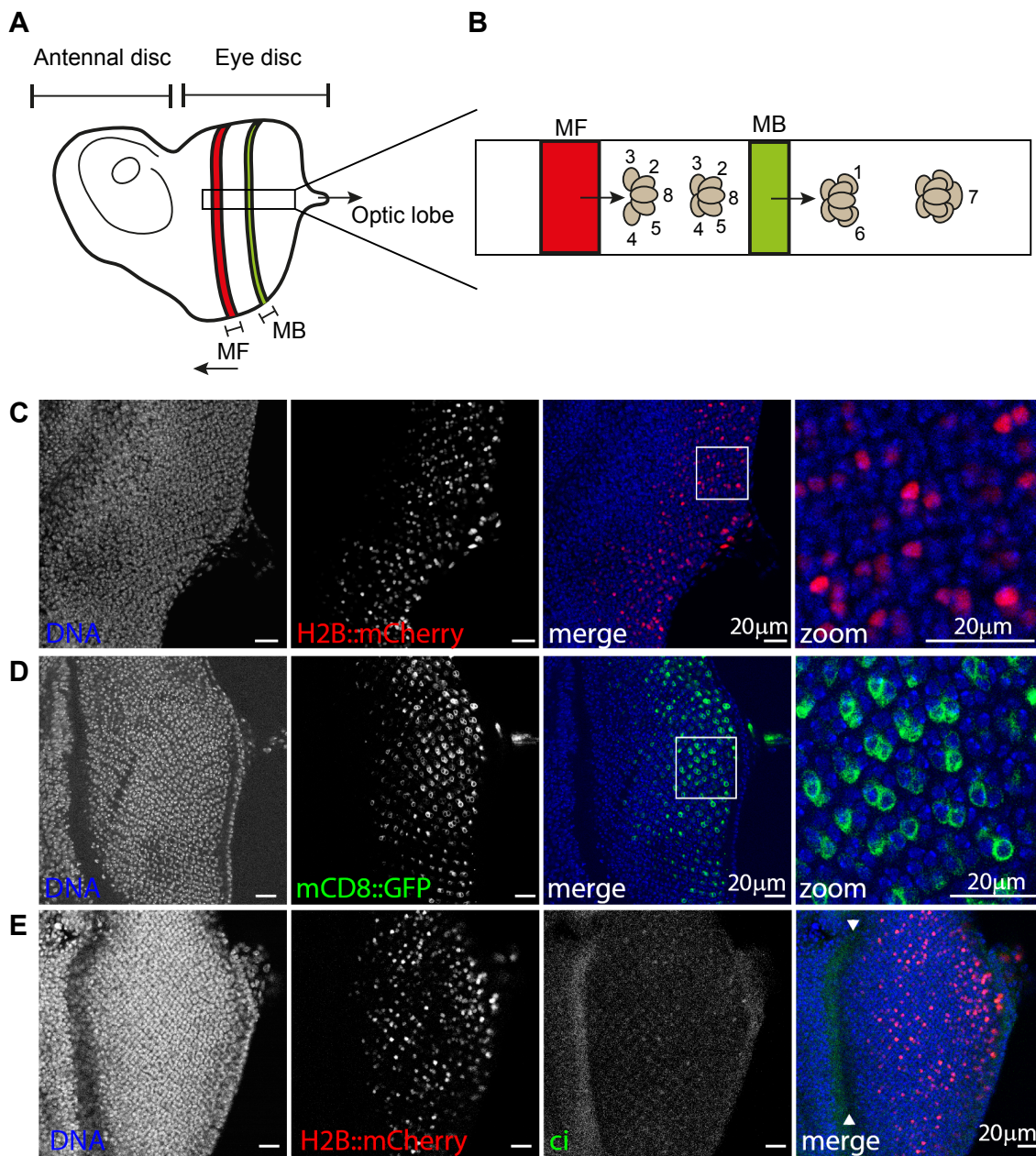


Figure 5.4 Reporter expression was observed in the eye imaginal disc using two different reporter lines. (A) Schematic of left eye imaginal disc and attached antennal disc. (B) Schematic showing development of photoreceptors within ommatidial clusters. MF = Morphogenetic Furrow, MB = Mitotic band. Adapted from Held (2002). (C) mCherry nuclei were observed at the posterior tip of eye imaginal discs. (D) An equivalent pattern was observed using membrane bound GFP. (E) All reporter positive nuclei are posterior to the morphogenetic furrow (MF) marked by Ci which is indicated with arrows.

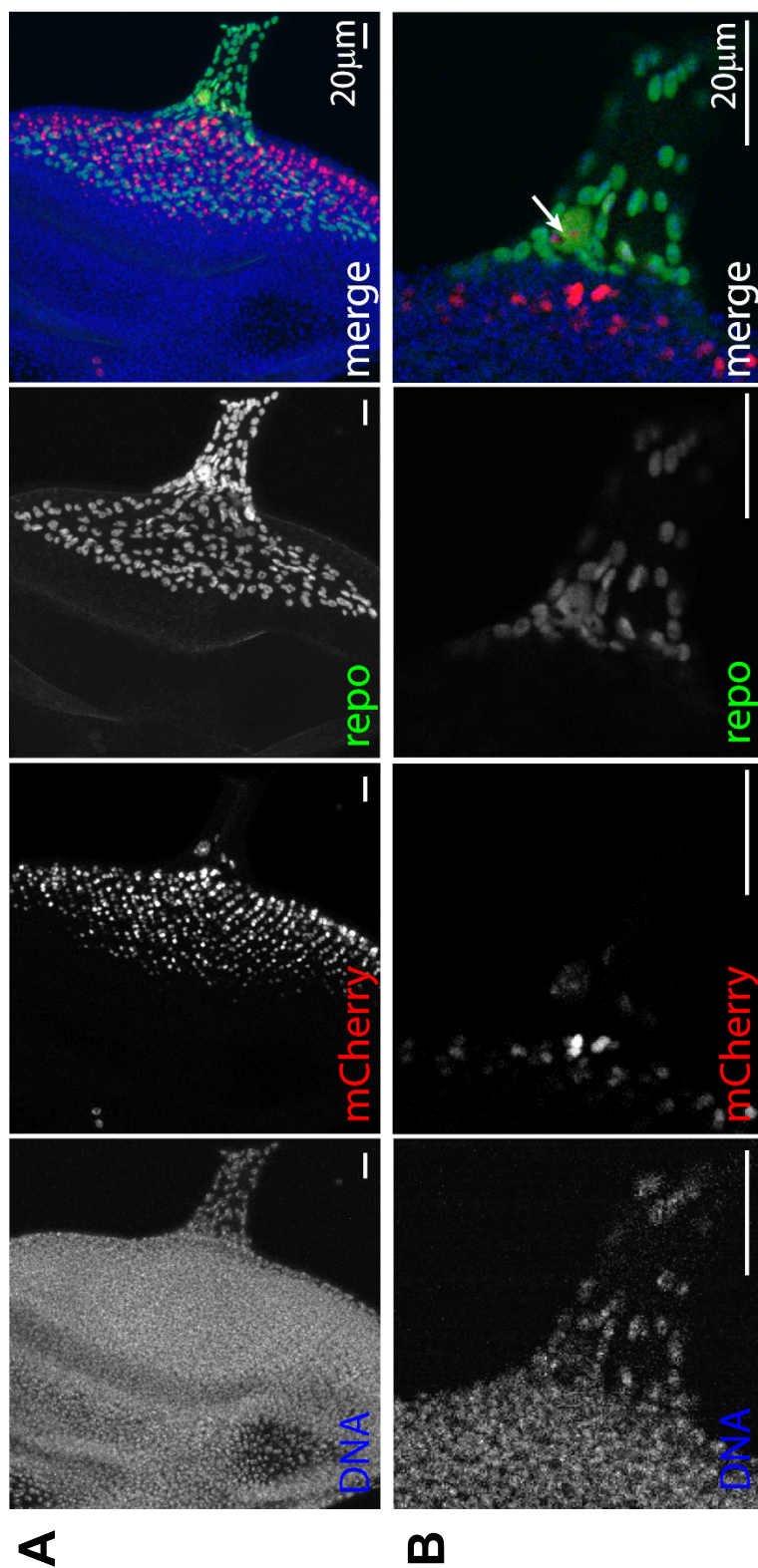


Figure 5.5 Reporter positive nuclei in the eye disc are not glia. (A) Z-projection of eye imaginal disc expressing mCherry under control of *Inc703gal4* and stained with anti-repo antibody, no cells in the eye disc were both mCherry and Repo positive. (B) Single plane image of optic stalk, arrow indicates a single large mCherry (red) and Repo (green) positive nucleus.

cells, positive for both repo and mCherry were observed in the optic stalk (Figure 5.5b, indicated with arrows), the size and location of these cells suggested they might be carpet glia (Silies et al., 2007).

The neuronal marker *embryonic lethal abnormal vision (elav)* is present in photoreceptor but not cone cells. All mCherry positive cells were also Elav positive (Figure 5.6a) indicating that they represent a subpopulation of photoreceptors. Markers available for specific photoreceptors include *prospero (pros)*, a marker of R8 photoreceptors, and *deadpan (dpn)*, a marker of R3/4 and R7 photoreceptors. No overlap was seen between mCherry and Pros (Figure 5.6b) or Dpn (Figure 5.6c). Together these results suggest that the mCherry positive cells are R1, R2, R5 or R6 photoreceptors. Given the presence of mCherry positive cells in ommatidial clusters containing five Elav positive nuclei (Figure 5.6a) which are R8, R2, R5, R3, and R4 photoreceptors (Figure 5.4b), mCherry positive cells are most likely to be R2 or R5 photoreceptors. However it was not possible to confirm this due to the absence of available antibody markers for R1/2/5/6.

5.3.4 Reporter expression is also observed in other larval tissues

Reporter expression was also detected in a set of unidentified cells in the cuticle of third instar larvae, as well as in a small number of cells in the midgut and in the larval fat body. All cells of the fat body expressed the mCherry or GFP reporter, but no reporter expression was observed in the embedded gonads (flanked by arrows) in males (Figure 5.7a) or females (Figure 5.7b).

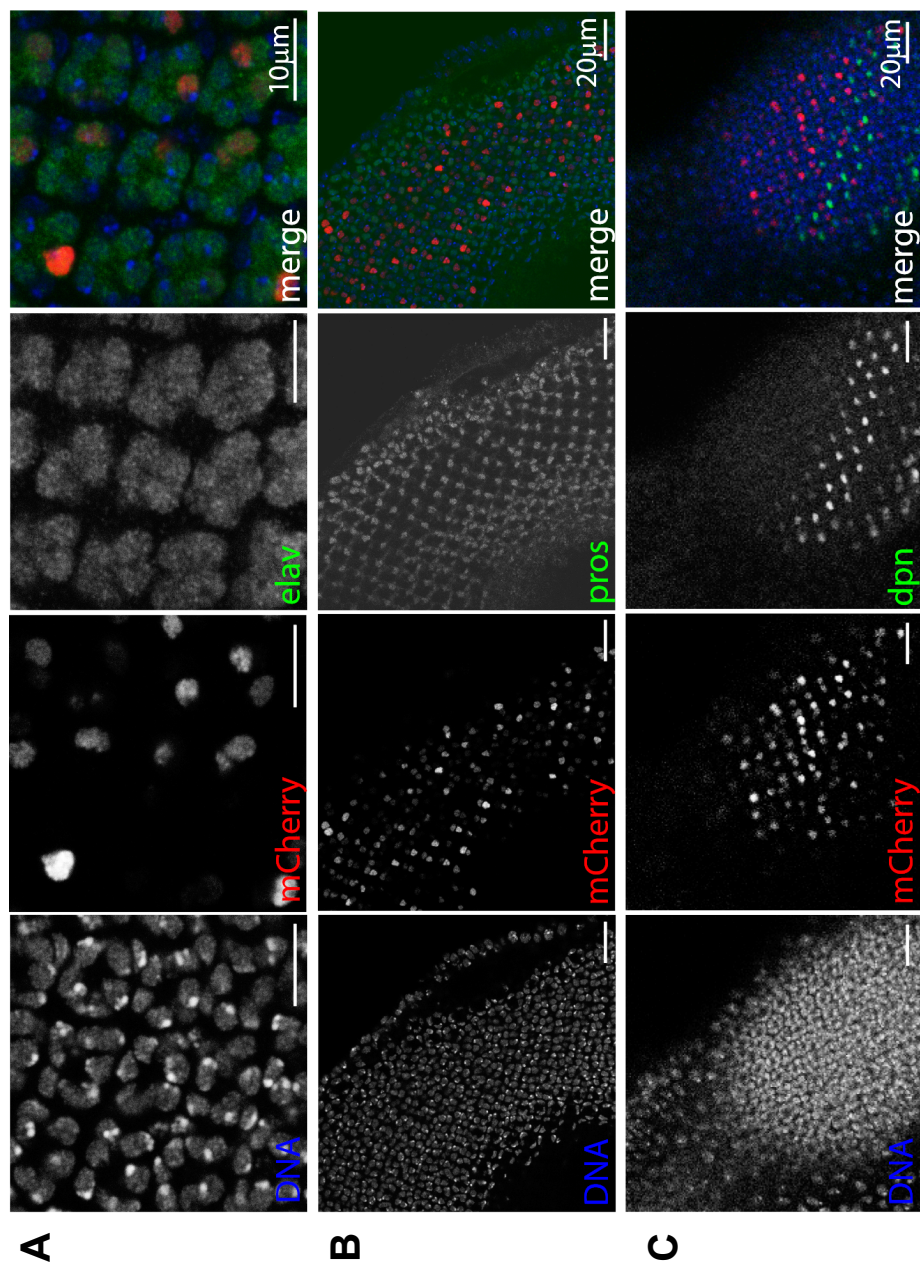


Figure 5.6 Reporter positive cells are an unidentified subset of photoreceptors. Single plane images of mCherry reporter eye discs stained with (A) anti-elav antibody that stains photoreceptors, one cell per ommatidial cluster is positive for mCherry (red) and Elav (green). (B) anti-pros antibody that stains R8 photoreceptors and cone cells, no cells are positive for mCherry (red) and Pros (green). (C) anti-dpn antibody that stains R3/4 and R7 photoreceptors, no cells are positive for mCherry (red) and Dpn (green).

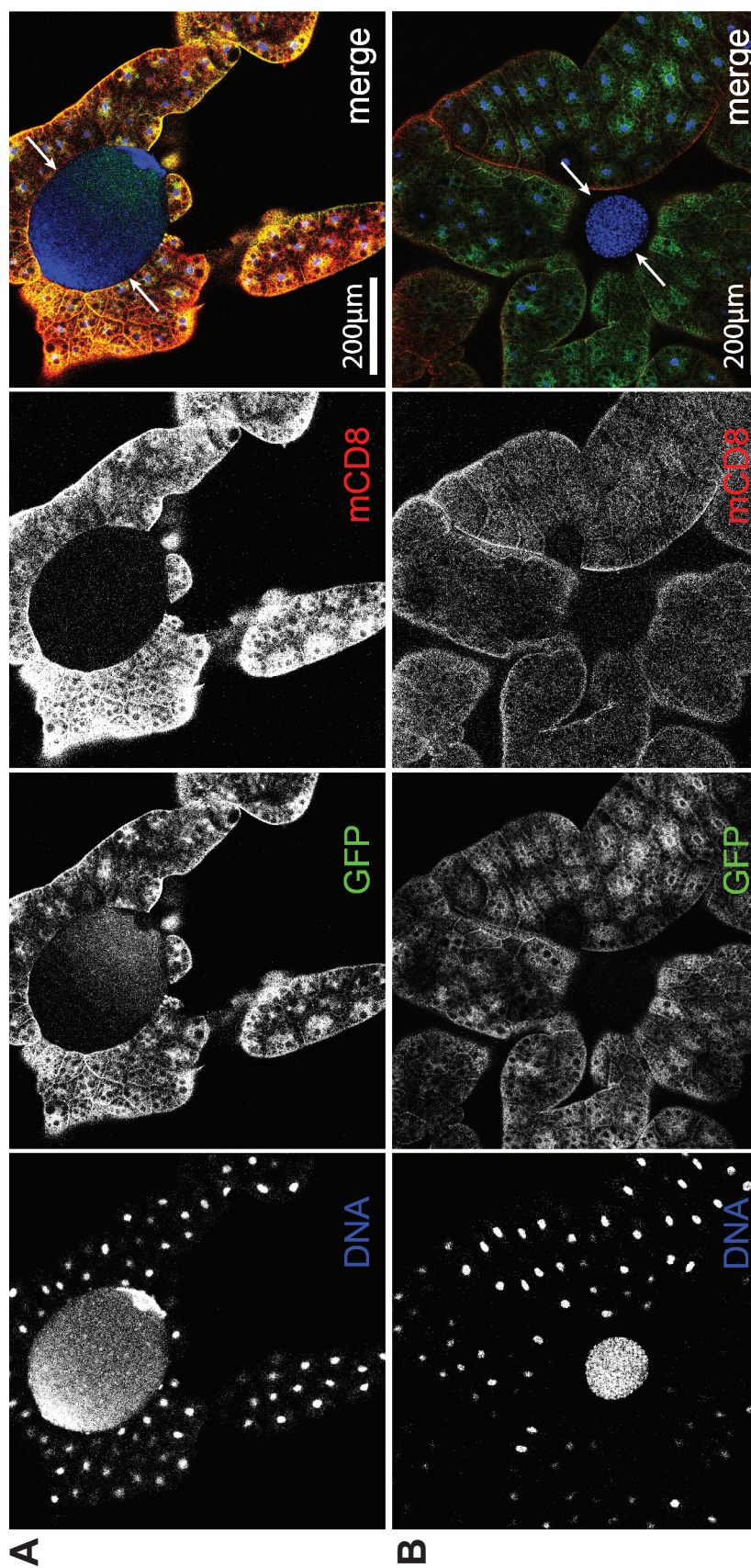


Figure 5.7 Larval fat body cells but not embedded gonads are positive for mCD8::GFP reporter in males and females. Single plane images third instar larval fat body from (A) males and (B) females, expressing mCD8::GFP (green) stained with Hoescht (blue) and anti-mCD8 antibody (red). Arrows indicate embedded gonads.

5.3.5 Confirmation of reporter expression by qRT-PCR

To corroborate the visualised expression profile, RNA was extracted from dissected tissues from larvae expressing *UAS-mCD8::GFP* under the control of one copy of the *Inc703^{gal4}* allele. The expression pattern of three genes with known tissue specificity were analysed compared to *actin5C* expression as a control (Figure 5.8a). As expected *Fat body protein 2 (Fbp2)* was detected in whole third instar larvae and highly enriched in the fat body and absent from other tissues. The carcass and imaginal discs were enriched for *papillote (pot)* and the CNS and to a lesser extent the eye imaginal discs were enriched for *repo*, confirming fidelity of the dissections.

Levels of *Inc703*, *mCD8* (the membrane targeting GFP conjugate) and *e(y)2b* were assessed in each tissue. mCD8 signal was detected in whole larvae and was very strongly enriched in the fat body (Figure 5.8b), consistent with the observed ubiquitous expression of GFP in the fat body (Figure 5.8c). Comparatively little signal was detected in other tissues. *Inc703* was also enriched in the fat body compared to whole larvae and showed lower levels in the other tissues (Figure 5.8b). However, the enrichment observed was far less than observed for mCD8. Levels of *Inc703* were unexpectedly similar in the eye and wing imaginal discs, whereas mCD8 transcript levels are very low in the wing disc (Figure 5.8b) and no GFP or anti-mCD8 signal was observed in the wing disc by confocal analysis (Figure 5.8c). It is possible that the *Inc703* is expressed in wing imaginal discs and the absence of mCD8::GFP is caused by the delay between *Inc703* expression and translation of mCD8::GFP protein. Levels of *e(y)2b* follow a similar expression profile to *Inc703*, however the enrichment in the fat body is relatively higher and more similar to mCD8 levels (Figure 5.8b).

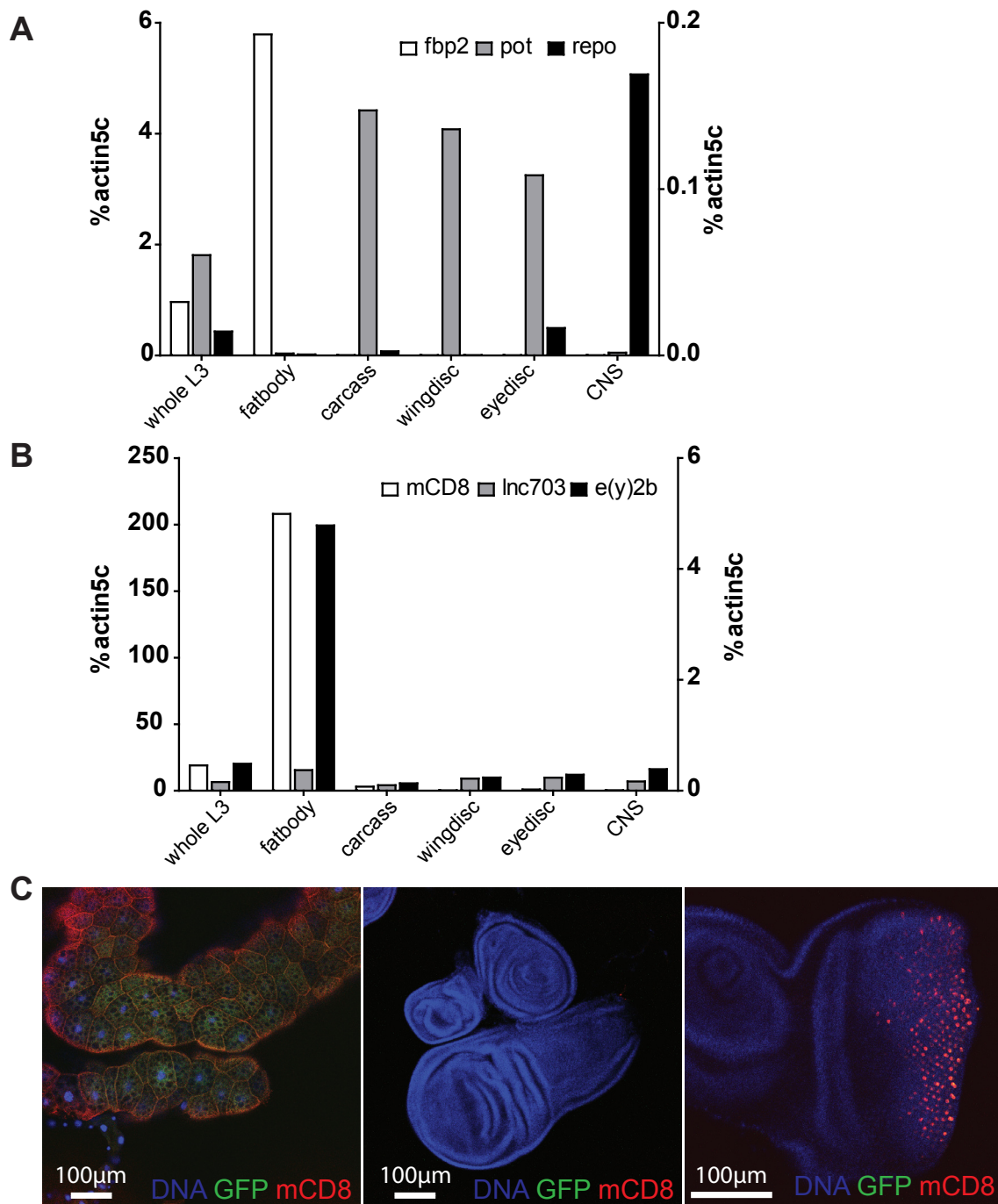


Figure 5.8 Validation of reporter expression profile confirms fat body expression of *Inc703*. (A) qRT-PCR results showing levels of markers for specific tissues types show the expected pattern. (B) qRT-PCR results from same samples showing enrichment of *mCD8*, *Inc703* and *e(y)2b* in the fat body. (C) Single plane confocal images of fat body, wing and eye imaginal discs of *UAS-mCD8::GFP; Inc703^{gal4}* third instar larva stained with Hoechst (blue) and anti-mCD8 (red).

5.3.6 *Inc703* is detected equally in the cytoplasmic and nuclear fractions of S2 cells

Subcellular localisation of lncRNAs can indicate their putative mode of action. Enrichment in the nucleus, particularly association with chromatin could indicate a lncRNA acts in *cis* to regulate transcription or binds to chromatin. Enrichment in the cytoplasm could suggest a *trans*-acting function of the lncRNA, or suggest that the lncRNA contains a translated ORF. Whilst the reporter system can identify cells that express *Inc703*, they cannot inform on the subcellular localisation of the transcript as the localisation of the reporter is determined by the protein that fluorescent protein is fused to.

FISH using probes complementary to *Inc703* showed non-specific signal in embryos (Personal Communication, A. Bassett). Therefore, the localisation of *Inc703* and two controls, with known localisation biases were assessed in fractionated S2 cells, in which *Inc703* is expressed at low but detectable levels (Figure 3.2). cDNA was generated from RNA extracted from either the cytoplasmic, salt extracted and chromatin fractions by Dr Charlotte Tibbit. I then used two controls to confirm the fidelity of the fractionation and as expected *GAPDH* was mostly highly represented in the cytoplasmic fraction and *roX2* enriched in the salt extract and chromatin fractions (Figure 5.9a). *Inc703* was detected equally in the cytoplasmic, salt extract and chromatin fractions (Figure 5.9b), suggesting that it does not have a specific subcellular localisation in S2 cells.

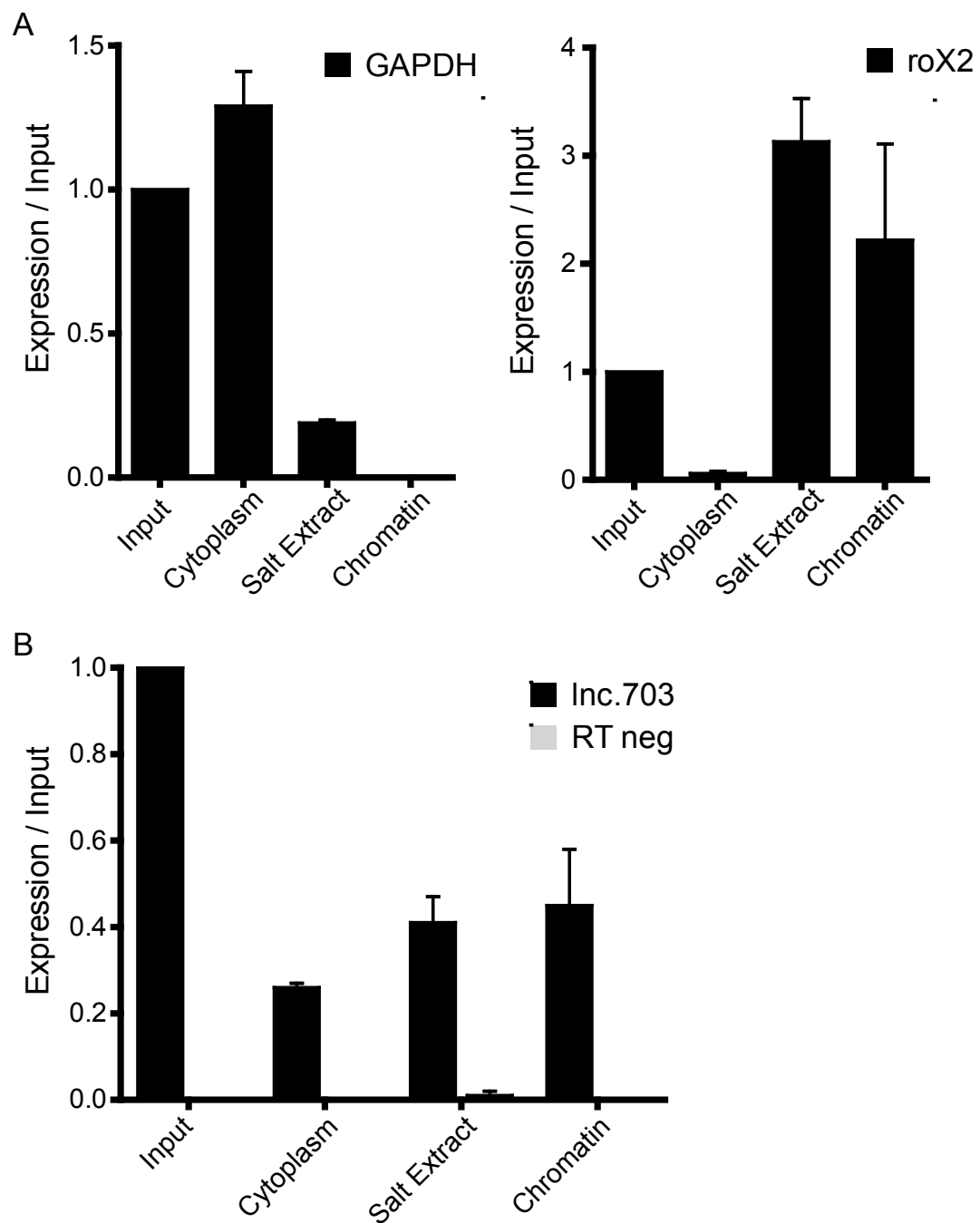


Figure 5.9 *Inc703* is detected equally in all subcellular fractions. (A) *GADPH*, (B) *roX2* and (C) *Inc703* levels in cytoplasmic, salt extracted and chromatin fractions of S2 cells as a proportion of the level in whole cells.

