

RESEARCH ARTICLE

Open Access



# Pleiotropic fitness effects of the lncRNA *Uhg4* in *Drosophila melanogaster*

Rebecca A. MacPherson<sup>1</sup>, Vijay Shankar<sup>1</sup>, Lakshmi T. Sunkara<sup>2</sup>, Rachel C. Hannah<sup>1</sup>, Marion R. Campbell III<sup>1</sup>, Robert R. H. Anholt<sup>1\*</sup>  and Trudy F. C. Mackay<sup>1\*</sup>

## Abstract

**Background:** Long noncoding RNAs (lncRNAs) are a diverse class of RNAs that are critical for gene regulation, DNA repair, and splicing, and have been implicated in development, stress response, and cancer. However, the functions of many lncRNAs remain unknown. In *Drosophila melanogaster*, *U snoRNA host gene 4* (*Uhg4*) encodes an antisense long noncoding RNA that is host to seven small nucleolar RNAs (snoRNAs). *Uhg4* is expressed ubiquitously during development and in all adult tissues, with maximal expression in ovaries; however, it has no annotated function(s).

**Results:** We used CRISPR-*Cas9* germline gene editing to generate multiple deletions spanning the promoter region and first exon of *Uhg4*. Females showed arrested egg development and both males and females were sterile. In addition, *Uhg4* deletion mutants showed delayed development and decreased viability, and changes in sleep and responses to stress. Whole-genome RNA sequencing of *Uhg4* deletion flies and their controls identified co-regulated genes and genetic interaction networks associated with *Uhg4*. Gene ontology analyses highlighted a broad spectrum of biological processes, including regulation of transcription and translation, morphogenesis, and stress response.

**Conclusion:** *Uhg4* is a lncRNA essential for reproduction with pleiotropic effects on multiple fitness traits.

**Keywords:** Noncoding RNA, Systems genetics, Oviposition, Stress response, CRISPR, snoRNA

## Background

Long noncoding RNAs (lncRNAs) are a diverse class of non-coding RNAs of at least 200 nucleotides in length. Although lncRNAs were initially thought to be “junk” due to their noncoding status, we now know that lncRNAs are critical for various biological processes, including, but not limited to, transcription and gene regulation [1–8], chromatin architecture [9–11], DNA damage response [12–14], and scaffolding and nuclear organization [15–18]. lncRNAs can also act as miRNA sponges [19–21], regulate gene splicing [17, 22–25], and be translated into functional peptides [26–28]. lncRNAs can be localized in the nucleus,

cytoplasm, or to a specific organelle [29–32], can act in *cis* or in *trans* [8, 23, 33], and may be conserved across taxa in sequence and/or function [34–36]. Across taxa, lncRNA dysregulation has been implicated in cancer [37–40], neurological disorders [6, 41–43], and immune and stress response [44–47]. Additional studies in *Drosophila melanogaster* have demonstrated critical roles for lncRNAs in development [48–53], gonadal function [48, 54–56], sleep [19], locomotion [57], and courtship behavior [58, 59].

Roles for lncRNAs in development, viability, and fertility have been identified in multiple model systems [4, 8, 34, 55, 60, 61]. However, fitness roles for some mammalian lncRNAs, including lncRNAs previously linked to fitness traits with gene-knockdown experiments, have not been replicated using CRISPR-based cell line or animal model knockouts [62–65]. Other recently discovered mammalian lncRNAs with expression limited to reproductive tissues do not affect reproductive phenotypes in

\*Correspondence: ranholt@clermson.edu; tmackay@clermson.edu

<sup>1</sup> Center for Human Genetics and Department of Genetics and Biochemistry, Clemson University, 114 Gregor Mendel Circle, Greenwood, SC 29646, USA  
Full list of author information is available at the end of the article



knockout mice [66, 67]. This controversy also extends beyond mammalian systems. *Drosophila* and zebrafish studies on CRISPR-mediated deletions of developmentally-expressed lncRNAs – some of which were previously implicated in development via RNA interference or morpholino-induced knockdown studies – also failed to identify roles for these lncRNAs in development, viability, or embryogenesis after no overt phenotypes were detected in knockout animals [68–70]. Inconsistencies across lncRNA studies may be due to the method of gene perturbation (e.g., RNAi interference, CRISPR-Cas9), discordant phenotyping, transcriptional noise competing with low expression of some lncRNAs, differences in the genetic background of organisms used across studies on the same lncRNA, and/or functional redundancy of the lncRNA [61, 69, 71, 72]. Furthermore, given the abundance of sex-specific and tissue-specific lncRNA expression, it is also possible that broader developmental pathways may overshadow effects due to loss of a lncRNA or be limited to a single tissue or behavioral phenotype [61, 72].

Here, we evaluate the effects of loss of function alleles of the *D. melanogaster* gene encoding *U snoRNA host gene 4* (*Uhg4*; FBgn0083124). *Uhg4* is a lncRNA with unknown function and is the host gene for seven intronic small nucleolar RNAs (snoRNAs), which guide posttranscriptional ribosomal RNA modification and processing [73, 74]. Some *U snoRNA* host genes interact with regulatory proteins controlling *piwi*-interacting RNAs (piRNAs) [75], which are noncoding RNAs involved in transposon silencing in germ cells [76, 77]. Although adult expression of *Uhg4* is highest in ovaries, it is expressed ubiquitously during development and in other adult tissues, including the brain and central nervous system, fat body, and trachea [78, 79]. *Uhg4* expression is correlated with modulation of expression of a subset of snoRNAs in response to developmental alcohol exposure in *Drosophila* females [80]. We used CRISPR/Cas9 germline gene editing to create deletions in the promoter region and first exon of *Uhg4*. These mutations have pleiotropic effects on fitness-related traits, stress responses, sleep and activity phenotypes, and transcript abundances of both non-coding and protein-coding RNAs.

## Results

### Generation of *Uhg4* Null Alleles

*Uhg4* is a long noncoding RNA that is host to seven small nucleolar RNAs (snoRNAs), and is expressed ubiquitously throughout development and in adults, with the highest adult expression in ovaries [78, 79]. We used CRISPR-Cas9 and a double guide RNA vector to target the deletion of a ~685 bp region that includes the promoter region upstream of *Uhg4* as well as the first exon

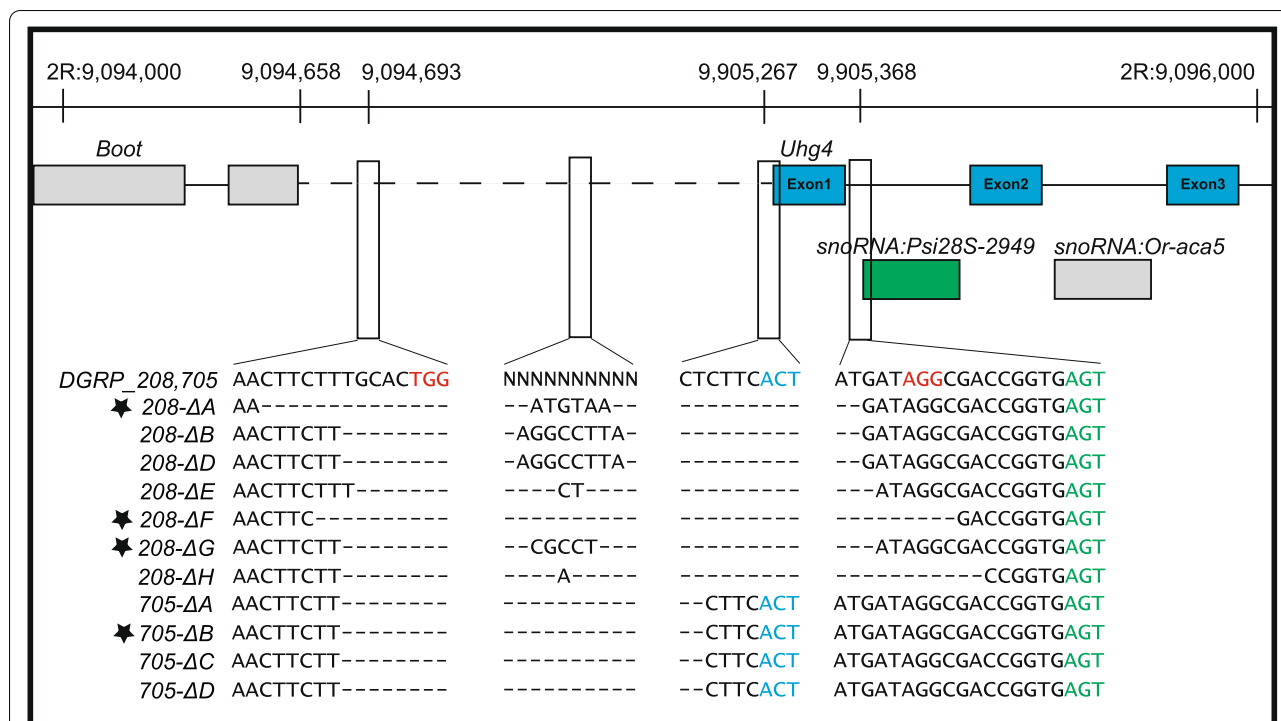
of *Uhg4* (Fig. 1) in two lines of the *Drosophila melanogaster* Genetic Reference Panel (DGRP [81]), DGRP lines, DGRP\_208 and DGRP\_705. We isolated seven independent deletion mutations in DGRP\_208, most of which varied from one another by a few base pairs (Fig. 1). The DGRP\_208 deletions spanned the promoter region and the first exon of *Uhg4*. We also isolated four independent identical mutations in DGRP\_705. The DGRP\_705 deletions removed the *Uhg4* promoter region and retained the first exon, and included a 44 bp AT-rich insertion in the first *Uhg4* intron upstream of the start of *snoRNA:Psi28S-2949* (Fig. 1). Sanger sequencing shows that the small nucleolar RNA (snoRNA) *snoRNA:Psi28S-2949* within the first exon of *Uhg4* is intact for all independently obtained deletions (Fig. 1).

Here, we focus on deletions *Uhg4*<sup>208-ΔA</sup>, *Uhg4*<sup>208-ΔF</sup>, *Uhg4*<sup>208-ΔG</sup>, and *Uhg4*<sup>705-ΔB</sup>, hereafter referred to as 208-ΔA, 208-ΔF, 208-ΔG, and 705-ΔB, respectively. We randomly selected these mutations for further study after several generations of backcrossing to the original genetic background. Both genetic backgrounds are highly inbred and free of inversions, and DGRP\_208 is free of *Wolbachia* infection [81]. Flies homozygous for the DGRP\_208 *Uhg4* deletions show changes in their resting wing position (File S1, S2). Compared to the control, *Uhg4* deletion flies carry their wings in an elevated and partially horizontally spread position.

