insects

MDPI

Article

Transposable Elements Contribute to the Regulation of Long Noncoding RNAs in *Drosophila melanogaster*

Yuli Gan [1,2], Lingyan Wang [1], Guoxian Liu [2], Xiruo Guo [1], Yiming Zhou [3], Kexin Chang [4], Zhonghui Zhang [3], Fang Yan [4], Qi Liu [2,*] and Bing Chen [1,5,*]

[1] College of Life Science, Hebei University, Baoding 071002, China; ganyuli123@163.com (Y.G.); wly1372@163.com (L.W.); guoxiruo@foxmail.com (X.G.)
[2] Rice Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China; guoxianliu0205@163.com
[3] Guangdong Provincial Key Laboratory of Biotechnology for Plant Development, School of Life Science, South China Normal University, Guangzhou 510631, China; 15137361521@163.com (Y.Z.); zhzhang@m.scnu.edu.cn (Z.Z.)
[4] Key Laboratory of Herbage and Endemic Crop Biotechnology, Ministry of Education, School of Life Science, Inner Mongolia University, Hohhot 010021, China; changkexincas@163.com (K.C.); yanfang@imu.edu.cn (F.Y.)
[5] Hebei Basic Science Center for Biotic Interaction, Hebei University, Baoding 071002, China
* Correspondence: rnainfor@gmail.com (Q.L.); chenbing@hbu.edu.cn (B.C.)

Simple Summary: This study focused on lncRNAs originating from TEs (TE-lncRNAs) in *Drosophila melanogaster* by integrating multi-omics data. We identified 2119 TE-lncRNAs (40.4% of all lncRNAs) using 271 RNA-seq, of which the LTR/Gypsy family was the most common transposon. In addition, transposons preferred certain genic regions. TE-lncRNAs had longer lengths, a lower conservation, and a specific expression. Multi-omics analysis showed positive correlations between transposon insertions and chromatin openness; some TE-lncRNAs provided transcription factor binding sites, rewired regulatory networks, and provided candidate small open reading frames through TE insertions. Thus, TEs contribute to lncRNAs, promoting transcriptional, post-transcriptional, and epigenetic regulation.

Abstract: Background: Transposable elements (TEs) and noncoding sequences are major components of the genome, yet their functional contributions to long noncoding RNAs (lncRNAs) are not well understood. Although many lncRNAs originating from TEs (TE-lncRNAs) have been identified across various organisms, their characteristics and regulatory roles, particularly in insects, remain largely unexplored. This study integrated multi-omics data to investigate TE-lncRNAs in *D. melanogaster*, focusing on the influence of transposons across different omics levels. **Results**: We identified 16,118 transposons overlapping with lncRNA sequences that constitute 2119 TE-lncRNAs (40.4% of all lncRNAs) using 256 public RNA-seq samples and 15 lncRNA-seq samples of *Drosophila* S2 cells treated with heavy metals. Of these, 67.2% of TE-lncRNAs contain more than one TE. The LTR/Gypsy family was the most common transposon insertion. Transposons preferred to insert into promoters, transcription starting sites, and intronic regions, especially in chromosome ends. Compared with lncRNAs, TE-lncRNAs showed longer lengths, a lower conservation, and lower levels but a higher specificity of expression. Multi-omics data analysis revealed positive correlations between transposon insertions and chromatin openness at the pre-transcriptional level. Notably, a total of 516 TE-lncRNAs provided transcriptional factor binding sites through transposon insertions. The regulatory network of a key transcription factor was rewired by transposons, potentially recruiting other transcription factors to exert regulatory functions under heavy metal stress. Additionally, 99 TE-lncRNAs were associated with m^6A methylation modification sites, and 115 TE-lncRNAs potentially provided candidate small open reading frames through transposon insertions. **Conclusions**: Our data analysis demonstrated that TEs contribute to the regulation of lncRNAs. TEs not only promote the transcriptional regulation of lncRNAs, but also facilitate their post-transcriptional and epigenetic regulation.

check for updates

Citation: Gan, Y.; Wang, L.; Liu, G.; Guo, X.; Zhou, Y.; Chang, K.; Zhang, Z.; Yan, F.; Liu, Q.; Chen, B. Transposable Elements Contribute to the Regulation of Long Noncoding RNAs in *Drosophila melanogaster*. *Insects* **2024**, *15*, 950. https://doi.org/10.3390/insects15120950

Academic Editor: David S. Haymer

Received: 12 October 2024
Revised: 28 November 2024
Accepted: 29 November 2024
Published: 30 November 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Keywords: long noncoding RNA; transposable element; *Drosophila*; TE-lncRNA; heavy metal

1. Introduction

Approximately 98% of the human genome is transcribed into noncoding RNAs [1], among which long noncoding RNAs (lncRNAs) constitute a significant portion of noncoding RNA molecules. LncRNAs are characterized by a low-coding potential and a length exceeding 200 nucleotides; they are broadly categorized into intronic lncRNA, intergenic lncRNA, and antisense lncRNA [2]. LncRNAs play diverse roles at different molecular levels, including transcriptional interference, chromatin remodeling and nucleosome modification, alternative splicing regulation, endogenous siRNA biogenesis, control of protein activity, etc. [3]. At the post-transcriptional level, lncRNAs contain m^6A modification sites related to disease [4,5] and small open reading frames (sORFs) that are often overlooked by traditional gene annotation methods. These sORFs can encode short peptides and some lncRNAs are even considered "dual-functional" RNAs, exerting regulatory roles in their RNA form and functional roles through peptide coding [6,7]. Several lncRNA-encoded short peptides have been discovered, such as SPAR [8], Minion/myomixer [9], HOXB-AS3 [10], NOBODY [11], and APPLE [12] in humans; MLN [13], and DWORF [14] in mice; Toddler [15] in zebrafish; and Tarsal-less/tal [16], Scl [17], and Pgc [18] in fruit flies, demonstrating significant regulatory functions. Recent analysis have revealed that lncRNAs play a significant role through the regulation of transposon activities [19]. Despite their diverse functions, the role of lncRNAs remains underexplored.

Transposable elements (TEs) are DNA sequences that are capable of being excised from their original location, replicated or not, and inserted elsewhere, thereby affecting adjacent genes. TEs are broadly classified into two categories: class I transposons, also known as retrotransposons, which employ a "copy and paste" mechanism, such as LTR, LINE, and SINE; and class II transposons, or DNA transposons, which utilize a "cut and paste" mechanism, of which there are two types of subclasses, subclass 1, e.g., TIR, and subclass 2, e.g., Helitrons, which is a recently discovered class of DNA transposons that transpose via a rolling-circle mechanism [20–24]. TEs constitute a 22% portion of genomes in *Drosophila melanogaster* [25] and 66% in humans [26]. Although TEs are often considered as "junk DNA", many of them also play various crucial roles in gene expression regulation and phenotypic evolution [26–28]. For example, Gypsy insertion in ducks caused bodyweight enlargement or white plumage phenotype formation, while the insertion of Alu elements led to the loss of their tail in humans [29,30]. Studies have highlighted the role of transposons in promoting the nuclear enrichment of lncRNAs, indicating their essential contribution to lncRNA functionality [31]. However, the intricate relationship between transposons and lncRNAs remains elusive.

Many lncRNAs overlap with TE sequences or originate from TEs, and are regarded as TE-lncRNAs. In recent years, numerous TE-lncRNAs have been identified and characterized across various species [32–34]. These TE-lncRNAs are widespread in both animals and plants [35–37]. The emergence and evolution of lncRNAs are closely intertwined with TEs [33,38]. However, the nature and regulatory functionality of TE-lncRNAs, especially in insects, largely remain elusive. Our understanding of the characteristics and features of TE-lncRNAs, as well as the impact of transposons on lncRNAs at different regulatory layers, is limited. A comprehensive investigation spanning transcriptional and post-transcriptional layers is imperative to elucidate the intricate mechanisms by which transposons shape the biogenesis and functional roles of lncRNAs.

Here, we hypothesized that TEs in lncRNAs contribute to biogenesis and regulatory functionality of lncRNAs. *D. melanogaster* serves as a model organism within the insect kingdom. This study comprehensively identified and analyzed the characteristics of TE-lncRNAs in *D. melanogaster* and explored TE-lncRNAs at the epigenetic, transcriptional, and post-transcriptional levels, investigating the contribution of transposons in the biogenesis,

characteristics, and functional roles of lncRNAs across these dimensions. Furthermore, the functional roles of TE-lncRNAs under stress conditions, such as heavy metal exposure, were explored. Thus, this research discovered the significance of transposons and noncoding sequences, which have previously dismissed as "junk sequences" or "dark matter," and established the intricate connection between transposons and noncoding sequences at multiple regulatory levels.

2. Materials and Methods

2.1. Data

A total of 256 RNA-seq samples were collected from the Berkeley *Drosophila* Genome Project (BDGP), with the NCBI accession ID SRA009364 [39,40]. A further 15 lncRNA-seq datasets from heavy metal-treated *Drosophila* S2 cells were sequenced by our lab [41]. Low-quality samples were filtered out, and all datasets consisted of paired-end reads. Detailed sample information is provided in Supplementary Table S2. Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) data from 28 *D. melanogaster* samples were downloaded from NCBI, with the corresponding sample information being listed in Supplementary Table S3. Methylated RNA Immunoprecipitation Sequencing (MeRIP-seq) data from 8 *D. melanogaster* samples were also obtained from NCBI, with details being provided in Supplementary Table S4. Furthermore, ribosome profiling sequencing (Ribo-seq) data from 72 *D. melanogaster* samples were downloaded from NCBI, and the associated sample information can be found in Supplementary Table S5. Lastly, *D. melanogaster* proteomics data were acquired from ProteomeXchange under the project ID PXD021022.