5.3.7 No alteration in TAG levels is observed in *Inc703* mutant larvae

Subsequent experiments will focus on the role of *Inc703* in the fat body because the presence of reporter signal in the fat body has been corroborated by enrichment of *Inc703* levels compared to whole larvae. Disruption of normal function of the larval fat body often manifests as a change in the level of triacylglyceride (TAG). TAG levels of dissected fat bodies from *Inc703^{gal4}* third instar larvae were compared to *w¹¹¹⁸* controls. No changes in TAG levels (normalised to protein content) were observed in males (Figure 5.10a) or females (Figure 5.10b). Although every effort was taken to ensure control and *Inc703^{gal4}* larvae were of equivalent stages, it is possible that differences in developmental timing between strains of different genotypes might confound results. I therefore used MARCM to generate clonal patches of mutant tissue in a wildtype background (Figure 5.10c and d) and assessed lipid droplet size as a proxy for lipid storage. No significant difference was observed in the lipid droplet size between mutant and wildtype cells in the fat bodies of male (Figure 5.10c') or females (Figure 5.10d').

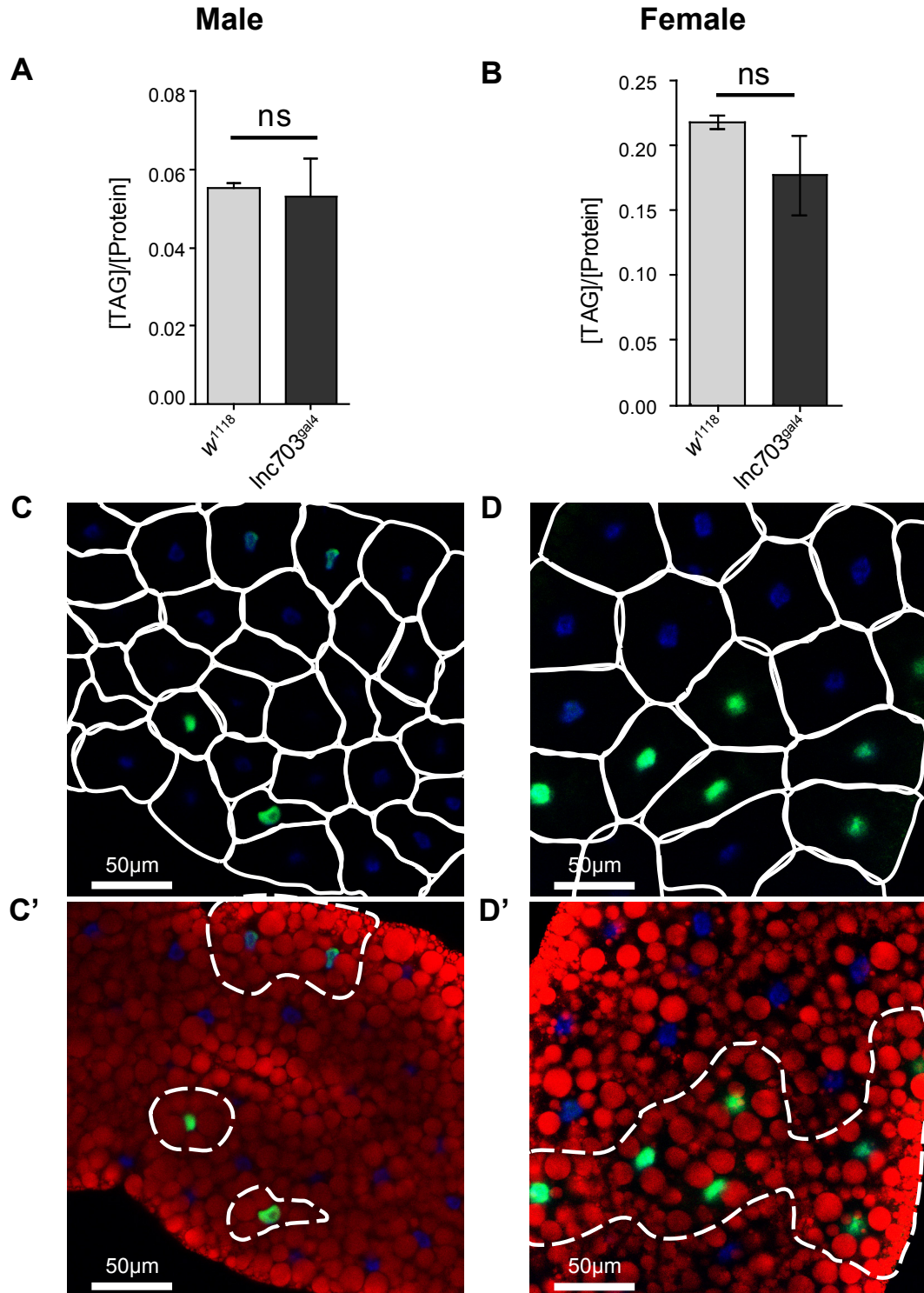


Figure 5.10 TAG levels and lipid droplet size in the fat body of *Inc703^{gal4}* mutant larvae. TAG levels normalised to protein content in larval fat body samples from (A) males and (B) females. (C and D) Images of larval fat body showing nuclei; cell boundaries have been highlighted in white. GFP positive cells are *Inc703^{gal4}* homozygous. (C' and D') Images showing lipid droplets stained with Nile red (red) with Hoechst (blue), nlsGFP is shown in green, *Inc703^{gal4}* mutant patches are outlined in dashed white lines.

5.3.8 Identification of genes differentially expressed in the fat body of isogenised *Inc703*^{gal4} larvae

The *Drosophila* fat body is considered analogous to the mammalian liver as well as being the major fat storage organ and controlling the immune response. In the absence of any gross changes in lipid storage, we used transcriptome profiling of the fat body to determine if any of the other functions of the fat body are disrupted in *Inc703*^{gal4} larvae. Samples were prepared and reads were mapped to the same transcript models used in Chapter 4. Differential gene expression was performed using DESeq2, making pairwise comparisons between *w*¹¹¹⁸ and each of the isogenised mutant strains (*Inc703*^{gal4}, *Inc703*^{stop} and BAC^{158D05}; *Inc703*^{gal4}).

Inc703 was not represented in the differential gene expression analysis as the read coverage of the gene did not meet the threshold, so loss of *Inc703* in *Inc703*^{gal4} and *Inc703*^{stop} samples and rescue of *Inc703* levels in BAC^{158D05}; *Inc703*^{gal4} samples was confirmed by qRT-PCR (Figure 5.11b). Loss of *Inc703* expression compared to *w*¹¹¹⁸ was observed in *Inc703*^{gal4} ($\log_2$ FC = -2.40, p = 0.0001) and *Inc703*^{stop} ($\log_2$ FC = -2.26, p = 0.0001) and comparable to control levels in BAC^{158D05}; *Inc703*^{gal4} ($\log_2$ FC = 1.10, p = 0.364).

117 genes (52 downregulated and 65 upregulated) were differentially expressed in the fat body of *Inc703*^{gal4} larvae compared to *w*¹¹¹⁸. Analysing all genes changed in *Inc703*^{gal4} using GO enrichment, a significant enrichment for genes involved in the biological process 'mitotic spindle elongation' (GO:0000022) was found (8 genes, p = 0.001). All eight of these genes were upregulated, so upregulated and downregulated genes were analysed separately. The enrichment for genes involved in mitotic spindle elongation was reproduced and an enrichment for genes involved in the biological process 'translation' (GO:0006412) was observed in upregulated genes (14 genes, p = 0.01). The enrichment for genes involved in translation was further supported by significant enrichment for the cellular

component term 'ribosomal subunit' (GO:0044391) (12 genes, $p = 1.7 \times 10^{-08}$) and the molecular function term 'structural constituent of ribosome' (GO:0003735) (12 genes, $p = 3.4 \times 10^{-05}$).

Differential gene expression analysis was also performed between w^{1118} controls and BAC^{158D05}; *Inc703*^{gal4} samples. If the BAC rescues *Inc703* mediated changes, it would be expected that very few genes would be differentially expressed, however a larger number of genes were differentially expressed than for *Inc703*^{gal4} alone, 557 in total (291 downregulated and 266 upregulated genes). To determine if these were the same genes that were changed in *Inc703*^{gal4} or independent changes caused by the BAC, I calculated the overlap between genes differentially expressed in both mutants (Figure 5.11a). Most genes differentially expressed in *Inc703*^{gal4} samples are changed in the same direction in BAC^{158D05}; *Inc703*^{gal4} samples (95 of 117 genes, $p < 0.0001$, Fisher's Exact Test). The correspondence between the two strains could indicate that the gene expression changes are due to loss of *Inc703* through a mechanism that is dependent on *Inc703* acting in its endogenous location or are caused by an element in the transposon insertion acting independently of *Inc703* levels, or finally mediated through increased levels of neighbouring gene *e(y)2b*.

5.3.9 BAC^{158D05} does not rescue *Inc703* expression levels in all tissues

The failure of BAC^{158D05} to rescue gene expression changes could be caused by the BAC not containing all the required regulatory elements to properly recapitulate endogenous *Inc703* expression or *Inc703* could function through the act of transcription either on a nearby gene not contained within the BAC, or by causing chromosomal looping. To test whether the expression pattern is comparable to endogenous wildtype levels, the expression of *Inc703* in BAC^{158D05}; *Inc703*^{gal4} were compared to controls at different time-points. Whilst *Inc703* levels are comparable to w^{1118} controls in BAC^{158D05}; *Inc703*^{gal4} larval fat body samples

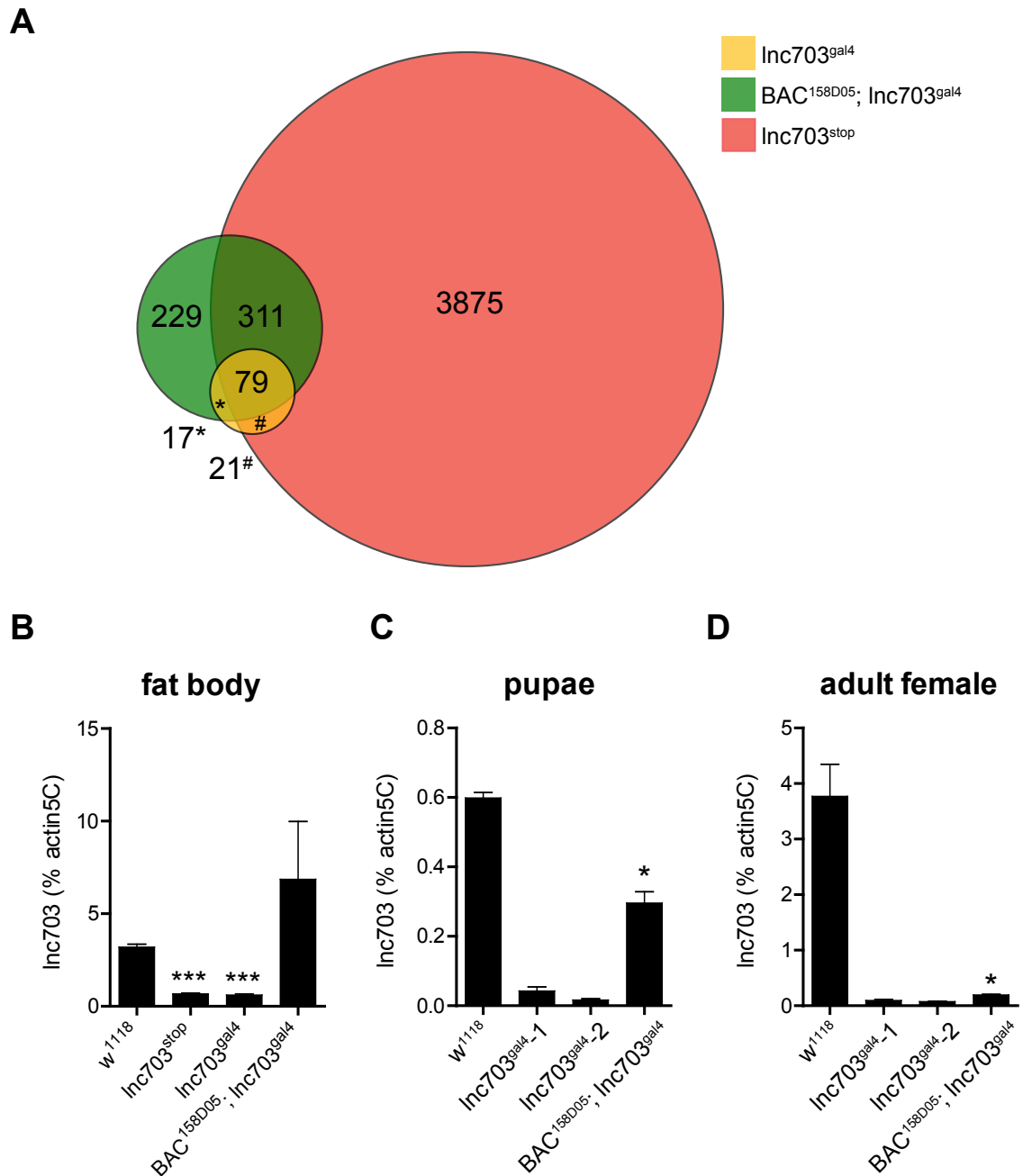


Figure 5.11 Overlap between the three mutant lines indicates common effects of *Inc703* are not rescued by BAC^{158D05}. (A) Significant overlap between differentially expressed genes is seen for each mutant, 79 genes are changed in all three mutants. (B) *Inc703* levels determined by qRT-PCR in the larval fat body as a proportion of *actin5C* levels in *w¹¹¹⁸*, *Inc703^{stop}*, *Inc703^{gal4}* and BAC^{158D05}; *Inc703^{gal4}*. (C and D) *Inc703* levels in (C) pupae one day APF and (D) two to four day old adult females in control, mutant and rescue. * $p < 0.05$, *** $p < 0.001$, two-tailed student T-test compared to *w¹¹¹⁸*.

(Figure 5.11b), only a partial rescue of levels is seen in pupae one day APF (Figure 5.11c) and no rescue is seen in adult females (Figure 5.11d). It is therefore possible that the gene expression changes observed in the fat body are non-cell autonomous effects of the loss of *Inc703* in another tissue, which is not rescued by the BAC. Furthermore, without direct visualisation of *Inc703* RNA, it is not possible to determine the contribution of the embedded gonads on the measured *Inc703* levels.

5.3.10 Gene expression changes in *Inc703*^{stop} correlate with *Inc703*^{gal4} and BAC^{158D05}; *Inc703*^{gal4}

Differentially expressed genes between *w*¹¹¹⁸ and *Inc703*^{stop} were also determined. A total of 4,286 genes were called as differentially expressed in *Inc703*^{stop} mutants (1,872 upregulated and 2,414 downregulated). The much higher number of differentially expressed genes could have been a result of the second site mutation in the *Inc703*^{stop} strain, so I analysed the intersection between *Inc703*^{stop} and *Inc703*^{gal4} to identify *Inc703* dependent changes. 100 of the 117 genes changed in *Inc703*^{gal4} were also changed in *Inc703*^{stop} (Figure 5.11a) ($p < 0.0001$, Chi-square). Of these a significant proportion were changed in the same direction in both mutants (89 genes, $p < 0.0001$, Fisher's exact) (Appendix F). A significant overlap was also found between *Inc703*^{stop} and BAC^{158D05}; *Inc703*^{gal4} (390 genes, $p < 0.0001$, Chi-square), within which a significant proportion changed in the same direction in both strains (349 genes, $p < 0.0001$ Fisher's exact). 79 genes were changed in the same direction in all three strains. Gene ontology enrichment analysis of the 25 genes downregulated in all strains found no enrichment. Analysis of the 48 upregulated genes, found enrichment for genes involved in the biological processes mitotic spindle elongation (GO:0000022) (6 genes, $p = 8.70 \times 10^{-4}$) and translation (GO:0006412) (12 genes, $p = 0.013$).

The failure of BAC^{158D05} to rescue gene expression changes could suggest that

the gene expression changes observed were caused by increased *e(y)2b* levels. Particularly considering the greater increase in *e(y)2b* levels in the BAC containing strain than *Inc703^{gal4}* alone ($\log_2$ FC = 0.694, p = 0.364; $\log_2$ FC = 1.32, p = 4.24×10^{-04} ; *Inc703^{gal4}* and BAC^{158D05}; *Inc703^{gal4}*, respectively) corresponds to a greater number of genes being differentially expressed. However, in whole *Inc703^{stop}* larvae, levels of *e(y)2b* were reduced (Figure 3.6i) and *e(y)2b* levels are also significantly decreased in the fat body ($\log_2$ FC = -1.26, p = 1.58×10^{-04}). Given that the same upregulation of genes involved in translation are seen when *e(y)2b* levels are reduced in *Inc703^{stop}* as in *Inc703^{gal4}* and BAC^{158D05}; *Inc703^{gal4}* when *e(y)2b* levels are increased, it is likely that this effect is independent of *e(y)2b* levels.

5.3.11 Gene expression changes in both *Inc703^{gal4}* and *Inc703^{stop}* are confirmed by qRT-PCR

The expression level changes of a subset of genes changed in the same direction in both *Inc703^{gal4}* and *Inc703^{stop}* samples including those changed in all three mutants were confirmed by qRT-PCR. All genes were changed in the same direction as calculated by DESeq2 for all genes tested excluding *e(y)2b* which significantly downregulated in *Inc703^{stop}* according to DESeq2 but was not significantly changed by qRT-PCR (Figure 5.12a). The magnitude of the change was generally higher for the same genes in *Inc703^{stop}* than *Inc703^{gal4}*. Using the RNA-seq data, and excluding all genes with an absolute $\log_2$ fold change less than 0.5 in either mutant, positive correlation is observed between the $\log_2$ fold changes between the two mutant alleles (Figure 5.12b) (r^2 = 0.228, p < 2.2×10^{-16} , Pearson's correlation). Indicating that the same genes are changed, but to a lesser extent in *Inc703^{gal4}* suggesting that *Inc703^{stop}* may be a stronger allele.

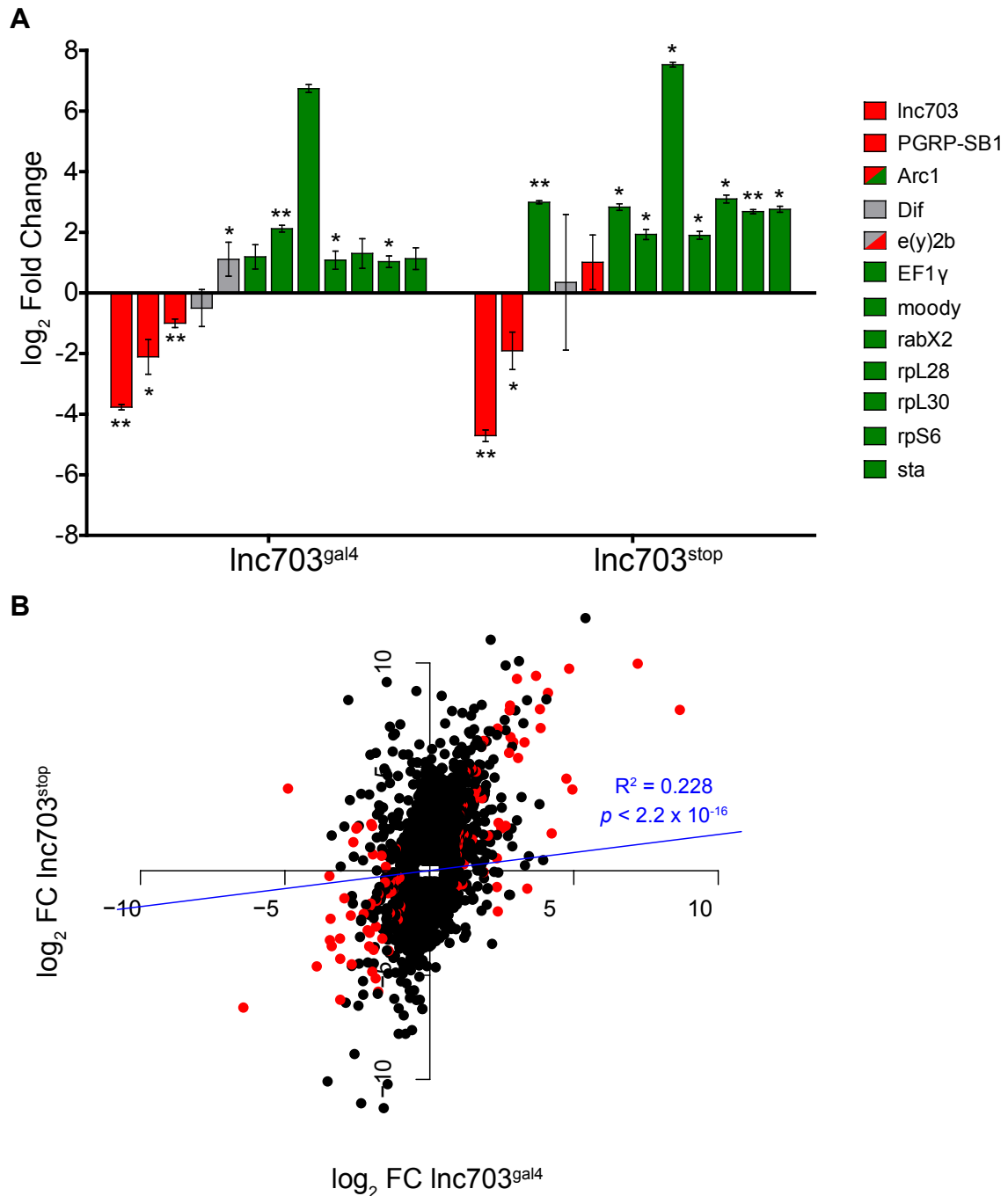


Figure 5.12 Gene expression changes in *Inc703*^{gal4} and *Inc703*^{stop} are confirmed by qRT-PCR. (A) Log₂ fold change of expression level in *Inc703*^{gal4} and *Inc703*^{stop} fat body samples compared to *w*¹¹¹⁸ controls determined by qRT-PCR. * *p* < 0.05; ** *p* < 0.01, two-tailed student t-test. (B) Positive correlation between log₂ fold changes compared to *w*¹¹¹⁸ of *Inc703*^{gal4} and *Inc703*^{stop} alleles. Red dots represent genes significantly changed in *Inc703*^{gal4}.

5.3.12 Ribosomal protein genes may also be increased at the protein level

Of the ribosomal proteins with increased expression levels in *Inc703^{gal4}*, *BAC^{158D05}*; *Inc703^{gal4}* and *Inc703^{stop}*, *rpS6* had an available antibody to test whether upregulation of the transcript levels results in increased protein abundance. RpS6 levels are slightly elevated compared to the loading control in protein extracts from fat bodies from both isogenised *Inc703^{gal4}* strains, and further increased in *BAC^{158D05}*; *Inc703^{gal4}* samples (Figure 5.13a and b) in which the increase in transcript levels is also higher. However, this experiment needs repeating to confirm the increase in RpS6 abundance.

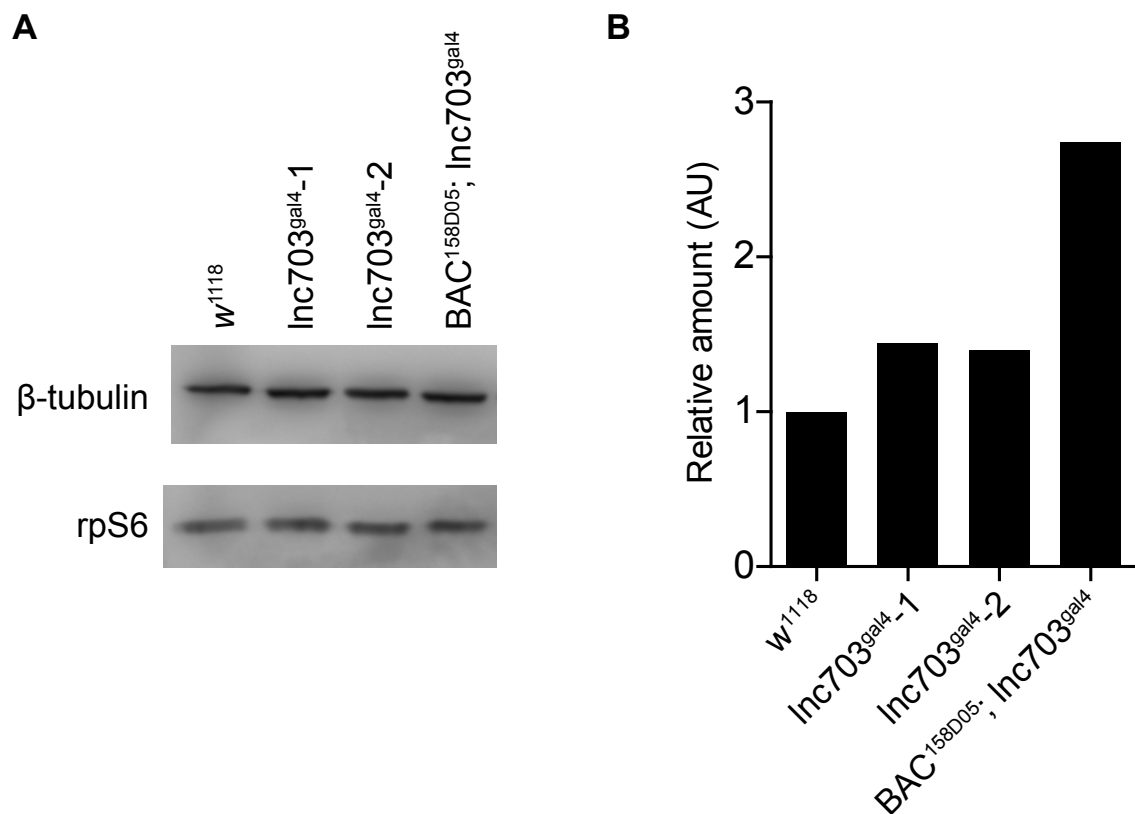


Figure 5.13 RpS6 levels are increased compared to β -tubulin loading controls in mutant and rescue fat body extracts. (A) Single bands are detected at ~57kDa and ~30kDa as expected for β -tubulin and RpS6 respectively. (B) Quantification of relative density of RpS6 and control bands indicate increased levels of RpS6 in mutant and rescue samples. N = 1.