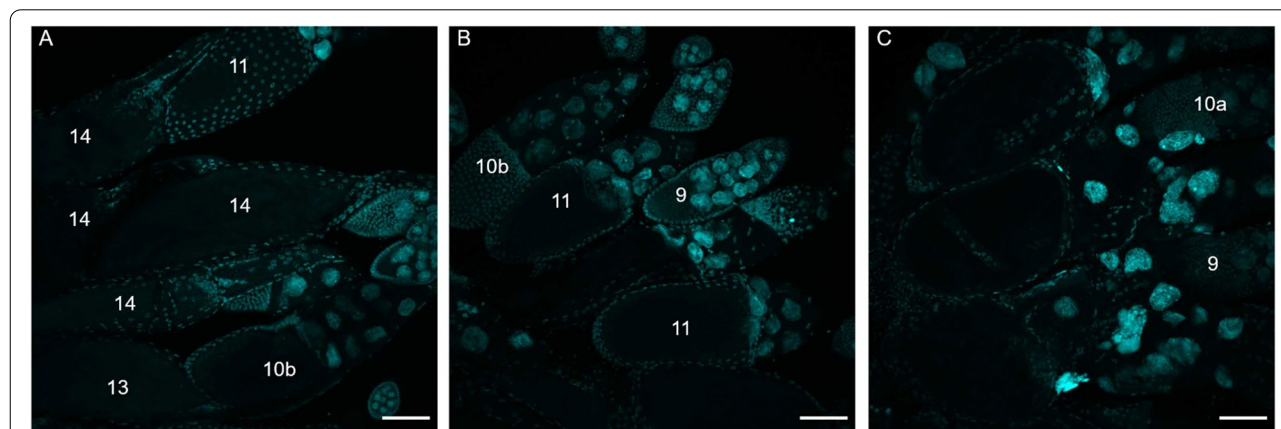
### Effects of *Uhg4* Deletions on Fitness Traits

#### Fertility and Mating Behavior

When we generated the *Uhg4* deletion fly stocks, we did not observe eggs in vials that contained only flies homozygous for a *Uhg4* deletion and we needed to use a *CyO* balancer chromosome to maintain the deletion lines (Figure S1). To test whether the lack of eggs could be due to a failure of the *Uhg4* deletion flies to mate, we assessed mating latency and copulation duration and found that, although deletion flies do not produce progeny when mated, they do exhibit normal mating latencies, copulation times, and proportion of flies mated (Figure S2, Tables S1A, S1B). After crossing flies with the deletion to wild-type DGRP\_208 flies of the opposite sex, we observed that *Uhg4* deletion females do not lay eggs, regardless of mating status or genotype of the male partner, and wild-type eggs fertilized by *Uhg4* deletion males do not develop past the embryo stage. To further probe why *Uhg4* deletion females are sterile, we dissected their ovaries after mating. Whereas ovaries in control DGRP\_208 flies contain late-stage oocytes with dorsal filaments (Fig. 2A), we did not observe any late-stage (>11) oocytes in ovaries of *Uhg4* deletion females (Fig. 2B and C). We did not perform these analyses for



**Fig. 1** *Uhg4* deletions. Diagram showing deletions across the *Uhg4* locus with Sanger sequencing data for the wildtype (DGRP\_208, DGRP\_705) and mutant genotypes. Genomic coordinates are shown at the top of the figure. The blue and green font colors indicate nucleotides of *Uhg4* and *snoRNA:Psi28S-2949* coding regions, respectively. Red font colors indicate the PAM sites. "-" indicates deleted nucleotides and N refers to multiple different nucleotides between the two wild-type DGRP lines. Genotypes used for phenotypic analyses are designated with a star



**Fig. 2** Late-stage oocytes are absent in *Uhg4* deletion lines. Maximum projection z-stack images (60 slices) of DAPI-stained developing egg chambers. **A** Control *Uhg4* (DGRP\_208) ovaries. **B** 208-Δ*F* ovaries. **C** 208-Δ*G* ovaries. Compared to wild type, late-stage oocytes (stage number > 11) and any dorsal filaments are absent in the ovaries of flies with a *Uhg4* deletion. Scale bars represent 75 μm. Numbers represent stages based on Jia et al. [110]

the 705-Δ*B* *Uhg4* allele because it is nearly lethal as a homozygote, with rare viable adults. However, the few escaper 705-Δ*B* females did not lay eggs. These results

show that *Uhg4* is critical for fertility in both sexes and that the sterility of *Uhg4* deletion in females may stem from a lack of fully developed eggs.

### Development Time and Viability

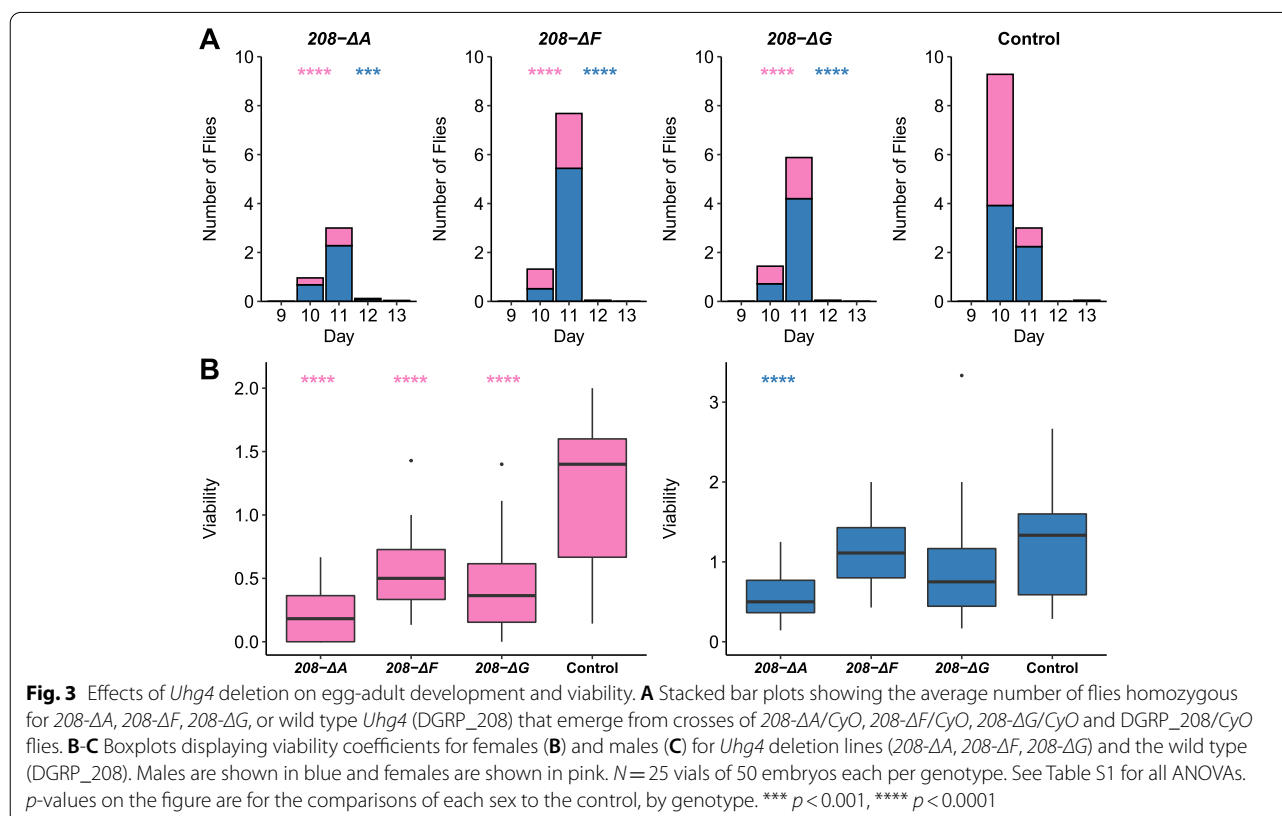
During maintenance of *CyO/Uhg4*-deletion flies, we observed delayed emergence and skewed non-Mendelian ratios of *CyO*/deletion heterozygotes and homozygous deletion progeny. We formally assessed egg-adult development time and viability for the DGRP\_208 *Uhg4* deletion lines compared to the wild-type control. We placed 50 eggs from matings of *208-ΔA/CyO*, *208-ΔF/CyO*, *208-ΔG/CyO* or DGRP\_208/*CyO* (wild type control) flies in 25 vials each and recorded the day each fly emerged, as well as the sex, balancer genotype (*Cy* or straight wing), and the total number of flies for each sex and balancer genotype that emerged. Compared to the wild-type controls, flies with *Uhg4* deletions have delayed development by about one day ( $p < 0.0001$ , Fig. 3A, Table S1C). Furthermore, *208-ΔA* males and all *Uhg4* deletion females show a 2.0- to 5.7-fold decrease in viability ( $p < 0.0001$  for all lines, Fig. 3B, Table S1D, S1E). These results suggest that *Uhg4* is necessary for the normal timing of egg-adult development and viability.

### Effects of *Uhg4* Deletions on Responses to Stressors and Sleep Traits

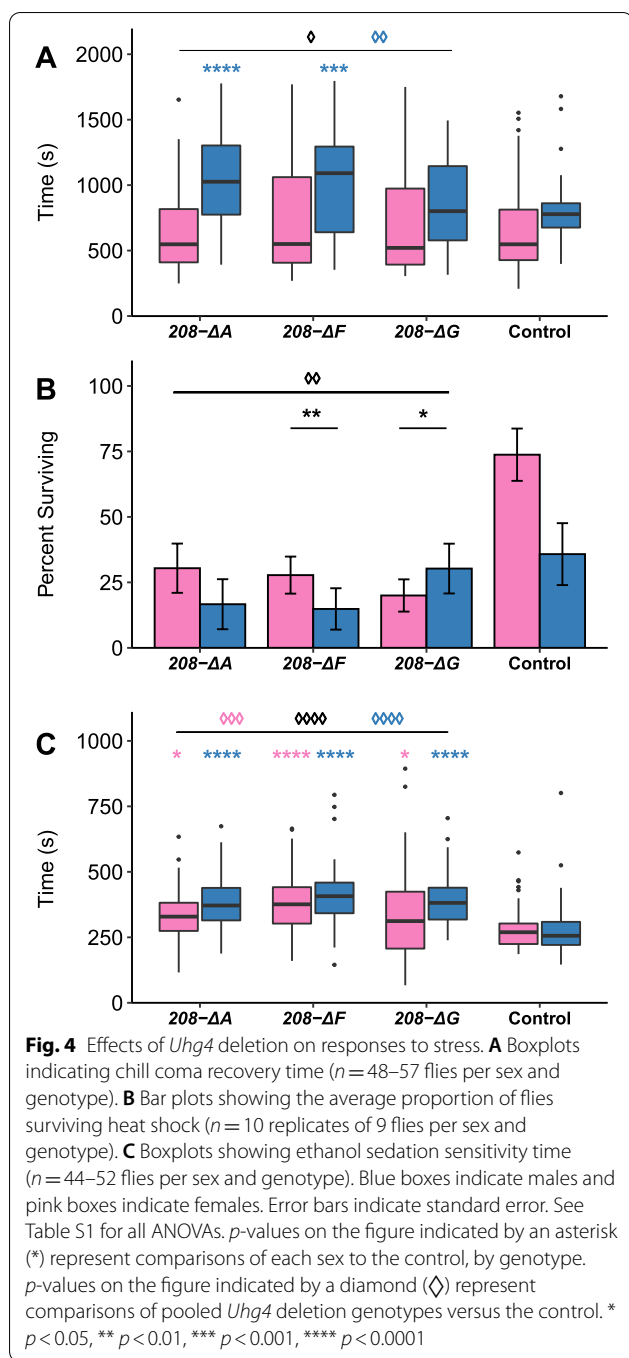
#### Stress Responses

We assessed the effect of several stress conditions (heat shock, chill coma recovery, and ethanol sedation) on

*Uhg4* deletion and wild-type control flies (Fig. 4). On average, *Uhg4* deletion lines take longer to recover from a chill-induced coma than the wild-type in analyses pooled across sexes and all deletion genotypes ( $p = 0.013$ , Table S1F), although there is no difference in the proportion of flies that recover from a chill-induced coma for the deletion lines compared to the control (Table S1G). However, the response to a chill-induced coma is both sex- and genotype-specific. The chill coma recovery time for all deletion lines compared to the wild type is significant for males ( $p = 0.0015$ ) but not females ( $p = 0.666$ ); and males are only significant for the *208-ΔA* ( $p < 0.0001$ ) and *208-ΔF* ( $p = 0.0007$ ) deletion genotypes (Fig. 4A). The *Uhg4* deletion lines also have reduced survival on average following a heat shock than the wild type in analyses pooled across sexes and all deletion genotypes ( $p = 0.0016$ , Table S1F). These effects were genotype-specific as well; only *208-ΔF* ( $p = 0.0013$ ) and *208-ΔG* ( $p = 0.034$ ) were formally significantly different from the wild type, although *208-ΔA* trended in the same direction ( $p = 0.093$ ) (Fig. 4B). In contrast to temperature-related stressors, *Uhg4* deletion flies show decreased susceptibility to ethanol-induced stress. The time to sedation in response to acute ethanol exposure is significantly increased averaged over all *Uhg4* deletion lines compared to the wild







type in the analyses pooled across sexes ( $p < 0.0001$ ) as well as in females ( $p = 0.0007$ ) and males ( $p < 0.0001$ ) (Table S1F). Although all *Uhg4* deletion sex/genotype comparisons had significantly increased sedation times relative to the control, there was heterogeneity in the magnitudes of effects among the deletion genotypes and sexes. In females, *208-ΔF* ( $p < 0.0001$ ) had larger effects than *208-ΔA* ( $p = 0.016$ ) and *208-ΔG* ( $p = 0.036$ );

while in males all *Uhg4* deletion lines had similar effects ( $p < 0.0001$  for all) (Fig. 4C).

