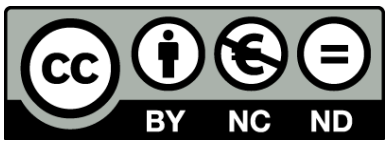
UNIVERSITAT DE BARCELONA

Development and regeneration of Drosophila wing imaginal discs: the role of lncRNAs and the stress sensor D-GADD45

Carlos Camilleri Robles



UNIVERSITAT DE BARCELONA

Doctoral thesis

Development and regeneration of *Drosophila* wing imaginal discs: the role of lncRNAs and the stress sensor D-GADD45

CARLOS CAMILLERI ROBLES

Doctoral thesis

May 2022

Doctoral thesis

CARLOS CAMILLERI ROBLES

Departament de Genètica, Microbiologia i Estadística

Programa de Doctorat de Genètica

Facultat de Biologia

Universitat de Barcelona

Development and regeneration of *Drosophila* wing imaginal discs: the role of lncRNAs and the stress sensor D-GADD45

Memòria presentada per en

Carlos Camilleri Robles

Per optar al grau de

Doctor

per la Universitat de Barcelona

UNIVERSITAT DE BARCELONA

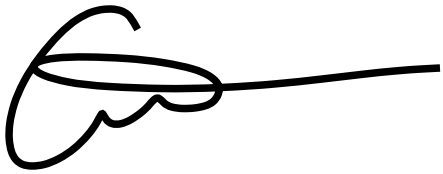
Tesi doctoral realitzada sota la direcció de la Dra. Montserrat Corominas Guiu

Realitzada al Departament de Genètica, Microbiologia i Estadística
de la Facultat de Biologia de la Universitat de Barcelona

La directora i tutora,

Dra. Montserrat Corominas Guiu

L'autor,



Carlos Camilleri Robles

Barcelona, Abril 2022

The wing images shown in the front cover represent the most common phenotypes observed upon the induction of cell death in the larval stages of wild type animals (top image) and lncRNA*CR40469* knock-out mutants (bottom image).

ABSTRACT

Long non-coding RNAs (lncRNAs) are defined as transcripts longer than 200 nucleotides that lack protein-coding potential. Although multiple examples of functional lncRNAs have been described, particularly regulating gene expression at different levels, the function of the vast majority of them remains to be elucidated. Here, we use the *Drosophila* wing imaginal discs as a model system to study the involvement of lncRNAs in development and regeneration. Additionally, we also studied the role of the stress sensor protein *Drosophila Growth Arrest and DNA Damage 45* (*D-GADD45*) in the wing disc.

For the study of lncRNAs, we used transcriptomic data from developing and regenerating wing discs. We identified a set of ~200 lncRNAs expressed in development, as well as 131 differentially-expressed (DE) lncRNAs in regeneration. Among them, we focused on the study of lncRNAs *CR40469* and *CR34335*, which share 99.1% sequence identity, however, their expression pattern is far from similar: while *CR40469* is upregulated in regeneration, the expression of *CR34335* is inhibited upon damage.

We generated a *CR40469* knock-out mutant, for which no phenotypes were detected in normal conditions. Nevertheless, upon the induction of cell death, these mutants lost their regeneration capacity, suggesting a putative function of *CR40469* in regeneration. Moreover, we characterised the molecular changes occurring in the mutant, revealing a set of 95 DE genes compared to controls. On the other hand, no phenotypes were detected for *CR34335* mutants, neither in development nor in regeneration, suggesting that this lncRNA is dispensable for both processes.

Regarding the study of *D-GADD45*, we activated its expression ectopically in the wing discs, resulting in increased apoptosis. Through genetic interaction experiments, we described the *D-GADD45*-induced cell death as dependent on the activation of the JNK signalling pathway. Additionally, we described a JNK-independent decrease in cell proliferation upon sustained activation of *D-GADD45*. Finally, we identified *D-GADD45* as an essential gene for the regeneration of wing discs, as the use of RNAi constructs against *D-GADD45* severely impairs the recovery process after the induction of cell death.

TABLE OF CONTENTS

FIGURE INDEX

INTRODUCTION

MATERIALS AND METHODS

RESULTS

Chapter 1: The role of lncRNAs during wing disc development and regeneration

DISCUSSION

TABLE INDEX

INTRODUCTION

MATERIALS AND METHODS

RESULTS

ANNEX I

ABREVIATIONS AND ACRONYMS

3C	Chromosome conformation capture
aa	aminoacids
AEL	After egg laying
Ask1	Apoptosis signal regulating kinase 1
ATAC-seq	Assay for Transposase-Accessible Chromatin followed by sequencing
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BSA	Bovine Serum Albumin
bsk	*basket*
cDNA	Complementary DNA
CE	Columnar epithelium
ci	*cubitus interruptus*
CP	Core Promoter
Ctrl	Control
DE	Differentially expressed
DNA	Deoxyribonucleic Acid
Down	Downregulated
DP	Distal Promoter
DRRE	Damage Responsive Regulatory Element
EDTA	Ethylene diaminetetraacetic acid
en	*engrailed*
EP	Early pupae
Fc	Fold change
FISH	Fluorescent *In Situ* Hybridization
GADD45	Growth Arrest and DNA Damage-inducible 45
Gal4	Galactosidase-induced gene 4
Gal80TS	Gal80 temperature sensitive
gDNA	Genomic DNA
GFP	Green Fluorescent Protein
GO	Gene Ontology
h	hours
Hi-C	Chromosome conformation capture followed by high throughput sequencing
JNK	c-Jun-NH2-terminal kinase
Kb	Kilobase
L1, L2, L3	First, second and third instar larvae
LexO	LexA Operator
LHG	LexA-hinge-Gal4 activation domain
lncRNA	Long non-coding RNA
LP	Late pupae
MAPK	Mitogen Activated Protein Kinase
Mb	Megabase

MEF	Minimum Free Energy
Mekk1	Mitogen-activated kinase kinase kinase 1
min	minutes
miRNA	micro RNA
Mmp1	Matrix metalloprotease 1
mRNA	messenger RNA
N	number of samples
NDE	Not Differentially Expressed
nt	Nucleotides
ORF	Open Reading Frame
PCG	Protein-coding gene
PCR	Polymerase Chain Reaction
PM	Peripodial membrane
PP	Proximal Promoter
qPCR	Quantitative Polymerase Chain Reaction
Regen	Regeneration
RNA	Ribonucleic Acid
RNAi	RNA of interference
ROS	Reactive Oxygen Species
rpr	*reaper*
rRNA	ribosomal RNA
sal/salm	*spalt major*
sal$^{E/Pv}$	*spalt E/Pv*
SD	Standard deviation
smORF	Small Open Reading Frame
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
sply	*sphingosine-1-phosphate lyase*
TF	Transcription Factor
TPM	Transcripts Per kilobase Million
tRNA	transfer RNA
TSS	Transcription Start Site
tub	*tubulin*
TUNEL	Terminal deoxynucleotidyl transferase dUTP nick end labelling
UAS	Upstream Activating Sequence
UCSC	University of California Santa Cruz
Up	Upregulated
w	*white*
WT	Wild type

INTRODUCTION

LONG NON-CODING RNAs

It is currently estimated that protein-coding genes only account for a small fraction of genomes, which is 1-2% in the case of the human genome. However, around 75% of the human genome is actively transcribed, manifesting the importance of the non-coding transcriptome (Djebali et al. 2012). Initially, non-coding genes were usually dubbed as 'junk' DNA, as they were thought to be non-functional byproducts of transcription or transcriptional noise. Growing evidence on the functionalities of these non-coding transcripts changed this initial perception, increasing the interest in their study.

The non-coding RNAs (ncRNAs), encoded by non-coding genes, form a very heterogeneous group of transcripts that lack protein-coding potential. They are classified depending on their size and function. Short ncRNAs, spanning less than 200 nucleotides (nt) in length, are composed by ribosomal RNA (rRNA), transfer RNA (tRNA), micro RNA (miRNA), piwi-interacting RNA (piRNA), small nuclear RNA (snRNA) or small nucleolar RNA (snoRNA). They participate in a variety of biological processes, including the translation of mRNAs (rRNAs and tRNAs), the targeting and inhibition of mRNAs and transposable elements (miRNAs and piRNAs), the splicing of pre-mRNAs (snRNAs), or the addition of RNA modifications (snoRNAs) (reviewed in Watson et al. 2019). The main features of the short ncRNA class are summarised in **Table 1**.

ncRNA class	Size (nt)	Annot. human genes	Localization	Function
rRNA	120 - 4.500	47	Ribosomes	Translation: binding to mRNA and tRNA
tRNA	76 - 90	610	Cytoplasm / Ribosomes	Translation: link between mRNA and aa
miRNA	~ 22	1.879	miRISC complex	Post-transcriptional silencing
piRNA	24 - 32	~ 6.000	RISC complex	Post-transcriptional silencing, mainly TEs
snRNA	100 - 300	1.901	Spliceosome	Splicing / Histone pre-mRNA processing
snoRNA	60 - 300	943	Nucleolus	RNA modifications

Table 1. Principal features of short ncRNAs. Number of annotated human genes as of GENCODE v39, with the exception of piRNA genes, for which no gene biotype is described in GENCODE. Instead, estimated data from Ha et al. 2014 was considered.

The non-coding transcripts longer than 200 nt and that do not belong to any of the previously mentioned ncRNA classes, are considered as long non-coding RNAs (lncRNAs). Although arbitrary, the 200 nt cut-off was mainly set to exclude most canonical ncRNAs, including tRNAs, miRNAs, snRNAs or snoRNAs. LncRNAs are the most abundant class of ncRNAs in the genome, although the putative function of the large majority of annotated lncRNAs remains to be elucidated. In humans, there are currently 18,811 annotated lncRNA genes, giving rise to 53,009 transcripts (GENCODE v39).

The transcription and initial processing of lncRNAs is similar to that of mRNAs, as they are usually transcribed by the RNA polymerase II (Pol II), their 5' end is capped by the addition of a methyl-7-guanosine, their 3' end is polyadenylated and, in the case of multi-exonic genes, their introns are spliced (reviewed in Quinn and Chang, 2016). Compared to protein-coding genes, the expression of lncRNAs is less abundant and more restricted to specific tissues and developmental stages (Derrien et al. 2012; Djebali et al. 2012;

Brown et al. 2014). Many lncRNAs seem to be expressed particularly in the testes. In humans, an unusually high expression was observed for lncRNAs in testes compared to any other tissue (Derrien et al. 2012). Similar results were also observed in flies, as it is estimated that around 30% of the annotated lncRNAs have peak expression in testes, some of which are only detected there (Brown et al. 2014).

Genomic context of lncRNAs

According to NONCODE, an integrated database of lncRNAs across 16 different species (Zhao et al. 2016), there are currently almost 550,000 annotated transcripts from 355,000 different lncRNA genes. However, despite the growing number of annotated lncRNAs, the vast majority remain uncharacterized. For this reason, to better understand their putative functions, multiple lncRNA classifications have been used according to different parameters.

One of the more basic, yet useful ways to classify lncRNAs is based on their genomic context. In this way, lncRNAs can be classified into genic or intergenic lncRNAs. **Genic** lncRNAs are located overlapping the sequence of other protein-coding or non-coding genes. When lncRNAs are located within an intron of their overlapping gene, they are considered as **intronic** (**Figure 1A**). On the other hand, lncRNAs overlapping entirely or partially the exonic sequence of other genes are considered as **exonic** (**Figure 1B**). Both intronic and exonic lncRNAs can be further classified depending on the strand from which they are transcribed in regard to their overlapping gene. In this way, **sense** lncRNAs are transcribed from the same strand as their overlapping protein-coding gene, while **antisense** lncRNAs are transcribed from the opposite strand. Antisense lncRNAs are often referred to as natural antisense transcripts (NATs), and have been described in most eukaryotic genomes, including humans, mice and flies (Pelechano and Steinmetz, 2013).

Alternatively, lncRNAs positioned in intergenic regions are referred to as **intergenic** lncRNAs or lincRNAs (**Figure 1C**). Depending on their orientation and configuration relative to their closest gene, lincRNAs are classified as convergent (when their orientation is opposed and are configured head to head), divergent (when their orientation is opposed and are configured tail to tail), or same strand (when both genes are transcribed in the same orientation) (Derrien et al. 2012).

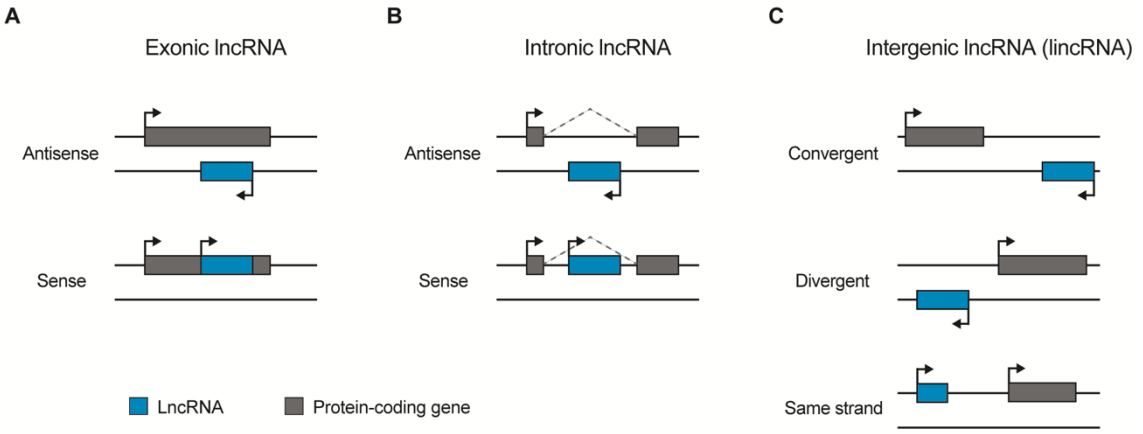


Figure 1. Classification of lncRNAs depending on their genomic context. (**A**) Representation of an antisense (top) and sense (bottom) exonic lncRNA. (**B**) Representation of an antisense (top) and sense (bottom) intronic lncRNA. (**C**) Representtion of the different genomic conformations of intergenic lncRNAs (lincRNAs).

Features of lncRNAs across different species

The number of lncRNAs varies greatly from one species to another. While in humans and mice the number of annotated protein-coding and long non-coding genes is quite similar (19,982 coding and 18,811 long non-coding genes in humans, GENCODE v39; 21,833 coding and 13,186 long non-coding genes in mice, GENCODE M28), in *Drosophila melanogaster* or *Caenorhabditis elegans* the number of long non-coding genes is considerably lower compared to the number of annotated protein-coding genes (13,969 coding and 2,545 long non-coding genes in flies, FlyBase r6.39; 19,987 coding and 1,698 long non-coding genes in nematode, WS281) (**Figure 2A**).

Regarding their number of exons, lncRNAs tend to contain less exons compared to protein-coding genes (Derrien et al. 2012). However, the number of exons of lncRNAs is different between species. For instance, the majority of lncRNAs annotated in *Drosophila* or *C. elegans* are either mono-exonic or composed of 2 exons (> 90% in both species), with only a few exceptions containing 3 or more exons. On the contrary, around 60% of lncRNAs in humans and mice contain 1 or 2 exons, and up to ~20% contain 4 or more exons (**Figure 2B**) (Camilleri-Robles et al. 2022).

As with protein-coding genes, the mean transcript size of lncRNAs also varies across different species. For instance, *Drosophila* lncRNAs are shorter than human and mouse lncRNAs (average length of 962 nt in flies compared to 1,230 and 1,456 nt in humans and mice, respectively). Moreover, less than 4% of *Drosophila* lncRNAs span more than 3 kb compared to 7.78% in humans and 10.35% in mice (**Figure 2C**) (Camilleri-Robles et al. 2021). On the other hand, more than 90% of annotated lncRNAs in *C. elegans* span less than 500 nt, exposing how variable the size of lncRNAs is (**Figure 2C**) (Camilleri-Robles et al. 2021).

Despite the differences in genome compactness between different species, ranging from less than 6-8 annotated protein-coding genes per Mb in humans and mice to 100-200 genes per Mb in flies and nematodes, the proportion of intronic, exonic and intergenic

lncRNAs is strikingly similar. Regardless of genome compactness, the lincRNAs are the most abundant class of lncRNAs, representing 50-55% of all annotated lncRNAs. On the other hand, intronic and exonic lncRNAs made up for ~20% and ~25% of the annotated lncRNAs, respectively (**Figure 2D**) (Camilleri-Robles et al. 2021).

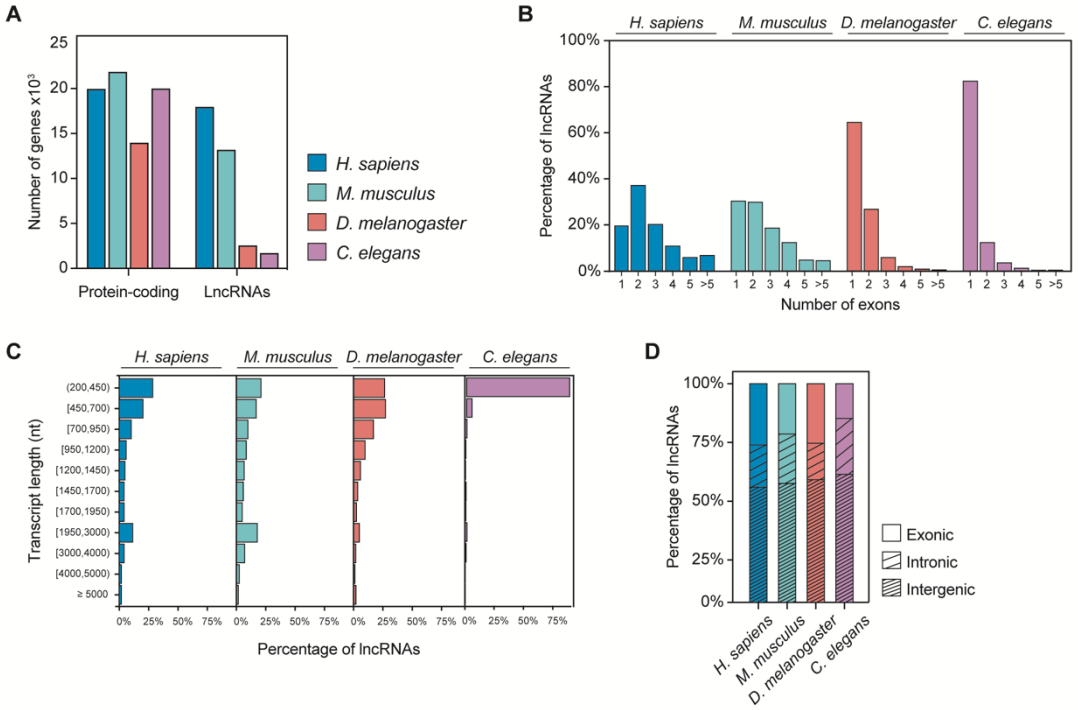


Figure 2. Features of lncRNAs across different species. (**A**) Number of annotated protein-coding and long non-coding genes. (**B**) Distribution of lncRNAs depending on their number of exons. (**C**) Distribution of lncRNAs depending on the size (nucleotides, nt) of their longest transcript. (**D**) Classification of annotated lncRNAs into exonic, intronic and intergenic. Data from GENCODE v39 (humans), GENCODE M28 (mice), FlyBase r6.39 (flies), WS281 (nematode). Adapted from Camilleri-Robles et al. 2022.

Subcellular localization of lncRNAs

As the ultimate product of the long non-coding genes is the RNA itself, their localization determines their putative binding partners and target sites. Thus, the study of their localization within the cell is critical to understand their putative functions. LncRNAs have been found in the nucleus, the cytoplasm, and virtually in any organelle and membraneless cellular structure (reviewed in Carlevaro-Fita and Johnson, 2019). In the nucleus, they can be found in the nucleoli or forming ribonucleoprotein aggregates in close proximity with chromatin regions, in occasions serving as reservoirs of pre-mRNA processing factors. On the other hand, cytoplasmic lncRNAs can be found in the ribosomes, mitochondria, exosomes, or even in the extracellular membrane.

Initially, it was suggested that lncRNAs were located preferentially in the nucleus (Derrien et al. 2012). However, growing evidence suggests that a large fraction of the expressed lncRNAs is cytoplasmic (Ingolia et al. 2011; van Heesch et al. 2014; Carlevaro-Fita et al. 2016). It is estimated that around 75% of lncRNAs in human and fly cells are detected in the cytoplasm (Benoit-Bouvrette et al. 2018), while fluorescent *in situ* hybridization (FISH) experiments in *Drosophila* tissues estimate that around 40% of

expressed lncRNAs are restricted to the cytoplasm, as opposed to only 4% only found in the nucleus (Wilk et al. 2016).

Nuclear lncRNAs

Similar to the nuclear localization signal (NLS) present in the amino acid sequence of proteins that must be imported to the nucleus, the presence of some sort of nuclear retention signal is believed to prevent the exportation of nuclear lncRNAs to the cytoplasm. Linking with this, multiple studies on individual lncRNA genes revealed the presence of specific sequences driving their nuclear localization. For instance, the presence of a conserved repeating sequence spanning 156 bp within the exonic sequence of lncRNA *FIRRE* is essential for their nuclear localization through their binding to the nuclear ribonucleoprotein hnRNPU (Hacisuleyman et al. 2014). Similarly, two large independent regions within the sequence of *MALAT1* were reported to promote their nuclear enrichment, as their deletion resulted in *MALAT1* exportation to the cytoplasm (Miyagawa et al. 2012). A more general approach on multiple lncRNAs identified the presence of 42-nt motifs antisense to Alu sequences that were able to drive the accumulation of transcripts in the nucleoplasm (Lubelsky and Ulitsky, 2018). These motifs were called SIRLOIN (SINE-derived nuclear RNA localization), and were able to recruit and bind to the heterogeneous nuclear ribonucleoprotein K (hnRNPK), driving the nuclear enrichment of their associated transcripts. Indeed, the presence of SIRLOIN-like elements in mouse RNAs also results in their nuclear retention, suggesting that this mechanism of RNA nuclear enrichment is conserved (Lubelsky and Ulitsky, 2018). A similar study identified longer sequences (> 300 nt) that were highly conserved and over-represented in nuclear transcripts, and whose presence was sufficient to partially redirect RNA subcellular localization to the nucleus (Shukla et al. 2018).

Despite the growing evidence pointing towards the presence of nuclear retention signals in the sequence of lncRNAs, alternative methods favouring the presence of lncRNAs in the nucleus have also been validated. For instance, it has been reported that mRNAs lacking poly(A)-tails are not exported to the cytoplasm (Dias et al. 2010). Similarly, lncRNAs produced by alternative processing pathways that omit the 3' polyadenylation step, may be restricted to the nucleus. Supporting this hypothesis, it was shown that RNAs lacking a poly(A)-tail are retained in the nucleus (Akef et al. 2013). Nonetheless, other experiments suggest that 3' polyadenylation is not required for nuclear exportation, as circular RNAs or histone mRNAs, both lacking poly(A)-tails, are efficiently exported to the cytoplasm (Huang et al. 2018; Erkmann et al. 2005).

Another possible mechanism driving lncRNA nuclear retention is the exaptation of transposable elements (TEs) to inhibit their nuclear export. Actually, approximately 83% of lincRNAs carry at least one TE in their exonic sequence, as opposed to ~40% of protein-coding genes (Kelley and Rinn, 2012). Genome-wide studies revealed the presence of specific TEs such as L2b, MIRb or MIRc within the exons of lncRNAs, which promote their nuclear enrichment (Carlevaro-Fita et al. 2018). It has been proposed that these TEs are recognized by other factors that prevent their recognition by the nuclear export machinery, however, the exact mechanism by which these TEs promote nuclear retention remains to be elucidated.

Cytoplasmic lncRNAs

Although lncRNAs are enriched in the nucleus compared to mRNAs, a large fraction of lncRNAs is actually exported to the cytoplasm, presumably sharing the same export pathways with mRNAs. In the cytoplasm, lncRNAs have been found virtually anywhere although they seem particularly enriched at the ribosomes (van Heesch et al. 2014; Carlevaro-Fita et al. 2016; Zeng et al. 2018). The fact that some non-coding genes produce functional peptides (Galindo et al. 2007; Magny et al. 2013; Anderson et al. 2015; Nelson et al. 2016), tempts to speculate that the interaction of lncRNAs with the ribosomes implies its translation. However, it is probable that most of these interactions reflect other processes.

For instance, *lincRNA-p21*, which has been found in polysomal fractions, is described to inhibit the translation of target mRNAs through their interaction with the translational repressor Rck (Yoon et al. 2012). It has also been suggested that ribosome-bound lncRNAs are subject to degradation, as part of a regulatory mechanism to control their transcript levels (Carlevaro-Fita et al. 2016; Zeng et al. 2018). Indeed, lncRNAs interacting with ribosomes are more sensitive to degradation by the nonsense-mediated decay (NMD) pathway than ribosome-free lncRNAs (Zeng et al. 2018). Also, mutations in the NMD pathway particularly promote the upregulation of many lncRNAs in plants, further suggesting that it could be a conserved mechanism of lncRNA surveillance (Kurihara et al. 2009).

LncRNAs encoded in the nucleus have also been found in the mitochondria, evidencing the trafficking of transcripts from the nucleus to the mitochondria (reviewed in Dong et al. 2017). A clear example is the lncRNA *RMRP*, which is transported first from the nucleus to the cytoplasm through the action of HuR, and finally to the mitochondria by binding to PNPASE and GRSF1 (Wang et al. 2012; Noh et al. 2016). There, *RMRP* is implicated in the replication of mitochondrial DNA (mtDNA) and the processing of RNAs (Dong et al. 2017).

Other lncRNA cytoplasmic locations include the exosomes, from which it is speculated that lncRNAs might be exported from one cell to another (Gezer et al. 2014), or cellular membranes, through their direct binding with membrane phospholipids (Lin et al. 2017).

Cis-acting and trans-acting lncRNAs

Similar to other regulatory elements, lncRNAs are also classified as cis-acting and trans-acting depending on the location at which they function relative to their transcription site. LncRNAs whose function is based at, and dependent on their site of transcription are classified as **cis-acting** lncRNAs (**Figure 3A**) (reviewed in Gil and Ulitsky, 2019). However, the function of cis-acting lncRNAs is not necessarily limited to nearby genes, in fact, their distance and orientation relative to their target genes is highly variable and can range from few kilobases to whole chromosomes. The most radical example of their long acting range are the repressive effects of lncRNA *Xist* (*X-inactive specific transcript*), spanning the entire X chromosome from which they are transcribed in female mammals (Engreitz et al. 2013).

On the contrary, **trans-acting** lncRNAs are transcribed, processed and transported from their sites of transcription to wherever they exert their function, which includes both nuclear and cytoplasmic locations (**Figure 3A**). As the final destination of trans-acting lncRNAs does not depend on their transcription site, loss of function phenotypes can be rescued by their expression from exogenous locations. Although a great portion of the functionally described lncRNAs act in cis, a mechanism that is believed to be favoured in detriment to a trans-activity due to the generally low expression of lncRNAs, as their transport to other locations may dilute in excess these transcripts to mediate any function (Gil and Ulitsky, 2019). Several examples of lncRNAs acting in trans have also been characterised. For instance, the highly conserved lncRNA NORAD was described to maintain genomic stability by sequestering PUMILIO proteins (Lee et al. 2016; Tichon et al. 2016). This family of proteins recognize specific 3' UTR sequences of target mRNAs, whose function is particularly enriched in chromosomal segregation during cell division, and promote their deadenylation and decapping, resulting in the repression of mRNA stability and translation (Miller and Olivas, 2011). Upon the absence of NORAD, PUMILIO proteins are hyperactivated, inhibiting genes related with mitosis, DNA repair or DNA replication, resulting in chromosomal instability (Lee et al. 2016; Tichon et al. 2016). Thus, through their binding with PUMILIO proteins, NORAD prevents the repression of several hundreds of genes in trans, maintaining genomic stability (Lee et al. 2016).

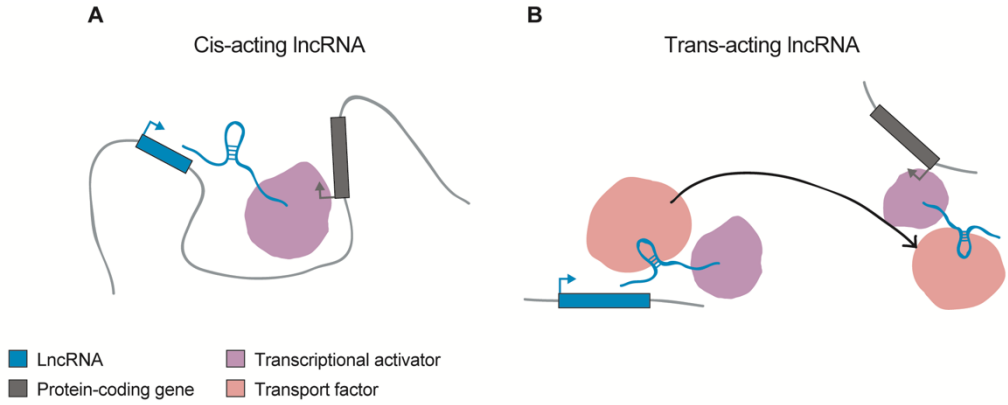


Figure 3. Representation of cis-acting and trans-acting lncRNAs. (**A**) LncRNAs acting in cis depend on their transription site to perform their functions. In this case, the lncRNA is bound by a transcriptional activator, promoting the transcription of a nearby protein-coding gene. (**B**) Trans-acting lncRNAs exert their function independently from their transcription site. They are often bound by proteins that mediate their transport to their sites of action.

Some lncRNAs have been described to function both in cis and in trans. For instance, the lincRNA-p21, which is activated by p53 in humans and mice, was initially characterised as a transcriptional repressor of a large subset of the p53-activated genes in trans, through the interaction with hnRNPK (Huarte et al. 2010). Additional functions in trans were also proposed for lincRNA-p21, including the repression of mRNA translation (Yoon et al. 2012) and the stabilisation of target proteins (Yang et al. 2014). However, studies on lincRNA-p21 knock-out mice also revealed its participation in the activation in cis of its neighbouring gene p21 (Dimitrova et al. 2014).

Functions of lncRNAs

The enormous heterogeneity of lncRNAs in terms of binding partners, subcellular localizations, sequence length or secondary structures confers them the ability to perform a wide variety of functions. Although only a small portion of all annotated lncRNAs have been functionally characterised, they have been described to participate in almost every step of gene expression, including chromatin remodelling, control of transcription, splicing, mRNA stability or translation (reviewed in Statello et al. 2021).

LncRNAs regulating chromatin states

Polycomb (PcG) and Trithorax (TrxG) group proteins are key modulators of an evolutionarily conserved gene regulatory system. They are chromatin regulators that tend to operate antagonistically, maintaining active (TrxG) or repressed (PcG) gene expression states. (Geisler and Paro, 2015; Schuettengruber et al. 2017). Many PcG/TrxG chromatin binding sites, referred as Polycomb response elements (PREs) and Trithorax response elements (TREs) in *Drosophila*, give rise to non-coding transcripts (reviewed in Ringrose, 2017). For instance, forward and reverse transcription has been detected from the *Drosophila vestigial* (*vg*) PRE/TRE, which switches the status of the element between silencing, induced by transcription from the forward strand, and activation, induced by transcription from the reverse strand. Moreover, the transcripts from the reverse strand are able to bind to the Polycomb repressive complex 2 (PRC2), inhibiting its enzymatic activity (Herzog et al. 2014).

Another example is the well-known human lncRNA HOTAIR, which also interacts with PRC2, promoting the trimethylation of the lysine 27 of histone 3 (H3K27me3) at the *HOXD* locus (Rinn et al. 2007). Additionally, HOTAIR also serves as a scaffold for the binding of the LSD1/CoREST/REST demethylase complex, thus coordinating not only the addition of H3K27me3 marks, but also promoting the demethylation of lysine 4 of histone 3 (Tsai et al. 2010).

One of the best examples to illustrate the functions that lncRNAs can exert in chromatin states, is their participation in dosage compensation mechanisms. In species in which sex is dictated by the XY sex-determination mechanism, such as mammals or flies, the imbalance in the expression of X-linked genes between females (XX) and males (XY) is corrected by a dosage compensation mechanism involving lncRNAs. In female mammals, the lncRNA *Xist* is upregulated in one of the X chromosomes at early embryonic stages and rapidly spreads along the X chromosome from which it is transcribed (Brockdorff et al. 1992; Brown et al. 1991, 1992). PRC2 is recruited by *Xist* and mediates H3K27me3. This signal triggers the heterochromatinization of the *Xist*-bound X chromosome, resulting in its inactivation (**Figure 4A**) (Lee et al. 1996; Penny et al. 1996; Wutz and Jaenisch, 2000). The opposite mechanism is observed in flies, where lncRNAs *roX1* and *roX2,* together with the MSL proteins, form the male-specific lethal complex (MSL) that overactivate the genes located in the X chromosome of male flies. Although very different in size and sequence, *roX1* and *roX2* (*RNA on the X 1 and 2*) act redundantly to allow the binding of MSL2 and the other subunits of the complex, which targets the X chromosome (Meller and Rattner, 2002). The MSL subunits mediate the activation of the X chromosome genes through the acetylation of lysine 16 in histone

H4 (H4K16ac) (**Figure 4B**) (Bone et al. 1994; Gelbart et al. 2009). In female flies, the *Sex lethal* (*sxl*) gene is upregulated, and the female-specific RNA-binding protein it encodes interacts with the mRNA of MSL2 to inhibit its translation, preventing the assembly of the MSL complex and the subsequent dosage compensation (Beckman et al. 2005; Gebauer et al. 1998).

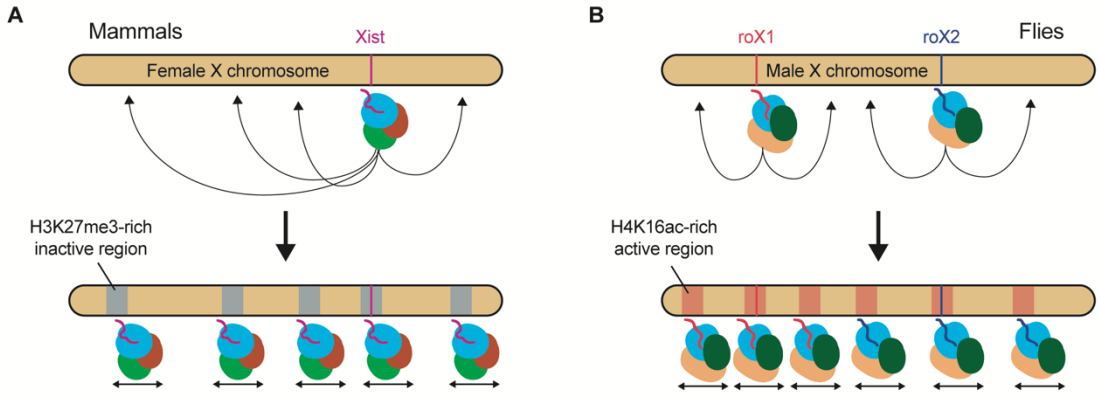


Figure 4. Participation of lncRNAs in dosage compensation mechanisms. (**A**) During X chromosome inactivation in female mammals, *Xist* is transcribed and accumulates in one of the X chromosomes. *Xist* is then bound, among others, by PRCs, and the ribonucleoprotein binds to different chromatin sites that are subsequently inactivated by H3K27me3 inactive marks, which spread along the X chromosome. (**B**) During X chromosome overactivation in male flies, *roX1* and *roX2* are bound by MSL proteins, forming the MSL complex. This complex targets multiple chromatin sites, promoting its activation by placing H4K16ac active marks, which spread along the X chromosome.

Although both strategies produce completely opposite outcomes, they share some mechanistic similarities. In both cases, lncRNAs are responsible for recruiting the chromatin-modifying complexes that drive the inactivation of the X-chromosome in female mammals, or the overactivation of the X-chromosome in male flies.

LncRNAs transcribed from active enhancers

Transcription from enhancers has been detected for multiple species, including mammals (Andersson et al. 2014; De Santa et al. 2010; Kim et al. 2010), *Drosophila* (Henriques et al. 2018; Meers et al. 2018) or *Caenorhabditis elegans* (Chen et al. 2013). These enhancer-derived transcripts can be classified into two groups: enhancer RNAs (eRNAs) and enhancer-associated lncRNAs (elncRNAs). While eRNAs are short RNAs transcribed bidirectionally, 5' capped, non-polyadenylated and generally unstable, elncRNAs are longer unidirectional transcripts, which are polyadenylated, and often spliced (reviewed in Arnold et al. 2020). Nevertheless, the distinction between the two enhancer-derived ncRNAs is not always clear and they are often confounded in the literature.

Although these RNAs are not transcribed from all enhancer regions, a correlation has been observed between enhancer activity and enhancer transcription both in mammals and flies (Hah et al. 2013; Mikhaylichenko et al. 2018). A growing number of studies demonstrate that specific enhancer-derived RNAs are required to properly activate the expression of their target genes (Lai et al. 2013; Lam et al. 2013; Ivaldi et al. 2018; Tsai et al. 2018). In mammals, eRNAs have been associated with the regulation of transcription through different mechanisms, including the interaction and enhancement of the activity of chromatin regulators, such as CBP, PRC2 or CTCF, the influence of

enhancer-promoter looping, or the alteration of Pol II elongation by the interaction with proteins that either induce or inhibit elongation (reviewed in De Lara et al. 2019). Similar mechanisms are described for elncRNAs, such as *CRED9*, whose transcription positively regulates the acetylation of the lysine 27 of histone H3 of the CEBPA enhancer region, thereby promoting CEBPA expression (Setten et al. 2021).

However, it is not clear whether these transcripts play an active role in enhancer activity, or if they are just transcriptional noise arising from the presence of the RNA polymerase machinery.

Regulators of isoform usage

A growing number of lncRNAs has also been linked to the modulation of alternative splicing in both plants and animals (reviewed in Romero-Barrios et al. 2018). For instance, the well-known lncRNA *MALAT1*, which acts as an oncogene and whose aberrant expression is involved in the development and progression of many cancer types (Malakar et al. 2016; Wang et al. 2016), is known to participate in the regulation of splicing. Particularly, *MALAT1* modulates the distribution and phosphorylation of serine/arginine-rich (SR) proteins (Tripathi et al. 2010), a conserved family of proteins involved in splicing. SR proteins recognize and bind to nascent pre-mRNAs and participate in the assembly of the spliceosome (Zhou and Fu, 2013). Moreover, their function is largely dependent on the phosphorylation or dephosphorylation of their SR domain, as it impacts not only their binding to their target pre-mRNAs, but also their localization. Depletion of *MALAT1* is described to increase the pool of dephosphorylated SR proteins, which display a more homogeneous nuclear distribution, and ultimately result in changes in the alternative splicing of pre-mRNAs (Tripathi et al. 2010).

On the other hand, some lncRNAs have also been described to modulate the splicing of particular target genes. In general, these lncRNAs are natural antisense transcripts that regulate the isoform usage of their overlapping protein-coding genes. In *Drosophila*, the lncRNA *blistered antisense* (*bsAS*), a natural antisense transcript of the *blistered* (*bs*) gene was recently described to regulate *bs* isoform usage (Pérez-Lluch et al. 2020). The *bs* gene encodes for the *Drosophila* serum response factor (DSRF) and is required for wing development and the formation of the wing vein patterning (Fristrom et al. 1994; Montagne et al. 1996; Roch et al. 1998). Transcription of *bsAS* occurs specifically in the intervein regions of the wing, and impairs the expression of the *bs* long isoform, thereby promoting the expression of the short isoform. In *bsAS* mutants, the overexpression of *bs* long isoform induces the formation of extra veins in the adult wings (Pérez-Lluch et al. 2020). The molecular mechanism driving the isoform regulation is presumably based on the formation of a genomic loop between *bs* and *bsAS* promoters, which impairs the transcription of the *bs* long isoform. This mechanism is independent of the presence of *bsAS* transcripts, as the ectopic expression of *bsAS* does not rescue the expression of *bs* long isoform in *bsAS* a mutant background (Pérez-Lluch et al. 2020).

Precursors and sponges of miRNAs

Some lncRNAs are also known to serve as microRNA (miRNA) precursors. Traditionally, miRNAs are thought to negatively regulate mRNA translation and stability by binding in

a sequence-specific manner to their 5' and 3' untranslated regions (UTRs), and in some cases even to their coding sequences (Forman et al. 2008). The canonical miRNA biogenesis includes the transcription of a larger non-coding precursor, which is cleaved in the nucleus by the Drosha/DGCR8 complex, then it is exported to the cytoplasm where it is further processed into mature miRNAs spanning ~ 22 nt (reviewed in Bartel, 2018). These mature miRNAs are loaded into the Argonaute (AGO) family of proteins, forming the miRNA-induced silencing complex (miRISC), which silences their target mRNAs by recruiting effector factors such as poly(A)-deadenylases or exoribonucleases (Jonas and Izurralde, 2015).

In most cases, miRNAs are derived from introns and exons of larger protein-coding and non-coding genes. In *Drosophila*, one of these non-coding transcripts is iab-8, which is transcribed primarily from the posterior central nervous system. It spans over 90 kb and is both spliced and polyadenylated (Bender et al. 2008; Garaulet et al. 2014). Once transcribed, it is processed into three miRNAs called *miR-iab-8*, encoded within its intronic sequence. These miRNAs target and downregulate the homeotic genes *abdominal A* (*abd-A*) and *ultrabithorax* (*Ubx*), as well as their cofactors *homothorax* (*hth*) and *extradenticle* (*exd*) (Garaulet et al. 2014; Gummalla et al. 2012). Loss of *iab-8* results in the increase in the level of their targeted transcripts, leading to male and female sterility due to a defective innervation of the abdominal and reproductive tract muscles of the fly (Maeda et al. 2018).

Another example is the maternally imprinted *H19* gene, which encodes the first lncRNA to be identified, which was later discovered to serve as the precursor of *miR-675* in humans and mice (Cai and Cullen, 2007). *H19* is highly expressed in foetal tissues, where it is processed into *miR-675*, limiting placental growth by targeting, among others, the growth promoting gene *Igf1r* (Keniry et al. 2012). In parallel, *H19* is also expressed in the adult skeletal muscle where, instead of being processed into *miR-675*, acts as a molecular sponge for the *let-7* family of miRNAs, sequestering them, thus blocking the inhibition of their target mRNAs (Kallen et al. 2013; Onyango and Feinberg, 2011).

Another lncRNA that is processed into smaller RNAs is *acal* (Ríos-Barrera et al. 2015). *acal* is one of the few lncRNAs showing sequence conservation in *Drosophila*. In particular, a 296-nt fragment is 80% sequence identical in *Drosophila melanogaster* and *Drosophila bipectinata*. Also, a similar-sized lncRNA is found in humans, showing a considerable 48% sequence identity to *Drosophila acal* (Murillo-Maldonado and Riesgo-Escovar, 2019). It was found that *acal*, through the regulation of two JNK modulators, *Connector of kinase to AP1* (*Cka*) and *anterior open* (*aop*), is able to modulate JNK activity (Ríos-Barrera et al. 2015). Remarkably, *acal* is transcribed from a mono-exonic gene into a 2.3-kb long transcript that, throughout the life cycle of the fly, particularly during pupal stages, is processed into smaller transcripts spanning 50 to 120 nt. The function of these small RNAs is yet to be investigated, but the differences in size with respect to the typical ~22 nt miRNAs indicate that processed *acal* does not act as a typical miRNA (Ríos-Barrera et al. 2015).

Regulators of translation

In addition, multiple pieces of evidence suggest that lncRNAs can regulate every step of translation by regulating the expression and function of translation factors (reviewed in

Karakas and Ozpolat, 2021). For instance, the lncRNA treRNA interacts with several ribonucleoproteins to form the treRNA-RNP complex, which binds to the translation initiation factor eIF4G1, suppressing the translation of E-cadherin (Gumireddy et al. 2013).

On the other hand, lncRNAs can also regulate translation by interacting with the ribosome. LncRNA lncNB1 is capable of binding to the ribosomal protein RPL35, promoting the translation of E2F1, which in turn stabilises N-Myc in cancer cells (Liu et al. 2019). Another example is the lncRNA ZFAS1, which associates with the ribosomal subunit 40S, regulating the production and assembly of ribosomes, thus indirectly influencing translation (Hansji et al. 2016).

LncRNAs encoding small ORFs

Although by definition lncRNAs lack protein-coding potential, it is estimated that roughly 98% of all annotated lncRNAs in humans, mice and flies contain small ORFs (smORFs) of 10 to 100 aa that may code for short peptides (Couso and Patraquim, 2017). Translation of smORFs is observed in many eukaryotes (Couso and Patraquim, 2017; Andrews and Rothnagel, 2014), however, the putative function of these peptides is often neglected, and the genes that encode them remain listed as non-coding.

Examples of functional peptides encoded by smORFs have been described primarily in humans (Andersson et al. 2015; Nelson et al. 2016) and insects (Galindo et al. 2007; Magny et al. 2013). For instance, in *Drosophila*, the *tarsal-less* (*tal*) gene encodes a polycistronic mRNA that is translated into 4 small peptides of 11 aa each. One of these peptides actively participates in the development of the leg imaginal disc by regulating gene expression, tissue folding, and Notch signalling (Galindo et al. 2007; Pueyo and Couso, 2011). Moreover, similar smORFs are present in *tal* homologues across different insect species, suggesting that their function could be conserved (Galindo et al. 2007).

Multiple studies using ribosome profiling techniques (Ribo-seq), which selectively identify the transcripts bound to the ribosomes, corroborated that a fraction of lncRNAs had a strong affinity for ribosomes (Aspden et al. 2014; Carlevaro-Fita et al. 2016; Ruiz-Orera et al. 2014). However, this association to the ribosomes does not necessarily imply that lncRNAs are actively translated, since some lncRNAs are known to regulate mRNA translation through ribosome binding (Carrieri et al. 2012; Hansji et al. 2016; Liu et al. 2019). To overcome this limitation, further studies on ribosome-bound lncRNAs should be considered, such as peptide tagging or in vitro translation assays to identify the coding potential of smORFs, while their functionally should be tested by mutating or inhibiting the transcription of the candidate lncRNAs.

The increasing number of functional smORF encoded by genes annotated as lncRNAs challenges the already discussed definition of lncRNAs. The fact that the vast majority of the annotated lncRNAs contain at least one predicted smORF, makes it impossible to rule them out just by the smORF presence. However, the lncRNA status of genes coding for functional peptides should be revised or, alternatively, the lncRNA definition should be updated to include the genes encoding for functional smORFs.

Conservation of lncRNAs

A main drawback in the study of lncRNAs is the difficulty of identifying orthologous genes in other species. Unlike mRNAs, whose translation can be truncated by a single point mutation, lncRNAs do not seem to depend on their sequence to perform their functions, leading to their rapid evolution and sequence degeneration compared to mRNAs. Nevertheless, lncRNAs are significantly more conserved than neutrally evolving sequences (Ponjavic et al. 2007; Guttman et al. 2009; Derrien et al. 2012), indicating that they are actually subject to purifying selection. Indeed, few examples of sequence-conserved lncRNAs have been described for evolutionarily closer species. For instance, the sequence of MALAT1 is exceptionally well conserved in mammals, including a remarkable sequence identity of > 70% between human and mouse (Ma et al. 2015; Hutchinson et al. 2017). Another example is the *yellow-achaete intergenic RNA* (*yar*), which is a lncRNA involved in fly sleep regulation. Several motifs ranging from 40 to 111 bp located in the TSS, the exons and the 3' end of *yar* are conserved in several *Drosophila* species separated by as much as 40-60 million years of evolution (Soshnev et al. 2011).

Despite this, the high substitution rate of lncRNAs usually impedes the identification of orthologous genes in other species. For this reason, alternative approaches are often used to search for orthologous lncRNAs, such as the conservation of their secondary structure, the conservation of their position in the linear genome, or the presence of small conserved regions.

Secondary structure

Conservation analysis of lncRNAs secondary structures is an alternative approach in determining lncRNA orthology. As the function of some lncRNAs is dependent on their folding conformation, it is thought that their secondary structure should be more conserved than their primary sequence (Smith et al. 2013; Graf and Kretz 2020).

It is possible to experimentally validate the RNA secondary structure with increased accuracy, generally following the selective 2' hydroxyl acylation analysed by primer extension (SHAPE) and SHAPE-derived methods (Deigan et al. 2009; Wilkinson et al. 2009; Smola and Weeks, 2018). However, these methods cannot be used to systematically analyse the structure of dozens of RNAs, as they are costly and time-consuming. Instead, in silico RNA folding prediction tools are often used to infer the secondary structure of RNAs. Currently, these tools are mainly based on the minimum free energy (MFE) algorithm, which finds the optimal secondary structure using an iterative method to meet the MEF criteria (Hofacker et al. 1994; Zuker, 2003). However, it has been demonstrated experimentally that the RNA is generally folded *in vivo* in suboptimal free energy structures (Zou et al. 2008), questioning the utility of these tools. In fact, the accuracy of the MEF algorithm varies greatly depending on the size of the RNA molecules analysed, as the biological folding of shorter RNA sequences approximates to the minimum free energy constraint, while the accuracy of this method decreases for longer RNAs (Zhang et al. 2019). Thus, secondary structure prediction methods do not seem accurate enough to consider the RNA folding form as the primary source for the identification of lncRNA orthologs.

Synteny

The positional conservation of genes across different species, or synteny conservation, emerged recently as a valuable tool in the identification of orthologous non-coding genes (Bryzghalov et al. 2021; Pegueroles et al. 2019; Herrera-Ubeda et al. 2019; Rolland et al. 2019). This analysis relies on the presence of orthologous genes located in the same order in the linear genome of different species. Syntenic conservation of the region surrounding the lncRNA locus could be an indicator of lncRNA orthology, as it is more probable that the lncRNA arose in their common ancestor, rather than appearing in different species as a consequence of independent evolutionarily events.

However, the presence of a lncRNA conserved by synteny in different species does not necessarily imply orthology. Particularly, the presence of large intergenic regions containing multiple lncRNAs increases the rate of false positives (Young et al. 2012). In addition, since the analysis of synteny depends on the presence of orthologous genes, it works better in evolutionarily closer species and becomes less useful as the evolutionarily distance increases between the species being compared.

Microhomology regions

Despite lacking sequence conservation, smaller regions of homology among different species have been observed for lncRNAs (Ulitsky et al. 2011; Hezroni et al. 2015; Ruiz-Orera and Albà, 2019; Ross et al. 2021). These microhomologous regions are thought to correspond to functional elements that are essential for the function of the lncRNA, such as RBP motifs or miRNA binding sites. It is important to note that these binding sites are very short (4-12 nt) and individual matches between different species can be found purely by chance. An interesting approach to bypass the rate of false positive hits is the addition of order to these elements (Ross et al. 2021). In this way, not only the presence of these motifs is considered, but also the order in which they are found in putatively orthologous lncRNAs.

Although this method has not been tested for distantly related species, finding small regions of homology should be more achievable than finding orthologous lncRNAs using the current methods based on whole-sequence similarity or secondary structure predictions.

DROSOPHILA IMAGINAL DISCS AS A MODEL FOR THE STUDY REGENERATION

Imaginal discs are the epithelial sacs determined in early embryogenesis that give rise to the adult structures after the differentiation during metamorphosis. A total of 19 imaginal discs are present in the *Drosophila* larvae, including the clypeolabral, eye-antennal, labial, humeral, leg, haltere, wing and genitalia discs, all of which are duplicated except for the leg (6 total discs) and genitalia (1 disc) discs (**Figure 5**) (reviewed in Beira and Paro, 2016). During the larval stages, imaginal discs grow and proliferate, and during metamorphosis, they differentiate to form the adult structures. In this work, we will focus on the use of wing imaginal discs as a model system for the study of regeneration.

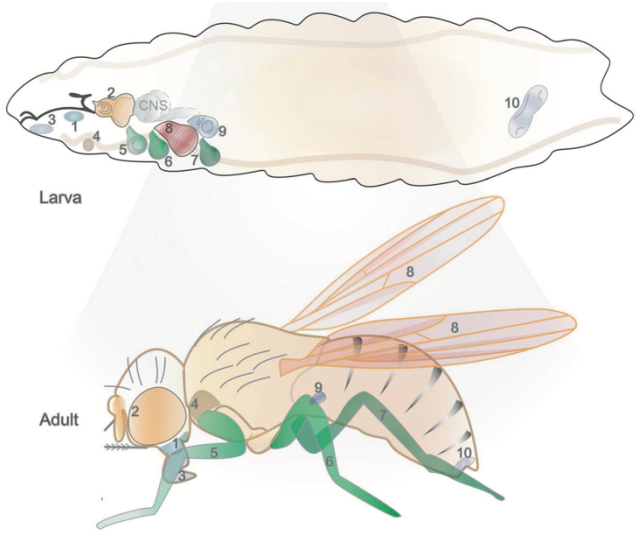


Figure 5. Imaginal discs and their derived adult structures. Distribution of the different imaginal discs types present in the larval stage (top picture) and the corresponding structures in the adult fly (bottom picture). Representation of the clypeolabral (1), eye-antennal (2), labial (3), humeral (4), leg (5-7), wing (8), haltere (9) and genitalia (10) discs. Adapted from Beira and Paro, 2016.

Wing imaginal discs derive from the embryonic wing primordia, which originates from a cluster of 20-70 cells from polyclonal origin that isolate by invagination from the embryonic ectoderm (Madhavan and Schneiderman, 1977; Cohen et al. 1993). By the end of the first instar larval stage, the wing primordia starts to proliferate, ending up with around 50,000 cells prior to pupariation (Garcia-Bellido and Merriam, 1971). Since the division rate is rather constant and apoptosis is scarce, the final wing size and shape is controlled by proliferation, which depends not only on the growth rate, but also on the duration of growth. Both parameters are influenced by genetics, but also by other variables such as temperature, diet, crowding or infections (reviewed in Edgar, 2006).

The pattern of the wing primordium is constituted by two major axes, the antero-posterior (AP) and the dorso-ventral (DV) axis (**Figure 6 A**). Cell fate determination depends on the presence of selector genes, which establish lineage restriction barriers within the AP and DV boundaries, impeding the mixture of population cells from different compartments (Garcia-Bellido et al. 1973). Morphologically, wing imaginal discs are a continuous epithelial monolayer forming a two-sided epithelial sac. One of the disc sides

is a columnar pseudostratified epithelium, or **columnar epithelium**, while the other side of the disc is composed by a squamous epithelium of wide and flatten cells, the **peripodial membrane** (**Figure 6 B**).

At the end of the third instar larval stage, the 20-hydroxyecdysone hormone induces the entry of the larvae into the pre-pupal stage, and major morphogenetic events place in the mature imaginal discs (Fristrom and Fristrom, 1993; Riddiford and Truman, 1993). While the columnar epithelium everts from the wing pouch and gives rise to the adult wing, notum and hinge (**Figure 6 C**), the function of the peripodial membrane is to facilitate the process of eversion and thorax fusion, and does not contribute to the formation of the wing (Pastor-Pareja et al. 2004; Tripura et al. 2011).

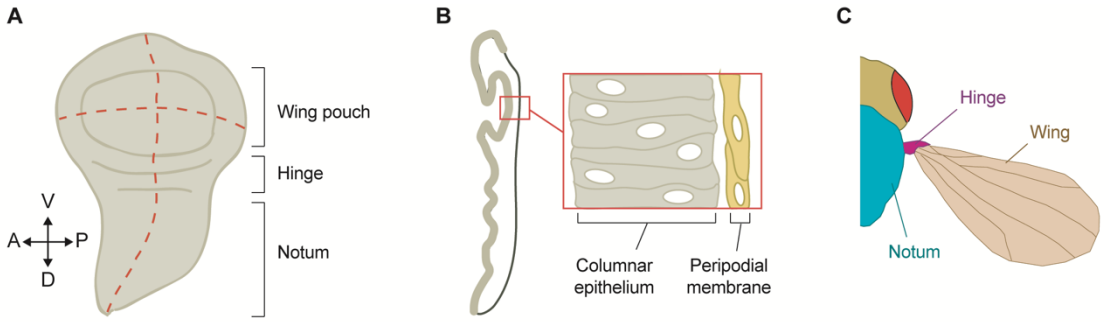


Figure 6. *Drosophila* wing imaginal discs. (**A**) Morphology of wing imaginal discs, showing the Antero-Posterior (AP) and Dorso-Ventral (DV) axes and the different parts of the disc. (**B**) Lateral view of wing discs, zooming into the cellular organization of the columnar epithelium and the peripodial membrane. (**C**) Diagram of an adult fly depicting every part deriving from wing discs.

Regeneration

Regeneration is the ability to reconstruct the original shape, size and function of body parts that have been physically or functionally lost or damaged. First defined by Thomas Morgan in 1901 as the replacement of missing structures upon injury, the study of regeneration has taken major steps towards the understanding of the regenerative process, pointing towards an ideal future in which we could confer the regeneration ability to non-regenerating tissues, organs or organisms.

The capacity to regenerate is widely diverse across the animal kingdom, and it can range from the ability to recover specific body structures, such as limb regeneration in axolotl or heart and fish regeneration in zebrafish, to the ability to regenerate all body parts, as it is the case for planarians, which are able to reconstruct a complete organism from small body fragments.

According to the biological level at which the recovery process occurs, regeneration can be classified into cellular regeneration, tissue regeneration, organ regeneration, structural regeneration, or extensive regeneration (reviewed in Bely and Nyberg, 2009). An example of **cellular regeneration** is the process of axonal growth upon the loss of the distal portion of nerve axons. **Tissue regeneration** is clearly exemplified by the recovery of epithelial injuries, where a homogeneous cell population proliferates to fill the gap left by the injured cells. Examples of **organ regeneration** include the recovery

of the human liver upon partial hepatectomy, or the regeneration of the zebrafish heart, recovering the original size and function of the affected organ. **Structural regeneration** refers to the recovery of missing body parts, such as tails, limbs, heads, tentacles or fins. Finally, **extensive regeneration** or whole-body regeneration is the ability to reconstruct a whole organism from small body fragments, as it occurs in *Hydra* or planarians.

Despite the differences in the regeneration ability across different tissues, organs and species, the molecular mechanisms underlying the recovery process appear to be conserved. The action of quick bioelectric signals upon injury is required to sense the damage (Levin, 2009), triggering the subsequent steps in the recovery process, which include the healing of the wound (Gurtner et al. 2008), the proliferation of the living cells to replace the death cells (Tanaka and Reddien, 2011), and the proper remodelling mechanisms to reconstruct the damaged area (Iisma et al. 2018).