5.3.13 Lifespan

Given the elevated levels of ribosomal protein genes in the fat body of *Inc703* mutants and the known interaction between ribosomal levels and lifespan (Houtkooper et al., 2013; Steffen et al., 2008), the lifespan of two independently isogenised *Inc703^{gal4}* strains was compared to *w¹¹¹⁸* controls (Figure 5.14). It was not possible to assess the lifespan of *Inc703^{stop}* flies as they die during pupal development. The median lifespan of one isogenised line was decreased in both males (58 days) and females (54 days) compared to controls (78 days and 66 days, respectively) ($p = 5.38 \times 10^{-16}$, 2.12×10^{-05} , Wilcoxon test, $n \geq 150$). The discrepancy between the two strains suggests the effect on lifespan is influenced by genetic background. The median lifespan of male but not female *Inc703^{gal4}* flies containing two copies of the BAC^{158D05} rescue construct was further reduced below levels observed for *Inc703^{gal4}* (38 days, $p < 2.2 \times 10^{-16}$, Wilcoxon test).

Given that BAC^{158D05}; *Inc703^{gal4}* rescue strains show elevated levels of both *Inc703* and neighbouring gene *e(y)2b*, the lifespan of flies overexpressing only *Inc703* or *e(y)2b* was assessed. Transgenic flies containing UAS-driven *Inc703* cDNA transgenes were crossed to *actin5C-GAL4* and *1032-GAL4* drivers to achieve ubiquitous overexpression. In parallel, male flies containing the P[EP]G3791 element containing UAS sites correctly orientated upstream of *e(y)2b* to allow overexpression from the endogenous locus were crossed to both driver lines. Progeny were collected, mated for 48 hours and then separated by sex.

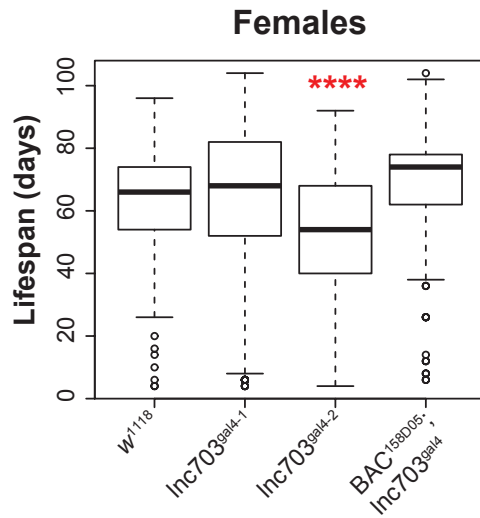
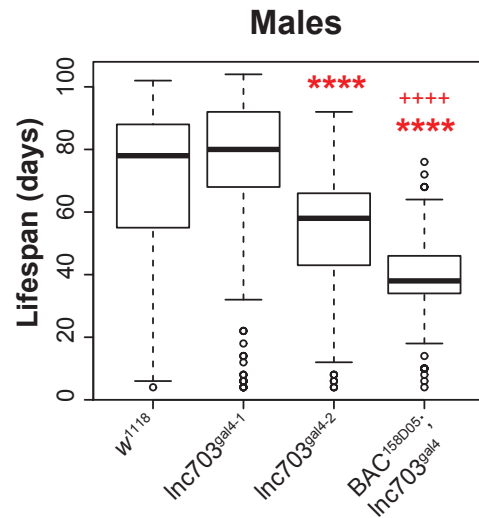
A**B**

Figure 5.14 Median lifespan of *w¹¹¹⁸* control, isogenised *Inc703^{gal4}* and *BAC^{158D05}; *Inc703^{gal4}* flies.* Lifespan of control, mutant and rescue (A) female and (B) male flies. **** $p < 0.0001$, Wilcoxon Sign Rank Test compared to *w¹¹¹⁸*; ++++ $p < 0.0001$, Wilcoxon Sign Rank Test compared to *Inc703^{gal4-2}*.

5.3.14 *Inc703* and *e(y)2b* are independently overexpressed in UAS-driven transgenic strains

qRT-PCR analysis of four day old mated flies showed significantly increased *Inc703* levels in all *Inc703* cDNA transgenic overexpression strains (Figure 5.15a and c). The largest increase in upregulation of *Inc703* was observed in 1032-GAL4 males ($\log_2$ FC = 4.05 and 4.49 for *UAS-Inc703-A* and *UAS-Inc703-B* respectively), the level of upregulation in 1032-GAL4 females is much lower ($\log_2$ FC = 1.11 and 1.01, *UAS-Inc703-A* and *UAS-Inc703-B*, respectively). *actin5C*-GAL4 driven overexpression is more equal between females ($\log_2$ FC = 1.40 and 1.88, *UAS-Inc703-A* and *UAS-Inc703-B*, respectively) and males ($\log_2$ FC = 2.58 and 2.40, *UAS-Inc703-A* and *UAS-Inc703-B*, respectively). *Inc703* levels are also changed in males in both *e(y)2b* overexpression strains. 1032-GAL4 driven overexpression of *e(y)2b* results in a small upregulation of *Inc703* ($\log_2$ FC = 0.71, $p = 0.03$) (Figure 5.15a) whereas *actin5C*-GAL4 driven overexpression results in downregulation of *Inc703* ($\log_2$ FC = -2.08, $p < 0.0001$) (Figure 5.15c). The inconsistency suggests the changes are independent of *e(y)2b* levels and may be caused by the insertion disrupting regulatory sequences for *Inc703* or by binding of GAL4 upstream of *Inc703* affecting its expression level.

Significant increases in *e(y)2b* expression was detected in flies containing the P[EP] element and either 1032-GAL4 (Figure 5.15b) or *actin5C*-GAL4 (Figure 5.15d). *e(y)2b* levels are also changed in males overexpressing both isoforms of *Inc703* (Figure 5.15b and d). Although statistically significant in three of the four *Inc703* overexpression strains, the changes are small (< 1.5-fold) and are not consistent in direction and may reflect innate variability in *e(y)2b* levels.

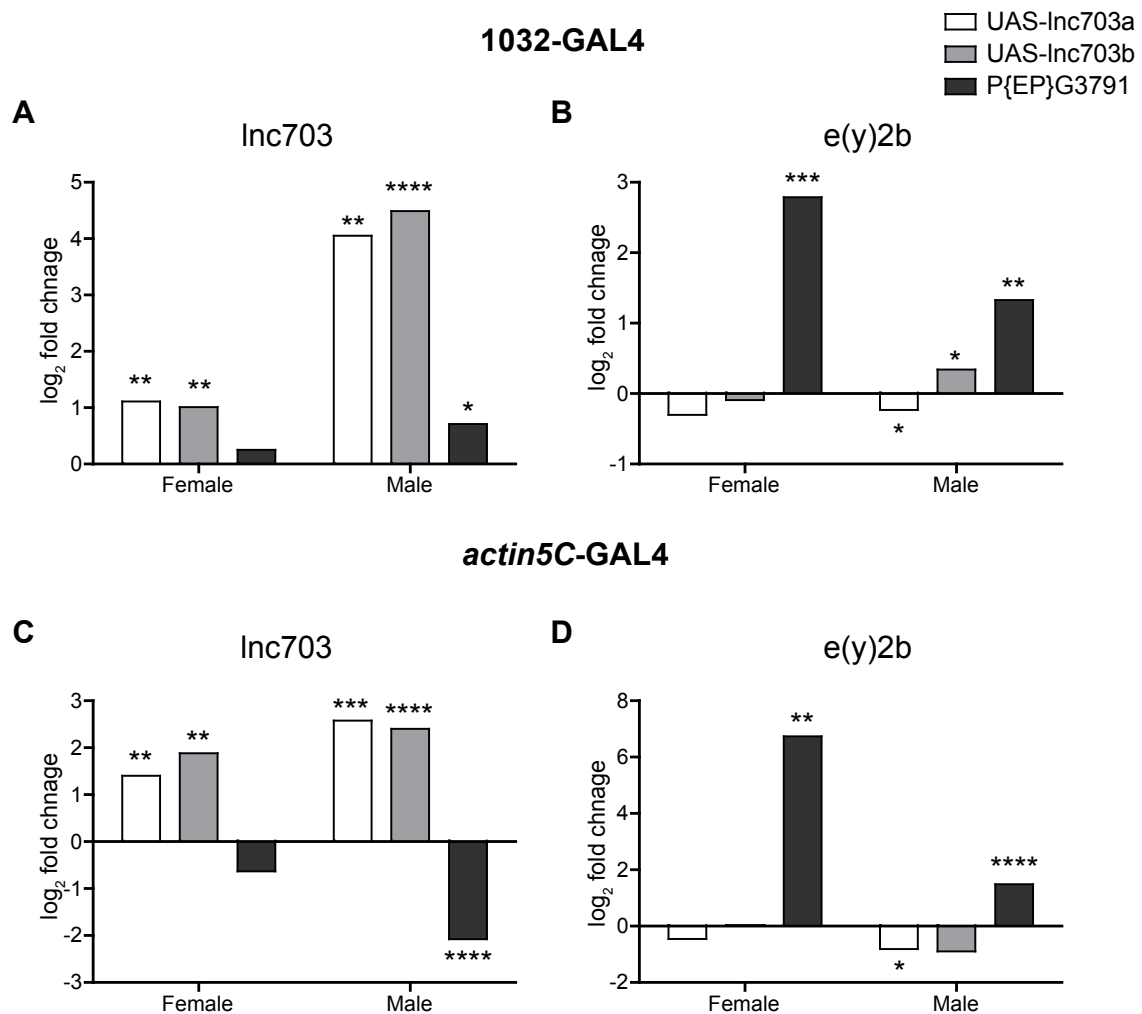


Figure 5.15 Transgenic *Inc703* and *e(y)2b* overexpression in two to four day old adult flies was confirmed by qRT-PCR. (A and B) Transgenes are driven by ubiquitous 1032-GAL4 driver. (C and D) Transgenes are driven by ubiquitous *actin5C*-GAL4 driver. (A and C) *Inc703* levels are significantly increased in males and females of both *Inc703* overexpression lines. (B and D) *e(y)2b* levels are significantly increased in males and females containing the P{EP}G3791 element. Significance is shown relative to *mCD8::GFP* overexpression * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, two-tailed student T-test.

5.3.15 Lifespan of transgenic overexpression strains

It would be expected that if the decreased lifespan of *Inc703^{gal4}-2* and BAC^{158D05}; *Inc703^{gal4}* males was due to increased levels of *e(y)2b*, males overexpressing *e(y)2b* under the control of 1032-GAL4 or *actin5C*-GAL4 would also have decreased lifespan. The median lifespan of males overexpressing *e(y)2b* (83 days) or GFP (73 days) under the control of 1032-GAL4 was not significantly different (Figure 5.16a) ($\text{adj}P = 0.101$, Wilcoxon test, Benjamini-Hochberg correction). Similar results were obtained using the *actin5C*-GAL4 driver (Figure 5.16c) (85 days, 80 days, $\text{adj}P = 0.226$, *e(y)2b* and GFP overexpression respectively). Conversely, *actin5C*-GAL4 driven *e(y)2b* expression increased lifespan compared to GFP overexpression in females (Figure 5.16d) (116 days, $\text{adj}P = 1.95 \times 10^{-13}$).

Consistent with the reduction in lifespan of *Inc703^{gal4}-2* males and females that have reduced *Inc703* levels, overexpression of either isoform of *Inc703* under the control of *actin5C*-GAL4 resulted in an increased lifespan compared to GFP overexpression in males (Figure 5.16c) (101 days, 99 days, 80 days, $\text{adj}P = 0.006$, $\text{adj}P = 0.002$, *Inc703-A*, *Inc703-B* and GFP, respectively) and females (Figure 5.16d) (125 days, 131 days, 94 days, $\text{adj}P = 2.79 \times 10^{-13}$, $\text{adj}P = 2.79 \times 10^{-13}$, *Inc703-A*, *Inc703-B* and GFP, respectively). Overexpression of either *Inc703* isoform under the control of 1032-GAL4 did not significantly affect lifespan in males or females (Figure 5.16a and b). The effect using *actin5C*-GAL4 but not 1032-GAL4 could be caused by the different expression profiles of the two drivers, alternatively the changes in lifespan observed could be the result of differences in the genetic background in these two lines.

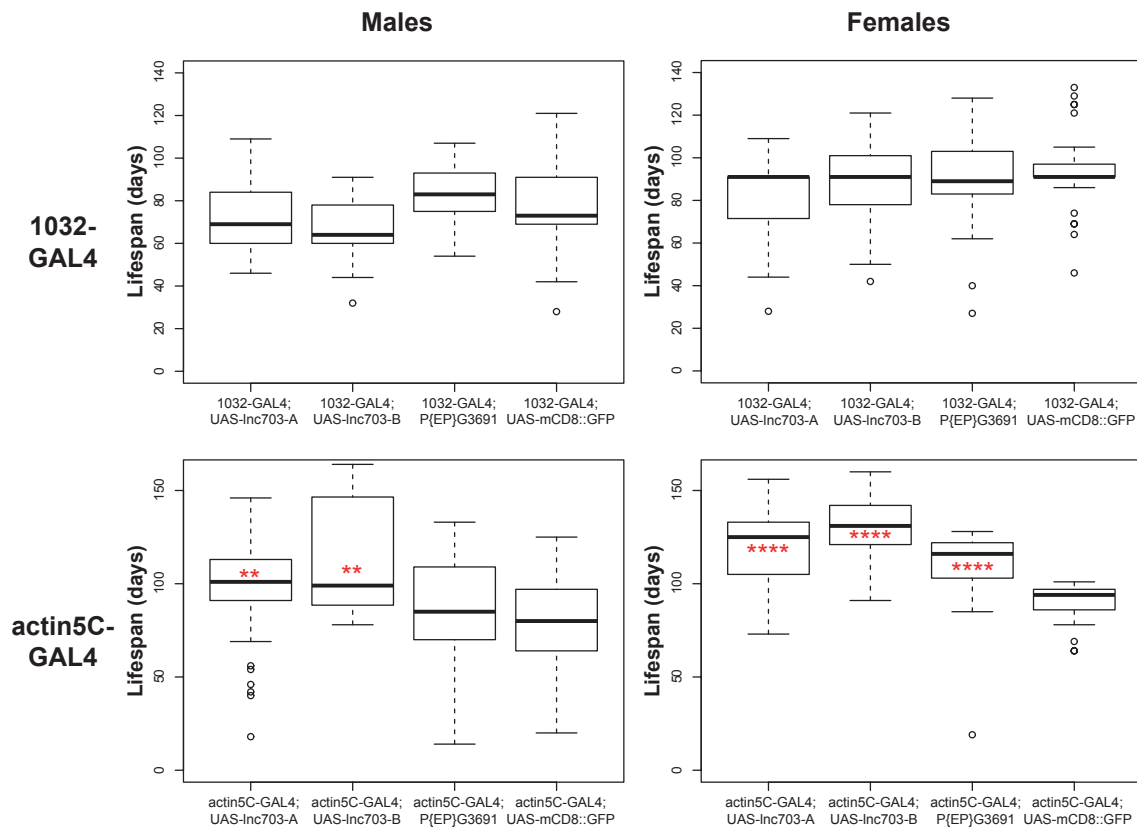


Figure 5.16 Median lifespan of *Inc703* or *e(y)2b* overexpressing flies. (A and B) Lifespan of males (A) and females (B) overexpressing *Inc703*, *e(y)2b* or *mCD8::GFP* under the control of 1032-GAL4. (C and D) Lifespan of males (C) and females (D) overexpressing *Inc703*, *e(y)2b* or *mCD8::GFP* under the control of *actin5C-GAL4*. ** $p < 0.01$, **** $p < 0.0001$, Wilcoxon Sign Rank Test with Benjamini-Hochberg multiple test correction.

5.4 Discussion

5.4.1 Visualising RNA *in vivo*

Two types of information are required in order to design the most appropriate strategy to study loss of *Inc703 in vivo*. Firstly, it is important to understand in which cell type or tissue the lncRNA is mostly likely to act. The availability of transcriptome data from different developmental stages and tissues (Graveley et al., 2011) is the simplest way to identify areas of interest by identifying when or where *Inc703* is most highly expressed. Although comprehensive, the available data is not exhaustive but can be used to guide more detailed analysis. Secondly, the strategy should reflect the mechanism of action of the lncRNAs which can be predicted by information including, cellular localisation, protein, RNA and DNA interactions.

The most direct method for visualisation of RNA in tissues is Fluorescent *In Situ* Hybridisation (FISH), using oligonucleotide probes complementary to the RNA of interest containing fluorescently labelled nucleotides. FISH can help to determine both the cell type specificity and subcellular localisation of lncRNAs. However, in the case of lncRNAs, FISH may be problematic due to the generally lower expression levels of lncRNAs. It may be possible to circumvent the issue of low expression levels by performing In Situ PCR (Personal Communication, Dr M. Alló). Alternatively, ribonucleoprotein (RNP) complexes may protect lncRNAs, making them inaccessible for probe binding. Other strategies include the generation of transgenic flies containing the RNA of interest tagged by natural bacteriophage sequences MS2 (Gong et al., 2012) or PP7 (Hogg and Goff, 2010), or synthetic sequences such as Streptavidin (S1) aptamers. However, little enrichment for *Inc703* was achieved by PP7 or MS2 pull-down of *in vitro* transcribed *Inc703* (Personal Communication, A. Bassett), suggesting that this strategy is unlikely to work in this context.

Recently a number of techniques have been developed to identify the interaction partners of RNA moieties. RNA-DNA interactions can be detected for specific RNAs using capture hybridization analysis of RNA targets (CHART) or Chromatin isolation by RNA purification (ChIRP) which involve pulling down RNA using biotinylated probes and detecting associated DNA using qRT-PCR or DNA sequencing. Whilst these techniques are established in human, mammalian and insect cell lines (Chu et al., 2011; Simon et al., 2011), their application *in vivo* is as yet unproven.

Whilst informative, the strategy of using a GAL4/UAS-driven fluorescent reporter utilised in this study has some caveats. Firstly, the basal promoters used in UAS constructs are known to have 'leaky' expression influenced by their genomic context and may not be equally active in different cell types. The impact of this was reduced by the use of two different reporters on different chromosomes. Secondly, no information about the subcellular localisation of *Inc703* can be inferred as the localisation of the fluorescent reporters was dictated by the fluorescent fusion protein that was expressed. This was localised to the membrane in the case of mCD8::GFP and the nucleus in the case of H2B::mCherry. Thirdly, the reporter system requires two rounds of translation, firstly of the GAL4 transcriptional activator and secondly of the fluorescent protein. This will lead to a delay between transcription of *Inc703* and the appearance of the fluorescent marker, which might explain the absence of GFP signal in wing disc despite low levels of *Inc703* being detected in wing discs by qRT-PCR (Figure 5.8). Finally, differences in the translation efficacy of different proteins and in different cell types could affect the levels of different reporters, and apparent relative levels in different tissues. The persistence of the fluorescent signal will also differ in different cell types and between different fluorescent markers. This is exemplified by the level of mCD8 transcript detected in the larval fat body being disproportionately higher compared to other tissues than *Inc703* levels. Therefore the reporter system was only used

as a guide to identify cells that were expressing or had recently expressed *Inc703*.

A further confounding factor is that the enhancer elements driving the expression of GAL4 may not be enhancers for *Inc703*. The location of the insertion relative to the surrounding genes makes it unlikely to be an enhancer for *atub84B* or *Alh*, however a subset of enhancers have been shown to act on distal genes (Kvon et al., 2014). The expression profile of the nearest neighbour, *e(y)2b*, had previously been characterised and was shown by northern blot to be male specific (Krasnov et al., 2005). Using a lacZ reporter that contained 950bp upstream of the *e(y)2b* TSS, that encompassed the start of *Inc703* and the insertion site of the transposon used to report *Inc703* expression, expression of *e(y)2b* was determined to be restricted to differentiating spermatocytes (Krasnov et al., 2005). The expression of the fluorescent reporter outside these cells in my experiments suggested that it was not being driven by *e(y)2b* enhancers. Whilst qRT-PCR experiments reproduced highest expression of *e(y)2b* in males, especially in the testes, a basal level of expression above that seen in negative controls was seen in at all developmental time points, and in adult females. In qRT-PCR experiments of the reporter lines and in RNA-seq data from modENCODE, high expression of *e(y)2b* is observed in the fat body, as both these experiments contained a mix of male and females, this signal could come from either the fat body or the embedded gonads. It was not possible to obtain an antibody specific for *E(y)2b*, however, it would be possible to compare expression in male and female fat bodies by qRT-PCR to determine which tissue is contributing to the signal.

In a number of tissues, it was not possible to identify the specific cell type in which the reporter was expressed, i.e. the photoreceptor subtype in the eye disc. Cell type identification can be achieved using the large number of GAL4 “enhancer-trap” lines; however this was not possible whilst relying on a GAL4/UAS reporter system to identify *Inc703* expressing cells. Use of a direct RNA visualisation

method could be combined with GAL4/UAS cell type identification would allow for identification of *Inc703* expressing cell types.

5.4.2 Levels of ribosomal protein genes are increased in *Inc703* mutants

Gene expression changes in the fat body of third instar larvae from two independent *Inc703* mutant alleles revealed 100 genes differentially expressed in both mutants, 88 of which are changed in the same direction. These account for 75% of all genes differentially expressed in *Inc703^{gal4}* but less than 5% of genes differentially expressed in *Inc703^{stop}*. The large number of genes differentially expressed exclusively in *Inc703^{stop}* could be caused by the second site mutation. However the correlation between the log₂ fold changes compared to wildtype between the two alleles suggests that *Inc703^{stop}* and *Inc703^{gal4}* affect the same genes, and that *Inc703^{stop}* is a stronger allele. The overlap is enriched for ribosomal proteins (12 ribosomal genes are upregulated in both) and 86 cytosolic ribosomal components are upregulated in *Inc703^{stop}*.

The extent of ribosomal protein expression changes correlates with the upregulation of the *Drosophila* Myc homolog *dm*. Levels of *dm* are slightly but not significantly elevated in *Inc703^{gal4}* (log₂ FC = 0.537, *p* = 1) and more highly upregulated in *Inc703^{stop}* (log₂ FC = 2.76, *p* = 2.37 × 10⁻⁰⁷). Increased levels of *dm* are known to increase ribosome biogenesis through the coordinated upregulation of ribosomal RNA and ribosomal protein levels (Grewal et al., 2005). For the levels of functional ribosomes to be increased rRNA levels must also be increased, as rRNA levels are the rate-limiting step in ribosome biogenesis (Laferte et al., 2006). However due to the abundance of these transcripts they are depleted from cDNA sequencing libraries so it will be necessary to perform further analysis to determine if rRNA levels are also increased in *Inc703* mutants. As well as controlling levels of ribosomal components it is thought that Dm might control up to 15% of coding gene (Orian et al., 2003). As well as GO terms involved in

translation, GO terms associated with the mitotic spindle were also significantly enriched. However, all the genes that contribute to the mitotic spindle terms also contribute to the translation terms, suggesting that although some ribosomal proteins have a function in the mitotic spindle, their primary role is in translation.

Curiously, gene expression changes in *Inc703*^{gal4} were not rescued by two copies of BAC^{158D05}, suggesting that *Inc703* only functions in its endogenous locus. It could simply be that the BAC does not contain all the required regulatory sequence to express *Inc703* in the endogenous pattern. Alternatively, it could also be that *Inc703* acts through chromosomal looping like *HOTTIP* (Wang et al., 2011) or that the act of transcription in the endogenous locus is required to regulate a neighbouring gene as for *SER3* (Martens et al., 2004).

5.4.3 Association of ribosomal protein levels, *dm* and cell competition

Decreased levels of ribosomal proteins globally results in the *Minute* phenotype, exhibiting thinned bristles and developmental delay. However, when *Minute/+* cells are generated in a wildtype background they fail to compete with wildtype cells and are lost through apoptosis. Conversely, animals with higher levels of *dm* and consequently increased ribosomal protein levels are larger and *dm* overexpressing clones are super-competitors which can induce apoptosis of surrounding less fit cells. It might therefore be expected that *Inc703* mutant cells, which have higher levels of ribosomal proteins, may have a competitive advantage.

The clonal analyses performed in this chapter were generated by MARCM (Lee and Luo, 2001), and consequentially the “twin-spot”, the corresponding cells which are fully wildtype, were indistinguishable from surrounding heterozygous cells. It is therefore not possible to infer anything from the mutant clone size in these experiments. However it would be possible to generate mitotic clones, in which both the homozygous mutant and corresponding homozygous wildtype

cells are differentially marked (Griffin et al., 2009; Tian and Jiang, 2014). The relative size of individual cells or proportion of mutant and wildtype cells in mosaic tissue generated by this method could be used to determine if the increased levels of Dm and ribosomal proteins in *Inc703* mutant cells confer any cellular advantage.

It is not known what is driving the increased *dm* levels in *Inc703* mutant fat bodies as neither *yorkie* (*yki*) or *raspberry* (*ras*), which are known to activate *dm* resulting in increased ribosome levels are upregulated. In fact, *ras* levels are significantly reduced in *Inc703^{stop}* mutants ($\log_2$ FC = -0.753, $p = 0.023$). It is possible that *dm* levels are being increased by a novel mechanism. However, because the fat body receives signals from other larval tissues it is possible that *dm* levels are being regulated in a non-cell autonomous manner. Currently, the increase in expression of ribosomal proteins has only been shown in the larval fat body, and it is unclear if the increase in ribosomal protein levels are observed in other tissues and at other developmental stages.

5.4.4 *arc1*, RpS6 and starvation response

One gene that is changed in opposing directions between the *Inc703^{gal4}* and *Inc703^{stop}* mutants is *arc1*, which is involved in response to starvation (Mattaliano et al., 2007). *arc1* is downregulated in *Inc703^{gal4}* ($\log_2$ FC = -1.98, $p = 2.13 \times 10^{-09}$) and significantly upregulated in *Inc703^{stop}* ($\log_2$ FC = 2.17, $p = 2.46 \times 10^{-13}$). Both mutants have elevated *rpS6* levels so it would be interesting to determine if the phosphorylation state differs between the two mutants, given that the dephosphorylation of RpS6 is known to induce autophagy in response to amino acid deprivation (Scott et al., 2004).

5.4.5 Ribosomal protein levels can influence lifespan

Changes in levels of ribosomal proteins have been shown to affect lifespan in *C.*

C. elegans and yeast (Houtkooper et al., 2013; Steffen et al., 2008). It is therefore tempting to speculate that the reduced lifespan of *Inc703*^{gal4} and BAC^{158D05}; *Inc703*^{gal4} strains is caused by increased ribosomal protein levels. Given that overexpression of *Inc703* cDNA transgenes driven by *actin5C*-GAL4 increased lifespan it would be interesting to see if ribosomal levels are decreased in these strains. Recently a lncRNA has been implicated in lifespan extension associated with reduced ribosomal protein levels in *C. elegans* (Essers et al., 2015). The lncRNA *tts-1* is induced in response to bacteria, on formation of the developmentally arrested dauer and in the long lived *daf-2*, *clk-1* and *eat-2* mutants in which ribosome levels are decreased. Reduction of *tts-1* reduces the long lived phenotype of *daf-2* and *clk-1* mutants by increasing ribosome levels.

5.4.6 A putative role for *Inc703* in controlling ribosome biogenesis

Using the currently available *Inc703* mutant strains it is not possible to conclude a definitive function for *Inc703*, however *Inc703* may have a direct or indirect role in the regulation of expression of a number of genes including multiple ribosomal protein genes, which is dependent on *Inc703* function at its endogenous locus. This will need to be confirmed by an independent method, for example ubiquitous or tissue specific shRNA-mediated knockdown of *Inc703* using the reagents described in Chapter 4 or by rescue of the gene expression changes using a cDNA rescue. The level of knockdown by shRNA may not be great enough and would only recapitulate the gene expression changes if the RNA molecule is functional. Given that it is not yet clear if the effects are cell autonomous, it is not known which tissues the cDNA should be driven in. However, expression of the UAS-driven *Inc703* cDNA transgenes used in lifespan experiments in this chapter could be achieved under the control of the GAL4 within the *Inc703*^{gal4} locus. This would be expected to achieve physiological levels of *Inc703* expression in the correct expression pattern.