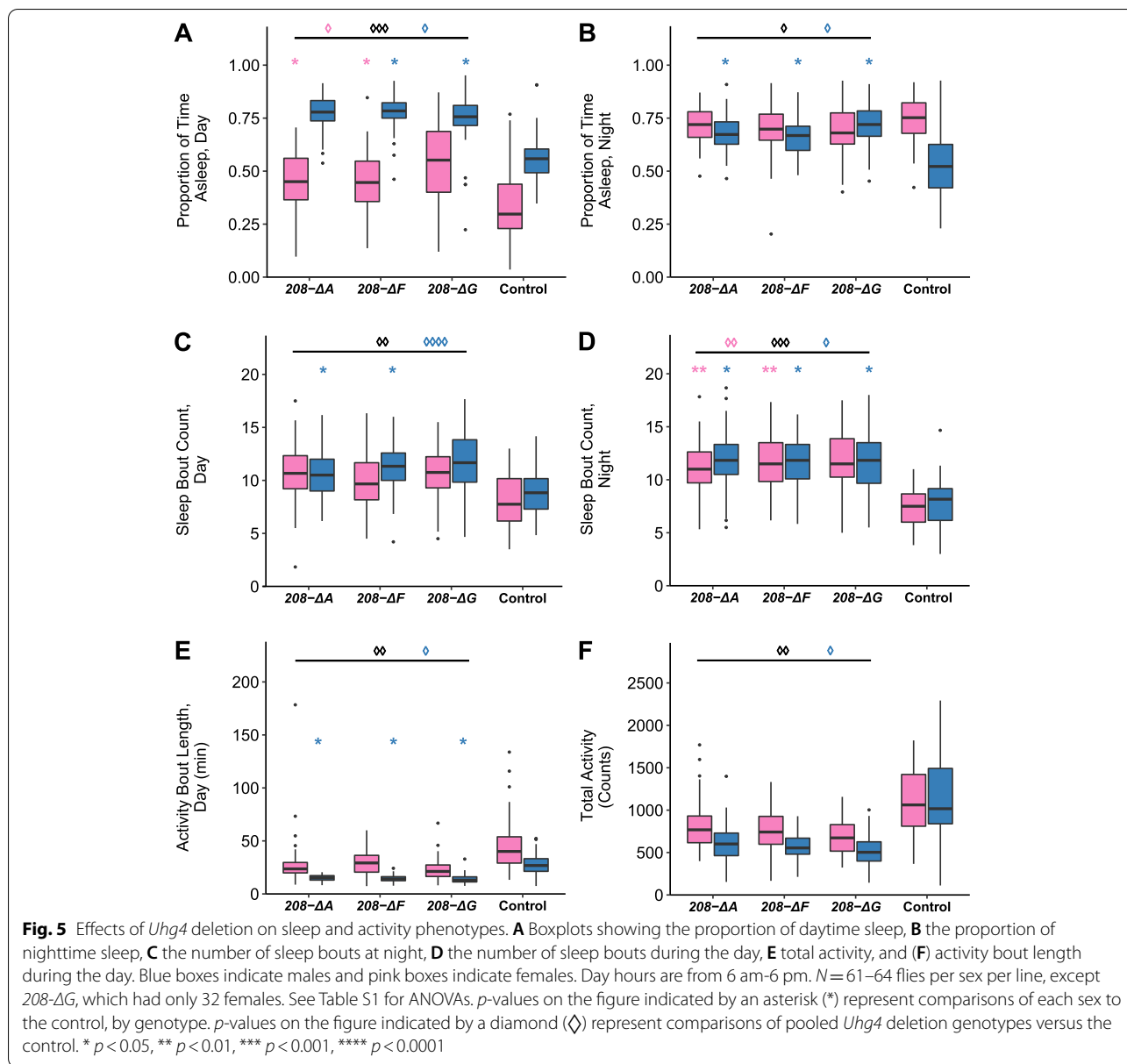
### Sleep and Activity

We used the Drosophila Activity Monitor (DAM) System to assess the effects of *Uhg4* deletions on sleep and activity traits.

*ΔA*, *208-ΔF*, and *208-ΔG* flies sleep more during the day ( $p = 0.0009$  from the analysis of all deletion lines pooled across sexes compared to the wild type, Fig. 5A, Table S1H) and night ( $p = 0.031$  from the analysis of all deletion lines pooled across sexes compared to the wild type, Fig. 5B, Table S1H). The effects of the deletions are much greater on day than on night sleep. In addition, day sleep is significant averaged over all deletion lines for males ( $p = 0.026$ ) and females ( $p = 0.014$ ); while the effect on night sleep is male-specific ( $p = 0.011$ ) (Table S1H). Although the *Uhg4* deletions sleep longer than the wild type during the day and night, they also have more fragmented sleep, as the number of sleep bouts increases both during the day ( $p = 0.002$  from the analysis of all deletion lines pooled across sexes compared to the wild type, Fig. 5C, Table S1H) and night ( $p = 0.0001$  from the analysis of all deletion lines pooled across sexes compared to the wild type, Fig. 5D, Table S1H). Concomitant with increased day and night sleep, the *Uhg4* deletions on average have decreased length of activity bouts during the day ( $p = 0.0061$ , Fig. 5E, Table S1H) as well as decreased total locomotor activity ( $p = 0.0062$  from the analysis of all deletion lines pooled across sexes compared to the wild type, Fig. 5F, Table S1H).

### Effects of *Uhg4* Deletions on Genome-Wide Gene Expression

To gain insights into the mechanisms by which *Uhg4* exerts its pleiotropic effects, we assessed the consequences of deletion of *Uhg4* on genome-wide gene expression, and performed RNA-sequencing on *208-ΔA*, *208-ΔF*, *208-ΔG*, and DGRP\_208 whole flies, separately for males and females. We performed factorial fixed effect ANOVAs for each of the 16,212 genes expressed in young adult flies that evaluate the significance of the main effects of the four *Uhg4* genotypes (Line), Sex, and the Line by Sex interaction. Plotting ordered raw  $p$ -values and adjusted  $p$ -values using a false discovery rate (FDR) correction using the Benjamini–Hochberg procedure (BH-FDR) against the number of tests revealed a non-monotonic relationship between raw  $p$ -values and adjusted  $p$ -values (Figure S3a). This relationship caused an artificial inflation in the number of differentially expressed genes at BH-FDR  $< 0.1$ . Therefore, we used a BH-FDR thresholding approach to identify statistically



significant genes at BH-FDR < 0.1. Briefly, after ordering the genes based on ascending raw *p*-values, we compared each gene's raw *p*-value to its BH-FDR critical value calculated as  $rank \times \frac{Q}{\text{number of tests}}$  at both *Q* = 0.05 and *Q* = 0.1 (Figure S3b). For both critical values, *p*-value thresholds were determined as the first occurrence of the raw *p*-value greater than critical values. Using this method, we identified 17 differentially expressed genes at a BH-FDR < 0.05 for the Line and/or Line by Sex terms (Table S2, Table S3). The top three differentially expressed genes were *Uhg4*, *snoRNA:Psi28S-2949*, and *snoRNA:Or-aca5*. The near-complete loss of *Uhg4* expression in the deletion genotypes is expected due to the deletions of the

promoter region and exon 1. The two snoRNAs are located in the first two introns of *Uhg4*. Decreased expression of *snoRNA:Psi28S-2949* and *snoRNA:Or-aca5* suggests the *Uhg4* deletions affect regulatory sequences common to both snoRNAs, since their coding sequences are not altered (Fig. 1). Ten of the 17 differentially expressed genes are computationally predicted genes and/or noncoding RNAs, including *Uhg4*, and have limited to no information on gene function. One of the 17 significantly differentially expressed genes was *insulin-like peptide 6 (Ilp6)*.

A total of 17 genes that are differentially expressed in *Uhg4* null genotypes is not sufficient to infer the function

of *Uhg4* from the enrichment of Gene Ontologies (GO) and networks of the co-regulated genes. Therefore, we relaxed the significance threshold to BH-FDR < 0.1 for the Line and Line by Sex terms of the ANOVA models. This resulted in 180 differentially expressed genes (Table S2). Notably, all genes significant for the Line by Sex terms also had significant Line effects (Table S3). For these 180 genes, only one GO term (humoral immune response GO:0,006,959) was enriched at BH-FDR < 0.05 (Table S4). However, 47 of the 180 genes did not map to a GO term. These analyses suggest that although deletion of *Uhg4* does change the transcriptome, many of these changes are associated with genes about which little, if anything, is known.

To fully quantify transcriptomic changes in the *Uhg4* deletion flies, we also assessed changes in expression of novel transcribed regions (NTRs) previously identified in the DGRP (including DGRP\_208) [82]. Of the 18,581 NTRs analyzed, we found three (12) were differentially expressed at BH-FDR < 0.05 (0.1) for the Line term, and one NTR was differentially expressed for the Line by Sex term at BH-FDR < 0.1. Together with the protein-coding genes, we have a total 193 differentially expressed genes/NTRs in the *Uhg4* deletion lines (Table S5).

We then performed k-means clustering to assess patterns of expression across genotypes, using  $k=8$ , as it offered the largest number of clusters without redundancy of expression patterns across clusters. GO analysis of these k-means clusters emphasizes the sparsity of information currently known about the genes that are coregulated with *Uhg4*. Of the eight clusters, only four clusters (clusters 1, 3, 6, 7) have enriched GO terms (BH-FDR < 0.05). Cluster 1 is enriched for genes involved in immune response, cluster 3 is enriched for genes involved in response to stress and temperature, and clusters 6 and 7 are enriched for genes involved in cuticle structure (Table S4). Cluster 1 consists of genes largely downregulated in deletion lines compared to the control, including *Uhg4*, as well as other noncoding RNAs and two NTRs (Figure S4). Broadly, cluster 2 contains sexually dimorphic genes, whereas clusters 3 and 5 show genes upregulated in *208-ΔA* and *208-ΔG*, respectively, compared to other lines. *208-ΔF* displays intermediate expression patterns compared to *208-ΔA* and *208-ΔG* in clusters 3, 6, and 8 (Figure S4).

