

Research



Cite this article: Salces-Ortiz J, Vargas-Chavez C, Guio L, Rech GE, González J. 2020 Transposable elements contribute to the genomic response to insecticides in *Drosophila melanogaster*. *Phil. Trans. R. Soc. B* **375**: 20190341.
<http://dx.doi.org/10.1098/rstb.2019.0341>

Accepted: 26 September 2019

One contribution of 15 to a discussion meeting issue 'Crossroads between transposons and gene regulation'.

Subject Areas:
evolution

Keywords:
ATAC-seq, RNA-seq, cap-n-collar (cnc), adaptation, transcription factor-binding sites

Author for correspondence:
Josefa González
e-mail: josefa.gonzalez@ibe.upf-csic.es

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.4795821>.

Transposable elements contribute to the genomic response to insecticides in *Drosophila melanogaster*

Judit Salces-Ortiz, Carlos Vargas-Chavez, Lain Guio, Gabriel E. Rech and Josefa González

Institute of Evolutionary Biology (IBE), CSIC-Universitat Pompeu Fabra, Barcelona, Spain

JG, 0000-0001-9824-027X

Most of the genotype–phenotype analyses to date have largely centred attention on single nucleotide polymorphisms. However, transposable element (TE) insertions have arisen as a plausible addition to the study of the genotypic–phenotypic link because of their role in genome function and evolution. In this work, we investigate the contribution of TE insertions to the regulation of gene expression in response to insecticides. We exposed four *Drosophila melanogaster* strains to malathion, a commonly used organophosphate insecticide. By combining information from different approaches, including RNA-seq and ATAC-seq, we found that TEs can contribute to the regulation of gene expression under insecticide exposure by rewiring *cis*-regulatory networks.

This article is part of a discussion meeting issue 'Crossroads between transposons and gene regulation'.

1. Background

Understanding the link between genotype and phenotype is one of the major goals in evolutionary biology [1,2]. Even though substitutions of single nucleotide polymorphisms (SNPs) in the coding and regulatory regions of the genome can cause major changes on phenotypes [3,4], SNPs alone can only explain a fraction of the existing phenotypic variation [5–7]. Other types of mutations such as inversions, segmental duplications, transposable elements (TEs) and other structural variants, are also important sources of phenotypic variation [8–11]. Among these structural variants, TE insertions are likely to play a major role owing to their abundance and activity. TEs make up a sizeable proportion of virtually all genomes analysed to date [12–15]. Moreover, they have the ability to move around the genome generating a wide range of mutations during the process; from gene disruptions to variations in gene regulation [12,16]. Although most TE insertions are expected to be either deleterious or neutral, TEs have also been associated with fast and beneficial changes in phenotypes, fostering rapid adaptation [9,17,18].

There is increasing evidence on the functional role of some TE insertions [16,19–23]. One iconic example is the industrial melanism in the peppered moth. The dark form of the peppered moth, which rapidly increased in frequency in response to coal-polluted environments, is owing to a TE insertion in the first intron of the gene *cortex* [24]. However, we are still far from fully understanding the role TEs play in genotype–environment interactions. *Drosophila melanogaster* represents an excellent model for addressing this question. First, it has a small and well-annotated genome, which allows the putative genomic regions involved in the phenotypic changes to be more easily characterized [25]. Second, *D. melanogaster* has colonized many different environments in recent evolutionary time [26–28], demonstrating its ability to adapt very quickly, which has led to extensive phenotypic variation within and between

populations [29]. Finally, *D. melanogaster* contains a wide repertoire of TE families, many of which contain polymorphic copies, suggesting that they are highly active [30–34]. Indeed, for some families, there is experimental evidence showing that they are active [35,36]. Moreover, several of these polymorphic TEs have been involved in phenotypic changes, probably linked to adaptive processes [37–44].

Besides works linking individual copies to phenotypic changes, an increasing number of investigations suggest that TEs have contributed to the rewiring of *cis*-regulatory networks during evolution, including pathways underlying processes such as pregnancy, brain development, innate immunity and stress response [17,45–49]. Owing to their repetitive and dynamic nature, TEs can distribute regulatory sequences such as promoters, transcription factor-binding sites and insulators across the genome altering gene expression [50,51]. In this context, it is reasonable to think that TEs can act as powerful agents to modify biological processes by creating *cis*-regulatory networks and rewiring already existing ones.

Pesticide resistance is an example of a rapid adaptive process that has resulted from a novel selective pressure [52]. In *Drosophila*, there is evidence to suggest that insecticides have played a large role in recent selection [53–58]. There are several examples of insecticide resistance in this species which has resulted from a TE affecting the expression of a relevant gene [38,40,43,44,59]. However, to date, to our knowledge, no genome-wide analysis of the potential role of TEs on the transcriptional response to insecticides has been performed.

In this work, we analysed gene expression profiles (RNA-seq), chromatin accessible regions (ATAC-seq), binding site predictions for the major transcriptional regulator of xenobiotic detoxification, cap-n-collar (*cnc*), [49,60] and signatures of selection in regions flanking TE insertions [33,49], in order to investigate the contribution of TEs to the regulation of gene expression in response to insecticide exposure. For this purpose, we exposed four *D. melanogaster* strains to malathion, a commonly used organophosphate insecticide (<http://www.epa.gov>). Our results suggest that TEs can contribute to the regulation of gene expression under insecticide exposure by rewiring *cis*-regulatory networks.

2. Methods

(a) Fly stocks

Four *D. melanogaster* strains were used: *SE_Sto_11_22* (*SE-Sto*), an iso-female strain collected in Stockholm, Sweden, in 2011 [61]; *RAL-375* and *RAL-177*, two strains from the *Drosophila* Genetic Reference Panel (DGRP) collected in Raleigh, North Carolina, USA [62], and the reference sequenced strain *iso-1* [63]. Flies were reared on fly food medium in a 12 : 12 h light/dark cycle at 25°C.

(b) Xenobiotic exposure

To induce xenobiotic response, we used malathion, an organophosphate insecticide commonly used to control a variety of insects that attack fruits (<http://www.epa.gov>). Malathion was dissolved in 2-propanol and added to agar–sucrose medium to a final concentration of 20 µM. We used a unique dose for all strains as this is more similar to what flies experience in nature. For non-stress conditions, we used agar–sucrose food. We transferred the flies to new tubes with or without malathion and kept them for 9 h at 25°C before gut dissection. Three biological replicates of 30 females each were performed for each condition and strain. No mortality was observed in any of

the control replicates. After 9 h of malathion exposure, no mortality was observed in *SE-Sto* and *RAL-375*, while 15% and 25% mortality was observed in *RAL-177* and *iso-1*, respectively.

(c) RNA isolation, library preparation and sequencing

Guts from 4- to 6-day-old females were dissected in 1× phosphate buffered saline (PBS) either under non-stress conditions or after xenobiotic exposure (stress). Total RNA was isolated using GenElute Mammalian genomic total RNA miniprep kit from SIGMA following the manufacturer's instructions. A 1.5 µg of total RNA from each sample was used for subsequent library preparation and sequencing. Briefly, library preparation was performed using the Truseq Stranded mRNA Sample Prep kit from Illumina following the manufacturer's instructions. Libraries were sequenced using Illumina 125 bp paired-end reads (25.4–57.8 million per sample, electronic supplementary material, table S1A).

(d) Analysis of RNA-seq data

Quality of the fastq sequencing files was assessed using FASTQC v. 0.11.8 (www.bioinformatics.babraham.ac.uk/projects/fastqc). TRIMGALORE v. 0.5.0 (www.bioinformatics.babraham.ac.uk/projects/trim_galore) was used for adapter contamination removal and CUTADAPT v. 1.18 (default parameters) was used for low-quality trimming [64]. Trimmed reads were mapped to the *D. melanogaster* genome r6.15 using HISAT2 v. 2.1.0 [65]. On average, 93.3% of the reads were uniquely mapped to the genome. We explored technical duplications in our samples using DUPRADAR [66]. Overall, we found no bias towards high number of duplicates at low read counts, so we did not remove duplicates from the alignments. We used FEATURECOUNTS v. 1.6.2 [67] for counting the number of reads mapping to genes for each sample (reverse-stranded parameter). The matrix of counting data was then imported into DESeq2 [68], an R bioconductor package, using the DESeqDataSetFromMatrix function. The DESeq main function, with default parameters, was used to obtain the library size-normalized read counts, which is used to identify differentially expressed genes (DEGs) and to generate the heat map and the principal component plots. DEGs were identified in pairwise comparisons modelling the samples as: ~strain + treatment + strain:treatment and adjusted using the Benjamini–Hochberg method to control for false discovery. The significant DEGs were identified after applying significance cut-offs (adjusted *p*-value ≤ 0.05 and fold-change ≥ 1.5). Note that DESeq2 normalization makes the expression of all genes comparable between samples, independently of sequencing depth and RNA composition [68].