Alternatively, CRISPR/Cas9 could be used to create an indel, possibly deleting the splice acceptor site of second exon, which would recapitulate the premature termination in the *Inc703*^{stop} and *Inc703*^{gal4} strains. Finally, the generation of a transgenic strain in which transcription of *Inc703* but not its neighbouring genes is blocked may be achievable by targeting a catalytically inactive Cas9 to the transcriptional start site of *Inc703* to block access of RNA Polymerase II (CRISPRi) and would allow for testing of this hypothesis. The success of this technique to block transcription of lncRNAs is as yet unclear, although knockdown of roX1 and roX2 has been achieved in cell culture (Personal Communication, S. Ghosh). It would be expected that gene expression changes could be recapitulated by this method irrespective of whether the RNA molecule or the act of transcription of *Inc703* is important.

Chapter 6 - Using transcriptome profiling to investigate tissue specific roles of miR-8

Declaration

The work presented in this chapter was carried out in collaboration with Dr Markus Toegel, supervised by Dr Tudor Fulga. Credit will be given in the text where appropriate.

Abstract

miR-8 is the *Drosophila* homolog of the highly conserved miR-200 family, found in bilaterian mammals. miR-8 has a complex expression pattern, and mutations in the gene result in numerous phenotypes across different tissues and developmental stages. Previous studies have suggested multiple target genes whose derepression may be responsible for these phenotypes. Using transcriptome profiling of the larval CNS and larval musculature of wildtype and miR-8 null flies, we show that miR-8 has different targets in two larval tissues in which it is expressed. Differentially expressed genes between wildtype and miR-8 null samples contain a mixture of direct and indirect effects, a subset of which exhibit the same change in both tissues. Candidates with potential to be direct targets of miR-8 were identified using two MRE prediction algorithms and enrichment for genes containing predicted miR-8 MREs was found genes upregulated in the muscle but not in the CNS. Three *Drosophila* insulin like peptides (DILPs), one of which (*ilp5*) contains a predicted miR-8 binding site, were highly upregulated in the CNS of miR-8 null larvae. Tissue specific knock down of miR-8 using a specific EGFP-sponge showed that the effect of miR-8 on *ilp5* is non-cell autonomous and likely a consequence of increased u-shaped (USH) levels affecting insulin signalling transduction.

6.1 Introduction

6.1.1 dme-miR-8 is a homolog of the conserved miR-200 family

The *Drosophila* microRNA dme-miR-8 (miR-8) is the only homolog of the highly conserved miR-200 family that is conserved in bilaterian mammals. Consequently, miR-8 has been used as a model to understand the biological function of the miR-200 family. Northern blots have shown miR-8 is detectable throughout development (Aravin et al., 2003) and various reporters have shown a complex expression pattern. The tissue specificity, as determined by miR8-GAL4 driven GFP, shows expression in a banded pattern in the ectoderm of stage 13 embryos, embryonic muscle expression from stage 15 and expression in the 3rd larval instar central nervous system (CNS) and wing and leg imaginal discs (Karres et al., 2007). Small RNA-seq shows that miR-8 accounts for approximately 10% of microRNA expression in both adult heads and ovaries (Berezikov et al., 2011).

6.1.2 The human miR-200 family controls epithelial to mesenchymal transition

Studies on the human miR-200 family have focused on its role in the epithelial-mesenchymal transition (EMT) (Christoffersen et al., 2007; Hurteau et al., 2007). The expression of the miR-200 family miRNAs is strongly correlated with the epithelial phenotype and its absence is sufficient to drive EMT (Gregory et al., 2008; Korpál et al., 2008; Park et al., 2008). Conversely, ectopic expression of miR-200 can drive the mesenchymal-epithelial transition (MET) and reduce motility and aggressiveness of breast carcinoma derived cell lines (Korpál et al., 2008; Park et al., 2008). The miR-200 family is therefore useful as a biomarker and is potentially an attractive therapeutic target to prevent metastasis (Chen et al., 2011).

6.1.3 Manipulation of miR-8 levels causes multiple tissue specific effects, mediated through distinct targets

6.1.3.1 miR-8 null flies have reduced survival and wing and leg defects

Karres et al. (2007) described three miR-8 loss of function, mutant alleles. miR-8^{Δ2} and miR-8^{Δ3} were generated by imprecise excision of the P element transposons P(EP)2269 and P(EP)2239 generating deletions of 1.8kb and 26kb, respectively. In miR-8^{Δ1}, a 400bp region of the miR-8 locus that includes the region encoding for the hairpin was replaced by a mini-*white* gene by homologous recombination. Northern blot analyses show that no mature miR-8 is produced in any allelic combination (Karres et al., 2007). Although flies lacking miR-8 develop normally through embryonic and larval stages, most fail to emerge from pupal cases and emergent adults have leg and wing extension defects and fail to survive longer than 24 hours after eclosion. miR-8 has been shown to be expressed in the developmental precursors of these tissues, the leg and wing imaginal discs. However, similar developmental phenotypes could not be recapitulated by generation of miR-8 null clones in the leg or wing imaginal discs (Karres et al., 2007). It has previously been suggested that, as both wing expansion and leg extension are driven by contraction of the abdominal muscles (Fortier et al., 2003), defects could be caused by a neuromuscular defect in the miR-8 mutant flies.

Karres et al. (2007) identified four genes that were upregulated greater than two-fold and contain miR-8 MREs conserved in *Drosophila pseudoobscura* in microarray analysis of whole pupae. One of these was the *Drosophila atrophin* gene, which is the orthologue of the mammalian genes atrophin-1 and RERE. This gene contains four predicted miR-8 MREs in its 3'UTR, two of which are conserved in *D. pseudoobscura*. Luciferase assays in which the 3'UTR of *atrophin* or its mammalian homologue, RERE, were cloned downstream of a

luciferase gene showed that mutation of the predicted miR-8 or miR-200 family member (miR-429 and miR-200b) binding sites led to an increase in reporter gene expression. *In vivo* reduction of *atrophin* levels in combination with loss of miR-8, reduced the penetrance of the survival and leg defects (50% and 40% reductions respectively), suggesting that *atrophin* is in part responsible for the miR-8 mutant phenotypes.

6.1.3.2 miR-8 negatively regulates Wnt signalling

Kennell et al. (2008) looked at the effects of misexpression, rather than loss, of miR-8 to identify pathways that may contribute to the leg and wing abnormalities seen in miR-8 null mutants. Ectopic expression of miR-8 in specific cells in the leg and wing discs reduced *Wg* signalling. miR-8 has been shown by lacZ-3'UTR fusion sensors to target sites in the 3'UTR of *wls* and *CG32767*. *Wls* is required for *Wg* secretion and *CG32767* has been identified as a positive regulator of *Wg* signalling. Moreover, ectopic miR-8 also reduces pangolin (Pan) protein levels, independent of pan mRNA levels (Kennell et al., 2008). Pan confers the DNA binding function of the bipartite Arm/Pan transcriptional activator complex downstream of the *Wg* signalling pathway (van de Wetering et al., 1997); miR-8 therefore acts to repress *Wg* signalling at multiple stages of the pathway. Kennell et al. (2008) speculate that because *Wg* signalling is important in the development of the wing and leg, perturbation of *Wg* signalling may contribute to the phenotypes observed in the miR-8 null mutants described by Karres et al. (2007). However, caution must be taken in interpreting the effects of overexpression of miRNAs as they may result in targeting that is not physiologically relevant due to improper stoichiometry of the miRNA and targeting transcripts. The presence of an excess of miR-8 or the absence or lower abundance of a true target upon ectopic expression might result in binding to targets for which binding would not be favourable in the physiological setting.

6.1.3.3 miR-8 null flies also have reduced body weight and size

A second paper characterising the miR-8^{Δ2} mutant showed a reduction in body weight and size of adults (Hyun et al., 2009). This phenotype was attributed to *u-shaped* (USH) the *Drosophila* homolog of Friend of GATA 2 (FOG2), which was identified as a direct target of miR-8 due to conservation of predicted miR-8/miR-200 binding sites between *ush* and FOG2. Targeting of *ush* by miR-8 was confirmed by a luciferase reporter assay as previously described. The steroid hormone Ecdysone was subsequently shown to suppress miR-8 expression and consequently levels of its target *ush*, via transcriptional repression of the miR-8 locus by E74 and BR-C (Jin et al., 2012). Constant transcriptional repression of miR-8 by Ecdysone throughout the third instar larval stage moderates insulin signalling, and therefore body size. No cooperation was shown between the *ush* and *atrophin* mediated phenotypes and *ush* levels were most significantly upregulated in the fat body leading Hyun et al. (2009) to conclude distinct tissue specific effects of loss of miR-8.

6.1.3.4 miR-8 regulates immune homeostasis

The basal level of the AMPs *Drosomycin* and *Diptericin* are elevated in miR-8 null larvae and adults. Although basal expression of AMPs in the absence of an immune challenge is observed in multiple tissues in larvae, the increase in *Drosomycin* (*Drs*) occurs specifically in the fat body in a tissue autonomous manner. *Drosomycin* is activated by the Toll pathway and both the *Toll* receptor (*Tl*) and the transcriptional activator *Dorsal* (*Dl*) activated by the Toll pathway were shown to be direct targets of miR-8 in luciferase reporter assays. Overexpression of *Dl* in the larval fat body was sufficient to reproduce the increase in *Drs* levels observed in miR-8 null larvae. Perturbation of the level of *Dl* or the Toll ligand *Spatzle* (*Spz*) increases the survival of miR-8 null flies, suggesting that the increased basal level of AMPs is partially responsible for the reduced survival

of miR-8 null flies. Interestingly, overexpression of previously identified miR-8 target *Ush* in the larval fat body also resulted in elevated *Drs* levels, although to a lesser extent than *Dl* overexpression. Mutation of *Ush* completely abrogated the increase in *Drs* expression in miR-8 null larvae suggesting a previously unknown role for *Ush* in immune homeostasis.

6.1.3.5 miR-8 positively regulates pigmentation

Loss of miR-8 globally, or specifically in the epidermis, causes alteration in the spatial patterning of pigment in the female dorsal abdomen (Kennell et al., 2012). No alteration is observed when miR-8 is reduced in the fat body suggesting this is independent of effects of body size. An independent miR-8 null mutant miR-8^{jk22} shows the same reduction in the number of flies that eclose as observed previously (Karres et al., 2007). Kennell et al. (2012) further show that both the pigmentation and eclosion defects are more severe at higher temperatures.

6.1.3.6 miR-8 defines a subset of glia

Whilst this work was being conducted, Morante et al. (2013) published work showing that miR-8 can define a subset of glial cells, distinct from the perineurial and subperineurial glia, and is required in these cells for correct neuroepithelial expansion and neuroblast transition. miR-8^{Δ2} homozygotes exhibit an overgrown neuroepithelium at the third instar larval stage and ectopic and precocious neuroblasts at later stages. The authors identified a miR-8 MRE in the 3'UTR of the EGFR ligand *spitz* (*Spi*) and used luciferase assays and a GFP reporter fused to the 3'UTR to confirm it as a direct target of miR-8 *in vitro* and *in vivo* respectively. Furthermore, overexpression of miR-8 in the miR-8 expressing glia resulted in increased size of these cells which the authors speculate is a result of repression of a negative regulator of endoreplication targeted by miR-8 (Morante et al., 2013).

6.1.4 Experimental aims

The numerous distinct tissue specific effects of the loss of miR-8 described in the literature led us to believe transcriptome profiling of individual tissues in which miR-8 is expressed may be more informative than the previous transcriptome profiling experiments on whole pupae. Moreover, different genes have been implicated in the different phenotypes observed in miR-8 mutants, and overexpression of *atrophin* only partially rescued the survival, leg and wing defects, suggesting that other target genes may play a role in these phenotypes (Karres et al., 2007). RNA-seq is an unbiased and highly sensitive technique which will allow identification of previously unknown miR-8 targets. Furthermore, using RNA-seq rather than microarray analyses will allow for the identification of lncRNAs, including novel transcripts, targeted by miR-8. Using shared MREs and correlations in expression between lncRNA and mRNA targets may allow us to identify mRNAs and lncRNAs involved in ceRNA networks of miR-8 targets.

6.2 Methods

6.2.1 Sample preparation and sequencing

Samples were prepared by Dr Markus Toegel. Either the musculature or central nervous system of either wildtype or miR-8^{Δ1}/miR-8^{Δ2} transheterozygous third instar larvae were removed by manual dissection and flash frozen in liquid nitrogen. Total RNA was extracted using mirVana miRNA Isolation kit (Life Technologies) according to manufacturer's protocols. 15 larvae were pooled for each biological replicate for the CNS and five larvae were pooled for each biological replicate for the musculature. 1.6 µg of RNA per sample was sent to the Wellcome Trust Centre for Human Genetics (Oxford) for polyA-selected, 100bp, paired-end RNA-seq.

6.2.2 Remapping, transcript assembly and defining lncRNAs

The 12 libraries ranged in size from 26 - 41 million reads and > 95% of reads mapped to the BDGP5.68 assembly of the *D. melanogaster* genome. Unmapped reads were extracted and remapped using a reference containing protein-coding and pseudogenes from the BDGP5.38 assembly and long non-coding RNAs from Young et al. (2012) which are not included in the BDGP5.38 assembly using TopHat2 (Kim et al., 2013). Newly mapped reads were added to those already mapped, biological replicates were pooled and transcripts were assembled using Cufflinks (Trapnell et al., 2010) and by supplying all protein-coding and non-coding exon junctions. A complete set of transcripts from all four biological conditions was determined using Cuffcompare. To identify lncRNAs, the Bedtools suite was used to extract transcribed genomic intervals that do not overlap on either strand with any known protein-coding or pseudogenes to give 8932 genomic intervals. Coding potential was assessed using CPC and only multi-exonic genomic regions with all transcripts within the region called as non-coding were considered. 771

multi-exon long non-coding RNA loci were predicted, including ten that had not previously been identified.

6.2.3 Calling differential gene expression

Differentially expressed genes were identified using DESeq2 (Love et al., 2014) as previously described. The samples were divided into two groups based on tissue of origin. Pairwise comparisons were made between wildtype and miR-8 null samples. Results are presented as $\log_2$ fold-change; p-values were adjusted to 5% FDR using the Benjamini-Hochberg method.

6.2.4 miRNA Recognition Element prediction

TargetScan (version 6.0) was used to predict MREs for full length lncRNA and protein-coding transcripts. Predictions for MREs within 3' UTRs of coding-genes were obtained from the TargetScanFly website and predictions for MREs within full length genes were downloaded for the miRanda website (www.mirna.org).

6.2.5 Gene Ontology analysis

To discover common biological processes (BP), molecular functions (MF), or cellular compartmentalisation (CC) of proteins for genes differentially expressed in miR-8 null flies, the Database for Annotation, Visualization and Integrated Discovery (DAVID; v6.7) or Flymine online resources were used. Genes with ≥ 0.5 reads/base pair in wildtype tissues were defined as expressed in that tissue and these gene sets were used as the background to eliminate any enrichment caused by the tissue of origin.

6.3 Results

6.3.1 Four distinct sample populations can be distinguished within the RNA-seq results

Samples could be partitioned by genotype and tissue of origin by Principal Component Analysis (PCA) of all of the mapped reads. The first component (PC1) is able to very clearly separate the samples based on tissue of origin and accounts for > 90% of the variance (Figure 6.1a). The ten genes that contribute most to PC1 are all highly expressed in the carcass including *Lcp1*, *Lcp2* and *Lcp4*. The top 100 genes are enriched for GO terms involved in cuticle development [GO:0042335] and muscle cell development [GO:0055001]. The second component (PC2) is able to separate between wildtype and mutant samples and accounts for approximately 2.3% of the variance (Figure 6.1a). No gene ontology enrichments were observed for the top 100 genes contributing to the second principal component; however the single gene with the largest contribution (4.2%) was the *white* gene, which is the visible marker of the miR-8^{Δ1} allele.

The clustering of samples by PC2 is less tight in muscle than in CNS samples, indicating that there is more variation within the muscle than in the CNS, and comparing wildtype to mutant, greater variability within biological replicates for the wildtype muscle (Figure 6.1a). The third principal component (PC3) accounts for 1.5% total variance (Figure 6.1b), and may explain the increased variability of muscle samples, particularly the wildtype muscle. Of the top ten genes that contribute to PC3, nine are highly expressed in the salivary gland including (*Sgs3*, *Sgs5*, *Sgs7* and *Sgs8*). Whilst the CNS samples all cluster tightly based on PC3, muscle samples, particularly wildtype muscle samples do not (Figure 6.1b), suggesting varying amounts of salivary gland contamination in these samples.

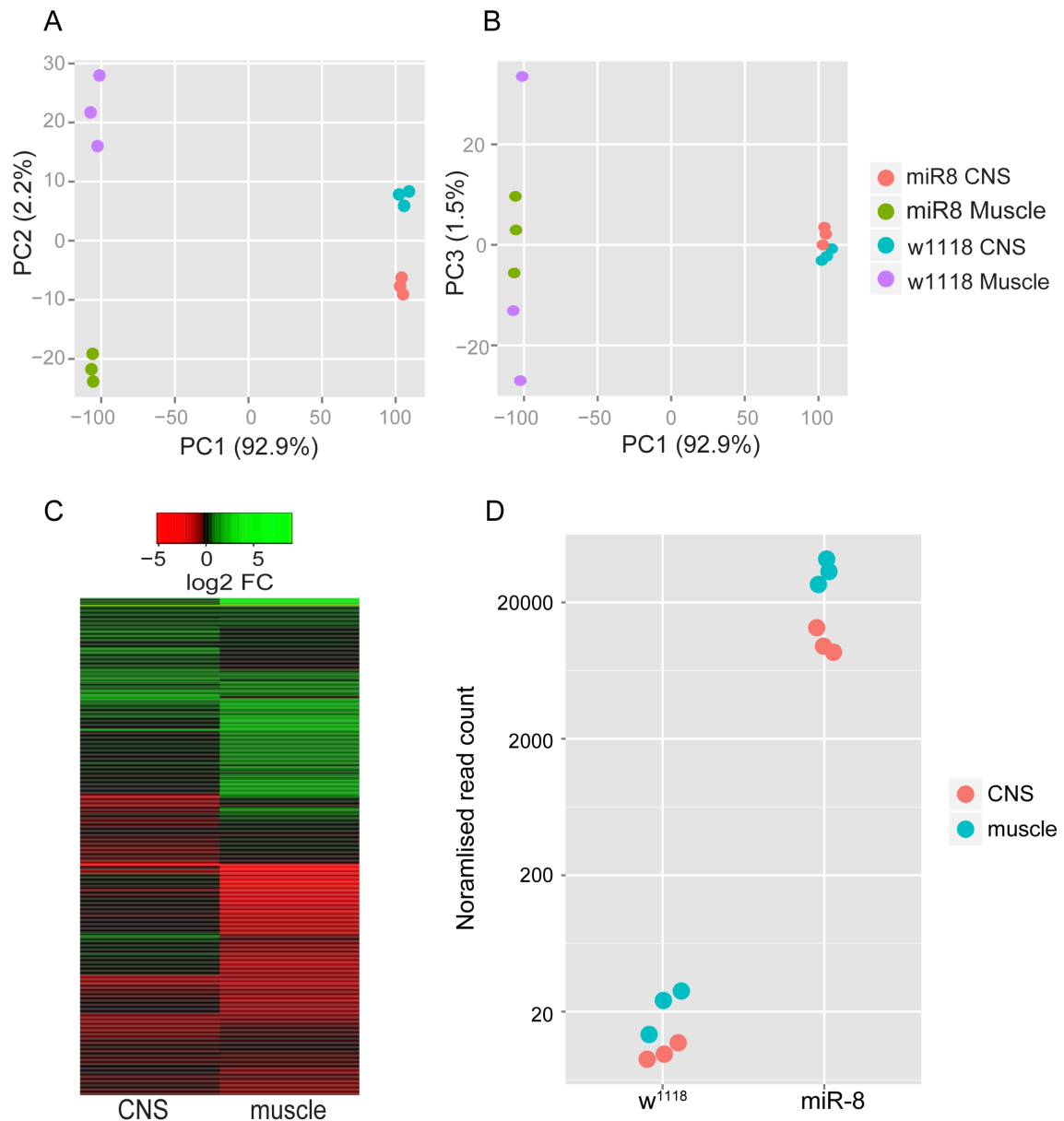


Figure 6.1 Distinct sets of genes are differentially expressed in the CNS and muscle of miR-8 null flies. (A) Considering all libraries, samples segregate into the triplicate measurements from different tissues (PC1) and genotypes (PC2). (B) The third principal component shows variability within the muscle samples. (C) In total 1,160 genes are differentially expressed in either mutant CNS or muscle, but changes are largely distinct in the different tissues. (D) The *white* gene acts as a positive control, showing upregulation in the mutant which contains a *mini-white* gene.

6.3.2 Differentially expressed genes include the visible marker of the miR-8^{Δ1} allele and represent both direct and indirect effects of loss of miR-8

DESeq2 calls 641 and 684 genes as differentially expressed between wildtype and mutant samples, in the muscle and CNS respectively (Figure 6.1c). The *white* (*w*) gene was used as a positive control for differential gene expression. Both the wildtype and mutant flies contain the *w*¹¹¹⁸ mutation that deletes the promoter and first exon of the *white* gene, the miR-8^{Δ1} allele contains a *mini-white* gene as a visible marker. In both the CNS and muscle, *w* has the second largest log₂ fold-change of all upregulated genes, with fewer than 20 sequencing reads mapped to it in wildtype tissues and more than 13,000 and 20,000 reads in the mutant CNS and muscle respectively (adj P = 1.05 x 10⁻⁶⁷, adj P = 1.58 x 10⁻⁶⁰) (Figure 6.1d).

The expectation is that direct miR-8 target genes will be derepressed and thus upregulated in the absence of the microRNA in the mutant samples. However, similar numbers of genes are downregulated and upregulated in both tissues (46% up in muscle, 43% up in CNS) (Figure 6.1c and Table 6.1), indicating that both direct and a large number of indirect changes are being detected.

Table 6.1 miR-8 MRE predictions made by TargetScan or miRanda. The number of differentially expressed genes containing one or more predicted MRE in each class.

		TargetScan		miRanda
	No. Genes	MREs (%)	3'UTR only (%)	MREs (%)
CNS up	292	53 (18.2)	10 (3.4)	29 (9.9)
CNS down	392	76 (19.4)	17 (4.3)	27 (6.9)
Muscle up	277	69 (24.9)	28 (10.1)	50 (18.0)
Muscle down	364	60 (16.5)	5 (1.4)	16 (4.4)
Both up	82	10 (12.3)	1 (1.2)	11 (13.4)
Both down	72	23 (32.4)	4 (5.6)	2 (2.8)

6.3.3 Muscle samples have higher variation but differentially expressed genes are not more variable

Filtering results only to include genes with a $\log_2$ fold change of greater than 0.5, left 605 differentially expressed genes in the muscle and 260 in the CNS. The higher number of differentially expressed genes with changes in expression less than 0.5 $\log_2$ fold change in the CNS compared to the muscle may be due to a true biological difference; alternatively, it may be a result of the higher variance in muscle samples, reducing the ability to detect small changes. To ensure the genes called as differentially expressed were not just lowly expressed and highly variable transcripts, expression levels were calculated for each gene in each condition.

In wildtype CNS samples, the median expression level of genes upregulated in the miR-8 mutant CNS (284.7 reads/kb) but not downregulated genes (416.1 reads/kb) is lower than unchanged genes (418.6 reads/kb) (Figure 6.2a, $p = 1.153 \times 10^{-05}$, Wilcoxon Test). Similarly, the median coefficient of variation (CV) is higher in upregulated genes (0.215) but not downregulated genes (0.194) than unchanged genes (0.200) (Figure 6.2c, $p = 0.012$, Wilcoxon test). The median expression level of genes upregulated (226.7 reads/kb) or downregulated (242.9 reads/kb) in mutant muscle was equivalent to that for genes unchanged in expression (231.7 reads/kb) (Figure 6.2b, $p = 0.849$, $p = 0.600$, Wilcoxon test). The median CV was also not significantly different for upregulated genes (0.231) than for unchanged genes (0.243) (Figure 6.2d, $p = 0.703$, Wilcoxon test), but was significantly higher for downregulated genes (0.278) (Figure 6.2d, $p = 0.0005$, Wilcoxon test). As expected, the CV in wildtype samples is higher in muscle samples than CNS samples (Figure 6.2c and d). However that only one class of differentially expressed genes in each tissue has greater variability suggests that genes called as differentially expressed are not just lowly expressed, highly variable genes.

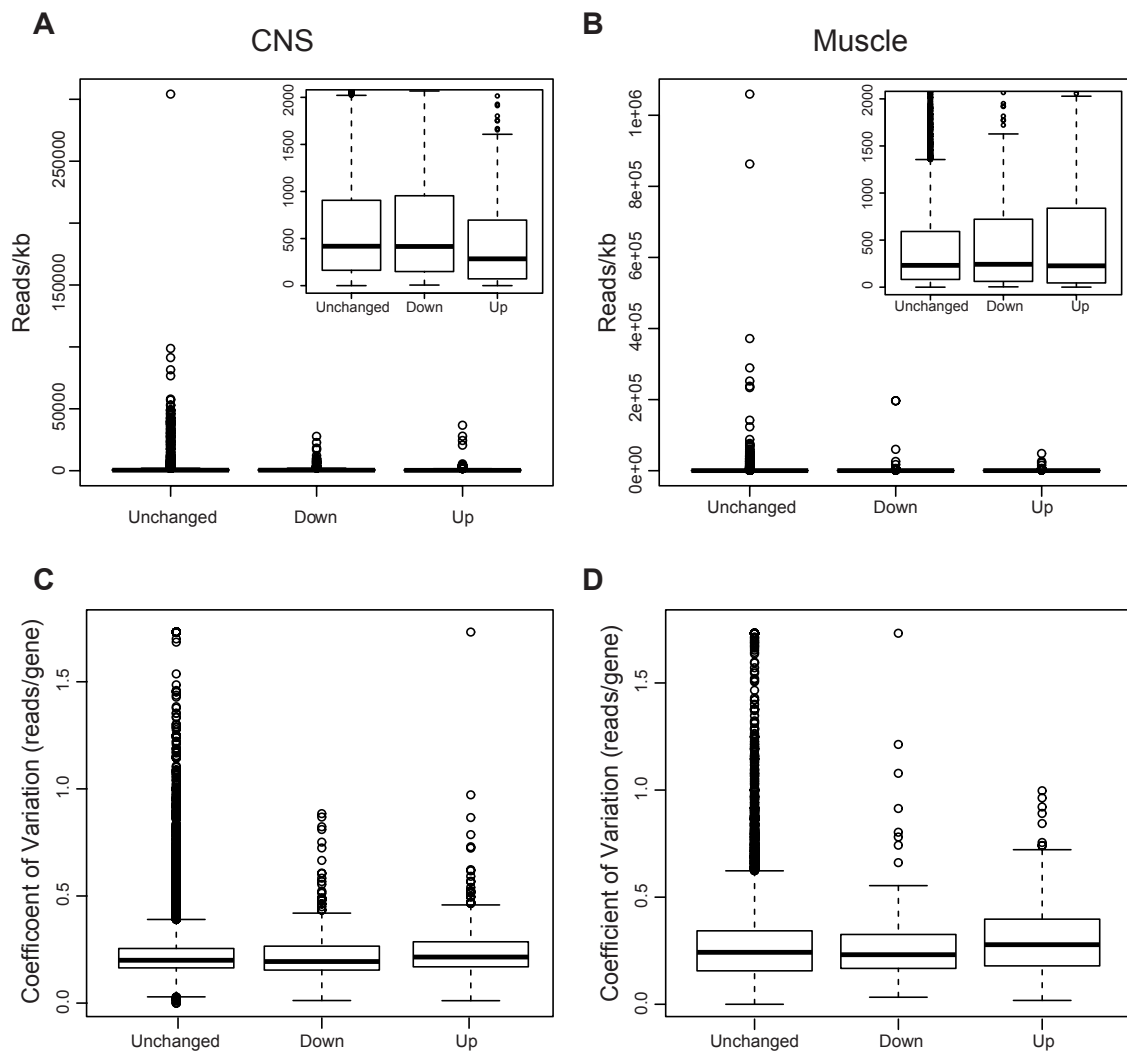


Figure 6.2 The expression level and variability of differentially expressed genes. Upregulated genes have lower median expression levels (A) and lower variability (C) in the CNS of wildtype larvae. (B) Downregulated and upregulated genes have expression levels equivalent to unchanged genes in the musculature of wildtype larvae. (D) The variability of upregulated genes but downregulated genes is lower in wildtype musculature.