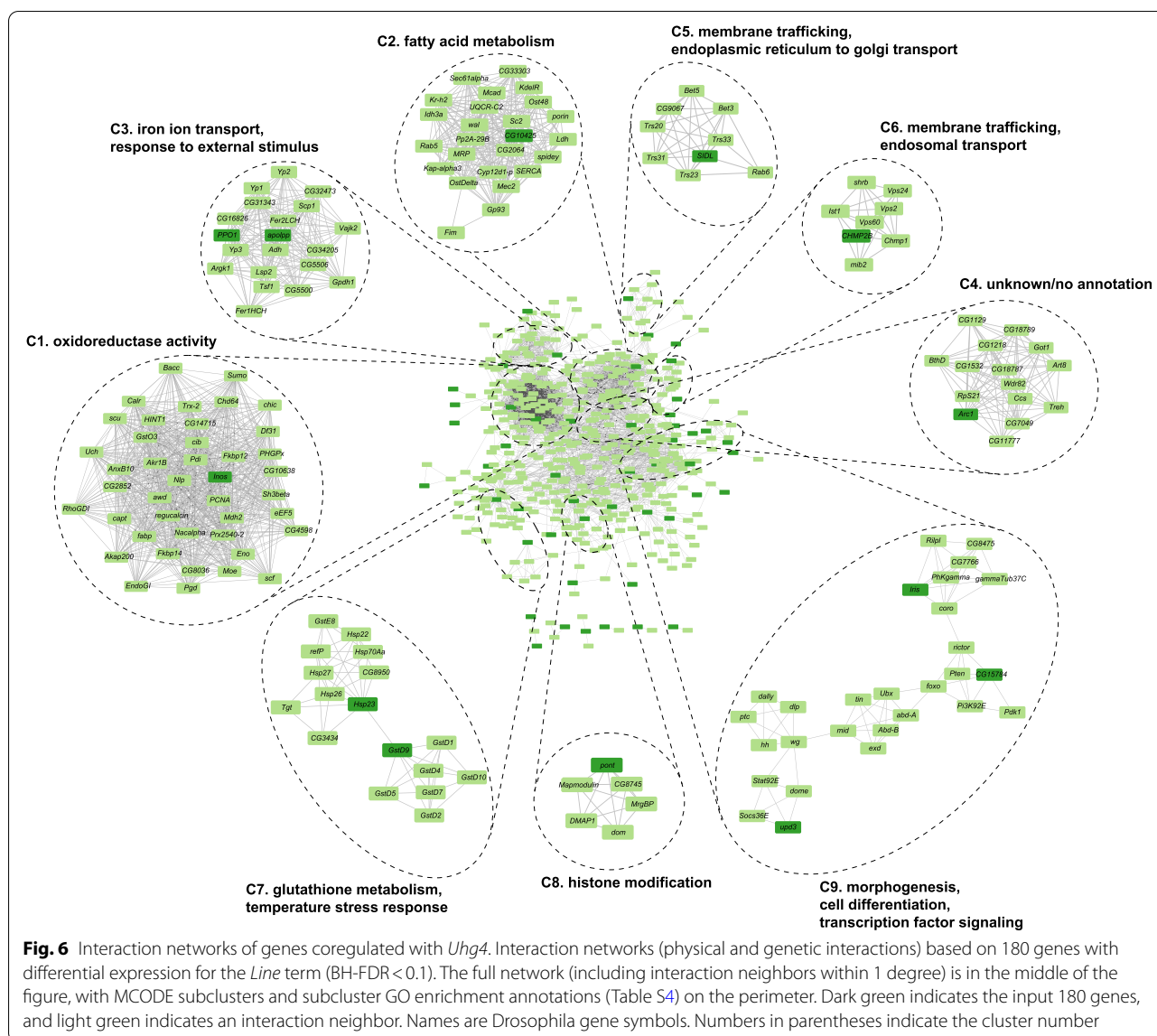
In an effort to generate further hypotheses about the role of *Uhg4* and its co-regulated genes, we used the FlyBase database of known genetic and physical interactions within *D. melanogaster* [79] to generate interaction networks for the subset of 180 co-regulated genes at BH-FDR < 0.1. We also included first-degree interaction neighbors, genes, or proteins that are recorded in the database as directly interacting with at least one of the

180 focal genes. These networks revealed nine subclusters containing genes enriched for a broad spectrum of biological processes, including iron ion transport, fatty acid metabolism, temperature stress response, membrane trafficking, and morphogenesis (Fig. 6, Table S4). Cluster 9, which contains the genes *Ubx*, *dlp*, and *Pten*, among others, was enriched for hundreds of GO terms, far more than any other cluster, suggesting genes in this cluster are critical for a wide range of processes, including morphogenesis, cell differentiation, transcription factor signaling, sleep and activity, reproduction, stress response, and metabolism (Table S4). Cluster 4 was not enriched for any GO terms and attempts at manual annotation did not reveal related functions for genes in this cluster (Table S4). These results indicate that the lncRNA *Uhg4* contributes to diverse cellular functions.

## Discussion

We used *Uhg4*-knockout flies to assess the role of the lncRNA *Uhg4* across multiple fitness traits. We present evidence that *Uhg4* is critical for egg development and fertility and has pleiotropic effects on viability, development, stress responses, and sleep. Genome-wide gene expression data support the finding that *Uhg4* has pleiotropic effects, as *Uhg4* is coregulated with genes involved in a wide range of biological processes, including development, trafficking, metabolism, and stress response. We also observed that many of the genes differentially expressed upon loss of *Uhg4* are also noncoding RNAs in addition to predicted genes. We can speculate that *Uhg4* exerts its pleiotropic effects through broad-based gene regulatory networks. Further studies will be needed to obtain more detailed mechanistic insights into the effects on fitness traits of noncoding RNAs in *D. melanogaster*.

Although many of the coregulated genes themselves do not have associated GO terms, many biological processes and functions implicated by GO analysis align with observed changes in organismal phenotypes. Terms involving response to external stimuli such as temperature and ethanol, as well as terms relating to immune system response, are enriched in multiple k-means and interaction network clusters. These terms align with the increased susceptibility of the *Uhg4* deletion flies to extreme hot and cold temperatures, as well as the decreased susceptibility to ethanol-mediated stressors. Interestingly, *Alcohol dehydrogenase (Adh)* interacts with at least two genes coregulated with *Uhg4*, providing a possible mechanistic link between *Uhg4* and ethanol response; *Uhg4* has been previously implicated in developmental ethanol exposure, which also results in decreased susceptibility to ethanol exposure in adult flies [80]. Although not annotated in Fig. 6, interaction cluster nine is enriched for genes involved in wing morphogenesis and gamete generation (Table S4), which



could explain the changes to wing position and sterility phenotypes, respectively, observed in *Uhg4* deletion flies. We constructed interaction networks from an input of 20 differentially expressed genes and NTRs (BH-FDR < 0.05) including neighbors within at least 1 degree (Figure S5A) and less stringent networks with input of 180 coregulated genes (BH-FDR < 0.1) including neighbors within at least 2 degrees (Figure S5B). Cluster 9 in Table S4, and its more stringent 2-degree interaction counterpart (Cluster 3 in Figure S5B), are enriched for hundreds of GO terms (Table S4), indicating that these genes have wide-ranging impacts. Thus, the effects of *Uhg4* deletion may extend to additional traits that we did not assess, such as iron ion transport, or other intracellular phenotypes.

Based on the DAPI-stained ovary images (Fig. 2) showing a lack of late-stage oocytes, we hypothesize that *Uhg4*-deletion females are capable of laying eggs but do not develop late-stage oocytes. This oocyte development phenotype is supported by the high expression of *Uhg4* in ovaries. However, this does not explain why embryos from wild-type females and *Uhg4* deletion males fail to develop.

*Bootlegger (Boot)*, a gene located immediately upstream of and in opposite orientation to *Uhg4*, is also critical for proper egg development [83, 84]. Our Sanger sequencing data indicate that *Boot* is intact in all deletions and our RNAseq data do not show *Boot* to be differentially expressed. Furthermore, unlike *Uhg4*, *Boot* is minimally expressed in males and would unlikely be responsible for



the sterility observed in *Uhg4*-deficient males. We are therefore confident that the phenotypes observed in our *Uhg4* deletion flies can be attributed to *Uhg4*, though it is possible that *Uhg4* and *Boot* share promoter elements in their intergenic region.

The role of some snoRNA host genes such as *Uhg4* was thought to facilitate transcription and splicing of snoRNAs [85–87]. Based on our results, we hypothesize that *Uhg4* has roles independent of hosting snoRNAs, as most of its snoRNAs are not differentially expressed in *Uhg4* deletion flies. The abundance of genes/NTRs coregulated with *Uhg4* that are noncoding RNAs and/or have no known function makes speculation about the possible functional mechanisms by which *Uhg4* affects a wide range of pleiotropic phenotypes challenging. *Uhg4* could bind directly to DNA or transcription factors to modulate transcriptional regulation, acting in *cis* to regulate the expression of *snoRNA:Psi28S-2949* and *snoRNA:Oraca5*, or in *trans* to regulate the expression of other genes we observed to be differentially expressed. *Uhg4* may also serve an oncogenic role, as overexpression of *Uhg4* in a *Drosophila* cell line is associated with tumor growth, and *Uhg4* is a downstream target of *Myc* [88]. *Uhg4* may also act to regulate gene expression at an epigenetic level via histone modifications (Cluster 8, Fig. 6). Other lncRNAs in *Drosophila* that are important for thermotolerance are essential for remobilization of heterogeneous ribonucleoprotein particles (hnRNPs) [41, 44]. *Uhg4*, as a host gene for snoRNAs, which also form ribonucleoproteins, might also modulate response to thermal stressors via ribonucleoproteins. *Uhg4* could also act in a similar manner to the transcript of *oskar*, which is critical for oogenesis in *Drosophila* [54, 89, 90]. *oskar* RNA facilitates oogenesis through multiple mechanisms, as it binds a translational regulator at one locus, has a separate 3' region critical for proper egg-laying, and may also be involved in scaffolding of ribonucleoproteins [89].

## Conclusion

*Uhg4* is a unique example of a lncRNA with pleiotropic functions that is indispensable for viability and reproduction in *D. melanogaster*.

## Methods

### Generation of *Uhg4* null alleles

We used the flyCRISPR target finder [91] to design gRNAs flanking *Uhg4* Exon 1 (upstream: 5'- GAAGTAAACTTCTTTGCACTGG -3'; downstream: 5'- GTAAGTATTATAGATATGATAGG -3') that did not overlap with known genes and did not have predicted off-target effects, resulting in a 685 bp deletion. We then synthesized a single guide RNA (gRNA) vector containing both gRNAs [92]. Briefly, synthesized phosphorylated

gRNAs with complementary sequences were ligated to form double-stranded gRNAs with CTTC overhangs. One double-stranded gRNA was cloned into *pBFV-U6.2* (Addgene #138,400), and the other double-stranded gRNA was cloned into *pBFV-U6.2B* (Addgene #138,401). Using flanking *EcoRI* and *NotI* sites, the resulting *U6* promoter and gRNA within *pBFV-U6.2* were excised and ligated into the *U6.2B* vector, creating a double-gRNA vector. Sanger sequencing confirmed the proper insertion of gRNAs.

The completed dual gRNA vector and *pBFV-nosP-Cas9* (Addgene #138,402) were purified and co-injected (BestGene Inc., Chino Hills, CA) into at least 300 embryos from two different *D. melanogaster* Genetic Reference Panel (DGRP) lines [81, 93], DGRP\_208 and DGRP\_705. These lines have minimal heterozygosity and are homozygous for the standard karyotype for all common inversions; DGRP\_208 is also free of *Wolbachia* infection [81].

To preserve the inbred genetic background of the DGRP lines, we screened flies for the presence of a deletion by individually isolating DNA from clipped wings of virgin female and male flies [94]. Fly wings were covered with 10  $\mu$ L of 400  $\mu$ g/mL protease K in a 10 mM Tris–Cl at pH 8.2, 1 mM EDTA and 25 mM NaCl buffer and incubated at 37 °C for 2 h followed by 95 °C for 2 min. We used 2  $\mu$ L of the resulting DNA mixture in a PCR reaction with primers (Left: 5'- CTAGCACGGAACCTTGAAAT -3'; Right: 5'- GCAGCGCCTAGTAATCACAGA -3') according to ApexRedTaq (Genesee Scientific, El Cajon, CA) manufacturer instructions, with a 61 °C annealing temperature for 30 s and 30 cycles. The deletion mutations *Uhg4*<sup>208- $\Delta$ A</sup>, *Uhg4*<sup>208- $\Delta$ F</sup>, *Uhg4*<sup>208- $\Delta$ G</sup>, and *Uhg4*<sup>705- $\Delta$ B</sup> (hereafter referred to as 208- $\Delta$ A, 208- $\Delta$ F, 208- $\Delta$ G and 705- $\Delta$ B, respectively) were isolated and placed over a *CyO* balancer chromosome in the appropriate genetic background. *Uhg4* wild-type lines (DGRP\_208 and DGRP\_705) are used as controls in this study.