The early signals of regeneration

A rapid response upon damage is essential to trigger in the living cells the regeneration programs needed for the recovery of the injured area. These early signals include bioelectrical stimulus, calcium waves and reactive oxygen species (ROS), and their ultimate function is the activation of multiple signalling pathways required for regeneration, including the JNK and p38 pathways, the Hippo pathway, the Wnt pathway or the Jak-STAT pathway (**Figure 7**) (Bergantiños et al. 2010; Whyte et al. 2012; Repiso et al. 2013; Sun and Irvine, 2013; Santabárbara-Ruiz et al. 2015; López-Luque et al. 2016).

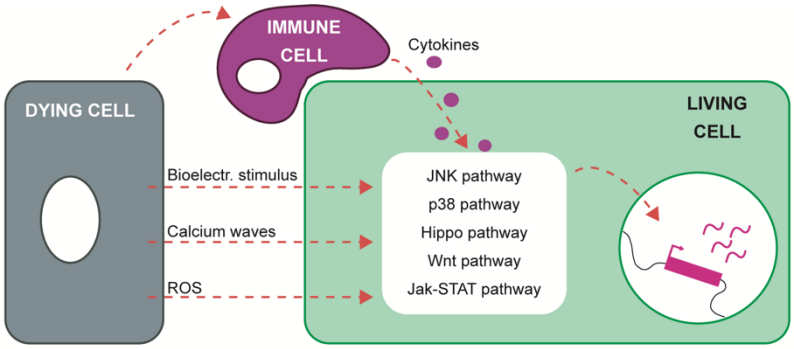


Figure 7. Early regeneration signals. The dying cells release a series of bioelectrical stimulus, calcium waves and reactive oxygen species (ROS), attracting immune cells, which release cytokines. These early signals are sensed by the neighboring living cells, activating a series of signalling pathways that activate the transcriptional programs needed for the recovery process. Adapted from Vizcaya-Molina et al. 2019.

Bioelectrical signals are generated by specific ion channels and pumps within cell membranes, changing their voltage. These changes in membrane polarity are transduced by membrane molecules, such as the phosphoinositide phosphatase, which is able to activate, among others, Notch, PTEN or NF-kB (Li et al. 2002; Murata et al. 2005; Tao et al. 2006).

The presence of quick calcium waves upon wounding have been described in multiple organisms, including *Xenopus* (Tu and Borodinsky, 2014), zebrafish (Yoo et al. 2012) or *Drosophila* (Razzell et al. 2013). It has been proposed that these calcium waves originate in the dying cells and propagate to the living neighbouring cells via gap junctions,

activating the oxidase DUOX and leading to the production of ROS (Santabárbara-Ruiz et al. 2015; Narciso et al. 2015; Restrepo and Basler, 2016). In turn, ROS are required for the activation of redox-sensitive signalling cascades, such as the Mitogen-Activated Protein Kinases (MAPKs) p38 and JNK (McCubrey et al. 2006; Jiand et al. 2011; Santabárbara-Ruiz et al. 2015). The presence of ROS also mediates the recruitment of inflammatory cells, which release cytokines that are sensed by the neighbouring living cells, promoting the activation of the Jak-STAT signalling pathway (Moreira et al. 2010; Niethammer et al. 2009; Santabárbara-Ruiz et al. 2015). Upon activation, these signalling cascades translocate the signal to the nucleus, where their effector transcription factors (TFs) mediate the activation of regeneration programs.

Production of new cells to close the wound

The replacement of a missing or injured structure requires the production of new cells, which is stimulated by the early signalling events triggered upon damage. The origin of these new cells is variable, and includes the proliferation of stem cells, the division of differentiated cells, or the dedifferentiation or transdifferentiation of mature cells into precursor cells (reviewed in Tanaka and Reddien, 2011). The interplay between one or more of these mechanisms depends on the species and the tissue to be reconstructed (**Figure 8**).

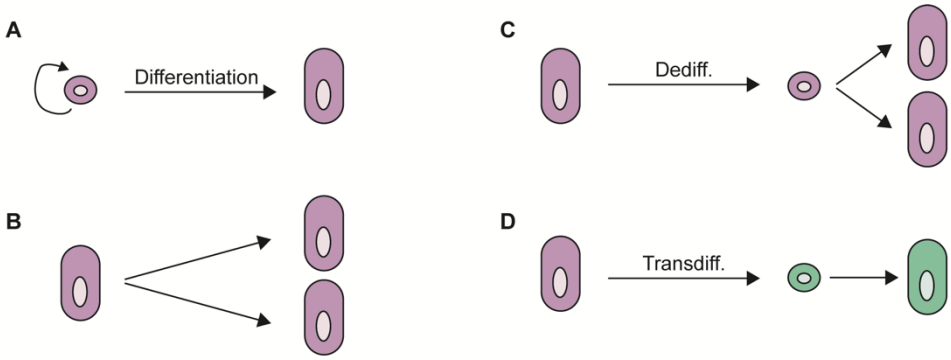


Figure 8. Sources of new cells in regeneration. (**A**) Stem cells self-renew to produce more differentiated cells. (**B**) Differentiated cells can proliferate to recover the missing tissue. (**C**) Differentiated cells dedifferentiate into precursor cells, which proliferate and produce more differentiated cells. (**D**) Differentiated cells can transdifferentiate into pluripotent cells, which differentiate into other cell types. Adapted from Tanaka and Reddien, 2011.

For instance, **stem cell-based proliferation** is widely described in *Hydra* and planarians, both of which can regenerate whole individuals from small cut fragments thanks to the existence of distinct stem cell populations throughout their organism (Baguñà et al. 1989; Galliot et al. 2006). In other cases, the **division of differentiated cells** is enough to recover the missing tissue and close the wound, as it happens with the proliferation of hepatocytes upon liver injury or partial hepatectomy (Michalopoulos and Bhushan, 2021). Another mechanism to produce new cells is the **dedifferentiation** of cells located near the wound. It contributes to the appendage regeneration in newts and axolotls (Lo et al. 1993; Laube et al. 2006; Lin et al. 2021), or to the heart and fin regeneration in zebrafish, where cardiomyocytes and osteoblasts, respectively, dedifferentiate and proliferate to recover the missing tissue (Kikuchi et al. 2010; Knopf et al. 2011). **Transdifferentiation**, the process by which differentiated cells regress to a

point where they can switch lineages and differentiate into another cell type, has been reported in newt lens regeneration (Henry and Tsonis, 2010; Jopling et al. 2011).

Recovery of the original pattern

Traditionally, two different methods to regenerate were described: epimorphosis and morphallaxis. **Epimorphosis** is based on the dedifferentiation of adult structures to form the blastema, a mass of undifferentiated pluripotent cells that drives the regeneration of the lost structures through active proliferation. Examples of regeneration by epimorphosis include the regeneration of planarians (Baguñà, 1976; Alvarado, 2006), limb regeneration in amphibians (Brockes, 1997), heart and fin regeneration in zebrafish (Poss et al. 2002; Pfefferli and Jazwinska, 2015), or the regeneration of digital tips in mammals (Muller et al. 1999). On the other hand, in regeneration by **morphallaxis** most of the regenerated tissue comes from the remodelling of pre-existing tissue, as very little cell proliferation is involved. For instance, upon cutting a *Hydra* in half, both parts can regenerate a fully functional individual from the rearrangement of pre-existing cells, resulting in two smaller *Hydra* (Bosch, 2006).

Nevertheless, the classification of regeneration into two different categories, misleadingly suggests that regeneration in different animals may be controlled by different principles. However, in both cases, the **positional information** integrated by the living cells surrounding the wound is essential for the recovery of the original pattern. Evidence of the importance of the positional information was already revealed in the 1970s from the results of grafting and regeneration experiments using larval cockroach legs (French et al. 1976). The experiments showed that the joining of normally non-adjacent positions within the leg segment resulted in intercalary growth of the intermediate structures (**Figure 9**). Similar results were obtained during the regeneration of amphibian appendages (Maden, 1980) or *Drosophila* imaginal discs (Haynie and Bryant, 1977), suggesting a conserved mechanism of intercalary regeneration.

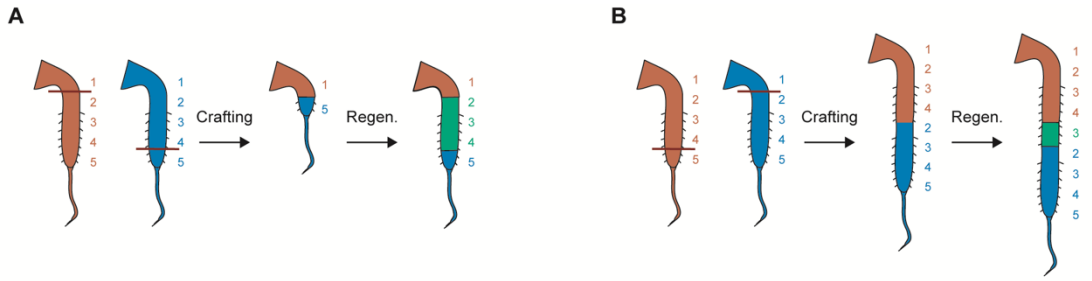


Figure 9. Intercalary regeneration of cockroach limbs after crafting. (**A**) When the distal fragment at level 5 is joined to the proximal stump at level 1, the missing parts can be regenerated (2-4) and legs with normal size and structure are restored. (**B**) When the distal fragment at level 2 is joined to the proximal stump at level 4, an intercalary structure between the two levels is regenerated (3). Adapted from Agata et al. 2003.

Regeneration of imaginal discs

The numerous advantages of working with *Drosophila* have been extensively described in the literature, and include their short generation times, high number of progenies, low maintenance costs, or the easy handling methods (Hales et al. 2015). Moreover, the

genome of *Drosophila* is much smaller, more compact and less redundant than the human genome. Indeed, roughly 60% of protein-coding genes annotated in flies have human orthologs (Wangler et al. 2015), while around 75% of human disease-related genes have a fly ortholog (Reiter et al. 2001). For this, *Drosophila* is widely used as a model for many human genetic disorders, including the transference of patient mutations into 'avatar' flies, which are then used to infer the response of patients to specific drugs (Bangi et al. 2021).

The use of *Drosophila* as a model system for the study of regeneration dates back to the 1940s (reviewed in Hariharan and Serras, 2017), when pioneering studies based on cutting fragments of larval imaginal discs and subsequently culturing them in adult female abdomens, resulted in the recovery of the missing parts thanks to local cell proliferation (Hadorn et al. 1949; Ursprung et al. 1959). The reimplantation of these regenerated discs into larvae and their successive differentiation during metamorphosis proved that the functionality of the cut tissue was restored (Hadorn and Buck, 1962; Schubiger, 1971). Later, the development of culture media mimicking the larvae hemolymph replaced these techniques by the *ex vivo* culture of imaginal discs (Handke et al. 2014; Restrepo et al. 2016).

An alternative and more precise approach to study regeneration arose thanks to the development of genetic transactivation systems such as the Gal4/UAS system (Brand and Perrimon, 1993). The addition of a temperature-sensitive mutant of the Gal80 repressor, allowed the genetic induction of cell death by overexpressing proapoptotic genes like *reaper* (*rpr*), *head involution defective* (*hid*) or *eiger* (*egr*) in specific regions of imaginal discs during specific periods of time (Smith-Bolton et al. 2009; Bergantiños et al. 2010; Herrera et al. 2013). In this way, the selection of specific *Gal4* constructs under the control of different gene promoters was used to select the region of cell death induction. Parallelly, the temperature shift from 17ºC to 29ºC allowed the expression of the proapoptotic genes for specific periods of time, and then larvae were shifted back to 17ºC to inhibit the Gal4 again and allow the tissue recovery.

The early signals triggered upon damage of wing imaginal discs include the propagation of calcium waves and a burst of ROS produced from the dying cells, recruiting haemocytes to the wound site, where they release cytokines (Santabárbara-Ruiz et al. 2015; Fogarty et al. 2016). These early signals are propagated to the living cells near the wound, activating the JNK, p38 or Jak-STAT pathways, which are required for wound healing and regenerative growth (Pastor-Pareja et al. 2008; Katsuyama et al. 2015; Santabárbara-Ruiz et al. 2015).

The activation of these signalling pathways leads to the transcription of regeneration programs that drive a burst of active transcription at the early stages of the recovery process (Vizcaya-Molina et al. 2018). Among these activated genes, *wingless* (*wg*) is essential for the activation of the growth-promoting Myc (Gibson and Schubinger, 1999; Herranz et al. 2008; Smith-Bolton et al. 2009). Additionally, as a result of the changes in cell tension, the Hippo pathway is inhibited, activating Yorkie (Yki), which collaborates with Scalloped (sd) in growth promotion (Sun and Irvine, 2011; Repiso et al. 2013).

To recover the original patterning, the respecification and intercalary growth of the cells surrounding the wound is also necessary (Smith-Bolton et al. 2009, Herrera et al. 2013; Repiso et al. 2013). Particularly, the cells near the AP and DV boundaries are described to change their cell fate to adopt new compartmental identities (Herrera and Morata, 2014).

One of the aspects influencing the regeneration capacity of wing imaginal discs is their developmental stage: early L3 larvae regenerate efficiently, while late L3 larvae cannot regenerate (Smith-Bolton et al. 2009). The loss of the regenerative capacity is mainly due to the epigenetic silencing of key regulatory elements whose activation is necessary for proper regeneration, such as the aforementioned *wg* (Harris et al. 2016). With the aim to elongate the early larval stages in which the fly is able to regenerate, damaged discs release the *Drosophila* Insulin-like peptide 8 (Dilp-8) which, together with the action of retinoids, mediates a pupariation delay to allow the recovery of the injured tissue (Halme et al. 2010; Colombani et al. 2012; Boone et al. 2016).

Recently, genome-wide studies interrogating transcriptional response and chromatin activity during *Drosophila* wing disc regeneration revealed the presence of damage-responsive regulatory elements or DRREs (Vizcaya-Molina et al. 2018). They are defined as conserved genomic regions that become activated upon the induction of cell death. These DRREs contain conserved binding sites for TFs that are downstream from the signalling pathways that are activated upon damage, such as the Toll, Jak-STAT, JNK or Hippo pathways. Two types of DRREs were defined: increasing DRREs (iDRREs) are those chromatin regions that become more accessible in regeneration than in control samples, and emerging DRREs (eDRREs) are those regions that are only accessible in regeneration. Additionally, it was found that some of these eDRREs were actually reused, as around half of all identified eDRREs were also active in other tissues or developmental stages (Vizcaya-Molina et al. 2018). Similar studies on heart regeneration in zebrafish also revealed the presence of emerging regulatory programs that are only accessible upon damage (Goldman et al. 2017).

GROWTH ARREST AND DNA DAMAGE 45 (GADD45)

One of the protein coding genes found upregulated at the early stages of wing imaginal disc regeneration is *D-GADD45,* the fly homolog of Growth Arrest and DNA Damage-inducible 45 (GADD45) family of proteins, which acts as stress sensors in response to various stimuli. The first *GADD45* gene was identified in mammals on the basis of its increased expression after growth cessation signals or treatment with DNA-damaging agents (Fornace et al. 1988, 1989). This gene, renamed as *GADD45α*, together with *GADD45β* and *GADD45γ,* forms the highly conserved *GADD45* family in mammals. These genes code for small (18 kDa) and highly acidic proteins that are mainly located in the nucleus (Zhan et al. 1994; Carrier et al. 1999).

The expression of all *GADD45* genes is induced after exposure to several genotoxic or oxidative agents, such as ultraviolet radiation, methyl methanesulfonate or hydrogen peroxide (Fornace et al. 1988). The different GADD45 family members have been implicated in a variety of responses to cell injury, including cell cycle checkpoints, apoptosis or DNA repair (reviewed in Liebermann and Hoffman, 2008). Since GADD45 proteins do not have enzymatic properties, it has been suggested that they function by physically interacting with partner proteins (**Figure 10**). Thus, upon stress induction, GADD45 interacts with different proteins involved in the different stress responses (Liebermann and Hoffman, 2008; Niehrs and Schafer, 2012; Salvador et al. 2013).

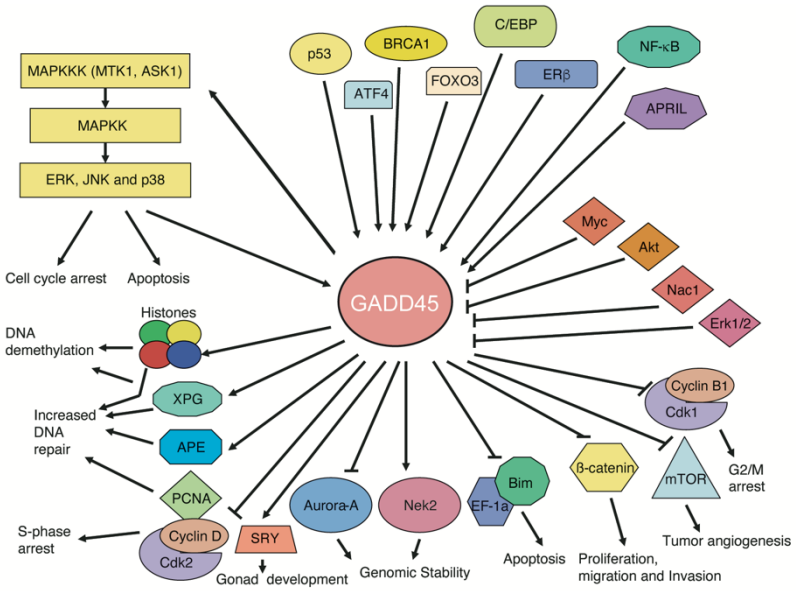


Figure 10. Described regulators and effectors of GADD45. Arrows indicate positive regulation, while capped lines indicate negative regulation. Adapted from Salvador et al. 2013.

GADD45 and the JNK and p38 signalling pathways

Although GADD45 proteins show complex regulation and numerous effectors, many of the prominent roles for the GADD45 proteins are associated with signalling mediated by

MAPKs (Papa et al. 2004, 2007; Miyake et al. 2007; Salvador et al. 2013). This association is rather complex for JNK and p38 proteins, members of the MAPKs pathways, which can contribute to upregulate *GADD45* and, at the same time, are effectors of GADD45 signalling (Salvador et al. 2013). JNK and p38 are also activated upon cellular stress, initiating the signalling modules that lead to the transcription of target genes involved in growth, differentiation, survival and apoptosis (reviewed in Hotamisligil et al. 2016). JNK and p38 can exert antagonistic effects on cell proliferation and survival, which depend on cell type-specific differences, as well as on the intensity and duration of the signal (Wagner and Nebreda, 2009).

In mammals, GADD45 proteins directly bind to MTK1/MEKK4, a MAP Kinase Kinase Kinase (MAP3K) upstream of both JNK and p38, activating its kinase catalytic domain and thus promoting the activation of JNK and p38 signalling pathways (De Smaele et al. 2001; Miyake et al. 2007). Additionally, co-immunoprecipitation studies aimed to identify putative binding partners of GADD45β within the MAPK cascades revealed another MAP3K upstream of JNK and p38, ASK1, to be bound by GADD45β (Papa et al. 2004). However, the effects of this interaction are unknown, as no differences in the ability of ASK1 to activate its downstream effectors were reported (Papa et al. 2004). Moreover, GADD45β has also been proposed to directly bind to MKK7, a MAP Kinase Kinase (MAP2K) downstream of MTK1 and ASK1, inhibiting its kinase activity in mouse fibroblasts (Papa et al. 2004, 2007; Tornatore et al. 2008; Ueda et al. 2017).

As the described interactions between GADD45 and the different members of the MAPK cascade can lead to the activation or inhibition of the downstream effectors, it has been suggested that the effect of GADD45 on these signalling pathways might be tissue-specific and context-dependent.

GADD45 in flies

D-GADD45 contains a single exon encoding a 163-aa protein. Opposite to their mammalian counterparts, *D-GADD45* is not upregulated upon different genotoxic stresses, including the exposure to arsenite, UV-light, X-rays or paraquat (Peretz et al. 2007). Instead, the induction of *D-GADD45* seems strongly dependent on the inflammatory response, as it was strongly activated upon septic injury (Peretz et al. 2007). In agreement with this hypothesis, experiments on fly embryos revealed that *D-GADD45* was not activated upon wounding in mutants defective for the inflammatory response (**Figure 11 A-C**) (Stramer et al. 2008). Similar observations were also reported for *GADD45α* in mice (**Figure 11 D-E**) (Stramer et al. 2008), suggesting a conserved mechanism of *GADD45* activation upon inflammation.

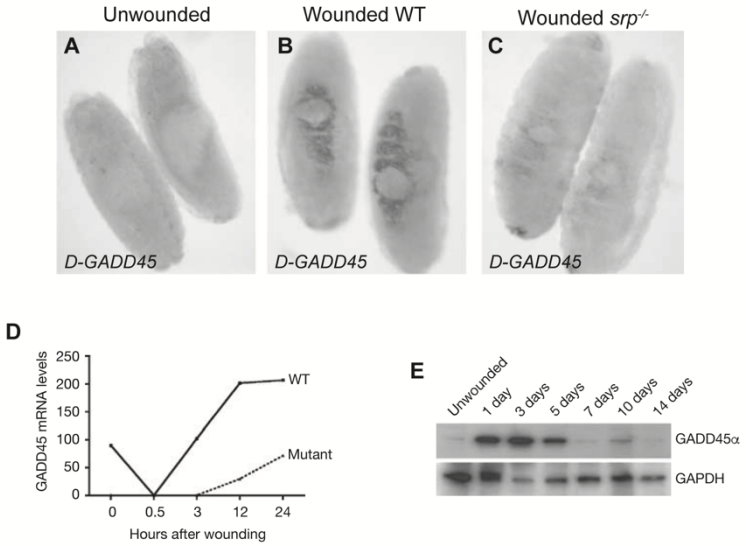


Figure 11. Upregulation of *D-GADD45* upon wounding is dependent on the inflammatory response. (**A-C**) *In situ* hybridization of *D-GADD45*, showing its upregulation in wounded wild type (WT) fly embryos but not in mutants defective for the inflammatory response (*serpent* null mutants, *srp*$^{-/-}$). (**D**) Transient increase in the mRNA levels of *GADD45α* in wild type and inflammation-defective mutant murine skin. (**E**) Western blot showing the transient upregulation of *GADD45α* upon wounding of murine skin. Adapted from Stramer et al. 2008

Similar to mammals, the effects of inducing the expression of *D-GADD45* seem to be tissue-specific also in the fly. For instance, overexpression of *D-GADD45* in the nervous system increases the lifespan of flies by a mechanism supposed to involve a protection against oxidative stress (Plyusnina et al. 2011; Moskalev et al. 2012). On the other hand, the activation of *D-GADD45* in the germline causes several patterning and polarity defects, as it affects the localization of several mRNAs involved in the establishment of the anterior-posterior axis, such as *gurken* and *oskar*, by a mechanism still unknown (Peretz et al. 2007).

Contrarily, the upregulation of *D-GADD45* in somatic cells causes apoptosis (Peretz et al. 2007). In line with these findings, the stability of *D-GADD45* transcripts was described to be regulated by the nonsense-mediated decay (NMD) pathway, a surveillance mechanism that degrades transcripts, primarily abnormal mRNAs containing premature stop codons (Celik et al. 2015). It was found that the viability of flies mutated for NMD genes was compromised due to an increase in the levels of *D-GADD45*, which led to increased cell death (Nelson et al. 2016). Moreover, the downregulation of *D-GADD45* clearly recovered the viability of NMD-mutant flies (Nelson et al. 2016), indicating that its mRNA levels are tightly regulated by the NMD pathway as higher *D-GADD45* levels induce apoptosis.

The expression of *D-GADD45* is rapidly increased upon cell-death induced damage, but after the initial steps of the stress response, when the tissue still has not completely recovered, the expression of *D-GADD45* returns to the levels observed prior to the damage (**Figure 12**) (Stramer et al. 2008; Blanco et al. 2010; Vizcaya-Molina et al. 2018).

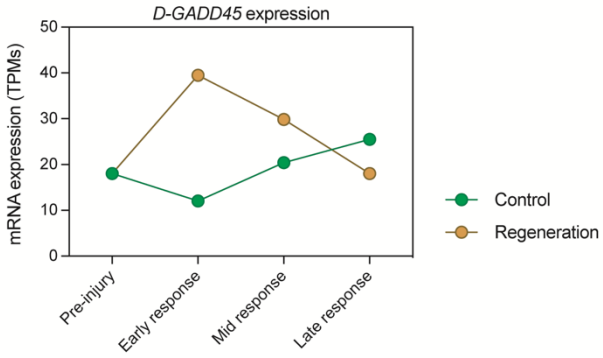


Figure 12. Transient upregulation of *D-GADD45* in regeneration. mRNA expression (in TPMs) of *D-GADD45* in wing imaginal discs previous to the injury, and at three different stages of the recovery process after the induction of cell death by expressing the proapoptotic gene *rpr* for 11 h. Early response corresponds to 0 h after stopping the *rpr* expression, mid response corresponds to 15 h, and late response corresponds to 25 h after cell *rpr* induction. The expression in control non-induced discs is represented in green, induced discs are represented in yellow. Data from Vizcaya-Molina et al. 2018.

This expression pattern is in line with the stress sensor function of GADD45 proteins. Indeed, at the initial steps of regeneration, p38 and JNK are also activated and are essential for wound healing (Bosch et al. 2005; Bergantiños et al. 2010; Santabárbara-Ruiz et al. 2015). For this, it is tempting to speculate about the relationship between the upregulation of *D-GADD45* and the activation of p38 and JNK during the early response to damage.

OBJECTIVES

Despite the growing number of functionally described lncRNAs, only a few of them have been characterized in *Drosophila*. The **main objective** of this thesis is to identify and characterize lncRNAs in the context of wing regeneration. Similarly, the role in regeneration of several genes encoding for stress-induced proteins remains to be elucidated. For these reasons, we propose the following **specific objectives**:

1. Characterize the set of lncRNAs expressed in the wing disc in development and regeneration.
2. Describe the role of duplicated lncRNAs *CR40469* and *CR34335* in development and regeneration.
3. Study the role of the protein-coding gene *D-GADD45* in JNK-dependent apoptosis and regeneration.

MATERIALS AND METHODS

MATERIALS

Drosophila melanogaster lines

The list of fly strains and their resource information is depicted in **Table 2**.

Strain	Source	Code
Canton S	Bloomington Drosophila Stock Centre	64349
lexO-rpr	Santabárbara-Ruiz et al. 2015	NA
$saf^{E/Pv}$-LHG	Santabárbara-Ruiz et al. 2015	NA
tub-$Gal80^{TS}$	Bloomington Drosophila Stock Centre	7017
$CR34335^{-/-}$	Bloomington Drosophila Stock Centre	23825
$CR40469^{-/-}$	Generated in this work	NA
UAS-rpr	Wing et al. 1998	NA
sal-Gal4	Barrio and De Celis 2004	NA
ci-Gal4	Martin and Morata 2006	NA
en-Gal4	Bloomington Drosophila Stock Centre	83351
UAS-GFP	Bloomington Drosophila Stock Centre	4776
UAS-RNAi-Ask1	Bloomington Drosophila Stock Centre	35331
UAS-RNAi-Mekk1	Bloomington Drosophila Stock Centre	35402
UAS-bsk^{DN}	Bloomington Drosophila Stock Centre	9311
UAS-D-GADD45	Bloomington Drosophila Stock Centre	81038
UAS-RNAi-D-GADD45	Bloomington Drosophila Stock Centre	35023
hh-Gal4	Santabárbara-Ruiz et al. 2015	NA

Table 2. Drosophila melanogaster strains. The strain, source and code is shown. NA (Not Associated).

Reagents

A detailed list of the antibodies, dies, enzymes and kits used in this work are shown in **Table 3**.

Reagent	Type	Used for	Source
α-Cubitus interruptus (Ci)	Antibody	Immunohistochemistry	DSHB
α-Matrix metalloproteinase 1 (Mmp-1)	Antibody	Immunohistochemistry	DSHB
α-P-histone-H3 (PH3)	Antibody	Immunohistochemistry	Millipore
α-P-JNK	Antibody	Immunohistochemistry	Promega
Anti-Digoxigenin-POD, Fab fragments	Antibody	RNA probe synthesis	Roche
TO-PRO-3	Dye	DNA staining	Life Technologies
YO-PRO-1	Dye	DNA staining	Invitrogen
T4 DNA Ligase	Enzyme	3C-qPCR	New England Biolabs
Taq DNA Polymerase recombinant	Enzyme	PCR amplification	Invitrogen
FastStart Universal SYBR Green Master (Rox)	Enzyme	qPCR	Roche
DIG RNA Labeling mix	Enzyme	RNA probe synthesis	Roche
KpnI	Enzyme	RNA probe synthesis	Promega
T3 RNA Polymerase	Enzyme	RNA probe synthesis	Promega
T7 RNA Polymerase	Enzyme	RNA probe synthesis	Promega
EcoRI, High-Fidelity	Enzyme	RNA probe synthesis, 3C-qPCR	New England Biolabs
RQ1 RNase-free DNase	Enzyme	RNA purification	Promega
M-MLV Reverse Transcriptase	Enzyme	RT-PCR	Invitrogen
EZNA Cycle Pure kit	Kit	DNA purification	Omega Bio-tek
MinElute Cleanup kit	Kit	DNA purification	Qiagen
MinElute PCR Purification kit	Kit	DNA purification	Qiagen
TSA-Plus TMR	Kit	FISH signal amplification	Perkin Elmer
Quick-RNA Microprep	Kit	RNA purification	Zymo Research
RNA Clean and Concentrator-5	Kit	RNA purification	Zymo Research

Table 3. List of reagents. The type of reagent, the experiments in which it was used and its source is shown.

Oligonucleotides

A detailed list of all oligonucleotides used in this work is shown in **Table 4**.

Region	Experiment	Forward sequence	Reverse sequence
Homology arm 1	CR40469 deletion	GAGGAATTCGAACTGGACATCATGCGTTG	GAGGCGGCCGCTGCTGGCATTATGTTGGTGT
Homology arm 2		GAGACTAGTTGGCAACAACATCTTGAACC	GAGAGATCTTGTAACACGCACAAGGAAGC
CR40469 deleted region	PCR	TTGTGCGCAGAGCATTGACTT	GGAAAATCTCCCAGCACACAAA
CR34335	PCR	CATTGCATGAGACGATGGCG	TTTCCTCTCCCTACTGCACA
CR40469 flanked by RE and T7/T3 promoters	RNA probe synthesis	TAATACGACTCACTATAGGGGGAATT CCCACGTTCCTCACTAATTGTGGCTA	AATTAACCCTCACTAAAGGGGGGTA CCCCGGATTTTTTGCTTGTTTTGTTTGA
sply	qPCR	CTTTCCCGATTCCCGTA	TGACGGGCTTAAGGCAATC
CR40469/CR34335	qPCR	TTGAGTTCCTCCAACACAGCGT	TTGCGCCATCGTCTCATGCAA
CR34335	qPCR	TTGCATGAGACGATGGCGCA	CGGTCGAGTGCCTCACAGTGTA
CR34335 (long isoform)	qPCR	GAATAACCACTCCCCCACGC	TGTACCCGTGAATTGATGAAAATTG
DIP-a	qPCR	TTTCGTTGCCATCTTCGCGGC	AGGGCAGTTCCGTGCGATTGA
DNA control	3C-qPCR	CCTTCGTCGAATCACAAAATTTCG	GGAGTCATCTTCATCAGCATCTCC
apterous	3C-qPCR	GTGAGTTTTACCCTCTGATCTCG	GCTTGCTTATTAACTAGATAAACCGG
CR40469-karl	3C-qPCR	GTGAGGAACGTGTTTTAACTTTGC	AGGCGATCATCCATCCATCC
CR40469-CG2898	3C-qPCR	GTGAGGAACGTGTTTTAACTTTGC	GAACACGACCAGCAGCAAG
CR40469-CG3168	3C-qPCR	GTGAGGAACGTGTTTTAACTTTGC	CGACAGCCAGTCATTTCGC
CR40469-CG32640	3C-qPCR	GTGAGGAACGTGTTTTAACTTTGC	GCAGCGGACAAGCTTGGA

Table 4. List of oligonucleotides. All primers used are shown in the table, including the region to which they hybridize, the experiments in which they were used, and their sequence.

Genome-wide data

A detailed list of all the data used from RNA-seq and ATAC-seq experiments is shown in **Table 5**.

Data produced in this thesis

Technique	Organism	Tissue	Condition	Accession	Reference
RNA-seq	*D. melanogaster*	Wing disc Early	WT Injured	TBA	-
RNA-seq	*D. melanogaster*	Wing disc Early	CR40469-/- Injured	TBA	-

Data used in this thesis

Technique	Organism	Tissue	Condition	Accession	Reference
RNA-seq	*D. melanogaster*	Wing disc Early	Injured and uninjured	GSE102841	Vizcaya-Molina et al. 2018
RNA-seq	*D. melanogaster*	Wing disc Mid	Injured and uninjured	GSE102841	Vizcaya-Molina et al. 2018
RNA-seq	*D. melanogaster*	Wing disc Late	Injured and uninjured	GSE102841	Vizcaya-Molina et al. 2018
ATAC-seq	*D. melanogaster*	Wing disc Early	Injured and uninjured	GSE102841	Vizcaya-Molina et al. 2018
ATAC-seq	*D. melanogaster*	Wing disc Mid	Injured and uninjured	GSE102841	Vizcaya-Molina et al. 2018
ATAC-seq	*D. melanogaster*	Wing disc Late	Injured and uninjured	GSE102841	Vizcaya-Molina et al. 2018
RNA-seq	*D. melanogaster*	Wing disc	L3, Early Pupae, Late Pupae	E-MTAB-7653	Pérez-Lluch et al. 2020
RNA-seq	*D. melanogaster*	Antenna disc	L3, Early Pupae, Late Pupae	E-MTAB-7654	Pérez-Lluch et al. 2020
RNA-seq	*D. melanogaster*	Eye disc	L3, Early Pupae, Late Pupae	E-MTAB-7655	Pérez-Lluch et al. 2020
RNA-seq	*D. melanogaster*	Leg disc	L3, Early Pupae, Late Pupae	E-MTAB-7656	Pérez-Lluch et al. 2020
RNA-seq	*D. melanogaster*	Embryo	0h to 24h	SRA009364	Graveley et al. 2012
RNA-seq	*D. melanogaster*	Whole larvae	L1, L2 and L3	SRA009364	Graveley et al. 2012
RNA-seq	*D. melanogaster*	Whole pupae	Early Pupae to 4-days pupae	SRA009364	Graveley et al. 2012

Table 5. List of genome-wide data. The genome-wide data produced in this work (top list) and the data used from other sources (bottom list) is shown in the table. The technique, organism, tissue and condition of each sample is shown. The accession and reference of the genome-wide samples used in this work are shown.

Software and web-based tools

Every software and web-based tools used in this work is listed in **Table 6**.

Tools	Used for	Reference
Adobe Ilustrator 2017	Figure preparation	RRID:SCR_010279
ClustalOmega	Sequence alignment	RRID:SCR_001591
DESeq2	Differential expression analysis	RRID:SCR_015687
FEELnc	Classification of lncRNAs	Wucher et al. 2017
Fiji	Image analysis	RRID: SCR_002285
Genesis	Heatmap design	Sturn et al. 2002
Grape-nf	Processing of transcriptomic data	https://github.com/guigolab/grape-nf
GraphPad Prism 8	Plot preparation	RRID: SCR_002798
MEME suite (FIMO)	Search for miRNA binding sites	RRID:SCR_001783
mFold	Prediction of RNA secondary structure	RRID:SCR_008543
Microsoft Excel 2016	Data filtering	RRID:SCR_016137
miRBase	miRNA seed sequence identification	RRID:SCR_003152
ORFinder	Search of ORFs	RRID:SCR_016643
Panther Gene Ontology	Gene ontology analysis	RRID:SCR_004869
PROMO	Search for TF motifs	Messeguer et al. 2012
ProtParam (ExPasy)	Scoring of peptide parameters	RRID:SCR_018087
RSEM	Quantification of RNA-seq reads	RRID:SCR_013027
SnapGene	Cloning set up	RRID:SCR_015052
STAR	Alignment of RNA-seq reads	Dobin et al. 2012
UCSC Genome Browser	Data visualization	RRID:SCR_005780

Table 6. Software and web-based tools. Every software and web-based tool used in this work are shown, as well as the process for which they were used and their reference.

METHODS

DROSOPHILA IN VIVO ASSAYS

Embryo to adult viability and pupariation assays

The collection of embryos for the embryo to adult viability and pupariation assays was the following: flies aged 3-5 days were allowed to lay eggs on standard medium plates containing yeast paste for 3 hours (h) at 25ºC. Eggs were incubated 16 h at 25ºC. Then, unhatched embryos were transferred to standard medium vials in four replicates containing exactly 25 embryos each, for a total of 100 embryos per condition.

For the analysis of pupariation, vials were incubated at 25ºC and the number of pupae was counted every 12 h starting from 108 h after egg laying. The pupariation percentage at each timepoint was calculated according to the total number of larvae capable of transitioning to the next developmental phase.

For the embryo to adult viability assay, vials were incubated at 25ºC for 15 days and the number of hatched adults per vial was counted. Embryo to adult viability was calculated as the proportion of adults hatched from deposited embryos. The proportion of hatched females was calculated as the number of females among all hatched adult flies.

Control of the UAS/Gal4 system

For the induction of *UAS* constructs, we employed the appropriate *Gal4* line for each case. For experiments requiring the sustained expression of *UAS* constructs, we kept the vials of flies containing the *UAS* and *Gal4* constructs at 25ºC from early embryos until the dissection of wing discs at L3 (96 h after egg laying - AEL) or until adult flies emerged (around 10-12 days AEL).

For experiments requiring the transient induction of UAS constructs, an additional construct was added: *tub-Gal80^{TS}*. The Gal80 protein acts as a repressor of Gal4 by binding to its transcriptional activation domain, blocking the recruitment of Pol II and preventing the activation of the *UAS* constructs. Particularly, we used a thermosensitive mutant of *Gal80* which is not stable at 29ºC. In this way, we controlled the transient expression of *UAS* constructs by shifting the temperature from 17ºC (Gal80 is active and *UAS* construct is not transcribed) to 29ºC (Gal80 is inactive and *UAS* construct is transcribed) (**Figure 13**).

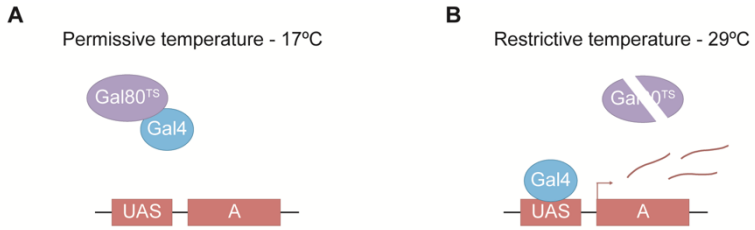


Figure 13. Control of the UAS/Gal4 system. (**A**) Inactivation of the system at 17ºC due to the action of the repressive Gal80, which inhibits the Gal4-driven activation of gene A. (**B**) At 29ºC, the thermo sensitive repressor Gal80 is unstable and the Gal4 can promote the expression of gene A by binding to the *UAS* sequence. The exact same mechanism is used to control the LexA/LexO system by using a Gal80-repressive LexA (LHG), and the *LexO* sequence upstream of the target gene.

Genetic ablation and dual Gal4/LexA transactivation system

Cell death was genetically induced as previously described by activating the expression of the constructs *UAS-rpr* or *lexO-rpr* under the control of *sal$^{E/Pv}$-Gal4* and *sal$^{E/Pv}$-LHG*, respectively. The generation of the *sal$^{E/Pv}$-LHG* construct, which is suppressible by Gal80, was described in Santabárbara-Ruiz et al. 2015. The system was controlled by the thermo sensitive repressor *tub-Gal80^{TS}.*

Embryos were kept at 17ºC for 192 h (8 days) to prevent *rpr* expression. They were subsequently moved to 29ºC for 11 h to allow *rpr* expression, then they were moved back to 17ºC until adulthood. In the experiments requiring the dissection of wing discs, *rpr* induction was prolonged to 16 h and discs were immediately dissected. Controls without *rpr* were always treated in parallel (**Figure 14**).

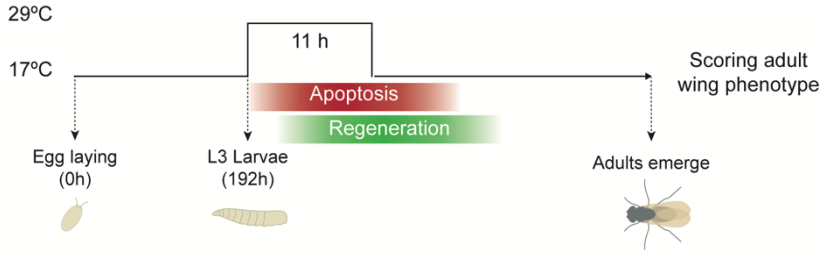


Figure 14. Genetic induction of cell death. The expression of the proapoptotic gene *rpr* is induced 196 h after egg laying (AEL) for 11 h by shifting the temperature from 17ºC to 29ºC. Then, the temperature is shifted back to 17ºC until adults emerge, when the adult wing phenotypes are scored. For experiments requiring the dissection of wing imaginal discs, the induction is prolonged to 16 h and discs are immediately dissected.

In dual transactivation experiments, we used *sal$^{E/Pv}$-LHG* and *lexO-rpr* to ablate the *spalt* domain (*sal$^{E/Pv}$> rpr*), whereas *ci-Gal4* was used to express different transgenes in the anterior compartment.

Analysis of the adult wing phenotypes

Female adult flies were fixed in 1:2 glycerol:ethanol for 24 h. Wings were dissected on water and then washed with ethanol. Then, they were mounted on 6:5 lactic acid:ethanol, and imaged under a microscope. We considered wings as aberrant or non-regenerated when missing complete veins or crossveins and/or when notches were present in the wing blade. Wing size was considered as the area inside the wing blade perimeter. The wild type patterning of an adult wing is shown in **Figure 15**.

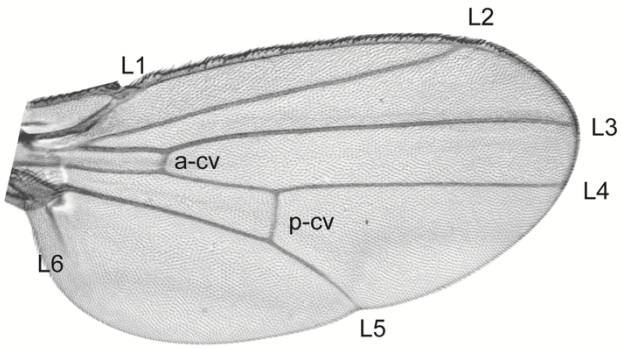


Figure 15. Adult wing wild type patterning. Sample image of a wild type wing showing the 6 longitudinal veins (labelled L1 - L6), the anterior crossvein (a-cv) and the posterior crossvein (p-cv).

Induction of cold and oxidative stresses

The process of embryo collection was identical to the process followed for the embryo to adult viability assay. A total of 6 vials containing 25 embryos each were prepared. Control non-stressed vials were incubated at 25ºC until adulthood. For the induction of cold stress, vials were transferred 96 h AEL from 25ºC to 0ºC (ice bath) for 4 h, and then they were transferred back to 25ºC until adulthood. For the induction of oxidative stress, embryos were transferred to vials containing 20 mM tBH in the fly food and incubated at 25ºC until adulthood.

Embryo to adult survival was calculated as the proportion of adults hatched from deposited embryos.

MOLECULAR TECHNIQUES

Immunostaining of wing imaginal discs

Wing discs were dissected in Schneider's medium (Sigma Aldrich) and fixed at room temperature for 30 min in 4% paraformaldehyde diluted in PBS. Discs were then washed twice in PBS, twice in PBT (0.5% Triton-X-100 in PBS), and incubated 1 h in blocking solution (2% BSA in PBT). Then, samples were incubated overnight at 4ºC with the primary antibody diluted in blocking solution at the appropriate concentration (**Table 7**). Discs were washed in PBT for 20 min 4 times to remove the excess of primary antibody and were incubated with the secondary antibody diluted 1:200 in PBT for 2 h at room temperature. After that, discs were washed in PBT for 10 min 3 times. Finally, nuclei were stained incubating 10 min with either YO-PRO-1 (Life Technologies) diluted 1:10.000 in PBS or TO-PRO-3 (Roche) diluted 1:1.000 in PBS, and mounted in SlowFade (Life Technologies).

Primary antibody	Organism	Dilution
α-Cubitus interruptus (Ci)	Rat	1:25
α-Matrix metalloproteinase 1 (Mmp-1)	Mouse	1:100
α-P-histone-H3 (PH3)	Rabbit	1:1,000
α-P-JNK	Rabbit	1:100

Table 7. Primary antibodies. The list of all primary antibodies used in this work, as well as the organism from which they are produced and their working solution is shown.

To calculate the mitotic index in samples incubated with P-histone-H3 (PH3), we manually counted the number of PH3-positive cells in the anterior and posterior compartment of the wing disc. The mitotic index for each compartment was calculated as the fraction of mitotic cells in that compartment multiplied by 100.

TUNEL assay

For apoptotic cell detection, we used the terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) assay. After fixation of wing imaginal discs, apoptotic cells were detected using the fluorescently labelled Alexa Fluor 647-aha-dUTP (Thermo Fisher Scientific) and incorporated using terminal deoxynucleotidyl transferase (Roche). The mixture was added to the sample following the manufacturer's instructions and it was incubated in the dark for 1h30 at 37ºC. EDTA was added to stop the reaction, the discs were washed and mounted as stated for immunostaining.

To calculate the apoptotic index, we manually counted the number of cells positive for TUNEL in each disc compartment (anterior and posterior). For each compartment, the apoptotic index was calculated as the fraction of apoptotic cells in that compartment multiplied by 100.

Generation of *CR40469* knock-out mutant

The CR40469 knock-out mutant was generated by the group of Aurelio Teleman.

The deletion of *CR40469* was performed using the ends-out homology recombination technique. Long homology arms for the upstream and downstream regions of the

CR40469 gene were designed, and a sequence spanning 1,089 bp containing the entire *CR40469* locus and 663 bp of its promoter was excised and replaced by an *mCherry* cassette (**Figure 16 A**). We validated the absence of the *CR40469* locus by PCR using primers hybridising specifically to the deleted region, and primers hybridising specifically to the *CR34335* gene were used as control (*see* ***Table 4*** *for primer sequences*). We used gDNA from two candidate homozygous mutants and control gDNA from *Canton S* flies and performed the following PCR reaction: 2 min at 95ºC, followed by 35 cycles of 30 sec at 95ºC, 30 sec at 63ºC and 30 sec at 72ºC, and a final elongation step of 10 min at 72ºC. We confirmed the *CR40469* deletion running the PCR products on an agarose gel (**Figure 16 B**).

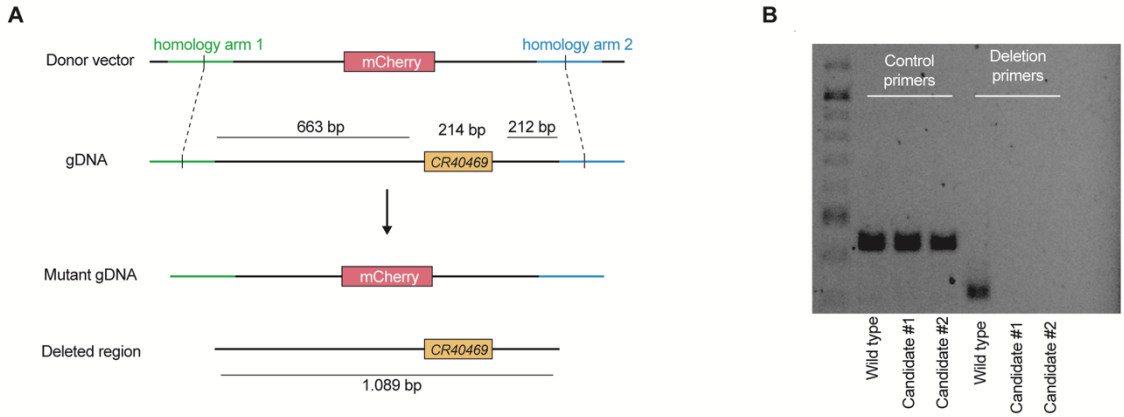


Figure 16. Generation and validation of *CR40469* knock-out mutant. (**A**) Ends-Out homologous recombination scheme of the *CR40469* deleted region. (**B**) PCR validation of the *CR40469* deleted region.

Fluorescent *in situ* hybridization (FISH)

To analyse the localization of *CR40469* and *CR34335* transcripts, we synthesised a digoxigenin-labelled RNA probe. For this, we selected the 214-bp sequence of the *CR40469* transcript, which is 99.1% identical to the *CR34335* transcript, for PCR amplification. As template for probe synthesis, we used gDNA from *Canton S* larvae. Fragments were amplified by PCR using the primer sets reported in **Table 4**. To increase the cloning efficiency, the T3 promoter sequence and an EcoRI target site were added at the 5' end of the forward primer, while the T7 promoter sequence and a KpnI target site were added at the 5' end of the reverse primer. Amplicons were purified using the MinElute PCR purification kit (Qiagen). Then, we digested the fragments using EcoRI (antisense probe) or KpnI (sense probe) and added calf intestinal alkaline phosphatase (Promega) to prevent religation. DIG-labelled probes were prepared using the DIG RNA labelling mixture (Roche) and T7 (antisense probe) or T3 RNA polymerase (sense probe). The probe synthesis mixture was prepared following manufacturer's instructions and was incubated for 2 h at 37ºC, then EDTA was added to stop the reaction. Synthesised probes were purified using RNA Clean and Concentrator-5 kit (Zymo Research). The size of the probes was tested running an agarose gel.

The FISH protocol used in this work is a slightly modified version of the protocol described in Jandura et al. 2017. All solutions were filtered through a 0.22 µm filter prior to use, and PBS was prepared using diethylpyrocarbonate (DEPC)-treated water. Freshly dissected wing imaginal discs were fixed for 30 min in 4% formaldehyde and

0.1% picric acid and washed in PBTT (PBS with 0.1% Tween-20, 0.3% Triton-X-100 and 0.1% picric acid) 3 times for 5 min each. To quench endogenous peroxidase activity, samples were incubated twice for 15 min with 0.3% hydrogen peroxide in PBS and washed with PBTT 4 times for 5 min each. Then, samples were incubated with 80% acetone in PBS for 10 min at -20ºC, washed with PBTT 3 times for 10 min each, rinsed with 1:1 PBTT:hybridization solution (50% formamide, 5X SSC buffer, 0.1% Tween-20, 0.3% Triton-X-100, 0.1 mg/ml heparin and 1% salmon sperm ssRNA in DEPC water). Then, we incubated the samples with tempered pre-boiled hybridization solution for 3 hours at 56ºC. Sense or antisense probe was diluted in hybridization solution (50 ng probe per 100 µl hybridization solution), and the mixture was denatured for 5 min at 85ºC prior to the incubation overnight at 56ºC. Then, probes were removed and samples were washed 15 min at 56ºC with decreasing concentration of hybridization solution:PBTT (3:1 twice, 1:1 and 1:3), followed by 3 washes of 5 min with PBTT. Samples were incubated with blocking solution (1% BSA in PBTT) 20 min at room temperature prior to the incubation with anti-digoxigenin (1:2,000 in blocking solution) for 2 h. Then, anti-digoxigenin was washed with blocking solution 8 times for 10 min each. Samples were treated using the Tyramide Signal Amplification (TSA) system (Perkin Elmer), by incubating them in the dark for 2 h with TSA working solution, and then washed 12 times for 15 min each with PBTT. Finally, samples were incubated 15 min with YO-PRO-1 (Invitrogen, 1:10,000 in PBS), mounted on SlowFade Antifade (Life Technologies) and imaged under a fluorescent microscope.

Extraction and purification of total RNA

For RNA extraction, 40 wing discs per sample were dissected in cold Schneider's medium (Sigma Aldrich) for less than 30 min. To isolate the RNA, the Quick-RNA Microprep kit (Zymo Research) was used following manufacturer's instructions, followed by an incubation with DNAse I (Promega) at 37ºC for 30 min. To further purify the RNA and clean the DNAse, the RNA Clean and Concentrator-5 kit (Zymo Research) was used. Purity and concentration of the RNA samples was assessed using Nanodrop.

Reverse transcription and quantitative PCR

A total of 1 µg of total RNA was used as a template for cDNA synthesis using Moloney Murine Leukaemia Virus reverse transcriptase (M-MLV RT) (Invitrogen), the resulting cDNA was diluted 2.5 times in RNase-free water. 20 µl volume reactions containing FastStart Universal SYBR Green Master (Rox) (Roche) were run in a 7500 Real-Time PCR System (Applied Biosystems). We used the following PCR program: 10 min at 95ºC followed by 40 cycles of 15 sec at 95ºC and 1 min at 60ºC. Melting curves were used to discriminate unspecific amplifications. Samples were normalised to the levels of *sply* and fold changes were calculated using the ddCt method. Three technical replicates were used for each reaction, and three separate biological replicates were collected for each experiment. The designed primer sequences are shown in **Table 4**.

RNA-seq sample collection and RNA isolation

For RNA-seq sample collection, flies were allowed to lay eggs on standard medium vials for 6 h at 17ºC. Embryos were incubated at 17ºC for 192 h (8 days), then vials were transferred to 29ºC for 16 h. L3 larvae from the desired genotypes were selected under

a microscope by the absence of visible markers. 50 wing discs per sample were dissected in cold Schneider's medium (Sigma Aldrich), then Quick-RNA Microprep kit (Zymo Research) was used, followed by DNAse I (Promega) incubation at 37ºC for 30 min, and RNA Clean and Concentrator-5 kit (Zymo Research). The purity and concentration of the resulting RNA was assessed using Nanodrop and Qubit.

RNA-seq library preparation and sequencing

For library preparation, 500 ng of total RNA were used for reverse transcription. Ribosomal RNA (rRNA) was depleted by selecting the poly-A transcripts. All libraries were sequenced on Illumina HiSeq2500 sequencer, using 50-bp paired-end reads. Library preparation and sequencing was performed by the Genomics Unit of the Centre for Genomic Regulation (CRG).

Chromatin conformation capture (3C)

A total of 100 wing discs per condition were dissected and resuspended in Buffer A containing 60 mM KCl, 15 mM NaCl, 4 mM MgCl2, 15 mM HEPES, 0.5% Triton-X-100, 0.5 mM DTT and 1X Protease Inhibitor Cocktail (Roche). Discs were homogenised with 2% formaldehyde in a tenbrock and incubated 10 min rotating at room temperature for crosslinking. Glycine was added for 5 min at a final concentration of 1 M to quench the formaldehyde. Nuclei were pelleted at 4ºC by centrifuging 2 min at 2,000 G, and washed with Buffer A. Nuclei were pelleted again, resuspended in 1X EcoRI buffer with 0.3% SDS, and incubated 30 min at 37ºC followed by 1 h at 65ºC. Then, we added Triton-X-100 to a final concentration of 1.8% and incubated the mixture 1 h at 37ºC to quench the SDS. 100 units of EcoRI were added and digestion was performed rotating overnight at 37ºC. EcoRI was inactivated by heating the reaction for 20 min at 65ºC, then the digested DNA was pelleted and resuspended in 1X T4 DNA Ligase buffer, 1% Triton-X-100 and 0.25 units/µl of T4 DNA Ligase (NEB). Ligation was performed rotating overnight at 16ºC. Then, samples were decroslinked for 4 h at 65ºC with 0.1 µg/µl Proteinase K (Qiagen) and 1% SDS. DNA was purified with standard phenol:chloroform protocol using sodium acetate and acrylamide to precipitate the DNA, which was resuspended in Milli-Q water.

To quantify the ligation products, we performed quantitative PCR and ran an agarose gel of the PCR products. To normalise each sample for the amount of DNA, we designed a primer pair hybridising within the same fragment. As a positive control, we used primers for the *apterous* enhancer, a region that is known to interact in the wing disc (Bieli et al. 2015). The list of primers used is found in **Table 4**.

IN SILICO ANALYSES

The processing, mapping, quality controls and the differential expression analysis using DESeq2 was performed by Dr. Raziel Amador-Rios.

Differential expression analysis of control and regeneration samples

Differential gene expression analyses between control and regeneration samples were performed separately on each time-point. Genes were filtered per time point (early, mid and late), removing all genes with a gene expression < 1 Transcript per Million (TPM). Analyses were run using R version 3.6.2, DESeq2 (Love et al. 2014), and a fold change approach (Vizcaya-Molina et al. 2018). All genes with an absolute fold change > 1.7 in both methods were considered differentially expressed.

LncRNA classification and pairing to protein-coding genes

LncRNA genes were classified with respect to their genome location using the classification module of the FEELnc (Wucher et al. 2017) pipeline. FEELnc received as input the 2,455 annotated lncRNAs from the FlyBase genome annotation r6.29, classifying the lncRNAs in three broad groups: intergenic, genic intronic, and genic exon. The classification was mutually exclusive in the following rank: genic exonic > genic intronic > intergenic.

Intergenic lncRNAs were paired with their closest protein-coding gene (PCG) based on the end-to-end distance. Genic intronic and genic exonic lncRNAs were paired with their overlapping PCG. For lncRNAs overlapping two or more PCGs, only the PCG showing the highest number of overlapping nucleotides was considered.

Developmental transcriptomic profile

Developmental RNA-seq data of *D. melanogaster* was obtained from Pérez-Lluch et al. 2020 and Graveley et al. 2011. Wing, antenna, eye and leg imaginal discs at 3 distinct developmental stages (third instar larvae - L3, early pupae - EP, and late pupae - LP), as well as 21 samples from whole individuals at the embryo (12 total samples divided in 2 h intervals from 0 h to 24 h), larval (3 total samples), and pupal (6 total samples) stages were analysed. The mean of 3 biological replicates was used for each stage. The developmental heatmap was based on gene expression (log_{10} transformation of TPMs plus a pseudo-count of 0.1) using GENESIS (Sturn et al. 2002).