2.2. LncRNA Identification Pipeline

The datasets were downloaded and converted to FASTQ format using sratoolkit. The reads were then cleaned with fastp [42] and mapped to the reference genome (*D. melanogaster*, Release 6 plus ISO1 MT, NCBI) using HISAT2 [43]. The mapped reads were assembled and merged with StringTie [44] to generate a merged GTF file, which was then compared to the reference annotation using GffCompare [45]. StringTie compares the assembled transcripts against the reference gene annotations (usually in a GTF file). It assesses the transcript's exon boundaries, splice sites, genomic coordinates, and strand information to identify different types of transcripts. If there is no continuous splice junction information between exons, StringTie is more likely to treat these exons as belonging to different transcripts, avoiding misassembly. The transcripts were filtered through the following three steps: (i) retaining only those annotated with class codes u, x, i, j, or o, where "i": fully contained within a reference intron, "j": multi-exon with at least one junction match, "o": other same-strand overlap with reference exons, "x": exonic overlap on the opposite strand, "u": unknown and intergenic; (ii) ensuring transcript length was at least 200 bp; and (iii) including only transcripts with at least one exon. For coding potential filtering, we employed a three-step process as follows: first, the coding potential was predicted using CNCI and CPC2 [46], and the results were intersected to identify noncoding sequences; second, these sequences were mapped to *D. melanogaster* protein sequences using BLAST [47], filtering out protein-coding sequences to identify candidate lncRNAs; third, the expression levels of candidate lncRNAs were calculated using FeatureCounts [48] and R scripts, retaining only those expressed in at least one sample. For each lncRNA, we analyzed its expression across all 271 samples. If the total sum of FPKM values for a given lncRNA across these samples was greater than zero, we retained that lncRNA, considering it to be expressed in at least one of the samples.

2.3. TE Annotation

TEs were annotated for genome (*D. melanogaster* reference genome, Release 6 plus ISO1 MT, NCBI), CDS, and lncRNA sequences using RepeatMasker [49,50], and RepBase 20181026 [51,52] as the library. The following parameters were used: -libdir Libraries

-species "*Drosophila melanogaster*" -e ncbi -nolow -pa 20. The low-complexity and simple repeat sequences were removed; only the TE sequences were retained. Then, makeTEGTF.pl was used to convert the output file of RepeatMasker into gtf format; the following parameters were used: -c 5 -s 6 -e 7 -o 9 -t 10 -f 11 -S. When the lncRNA sequence overlapped with the TE sequence, it was regarded as a TE-lncRNA.

2.4. General Bioinformatical Analysis

The ChIPseeker [53] R package was used to statistically analyze and visualize the positional distribution of TE insertions, coding genes, TE-lncRNAs, and Non-TE-lncRNAs. Conservation was assessed using PhyloP and PhastCons scores, which were obtained from the "dm6.phyloP124way" and "dm6.phastCons124way" datasets. The score values refer to evolutionary conservation scores based on a multiple sequence alignment of the fruit fly genome (*D. melanogaster*, version dm6) with a set of 124 species. These species are primarily chosen for their evolutionary proximity to *D. melanogaster*, providing insight into both functional conservation and evolutionary pressures. The 124 species typically include *Drosophila* species, as well as close relatives such as *Drosophila simulans*, *Drosophila yakuba*, *Drosophila pseudoobscura*, etc. All the 124 species are shown in the txt file from UCSC (http://hgdownload.soe.ucsc.edu/goldenPath/dm6/phyloP124way/assemblyInformation.txt, accessed on 29 November 2018). The levels and specificity of gene expression were analyzed using TBtools [54]. The TAU (tissue-specific gene expression) index indicates how specific or broadly expressed a gene or transcript is within studied tissues [55]. TBtools uses the expression matrix of genes and lncRNAs as input data to compute the TAU index for each gene and lncRNA. The TAU index has a range of 0–1, whereby a value closer to 1 indicate more expression specificity. The expression specificity calculation formula is $\tau = \frac{\sum_{i=1}^{N}(1-x_i)}{N-1}$. We also constructed a co-expression network using the WGCNA [56] R package. Finally, coding genes were selected as input data for Gene Ontology (GO) functional enrichment analysis, which was performed using the clusterProfiler [57] R package.

2.5. ATAC-Seq Analysis

ATAC-seq data were downloaded and extracted using sratoolkit to obtain data in FASTQ format. Quality control was performed using fastp; then, the quality-controlled data were aligned to the reference genome (*D. melanogaster*, Release 6 plus ISO1 MT, NCBI) using STAR. The alignment results (BAM files) were converted to bigWig (BW) files using deepTools [58], and the data were visualized using IGV. Chromatin-accessible regions were identified by calling peaks with MACS2 [59] based on the alignment results, and BEDTools [60] was used to intersect lncRNA locations with ATAC-seq peak regions to determine the chromatin accessibility of lncRNAs.

2.6. MeRIP-Seq Analysis

MeRIP-seq data were downloaded and extracted using sratoolkit to obtain data in FASTQ format. Quality control was performed using fastp; then; the quality-controlled data were aligned to the fruit fly reference genome (*D. melanogaster*, Release 6 plus ISO1 MT, NCBI) using STAR. Based on the alignment results (BAM files), the R packages exomePeak2 [61] and RNAmod [62] were used to identify m^6A modification sites. Then, BEDTools was used to intersect the locations of lncRNAs with m^6A peak regions to determine the m^6A methylation modification sites on lncRNAs.

2.7. Ribo-Seq Analysis

First, candidate ORFs were annotated within lncRNAs; then, these annotations with Ribo-seq data were integrated to identify translated ORFs within lncRNAs. The Ribo-seq data were downloaded from NCBI using sratoolkit, which was then extracted into FASTQ format, and preprocessed by removing adapters with cutadapt, filtering with the fastq_quality_filter, and converting to FASTA format. Next, Bowtie [63] and SAMtools [64] were used to remove tRNA and rRNA sequences; then, the cleaned data were aligned

with STAR. Finally, translated lncRNAs were identified and verified using tools such as Ribo-TISH [65], riboWaltz [66], ribORF [67], and RiboToolkit [68] and were quantified using FeatureCounts. When we ran the pipeline of the Ribo-seq, we set the length at 26–34 RPF.

2.8. Proteomic Analysis

Using the annotated candidate ORFs on lncRNAs as a database, a database search analysis was performed on the downloaded raw proteomic mass spectrometry files using MaxQuant [69]. Based on the search results, the peptide information encoded by ORFs and secondary mass spectra of the peptides was visualized using PDV [70].

3. Results

3.1. Identification of lncRNAs and TEs in D. melanogaster

We compiled a total of 271 transcriptome datasets, distributed across various tempo-spatial scales and experimental conditions in *D. melanogaster* (Table S1). These datasets comprised 256 paired-end RNA-seq datasets from the BDGP (Berkeley *Drosophila* Genome Project) database, and 15 paired-end lncRNA-seq datasets generated in our laboratory from *Drosophila* S2 cells exposed to four distinct heavy metal treatments (Cd, Cu, Pb, and Zn) [41].

We established a robust pipeline tailored to the identification of TE-lncRNAs in *D. melanogaster*, delineated into three key steps (Figure 1A). First, we rigorously cleaned the 271 RNA-seq datasets using fastp [42], and aligned them to the reference genome with HISAT2 [43]. Next, transcripts were assembled and compared using StringTie [44] in combination with GffCompare [45].

We then applied stringent criteria, retaining only transcripts longer than 200 bp with class codes u, x, i, j, or o, and that contained at least one exon. Last, we assessed protein-coding potential using CPC2 [46] and CNCI [71], filtering out any transcripts overlapping with *D. melanogaster* protein sequences. This process identified 14,146 putative lncRNAs in *D. melanogaster*. Further refinement revealed 5246 lncRNAs expressed in at least one sample (Figure 1B).

Additionally, we identified 27,642 mRNAs and 39,051 TEs in the fly genome using RepeatMasker (Figure 1B). Notably, our analysis revealed that a substantial proportion, constituting 40.39% (2119) of lncRNAs, were TE-lncRNAs, characterized by their overlap with TE sequences (Figure 1C). In contrast, only 20.69% of mRNAs were classified as TE-mRNAs which are mRNAs overlapping with transposons (Figure 1C).

3.2. Characterization of TE-lncRNAs

We performed a comprehensive comparison of TE-lncRNAs, Non-TE-lncRNAs, and coding genes, focusing on their length, location, conservation, and expression profiles to uncover the distinct characteristics of TE-lncRNAs. Our analysis revealed that TE-lncRNAs have an average length of 12,377.92 base pairs (bp), which is significantly longer than that of Non-TE-lncRNAs (3580.13 bp on average; *t*-test: $p < 0.001$) and coding genes (6959.49 bp on average; *t*-test: $p < 0.01$) (Figure 2A).

Next, we assessed the conservation levels of coding genes, Non-TE-lncRNAs, TE-lncRNAs, and TE insertion sequences using Phastcons and Phylop conservation scores. Notably, coding genes exhibit the highest conservation, followed by Non-TE-lncRNAs and TE-lncRNAs (Figure 2B). Both TE-lncRNAs and TE insertion sequences have a significant proportion of sequences with conservation scores of zero, indicating lower conservation levels (Figures S1 and S2).