### *Drosophila* culture

Flies were reared at 25 °C with 50% humidity on standard cornmeal-molasses-agar medium (Genesee Scientific, El Cajon, CA), supplemented with yeast. Flies were maintained at controlled density on a 12-h light–dark cycle (lights on at 6 am). Unless otherwise indicated, all behavioral assays were performed on 3–5 day old homozygous flies from 8:30 am to 11:30 am.

### Sanger sequencing

DNA from homozygous mutant flies was sequenced by the Sanger chain termination method using a BigDye Terminator Kit v3.1 (ThermoFisher Scientific, Waltham, MA), with the same primers used to identify deletions.



Sequencing was performed on an Applied Biosystems 3730xl DNA Analyzer (ThermoFisher Scientific, Waltham, MA).

#### Quantitative real-time PCR

Flies were flash-frozen for RNA extraction at 3–5 days of age. Each sample contained 30 whole flies and was homogenized using a Fisherbrand™ Bead Mill (ThermoFisher Scientific, Waltham MA). RNA was extracted using the Direct-zol RNA Miniprep kit (Zymo Research, Irvine, CA), resuspended in RNase-free water, and kept at -80 °C until further use. cDNA was synthesized using the iScript™ Reverse Transcription Supermix for RT-qPCR (Bio-Rad Laboratories, Inc., Hercules, CA) according to the manufacturer's instructions. Quantitative real-time PCR to detect expression was performed on three biological replicates and two technical replicates per sample (except *208-ΔA*, which had two biological replicates), using SYBR™ Green PCR Master Mix (ThermoFisher Scientific, Waltham, MA) and primers (*Uhg4*-Forward: 5'-TCGGTCTTTCGATTTGGATT-3'; *Uhg4*-Reverse: 5'-TGTGTTAGTGAGCCACGTTTG-3', spanning exons 4–5 of *Uhg4*; GAPDH-Left: 5'-CTTCTTCAGCGACACCATT-3'; GAPDH-Right: 5'-ACCGAACTCGTTGTCGTACC-3') according to the manufacturer's instructions. Percent knockdown was assessed using  $\Delta\Delta Ct$  [95]. No amplification was observed in non-template negative controls.

#### Viability and development time

We placed 75 male–female pairs of flies from each of the DGRP\_208 *Uhg4* genotypes (DGRP\_208/*CyO*; *208-ΔA/CyO*, *208-ΔF/CyO*, *208-ΔG/CyO*) in large embryo collection cages (Genesee Scientific, San Diego, CA), supplemented with fresh yeast paste and grape juice-agar plates containing 3% agar and 1.2% sucrose in a 25% Welch's Grape Juice concentrate solution every 12 h. After 36 h, we placed 50 0–12 h embryos in each of 25 vials per genotype. From days 9–13 after egg-laying, we collected adult flies once daily and recorded the sex, wing phenotype, and day of emergence for each fly, including flies that eclosed, but died in the food. We calculated relative egg–adult viability ( $\nu$ ) for each vial as the proportion of homozygous *Uhg4* adults ( $1-r$ ) relative to the proportion of *Uhg4/CyO* heterozygote adults ( $r$ ) as  $\nu = 2(1-r)/r$  [96].

#### Fertility

We assessed the fertility of homozygous males and females from the DGRP\_208 *Uhg4* deletion lines by crossing virgin deletion mutant flies to virgin DGRP\_208 flies, respectively, of the opposite sex. In addition, we also performed crosses of males and females within each

*Uhg4* deletion line. We set up four vials each with four males and four virgin females for each genotype. For each cross, qualitative observations (presence/absence) were made for each stage of development (embryos, first instar larvae, second instar larvae, third instar larvae, pupae, adult flies). We did not perform a formal fertility assay with the *705-ΔB* deletion line due to the very low viability of *705-ΔB* homozygotes.

#### Mating behavior

For each DGRP\_208 *Uhg4* genotype (*208-ΔA*, *208-ΔF*, *208-ΔG*, DGRP\_208), we placed 22–24 pairs of virgin flies in separate mating chambers [97, 98] to acclimate overnight. Between 8 and 10 am, fly pairs were united and video-recorded for 30 min. Copulation duration and mating latency were recorded in seconds for each fly pair. We only included flies that mated in the analyses of mating behaviors and recorded the number of pairs that did not mate during the 30-min testing period. We selected *208-ΔG* as a representative *Uhg4* mutant line and assessed the mating behaviors of crosses of *208-ΔG* males to DGRP\_208 virgin females and *208-ΔG* virgin females to DGRP\_208 males.

#### Survival after heat shock

Vials with 2 ml of medium and 9 homozygous flies per sex and *Uhg4* genotype were placed in a 37 °C incubator for 2 or 3 h beginning at 1 pm, with five replicate vials per sex per genotype per exposure time. After the heat shock, flies were transferred to fresh vials with 2 ml of medium and allowed to recover overnight. Twenty-four hours after the heat shock, the number of surviving flies in each vial was recorded.

#### Chill coma recovery time

For each genotype we placed four vials containing 15 flies without medium on ice for 3 h, sexes separately, and allowed the flies to recover in a 6-well cell culture plate (five flies per well, two genotypes per plate) for 30 min. We recorded the time until each fly righted itself by standing up [99]. Only flies that recovered within 30 min of being removed from the ice were included in the analysis ( $n = 48$ – $57$  flies per sex per line), although we also recorded the number of flies that did not recover.

#### Ethanol sensitivity

We assessed the time to sedation in response to acute ethanol exposure [100] on 44–52 flies per sex per *Uhg4* genotype. Briefly, flies were aspirated into a 24-well cell culture plate and placed opposite a 24-well cell culture plate containing 100% ethanol, separated by a layer of fine

screen mesh. We recorded the time to sedation, defined as the moment each fly loses postural control.

### Sleep and locomotor activity

We collected sleep and locomotor activity data using *Drosophila* Activity Monitors (DAM) (TriKinetics, Waltham, MA). Briefly, we placed 1–2 day-old flies into DAM monitor tubes containing 2% agar with 5% sucrose, with two DAM monitors per sex per *Uhg4* genotype, for a total of 64 flies per sex and genotype. Sleep and activity data were recorded on days 3–8 of the fly lifespan on a 12-h light–dark cycle. Sleep was defined as at least 5 min of inactivity. Only flies that survived the entire testing period were included in analyses, resulting in 61–64 flies per sex per *Uhg4* genotype (except for *208-ΔG* females, for which there was only one replicate of 32 individuals). The DAM data were initially processed with ShinyRDAM [101] and resulting raw sleep and activity output files were downloaded for further statistical analysis.

### Statistical analyses

Unless otherwise indicated, all behavioral assays were analyzed using the “PROC MIXED” command (for a mixed-effects model) or “PROC GLM” command (for a pure fixed-effects model) in SAS v3.8 as a Type III Analysis of Variance (ANOVA). Where appropriate, flies were randomized to avoid positional or time-related effects.

Development time (day of eclosion) was analyzed according to the model  $Y = \mu + L + S + G + L \times S + L \times G + S \times G + L \times S \times G + Rep(L) + S \times Rep(L) + G \times Rep(L) + S \times G \times Rep(L) + \varepsilon$ . Heat shock (percent of flies that survived) and all sleep and activity phenotypes (e.g. locomotor activity, bout count, percent of time asleep) were analyzed according to the model  $Y = \mu + L + S + L \times S + Rep(L \times S) + \varepsilon$ . Chill coma recovery time (time to recover, in seconds) and ethanol sensitivity (time to sedation, in seconds) were analyzed according to the model  $Y = \mu + L + S + L \times S + \varepsilon$ . Mating latency (in seconds), mating duration (in seconds), and viability (viability coefficient  $v$ ) were assessed according to the ANOVA model  $Y = \mu + L + \varepsilon$ . In these models,  $Y$  is the phenotype of interest,  $\mu$  is the overall mean,  $L$  is the fixed effect of Line (DGRP\_208, *208-ΔA*, *208-ΔF*, *208-ΔG*),  $S$  is the fixed effect of sex (male, female),  $G$  is the fixed effect of balancer genotype (*Uhg4/CyO* heterozygote, *Uhg4* homozygote),  $Rep$  is replicate nested within lines, and  $\varepsilon$  is the residual error. We performed pairwise comparisons of each *Uhg4* deletion line (*208-ΔA*, *208-ΔF*, or *208-ΔG*) with the *Uhg4* wild type control, as well as a pooled comparison across all deletion lines compared to the control. For mating phenotypes, we also compared *208-ΔG* females × DGRP\_208 males and DGRP\_208 females × *208-ΔG* males to the control. We used Fisher’s

exact tests to assess the proportion of flies that mated and the proportion of flies that recovered from a chill-induced coma in mating and chill coma recovery time, respectively using the R *stats* package. Models were also run separately by sex. For development time, models were also run separately for *Uhg4/CyO* heterozygotes and *Uhg4* homozygotes.

### Ovary dissection

We placed mated females from representative *Uhg4* deletion lines and their controls (*208-ΔF*, *208-ΔG*, DGRP\_208) in fresh food vials supplemented with yeast paste every 12 h for 36 h prior to dissection. Flies were dissected in 1X PBS and ovarioles were gently separated. Ovaries were fixed in 4% paraformaldehyde for 15 min, followed by three 15-min washes in PBS with 0.2% Triton X-100. Following a final 15-min wash in PBS, ovaries were stained with DAPI (Invitrogen) (1 μg/mL) for 10 min and mounted with ProLong Gold (Invitrogen) immediately after the final PBS wash. Ovaries were imaged on an Olympus Fluoview FV3000 microscope at 20 × magnification. Images were processed in Fiji [102].

### RNA sequencing

We prepared libraries for RNA sequencing from each RNA sample used in the RT-qPCR analyses according to Universal RNA-Seq with NuQuant+UDI (Tecan Genomics, Inc., Redwood City, CA) manufacturer instructions. Specifically, 100 ng of total RNA was converted into cDNA via integrated DNase treatment. Second strand cDNA was fragmented using a Covaris ME220 Focused-ultrasonicator (Covaris, Woburn, MA) to 350 bp. *Drosophila* AnyDeplete probes were used to deplete remaining ribosomal RNA and final libraries were amplified using 17 PCR cycles. We used TapeStation High Sensitivity D1000 Screentape (Agilent Technologies, Inc., Santa Clara, CA) and a Qubit™ 1X dsDNA HS Assay kit (Invitrogen, Carlsbad, CA) to measure the final library insert sizes and concentration, respectively. Final libraries were diluted to 4 nM and sequenced on a NovaSeq6000 (Illumina, Inc., San Diego, CA) using an S1 flow cell. We sequenced two biological replicates per sex per line (*208-ΔA*, *208-ΔF*, *208-ΔG*, DGRP\_208), with ~20–74 million reads per sample.