Analysis of promoter regions

For the analysis of the *CR40469* and *CR34335* promoters, we defined the core promoter (CP) as the region located ± 100 bp from the transcription start site (TSS), the proximal promoter (PP) as the region ± 2 kb around the CP, and the distal promoter (DP) as the region 10 kb upstream of the PP. We used the chromatin peaks found in control and regenerating ATAC-seq data from Vizcaya-Molina et al. 2018. The sequence of all peaks found within the promoter (CP+PP+DP) of both genes was retrieved using the UCSC Genome Browser, and the enrichment for transcription factor (TF) motifs was addressed using the PROMO tool (Messeguer et al. 2002).

Analysis of miRNA target sites

For the analysis of miRNA target sites within the *CR40469* and *CR34335* RNA sequences we retrieved the seed sequences of the 161 mature miRNAs annotated in *Drosophila*. Then, using the Find Individual Motif Occurrences (FIMO) tool from the MEME suite, we got the number of miRNA binding sites.

Prediction of RNA secondary structure

To predict the secondary structures of *CR40469* and *CR34335* transcripts, we used the mFold server (Zuker et al. 2003) and the default parameters.

Identification of smORFs and analysis of peptide properties

To identify putative small ORFs within the sequence of CR40469 and CR34335, we used ORFinder. We analysed the whole sequence of each gene, and only considered the predicted ORFs containing both a start and stop codon, spanning at least 10 amino acids, and whose directionality matched with the transcription sense of each gene. To analyse the properties of the smORF predicted peptide, we used the ProtParam tool from the ExPasy resource portal.

Mapping and assembling pipeline of *CR40469*$^{-/-}$ RNA-seq

Transcriptomic data was processed using grape-nf (github.com/guigolab/grape-nf). Reads were aligned to the fly genome (dm6) using STAR (Dobin et al. 2013) v2.4.0j with up to 4 mismatches per paired alignment, using the FlyBase genome annotation r6.29. Only alignments for reads mapping to ten or fewer loci were reported. Genes and transcripts were quantified in TPMs using RSEM (Li et al. 2011) v1.2.21. The GTF version r6.29 contains a total of 16,412 genes; 13,957 protein coding genes (PCGs) and 2,455 long noncoding RNAs (lncRNAs). In our study the lncRNAs were defined as genes > 200 bp and aligned to canonical chromosomes.

Quality control of alignment sequencing data was performed using QualiMap (García-Alcalde et al. 2012) v.2.2.1 and Picard v.2.6.0 (http://broadinstitute.github.io/picard). Using QualiMap we obtained: number of reads, number of mapped reads, duplication rate, and GC percentage; and using Picard we obtained: dropout, and GC dropout. Assessment of replicates reliability was measured with weighted correlation network analysis (WGCNA). WGCNA was implemented with the R package WGCNA (Langfelder and Horvath, 2008) version 1.69. A cut-off of less than 2 standard deviations from a normal distribution was implemented to use a replicate.

Filtering and differential expression analysis of *CR40469*$^{-/-}$ RNA-seq

We used the statistical methods implemented in the DESeq2 (Love et al. 2014) package version 1.26.0. Only genes expressed at least 1 TPM in at least one sample were selected for this analysis. The two factors with interaction approach was implemented, considering the following design matrix: genotype + condition + genotype:condition. Where genotype is wild type or *CR40469*$^{-/-}$ and condition is regeneration or control. All genes with an absolute fold change > 1.7 and a Benjamini-Hochberg adjusted p-value < 0.05 were considered differentially expressed.

Image processing and analysis

Bright field images were taken using Leica DMLB optic microscope, while fluorescent images were taken using Leica SPE confocal microscope. All images were processed using Fiji and Adobe Illustrator software.

Statistics

Prior to any statistical analysis, the Shapiro-Wilk normality test was used to check the distribution of the data. According to the result, the subsequent tests were parametric or non-parametric.

For comparisons of 2 groups, we used either the t-test or Mann-Whitney U test depending on the results of the previous Shapiro-Wilk test. For comparisons of > 2 groups, we used either a one-way ANOVA or Kruskal-Wallis test, again depending on the Shapiro-Wilk test results, followed by Dunn's multiple comparison test. For comparisons of > 2 groups considering 2 factors, we used a 2-way ANOVA followed by Dunnett's multiple comparison test.

All statistical tests were two-tailed. Differences were considered significant when p-values were less than 0.05 (*), 0.01 (**), or 0.001 (***). Every test was calculated using GraphPad Prism 8 software.

RESULTS

CHAPTER 1:

The role of lncRNAs during wing disc development and regeneration

Considering that lncRNAs are generally expressed with high tissue- and stage-specificity, we interrogated currently available transcriptomic data of *Drosophila* imaginal discs (Pérez-Lluch et al. 2020). Then, we analyzed the expression of lncRNAs in RNA-seq data obtained after the induction of cell death in wing imaginal discs (Vizcaya-Molina et al. 2018). Finally, we performed functional experiments to determine the contribution of selected lncRNAs to development and regeneration.

LncRNAs expressed in the wing disc

Currently, there are 2.545 lncRNAs annotated in the *Drosophila* genome (FlyBase r6.45), but only a handful of them have been functionally characterized. With the aim to get insight into the hypothetical functions of lncRNAs during the development of the wing disc, we analyzed their expression patterns, as well as their putative contribution in the regulation of the expression of their nearby genes.

Set of lncRNAs expressed in the wing disc

We used available transcriptomic data from RNA-seq samples of wing discs at three different developmental stages: third instar larvae (L3), early pupae (EP) and late pupae (LP) (Pérez-Lluch et al. 2020). We found a total of 203 lncRNAs expressed in L3 wing discs (**Figure 17 A**), corresponding to merely 8.3% of all annotated lncRNAs in *Drosophila*. In contrast, up to 7.516 protein-coding genes (PCGs) were expressed at the same stage (**Figure 17 A**), accounting for 53.9% of all annotated PCGs. Different trends between PCGs and lncRNAs were also observed over time, as the number of expressed PCGs increased, topping at 8.208 genes at LP stage, opposed to the number of expressed lncRNAs, barely changing from 203 at L3 to 207 at LP (**Figure 17 A**).

Although the total number of expressed lncRNAs in the wing disc across development was almost identical, we checked whether the set of expressed lncRNAs was the same or if it changed from one stage to the other. We found that only 35.65% of the expressed lncRNAs were common in L3, EP and LP, as opposed to the 76.09% for PCGs (**Figure 17 B**). Focusing on the stage-specific lncRNAs, 9.15% were only expressed at L3, 7.89% were only expressed at EP, and up to 23.66% were only expressed at LP (**Figure 17 B**). In accordance with the higher stage-specificity of lncRNAs (Djebali et al. 2012; Derrien et al. 2012), these numbers were remarkably lower for PCGs: 1.36% L3-specific, 1.82% EP-specific, and 11.58% LP-specific (**Figure 17 B**).

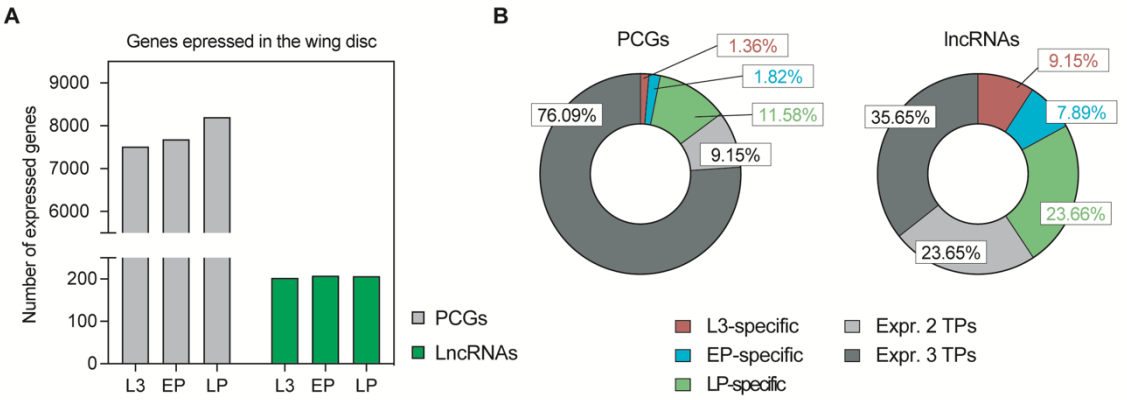


Figure 17. Protein-coding genes and lncRNAs expressed in the wing disc. (**A**) Number of protein coding genes (PCGs) and lncRNAs expressed at L3, EP and LP stages in the wing disc. (**B**) Percentage of stage-specific PCGs and lncRNAs expressed in the wing disc.

Stage-specificity of lncRNAs is maintained for different tissues

We next performed similar analysis in RNA-seq samples from antenna, eye and leg discs at L3, EP and LP stages to determine whether an increased percentage of lncRNAs was also found in other imaginal discs at the LP stage.

The total number of expressed PCGs and lncRNAs in every analyzed tissue was similar to the number of expressed genes in the wing disc (**Figure 18 A**). Regarding the stage-specificity, the same tendencies observed for wing discs were observed for antenna, eye and leg discs: (1) lncRNAs showed a higher stage-specificity than PCGs and (2) there was an increase in the proportion of lncRNAs and PCGs expressed only at the LP stage (**Figure 18 B**). Actually, the percentage of L3 and EP-specific lncRNAs was below 20% in any of the analyzed disc types, while this percentage increased to ~40% at LP (**Figure 18 B**).

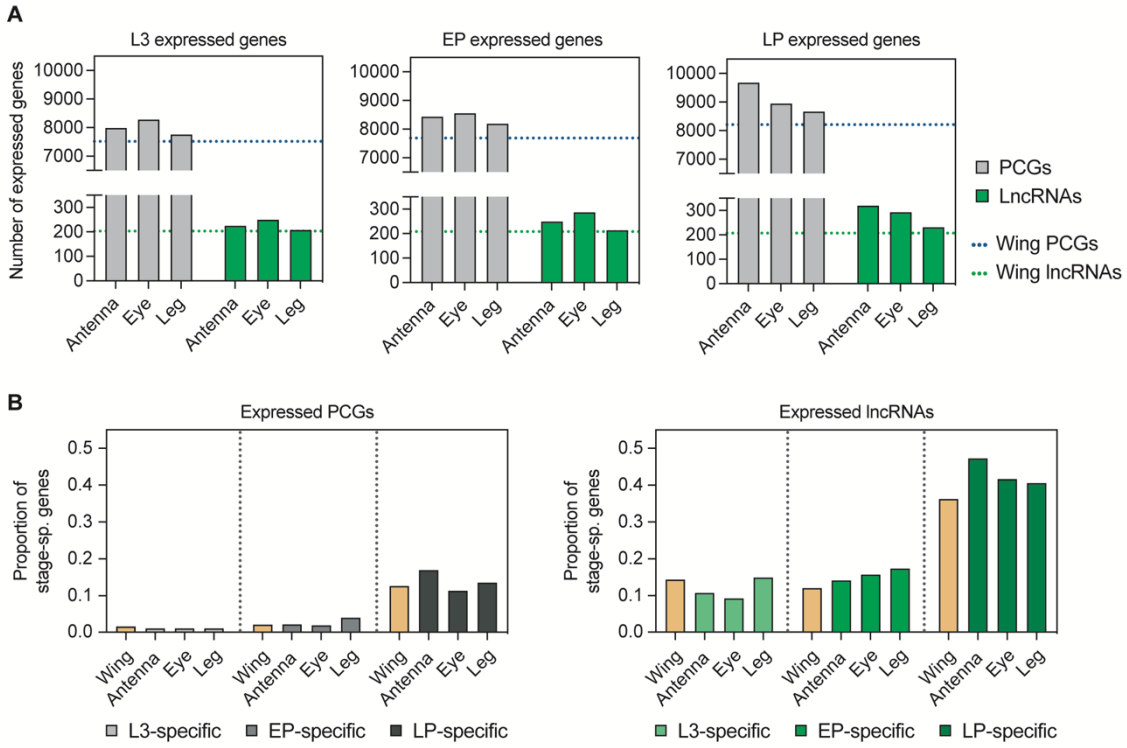


Figure 18. Number and stage-specific genes expressed in different imaginal disc types. (**A**) Total number of PCGs (grey) and lncRNAs (green) expressed in each tissue at L3, EP and LP. Dotted line indicates the number of expressed PCGs (blue) and lncRNAs (green) in the wing disc. (**B**) Comparison of the stage-specificity of PCGs and lncRNAs expressed in each disc type and stage.

Altogether, these findings indicate that the increased stage-specificity of lncRNAs and PCGs at LP was not a specific feature of the wing disc, as it was also maintained for antenna, eye and leg discs. Thus, most probably it could be related with the dramatic morphological changes that imaginal discs suffer during the metamorphosis occurring at the pupal stage.

Wing expression correlation between different developmental stages

To get more insight into the molecular changes taking place in the wing disc across development, we analyzed the differences in gene expression between L3, EP and LP stages. We found minor changes in the expression profile of lncRNAs between L3 and EP (**Figure 19 A**), as we found a relatively high correlation in their expression (r^2=0.68). On the contrary, when we compared the expression profiles between EP and LP (**Figure 19 B**), or between L3 and LP (**Figure 19 C**), we obtained poor correlations (r^2≈0.10 in both cases). These results were also observed for PCGs (**Figure 19 A'-C'**), suggesting that gene expression changes occurring during wing disc development affect both, lncRNAs and PCGs, similarly.

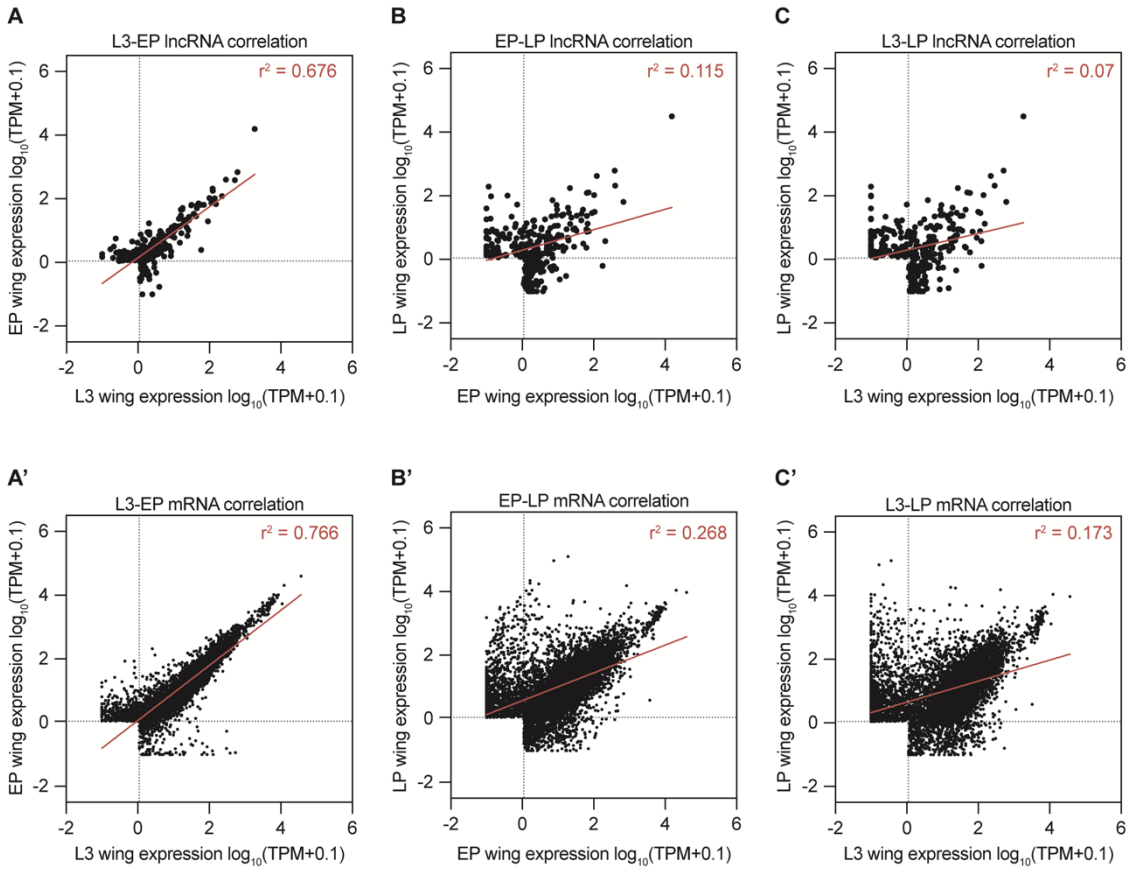


Figure 19. Gene expression correlation between different developmental stages in the wing disc. Expression correlation of (**A-C**) lncRNAs and (**A'-C'**) protein-coding genes between (**A,A'**) L3 and EP, (**B,B'**) EP and LP, or (**C,C'**) L3 and LP. r^2 represents the coefficient of determination of the linear regression for each comparison.

These findings were in line with our results on stage-specificity, as the number of lncRNAs and PCGs expressed only in LP wing discs was higher than that of L3 and EP (**Figure 18 B**).

Effects of wing-expressed lncRNAs in the expression of nearby genes

One of the most extensively used methods to predict gene function *in silico* is the analysis of their ontology. Gene ontology (GO) analysis requires the association of genes with ontology terms, however, most lncRNAs in flies remain uncharacterized and they lack an associated ontology term. For this reason, we analyzed their putative role based on the functions of their neighbouring PCGs. Multiple lncRNAs have been described to regulate the expression of overlapping or nearby genes. For instance, in flies, the lncRNA *CR44811* was recently described to regulate the isoform usage of the overlapping coding gene *blistered* (*bs*), as the expression of *CR44811* reduced the presence of *bs* long isoform in favour of a shorter transcript, with implications in the vein patterning of the wing (Pérez-Lluch et al. 2020).

To test the putative functions of wing lncRNAs in their nearby genes, we paired each expressed lncRNA with their closest PCG neighbour, including overlapping genes. In this way, intergenic lncRNAs were paired with their closest neighbour, while genic lncRNAs were paired with their overlapping PCG. We next categorized their expression patterns depending on their changes from L3 to EP and from EP to LP. Thereby, we

ended up with 4 possible subgroups: increasing, decreasing, peak and valley (**Figure 20 A**). The same categorization was applied for their associated PCGs. Depending on their categories, we classified each lncRNA-PCG pair into concordant pairs if their categories matched, into discordant pairs if their categories were opposed, and into unrelated pairs if their categories were neither concordant nor discordant. The rationale behind this classification was based on two opposed regulation mechanisms exerted by lncRNAs on the expression of their associated PCGs: (1) an activator effect in which the expression of the lncRNA promotes the expression of the PCG (concordant pairs), or (2) an inhibitory effect in which the expression of the lncRNA interferes with the expression of the PCG (discordant pairs).

As expected, we found a higher correlation in the expression of lncRNA-PCG pairs among genic lncRNAs compared to intergenic (**Figure 20 B**). Indeed, almost half of the genic pairs had concordant or discordant expression patterns (**Figure 20 C-D**), which could indicate a possible role of these lncRNAs in the regulation of their overlapping PCGs. However, these results could also reflect the sharing of the same regulatory elements. With the aim to get insight into the processes these lncRNAs could be involved in, we represented the putative functions of their PCG pairs by analysing their GO. Initially, we analyzed individually the sets of concordant and discordant PCGs for each expression subgroup (increasing, decreasing, peak and valley). However, most probably due to the reduced number of genes in each category, no GO term enrichment was found for any of the groups analyzed. To solve this issue, we performed the GO analysis on two main groups: concordant (80 genes) and discordant (50 genes) PCGs. Thereby, for concordant PCGs, we found an enrichment in GO terms related with wing development, such as wing disc morphogenesis, pattern specification or cell differentiation (**Figure 20 E**). We also found an enrichment in genes involved in eye morphogenesis, which may indicate a degree of overlapping in the genes expressed in the wing disc with those expressed in the eye disc. On the other hand, no significant GO term enrichment was found for discordant PCGs, which could be related with the smaller number of discordant pairs compared to the number of concordant.

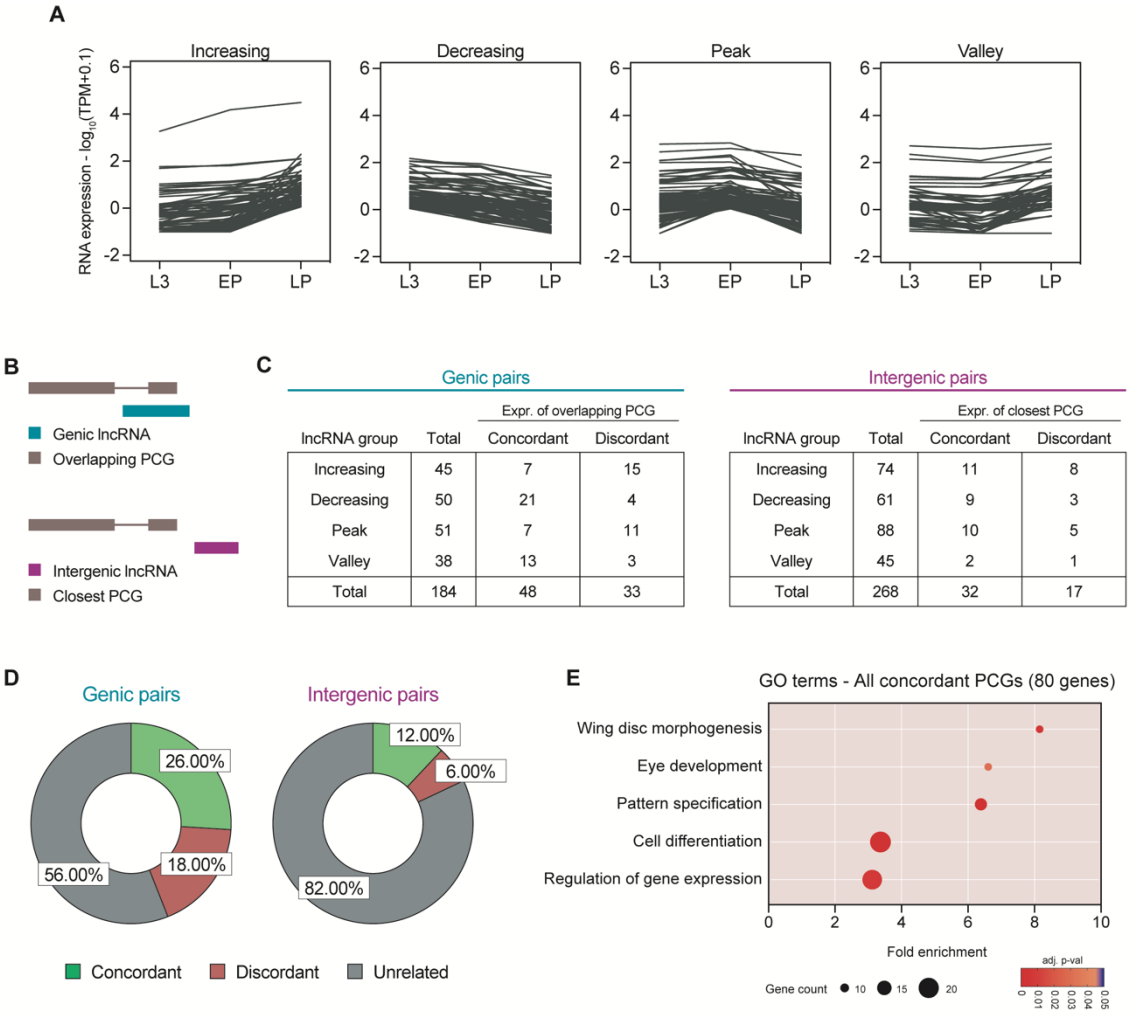


Figure 20. Putative effects of wing-expressed lncRNAs in the expression of nearby genes. (**A**) Expression plots of lncRNAs categorized depending on their L3 to EP and EP to LP patterns. (**B**) Scheme of genic and intergenic lncRNAs and their overlapping and closest PCG, respectively. (**C**) Tables representing the total number of genic and intergenic lncRNAs of each group (increasing, decreasing, peak and valley), as well as the number of concordant and discordant lncRNA-PCG pairs. (**D**) Percentage of concordant, discordant and unrelated lncRNA-PCG pairs. (**E**) Gene Ontology enrichment of the 80 concordant pairs (including 48 genic and 32 intergenic).

Tissue-specificity of wing disc lncRNAs

Apart from their stage-specificity, the expression of lncRNAs is also characterized by their generally high tissue-specificity (Katayama et al. 2005; Derrien et al. 2012; Brown et al. 2014; de Goede et al. 2021). To better characterize the lncRNAs expressed in developmental tissues, we analyzed their expression in the antenna, eye and leg imaginal discs. We found that the majority of lncRNAs expressed in the wing were also expressed in all the other tissues (72-75%) regardless of the developmental stage (**Figure 21 A**). On the other hand, 10% of lncRNAs at L3, 4% at EP, and 8% at LP were only expressed in the wing, thus we called them wing-specific (**Figure 21 A**). Even though the percentage of wing-specific lncRNAs seemed quite low, it was even lower for PCGs, as roughly 1% of wing-expressed PCGs were not expressed in antenna, eye nor leg discs (**Figure 21 A**). These results indicate strong similarities in the pool of expressed PCGs and lncRNAs between the different imaginal disc types.

Next, we associated each wing-specific lncRNA to their overlapping or closest PCG (40 pairs). Applying on these pairs the same categorization applied in **Figure 20 A**, we obtained 22.5% of concordant, and 15% of discordant pairs (**Figure 21 B**). As the total number of gene pairs was too low, their GO analysis did not reveal any enriched ontology term. Nevertheless, through manual curation of these PCGs, we found two wing transcription factors (TFs), *rotund* (*rn*) and *caupolican* (*caup*). While the expression pattern of *rn* through development was concordant to that of the intergenic lncRNA *CR44334* (located 2.7 kb upstream), *caup* and the intergenic lncRNA *CR43431* (located 27.1 kb downstream) showed opposite expression patterns (**Figure 21 C**). These results may indicate that wing-specific lncRNAs could regulate the expression of key TFs such as *rn* and *caup*, which are involved in cell fate specification and pattern formation in the developing wing disc (Gómez-Skarmeta and Modolell, 1996; St Pierre et al. 2002).

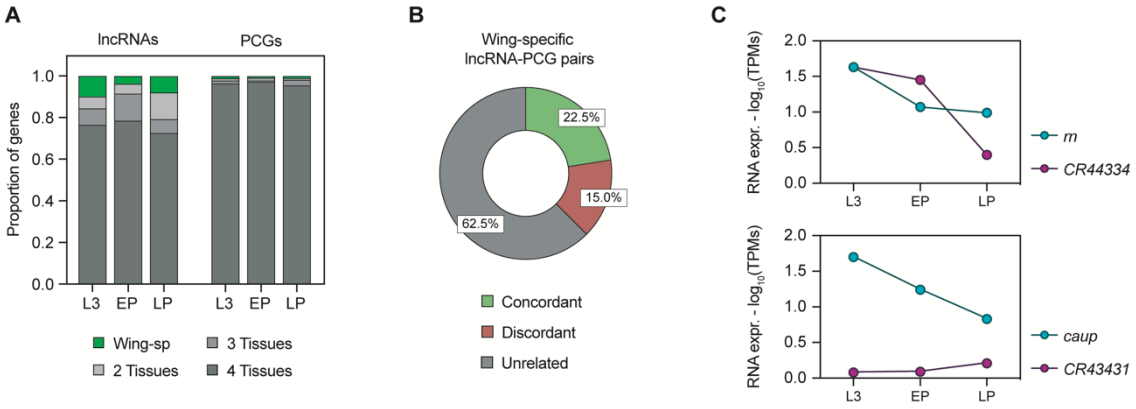


Figure 21. Tissue-specificity of wing disc lncRNAs. (**A**) LncRNAs and PCGs expressed only in the wing disc (wing-sp; green) or expressed also in other tissues (antenna, eye and leg discs). (**B**) LncRNA-PCG pairs for all wing-specific lncRNAs. Percentage of pairs with concordant, discordant or unrelated expression pattern is shown. (**C**) Example of concordant (top plot) and discordant (bottom plot) lncRNA-PCG pairs.

LncRNAs differentially-expressed in regeneration

LncRNA expression changes occurring in regeneration

To determine the contribution of lncRNAs during the regenerative response occurring in the wing disc upon damage, we used previously published RNA-seq data from our laboratory (Vizcaya-Molina et al. 2018). We studied the transcriptome of control and regenerating discs after the induction of the proapoptotic gene *rpr* (White et al. 1996). Data from three different time-points was used: early regeneration (0 h after stopping *rpr* induction), mid regeneration (15 h after stopping *rpr* induction), and late regeneration (25 h after stopping *rpr* induction) (**Figure 22**). These time-points recapitulate three stages of the regenerative process: at the early stage apoptotic cells are extruded to the basal part of the wing disc and the patterning is clearly disrupted, while mitotic cells are already present at the wound edges and some of the early signals triggering regeneration are activated (Smith-Bolton et al. 2009; Bergantiños et al. 2010; Repiso et al. 2013; Santabárbara-Ruiz et al. 2015); at the mid stage the wound has almost been closed but the patterning is still disrupted; at the late stage the wound is completely closed and the size and patterning are mostly recovered (Vizcaya-Molina et al. 2018).

In the original publication in which these RNA-seq datasets were analyzed, the differential expression analysis was based on applying a 1.7-fold change between control and regenerating samples for each time-point (Vizcaya-Molina et al. 2018). As we focused on lncRNAs and their expression is generally very low, we applied a more restrictive approach. To get a robust set of lncRNAs differentially-expressed (DE) in regeneration, we applied three different constraints: (1) a minimum expression of 1 TPM, (2) a minimum fold change of 1.7 between control and regeneration samples, and (3) an adjusted p-value of less than 0.05 after DESeq2 analysis (**Figure 22**).

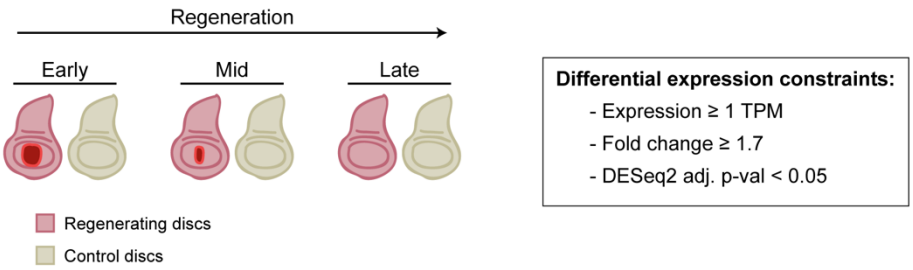


Figure 22. Scheme of the RNA-seq samples of control and regenerating discs. Three different stages of regeneration were compared: early, mid and late. Pairwise comparisons of control and regenerating discs for each stage were performed. We set 3 different constraints for differential expression: expression of at least 1 TPM, absolute fold change of at least 1.7 and a DESeq2 adj. p-val < 0.05.

Applying these parameters, we obtained an initial list of 201 DE lncRNAs in regeneration. However, the RNA-seq method used in the previous study did not provide information of the genomic strand, and we found that many of the DE lncRNAs were located overlapping exons of protein-coding genes (exonic lncRNAs; 109 genes, 54.2%). Thus, there was a high probability that the expression of some exonic lncRNAs was misattributed with reads corresponding to the overlapping protein-coding genes. For this reason, we used the UCSC Genome Browser to manually validate the expression of the 109 exonic lncRNAs using the RNA-seq bigwig tracks (**Figure 23 A-B**). After manual curation, we removed 70 lncRNAs, ending up with a robust list of 131 DE lncRNAs in regeneration (see ***Table 9*** *in Annex 2*), corresponding to 5.3% of the total annotated lncRNAs.

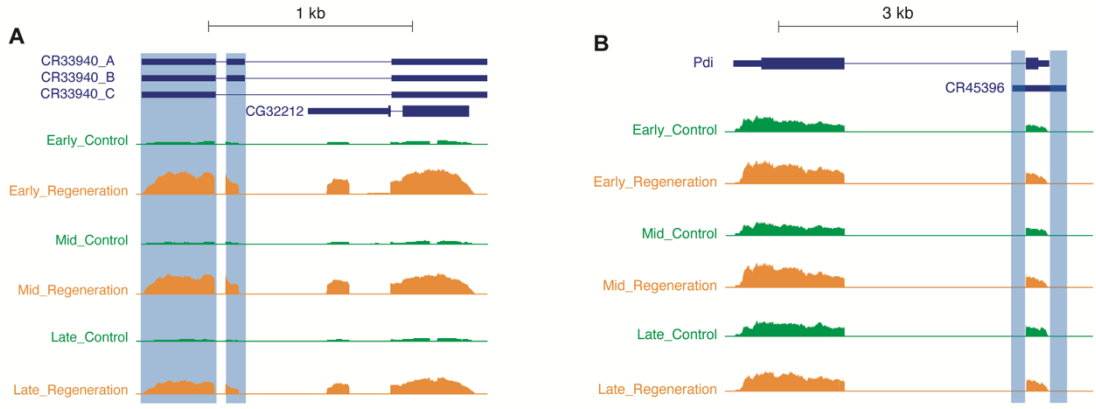


Figure 23. Screenshot of exonic lncRNAs. (**A**) Example of manually-validated exonic lncRNA: *CR33940*. (**B**) Example of manually-excluded exonic lncRNA: *CR45396*. Blue highlight corresponds to the lncRNA-exclusive reads used for curation.

We found a similar number of DE lncRNAs at the early stage (29 upregulated and 33 downregulated), but more downregulation was observed at the mid (11 upregulated and 56 downregulated) and late (22 upregulated and 30 downregulated) stages (**Figure 24**

A), indicating that transcription of lncRNAs during regeneration is generally inhibited, especially after the early stage. On the contrary, expression of PCGs is generally activated upon damage, mostly at the early time point (Vizcaya-Molina et al. 2018). Moreover, most lncRNAs were only DE in one particular time-point (94 genes, 71.8%), while very few (13 genes, 9.9%) were either upregulated or downregulated throughout the regeneration process (**Figure 24 B-C**). Similar numbers were reported for PCGs, as 78.5% were only DE in one particular time-point, while 3.44% were always DE in regeneration (Vizcaya-Molina et al. 2018). It is also worth mentioning that only 7 lncRNAs (5.3%), changed from upregulated to downregulated or vice versa.

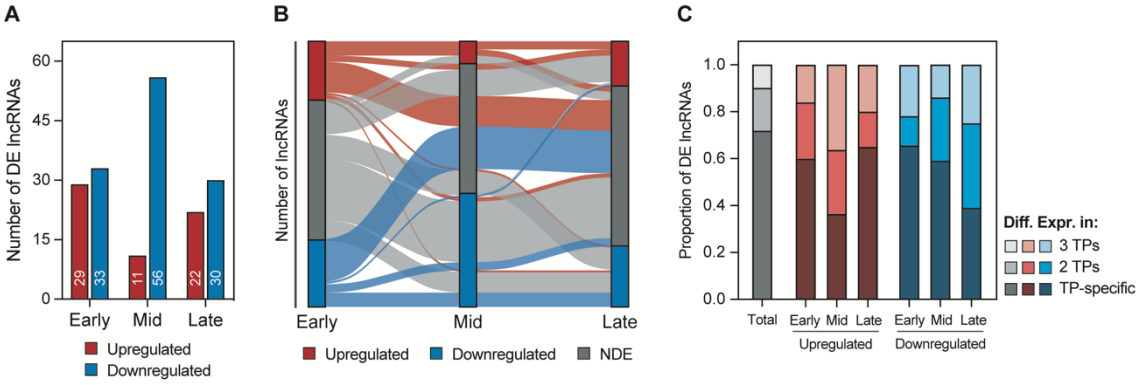


Figure 24. Set of differentially-expressed (DE) lncRNAs in regeneration. (**A**) Total number of upregulated and downregulated lncRNAs at each regeneration stage. (**B**) Alluvial plot of DE lncRNAs showing their behaviour across time. Upregulated and downregulated lncRNAs at the early stage are colored in red and blue, respectively. (**C**) Proportion of lncRNAs upregulated (red) or downregulated (blue) at the 3 different timepoints (TP), at 2 TPs or only at one TP (TP-specific).

Regeneration-specific lncRNAs

To address the possibility that some lncRNAs upregulated in regeneration would have never been expressed in the wing disc under normal conditions, we analyzed the expression of lncRNAs upregulated in at least one regeneration time-point (48 genes) in developing wing discs at L3, EP and LP stages. We identified a total of 14 upregulated lncRNAs with no expression detected during normal wing disc development (**Figure 25 A**; *see* ***Table 9*** *in Annex 2*). Among them, 10 were upregulated at the early time-point, 6 at the mid and 4 at the late (**Figure 25 B**). We defined these genes as regeneration-specific, as their expression in the wing disc was only activated upon damage. For instance, the regeneration-specific *CR43611* was upregulated at the three regeneration time-points, but not found expressed in the wing disc at L3, EP or LP (**Figure 25 C**). On the other hand, 15 upregulated lncRNAs in regeneration were always expressed during wing disc development (**Figure 25 A**).

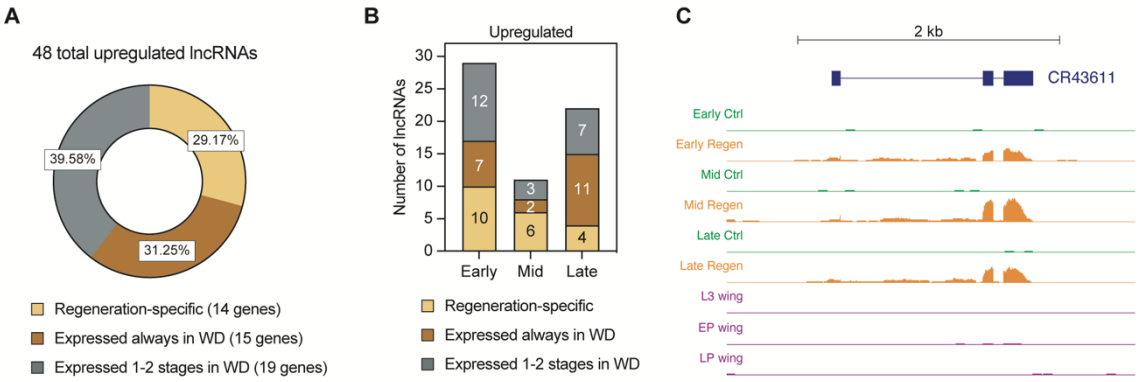


Figure 25. Regeneration-specific lncRNAs. (**A**) Proportion of total upregulated lncRNAs that are only expressed in the wing disc (WD) upon regeneration (regeneration-specific), that are always expressed in the WD, or that are expressed at 1-2 developmental stages in the WD. (**B**) Number of lncRNAs upregulated at early, mid or late regeneration that are regeneration-specific, that are always expressed in the WD, or that are expressed at 1 or 2 developmental stages in the WD. (**C**) Screenshot of the RNA-seq tracks of regeneration-specific lncRNA *CR43611*.

Thus, although the expression of lncRNAs was mainly inhibited during regeneration (**Figure 24 A**), we managed to identify at least 14 regeneration-specific lncRNAs, corresponding to 29.17% of all upregulated lncRNAs in regeneration.

Expression of lncRNAs DE in regeneration in other tissues and stages

Previously, we showed that most lncRNAs expressed in the wing disc were also expressed in other imaginal discs, and that only 5-10% of expressed lncRNAs were wing-specific (**Figure 21 A**). To elucidate the tissue-specificity of lncRNAs found differentially expressed in regeneration, we analyzed their expression in the wing, antenna, eye and leg discs at different developmental stages (L3, EP and LP). Of the 131 DE lncRNAs, 34 (25.9%) were not expressed in any of the analyzed tissues at L3, while 60 (45.8%) were found expressed in all tissues at L3 (**Figure 26 A**). Similar proportions were also observed at EP and LP stages (**Figure 26 A**), indicating that the developmental stage was not a factor in their tissue-specificity. Also, 21 genes (16%) were not detected in any tissue, while up to 33 genes (25.2%) were expressed in all tissues at the three developmental time-points (**Figure 26 B**). Among the DE lncRNAs whose expression has not been detected in imaginal discs, 1 was found upregulated in regeneration at all time-points (*CR46123*) (**Figure 26 C**), 5 were upregulated at the early stage, and 3 were upregulated at mid. Thus, 9 lncRNAs upregulated in regeneration were never expressed during the normal development of the wing, antenna, eye or leg disc. Additionally, the regeneration RNA-seq tracks from *CR46123*, transcribed from the negative strand, show a misannotation of its TSS, as we found a significant number of reads mapping up to 500 nt downstream from its TSS (**Figure 26 C**).

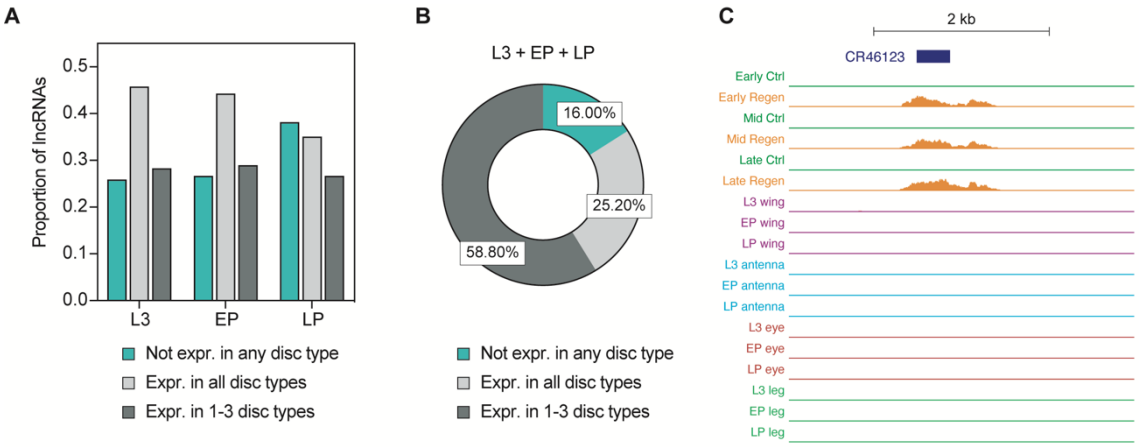


Figure 26. Tissue-specificity of lncRNAs DE in regeneration. (**A**) Proportion of DE lncRNAs not expressed in wing, antenna, eye and leg discs (turquoise), expressed in all disc types (grey), or only expressed in 1-3 disc types (dark grey). Proportion of lncRNAs shown at L3, EP and LP. (**B**) Proportion of DE lncRNAs never expressed in any disc type, expressed always in all disc types, or expressed in 1-3 disc types. (**C**) Screenshot of regeneration upregulated lncRNA *CR46123*, which is not expressed in any disc type at L3, EP nor LP.

Next, we analyzed the expression levels of DE lncRNAs in regeneration in other developmental stages. For this, we used transcriptomic data from whole animals at embryo, larval and pupal stages. Our findings showed that DE lncRNAs could be divided into two groups: (1) genes generally expressed from embryo to pupae and in every imaginal disc, and (2) genes generally non-expressed or expressed at low levels in all samples (**Figure 27**). The first group was composed of a small proportion of the 131 lncRNAs DE in regeneration (36 genes, 26% of total). Indeed, most of these lncRNAs were downregulated in regeneration (27 genes), while only 9 were found upregulated in at least one regeneration time-point (**Figure 27**). In contrast, the majority of lncRNAs DE in regeneration were generally non-expressed or expressed at low levels regardless of the tissue and developmental stage, which was in line with the predominantly low expression of lncRNAs. In this second group, 39 lncRNAs were upregulated and 62 downregulated in at least one time-point in regeneration (**Figure 27**).

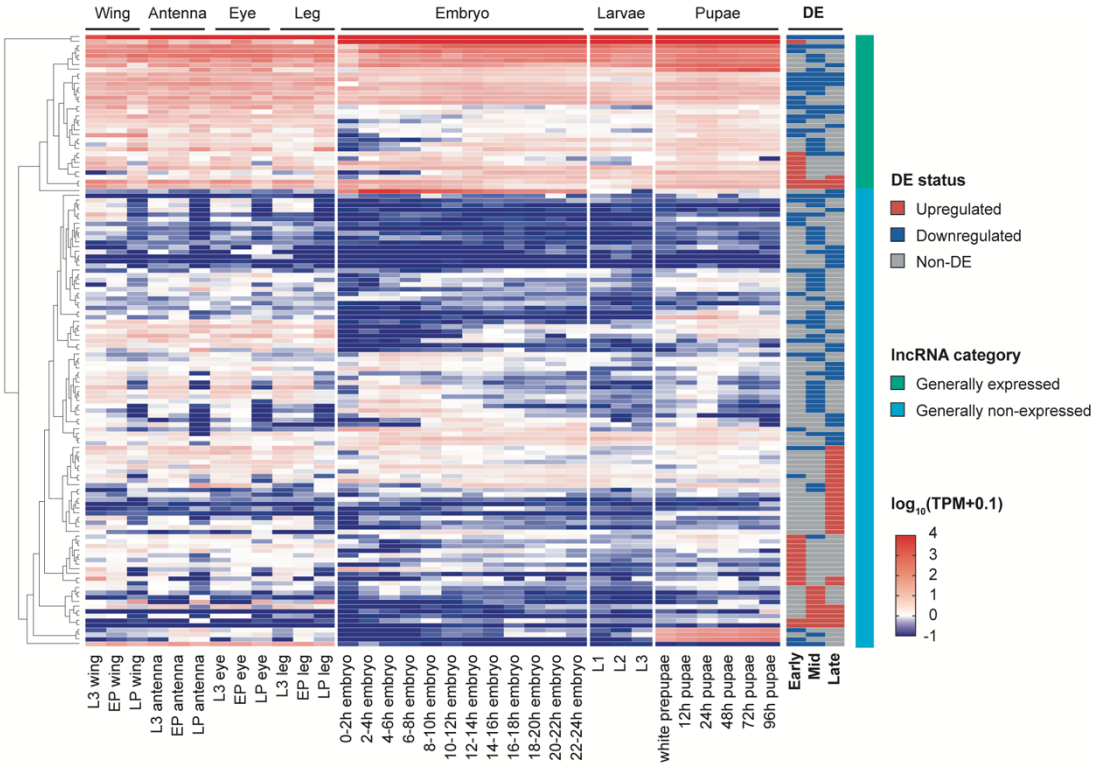


Figure 27. Heatmap of DE lncRNAs in regeneration. Heatmap representating the 131 DE lncRNAs in regeneration. Their expression is represented for RNA-seq samples from wing, antenna, eye and leg imaginal discs, as well as for whole animals from 12 embryo stages, 3 larval stages and 6 pupal stages. Their upregulated (red), downregulated (blue) or non-differential expressed (grey) status is shown in the panel on the right. LncRNAs are categorized into generally expressed (green) and generally non-expressed (cyan).

Features of lncRNAs DE in regeneration

To better characterize the set of lncRNAs DE in regeneration, we classified them depending on their genomic position relative to their overlapping or closest neighbouring protein-coding gene. LncRNAs overlapping exonic sequences from PCGs were classified as exonic, lncRNAs placed within introns of PCGs were classified as intronic, while lncRNAs not overlapping any PCG were classified as intergenic. Using this criterion, we first classified the annotated lncRNAs, and found that 26.2% of them were exonic, 17.3% were intronic, and 56.5% were intergenic. These proportions were generally maintained for DE lncRNAs in regeneration, as 29.8% were exonic, 21.4% were intronic, and 48.8% were intergenic (**Table 8**). Thus, no significant differences comparing these proportions were found between DE and non-DE lncRNAs in regeneration (**Figure 28 A**).

Classification of lncRNAs	*Drosophila* genome	Regeneration Total DE	Upregulated			Downregulated		
			Early	Mid	Late	Early	Mid	Late
Exonic	643 (26.2%)	39 (29.8%)	12 (41.4%)	2 (18.2%)	7 (31.8%)	4 (12.1%)	16 (28.6%)	9 (30.0%)
Intronic	424 (17.3%)	28 (21.4%)	4 (13.8%)	5 (45.4%)	6 (27.3%)	10 (30.3%)	13 (23.2%)	5 (16.7%)
Intergenic	1.388 (56.5%)	64 (48.8%)	13 (44.8%)	4 (36.4%)	9 (40.9%)	19 (57.6%)	27 (48.2%)	16 (53.3%)
lncRNAs Total	2.455	131	29	11	22	33	56	30
mRNAs								

Table 8. Classification of lncRNAs depending on their genomic context. Classification of lncRNAs into exonic, intronic and intergenic based on their position respect to their closest neighboring protein-coding gene. Total number and proportions for all annotated lncRNAs and for all DE lncRNAs in regeneration are represented.

We also characterized different features of DE lncRNAs, such as their transcript length, the exon number, or their GC content. Regarding their number of exons, no differences were observed between DE and non-DE lncRNAs, as the vast majority of lncRNAs presented 1 or 2 exons regardless of their classification (**Figure 28 B**). In terms of the length of their longest transcript, we found that exonic lncRNAs were the largest (median of 7.6 and 7.4 kb for DE and non-DE lncRNAs, respectively), followed by intergenic (median of 3.8 and 4.1 kb for DE and non-DE lncRNAs, respectively), and intronic (median of 2.8 and 4.0 kb for DE and non-DE lncRNAs, respectively) (**Figure 28 C**). However, no significant differences were observed between DE and non-DE for any of the classified groups (**Figure 28 C**). Finally, their GC content was quite regular, as the median for all groups ranged from 40.85% to 46.59% (**Figure 28 D**). The GC content of DE lncRNAs was lower than that of non-DE lncRNAs for the three groups, although it was only significant for intergenic lncRNAs (**Figure 28 D**). However, the implications of these differences are yet to be elucidated.

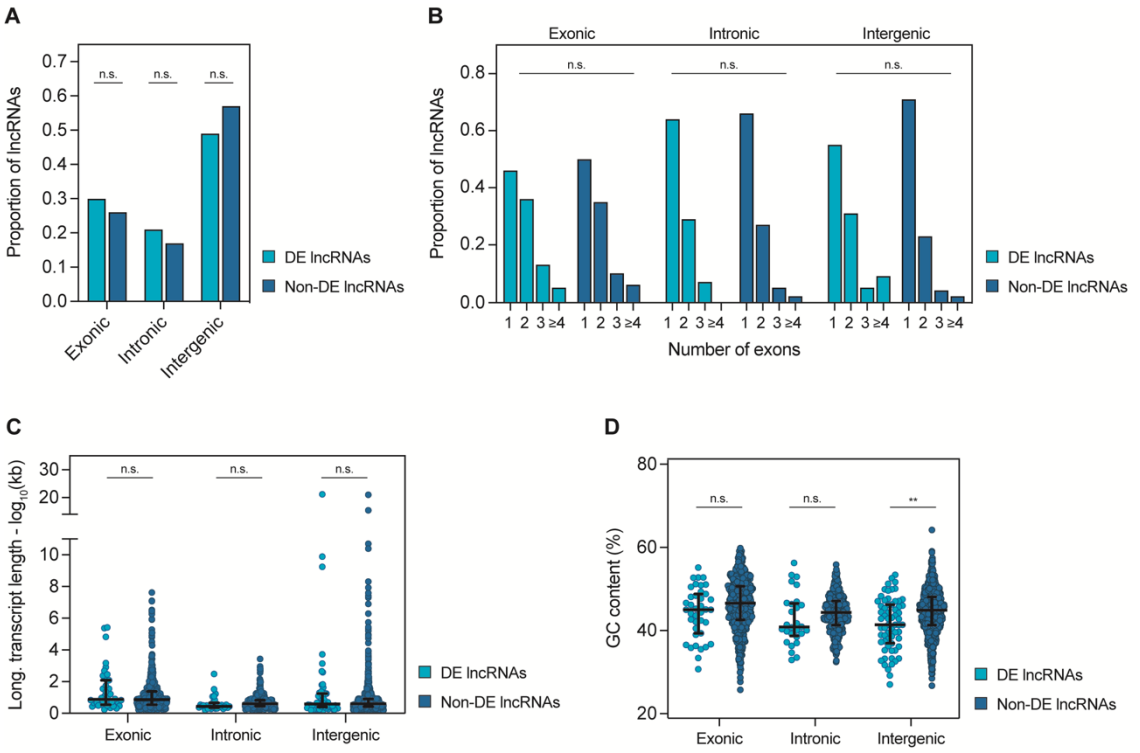


Figure 28. Features of DE lncRNAs in regeneration compared to non-DE. (**A**) Proportion of exonic, intronic and intergenic lncRNAs. (**B**) Classification of exonic, intronic and intergenic lncRNAs depending on the number of exons of their longest transcript. (**C**) Longest transcript length of exonic, intronic and intergenic lncRNAs. (**D**) GC content of the longest transcript of exonic, intronic and intergenic lncRNAs. Light blue represents DE lncRNAs, while dark blue corresponds to non-DE lncRNAs. Median and interquartile range are represented in (**C**) and (**D**). p<0.01 (**).

Overall, very few differences were observed comparing the exonic, intronic and intergenic groups between DE and non-DE lncRNAs, indicating that the group of lncRNAs DE in regeneration was a good representation of the annotated lncRNAs in flies, as their classification, number of exons, transcript length and GC content was very similar to that found for non-DE lncRNAs.

CHAPTER 2:
Characterization of lncRNAs *CR40469* and *CR34335*

Out of the 131 DE lncRNAs in regeneration, we selected one candidate for functional analysis using a loss of function mutation. To minimize the probability to affect the expression or function of overlapping PCGs, we focused on intergenic lncRNAs. We only considered lncRNAs upregulated during the early stage of regeneration, as we supposed that their absence could more likely compromise the wing recovery process. In total, we identified 13 intergenic lncRNAs upregulated in early regeneration (*see* ***Table 9*** *in Annex 2*). As some of these lncRNAs were lowly expressed in regenerating discs, we based our selection on the most expressed ones. Among them, *CR32207* showed the highest expression in early regenerating discs (**Figure 29**), however, although it did not overlap with any protein-coding gene and thus was classified as intergenic, it did overlap with another lncRNA (*CR42842*). Moreover, *CR32207* was already described as a hairpin RNA (hpRNA) regulating the *825-Oak* gene family at the post-transcriptional level (Wen et al. 2019). Consequently, we decided to focus on *CR40469*, the second most expressed intergenic lncRNA in early regeneration, which was completely uncharacterized.

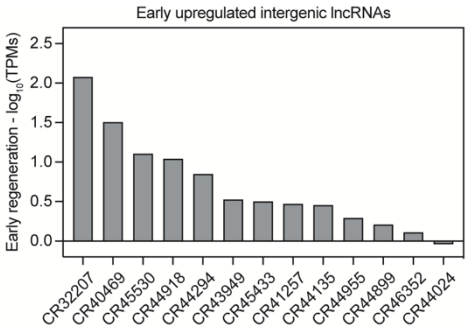


Figure 29. Expression of early upregulated intergenic lncRNAs. Expression in early regenerating discs of the 13 intergenic lncRNAs upregulated in the early stage of regeneration.

CR40469 is duplicated in the genome of *Drosophila*

One of the major drawbacks in the study of lncRNAs is finding orthologous RNAs across distant species, as lncRNAs usually lack sequence conservation. To analyse sequence similarities to *CR40469*, we performed a BLASTn search, but no matches were found for any other species, not even for other *Drosophilidae* species. However, we discovered that *CR40469* was actually duplicated in the genome of *Drosophila melanogaster*. The entire *CR40469* sequence was found inverted with 99.1% similarity (212/214 identities) within the sequence of the lncRNA *CR34335* (**Figure 30**). Also, while *CR40469* only got one described mono-exonic transcript of 214 nucleotides, two different mono-exonic transcripts with different poly-adenilation sites were annotated for *CR34335*, spanning 249 (short isoform) and 361 (long isoform) nucleotides each (**Figure 31 A**). Both *CR34335* isoforms include the *CR40469* sequence and 35 extra nucleotides at the 5' end, while the long isoform also presents 112 extra nucleotides at the 3' end. These extra sequences were not found anywhere else in the *Drosophila* genome nor in any other annotated genome. It is worth mentioning that, while *CR40469* was transcribed from the positive strand, *CR34335* was transcribed from the negative strand, thus, as the *CR40469* sequence found within *CR34335* was inverted, the transcripts of both genes shared the same sequence instead of being complementary to each other.

```
CR40469             ------------------------------------CACGTTCCTCACTAATTGTGGCTA  24
CR34335-short       GGCGGTCGAGTGCCTCACAGTGTATCAAGGGTTGGCCACGTTCCTCACTAATTGTGGCTA  60
CR34335-long        GGCGGTCGAGTGCCTCACAGTGTATCAAGGGTTGGCCACGTTCCTCACTAATTGTGGCTA  60
                                                    ************************

CR40469             TTTGCGCCATCGTCTCATGCAATGTTATTTGAGAGATGGCAAAATATATAGTATGTTTGT  84
CR34335-short       TTTGCGCCATCGTCTCATGCAATGTTATTTGAGAGATGGCAAAATATATAGTATGTTTGT  120
CR34335-long        TTTGCGCCATCGTCTCATGCAATGTTATTTGAGAGATGGCAAAATATATAGTATGTTTGT  120
                    ************************************************************

CR40469             CTCCAATGTGTTGAGACTGAGAAGATATTGTACCCGTGAATTGATGAAAATTGATTGATT  144
CR34335-short       CTCCAATGTGTTGAGACTGAGAAGATATTGTACCCGTGAATTGATGAAAATTGATTGATT  180
CR34335-long        CTCCAATGTGTTGAGACTGAGAAGATATTGTACCCGTGAATTGATGAAAATTGATTGATT  180
                    ************************************************************

CR40469             ATATTGTAATGTTGATTTCATGAAAAACACGCTGTGTTGGAGGAACTCAAACAAAACAAG  204
CR34335-short       ATAT-GTAATGTTGATTTCATGAAAAACACGCTGTGTTGGAGGAACTCAAACAAAACAAG  239
CR34335-long        ATAT-GTAATGTTGATTTCATGAAAAACACGCTGTGTTGGAGGAACTCAAACAAAACAAG  239
                    **** ********************************************************

CR40469             CAAAAAATCC----------------------------------------------------  214
CR34335-short       CATAAAATCC----------------------------------------------------  249
CR34335-long        CATAAAATCCAAAAAAAAAAAAAAAAAAAACAAATCAAATTTTAACAAACAATAATAATAC  299
                    ** *******

CR40469             ----------------------------------------------------------------  214
CR34335-short       ----------------------------------------------------------------  249
CR34335-long        TGTGTGGTGCGCTGGCGTGGGGGAGTGGTTATTCTCACATTCCAGTCTGATCGCTTTTTT  359


CR40469             --  214
CR34335-short       --  249
CR34335-long        TT  361


Identities 212/214(99%)
Gaps 1/214(0%)
```

Figure 30. Sequence alignment of *CR40469* and *CR34335* transcripts. The single transcript of *CR40469* and the short and long isoforms of *CR34335* were used for the alignment. 212 of the 214 nt of the *CR40469* transcript are identical in both *CR34335* isoforms. Identities are marked with an asterisk (*).