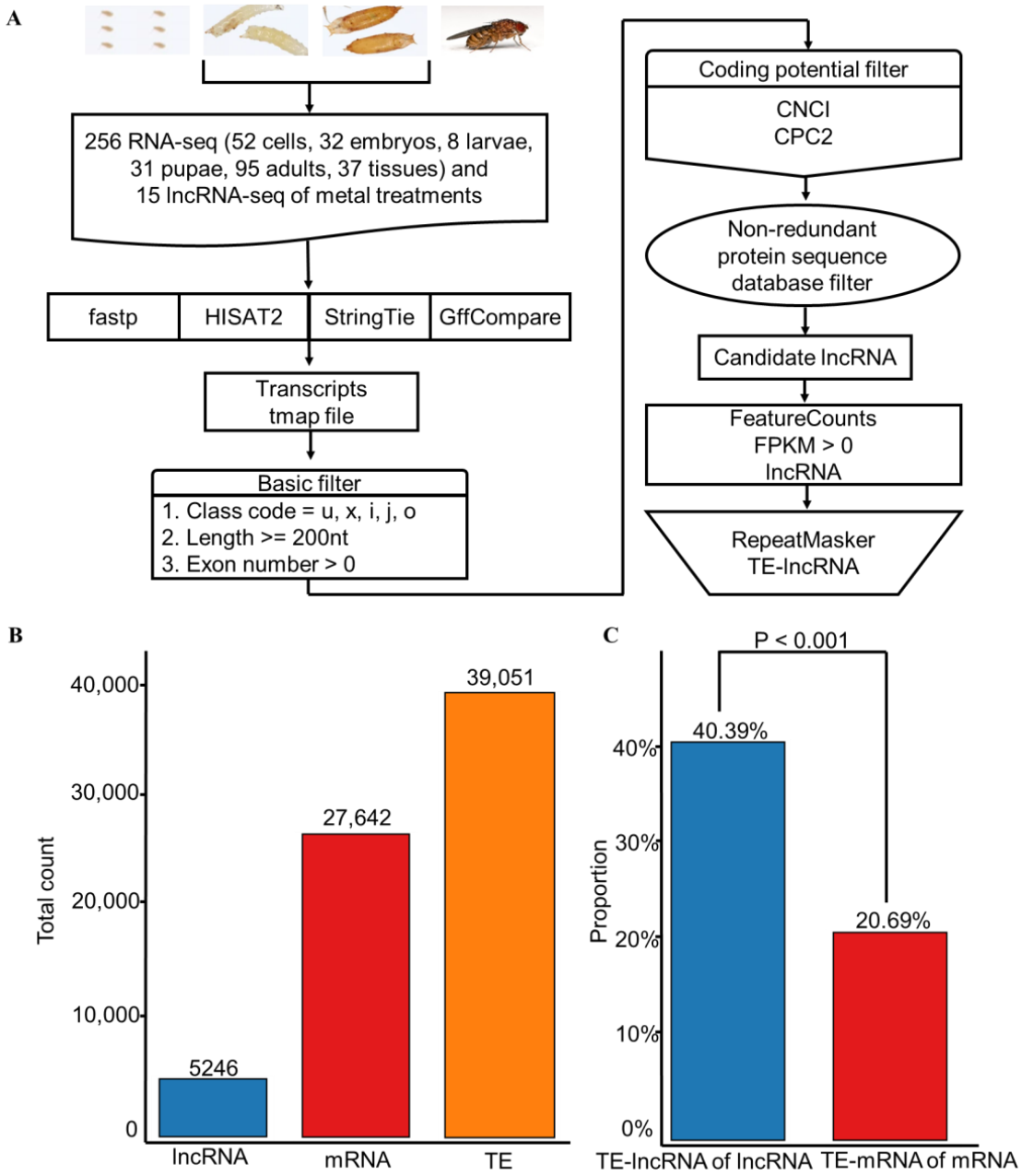

Figure 1. The pipeline of identifying TE-lncRNAs from RNA-seq data. (**A**) The pipeline of identifying TE-lncRNAs. (**B**) A total of 27,642 mRNAs, 5246 lncRNAs, and 39,051 TEs were identified in *D. melanogaster*. (**C**) The proportion of TE-lncRNAs in lncRNAs and TE-mRNAs in mRNAs.

Additionally, we analyzed the expression levels and specificity of coding genes, Non-TE-lncRNAs, and TE-lncRNAs across the 15 lncRNA-seq datasets, which includes samples treated with various heavy metals. This allowed us to examine their expression patterns under different conditions. Principal component analysis (PCA) revealed a distinct expression profile associated with TE-lncRNAs in response to heavy metal exposure in comparison to Non-TE-lncRNAs and coding genes (Figure S3). Expression heatmaps showed that most TE-lncRNAs were highly induced by at least one heavy metal (Figure S4). Genes with element-specific expression likely play key roles in cellular responses to heavy metal exposure. The results indicated that coding genes had the highest expression levels, while TE-lncRNAs had the lowest expression levels. TE-lncRNAs presented higher expression specificity than coding genes (Figure 2C,D). However, the difference in specificity between TE-lncRNAs and Non-TE-lncRNAs was not substantial.

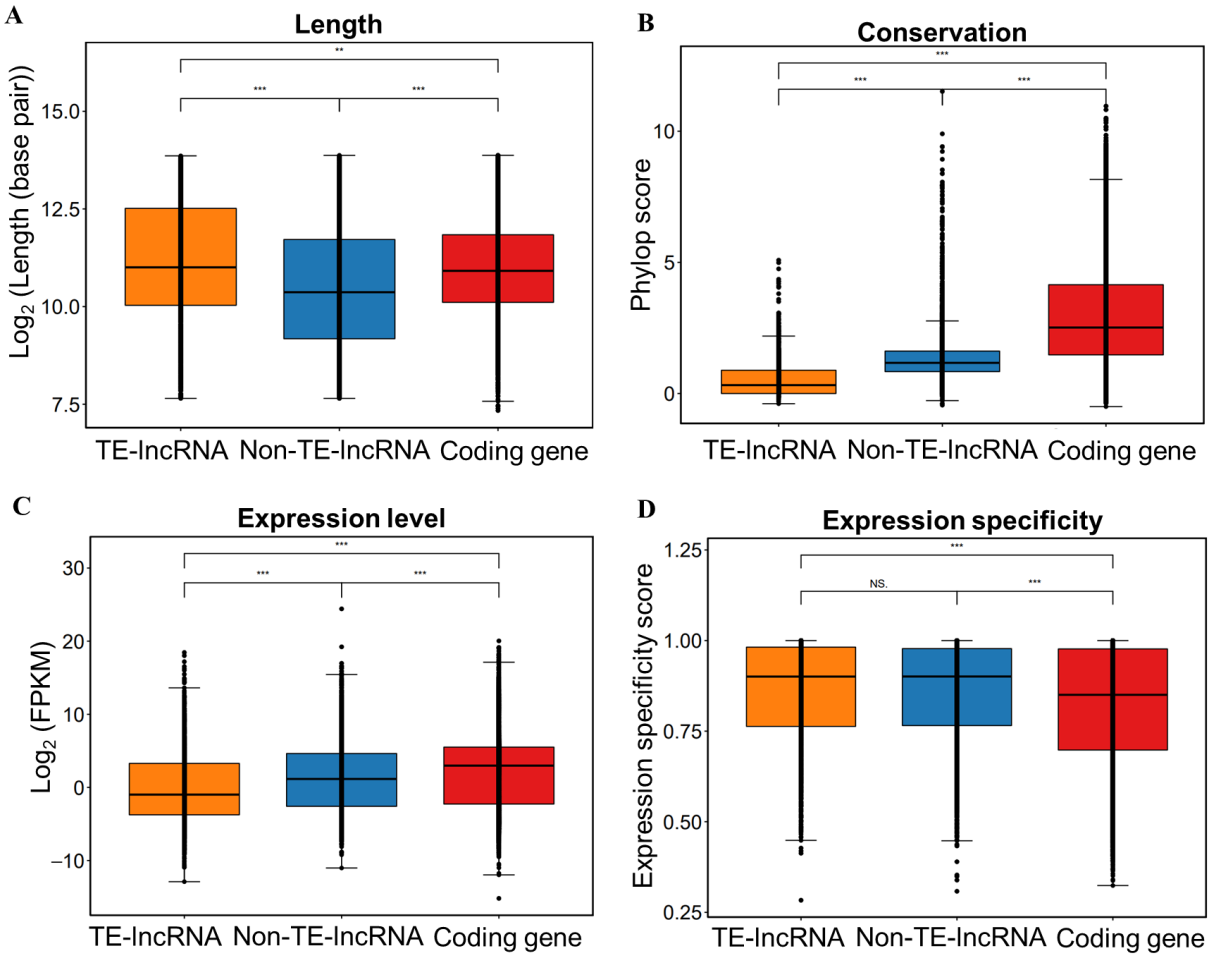

Figure 2. The characteristics of the lncRNAs derived from TEs. (**A**) The length of TE-lncRNAs, Non-TE-lncRNAs, and coding genes. (**B**) Phylop scores among TE-lncRNA, Non-TE-lncRNA, and coding genes sequence. (**C**) The expression levels among TE-lncRNAs, Non-TE-lncRNAs, and coding genes. (**D**) The expression specificity of TE-lncRNAs, Non-TE-lncRNAs, and coding genes. Statistical significance: **: $p < 0.01$; ***: $p < 0.001$; NS.: $p \geq 0.05$.

3.3. Characteristics of TE Insertions in lncRNAs

We conducted a comprehensive analysis of the impact of TEs on lncRNAs in *D. melanogaster*. Our findings revealed 16,118 TEs overlapping with lncRNA sequences, averaging 7.6 TEs per lncRNA, compared to just 0.17 TEs per coding gene (Figure 3A). A significant proportion of TE-lncRNAs harbored multiple transposon sequences, with 67.2% containing more than one TE. Some TE-lncRNAs hosted up to 20 or more TEs (Figure 3B). The percentage of TEs inserted in lncRNA is 41.27% (16,118/39,051), while only 5.30% (2071/39,051) are in CDS (Fisher's exact test: $p < 0.001$). These results indicate that noncoding regions have abundant transposon insertions compared to coding regions.

Class I TEs constitute the majority of TEs associated with lncRNAs (Figure 3A). The Gypsy family has a proportion of 37.32% in the genome, 40.36% in lncRNA, and only 3.04% in CDS (Fisher's exact test: $p < 0.001$). However, the Pao family has a proportion of 12.26% in the genome, 10.71% in lncRNA, and highly 90.39% in CDS (Fisher's exact test: $p < 0.001$). Notably, the LTR/Gypsy family within Class I is the most prevalent, comprising 6505 Gypsy elements and accounting for 39% of the total TEs from lncRNAs (Figures 3C and S5A,B). Other significant contributors include LINE/I-Lockey, RC/Helitron, and LTR/Pao.