Barcoded reads were demultiplexed using the NovaSeq Illumina BaseSpace sequencing pipeline and merged across S1 flow cell lanes. We used the *AfterQC* pipeline (version 0.9.7) [103] to filter out low-quality, short, and adapter reads and the *bbduk* command within the BBTools package [104] to detect levels of rRNA contamination. We used GMAP-GSNAP [105] to align filtered reads to the *D. melanogaster* reference genome v6.13 and the *featurecounts* pipeline from the Subread

package [106] to count unique alignments to *Drosophila* genes. Expression for novel transcribed regions (NTRs) was estimated by first compiling a list of NTRs detected from RNA sequencing of young adult DGRP flies [82]. The coordinates of these NTRs were converted from R5 to R6 using the Coordinate Converter tool on FlyBase. A new gene transfer format file was constructed using the coordinate-converted NTR gene models and was used in conjunction with the alignment files for expression estimation. Counts data for each sample were filtered to omit genes for which the median count was less than two, as well as genes for which the proportion of null values across all samples was less than 0.25. The data were then normalized for gene length and library size using Ge-TMM [107]. Filtered and normalized data were analyzed for differential expression using the “PROC glm” command in SAS v3.8 (Cary, NC) according to the ANOVA model  $Y = \mu + L + S + L \times S + \varepsilon$ , where Line ( $L$ ), Sex ( $S$ ), and Line  $\times$  Sex ( $L \times S$ ) are fixed effects,  $Y$  is gene expression,  $\mu$  is the overall mean, and  $\varepsilon$  is residual error. A false discovery rate (FDR) correction using the Benjamini–Hochberg procedure (BH-FDR) for multiple tests was applied across all genes to determine the subset of differentially expressed genes significant at BH-FDR < 0.05 and BH-FDR < 0.1 for either the Line or Line  $\times$  Sex terms. NTR expression counts were analyzed using the same approach described above but as a separate dataset. For the bulk RNA sequencing data, plotting ordered raw  $p$ -values and BH-FDR adjusted  $p$ -values against the number of tests revealed a non-monotonic relationship between raw  $p$ -values and adjusted  $p$ -values. Therefore, we used a BH-FDR thresholding approach to identify genes with statistically significant  $p$ -values from the ANOVA model. Briefly, after ordering the genes based on ascending raw  $p$ -values, we compared each gene’s raw  $p$ -value to FDR critical values calculated as  $\text{rank} \times Q / \text{number of tests}$  at both  $Q = 0.05$  and  $Q = 0.1$ . For both critical values,  $p$ -value thresholds were determined as the first occurrence of the raw  $p$ -values greater than critical values. Genes with raw  $p$ -values below the  $p$ -value threshold at critical values  $Q = 0.05$  and  $Q = 0.1$  were considered for downstream analyses. We used the resulting 180 genes significant at FDR < 0.1 for network construction and included the 13 differentially expressed NTRs (BH-FDR < 0.1) with these 180 genes for k-means clustering.

We performed k-means clustering ( $k = 8$ , average linkage algorithm) on the Ge-TMM-normalized least squares means of the 193 coregulated genes. We performed iterative k-means clustering with different  $k$  values to determine the largest number of clusters without redundant expression patterns across clusters. We used Cytoscape v3.9.1 and the interaction networks from FlyBase [79] (FB\_2021\_05) to create protein and genetic interaction

networks including first-degree neighbors, clustered via MCODE score [108] where applicable. Cluster annotations are based on significantly enriched Gene Ontology (GO) terms. We performed GO statistical overrepresentation analyses with GO: Biological Process Complete, Molecular Function Complete, and Reactome Pathway terms (GO Ontology database released 2021–11–16) using Panther db v16.0 [109] using Fisher Exact tests with BH-FDR correction.

#### Abbreviations

BH-FDR: Benjamini–Hochberg procedure for obtaining adjusted P values based on False Discovery Rate; DGRP: *Drosophila melanogaster* Genetic Reference Panel; gRNA: Guide RNA; lncRNA: Long noncoding RNA; snoRNA: Small nucleolar RNA; Uhg4: Unknown host gene 4.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08972-0>.

**Additional file 1: Table S1.** Analyses of behavioral phenotypes. Analyses of variance (ANOVA) and Fisher’s exact tests comparing Uhg4 deletion flies to their respective controls.

**Additional file 2: Table S2.** Differentially expressed genes. (A) BH-FDR correction for genes based on *Line* and (B) *Line* $\times$ *Sex* model terms; (C) Normalized average read counts values for the 180 coregulated genes and 13 NTRs; (D) FDR correction for NTRs genes based on *Line* and *Line* $\times$ *Sex* model terms. See corresponding Figure S3.

**Additional file 3: Table S3.** Analysis of differential expression. For each gene (A) and NTR (B), the FBgn number or NTR ID, raw  $p$ -values, BH-FDR adjusted  $p$ -values for the effects of *Line* ( $L$ ), *Sex* ( $S$ ), and *Line* $\times$ *Sex* ( $L \times S$ ) according to the ANOVA model  $Y = \mu + L + S + L \times S + \varepsilon$ .

**Additional file 4: Table S4.** Gene Ontology enrichment analyses. Enriched Gene Ontology (GO) Biological Process, Molecular Function and REACTOME terms and their raw and BH-FDR adjusted  $p$ -values. Groups of genes analyzed include 20 genes/NTRs with differential expression for the *Line* or *Line* $\times$ *Sex* terms (BH-FDR < 0.05), 20 genes/NTRs and their 1-degree neighbors, 193 coregulated genes/NTRs (BH-FDR < 0.1) due to a significant *Line* or *Line* $\times$ *Sex* effect, 193 genes/NTRs and their 1-degree neighbors, genes within each MCODE cluster (Figure 4, Figure S5 A and B), and genes in each k-means cluster (Figure S4).

**Additional file 5: Table S5.** Gene lists. Lists of gene names, symbols, and Flybase IDs (FBID) for each network cluster and k-means cluster.

**Additional file 6: File S1.** Video of a representative Uhg4 deletion (208- $\Delta$ G) female fly displaying an elevated, splayed resting wing position compared to a wildtype female fly. The Uhg4 deletion fly is walking during the first part of the video and the wildtype fly is walking at the end of the video.

**Additional file 7: File S2.** Video of a representative Uhg4 deletion (208- $\Delta$ G) male fly displaying an elevated, splayed resting wing position compared to a wildtype male fly. The Uhg4 deletion fly is to the right of the wildtype fly throughout the video.

**Additional file 8: Figure S1.** Crossing scheme to generate Uhg4 deletion CRISPR mutants. Following injections of *Cas9* and gRNA vectors into embryos of each DGRP line, resulting progeny were screened for presence of a deletion around Uhg4, and if a deletion was present, crossed to the original genetic background to generate additional flies heterozygous for the original deletion. Backcrossing siblings heterozygous for the same mutation resulted in homozygous flies that were sterile for all isolated deletion mutations. Virgin female flies heterozygous for a specific deletion were then crossed to male flies containing the CyO balancer. Resulting progeny were screened for the presence of the CyO balancer

chromosome as well as the respective mutation and crossed to full siblings to establish the stock. All females were crossed as virgin flies. “Δ” refers to a specific Uhg4 deletion.

**Additional file 9: Figure S2.** Mating phenotypes of Uhg4 deletion flies. Boxplots displaying time, in seconds, of (A) mating latency (time until mating begins) and (B) mating duration (length of copulation) for each of the Uhg4 mutant lines (208-ΔA, 208-ΔF, 208-ΔG), the control line DGRP\_208 (208wt), and a representative mutant (208-ΔG) versus DGRP\_208 wildtype pairings (208-ΔG females x 208wt males, 208wt females x 208-ΔG males). N = 22–24 pairings of 3–5 day old virgin flies per line. Only flies which successfully initiated or completed mating within 30 minutes were included in analysis for mating latency and mating duration, respectively. See Table S1A. \*  $p < 0.05$

**Additional file 10: Figure S3.** Non-monotonic relationship between raw and Benjamini-Hochberg adjusted  $p$ -values. (A) raw  $p$ -values plotted against number of tests; (B) BH-FDR thresholds on raw  $p$ -values. Raw and adjusted  $p$ -values are shown in blue and red, respectively. “r”: rank, “m”: number of tests. See corresponding Table S2.

**Additional file 11: Figure S4.** K-means clusters. K-means clusters derived from genes and NTRs with differential expression in a global transcriptomic analysis using genes significant for the *Line* or *Line* × *Sex* terms (BH-FDR < 0.1). Uhg4 is shown in bold and indicated with a star symbol. Magenta indicates a relatively higher degree of expression, blue indicates a relatively lower degree of expression.

**Additional file 12: Figure S5.** Additional interaction networks. Interaction networks containing physical and genetic interactions with genes with a significant *Line* effect. (A) Networks generated from an input of 20 differentially expressed genes/NTRs (BH-FDR < 0.05) including neighbors within at least 1 degree. (B) Networks generated from an input of 180 coregulated genes (BH-FDR < 0.1) including neighbors within at least 2 degrees. Annotation is based on enriched Gene Ontology terms. Dark green indicates the genes in the input data set and light green indicates interaction neighbors. Names are *Drosophila* gene symbols. See Table S4.

#### Acknowledgements

We thank Miller Barksdale and Patrick Freymuth for their technical assistance, Dr. Sneha Mokashi for experimental advice and Jonathon Thomalla and Drs. Mariana Wolfner, Melissa White, and Deepthiman Chatterjee for their insights into ovary biology. Additionally, we would like to thank the Greenwood Genetic Center and Dr. Heather Flanagan-Steet for the use of their Sanger sequencing and microscope facilities.

#### Authors' contributions

RAM – conceptualization, formal analysis, funding acquisition, investigation, visualization, writing – original draft preparation, writing – review & editing; VS – data curation, formal analysis, software; LS – investigation; RCH – investigation; MRCIII – investigation; RRHA – conceptualization, writing – review & editing, funding acquisition, supervision; TFCM – conceptualization, writing – review & editing, funding acquisition, supervision, project administration. All authors have read and approved the manuscript.

#### Funding

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number P20GM139769 and Award Number R01GM128974 to RRHA and TFCM and by the National Institute of Child Health and Development under Award Number F31HD106719 to RAM. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

#### Availability of data and materials

All high throughput sequencing data are deposited in GEO GSE199865. Sanger sequencing, qPCR, and raw behavioral data are available on GitHub at [rebeccamacpherson/Pleiotropic\\_fitness\\_effects\\_Uhg4\\_rawdata](https://github.com/rebeccamacpherson/Pleiotropic_fitness_effects_Uhg4_rawdata).

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The authors do not disclose competing financial interests.

##### Author details

<sup>1</sup>Center for Human Genetics and Department of Genetics and Biochemistry, Clemson University, 114 Gregor Mendel Circle, Greenwood, SC 29646, USA.

<sup>2</sup>Present address: Clemson Veterinary Diagnostic Center, Livestock Poultry Health, Clemson University, 500 Clemson Road, Columbia, SC 29229, USA.

Received: 19 July 2022 Accepted: 26 October 2022

Published online: 30 November 2022

#### References

1. Brannan CI, Dees EC, Ingram RS, Tilghman SM. The product of the H19 gene may function as an RNA. *Mol Cell Biol*. 1990;10:28–36. <https://doi.org/10.1128/mcb.10.1.28-36.1990>.
2. Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, et al. 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*. 1992;71:515–26. [https://doi.org/10.1016/0092-8674\(92\)90519-1](https://doi.org/10.1016/0092-8674(92)90519-1).
3. Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, et al. 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*. 1992;71:527–42. [https://doi.org/10.1016/0092-8674\(92\)90520-M](https://doi.org/10.1016/0092-8674(92)90520-M).
4. Kung JTY, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics*. 2013;193:651–69. <https://doi.org/10.1534/genetics.112.146704>.
5. McHugh CA, Chen C, Chow A, Surka CF, Tran C, McDonel P, et al. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*. 2015;521:232–6. <https://doi.org/10.1038/nature14443>.
6. Rom A, Melamed L, Gil N, Goldrich MJ, Kadir R, Golan M, et al. Regulation of CHD2 expression by the Chaserr long noncoding RNA gene is essential for viability. *Nat Commun*. 2019;10:5092. <https://doi.org/10.1038/s41467-019-13075-8>.
7. Pandya-Jones A, Markaki Y, Serizay J, Chitashvili T, Mancia Leon WR, Damianov A, et al. A protein assembly mediates Xist localization and gene silencing. *Nature*. 2020;587:145–51. <https://doi.org/10.1038/s41586-020-2703-0>.
8. Statello L, Guo C, Chen L, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol*. 2021;22:96–118. <https://doi.org/10.1038/s41580-020-00315-9>.
9. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, et al. The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science*. 2013;341:1237973. <https://doi.org/10.1126/science.1237973>.
10. Li X, Zhou B, Chen L, Gou L, Li H, Fu X. GRID-seq reveals the global RNA-chromatin interactome. *Nat Biotechnol*. 2017;35:940–50. <https://doi.org/10.1038/nbt.3968>.
11. Mumbach MR, Granja JM, Flynn RA, Roake CM, Satpathy AT, Rubin AJ, et al. HiChIRP reveals RNA-associated chromosome conformation. *Nat Methods*. 2019;16:489–92. <https://doi.org/10.1038/s41592-019-0407-x>.
12. Sánchez Y, Segura V, Marín-Béjar O, Athie A, Marchese FP, González J, et al. Genome-wide analysis of the human p53 transcriptional network unveils a lncRNA tumour suppressor signature. *Nat Commun*. 2014;5:5812. <https://doi.org/10.1038/ncomms6812>.
13. Michelini F, Pitchiaya S, Vitelli V, Sharma S, Gioia U, Pessina F, et al. Damage-induced lncRNAs control the DNA damage response through interaction with DDRNAs at individual double-strand breaks. *Nat Cell Biol*. 2017;19:1400–11. <https://doi.org/10.1038/ncb3643>.



