



Genomic and functional conservation of lncRNAs: lessons from flies

Carlos Camilleri-Robles¹ · Raziel Amador^{1,2} · Cecilia C. Klein^{1,2} · Roderic Guigó^{2,3} · Montserrat Corominas¹ · Marina Ruiz-Romero²

Received: 17 June 2021 / Accepted: 9 December 2021
© The Author(s) 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

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 XO 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 lncRNA 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

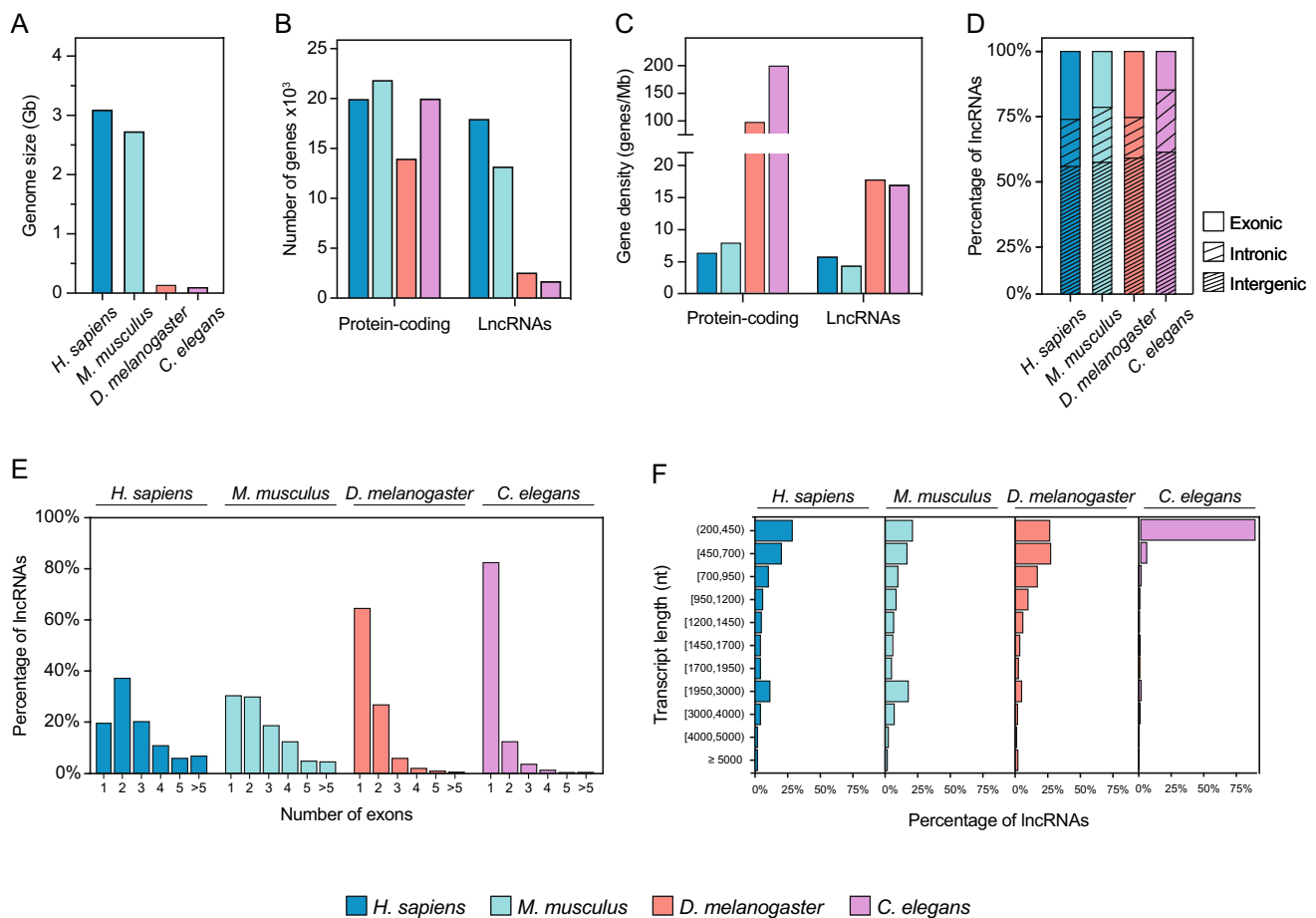


Fig. 1 Genomic and transcriptomic comparison of humans, mice and flies. **a** Barplot showing the size (gigabases—Gb) of the human (*H. sapiens*), mouse (*M. musculus*), fly (*D. melanogaster*) and worm (*C. elegans*) genome. **b** Number of annotated protein-coding and long non-coding RNA (lncRNA) genes 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 API (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%)