6.3.4 miR-8 controls potentially overlapping pathways in different larval tissues

Previous studies have shown miR-8 induced gene expression changes specifically in one tissue. It might therefore be expected that the gene expression changes observed in the CNS and the muscle will be distinct. However, there are 165 genes differentially expressed in both tissues and there is a significant portion of these that are changed in expression level in the same direction in both (152, Fisher's exact test, $p < 0.0001$). This suggests that common pathways are regulated by miR-8 in the CNS and muscle. Alternatively, it could represent differences in the background of the mutant and wildtype strains.

Using GO analysis with a background of genes expressed in muscle and CNS, the biological process 'GO:0055114: oxidation reduction' was significantly enriched (15 genes, $\text{adj}P = 9.8 \times 10^{-05}$) with both up and down regulated genes contributing to the enrichment. However, these enzymes are not in a common pathway therefore suggesting a more general alteration in metabolism of miR-8 null larvae. To identify direct targets within the overlapping gene set, MRE predictions were made using either TargetScan (Ruby et al., 2007) or miRanda (Enright et al., 2003) algorithms (Table 6.1). Of the genes whose expression is changed in the same direction in both tissues, 23 contain miR-8 MREs predicted by TargetScan (10 upregulated genes, 13 downregulated genes) and 13 contain miR-8 MREs predicted by miRanda (11 upregulated, 2 downregulated). The transcription factor *shaven*, which is involved in muscle and neuron development, contains four miR-8 MREs predicted by TargetScan, two of which are also predicted by miRanda. However, *shaven* is downregulated in both the muscle and CNS, which is not consistent with targeting by miR-8. Table 6.2 lists the 14 genes that were identified as upregulated in both tissues and contain one or more predicted miR-8 MREs.

Table 6.2 Putative direct targets upregulated in both tissues. Genes are listed with the log₂ fold change (FC) compared to wildtype in muscle and CNS and the algorithm predicting the miR-8 MRE is indicated.

Gene	Muscle log₂ FC	CNS log₂ FC	MRE Predictions
<i>Tes</i>	1.046	0.261	TS + M
<i>Lkr</i>	1.115	0.621	TS + M
<i>CG4630</i>	1.431	0.874	M
<i>Rtet</i>	1.772	1.203	TS + M
<i>CG14464</i>	0.435	0.392	TS + M
<i>Wls</i>	1.276	0.261	TS + M
<i>Orct2</i>	0.834	0.437	TS + M
<i>CG14291</i>	0.652	1.208	M
<i>VhaSFD</i>	0.564	0.210	M
<i>GstD3</i>	0.895	0.448	M
<i>Caps</i>	1.415	0.614	TS + M
<i>P5CDh1</i>	0.667	0.201	TS
<i>CG6084</i>	0.582	0.237	TS
<i>Cyp9f3Psi</i>	1.483	1.454	TS

TS = TargetScan, M = miRanda.

6.3.5 CNS specific effects of miR-8

Two previous papers have shown different phenotypic changes in the pupal (Karres et al., 2007) and larval CNS (Morante et al., 2013), predicted to be mediated by different genes, *atrophin* and *Spi*, respectively. We sought to corroborate these observations using our data set.

6.3.5.1 *Atrophin* levels are not affected in the larval CNS of miR-8 null flies

Using our gene expression data to look specifically at *atrophin*, the gene previously implicated in the leg and wing phenotypes (Karres et al., 2007), no change in expression was seen in our data in CNS ($\log_2$ FC = 0.025, $\text{adj}P = 1$) or muscle ($\log_2$ FC = 0.201, $\text{adj}P = 1$) (Table 6.4), despite being expressed highly in the larval CNS (>14,000 reads in wildtype and mutant). Using the modENCODE data, expression is comparatively lower in pupae than the larvae (Graveley et al., 2011). The lack of effect on *atrophin* in the larval CNS suggests either that miR-8 and *atrophin* are not expressed in the same cells at this earlier time point, or that there is another level of regulation of the interaction between miR-8 and *atrophin* that prevents miR-8 targeting *atrophin* in the larval CNS. For example, the miR-8 MRE may be blocked by binding of other proteins, or a miR-8 target expressed in the larval CNS that is not expressed in the pupal CNS may compete for binding, or transcriptional feedback loops may result in homeostasis of *atrophin* gene expression in the larval CNS.

6.3.5.2 *Spi* levels are not changed in the CNS of miR-8 null larvae

Levels of *Spi*, the gene implicated in the ectopic neuroblast phenotype, are unchanged in CNS samples ($\log_2$ FC = 0.029, $\text{padj} = 1$) and not significantly changed in muscle samples ($\log_2$ FC = 0.575, $\text{adj}P = 0.24$) in our data (Table 6.4). However, the magnitude of the change observed in the musculature is similar to that observed in the brain by qRT-PCR by Morante et al. (2013) ($\log_2$ FC = 0.58). The

allelic combination of mutants used in the Morante study differed from that in our experiment, miR-8^{Δ2}/miR-8^{Δ2} compared with miR-8^{Δ1}/miR-8^{Δ2} transheterozygotes. However, because the initial Karres et al. (2007) paper describing these alleles showed no detectable mature miR-8 in any allelic combination, the phenotypes would be expected to be comparable. It is therefore unclear why it is not possible to detect the same expression changes in the CNS. Small differences in the developmental stage between the two experiments could explain the lack of reproducibility as neuronal development is very dynamic during the third instar larval stages. It is also possible that a background mutation in the miR-8^{Δ1} strain is masking or compensating for the change.

6.3.5.3 Identification of putative regulators of endoreplication

Morante et al. (2013) predicted that miR-8 affects glial size via repression of an unknown negative regulator of endoreplication. It was not possible to identify a direct regulator using our data set. The only two genes listed under the gene ontology term 'negative regulation of DNA endoreduplication' (GO:0032875) are *wg* and *esg*, neither of which was detected as differentially expressed in our analyses of miR-8 null larval CNS ($\log_2$ fold change = -0.20 and 0.01 respectively). However, *W/s*, which is required for Wg secretion and has previously been identified as a target of miR-8 (Kennell et al., 2008), is upregulated in both the CNS and muscle of miR-8 null larvae ($\log_2$ fold change 0.26 and 1.28, $\text{adj}P = 0.0036$ and 1.95×10^{-19} , respectively). Therefore, it is possible that miR-8 is controlling endoreplication in miR-8 expressing glia through control of Wg secretion by repression of *W/s*. However, further investigation will be required to determine whether the effect of miR-8 on endoreplication is mediated through *W/s*.

6.3.5.4 Genes upregulated in the CNS are not more likely to contain predicted miR-8 MREs

To identify novel miR-8 target genes in the CNS, miRanda and TargetScan were used to predict miR-8 MREs in differentially expressed genes. Despite the known problems with both high false negative and false positive discovery rates of MRE prediction algorithms, the expectation is that miR-8 targets would be enriched in genes upregulated in the CNS of miR-8 null larvae. Enrichment for miR-8 MREs was determined by the proportion of the class of genes that contain miR-8 MREs (Table 6.1). Comparing the number of genes containing miR-8 MREs in the full length transcripts of genes either upregulated or downregulated in CNS samples, no significant enrichment was found using predictions by TargetScan (up = 18.2%, down = 19.4%, $p = 0.694$, Fisher's exact) (Figure 6.3b). However, using the miRanda algorithm, a modest enrichment was observed for downregulated genes (up = 8.9%, down = 14.3%, $p = 0.033$, Fisher's exact). This suggests that most of the differentially expressed genes detected in the CNS are indirect targets of miR-8, or direct targets with non-canonical binding sites.

6.3.5.5 Gene Ontology (GO) analyses of genes differentially expressed in the CNS did not discover functional clusters

Gene ontology enrichment was tested for all genes differentially expressed in the CNS and upregulated and downregulated genes separately. No significant enrichment for a particular biological process, molecular function or cellular compartment was detected compared to the background list in any group of differentially expressed genes. Manual inspection of the genes differentially expressed specifically in CNS identified small clusters of functionally similar genes. For example, the three *Drosophila* insulin like peptides (*ilp2*, *ilp3* and *ilp5*) that are produced in the insulin-like peptide producing cells (IPCs) in the CNS were upregulated ($\log_2$ fold change = 2.6, 1.4, 1.3; $\text{adj}P = 3.91 \times 10^{-25}$, 6.51×10^{-25} , 1.1×10^{-25}).

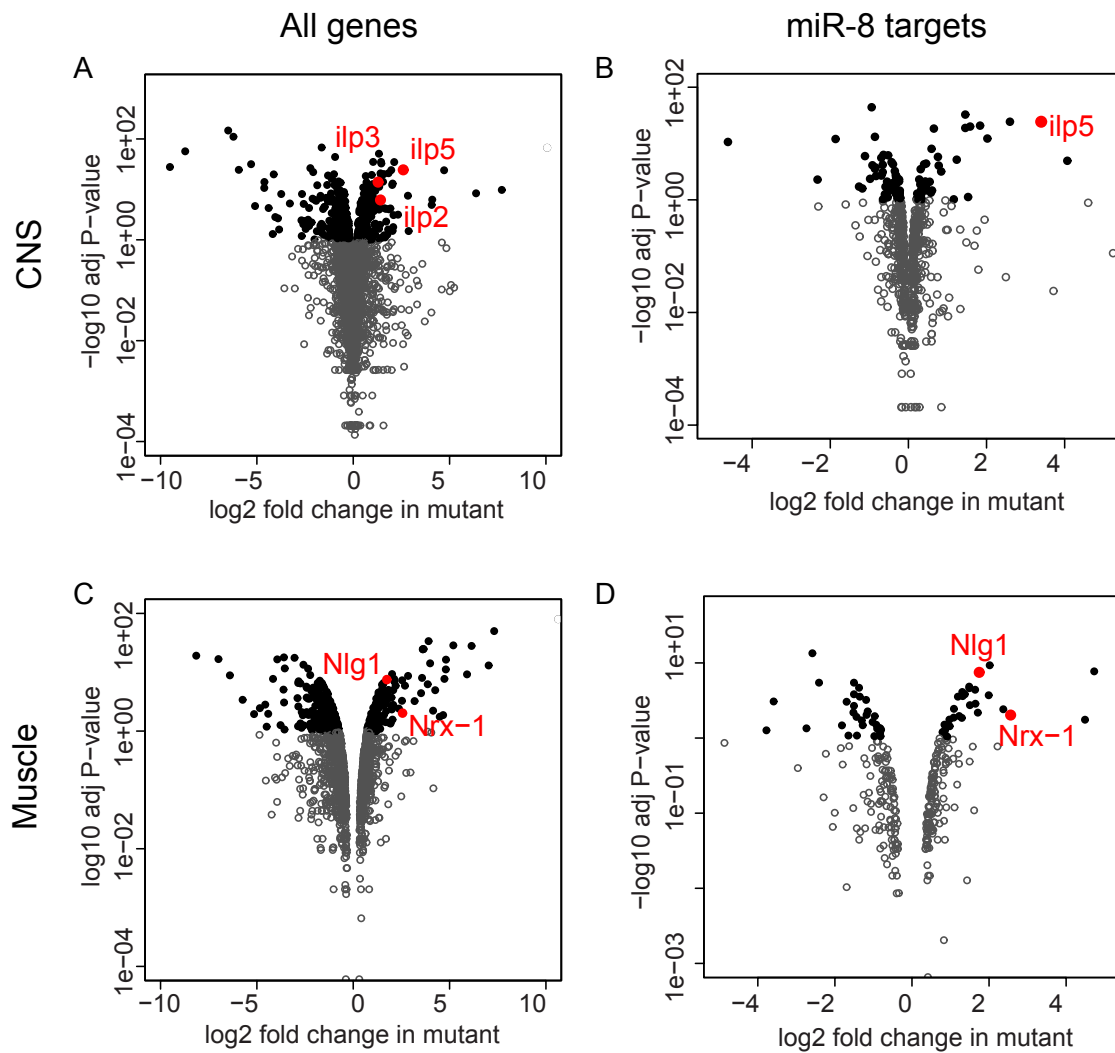


Figure 6.3 All differentially expressed genes and differentially expressed genes containing miR-8 MREs can be either upregulated or downregulated. Similar numbers of genes are upregulated or downregulated in miR-8 null CNS (A) and musculature (C). Genes containing TargetScan predicted miR-8 MREs are enriched in genes upregulated in the muscle (D) but not the CNS (B). Significantly differentially expressed genes are shown in black. Genes of interest are shown in red.

⁰⁷, 1.24×10^{-14}) (Figure 6.3a). Of these, only *ilp5* contains a predicted miR-8 MRE (Figure 6.3b).

6.3.5.6 Loss of miR-8 increases *ilp5* levels by an indirect, non-cell autonomous mechanism

The effect of miR-8 on *ilp5* was investigated further by Markus Toegel. Increase in *ilp5* levels in the CNS of miR-8^{Δ1}/miR-8^{Δ2} was confirmed by qRT-PCR (Figure 6.4a). Increased *ilp5* levels were also detected in whole third instar larvae in miR-8^{Δ1}/miR-8^{Δ2}, miR-8^{Δ2}/miR-8^{Δ2} and hemizygotes of both mutant alleles (Figure 6.4b) indicating that this effect is caused by absence of miR-8. Given the known effect of miR-8 on *ush* levels in the fat body, and subsequently on the insulin signalling pathway, miR-8 was knocked down in a tissue specific manner using a sponge containing ten miR-8 MREs in the 3'UTR of EGFP. Upregulation of *ilp5* was seen when the sponge was driven with two independent fat body drivers (*cg*-GAL4 and *ppl*-GAL4) but not by *InsP3*-GAL4 that expresses in the insulin producing cells in the CNS using the *ilp3* promoter. *Cg*-GAL4 is known to express in the glial cells of the CNS as well as in the fat body, whereas *ppl*-GAL4 is expressed in the fat body but not in the CNS (Figure 6.4c) indicating that the upregulation of *ilp5* occurs in a non-cell autonomous manner, likely as a downstream effect of altered USH signalling in the fat body in the absence of miR-8 (Table 6.4).

6.3.5.7 More neuron enriched upregulated genes contain miR-8 MREs than neuroblast enriched or downregulated genes

A GFP-sensor of miR-8 shows that its expression in the larval CNS is restricted to neuroblasts (Personal Communication, Dr Tudor Fulga) suggesting genes would be repressed by miR-8 in these cells. To further classify the genes differentially expressed in the CNS samples, genes were defined as either neuroblast or neuron enriched using RNA-seq data from FACS sorted cells from the larval CNS (Berger

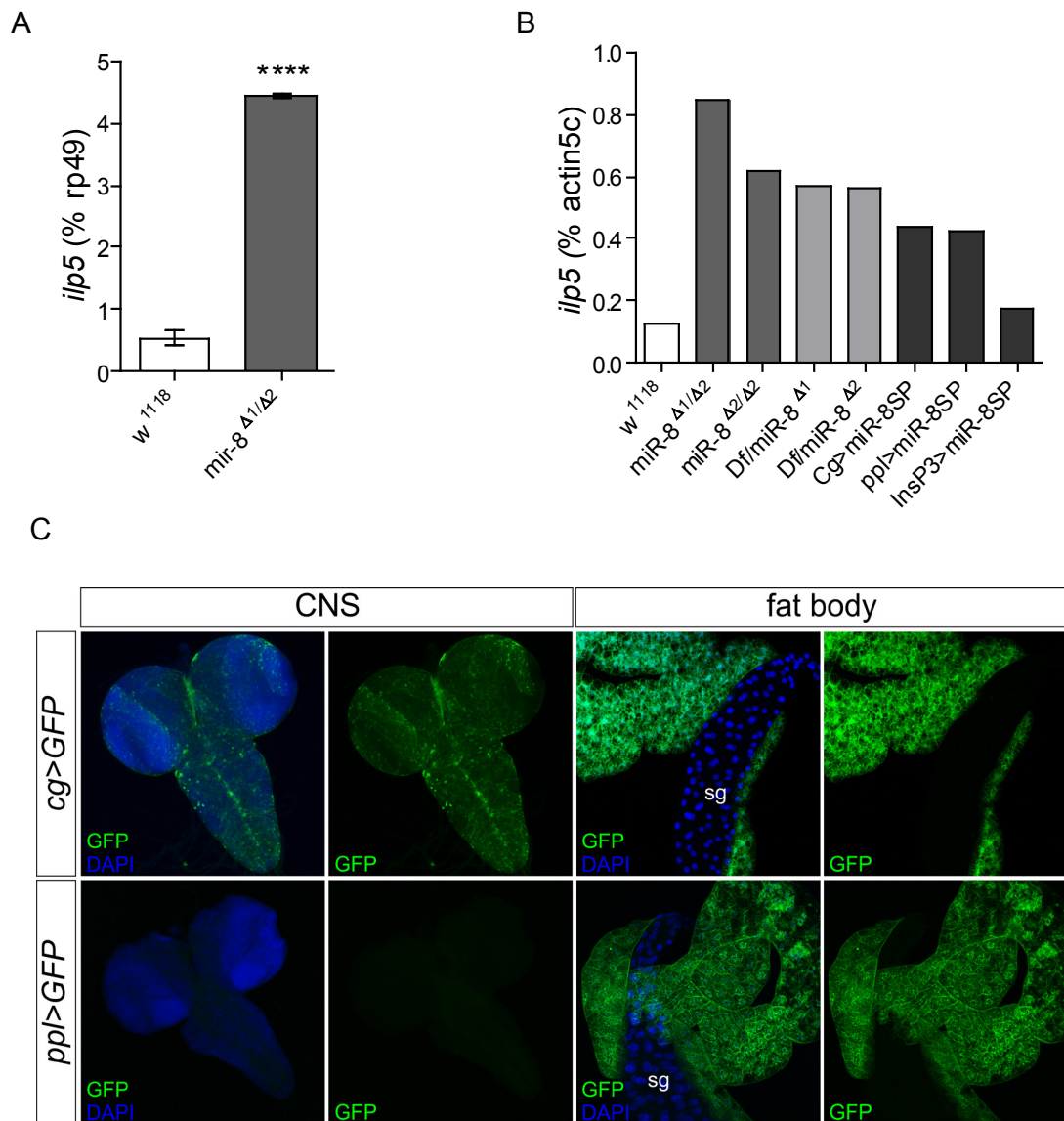


Figure 6.4 Upregulation of *ilp5* levels in miR-8 null larvae is mediated through the fat body. (A) qRT-PCR confirmed a > 8-fold increase in *ilp5* levels in miR-8 null transheterozygotes. (B) Upregulation is also seen in other miR-8 null mutants and by miR-8 knockdown in the fat body but not CNS. (C) *Cg*-GAL4 driven GFP shows expression in the glial cells of the CNS as well as the fat body, whereas *ppl*-GAL4 is expressed in the fat body but not the CNS. Neither are expressed in the salivary gland (sg). **** $p < 0.0001$ Student t-test. The data and images presented in this figure were generated by Dr Markus Toegel.

et al., 2012). Similar numbers of upregulated genes were found enriched in the neuroblast and neuron (25 and 24, respectively). However, significantly more neuron-enriched, upregulated genes contained predicted miR-8 MREs (Table 6.3) ($p = 0.01$, Fisher's Exact test). The enrichment for miR-8 targets in neuron-expressed genes is consistent with the hypothesis that miR-8 is expressed in the neuroblast ensuring repression of neuronal genes and that neuronal genes would be derepressed in the neuroblasts of mutant flies. No significant enrichment for specific gene ontologies or pathways was found for any subset of neuron-enriched or neuroblast-enriched genes.

Table 6.3 Genes significantly upregulated in CNS of miR-8 null larvae that contain predicted miR-8 MREs and show cell-type enrichment.

id	Gene name	log ₂ FC	Cell type enrichment
FBgn0001085	<i>fz</i>	4.02	neuron
FBgn0000529	<i>bsh</i>	3.65	neuron
FBgn0053653	<i>Caps</i>	9.05	neuron
FBgn0036141	<i>wls</i>	3.82	neuron
FBgn0032506	<i>CG9395</i>	5.27	neuron
FBgn0264753	<i>Rgk1</i>	4.33	neuron
FBgn0001229	<i>Hsp67Bc</i>	4.70	neuron
FBgn0039900	<i>Syt7</i>	6.62	neuron
FBgn0036279	<i>Ncc69</i>	4.01	NB

6.3.6 Muscle specific effects of miR-8

Whilst this work was being conducted two papers were published about the role of miR-8 in the development and function of the larval neuromuscular junction (NMJ).

6.3.6.1 *Lar* but not *wg* is upregulated in the muscle

The first identified miR-8 with miR-289 and miR-958, as regulators of activity dependent synaptic growth (Nesler et al., 2013). The authors identify sets of genes that are potentially coregulated by miR-8 and miR-289, or miR-8, miR-289 and miR-958. None of the listed genes are upregulated in the CNS of miR-8 null larvae. Two genes, *Lar* and *wg*, were shown to be targeted by miR-8 in *in vitro* assays. *Lar*, which is predicted to be coregulated by all three miRNAs, is significantly upregulated in the muscle of miR-8 null larvae ($\log_2$ FC = 1.062, p = 0.0004), supporting that *Lar* but not *wg* is a direct target of miR-8 at the NMJ *in vivo*. Nevertheless, it is not possible to rule out a role of miR-8 regulating *Wg* through translational repression, or that *Wg* levels are regulated in other ways, for example through transcriptional regulatory loops.

6.3.6.2 *Ena* is not changed at the transcriptional level

The second publication showed *Enabled* (*Ena*) to be both sufficient and necessary to account for miR-8-mediated regulation of subsynaptic reticulum architecture (Loya et al., 2014). Our data did not find any changes in *Ena* expression in the CNS ($\log_2$ FC = -0.047, p = 0.900) or the muscle ($\log_2$ FC = -0.074, p = 0.980). Furthermore, the fusion reporter method used to show *in vivo* targeting has been shown not to reflect physiological targeting of *Ena* by the miRNA *bantam* (Bassett et al., 2014b). Although, it is possible that *Ena* is regulated by miR-8 solely at the translational level, it is also possible that *Ena* function is required indirectly, through interaction with an unknown direct target of miR-8.

6.3.6.3 Detection of enrichment of miR-8 MREs in genes upregulated in the muscle is dependent on the prediction method

In muscle samples, significantly more upregulated genes (24.9%) than downregulated genes (16.5%) contained miR-8 MREs predicted by TargetScan ($p = 0.009$) but not the miRanda algorithm (10.1%, 12.4%, $p = 0.383$, Fisher's Exact test) which predicted 1.6 times fewer miR-8 MREs across all classes of genes. However, the best correlation between the two predictions methods was found for genes upregulated in the muscle (47% of predictions were made by both). It is still unclear what influence the location of an MRE, in the 5'UTR, CDS or 3'UTR of a gene, has on its efficacy of targeting. Therefore, I performed enrichment analysis considering only miR-8 MRE predictions within the 3'UTR of genes. The enrichment for MREs in genes upregulated in the muscle was stronger (10.1%, 1.4%, $p < 0.0001$, Fisher's Exact test). The distribution of expression level changes of genes with predicted miR-8 MREs can be seen in Figure 6.2d.

6.3.6.4 Gene Ontology (GO) analyses of differentially expressed genes in the musculature did not discover functional clusters

Gene ontology enrichment was tested for all genes differentially expressed in the musculature and upregulated and downregulated genes individually. Again, no significant enrichments were detected compared to the background list in any group of differentially expressed genes. No impact of the potential contamination of muscle samples with salivary gland tissue could be observed using gene ontology analysis. Manual inspection of the genes differentially expressed only in the muscle, identified *Nrx-1* and *Nlg1*, both of which contain five predicted miR-8 MREs, consistent with the published effect of miR-8 on the neuromuscular junction (Loya et al., 2009). Further experiments are required to determine whether *Nrx-1* and *Nlg1* transcripts are targeted by miR-8 *in vivo*.

6.3.7 Identification of tissue specificity of published changes

Other predicted targets of miR-8 have been reported in the literature associated with the numerous phenotypes described previously. Of particular interest are the studies carried out in larval tissues (Choi and Hyun, 2012; Hyun et al., 2009; Jin et al., 2012; Lee and Hyun, 2014; Morante et al., 2013), which can be compared to assess the tissue specificity of the changes using our data. The published change in *4EBP* expression in whole third instar larvae could both be replicated in both muscle and CNS samples. However, changes in *Drs* could only be replicated in muscle samples. *Drs* had previously been shown to be specific to the fat body (Choi and Hyun, 2012) suggesting that *Drs* is also elevated in the muscle or the epidermis to a lesser extent than the fatbody (Table 6.4). None of the published gene expression changes reported in the fat body could be replicated in the musculature or CNS (Table 6.4), including *Ush*, *Tl* and *DI* which have been shown to be responsible for the increase in basal expression of AMPs such as *Drs*, supporting the hypothesis that these effects caused by action of miR-8 specifically in the fat body. Together this supports previous assertions that miR-8 has distinct targets in different tissues but also targets some genes in multiple tissues.

6.3.8 Identifying ceRNA networks

Long non-coding RNAs were represented in upregulated and downregulated gene sets in the muscle and CNS. Of all the differentially expressed lncRNAs, only one gene, *XLOC_016468*, contained any predicted miR-8 binding sites. *XLOC_016468* was assembled from the RNA-seq data and corresponds to the three exon isoform of *CR43283*, albeit with a shorter first exon. *XLOC_016468* was downregulated in the CNS ($\log_2$ FC = -4.28 , $p = 0.0007$) which is inconsistent with the expected change for a gene functioning as a ceRNA suggesting that lncRNAs are not involved in controlling miR-8 networks in the CNS or muscle.

Table 6.4 Comparison of gene expression level changes in miR-8 null flies between our data and published studies. Significant changes are highlighted in red.

	L3 CNS		L3 muscle		Published data			
Gene	log ₂ FC	adjP	log ₂ FC	adjP	log ₂ FC	p	Stage/ Tissue	Ref
<i>Dpt</i>	-0.02	NA	0.88	0.08	2.58	<0.05	Adult males	1
<i>Drs</i>	-0.03	0.97	1.73	9.9 x 10 ⁻⁰⁵	3.91	<0.05	whole L3	1
<i>4EBP</i>	0.70	2.3 x 10 ⁻⁰⁹	1.14	0.01	1.32	Nr	whole L3	2
<i>Ush</i>	-0.08	0.84	0.69	0.26	1.00	Nr	L3 Fat body	2
					0.38	Nr	whole L3	2
<i>Tl</i>	-0.04	0.88	0.03	0.99	1.58	<0.05	L3 Fat body	3
<i>DI</i>	-0.08	0.82	0.14	0.94	1.68	<0.05	L3 Fat body	3
<i>Step</i>	-0.01	0.99	-0.01	1	0.77	Nr	L3 Fat body	4
<i>InR</i>	0.07	0.86	0.22	0.86	0.77	Nr	L3 Fat body	4
<i>Atrophin</i>	0.02	0.93	0.15	0.94	1.00	<0.0001	Pupae	5
<i>CG13060</i>	-0.15	NA	-1.42	2.7 x 10 ⁻⁰⁵	1.6	<0.0001	Pupae	5
<i>CG8420</i>	-0.21	0.58	-0.28	0.78	1.5	<0.0001	Pupae	5
<i>CG9036</i>	-0.08	NA	-0.37	0.46	1.4	<0.0001	Pupae	5
<i>Spi</i>	0.03	1	0.58	0.24	0.58	Nr	L3 CNS	6

References: ¹Choi and Hyun (2012); ²Hyun et al. (2009); ³Lee and Hyun (2014); ⁴ Jin et al. (2012); ⁵Karres et al. (2007); ⁶Morante et al. (2013).