CR40469 and *CR34335* were already annotated as two lncRNAs positioned 3.2 Mb apart in the X chromosome (**Figure 31 A**). Also, it is important to mention that *CR34335* is located within a long intron of the protein-coding gene *Dpr interacting protein α* (*DIP-α*), which is involved in synapse organization, and is not expressed in the wing disc. Intriguingly, both lncRNAs were differentially-expressed in regeneration: while *CR40469* was upregulated in the early and late stages (**Figure 31 B**), *CR34335* was downregulated throughout all regeneration time-points (**Figure 31 B**). These opposing expression patterns in regeneration prompted us to analyse their expression in other tissues and stages to elucidate whether their expression was always antagonistic.

We used modENCODE data from whole animals at 21 different developmental stages, ranging from early embryo to late pupae. Indeed, *CR40469* was barely expressed, while *CR34335* was expressed at very high levels in all samples (**Figure 31 C**). Then, we analyzed their expression in available transcriptomic data from particular disc types (wing, antenna, eye and leg) at L3, EP and LP stages (Pérez-Lluch et al. 2020). Similar to what we observed in whole animals, *CR40469* was very lowly expressed, while *CR34335* was highly expressed everywhere (**Figure 31 D**). Moreover, we noticed that the lowest *CR34335* expression correlated with the highest *CR40469* expression, at the L3 wing disc (**Figure 31 D**).

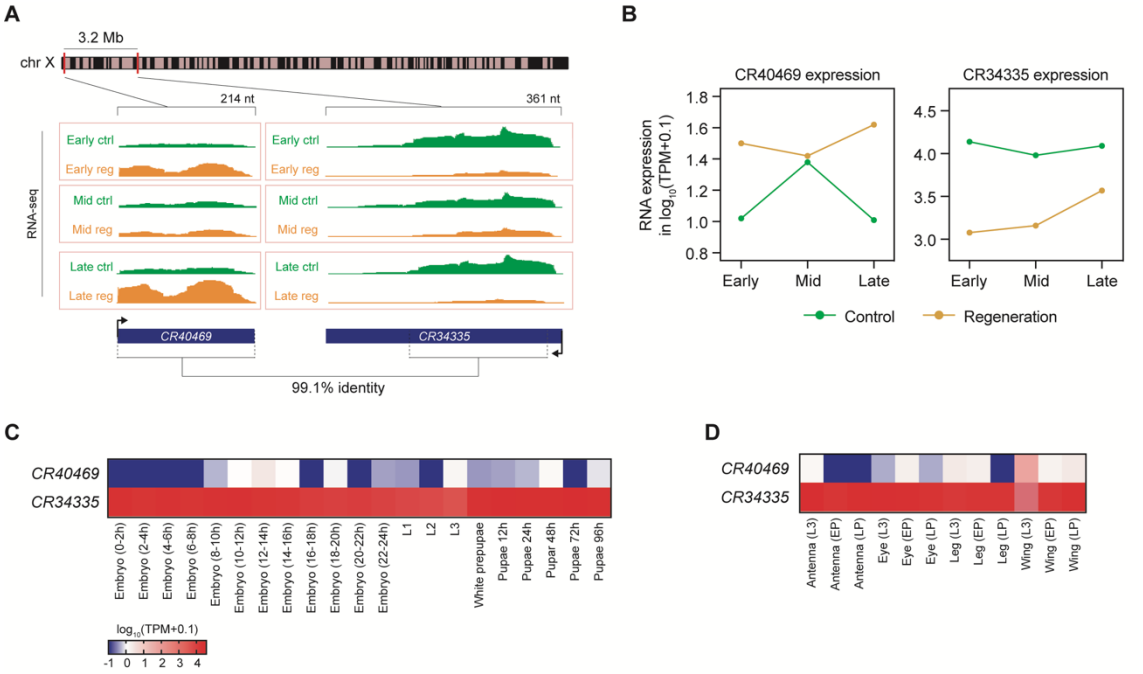


Figure 31. Expression pattern of *CR40469* and *CR34335*. (**A**) Scheme showing the positioning of *CR40469* and *CR34335*. RNA-seq tracks are represented in green (control) and orange (regeneration). (**B**) RNA-seq expression of *CR40469* and *CR34335* at each time-point. (**C**) Heatmap of *CR40469* and *CR34335* expression in different developmental stages. (**D**) Heatmap of *CR40469* and *CR34335* expression in different imaginal discs (antenna, eye, leg and wing) at L3, EP and LP stages.

Sequence analyses of *CR40469* and *CR34335*

To better characterize *CR40469* and *CR34335*, we performed *in silico* analyses based on their sequences. We analyzed their promoter regions, their putative RBP and miRNA binding sites, their predicted secondary structures, and the presence of smORFs within their sequences.

Analysis of their promoter regions

First, we analyzed the promoter sequence of *CR40469* and *CR34335* searching for transcription factor binding sites. We defined the core promoter (CP) as the region located ± 100 bp around the TSS, the proximal promoter (PP) as the region ± 2 kb around the CP, and the distal promoter (DP) as the region 10 kb upstream of the PP (**Figure 32 A**). Then, we used ATAC-seq data from control and regenerating discs to search for open chromatin regions within the CP, PP and DP of these lncRNAs.

Up to 6 open chromatin peaks were found at the *CR40469* promoter (4 distal and 2 proximal), all of which were significantly more open in regeneration than in control (**Figure 32 B**). Then, we analyzed the TF binding sites present in these open chromatin regions, and we found an enrichment for several TFs, including *Adh transcription factor 1* (*Adf-1*), *dorsal* (*Dl*), *E2F transcription factor 1* (*E2f1*), *Mothers against dpp* (*Mad*), *Trithorax-like* (*Trl*) or *zeste* (*z*), all of which were expressed in regenerating wing discs (**Figure 32 C**). On the other hand, in line with its high expression levels, the 2 open chromatin regions found in the *CR34335* promoter (1 proximal and 1 at the core promoter) were accessible in control and regeneration discs at all time-points (**Figure 32**

D). These 2 regions were enriched for multiple TF binding sites, including *Adf-1*, *Dl*, *Mad, Trl and z* (**Figure 32 E**).

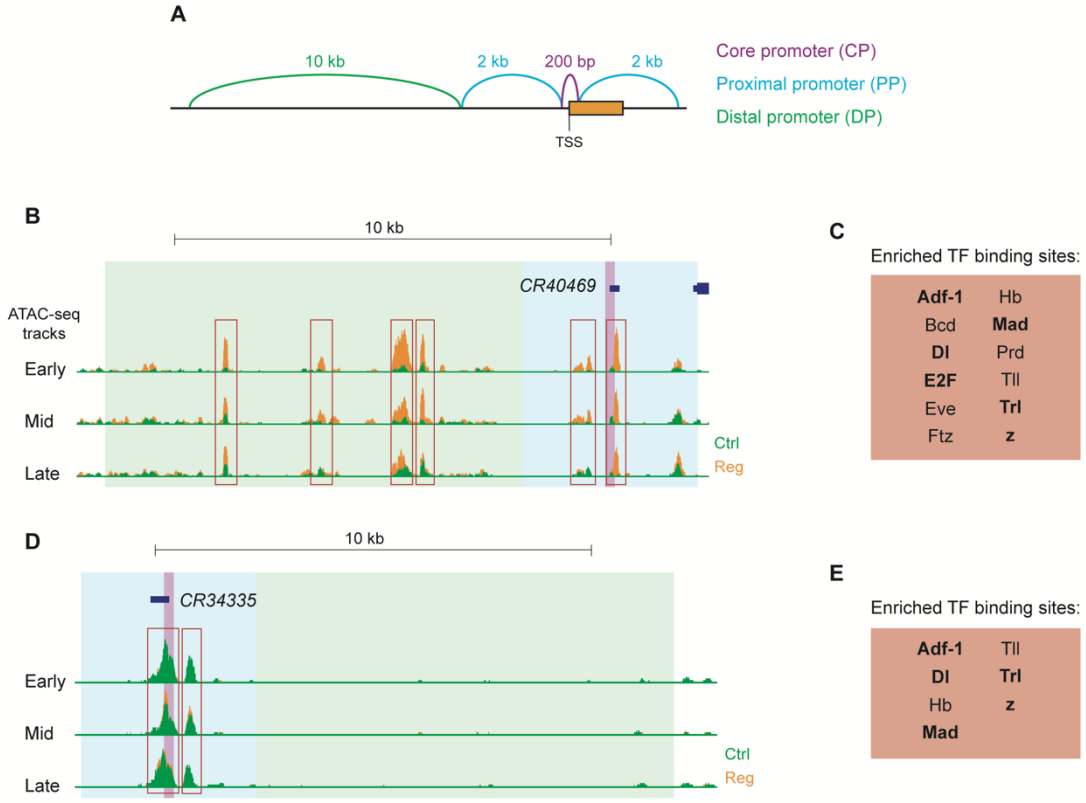


Figure 32. Open chromatin regions in the promoter region of *CR40469* and *CR34335*. (**A**) Scheme representing the core promoter (CP), proximal promoter (PP) and distal promoter (DP) around the TSS. (**B**) Screenshot showing the ATAC-seq peaks present in the promoter of *CR40469*. (**C**) TF binding sites enriched at the open chromatin regions present in the promoter of *CR40469*. (**D**) Screenshot showing the ATAC-seq peaks present in the promoter of *CR34335*. (**E**) TF binding sites enriched at the open chromatin regions present in the promoter of *CR34335*. The name of TFs expressed in regeneration are represented in bold (in C and E).

The set of TF binding sites enriched in the open chromatin regions present at the *CR40469* and *CR34335* promoters was quite similar. Indeed, among the TF expressed in the wing disc, motifs for 5 of them were enriched in the promoters of both lncRNAs, while *E2f1* binding sites were enriched only in the promoter of *CR40469*. Thus, it is probable that the same TFs were targeting both lncRNAs. Nevertheless, the differences in the availability of these open chromatin regions was probably a factor driving their antagonistic expression patterns despite being targeted by similar TFs.

Analysis of miRNA target sites

Apart from differences in the regulation at the transcriptional level, the levels of *CR40469* and *CR34335* could also be regulated post-transcriptionally, for instance, by binding with different preferences to different partners. For instance, there are many examples of RBPs whose binding to lncRNAs promotes their stability, such as HuR, which blocks the target sites recognized by RNA-degrading enzymes (Ripin et al. 2019), or SRSF1, which stabilizes the lncRNA NEAT1 in certain types of cancer (Zhou et al. 2019). Unfortunately, as RBPs do not bind to specific RNA sequences, there are currently no mechanisms to predict the binding partners of a given RNA *in silico*. Another type of lncRNA putative binding partners are microRNAs (miRNAs). Traditionally, miRNAs are thought to

negatively regulate mRNA translation and stabilization by binding in a sequence-specific manner to their 3' untranslated region (UTR) (Bartel, 2018). Some lncRNAs are actually known to act as miRNA precursors, while others are described to act as miRNA sponges, competing with miRNA target mRNAs for their binding, thus regulating their availability.

Regarding *CR40469* and *CR34335*, there was no evidence pointing towards the possibility of these lncRNAs acting as miRNA precursors. Genome-wide studies searching for novel pre-miRNAs did not find *CR40469* nor *CR34335* as putative targets of *Drosha*, which actively participates in the canonical pre-miRNA cleavage (Kadener et al. 2009). Moreover, dedicated BLAST search did not reveal significant similarities to any of the known pre-miRNAs. On the other hand, to test whether *CR40469* and *CR34335* could act as miRNA sponges, we used the *Find Individual Motif Occurrences* (*FIMO*) tool from the *MEME* suite. We searched for the presence of the seed sequence of the 161 miRNAs annotated in *Drosophila* within the *CR40469* and *CR34335* RNA sequences. Complementary regions to the seed sequences of 4 different miRNAs were localized in the shared sequence between *CR40469* and *CR34335* (*miR-210*, *miR-252*, *miR-307* and *miR-375*), but no sites were observed for the extra sequence specific to *CR34335* (**Figure 33 A**). All predicted miRNA target sites were clustered in 35 nucleotides near the 3' end of *CR40469* (**Figure 33 B**).

A

miRNA	Seed sequence	Target	Start pos.	Stop pos.
miR-210	UUGUGCG	*CR40469*	170	176
		CR34335	205	211
miR-252	AAGUACU	*CR40469*	161	167
		CR34335	196	202
miR-307	CACAACC	*CR40469*	178	184
		CR34335	213	219
miR-375	UUGUUCG	*CR40469*	199	205
		CR34335	234	240

B

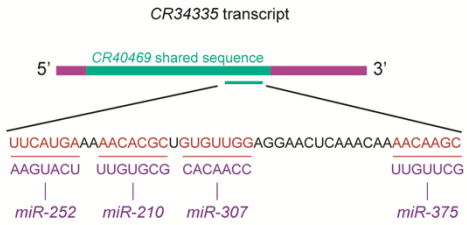


Figure 33. Predicted miRNA target sites within *CR40469* and *CR34335* lncRNAs. (**A**) Table showing the miRNAs with predicted target sites, their seed sequence, and their position within the target sequences. (**B**) Scheme showing the position of the miRNA seed sequences within the *CR34335* transcript sequence.

As we observed no differences in the predicted miRNA target sites between *CR40469* and *CR34335*, it is highly unlikely that their antagonistic expression patterns were the result of miRNA regulation.

Predicted secondary structures

The function of some lncRNA is known to rely mostly on their secondary structure. The specific folding of some lncRNAs is necessary to allow the binding of multiple proteins, such as the well-known HOTAIR, which serves as a scaffold to the histone-modifying complex PRC2 and the histone demethylase LSD1 through different structural motifs (Tsai et al. 2010; Somarowthu et al. 2015). Moreover, lncRNAs that act through the direct sequence-specific binding to the DNA or to other RNA molecules, fold in a specific manner so that their sequence recognizing their target DNA or RNA regions remains accessible for binding (Li et al. 2016; Kuo et al. 2019).

We used the mFold predicting tool based on the MEF algorithm to predict the secondary structures of *CR40469* and *CR34335* (Zuker, 2003). The predicted structure of *CR40469* highlights 3 stem loops, however, the rest of the structure had very poor probability scores (**Figure 34 A**). If these high probability stem loops are formed *in vivo*, they could

facilitate the binding to RNA binding proteins (RBPs) (Corley et al. 2020). Nevertheless, there is little knowledge about the RNA sequence preferences of RBPs, thus it is currently impossible to infer the putative *CR40469* binding partners just by the RNA sequence. Regarding the secondary structure of *CR34335*, despite the fact that the short isoform only adds 35 extra nucleotides at the 5' end with respect to *CR40469* (**Figure 31 A**), clear differences between their predicted structures can be observed, not only in the number of stem loops (4 for *CR34335*), but also in their length, as the number of complementary bases was notably higher for *CR34335* stem loops (**Figure 34 B**). These differences may indicate that, despite their striking sequence similarity, their functions could be completely different, in line with their highly dissimilar expression patterns (**Figure 31 B-D**). When comparing the short and long isoforms of *CR34335*, 3 main stem loops were maintained at the same position, but the rest of the sequence was notably different (**Figure 34 B-C**).

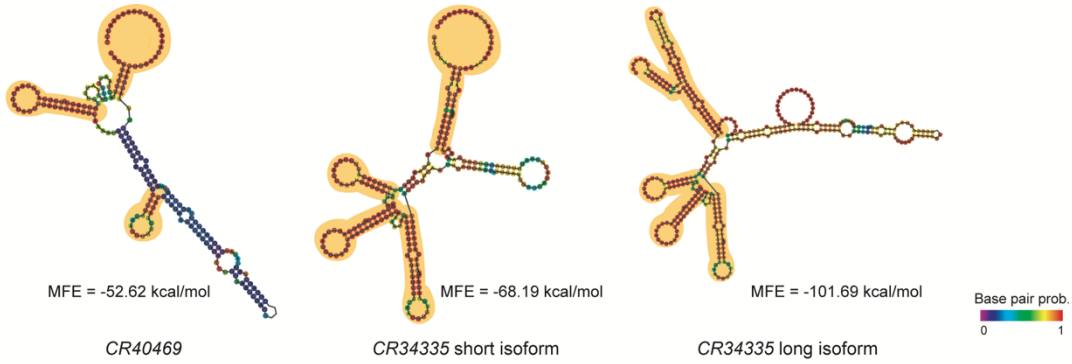


Figure 34. Predicted secondary structure of *CR40469* and *CR34335* isoforms. Secondary structures of *CR40469*, *CR34335* short and *CR34335* long isoforms based on the minimum free energy (MFE) prediction method. High probability structures marked in yellow.

Altogether, despite remarkable differences between *CR40469* and *CR34335* predicted secondary structures were observed, further studies are required to corroborate their folding *in vivo*. In any case the stem loops present in both lncRNAs were predicted with very high probability, thus it is tempting to speculate that their presence *in vivo* is rather likely. If that is true, these differences could imply different protein binding preferences, resulting in completely different functions.

Identification of smORFs

Although by definition lncRNAs lack protein-coding potential, some of them have been described to encode small functional peptides. For this reason, we searched for putative smORFs within the *CR40469* and *CR34335* RNA sequences. A predicted smORF encoding a peptide spanning 33 aa was found within the common sequence between both lncRNAs, meaning that both transcripts could be potentially translated (**Figure 35**). Next, we analyzed the peptide sequence, but we did not find sequence similarity to any other protein, nor any known functional domain.

Nonetheless, it is not rare to find smORFs within lncRNA sequences, in fact, it is estimated that around 98% of the annotated lncRNAs contain at least one smORF (Couso and Patraquim, 2017). For this, further experiments are required to determine whether this particular smORF is actually translated and, if that is the case, if the encoded peptide is functional.

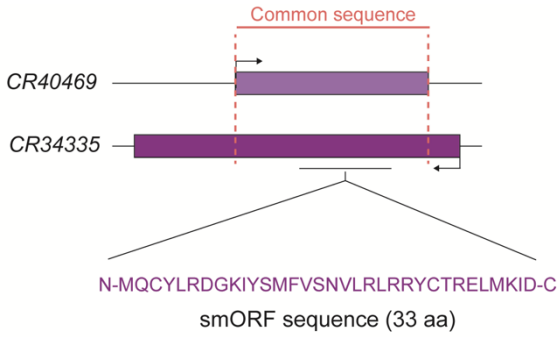


Figure 35. Predicted smORF within *CR40469* and *CR34335* RNA sequence. Localization and amino acid sequence of the predicted smORF present in *CR40469* and *CR34335* sequences.

Deletion of *CR40469* does not affect development

Reverse genetic assays, in which targeted mutations are created for specific candidate genes, have been widely used to test the functionality of lncRNAs. In our case, as *CR40469* was upregulated upon damage, we were interested in knock-out mutants. Since no mutants were available, a *CR40469* knock-out was generated at the laboratory of Aurelio Teleman group (*see Materials and Methods section for description and characterization of the mutation)*.

We first checked whether *CR40469* deletion affected the expression of nearby genes by analysing the expression of all genes located 100 kb away from the *CR40469* locus. No significant differences were detected in the expression of any of the neighbouring genes studied (**Figure 36**). We also investigated whether *CR40469* deletion had any effects on the expression of *CR34335.* Indeed, a significant increase in the expression of *CR34335* was found in *CR40469* mutants (**Figure 36**). This finding may imply an antagonistic regulation at the transcriptional or post-transcriptional level of both lncRNAs, which could explain the coincidence of the lowest *CR34335* expression and the highest *CR40469* expression at the L3 wing disc.

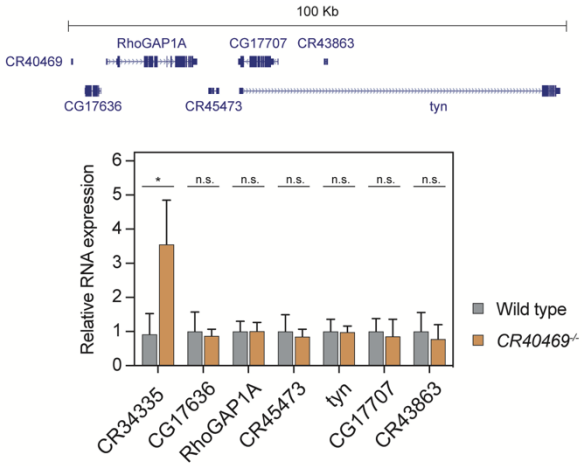


Figure 36. Molecular characterization of *CR40469* mutant. Analysis of the expression by qPCR of *CR34335* and the 6 genes located at a distance of < 100 kb from the *CR40469* locus. The plot represents the fold change enrichment between wild type (grey) and *CR40469* mutants (orange). Error bars indicate the standard error of the mean of 3 biological replicates.

Prior to the study of the *CR40469* mutant in the context of regeneration, we studied its putative effects under normal laboratory conditions. First, we analyzed the embryo to adult viability of *CR40469* homozygous mutants compared to wild types. However, we did not observe any differences in their viability (**Figure 37 A**), nor in the proportion of hatched males and females (**Figure 37 B**), indicating that the deletion of *CR40469* did not affect the normal development of both, male and female adult flies. These results were in line with the almost negligible expression levels of *CR40469* throughout development from embryo to late pupae (**Figure 31 C**).

A key developmental feature during development is pupariation, the transition from larvae to pupae. The process of pupariation is triggered by an increase in the levels of the steroid hormone 20-hydroxyecdysone at the end of the L3 stage, and it marks the moment in which the larvae stops feeding, attaches to the substrate and produces the appropriate changes in the larval cuticle to form the puparium (Heredia et al. 2021). Indeed, it is also useful to study developmental delays, as it recapitulates any delay in the embryo and larval stages. To detect putative alterations in development caused by the deletion of *CR40469*, we conducted a pupariation assay. Our results did not show any differences in pupariation between control and mutant flies, indicating, again, that the complete absence of *CR40469* was not necessary for their proper developmental timing (**Figure 37 C**). This feature was also important for further regeneration experiments, as the capacity of wing discs to regenerate is dependent on the developmental stage: the later we induce regeneration, the less proportion of regenerated wings are obtained (Smith-Bolton et al. 2009).

Next, we investigated whether *CR40469* deletion had any effect on the adult fly body size, but we were unable to observe any differences between *CR40469* homozygous mutants and controls, neither in males, nor in females (**Figure 37 D**). Likewise, no differences were observed in the size or pattern of adult wings (**Figure 37 E-F**). These results show that *CR40469* was completely dispensable for the proper development of adult wings, albeit *CR40469* was clearly expressed in L3 wing discs (**Figure 31 D**).

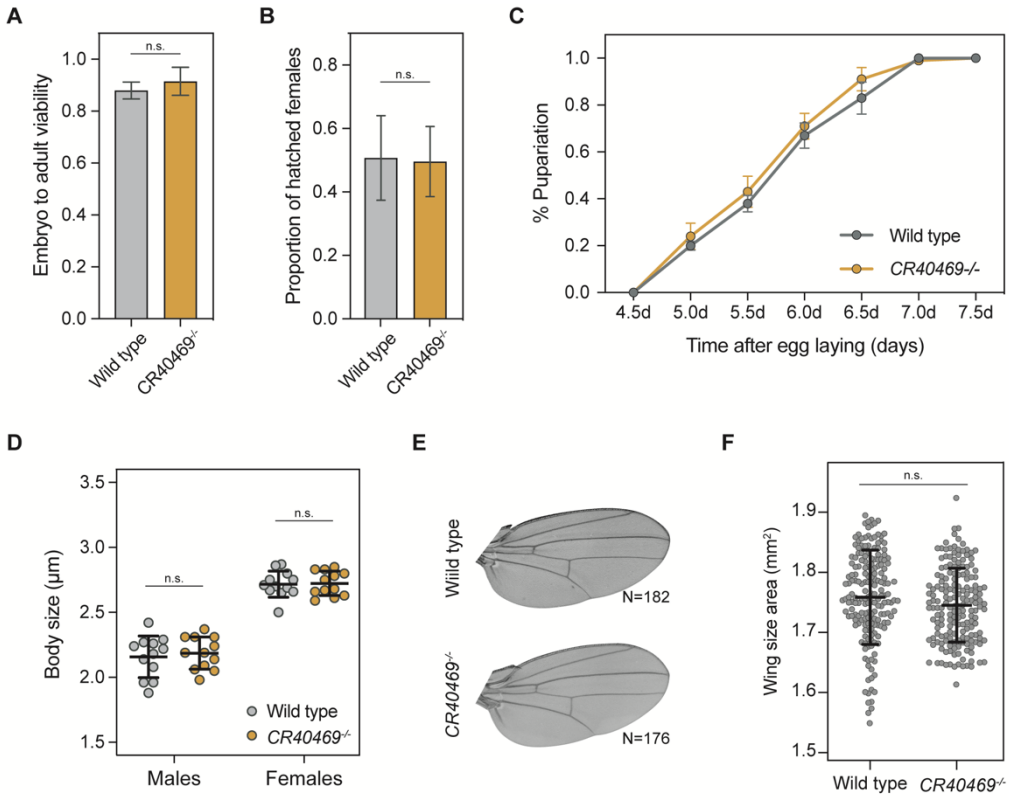


Figure 37. Characterization of *CR40469* homozygous mutant. (**A**) Plot representing the embryo to adult viability of wild type and *CR40469*$^{-/-}$ flies. (**B**) Plot showing the proportion of females among all hatched flies. (**C**) Pupariation analysis showing the % of flies that reach pupariation at each timepoint. (**D**) Scattered plot showing the body size of male and female wild type and *CR40469*$^{-/-}$ adult flies. (**E**) Sample images of adult female wings of wild type and *CR40469*$^{-/-}$ mutants. (**F**) Scattered dot plot showing the adult wing sizes of wild type and *CR40469*$^{-/-}$ mutants. Mean and standard deviation (SD) are represented in A, B, D and F. Mean and standard error of the mean (SEM) is represented in C.

Altogether, we did not observe any differences between wild types and *CR40469* homozygous mutants for any of the parameters tested, suggesting that *CR40469* was not required for the proper development of adult flies under normal laboratory conditions.

Absence of *CR40469* severely compromises regeneration

These observations during normal development could not rule out the possibility of *CR40469* participating in wing development specifically upon damage, as this lncRNA was clearly upregulated in regenerating wing discs (**Figure 31 B**). To analyse its putative involvement in regeneration, we compared the proportion of regenerated wings between control and *CR40469* homozygous mutants. For this, we used the *lexO/lexA* transactivator system to induce the expression of the proapoptotic gene *rpr* in the *spalt* domain of control and mutant wing discs (*see **Figure 14** in Methods section*). Strikingly, we observed a significant decrease in the proportion of regenerated wings among *CR40469* mutants compared to controls (**Figure 38 A-C**). Aberrant non-regenerated wings were characterized by the absence of complete veins and crossveins, mainly the absence of L3 and L4 veins and the anterior crossvein, by the presence of notches in the wing blade, and by the smaller wing size (**Figure 38 A-C**).

These findings posed *CR40469* as the first lncRNA described to participate in the regeneration of wings in *Drosophila*, but not required for normal development.

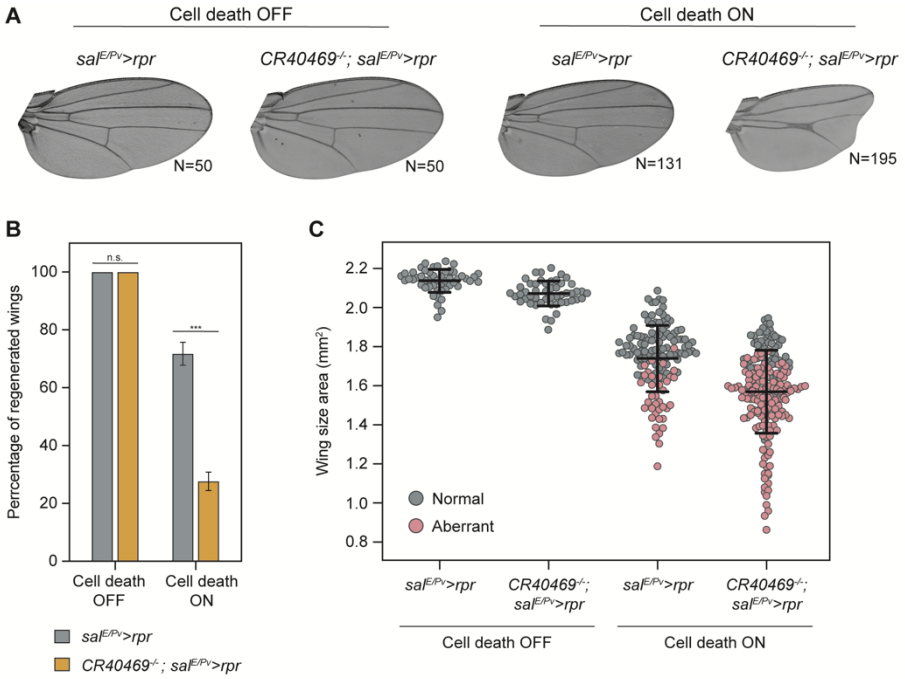


Figure 38. Regeneration capacity of *CR40469*$^{-/-}$ mutants. (**A**) Sample adult wings of control (*sal*$^{E/Pv}$*>rpr)* and *CR40469* homozygous mutants (*CR40469*$^{-/-}$*;sal*$^{E/Pv}$*>rpr)* without cell death induction (cell death OFF, grown at 17ºC) and after the induction of cell death (cell death ON, *rpr* activation for 11 h). (**B**) Percentage of regenerated wings per condition. (**D**) Adult wing size of each genotype with and without cell death. Grey dots represent normal wings, pink dots represent aberrant wings. Mean and standard deviation (SD) are represented in B and C.

Transcriptomic changes of *CR40469* mutants in regeneration

To characterize the molecular changes occurring in *CR40469* mutants in regeneration, we investigated their transcriptome. For this, we performed RNA-seq of wild type and *CR40469* homozygous mutants in regenerating conditions (**Figure 39**).

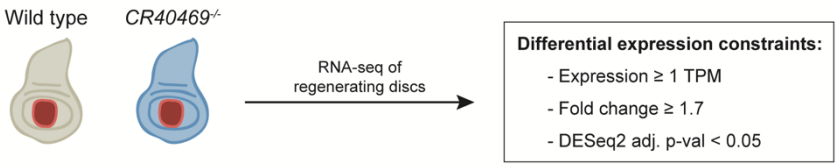


Figure 39. Scheme of the RNA-seq samples from regenerating controls and *CR40469*$^{-/-}$ mutants. Scheme showing the samples analysed and the differential expression analysis constraints.

For the differential expression analysis, we performed DESeq2 and set up a minimum fold change of 1.7, a minimum adjusted p-value of 0.05, and a minimum expression of 1 TPM. Applying these constraints, we found a total of 95 DE genes: 48 upregulated and 47 downregulated genes in *CR40469* mutants compared to wild types (**Figure 40 A**). Next, we analyzed the expression pattern of these misexpressed genes during the normal regeneration process taking place in wild type discs. To do so, we checked whether they were differentially-expressed in any time-point during regeneration. Actually, 45.8% of genes upregulated in the mutant (22 genes) were inhibited in at least

one regeneration time-point, 7 of which were inhibited at the 3 stages (**Figure 40 B**). On the other hand, 55.3% of downregulated genes in the mutant (26 genes) were activated in at least one regeneration time-point in the wild type, 8 of which were activated at the 3 stages (**Figure 40 B**).

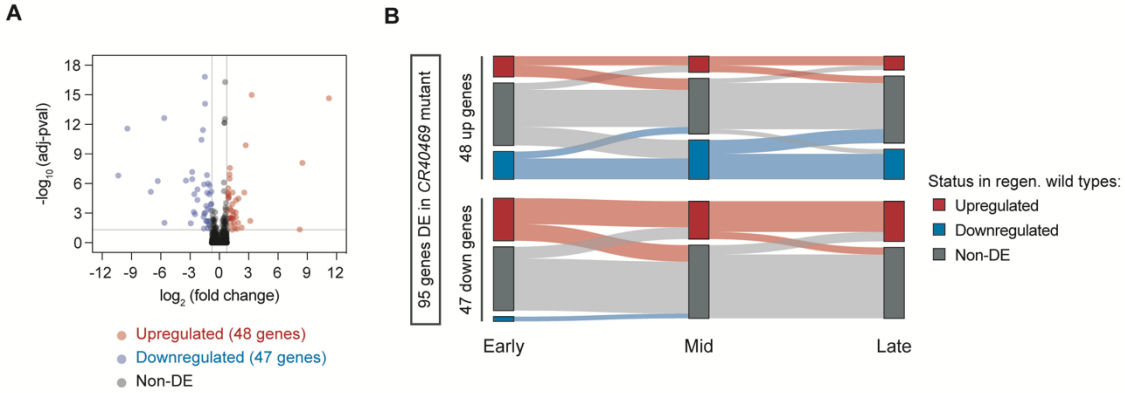


Figure 40. Differentially-expressed genes in *CR40469* mutants in regeneration. (**A**) Volcano plot displaying the upregulated (red), downregulated (blue) and non differentially-expressed genes (grey) in *CR40469* mutants compared to wild types in regenerating conditions. (**B**) Alluvia plot for upregulated (top plot) and downregulated (bottom plot) genes in *CR40469* mutants. The DE status (upregulated, downregulated and NDE) of each gene in regenerating wild types for each time-point is represented.

To get an overview of the genomic location of these DE genes, we distributed them into the fly chromosomes. We found a non-significant increase in the proportion of DE genes placed in the X chromosome (**Figure 41 A**), which is the chromosome where *CR40469* is located. Although the number of DE genes was too low to reach significance, the higher proportion of genes located in the X chromosome was maintained for both upregulated and downregulated genes compared to the proportion of all annotated genes.

To determine if the deletion of *CR40469* caused a more severe dysregulation in the DE genes located in the X chromosome compared to those from other chromosomes, we compared the $\log_2$(fold change) of DE genes from each chromosome, but no significant differences were observed (**Figure 41 B**). Next, we investigated the distribution of DE genes located in the X chromosome and we did not find any clustering in specific hotspots (**Figure 41 C**). Also, we did not observe a local effect around the deleted region, as the nearest DE gene was located ~1 Mb away from the deletion. Thus, these findings indicate that *CR40469* is most probably acting *in trans*, although further studies are required to describe its mechanism of action.

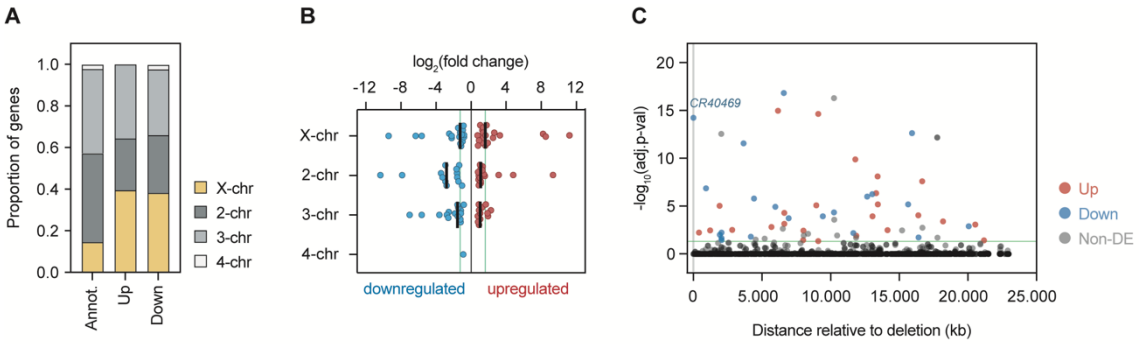


Figure 41. Chromosome distribution of DE genes in *CR40469* mutants. (**A**) Chromosome distribution of the upregulated and downregulated genes in *CR40469* mutants compared to the proportion of all annotated genes in flies. (**B**) Comparison by chromosomes of the $\log_2$ (fold change) of upregulated and downregulated genes. (**C**) Position of genes located in the X chromosome relative to the *CR40469* deletion. The $-\log_{10}$(adj.p-val) is shown in X-axis. Upregulated genes in the mutant are shown in red, downregulated genes in blue, and non-DE genes in grey.

Analysis of *CR34335* mutant phenotypes

Considering that *CR40469* was duplicated in the *Drosophila* genome, we also studied its duplication, the lncRNA *CR34335*. A mutant with a transposable element inserted within the *CR34335* single exon was commercially available. This mutant has a *Minos* transposon spanning 7.7 kb in the middle of the *CR34335* exon, affecting both isoforms. Mutant flies were perfectly viable in homozygosis, allowing us to study the expression of both, *CR34335* and their nearby genes (**Figure 42 A**). We confirmed that the expression of *CR34335* was inhibited by the presence of the transposon, while the expression of the overlapping gene *DIP-α* was not affected (**Figure 42 B**). However, the expression of 4 nearby genes located upstream and downstream of the mutant was also downregulated (**Figure 42 B**), suggesting that the presence of the transposable element may disrupt regulatory elements, leading to a dysregulation of these genes.

Next, we analyzed the phenotypes of adult wings in *CR34335* homozygous mutants, and no defects were observed in the pattern formation, as the veins and crossveins were perfectly formed (**Figure 42 C**). However, mutant wings were smaller compared to wild types (**Figure 42 C**), which could either indicate an effect of *CR34335* in wing size, or an effect of the downregulated genes located near the transposon insertion site.

We also tested the regeneration capacity of *CR34335* homozygous mutants, revealing a small but significant decrease in the percentage of regenerated wings upon damage (**Figure 42 D**). Since *CR34335* was downregulated in regeneration, this observation suggests that a basal expression of *CR34335* is necessary for the regeneration process. Nevertheless, it could not be ruled out that it is the downregulation of adjacent genes, rather than the *CR34335* inhibition, that compromised the process.

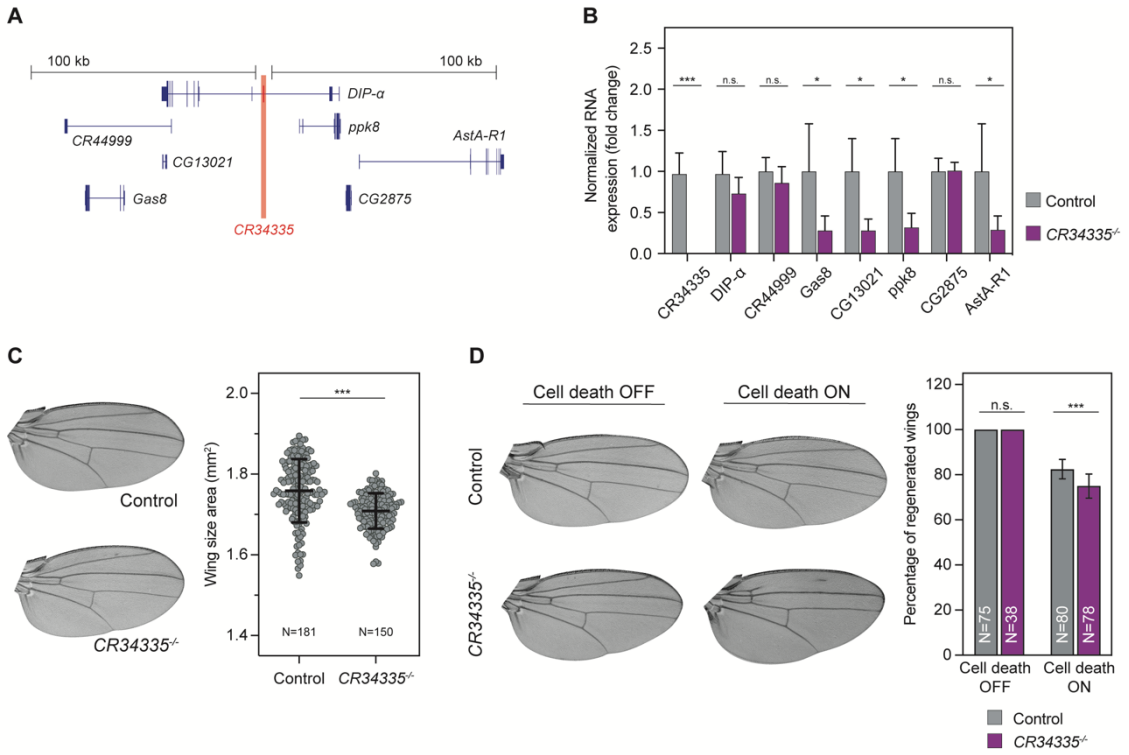


Figure 42. Characterizaztion of *CR34335*$^{-/-}$ mutant. (**A**) Screenshot showing the genes surrounding the *CR34335* locus. (**B**) Expression of control and *CR34335* mutant wing discs measured by qPCR. Fold change between control and *CR34335* mutants is represented. Error bars represent the standard error of the mean (SEM) of 3 biological replicates. (**C**) Adult wings of control and *CR34335* mutant flies. Plot showing the wing size of adult wings. (**D**) Adult wings of control and *CR34335* mutants without cell death (cell death OFF) and after the induction of apoptosis (cell death ON). The plot shows the percentage of regenerated wings per condition. Mean and standard deviation (SD) is represented in C and D. $p<0.05$ (*). $p<0.01$ (**). $p<0.001$ (***).

Subcellular localization of *CR40469* and *CR34335*

One of the major factors for the prediction of lncRNA functionality is their subcellular localization. To get insight into the putative roles of *CR40469* and *CR34335*, we studied their subcellular localization in the wing disc. For this, we used fluorescent in situ hybridization (FISH) to detect *CR40469* and *CR34335* transcripts. Unfortunately, it was impossible to design specific probes for *CR40469* and for *CR34335*, as both lncRNAs share the same sequence, and the extra sequence present only in *CR34335* was not long enough to effectively design an appropriate probe. In consequence, we could only design a single probe to detect both lncRNAs at the same time.

In order to distinguish each lncRNA, we used different genetic backgrounds: we used wing discs of *CR34335* homozygous mutants to detect *CR40469*, and *CR40469* homozygous mutant discs to detect *CR34335*. Probably due to the much higher expression of *CR34335* compared to *CR40469*, the expression pattern observed in the wild type and *CR40469* mutants was undistinguishable (**Figure 43 A,C**), meaning that the FISH signal observed in the wild type probably corresponded to *CR34335* transcripts, which localize specifically in the cytoplasm of wing discs (**Figure 43 A,C**). On the other hand, we were not capable of detecting *CR40469* transcripts, as no differences were observed between the antisense and the control sense probe (**Figure 43 E-F**).

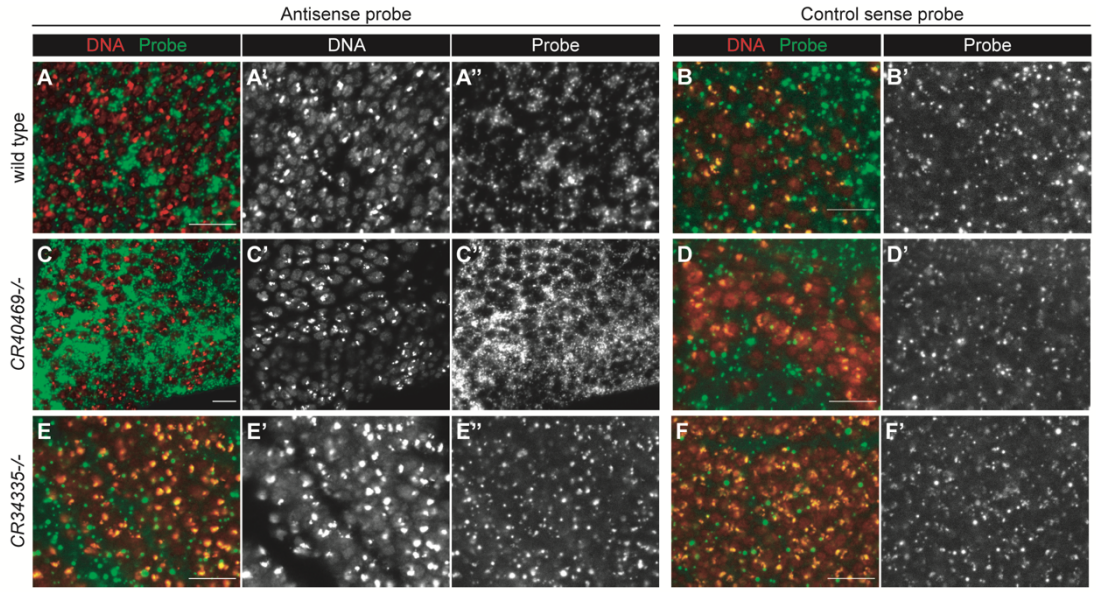


Figure 43. Subcellular localization of *CR40469* and *CR34335* in the wing disc. FISH imaging of (**A**) wild type, (**C**) *CR40469* homozygous mutants, or (**E**) *CR34335* homozygous mutants hybridized with an antisense probe designed to detect *CR40469* and *CR34335* transcripts. FISH negative controls of (**B**) wild type, (**D**) *CR40469* homozygous mutant, and (**F**) *CR34335* homozygous mutant wing discs hybridized with the control sense probe. N ≥ 10 per condition. Scale bar = 10 µm.

As *CR40469* was upregulated in regeneration, we repeated the hybridization protocol on *CR34335*$^{-/-}$ discs upon the induction of cell death. Nevertheless, the apoptotic cells acted as a sponge and sequestered the RNA probe, making it impossible to properly detect the subcellular localization of *CR40469* in the living cells (**Figure 44 A-B**). To overcome this technical problem, we followed an alternative approach to induce regeneration: we performed a small cut in the wing disc and cultured it under specific conditions of medium and temperature to promote wound healing (Fristrom et al. 1973; Tsao et al. 2016). Nonetheless, we were not able to detect *CR40469* in any case (**Figure 44 C-D**).

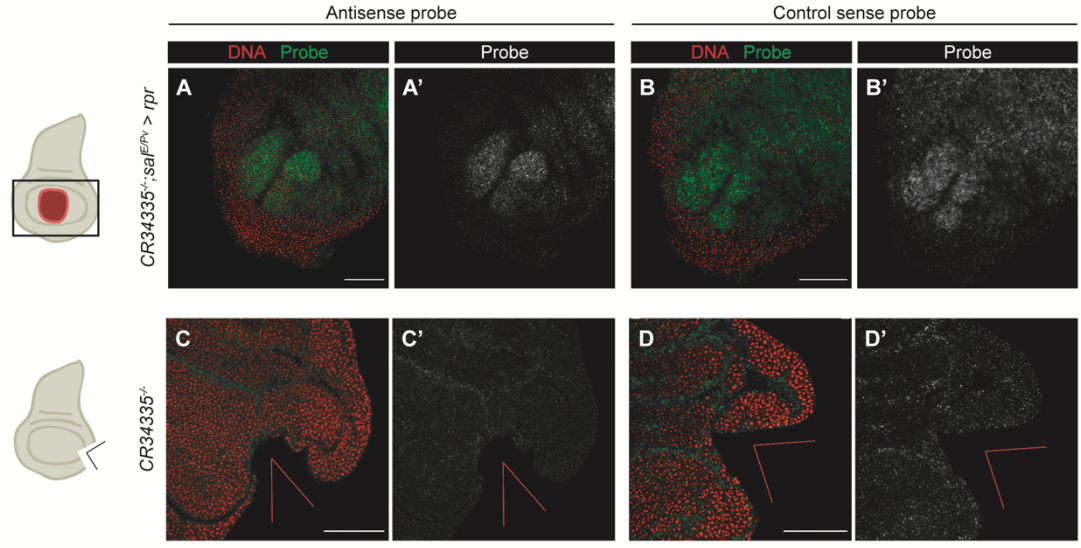


Figure 44. FISH images of regeneration-induced wing discs. (**A-B**) Images of *CR34335* mutant wing discs after the induction of cell death (*sal*$^{E/Pv}$*>rpr*) hybridized with *CR40469* antisense probe (A) or control sense probe (B). (**C-D**) Images of *CR34335* mutant wing discs cut and cultured for 6 hours and hibridized with *CR40469* antisense probe (C) or control sense probe (D). N ≥ 5 per condition. Scale bar = 50 µm.

CR34335 may participate in in translation regulation

With the aim to infer the putative biological processes in which this lncRNA might be involved in, we investigated the function of genes with similar expression patterns. Given the high expression levels of *CR34335* across all tissues (**Figure 31 B-D**), we plotted the expression of the most expressed genes in control wing discs. The results revealed that only the expression of genes encoding for ribosomal subunits was on par with the expression levels of *CR34335* (**Figure 45 A**).

We further explored if the expression profile of *CR34335* and that of genes coding for ribosomal subunits was also similar across the regeneration samples. We observed that, while their expression was similar in control discs, the downregulation of *CR34335* in regeneration was not observed for any of the ribosomal genes tested (**Figure 45 B**). Then, we compared their expression in wing, antenna, eye, and leg discs at different developmental stages (L3, EP and LP) and, again, their expression levels were much higher than the median of expressed mRNAs (**Figure 45 C**). Nevertheless, while the expression of ribosomal genes decreased at the LP stage in all disc types, the expression of *CR34335* was more stable, with the exception of L3 wing discs, in which *CR34335* expression was the lowest (**Figure 45 C**).

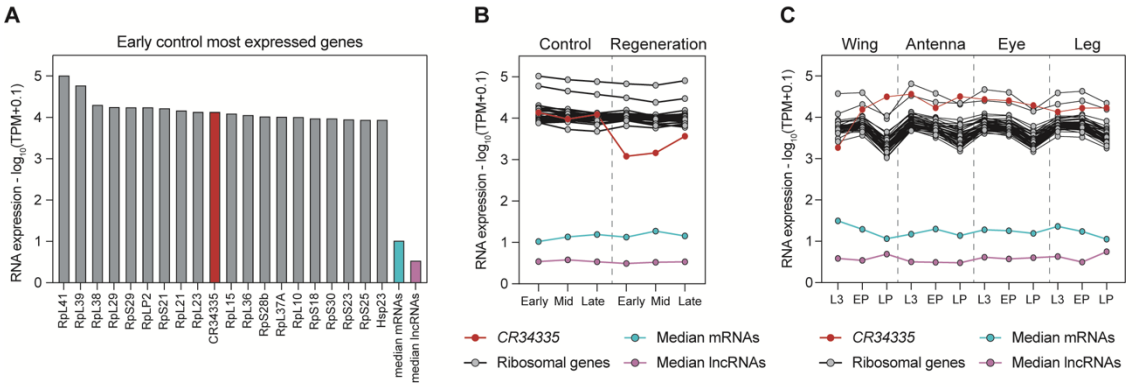


Figure 45. *CR34335* is expressed at the levels of genes encoding for ribosomal subunits. (**A**) Expression of top 20 most expressed genes in early control discs. (**B**) Expression of *CR34335* compared to ribosomal genes across control and regeneration samples at early, mid and late stages. (**C**) Expression of *CR34335* and ribosomal genes at L3, EP and LP stages of wing, antenna, eye and leg discs.

Next, we explored the possibility that the putative peptide encoded by the smORF present in the exonic sequence of *CR34335* and *CR40469* (**Figure 35**), was similar to the ribosomal subunits encoded by ribosomal genes. For this, we compared the predicted smORF to 137 ribosomal subunits encoded by 49 different genes. First, we observed that the smORF-encoded peptide was smaller than the ribosomal subunits (**Figure 46 A**). In terms of molecular weight, ribosomal subunits showed a wide weight range, and the predicted peptide fell within the range (**Figure 46 B**). We also investigated the isoelectric point (pI), which is defined as the pH at which a given molecule has a net charge of zero. It is known that ribosomal subunits generally carry net positive charges that facilitate the interaction with the negatively charged ribosomal RNA (rRNA) (Baker et al. 2001; Klein et al. 2004). Subsequently, 96.3% of ribosomal peptides showed a pI ranging from 9 to 13 (the pH within *Drosophila* cells is 6.5-7.5 (Massie et al. 1981)), but the smORF-encoded peptide presented a pI lower than 6 (**Figure 46 C**). Thus, it is improbable that this putative peptide interacts with rRNA or other negatively charged nucleic acids.

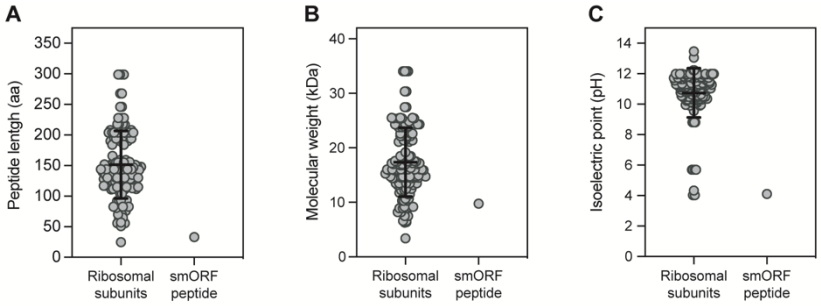


Figure 46. Comparison of the putative smORF-encoded peptide and ribosomal subunits. (**A**) Length (aminoacids, aa), (**B**) molecular weight in kDa, and (**C**) isoelectric point (pH) of ribosomal subunit peptides and *CR40469-CR34335* putative smORF-encoding peptide.

Although the *CR34335* predicted peptide was within the range of the ribosomal peptides in every analyzed feature, it was clearly an outlier. Thus, in the eventual case that *CR34335* was actually translated, the resulting peptide does not show the properties of the typical ribosomal subunits, indicating that most probably it is not a compound of the ribosomes.

CHAPTER 3:

The role of D-GADD45 in JNK-dependent apoptosis and regeneration

Transcriptomic analyses of regenerating wing discs have revealed that the stress sensor *D-GADD45* is transiently upregulated upon damage, suggesting a putative role in the initial steps of regeneration (Blanco et al. 2010; Vizcaya-Molina et al. 2018). One of the early signals activated upon damage is the JNK signaling pathway (Bosch et al. 2005; Bergantiños et al. 2010; Santabárbara-Ruiz et al. 2015), which is essential for wound healing. In mammals, GADD45 proteins have been described to interact with some members of this pathway, such as MTK1 and ASK1 (De Smaele et al. 2001; Papa et al. 2004; Miyake et al. 2007). Thus, we investigated the putative role of D-GADD45 in the activation of the JNK pathway in wing imaginal discs, as well as if its presence is necessary for the recovery upon injury.

Sustained activation of *D-GADD45* induces JNK-dependent cell death

We first determined the effect of increased *D-GADD45* expression in wing imaginal discs by employing the *Gal4/UAS* system (Brand and Perrimon, 1993) to induce the exogenous expression of *UAS-D-GADD45* from early development to L3 under the control of *hedgehog (hh)-Gal4*. As *hh* is expressed in the posterior compartment of imaginal discs, we could compare the effects on the autonomous (posterior) and non-autonomous (anterior) compartments within the same disc.

The activation of *D-GADD45* resulted in a reduction of the size of both compartments (**Figure 47 A-C**). To elucidate whether this reduction was due to an increase in cell death, we measured the apoptotic index in the posterior (GFP-positive) and anterior (GFP-negative) compartments. A clear increase in apoptosis was observed in both compartments upon ectopic expression of *D-GADD45* (**Figure 47 A-B,D**). Since sustained activation of the JNK cascade has proapoptotic effects (Lin, 2003; Liu and Lin,. 2005), we investigated whether high levels of *D-GADD45* could activate the JNK pathway. Indeed, we observed a clear increase in the activity of JNK upon increased *D-GADD45* expression, indicating that D-GADD45 can activate JNK in the wing disc (**Figure 47 E-H**)

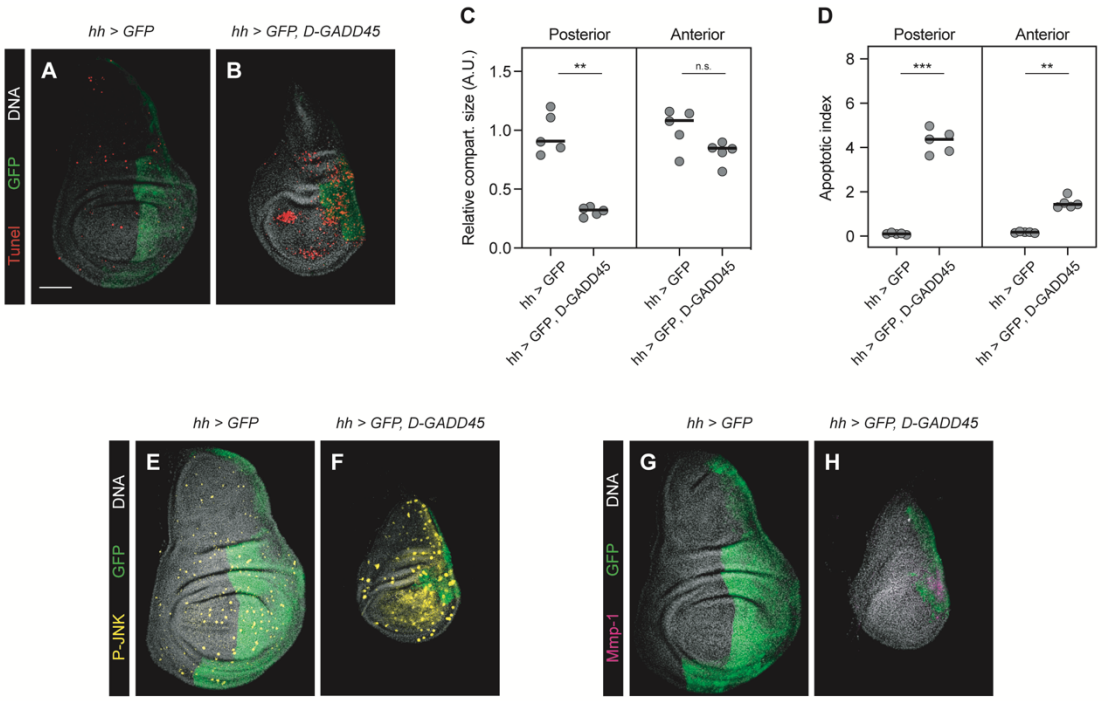


Figure 47. Ectopic expression of *D-GADD45* leads to increased apoptosis and JNK activation. (**A-B**) TUNEL assay of wing discs labelling apoptotic cells after sustained expression of each construct in the posterior (*hedgehog, hh*) compartment. Apoptotic cells (red), posterior compartment (green) and DNA (white). (**C**) Relative size area (in aribtrary units, A.U.) of disc compartments for each genotype. The median is represented for each group. (**D**) Apoptotic index for each genotype in the posterior (GFP positive) and anterior (GFP negative) compartments. The median is represented for each group. (**E-F**) Immunostaining labeling phosphorylated-JNK (marker of JNK activation) (yellow), the posterior compartment (green) and DNA (white). (**G**) Immunostaining labeling Mmp-1 (effector of JNK) (magenta), the posterior compartment (green) and DNA (white). Scale bar: 50 μm. N ≥ 5 discs for each genotype. $^{**}p < 0.01$. $^{***}p < 0.001$.