(e) Gene ontology analysis

Gene ontology (GO) enrichment analysis for biological process of the DEGs was performed using the functional analysis and clustering tool DAVID v. 6.8 (<https://david.ncifcrf.gov>) with default options [69,70]. All biological clusters above a score of 1.3 were considered as significantly enriched (electronic supplementary material, table S2) [70].

(f) Generation of protein–protein interaction networks and identification of hub genes

Protein–protein interaction (PPI) networks were constructed with the DEGs of each strain using STRING v. 11.0 [71] taking into account experimental data, databases and gene co-expression data. The minimum required interaction score cut-off was set at 0.5. The PPI database generated by STRING was used in CYTOSCAPE to rank nodes in the network by their network features. Hub genes (genes with a high number of edges) were also explored using the CytoHubba plugin r1.6 (<http://hub.iis>).

sinica.edu.tw/cytoHubba/index.html) in CYTOSCAPE. In this work, we first compared Degree (the most typically used methodology for PPI analysis) and Maximal Centrality Clique (MCC) algorithms to look for significant hub genes. According to the CytoHubba developer site (<http://hub.iis.sinica.edu.tw/cytoHubba/supplementary/index.htm>), Degree rank genes based on its number of interactions, while MCC captures proteins that are tightly connected to others and it is able to discover new feature nodes. The overlap between Degree and MCC methodologies in our work was between 78.4% and 100% when analysing the top 15% hub genes from each methodology. Thus, we used the MCC score to rank the top 15% hub genes from the four predicted PPI networks (electronic supplementary material, table S3A–D).

(g) Transposable element dataset

The release 6 of the reference genome (*iso-1* strain) contains 5416 TE insertions [25]. We used *T-lex3* [72] to genotype these insertions in *SE-Sto*, *RAL-375* and *RAL-177* using whole-genome sequencing data [61,62]. Note that *T-lex3* cannot accurately estimate frequencies for TEs that are nested or part of segmental duplications (1552 TEs [33]) and failed to genotype another 65 insertions. Thus, we were able to genotype 3784 in at least one of the three strains (electronic supplementary material, table S4A). Most of these insertions, 78.6% (2975 out of 3784) were present in the four strains analysed, 11.9% (449 out of 3784) were only present in the reference genome (*iso-1*) and 9.5% (360 out of 3784) were polymorphic in the four strains (electronic supplementary material, table S4A).

(h) DNA isolation for ATAC-seq

The protocol used for ATAC-seq was adapted from Buenrostro *et al.* [73] to be used for *D. melanogaster* tissue. Thirty guts from *Drosophila* females between 4 and 6 days old were isolated and immediately placed in 200 μ l ice-cold PBS. Guts were lysed and homogenized to isolate between 50 000 and 100 000 cells using a Neubauer camera. The cells were washed and lysed to obtain a pellet with crude nuclei. The nuclei were resuspended in 47.5 μ l of tagmentation buffer and incubated for 30 min at 37°C with 2.5 μ l of Nextera Tn5 Transposase. DNA was purified using a Qiagen MinElute PCR Purification Kit. Finally, a library was generated using the Nextera kit and amplified by polymerase chain reaction. All libraries were sequenced using *Illumina* 50 bp paired-end reads (25 to 51.9 million per sample, electronic supplementary material, table S5A).

(i) Analysis of ATAC-seq data

The general quality of the raw reads was evaluated using FASTQC v. 0.11.8 (www.bioinformatics.babraham.ac.uk/projects/fastqc). The adapters were identified and removed from the reads using *bbduk* from the BBTOOLS suite v. 38.00 (<https://sourceforge.net/projects/bbmap/>). Next, the reads were trimmed using TRIMMOMATIC [74] with the following parameters: *LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36*. The filtered reads were mapped against the *D. melanogaster* genome r6.15 using BOWTIE v. 1.2.2 [75] with the following parameters: *-S -v 3 -a -m 100 -best -strata -X 2000 -p 10 -interleaved*. The output files were processed with CSEM v. 2.4 [76] and filtered using *piPipes_bam_ZW_filter* from PIPIPES [77]. Finally, only pairs of reads with an insert size smaller or equal to 100 bp were conserved.

The peaks were predicted following the strategy described in the ENCODE ATAC-seq pipeline (<https://github.com/ENCODE-DCC/atac-seq-pipeline>). The *bam* files were first deduplicated. Next, they were converted into *tagAlign* (*bed*) format. The *tagAlign* files for the three replicates for each strain under the same conditions were merged and then split into two sets randomly. Peaks were predicted in each of the two sets using MACS2 v. 2.1.2 [78] with the following parameters:

-p .01 -shift 75 -extsize 150 -nomodel -B -SPMR -keep-dup all -call-summits. Only peaks with an overlap greater than 50% between both sets of predicted peaks were conserved. For this set of peaks, the irreproducible discovery rate (IDR) was calculated and only peaks with an IDR less than 0.05 were conserved. The peaks were rescaled around the predicted summits by extending 100 bp to both sides of the summit. Finally, the predictions for all strains in both conditions were merged into a single-bed file comprising the universal set of peaks. To define whether a peak was open or not, the background noise was calculated generating windows spread randomly throughout the genome and calculating the median of the coverage in each window. If the median coverage in a peak was above the background noise, it was considered to be open.

(j) Expression analysis of transposable element families

For both mapping and analysing the expression of TEs, we used the TETOOLS pipeline [79]. Briefly, data were trimmed using URQT [80] in order to remove low-quality nucleotides. The resulting trimmed reads were aligned to a TE library using BOWTIE2 v. 2.3.4.1 [81]. The TE library was created from the TE family information obtained from Repbase. The read count step was computed per TE family, adding all reads mapped on copies of the same family. Finally, we performed the differential expression analysis between TE families using the R Bioconductor package DESeq2 [68] on the raw read counts retaining the TE families with a *p*-value ≤ 0.05 and fold-change ≥ 3 .

3. Results

(a) Strains tolerant to malathion showed a lower number of differentially expressed genes compared with sensitive strains

To investigate the effect that malathion has on gene expression and the potential role of TEs in these changes, we performed RNA-seq analysis on four *D. melanogaster* strains that differed in their tolerance to malathion. After 9 h of malathion exposure, no mortality was observed in *SE-Sto* and *RAL-375*, while 15% and 25% mortality was observed in *RAL-177* and *iso-1*, respectively. Overall, the number of DEGs ranged from 153 to 2778, similar to other transcriptomic studies performed with xenobiotics (figure 1 and table 1; electronic supplementary material, table S1B–E) [58,82]. Among these DEGs, we found several xenobiotic-related genes including *Cyp12d1-d*, *Cyp6g1*, *Jheh1* and *Jheh2*, that have previously been identified as major malathion resistance loci (electronic supplementary material, table S1F) [83].

The number of DEGs was lower in the more tolerant strains compared with the more sensitive ones (figure 1 and table 1). Fifty-one per cent of the DEGs in the *SE-Sto* tolerant strain were also found to be differentially expressed in the *iso-1* sensitive strain (table 1). However, the majority of these genes (75 of 78) showed opposite expression patterns, consistent with the different phenotyping profiles of these strains (electronic supplementary material, figure S1). These 75 genes were enriched for ‘cell cycle’ biological processes (electronic supplementary material, table S2A). Moreover, the most sensitive strains (*RAL-177* and *iso-1*) shared 41.2% of the DEGs and most of them (1054 of 1096) showed the same pattern of expression, which probably reflects that both strains are highly stressed and struggling to survive (table 1; electronic supplementary material, figure S1). Indeed, the shared upregulated DEGs are mainly related with ‘cellular response to stimulus’,