6.3.9 The miR-8 primary transcript is upregulated in miR-8 null larvae

Reads which map to the 10kb region of the miR-8 primary transcript (pri-miR8) are detected at a low level in both the muscle (2,668 reads) and CNS (151 reads) of *w¹¹¹⁸* larvae. The read coverage was not sufficient to assemble a full length transcript in either tissue; however shorter transcripts that span the region were assembled (Figure 6.5). In the mutant samples, the number of reads mapping to the primary transcript is greatly increased in both muscle (16,121 reads) and CNS (6,515 reads) samples but as expected no reads map to the miR-8 hairpin which is absent in both mutant alleles (Figure 6.5). The apparently low steady state levels of the primary transcript in wildtype tissues are likely a result of degradation of the transcript during processing of the miRNA from the longer transcripts, or direct targeting of the primary transcripts by the mature miRNA. In the absence of a hairpin in mutant samples, the primary transcript can persist.

The pri-miR8 transcripts contain predicted 155 MREs for 87 different miRNAs predicted by TargetScan. No miR-8 binding sites were predicted by TargetScan. It is possible that the presence of high levels of this large transcript may mediate some of the indirect gene expression changes observed in our data. To investigate this further Dr Markus Toegel is generating a miR-8 null mutant, in which the hairpin is replaced by a non-targeting hairpin, to remove the mature miR-8 but maintain proper processing of the primary transcript.

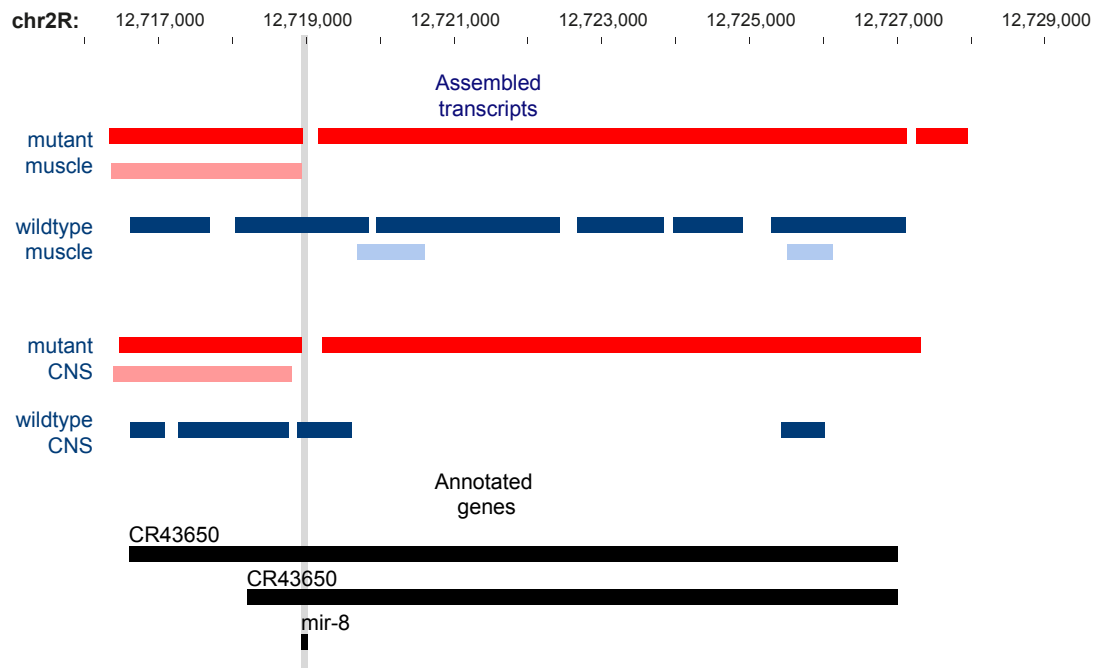


Figure 6.5 Assembled transcripts at the miR-8 locus in wildtype and mutant samples.

The annotated miR-8 primary transcripts and hairpin are shown in black. Transcripts assembled from the mutant samples are shown in dark red for sense transcripts and light red for antisense transcripts. Transcripts assembled from wildtype samples are shown in dark blue for sense transcripts and light blue for antisense.

6.4 Discussion

6.4.1 Levels of the *Drosophila* insulin like peptides and body size are uncoupled in *miR-8* null larvae

The ability of *Drosophila* to control their body-size according to nutrient availability is mediated through a family of diffusible molecules, the *Drosophila* insulin-like peptides (DILPs). In starvation conditions, DILPs are reduced in level, contributing to the reduction in body size. Hyun et al. (2009) describe a reduction in body size in *miR-8* null larvae, but dismiss the role of DILPs because they did not observe a reduction in levels of any *Ilp* genes in the *miR-8* mutant. However, *Ilp5*, *Ilp2* and *Ilp3* are all significantly upregulated in the larval CNS of *miR-8* null flies. *Ilp5* contains a *miR-8* MRE predicted by TargetScan and has the seconded highest fold change of upregulated, predicted *miR-8* targets.

Analysis performed by Dr Toegel indicates the effect of *miR-8* on *ilp5* levels is non-cell autonomous (Figure 6.3b) and therefore indirect. Furthermore, it is surprising to see increased DILP levels in flies with a reduced body size. However, the absence of the expected correlation between DILP levels and body size can be explained by the relief of repression of *Ush* specifically in the fat body of *miR-8* null larvae. Increased USH levels in the absence of *miR-8* blocks the signal transduction through the insulin receptor (Hyun et al., 2009), resulting in an uncoupling of DILP levels and activation of downstream signalling pathways. It would be possible to further confirm the *miR-8* dependence of increased DILP levels by expressing *miR-8* specifically in the fat body. To confirm if the increase in DILP levels are caused by *Ush*, *Ush* RNAi or the dominant negative *Ush*¹⁵¹³ could be expressed specifically in the fat body.

6.4.2 Comparison with published data reveals challenges in identifying miRNA targets

We had hoped to identify a high confidence set of miR-8 targets that were differentially expressed, contained predicted miR-8 MREs and were corroborated by a second data set. Whilst some published gene expression changes were reproduced, and some additional information about the tissue specificity of some changes was identified in the case of *Drs* levels changing in musculature but not CNS samples, other changes could not be reproduced. One of the challenges was the inability to make meaningful comparisons between our data and the only other genome-wide screen generated by Karres et al. (2007) because the two experiments were conducted at different developmental stages. Of the four genes identified as upregulated and containing conserved miR-8 MREs during pupal development, only one, *CG13060*, is also changed in expression in the musculature but not CNS during larval development. The magnitude and direction of the change seen in the larval muscle ($\log_2$ FC = 1.42, $\text{adj}P = 2.7 \times 10^{-05}$) is equivalent to the effect in pupal development ($\log_2$ FC = 1.6, $\text{adj}P < 0.0001$), suggesting that miR-8 has a consistent role in the regulation of *CG13060* but not *atrophin*, *CG8420* or *CG9036* during larval and pupal development (Table 6.4).

6.4.3 Methods of identifying true miR-8 targets

It is expected that direct targets of miR-8 will be derepressed in the absence of miR-8 which could be detected as upregulation by differential expression analysis as described in this chapter. However, equal numbers of genes were found to be upregulated and downregulated in both CNS and muscle samples from miR-8 null flies, suggesting that both primary and second effects are being detected. Algorithms that predict miR-8 MREs can identify the genes within this

data that are putatively direct miR-8 targets. However, these algorithms rely on the principle that miRNA-target interactions are mediated through seed matches which is not the case for approximately 19% of all such interactions (Helwak et al., 2013) and therefore will not be predicted by the algorithms. Furthermore, most published data generated using the algorithms only uses 3'UTR sequences, but 42.7% of interactions identified by CLASH were within the CDS (Helwak et al., 2013). However, it is not currently clear whether miRNA-target interactions that occur in the coding and non-coding regions of transcripts illicit the same effects. Although, there is some evidence for functional sites within coding regions (Tay et al., 2008). Therefore, predictions reported in this chapter have been based on both 3'UTR sites only and sites within the full transcript including 5'UTRs, coding exons and 3'UTRs.

One of the major challenges of finding direct miRNA targets using gene expression data, and equally confirming changes identified by alternative methods, lies in the unpredictability of gene expression changes resulting from absence of a miRNA. For example, *shaven* as a master regulator of both CNS and muscle development which contains two MREs predicted by both algorithms and is changed in both the CNS and muscle, is downregulated. Given that transcriptome profiling captures the steady state levels of transcripts, it is not possible to determine the regulatory dynamics that result in the reduction of *shaven* levels in miR-8 null larvae. This is particularly so when using genomic lesions such as the two alleles used in this study, because the miRNA is absent in all tissues and throughout development allowing for compensatory mechanisms and resulting in the detection of downstream effects. The advent of inducible miRNA sponges allows for tissue specific knockdown of miRNAs which can help distinguish between direct targeting and downstream effects (Loya et al., 2009), when the temporal and spatial expression of the drivers is known. As is the case when miR-8 is knocked down in the fat body; resulting in upregulation of DILPs in the CNS.

Furthermore, it is currently believed the miRNAs act to fine-tune or buffer expression levels of genes whose level must be tightly controlled (Friedman et al., 2009), meaning that the magnitude of change in a target might only be small, especially relative to the downstream effects this dysregulation may have. In some cases, known targets have failed to be identified in screens because their expression level did not change (Nicolas et al., 2008). Of the five studies using qRT-PCR to detect gene expression changes of targets shown in Table 6.4, only two reported the significance of the gene expression change, suggesting the effect of miRNA on their targets may not reach statistical significance. Furthermore, the two studies looking at the role of miR-8 in NMJ function infer direct targeting without measuring the change in RNA levels of any of the predicted targeted genes. Whilst RNA-seq offers the advantage of sensitivity over microarray technology, which has a limited ability to detect small gene expression changes, all genome wide techniques rely on rigorous multiple testing corrections and may struggle to detect small changes dependent on background noise or increased variability of gene expression levels (Schmiedel et al., 2015). Moreover, a recent study suggested that whilst three replicates is sufficient to detect changes with a $\log_2$ fold change ≥ 2 , a minimum of 12 replicates may be required to reliably detect >90% of changes $\geq 0.3 \log_2$ fold change which might be expected of direct miRNA targets (Schurch et al., 2015).

6.4.4 Better methods of predicting targets

Better determining the direct targets of miRNAs requires not only the improvement of the prediction algorithms used by better understanding of the factors that determine binding, but also the utilisation of new genetic tools to confirm *in vivo* targeting. Techniques to crosslink then sequence miRNAs and their targets have increased the understanding of the rules governing miRNA-target interaction

(Spitzer et al., 2014) and need to be combined with assessment of the gene expression changes in targets (Helwak et al., 2013) to better the understanding of the relationship between targeting and degradation or translational repression. Furthermore, the use of *in vitro* luciferase-3'UTR assays and *in vivo* lacZ-3'UTR or GFP-3'UTR fusion reporters to determine targeting need to be supplemented with deletion or manipulation of the putative miRNA binding sites at the endogenous locus which can now be achieved using genome engineering technologies (Bassett et al., 2014b).

6.4.5 Ongoing experiments

Dr Markus Toegel is currently investigating the effect of the persistence of the miR-8 primary transcripts from the miR-8^{Δ1} and miR-8^{Δ2} alleles on the phenotypes and gene expression changes observed in different allelic combinations. Preliminary evidence suggests different allelic combinations of miR-8 null alleles have different phenotypes supporting the idea that absence of mature miR-8 is not the only factor causing the mutant phenotypes.

Chapter 7 - Discussion

7.1 Summary of findings

In Chapter 3 of this thesis, I describe the characterisation of six putative long non-coding RNA loci which each contain a P-element or *Piggybac* transposon insertion. The effect of a transposon insertion within a lncRNA was shown to have varied effects, partially dependent on the modular elements in the transposon insertion and the specific genetic context. This work provided evidence to support why lncRNAs have rarely been identified through forward genetic screens and highlighted other challenges associated with the study of lncRNAs. However, I did identify a good candidate for further study, *Inc703*, which had two independent insertions that reduced levels of the lncRNA.

In Chapter 4 I showed the lethal phenotype of the *Inc703*^{stop} strain was caused by a tightly linked second mutation and showed transcriptional changes in the two mutant strains. In depth analysis of a subset of these changes in immune response genes using multiple complimentary techniques highlighted the importance of rigorous analysis and the challenges of discovering lncRNA associated phenotypes. Work presented in Chapter 5 shows *Inc703* to be expressed in the larval fat body and suggests that *Inc703* may have a functional role in the control of ribosome biogenesis associated with increased dm levels in the fat body. It may be that the disruption of *Inc703* expression increases ribosomal biogenesis which in turn decreases life span. These findings are particularly exciting given that a lncRNA *tts-1*, has been shown to alter lifespan through control of ribosome levels in *C. elegans*. However, due to the observed influence of genetic background on lifespan alterations, it will be necessary to confirm the effects reported in this thesis by the generation of independent mutants.

Work carried out in collaboration with Dr Markus Toegel presented in Chapter 6, identified many potentially direct targets of the evolutionarily conserved miRNA, miR-8 in the larval muscle and CNS. Through the use of complementary techniques it was shown that *ilp5*, which is upregulated in the CNS of miR-8 null larvae and contains a predicted miR-8 MRE was not a direct target of miR-8 but is upregulated in response to loss of miR-8 in the fat body. Furthermore, Wg signalling has previously been implicated in loss of miR-8 phenotypes, and I present evidence supporting that this is not through direct targeting of *wg* but may be mediated by targeting of *wls* which controls Wg export and Wg ligand *fz*.

7.2 Challenges using current techniques to study lncRNA function

The work presented in Chapter 3 highlighted a number of issues using existing methods to study lncRNA function. Firstly, it is clear that the existing lncRNA models only represent incomplete regions of transcription and that it is necessary to properly map the putative transcripts and where possible confirm the presence of the transcript by direct visualisation. Confirmation of transcript structure using RACE is difficult for lowly expressed lncRNAs, although if the expression pattern is known samples could be enriched for cells expressing the lncRNA. However, detection of specific cell populations that express lncRNAs requires visualisation of the lncRNA. Visualisation of lncRNAs poses challenges itself, current efforts by my colleagues have failed to achieve specific staining of any tested lncRNA (including the six characterised in this thesis) whilst robustly staining was achieved for other mRNAs with similar expression levels. Therefore, it is not yet clear if this is a problem of expression level, transcript turnover or caused by a quirk of the lncRNAs being studied. Single molecule FISH has been used to successfully detect lncRNAs in a number of mammalian cell culture systems (Cabili et al., 2015; Dunagin et al., 2015; Raj et al., 2008) and the Fly-FISH resource (Lecuyer

et al., 2007) is being expanded to include lncRNAs.

Secondly, evidence presented in Chapter 3 indicates that the effect of a transposon within a non-coding region is unpredictable and does not always result in termination of transcription. Indeed, two of the lncRNAs that I studied demonstrated increased expression levels in the insertion lines. That very few lncRNAs have been identified using traditional EMS and transposon mediated mutagenesis has led some to speculate that lncRNAs are not likely to be functional. However, in light of the inconsistent effects of transposon insertions on lncRNA loci and the different insertional biases of transposons which may preclude them from inserting in non-coding regions, this observation can be interpreted as the unsuitableness of current transposon insertion mutants for the study of lncRNAs, rather than a lack of functionality of lncRNAs.

The recent progress of genome engineering through the development of TALE and CRISPR/Cas9 technologies will facilitate the precise genetic modifications required to probe the function of long non-coding RNAs and remove the constraint of studying only lncRNAs amenable to shRNA mediated knockdown or with existing associated mutations. However, a major challenge still remains to identify potentially functional lncRNAs. Without the constraint of requiring a pre-existing mutation to study a lncRNA it will be possible to identify lncRNAs associated with particular physiological responses or developmental transitions. For example, the mammalian *lincRNA-Cox2* was identified as upregulated upon LPS stimulation and found to have a role in regulation of gene expression changes upon immune response (Carpenter et al., 2013). Currently no RNA-seq data for immune challenged flies is available as published studies have used microarrays lacking lncRNAs to identify gene expression changes. Nevertheless, increasing numbers of RNA-seq datasets are publically available and could be mined to identify candidate lncRNAs. For example the cell type specific data used

in chapter six of this thesis to define neuroblast or neuron specific genes could be used to identify lncRNAs exclusively expressed in neuroblasts which then could be perturbed and screened for defects in neuroblast development, expansion or transition.

The best approach to perturbation of lncRNAs is still not clear, however the insertion of a terminator sequence using CRISPR/Cas9 deletion combined with homologous recombination may stop transcription with minimal impact on the underlying DNA sequence. However, analysis of multiple transcriptional terminators in a cell culture based system indicated traditional terminators are not all effective in lncRNA loci (Personal communication, A. Bassett) and efficiency *in vivo* may be context dependent (Latos et al., 2012). Alternatively, catalytically inactive Cas9 can be recruited to genomic DNA without causing cleavage of the DNA, and recruitment to the transcriptional start site has been shown to block transcription. This can be modified by the fusion of repressor domains to produce stronger downregulation or activator domains to achieved upregulation in the endogenous locus (Gilbert et al., 2013). If these methods prove effective for lncRNAs, they may provide a platform for a new generation of forward genetic screens for lncRNA function. Similarly, the inclusion of lncRNAs in RNAi screening may provide the opportunity to identify functional lncRNAs. The incorporation of these techniques in modifier screens may prove particularly fruitful given that lncRNAs with subtle phenotypes may have a stronger effect in a sensitised background.

7.3 Future work to confirm *Inc703* function

To confirm that increases in ribosomal levels and decreases in lifespan are caused by disruption of *Inc703* expression, we would first like to generate a

genetic deletion of the splice acceptor site at the beginning of the second exon to recapitulate the premature termination of *Inc703* achieved in the *Inc703^{gal4}* and *Inc703^{stop}* alleles. However, this may result in the usage of cryptic splice sites so a larger deletion may be required. If changes in ribosomal levels are reproduced in a novel *Inc703* mutant, it would be possible to reduce *dm* levels using RNAi to determine if these effects are mediated through increased *dm*. It will also be interesting to determine if ribosome levels are decreased in the long lived flies overexpressing *Inc703* cDNA under the control of *actin5C-GAL4*. Furthermore, as the BAC rescue of *Inc703* fails to reproduce the same expression pattern as wildtype, it will be interesting to use the cDNA transgenes to try and rescue changes in ribosomal levels or gene expression changes in the mutant, to understand the tissue that is critical for these effects.

The assumption has been made throughout this thesis that *Inc703* does not encode for a protein, however it has not been conclusively shown that putative small open reading frames (smORFs) within *Inc703* are not translated. If causality between *Inc703* and ribosomal protein levels is established it will be necessary to mutate the start codons of the four putative smORFs identified by Aspden et al. (2014) to ensure *Inc703* does not act through encoded small peptides.

7.4 Challenges in identifying true miRNA targets

In Chapter six of this thesis, I described our efforts to identify tissue-specific targets of miR-8 using transcriptome profiling of muscle and CNS by determining gene expression changes in the absence of miR-8. Whilst, some previously identified expression changes could be reproduced (i.e. *Drs* and *4EBP*), others could not (i.e. *Spi*, *Wg* and *atrophin*). Furthermore, the expected enrichment for genes upregulated in the miR-8 null mutant to contain MREs was not observed in the

CNS suggesting one of the assumptions: that miR-8 targets would be reduced in expression, that miR-8 interacts with targets through seed matches or that miR-8 has multiple targets in the CNS, is incorrect.

Although there is mounting evidence that the primary mechanism of regulation by miRNA occurs by transcript destabilisation and degradation (Eichhorn et al., 2014; Guo et al., 2010), when using transcriptome profiling methods it is necessary to acknowledge that it may not capture all of the effects of a miRNA. Particularly the effects on targets where the miRNA acts through translational repression; but also genes in which the effect is an attenuation of noise rather than a reduction of steady state levels (Schmiedel et al., 2015). Furthermore, a major challenge still remains in accurate identification of *in vivo* miRNA-target interactions. A number of studies have tried to improve our understanding of the nature of miRNA-target interactions by identifying direct binding between miRNA and target mRNAs. Yet these have revealed an even more complex relationship than previously thought, including interactions in the absence of any seed sequences, and a prevalence of binding in coding sequences and 5'UTRs that contradicts the common idea of miRNA binding to 3'UTRs. Using conditional knockdown or knockout methods, followed by analysis of transcript levels at multiple timepoints may allow for the separation of direct and indirect targets.

It is also possible that some of the gene expression changes detected in this work were not caused by loss of miR-8. I identified an accumulation of the primary miR-8 transcript in the mutant lines, which may be the result of lack of processing of the primary transcript in the absence of the hairpin. The presence of high levels of a long transcript not expressed in wildtype conditions may have a confounding effect the gene expression changes observed and calls in to question the appropriateness of using hairpin deletions to study miRNA function. Furthermore, the presence of antisense transcripts, potentially driven

by the promoter of the *mini-white* marker gene that have the potential to form siRNAs, further complicates the interpretation of this genetic strain. Whilst the full impact of the primary transcript and antisense transcripts is unclear, both miR-8^{Δ1} and miR-8^{Δ2} are hemizygous viable over deficiency chromosomes (Personal communication, Dr Markus Toegel) suggesting the phenotypes observed are not solely caused by the absence of miR-8.

It is not clear if the persistence of the primary transcript is a common occurrence in miRNA mutants. However, in this case it might be more appropriate to delete a large region, including the promoter region or replace the hairpin with a non-targeting hairpin to achieve a more specific mutant. The diverse nature of miRNA biogenesis requires an individualised strategy as promoter deletion would be inappropriate for miRNAs derived from the introns or exons of mRNAs. As described for the study of lncRNAs, the best strategy for understanding *in vivo* functions of miRNAs requires a combinatorial approach. Using available mutant strains (Chen et al., 2014), supported by temporal or spatial specific miRNA sponges (Ebert et al., 2007) and deletion of the predicted effector MREs using genome editing, it may be possible to identify miRNA-target-phenotype interactions (Bassett et al., 2014b).

7.5 Future work to identify true miR-8 targets

I identified more than 100 genes which are upregulated in one or both of the tissues studied and contain one or more predicted miR-8 MRE. However, as was the case with *ilp5*, not all of these genes are likely to be direct targets of miR-8. The intersection of the gene expression changes observed in our data, with equivalent tissue specific PAR-CLiP or CLASH would provide a powerful tool to identify which upregulated transcripts are bound by miR-8 *in vivo* and provide

information about the nature of miRNA-target interaction to improve prediction of miR-8 targets. Dr Markus Toegel is also continuing to study the effects of the persistence of the miR-8 primary transcript and is generating a mutation in which the miR-8 hairpin is mutated into a non-targeting hairpin, so gene expression changes can be confirmed.

7.6 Concluding remarks

In this thesis, I have described one of the first attempts to use a reverse genetics approach to discover the *in vivo* functions of long non-coding RNAs. The success of this approach was limited by the ineffectiveness of transposon insertions to perturb lncRNAs and the background dependency of subtle phenotypes. However, given the growing number of lncRNAs that have functions in cell culture, it is still incredibly important to continue to determine if lncRNAs have functional roles *in vivo*. This work has highlighted the need for the use of multiple complementary approaches when studying lncRNAs. Moving forward, more sophisticated genome engineering techniques are removing the barriers to probe lncRNA function on a larger scale.

References

- (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74.
- Aguila, J.R., Suszko, J., Gibbs, A.G., and Hoshizaki, D.K. (2007). The role of larval fat cells in adult *Drosophila melanogaster*. *The Journal of experimental biology* 210, 956-963.
- Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during *Drosophila melanogaster* development. *Developmental cell* 5, 337-350.
- Aspden, J.L., Eyre-Walker, Y.C., Phillips, R.J., Amin, U., Mumtaz, M.A., Brocard, M., and Couso, J.P. (2014). Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *eLife* 3, e03528.
- Bassett, A.R., Akhtar, A., Barlow, D.P., Bird, A.P., Brockdorff, N., Duboule, D., Ephrussi, A., Ferguson-Smith, A.C., Gingeras, T.R., Haerty, W., *et al.* (2014a). Considerations when investigating lncRNA function in vivo. *eLife* 3, e03058.
- Bassett, A.R., Azzam, G., Wheatley, L., Tibbit, C., Rajakumar, T., McGowan, S., Stanger, N., Ewels, P.A., Taylor, S., Ponting, C.P., *et al.* (2014b). Understanding functional miRNA-target interactions in vivo by site-specific genome engineering. *Nature communications* 5, 4640.
- Bassett, A.R., Tibbit, C., Ponting, C.P., and Liu, J.L. (2013). Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell reports* 4, 220-228.
- Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes & development* 20, 1885-1898.
- Bellen, H.J., Levis, R.W., He, Y., Carlson, J.W., Evans-Holm, M., Bae, E., Kim, J., Metaxakis, A., Savakis, C., Schulze, K.L., *et al.* (2011). The *Drosophila* gene disruption project: progress using transposons with distinctive site specificities. *Genetics* 188, 731-743.
- Bellen, H.J., Levis, R.W., Liao, G., He, Y., Carlson, J.W., Tsang, G., Evans-Holm, M., Hiesinger, P.R., Schulze, K.L., Rubin, G.M., *et al.* (2004). The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics* 167, 761-781.
- Bellen, H.J., O'Kane, C.J., Wilson, C., Grossniklaus, U., Pearson, R.K., and Gehring, W.J. (1989). P-element-mediated enhancer detection: a versatile method to study development in *Drosophila*. *Genes & development* 3, 1288-1300.
- Berezikov, E., Robine, N., Samsonova, A., Westholm, J.O., Naqvi, A., Hung, J.H., Okamura, K., Dai, Q., Bortolamiol-Becet, D., Martin, R., *et al.* (2011). Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence. *Genome research* 21, 203-215.
- Berger, C., Harzer, H., Burkard, T.R., Steinmann, J., van der Horst, S., Laurenson, A.S., Novatchkova, M., Reichert, H., and Knoblich, J.A. (2012). FACS purification and transcriptome analysis of *drosophila* neural stem cells reveals a role for Klumpfuss in self-renewal. *Cell reports* 2, 407-418.
- Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118, 401-415.
- Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Molecular cell* 44, 120-133.
- Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. *RNA* 19, 429-442.
- Brown, J.B., Boley, N., Eisman, R., May, G.E., Stoiber, M.H., Duff, M.O., Booth, B.W., Wen, J., Park, S., Suzuki, A.M., *et al.* (2014). Diversity and dynamics of the *Drosophila* transcriptome. *Nature* 512, 393-399.