To assess whether the increased apoptosis upon *D-GADD45* overexpression was JNK-dependent, we combined the expression of *D-GADD45* with the inhibition of the JNK pathway using the dominant negative form of Basket (Bsk). Although we still observed an increase in cell death (**Figure 48 A-B,G**), apoptosis was significantly lower compared to the activation of *D-GADD45* alone (**Figure 48 G**), indicating that the sustained expression of *D-GADD45* leads to a JNK-dependent increase in apoptosis.

We next assessed whether the interaction between GADD45 and members of the JNK pathway was conserved in *Drosophila*. For this, we examined the genetic interaction between D-GADD45 and Mekk1, the *Drosophila* ortholog of mammalian MTK1, and Ask1. In these experiments, we combined the expression of *D-GADD45* and the expression of RNA interference (RNAi) constructs against either *Mekk1* or *Ask1*. Depletion of *Mekk1* or *Ask1* completely blocked apoptosis caused by *D-GADD45* overexpression (**Figure 48 C-G**), suggesting that D-GADD45 mediates the activation of the JNK signaling pathway through, at least, Mekk1 and Ask1.

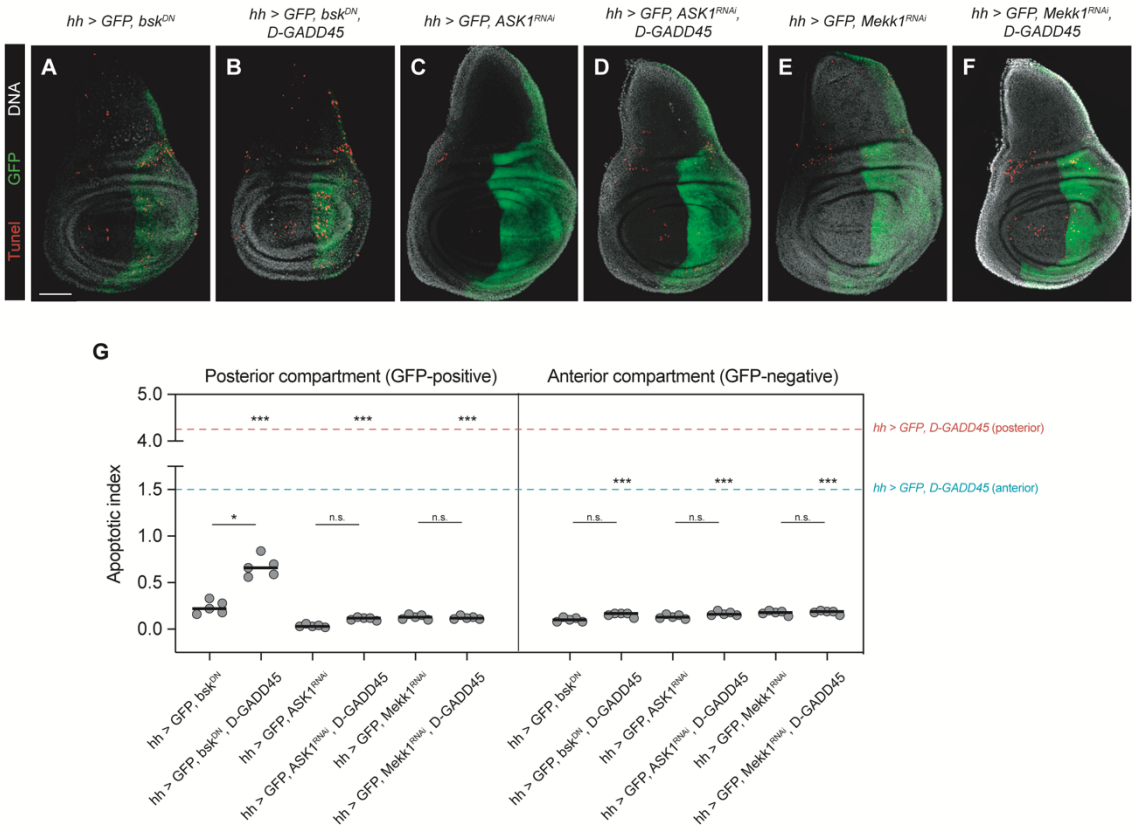


Figure 48. Sustained expression of *D-GADD45* induces JNK-dependent apoptosis. (**A-D**) TUNEL assay of wing discs labeling apoptotic cells after sustained expression of each construct in the posterior (*hedgehog, hh*) compartment. Apoptotic cells (red), posterior compartment (green) and DNA (white). Scale bar: 50 μm. (**G**) Apoptotic index for each genotype in the posterior compartment (autonomous effect) and anterior compartment (non-autonomous effect). The median is represented for each group. N = 5 discs for each genotype. $^{**}p < 0.01$. $^{***}p < 0.001$.

Next, we scored the adult wing phenotypes of flies in which *D-GADD45* was overexpressed. Consistent with the phenotypes observed in the imaginal discs, the expression of *D-GADD45* resulted in aberrant and smaller wings compared to controls (**Figure 49 A-B,I-J**), while the inhibition of the JNK cascade, either by expressing the mutant form of *Basket* or by depleting *Mekk1* or *Ask1* by RNAi, was sufficient to rescue the normal phenotypes, recovering the size and pattern of control wings (**Figure 49 C-J**).

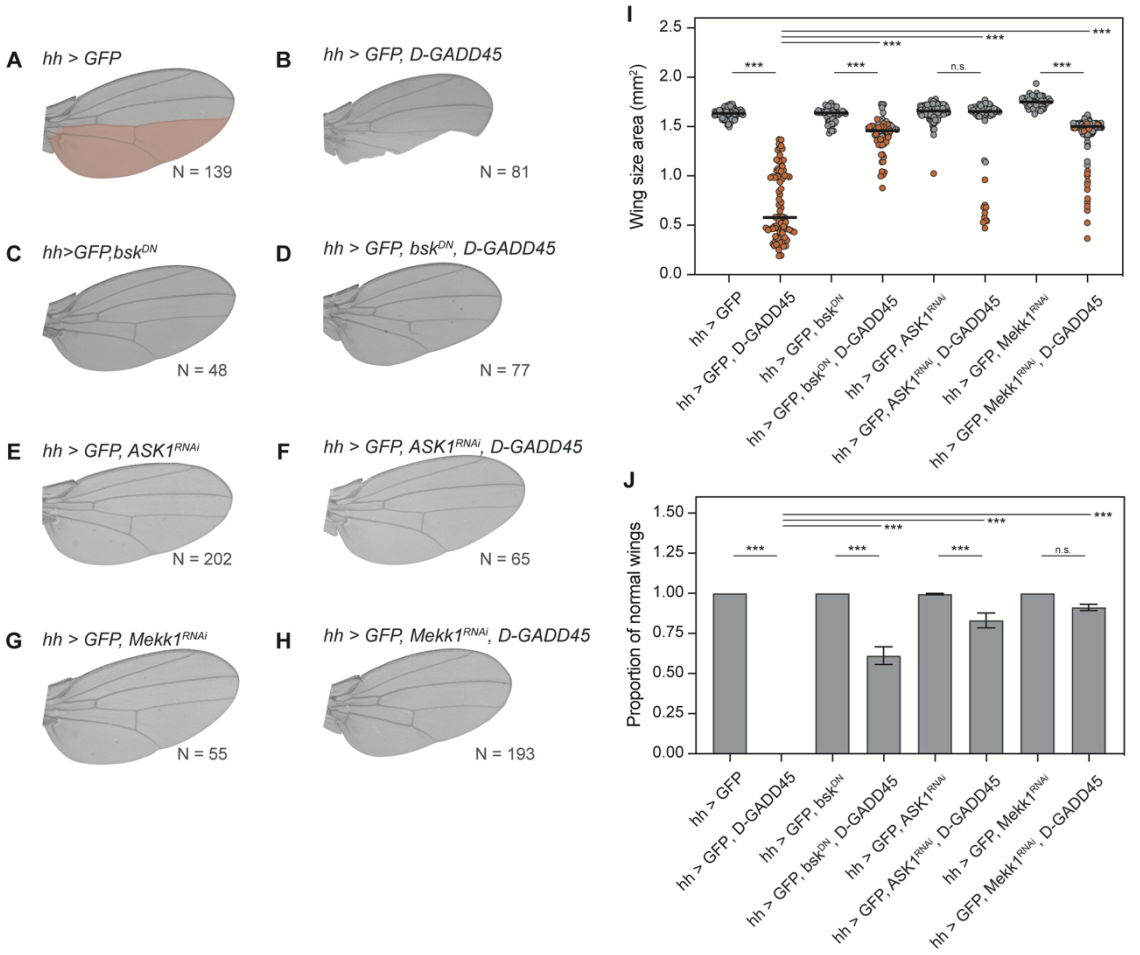


Figure 49. Sustained expression of *D-GADD45* results in aberrant, smaller adult wings. (**A-H**) Adult wings showing the predominant phenotype observed after sustained expression of each construct in the posterior compartment (marked in red). (**I**) Wing size area (mm^2) for each genotype. Each dot represents one wing: normal wings in size and pattern (grey) and aberrant wings in size or pattern (orange). Median for each condition is represented. (**J**) Proportion of normal non-aberrant wings for each genotype. Error bars represent the standard error of the proportion. ***$p < 0.001$.

Transient expression of *D-GADD45* is not sufficient to induce apoptosis

As sustained expression of *D-GADD45* induces JNK-dependent cell death, we analyzed whether its transient expression was also sufficient to induce apoptosis. For this, we transiently activated the expression of *D-GADD45* in the posterior compartment of the wing disc. After 6 and 8 h of *D-GADD45* activation, no differences were observed in terms of disc size and cell death compared to controls (**Figure 50 A-D,G-H**). However, after 11 h of induction, we observed an increase in the number of apoptotic cells in the posterior compartment (**Figure 50 E-F,G**). In line with these results, no patterning defects were observed in the adult wings upon transient expression of *D-GADD45* (**Figure 50 I-O**), only a small increase in the wing size was observed upon 6 h of *D-GADD45* induction (**Figure 50 I-O**).

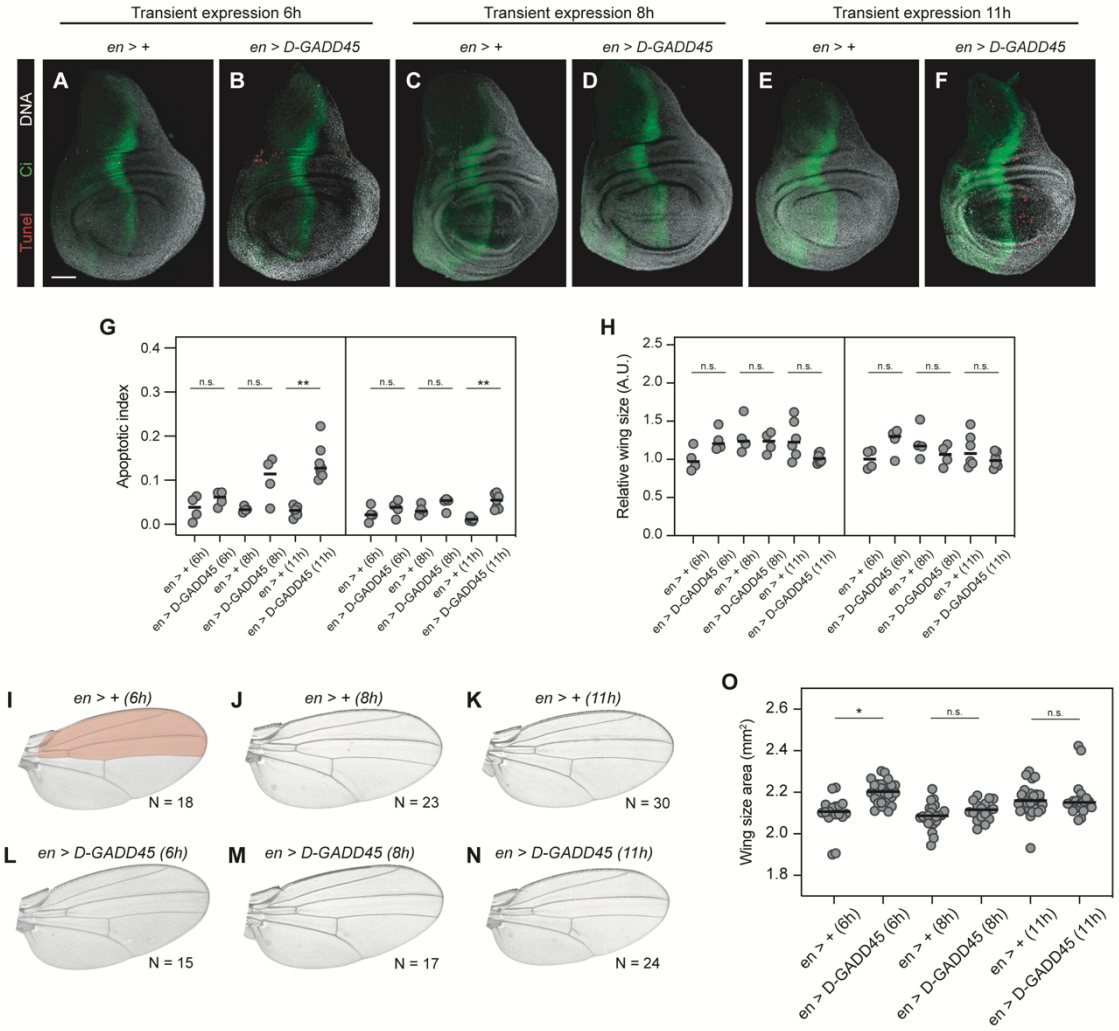


Figure 4. Transient expression of *D-GADD45* is not sufficient to induce cell death. (**A-F**) TUNEL assay of wing discs labeling apoptotic cells (red), the anterior compartment (Ci: Cubitus interruptus, green) and DNA (white). Scale bar: 50 µm. N ≥ 5 for each condition. (**G**) Apoptotic index for each genotype in the posterior (GFP negative) and anterior (GFP positive) compartments. Median is represented for each condition. (**H**) Relative size area (in arbitrary units, A.U.) of disc compartments for each genotype. Median is represented for each condition. (**I-N**) Adult wings showing the predominant phenotype observed after transient expression of each construct in the posterior (*engrailed, en*) compartment (labelled in orange). (**O**) Wing size area (mm^2) for each genotype. Each dot represents one wing. Median is represented for each condition. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

D-GADD45 expression decreases cell proliferation

To test whether higher levels of *D-GADD45* affect proliferation in fly wing discs, we overexpressed *D-GADD45* in the posterior compartment and calculated the mitotic index using PH3 (histone H3 phosphorylated) as a mitotic marker.

We found that increased *D-GADD45* expression in the posterior compartment reduced the size of the whole disc (**Figure 51 A-B,E**). Moreover, it significantly reduced mitosis in the posterior compartment, while proliferation in the anterior compartment was comparable to controls (**Figure 51 A-B,F**). To test whether this reduction in cell proliferation was dependent on the activation of JNK, we inhibited the JNK cascade by expressing the dominant negative form of Basket. No differences were observed in cell proliferation when comparing the overexpression of *D-GADD45* with or without the

inhibition of JNK (**Figure 51 C-F**), indicating that, besides apoptosis, D-GADD45 induces an autonomous decrease in proliferation independently of the activation of the JNK signaling pathway.

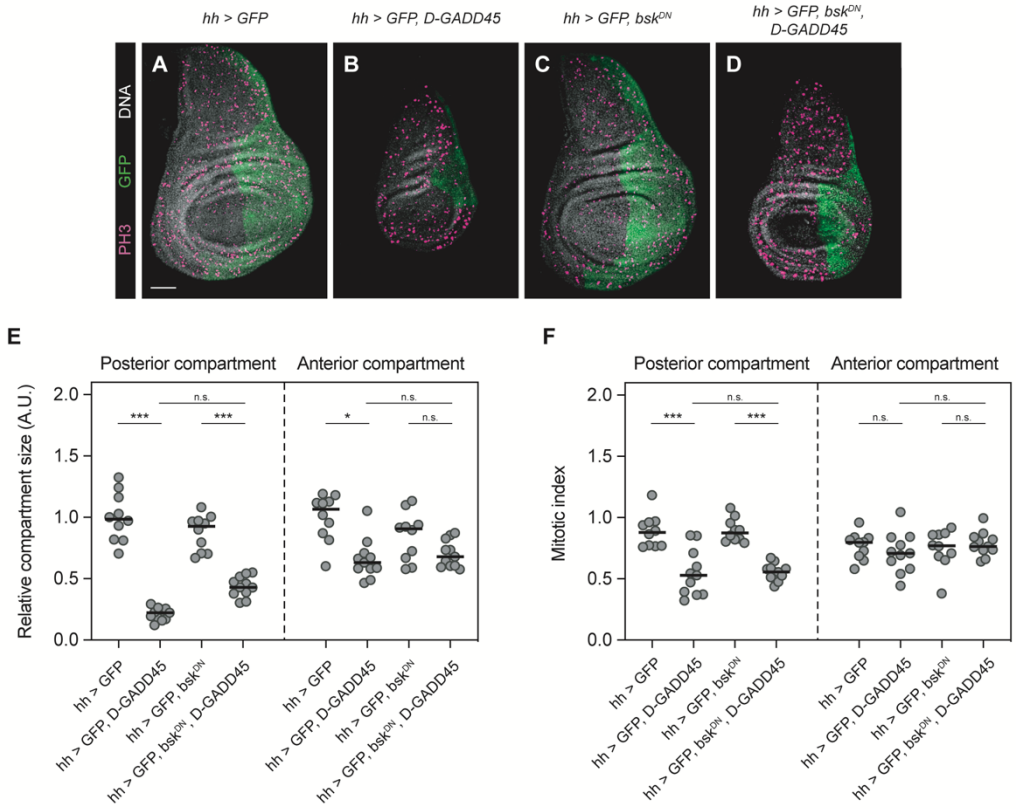


Figure 51. Sustained expression of *D-GADD45* decreases cell proliferation independently of JNK activation. (**A-D**) Immunostaining of wing discs with P-histone-H3 (PH3) labelling mitosis (magenta), the posterior compartment (green) and DNA (white). Scale bar: 50 µm. (**E**) Relative size in aribitrary units (A.U.) of the posterior and the anterior compartment of the wing discs for each genotype. The median is represented for each group. (**F**) Mitotic index for each genotype in the posterior compartment (autonomous effect) and anterior compartment (non-autonomous effect). Median is represented for each group. N ≥ 10 discs for each genotype. **p* < 0.05; ****p* < 0.001.

D-GADD45 is required for wing regeneration

Considering that *D-GADD45* is transiently upregulated upon damage (Blanco et al. 2010; Vizcaya-Molina et al. 2018), we analyzed the ability to regenerate of *D-GADD45*-depleted discs. For this, we combined the *Gal4/UAS* system to induce the expression of an RNAi against *D-GADD45* in the anterior compartment, and the *lexO/lexA* system to induce the expression of the proapoptotic gene *rpr* in the *spalt* domain of the wing disc (**Figure 52**).

Transient depletion of *D-GADD45,* without the induction of cell death, produced visible defects in wing size and patterning in around half of the analyzed wings compared to controls (**Figure 52 A-B,E-F**). Upon *rpr* induction, the inhibition of *D-GADD45* completely impaired regeneration, resulting in aberrant wings lacking the size and pattern of controls (**Figure 52 C-F**). Moreover, the presence of notches in the wing blade also indicated that the missing tissue had not been recovered, suggesting that activation of *D-GADD45* is required for wing repair after cell death. These findings show that *D-GADD45* contributes to regeneration.

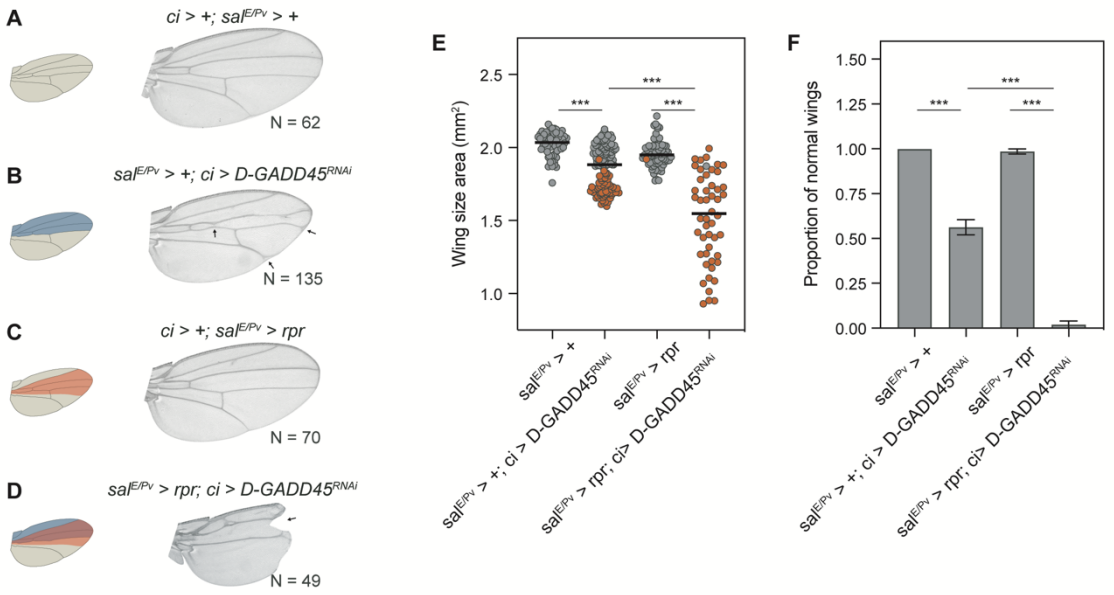


Figure 52. Depletion of *D-GADD45* severely impairs wing regeneration. (**A-D**) Adult wings showing the predominant phenotypes observed in each condition. The number of wings analyzed is indicated for each condition. (**E**) Wing size area in mm^2 of each genotype. Each dot represents one wing; normal pattern (grey) and aberrant pattern (orange). Median for each genotype is represented. (**F**) Proportion of normal wings in size and pattern for each genotype. Error bars indicate the standard error of the proportion. ***$p < 0.001$.

DISCUSSION

The results of this work contribute to the understanding of the role of lncRNAs during the development and regeneration of *Drosophila* imaginal discs. We have described the set of differentially-expressed lncRNAs throughout the wing disc development and during the recovery process after damage. Moreover, we have characterised *CR40469*, the first lncRNA whose presence is essential to maintain the regeneration capacity of wing discs. Additionally, we have investigated the putative functions of the protein-coding gene *D-GADD45*, which is transiently upregulated upon damage. We describe two distinct mechanisms of action of D-GADD45 in the wing disc, and validate that its presence upon damage is required for the recovery process.

The role of lncRNAs in wing disc development and regeneration

The relevance of lncRNAs is increasing over the last few years, as multiple mechanisms of action of lncRNAs have been extensively described (reviewed in Statello et al. 2021). Nevertheless, the function, if any, of the vast majority of lncRNAs remains to be elucidated. In flies, just few lncRNAs have been functionally characterised (reviewed in Camilleri-Robles et al. 2022). We have focused our study on the duplicated lncRNAs *CR40469* and *CR34335*, upregulated and downregulated in regeneration, respectively.

LncRNAs transcribed during the wing disc development

Using transcriptomic data from developing wing discs (Pérez-Lluch et al. 2020), we have identified ~200 lncRNAs expressed in the wing disc during development. We have observed clear expression changes across time, as only one third of all the expressed lncRNAs in the wing disc were expressed throughout the whole disc development. Despite this, we have described that around 75% of lncRNAs expressed in the wing disc are also expressed in other imaginal discs (antenna, eye and leg discs), and only ~10% (corresponding to ~20 lncRNAs) are specifically expressed in the wing disc.

With the aim to address the putative function of the lncRNAs expressed in the wing disc, we have analysed the expression of their overlapping or closest protein-coding genes. Changes in the expression of 26% genic lncRNAs were concordant with those of their overlapping protein-coding genes across time, while discordant expression patterns were observed for 18% of genic lncRNAs. These numbers could imply either a function of these lncRNAs in the regulation of the expression of the overlapping protein-coding genes, or simply the sharing of the same regulatory elements. On the other hand, intergenic lncRNAs showed a much lower number of concordant and discordant expression patterns, highlighting that in most cases they may not participate in the regulation of their closest protein-coding gene. In any case, among the protein-coding genes showing concordant expression with their closest lncRNA, we have found an enrichment in gene ontology terms related with wing disc morphogenesis, pattern specification or cell differentiation. Although the functionality of these lncRNAs was not addressed in this work, it is tempting to speculate about their participation in the regulation of genes involved in key biological processes taking place during the development of the wing disc. Reverse genetic assays by mutating some of these lncRNAs would be needed to address their function.

Differentially-expressed lncRNAs in wing disc regeneration

Using RNA-seq data from regenerating wing discs (Vizcaya-Molina et al. 2018), we have identified a subset of 131 lncRNAs differentially-expressed (DE) during regeneration. Opposed to the burst of active transcription observed for protein-coding genes in regeneration (Vizcaya-Molina et al. 2018), the majority of the DE lncRNAs were downregulated, particularly at mid regeneration stage. Also, in line with the high specificity of lncRNA expression, most of the 131 DE lncRNAs in regeneration are found DE only in one particular time-point.

Although the majority of the DE lncRNAs in regeneration are also expressed at some point during the development of the wing disc, we managed to identify 14 regeneration-specific lncRNAs, whose expression in the wing disc is only activated upon damage. It would be interesting to address their function in regeneration, but the rather low expression levels of many of them makes it difficult to analyse.

To characterise this set of DE lncRNAs, we have classified them into exonic, intronic and intergenic lncRNAs, depending on their genomic context. The proportions of each group for DE lncRNAs are comparable to those found for all the annotated lncRNAs. Similarly, we have not found any differences in terms of number of exons, transcript length or GC content between DE and non-DE lncRNAs. These results suggest that the set of DE lncRNAs in regeneration is a good representation of the lncRNAs annotated in *Drosophila*, as they do not show any particular feature.

It is highly probable that the number of lncRNAs DE involved in regeneration is higher, since we have not looked for novel unannotated lncRNAs. In fact, it is not rare to find novel lncRNAs from transcriptomic data. For instance, 179 novel lncRNA genes were identified from developing embryos (Schor et al. 2018). Possibly, the use of specific scripts aimed to identify novel lncRNAs could yield a number of unannotated lncRNAs expressed specifically in the wing disc upon damage.

CR40469 and *CR34335*: a duplication in the *Drosophila* genome

Among the set of 48 lncRNAs upregulated in the early stages of regeneration, we focused our study on the lincRNA *CR40469*. We reported a duplication of *CR40469*: the lncRNA *CR34335*, located 3.2 Mb downstream of the *CR40469* locus in the X chromosome.

Different mechanisms by which lncRNAs can arise have been proposed, including the duplication of existing DNA or RNA sequences, the exaptation of transposable elements, or the metamorphosis of protein-coding genes by the loss of their coding potential (Marques and Pointing, 2014; Jarroux et al. 2017). In the case of *CR40469* and *CR34335*, the most probable scenario is that one of the two genes arose as a consequence of the duplication of the other one. The huge genomic distance between both genes tempts us to speculate about the possibility that, instead of the duplication of the DNA sequence, it might be the retrotransposition of the RNA, the mechanism by which the duplication arose.

Indeed, the presence of TE sequences within lncRNAs is well described. It is estimated that around 60% of human lncRNAs contain TE sequences accounting for as much as

30% of their sequence (Kapusta et al. 2013; reviewed in Lee et al. 2019). Also, the presence of specific ancient TE sequences were shown to drive a nuclear enrichment of the lncRNAs containing them (Carlevaro-Fita et al. 2016; Lubelsky and Ulitsky, 2018). For this, we scanned the genomic sequences of both *CR40469* and *CR34335*, as well as their surrounding sequences searching for reminiscents of transposase or integrase, but we found no evidence of transposable element (TE) exaptation. However, we cannot rule out the possibility that their sequences diverged to a point where no traces of TEs can be found. Moreover, when analysing closely the sequence of *CR40469* and *CR34335*, we found a poly-A region spanning 19 nucleotides in the genomic sequence of *CR34335*, coinciding with the transcription termination site of the *CR34335* short isoform, which also marks the end of the sequence identical to *CR40469*. This poly-A sequence might be remnant of the retrotransposition of the *CR40469* RNA. In this way, we propose that the lncRNA *CR40469* might actually derive from an exaptated TE, which was retrotransposed inverted into the X chromosome, deriving into the lncRNA *CR34335* (**Figure 53**).

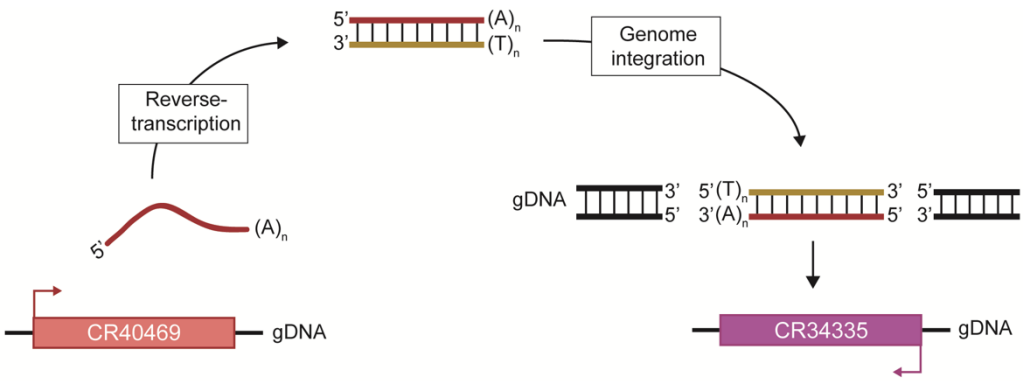


Figure 53. Proposed model of *CR40469* retrotransposition. Putative mechanism by which the duplication of *CR40469* arose. In this model, we propose that *CR40469* was retrotranscribed and then integrated in a inverted position in the genome, giving rise to the current *CR34335* gene, which is transcribed in the opposite direction.

In line with this hypothesis, the *CR40469* locus is surrounded by retrotransposons. Specifically, a TAHRE element is found merely 18 bp downstream of the *CR40469* termination site, while a HeT-A element is located immediately upstream of the *CR40469* TSS (**Figure 54**). Also, the 120 kb region separating the *CR40469* locus from the chromosome end is filled with HeT-A, TAHRE and TART elements, which are the three retrotransposons responsible for the elongation of the fly telomeres (George et al. 2006, 2010; Casacuberta, 2017). This particular genomic context suggests that the *CR40469* locus is located at the end of the sub-telomeric region of the X chromosome, strengthening the possibility that *CR40469* derived from a retrotransposon.

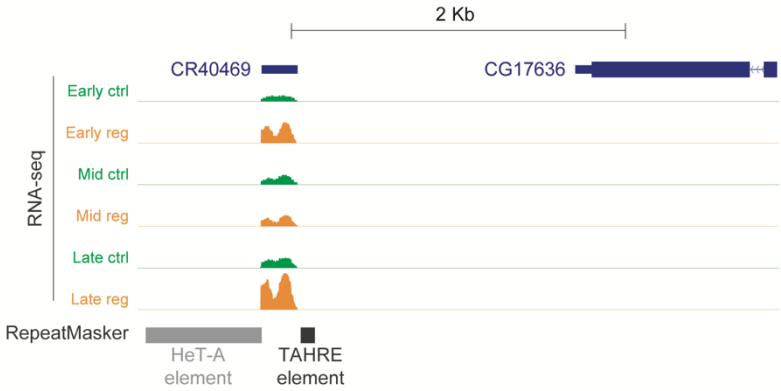


Figure 54. *CR40469* locus is surrounded by retrotransposons. Screenshot of the UCSC Genome Browser showing the repetitive sequences surrounding *CR40469*. Also, the RNA-seq track is shown, with control samples labelled in green, and regenerating samples labelled in orange.

Conservation of *CR40469* and *CR34335*

We have not been able to find any similar sequence to *CR40469* or *CR35335* in other species of *Drosophila*. Although it might imply that both lncRNAs appeared later in evolution, another possibility would be that the *CR40469* sequence diverged too much to find any sequence similarity, even in evolutionarily closer species.

Other types of conservation could also be addressed, such as the presence of micro-homologous regions (Ulitsky et al. 2011; Hezroni et al. 2015; Ruiz-Orera and Albà, 2019; Ross et al. 2021). However, the *CR40469* sequence might not be long enough (214 nt) to effectively find these small conserved motifs. On the other hand, synteny conservation could also be assessed (Bryzghalov et al. 2021; Pegueroles et al. 2019; Herrera-Ubeda et al. 2019; Rolland et al. 2019), particularly comparing *Drosophila melanogaster* with other *Drosophila* species. The major drawback in this case would be the fact that *CR40469* is located in the first position of the X chromosome and thus the synteny conservation could only be assessed by the positional conservation of the genes located downstream of *CR40469*. Alternatively, *CR34335* could be used to test the conservation by synteny, but it would only be reliable if the *CR40469* retrotransposition actually appeared in the common ancestor between the species compared.

Co-regulation of both lncRNAs

Despite being almost identical in sequence, the expression pattern of *CR40469* and *CR34335* is far from similar: while *CR40469* is lowly expressed in all tissues and developmental stages, *CR34335* is very highly expressed regardless of the tissue and stage. The minimum expression level of *CR34335* was coincident with the maximum *CR40469* expression at the L3 wing disc. Additionally, their expression in regeneration is opposed: *CR40469* is upregulated at early and late regeneration, while *CR34335* is downregulated during the whole recovery process. Also, we clearly observed a significant *CR34335* upregulation in *CR40469* homozygous mutants. For these reasons, it is tempting to speculate about a putative co-regulation of both lncRNAs.

The huge differences in the expression levels of both lncRNAs suggest that they are differentially regulated at the transcriptional level. We have not found strong differences in the TF binding sites present in the promoters of *CR40469* and *CR34335*. Regarding their chromatin accessibility, we have found that the open chromatin peaks present in

their promoter regions reflect their expression pattern: while *CR34335* peaks are slightly more accessible in control than regeneration samples, *CR40469* peaks are only accessible in regeneration. At the post-transcriptional level, we have investigated a putative regulation by miRNAs, but we did not find differences in the miRNA binding sites between both lncRNAs.

CR40469 as a putative enhancer RNA

The upregulation of *CR40469* in the early and late stages of regeneration suggests that it might participate in the recovery process after injury. A knock-out mutant was generated for *CR40469,* but we did not observe significant differences between *CR40469* homozygous mutants and control flies in several parameters, including embryo to adult viability, timing of pupariation, adult body size or development of the adult wings.

The lack of phenotypes upon lncRNA mutation is not unusual, as it is proposed that many functional lncRNAs participate in the regulation of tightly regulated biological processes, and the loss of the lncRNA is overcome by other factors, resulting in no visible phenotypes (reviewed in Gao et al. 2020). This lack of distinguishable phenotypes is actually one of the major drawbacks in the study of lncRNAs, as it renders the definition of lncRNAs as possibly non-functional.

The remarkable reduction in the ability to regenerate of *CR40469* mutants makes this gene the first lncRNA essential for the regenerative process. The analysis of the transcriptome of *CR40469* mutants compared to wild type regenerating discs, has revealed a set of 95 differentially-expressed (DE) genes, 48 upregulated and 47 downregulated. However, whether these 95 genes are directly or indirectly regulated by *CR40469* remains to be elucidated, as well as the exact mechanism by which *CR40469* mediates the upregulation or downregulation of these genes.

One possibility is that *CR40469* is acting as a regeneration-specific enhancer-derived RNA, as it actually overlaps with an emerging damage-responsive element (eDRRE), an open chromatin region only accessible upon damage and not in normally developing wing discs (Vizcaya-Molina et al. 2018). Transcription of enhancer-derived RNAs has been linked with enhancer activity (Hah et al. 2013; Mikhaylichenko et al. 2018). In line with this hypothesis, among the 95 DE genes in the *CR40469* mutant background, a clear enrichment for the X chromosome is detected, which is precisely the chromosome in which the *CR40469* locus is located. However, the closest DE gene in the X chromosome is located more than 1 Mb away from the *CR40469* locus, thus, if *CR40469* is acting as an elncRNA, it should act through large chromatin loops. Furthermore, we did not observe any particular clustering among the DE genes, as they were located throughout the whole chromosome. For this reason, in the case that *CR40469* directly participates in their regulation as an elncRNA, multiple chromatin loops involving large genomic distances should be formed to approximate the *CR40469* locus and the promoter of the DE genes. Additionally, the fact that we could not detect the transcripts of *CR40469* by FISH also points towards the possibility that it might act as an elncRNA, as they have been detected by FISH only in cultured cells, but not in tissues (Shibayama et al. 2017; Rahman et al. 2017; Tsai et al. 2018).

Further experiments are required to validate whether *CR40469* actually acts as a regeneration-specific enhancer-derived RNA. For instance, the circular chromosome conformation capture (4C) technique allows to capture all genomic interactions of a specific chromatin region (Krijger et al. 2020). Performing 4C experiments in regenerating discs would allow detecting whether the *CR40469* locus contacts with any of the DE genes present in the mutant background. In preliminary experiments using the chromatin conformation capture (3C) technique we could not detect physical interaction neither in control nor in regenerating wing discs between the *CR40469* locus and 4 selected candidate genes.

Alternatively, the putative enhancer activity of *CR40469* could also be tested. For this, the *CR40469* genomic sequence should be cloned upstream of a minimal promoter, followed by a Gal4-encoded sequence. In this way, upon the addition of a *UAS-GFP* construct, the enhancer capacity of the *CR40469* sequence in regenerating discs can be monitored: if GFP is observed, it is because the *CR40469* acts as an enhancer to drive the expression of the *Gal4* gene, which in turn activates the expression of the GFP. This approach has been used in enhancer screenings (Tokusumi et al. 2017), and to confirm enhancer activity of DRREs during regeneration (Vizcaya-Molina et al. 2018), and it can be useful also to test putative enhancer-derived RNAs. Nonetheless, this assay would only test the ability of the lncRNA genomic sequence to serve as enhancer, while the actual function of the transcript, if any, would remain to be elucidated.

Possible role of *CR40469* in responses to other types of stress

To investigate whether *CR40469* participates only in the response to cell death, or if it also participates in the response to other types of stresses, several experiments can be done with the use of mutant flies. Preliminary experiments with flies subjected to cold and oxidative stress did not show viability differences between *CR40469* mutants and control flies (**Figure 55**). However, it is important to consider that these assays involve systemic responses, as the whole organism is subjected to stress. Because of the use of specific drivers, genetically-induced cell death, on the contrary, acts locally, just in wing imaginal discs. For this reason, although these results might imply that *CR40469* does not participate in the response to cold and oxidative stress, it could be that this lncRNA is only required in imaginal discs.

Additionally, transcriptomic data from *Drosophila* adults and L3 larvae treated with a plethora of compounds, including ethanol, caffeine, rotenone or paraquat, did not reveal particular expression changes of *CR40469* (Brown et al. 2014), suggesting that it might only be activated upon cell death, or maybe that its expression is constrained to the wing imaginal discs.

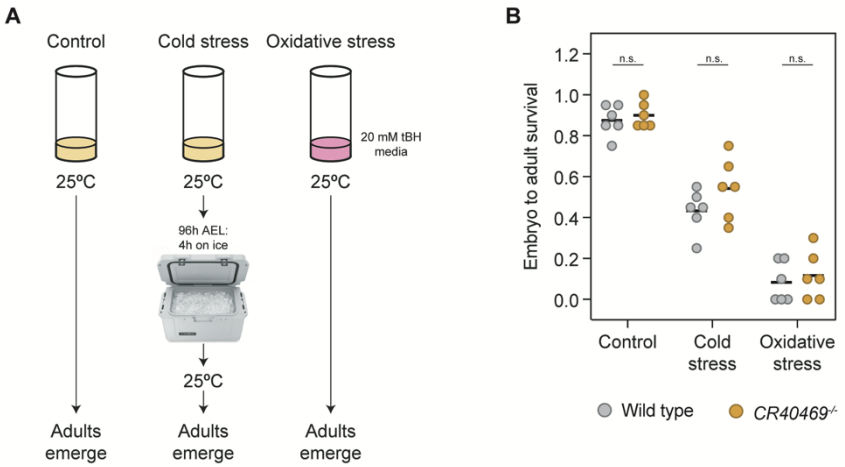


Figure 55. Effect of cold and oxidative stress on *CR40469*$^{-/-}$ mutants. (**A**) Scheme showing the induction of cold and oxidative stress. (**B**) Embryo to adult survival of wild type (grey) or *CR40469*$^{-/-}$ (orange) in control or after the induction of cold or oxidative stress. Proportion of hatched adults per each replicate and the mean of all replicates are represented.

The putative function of *CR34335*

LncRNA *CR34335* is expressed at very high levels in every sample tested. Since only genes encoding for ribosomal subunits are expressed at such high levels in *Drosophila*, we hypothesised about a putative function of *CR34335* related with the ribosomes. Analysis of several features of the putative peptide encoded by *CR34335* indicate that it is highly unlikely that this small peptide functions as a ribosomal subunit. Further experiments, such as polysome profiling to analyse whether *CR34335* transcripts are bound to the ribosomes, or the use of antibodies against the putative peptide encoded by *CR34335* to test whether it is actively translated, could shed light into a putative function of *CR34335*. Alternatively, since we could not detect any visible phenotype in *CR34335* mutants in development, even considering that its expression is ubiquitous, there is also the possibility that this lncRNA has no function at all.

In the regeneration context, *CR34335* is still expressed at very high levels, but it is significantly downregulated compared to control wing discs. Around 80% of the analysed wings were able to regenerate in the mutant, indicating that the presence of *CR34335* is probably not essential in the recovery process.

Altogether, our results suggest that, despite its high expression in every tissue and developmental stage, CR34335 is not functional, and thus dispensable during development and regeneration. The subcellular localization of CR34335 does not provide evidence of functionality either. Nuclear export seems to be the default pathway of transcribed RNAs (Palazzo and Lee, 2018; Lubelsy and Ulitsky, 2018; Shukla et al. 2018), thus the localization of *CR34335* transcripts in the cytoplasm does not imply that they have been actively exported to perform any function there. It is also possible that the high levels of *CR34335* in the cytoplasm cause some degree of cellular stress, as these transcripts possess a smORF that could potentially be translated, occupying ribosomes that otherwise would be translating other mRNAs (Palazzo and Lee, 2015).

In any case, analysis of tissue-specific and condition-specific expression patterns, as well as the production of a clean mutant (not affecting the expression of neighbouring genes) are needed to properly address the possible functionality of *CR34335*.

D-GADD45 role in JNK-dependent apoptosis and regeneration

In this work, we have described a role for the stress sensor D-GADD45 in *Drosophila*. Our results, together with previous observations (Peretz et al. 2007; Nelson et al. 2016), indicate that the levels of *D-GADD45* must be tightly controlled both, during normal development as well as after the induction of cell death.

We have also revealed the relationship between D-GADD45 and the JNK signalling pathway during development. A previous study in flies already showed a genetic interaction in the germline between D-GADD45 and the MAP Kinase Kinases (MAP2Ks) Hemipterous and Licorne, members of the JNK and p38 pathways, respectively (Peretz et al. 2007). Our data show a genetic interaction of D-GADD45 and the JNK signalling pathway at the level of MAP3Ks. In mammals, the binding of GADD45 proteins to MTK1, ortholog of *Drosophila* Mekk1, leads to the auto-phosphorylation of its kinase domain, allowing MTK1 to trigger the JNK signalling cascade (Takekawa et al. 1998; Miyake et al. 2007). Previous studies also revealed a physical interaction between ASK1 and GADD45 in human cells, although this interaction was thought to be non-functional (Papa et al. 2004). Further analyses are required to uncover whether the molecular mechanism of GADD45-mediated activation of the JNK pathway is conserved in *Drosophila*.

In mammals, *GADD45* overexpression was found to induce G2/M phase arrest in numerous cell lines (Zhan et al. 1994; Wang et al. 1999; Nakayama et al. 1999; Zhang et al. 1999; Jin et al. 2002). Here, we have shown that a sustained increase of *D-GADD45* levels results in an increase in apoptosis and a decrease in cell proliferation. Although several studies implicated JNK signalling in G2/M phase arrest (Zhu et al. 2009; Cosolo et al. 2019), we still observed a decrease in mitosis when activating *D-GADD45* and inhibiting the JNK cascade, suggesting that the activation of JNK is dispensable in *Drosophila* for the D-GADD45-mediated effect in cell proliferation. On the other hand, it was described that wing size in Drosophila is regulated by JNK signalling during development (Willsey et al. 2016). Thus, we cannot discard that a short induction of *D-GADD45* (6 h) may activate a JNK-mediated proliferative response.

One of the early responses to damage in the wing imaginal discs is the activation of the JNK signalling cascade, which is required for regeneration (Bosch et al. 2005; Mattila et al. 2005; Bergantiños et al. 2010). In this system, the induction of cell death activates p38 and induces tolerable levels of JNK, which are essential for wound healing (Santabárbara-Ruiz et al. 2015). Moreover, it was recently demonstrated the requirement for Ask1 during wing regeneration (Santabárbara-Ruiz et al. 2019). We hypothesise here that D-GADD45 could be required to regulate the activity of JNK by activating Ask1 and Mekk1. However, we cannot rule out the possibility that other members of the JNK signalling pathway could also interact with D-GADD45 during the stress response in *Drosophila*, as observed in other systems (Papa et al. 2004, 2007; Tornatore et al. 2008; Ueda et al. 2017). Similar to mammals (Salvador et al. 2013), it is likely that members of the MAPK family that are effectors of GADD45 signalling, contribute at the same time to GADD45 induction.

Finally, it is tempting to speculate about the possible mechanisms behind the tight regulation of *D-GADD45* levels during regeneration. On the one hand, *D-GADD45* could be a direct downstream target of the JNK pathway. Both p38 and JNK are required only

during the early response (Santabárbara-Ruiz et al. 2015), while the expression of *D-GADD45* shows an increase/decrease pattern during regeneration (Blanco et al. 2010; Vizcaya-Molina et al. 2018). In addition, the promoter region of *D-GADD45* contains putative binding sites for AP1 (Activator Protein-1), the transcription factor downstream of the stress-responding JNK pathway (Blanco et al. 2010). On the other hand, a degradation of *D-GADD45* mRNA by the nonsense-mediated decay (NMD) pathway was described as essential for viability in flies (Nelson et al. 2016). This observation suggests that when *D-GADD45* mRNA levels reach a certain threshold, the NMD pathway would destroy these transcripts, reducing the amount of the D-GADD45 protein to the appropriate levels to prevent D-GADD45-mediated apoptosis.

CONCLUSIONS

1. LncRNAs are not particularly expressed in the wing disc; only the expression of ~200 lncRNAs was detected, 75% of which were also found expressed in other imaginal discs.

2. 131 lncRNAs are differentially expressed in the context of regeneration, 14 of which are only expressed in the wing imaginal disc upon damage.

3. The sequence of lncRNAs *CR40469* and *CR34335* is 99.1% identical in the genome of *Drosophila*. Upon damage, *CR40469* is upregulated while *CR34335* is downregulated.

4. LncRNA *CR40469* is dispensable for fly development, but it becomes essential upon damage, as knock-out mutants lose the regeneration capacity of wing imaginal discs.

5. LncRNA *CR34335* is expressed at unusually high levels in most tissues and developmental stages and is particularly enriched in the cytoplasm.

6. Sustained activation of *D-GADD45* leads to a JNK-dependent increase in apoptosis that is mediated, at least, by Mekk1 and Ask1. On the contrary, transient expression of *D-GADD45* is not sufficient to induce cell death.

7. Sustained activation of *D-GADD45* decreases cell proliferation independently of the activation of the JNK signaling pathway.

8. Activation of *D-GADD45* upon damage is essential to maintain the regenerative capacity of wing imaginal discs.

BIBLIOGRAPHY

Agata, K., Tanaka, T., Kobayashi, C., Kato, K. & Saitoh, Y. Intercalary regeneration in planarians. *Dev. Dyn.* **226**, 308–316 (2003).

Agata, K., Saito, Y. & Nakajima, E. Unifying principles of regeneration I: Epimorphosis versus morphallaxis. *Dev. Growth Differ.* **49**, 73–78 (2007).

Akef, A., Zhang, H., Masuda, S. & Palazzo, A. F. Trafficking of mRNAs containing ALREXpromoting elements through nuclear speckles. *Nucleus* **4**, 326–340 (2013).

Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).

Andrews, S. J. & Rothnagel, J. A. Emerging evidence for functional peptides encoded by short open reading frames. *Nature Reviews Genetics* **15**, 193–204 (2014).

Arnold, P. R., Wells, A. D. & Li, X. C. Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Front. Cell Dev. Biol.* **7**, 1–14 (2020).

Aspden, J. L. *et al.* Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *Elife* **3**, 1–19 (2014).

Baguna, J., Salo, E. & Auladell, C. Regeneration and pattern formation in planarians. III. Evidence that neoblasts are totipotent stem cells and the source of blastema cells. *Development* **107**, 77–86 (1989).

Baguñà, J. Mitosis in the intact and regenerating planarian Dugesia mediterranea n.sp. II. Mitotic studies during regeneration, and a possible mechanism of blastema formation. *J. Exp. Zool.* **195**, 65–79 (1976).

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–10041 (2001).

Bangi, E. *et al.* A Drosophila platform identifies a novel, personalized therapy for a patient with adenoid cystic carcinoma. *iScience* **24**, (2021).

Barrio, R. & De Celis, J. F. Regulation of spalt expression in the Drosophila wing blade in response to the Decapentaplegic signaling pathway. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6021–6026 (2004).

Bartel, D. P. Metazoan MicroRNAs. *Cell* **173**, 20–51 (2018).

Beckmann, K., Grskovic, M., Gebauer, F. & Hentze, M. W. A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in Drosophila. *Cell* **122**, 529–540 (2005).

Beira, J. V. & Paro, R. The legacy of Drosophila imaginal discs. *Chromosoma* **125**, 573–592 (2016).

Bely, A. E. & Nyberg, K. G. Evolution of animal regeneration: re-emergence of a field. *Trends Ecol. Evol.* **25**, 161–170 (2010).

Bender, W. MicroRNAs in the Drosophila bithorax complex. *Genes Dev.* **22**, 14–19 (2008).

Bergantinos, C., Corominas, M. & Serras, F. Cell death-induced regeneration in wing imaginal discs requires JNK signalling. *Development* **137**, 1169–1179 (2010).

Bergantiños, C., Vilana, X., Corominas, M. & Serras, F. Imaginal discs: Renaissance of a model for regenerative biology. *BioEssays* **32**, 207–217 (2010).

Bieli, D. *et al.* Establishment of a Developmental Compartment Requires Interactions between Three Synergistic Cis-regulatory Modules. *PLoS Genet.* **11**, 1–30 (2015).

Blanco, E. *et al.* Gene expression following induction of regeneration in Drosophila wing imaginal discs. Expression profile of regenerating wing discs. *BMC Dev. Biol.* **10**, 1–14 (2010).

Bone, J. R. *et al.* Acetylated histone H4 on the male X chromosome is associated with dosage compensation in Drosophila. *Genes Dev.* **8**, 96–104 (1994).

Boone, E., Colombani, J., Andersen, D. S. & Leópold, P. The Hippo signalling pathway coordinates organ growth and limits developmental variability by controlling dilp8 expression. *Nat. Commun.* **7**, 1–8 (2016).

Bosch, M., Serras, F., Martín-Blanco, E. & Baguñà, J. JNK signaling pathway required for wound healing in regenerating Drosophila wing imaginal discs. *Dev. Biol.* **280**, 73–86 (2005).

Bosch, T. C. G. Why polyps regenerate and we don't: Towards a cellular and molecular framework for Hydra regeneration. *Dev. Biol.* **303**, 421–433 (2007).

Bouvrette, L. P. B. *et al.* CeFra-seq reveals broad asymmetric mRNA and noncoding RNA distribution profiles in Drosophila and human cells. *Rna* **24**, 98–113 (2018).

Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–15 (1993).

Brockdorff, N. *et al. The Product of the Mouse Xist Gene Is a 15 kb Inactive X-Specific Transcript Containing No Consewed ORF and Located in the Nucleus*. *Cell* **71**, (1992).

Brockes, J. P. Amphibian limb regeneration: Rebuilding a complex structure. *Science.* **276**, 81–87 (1997).

Brown, C. J. *et al. The Human X/ST Gene: Analysis of a 17 kb Inactive X-Specific RNA That Contains Conserved Repeats and Is Highly Localized within the Nucleus*. *Cell* **71**, (1992).

Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44 (1991).

Brown, J. B. *et al.* Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**, 393–399 (2014).

Bryzghalov, O., Makałowska, I. & Szcześniak, M. W. lncEvo: automated identification and conservation study of long noncoding RNAs. *BMC Bioinformatics* **22**, (2021).

Cai, X. & Cullen, B. R. The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA* **13**, 313–316 (2007).

Camilleri-Robles, C. *et al.* Genomic and functional conservation of lncRNAs: lessons from flies. *Mamm. Genome* (2022). doi:10.1007/s00335-021-09939-4

Carlevaro-Fita, J. & Johnson, R. Global Positioning System: Understanding Long Noncoding RNAs through Subcellular Localization. *Mol. Cell* **73**, 869–883 (2019).

Carlevaro-Fita, J. *et al.* Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. *Genome Res.* **29**, 208–222 (2019).

Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L. A. & Johnson, R. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *Rna* **22**, 867–882 (2016).

Carrier, F. *et al.* Gadd45, a p53-responsive stress protein, modifies DNA accessibility on damaged chromatin. *Mol. Cell. Biol.* **19**, 1673–85 (1999).

Carrieri, C. *et al.* Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* **491**, 454–457 (2012).

Casacuberta, E. Drosophila: Retrotransposons making up telomeres. *Viruses* **9**, (2017).

Celik, A., Kervestin, S. & Jacobson, A. NMD: At the crossroads between translation termination and ribosome recycling. *Biochimie* **114**, 2–9 (2015).

Chen, R. A. J. *et al.* The landscape of RNA polymerase II transcription initiation in C. elegans reveals promoter and enhancer architectures. *Genome Res.* **23**, 1339–1347 (2013).

Cohen, B., Simcox, A. A. & Cohen, S. M. Allocation of the thoracic imaginal primordia in the Drosophila embryo. *Development* **117**, 597–608 (1993).

Colombani, J., Andersen, D. S. & Léopol, P. Secreted peptide dilp8 coordinates Drosophila tissue growth with developmental timing. *Science.* **336**, 582–585 (2012).

Corley, M., Burns, M. C. & Yeo, G. W. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Mol. Cell* **78**, 9–29 (2020).

Cosolo, A. *et al.* JNK-dependent cell cycle stalling in G2 promotes survival and senescence-like phenotypes in tissue stress. *Elife* **8**, 1–27 (2019).

Couso, J. P. & Patraquim, P. Classification and function of small open reading frames. *Nature Reviews Molecular Cell Biology* **18**, 575–589 (2017).

de Goede, O. M. *et al.* Population-scale tissue transcriptomics maps long non-coding RNAs to complex disease. *Cell* **184**, 2633-2648.e19 (2021).

de Lara, J. C. F., Arzate-Mejía, R. G. & Recillas-Targa, F. Enhancer RNAs: Insights Into Their Biological Role. *Epigenetics Insights* **12**, (2019).

de Santa, F. *et al.* A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, (2010).

Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 97–102 (2009).

Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).

Dias, A. P., Dufu, K., Lei, H. & Reed, R. A role for TREX components in the release of spliced mRNA from nuclear speckle domains. *Nat. Commun.* **1**, (2010).

Dimitrova, N. *et al.* LincRNA-p21 Activates p21 In cis to Promote Polycomb Target Gene Expression and to Enforce the G1/S Checkpoint. *Mol. Cell* **54**, 777–790 (2014).

Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).

Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

Dong, Y., Yoshitomi, T., Hu, J. F. & Cui, J. Long noncoding RNAs coordinate functions between mitochondria and the nucleus. *Epigenetics and Chromatin* **10**, 1–11 (2017).

E, D. S. *et al.* Induction of gadd45beta by NF-kappaB downregulates pro-apoptotic JNK signalling. *Nature* **414**, 308–313 (2001).

Edgar, B. A. How flies get their size: Genetics meets physiology. *Nat. Rev. Genet.* **7**, 907–916 (2006).

Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science.* **341**, 1–9 (2013).

Erkmann, J. A., Sànchez, R., Treichel, N., Marzluff, W. F. & Kutay, U. Nuclear export of metazoan replication-dependent histone mRNAs is dependent on RNA length and is mediated by TAP. *Rna* **11**, 45–58 (2005).

Fan, Y. & Bergmann, A. Apoptosis-induced compensatory proliferation. The Cell is dead. Long live the Cell! *Trends Cell Biol.* **18**, 467–473 (2008).

Fogarty, C. E. *et al.* Extracellular Reactive Oxygen Species Drive Apoptosis-Induced Proliferation via Drosophila Macrophages. *Curr. Biol.* **26**, 575–584 (2016).

Forman, J. J., Legesse-Miller, A. & Coller, H. A. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14879–14884 (2008).

Fornace, A. J., Alamo, I. & Hollander, M. C. DNA damage-inducible transcripts in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8800–4 (1988).

Fornace, A. J. *et al.* Mammalian genes coordinately regulated by growth arrest signals and DNA-damaging agents. *Mol. Cell. Biol.* **9**, 4196–4203 (1989).

French, V. Leg regeneration in the cockroach, Blattella germanica - I. Regeneration from a congruent tibial graft/host junction. *Wilhelm Roux's Arch. Dev. Biol.* **179**, 57–76 (1976).