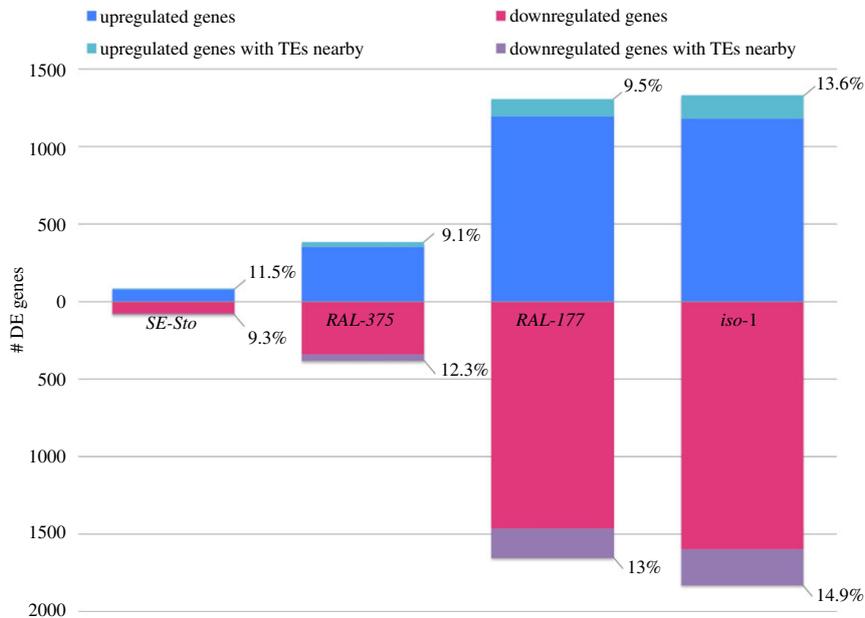


Figure 1. Number of DEGs in response to malathion. DEGs in the *SE-Sto* strain collected in Stockholm (Sweden), two North American strains from the DGRP panel, *RAL-375* and *RAL-177*, and the sequenced strain *iso-1*. The percentage of DEGs with TEs nearby per strain is also shown. Note that while no mortality after 9 h of malathion exposure was found for *SE-Sto* and *RAL-375*, 15% and 25% mortality was found for *RAL-177* and *iso-1*, respectively.

Table 1. Number of DEGs in response to malathion exposure per strain, and number of DEGs shared among strains. (Percentages in parentheses correspond to the strain in the column.)

no. DEGs	<i>SE-Sto</i>	<i>RAL-375</i>	<i>RAL-177</i>	<i>iso-1</i>
<i>SE-Sto</i>	153			
<i>RAL-375</i>	17 (11.1%)	694		
<i>RAL-177</i>	40 (26.1%)	311 (44.8%)	2659	
<i>iso-1</i>	78 (51%)	304 (43.8%)	1096 (41.2%)	2778

‘cellular transport’ and several ‘metabolic processes’, while the shared downregulated DEGs have mostly functions related with ‘circadian rhythm’, ‘ageing’ or ‘molting’ (electronic supplementary material, table S2B).

(b) Only metabolic and stress response genes are upregulated after malathion exposure in the most tolerant strains

Besides analysing the GO enrichment of DEGs shared between strains, we also analysed each strain individually. In the *SE-Sto*, only ‘metabolic processes’ were upregulated, while in the *RAL-375*, ‘response to insecticide’ was also upregulated, suggesting that this strain was more affected by stress (figure 2; electronic supplementary material, table S2C,D). In both cases, the upregulation of metabolic-related genes suggests the potential detoxification ability via metabolic processes of these strains [58]. In *RAL-177*, besides ‘response to stress’ and ‘metabolism’, ‘cell cycle’, ‘chromatin organization’ and ‘nuclear division’ were also upregulated (figure 2; electronic supplementary material, table S2E). In addition to these, ‘negative regulation of transcription’ and ‘negative ribosome biogenesis’ are upregulated in the *iso-1* strain, suggesting that this strain is entering apoptosis (electronic supplementary material, table S2F). Finally, the

number of downregulated genes involved in ‘lipid and carbohydrate metabolism’ was higher in sensitive strains as previously reported by Riahi *et al.* [58] (figure 2; electronic supplementary material, table S2E,F). Overall, the patterns of GO enrichment were in agreement with the differences in malathion tolerance observed across strains.

(c) Up to 14.4% of differentially expressed genes are located nearby transposable elements

To identify the role that TEs could have in response to malathion, we analysed the proportion of DEGs that were located nearby annotated TEs (less than 1 kb or inside the gene). For the *iso-1* strain, we analysed all the annotated TEs (5416 TEs [25]) and for the other three strains, we analysed the TEs that could be genotyped using *T-lex3* [72] (3741 in *SE-Sto*, 3637 in *RAL-375* and 3651 in *RAL-177*, electronic supplementary material, table S4A, see Material and methods). We found that up to 14.4% of DEGs are located nearby annotated TEs (figure 1; electronic supplementary material, table S4B). DEGs were enriched nearby TE insertions compared with the overall distribution of genes nearby TEs in the genome for the four strains tested, although this enrichment was only significant in *RAL-177* and *iso-1* (χ^2 test, p -value = 0.00019 and 0.005294, respectively; electronic supplementary material, table S4B). Among those TEs found in proximity to DEGs, we found *FBti0019430*, *FBti0018880* and *FBti0019627*, which have been shown to play a role in insecticide resistance [38,40,43,84].

To identify those TEs that are likely to have a greater effect in the response to malathion exposure, we used DEGs to construct PPI networks. We focused on the top 15% of genes according to the MCC score, which captures proteins that are tightly connected to other proteins. Only 76 of the 307 unique hub genes (approx. 25%) have previously been identified as xenobiotic stress response candidates (electronic supplementary material, table S3E). We found that 38 of the hub genes have TEs nearby (electronic supplementary

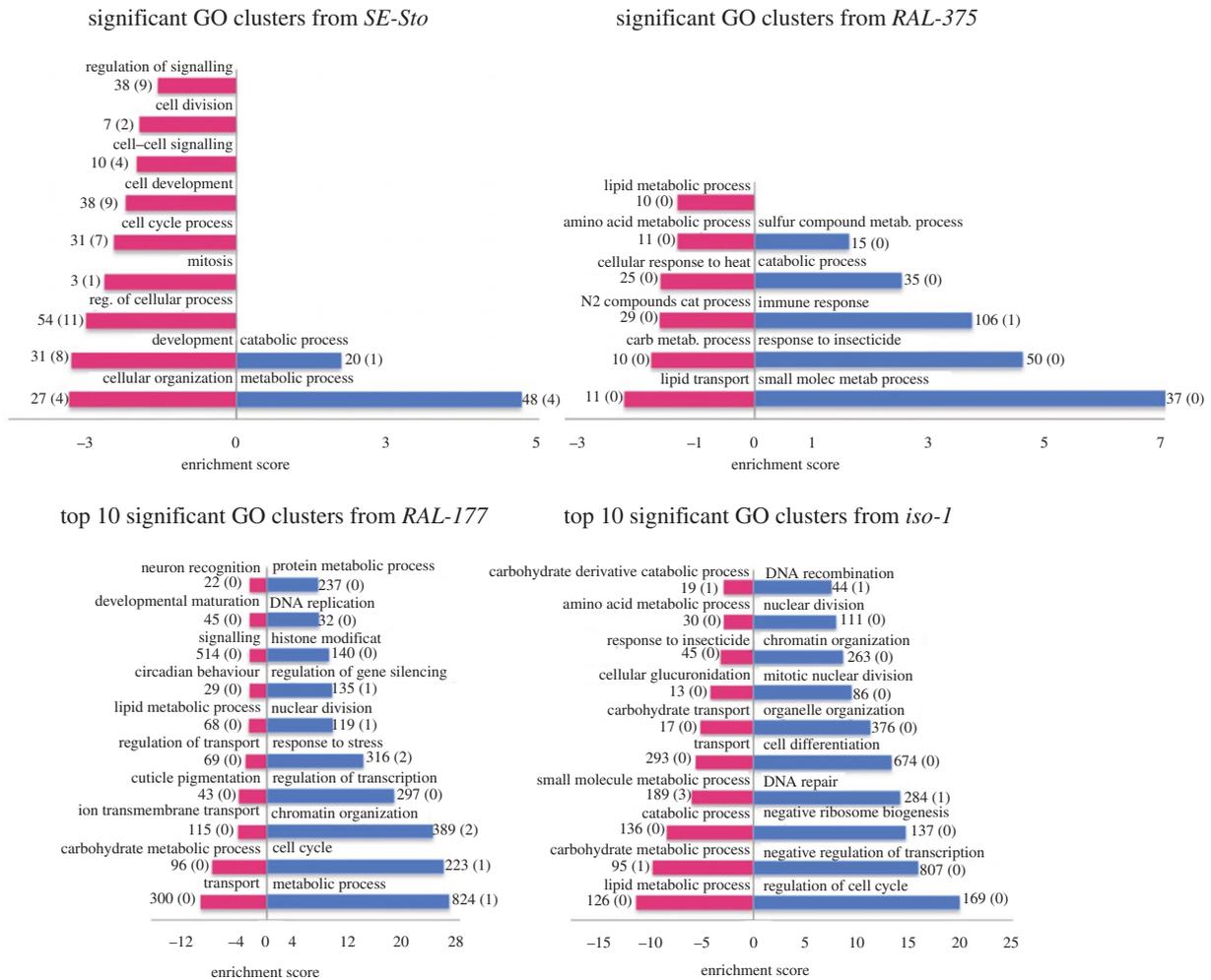


Figure 2. Significant gene ontology (GO) clusters according to DAVID functional annotation tool for upregulated (blue) and downregulated (red) genes. For strains with more than 10 significant GO clusters, only the top 10 are shown (enrichment score greater than 1.3). The horizontal axis represents the enrichment score. For each cluster, the number of DEGs and the number of DEGs with TEs nearby (in parentheses) are given.