- Bulet, P., Hetru, C., Dimarcq, J.L., and Hoffmann, D. (1999). Antimicrobial peptides in insects; structure and function. *Developmental and comparative immunology* 23, 329-344.
- Cabili, M.N., Dunagin, M.C., McClanahan, P.D., Bialesch, A., Padovan-Merhar, O., Regev, A., Rinn, J.L., and Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome biology* 16, 20.
- Cantera, R., Roos, E., and Engstrom, Y. (1999). Dif and cactus are colocalized in the larval nervous system of *Drosophila melanogaster*. *Journal of neurobiology* 38, 16-26.
- Carpenter, S., Aiello, D., Atianand, M.K., Ricci, E.P., Gandhi, P., Hall, L.L., Byron, M., Monks, B., Henry-Bezy, M., Lawrence, J.B., *et al.* (2013). A long noncoding RNA mediates both activation and repression of immune response genes. *Science* 341, 789-792.
- Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 147, 358-369.
- Chalei, V., Sansom, S.N., Kong, L., Lee, S., Montiel, J.F., Vance, K.W., and Ponting, C.P. (2014). The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. *eLife* 3, e04530.
- Chekulaeva, M., Filipowicz, W., and Parker, R. (2009). Multiple independent domains of dGW182 function in miRNA-mediated repression in *Drosophila*. *RNA* 15, 794-803.
- Chen, L.L., and Carmichael, G.G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Molecular cell* 35, 467-478.
- Chen, J., Wang, L., Matyunina, L.V., Hill, C.G., and McDonald, J.F. (2011). Overexpression of miR-429 induces mesenchymal-to-epithelial transition (MET) in metastatic ovarian cancer cells. *Gynecologic oncology* 121, 200-205.
- Chen, Y.W., Song, S., Weng, R., Verma, P., Kugler, J.M., Buescher, M., Rouam, S., and Cohen, S.M. (2014). Systematic study of *Drosophila* microRNA functions using a collection of targeted knockout mutations. *Developmental cell* 31, 784-800.
- Chi, S.W., Hannon, G.J., and Darnell, R.B. (2012). An alternative mode of microRNA target recognition. *Nature structural & molecular biology* 19, 321-327.
- Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 460, 479-486.
- Choi, I.K., and Hyun, S. (2012). Conserved microRNA miR-8 in fat body regulates innate immune homeostasis in *Drosophila*. *Developmental and comparative immunology* 37, 50-54.
- Choi, W.Y., Giraldez, A.J., and Schier, A.F. (2007). Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430. *Science* 318, 271-274.
- Christoffersen, N.R., Silahtaroglu, A., Orom, U.A., Kauppinen, S., and Lund, A.H. (2007). miR-200b mediates post-transcriptional repression of ZFX1B. *RNA* 13, 1172-1178.
- Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular cell* 44, 667-678.
- Clark, R.I., Tan, S.W., Pean, C.B., Roostalu, U., Vivancos, V., Bronda, K., Pilatova, M., Fu, J., Walker, D.W., Berdeaux, R., *et al.* (2013). MEF2 is an in vivo immune-metabolic switch. *Cell* 155, 435-447.
- Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Molecular cell* 33, 717-726.
- Colombani, J., Raisin, S., Pantalacci, S., Radimerski, T., Montagne, J., and Leopold, P. (2003). A

nutrient sensor mechanism controls *Drosophila* growth. *Cell* **114**, 739-749.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., *et al.* (2015). Ensembl 2015. *Nucleic acids research* **43**, D662-669.

Daines, B., Wang, H., Wang, L., Li, Y., Han, Y., Emmert, D., Gelbart, W., Wang, X., Li, W., Gibbs, R., *et al.* (2011). The *Drosophila melanogaster* transcriptome by paired-end RNA-seq. *Genome research* **21**, 315-324.

Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K.J., Gooding, A.R., Lee, J.T., and Cech, T.R. (2015). Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Molecular cell* **57**, 552-558.

Delanoue, R., Slaidina, M., and Leopold, P. (2010). The steroid hormone ecdysone controls systemic growth by repressing dMyc function in *Drosophila* fat cells. *Developmental cell* **18**, 1012-1021.

Denzler, R., Agarwal, V., Stefano, J., Bartel, D.P., and Stoffel, M. (2014). Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Molecular cell* **54**, 766-776.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., *et al.* (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* **22**, 1775-1789.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., *et al.* (2012). Landscape of transcription in human cells. *Nature* **489**, 101-108.

Dunagin, M., Cabili, M.N., Rinn, J., and Raj, A. (2015). Visualization of lncRNA by single-molecule fluorescence in situ hybridization. *Methods Mol Biol* **1262**, 3-19.

Ebert, M.S., Neilson, J.R., and Sharp, P.A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature methods* **4**, 721-726.

Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S.H., Ghoshal, K., Villen, J., and Bartel, D.P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Molecular cell* **56**, 104-115.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in *Drosophila*. *Genome biology* **5**, R1.

Essers, P.B., Nonnekens, J., Goos, Y.J., Betist, M.C., Viester, M.D., Mossink, B., Lansu, N., Korswagen, H.C., Jelier, R., Brenkman, A.B., *et al.* (2015). A Long Noncoding RNA on the Ribosome Is Required for Lifespan Extension. *Cell reports*.

Fortier, T.M., Vasa, P.P., and Woodard, C.T. (2003). Orphan nuclear receptor betaFTZ-F1 is required for muscle-driven morphogenetic events at the prepupal-pupal transition in *Drosophila melanogaster*. *Developmental biology* **257**, 153-165.

Franke, A., and Baker, B.S. (1999). The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Molecular cell* **4**, 117-122.

Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome research* **19**, 92-105.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., *et al.* (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451.

Gilbert, W. (1993). A vision of the Grail. In *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, D.J. Kevles, and L.E. Hood, eds. (Cambridge: Harvard University Press),

pp. 83-87.

Gong, C., Popp, M.W., and Maquat, L.E. (2012). Biochemical analysis of long non-coding RNA-containing ribonucleoprotein complexes. *Methods* 58, 88-93.

Govind, S. (1999). Control of development and immunity by rel transcription factors in *Drosophila*. *Oncogene* 18, 6875-6887.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., *et al.* (2011). The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471, 473-479.

Gregory, P.A., Bert, A.G., Paterson, E.L., Barry, S.C., Tsykin, A., Farshid, G., Vadas, M.A., Khew-Goodall, Y., and Goodall, G.J. (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nature cell biology* 10, 593-601.

Grewal, S.S., Li, L., Orian, A., Eisenman, R.N., and Edgar, B.A. (2005). Myc-dependent regulation of ribosomal RNA synthesis during *Drosophila* development. *Nature cell biology* 7, 295-302.

Griffin, R., Sustar, A., Bonvin, M., Binari, R., del Valle Rodriguez, A., Hohl, A.M., Bateman, J.R., Villalta, C., Heffern, E., Grunwald, D., *et al.* (2009). The twin spot generator for differential *Drosophila* lineage analysis. *Nature methods* 6, 600-602.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Wahrisch, S., Beisaw, A., Macura, K., Blass, G., Kellis, M., Werber, M., *et al.* (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Developmental cell* 24, 206-214.

Grun, D., Wang, Y.L., Langenberger, D., Gunsalus, K.C., and Rajewsky, N. (2005). microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLoS computational biology* 1, e13.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835-840.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., *et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223-227.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., *et al.* (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477, 295-300.

Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., and Lander, E.S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154, 240-251.

Hacker, U., Nystedt, S., Barmchi, M.P., Horn, C., and Wimmer, E.A. (2003). piggyBac-based insertional mutagenesis in the presence of stably integrated P elements in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* 100, 7720-7725.

Haerty, W., and Ponting, C.P. (2013). Mutations within lncRNAs are effectively selected against in fruitfly but not in human. *Genome biology* 14, R49.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., *et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129-141.

Halton, M.S., and Keshishian, H. (1998). The Toll pathway is required in the epidermis for muscle development in the *Drosophila* embryo. *Developmental biology* 199, 164-174.

Hayes, J., Peruzzi, P.P., and Lawler, S. (2014). MicroRNAs in cancer: biomarkers, functions and therapy. *Trends in molecular medicine* 20, 460-469.

- Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153, 654-665.
- Hogg, J.R., and Goff, S.P. (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. *Cell* 143, 379-389.
- Hoshizaki, D.K., Lunz, R., Ghosh, M., and Johnson, W. (1995). Identification of fat-cell enhancer activity in *Drosophila melanogaster* using P-element enhancer traps. *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada* 38, 497-506.
- Houtkooper, R.H., Mouchiroud, L., Ryu, D., Moullan, N., Katsyuba, E., Knott, G., Williams, R.W., and Auwerx, J. (2013). Mitonuclear protein imbalance as a conserved longevity mechanism. *Nature* 497, 451-457.
- Hummel, T., and Klambt, C. (2008). P-element mutagenesis. *Methods Mol Biol* 420, 97-117.
- Hurteau, G.J., Carlson, J.A., Spivack, S.D., and Brock, G.J. (2007). Overexpression of the microRNA hsa-miR-200c leads to reduced expression of transcription factor 8 and increased expression of E-cadherin. *Cancer research* 67, 7972-7976.
- Hyun, S., Lee, J.H., Jin, H., Nam, J., Namkoong, B., Lee, G., Chung, J., and Kim, V.N. (2009). Conserved MicroRNA miR-8/miR-200 and its target USH/FOG2 control growth by regulating PI3K. *Cell* 139, 1096-1108.
- Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S., *et al.* (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nature genetics* 47, 199-208.
- Ji, P., Diederichs, S., Wang, W., Boing, S., Metzger, R., Schneider, P.M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., *et al.* (2003). MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 22, 8031-8041.
- Ji, Q., Zhang, L., Liu, X., Zhou, L., Wang, W., Han, Z., Sui, H., Tang, Y., Wang, Y., Liu, N., *et al.* (2014). Long non-coding RNA MALAT1 promotes tumour growth and metastasis in colorectal cancer through binding to SFPQ and releasing oncogene PTBP2 from SFPQ/PTBP2 complex. *British journal of cancer* 111, 736-748.
- Jia, H., Osak, M., Bogu, G.K., Stanton, L.W., Johnson, R., and Lipovich, L. (2010). Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA* 16, 1478-1487.
- Jin, H., Kim, V.N., and Hyun, S. (2012). Conserved microRNA miR-8 controls body size in response to steroid signaling in *Drosophila*. *Genes & development* 26, 1427-1432.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821.
- Johnston, L.A., Prober, D.A., Edgar, B.A., Eisenman, R.N., and Gallant, P. (1999). *Drosophila* myc regulates cellular growth during development. *Cell* 98, 779-790.
- Kaiser, K. (1993). Second generation enhancer traps. *Current biology : CB* 3, 560-562.
- Karres, J.S., Hilgers, V., Carrera, I., Treisman, J., and Cohen, S.M. (2007). The conserved microRNA miR-8 tunes atrophin levels to prevent neurodegeneration in *Drosophila*. *Cell* 131, 136-145.
- Kennell, J.A., Cadigan, K.M., Shakhmantsir, I., and Waldron, E.J. (2012). The microRNA miR-8 is a positive regulator of pigmentation and eclosion in *Drosophila*. *Developmental dynamics : an official publication of the American Association of Anatomists* 241, 161-168.
- Kennell, J.A., Gerin, I., MacDougald, O.A., and Cadigan, K.M. (2008). The microRNA miR-8 is a conserved negative regulator of Wnt signaling. *Proceedings of the National Academy of Sciences*

of the United States of America *105*, 15417-15422.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* *14*, R36.

Kong, L., Zhang, Y., Ye, Z.Q., Liu, X.Q., Zhao, S.Q., Wei, L., and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic acids research* *35*, W345-349.

Korpala, M., Lee, E.S., Hu, G., and Kang, Y. (2008). The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *The Journal of biological chemistry* *283*, 14910-14914.

Krasnov, A.N., Kurshakova, M.M., Ramensky, V.E., Mardanov, P.V., Nabirochkina, E.N., and Georgieva, S.G. (2005). A retrocopy of a gene can functionally displace the source gene in evolution. *Nucleic acids research* *33*, 6654-6661.

Kumar, V., Westra, H.J., Karjalainen, J., Zhernakova, D.V., Esko, T., Hrdlickova, B., Almeida, R., Zhernakova, A., Reinmaa, E., Vosa, U., *et al.* (2013). Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS genetics* *9*, e1003201.

Kutter, C., Watt, S., Stefflova, K., Wilson, M.D., Goncalves, A., Ponting, C.P., Odom, D.T., and Marques, A.C. (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS genetics* *8*, e1002841.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yanez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* *512*, 91-95.

Laferte, A., Favry, E., Sentenac, A., Riva, M., Carles, C., and Chedin, S. (2006). The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components. *Genes & development* *20*, 2030-2040.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science* *294*, 853-858.

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Current biology : CB* *14*, 2162-2167.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature protocols* *8*, 2180-2196.

Latos, P.A., Pauler, F.M., Koerner, M.V., Senergin, H.B., Hudson, Q.J., Stocsits, R.R., Allhoff, W., Stricker, S.H., Klement, R.M., Warczok, K.E., *et al.* (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* *338*, 1469-1472.

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

Lee, G.J., and Hyun, S. (2014). Multiple targets of the microRNA miR-8 contribute to immune homeostasis in *Drosophila*. *Developmental and comparative immunology* *45*, 245-251.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* *75*, 843-854.

Lee, T., and Luo, L. (2001). Mosaic analysis with a repressible cell marker (MARCM) for *Drosophila* neural development. *Trends in neurosciences* *24*, 251-254.

- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., *et al.* (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-419.
- Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *The EMBO journal* **21**, 4663-4670.
- Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. (2004a). MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal* **23**, 4051-4060.
- Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004b). Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69-81.
- Lemaitre, B., and Hoffmann, J. (2007). The host defense of *Drosophila melanogaster*. *Annual review of immunology* **25**, 697-743.
- Letsou, A., Alexander, S., Orth, K., and Wasserman, S.A. (1991). Genetic and molecular characterization of tube, a *Drosophila* gene maternally required for embryonic dorsoventral polarity. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 810-814.
- Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20.
- Lin, M.F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275-282.
- Lin, R., Maeda, S., Liu, C., Karin, M., and Edgington, T.S. (2007). A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene* **26**, 851-858.
- Livvyatan, I., Harikumar, A., Nissim-Rafinia, M., Dutttagupta, R., Gingeras, T.R., and Meshorer, E. (2013). Non-polyadenylated transcription in embryonic stem cells reveals novel non-coding RNA related to pluripotency and differentiation. *Nucleic acids research* **41**, 6300-6315.
- Long, D., Lee, R., Williams, P., Chan, C.Y., Ambros, V., and Ding, Y. (2007). Potent effect of target structure on microRNA function. *Nature structural & molecular biology* **14**, 287-294.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550.
- Loya, C.M., Lu, C.S., Van Vactor, D., and Fulga, T.A. (2009). Transgenic microRNA inhibition with spatiotemporal specificity in intact organisms. *Nature methods* **6**, 897-903.
- Loya, C.M., McNeill, E.M., Bao, H., Zhang, B., and Van Vactor, D. (2014). miR-8 controls synapse structure by repression of the actin regulator enabled. *Development* **141**, 1864-1874.
- Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. *Science* **303**, 95-98.
- Marques, A.C., Hughes, J., Graham, B., Kowalczyk, M.S., Higgs, D.R., and Ponting, C.P. (2013). Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. *Genome biology* **14**, R131.
- Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* **429**, 571-574.
- Mattaliano, M.D., Montana, E.S., Parisky, K.M., Littleton, J.T., and Griffith, L.C. (2007). The *Drosophila* ARC homolog regulates behavioral responses to starvation. *Molecular and cellular neurosciences* **36**, 211-221.
- Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P.,

- Sandstrom, R., Qu, H., Brody, J., *et al.* (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195.
- McKinlay, A., Araya, C.L., and Fields, S. (2011). Genome-Wide Analysis of Nascent Transcription in *Saccharomyces cerevisiae*. *G3 (Bethesda)* **1**, 549-558.
- Meador, S., Ponting, C.P., and Lunter, G. (2010). Massive turnover of functional sequence in human and other mammalian genomes. *Genome research* **20**, 1335-1343.
- Meller, V.H., and Rattner, B.P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *The EMBO journal* **21**, 1084-1091.
- Metaxakis, A., Oehler, S., Klinakis, A., and Savakis, C. (2005). Minos as a genetic and genomic tool in *Drosophila melanogaster*. *Genetics* **171**, 571-581.
- Morante, J., Vallejo, D.M., Desplan, C., and Dominguez, M. (2013). Conserved miR-8/miR-200 defines a glial niche that controls neuroepithelial expansion and neuroblast transition. *Developmental cell* **27**, 174-187.
- Morata, G., and Ripoll, P. (1975). Minutes: mutants of *drosophila* autonomously affecting cell division rate. *Developmental biology* **42**, 211-221.
- Moreno, E., and Basler, K. (2004). dMyc transforms cells into super-competitors. *Cell* **117**, 117-129.
- Moreno, E., Basler, K., and Morata, G. (2002). Cells compete for decapentaplegic survival factor to prevent apoptosis in *Drosophila* wing development. *Nature* **416**, 755-759.
- Nakagawa, S., Naganuma, T., Shioi, G., and Hirose, T. (2011). Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *The Journal of cell biology* **193**, 31-39.
- Nam, J.W., and Bartel, D.P. (2012). Long noncoding RNAs in *C. elegans*. *Genome research* **22**, 2529-2540.
- Nesler, K.R., Sand, R.I., Symmes, B.A., Pradhan, S.J., Boin, N.G., Laun, A.E., and Barbee, S.A. (2013). The miRNA pathway controls rapid changes in activity-dependent synaptic structure at the *Drosophila melanogaster* neuromuscular junction. *PloS one* **8**, e68385.
- Neufeld, T.P. (2008). Genetic manipulation and monitoring of autophagy in *Drosophila*. *Methods in enzymology* **451**, 653-667.
- Nicolas, F.E., Pais, H., Schwach, F., Lindow, M., Kauppinen, S., Moulton, V., and Dalmay, T. (2008). Experimental identification of microRNA-140 targets by silencing and overexpressing miR-140. *RNA* **14**, 2513-2520.
- Ning, S., Zhao, Z., Ye, J., Wang, P., Zhi, H., Li, R., Wang, T., Wang, J., Wang, L., and Li, X. (2014). SNP@lincTFBS: an integrated database of polymorphisms in human lincRNA transcription factor binding sites. *PloS one* **9**, e103851.
- Ntini, E., Jarvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jorgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R., *et al.* (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nature structural & molecular biology* **20**, 923-928.
- O'Connell, M.R., Oakes, B.L., Sternberg, S.H., East-Seletsky, A., Kaplan, M., and Doudna, J.A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**, 263-266.
- Okamura, K., Liu, N., and Lai, E.C. (2009). Distinct mechanisms for microRNA strand selection by *Drosophila* Argonautes. *Molecular cell* **36**, 431-444.
- Okamura, K., Phillips, M.D., Tyler, D.M., Duan, H., Chou, Y.T., and Lai, E.C. (2008). The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. *Nature structural & molecular biology* **15**, 354-363.

Olson, E.N. (2014). MicroRNAs as therapeutic targets and biomarkers of cardiovascular disease. *Science translational medicine* 6, 239ps233.

Orian, A., van Steensel, B., Delrow, J., Bussemaker, H.J., Li, L., Sawado, T., Williams, E., Loo, L.W., Cowley, S.M., Yost, C., *et al.* (2003). Genomic binding by the Drosophila Myc, Max, Mad/Mnt transcription factor network. *Genes & development* 17, 1101-1114.

Parisi, F., Riccardo, S., Zola, S., Lora, C., Grifoni, D., Brown, L.M., and Bellosta, P. (2013). dMyc expression in the fat body affects DILP2 release and increases the expression of the fat desaturase Desat1 resulting in organismal growth. *Developmental biology* 379, 64-75.

Park, S.M., Gaur, A.B., Lengyel, E., and Peter, M.E. (2008). The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes & development* 22, 894-907.

Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A., *et al.* (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome research* 22, 577-591.

Pease, B., Borges, A.C., and Bender, W. (2013). Noncoding RNAs of the Ultrabithorax domain of the Drosophila bithorax complex. *Genetics* 195, 1253-1264.

Petruk, S., Sedkov, Y., Riley, K.M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J.B., Brock, H.W., and Mazo, A. (2006). Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell* 127, 1209-1221.

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J., and Pandolfi, P.P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033-1038.

Ponting, C.P., and Hardison, R.C. (2011). What fraction of the human genome is functional? *Genome research* 21, 1769-1776.

Prasanth, K.V., Rajendra, T.K., Lal, A.K., and Lakhota, S.C. (2000). Omega speckles - a novel class of nuclear speckles containing hnRNPs associated with noncoding hsr-omega RNA in Drosophila. *Journal of cell science* 113 Pt 19, 3485-3497.

Puerta-Fernandez, E., Barrick, J.E., Roth, A., and Breaker, R.R. (2006). Identification of a large noncoding RNA in extremophilic eubacteria. *Proceedings of the National Academy of Sciences of the United States of America* 103, 19490-19495.

Pueyo, J.I., and Couso, J.P. (2008). The 11-aminoacid long Tarsal-less peptides trigger a cell signal in Drosophila leg development. *Developmental biology* 324, 192-201.

Qiu, P., Pan, P.C., and Govind, S. (1998). A role for the Drosophila Toll/Cactus pathway in larval hematopoiesis. *Development* 125, 1909-1920.

Quinn, J.J., Ilik, I.A., Qu, K., Georgiev, P., Chu, C., Akhtar, A., and Chang, H.Y. (2014). Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nature biotechnology* 32, 933-940.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods* 5, 877-879.

Rajewsky, N. (2006). microRNA target predictions in animals. *Nature genetics* 38 Suppl, S8-13.

Rands, C.M., Meader, S., Ponting, C.P., and Lunter, G. (2014). 8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS genetics* 10, e1004525.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., *et al.* (2007). Functional demarcation of active and silent chromatin

domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311-1323.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L., and Bradley, A. (2004). Identification of mammalian microRNA host genes and transcription units. *Genome research* 14, 1902-1910.

Ruby, J.G., Stark, A., Johnston, W.K., Kellis, M., Bartel, D.P., and Lai, E.C. (2007). Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome research* 17, 1850-1864.

Rynes, J., Donohoe, C.D., Frommolt, P., Brodesser, S., Jindra, M., and Uhlirova, M. (2012). Activating transcription factor 3 regulates immune and metabolic homeostasis. *Molecular and cellular biology* 32, 3949-3962.

Sahu, N.K., Shilakari, G., Nayak, A., and Kohli, D.V. (2007). Antisense technology: a selective tool for gene expression regulation and gene targeting. *Current pharmaceutical biotechnology* 8, 291-304.

Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Hacisuleyman, E., Li, E., Spence, M., *et al.* (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* 2, e01749.

Schmiedel, J.M., Klemm, S.L., Zheng, Y., Sahay, A., Bluthgen, N., Marks, D.S., and van Oudenaarden, A. (2015). Gene expression. MicroRNA control of protein expression noise. *Science* 348, 128-132.

Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T., *et al.* (2015). Evaluation of tools for differential gene expression analysis by RNA-seq on a 48 biological replicate experiment. *arXiv*.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115, 199-208.

Scott, R.C., Schuldiner, O., and Neufeld, T.P. (2004). Role and regulation of starvation-induced autophagy in the *Drosophila* fat body. *Developmental cell* 7, 167-178.

Silies, M., Yuva, Y., Engelen, D., Aho, A., Stork, T., and Klambt, C. (2007). Glial cell migration in the eye disc. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27, 13130-13139.

Simon, M.D., Wang, C.I., Kharchenko, P.V., West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). The genomic binding sites of a noncoding RNA. *Proceedings of the National Academy of Sciences of the United States of America* 108, 20497-20502.

Slaidina, M., Delanoue, R., Gronke, S., Partridge, L., and Leopold, P. (2009). A *Drosophila* insulin-like peptide promotes growth during nonfeeding states. *Developmental cell* 17, 874-884.

Spitzer, J., Hafner, M., Landthaler, M., Ascano, M., Farazi, T., Wardle, G., Nusbaum, J., Khorshid, M., Burger, L., Zavolan, M., *et al.* (2014). PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods in enzymology* 539, 113-161.

Standaert, L., Adriaens, C., Radaelli, E., Van Keymeulen, A., Blanpain, C., Hirose, T., Nakagawa, S., and Marine, J.C. (2014). The long noncoding RNA Neat1 is required for mammary gland development and lactation. *RNA* 20, 1844-1849.

Steffen, K.K., MacKay, V.L., Kerr, E.O., Tsuchiya, M., Hu, D., Fox, L.A., Dang, N., Johnston, E.D., Oakes, J.A., Tchao, B.N., *et al.* (2008). Yeast life span extension by depletion of 60s ribosomal subunits is mediated by Gcn4. *Cell* 133, 292-302.

- Steller, H., and Pirrotta, V. (1985). A transposable P vector that confers selectable G418 resistance to *Drosophila* larvae. *The EMBO journal* 4, 167-171.
- Steward, R., McNally, F.J., and Schedl, P. (1984). Isolation of the dorsal locus of *Drosophila*. *Nature* 311, 262-265.
- Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508, 66-71.
- Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S., and Spector, D.L. (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome research* 19, 347-359.
- Tan, J.Y., Sirey, T., Honti, F., Graham, B., Piovesan, A., Merckenschlager, M., Webber, C., Ponting, C.P., and Marques, A.C. (2015). Extensive microRNA-mediated crosstalk between lncRNAs and mRNAs in mouse embryonic stem cells. *Genome research* 25, 655-666.
- Tan, J.Y., Vance, K.W., Varela, M.A., Sirey, T., Watson, L.M., Curtis, H.J., Marinello, M., Alves, S., Steinkraus, B.R., Cooper, S., *et al.* (2014). Cross-talking noncoding RNAs contribute to cell-specific neurodegeneration in SCA7. *Nature structural & molecular biology* 21, 955-961.
- Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan, S.M., Ala, U., Karreth, F., Poliseno, L., Provero, P., Di Cunto, F., *et al.* (2011). Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* 147, 344-357.
- Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 455, 1124-1128.
- Teixeira, L. (2012). Whole-genome expression profile analysis of *Drosophila melanogaster* immune responses. *Briefings in functional genomics* 11, 375-386.
- Tian, A., and Jiang, J. (2014). Intestinal epithelium-derived BMP controls stem cell self-renewal in *Drosophila* adult midgut. *eLife* 3, e01857.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105-1111.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* 28, 511-515.
- Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689-693.
- Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., Seal, R., *et al.* (2009). FlyBase: enhancing *Drosophila* Gene Ontology annotations. *Nucleic acids research* 37, D555-559.
- Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26-46.
- Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147, 1537-1550.
- Unckless, R.L., Rottschaefer, S.M., and Lazzaro, B.P. (2015). The complex contributions of genetics and nutrition to immunity in *Drosophila melanogaster*. *PLoS genetics* 11, e1005030.
- van de Wetering, M., Cavallo, R., Dooijes, D., van Beest, M., van Es, J., Loureiro, J., Ypma, A., Hursh, D., Jones, T., Bejsovec, A., *et al.* (1997). Armadillo coactivates transcription driven by the

product of the *Drosophila* segment polarity gene dTCF. *Cell* 88, 789-799.

Vance, K.W., Sansom, S.N., Lee, S., Chalei, V., Kong, L., Cooper, S.E., Oliver, P.L., and Ponting, C.P. (2014). The long non-coding RNA Paupar regulates the expression of both local and distal genes. *The EMBO journal* 33, 296-311.

Vert, J.P., Foveau, N., Lajaunie, C., and Vandenbrouck, Y. (2006). An accurate and interpretable model for siRNA efficacy prediction. *BMC bioinformatics* 7, 520.

Wang, F., Ren, S., Chen, R., Lu, J., Shi, X., Zhu, Y., Zhang, W., Jing, T., Zhang, C., Shen, J., *et al.* (2014). Development and prospective multicenter evaluation of the long noncoding RNA MALAT-1 as a diagnostic urinary biomarker for prostate cancer. *Oncotarget* 5, 11091-11102.

Wang, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. *Molecular cell* 43, 904-914.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., *et al.* (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472, 120-124.

Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., Wu, M., Xiong, J., Guo, X., and Liu, H. (2013). Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Developmental cell* 25, 69-80.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B., *et al.* (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121, 725-737.

Xu, C., Yang, M., Tian, J., Wang, X., and Li, Z. (2011). MALAT-1: a long non-coding RNA and its important 3' end functional motif in colorectal cancer metastasis. *International journal of oncology* 39, 169-175.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033-1037.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & development* 17, 3011-3016.

Young, R.S., Marques, A.C., Tibbit, C., Haerty, W., Bassett, A.R., Liu, J.L., and Ponting, C.P. (2012). Identification and properties of 1,119 candidate lincRNA loci in the *Drosophila melanogaster* genome. *Genome biology and evolution* 4, 427-442.

Yuva-Aydemir, Y., Xu, X.L., Aydemir, O., Gascon, E., Sayin, S., Zhou, W., Hong, Y., and Gao, F.B. (2015). Downregulation of the Host Gene *jigr1* by miR-92 Is Essential for Neuroblast Self-Renewal in *Drosophila*. *PLoS genetics* 11, e1005264.