14. Uroda T, Anastasakou E, Rossi A, Teulon J, Pellequer J, Annibale P, et al. Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway. *Mol Cell*. 2019;75:982–95.e9. <https://doi.org/10.1016/j.molcel.2019.07.025>.
15. Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, et al. An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol Cell*. 2009;33:717–26. <https://doi.org/10.1016/j.molcel.2009.01.026>.
16. Hacisuleyman E, Shukla CJ, Weiner CL, Rinn JL. Function and evolution of local repeats in the Firre locus. *Nat Commun*. 2016;7:11021. <https://doi.org/10.1038/ncomms11021>.
17. Wu H, Yin Q, Luo Z, Yao R, Zheng C, Zhang J, et al. Unusual Processing Generates SPA LncRNAs that Sequester Multiple RNA Binding Proteins. *Mol Cell*. 2016;64:534–48. <https://doi.org/10.1016/j.molcel.2016.10.007>.
18. Fei J, Jadalaha M, Harmon TS, Li ITS, Hua B, Hao Q, et al. Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *J Cell Sci*. 2017;130:4180–92. <https://doi.org/10.1242/jcs.206854>.
19. Soshnev AA, Ishimoto H, McAllister BF, Li X, Wehling MD, Kitamoto T, et al. A Conserved Long Noncoding RNA Affects Sleep Behavior in *Drosophila*. *Genetics*. 2011;189:455–68. <https://doi.org/10.1534/genetics.111.131706>.
20. Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. *Nat Rev Genet*. 2016;17:272–83. <https://doi.org/10.1038/nrg.2016.20>.
21. Hu WL, Jin L, Xu A, Wang YF, Thorne RF, Zhang XD, et al. GUARDIN is a p53-responsive long non-coding RNA that is essential for genomic stability. *Nat Cell Biol*. 2018;20:492–502. <https://doi.org/10.1038/s41556-018-0066-7>.
22. Yin Q, Yang L, Zhang Y, Xiang J, Wu Y, Carmichael G, et al. Long Noncoding RNAs with snoRNA Ends. *Mol Cell*. 2012;48:219–30. <https://doi.org/10.1016/j.molcel.2012.07.033>.
23. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*. 2016;539:452–5. <https://doi.org/10.1038/nature20149>.
24. Yap K, Mukhina S, Zhang G, Tan JSC, Ong HS, Makeyev EV. A Short Tandem Repeat-Enriched RNA Assembles a Nuclear Compartment to Control Alternative Splicing and Promote Cell Survival. *Mol Cell*. 2018;72:525–40.e13. <https://doi.org/10.1016/j.molcel.2018.08.041>.
25. Pisignano G, Ladomery M. Epigenetic Regulation of Alternative Splicing: How LncRNAs Tailor the Message. *Noncoding RNA*. 2021;7:21. <https://doi.org/10.3390/ncrna7010021>.
26. Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S, Kageyama Y. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat Cell Biol*. 2007;9:660–5. <https://doi.org/10.1038/ncb1595>.
27. Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP. Peptides Encoded by Short ORFs Control Development and Define a New Eukaryotic Gene Family. *PLoS Biol*. 2007;5: e106. <https://doi.org/10.1371/journal.pbio.0050106>.
28. Hartford CCR, Lal A. When Long Noncoding Becomes Protein Coding. *Mol Cell Biol*. 2020;40:e00528–19. <https://doi.org/10.1128/MCB.00528-19>.
29. Noh JH, Kim KM, Abdelmohsen K, Yoon J, Panda AC, Munk R, et al. HuR and GRSF1 modulate the nuclear export and mitochondrial localization of the lncRNA RMRP. *Genes Dev*. 2016;30:1224–39. <https://doi.org/10.1101/gad.276022.115>.
30. Krause HM. New and Prospective Roles for lncRNAs in Organelle Formation and Function. *Trends Genet*. 2018;34:736–45. <https://doi.org/10.1016/j.tig.2018.06.005>.
31. Lubelsky Y, Ulitsky I. Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature*. 2018;555:107–11. <https://doi.org/10.1038/nature25757>.
32. Azam S, Hou S, Zhu B, Wang W, Hao T, Bu X, et al. Nuclear retention element recruits U1 snRNP components to restrain spliced lncRNAs in the nucleus. *RNA Biol*. 2019;16:1001–9. <https://doi.org/10.1080/15476286.2019.1620061>.
33. Yan P, Luo S, Lu JY, Shen X. Cis- and trans-acting lncRNAs in pluripotency and reprogramming. *Curr Opin Genet Dev*. 2017;46:170–8. <https://doi.org/10.1016/j.gde.2017.07.009>.
34. Ulitsky I, Shkumatava A, Jan C, Sive H, Bartel D. Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. *Cell*. 2012;151:684–6. <https://doi.org/10.1016/j.cell.2012.10.002>.
35. Kirk JM, Kim SO, Inoue K, Smola MJ, Lee DM, Schertzer MD, et al. Functional classification of long non-coding RNAs by k-mer content. *Nat Genet*. 2018;50:1474–82. <https://doi.org/10.1038/s41588-018-0207-8>.
36. Guo C, Ma X, Xing Y, Zheng C, Xu Y, Shan L, et al. Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell*. 2020;181:621–36.e22. <https://doi.org/10.1016/j.cell.2020.03.006>.
37. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discov*. 2011;1:391–407. <https://doi.org/10.1158/2159-8290.CD-11-0209>.
38. Peng W, Koirala P, Mo Y. LncRNA-mediated regulation of cell signaling in cancer. *Oncogene*. 2017;36:5661–7. <https://doi.org/10.1038/onc.2017.184>.
39. Carlevaro-Fita J, Lanzós A, Feuerbach L, Hong C, Mas-Ponte D, Pedersen JS, et al. Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun Biol*. 2020;3:56. <https://doi.org/10.1038/s42003-019-0741-7>.
40. Yang M, Lu H, Liu J, Wu S, Kim P, Zhou X. lncRNAfunc: a knowledgebase of lncRNA function in human cancer. *Nucleic Acids Res*. 2022;50:D1295–306. <https://doi.org/10.1093/nar/gkab1035>.
41. Lo Piccolo L, Yamaguchi M. RNAi of arcRNA hsrw affects sub-cellular localization of *Drosophila* FUS to drive neurodegeneration. *Exp Neurol*. 2017;292:125–34. <https://doi.org/10.1016/j.expneurol.2017.03.011>.
42. Muraoka Y, Nakamura A, Tanaka R, Suda K, Azuma Y, Kushimura Y, et al. Genetic screening of the genes interacting with *Drosophila* FIG4 identified a novel link between CMT-causing gene and long noncoding RNAs. *Exp Neurol*. 2018;310:1–13. <https://doi.org/10.1016/j.expneurol.2018.08.009>.
43. Wei C, Luo T, Zou S, Wu A. The Role of Long Noncoding RNAs in Central Nervous System and Neurodegenerative Diseases. *Front Behav Neurosci*. 2018;12:175. <https://doi.org/10.3389/fnbeh.2018.00175>.
44. Lakhota SC, Mallik M, Singh AK, Ray M. The large noncoding hsrw-n transcripts are essential for thermotolerance and remodeling of hnRNPs, HP1 and RNA polymerase II during recovery from heat shock in *Drosophila*. *Chromosoma*. 2012;121:49–70. <https://doi.org/10.1007/s00412-011-0341-x>.
45. Kukharsky MS, Ninkina NN, An H, Telezkin V, Wei W, Meritens CRd. Long non-coding RNA Neat1 regulates adaptive behavioural response to stress in mice. *Transl Psychiatry*. 2020;10:171. <https://doi.org/10.1038/s41398-020-0854-2>.
46. Gao J, Chow EWL, Wang H, Xu X, Cai C, Song Y, et al. LncRNA DINOR is a virulence factor and global regulator of stress responses in *Candida auris*. *Nat Microbiol*. 2021;6:842–51.
47. Zeni PF, Mraz M. LncRNAs in adaptive immunity: role in physiological and pathological conditions. *RNA Biol*. 2021;18:619–32. <https://doi.org/10.1080/15476286.2020.1838783>.
48. Nakamura A, Amikura R, Mukai M, Kobayashi S, Lasko PF. Requirement for a noncoding RNA in *Drosophila* polar granules for germ cell establishment. *Science*. 1996;274:2075–9. <https://doi.org/10.1126/science.274.5295.2075>.
49. Hardiman KE, Brewster R, Khan SM, Deo M, Bodmer R. The bereft Gene, a Potential Target of the Neural Selector Gene cut. Contributes to Bristle Morphogenesis. *Genetics*. 2002;161:231–47.
50. Pease B, Borges AC, Bender W. Noncoding RNAs of the Ultrabithorax Domain of the *Drosophila* Bithorax Complex. *Genetics*. 2013;195:1253–64. <https://doi.org/10.1534/genetics.113.155036>.
51. Chen B, Zhang Y, Zhang X, Jia S, Chen S, Kang L. Genome-wide identification and developmental expression profiling of long noncoding RNAs during *Drosophila* metamorphosis. *Sci Rep*. 2016;6:23330.
52. Kim M, Faucillion M, Larsson J. RNA-on-X 1 and 2 in *Drosophila* melanogaster fulfill separate functions in dosage compensation. *PLoS Genet*. 2018;14: e1007842. <https://doi.org/10.1371/journal.pgen.1007842>.
53. Xu M, Xiang Y, Liu X, Bai B, Chen R, Liu L, et al. Long noncoding RNA SMRG regulates *Drosophila* macrochaetes by antagonizing scute through E(sp1)mβ. *RNA Biol*. 2019;16:42–53. <https://doi.org/10.1080/15476286.2018.1556148>.