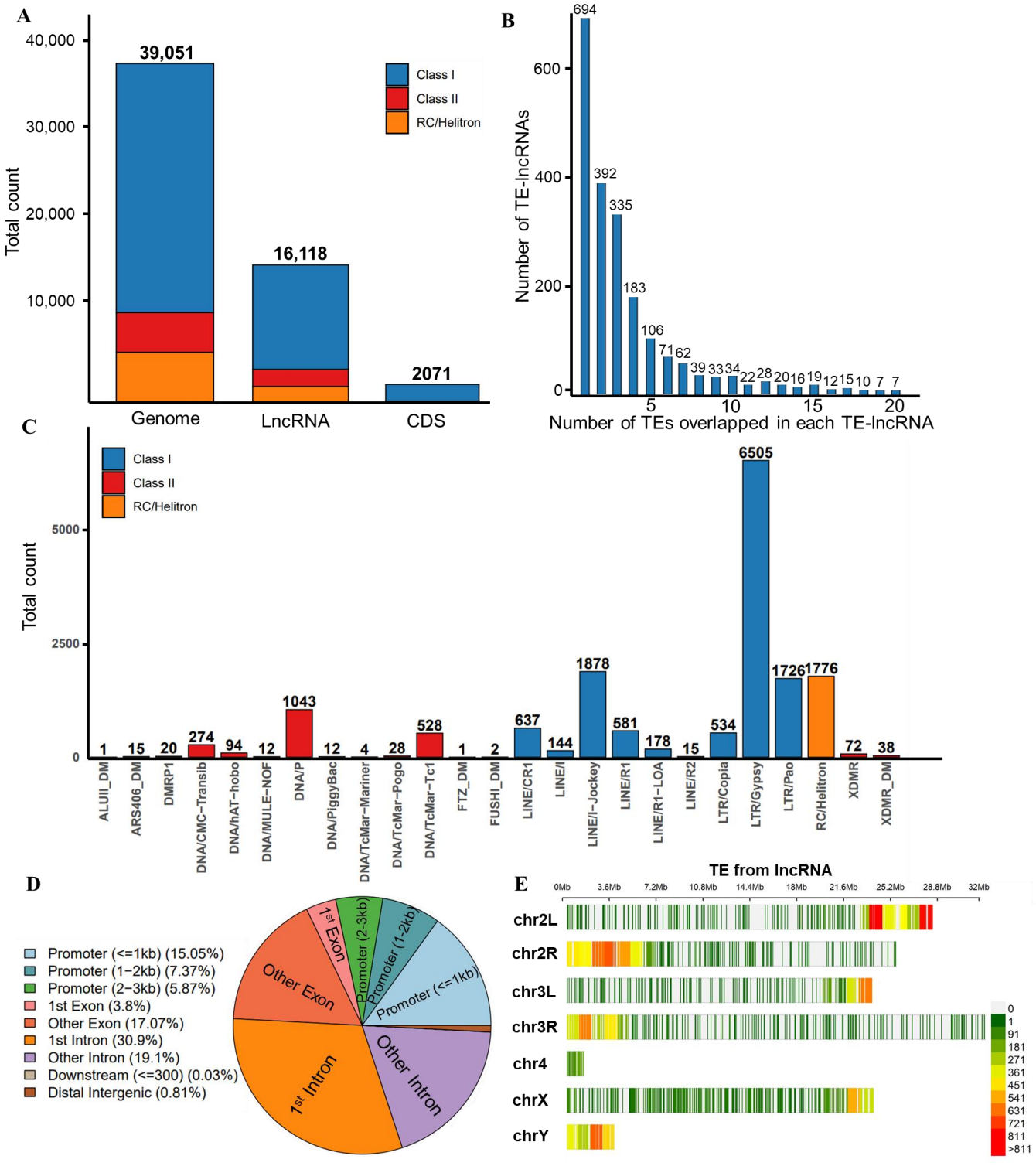

Figure 3. The contribution of TEs to lncRNAs. (**A**) The quantities of various types of TEs that overlap with genome, lncRNA, and CDS sequences. (**B**) Number of TEs overlapping with each TE-lncRNA. (**C**) The quantities of various types of TEs that overlap with lncRNAs. (**D**) Positional preference of TE insertions in lncRNAs. (**E**) Hotspots of TE insertion on chromosomes.

Our investigation into the length distribution of transposon sequences that overlapped with lncRNAs revealed that they predominantly fell within the range of 0–15,000 bp. Additionally, the coverage rate of transposon insertions within lncRNAs was primarily concentrated within the 0–25% and 75–100% intervals (Figure S7).

We analyzed the spatial distribution of transposon insertions within lncRNAs. The results showed that transposons preferred to insert into the intronic and promoter regions (Figures 3D and S1E). Furthermore, our investigation into the contribution of transposons that overlap with lncRNAs to transcription start sites (TSSs) revealed a predominant involvement of TE-lncRNAs, indicating that transposons may contribute to the formation of essential gene features (Figures S5C,D and S6). Intriguingly, our analysis of transposon insertion hotspots along chromosomes revealed a bias towards chromosome ends in *D. melanogaster* (Figures 3E and S5F).

3.4. Epigenetic Regulation Involving TE-lncRNAs

Here, we utilized ATAC-seq data to assess chromatin accessibility across 28 datasets, and identified a total of 372,119 peaks indicative of open chromatin states within lncRNA regions by merging these datasets. We hypothesized that transposon insertions are more prevalent in regions with open chromatin, which facilitates insertion and transcription factor recruitment, while closed and condensed chromatin limits these processes. Additionally, an open chromatin configuration is conducive to the recruitment of transcription factors, thereby initiating the transcriptional process. Our analysis revealed a significant difference in the number of open chromatin peaks between TE-lncRNAs and Non-TE-lncRNAs in *D. melanogaster*. Specifically, 1375 TE-lncRNAs contained 240,354 ATAC peaks, whereas 2774 Non-TE-lncRNAs had only 131,765 ATAC peaks. Remarkably, TE-lncRNAs had twice as many open chromatin regions as Non-TE-lncRNAs. Statistical analysis showed that TE-lncRNAs had significantly more ATAC peaks than Non-TE-lncRNAs (Figure 4A). Additionally, TE-lncRNAs with ATAC peaks exhibited higher expression levels compared to those lacking ATAC peaks (Figure 4B).

We further assessed the transposon coverage length, transposon count, ATAC peak count, transcription factor binding sites (TFBSs), and transposon coverage rate for each lncRNA (Table S6). Correlation analyses among these variables revealed the following notable associations: a positive correlation between transposon quantity and coverage length (Figure S8H), as well as a positive correlation between transposon presence and coverage with the quantity of ATAC peaks in lncRNAs (Figures 4C and S8I). This indicates that a higher transposon content in lncRNAs is associated with an increased ATAC signal intensity. Moreover, we found a positive correlation between the number of ATAC peaks and TFBSs (Figure S8G).

Our examination of transposon and open chromatin state distribution in *Drosophila* lncRNA regions showed striking similarities in the frequency density of ATAC peaks and transposon distribution areas (Figure 4D). Examples from two lncRNAs, *MSTRG.6363.1* and *MSTRG.6368.8*, illustrate the contribution of TEs to ATAC peaks (Figure 4E). These findings indicate that transposon insertion is closely associated with chromatin accessibility, with transposons favoring regions of open chromatin.

3.5. Transcriptional Regulation and Co-Expression Analysis for TE-lncRNAs

To explore the functional roles of TE-lncRNAs at the transcriptional level, we performed a differential expression analysis to identify the TE-lncRNAs, transcription factors, and protein-coding genes involved in responses to heavy metal exposure (Figure 5A), and scanned the TFBSs for lncRNAs; the most enriched TFBSs are listed in Supplementary Table S7. We identified 3580 coding genes, 1161 TE-lncRNAs, and 49 transcription factors that were significantly induced. The number of TE-lncRNAs either upregulated or downregulated by exposure to the four heavy metals varied greatly compared to coding genes (Figure 5A).

We assessed the abundance of TFBSs within the promoter regions of lncRNAs by categorizing TE-lncRNAs into two groups: those with transposons in their promoters (TE-pro-lncRNAs) and those without (Non-TE-pro-lncRNAs). The results showed that TE-lncRNAs with transposon promoters had a higher density of TFBSs compared to those without transposon promoters and Non-TE-lncRNAs (Figure 5B). Additionally, we iden-

tified 516 distinct TE-lncRNAs where transposons contributed TFBSs within the lncRNA sequences (Figure S9A). Among these, 38 TE-lncRNAs exhibited heightened expression specifically under heavy metal treatment in S2 cells, indicating some degree of heavy metal specificity (Figure S9B). The main contributors to these TFBS-containing TE-lncRNAs were the LTR/Gypsy, LTR/Pao, and LINE/I-Jockey transposon families (Figure S9B).