Zekri, L., Kuzuoğlu-Öztürk, D., and Izaurralde, E. (2013). GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. *The EMBO journal* 32, 1052-1065.

Zhang, B., Arun, G., Mao, Y.S., Lazar, Z., Hung, G., Bhattacharjee, G., Xiao, X., Booth, C.J., Wu, J., Zhang, C., *et al.* (2012). The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell reports* 2, 111-123.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Molecular cell* 40, 939-953.

Appendices

Appendix A – Primers used for cloning, colony PCR and sequencing

Primer	Sequence 5' – 3'	Product size	Application
M13F	TGTAAAACGACGGCCAGT		Used to sequence inserts in pGEM-T vector
M13R	CAGGAAACAGCTATGACC		Used to sequence inserts in pGEM-T vector
UASp-F	TTGATTCTCAATTGTGCTCGGCAACAG	674	To amplify UASp from pWALIUM22
UASp-R	TTGCGGCCGCGCGGCTGAATATGGGAT-GTT		
SV40_F	TCTGATCTGCTAGACAATTGTTGGCA		To amplify SV40 sequence from mENTRY-SV40 vector
SV40_R	GTTTGTCCAAACTCATCAATGTA		
pBID-F	AATTGTGCTCGGCAACAGTA		For colony PCR to confirm insertions
pBID-R	TTTGCCGAAAATATGCAATG		
pBID_seq	GGATCCTGGCCACGTAATAA		To sequence pBID to confirm insertions
Inc703_del45_F	AAGGTTCCCTTTTTATTTGGCTTAC-GAAAACTGTCTTCCGATGTC		To make 323bp deletion in <i>Inc703</i> cDNA transgene
Inc703_del45_R	CACATCGGAAGACAGTTTTCGTAAGC-CAAATAAAAAGGGAACCTT		
Del_colony_F	CGCAGAGGTGCGTCAGAAAGGT	232 or 555	To screen for deletions in <i>Inc703</i> cDNA
Del_colony_R	TCCGTTGATTTGCGGGGCTA		
shRNA2_sense	CTAGCAGTGGTTCACCTGTGCTACGG-TATAGTTATATTCAAGCATATACCGTAGCA-CAGTGAACCGCG		Sense and antisense strands for shRNA2, targets base 448 of <i>Inc703</i> .
shRNA2_anti	AATTCGCGGGTTCACCTGTGCTACGGTATAT-GCTTGAATATAACTATACCGTAGCACAGT-GAACCACTG		
shRNA10_sense	CTAGCAGTGAA-CAAATATATATTGAAATAGT-TATATTCAAGCATATTTCAATATATATTT-GTTCGCG		Sense and antisense strands for shRNA10, targets base 1155 of <i>Inc703</i> .
shRNA10_anti	AATTCGCGAACAATATATATTGAAATAT-GCTTGAATATAACTATTTCAATATATATTT-GTTCACCTG		

shRNA11_ sense	CTAGCAGTCCA ACTATGTGGAGTC- CA ATAGTTATATTCAAGCATATTGGACTC- CAC ATAGTTGGGCG		Sense and antisense strands for shRNA11, targets base 752 of <i>Inc703</i> .
shRNA11_ anti	AATTCGCCCC AA CTATGTGGAGTCCAATAT- GCTTGAATATAACTATTGGACTCCA- CAT AGTTGGACTG		
pWALIUM22- seq	GGTGATAGAGCCTGAACCAG		To sequence pWALIUM22 to confirm insertion of shRNA sequences

Appendix B – Primers used for qRT-PCR experiments

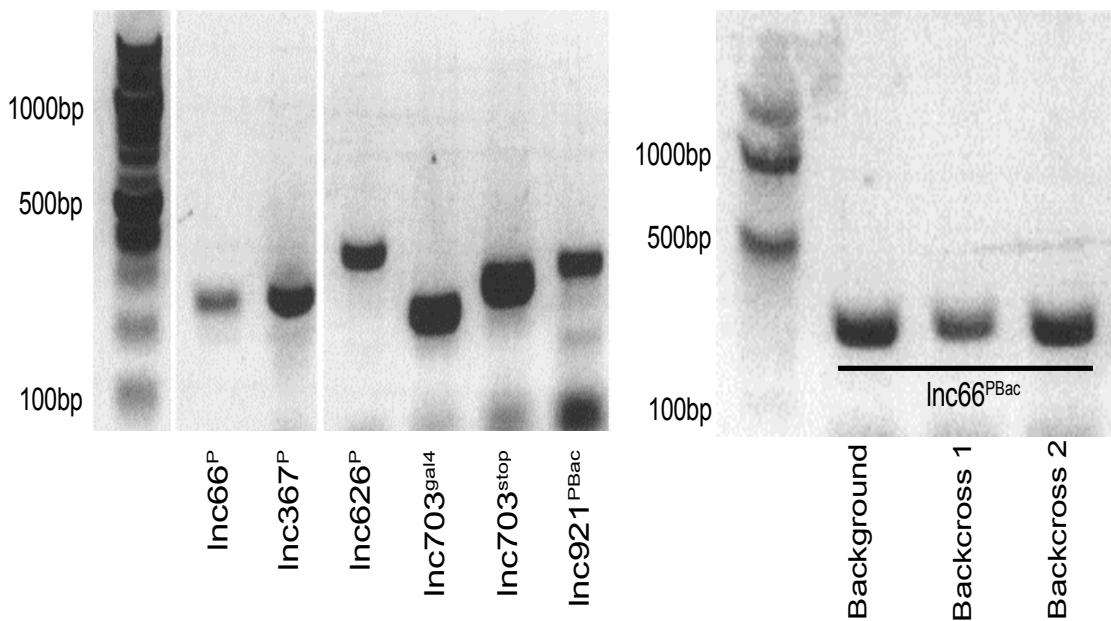
Gene	Sequence 5' – 3'	Product size
actin5C_F	TCCGCTGCCCCGAGGCCCTG	147
actin5C_R	TGGTGCCACCGGACAGCACGG	
Alh_F	CTGCCACGACCACTGAGAGC	174
Alh_R	GAGATGTTGTCGCGGGCACT	
αtub84B_F	GCCAAGCGTGCTTCGTCCA	202
αtub84B_R	CCGCCCACCAATGACGCTCC	
Arc1_F	ATGGCCCAGCTTACACAGATG	138
Arc1_R	GGAGAAGTTGCCTTTGCCTC	
attA_iQ_F	TACCGAGGCACTTCCCACAACA	Rynes et al. 2012
attA_iQ_R	AAAGAGGCACCATGACCAGCAT	
attD_iQ_F	TAAGGGTTCGGTGATGATCTTGC	Rynes et al. 2012
attD_iQ_R	GAAATGCAGTGGCATTGAGAGC	
bip2_F	TGCCACCAATGCAGGACCCC	228
bip2_R	CCACGCTGCTCATTTGCGGG	
CaMKI_F	ACAGCAGCGACTGGAGCGTG	102
CaMKI_R	TGCATTTGTGCTGCCACCCGA	
CG6933_F	ACTGCCATGCCTGGGGATAC	138
CG6933_R	TCTCAACGGCACTGGTGGAG	
CG7429_F	TCCCACCTGGCTTTTGCTCA	144
CG7429_R	AAGCTGGCGGTGAATCCCTT	
CG9300_F	GCAAGCCCAACAGTCCACGCA	125
CG9300_R	GTACTTCCGCCGGTCAGCCG	
Cht4_F	CCACATGGCTCTGGATGGCT	96
Cht4_R	TCAGCGGGAGCACCATGAGA	
Dif_F	GGAGCCGACAAGCAATATAATCC	127
Dif_R	GTAGTTGCACACTTCGATGGT	
drs_iQ_F	AAGTACTTGTTGCCCCTCTTCGCT	Rynes et al. 2012
drs_iQ_R	ATCCTTCGCACCAGCACTTCAGA	
e(y)2b_F	AACGAATTTTACCGACTCGGGCTT	234
e(y)2b_R	ACGCGTGTGCAGCAGTTCCT	
EF1gamma_F	TTGGCCGACATTGTGGTCTTC	108
EF1gamma_R	GATGGTGACGAACCAGCGG	
elav_F	CAGTGCAAGGGCTACGGCTT	112
elav_R	TGAAGCTGACCTGCAGCACC	
Fbp2_F	CGCTGGACATCTGGAATCGGT	118
Fbp2_R	TTGGGGTTGACGTGGCTTGG	
IM10_F	TCTTCGGTTCACCGGGCAAT	183
IM10_R	TGATGGTGTGGGAAGCTGGT	

IM2_F	CTCTGGCCAACGCTGTTCC	139
IM2_R	GCCCATCTTCGAAATTAAGGCGA	
IM23_F	GGCTACGGCCGGAAATGTGA	173
IM23_R	GCCGAATGGGACCATCGAGA	
Insc_F	GCAGTGCGAACCCTGCTCGT	329
Insc_R	TGGGTGCAGCATTCCGGCTG	
insc_F2	CAGCCGGAATGCTGCACCCA	331
insc_R2	TCTTGATGGTGCGGCGCTGG	
Lcp3_F	TGTTCAAGATCCTGCTTGTCTGTT	255
Lcp3_R	CACTCTGGGGCTGGTATCCG	
Lip1_F	TCGCCAAGTACGGCTACGAG	179
Lip1_R	CAGCAAGTAGGCCAGGCTCA	
linc367_F2	CCCAGCTCCTAGCTCCTAGAGCC	113
linc367_R2	GCTGCGAGGCTCGTGCAATTC	
Inc458_F	GCTGAAAAGCAGAAGCAACC	91
Inc458_R2	TTAATAAACTGCGCGACGTG	
Inc66_F	TGTCGGATCGGGGCCACAAA	178
Inc66_R	CACTCCGCTACGCTCCTCACCT	
Inc626_F	TCAAAACTGTACCAGCTGCCTGGT	80
Inc626_R	TGGTCGCTTGTGCTCGGATCG	
Inc703_F	GTTTTGTGACGGCTTCCAAC	377
Inc703_R	CAGCGTTTCAACTATTGCA	
Inc921_F	GGATTGCTGCGAATCTTGTT	299
Inc921_R	AGTGCGTTGGTTTTAATGGC	
linc927_F2	TTCTGCGTAGGTGGAAGAATTTTCA	93
linc927_R3	TTTACAAAAATGCTCCGCACATCC	
LysBCDE_iQ_F	ATGAAGGCTTTCATCGTTCTGGT	Rynes et al. 2012
LysBCDE_iQ_R	ACTCCGGTGCGGTAGGAGGA	
mCD8_F3	CTACCAAGCCAGTGCTGCGA	117
mCD8_R3	CGAAGTCCAATCCGGTCCCC	
Mlp84_F	GACCGATGGCTTGACCGAGG	118
Mlp84_R	GAATACGGCTCCTCCGCACC	
moody_F	GAGACAACCTATTAGCCTGGAGGA	141
moody_R	GCCCACGATCATTATGAGAAAGG	
Mthl8_F	CAGCCTTCCCTCCCATAGCC	177
Mthl8_R	TGGCTGATGGTTCCGTTCTCA	
PGRP-SB1_F	TTGCTGCTTAGCTCTATCCGCC	162
PGRP-SB1_R	TTGCACTGCTCGGATGTGGA	
pot_F	ACTGGATGGCAAGGTTACCAA	125
pot_R	GGAACGGATCGTTGAACTTGAG	

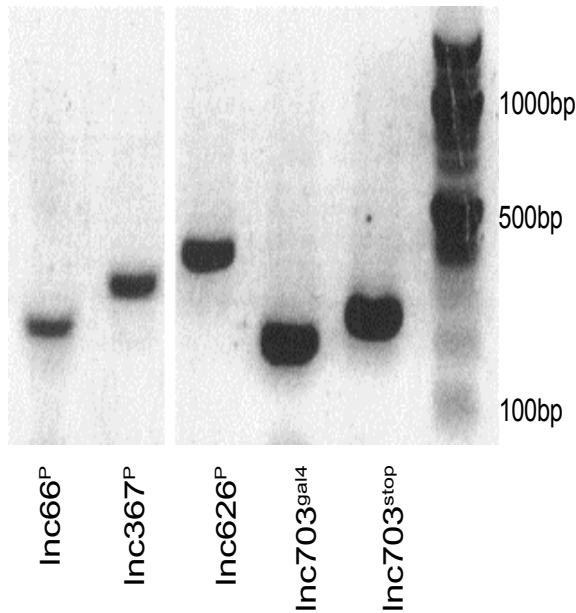
RabX2	ATGAGCCACTTCGACTATCTGT	118
RabX2	TCGTTCTCAGGTATTTCTCGGT	
rel_iQ_F	AGCTACAGGAACTGCATCAGGA	Rynes et al. 2012
rel_iQ_R	GTTTGTGCCGACTTGCGGTTAT	
repo_F	CGCCAAGTAGTTCCTCCACGG	216
repo_R	GGTAGGCAGTAAAGGTGGTTCTCG	
Rp49F	TACAGGCCCAAGATCGTGAA	LaLonde et al. 2006
Rp49R	TCTCCTTGCGCTTCTTGGA	
RpL28_F	GCAACCTCTTCGCACCTCAAT	165
RpL28_R	CACGACTCCCAGGGTCTTCT	
RpL30_F2	ATGGTGGCCGTTAAGAAACAA	84
RpL30_R2	GCAGTATTTGCCGGACTTCAT	
RpS6_F	TGAAGCAGGGTGTCTTGACC	107
RpS6_R	ACGCACAGACTTGCGCTTA	
skittles_F	GCCACCCAAACGCCGGCCTCG	133
skittles_R	CACCGCCCTCCGCCACTCGC	
sta_F	CTGGGTTCCGAGAACGTCAAC	154
sta_R	CGAAGATGTCCGAGGGGTTATC	
TotA_F	ACTGCTCTTATGTGCTTTGCACTG	128
TotA_R	TTGGAGTCATCGTCCTGGGC	
toyF2	GGAGCAGCAAGCTCAGCAACAAAA	112
toyR2	CAAATGTGCCGTCGTTGTATTGCT	
vvIF2	TCGTTAAACTGAGCCCCAACTTGG	131
vvIR2	TCCGAGGGCCATGTCCAGAT	
wg_F2	CCAAGTCGAGGGCAAACAGAA	117
wg_R2	TGGATCGCTGGGTCCATGTA	
wnt6_F2	CGAAATCTGTGCGGCACGATTC	169
wnt6_R2	GAAACCCGTCTCTCGGGAA	

Appendix C - Images of gels showing PCR products, confirming presence of transposon in IncRNA loci after isogenisation.

Backcross 1

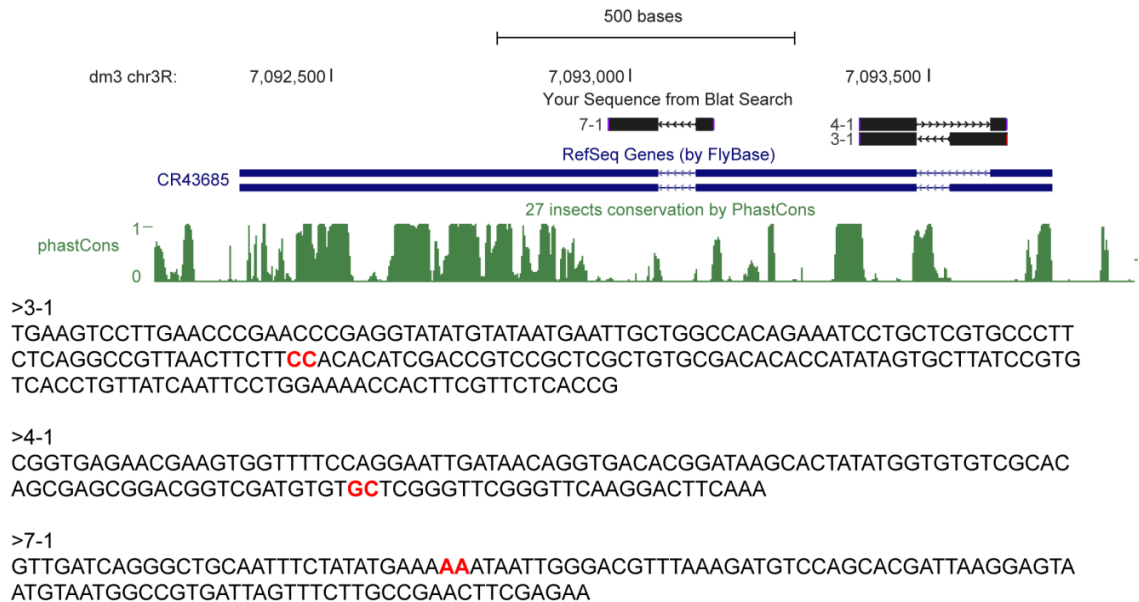


Backcross 2

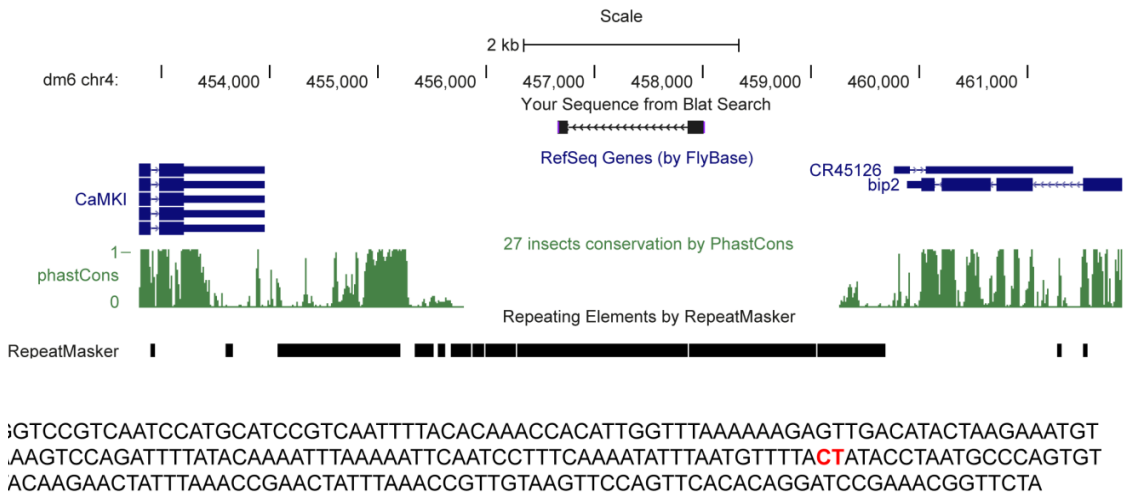


Appendix D - Sequencing results confirm introns in *Inc703* and *Inc921* aligned to BDGP6. Exon/exon boundaries are highlighted in red.

Sequences and BLAT results confirming *Inc703* introns aligned to chr3R of BDGP Release 6



Sequences and BLAT results confirming *Inc921* putative intron aligned to chr4 of BDGP Release 6



Appendix E – Genes changed in the same direction one day AFP in *Inc703*^{gal4} and *Inc703*^{stop} strains

id	Gene name	Change
FBgn0033830	<i>CG10814</i>	+
FBgn0035022	<i>CG11413</i>	+
FBgn0037503	<i>CG14598</i>	+
FBgn0054057	<i>CG34057</i>	+
FBgn0040984	<i>CG4440</i>	+
FBgn0040723	<i>CG5011</i>	+
FBgn0053319	<i>CR33319</i>	+
FBgn0058469	<i>CR40469</i>	+
FBgn0033292	<i>Cyp4ad1</i>	+
FBgn0010039	<i>GstD3</i>	+
FBgn0002534	<i>Lcp3</i>	+
FBgn0016675	<i>Lectin-galC1</i>	+
FBgn0004425	<i>LysB</i>	+
FBgn0004426	<i>LysC</i>	+
FBgn0004427	<i>LysD</i>	+
FBgn0004428	<i>LysE</i>	+
FBgn0013323	<i>Ptth</i>	+
FBgn0040069	<i>vanin-like</i>	+
FBgn0011559	<i>Acp36DE</i>	-
FBgn0035234	<i>CG12003</i>	-
FBgn0036102	<i>CG12523</i>	-
FBgn0033957	<i>CG12865</i>	-
FBgn0263376/FBgn0032211	<i>CG13138</i>	-
FBgn0029518	<i>CG13376</i>	-
FBgn0034966	<i>CG13563</i>	-
FBgn0035845	<i>CG13675</i>	-
FBgn0036716	<i>CG13728</i>	-
FBgn0039028	<i>CG13840</i>	-
FBgn0032824	<i>CG13962</i>	-
FBgn0036858	<i>CG14088</i>	-
FBgn0029573	<i>CG14770</i>	-
FBgn0031088	<i>CG15322</i>	-
FBgn0039806	<i>CG15545</i>	-
FBgn0038131	<i>CG15888</i>	-
FBgn0034624	<i>CG17974</i>	-
FBgn0031929	<i>CG18585</i>	-
FBgn0029990	<i>CG2233</i>	-

FBgn0050076	<i>CG30076</i>	-
FBgn0052318	<i>CG32318</i>	-
FBgn0053128	<i>CG33128</i>	-
FBgn0053296	<i>CG33296</i>	-
FBgn0053758	<i>CG33758</i>	-
FBgn0085250	<i>CG34221</i>	-
FBgn0085489	<i>CG34460</i>	-
FBgn0250833	<i>CG34461</i>	-
FBgn0259233	<i>CG42331</i>	-
FBgn0263040	<i>CG43335</i>	-
FBgn0032585	<i>CG4580</i>	-
FBgn0027586	<i>CG5867</i>	-
FBgn0033866	<i>CG6280</i>	-
FBgn0039027	<i>CG7031</i>	-
FBgn0031940	<i>CG7214</i>	-
FBgn0034623	<i>CG9822</i>	-
FBgn0033867	<i>Cpr50Ca</i>	-
FBgn0035875	<i>Cpr66Cb</i>	-
FBgn0036619	<i>Cpr72Ec</i>	-
FBgn0262616	<i>CR43142</i>	-
FBgn0262996	<i>CR43304</i>	-
FBgn0005664	<i>Cry</i>	-
FBgn0262111	<i>F</i>	-
FBgn0005633	<i>Fln</i>	-
FBgn0063498	<i>GstE2</i>	-
FBgn0064237	<i>Idgf5</i>	-
	<i>lincRNA.118</i>	-
	<i>lincRNA.205</i>	-
	<i>lincRNA.703</i>	-
	<i>lincRNA.930</i>	-
FBgn0032891	<i>Oseg5</i>	-
FBgn0037406	<i>Osi1</i>	-
FBgn0037418	<i>Osi11</i>	-
FBgn0037419	<i>Osi12</i>	-
FBgn0032359	<i>Osi21</i>	-
FBgn0038133	<i>Osi22</i>	-
FBgn0037411	<i>Osi3</i>	-
FBgn0037412	<i>Osi4</i>	-
FBgn0037415	<i>Osi8</i>	-
FBgn0043578	<i>PGRP-SB1</i>	-
FBgn0259210	<i>Prom</i>	-

FBgn0034157	<i>resilin</i>	-
FBgn0020908	<i>Scp1</i>	-
FBgn0035638	<i>Tektin-C</i>	-
FBgn0028396	<i>TotA</i>	-
FBgn0013348	<i>TpnC41C</i>	-
FBgn0037227	<i>TwdIV</i>	-
FBgn0061188	<i>Yu</i>	-

Appendix F – Genes changed in the same direction in larval fat bodies of isogenised *Inc703^{gal4}* and *Inc703^{stop}* strains

id	Gene name	Change	GO term
FBgn0015570	<i>alpha-Est2</i>	+	
FBgn0261393	<i>alpha-Est5</i>	+	
FBgn0034440	<i>CG10073</i>	+	
FBgn0035517	<i>CG1265</i>	+	
FBgn0032810	<i>CG13077</i>	+	
FBgn0032896	<i>CG14400</i>	+	
FBgn0037819	<i>CG14688</i>	+	
FBgn0028526	<i>CG15293</i>	+	
FBgn0040993	<i>CG17325</i>	+	
FBgn0051769	<i>CG31769</i>	+	
FBgn0052280	<i>CG32280</i>	+	
FBgn0259201	<i>CG42305</i>	+	
FBgn0262607	<i>CG43133</i>	+	
FBgn0037548	<i>CG7900</i>	+	
FBgn0037684	<i>CG8129</i>	+	
FBgn0038214	<i>CG9616</i>	+	
FBgn0042701	<i>CR12628</i>	+	
FBgn0037560	<i>CR18228</i>	+	
FBgn0028940	<i>Cyp28a5</i>	+	
FBgn0033065	<i>Cyp6w1</i>	+	
FBgn0086687	<i>Desat1</i>	+	
FBgn0027610	<i>Dic1</i>	+	
FBgn0028737	<i>Ef1beta</i>	+	Translation [GO:0006412]
FBgn0029176	<i>Ef1gamma</i>	+	Translation [GO:0006412]
FBgn0027571	<i>FASN1</i>	+	
FBgn0038074	<i>Gnmt</i>	+	
FBgn0031117	<i>GstT3</i>	+	
FBgn0031305	<i>Iris</i>	+	
FBgn0034540	<i>Lrt</i>	+	
FBgn0025814	<i>Mgstl</i>	+	
FBgn0025631	<i>moody</i>	+	
FBgn0051052	<i>Nha2</i>	+	
FBgn0013307	<i>Odc1</i>	+	
FBgn0030200	<i>RabX2</i>	+	
FBgn0020618	<i>Rack1</i>	+	
FBgn0032987	<i>RpL21</i>	+	Translation [GO:0006412]

FBgn0015288	<i>RpL22</i>	+	Translation [GO:0006412]
FBgn0026372	<i>RpL23A</i>	+	Translation [GO:0006412]
FBgn0036825	<i>RpL26</i>	+	Translation [GO:0006412]
FBgn0020910	<i>RpL3</i>	+	Translation [GO:0006412]
FBgn0037686	<i>RpL34b</i>	+	Translation [GO:0006412]
FBgn0066084	<i>RpL41</i>	+	Translation [GO:0006412]
FBgn0064225	<i>RpL5</i>	+	Translation [GO:0006412]
FBgn0261592	<i>RpS6</i>	+	Translation [GO:0006412]
FBgn0038257	<i>smp-30</i>	+	
FBgn0003517	<i>sta</i>	+	Translation [GO:0006412]
FBgn0026315	<i>Ugt35a</i>	+	
FBgn0003961	<i>Uro</i>	+	
FBgn0003996	<i>w</i>	+	
FBgn0261113	<i>Xrp1</i>	+	
XLOC_025617		+	
FBgn0041707	<i>7B2</i>	-	
FBgn0005592	<i>btl</i>	-	
FBgn0032797	<i>CG10186</i>	-	
FBgn0040364	<i>CG11378</i>	-	
FBgn0036364	<i>CG14109</i>	-	
FBgn0029898	<i>CG14439</i>	-	
FBgn0032449	<i>CG17036</i>	-	
FBgn0032124	<i>CG17855</i>	-	
FBgn0051522	<i>CG31522</i>	-	
FBgn0052228	<i>CG32228</i>	-	
FBgn0040384	<i>CG32795</i>	-	
FBgn0038878	<i>CG3301</i>	-	
FBgn0038912	<i>CG6656</i>	-	
FBgn0036416	<i>CG7924</i>	-	
FBgn0038466	<i>CG8907</i>	-	
FBgn0030594	<i>CG9509</i>	-	
FBgn0030158	<i>CG9686</i>	-	
FBgn0023023	<i>CRMP</i>	-	
FBgn0038194	<i>Cyp6d5</i>	-	
FBgn0004910	<i>Eip63F-1</i>	-	
FBgn0011554	<i>etaTry</i>	-	
FBgn0001149	<i>GstD1</i>	-	
FBgn0037950	<i>HisCl1</i>	-	
FBgn0031261	<i>nAChRbeta3</i>	-	
FBgn0015774	<i>NetB</i>	-	

FBgn0043578	<i>PGRP-SB1</i>	-	
FBgn0042086	<i>Tsp42Eb</i>	-	
XLOC_010527		-	