Fristrom, D. *et al.* Blistered: a gene required for vein/intervein formation in wings of Drosophila. *Development* **120**, 2661–2671 (1994).

Fristrom, D. & Fristrom, J. W. The metamorphic development of the adult epidermis. *Bate, Martinez Arias*, 843-897 (1993).

Fristrom, J. W., Logan, W. R. & Murphy, C. The synthetic and minimal culture requirements for evagination of imaginal discs of Drosophila melanogaster in vitro. *Dev. Biol.* **33**, 441–456 (1973).

Galindo, M. I., Pueyo, J. I., Fouix, S., Bishop, S. A. & Couso, J. P. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.* **5**, 1052–1062 (2007).

Galliot, B., Miljkovic-Licina, M., de Rosa, R. & Chera, S. Hydra, a niche for cell and developmental plasticity. *Semin. Cell Dev. Biol.* **17**, 492–502 (2006).

Gao, F., Cai, Y., Kapranov, P. & Xu, D. Reverse-genetics studies of lncRNAs-what we have learnt and paths forward. *Genome Biology* **21**, (2020).

Garaulet, D. L. *et al.* Homeotic Function of Drosophila Bithorax-Complex miRNAs Mediates Fertility by Restricting Multiple Hox Genes and TALE Cofactors in the CNS. *Dev. Cell* **29**, 635–648 (2014).

García-Alcalde, F. *et al.* Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics* **28**, 2678–2679 (2012).

Garcia-Bellido, A. & Merriam, J. R. Parameters of the wing imaginal disc development ofDrosophila melanogaster. *Dev. Biol.* **24**, 61–87 (1971).

Garcia-Bellido, A., Ripoll, P. & Morata, G. Developmental Compartimentalisation of the Wing Disk of Drosophila. *Nat. New Biol.* **245**, 251–253 (1973).

Gebauer, F., Merendino, L., Hentze, M. W. & Valcárcel, J. The Drosophila splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. *RNA* **4**, 142–150 (1998).

Geisler, S. J. & Paro, R. Trithorax and polycomb group-dependent regulation: A tale of opposing activities. *Development* **142**, 2876–2887 (2015).

Gelbart, M. E., Larschan, E., Peng, S., Park, P. J. & Kuroda, M. I. Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nat. Struct. Mol. Biol.* **16**, 825–832 (2009).

George, J. A., DeBaryshe, P. G., Traverse, K. L., Celniker, S. E. & Pardue, M. Lou. Genomic organization of the Drosophila telomere retrotransposable elements. *Genome Res.* **16**, 1231–1240 (2006).

George, J. A., Traverse, K. L., DeBaryshe, P. G., Kelley, K. J. & Pardue, M. Lou. Evolution of diverse mechanisms for protecting chromosome ends by Drosophila TART telomere retrotransposons. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21052–21057 (2010).

Gezer, U., Özgür, E., Cetinkaya, M., Isin, M. & Dalay, N. Long non-coding RNAs with low expression levels in cells are enriched in secreted exosomes. *Cell Biol. Int.* **38**, 1076–1079 (2014).

Gibson, M. C. & Schubiger, G. Hedgehog is required for activation of engrailed during regeneration of fragmented Drosophila imaginal discs. *Development* **126**, 1591–1599 (1999).

Gil, N. & Ulitsky, I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* **21**, 102–117 (2020).

Goldman, J. A. *et al.* Resolving Heart Regeneration by Replacement Histone Profiling. *Dev. Cell* **40**, 392-404.e5 (2017).

Gómez-Skarmeta, J. L. & Modolell, J. araucan and caupolican provide a link between compartment subdivisions and patterning of sensory organs and veins in the Drosophila wing. *Genes Dev.* **10**, 2935–2945 (1996).

Graf, J. & Kretz, M. From structure to function: Route to understanding lncRNA mechanism. *BioEssays* **42**, (2020).

Graveley, B. R. *et al.* The developmental transcriptome of Drosophila melanogaster. *Nature* **471**, 473–479 (2011).

Gumireddy, K. *et al.* Identification of a long non-coding RNA-associated RNP complex regulating metastasis at the translational step. *EMBO J.* **32**, 2672–2684 (2013).

Gummalla, M. *et al.* Abd-A regulation by the iab-8 noncoding RNA. *PLoS Genet.* **8**, (2012).

Gurtner, G. C., Werner, S., Barrandon, Y. & Longaker, M. T. Wound repair and regeneration. *Nature* **453**, 314–321 (2008).

Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).

Ha, H. *et al.* A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. *BMC Genomics* **15**, 1–16 (2014).

Hacisuleyman, E. *et al.* Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* **21**, 198–206 (2014).

Hadorn, E., Bertani, G. & Gallera, J. Regulative capacity and field organization of male genital discs in Drosophila melanogaster. *Roux's Arch Dev Biol* **144**, 31–70 (1949).

Hadorn, E. & Buck, D. On the differentiation of transplanted wing imaginal disc fragments of Drosophila melanogaster. *Rev. Suisse Zool* **69**, 302-310 (1962).

Hah, N., Murakami, S., Nagari, A., Danko, C. G. & Lee Kraus, W. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res.* **23**, 1210–1223 (2013).

Hales, K. G., Korey, C. A., Larracuente, A. M. & Roberts, D. M. Genetics on the fly: A primer on the drosophila model system. *Genetics* **201**, 815–842 (2015).

Halme, A., Cheng, M. & Hariharan, I. K. Retinoids Regulate a Developmental Checkpoint for Tissue Regeneration in Drosophila. *Curr. Biol.* **20**, 458–463 (2010).

Handke, B., Szabad, J., Lidsky, P. V., Hafen, E. & Lehner, C. F. Towards long term cultivation of Drosophila wing imaginal discs in vitro. *PLoS One* **9**, (2014).

Hansji, H. *et al.* ZFAS1: a long noncoding RNA associated with ribosomes in breast cancer cells. *Biol. Direct* **11**, 62 (2016).

Harris, R. E., Setiawan, L., Saul, J. & Hariharan, I. K. Localized epigenetic silencing of a damage-activated WNT enhancer limits regeneration in mature Drosophila imaginal discs. *Elife* **5**, 1–28 (2016).

Haynie, J. L. & Bryant, P. J. The effects of X-rays on the proliferation dynamics of cells in the imaginal wing disc of Drosophila melanogaster. *Wilhelm Roux's Arch. Dev. Biol.* **183**, 85–100 (1977).

Henriques, T. *et al.* Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev.* **32**, 26–41 (2018).

Henry, J. J. & Tsonis, P. A. Molecular and cellular aspects of amphibian lens regeneration. *Prog. Retin. Eye Res.* **29**, 543–555 (2010).

Heredia, F. *et al.* The steroid-hormone ecdysone coordinates parallel pupariation neuromotor and morphogenetic subprograms via epidermis-to-neuron Dilp8-Lgr3 signal induction. *Nat. Commun.* **12**, 1–20 (2021).

Herranz, H., Pérez, L., Martín, F. A. & Milán, M. A Wingless and Notch double-repression mechanism regulates G1-S transition in the Drosophila wing. *EMBO J.* **27**, 1633–1645 (2008).

Herrera-Úbeda, C. *et al.* Microsyntenic clusters reveal conservation of lncRNAs in chordates despite absence of sequence conservation. *Biology (Basel).* **8**, (2019).

Herrera, S. C. & Morata, G. Transgressions of compartment boundaries and cell reprogramming during regeneration in Drosophila. *Elife* **3**, 1–15 (2014).

Herrera, S. C., Martín, R. & Morata, G. Tissue Homeostasis in the Wing Disc of Drosophila melanogaster: Immediate Response to Massive Damage during Development. *PLoS Genet.* **9**, (2013).

Herzog, V. A. *et al.* A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element. *Nat. Genet.* **46**, 973–981 (2014).

Hezroni, H. *et al.* Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Rep.* **11**, 1110–1122 (2015).

Hofacker, I. L. *et al.* Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie Chem. Mon.* **125**, 167–188 (1994).

Hotamisligil, G. S. & Davis, R. J. Cell Signaling and Stress Responses. *Cold Spring Harb. Perspect. Biol.* **8**, 1–20 (2016).

Huang, C., Liang, D., Tatomer, D. C. & Wilusz, J. E. A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs. *Genes Dev.* **32**, 639–644 (2018).

Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).

Hutchinson, J. N. *et al.* A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 1–16 (2007).

Iismaa, S. E. *et al.* Comparative regenerative mechanisms across different mammalian tissues. *npj Regen. Med.* **3**, 1–20 (2018).

Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).

Ivaldi, M. S. *et al. Fetal g-globin genes are regulated by the BGLT3 long noncoding RNA locus*. *Blood* **132**, (2018).

Jandura, A., Hu, J., Wilk, R. & Krause, H. M. High resolution fluorescent in situ hybridization in Drosophila embryos and tissues using tyramide signal amplification. *J. Vis. Exp.* **2017**, 1–10 (2017).

Jarroux, J., Morillon, A. & Pinskaya, M. *Long Non-Coding RNA Biology Chapter 1 History, Discovery, and Classification of lncRNAs*. *Advances in experimental medicine and biology* **1008**, (2017).

Jiang, H., Grenley, M. O., Bravo, M. J., Blumhagen, R. Z. & Edgar, B. A. EGFR/Ras/MAPK signaling mediates adult midgut epithelial homeostasis and regeneration in drosophila. *Cell Stem Cell* **8**, 84–95 (2011).

Jin, S. *et al.* GADD45-induced cell cycle G2-M arrest associates with altered subcellular distribution of cyclin B1 and is independent of p38 kinase activity. *Oncogene* **21**, 8696–8704 (2002).

Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16**, 421–433 (2015).

Jopling, C., Boue, S. & Belmonte, J. C. I. Dedifferentiation, transdifferentiation and reprogramming: Three routes to regeneration. *Nat. Rev. Mol. Cell Biol.* **12**, 79–89 (2011).

Kadener, S. *et al.* Genome-wide identification of targets of the drosha-pasha/DGCR8 complex. *Rna* **15**, 537–545 (2009).

Kallen, A. N. *et al.* The Imprinted H19 LncRNA Antagonizes Let-7 MicroRNAs. *Mol. Cell* **52**, 101–112 (2013).

Kapusta, A. *et al.* Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *PLoS Genet.* **9**, (2013).

Katayama, S. *et al.* Antisense Transcription in the Mammalian Transcriptome. *Science.* **309**, 1564–1566 (2005).

Katsuyama, T., Comoglio, F., Seimiya, M., Cabuy, E. & Paro, R. During Drosophila disc regeneration, JAK/STAT coordinates cell proliferation with Dilp8-mediated developmental delay. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E2327–E2336 (2015).

Kelley, D. & Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* **13**, R107 (2012).

Keniry, A. *et al.* The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat. Cell Biol.* **14**, 659–665 (2012).

Kikuchi, K. *et al.* Primary contribution to zebrafish heart regeneration by gata4+ cardiomyocytes. *Nature* **464**, 601–605 (2010).

Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).

Klein, D. J., Moore, P. B. & Steitz, T. A. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.* **340**, 141–177 (2004).

Knopf, F. *et al.* Bone regenerates via dedifferentiation of osteoblasts in the zebrafish fin. *Dev. Cell* **20**, 713–724 (2011).

Krijger, P. H. L., Geeven, G., Bianchi, V., Hilvering, C. R. E. & de Laat, W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **170**, 17-32 (2020).

Kuo, C. C. *et al.* Detection of RNA-DNA binding sites in long noncoding RNAs. *Nucleic Acids Res.* **47**, (2019).

Kurihara, Y. *et al.* Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2453–2458 (2009).

Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013).

Lam, M. T. Y. *et al.* Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498**, 511–515 (2013).

Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, (2008).

Laube, F., Heister, M., Scholz, C., Borchardt, T. & Braun, T. Re-programming of newt cardiomyocytes is induced by tissue regeneration. *J. Cell Sci.* **119**, 4719–4729 (2006).

Lee, H., Zhang, Z. & Krause, H. M. Long Noncoding RNAs and Repetitive Elements: Junk or Intimate Evolutionary Partners? *Trends in Genetics* **35**, 892–902 (2019).

Lee, J. T. & Strauss, W. M. *A 450 kb Transgene Displays Properties of the Mammalian X-Inactivation Center*. *Cell* **86**, (1996).

Lee, S. *et al.* Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* **164**, 69–80 (2016).

Levin, M. Bioelectric mechanisms in regeneration: Unique aspects and future perspectives. *Semin. Cell Dev. Biol.* **20**, 543–556 (2009).

Li, L. *et al.* PTEN in neural precursor cells: Regulation of migration, apoptosis, and proliferation. *Mol. Cell. Neurosci.* **20**, 21–29 (2002).

Li, Y., Syed, J. & Sugiyama, H. RNA-DNA Triplex Formation by Long Noncoding RNAs. *Cell Chem. Biol.* **23**, 1325–1333 (2016).

Liebermann, D. a & Hoffman, B. Gadd45 Stress Sensor Genes. *Adv. Exp. Med. Biol.* 1–130 (2013). doi:10.1007/978-1-4614-8289-5

Liebermann, D. A. & Hoffman, B. Gadd45 in stress signaling. *J. Mol. Signal.* **3**, 1–8 (2008).

Lin, A. *et al.* The LINK-A lncRNA interacts with PtdIns(3,4,5)P3 to hyperactivate AKT and confer resistance to AKT inhibitors. *Nat. Cell Biol.* **19**, 238–251 (2017).

Lin, A. Activation of the JNK signaling pathway: Breaking the brake on apoptosis. *BioEssays* **25**, 17–24 (2003).

Lin, T. Y. *et al.* Fibroblast dedifferentiation as a determinant of successful regeneration. *Dev. Cell* **56**, 1541-1551.e6 (2021).

Liu, J. & Lin, A. Role of JNK activation in apoptosis: A double-edged sword. *Cell Res.* **15**, 36–42 (2005).

Liu, P. Y. *et al.* The long noncoding RNA lncNB1 promotes tumorigenesis by interacting with ribosomal protein RPL35. *Nat. Commun.* **10**, 5026 (2019).

Lo, D. C., Allen, F. & Brockes, J. P. Reversal of muscle differentiation during urodele limb regeneration. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 7230–7234 (1993).

López-Luque, J. *et al.* Dissecting the role of epidermal growth factor receptor catalytic activity during liver regeneration and hepatocarcinogenesis. *Hepatology* **63**, 604–619 (2016).

Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

Lubelsky, Y. & Ulitsky, I. Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature* **555**, 107–111 (2018).

Ma, X. Y. *et al.* Malat1 as an evolutionarily conserved lncRNA, plays a positive role in regulating proliferation and maintaining undifferentiated status of early-stage hematopoietic cells. *BMC Genomics* **16**, 1–11 (2015).

Maden, M. Intercalary regeneration in the amphibian limb and the rule of distal transformation. *J. Embryol. Exp. Morphol.* **Vol. 56**, 201–209 (1980).

Maeda, R. K. *et al.* The lncRNA male-specific abdominal plays a critical role in Drosophila accessory gland development and male fertility. *PLoS Genet.* **14**, (2018).

Malakar, P. *et al.* Long noncoding RNA MALAT1 promotes hepatocellular carcinoma development by SRSF1 upregulation and mTOR activation. *Cancer Res.* **77**, 1155–1167 (2017).

Mandaravally Madhavan, M. & Schneiderman, H. A. Histological analysis of the dynamics of growth of imaginal discs and histoblast nests during the larval development of Drosophila melanogaster. *Wilhelm Roux's Arch. Dev. Biol.* **183**, 269–305 (1977).

Marques, A. C. & Ponting, C. P. Intergenic lncRNAs and the evolution of gene expression. *Curr. Opin. Genet. Dev.* **27**, 48–53 (2014).

Martín, F. A. & Morata, G. Compartments and the control of growth in the Drosophila wing imaginal disc. *Development* **133**, 4421–4426 (2006).

Massie, H. R., Williams, T. R. & Colacicco, J. R. Changes in pH with age in Drosophila and the influence of buffers on longevity. *Mech. Ageing Dev.* **16**, 221–231 (1981).

Mattila, J., Omelyanchuk, L., Kyttälä, S., Turunen, H. & Nokkala, S. Role of Jun N-terminal Kinase (JNK) signaling in the wound healing and regeneration of a Drosophila melanogaster wing imaginal disc. *Int. J. Dev. Biol.* **49**, 391–399 (2005).

McCubrey, J. A., LaHair, M. M. & Franklin, R. A. Reactive Oxygen Species-Induced Activation of the MAP Kinase Signaling Pathways. *Antioxid. Redox Signal.* **8**, 1775–1789 (2006).

Meers, M. P. *et al.* Transcription start site profiling uncovers divergent transcription and enhancer-associated RNAs in Drosophila melanogaster. *BMC Genomics* **19**, (2018).

Meller, V. H. & Rattner, B. P. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *EMBO J.* **21**, 1084–1091 (2002).

Messeguer, X. *et al.* PROMO: Detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* **18**, 333–334 (2002).

Michalopoulos, G. K. & Bhushan, B. Liver regeneration: biological and pathological mechanisms and implications. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 40–55 (2021).

Mikhaylichenko, O. *et al.* The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev.* **32**, 42–57 (2018).

Milán-Rois, P., Quan, A., Slack, F. J. & Somoza, Á. The role of lncrnas in uveal melanoma. *Cancers (Basel).* **13**, 1–14 (2021).

Miller, M. A. & Olivas, W. M. Roles of Puf proteins in mRNA degradation and translation. *Wiley Interdiscip. Rev. RNA* **2**, 471–492 (2011).

Miyake, Z., Takekawa, M., Ge, Q. & Saito, H. Activation of MTK1/MEKK4 by GADD45 through Induced N-C Dissociation and Dimerization-Mediated trans Autophosphorylation of the MTK1 Kinase Domain. *Mol. Cell. Biol.* **27**, 2765–2776 (2007).

Montagne, J. *et al.* The Drosophila Serum Response Factor gene is required for the formation of intervein tissue of the wing and is allelic to blistered. *Development* **122**, 2589–2597 (1996).

Moskalev, A. *et al.* The role of D-GADD45 in oxidative, thermal and genotoxic stress resistance. *Cell Cycle* **11**, 4222–4241 (2012).

Muller, T. L. *et al.* Regeneration in higher vertebrates: Limb buds and digit tips. *Semin. Cell Dev. Biol.* **10**, 405–413 (1999).

Murata, Y., Iwasaki, H., Sasaki, M., Inaba, K. & Okamura, Y. Phosphoinositide phosphatase activity coupled to an intrinsic voltage sensor. *Nature* **435**, 1239–1243 (2005).

Murillo-Maldonado, J. M. & Riesgo-Escovar, J. R. The various and shared roles of lncRNAs during development. *Developmental Dynamics* **248**, 1059–1069 (2019).

Nakayama, K. *et al.* A Novel Oncostatin M-inducible Gene OIG37 Forms a Gene Family with MyD118 and GADD45 and Negatively Regulates Cell Growth. *J. Biol. Chem.* **274**, 24766–24772 (1999).

Narciso, C. *et al.* Patterning of wound-induced intercellular Ca2+ flashes in a developing epithelium. *Phys. Biol.* **12**, 56005 (2015).

Nelson, B. R. *et al.* A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science.* **351**, 271–275 (2016).

Nelson, J. O., Moore, K. A., Chapin, A., Hollien, J. & Metzstein, M. M. Degradation of Gadd45 mRNA by nonsense-mediated decay is essential for viability. *Elife* **5**, 1–13 (2016).

Niehrs, C. & Schäfer, A. Active DNA demethylation by Gadd45 and DNA repair. *Trends Cell Biol.* **22**, 220–227 (2012).

Niethammer, P., Grabher, C., Look, A. T. & Mitchison, T. J. A tissue-scale gradient of hydrogen peroxide mediates rapid wound detection in zebrafish. *Nature* **459**, 996–999 (2009).

Noh, J. H. *et al.* HuR and GRSF1 modulate the nuclear export and mitochondrial localization of the lncRNA RMRP. *Genes Dev.* **30**, 1224–1239 (2016).

Onyango, P. & Feinberg, A. P. A nucleolar protein, H19 opposite tumor suppressor (HOTS), is a tumor growth inhibitor encoded by a human imprinted H19 antisense transcript. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16759–16764 (2011).

Palazzo, A. F. & Lee, E. S. Sequence determinants for nuclear retention and cytoplasmic export of mRNAs and lncRNAs. *Front. Genet.* **9**, 1–16 (2018).

Palazzo, A. F. & Lee, E. S. Non-coding RNA: What is functional and what is junk? *Front. Genet.* **5**, 1–11 (2015).

Papa, S. *et al.* Insights into the Structural Basis of the GADD45β-mediated Inactivation of the JNK Kinase, MKK7/JNKK2. *J. Biol. Chem.* **282**, 19029–19041 (2007).

Papa, S. *et al.* Gadd45β mediates the NF-κB suppression of JNK signalling by targeting MKK7/JNKK2. *Nat. Cell Biol.* **6**, 146–153 (2004).

Parrish, N., Hormozdiari, F. & Eskin, E. Assembly of non-unique insertion content using next-generation sequencing. *Bioinforma. Impact Accurate Quantif. Proteomic Genet. Anal. Res.* 21–40 (2014). doi:10.1201/b16589

Pastor-Pareja, J. C., Grawe, F., Martín-Blanco, E. & García-Bellido, A. Invasive cell behavior during Drosophila imaginal disc eversion is mediated by the JNK signaling cascade. *Dev. Cell* **7**, 387–399 (2004).

Pastor-Pareja, J. C., Ming, W. & Tian, X. An innate immune response of blood cells to tumors and tissue damage in Drosophila. *DMM Dis. Model. Mech.* **1**, 144–154 (2008).

Pegueroles, C. *et al.* Transcriptomic analyses reveal groups of co-expressed, syntenic lncRNAs in four species of the genus Caenorhabditis. *RNA Biol.* **16**, 320–329 (2019).

Pelechano, V. & Steinmetz, L. M. Gene regulation by antisense transcription. *Nat. Rev. Genet.* **14**, 880–893 (2013).

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137 (1996).

Peretz, G., Bakhrat, A. & Abdu, U. Expression of the Drosophila melanogaster GADD45 homolog (CG11086) affects egg asymmetric development that is mediated by the c-Jun N-terminal kinase pathway. *Genetics* **177**, 1691–1702 (2007).

Pérez-Lluch, S. *et al.* RESEARCH ARTICLE bsAS, an antisense long non-coding RNA, essential for correct wing development through regulation of blistered/DSRF isoform usage. *PLoS Genet.* **16**, (2020).

Pfefferli, C. & Jaźwińska, A. The art of fin regeneration in zebrafish. *Regeneration* **2**, 72–83 (2015).

Plyusnina, E. N., Shaposhnikov, M. V. & Moskalev, A. A. Increase of Drosophila melanogaster lifespan due to D-GADD45 overexpression in the nervous system. *Biogerontology* **12**, 211–226 (2011).

Ponjavic, J., Ponting, C. P. & Lunter, G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* **17**, 556–565 (2007).

Poss, K. D., Wilson, L. G. & Keating, M. T. Heart regeneration in zebrafish. *Science.* **298**, 2188–2190 (2002).

Pueyo, J. I. & Couso, J. P. Tarsal-less peptides control Notch signalling through the Shavenbaby transcription factor. *Dev. Biol.* **355**, 183–193 (2011).

Quinn, J. J. & Chang, H. Y. Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* **17**, 47–62 (2016).

R., M. *et al.* Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *Rna* **18**, 738–751 (2012).

Rahman, S. *et al.* Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription. *Nucleic Acids Res.* **45**, 3017–3030 (2017).

Razzell, W., Evans, I. R., Martin, P. & Wood, W. Calcium flashes orchestrate the wound inflammatory response through duox activation and hydrogen peroxide release. *Curr. Biol.* **23**, 424–429 (2013).

Reiter, L. T., Potocki, L., Chien, S., Gribskov, M. & Bier, E. A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster. *Genome Res.* **11**, 1114–1125 (2001).

Repiso, A., Bergantinos, C. & Serras, F. Cell fate respecification and cell division orientation drive intercalary regeneration in Drosophila wing discs. *Development* **140**, 3541–3551 (2013).

Restrepo, S. & Basler, K. Drosophila wing imaginal discs respond to mechanical injury via slow InsP 3 R-mediated intercellular calcium waves. *Nat. Commun.* **7**, 1–9 (2016).

Restrepo, S., Zartman, J. J. & Basler, K. Cultivation and live imaging of drosophila imaginal discs. *Methods Mol. Biol.* **1478**, 203–213 (2016).

Riddiford, L. M. & Truman, J. W. Hormone Receptors and the Regulation of Insect Metamorphosis. *Amer. Zool.* **33**, 340–347 (1993).

Ringrose, L. Noncoding RNAs in Polycomb and Trithorax Regulation: A Quantitative Perspective. *Annu. Rev. Genet.* (2017). doi:10.1146/annurev-genet-120116

Rinn, J. L. *et al.* Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* **129**, 1311–1323 (2007).

Ríos-Barrera, L. D., Gutiérrez-Pérez, I., Domínguez, M. & Riesgo-Escovar, J. R. acal is a Long Non-coding RNA in JNK Signaling in Epithelial Shape Changes during Drosophila Dorsal Closure. *PLoS Genet.* **11**, 1–27 (2015).

Ripin, N. *et al.* Molecular basis for AU-rich element recognition and dimerization by the HuR C-terminal RRM. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 2935–2944 (2019).

Roch, F., Baonza, A., Martin-Blanco, E. & Garcia-Bellido, A. Genetic interactions and cell behaviour in blistered mutants during proliferation and differentiation of the Drosophila wing. *Development* **125**, 1823–1832 (1998).

Rolland, A. D. *et al.* RNA profiling of human testicular cells identifies syntenic lncRNAs associated with spermatogenesis. *Hum. Reprod.* **34**, 1278–1290 (2019).

Romero-Barrios, N., Legascue, M. F., Benhamed, M., Ariel, F. & Crespi, M. Survey and summary Splicing regulation by long noncoding RNAs. *Nucleic Acids Res.* **46**, 2169–2184 (2018).

Ross, C. J. *et al.* Uncovering deeply conserved motif combinations in rapidly evolving noncoding sequences. *Genome Biol.* **22**, (2021).

Ruiz-Orera, J., Messeguer, X., Subirana, J. A. & Alba, M. M. Long non-coding RNAs as a source of new peptides. *Elife* **3**, 1–24 (2014).

Sánchez Alvarado, A. Planarian regeneration: Its end is its beginning. *Cell* **124**, 241–245 (2006).

Santabárbara-Ruiz, P. *et al.* Ask1 and Akt act synergistically to promote ROS-dependent regeneration in Drosophila. *PLoS Genet.* **15**, e1007926 (2019).

Santabárbara-Ruiz, P. *et al.* ROS-Induced JNK and p38 Signaling Is Required for Unpaired Cytokine Activation during Drosophila Regeneration. *PLoS Genet.* **11**, 1–26 (2015).

Schor, I. E. *et al.* Non-coding RNA Expression, Function, and Variation during Drosophila Embryogenesis. *Curr. Biol.* **28**, 3547-3561.e9 (2018).

Schubiger, G. Regeneration, duplication and transdetermination in fragments of the leg disc of Drosophila melanogaster. *Dev. Biol.* **26**, 277–295 (1971).

Schuettengruber, B., Bourbon, H. M., Di Croce, L. & Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34–57 (2017).

Setten, R. L., Chomchan, P., Epps, E. W., Burnett, J. C. & Rossi, J. J. CRED9: A differentially expressed elncRNA regulates expression of transcription factor CEBPA. *Rna* **27**, 891–906 (2021).

Shibayama, Y., Fanucchi, S. & Mhlanga, M. M. *Enhancer RNAs: Methods and Protocols*. **1468**, (Springer New York, 2017).

Shukla, C. J. *et al.* High-throughput identification of RNA nuclear enrichment sequences. *EMBO J.* **37**, 1–11 (2018).

Silva, F. M. da, Marques, A. & Chaveiro, A. Reactive Oxygen Species: A Double-Edged Sword in Reproduction. *Open Vet. Sci. J.* **4**, 127–133 (2014).

Smith-Bolton, R. K., Worley, M. I., Kanda, H. & Hariharan, I. K. Regenerative Growth in Drosophila Imaginal Discs Is Regulated by Wingless and Myc. *Dev. Cell* **16**, 797–809 (2009).

Smith, M. A., Gesell, T., Stadler, P. F. & Mattick, J. S. Widespread purifying selection on RNA structure in mammals. *Nucleic Acids Res.* **41**, 8220–8236 (2013).

Smola, M. J. & Weeks, K. M. In-cell RNA structure probing with SHAPE-MaP. *Nat. Protoc.* **13**, 1181–1195 (2018).

Somarowthu, S. *et al.* HOTAIR Forms an Intricate and Modular Secondary Structure. *Mol. Cell* **58**, 353–361 (2015).

Soshnev, A. A. *et al.* A conserved long Noncoding RNA affects sleep behavior in Drosophila. *Genetics* **189**, 455–468 (2011).

St. Pierre, S. E., Galindo, M. I., Couso, J. P. & Thor, S. Control of Drosophila imaginal disc development by rotund and roughened eye: Differentially expressed transcripts of the same gene encoding functionally distinct zinc finger proteins. *Development* **129**, 1273–1281 (2002).

Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology* **22**, 96–118 (2021).

Stramer, B. *et al.* Gene induction following wounding of wild-type versus macrophage-deficient Drosophila embryos. *EMBO Rep.* **9**, 465–471 (2008).

Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics* **18**, 207–208 (2002).

Sun, G. & Irvine, K. D. Ajuba family proteins link JNK to hippo signaling. *Sci. Signal.* **6**, 1–10 (2013).

Tanaka, E. M. & Reddien, P. W. The Cellular Basis for Animal Regeneration. *Dev. Cell* **21**, 172–185 (2011).

Tao, Y., Yan, D., Yang, Q., Zeng, R. & Wang, Y. Low K + Promotes NF-κB/DNA Binding in Neuronal Apoptosis Induced by K + Loss. *Mol. Cell. Biol.* **26**, 1038–1050 (2006).

Tichon, A. *et al.* A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat. Commun.* **7**, 1–10 (2016).

Tokusumi, T. *et al.* Screening and analysis of Janelia Flylight Project enhancer-GAl4 strains identifies multiple gene enhancers active during hematopoiesis in normal and wasp-challenged Drosophila larvae. *G3 Genes, Genomes, Genet.* **7**, 437–448 (2017).

Tornatore, L. *et al.* Gadd45β forms a Homodimeric Complex that Binds Tightly to MKK7. *J. Mol. Biol.* **378**, 97–111 (2008).

Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39**, 925–938 (2010).

Tripura, C., Chandrika, N. P., Susmitha, V. N., Noselli, S. & Shashidhara, L. Regulation and activity of JNK signaling in the wing disc peripodial membrane during adult morphogenesis in drosophila. *Int. J. Dev. Biol.* **55**, 583–590 (2011).

Tsai, M. *et al.* Modification Complexes. *Science.* **329**, 689–693 (2010).

Tsai, P. F. *et al.* A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and Regulates Transcription in Trans. *Mol. Cell* **71**, 129-141.e8 (2018).

Tsao, C. K., Ku, H. Y., Lee, Y. M., Huang, Y. F. & Sun, Y. H. Long term ex vivo culture and live imaging of Drosophila larval imaginal discs. *PLoS One* **11**, 1–19 (2016).

Tu, M. K. & Borodinsky, L. N. Spontaneous calcium transients manifest in the regenerating muscle and are necessary for skeletal muscle replenishment. *Cell Calcium* **56**, 34–41 (2014).

Ueda, T., Kohama, Y., Kuge, A., Kido, E. & Sakurai, H. GADD45 family proteins suppress JNK signaling by targeting MKK7. *Arch. Biochem. Biophys.* **635**, 1–7 (2017).

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. & Bartel, D. P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550 (2011).

Ursprung, H. Fragmentation and radiation experiments to determine the determination and fate map of the Drosophila genital disc. *Roux's Arch Dev Biol* **151**, 504–558 (1959).

van Heesch, S. *et al.* Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol.* **15**, R6 (2014).

Vizcaya-Molina, E., Klein, C. C., Serras, F. & Corominas, M. Chromatin dynamics in regeneration epithelia: Lessons from Drosophila imaginal discs. *Semin. Cell Dev. Biol.* **97**, 55–62 (2020).

Vizcaya-Molina, E. *et al.* Damage-responsive elements in Drosophila regeneration. *Genome Res.* **28**, 1852–1866 (2018).

Wagner, E. F. & Nebreda, Á. R. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat. Rev. Cancer* **9**, 537–549 (2009).

Wang, G., Shimada, E., Koehler, C. M. & Teitell, M. A. PNPASE and RNA trafficking into mitochondria. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819**, 998–1007 (2012).

Wang, X., Gorospe, M. & Holbrook, N. J. gadd45 is not required for activation of c-Jun N-terminal kinase or p38 during acute stress. *J. Biol. Chem.* **274**, 29599–29602 (1999).

Wang, X. *et al.* Lncrna malat1 promotes development of mantle cell lymphoma by associating with ezh2. *J. Transl. Med.* **14**, 1–14 (2016).

Wangler, M. F., Yamamoto, S. & Bellen, H. J. Fruit flies in biomedical research. *Genetics* **199**, 639–653 (2015).

Watson, C. N., Belli, A. & Di Pietro, V. Small non-coding RNAs: New class of biomarkers and potential therapeutic targets in neurodegenerative disease. *Front. Genet.* **10**, 1–14 (2019).

Wen, K. *et al.* Critical roles of long noncoding RNAs in Drosophila Spermatogenesis. *Genome Res.* **26**, 1233–1244 (2016).

White, K., Tahaoglu, E. & Steller, H. Cell Killing by the Drosophila Gene reaper. *Science.* **271**, 805–807 (1996).

Whyte, J. L., Smith, A. A. & Helms, J. A. Wnt signaling and injury repair. *Cold Spring Harb. Perspect. Biol.* **4**, 1–13 (2012).

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

Wilkinson, K. A. *et al.* Influence of nucleotide identity on ribose 2′-hydroxyl reactivity in RNA. *Rna* **15**, 1314–1321 (2009).

Willsey, H. R. *et al.* Localized JNK signaling regulates organ size during development. *Elife* **5**, 1–18 (2016).

Wing, J. P., Zhou, L., Schwartz, L. M. & Nambu, J. R. Erratum: Distinct cell killing properties of the Drosophila reaper, head involution defective, and grim genes (Cell Death and Differentiation (1998) 5 (930-939). *Cell Death Differ.* **6**, 212–213 (1999).

Wucher, V. *et al.* FEELnc: A tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* **45**, 1–12 (2017).

Wutz, A. & Jaenisch, R. *A Shift from Reversible to Irreversible X Inactivation Is Triggered during ES Cell Differentiation function of the Xic can be reconstituted on autosomes by the insertion of YAC transgenes in multiple copies (Lee et al*. *Lee and Jaenisch* **5**, (2000).

Yamanaka, N. *et al.* Neuroendocrine control of Drosophila larval light preference. *Science.* **341**, 1113–1116 (2013).

Yang, F., Zhang, H., Mei, Y. & Wu, M. Reciprocal Regulation of HIF-1α and LincRNA-p21 Modulates the Warburg Effect. *Mol. Cell* **53**, 88–100 (2014).

Yoo, S. K., Freisinger, C. M., LeBert, D. C. & Huttenlocher, A. Early redox, Src family kinase, and calcium signaling integrate wound responses and tissue regeneration in zebrafish. *J. Cell Biol.* **199**, 225–234 (2012).

Yoon, J. H. *et al.* LincRNA-p21 Suppresses Target mRNA Translation. *Mol. Cell* **47**, 648–655 (2012).

Young, R. S. *et al.* Identification and Properties of 1,119 Candidate LincRNA Loci in the Drosophila melanogaster Genome. *Genome Biol. Evol.* **4**, 427–442 (2012).

Zeng, C., Fukunaga, T. & Hamada, M. Identification and analysis of ribosome-associated lncRNAs using ribosome profiling data. *BMC Genomics* **19**, 1–14 (2018).

Zhan, Q. *et al.* The gadd and MyD genes define a novel set of mammalian genes encoding acidic proteins that synergistically suppress cell growth. *Mol. Cell. Biol.* **14**, 2361–2371 (1994).

Zhang, H. *et al.* A new method of RNA secondary structure prediction based on convolutional neural network and dynamic programming. *Front. Genet.* **10**, 1–12 (2019).

Zhang, W. *et al.* CR6: A third member in the MyD118 and Gadd45 gene family which functions in negative growth control. *Oncogene* **18**, 4899–4907 (1999).

Zhao, Y. *et al.* NONCODE 2016: An informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res.* **44**, D203–D208 (2016).

Zhou, X. *et al.* The RNA-binding protein SRSF1 is a key cell cycle regulator via stabilizing NEAT1 in glioma. *Int. J. Biochem. Cell Biol.* **113**, 75–86 (2019).

Zhou, Z. & Fu, X. D. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* **122**, 191–207 (2013).

Zhu, N., Shao, Y., Xu, L., Yu, L. & Sun, L. Gadd45-α and Gadd45-γ utilize p38 and JNK signaling pathways to induce cell cycle G2/M arrest in Hep-G2 hepatoma cells. *Mol. Biol. Rep.* **36**, 2075–2085 (2009).

Zou, Z. *et al.* Hyper-acidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in Escherichia coli. *J. Biotechnol.* **135**, 333–339 (2008).

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

ANNEX

ANNEX I:
Gene lists

Table 9. List of 131 differentially-expressed lncRNAs in regeneration. The status of each lncRNA at each time-point (early, mid and late) is shown, as well as their classification depending on their genomic context. NDE stands for Non Differentially-Expressed.

Gene name	Early timepoint	Mid timepoint	Late timepoint	Type
CR31514	Upregulated	NDE	NDE	exonic
CR32027	NDE	NDE	Upregulated	intronic
CR32661	NDE	NDE	Downregulated	intergenic
CR34046	Downregulated	NDE	NDE	intergenic
CR34335	Downregulated	Downregulated	Downregulated	intronic
CR40469	Upregulated	NDE	Upregulated	intergenic
CR41257	Upregulated	NDE	Upregulated	intergenic
CR42549	NDE	NDE	Downregulated	exonic
CR42862	NDE	Downregulated	NDE	exonic
CR42868	NDE	NDE	Upregulated	intergenic
CR43144	Downregulated	Downregulated	Downregulated	intergenic
CR43157	NDE	Downregulated	NDE	intergenic
CR43264	Downregulated	NDE	NDE	intronic
CR43436	NDE	Downregulated	Downregulated	intergenic
CR43461	Downregulated	Downregulated	Downregulated	intronic
CR43496	Downregulated	Downregulated	Downregulated	intronic
CR43611	Upregulated	Upregulated	Upregulated	intronic
CR43644	NDE	NDE	Upregulated	intergenic
CR43684	Upregulated	Upregulated	NDE	intronic
CR43837	Downregulated	NDE	NDE	intergenic
CR43887	Downregulated	NDE	NDE	intergenic
CR43900	Upregulated	NDE	Downregulated	exonic
CR43918	NDE	Downregulated	NDE	intronic
CR43927	NDE	Downregulated	NDE	intronic
CR43943	NDE	Downregulated	NDE	intergenic
CR43949	Upregulated	NDE	NDE	intergenic
CR43956	NDE	Downregulated	NDE	intronic
CR43963	Downregulated	NDE	NDE	exonic
CR43971	NDE	Downregulated	NDE	exonic
CR43995	Downregulated	Downregulated	Downregulated	intergenic
CR44024	Upregulated	Downregulated	NDE	intergenic
CR44029	NDE	Downregulated	Downregulated	exonic
CR44042	NDE	NDE	Downregulated	intergenic
CR44043	NDE	Upregulated	NDE	intronic
CR44078	NDE	Upregulated	NDE	intergenic
CR44119	Upregulated	NDE	NDE	exonic
CR44127	Downregulated	Downregulated	NDE	intergenic

CR44135	Upregulated	Upregulated	NDE	intergenic
CR44172	Downregulated	NDE	NDE	intergenic
CR44209	Downregulated	NDE	NDE	intergenic
CR44294	Upregulated	NDE	NDE	intergenic
CR44299	NDE	NDE	Downregulated	intronic
CR44417	NDE	Downregulated	NDE	intronic
CR44440	NDE	Downregulated	Downregulated	intergenic
CR44458	NDE	Downregulated	NDE	intergenic
CR44489	NDE	Downregulated	Downregulated	intergenic
CR44499	Downregulated	NDE	NDE	intergenic
CR44552	NDE	NDE	Downregulated	intergenic
CR44589	NDE	Downregulated	NDE	intergenic
CR44677	NDE	Downregulated	NDE	intergenic
CR44749	Downregulated	NDE	NDE	intronic
CR44761	Upregulated	NDE	NDE	exonic
CR44784	NDE	NDE	Upregulated	intergenic
CR44811	NDE	NDE	Downregulated	intronic
CR44831	NDE	Downregulated	NDE	exonic
CR44899	Upregulated	NDE	NDE	intergenic
CR44917	NDE	NDE	Upregulated	intergenic
CR44918	Upregulated	NDE	NDE	intergenic
CR44955	Upregulated	Upregulated	NDE	intergenic
CR44964	NDE	Upregulated	NDE	exonic
CR44970	Downregulated	NDE	NDE	intergenic
CR44971	Downregulated	NDE	NDE	intergenic
CR44974	Upregulated	NDE	NDE	exonic
CR44987	NDE	Downregulated	Downregulated	exonic
CR44993	Upregulated	Downregulated	NDE	intronic
CR45000	NDE	NDE	Upregulated	exonic
CR45009	Downregulated	NDE	NDE	exonic
CR45028	Upregulated	NDE	NDE	exonic
CR45036	NDE	Downregulated	NDE	exonic
CR45102	NDE	Downregulated	Downregulated	intergenic
CR45128	Downregulated	NDE	NDE	intronic
CR45171	Upregulated	Downregulated	Downregulated	exonic
CR45176	Upregulated	NDE	NDE	exonic
CR45181	NDE	Downregulated	NDE	intergenic
CR45182	Downregulated	NDE	NDE	exonic
CR45187	Downregulated	Downregulated	NDE	intergenic
CR45232	NDE	Downregulated	NDE	intergenic
CR45310	NDE	Downregulated	Downregulated	intergenic
CR45311	NDE	Downregulated	NDE	intergenic
CR45330	Downregulated	Downregulated	Downregulated	intergenic

CR45433	Upregulated	NDE	NDE	intergenic
CR45466	NDE	Downregulated	NDE	exonic
CR45473	NDE	Downregulated	Downregulated	intergenic
CR45479	Upregulated	NDE	NDE	exonic
CR45485	NDE	Downregulated	NDE	exonic
CR45501	NDE	NDE	Upregulated	exonic
CR45517	NDE	Downregulated	NDE	intergenic
CR45530	Upregulated	NDE	NDE	intergenic
CR45550	NDE	Downregulated	NDE	intergenic
CR45566	NDE	NDE	Upregulated	exonic
CR45600	NDE	Downregulated	Downregulated	exonic
CR45627	Downregulated	NDE	NDE	intergenic
CR45638	Downregulated	NDE	NDE	intergenic
CR45640	NDE	NDE	Upregulated	intergenic
CR45641	Downregulated	NDE	NDE	intergenic
CR45721	NDE	Downregulated	NDE	intergenic
CR45734	Downregulated	Downregulated	NDE	intergenic
CR45822	NDE	Downregulated	Downregulated	exonic
CR45895	NDE	Downregulated	NDE	exonic
CR45908	NDE	Downregulated	NDE	exonic
CR45924	Downregulated	NDE	NDE	exonic
CR45925	NDE	Downregulated	Upregulated	exonic
CR46006	NDE	Downregulated	NDE	intergenic
CR46044	NDE	Downregulated	NDE	exonic
CR46056	NDE	Downregulated	NDE	intronic
CR46057	NDE	NDE	Upregulated	intronic
CR46093	Upregulated	NDE	NDE	exonic
CR46123	Upregulated	Upregulated	Upregulated	intronic
CR46139	NDE	NDE	Downregulated	exonic
CR46194	Upregulated	NDE	Upregulated	exonic
CR46252	Downregulated	NDE	NDE	intergenic
CR46352	Upregulated	NDE	NDE	intergenic
CR46410	NDE	Upregulated	NDE	intronic
CR9284	NDE	NDE	Downregulated	intergenic
dntRL	NDE	NDE	Downregulated	intergenic
hpRNA:CR32205	NDE	NDE	Upregulated	intergenic
hpRNA:CR32207	Upregulated	Upregulated	Upregulated	intergenic
hpRNA:CR33940	Upregulated	Upregulated	Upregulated	exonic
Hsr-omega	NDE	Downregulated	NDE	intergenic
let7C	NDE	Downregulated	NDE	exonic
noe	NDE	Downregulated	NDE	intronic
RNaseP:RNA	NDE	NDE	Upregulated	intronic
roX1	NDE	NDE	Downregulated	exonic

roX2	Downregulated	Downregulated	Downregulated	intergenic
sisRNA:1	Downregulated	Downregulated	Upregulated	intronic
sisRNA:2	Downregulated	NDE	NDE	intronic
sisRNA:3	Downregulated	Downregulated	NDE	intronic
sisRNA:4	NDE	Downregulated	NDE	intronic
sisRNA:CR46364	Downregulated	NDE	NDE	intronic
Uhg3	NDE	NDE	Downregulated	intergenic
Uhg8	NDE	NDE	Upregulated	exonic

Table 10. List of 14 regeneration-specific lncRNAs.

Gene name	Early timepoint	Mid timepoint	Late timepoint	Type
CR31514	Upregulated	NDE	NDE	exonic
CR43611	Upregulated	Upregulated	Upregulated	intronic
CR43684	Upregulated	Upregulated	NDE	intronic
CR43949	Upregulated	NDE	NDE	intergenic
CR44043	NDE	Upregulated	NDE	intronic
CR44078	NDE	Upregulated	NDE	intergenic
CR44761	Upregulated	NDE	NDE	exonic
CR44784	NDE	NDE	Upregulated	intergenic
CR44955	Upregulated	Upregulated	NDE	intergenic
CR44974	Upregulated	NDE	NDE	exonic
CR44993	Upregulated	Downregulated	NDE	intronic
CR45925	NDE	Downregulated	Upregulated	exonic
CR46123	Upregulated	Upregulated	Upregulated	intronic
CR46352	Upregulated	NDE	NDE	intergenic

ANNEX II:
Publications

genes

MDPI

Article

Role of *D-GADD45* in JNK-Dependent Apoptosis and Regeneration in *Drosophila*

Carlos Camilleri-Robles, Florenci Serras and Montserrat Corominas *

Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia and Institut de Biomedicina (IBUB), Universitat de Barcelona, Barcelona 08028, Spain; carloscamilleri@ub.edu (C.C.-R.); fserras@ub.edu (F.S.)
* Correspondence: mcorominas@ub.edu

Received: 28 March 2019; Accepted: 16 May 2019; Published: 18 May 2019

Abstract: The GADD45 proteins are induced in response to stress and have been implicated in the regulation of several cellular functions, including DNA repair, cell cycle control, senescence, and apoptosis. In this study, we investigate the role of D-GADD45 during *Drosophila* development and regeneration of the wing imaginal discs. We find that higher expression of *D-GADD45* results in JNK-dependent apoptosis, while its temporary expression does not have harmful effects. Moreover, D-GADD45 is required for proper regeneration of wing imaginal discs. Our findings demonstrate that a tight regulation of *D-GADD45* levels is required for its correct function both, in development and during the stress response after cell death.

Keywords: GADD45; JNK; p38; development; regeneration; imaginal discs

1. Introduction

The Growth Arrest and DNA Damage-inducible 45 (GADD45) family of proteins acts as stress sensors in response to various stimuli. The first *GADD45* gene was identified in mammals on the basis of its increased expression after growth cessation signals or treatment with DNA-damaging agents [1,2]. This gene, renamed as *GADD45α*, is a member of a highly conserved family, together with *GADD45β* and *GADD45γ*. *GADD45* genes encode for small (18 kd) and highly acidic proteins that can be both nuclear and cytoplasmic [3,4]. The expression of all *GADD45* genes is induced after exposure to several genotoxic or oxidative agents, such as ultraviolet radiation, methyl methanesulfonate (MMS) or hydrogen peroxide [1] and different family members have been implicated in a variety of responses to cell injury, including cell cycle checkpoints, apoptosis, and DNA repair (reviewed in [5,6]). GADD45 is also known to participate in the regeneration of zebrafish fin and retina [7–9].

Since GADD45 proteins do not have enzymatic properties, it has been suggested that they perform their functions by physically interacting with partner proteins (reviewed in [5]). Thus, upon stress induction, GADD45 interacts with different proteins involved in the different stress responses [5,10,11]. Although GADD45 proteins show complex regulation and numerous effectors, many of the prominent roles for the GADD45 proteins are associated with signaling mediated by mitogen-activated protein kinases (MAPK) [11–16]. This association is rather complex for c-Jun N-terminal kinase (JNK) and p38 proteins, members of the MAPKs pathways, which can contribute to GADD45 induction and at the same time are effectors of GADD45 signaling [11]**.** JNK and p38 pathways are also activated upon cellular stress, initiating the signaling modules that lead to the transcription of target genes involved in growth, differentiation, survival and apoptosis (reviewed in [17]). JNK and p38 can exert antagonistic effects on cell proliferation and survival, which depend on cell type-specific differences, as well as on the intensity and duration of the signal [18].

In mammals, GADD45 proteins directly bind to and activate MTK1/MEKK4, a MAP Kinase Kinase Kinase (MAP3K) upstream of JNK and p38 [12,16]. Other studies have revealed a putative

interaction between GADD45ß, and ASK1, another MAP3K upstream of JNK and p38 [14]. However, it has been proposed that GADD45ß also interacts with MKK7, a MAP2K downstream of MTK1 and ASK1, inhibiting its kinase activity in mouse fibroblasts [14,15,19,20]. All these observations suggest that the effect of GADD45 on JNK signaling might be tissue-specific [21].

D-GADD45 is the only member of the GADD45 family found in *Drosophila*, and it contains a single exon encoding a 163-amino acid protein. *D-GADD45* expression seems strongly dependent on the inflammatory response. *D-GADD45* was found to be upregulated upon activation of the immune response, but not following different stress stimuli including genotoxic stress [22]. By using a microarray screen to compare gene expression after laser wounding, *D-GADD45* was also identified as an inflammation-associated gene [23]. The effects of inducing the expression of *D-GADD45* seem to be tissue-specific also in the fly: overexpression of *D-GADD45* in the nervous system increases the lifespan of flies [24,25]. However, increased expression in somatic cells leads to apoptosis and in the germline causes several patterning and polarity defects [22]. *D-GADD45* was also found to be strongly upregulated in imaginal discs during regeneration. The expression of *D-GADD45* is rapidly increased upon damage, but after the initial steps of the stress response, when the tissue has still not completely regenerated, the expression of *D-GADD45* is recovered to the levels observed prior to damage [23,26,27]. This suggests a putative role of D-GADD45 only in the initial steps of regeneration. Moreover, damage also activates p38 and induces tolerable levels of JNK, which are essential for wound healing [28–30].

Here, we use *Drosophila* wing imaginal discs to study the role of *D-GADD45* during development and in the activation of the JNK signaling pathway. We found that a sustained activation of *D-GADD45* leads to JNK-dependent cell death, whereas transient expression of *D-GADD45* does not have detrimental effects. Moreover, the activation of *D-GADD45* also induces a decrease in proliferation, which is independent of the activation of the JNK signaling pathway. We also found that, while *D-GADD45* is dispensable during wing development under normal conditions, it becomes essential for regeneration. These findings suggest that *D-GADD45* could act as an in vivo stress sensor upstream of the MAP3K signaling cascade in *Drosophila*.

2. Materials and Methods

2.1. Drosophila Strains

The following Drosophila melanogaster strains were used: *TRE-DsRed.T4* [31], *ci-Gal4* (from R. Holmgren), *sal$^{E/Pv}$-LHG* [30], *lexO-rpr* [30], *en-Gal4* (from G. Morata), and *UAS-GFP (RRID:BDSC_4776)*, *UAS-RNAi-Ask1 (RRID:BDSC_35331)*, *UAS-RNAi-Mekk1 (RRID:BDSC_35402)*, *UAS-bsk^{DN} (RRID:BDSC_9311)*, *UAS-D-GADD45 (RRID:BDSC_81038)*, *UAS-RNAi-D-GADD45 (RRID:BDSC_35023)* and *tubGal80^{TS} (RRID:BDSC_7017)* from the Bloomington Drosophila Stock Center. *Hh-Gal4* is described in FlyBase (https://flybase.org/).

2.2. Activation of Transgenes Using the Gal4/UAS System

The Gal4 and UAS lines used are indicated for each experiment. For sustained activation of transgenes, flies were kept at 25 °C until 96 h after egg laying, when they were dissected and stained. To study adult phenotypes, flies were kept at 25 °C until adulthood.

For transient experiments, the expression of *Gal4* was controlled by the thermo sensitive repressor *tubGal80^{TS}*. Flies were kept at 17 °C until 192 h after egg laying (equivalent to 96 h at 25 °C) to prevent the expression of the constructs. They were subsequently kept at 29 °C for 6, 8 or 11 h. After that time, wing discs were immediately dissected and stained. To study the size and patterning of adult wings the vials were kept at 17 °C until adulthood.

2.3. Genetic Ablation and Dual Gal4/LexA Transactivation System

Cell death was induced as previously described [29,32]. We used the *sal$^{E/Pv}$-LHG* driver [30], which contains the binding domain of *LexA* and the activator domain of *Gal4*, which is recognized by

the inhibitor $Gal80^{TS}$, to induce the expression of the pro-apoptotic gene *reaper* (*rpr*) cloned downstream of the *LexA* operator *LexO*. We simultaneously used *Gal4* to express either *UAS-D-GADD45* or *UAS-RNAi-D-GADD45* under the control of *ci-Gal4*. The system was controlled by the thermo sensitive *Gal4*-repressor $tubGal80^{TS}$.

Embryos were kept at 17 °C until the 192 h after egg laying (equivalent to 96 h at 25 °C) to prevent *rpr* expression. They were subsequently moved to 29 °C for 11 h and then back to 17 °C until adulthood. Controls without *rpr* expression were always treated in parallel.

2.4. Immunohistochemistry and Statistics

Immunostaining was performed using standard protocols. Primary antibodies were P-Histone-H3 (rabbit 1:1000; Millipore, Burlington, MA, USA), Mmp-1 cocktail of three antibodies (3A6B4, 3B8D12 and 14A3D2, mouse 1:100 each; Developmental Studies Hybridoma Bank, Iowa, IA, USA), Cubitus interruptus (rat 1:25, Developmental Studies Hybridoma Bank) and P-JNK (rabbit 1:100; Promega, Madison, WI, USA).

Fluorescently labeled secondary antibodies were from Life Technologies (Carlsbad, CA, USA). Discs were mounted on SlowFade (Life Technologies). To label nuclei we used TO-PRO-3 (1:1000, Life Technologies), YO-PRO-1 (1:1000, Molecular Probes, Eugene, OR, USA) or Sytox Orange (1:10.000, Molecular Probes).

For apoptotic cell detection, we used the terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) assay, for which we used the fluorescently labelled Alexa Fluor 647-aha-dUTP (Thermo Fisher Scientific, Waltham, MA, USA) and incorporated using terminal deoxynucleotidyl transferase (Roche Diagnostics, Mannheim, Germany).

The apoptotic index was calculated after manually counting the number of cells positive for TUNEL in the anterior and posterior compartments of the disc, for all genotypes shown. The apoptotic index for each compartment was calculated as the fraction of apoptotic cells in that compartment multiplied by 100. Error bars indicate standard deviation of the mean. To compare apoptotic indexes, we used two-way analysis of variance (ANOVA) followed by Tukey post-hoc test on variables transformed to logarithmic scale.

The mitotic index was calculated after manually counting the number of P-Histone-H3 (PH3) positive cells in the anterior and posterior compartments, for all genotypes shown. The mitotic index for each compartment was calculated as the fraction of mitotic cells in that compartment multiplied by 100. Error bars indicate standard deviation of the mean. To compare mitotic indexes, we used two-way ANOVA followed by Tukey post-hoc test.

2.5. Test for Adult Wing Phenotypes

Female adult flies were fixed in glycerol:ethanol (1:2) for 24 h. Wings were dissected on water and then washed with ethanol. Then they were mounted on 6:5 lactic acid:ethanol and analyzed and imaged under a microscope. We considered wings as aberrant when missing veins or crossveins and/or when notches were present in the wing blade. Wing size was considered as the area inside the wing blade perimeter, as represented in Results section hh>GFP image.

Wing size and phenotype for each genotype is represented as a boxplot. Boxes represent the 1st, 2nd and 3rd quartiles. Whiskers represent interquartile range.

3. Results

3.1. Sustained Activation of D-GADD45 Induces JNK-Dependent Apoptosis

To get insight into the relationship between *GADD45* levels and their putative role, we first determined the effect of increased expression of *D-GADD45* in wing imaginal discs by employing the Gal4/UAS system [33]. For this, we used *UAS-D-GADD45* under the control of *hedgehog* (*hh)-Gal4*, which is expressed in the posterior compartment of the imaginal discs, allowing us to compare the

effects on the posterior (autonomous) and anterior (non-autonomous) compartments within the same disc. We activated the Gal4/UAS system throughout development until the third instar larval (L3) stage, when wing discs were dissected. We found a reduction in the whole disc size, including a reduction in size of the posterior compartment (Figure 1A). We also measured the apoptotic index in the posterior (GFP-positive) and anterior (GFP-negative) compartments and we observed a significant increase in the apoptotic index of both compartments when *D-GADD45* was upregulated compared to control discs (Figure 1A).