material, table S3E). In most cases, these genes are within the ones with the highest MCC values (figure 3). Moreover, some of the hub genes that connect subnetworks are located nearby TEs, such as *LysS*, *arp2* and *Mmp2* in *RAL-177* (figure 3). Thirty of these hub genes have a single TE nearby, which could be considered the most likely candidate to affect their expression (electronic supplementary material, table S3F). Indeed, seven of these 30 TEs have previously been identified as showing signatures of selection, and most of them have been related to stress response (table 2; electronic supplementary material, table S3F) [33,49]. We checked whether any of the 17 TEs located nearby upregulated hub genes contained binding sites for *cnc*, the major transcription factor in xenobiotic and oxidative stress response [60]. We found that five TEs contained *cnc*-binding sites, including three TEs with evidence of selection (table 2; electronic supplementary material, table S3F) [33,49]. Overall, we identified 30 TEs located nearby hub DEGs and thus, likely to have a bigger effect in malathion response. Nine of these TEs have evidence of selection and/or *cnc*-binding sites (table 2).

(d) Most of the differentially expressed genes have ATAC-seq peaks

We analysed the chromatin accessibility genome-wide in non-stress and stress conditions. The total number of ATAC-seq

peaks identified was similar among strains, 23 562 on average, and comparable with the previous studies (electronic supplementary material, table S5B) [90–92]. As expected, most of the peaks were located in gene bodies or promoter regions (approx. 92.2%) (electronic supplementary material, table S5B) [90]. If we focus on DEGs, approximately 83.6% of them had peaks assigned to their gene bodies or promoter regions. Although some upregulated genes showed increased accessibility in stress compared with non-stress conditions, most of the genes did not, as has been previously reported (figure 4; electronic supplementary material, table S5C) [93]. Similarly, most downregulated genes did not show decreased accessibility in stress versus non-stress conditions (electronic supplementary material, table S5C). Finally, we also tested whether peaks located nearby upregulated genes had *cnc* motifs. We found that, on average, approximately 15% of the peaks contained *cnc* motifs except in *SE-Sto*, where this percentage was 6.5%. This lower percentage in the *SE-Sto* strain is consistent with the lower number of DEGs and the lack of mortality, suggesting that this strain is experiencing lower levels of stress.

(e) Half of the transposable elements located in open chromatin regions are distal to genes

We identified 199 TEs across strains that overlap with ATAC-seq peaks (electronic supplementary material, table S5D).

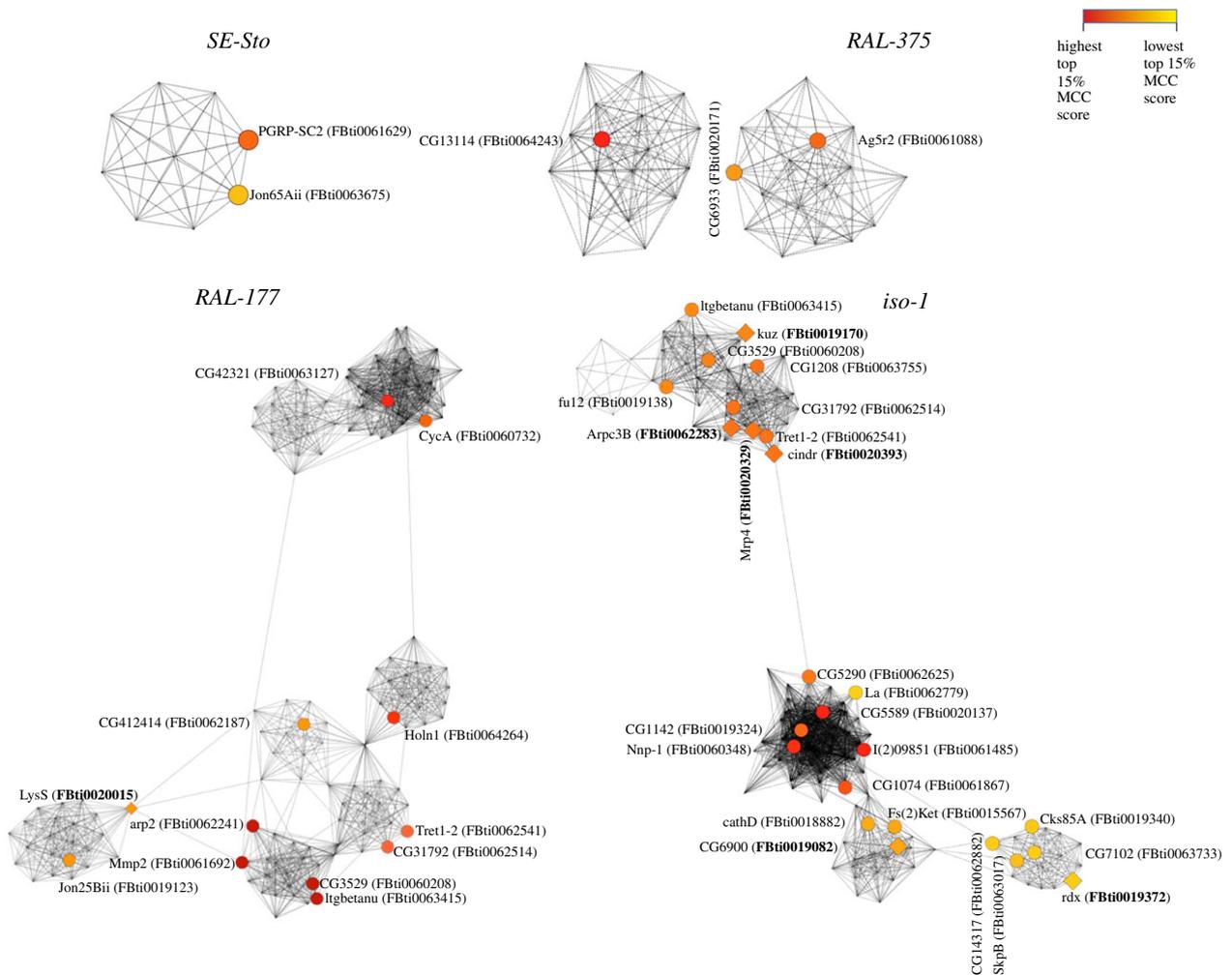


Figure 3. PPI networks per strain are shown for the DEGs with the top 15% MCC scores. Colour gradient indicates the MCC values: from higher MCC scores (red) to lower MCC scores (yellow). Circular nodes represent genes with TEs nearby, while diamond-shaped nodes represent genes with TEs nearby with evidence of selection (in bold) [33,49]. Note that four hub genes nearby TEs are present both in *RAL-177* and in *iso-1*.