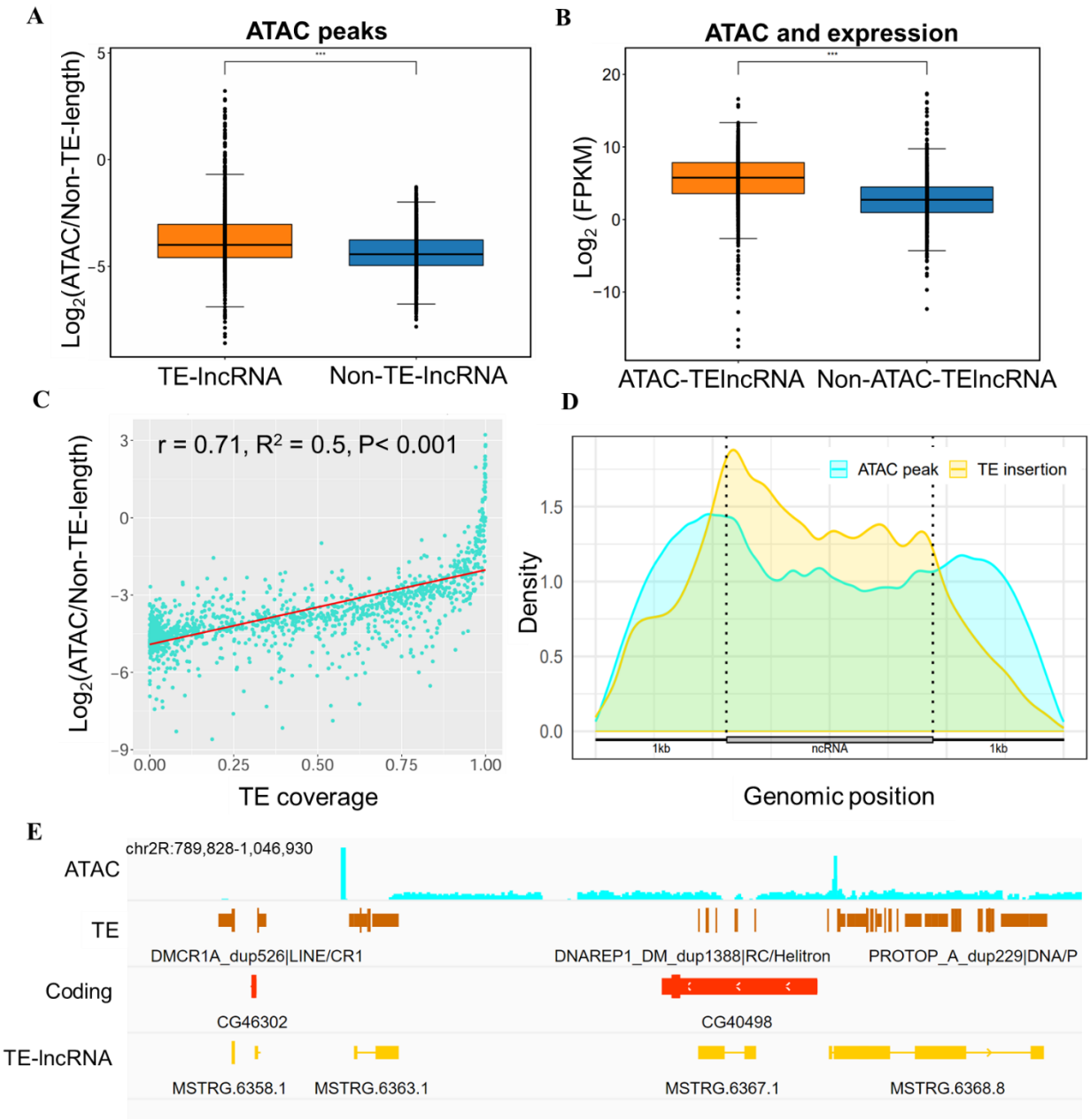

Figure 4. Impact of TE-lncRNAs on epigenetic levels in *Drosophila*. (**A**) ATAC peaks in lncRNAs including TE-lncRNA and Non-TE-lncRNA. (**B**) Expression of ATAC-TE-lncRNA and Non-ATAC-TE-lncRNA. (**C**) The correlation between TE coverage and ATAC peaks. (**D**) The distribution of TE and ATAC peaks in lncRNA. (**E**) IGV view of the ATAC peaks in TE-lncRNA. Statistical significance: ***: $p < 0.001$.

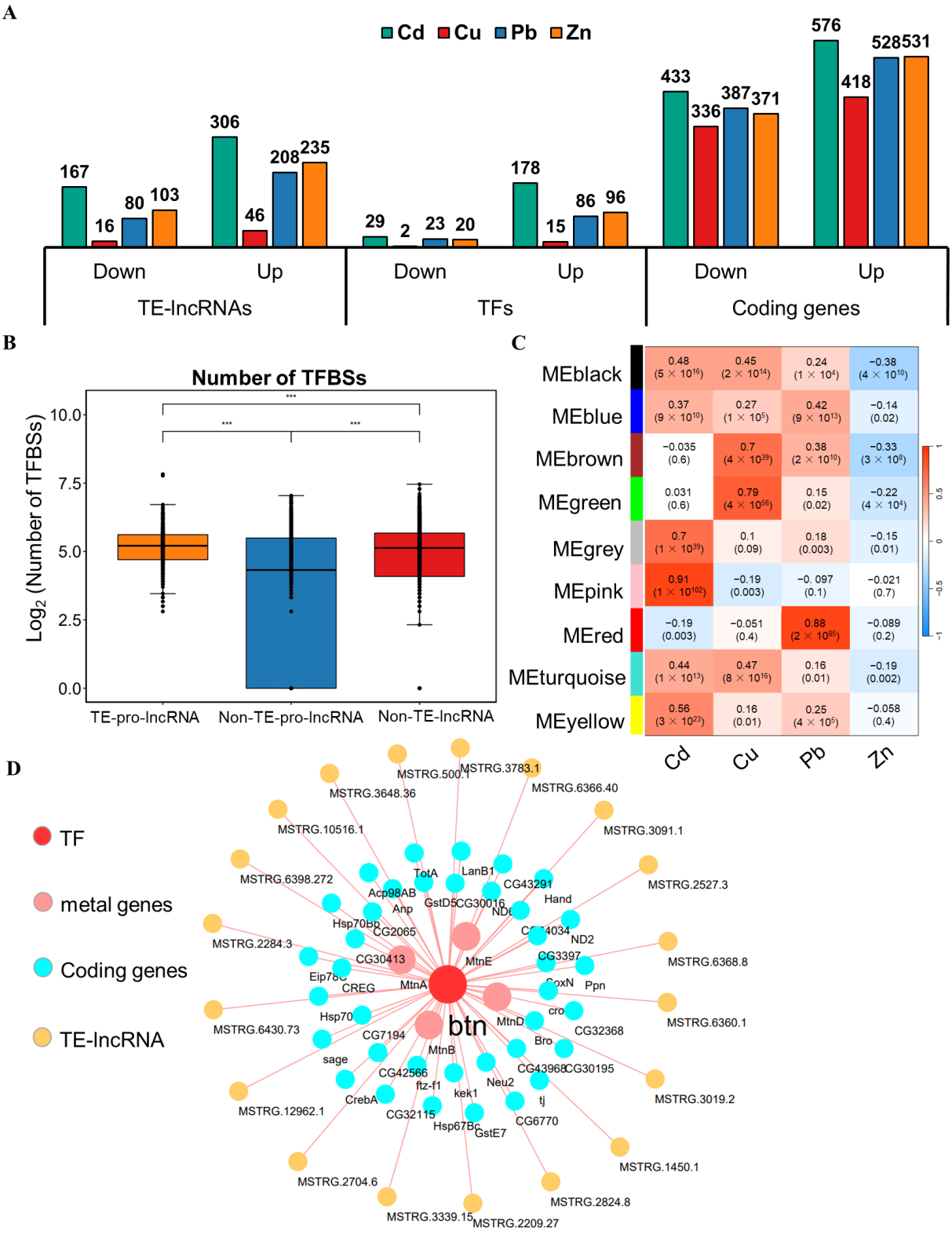

Figure 5. Co-expressed regulatory networks of TE-lncRNAs at the transcriptional level. (**A**) Differentially expressed genes under heavy metal conditions. (**B**) The quantities of TFBSs between TE-lncRNAs and Non-TE-lncRNAs. (**C**) The correlation between gene modules and heavy metal samples; each ME color represents a different expression module. (**D**) The TF regulatory networks mediated by transposons. Statistical significance: ***: $p < 0.001$.

We next validated the roles of TE-lncRNAs in regulatory networking mediated by transposons. By merging expression matrices from 271 samples encompassing TE-lncRNAs, transcription factors, and differentially expressed protein-coding genes, we constructed a comprehensive co-expression network (Figure S9C). Within this network, we identified known heavy metal responsive genes such as *MtnA*, *MtnB*, *MtnD*, and *MtnE*, which were clustered in a grey subnetwork module highly correlated with Cd stress samples (Figure 5C). Further analysis revealed the distribution of TFBSs within TE-lncRNAs and chromatin-accessible regions within the grey module, leading to the delineation of a regulatory network centered around the *btn* transcription factor (Figure 5D). For instance, *btn* exhibited binding sites on 19 TE-lncRNAs, including the promoter region of the TE-lncRNA, which displayed an open chromatin peak possibly facilitated by the transposon insertion, thereby providing a binding site for *btn* on this TE-lncRNA. This transposon-mediated formation of a transcriptional regulatory network contributes to the regulatory functions (Figure 5D).

We performed a further functional enrichment analysis of the protein-coding genes within the network. The analysis indeed revealed enrichment in functions associated with heavy metal response, such as response to metal ion, chaperone-mediated protein folding, and positive regulation of transcription (Figure 6A). Thus, in the transposon-mediated *btn* regulatory network, transposons may furnish binding sites for TE-lncRNAs, enabling the recruitment of transcription factors for transcriptional regulation related to heavy metal stress response. An examination of the TE-lncRNA expression within the co-expression regulatory network heatmap highlighted the highly induced expression level of *MSTRG.3783.1* (Figure 6B). Subsequent quantitative expression measurement validated the specific expression patterns of both *btn* and *MSTRG.3783.1* under heavy metal stress conditions (Figure 6C,D).

3.6. Post-Transcriptional Regulation of TE-lncRNAs

To investigate the association between TEs and m^6A modification sites in lncRNAs, we performed a comparative analysis of m^6A sites in TE-lncRNAs versus Non-TE-lncRNAs. We observed no significant difference in m^6A abundance between the two groups (Figure S10A). However, we identified 99 lncRNAs with m^6A modifications linked to TEs, with 16 of these showing elevated expression in heavy metal-treated S2 cells, suggesting a specific response to heavy metal stress (Figure S10B). Notably, the LTR/Pao and LTR/Gypsy transposon families were significant contributors to the m^6A-modified TE-lncRNAs (Figure S10B).

We also annotated open reading frames (ORFs) within lncRNAs and found a higher prevalence of candidate ORFs in TE-lncRNAs compared to Non-TE-lncRNAs, indicating that transposons may contribute to ORF formation within lncRNAs (Figure 7A). Further analysis identified 115 lncRNAs with TE-derived ORFs (TE-ORF-lncRNAs). Among these, 18 TE-ORF-lncRNAs showed increased expression following heavy metal treatment in S2 cells, with significant contributions from the LTR/Copia and RC/Helitron transposon families (Figure 7B). Additionally, Ribo-seq data revealed 170 translated TE-lncRNAs and nine translated TE-ORF-lncRNAs, while proteomics data identified 26 translated TE-lncRNAs and one translated TE-ORF-lncRNA (Figure 7C). Notably, the lncRNA *MSTRG.9753.7*, which contains an ORF derived from a transposon, was confirmed to be translated, while the peptide encoded by *MSTRG.9753.7* was validated using mass spectrometry (Figure 7D).