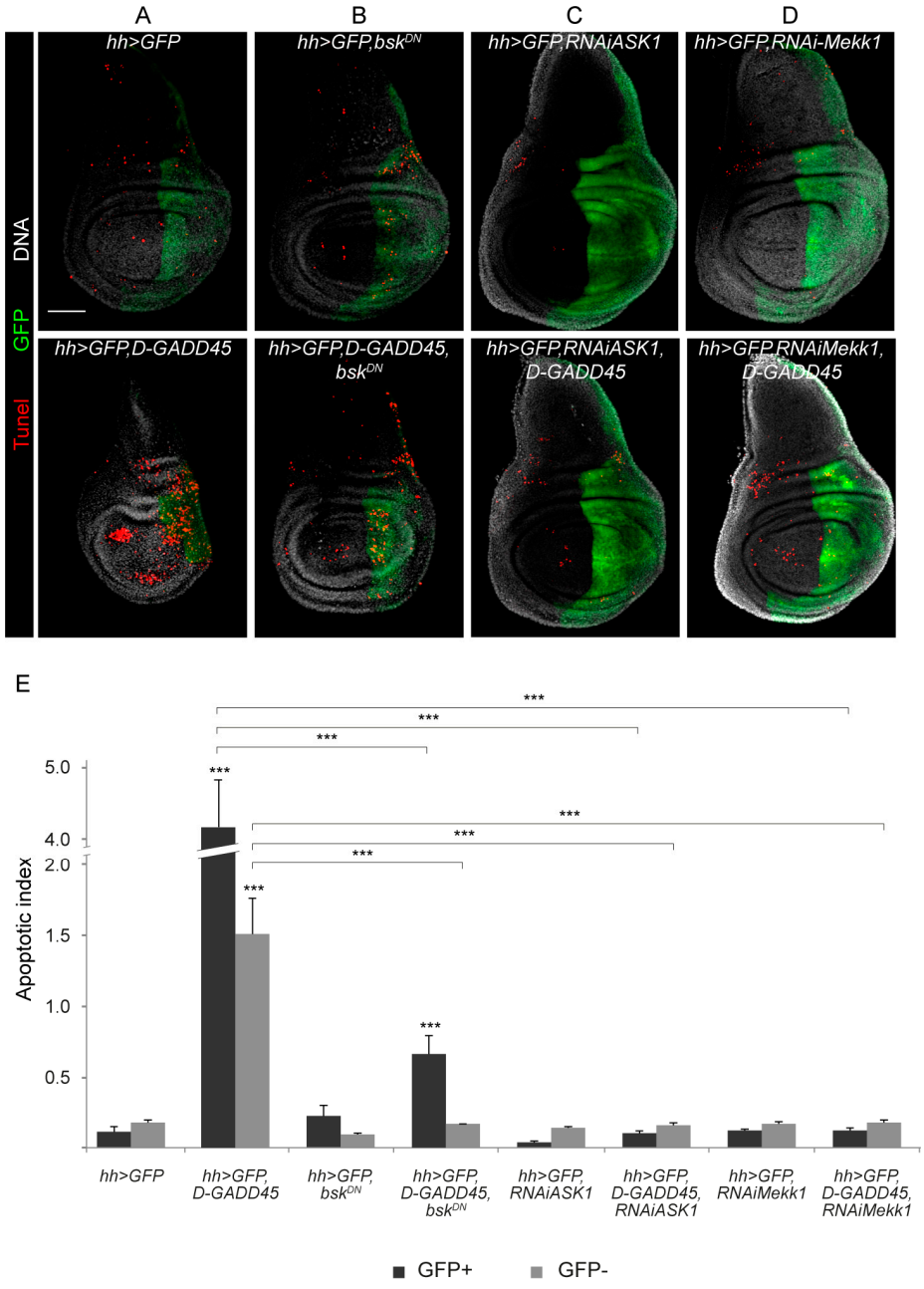


Figure 1. Sustained expression of *D-GADD45* induces JNK-dependent cell death. (**A**–**D**) TUNEL assay of wing discs labeling apoptotic cells after sustained expression of each construct in the posterior (*hedgehog*, *hh*) compartment. Apoptotic cells (red), posterior compartment (green) and DNA (white). Scale bar: 50 µm. (**E**) Histogram showing the apoptotic index for each genotype in the green fluorescent protein (GFP)-positive and GFP-negative compartments. Error bars indicate standard deviation. $N \geq 4$ discs for each genotype. $^{**}p < 0.01$. $^{***}p < 0.001$.

As mentioned before, it is well known that mammalian GADD45 proteins are able to interact to and activate the JNK signaling pathway [12,16]. Since sustained activation of the JNK cascade also has proapoptotic effects [34,35], we next checked whether sustained expression of *D-GADD45* activates the JNK pathway. We observed a clear increase in the activity of JNK upon increased *D-GADD45* expression, which indicates that D-GADD45 can activate JNK in the wing imaginal discs (Figure S1). To assess whether the increase in apoptosis is JNK dependent, we combined the expression of *D-GADD45* and the expression of a dominant negative form of the serine/threonine-protein kinase Basket (Bsk), a key downstream component of the JNK cascade, to inhibit the pathway. These discs showed a significant reduction of the apoptotic index in both compartments compared to discs expressing only *D-GADD45* (Figure 1A–B).

To gain further insight into the interaction between D-GADD45 and the JNK pathway, we examined the genetic interaction between two different MAP3Ks involved in the JNK pathway: the *Drosophila* ortholog of mammalian MTK1, Mekk1, and Ask1. In these experiments we combined the expression of *D-GADD45* and the expression of RNA interference (RNAi) constructs against either Mekk1 or Ask1. Similar to observations made when inhibiting the JNK pathway using the dominant negative form of Basket, depletion of Mekk1 or Ask1 significantly reduced the apoptosis caused by D-GADD45 (Figure 1C–D).

We also scored the adult wing phenotypes of flies in which *D-GADD45* was overexpressed in the posterior compartment and found aberrant and smaller size wings compared to controls (Figure 2A–B). Consistent with the phenotype observed in the imaginal discs, the inhibition of the JNK cascade, either by expressing the mutant form of *basket* or by depleting *Mekk1* or *Ask1* by RNAi, was sufficient to rescue the phenotypes, resulting in normal-sized and well-patterned wings (Figure 2C–H).

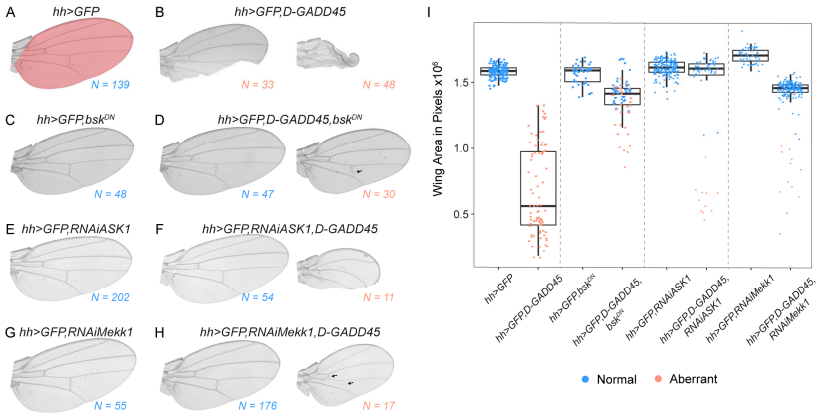


Figure 2. Sustained expression of D-GADD45 results in smaller and aberrant wings. (**A**–**H**) Adult wings showing the predominant phenotypes observed after sustained expression of each construct in the posterior (*hedgehog*, *hh*) compartment. Number of wings analyzed is indicated for each condition. Colored region in (**A**) represents the area selected to calculate wing size. (**I**) Box plot showing the average area of adult wings obtained after the expression of each combination of constructs in the posterior (*hedgehog*, *hh*) compartment. Each dot represents one wing; wild-type pattern (blue) and aberrant pattern (orange).

Since sustained expression of *D-GADD45* induces JNK-dependent cell death, we analyzed whether transient expression of *D-GADD45* is sufficient to induce apoptosis. To do so, we transiently activated the expression of *D-GADD45* in the posterior compartment of the wing disc (Figure 3A). After 6 and 8 h of *D-GADD45* activation, no differences were observed compared to control discs (Figure 3B–C). However, after 11 h of induction there was an increase in the number of apoptotic cells in the posterior compartment (Figure 3D). No wing patterning defects were observed after 6 or 8 h (Figure 3E–F), and minor vein alterations were detected after 11 h (Figure 3G). A slight increase in wing size was observed after 6 h of *D-GADD45* induction (Figure 3H).

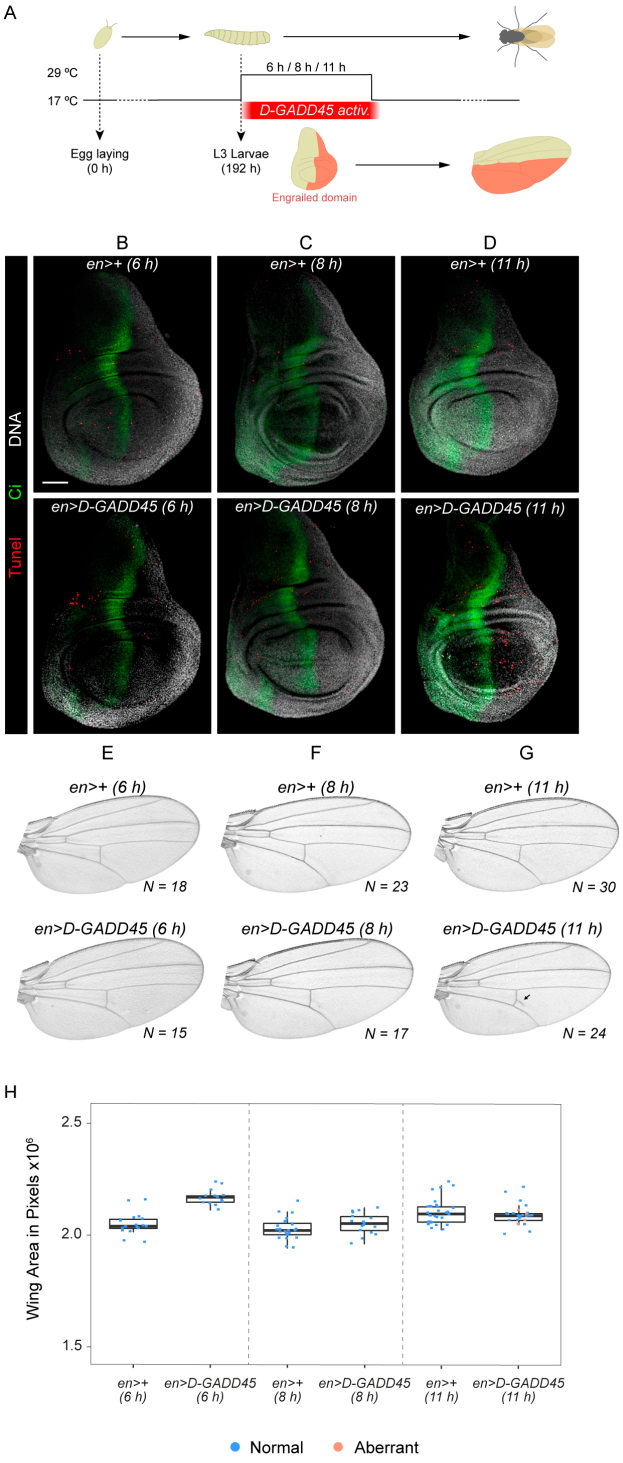


Figure 3. Transient expression of *D-GADD45* is not sufficient to induce cell death. (**A**) Schematic representation of *D-GADD45* transient activation in the posterior (*engrailed*, *en*) compartment. The affected domain is represented in orange. (**B**–**D**) TUNEL assay of wing imaginal discs labeling apoptotic cells (red), the anterior compartment (Ci: Cubitus interruptus, green) and DNA (white). $N = 5$ for each condition. Scale bar: 50µm. (**E**–**G**) Adult wings showing the predominant phenotype observed after transient expression of each construct in the posterior (*engrailed*, *en*) compartment. Number of wings analyzed is indicated for each condition. (**H**) Box plots showing the average area of adult wings. Each dot represents one wing; normal (blue) and aberrant pattern (orange).

Altogether, these results demonstrate that the detrimental effects of *D-GADD45* are due to the exposure to sustained levels of expression and that transient expression of *D-GADD45* is not sufficient to induce cell death, since we could not detect apoptosis after 6 and 8 h of induction. Our findings also suggest that D-GADD45 interaction with the JNK pathway could be mediated at the MAP3K level, by Mekk1 and Ask1 kinases.

3.2. Increased Expression of D-GADD45 Decreases Cell Proliferation Independently of the JNK Signaling Pathway

Overexpression of GADD45 results in cell growth suppression in numerous mammalian cell lines [3,36–38]. Therefore, we analyzed whether higher levels of *D-GADD45* would also affect proliferation in the imaginal discs. For this, we induced *D-GADD45* expression in the posterior compartment of the wing discs and calculated the mitotic index using P-histone-H3 (histone H3 phosphorylated) as a mitotic marker. We found that increased *D-GADD45* expression in the posterior compartment reduced the whole disc size (Figure 1A and 4C) and significantly reduced mitosis in this compartment (Figure 4C). The mitotic index in the anterior compartment was, however, comparable to that in control discs (Figure 4A).

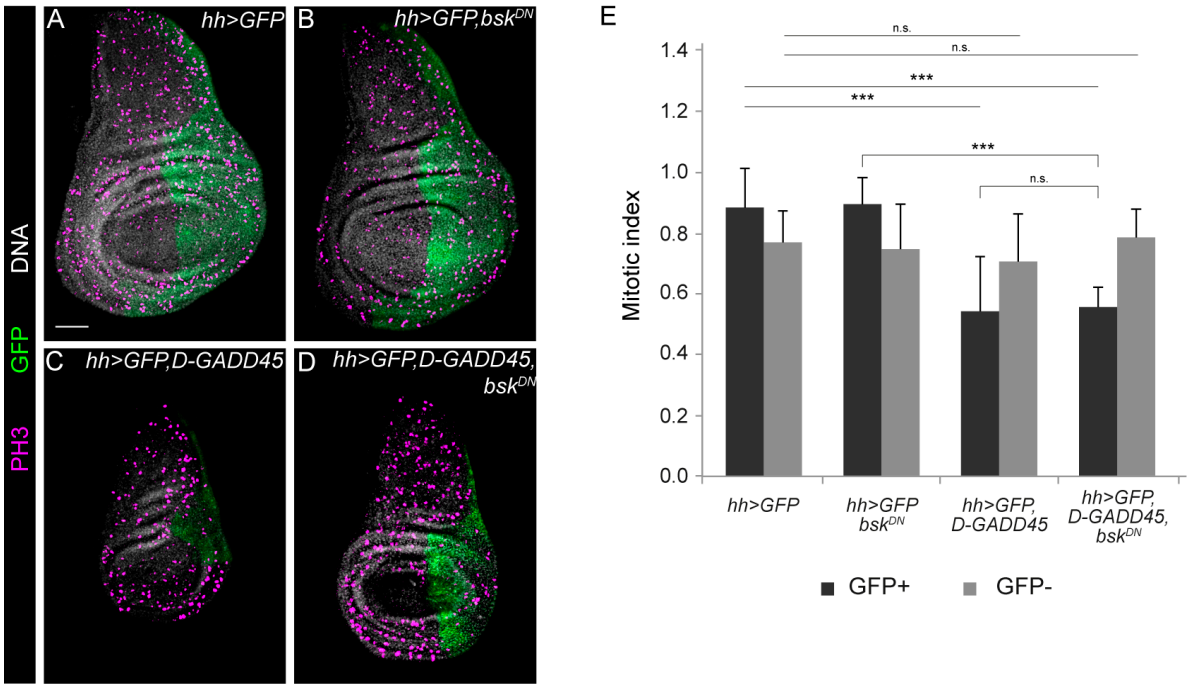


Figure 4. Sustained expression of *D-GADD45* decreases cell proliferation independently of JNK activation. (**A**–**D**) Immunostaining of wing discs with P-histone-H3 labeling mitosis (magenta), the posterior compartment (green) and DNA (white). Scale bar: 50 µm. (**E**) Histogram showing the mitotic index of PH3 labeling for each genotype in the GFP-positive and GFP-negative compartments. $N \geq 10$ for genotype. ***$p < 0.001$.

We next inhibited the JNK cascade by expressing the dominant negative form of *basket*, preventing *D-GADD45*-induced apoptosis. We found that, in these discs, mitosis in the anterior compartment was also equivalent to that in control discs (Figure 4B–D); however, proliferation in the posterior compartment was still significantly reduced but did not differ significantly from that found in discs in which only *D-GADD45* was activated (Figure 4C–D). Altogether, these results suggest that besides apoptosis, *D-GADD45* induces an autonomous decrease in proliferation independently of the activation of the JNK cascade.

3.3. D-GADD45 is Required for Regeneration of Wing Imaginal Discs

Because we had previously observed that *D-GADD45* was transiently upregulated following physical injury or genetic induction of cell death [26,27], we analyzed the ability of *D-GADD45*-depleted discs to regenerate. We scored wing regeneration after the induction of cell death by expressing the proapoptotic gene *reaper* ($sal^{E/Pv} > rpr$) while depleting *D-GADD45* by RNAi in the anterior compartment using the driver *cubitus interruptus* (*ci > RNAi-D-GADD45*) (Figure 5A). While control wings were normal in size and pattern, transient depletion of *D-GADD45* by RNAi produced visible defects in wing size and patterning in around half of the analyzed wings (Figure 5B–C). In regenerating animals, depletion of *D-GADD45* completely impaired regeneration, resulting in aberrant wings lacking the pattern and size of controls (Figure 5D–F). The presence of notches in the wing blade also indicated that the missing tissue had not been recovered, suggesting that activation of *D-GADD45* is required for wing repair after cell death. These findings show that *D-GADD45* contributes to regeneration.

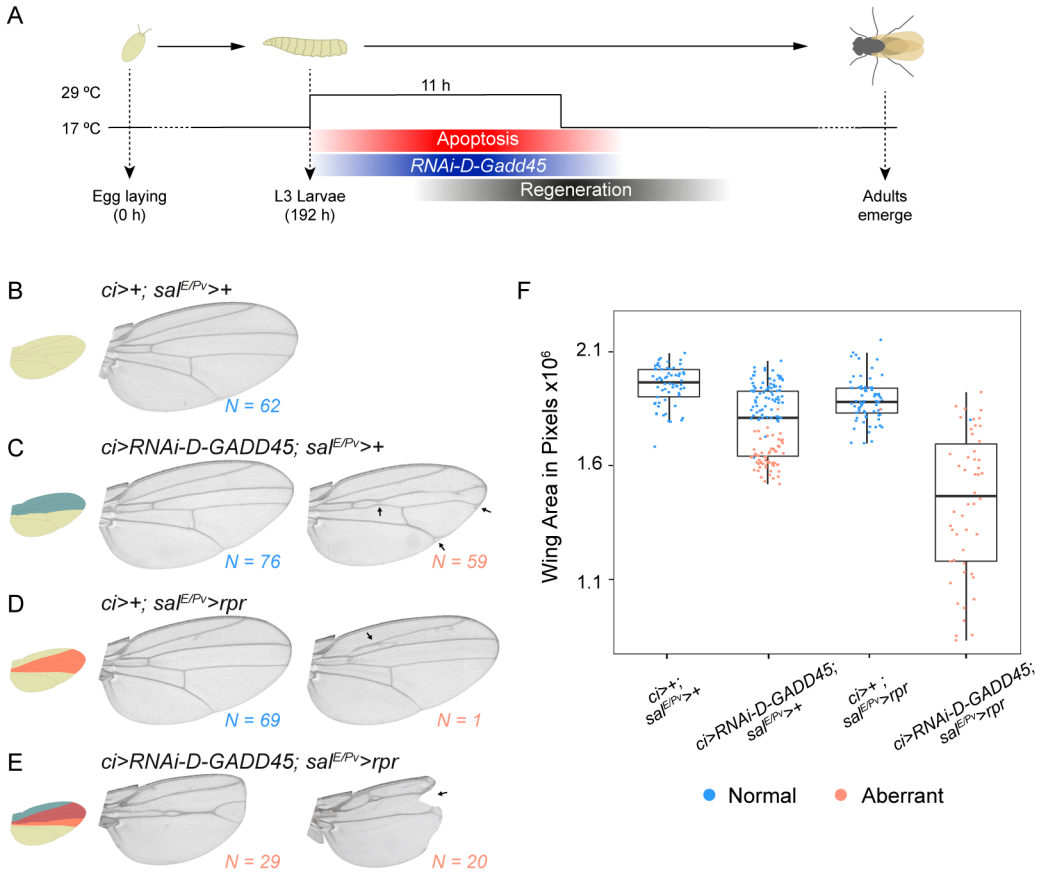


Figure 5. Depletion of *D-GADD45* severely impairs wing regeneration. (**A**) Schematic representation of the regeneration model used. (**B**–**E**) Adult wings showing the predominant phenotypes observed in each condition. Number of wings analyzed is indicated for each condition. (**F**) Box plot showing the average area of adult wings after 11 h of induction. Each dot represents one wing; wild-type pattern (blue) and aberrant pattern (orange).

4. Discussion

In this work, we describe a role for the stress sensor *D-GADD45* in *Drosophila*. Our results, together with previous observations, indicate that the levels of *D-GADD45* must be controlled both, during normal development as well as after the induction of cell death.

We also reveal the relationship between the D-GADD45 protein and the JNK signaling pathway during development. A previous study in flies already showed a genetic interaction in the germline

between D-GADD45 and MAP Kinase Kinases (Hemipterous and Licorne), members of the JNK and p38 signaling pathways, respectively [22]. Our data show a genetic interaction of D-GADD45 and the JNK signaling pathway at the level of MAP3Ks. In mammals, binding of GADD45 to MTK1, ortholog of *Drosophila* Mekk1, leads to the auto-phosphorylation of its kinase domain, allowing MTK1 to trigger the JNK signaling cascade [12,16]. Previous studies have also revealed physical binding between ASK1 and GADD45 in human cells, although this interaction was thought to be non-functional [14]. Further analyses are required to uncover whether the molecular mechanism of GADD45-mediated activation of the JNK pathway is conserved in *Drosophila*.

In mammals, overexpression of GADD45 has been found to induce G2/M phase arrest in numerous cell lines [3,36,37,39,40]. Here we show that a sustained increase of *D-GADD45* levels results in an increase in apoptosis and a decrease in cell proliferation. Although several studies have implicated JNK signaling in G2/M phase arrest [41,42], we still observed a decrease in mitosis when activating *D-GADD45* and inhibiting the JNK cascade through development, suggesting that the activation of JNK is dispensable in *Drosophila* for the *D-GADD45*-mediated effect in cell proliferation. On the other hand, it has been described that the *Drosophila* wing size is regulated by JNK signaling during development [43]. Thus, we cannot discard that a short induction of *D-GADD45* (6h) may activate a JNK-mediated proliferative response.

One of the early responses to damage in the wing imaginal discs is the activation of the JNK signaling cascade, which is required for regeneration [28,29,44]. In this system, induction of cell death activates p38 and induces tolerable levels of JNK, which are essential for wound healing [30]. Moreover, we have recently demonstrated the requirement for Ask1 during wing regeneration [45]. We hypothesize here that D-GADD45 could be required to regulate the activity of JNK by activating Ask1 and Mekk1. We cannot rule out the possibility, however, that other members of the JNK signaling pathway could also interact with D-GADD45 during the stress response in *Drosophila*, as observed in other systems [14,15,19,20]. Similar to mammals [11], it is likely that members of the MAPK family that are effectors of GADD45 signaling, contribute, at the same time to GADD45 induction.

Finally, it is tempting to speculate about the possible mechanisms behind the tight regulation of *D-GADD45* levels during regeneration. On the one hand, *D-GADD45* could be a direct downstream target of the JNK pathway. Both, p38 and JNK are required only during the early response [30] and the expression of *D-GADD45* shows an increase/decrease pattern during regeneration [26,27]. In addition, the promoter region of *D-GADD45* contains putative binding sites for the AP1 (activator protein-1) protein, the transcription factor downstream of the stress-responding JNK pathway [26]. On the other hand, a degradation of *D-GADD45* messenger RNA (mRNA) by nonsense-mediated decay (NMD) has been described as essential for viability in flies [46]. This observation suggests that when *D-GADD45* mRNA levels reach a certain threshold, the NMD pathway would destroy these transcripts, reducing the amount of the D-GADD45 protein to the appropriate levels to prevent D-GADD45-mediated apoptosis.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/10/5/378/s1, Figure S1: Sustained expression of *D-GADD45* activates JNK.

Author Contributions: Conceptualization, C.C.-R. and M.C.; methodology, C.C.-R. and M.C.; formal analysis, C.C.-R. and M.C.; investigation, C.C.-R.; resources, F.S. and M.C.; writing-original draft preparation, C.C.-R. and M.C.; writing-review and editing, C.C.-R., F.S. and M.C.; supervision, M.C.; project administration, F.S. and M.C.; funding acquisition, F.S. and M.C.

Funding: This research was funded by the following grants: BFU2012-36888 and BFU2015-67623-P from the Spanish Ministerio de Economía y Competitividad to F.S. and M.C. and the Institució Catalana de Recerca i Estudis Avançats (via an ICREA Academia award) to M.C.

Acknowledgments: We thank Elena Vizcaya-Molina for insightful comments and suggestions, Raziel Amador for technical advice, Miquel Calvo for advice on statistical analysis and J. Ramon for previous experiments. We thank the Confocal Unit of the CCiT-UB and Bloomington Drosophila Stock Center (http://flystocks.bio.indiana.edu/) for fly stocks.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Fornace, A.J.; Alamo, I.; Hollander, M.C. DNA damage-inducible transcripts in mammalian cells. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 8800–8804. [CrossRef] [PubMed]
2. Fornace, A.J.; Nebert, D.W.; Hollander, M.C.; Luethy, J.D.; Papathanasiou, M.; Papathanasiou, M.; Fargnoli, J.; Holbrook, N.J. Mammalian genes coordinately regulated by growth arrest signals and DNA-damaging agents. *Mol. Cell. Biol.* **1989**, *9*, 4196–4203. [CrossRef] [PubMed]
3. Zhan, Q.; Lord, K.A.; Alamo, I.; Hollander, M.C.; Carrier, F.; Ron, D.; Kohn, K.W.; Hoffman, B.; Liebermann, D.A.; Fornace, A.J. The *gadd* and *MyD* genes define a novel set of mammalian genes encoding acidic proteins that synergistically suppress cell growth. *Mol. Cell. Biol.* **1994**, *14*, 2361–2371. [CrossRef]
4. Carrier, F.; Georgel, P.T.; Pourquier, P.; Blake, M.; Kontny, H.U.; Antinore, M.J.; Gariboldi, M.; Myers, T.G.; Weinstein, J.N.; Pommier, Y.; et al. Gadd45, a p53-responsive stress protein, modifies DNA accessibility on damaged chromatin. *Mol. Cell. Biol.* **1999**, *19*, 1673–1685. [CrossRef] [PubMed]
5. Liebermann, D.A.; Hoffman, B. Gadd45 in stress signaling. *J. Mol. Signal.* **2008**, *3*, 1–8. [CrossRef]
6. Tamura, R.E.; de Vasconcellos, J.F.; Sarkard, D.; Libermann, T.A.; Fisher, P.B.; Zerbinia, L.F. GADD45 proteins: central players in tumorigenesis. *Curr. Mol. Med.* **2012**, *12*, 634–651. [CrossRef] [PubMed]
7. Powell, C.; Elsaeidi, F.; Goldman, D. Injury-dependent muller glia and ganglion cell reprogramming during tissue regeneration requires Apobec2a and Apobec2b. *J. Neurosci.* **2012**, *32*, 1096–1109. [CrossRef]
8. Hirose, K.; Shimoda, N.; Kikuchi, Y. Transient reduction of 5-methylcytosine and 5-hydroxymethylcytosine is associated with active DNA demethylation during regeneration of zebrafish fin. *Epigenetics* **2013**, *8*, 899–906. [CrossRef]
9. Chen, Z.; Wan, X.; Hou, Q.; Shi, S.; Wang, L.; Chen, P.; Zhu, X.; Zeng, C.; Qin, W.; Zhou, W.; et al. GADD45B mediates podocyte injury in zebrafish by activating the ROS-GADD45B-p38 pathway. *Cell Death Dis.* **2016**, *7*, e2068. [CrossRef]
10. Niehrs, C.; Schäfer, A. Active DNA demethylation by Gadd45 and DNA repair. *Trends Cell Biol.* **2012**, *22*, 220–227. [CrossRef]
11. Salvador, J.M.; Brown-Clay, J.D.; Fornace, A.J., Jr. *Gadd45 in Stress Signaling, Cell Cycle Control., and Apoptosis*; Springer: New York, NY, USA, 2013; ISBN 9781461482888.
12. Takekawa, M.; Saito, H. A family of stress-inducible GADD45-like proteins mediate activation of the stress-responsive MTK1/MEKK4 MAPKKK. *Cell* **1998**, *95*, 521–530. [CrossRef]
13. De Smaele, E.; Zazzeroni, F.; Papa, S.; Nguyen, D.U.; Jin, R.; Jones, J.; Cong, R.; Franzoso, G. Induction of gadd45beta by NF-kappaB downregulates pro-apoptotic JNK signalling. *Nature* **2001**, *414*, 308–313. [CrossRef] [PubMed]
14. Papa, S.; Zazzeroni, F.; Bubici, C.; Jayawardena, S.; Alvarez, K.; Matsuda, S.; Nguyen, D.U.; Pham, C.G.; Nelsbach, A.H.; Melis, T.; et al. Gadd45β mediates the NF-κB suppression of JNK signalling by targeting MKK7/JNKK2. *Nat. Cell Biol.* **2004**, *6*, 146–153. [CrossRef] [PubMed]
15. Papa, S.; Monti, S.M.; Vitale, R.M.; Bubici, C.; Jayawardena, S.; Alvarez, K.; De Smaele, E.; Dathan, N.; Ruvo, M.; Pedone, C.; et al. Insights into the Structural Basis of the GADD45β-mediated Inactivation of the JNK Kinase, MKK7/JNKK2. *J. Biol. Chem.* **2007**, *282*, 19029–19041. [CrossRef] [PubMed]
16. Miyake, Z.; Takekawa, M.; Ge, Q.; Saito, H. Activation of MTK1/MEKK4 by GADD45 through induced N-C dissociation and dimerization-mediated trans autophosphorylation of the MTK1 kinase domain. *Mol. Cell. Biol.* **2007**, *27*, 2765–2776. [CrossRef] [PubMed]
17. Hotamisligil, G.S.; Davis, R.J. Cell signaling and stress responses. *Cold Spring Harb. Perspect. Biol.* **2016**, *8*, 1–20. [CrossRef] [PubMed]
18. Wagner, E.F.; Nebreda, Á.R. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat. Rev. Cancer* **2009**, *9*, 537–549. [CrossRef]
19. Tornatore, L.; Marasco, D.; Dathan, N.; Vitale, R.M.; Benedetti, E.; Papa, S.; Franzoso, G.; Ruvo, M.; Monti, S.M. Gadd45β forms a homodimeric complex that binds tightly to MKK7. *J. Mol. Biol.* **2008**, *378*, 97–111. [CrossRef]

20. Ueda, T.; Kohama, Y.; Kuge, A.; Kido, E.; Sakurai, H. GADD45 family proteins suppress JNK signaling by targeting MKK7. *Arch. Biochem. Biophys.* **2017**, *635*, 1–7. [CrossRef] [PubMed]
21. Yang, Z.; Song, L.; Huang, C. Gadd45 proteins as critical signal transducers linking NF-κB to MAPK cascades. *Curr. Cancer Drug Targets* **2009**, *9*, 915–930. [CrossRef]
22. Peretz, G.; Bakhrat, A.; Abdu, U. Expression of the *Drosophila melanogaster* GADD45 homolog (CG11086) affects egg asymmetric development that is mediated by the c-Jun N-terminal kinase pathway. *Genetics* **2007**, *177*, 1691–1702. [CrossRef] [PubMed]
23. Stramer, B.; Winfield, M.; Shaw, T.; Millard, T.H.; Woolner, S.; Martin, P. Gene induction following wounding of wild-type versus macrophage-deficient *Drosophila* embryos. *EMBO Rep.* **2008**, *9*, 465–471. [CrossRef] [PubMed]
24. Plyusnina, E.N.; Shaposhnikov, M.V.; Moskalev, A.A. Increase of *Drosophila melanogaster* lifespan due to *D-GADD45* overexpression in the nervous system. *Biogerontology* **2011**, *12*, 211–226. [CrossRef]
25. Moskalev, A.A.; Smit-McBride, Z.; Shaposhnikov, M.V.; Plyusnina, E.N.; Zhavoronkov, A.; Budovsky, A.; Tacutu, R.; Fraifeld, V.E. Gadd45 proteins: Relevance to aging, longevity and age-related pathologies. *Ageing Res. Rev.* **2012**, *11*, 51–66. [CrossRef] [PubMed]
26. Blanco, E.; Ruiz-Romero, M.; Beltran, S.; Bosch, M.; Punset, A.; Serras, F.; Corominas, M. Gene expression following induction of regeneration in *Drosophila* wing imaginal discs. Expression profile of regenerating wing discs. *BMC Dev. Biol.* **2010**, *10*, 94. [CrossRef]
27. Vizcaya-Molina, E.; Klein, C.C.; Serras, F.; Mishra, R.K.; Guigó, R.; Corominas, M. Damage-responsive elements in *Drosophila* regeneration. *Genome Res.* **2018**, *28*, 1852–1866. [CrossRef] [PubMed]
28. Bosch, M.; Serras, F.; Martín-Blanco, E.; Baguñà, J. JNK signaling pathway required for wound healing in regenerating *Drosophila* wing imaginal discs. *Dev. Biol.* **2005**, *280*, 73–86. [CrossRef] [PubMed]
29. Bergantinos, C.; Corominas, M.; Serras, F. Cell death-induced regeneration in wing imaginal discs requires JNK signalling. *Development* **2010**, *137*, 1169–1179. [CrossRef]
30. Santabárbara-Ruiz, P.; López-Santillán, M.; Martínez-Rodríguez, I.; Binagui-Casas, A.; Pérez, L.; Milán, M.; Corominas, M.; Serras, F. ROS-induced JNK and p38 signaling is required for unpaired cytokine activation during *Drosophila* regeneration. *PLoS Genet.* **2015**, *11*, e1005595. [CrossRef]
31. Chatterjee, N.; Bohmann, D. A versatile φC31 based reporter system for measuring AP-1 and NRF2 signaling in *Drosophila* and in tissue culture. *PLoS ONE* **2012**, *7*, e34063. [CrossRef]
32. Repiso, A.; Bergantinos, C.; Serras, F. Cell fate respecification and cell division orientation drive intercalary regeneration in *Drosophila* wing discs. *Development* **2013**, *140*, 3541–3551. [CrossRef] [PubMed]
33. Brand, A.H.; Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **1993**, *118*, 289–295.
34. Lin, A. Activation of the JNK signaling pathway: Breaking the brake on apoptosis. *BioEssays* **2003**, *25*, 17–24. [CrossRef]
35. Liu, J.; Lin, A. Role of JNK activation in apoptosis: A double-edged sword. *Cell Res.* **2005**, *15*, 36–42. [CrossRef] [PubMed]
36. Wang, X.W.; Zhan, Q.; Coursen, J.D.; Khan, M.A.; Kontny, H.U.; Yu, L.; Hollander, M.C.; O'Connor, P.M.; Fornace, A.J.; Harris, C.C. GADD45 induction of a G2/M cell cycle checkpoint. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 3706–3711. [CrossRef] [PubMed]
37. Nakayama, K.; Hara, T.; Hirano, T.; Hibi, M.; Hirano, T.; Miyajima, A. A novel oncostatin m-inducible gene OIG37 forms a gene family with MyD118 and GADD45 and negatively regulates cell growth. *J. Biol. Chem.* **1999**, *274*, 24766–24772. [CrossRef] [PubMed]
38. Jin, S.; Antinore, M.J.; Lung, F.-D.T.; Dong, X.; Zhao, H.; Fan, F.; Colchagie, A.B.; Blanck, P.; Roller, P.P.; Fornace, A.J.; et al. The GADD45 inhibition of Cdc2 kinase correlates with GADD45 -mediated growth suppression. *J. Biol. Chem.* **2000**, *275*, 16602–16608. [CrossRef] [PubMed]
39. Zhang, W.; Bae, I.; Krishnaraju, K.; Azam, N.; Fan, W.; Smith, K.; Hoffman, B.; Liebermann, D.A. *CR6*: A third member in the MyD118 and Gadd45 gene family which functions in negative growth control. *Oncogene* **1999**, *18*, 4899–4907. [CrossRef] [PubMed]
40. Jin, S.; Tong, T.; Fan, W.; Fan, F.; Antinore, M.J.; Zhu, X.; Mazzacurati, L.; Li, X.; Petrik, K.L.; Rajasekaran, B.; et al. GADD45-induced cell cycle G2-M arrest associates with altered subcellular distribution of cyclin B1 and is independent of p38 kinase activity. *Oncogene* **2002**, *21*, 8696–8704. [CrossRef]

41. Zhu, N.; Shao, Y.; Xu, L.; Yu, L.; Sun, L. Gadd45-α and Gadd45-γ utilize p38 and JNK signaling pathways to induce cell cycle G2/M arrest in Hep-G2 hepatoma cells. *Mol. Biol. Rep.* **2009**, *36*, 2075–2085. [CrossRef]
42. Cosolo, A.; Jaiswal, J.; Csordás, G.; Grass, I.; Uhlirova, M.; Classen, A.-K. JNK-dependent cell cycle stalling in G2 promotes survival and senescence-like phenotypes in tissue stress. *Elife* **2019**, *8*, e41036. [CrossRef] [PubMed]
43. Willsey, H.R.; Zheng, X.; Pastor-Pareja, J.; Willsey, A.J.; Beachy, P.A.; Xu, T. Localized JNK signaling regulates organ size during development. *Elife* **2016**, *5*, e11491. [CrossRef]
44. Mattila, J.; Omelyanchuk, L.; Kyttälä, S.; Turunen, H.; Nokkala, S. Role of Jun N-terminal Kinase (JNK) signaling in the wound healing and regeneration of a *Drosophila melanogaster* wing imaginal disc. *Int. J. Dev. Biol.* **2005**, *49*, 391–399. [CrossRef] [PubMed]
45. Santabárbara-Ruiz, P.; Esteban-Collado, J.; Pérez, L.; Viola, G.; Abril, J.F.; Milán, M.; Corominas, M.; Serras, F. Ask1 and Akt act synergistically to promote ROS-dependent regeneration in *Drosophila*. *PLoS Genet.* **2019**, *15*, e1007926. [CrossRef] [PubMed]
46. Nelson, J.O.; Moore, K.A.; Chapin, A.; Hollien, J.; Metzstein, M.M. Degradation of Gadd45 mRNA by nonsense-mediated decay is essential for viability. *Elife* **2016**, *5*, e12876. [CrossRef]

Mammalian Genome
https://doi.org/10.1007/s00335-021-09939-4

Genomic and functional conservation of lncRNAs: lessons from flies

Carlos Camilleri-Robles[1] · Raziel Amador[1,2] · Cecilia C. Klein[1,2] · Roderic Guigó[2,3] · Montserrat Corominas[1] · Marina Ruiz-Romero[2]

Received: 17 June 2021 / Accepted: 9 December 2021

Abstract
Over the last decade, the increasing interest in long non-coding RNAs (lncRNAs) has led to the discovery of these transcripts in multiple organisms. LncRNAs tend to be specifically, and often lowly, expressed in certain tissues, cell types and biological contexts. Although lncRNAs participate in the regulation of a wide variety of biological processes, including development and disease, most of their functions and mechanisms of action remain unknown. Poor conservation of the DNA sequences encoding for these transcripts makes the identification of lncRNAs orthologues among different species very challenging, especially between evolutionarily distant species such as flies and humans or mice. However, the functions of lncRNAs are unexpectedly preserved among different species supporting the idea that conservation occurs beyond DNA sequences and reinforcing the potential of characterising lncRNAs in animal models. In this review, we describe the features and roles of lncRNAs in the fruit fly *Drosophila melanogaster,* focusing on genomic and functional comparisons with human and mouse lncRNAs. We also discuss the current state of advances and limitations in the study of lncRNA conservation and future perspectives.

Keywords LncRNAs · *Drosophila melanogaster* · Flies · Conservation · Comparative genomics · Development

Introduction

Long non-coding RNAs (lncRNAs), are DNA sequences encoding transcripts larger than 200 nt that lack protein-coding potential. Although many lncRNAs show low levels of expression, some are known to play a pivotal role in the regulation of several cellular processes. In recent years, the amount of available transcriptomic data has increased exponentially and has been crucial in demonstrating that genomes are extensively transcribed. Additionally, the emergence of tools to identify putative non-coding genes has led to the annotation of a large number of lncRNAs not only in humans, but also in mice, insects and plants (Brown et al. 2014; Derrien et al. 2012; Lagarde et al. 2017; Legeai and Derrien 2015; Paytuví Gallart et al. 2016; Pervouchine et al. 2015).

The most updated version of the human genome annotation contains 19,951 protein-coding genes and 17,948 lncRNA genes (GENCODE v37, March 2021). In contrast to protein-coding genes, whose molecular functions can often be inferred by the presence of protein domains, inferring the function, if any, of lncRNAs is a whole different story. Although some lncRNAs have been functionally characterised in humans (Rinn et al. 2007; Tripathi et al. 2010; Wutz et al. 2002; Zhou et al. 2007), the frequent absence of phenotypes after their mutation or deletion has raised questions about the proportion of annotated lncRNAs that are actually functional (Gao et al. 2020; Lee et al. 2019).

High-throughput screens of lncRNA knock downs affecting molecular phenotypes have been performed in human cells (Liu et al. 2017, 2018; Ramilowski et al. 2020). However, the difficulties in conducting functional genetic screens in humans and other vertebrates in vivo limit the ability to characterise the role of annotated lncRNAs in these species, pointing out the need to use less

✉ Marina Ruiz-Romero
marina.ruiz@crg.eu

[1] Departament de Genètica, Microbiologia I Estadística, Facultat de Biologia and Institut de Biomedicina (IBUB), Universitat de Barcelona, Barcelona, Catalonia, Spain

[2] Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology (BIST), Barcelona, Catalonia, Spain

[3] Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain

Published online: 31 January 2022

complex model organisms. One of the most useful animal models for genetic analyses is the fruit fly *Drosophila melanogaster*, whose genome contains 13,969 protein-coding and 2545 long non-coding RNA annotated genes (FlyBase r6.39, February 2021). A huge advantage of using *Drosophila* as an animal model is the availability of a great variety of genetic tools, resources and mutant collections that facilitate the undertaking of genetic screens. Forward genetic screens use mutagenesis to create random mutations in the search for the genotypes that underlie the resulting phenotypes. They have been instrumental in identifying the function of protein-coding genes (St Johnston 2002). On the contrary, reverse genetic assays are preferentially used to screen for lncRNAs, searching for phenotypes after creating targeted mutations in candidate genes (Wen et al. 2016). Compared to protein-coding genes, in which a single nucleotide deletion or insertion can abolish the production of the proteins they encode, the deletion of a large region encompassing the whole gene, smaller specific domains or the promoter, may be required to compromise the function of lncRNAs. The existence of different systems for the conditional expression of transgenes (both for inhibition and/or overexpression), widely used to assess protein-coding genes, may also contribute to understanding the function of lncRNAs.

It is well known that fundamental biological mechanisms and signalling pathways are conserved throughout evolution. An estimated 75% of genes related to human diseases have orthologs in the *Drosophila* genome (Bier 2005; Ji et al. 2019), endorsing the study of human diseases in flies. In this context, several lncRNAs have been associated with cancer (Dong et al. 2015; Li et al. 2014, 2015; Wu et al. 2014), and many neurological disorders, such as amyotrophic lateral sclerosis (Zu et al. 2013), Alzheimer's disease (Lee et al. 2015) and Huntington's disease (Johnson 2012). Although no fly orthologs have been identified for lncRNAs associated to human diseases, the study of lncRNAs in *Drosophila* could shed light into the regulation of disease-causing genes (Li et al. 2012; Lo Piccolo and Yamaguchi 2017; reviewed in Rogoyski et al. 2017).

In this review, we discuss lncRNAs in the fly genome and compare them with human and mouse lncRNAs. Furthermore, we provide an overview of the functions and mechanisms of action associated with lncRNAs in *Drosophila*, including similarities in the function of some lncRNAs between flies and humans. We also characterise developmentally dynamic fly lncRNAs that are differentially expressed during tissue development, and report resemblances among these lncRNAs and the ones identified in human and mouse organ development (Sarropoulos et al. 2019). Finally, we discuss the current status of identifying orthologues in evolutionarily distant species such as flies and humans.

Genomic and transcriptomic comparison between flies, mice and humans

Drosophila has four pairs of chromosomes: one pair of sexual chromosomes and three pairs of autosomes. In flies, similar to what happens in mice and humans, sex is determined by the XX/XY mechanism, with females carrying two X chromosomes and males carrying one X and Y chromosomes. In mammals, the presence of the Y chromosome determines the male sex, while its absence results in female individuals. However, the Y chromosome is not involved in sex determination in flies. Instead, the X:A ratio is responsible for the activation of the feminizing gene Sex-lethal (sxl). Hence, flies carrying XY or X0 are male, while flies carrying XX or XXY are female. The *Drosophila* genome is small, with approximately 120 megabases, (Adams et al. 2000) compared to the human and mouse genomes (3100 and 2700 megabases, respectively) (Lander et al. 2001; Venter et al. 2001; Consortium et al. 2002) (Fig. 1a). This is consistent with the reduced number of annotated genes in *Drosophila* (17,874) compared to humans and mice (45,468 and 39,923 genes, respectively). This trend is preserved for both protein-coding genes (13.969 genes in flies compared to 19,951 and 21,848 genes in humans and mice, respectively) and lncRNA genes (2545 genes in *Drosophila* compared to 17,948 and 13,186 genes in humans and mice, respectively) (Fig. 1b). Remarkably, the number of lncRNAs in *Drosophila* is considerably smaller compared to protein-coding genes, whereas in mice and humans the number of lncRNAs and protein coding-genes is similar (Fig. 1b). Furthermore, the *Drosophila* genome is much more compact, containing around 100 protein-coding and 18 lncRNA annotated genes per megabase compared to fewer than 10 protein-coding and lnRNA genes per megabase in the human and mouse genomes (Fig. 1c).

LncRNAs are pervasively distributed throughout the genome and can be found in intergenic regions (lincRNAs) or overlapping totally or partially with sequences of other genes transcribed in the same direction (sense) or in the opposite direction (antisense). Despite the differences in genome compactness among *Drosophila*, mice and humans, lincRNAs represent 50–55% of all annotated lncRNAs in the three species. Similarly, the proportion of lncRNAs found overlapping the introns (intronic) or exons (exonic) of other genes accounts for ~ 20% and ~ 25%, respectively, in all three species (Fig. 1d). Regarding the number of exons, the majority of lncRNAs found in *Drosophila* are either mono-exonic or composed of 2 exons, with only a few exceptions containing 3 or more exons. On the contrary, the number of exons in human and mouse lncRNAs is more diverse, with around 60% in each species

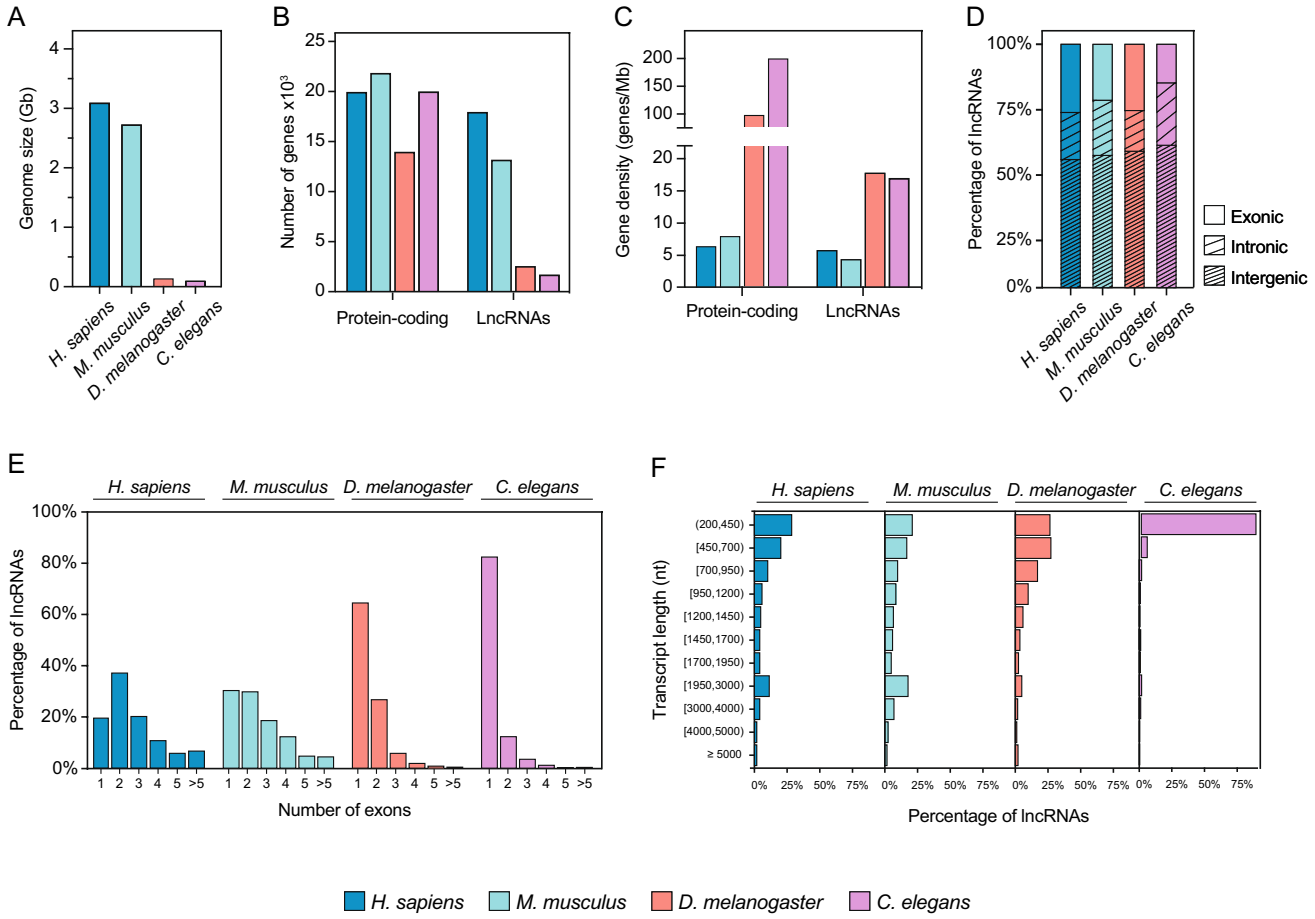


Fig. 1 Genomic and transcriptomic comparison of humans, mice and flies. **a** Barplot showing the size (gigabases—Gb) of the human (*H. sapiens*), mouse (*M. musculus*), fly (*D. melanogaster*) and worm (*C. elegans*) genome. **b** Number of annotated protein-coding and long non-coding RNA (lncRNA) genes in each species. **c** Gene density, measured as the number of genes per megabase, of protein-coding and lncRNA genes in each species. **d** Classification of the annotated lncRNA genes into exonic, intronic or intergenic groups. The longest annotated isoform of each lncRNA has been used for overlap analysis and classification. **e** Distribution of the lncRNA genes annotated in human, mouse, fly and worm depending on their number of exons. **f** Distribution of the long non-coding genes of humans, mice, flies and worms based on the size (nucleotides—nt) of their longest transcript. Human data from GENCODE v37 are shown in blue, mouse data from GENCODE M27 are shown in cyan, fly data from FlyBase r6.39 are presented in red and worm data from WS281 are shown in pink

containing 1–2 exons (compared to 90% in fly) and around 20% containing 4 or more exons (Fig. 1e). This is consistent with the proportion observed in protein-coding genes in flies (Graveley et al. 2011). In terms of transcript size, as with the protein-coding genes, *Drosophila* lncRNAs are shorter than human and mouse lncRNAs (average length of 962 nt in flies compared to 1230 and 1456 nt in humans and mice, respectively, Mann–Whitney–Wilcoxon test p-value < 1e−12 for all comparisons). Only 3.35% of *Drosophila* lncRNAs span more than 3 kb compared to 7.78% in humans and 10.35% in mice (Fig. 1f).

The small number of lncRNA genes annotated in *Drosophila*, with respect to humans and mice may not be a biological feature, but it might rather reflect two factors that hinder the identification of lncRNAs in *Drosophila*. First, the properties of most lncRNAs (low expression and high specificity in terms of time and tissue) might constraint their expression to a specific region during a very specific period of time. Thus, the identification of these lncRNAs may require transcriptomic analyses in specific tissues and developmental stages, which are not as common in flies as they are in humans and mice. Supporting this hypothesis, a recent publication producing transcriptomic data from *Drosophila* embryonic mesodermal cells collected at different developmental stages identified 179 novel lncRNA genes that could play a role in embryogenesis (Schor et al. 2018). Second, the bias towards mono-exonic genes found so far in *Drosophila* could also affect the identification of novel lncRNA genes, since most identifying pipelines often omit novel mono-exonic transcripts in favour of spliced, multi-exonic

transcripts to reduce false positives. To provide a global vision of the evolutionary trend we observed for lncRNAs in flies, we inspected the genome of the worm *Caenorhabditis elegans* (Fig. 1). Interestingly, the genomic features we interrogated indicate that lncRNAs in worms resemble fly lncRNAs compared to human and mouse lncRNAs.

Functions and mechanisms of action of *Drosophila* lncRNAs

Since the discovery of lncRNAs, several studies in mammals and flies have demonstrated that lncRNAs participate in a variety of cellular processes, such as development, differentiation and proliferation, and often contribute to the modulation of gene expression programmes (reviewed in Jandura and Krause 2017; Statello et al. 2021). Functional lncRNAs can be classified as *cis*-acting lncRNAs, when they act near their site of transcription within the same molecule, or *trans*-acting lncRNAs, which act far from their locus or in a different DNA molecule. In general, lncRNAs influence gene expression at three main levels: chromatin regulation, transcriptional regulation and post-transcriptional regulation (reviewed in Statello et al. 2021). On chromatin, some lncRNAs regulate the activity or localisation of chromatin regulatory complexes and transcription factors. These lncRNAs bind to specific chromatin regions and interact with proteins, facilitating or inhibiting their binding to targeted neighbouring genes, thereby promoting or repressing gene expression (Grote and Herrmann 2013; Jiang et al. 2015; Prensner et al. 2013; Rinn et al. 2007; Yap et al. 2010). At the transcriptional level, lncRNAs influence gene regulation directly by interacting with the transcriptional machinery, mediating or facilitating loops between promoters and enhancers (eRNAs) or, in some cases, the act of transcription or splicing of a lncRNA influence the transcription of nearby genes (reviewed in Statello et al. 2021). LncRNAs can also act at the post-transcriptional level by interacting with a plethora of RNA-binding proteins that contribute to mRNA stability, localisation, splicing or translation (Cao et al. 2017; Gumireddy et al. 2013; Lee et al. 2016; reviewed in He et al. 2019).

In the last few years, several lncRNAs have been characterised in *Drosophila*. Although most of them are not evolutionarily conserved across metazoans at the sequence level, some seem to participate in similar cellular processes as those in mammals, such as dosage compensation or Polycomb group (PcG)/Trithorax group (TrxG) regulation (reviewed in Murillo-Maldonado and Riesgo-Escovar 2019; Ringrose 2017; Samata and Akhtar 2018). In this section, we discuss the functions associated with fly lncRNAs, their level of conservation in mammals, and provide some specific examples.

LncRNAs influencing chromatin regulation

LncRNAs involved in dosage compensation mechanisms

As mentioned before, in *Drosophila*, sex is dictated by the XY sex-determination system. Comparable to that occurring in mammals, the imbalance in the expression of X-linked genes between females and males is corrected by a dosage compensation mechanism, involving lncRNAs, which result in similar levels of expression of the genes in the X chromosome. However, in female mammals, one of the X chromosomes is subjected to inactivation, whereas in *Drosophila*, the transcription rate of the male X chromosome is almost doubled. These strategies share some mechanistic similarities, including the involvement of lncRNAs. In both cases, a lncRNA is responsible for recruiting chromatin-modifying complexes that drive the inactivation (in female mammals) or overactivation (in male flies) of the X chromosome. Briefly, in mammals, the lncRNA *Xist* is upregulated in one of the X chromosomes of the females at early embryonic stages and rapidly spreads along the X chromosome from which it is transcribed (Brockdorff et al. 1992; Brown et al. 1991, 1992). Polycomb repressive complex 2 (PRC2), a chromatin regulatory complex, is recruited by *Xist* and mediates the trimethylation of lysine 27 in the histone H3 tail (H3K27me3). This triggers the heterochromatinisation of the *Xist*-bound X chromosome, resulting in X chromosome inactivation (Lee et al. 1996; Penny et al. 1996; Wutz and Jaenisch 2000).

On the contrary, in flies, the male-specific lethal complex (MSL), composed of MSL proteins and the lncRNAs *roX1* and *roX2*, is responsible for the overactivation of genes located in the X chromosome of *Drosophila* males. Although very different in size and sequence, *roX1* and *roX2* act redundantly to allow the binding of MSL2 and the other subunits of the complex, which target the X-chromosome in males (Meller and Rattner, 2002). The MSL subunits mediate the activation of the X-chromosome genes by the acetylation of lysine 16 in histone H4 (H4K16ac) (Bone et al. 1994; Gelbart et al. 2009). In female flies, the *Sex lethal* (*sxl*) gene, is upregulated and the female-specific RNA-binding protein it encodes interacts with the *msl2* mRNA to inhibit its translation, preventing the assembly of the MSL complex and the subsequent dosage compensation (Beckmann et al. 2005; Gebauer et al. 1998; Graindorge et al. 2013).