Table 2. Description of the TE insertions located nearby differentially expressed hub genes and having either evidence of selection, cnc-binding sites (BS), or ATAC-seq peaks.

candidate	nearby	evidence of	cnc	ATAC-seq			
TEs	TE family	gene	TE location	gene product/stress related	selection [33,49]	BS	peaks
<i>FBti0019170</i>	<i>F-element</i>	<i>kuz</i>	inside gene	metalloendopeptidase/zinc tolerance related [85]	fTE	0	—
<i>FBti0019372</i>	<i>S-element</i>	<i>rdx</i>	inside gene	ubiquitin ligase binding protein/-	H12	1	—
<i>FBti0019082</i>	<i>Rt1b</i>	<i>CG6900</i>	5' of gene	pseudogene /-	TajimaD	3	yes
<i>FBti0020393</i>	<i>1360</i>	<i>cindr</i>	Inside gene	adaptor protein/-	TajimaD	0	—
<i>FBti0020329</i>	<i>G5</i>	<i>Mrp4</i>	5' of gene	ABC transporter/response to oxidative stress [86]	TajimaD	0	—
<i>FBti0062283</i>	<i>ninja-Dsim-like</i>	<i>Arpc3B</i>	5' of gene	actin-related protein 2/3 complex/cold tolerance and bacterial infection	TajimaD	0	—
<i>FBti0020015</i>	<i>412</i>	<i>LysS</i>	5' of gene	lysozyme/immune response [87,88]	young and long	3	—
<i>FBti0015567</i>	<i>Tirant</i>	<i>Fs(2)Ket</i>	inside gene	nuclear protein import/-	—	1	—
<i>FBti0020137</i>	<i>S-element</i>	<i>CG5589</i>	3' of gene	adenosintriphosphate/-	—	1	—
<i>FBti0062779</i>	<i>1360</i>	<i>La</i>	5' of gene	la autoantigen-like/-	—	0	yes
<i>FBti0061088</i>	<i>INE-1</i>	<i>Ag5r2</i>	5' of gene	antigen 5-related 2/-	—	0	yes
<i>FBti0064264</i>	<i>INE-1</i>	<i>Holn1</i>	3' of gene	U5 small nuclear ribonucleoprotein particle/wound response [89]	—	0	yes

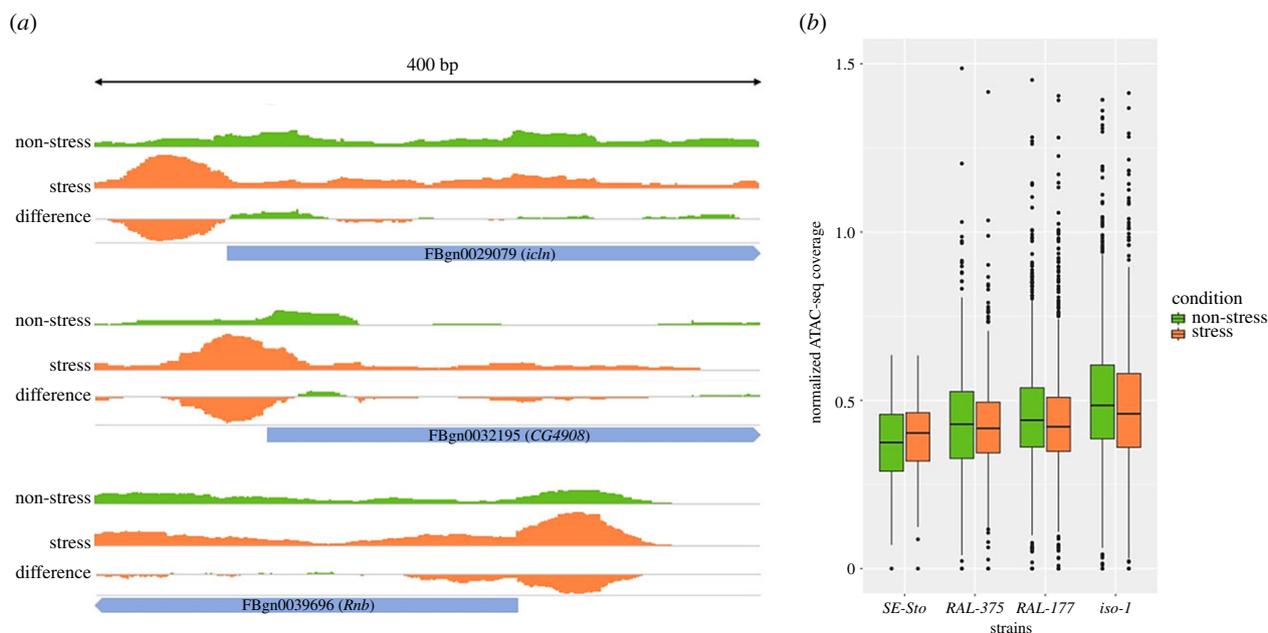


Figure 4. Changes in chromatin accessibility after malathion exposure. (a) Changes in the chromatin accessibility at the promoter region of three individual upregulated genes. (b) Global ATAC-seq coverage in all upregulated genes in each strain. The values are scaled, so that the median values of the first and 20th quantiles correspond to 0 and 1.

Most of these TEs showed peaks in both non-stress and stress conditions (74%). Still, on average, approximately 19.1% were only present in non-stress conditions and approximately 22.6% were only present in stress conditions. Ninety-five of these TEs were located inside genes or in promoter regions, and 30 of them were located inside DEGs or in their promoter regions (electronic supplementary material, table S5E). These 30 TEs are enriched for 1360 family (electronic supplementary material, table S5E). Three of these 30 TEs have evidence of selection (*FBti0019430*, *FBti0060443* and *FBti0019082*), two contain cnc-binding sites (*FBti0061018* and *FBti0062187*) and three are located in the promoter or inside hub genes (the aforementioned *FBti0019082* and *FBti0061088*, and *FBti0062779*) (table 2).

However, the other 104 TEs with peaks are distal to genes. These TEs are enriched for several families including *Rtlb* and *gypsy8* (electronic supplementary material, table S5D). Two of the 104 TEs have evidence of selection (*FBti0019419* and *FBti0019355*) and nine contained a cnc motif (electronic supplementary material, table S5F). Thus, our results suggest that half of the TEs located in open chromatin are distal to genes.

(f) Fifty transposable element families showed differences in expression after malathion exposure

Finally, we also checked whether the expression of TE families was affected by the malathion treatment. While no differentially expressed TE families were detected in *SE-Sto*, 50 families were differentially expressed between non-stress and stress conditions in at least one of the other three strains (figure 5; electronic supplementary material, table S6A). The pattern of expression was family dependent as has been previously suggested (figure 5) [94]. Similar to the differences in the number of DEGs across strains, we observed that the most tolerant strains have a lower number of differentially expressed TE families compared with the most sensitive strains (Pearson correlation $r = 0.8$).

Both upregulated and downregulated TE families were enriched for long terminal repeat (LTR) elements when compared with all the TEs annotated in the genome (χ^2 , p -value = 1.78×10^{-30} and 8.26×10^{-25} for upregulated and downregulated, respectively).

We checked whether copies from these differentially expressed TE families were located nearby DEGs. We found that *diver2*, *Bari1* and *accord* copies were mostly located nearby upregulated genes (electronic supplementary material, table S6B). Thus, we cannot discard that the upregulation observed for these three families was owing to the upregulation of genes located nearby.

Finally, we also analysed an RNA-seq dataset obtained from two *D. melanogaster* strains that differed in their dichlorodiphenyltrichloroethane tolerance [54]. We identified 36 differentially expressed TE families, 24 of them overlapped with our dataset of differentially expressed families (figure 5). However, this overlap is not statistically significant (χ^2 test, p -value = 0.3314), and most of the observed changes were in the opposite direction (electronic supplementary material, table S6C).

4. Discussion

In this work, we analysed the gene expression profiles (RNA-seq), the chromatin accessible regions (ATAC-seq), cnc-binding site predictions [49,60] and signatures of selection in regions flanking TE insertions [33,49], to identify TE insertions likely to contribute to the genomic response to malathion. Our analysis of four *D. melanogaster* strains has shown that the number, the pattern of expression (up- or downregulation) and the GO enrichment of DEGs in response to malathion were consistent with the differences among strains in their tolerance level to this insecticide (figures 1 and 2; electronic supplementary material, figure S1). Our results are consistent with a recent analysis of a mutant *D. melanogaster* strain sensitive to xenobiotics in which the number of DEGs was also higher than in the most tolerant



Figure 5. Heat map showing the differentially expressed TE families after malathion exposure. On the left, the three strains with differentially expressed TEs from our work. In bold, the TE families shared among the strains. Four families, *flea*, *gypsy-2*, *ldefix* and *micropia*, were differentially expressed in the three strains, 12 in two strains and 34 only in one strain. On the right, the expression profile of TE families using data from Seong *et al.* [54].

control strain [58]. While similar results have been found in the green peach aphid [95], more DEGs were found in the most resistant strains of *Aedes Aegypti* mosquitoes [96].

Our results also suggest that the control of energy consumption is relevant to stress response, as the number of downregulated genes involved in lipid and carbohydrate metabolism was higher in sensitive strains [58]. Finally, while we found that several of the DEGs have been previously identified as candidates for xenobiotic response,

including four of the five major genes, we failed to identify others. However, it is known that the expression of stress response genes is time dependent, with genes not actively expressed along the entire stress period [83,97–99].