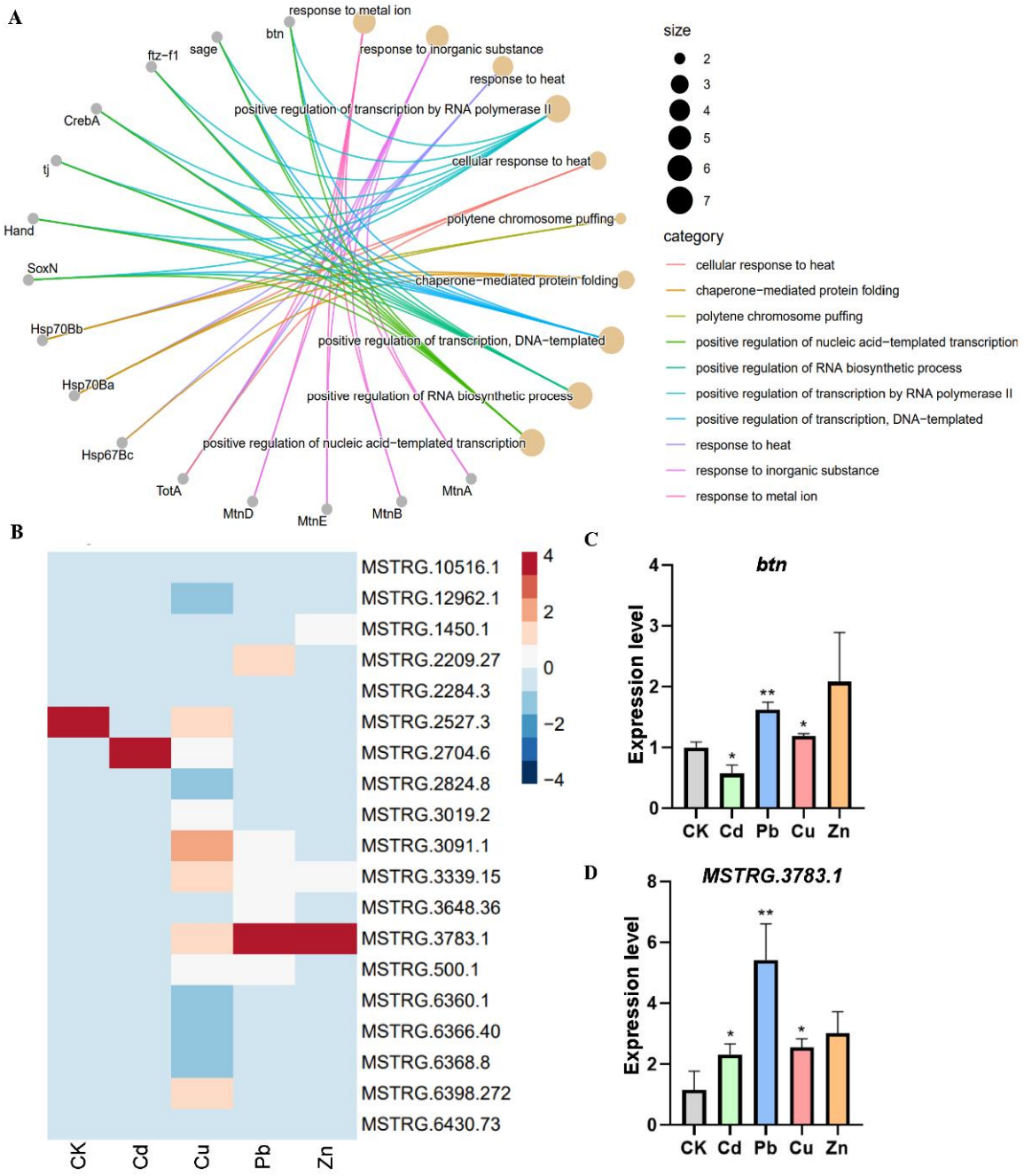

Figure 6. Validation of genes expressed in response to heavy metal stress in *Drosophila* S2 cells. (**A**) GO enrichment analysis of coding genes from those regulatory networks correlated with TE-lncRNAs; grey indicates genes, yellow indicates functional descriptions, and the line between grey circles and yellow circles indicates relations between genes and functions. (**B**) Expression heatmap of TE-*lncRNAs* from the *btn* network. (**C**) Expression validation of *btn* using quantitative PCR (qPCR). (**D**) Expression validation of *MSTRG.3783.1* using qPCR. Statistical significance: *: $p < 0.05$; **: $p < 0.01$.

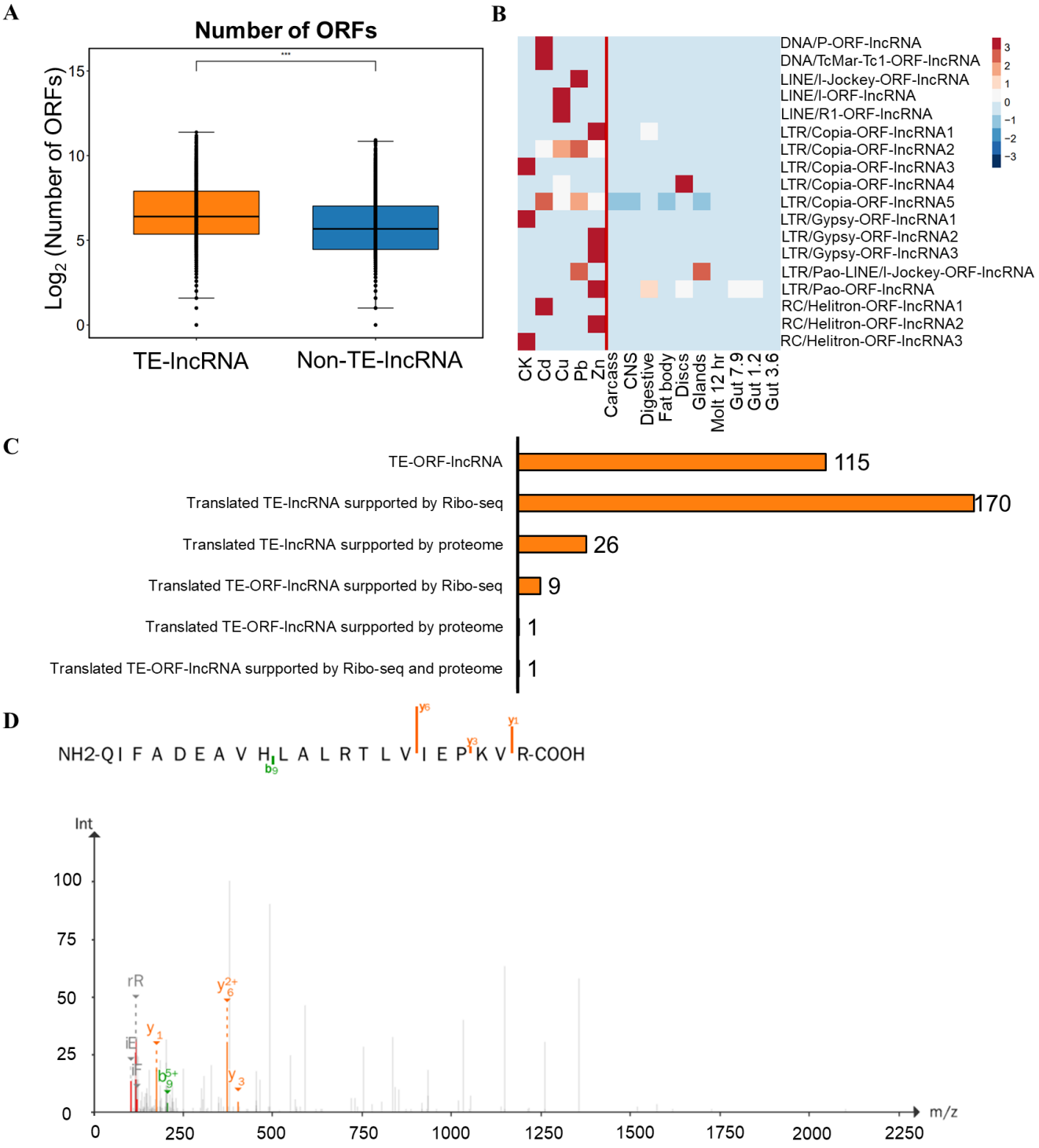

Figure 7. Effects of TE-lncRNAs on post-translational regulation in *Drosophila*. (**A**) Number of ORFs in TE-lncRNAs and Non-TE-lncRNAs. (**B**) Expression level of TE-ORF-lncRNAs in different samples. (**C**) Statistics of translated TE-lncRNAs. (**D**) Peptides encoded by TE-lncRNAs. Statistical significance: ***: $p < 0.001$.

4. Discussion

In this study, we comprehensively identified and functionally characterized a special type of lncRNA, known as TE-lncRNA, based on 256 public *Drosophila* RNA-seq datasets and 15 lncRNA-seq datasets obtained within our laboratory. We identified 5246 lncRNAs, from which 2119 TE-lncRNAs were further identified and overlapped with transposon

sequences in the *Drosophila* genome. We here identified many more lncRNAs than those (3085 lncRNAs) in a previous study [40], maybe because more lncRNA-seq transcriptome datasets are included, and a new pipeline and tools were used (see Methods). Previous research has demonstrated that TEs contribute to the origin, diversification, and regulation of lncRNAs in species such as humans, mice, zebrafish, chimpanzees, gorillas, and rhesus [32–34]. Our study highlighted the widespread contribution of TEs to lncRNAs in different species as well as in the model insect species, *D. melanogaster*.

Several recent studies have demonstrated the important functional roles of lncRNAs in *Drosophila*. Remarkably, the expression of the aal1 lncRNA in *Drosophila* boosts fly lifespan [72], and the dysfunction of lncRNA:CR43306 contributes to testicular aging [73]. For development and stress, bsAS, an antisense lncRNA, plays an essential role for correct wing development [74]; lncRNA CR40469, in trans responds to damage in the wing imaginal disk [75], and lncRNA NEAT1 is dramatically upregulated in stressed neurons [76]. LncRNAs are also involved in the regulation of obesity, immunity, and metabolisms in *Drosophila* [77]. For example, lncRNA-IRAR mediates the regulation of insulin receptor transcripts [78], and lincRNA-IBIN connects immunity and metabolism [79]. In addition, lncRNA VINR activates a non-canonical antimicrobial defense pathway in response to the VSR of *Drosophila* C virus [80]; CifA and CifB proteins alter lncRNA to establish a paternal-effect embryonic lethality [81]. From molecular aspects, roX lncRNAs are essential components of the chromatin modifying Dosage Compensation Complex (DCC) in *Drosophila* [82]. Meanwhile, lncRNAs play major roles in evolution by controlling transposable element activities, Y chromosome gene expression, and sperm construction [19]. Interestingly, some lncRNAs in *Drosophila* even could encode micropeptides [83].