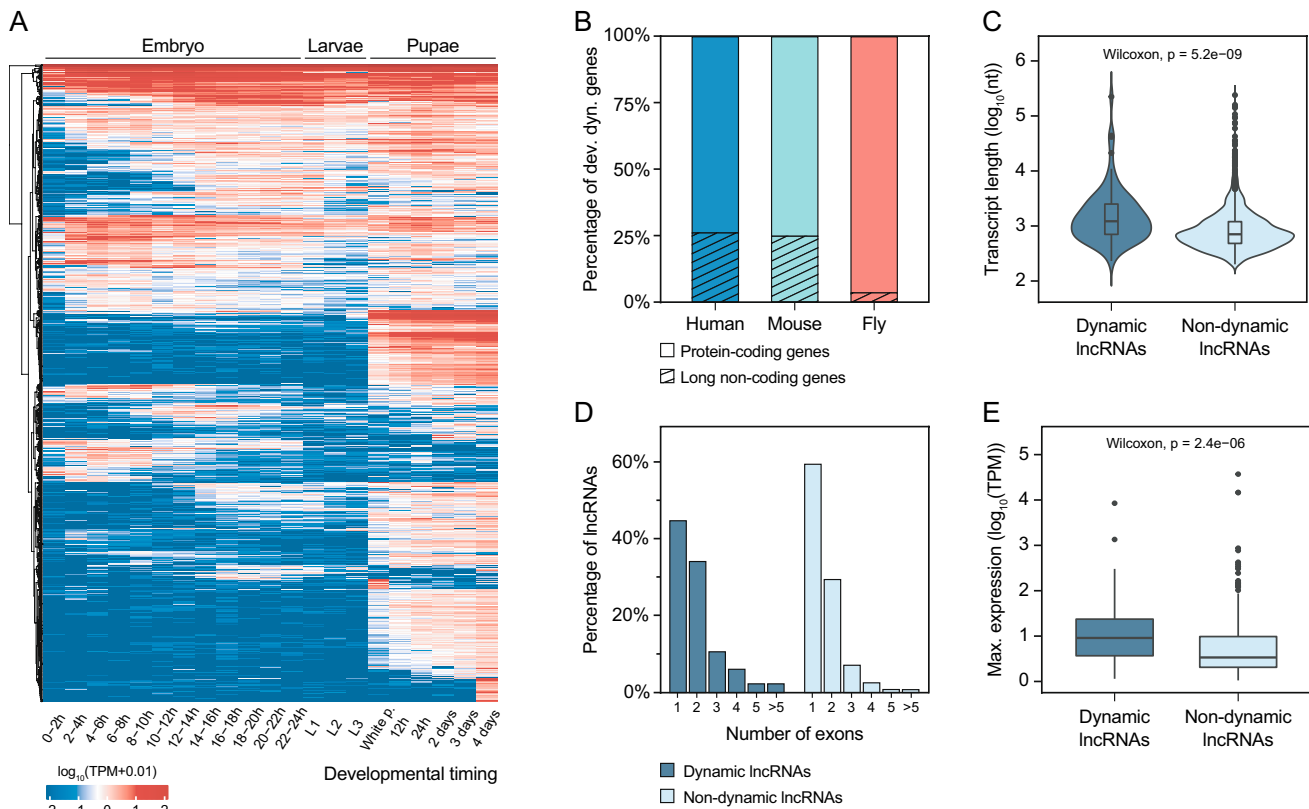


Fig. 2 Properties of *Drosophila* lncRNAs in development. **a** Heatmap showing the expression of all the annotated lncRNAs in *Drosophila* RNA-Seq samples from different developmental stages. **b** Percentage of developmentally dynamic genes corresponding to protein-coding and lncRNA genes in humans, 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** Box-plot 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). Sarpoulos 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 lncRNAs 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 orthologous 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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

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) catRAPID 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, Rønnerblad 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/emboj.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\(92\)90519-1](https://doi.org/10.1016/0092-8674(92)90519-1)
- 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](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 non-coding 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, Miettton 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-Fernández J (2019) Microsynthetic 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, Uszczyńska-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, Cesarini 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](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 insect 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 hsr ω affects sub-cellular localization of *Drosophila* FUS to drive neurodegenerative diseases. *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, Groppa 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>
- Rahmoun 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, Bruggmann 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) aca1 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
- Rogowski 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 non-coding 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>
- Schawkowitch 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](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>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.