We identified 38 hub genes located nearby TE insertions. However, only three of these TEs had an ATAC-seq peak and were located in the promoter or inside the hub gene. One possible explanation is that some TEs are bound by numerous transcription factors and other co-activators that could

prevent the Tn5 transposase from cutting in these regions [100,101]. Indeed, for some stress genes, it has been shown that *RNA polymerase II* is associated with their promoter regions prior to the induction of the stress [102]. In addition, the number of TEs with identifiable peaks might be underestimated. Given the repetitive nature of the TEs, there are limitations when attempting to accurately map the reads to their correct position. While using CSEM [76] increases the number of uniquely mapping reads and thus reveals additional peaks, the remaining discarded multi-mapping reads might have arisen from TEs located in regions with open chromatin. Combining ATAC-seq with histone mark information could further inform on the potential role of TEs as enhancers and promoters, although it has been shown that this is not always the case [49,103].

If we consider those TEs that are located nearby differentially expressed hub genes and have either ATAC-seq peaks, evidence of selection and/or cnc-binding sites, we identified a dataset of 12 insertions, seven of them located nearby genes not previously related to stress response (table 2). However, *rdx* regulates the *Hedgehog* signalling pathway involved in cell survival under stress conditions [104], and *cindr* is a multi-adaptor protein that has been related to the activation of the p38 pathway in response to oxidative stress [105,106]. Thus, our results suggest that TEs located nearby these genes could play an important role in xenobiotic stress response. Although there is not a clear association between the presence/absence of these TE insertions and the change in expression of their nearby genes (electronic supplementary material, table S4C), this result is consistent with previous analysis showing that the effect of TEs on the expression of nearby genes is background dependent [39,94]. More genetic backgrounds should be analysed to elucidate whether TEs have a specific effect restricted to particular backgrounds or whether the effect of TEs is more general.

5. Conclusion

We found that TEs can contribute to the genome-wide response to insecticide resistance as suggested by the association of TEs with differentially expressed hub genes. Other TEs identified in this work can also influence insecticide response, as exemplified by *FBti0019430*, *FBti0018880* and *FBti0019627* that have been previously reported to be involved in insecticide response [38,40,43,84]. Our results also suggest that the effect of TEs on gene expression in response to insecticides is background dependent. Functional validation of the candidate TEs in several backgrounds would help determine whether the effect of TEs in response to insecticide is global or restricted to particular backgrounds.

Data accessibility. Sequencing data are deposited at SRA with the accession number PRJNA551114. The additional datasets supporting this article have been uploaded as part of the supplementary material.

Authors' contributions. J.S.-O. participated in data analysis and drafted the manuscript; C.V.-C. participated in data analysis and drafted the manuscript; L.G. carried out the molecular laboratory work; G.E.R. participated in data analysis and helped draft the manuscript; J.G. designed the study, coordinated the study and drafted the manuscript. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by the European Commission (grant no. H2020-ERC-2014-CoG-647900), by the Ministerio de Ciencia, Innovación y Universidades/AEI/FEDER (grant no. BFU2017-82937-P) and by the 'Secretaria d'Universitats i Recerca. Departament d'Economia i Coneixement. Generalitat de Catalunya' (grant no. GRC 2017 SGR 880). J.S.-O. is funded by a Juan de la Cierva-Formación fellowship (no. FJCI-2016-28380).

Acknowledgements. We thank Llewellyn Green and Sara Guirao-Rico for comments on the manuscript, and María Bogaerts-Márquez for running *T-lex3* in the strains analysed in this work.

References

- Orgogozo V, Morizot B, Martin A. 2015 The differential view of genotype–phenotype relationships. *Front. Genet.* **6**, 179. (doi:10.3389/fgene.2015.00179)
- Taylor P, Lewontin R. 2017 The genotype/phenotype distinction. In *The Stanford encyclopedia of philosophy* (summer 2017 edition) (ed. EN Zalta). See <https://plato.stanford.edu/archives/sum2017/entries/genotype-phenotype/>.
- Shastri BS. 2009 SNPs: impact on gene function and phenotype. *Methods Mol. Biol.* **578**, 3–22. (doi:10.1007/978-1-60327-411-1_1)
- Merkulov VM, Leberfarb EY, Merkulova TI. 2018 Regulatory SNPs and their widespread effects on the transcriptome. *J. Biosci.* **43**, 1069–1075. (doi:10.1007/s12038-018-9817-7)
- Marigorta UM, Rodríguez JA, Gibson G, Navarro A. 2018 Replicability and prediction: lessons and challenges from GWAS. *Trends Genet.* **34**, 504–517. (doi:10.1016/j.tig.2018.03.005)
- Boyle EA, Li YI, Pritchard JK. 2017 An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186. (doi:10.1016/j.cell.2017.05.038)
- Korte A, Farlow A. 2013 The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* **9**, 29. (doi:10.1186/1746-4811-9-29)
- Zhou J, Lemos B, Dopman EB, Hartl DL. 2011 Copy-number variation: the balance between gene dosage and expression in *Drosophila melanogaster*. *Genome Biol. Evol.* **3**, 1014–1024. (doi:10.1093/gbe/evr023)
- Chuong EB, Elde NC, Feschotte C. 2017 Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86. (doi:10.1038/nrg.2016.139)
- Kapun M, Fabian DK, Goudet J, Flatt T. 2016 Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Mol. Biol. Evol.* **33**, 1317–1336. (doi:10.1093/molbev/msw016)
- Usher CL, McCarroll SA. 2015 Complex and multi-allelic copy number variation in human disease. *Brief. Funct. Genomics* **14**, 329–338. (doi:10.1093/bfpg/evl028)
- Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. 2011 The struggle for life of the genome's selfish architects. *Biol. Direct* **6**, 19. (doi:10.1186/1745-6150-6-19)
- Schnable PS *et al.* 2009 The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115. (doi:10.1126/science.1178534)
- Chalopin D, Naville M, Plard F, Galiana D, Volff J-N. 2015 Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* **7**, 567–580. (doi:10.1093/gbe/evv005)
- Gao B, Shen D, Xue S, Chen C, Cui H, Song C. 2016 The contribution of transposable elements to size variations between four teleost genomes. *Mob. DNA* **7**, 4. (doi:10.1186/s13100-016-0059-7)
- Elbarbary RA, Lucas BA, Maquat LE. 2016 Retrotransposons as regulators of gene expression. *Science* **351**, aac7247. (doi:10.1126/science.aac7247)
- Chuong EB, Elde NC, Feschotte C. 2016 Regulatory evolution of innate immunity through co-option of