Our analysis revealed that the LTR/Gypsy family represents the largest proportion (39%) of TEs within TE-lncRNAs in *D. melanogaster*, compared to its relatively minor presence (3%) in coding sequences. This suggests a preferential insertion of the LTR/Gypsy family into TE-lncRNAs. This finding is in line with previous studies across various species, where the LTR/Gypsy family is consistently the most abundant transposon family within lncRNAs [31,84–86]. In contrast, the ERV family is the most enriched for TE-lincRNAs in humans, while Alu elements predominate in mice TE-lincRNAs [32]. Among plants, different TE families are the primary contributors to TE-lincRNAs, namely Helitron in *Arabidopsis*, MITEs in rice, and Gypsy in maize [84]. These results suggest a potential role for these retained transposons in contributing to lncRNAs.

TE-lncRNAs differ from other gene types in several respects, including length, location, conservation, and expression levels. Notably, TE-lncRNAs are longer than Non-TE-lncRNAs. This finding suggests that transposon insertions contribute to the elongation of TE-lncRNAs, potentially increasing their resilience to transposon insertion. TE-lncRNAs also overlap with promoter and transcription start site (TSS) regions, suggesting they may function as regulatory elements. TE-lncRNAs are less conserved than other lncRNAs and coding genes. This variation in conservation among different gene types likely reflects the dynamic nature of transposon activity, leading to sequence variations and genomic instability.

Despite these potential functions, TE-lncRNAs are relatively poorly conserved, reflecting their origins from transposon insertions. They exhibit high specificity but low overall expression across various species, including humans, mice, zebrafish, and several plants [32,33,84,87,88]. TEs are often tightly regulated due to their potential disruptive impact on the genome and gene regulatory elements [27]. The piRNA system can control TE mobility by both transcriptional gene silencing and post-transcriptional gene silencing [89]. Thus, although the chromatin is accessible, active suppression mechanisms still could prevent TE-lncRNAs from being expressed at high levels.

We conducted a comparative analysis of the positioning of TEs genome-wide and in all TE-lncRNAs in *D. melanogaster*. A higher occurrence of TE-lncRNAs at transposon hot sites across the genome was observed. Notably, the majority of transposon insertions from TE-lncRNAs were concentrated within intronic regions, with a subset contributing to

promoter and TSSs regions, which is consistent with previous findings [32]. This preference suggests that the intronic region may exhibit greater tolerance to transposon insertion, while insertion in promoter regions could influence the regulation of nearby lncRNAs. This finding underscores the functional potential of transposon sequences, particularly when inserted into regulatory regions associated with biological activities. Of particular interest are several hotspots of transposon insertions on chromosomes, which are notably concentrated close to either end of some chromosomes or even within subtelomeric regions. Subtelomeric regions are also transcribed into coding and noncoding RNAs, such as ARRET, αARRET, subTERRA, and TERRA [90]. The establishment of open DNA replication-prone structures in subtelomeric regions could be mediated by these lncRNAs and even some TE-lncRNAs.

Epigenetic analysis revealed a significant positive correlation between the abundance of TEs in lncRNA regions and open chromatin regions. This supports the idea that transposons integrate into accessible chromatin, which may facilitate gene expression regulation by providing additional TFBSs. Our study also found that TE-lncRNAs with TEs in their promoter regions had a higher abundance of TFBSs compared to Non-TE-lncRNAs. This association suggests that transposons within lncRNAs may either provide functional TFBSs or facilitate their creation, facilitating transcription factor recruitment and enhancing the transcriptional regulatory functions of lncRNAs.

We used a heavy metal-induced *Drosophila* S2 cell model to explore the regulatory roles of transposon-mediated networks under stress. Differential expression analysis encompassing TE-lncRNAs, transcription factors, and coding genes revealed several well-known heavy metal responsive genes such as *MtnA, MtnB*, *MtnD*, and *MtnE* in a subnetwork module [91–93]. A regulatory network centered around transcription factor *btn* was delineated, in which *btn* exhibited binding sites on 19 TE-lncRNAs, including the promoter region of TE-lncRNA. Through meticulous screening for TE-lncRNAs exhibiting high correlation, this study also identified *MSTRG.3783.1* as a heavy metal-responsive TE-lncRNA. However, the specific functional role of this TE-lncRNA awaits further investigation. In terms of biological functions, *btn* is also the *Drosophila* homologue of MEOX2 in primary afferent nociceptor neurons that is proposed for the maintenance of a transcriptional program required for proper perception of acute and inflammatory noxious stimuli [94]. The *btn* contributes to cuticle pigmentation, and one biological process function for *btn* is "response to stimulus" such as noxious heat stimuli [94]. Thus, although no reports show a direct role of *btn* in response to metal, *btn* is implicated for getting involved in regulating heavy metal response and its specific regulatory role awaits further experimental validation.

At the post-transcriptional level, we examined m^6A modification sites within lncRNAs and found no significant difference between TE-lncRNAs and Non-TE-lncRNAs, indicating a weak connection between TEs and m^6A modification. However, translational analysis showed that TE-lncRNAs have more ORFs compared to Non-TE-lncRNAs, likely due to their longer length or the insertion of transposons with additional ORFs. Ribo-seq and proteomics data identified translated TE-lncRNAs, providing evidence that TEs contribute to the translation of lncRNAs and the encoding of short peptides. This suggests that TEs play a significant role in expanding the functional repertoire of lncRNAs.

The analysis findings regarding transposons providing TFBSs and ORF elements for lncRNAs corroborate the RIDL hypothesis, which posits that transposon insertion into lncRNAs can function akin to domains found in proteins [95]. Through a multi-omics approach, transposons are shown to facilitate the genesis of lncRNAs and exert regulatory influences by furnishing diverse functional elements for their functionality. By elucidating the presence and impacts of transposons at various omics tiers, we offered a new insight for exploring the intricate relationship between transposons and lncRNAs. One shortcoming might be the complexity of distinguishing the functional contributions of transposon-derived sequences from those of native genomic elements within lncRNAs. Additionally, the multi-omics approach, while comprehensive, could be influenced by dataset biases or technical limitations in detecting transposon activity at different omics levels. To address

these challenges, more experiments on transposon activity and regulatory impact may be needed in the future.

5. Conclusions

This study offers a comprehensive identification and analysis of TE-lncRNAs in *D. melanogaster*, elucidating their distinctive features and regulatory roles. A multi-omics integrated analysis approach helped unravel the complex regulatory networks of TE-lncRNAs. These findings collectively highlight the regulatory and functional significance of transposon-mediated mechanisms for shaping the landscape of lncRNA biology, opening avenues for future research on TE-lncRNAs.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/insects15120950/s1. Supplementary Table S1. RNA-seq samples. Table S2. Sample information of RNA-seq. Table S3. Sample information of ATAC -seq. Table S4. Sample information of MeRIP-seq. Table S5. Sample information of Ribo-seq. Table S6. TEs, TFBS, ATAC information in lncRNAs. Table S7. TFBS motifs in lncRNAs. Figure S1. Phastcons conservation assessment of coding genes, Non-TE-lncRNA, TE-lncRNA, and TE insertion sequences. (A) Phastcons score of coding genes. (B) Phastcons score of Non-TE-lncRNAs. (C) Phastcons score of TE-lncRNAs. (D) Phastcons score of TE insertions. Figure S2. Phylop conservation assessment of coding genes, Non-TE-lncRNA, TE-lncRNA, and TE insertion. (A) Phylop score of coding genes. (B) Phylop score of Non-TE-lncRNAs. (C) Phylop score of TE-lncRNAs. (D) Phylop score of TE insertions. Figure S3. Principal component analysis of coding genes, Non-TE-lncRNAs, and TE-lncRNAs in heavy metal samples. (A) PCA of coding genes. (B) PCA of Non-TE-lncRNAs. (C) PCA of TE-lncRNAs. Figure S4. Expression level of coding genes, Non-TE-lncRNAs, and TE-lncRNAs under heavy metal conditions. (A) Heatmap of expression for coding genes. (B) Heatmap of expression for Non-TE-lncRNAs. (C) Heatmap of expression for TE-lncRNAs. Figure S5. The contribution assessment of TEs in *D. melanogaster*. (A) Number of transposons from genome, CDS, and lncRNA. (B) Proportions of transposons from TE-lncRNA. (C) The TE insertions from lncRNA frequency near TSSs. (D) TEs from lncRNA insert into transcription factor binding loci relative to TSSs. (E) The gene feature of TE insertions from lncRNA. (F) The hotspot of TE insertions on the genome. Figure S6. Positions of TE insertions in the genome. (AB) The gene feature of TE insertions from the genome. (C) The TE insertions from genome frequency near TSSs. (D) TEs from genome insertion into transcription factor binding loci relative to TSSs. Figure S7. Length of TE insertions. (A) Length of TE derived for lncRNA. (B) TE insertion coverage in lncRNA. Figure S8. Correlations among TE number, TE coverage, TE coverage length, TFBSs, and ATAC peaks. (A) Correlations between TE coverage and TE number. (B) Correlations between TFBSs and TE number. (C) Correlations between TE coverage and TE coverage length. (D) Correlations between TFBSs and TE coverage length. (E) Correlations between ATAC peaks and TE coverage. (F) Correlations between TFBSs and TE coverage. (G) Correlations between TFBSs and ATAC peaks. (H) Correlations between TE coverage length and TE number. (I) Correlations between ATAC peaks and TE number. Figure S9. The transcriptional analysis of TE-lncRNA in *D. melanogaster*. (A) Number of TFBSs on TE-promoter-lncRNA. (B) Highly expressed TE-TF-lncRNAs under heavy metal conditions. (C) Co-expression network of heavy metal network. Figure S10. Post-transcriptional analysis of TE-lncRNA in *D. melanogaster*. (A) Comparison of m^6A peaks between Non-TE-lncRNA and TE-lncRNA. (B) Highly expressed TE-m^6A-lncRNAs under heavy metal conditions. References [39,41] are cited in the supplementary materials.