54. Jenny A, Hachet O, Závorszky P, Cyrklaff A, Weston MD, Johnston DS, et al. A translation-independent role of oskar RNA in early *Drosophila* oogenesis. *Development*. 2006;133:2827–33 (doi: 133/15/2827 [pii]).
55. Wen K, Yang L, Xiong T, Di C, Ma D, Wu M, et al. Critical roles of long noncoding RNAs in *Drosophila* spermatogenesis. *Genome Res*. 2016;26:1233–44. <https://doi.org/10.1101/gr.199547.115>.
56. Maeda RK, Sitnik JL, Frei Y, Prince E, Gligorov D, Wolfner MF, et al. The lncRNA male-specific abdominal plays a critical role in *Drosophila* accessory gland development and male fertility. *PLoS Genet*. 2018;14:e1007519. <https://doi.org/10.1371/journal.pgen.1007519>.
57. Li M, Wen S, Guo X, Bai B, Gong Z, Liu X, et al. The novel long non-coding RNA CRG regulates *Drosophila* locomotor behavior. *Nucleic Acids Res*. 2012;40:11714–27. <https://doi.org/10.1093/nar/gks943>.
58. Bender W. MicroRNAs in the *Drosophila* bithorax complex. *Genes Dev*. 2008;22:14–9. <https://doi.org/10.1101/gad.1614208>.
59. Dai H, Chen Y, Chen S, Mao Q, Kennedy D, Landback P, et al. The Evolution of Courtship Behaviors through the Origination of a New Gene in *Drosophila*. *PNAS*. 2008;105:7478–83. <https://doi.org/10.1073/pnas.0800693105>.
60. Aprea J, Prenninger S, Dori M, Ghosh T, Monasor LS, Wessendorf E, et al. Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. *EMBO J*. 2013;32:3145–60. <https://doi.org/10.1038/emboj.2013.245>.
61. Lee H, Zhang Z, Krause HM. Long Noncoding RNAs and Repetitive Elements: Junk or Intimate Evolutionary Partners? *Trends Genet*. 2019;35:892–902. <https://doi.org/10.1016/j.tig.2019.09.006>.
62. Amândio AR, Necsulea A, Joye E, Mascrez B, Duboule D. Hotaïr is dispensable for mouse development. *PLoS Genet*. 2016;12:e1006232. <https://doi.org/10.1371/journal.pgen.1006232>.
63. Bell CC, Amaral PP, Kalsbeek A, Magor GW, Gillinder KR, Tangermann P, et al. The *Evx1/Evx1as* gene locus regulates anterior-posterior patterning during gastrulation. *Sci Rep*. 2016;6:26657. <https://doi.org/10.1038/srep26657>.
64. Han X, Luo S, Peng G, Lu JY, Cui G, Liu L, et al. Mouse knockout models reveal largely dispensable but context-dependent functions of lncRNAs during development. *J Mol Cell Biol*. 2018;10:175–8. <https://doi.org/10.1093/jmcb/mjy003>.
65. Majumder S, Hadden MJ, Thieme K, Batchu SN, Niveditha D, Chowdhury S, et al. Dysregulated expression but redundant function of the long non-coding RNA HOTAIR in diabetic kidney disease. *Diabetologia*. 2019;62:2129–42. <https://doi.org/10.1007/s00125-019-4967-1>.
66. Li C, Shen C, Shang X, Tang L, Xiong W, Ge H, et al. Two novel testis-specific long noncoding RNAs produced by 1700121C10Rik are dispensable for male fertility in mice. *J Reprod Dev*. 2020;66:57–65. <https://doi.org/10.1262/jrd.2019-104>.
67. Zhu Y, Lin Y, He Y, Wang H, Chen S, Li Z, et al. Deletion of lncRNA5512 has no effect on spermatogenesis and reproduction in mice. *Reprod Fertil Dev*. 2020;32:706–13. <https://doi.org/10.1071/RD19246>.
68. Kok F, Shin M, Ni C, Gupta A, Grosse A, van Impel A, et al. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev Cell*. 2015;32:97–108. <https://doi.org/10.1016/j.devcel.2014.11.018>.
69. Schor IE, Bussotti G, Maleš M, Forneris M, Viales RR, Enright AJ, et al. Non-coding RNA expression, function, and variation during *Drosophila* embryogenesis. *Curr Biol*. 2018;28:3547–61.e9. <https://doi.org/10.1016/j.cub.2018.09.026>.
70. Goudarzi M, Berg K, Pieper LM, Schier AF. Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility. *eLife*. 2019;8:e40815. <https://doi.org/10.7554/eLife.40815>.
71. Selli L, Bartolomei MS, Bickmore WA, He L, Stubbs L, Reik W, et al. A Hox-embedded long noncoding RNA: Is it all hot air? *PLoS Genet*. 2016;12:e1006485. <https://doi.org/10.1371/journal.pgen.1006485>.
72. Rappaport Y, Falk R, Achache H, Tzur YB. Linc-20 and linc-9 do not have compensatory fertility roles in *C. elegans*. *MicroPubl Biol*. 2022; <https://doi.org/10.17912/micropub.biology.000524>.
73. Kufel J, Grzechnik P. Small nucleolar RNAs tell a different tale. *Trends Genet*. 2019;35:104–17. <https://doi.org/10.1016/j.tig.2018.11.005>.
74. Bratkovič T, Božič J, Rogelj B. Functional diversity of small nucleolar RNAs. *Nucleic Acids Res*. 2020;48:1627–51. <https://doi.org/10.1093/nar/gkz1140>.
75. Pritykin Y, Brito T, Schupbach T, Singh M, Pane A. Integrative analysis unveils new functions for the *Drosophila* Cutoff protein in noncoding RNA biogenesis and gene regulation. *RNA*. 2017;23:1097–109. <https://doi.org/10.1261/rna.058594.116>.
76. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet*. 2019;20:89–108. <https://doi.org/10.1038/s41576-018-0073-3>.
77. Onishi R, Yamanaka S, Siomi MC. piRNA- and siRNA-mediated transcriptional repression in *Drosophila*, mice, and yeast: new insights and biodiversity. *EMBO Rep*. 2021;22:e53062-n/a. doi: <https://doi.org/10.15252/embr.202153062>.
78. Leader DP, Krause SA, Pandit A, Davies SA, Dow JA. FlyAtlas 2: a new version of the *Drosophila* melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic Acids Res*. 2018;46:D809–15. <https://doi.org/10.1093/nar/gkx976>.
79. Larkin A, Marygold SJ, Antonazzo G, Attrill H, dos Santos G, Garapati PV, et al. FlyBase: updates to the *Drosophila* melanogaster knowledge base. *Nucleic Acids Res*. 2021;49:D899–907. <https://doi.org/10.1093/nar/gkaa1026>.
80. Morozova TV, Shankar V, MacPherson RA, Mackay TFC, Anholt RRH. Modulation of the *Drosophila* Transcriptome by Developmental Exposure to Alcohol. *BMC Genomics*. 2022;23:347. <https://doi.org/10.1186/s12864-022-08559-9>.
81. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, et al. The *Drosophila* melanogaster Genetic Reference Panel. *Nature*. 2012;482:173–8. <https://doi.org/10.1038/nature10811>.
82. Everett LJ, Huang W, Zhou S, Carbone MA, Lyman RF, Arya GH, et al. Gene expression networks in the *Drosophila* Genetic Reference Panel. *Genome Res*. 2020;30:485–96. <https://doi.org/10.1101/gr.257592.119>.
83. EIMaghaby MF, Andersen PR, Pühringer F, Hohmann U, Meixner K, Lendl T, et al. A heterochromatin-specific RNA export pathway facilitates piRNA production. *Cell*. 2019;178:964–79.e20. <https://doi.org/10.1016/j.cell.2019.07.007>.
84. Kneuss E, Munafò M, Eastwood EL, Deumer U, Preall JB, Hannon GJ, et al. Specialization of the *Drosophila* nuclear export family protein Nxf3 for piRNA precursor export. *Genes Dev*. 2019;33:1208. <https://doi.org/10.1101/716258>.
85. Tycowski KT, Steitz JA. Non-coding snoRNA host genes in *Drosophila*: expression strategies for modification guide snoRNAs. *Eur J Cell Biol*. 2001;80:119–25. <https://doi.org/10.1078/0171-9335-00150>.
86. Dieci G, Preti M, Montanini B. Eukaryotic snoRNAs: A paradigm for gene expression flexibility. *Genomics*. 2009;94:83–8. <https://doi.org/10.1016/j.ygeno.2009.05.002>.
87. Lykke-Andersen S, Chen Y, Ardal BR, Lilje B, Waage J, Sandelin A, et al. Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev*. 2014;28:2498–517. <https://doi.org/10.1101/gad.246538.114>.
88. Herter EK, Stauch M, Gallant M, Wolf E, Raabe T, Gallant P. snoRNAs are a novel class of biologically relevant Myc targets. *BMC Biol*. 2015;13:25. <https://doi.org/10.1186/s12915-015-0132-6>.
89. Kanke M, Jambor H, Reich J, Marches B, Gstir R, Ryu YH, et al. oskar RNA plays multiple noncoding roles to support oogenesis and maintain integrity of the germline/soma distinction. *RNA*. 2015;21:1096–109. <https://doi.org/10.1261/rna.048298.114>.
90. Kenny A, Morgan MB, Mohr S, Macdonald PM. Knock down analysis reveals critical phases for specific oskar noncoding RNA functions during *Drosophila* oogenesis. *G3 (Bethesda)*. 2021;11. doi: <https://doi.org/10.1093/g3journal/jkab340>.
91. Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, et al. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics*. 2014;196:961–71. <https://doi.org/10.1534/genetics.113.160713>.
92. Kondo S, Ueda R. Highly improved gene targeting by germline-specific Cas9 expression in *Drosophila*. *Genetics*. 2013;195:715–21. <https://doi.org/10.1534/genetics.113.156737>.
93. Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, et al. Natural variation in genome architecture among 205 *Drosophila* melanogaster Genetic Reference Panel lines. *Genome Res*. 2014;24:1193–208. <https://doi.org/10.1101/gr.171546.113>.

94. Carvalho GB, Ja WW, Benzer S. Non-lethal PCR genotyping of single *Drosophila*. *Biotechniques*. 2009;46:312–4. <https://doi.org/10.2144/000113088>.
95. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2(-\Delta \Delta C(T))$  method. *Methods*. 2001;25:402–8. <https://doi.org/10.1006/meth.2001.1262>.
96. Mackay TFC, Lyman RF, Jackson MS. Effects of P element insertions on quantitative traits in *Drosophila melanogaster*. *Genetics*. 1992;130:313–32.
97. Drapeau MD, Long AD. The Copulatron, a multi-chamber apparatus for observing *Drosophila* courtship behaviors. *Drosophila Information Service*;83: 194–6.
98. Gaertner BE, Ruedi EA, McCoy LJ, Moore JM, Wolfner MF, Mackay TFC. Heritable variation in courtship patterns in *Drosophila melanogaster*. *G3 (Bethesda)*. 2015;5: 531–9. doi: <https://doi.org/10.1534/g3.114.014811>.
99. Morgan TJ, Mackay TFC. Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. *Heredity*. 2006;96:232–42. <https://doi.org/10.1038/sj.hdy.6800786>.
100. Sass TN, MacPherson RA, Mackay TFC, Anholt RRH. High-Throughput Method for Measuring Alcohol Sedation Time of Individual *Drosophila melanogaster*. *J Vis Exp*. 2020. <https://doi.org/10.3791/61108>.
101. Cichewicz K, Hirsh J. ShinyR-DAM: a program analyzing *Drosophila* activity, sleep and circadian rhythms. *Commun Biol*. 2018;1:25. <https://doi.org/10.1038/s42003-018-0031-9>.
102. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9:676–82. <https://doi.org/10.1038/nmeth.2019>.
103. Chen S, Huang T, Zhou Y, Han Y, Xu M, Gu J. AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics*. 2017;18:80. <https://doi.org/10.1186/s12859-017-1469-3>.
104. Bushnell B. BBMap. SOURCEFORGE. 2014.
105. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. GMAP and GSNAP for genomic sequence alignment: Enhancements to speed, accuracy, and functionality. In: Mathé E, Davis S, editors. *Statistical Genomics: Methods and Protocols*. New York, NY: Springer New York; 2016. pp. 283–334.
106. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 2013;41: e108. <https://doi.org/10.1093/nar/gkt214>.
107. Smid M, Coebergh van den Braak, Robert R. J., van de Werken, Harmen J. G., van Riet J, van Galen A, de Weerd V, et al. Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons. *BMC Bioinformatics*. 2018;19: 236. doi: <https://doi.org/10.1186/s12859-018-2246-7>.
108. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003;4:2. <https://doi.org/10.1186/1471-2105-4-2>.
109. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res*. 2013;41:D377–86. <https://doi.org/10.1093/nar/gks1118>.
110. Jia D, Xu Q, Xie Q, Mio W, Deng W. Automatic stage identification of *Drosophila* egg chamber based on DAPI images. *Sci Rep*. 2016;6:18850. <https://doi.org/10.1038/srep18850>.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