- endogenous retroviruses. *Science* **351**, 1083–1087. (doi:10.1126/science.aad5497)
18. Casacuberta E, González J. 2013 The impact of transposable elements in environmental adaptation. *Mol. Ecol.* **22**, 1503–1517. (doi:10.1111/mec.12170)
 19. Davidson EH, Britten RJ. 1979 Regulation of gene expression: possible role of repetitive sequences. *Science* **204**, 1052–1059. (doi:10.1126/science.451548)
 20. Muotri AR, Gage FH. 2006 Generation of neuronal variability and complexity. *Nature* **441**, 1087–1093. (doi:10.1038/nature04959)
 21. Lu X, Sachs F, Ramsay L, Jacques P-É, Göke J, Bourque G, Ng H-H. 2014 The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.* **21**, 423–425. (doi:10.1038/nsmb.2799)
 22. Lanciano S, Mirouze M. 2018 Transposable elements: all mobile, all different, some stress responsive, some adaptive? *Curr. Opin. Genet. Dev.* **49**, 106–114. (doi:10.1016/j.gde.2018.04.002)
 23. Sahebi M *et al.* 2018 Contribution of transposable elements in the plant's genome. *Gene* **665**, 155–166. (doi:10.1016/j.gene.2018.04.050)
 24. Van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri JJ. 2016 The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**, 102–105. (doi:10.1038/nature17951)
 25. Gramates LS *et al.* 2017 FlyBase at 25: looking to the future. *Nucleic Acids Res.* **45**, D663–D671. (doi:10.1093/nar/gkw1016)
 26. Lachaise D, Cariou M-L, David JR, Lemeunier F, Tsacas L, Ashburner M. 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup. In *Evolutionary biology*, vol. 22 (eds MK Hecht, B Wallace, GT Prance), pp. 159–225. Boston, MA: Springer US.
 27. David JR, Capy P. 1988 Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet.* **4**, 106–111. (doi:10.1016/0168-9525(88)90098-4)
 28. Arguello JR, Laurent S, Clark AG. 2019 Demographic history of the human commensal *Drosophila melanogaster*. *Genome Biol. Evol.* **11**, 844–854. (doi:10.1093/gbe/evz022)
 29. Mackay TFC, Huang W. 2018 Charting the genotype-phenotype map: lessons from the *Drosophila melanogaster* Genetic Reference Panel. *Wiley Interdiscip. Rev. Dev. Biol.* **7**, e289. (doi:10.1002/wdev.289)
 30. Petrov DA, Fiston-Lavie A-S, Lipatov M, Lenkov K, González J. 2011 Population genomics of transposable elements in *Drosophila melanogaster*. *Mol. Biol. Evol.* **28**, 1633–1644. (doi:10.1093/molbev/msq337)
 31. González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. 2008 High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol.* **6**, e251. (doi:10.1371/journal.pbio.0060251)
 32. Charlesworth B, Langley CH. 1989 The population genetics of *Drosophila* transposable elements. *Annu. Rev. Genet.* **23**, 251–287. (doi:10.1146/annurev.ge.23.120189.001343)
 33. Rech GE *et al.* 2019 Stress response, behavior, and development are shaped by transposable element-induced mutations in *Drosophila*. *PLoS Genet.* **15**, e1007900. (doi:10.1371/journal.pgen.1007900)
 34. Lerat E, Goubert C, Guirao-Rico S, Merenciano M, Dufour A-B, Vieira C, González J. 2019 Population-specific dynamics and selection patterns of transposable element insertions in European natural populations. *Mol. Ecol.* **28**, 1506–1522. (doi:10.1111/mec.14963)
 35. Kim A, Terzian C, Santamaria P, Péllisson A, Purd'homme N, Bucheton A. 1994 Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **91**, 1285–1289. (doi:10.1073/pnas.91.4.1285)
 36. Leblanc P, Desset S, Giorgi F, Taddei AR, Fausto AM, Mazzini M, Dastugue B, Vauray C. 2000 Life cycle of an endogenous retrovirus, ZAM, in *Drosophila melanogaster*. *J. Virol.* **74**, 10 658–10 669. (doi:10.1128/jvi.74.22.10658-10669.2000)
 37. Chung H, Bogwitz MR, McCart C, Andrianopoulos A, French-Constant RH, Batterham P, Daborn PJ. 2007 Cis-regulatory elements in the Accord retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene Cyp6g1. *Genetics* **175**, 1071–1077. (doi:10.1534/genetics.106.066597)
 38. Mateo L, Ullastres A, González J. 2014 A transposable element insertion confers xenobiotic resistance in *Drosophila*. *PLoS Genet.* **10**, e1004560. (doi:10.1371/journal.pgen.1004560)
 39. Merenciano M, Ullastres A, de Cara MAR, Barrón MG, González J. 2016 Multiple independent retroelement insertions in the promoter of a stress response gene have variable molecular and functional effects in *Drosophila*. *PLoS Genet.* **12**, e1006249. (doi:10.1371/journal.pgen.1006249)
 40. Guio L, Barrón MG, González J. 2014 The transposable element Bari-Jeh mediates oxidative stress response in *Drosophila*. *Mol. Ecol.* **23**, 2020–2030. (doi:10.1111/mec.12711)
 41. Zhu C-T, Chang C, Reenan RA, Helfand SL. 2014 Indy gene variation in natural populations confers fitness advantage and life span extension through transposon insertion. *Aging (Albany NY)* **6**, 58–69. (doi:10.18632/aging.100634)
 42. González J, Macpherson JM, Petrov DA. 2009 A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. *Mol. Biol. Evol.* **26**, 1949–1961. (doi:10.1093/molbev/msp107)
 43. Aminetzach YT, Macpherson JM, Petrov DA. 2005 Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. *Science* **309**, 764–767. (doi:10.1126/science.1112699)
 44. Schmidt JM *et al.* 2010 Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. *PLoS Genet.* **6**, e1000998. (doi:10.1371/journal.pgen.1000998)
 45. Lynch VJ, Roth JJ, Takahashi K, Dunn CW, Nonaka DF, Stopper GF, Wagner GP. 2004 Adaptive evolution of HoxA-11 and HoxA-13 at the origin of the uterus in mammals. *Proc. R. Soc. B* **271**, 2201–2207. (doi:10.1098/rspb.2004.2848)
 46. Lynch VJ *et al.* 2015 Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. *Cell Rep.* **10**, 551–561. (doi:10.1016/j.celrep.2014.12.052)
 47. Notwell JH, Chung T, Heavner W, Bejerano G. 2015 A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nat. Commun.* **6**, 6644. (doi:10.1038/ncomms7644)
 48. Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM. 2015 Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* **11**, e1004915. (doi:10.1371/journal.pgen.1004915)
 49. Villanueva-Cañas JL, Horvath V, Aguilera L, González J. 2019 Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in *Drosophila melanogaster*. *Nucleic Acids Res.* **47**, 6842–6857. (doi:10.1093/nar/gkz490)
 50. Horvath V, Merenciano M, González J. 2017 Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends Genet.* **33**, 832–841. (doi:10.1016/j.tig.2017.08.007)
 51. Bourque G *et al.* 2018 Ten things you should know about transposable elements. *Genome Biol.* **19**, 199. (doi:10.1186/s13059-018-1577-z)
 52. Palumbi SR. 2001 Humans as the world's greatest evolutionary force. *Science* **293**, 1786–1790. (doi:10.1126/science.293.5536.1786)
 53. Beaver LM, Hooven LA, Butcher SM, Krishnan N, Sherman KA, Chow ES-Y, Giebultowicz JM. 2010 Circadian clock regulates response to pesticides in *Drosophila* via conserved Pdp1 pathway. *Toxicol. Sci.* **115**, 513–520. (doi:10.1093/toxsci/kfq083)
 54. Seong KM, Coates BS, Sun W, Clark JM, Pittendrigh BR. 2017 Changes in neuronal signaling and cell stress response pathways are associated with a multigenic response of *Drosophila melanogaster* to DDT selection. *Genome Biol. Evol.* **9**, 3356–3372. (doi:10.1093/gbe/evx252)
 55. Denecke S, Fusetto R, Batterham P. 2017 Describing the role of *Drosophila melanogaster* ABC transporters in insecticide biology using CRISPR-Cas9 knockouts. *Insect Biochem. Mol. Biol.* **91**, 1–9. (doi:10.1016/j.ibmb.2017.09.017)
 56. Garud NR, Petrov DA. 2016 Elevated linkage disequilibrium and signatures of soft sweeps are common in *Drosophila melanogaster*. *Genetics* **203**, 863–880. (doi:10.1534/genetics.115.184002)
 57. Kim JH, Moreau JA, Zina JM, Mazgaen L, Yoon KS, Pittendrigh BR, Clark JM. 2018 Identification and interaction of multiple genes resulting in DDT resistance in the 91-R strain of *Drosophila melanogaster* by RNAi approaches. *Pestic. Biochem. Physiol.* **151**, 90–99. (doi:10.1016/j.pestbp.2018.03.003)
 58. Riahi H *et al.* 2019 The histone methyltransferase G9a regulates tolerance to oxidative stress-induced