LncRNAs mediating PcG and TrxG gene regulation

PcG and TrxG proteins are key modulators of an evolutionarily conserved gene regulatory system. They are chromatin modifiers that operate antagonistically and were originally identified as part of an epigenetic cellular memory system that maintains repressed or active gene expression states. The

first identified target genes of PcG and TrxG regulation were the fly Hox genes (reviewed in Kassis et al. 2017). Hox genes encode transcription factors that determine the allocation of segmental identity along the anterior–posterior body axis and when mutated, typically, lead to homeotic transformations (reviewed in Mallo and Alonso 2013). In *Drosophila*, Hox genes are organized in two separate gene clusters: the Antennapedia and Bithorax complexes (ANT-C and BX-C, respectively); and their expression is activated by the segmentation gene products in early fly development. Further characterisation of Hox loci allowed the identification of several elements that respond to PcG and TrxG genes, named Polycomb response elements (PREs) and Trithorax response elements (TREs) (Chan et al. 1994; Simon et al. 1993). *Drosophila* PcG and TrxG proteins are recruited to chromatin by targeting these PREs and TREs, which are *cis*-regulatory DNA elements essential for the regulation of several hundred developmental genes beyond Hox genes. The PcG and TrxG proteins are able to regulate their target genes in a complex and dynamic manner, modifying local chromatin depending on the state of the promoters and maintaining active (TrxG) or repressive (PcG) states. (reviewed in Kassis and Brown, 2013; Geisler and Paro 2015; Grossniklaus and Paro 2014; Steffen and Ringrose, 2014; Ringrose 2017; Schuettengruber et al. 2017). Many PcG/TrxG binding sites give rise to non-coding transcripts (reviewed in Hekimoglu and Ringrose 2009 and Ringrose 2017). For instance, forward and reverse non-coding transcription has been detected from the *Drosophila melanogaster vestigial* (*vg*) PRE/TRE, which switches the status of the element between silencing (induced by transcription from the forward strand) and activation (induced by transcription from the reverse strand). Moreover, the non-coding transcripts from the reverse strand are able to bind to the PRC2 in vivo, inhibiting its enzymatic activity (Herzog et al. 2014). Additionally, since the initial discovery of lncRNA *Xist* targeting PcG to the inactive X chromosome in mammals (Plath et al. 2003), several lncRNAs in flies and mammals have been described to participate, not only in PcG-dependent silencing, but also in gene activation via disruption of PcG silencing or physical interaction with TrxG components (Geisler and Paro 2015; Schuettengruber et al. 2017). Altogether, the analyses of non-coding-mediated regulation of PcG and TrxG suggest that non-coding transcripts may be required to destabilize stable active and silent chromatin states, and to recruit or evict components of the PcG and TrxG complexes depending on their transcription rate (Ringrose 2017).

LncRNAs modulating gene expression

LncRNAs transcribed from active enhancers (eRNAs)

Transcription has been observed from multiple active enhancers in mammals (Andersson et al. 2014; Arner et al. 2015; De Santa et al. 2010; Kim et al. 2010), *Drosophila* (Henriques et al. 2018; Meers et al. 2018) and *Caenorhabditis elegans* (Chen et al. 2013). Although these enhancer RNAs (eRNAs) are not transcribed from all enhancer regions, a correlation has been observed between enhancer activity and the transcription of eRNAs both in mammals and flies (Hah et al. 2013; Mikhaylichenko et al. 2018). A growing number of studies demonstrate that specific eRNAs are required to properly activate the expression of their target genes (Ivaldi et al. 2018; Lai et al. 2013; Lam et al. 2013; Li et al. 2013; Rahnamoun et al. 2018; Schaukowitch et al. 2014; Tsai et al. 2018). In mammals, eRNAs have been associated with regulation of transcription through different mechanisms including: interaction and enhancement of the activity of chromatin regulators, like the acetyltransferase CREB binding protein(CBP), PRC2, MLL1 or CTCF; influencing enhancer-promoter looping or altering RNA polymerase II elongation by interaction with proteins that either induce or inhibit elongation (reviewed in De Lara et al. 2019). However, as with the other types of lncRNAs, further studies are required to distinguish the eRNAs that actually play an active role in enhancer activity from those that might just be transcriptional noise arising from the presence of the RNA polymerase machinery. Although few eRNAs have been functionally characterised in flies, identification of general properties of eRNAs in *Drosophila* shows that eRNAs in flies share many characteristics with mammalian eRNAs, for instance, directionality, low abundance, correlation between expression and enhancer activity, or the presence of promoter-like motifs like INR motif (Mikhaylichenko et al. 2018).

LncRNAs acting at post-transcriptional level

LncRNAs as a source of miRNAs

MicroRNAs (miRNAs) are small non-coding transcripts (about 22 nucleotides) that play a major role in the post-transcriptional regulation of gene expression. In most cases, miRNAs are derived from the introns or exons of larger protein-coding or non-coding genes. In *Drosophila*, one of these non-coding transcripts, *iab-8*, is transcribed primarily from the posterior central nervous system, beginning in early development (Bender 2008). It spans over 90 kb and is both spliced and polyadenylated (Bender 2008; Garaulet et al. 2014). Once transcribed, *iab-8* is processed into three miRNAs that altogether are called *miR-iab-8*, which are encoded within its intronic sequence. These miRNAs are known to target and downregulate the homeotic genes *abd-A* and *Ubx*, as well as their cofactors *hth* and *exd* (Garaulet et al. 2014; Gummalla et al. 2012). The consequence of the loss of *iab-8* is male and female sterility caused by the increase in the level of the transcripts targeted by miR-iab-8 that is thought

to elicit a defective innervation of the abdominal and/or reproductive tract muscles of the fly (Maeda et al. 2018). In mammals, several lncRNAs have been described as precursors of miRNAs, although none have been found to target the Hox genes. For instance, the maternally imprinted *H19* gene encodes one of the first lncRNAs described, which is a known precursor of *miR-675* (Cai and Cullen 2007). *H19* is highly transcribed in fetal tissues, where it is found to be processed into *miR-675*, which limits placental growth by targeting, among others, growth promoting *Igf1r* (Keniry et al. 2012). In parallel, *H19* is also expressed in the adult skeletal muscle of humans and mice, where, instead of being processed into *miR-675*, *H19* acts as a molecular sponge for the let-7 family of miRNAs (Kallen et al. 2013; Onyango and Feinberg 2011).

Another lncRNA that is processed into smaller RNAs is *acal*, which was described by Riesgo-Escovar and colleagues in 2015 (Ríos-Barrera et al. 2015). *acal* is one of the few *Drosophila* lncRNAs showing sequence conservation. In particular, a 296 nt-long fragment is 80% sequence identical in *Drosophila melanogaster* and *Drosophila bipectinata*. Also, a similar-sized lncRNA is found in humans, showing a considerable 48% sequence identity to *Drosophila acal* (Murillo-Maldonado and Riesgo-Escovar 2019). Mutations in *acal* are embryonic lethal and result in defects in dorsal closure, a JNK-dependent process that is essential for *Drosophila* embryogenesis. It was found that *acal*, through the regulation of two JNK modulators, *Connector of kinase to AP1* (*Cka*) and *anterior open* (*aop*), is able to modulate JNK activity (Ríos-Barrera et al. 2015). Remarkably, *acal* is transcribed from a mono-exonic gene into a 2.3-kb long transcript that, throughout the life cycle of the fly, particularly during pupal stages, is processed into smaller transcripts spanning from 50 to 120 nucleotides. The function of these small RNAs is yet to be investigated, but the differences in size with respect to the ~ 22 nucleotide miRNAs indicate that processed *acal* does not act as a typical miRNA (Ríos-Barrera et al. 2015).

LncRNAs regulating isoform usage

We recently identified *blistered antisense* (*bsAS*) as a natural antisense transcript of the *blistered* (*bs)* gene involved in the regulation of *bs* isoform usage in flies (Pérez-Lluch et al. 2020). The *bs* gene encodes the *Drosophila* serum response factor (DSRF) and is a well characterised gene required for wing development and formation (Fristrom et al. 1994; Montagne et al. 1996; Roch et al. 1998). We have found that the usage of *bs* isoforms is regulated in a tissue-specific manner by the expression of the *bsAS*. Transcription of *bsAS* occurs specifically in wing intervein regions and impairs the expression of the long isoforms of *bs*, thereby promoting the relative expression of the short isoform. Overexpression of the long isoform in *bsAS* mutants induces the formation of extra vein tissue in adult wings. The regulation of *bs* isoform usage is based on the formation of a genomic loop between *bs* and *bsAS* promoters that impairs transcription of the long isoform and potentiates short isoform presence. This regulatory mechanism is totally independent of the presence of the *bsAs* transcript, as *bsAS* overexpression does not affect *bs* transcription.

A growing number of lncRNAs has been linked to the modulation of alternative splicing in mammals (reviewed in Romero-Barrios et al. 2018). For example, a natural antisense transcript regulates Zeb2/Sip1 expression during epithelial-mesenchymal transition in mammalian cells by preventing splicing of the Zeb2 5′-UTR (Beltran et al. 2008). An evolutionarily conserved nuclear antisense lncRNA, generated from the human fibroblast growth factor receptor 2 (*FGFR2)* locus, promotes epithelial-specific alternative splicing of *FGFR2* (Gonzalez et al. 2015). This lncRNA impairs the binding of a repressive chromatin-splicing adaptor complex important for mesenchymal-specific splicing, by recruiting PcG proteins and the histone demethylase KDM2a. More recently, Singer and colleagues (Singer et al. 2019) characterised *Paupar,* a lncRNA that interacts with SR proteins to promote the alternative splicing of PAX6 in pancreatic glucagon-producing α cells and computational analysis of hepatocellular carcinoma RNA-Seq samples predicted hundreds of splicing-related lncRNAs (Wang et al. 2020).

Other mechanisms of action of lncRNAs

LncRNAs encoding small functional peptides

By definition, lncRNAs lack protein coding potential. Nevertheless, roughly 98% of the annotated lncRNAs in humans, mice and flies contain small open reading frames (smORFs) of 10 to 100 codons that may code for peptides (Couso and Patraquim 2017). The putative function of these peptides is, however, often neglected and the genes that encode them remain listed as non-coding. Translation of smORFs is observed in many eukaryotes (Andrews and Rothnagel 2014; Couso and Patraquim 2017), but examples of small functional peptides have been described primarily in humans (Anderson et al. 2015; D'Lima et al. 2017; Huang et al. 2017; Nelson et al. 2016; Slavoff et al. 2014; van Heesch et al. 2019) and insects (Galindo et al. 2007; Kondo et al. 2007; Magny et al. 2013). In *Drosophila*, the *tarsal-less* (*tal*) gene, previously classified as non-coding, encodes for a polycistronic mRNA that is translated into 4 small peptides of 11 amino acids. One of these peptides actively participates in leg development at the larval stage by regulating gene expression and tissue folding (Galindo et al. 2007) and at the pupal stage by modulating Notch signalling (Pueyo and Couso 2011). Moreover, the presence of

similar smORFs in *tal* homologues across different species of insects suggests the presence of a conserved family of functional peptides (Galindo et al. 2007).

Ribosome profiling techniques (Ribo-seq), which specifically identify ribosome-bound transcripts, have corroborated that a fraction of lncRNAs have a strong affinity for ribosomes (Aspden et al. 2014; Bazzini et al. 2014; Carlevaro-Fita et al. 2016; Ingolia et al. 2011; Ruiz-Orera et al. 2014; van Heesch et al. 2014). However, the association with ribosomes does not necessarily imply that these lncRNAs are actively translated, since lncRNAs are known to regulate the translation of mRNAs through ribosome binding (Carrieri et al. 2012; Hansji et al. 2016; Liu et al. 2019; Yoon et al. 2012). To overcome this limitation, further studies on ribosome-bound lncRNAs should be taken: (1) to confirm whether they are translated and (2) to test the functionality of the translated smORFs. While peptide tagging or in vitro translation assays can be used to identify the coding potential of smORFs (Galindo et al. 2007; Pueyo and Couso 2011; van Heesch et al. 2019), the generation of knock-out mutants or the inhibition of the lncRNA transcription or translation should be considered to study their functionality (Anderson et al. 2015; Pueyo and Couso 2011).

The increasing number of functional smORFs encoded by genes annotated as lncRNAs challenges the current definition of lncRNAs. The fact that almost the totality of annotated lncRNAs present at least one predicted smORF within its sequence makes it impossible to rule them out just because of the smORF presence. However, to our understanding, the lncRNA status of those genes encoding for functionally validated smORFs should be revised or, on the contrary, the definition of lncRNA should be revised to include the genes encoding for functional smORFs.

Expression of lncRNAs in development

The first evidence of the involvement of mammalian lncRNAs in development came from high-throughput expression analyses of different tissues (Grote et al. 2013). Cell-type and tissue specificities have been described for many lncRNAs and differential expression of lncRNAs has been reported in in vitro models of haematopoiesis, suggesting that they could have a role in the regulation of cell fate decisions (Briggs et al. 2015; Constanty and Shkumatava 2021; Perry and Ulitsky 2016; Schwarzer et al. 2017). Although most lncRNAs are still uncharacterised, a wide variety of functional activities have been associated with lncRNAs involved in development, such as the regulation of chromatin and DNA interactions, modulation of transcription factors, roles in mRNA stability and processing, and involvement in protein stability and function. Thus, an increasing number of human and mouse lncRNAs are being implicated as key regulators in a variety of cellular processes including proliferation, apoptosis and responses to stress. In agreement with observations in mammals, analyses based on the modENCODE RNA-Seq data from whole *Drosophila* animals have shown that a substantial number of lncRNAs are differentially expressed during development (referred to as developmentally dynamic lncRNAs), although some of the lncRNAs characterised were very lowly expressed (Chen et al. 2016; Brown et al. 2014; Lee et al. 2019; Li et al. 2019). Figure 2a shows the expression changes of the updated list of annotated lncRNAs in *Drosophila* (FlyBase r6.39, 2,545 lncRNAs) across fly development, using the modENCODE RNA-Seq data. Although different profiles of expression can be observed, a huge proportion of lncRNAs is upregulated towards the end of development, as previously reported (Graveley et al. 2011). Indeed, large changes of expression are detected for many genes specifically at the entrance of metamorphosis.

The expression patterns of developmentally dynamic lncRNAs in *Drosophila* are more restricted than those of protein-coding genes. Brown and colleagues reported that, on average, lncRNAs are expressed in a smaller number of stages and tissues compared to protein-coding genes (Brown et al. 2014). Remarkably, similarly restricted expression patterns have been reported for lncRNAs in humans and other mammals (Briggs et al. 2015; Constanty and Shkumatava 2021; Perry and Ulitsky 2016). Most studies characterising lncRNAs expression during development, either in *Drosophila* or in mammals, have been carried out using whole animals, which could be an important constraint considering the high level of tissue specificity that lncRNAs display. Interestingly, a recent publication from Kaessmann's group systematically described developmentally dynamic lncRNAs across several organs during mammalian development (Sarropoulos et al. 2019). After analysing the RNA-Seq data from seven species, the authors identified developmentally dynamic genes that displayed changes in expression during the development of mammalian organs, showing that the fraction of lncRNAs among this group of genes was substantially low considering the total proportion of lncRNAs in the human and mouse genomes (Sarropoulos et al. 2019). We took advantage of a previously published RNA-Seq data set from our group (Pérez-Lluch et al. 2020) containing the expression values for three tissues (eye, leg and wing) in three developmental stages (third instar larvae, early pupae and late pupae) to identify developmentally dynamic *Drosophila* genes, including lncRNA genes. We observed that the proportion of developmentally dynamic genes corresponding to lncRNAs is much lower in *Drosophila* (4%) than humans and mice (~25%), which correlates with the lower number of annotated lncRNAs in flies (Fig. 2b).

We observed that the proportion of lncRNAs within developmentally dynamic genes in flies was lower (3.3%)

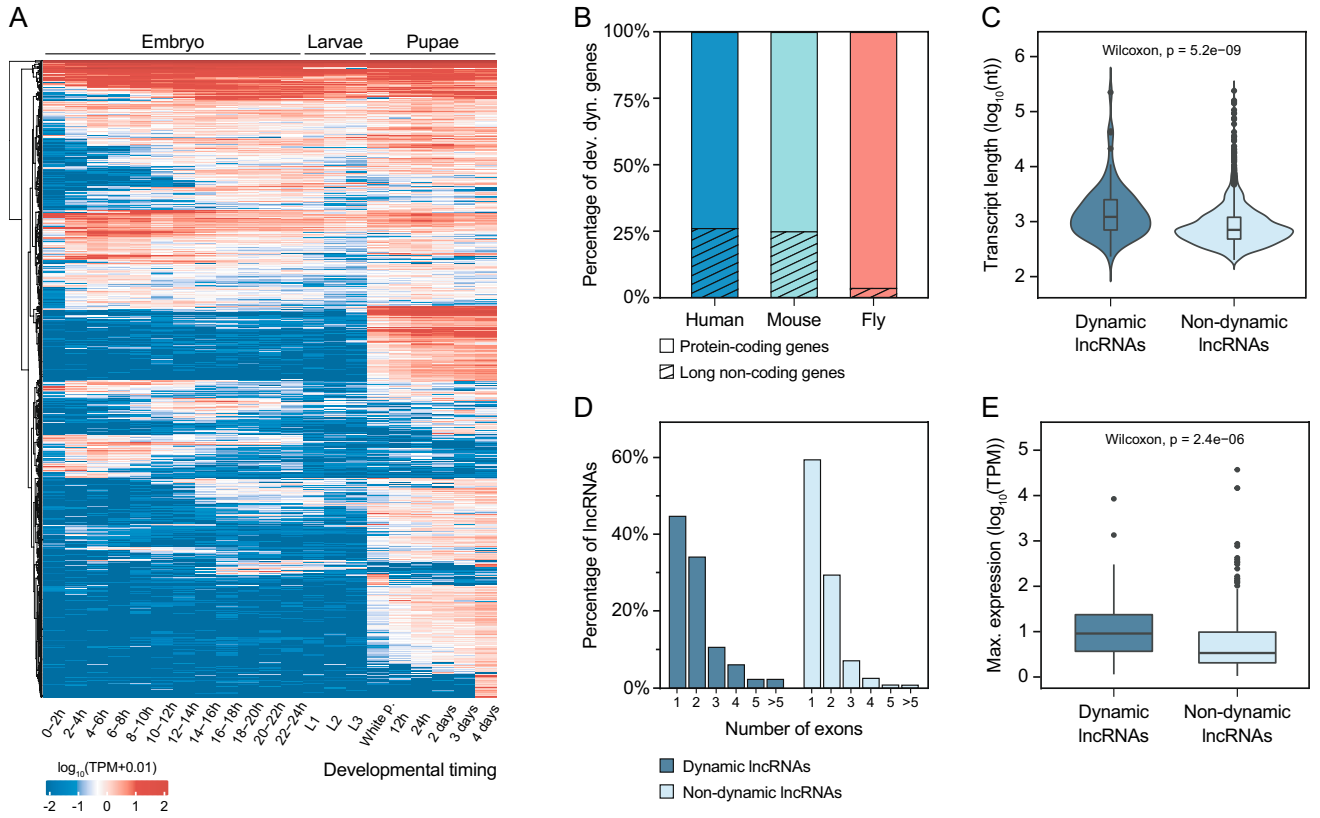


Fig. 2 Properties of *Drosophila* lncRNAs in development. **a** Heatmap showing the expression of all the annotated lncRNAs in *Drosophila* RNA-Seq samples from different developmental stages. **b** Percentage of developmentally dynamic genes corresponding to protein-coding and lncRNA genes in humans, mice and flies. Identification of developmentally dynamic genes was performed using edgeR on developmental RNA-Seq samples from Pérez-Lluch et al. (2020). **c** Violin plot showing the distribution of *Drosophila* dynamic and non-dynamic lncRNAs according to the length of their longest transcript. **d** Distribution of *Drosophila* dynamic and non-dynamic lncRNAs according to the number of exons in their longest transcript. **e** Boxplot showing the maximum expression across tissue development (expressed in $\log_{10}$TPMs) of dynamic and non-dynamic lncRNAs in *Drosophila*

than that of protein-coding genes (96.7%), a trend stronger than that observed in mammals, in which lncRNAs account for ~ 25% of developmentally dynamic genes (Fig. 2b). Sarropoulos and colleagues found some traits associated with developmentally dynamic lncRNAs. For example, the developmentally dynamic lncRNAs have a higher and broader expression than non-dynamic lncRNAs, they are in closer proximity to protein-coding genes, the transcripts are longer and they contain more exons than non-dynamic lncRNAs. To further characterise dynamic lncRNAs in *Drosophila,* and to compare them with mammalian ones, we analysed the lncRNA length, number of exons and level of expression during tissue development. Developmentally dynamic lncRNAs are longer (Fig. 2c, Mann–Whitney–Wilcoxon test *p*-value = 1.9e−9), contain more exons (Fig. 2d), and generally show higher expression across tissues during development than non-dynamic lncRNAs (Fig. 2e, Mann–Whitney–Wilcoxon test *p*-value = 2.4e−6). Our results indicate, therefore, that the properties identified previously for the mammalian lncRNAs with dynamic expression during organ development are conserved in *Drosophila* developmentally dynamic lncRNAs. Although it is difficult to identify conservation of lncRNAs in different species, the fact that their properties are conserved suggests that some of their roles in development could be conserved, as well.

Conservation

LncRNA sequences are generally not conserved across different species, which severely hinders the identification of conserved lncRNAs that are likely to be functional. While protein-coding genes are constrained by a strong selective pressure to maintain their reading frame and codon synonymy, lncRNAs do not seem to depend on their sequence to perform their function, leading to their rapid evolution and sequence degeneration. Nevertheless, a few examples of lncRNAs whose sequence is conserved between different species of *Drosophila* have been described. This is the case of the previously discussed lncRNA *acal* or the

yellow-achaete intergenic RNA (*yar*), which is a lncRNA involved in *Drosophila* sleep regulation. Several motifs ranging from 40 to 111 bp located in the TSS, the exons and the 3′-end of *yar* genomic sequence are conserved in different *Drosophila* species separated by as much as 40–60 million years of evolution (Soshnev et al. 2011). However, it is not possible to find sequence similarity for most lncRNAs, thus, other types of conservation analysis are often used to discover orthologous lncRNAs in different organisms.

Synteny, the positional conservation of neighbouring genes across different species, has emerged as a valuable in identifying orthologous non-coding genes (Bryzghalov et al. 2021; Herrera-Úbeda et al. 2019; Pegueroles et al. 2019; Rolland et al. 2019). This analysis relies on the presence of orthologous genes located in the same order in the linear genome of different species. Syntenic conservation of the region surrounding the lncRNA locus could be an indicator of lncRNA orthology. However, the presence of a lncRNA conserved by synteny in different species does not necessarily imply orthology. Particularly, the presence of large intergenic regions containing multiple lncRNAs increases the rate of false positives (Young et al. 2012). In addition, since the analysis of synteny depends on the presence of orthologous genes, it works better in evolutionarily closer species and becomes less useful as the evolutionary distance increases between the species being compared. Around 60% of protein-coding genes in *Drosophila melanogaster* have human homologues (Wangler et al. 2015), which is often not enough to find orthologous lncRNA genes consistently by the analysis of synteny. Nevertheless, the number of syntenic lincRNAs found in flies and mice is significantly higher than expected by chance, suggesting that a subset of those could be actual orthologs (Young et al. 2012), paving the way for further studies of lncRNAs in *Drosophila*.

Despite lacking sequence conservation, smaller regions of homology among different species have been observed for lncRNAs (Hezroni et al. 2015; Quinn et al. 2016; Ulitsky et al. 2011). These microhomologous regions are thought to correspond to functional elements that are essential for the function of the lncRNA, such as RNA-binding protein motifs or miRNA-binding sites. Recent studies have used a novel approach to identify orthologous lncRNAs based on the identification of these regions of microhomology. It is important to note that RNA-binding protein motifs or miRNA-binding sites are very short (between 4 and 12 nucleotides) and individual matches between different species can be found purely by chance (Bartel 2018). An interesting approach to bypass the rate of false positive hits is the addition of order to these elements (Ross et al. 2021). In this way, not only the presence of these motifs is considered, but also the order in which they are found in putatively orthologous lncRNAs. Although this method has not been tested for distantly related species, finding small regions of homology should be more achievable than finding orthologous lncRNAs using the current methods based on whole-sequence similarity or secondary structure predictions.

Another type of conservation analysis is the study of lncRNA secondary structures, which are thought to be more conserved than the primary sequence (Graf and Kretz 2020; Smith et al. 2013). Unfortunately, the currently available secondary structure predicting tools are not very accurate. Most of these programmes use a minimum free energy algorithm, which predicts the optimal secondary structure that requires the minimum energy to fold. However, features such as the presence of RNA-binding motifs, which should be located outside of the main structure to be properly recognised by their binding proteins, are not taken into consideration. Although it could be a reliable alternative for confirming the orthologous genes identified by other methods, the current secondary structure prediction tools do not seem accurate enough to consider the RNA folding form as the primary source to identify orthologous lncRNAs.

Another approach to identify potential lncRNAs orthologous is through the analysis of their pattern of binding to protein coding genes. If lncRNAs from different species bind to the same orthologous protein coding genes, they may exert a similar function. Methods have been developed to estimate the binding propensity of protein-RNA pairs in silico (Agostini et al. 2013; Armaos et al. 2021; Bellucci et al. 2011). However, to date, this approach has not been tested on a large scale to identify ortholog lncRNAs. To date, no reliable methods exist to systematically establish conservation among lncRNAs in evolutionarily distant species like flies and humans. However, the huge amount of effort made in that direction and the increasing number of annotated transcripts that will emerge in the coming years, hint to a promising perspective regarding lncRNA orthology. The fact that many functions and features associated with lncRNAs are conserved in *Drosophila* reinforces its extraordinary potential as a model organism to functionally characterise and model lncRNAs.

Finally, the characterisation of genomes across the tree of life will provide an incredible amount of data to perform comparative analyses. Advances in sequencing technologies that enable the identification of complete genomes have led to the emergence of the Earth BioGenome Project, an international collaboration that aims to sequence, catalogue and characterise the genomes of all eukaryotes on Earth (Lewin et al. 2018). One of the outcomes of this project is the production of new knowledge on the organisation and evolution of genomes, which could also have a major impact on the field of lncRNAs.

Author contributions CC-R, MC and MR-R conceived the idea of this article and performed the literature search. Data analysis was

performed by CC-R, RA, CCK and MR-R. The first draft of the manuscript was written by CC-R, MC and MR-R and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This project was funded by Grants: BFU2015-67623-P and PGC2018-099763-B100, 2017SGR1455 from AGAUR (Generalitat de Catalunya) and a Grant from the Institució Catalana de Recerca i Estudis Avançats (via an ICREA Academia award) to M.C.; as well as support from the European Community under the FP7 programme (ERC-2011-AdG-294653-RNA-MAPS), 2U41 HG007234-05G from NIH/NHGRI (USA) and the Grant PGC2018-094017-B-I00 from the Spanish Ministerio de Ciencia, Innovación y Universidades to R.G.. C.C.-R. holds a predoctoral FPI contract from the Spanish Government (Ministerio de Ciencia e Innovación). R.A. is a predoctoral fellow of CONACYT, the "Becas al Extranjero" Program of Mexico.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval Not applicable.

Consent to participate All authors consent to participate in this manuscript.

Consent for publication All authors consent for the publication of this manuscript.

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW et al (2000) The genome sequence of Drosophila melanogaster. Science 287:2185–2195

Agostini F, Zanzoni A, Klus P, Marchese D, Cirillo D, Tartaglia GG (2013) catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. Bioinformatics 29:2928–2930. https://doi.org/10.1093/bioinformatics/btt495

Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM et al (2015) A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell 160:595–606. https://doi.org/10.1016/j.cell.2015.01.009

Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X et al (2014) An atlas of active enhancers across human cell types and tissues. Nature 507:455–461. https://doi.org/10.1038/nature12787

Andrews SJ, Rothnagel JA (2014) Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet 15:193–204. https://doi.org/10.1038/nrg3520

Armaos A, Colantoni A, Proietti G, Rupert J, Tartaglia GG (2021) cat RAPID omics v2.0: going deeper and wider in the prediction of protein–RNA interactions. Nucleic Acids Res. https://doi.org/10.1093/nar/gkab393

Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drabløs F, Lennartsson A, Rönnerblad M et al (2015) Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science 347:1010–1014. https://doi.org/10.1126/science.1259418

Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MAS, Brocard M, Couso J-P (2014) Extensive translation of small Open reading frames revealed by poly-Ribo-Seq. Elife 3:1–19. https://doi.org/10.7554/eLife.03528

Bartel DP (2018) Metazoan microRNAs. Cell 173:20–51. https://doi.org/10.1016/j.cell.2018.03.006

Bazzini AA, Johnstone TG, Christiano R, MacKowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT et al (2014) Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J 33:981–993. https://doi.org/10.1002/embj.201488411

Beckmann K, Grskovic M, Gebauer F, Hentze MW (2005) A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in Drosophila. Cell 122:529–540. https://doi.org/10.1016/j.cell.2005.06.011

Bellucci M, Agostini F, Masin M, Tartaglia GG (2011) Predicting protein associations with long noncoding RNAs. Nat Methods 8:444–445. https://doi.org/10.1038/nmeth.1611

Beltran M, Puig I, Pena C, Garcia JM, Alvarez AB, Pena R, Bonilla F, de Herreros AG (2008) A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. Genes Dev 22:756–769. https://doi.org/10.1101/gad.455708

Bender W (2008) MicroRNAs in the Drosophila bithorax complex. Genes Dev 22:14–19. https://doi.org/10.1101/gad.1614208

Bier E (2005) Drosophila, the golden bug, emerges as a tool for human genetics. Nat Rev Genet 6:9–23. https://doi.org/10.1038/nrg1503

Bone JR, Lavender J, Richman R, Palmer MJ, Turner BM, Kuroda MI (1994) Acetylated histone H4 on the male X chromosome is associated with dosage compensation in Drosophila. Genes Dev 8:96–104. https://doi.org/10.1101/gad.8.1.96

Briggs JA, Wolvetang EJ, Mattick JS, Rinn JL, Barry G (2015) Mechanisms of long non-coding RNAs in mammalian nervous system development, plasticity, disease, and evolution. Neuron 88:861–877. https://doi.org/10.1016/j.neuron.2015.09.045

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S (1992) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71:515–526. https://doi.org/10.1016/0092-8674(92)90519-I

Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature 349:38–44. https://doi.org/10.1038/349038a0

Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is

highly localized within the nucleus. Cell 71:527–542. https://doi.org/10.1016/0092-8674(92)90520-M

Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, Booth BW, Wen J et al (2014) Diversity and dynamics of the Drosophila transcriptome. Nature 512:393–399. https://doi.org/10.1038/nature12962

Bryzghalov O, Makałowska I, Szcześniak MW (2021) lncEvo: automated identification and conservation study of long noncoding RNAs. BMC Bioinform. https://doi.org/10.1186/s12859-021-03991-2

Cai X, Cullen BR (2007) The imprinted H19 noncoding RNA is a primary microRNA precursor. RNA 13:313–316. https://doi.org/10.1261/rna.351707

Cao L, Zhang P, Li J, Wu M (2017) Last, a c-Myc-inducible long noncoding RNA, cooperates with CNBP to promote CCND1 MRNA stability in human cells. Elife. https://doi.org/10.7554/eLife.30433

Carlevaro-Fita J, Rahim A, Guigó R, Vardy LA, Johnson R (2016) Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. RNA 22:867–882. https://doi.org/10.1261/rna.053561.115

Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I et al (2012) Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature 491:454–457. https://doi.org/10.1038/nature11508

Chan CS, Rastelli L, Pirrotta V (1994) A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. EMBO J 13:2553–2564. https://doi.org/10.1002/j.1460-2075.1994.tb06545.x

Chen RAJ, Down TA, Stempor P, Chen QB, Egelhofer TA, Hillier LDW, Jeffers TE, Ahringer J (2013) The landscape of RNA polymerase II transcription initiation in *C. elegans* reveals promoter and enhancer architectures. Genome Res 23:1339–1347. https://doi.org/10.1101/gr.153668.112

Chen B, Zhang Y, Zhang X, Jia S, Chen S, Kang L (2016) Genome-wide identification and developmental expression profiling of long noncoding RNAs during Drosophila metamorphosis. Sci Rep 6:1–8. https://doi.org/10.1038/srep23330

Consortium MGS, Brent MR, Birren BW, Antonarakis SE, Alexandersson M, Zody M, Birney E, Baertsch R et al (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562. https://doi.org/10.1038/nature01262

Constanty F, Shkumatava A (2021) LncRNAs in development and differentiation: from sequence motifs to functional characterization. Development. https://doi.org/10.1242/dev.182741

Couso J-P, Patraquim P (2017) Classification and function of small open reading frames. Nat Rev Mol Cell Biol 18:575–589. https://doi.org/10.1038/nrm.2017.58

D'Lima NG, Ma J, Winkler L, Chu Q, Loh KH, Corpuz EO, Budnik BA, Lykke-Andersen J et al (2017) A human microprotein that interacts with the mRNA decapping complex. Nat Chem Biol 13:174–180. https://doi.org/10.1038/nchembio.2249

De Lara JCF, Arzate-Mejía RG, Recillas-Targa F (2019) Enhancer RNAs: insights into their biological role. Epigenet Insights. https://doi.org/10.1177/2516865719846093

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J et al (2010) A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. PLoS Biol. https://doi.org/10.1371/journal.pbio.1000384

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22:1775–1789. https://doi.org/10.1101/gr.132159.111

Dong Y, Liang G, Yuan B, Yang C, Gao R, Zhou X (2015) MALAT1 promotes the proliferation and metastasis of osteosarcoma cells by activating the PI3K/Akt pathway. Tumor Biol 36:1477–1486. https://doi.org/10.1007/s13277-014-2631-4

Fristrom D, Gotwals P, Eaton S, Kornberg TB, Sturtevant M, Bier E, Fristrom JW (1994) blistered: a gene required for vein/intervein formation in wings of Drosophila. Development 120:2661–2671

Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP (2007) Peptides encoded by short ORFs control development and define a new eukaryotic gene family. PLoS Biol 5:1052–1062. https://doi.org/10.1371/journal.pbio.0050106

Gao F, Cai Y, Kapranov P, Xu D (2020) Reverse-genetics studies of lncRNAs—what we have learnt and paths forward. Genome Biol 21:93. https://doi.org/10.1186/s13059-020-01994-5

Garaulet DL, Castellanos MC, Bejarano F, Sanfilippo P, Tyler DM, Allan DW, Sánchez-Herrero E, Lai EC (2014) Homeotic function of Drosophila bithorax-complex miRNAs mediates fertility by restricting multiple Hox genes and TALE cofactors in the CNS. Dev Cell 29:635–648. https://doi.org/10.1016/j.devcel.2014.04.023

Gebauer F, Merendino L, Hentze MW, Valcárcel J (1998) The Drosophila splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. RNA 4:142–150

Geisler SJ, Paro R (2015) Trithorax and Polycomb group-dependent regulation: a tale of opposing activities. Development 142:2876–2887. https://doi.org/10.1242/dev.120030

Gelbart ME, Larschan E, Peng S, Park PJ, Kuroda MI (2009) Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. Nat Struct Mol Biol 16:825–832. https://doi.org/10.1038/nsmb.1644

Gonzalez I, Munita R, Agirre E, Dittmer TA, Gysling K, Misteli T, Luco RF (2015) A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. Nat Struct Mol Biol 22:370–376. https://doi.org/10.1038/nsmb.3005

Graf J, Kretz M (2020) From structure to function: Route to understanding lncRNA mechanism. BioEssays. https://doi.org/10.1002/bies.202000027

Graindorge A, Carré C, Gebauer F (2013) Sex-lethal promotes nuclear retention of msl2 mRNA via interactions with the STAR protein HOW. Genes Dev 27:1421–1433. https://doi.org/10.1101/gad.214999.113

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, Van Baren MJ et al (2011) The developmental transcriptome of Drosophila melanogaster. Nature 471:473–479. https://doi.org/10.1038/nature09715

Grossniklaus U, Paro R (2014) Transcriptional silencing by polycomb-group proteins. Cold Spring Harb Perspect Biol 6:1–26. https://doi.org/10.1101/cshperspect.a019331

Grote P, Herrmann BG (2013) The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. RNA Biol 10:1579–1585. https://doi.org/10.4161/rna.26165

Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A, Macura K, Bläss G et al (2013) The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev Cell 24:206–214. https://doi.org/10.1016/j.devcel.2012.12.012

Gumireddy K, Li A, Yan J, Setoyama T, Johannes GJ, Ørom UA, Tchou J, Liu Q et al (2013) Identification of a long non-coding RNA-associated RNP complex regulating metastasis at the translational step. EMBO J 32:2672–2684. https://doi.org/10.1038/emboj.2013.188

Gummalla M, Maeda RK, Alvarez JJ, Gyurkovics H, Singari S, Edwards KA, Karch F, Bender W (2012) Abd-A regulation by the iab-8 noncoding RNA. PLoS Genet. https://doi.org/10.1371/journal.pgen.1002720

Hah N, Murakami S, Nagari A, Danko CG, Lee Kraus W (2013) Enhancer transcripts mark active estrogen receptor binding sites. Genome Res 23:1210–1223. https://doi.org/10.1101/gr.152306.112

Hansji H, Leung EY, Baguley BC, Finlay GJ, Cameron-Smith D, Figueiredo VC, Askarian-Amiri ME (2016) ZFAS1: a long non-coding RNA associated with ribosomes in breast cancer cells. Biol Direct 11:62. https://doi.org/10.1186/s13062-016-0165-y

He R-Z, Luo D-X, Mo Y-Y (2019) Emerging roles of lncRNAs in the post-transcriptional regulation in cancer. Genes Dis 6:6–15. https://doi.org/10.1016/j.gendis.2019.01.003

Hekimoglu B, Ringrose L (2009) Non-coding RNAs in polycomb/trithorax regulation. RNA Biol 6:129–137. https://doi.org/10.4161/rna.6.2.8178

Henriques T, Scruggs BS, Inouye MO, Muse GW, Williams LH, Burkholder AB, Lavender CA, Fargo DC et al (2018) Widespread transcriptional pausing and elongation control at enhancers. Genes Dev 32:26–41. https://doi.org/10.1101/gad.309351.117

Herrera-Úbeda C, Barba MM, Pérez EN, Gravemeyer J, Albuixech-Crespo B, Wheeler GN, Garcia-Fernàndez J (2019) Microsyntenic clusters reveal conservation of lncRNAs in chordates despite absence of sequence conservation. Biology (basel). https://doi.org/10.3390/biology8030061

Herzog VA, Lempradl A, Trupke J, Okulski H, Altmutter C, Ruge F, Boidol B, Kubicek S et al (2014) A strand-specific switch in noncoding transcription switches the function of a polycomb/trithorax response element. Nat Genet 46:973–981. https://doi.org/10.1038/ng.3058

Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I (2015) Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Rep 11:1110–1122. https://doi.org/10.1016/j.celrep.2015.04.023

Huang J-Z, Chen M, Chen D, Gao X-C, Zhu S, Huang H, Hu M, Zhu H et al (2017) A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth. Mol Cell 68:171-184.e6. https://doi.org/10.1016/j.molcel.2017.09.015

Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147:789–802. https://doi.org/10.1016/j.cell.2011.10.002

Ivaldi MS, Diaz LF, Chakalova L, Lee J, Krivega I, Dean A (2018) Fetal γ-globin genes are regulated by the BGLT3 long noncoding RNA locus. Blood 132:1963–1973. https://doi.org/10.1182/blood-2018-07-862003

Jandura A, Krause HM (2017) The new RNA world: growing evidence for long noncoding RNA functionality. Trends Genet 33:665–676. https://doi.org/10.1016/j.tig.2017.08.002

Ji J-Y, Han C, Deng W-M (2019) Understanding human diseases using Drosophila. J Genet Genomics 46:155–156. https://doi.org/10.1016/j.jgg.2019.04.001

Jiang W, Liu Y, Liu R, Zhang K, Zhang Y (2015) The lncRNA DEANR1 facilitates human endoderm differentiation by activating FOXA2 expression. Cell Rep 11:137–148. https://doi.org/10.1016/j.celrep.2015.03.008

Johnson R (2012) Long non-coding RNAs in Huntington's disease neurodegeneration. Neurobiol Dis 46:245–254. https://doi.org/10.1016/j.nbd.2011.12.006

Kallen AN, Zhou XB, Xu J, Qiao C, Ma J, Yan L, Lu L, Liu C et al (2013) The imprinted H19 lncRNA antagonizes Let-7 microRNAs. Mol Cell 52:101–112. https://doi.org/10.1016/j.molcel.2013.08.027

Kassis JA, Brown JL (2013) Polycomb group response elements in Drosophila and vertebrates. In: Kumar D (ed) Advances in genetics, 1st edn. Elsevier Inc., Cardiff, pp 83–118

Kassis JA, Kennison JA, Tamkun JW (2017) Polycomb and trithorax group genes in drosophila. Genetics 206:1699–1725. https://doi.org/10.1534/genetics.115.185116

Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, Smits G, Reik W (2012) The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nat Cell Biol 14:659–665. https://doi.org/10.1038/ncb2521

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M et al (2010) Widespread transcription at neuronal activity-regulated enhancers. Nature 465:182–187. https://doi.org/10.1038/nature09033

Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S, Kageyama Y (2007) Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. Nat Cell Biol 9:660–665. https://doi.org/10.1038/ncb1595

Lagarde J, Uszczynska-Ratajczak B, Carbonell S, Pérez-Lluch S, Abad A, Davis C, Gingeras TR, Frankish A et al (2017) High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. Nat Genet 49:1731–1740. https://doi.org/10.1038/ng.3988

Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R (2013) Activating RNAs associate with mediator to enhance chromatin architecture and transcription. Nature 494:497–501. https://doi.org/10.1038/nature11884

Lam MTY, Cho H, Lesch HP, Gosselin D, Heinz S, Tanaka-Oishi Y, Benner C, Kaikkonen MU et al (2013) Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature 498:511–515. https://doi.org/10.1038/nature12209

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K et al (2001) Correction: initial sequencing and analysis of the human genome. Nature 412:565–566. https://doi.org/10.1038/35087627

Lee JT, Strauss WM, Dausman JA, Jaenisch R (1996) A 450 kb transgene displays properties of the mammalian X-inactivation center. Cell 86:83–94. https://doi.org/10.1016/S0092-8674(00)80079-3

Lee DY, Moon J, Lee S-T, Jung K-H, Park D-K, Yoo J-S, Sunwoo J-S, Byun J-I et al (2015) Distinct expression of long non-coding RNAs in an Alzheimer's disease model. J Alzheimer's Dis 45:837–849. https://doi.org/10.3233/JAD-142919

Lee S, Kopp F, Chang TC, Sataluri A, Chen B, Sivakumar S, Yu H, Xie Y et al (2016) Noncoding RNA NORAD regulates genomic stability by sequestering Pumilio proteins. Cell 164:69–80. https://doi.org/10.1016/j.cell.2015.12.017

Lee H, Zhang Z, Krause HM (2019) Long noncoding RNAs and repetitive elements: junk or intimate evolutionary partners? Trends Genet 35:892–902. https://doi.org/10.1016/j.tig.2019.09.006

Legeai F, Derrien T (2015) Identification of long non-coding RNAs in insects genomes. Curr Opin Insect Sci 7:37–44. https://doi.org/10.1016/j.cois.2015.01.003

Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, Durbin R, Edwards SV et al (2018) Earth BioGenome project: sequencing life for the future of life. Proc Natl Acad Sci USA 115:4325–4333

Li M, Wen S, Guo X, Bai B, Gong Z, Liu X, Wang Y, Zhou Y et al (2012) The novel long non-coding RNA CRG regulates Drosophila locomotor behavior. Nucleic Acids Res 40:11714–11727. https://doi.org/10.1093/nar/gks943

Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J et al (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature 498:516–520. https://doi.org/10.1038/nature12210

Li H, Yu B, Li J, Su L, Yan M, Zhu Z, Liu B (2014) Overexpression of lncRNA H19 enhances carcinogenesis and metastasis of gastric

cancer. Oncotarget 5:2318–2329. https://doi.org/10.18632/oncotarget.1913
Li F, Cao L, Hang D, Wang F, Wang Q (2015) Long non-coding RNA HOTTIP is up-regulated and associated with poor prognosis in patients with osteosarcoma. Int J Clin Exp Pathol 8:11414–11420
Li K, Tian Y, Yuan Y, Fan X, Yang M, He Z, Yang D (2019) Insights into the functions of lncRNAs in Drosophila. Int J Mol Sci 20:4646. https://doi.org/10.3390/ijms20184646
Liu SJ, Horlbeck MA, Cho SW, Birk HS, Malatesta M, He D, Attenello FJ, Villalta JE et al (2017) CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science 355:7111. https://doi.org/10.1126/science.aah7111
Liu Y, Cao Z, Wang Y, Guo Y, Xu P, Yuan P, Liu Z, He Y et al (2018) Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. Nat Biotechnol 36:1203–1210. https://doi.org/10.1038/nbt.4283
Liu PY, Tee AE, Milazzo G, Hannan KM, Maag J, Mondal S, Atmadibrata B, Bartonicek N et al (2019) The long noncoding RNA lncNB1 promotes tumorigenesis by interacting with ribosomal protein RPL35. Nat Commun 10:5026. https://doi.org/10.1038/s41467-019-12971-3
Lo Piccolo L, Yamaguchi M (2017) RNAi of arcRNA hsrω affects sub-cellular localization of Drosophila FUS to drive neurodiseases. Exp Neurol 292:125–134. https://doi.org/10.1016/j.expneurol.2017.03.011
Maeda RK, Sitnik JL, Frei Y, Prince E, Gligorov D, Wolfner MF, Karch F (2018) The lncRNA male-specific abdominal plays a critical role in Drosophila accessory gland development and male fertility. PLoS Genet. https://doi.org/10.1371/journal.pgen.1007519
Magny EG, Pueyo JI, Pearl FMG, Cespedes MA, Niven JE, Bishop SA, Couso JP (2013) Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. Science 341:1116–1120. https://doi.org/10.1126/science.1238802
Mallo M, Alonso CR (2013) The regulation of Hox gene expression during animal development. Development 140:3951–3963. https://doi.org/10.1242/dev.068346
Meers MP, Adelman K, Duronio RJ, Strahl BD, McKay DJ, Matera AG (2018) Transcription start site profiling uncovers divergent transcription and enhancer-associated RNAs in Drosophila melanogaster. BMC Genomics. https://doi.org/10.1186/s12864-018-4510-7
Meller VH, Rattner BP (2002) The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. EMBO J 21:1084–1091. https://doi.org/10.1093/emboj/21.5.1084
Mikhaylichenko O, Bondarenko V, Harnett D, Schor IE, Males M, Viales RR, Furlong EEM (2018) The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. Genes Dev 32:42–57. https://doi.org/10.1101/gad.308619.117
Montagne J, Groppe J, Guillemin K, Krasnow MA, Gehring WJ, Affolter M (1996) The Drosophila serum response factor gene is required for the formation of intervein tissue of the wing and is allelic to blistered. Development 122:2589–2597
Murillo-Maldonado JM, Riesgo-Escovar JR (2019) The various and shared roles of lncRNAs during development. Dev Dyn 248:1059–1069. https://doi.org/10.1002/dvdy.108
Nelson BR, Makarewich CA, Anderson DM, Winders BR, Troupes CD, Wu F, Reese AL, McAnally JR et al (2016) A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science 351:271–275. https://doi.org/10.1126/science.aad4076
Onyango P, Feinberg AP (2011) A nucleolar protein, H19 opposite tumor suppressor (HOTS), is a tumor growth inhibitor encoded by a human imprinted H19 antisense transcript. Proc Natl Acad Sci USA 108:16759–16764. https://doi.org/10.1073/pnas.1110904108
Paytuví Gallart A, Hermoso Pulido A, Martínez A, de Lagrán I, Sanseverino W, Aiese Cigliano R (2016) GREENC: a Wiki-based database of plant lncRNAs. Nucleic Acids Res 44:D1161–D1166. https://doi.org/10.1093/nar/gkv1215
Pegueroles C, Iraola-Guzmán S, Chorostecki U, Ksiezopolska E, Saus E, Gabaldón T (2019) Transcriptomic analyses reveal groups of co-expressed, syntenic lncRNAs in four species of the genus Caenorhabditis. RNA Biol 16:320–329. https://doi.org/10.1080/15476286.2019.1572438
Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N (1996) Requirement for Xist in X chromosome inactivation. Nature 379:131–137. https://doi.org/10.1038/379131a0
Pérez-Lluch S, Klein CC, Breschi A, Ruiz-Romero M, Abad A, Palumbo E, Bekish L, Arnan C et al (2020) bsAS, an antisense long non-coding RNA, essential for correct wing development through regulation of blistered/DSRF isoform usage. PLoS Genet. https://doi.org/10.1371/journal.pgen.1009245
Perry RB-T, Ulitsky I (2016) The functions of long noncoding RNAs in development and stem cells. Development 143:3882–3894. https://doi.org/10.1242/dev.140962
Pervouchine DD, Djebali S, Breschi A, Davis CA, Barja PP, Dobin A, Tanzer A, Lagarde J et al (2015) Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. Nat Commun 6:5903. https://doi.org/10.1038/ncomms6903
Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP et al (2003) Role of histone H3 lysine 27 methylation in X inactivation. Science 300:131–135. https://doi.org/10.1126/science.1084274
Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, Vergara IA, Davicioni E et al (2013) The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat Genet 45:1392–1398. https://doi.org/10.1038/ng.2771
Pueyo JI, Couso JP (2011) Tarsal-less peptides control Notch signalling through the Shavenbaby transcription factor. Dev Biol 355:183–193. https://doi.org/10.1016/j.ydbio.2011.03.033
Quinn JJ, Zhang QC, Georgiev P, Ilik IA, Akhtar A, Chang HY (2016) Rapid evolutionary turnover underlies conserved lncRNA-genome interactions. Genes Dev 30:191–207. https://doi.org/10.1101/gad.272187.115
Rahnamoun H, Lee J, Sun Z, Lu H, Ramsey KM, Komives EA, Lauberth SM (2018) RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. Nat Struct Mol Biol 25:687–697. https://doi.org/10.1038/s41594-018-0102-0
Ramilowski JA, Yip CW, Agrawal S, Chang J-C, Ciani Y, Kulakovskiy IV, Mendez M, Ooi JLC et al (2020) Corrigendum: functional annotation of human long noncoding RNAs via molecular phenotyping. Genome Res 30:1377. https://doi.org/10.1101/gr.270330.120
Ringrose L (2017) Noncoding RNAs in polycomb and trithorax regulation: a quantitative perspective. Annu Rev Genet. https://doi.org/10.1146/annurev-genet-120116
Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA et al (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129:1311–1323. https://doi.org/10.1016/j.cell.2007.05.022
Ríos-Barrera LD, Gutiérrez-Pérez I, Domínguez M, Riesgo-Escovar JR (2015) acal is a long non-coding RNA in JNK signaling in epithelial shape changes during Drosophila dorsal closure. PLoS Genet 11:1–27. https://doi.org/10.1371/journal.pgen.1004927

Roch F, Baonza A, Martín-Blanco E, García-Bellido A (1998) Genetic interactions and cell behaviour in blistered mutants during proliferation and differentiation of the Drosophila wing. Development 125:1823–1832

Rogoyski OM, Pueyo JI, Couso JP, Newbury SF (2017) Functions of long non-coding RNAs in human disease and their conservation in Drosophila development. Biochem Soc Trans 45:895–904. https://doi.org/10.1042/BST20160428

Rolland AD, Evrard B, Darde TA, Le Beguec C, Le Bras Y, Bensalah K, Lavoue S, Jost B et al (2019) RNA profiling of human testicular cells identifies syntenic lncRNAs associated with spermatogenesis. Hum Reprod 34:1278–1290. https://doi.org/10.1093/humrep/dez063

Romero-Barrios N, Legascue MF, Benhamed M, Ariel F, Crespi M (2018) Splicing regulation by long noncoding RNAs. Nucleic Acids Res 46:2169–2184. https://doi.org/10.1093/nar/gky095

Ross CJ, Rom A, Spinrad A, Gelbard-Solodkin D, Degani N, Ulitsky I (2021) Uncovering deeply conserved motif combinations in rapidly evolving noncoding sequences. Genome Biol. https://doi.org/10.1186/s13059-020-02247-1

Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM (2014) Long noncoding RNAs as a source of new peptides. Elife 3:1–24. https://doi.org/10.7554/eLife.03523

Samata M, Akhtar A (2018) Dosage compensation of the X chromosome: a complex epigenetic assignment involving chromatin regulators and long noncoding RNAs. Annu Rev Biochem 87:323–350. https://doi.org/10.1146/annurev-biochem-062917-011816

Sarropoulos I, Marin R, Cardoso-Moreira M, Kaessmann H (2019) Developmental dynamics of lncRNAs across mammalian organs and species. Nature 571:510–514. https://doi.org/10.1038/s41586-019-1341-x

Schaukowitch K, Joo JY, Liu X, Watts JK, Martinez C, Kim TK (2014) Enhancer RNA facilitates NELF release from immediate early genes. Mol Cell 56:29–42. https://doi.org/10.1016/j.molcel.2014.08.023

Schor IE, Bussotti G, Maleš M, Forneris M, Viales RR, Enright AJ, Furlong EEM (2018) Non-coding RNA expression, function, and variation during Drosophila embryogenesis. Curr Biol 28:3547–3561. https://doi.org/10.1016/j.cub.2018.09.026

Schuettengruber B, Bourbon H-M, Di Croce L, Cavalli G (2017) Genome regulation by polycomb and trithorax: 70 years and counting. Cell 171:34–57. https://doi.org/10.1016/j.cell.2017.08.002

Schwarzer A, Emmrich S, Schmidt F, Beck D, Ng M, Reimer C, Adams FF, Grasedieck S et al (2017) The non-coding RNA landscape of human hematopoiesis and leukemia. Nat Commun 8:1–17. https://doi.org/10.1038/s41467-017-00212-4

Simon J, Chiang A, Bender W, Shimell MJ, O'Connor M (1993) Elements of the Drosophila bithorax complex that mediate repression by polycomb group products. Dev Biol 158:131–144. https://doi.org/10.1006/dbio.1993.1174

Singer RA, Arnes L, Cui Y, Wang J, Gao Y, Guney MA, Burnum-Johnson KE, Rabadan R et al (2019) The long noncoding RNA paupar modulates PAX6 regulatory activities to promote alpha cell development and function. Cell Metab 30:1091–1106. https://doi.org/10.1016/j.cmet.2019.09.013

Slavoff SA, Heo J, Budnik BA, Hanakahi LA, Saghatelian A (2014) A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining. J Biol Chem 289:10950–10957. https://doi.org/10.1074/jbc.C113.533968

Smith MA, Gesell T, Stadler PF, Mattick JS (2013) Widespread purifying selection on RNA structure in mammals. Nucleic Acids Res 41:8220–8236. https://doi.org/10.1093/nar/gkt596

Soshnev AA, Ishimoto H, Mcallister BF, Li X, Wehling MD, Kitamoto T, Geyer PK (2011) A conserved long noncoding RNA affects sleep behavior in Drosophila. Genetics 189(2):455–468. https://doi.org/10.1534/GENETICS.111.131706

St Johnston D (2002) The art and design of genetic screens: Drosophila melanogaster. Nat Rev Genet 3:176–188. https://doi.org/10.1038/nrg751

Statello L, Guo C-J, Chen L-L, Huarte M (2021) Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol 22:96–118. https://doi.org/10.1038/s41580-020-00315-9

Steffen PA, Ringrose L (2014) What are memories made of? How polycomb and trithorax proteins mediate epigenetic memory. Nat Rev Mol Cell Biol 15:340–356. https://doi.org/10.1038/nrm3789

Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF et al (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell 39:925–938. https://doi.org/10.1016/j.molcel.2010.08.011

Tsai PF, Dell'Orso S, Rodriguez J, Vivanco KO, Ko KD, Jiang K, Juan AH, Sarshad AA et al (2018) A muscle-specific enhancer RNA mediates cohesin recruitment and regulates transcription in trans. Mol Cell 71:129–141. https://doi.org/10.1016/j.molcel.2018.06.008

Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell 147:1537–1550. https://doi.org/10.1016/j.cell.2011.11.055

van Heesch S, Van Iterson M, Jacobi J, Boymans S, Essers PB, De Bruijn E, Hao W, MacInnes AW et al (2014) Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. Genome Biol. https://doi.org/10.1186/gb-2014-15-1-r6

van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, Kirchner M, Maatz H et al (2019) The translational landscape of the human heart. Cell 178:242-260.e29. https://doi.org/10.1016/j.cell.2019.05.010

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M et al (2001) The sequence of the human genome. Science 291:1304–1351. https://doi.org/10.1126/science.1058040

Wang J, Wang X, Bhat A, Chen Y, Xu K, Mo YY, Yi SS, Zhou Y (2020) Comprehensive network analysis reveals alternative splicing-related lncRNAs in hepatocellular carcinoma. Front Genet. https://doi.org/10.3389/fgene.2020.00659

Wangler MF, Yamamoto S, Bellen HJ (2015) Fruit flies in biomedical research. Genetics 199:639–653. https://doi.org/10.1534/genetics.114.171785

Wen K, Yang L, Xiong T, Di C, Ma D, Wu M, Xue Z, Zhang X et al (2016) Critical roles of long noncoding RNAs in Drosophila spermatogenesis. Genome Res 26:1233–1244. https://doi.org/10.1101/gr.199547.115

Wu Z-H, Wang X-L, Tang H-M, Jiang T, Chen J, Lu S, Qiu G-Q, Peng Z-H et al (2014) Long non-coding RNA HOTAIR is a powerful predictor of metastasis and poor prognosis and is associated with epithelial-mesenchymal transition in colon cancer. Oncol Rep 32:395–402. https://doi.org/10.3892/or.2014.3186

Wutz A, Jaenisch R (2000) A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell 5:695–705. https://doi.org/10.1016/S1097-2765(00)80248-8

Wutz A, Rasmussen TP, Jaenisch R (2002) Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat Genet 30:167–174. https://doi.org/10.1038/ng820

Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ et al (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol Cell 38:662–674. https://doi.org/10.1016/j.molcel.2010.03.021

Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, Huarte M, Zhan M et al (2012) LincRNA-p21 suppresses target mRNA translation. Mol Cell 47:648–655. https://doi.org/10.1016/j.molcel.2012.06.027

Young RS, Marques AC, Tibbit C, Haerty W, Bassett AR, Liu J-L, Ponting CP (2012) Identification and properties of 1,119 candidate LincRNA loci in the drosophila melanogaster genome. Genome Biol Evol 4:427–442. https://doi.org/10.1093/gbe/evs020

Zhou Y, Zhong Y, Wang Y, Zhang X, Batista DL, Gejman R, Ansell PJ, Zhao J et al (2007) Activation of p53 by MEG3 non-coding RNA. J Biol Chem 282:24731–24742. https://doi.org/10.1074/jbc.M702029200

Zu T, Liu Y, Banez-Coronel M, Reid T, Pletnikova O, Lewis J, Miller TM, Harms MB et al (2013) RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc Natl Acad Sci 110:E4968–E4977. https://doi.org/10.1073/pnas.1315438110