- energy consumption. *PLoS Biol.* **17**, e2006146. (doi:10.1371/journal.pbio.2006146)
59. Daborn PJ *et al.* 2002 A single p450 allele associated with insecticide resistance in *Drosophila*. *Science* **297**, 2253–2256. (doi:10.1126/science.1074170)
 60. Misra JR, Horner MA, Lam G, Thummel CS. 2011 Transcriptional regulation of xenobiotic detoxification in *Drosophila*. *Genes Dev.* **25**, 1796–1806. (doi:10.1101/gad.17280911)
 61. Mateo L, Rech GE, González J. 2018 Genome-wide patterns of local adaptation in western European *Drosophila melanogaster* natural populations. *Sci. Rep.* **8**, 16143. (doi:10.1038/s41598-018-34267-0)
 62. Mackay TFC *et al.* 2012 The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**, 173–178. (doi:10.1038/nature10811)
 63. Adams MD *et al.* 2000 The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195. (doi:10.1126/science.287.5461.2185)
 64. Martin M. 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10. (doi:10.14806/ej.17.1.200)
 65. Kim D, Langmead B, Salzberg SL. 2015 HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360. (doi:10.1038/nmeth.3317)
 66. Sayols S, Scherzinger D, Klein H. 2016 dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC Bioinf.* **17**, 428. (doi:10.1186/s12859-016-1276-2)
 67. Liao Y, Smyth GK, Shi W. 2014 featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930. (doi:10.1093/bioinformatics/btt656)
 68. Love MI, Huber W, Anders S. 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. (doi:10.1186/s13059-014-0550-8)
 69. Huang DW, Sherman BT, Lempicki RA. 2009 Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57. (doi:10.1038/nprot.2008.211)
 70. Huang DW, Sherman BT, Lempicki RA. 2009 Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13. (doi:10.1093/nar/gkn923)
 71. Szklarczyk D *et al.* 2019 STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613. (doi:10.1093/nar/gky1131)
 72. Bogaerts-Márquez M, Barrón MG, Fiston-Lavier AS, Vendrell-Mir P, Castanera R, Casacuberta JM, González J. 2019 T-lex3: an accurate tool to genotype and estimate population frequencies of transposable elements using the latest short-read whole genome sequencing data. *Bioinformatics* **btz727**. (doi:10.1093/bioinformatics/btz727)
 73. Buenostro JD, Wu B, Chang HY, Greenleaf WJ. 2015 ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21–29. (doi:10.1002/0471142727.mb2129s109)
 74. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
 75. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25. (doi:10.1186/gb-2009-10-3-r25)
 76. Chung D, Kuan PF, Li B, Sanalkumar R, Liang K, Bresnick EH, Dewey C, Keleş S. 2011 Discovering transcription factor binding sites in highly repetitive regions of genomes with multi-read analysis of ChIP-Seq data. *PLoS Comput. Biol.* **7**, e1002111. (doi:10.1371/journal.pcbi.1002111)
 77. Han BW, Wang W, Zamore PD, Weng Z. 2015 piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics* **31**, 593–595. (doi:10.1093/bioinformatics/btu647)
 78. Zhang Y *et al.* 2008 Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137. (doi:10.1186/gb-2008-9-9-r137)
 79. Lerat E, Fablet M, Modolo L, Lopez-Maestre H, Vieira C. 2017 TETools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. *Nucleic Acids Res.* **45**, e17. (doi:10.1093/nar/gkw953)
 80. Modolo L, Lerat E. 2015 UrQt: an efficient software for the unsupervised quality trimming of NGS data. *BMC Bioinf.* **16**, 137. (doi:10.1186/s12859-015-0546-8)
 81. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. (doi:10.1038/nmeth.1923)
 82. Moskalev A, Shaposhnikov M, Plyusnina E, Plyusnin S, Shostal O, Aliper A, Zhavoronkov A. 2014 Exhaustive data mining comparison of the effects of low doses of ionizing radiation, formaldehyde and dioxins. *BMC Genomics* **15**(Suppl 12), S5. (doi:10.1186/1471-2164-15-S12-S5)
 83. Battlay P, Leblanc PB, Green L, Garud NR, Schmidt JM, Fournier-Level A, Robin C. 2018 Structural variants and selective sweep foci contribute to insecticide resistance in the *Drosophila* genetic reference panel. *G3 (Bethesda)* **8**, 3489–3497. (doi:10.1534/g3.118.200619)
 84. Guio L, González J. 2015 The dominance effect of the adaptive transposable element insertion Bari-Jheh depends on the genetic background. *Genome Biol. Evol.* **7**, 1260–1266. (doi:10.1093/gbe/evv071)
 85. Le Manh H, Guio L, Merenciano M, Rovira Q, Barrón MG, González J. 2017 Natural and laboratory mutations in kuzbanian are associated with zinc stress phenotypes in *Drosophila melanogaster*. *Sci. Rep.* **7**, 42663. (doi:10.1038/srep42663)
 86. Huang H, Lu-Bo Y, Haddad GG. 2014 A *Drosophila* ABC transporter regulates lifespan. *PLoS Genet.* **10**, e1004844. (doi:10.1371/journal.pgen.1004844)
 87. Greene JC, Whitworth AJ, Andrews LA, Parker TJ, Pallanck LJ. 2005 Genetic and genomic studies of *Drosophila* parkin mutants implicate oxidative stress and innate immune responses in pathogenesis. *Hum. Mol. Genet.* **14**, 799–811. (doi:10.1093/hmg/ddi074)
 88. Štětina T, Poupardin R, Moos M, Šimek P, Šmilauer P, Košťál V. 2019 Larvae of *Drosophila melanogaster* exhibit transcriptional activation of immune response pathways and antimicrobial peptides during recovery from supercooling stress. *Insect Biochem. Mol. Biol.* **105**, 60–68. (doi:10.1016/j.ibmb.2019.01.006)
 89. Geiger JA, Carvalho L, Campos I, Santos AC, Jacinto A. 2011 Hole-in-one mutant phenotypes link EGFR/ERK signaling to epithelial tissue repair in *Drosophila*. *PLoS ONE* **6**, e28349. (doi:10.1371/journal.pone.0028349)
 90. Davie K, Jacobs J, Atkins M, Potier D, Christiaens V, Halder G, Aerts S. 2015 Discovery of transcription factors and regulatory regions driving *in vivo* tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. *PLoS Genet.* **11**, e1004994. (doi:10.1371/journal.pgen.1004994)
 91. Hannon CE, Blythe SA, Wieschaus EF. 2017 Concentration dependent chromatin states induced by the bicoid morphogen gradient. *eLife* **6**, e28275. (doi:10.7554/eLife.28275)
 92. Bozek M, Cortini R, Storti AE, Unnerstall U, Gaul U, Gompel N. 2019 ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the *Drosophila* blastoderm. *Genome Res.* **29**, 771–783. (doi:10.1101/gr.242362.118)
 93. Mueller B, Mieczkowski J, Kundu S, Wang P, Sadreyev R, Tolstorukov MY, Kingston RE. 2017 Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. *Genes Dev.* **31**, 451–462. (doi:10.1101/gad.293118.116)
 94. Romero-Soriano V, Modolo L, Lopez-Maestre H, Mugat B, Pessia E, Chambeyron S, Vieira C, Garcia Guerreiro MP. 2017 Transposable element misregulation is linked to the divergence between parental piRNA pathways in *Drosophila* hybrids. *Genome Biol. Evol.* **9**, 1450–1470. (doi:10.1093/gbe/evx091)
 95. Silva AX, Jander G, Samaniego H, Ramsey JS, Figueroa CC. 2012 Insecticide resistance mechanisms in the green peach aphid *Myzus persicae* (Hemiptera: Aphididae) I: a transcriptomic survey. *PLoS ONE* **7**, e36366. (doi:10.1371/journal.pone.0036366)
 96. Poupardin R, Riaz MA, Jones CM, Chandor-Proust A, Reynaud S, David J-P. 2012 Do pollutants affect insecticide-driven gene selection in mosquitoes? Experimental evidence from transcriptomics. *Aquat. Toxicol.* **114–115**, 49–57. (doi:10.1016/j.aquatox.2012.02.001)
 97. Zou S, Meadows S, Sharp L, Jan LY, Jan YN. 2000 Genome-wide study of aging and oxidative stress response in *Drosophila melanogaster*. *Proc. Natl*

- Acad. Sci. USA* **97**, 13 726–13 731. (doi:10.1073/pnas.260496697)
98. Zhang J, Marshall KE, Westwood JT, Clark MS, Sinclair BJ. 2011 Divergent transcriptomic responses to repeated and single cold exposures in *Drosophila melanogaster*. *J. Exp. Biol.* **214**, 4021–4029. (doi:10.1242/jeb.059535)
 99. Kuintzle RC, Chow ES, Westby TN, Gvakharia BO, Giebultowicz JM, Hendrix DA. 2017 Circadian deep sequencing reveals stress-response genes that adopt robust rhythmic expression during aging. *Nat. Commun.* **8**, 14529. (doi:10.1038/ncomms14529)
 100. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013 Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218. (doi:10.1038/nmeth.2688)
 101. Li Z, Schulz MH, Look T, Begemann M, Zenke M, Costa IG. 2019 Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45. (doi:10.1186/s13059-019-1642-2)
 102. Rougvie AE, Lis JT. 1988 The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. *Cell* **54**, 795–804. (doi:10.1016/S0092-8674(88)91087-2)
 103. Todd CD, Deniz Ö, Taylor D, Branco MR. 2019 Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. *eLife* **8**, e44344. (doi:10.7554/eLife.44344)
 104. Lin E-H, Kao Y-R, Lin C-A, Kuo T-Y, Yang S-P, Hsu C-F, Chou T-Y, Ho C-C, Wu C-W. 2016 Hedgehog pathway maintains cell survival under stress conditions, and drives drug resistance in lung adenocarcinoma. *Oncotarget* **7**, 24 179–24 193. (doi:10.18632/oncotarget.8253)
 105. Aissouni Y, Zapart G, Iovanna JL, Dikic I, Soubeyran P. 2005 CIN85 regulates the ability of MEKK4 to activate the p38 MAP kinase pathway. *Biochem. Biophys. Res. Commun.* **338**, 808–814. (doi:10.1016/j.bbrc.2005.10.032)
 106. Bettinger BT, Amberg DC. 2007 The MEK kinases MEKK4/Ssk2p facilitate complexity in the stress signaling responses of diverse systems. *J. Cell. Biochem.* **101**, 34–43. (doi:10.1002/jcb.21289)