Author Contributions: Conceptualization, B.C. and Q.L.; methodology, B.C. and Q.L.; software, G.L., Y.Z. and K.C.; validation, L.W., X.G., G.L., Y.Z. and K.C.; formal analysis, Y.G.; investigation, Y.G.; resources, B.C. and Q.L.; data curation, Y.G.; writing—original draft preparation, Y.G.; writing—review and editing, B.C. and Q.L.; visualization, Y.G.; supervision, Z.Z. and F.Y.; project administration, B.C. and Q.L.; funding acquisition, B.C., Q.L. and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of China (31872304, 32270698, 32170593, and 32470447), the Hebei Natural Science Foundation (C2022201042), the Elite Rice Plan of GDRRI (2022YG01), and the Guangdong Provincial Pearl River Talent Plan (2019QN01N108).

Data Availability Statement: The data presented in the study are available in the Supplementary Materials.

Acknowledgments: We are grateful to Boyang Guo and Xia Zhang for their experimental help.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Birney, E.; Stamatoyannopoulos, J.A.; Dutta, A.; Guigo, R.; Gingeras, T.R.; Margulies, E.H.; Weng, Z.; Snyder, M.; Dermitzakis, E.T.; Stamatoyannopoulos, J.A.; et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **2007**, *447*, 799–816. [CrossRef] [PubMed]
2. Scheuermann, J.C.; Boyer, L.A. Getting to the heart of the matter: Long non-coding RNAs in cardiac development and disease. *EMBO J.* **2013**, *32*, 1805–1816. [CrossRef] [PubMed]
3. Wilusz, J.E.; Sunwoo, H.; Spector, D.L. Long noncoding RNAs: Functional surprises from the RNA world. *Genes Dev.* **2009**, *23*, 1494–1504. [CrossRef] [PubMed]
4. Zuo, L.; Su, H.; Zhang, Q.; Wu, W.-Y.; Zeng, Y.; Li, X.-M.; Xiong, J.; Chen, L.-F.; Zhou, Y. Comprehensive analysis of lncRNAs N6-methyladenosine modification in colorectal cancer. *Aging* **2021**, *13*, 4182–4198. [CrossRef]
5. Liu, Y.; Shi, M.; He, X.; Cao, Y.; Liu, P.; Li, F.; Zou, S.; Wen, C.; Zhan, Q.; Xu, Z.; et al. LncRNA-PACERR induces pro-tumour macrophages via interacting with miR-671-3p and m6A-reader IGF2BP2 in pancreatic ductal adenocarcinoma. *J. Hematol. Oncol.* **2022**, *15*, 52. [CrossRef]
6. Choi, S.-W.; Kim, H.-W.; Nam, J.-W. The small peptide world in long noncoding RNAs. *Brief. Bioinform.* **2019**, *20*, 1853–1864. [CrossRef]
7. Kong, S.; Tao, M.; Shen, X.; Ju, S. Translatable circRNAs and lncRNAs: Driving mechanisms and functions of their translation products. *Cancer Lett.* **2020**, *483*, 59–65. [CrossRef]
8. Matsumoto, A.; Pasut, A.; Matsumoto, M.; Yamashita, R.; Fung, J.; Monteleone, E.; Saghatelian, A.; Nakayama, K.I.; Clohessy, J.G.; Pandolfi, P.P. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature* **2017**, *541*, 228–232. [CrossRef]
9. Zhang, Q.; Vashisht, A.A.; O'Rourke, J.; Corbel, S.Y.; Moran, R.; Romero, A.; Miraglia, L.; Zhang, J.; Durrant, E.; Schmedt, C.; et al. The microprotein Minion controls cell fusion and muscle formation. *Nat. Commun.* **2017**, *8*, 15664. [CrossRef]
10. Huang, J.-Z.; Chen, M.; Chen, D.; Gao, X.-C.; Zhu, S.; Huang, H.; Hu, M.; Zhu, H.; Yan, G.-R. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol. Cell* **2017**, *68*, 171–184. [CrossRef]
11. D'Lima, N.G.; Ma, J.; Winkler, L.; Chu, Q.; Loh, K.H.; Corpuz, E.O.; Budnik, B.A.; Lykke-Andersen, J.; Saghatelian, A.; Slavoff, S.A. A human microprotein that interacts with the mRNA decapping complex. *Nat. Chem. Biol.* **2017**, *13*, 174–180. [CrossRef] [PubMed]
12. Sun, L.; Wang, W.; Han, C.; Huang, W.; Sun, Y.; Fang, K.; Zeng, Z.; Yang, Q.; Pan, Q.; Chen, T.; et al. The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation. *Mol. Cell* **2021**, *81*, 4493–4508. [CrossRef] [PubMed]
13. Anderson, D.M.; Anderson, K.M.; Chang, C.-L.; Makarewich, C.A.; Nelson, B.R.; McAnally, J.R.; Kasaragod, P.; Shelton, J.M.; Liou, J.; Bassel-Duby, R.; et al. A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance. *Cell* **2015**, *160*, 595–606. [CrossRef] [PubMed]
14. Nelson, B.R.; Makarewich, C.A.; Anderson, D.M.; Winders, B.R.; Troupes, C.D.; Wu, F.; Reese, A.L.; McAnally, J.R.; Chen, X.; Kavalali, E.T.; et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* **2016**, *351*, 271–275. [CrossRef] [PubMed]
15. Pauli, A.; Norris, M.L.; Valen, E.; Chew, G.-L.; Gagnon, J.A.; Zimmerman, S.; Mitchell, A.; Ma, J.; Dubrulle, J.; Reyon, D.; et al. Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors. *Science* **2014**, *343*, 1248636. [CrossRef]
16. 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*, 660–665. [CrossRef]
17. Magny, E.G.; Pueyo, J.I.; Pearl, F.M.; Cespedes, M.A.; Niven, J.E.; Bishop, S.A.; Couso, J.P. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* **2013**, *341*, 1116–1120. [CrossRef]
18. Hanyu-Nakamura, K.; Sonobe-Nojima, H.; Tanigawa, A.; Lasko, P.; Nakamura, A. Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. *Nature* **2008**, *451*, 730–733. [CrossRef]
19. Shao, Z.; Hu, J.; Jandura, A.; Wilk, R.; Jachimowicz, M.; Ma, L.; Hu, C.; Sundquist, A.; Das, I.; Samuel-Larbi, P.; et al. Spatially revealed roles for lncRNAs in Drosophila spermatogenesis, Y chromosome function and evolution. *Nat. Commun.* **2024**, *15*, 3806. [CrossRef]
20. Lisch, D. How important are transposons for plant evolution? *Nat. Rev. Genet.* **2013**, *14*, 49–61. [CrossRef]
21. Piégu, B.; Bire, S.; Arensburger, P.; Bigot, Y. A survey of transposable element classification systems—A call for a fundamental update to meet the challenge of their diversity and complexity. *Mol. Phylogenetics Evol.* **2015**, *86*, 90–109. [CrossRef] [PubMed]
22. Arensburger, P.; Piégu, B.; Bigot, Y. The future of transposable element annotation and their classification in the light of functional genomics—What we can learn from the fables of Jean de la Fontaine? *Mob. Genet. Elem.* **2016**, *6*, e1256852. [CrossRef] [PubMed]
23. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [CrossRef] [PubMed]

24. Hua-Van, A.; Capy, P. *Transposable Elements and Genome Evolution*; John Wiley & Sons: Hoboken, NJ, USA, 2024.
25. Walser, J.-C.; Chen, B.; Feder, M.E. Heat-Shock Promoters: Targets for Evolution by P Transposable Elements in Drosophila. *PLoS Genet.* **2006**, *2*, e165. [CrossRef] [PubMed]
26. Chen, B.; Zhang, B.; Xu, L.; Li, Q.; Jiang, F.; Yang, P.; Xu, Y.; Kang, L. Transposable Element-Mediated Balancing Selection at Hsp90 Underlies Embryo Developmental Variation. *Mol. Biol. Evol.* **2017**, *34*, 1127–1139. [CrossRef]
27. Chen, B.; Walser, J.C.; Rodgers, T.H.; Sobota, R.S.; Burke, M.K.; Rose, M.R.; Feder, M.E. Abundant, diverse, and consequentialPelements segregate in promoters of small heat-shock genes inDrosophilapopulations. *J. Evol. Biol.* **2007**, *20*, 2056–2066. [CrossRef]
28. Chen, B.; Shilova, V.Y.; Zatsepina, O.G.; Evgen'ev, M.B.; Feder, M.E. Location of P element insertions in the proximal promoter region of Hsp70A is consequential for gene expression and correlated with fecundity in Drosophila melanogaster. *Cell Stress Chaperones* **2008**, *13*, 11–17. [CrossRef]
29. Wang, K.; Hua, G.; Li, J.; Yang, Y.; Zhang, C.; Yang, L.; Hu, X.; Scheben, A.; Wu, Y.; Gong, P.; et al. Duck pan-genome reveals two transposon insertions caused bodyweight enlarging and white plumage phenotype formation during evolution. *iMeta* **2023**, *3*, e154. [CrossRef]
30. Xia, B.; Zhang, W.; Zhao, G.; Zhang, X.; Bai, J.; Brosh, R.; Wudzinska, A.; Huang, E.; Ashe, H.; Ellis, G.; et al. On the genetic basis of tail-loss evolution in humans and apes. *Nature* **2024**, *626*, 1042–1048. [CrossRef]
31. Zhang, X.; Zhu, Y.N.; Chen, B.; Kang, L. A Gypsy element contributes to the nuclear retention and transcriptional regulation of the resident lncRNA in locusts. *Rna Biol.* **2022**, *19*, 206–220. [CrossRef]
32. Kelley, D.; Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* **2012**, *13*, R107. [CrossRef] [PubMed]
33. Kapusta, A.; Kronenberg, Z.; Lynch, V.J.; Zhuo, X.Y.; Ramsay, L.; Bourque, G.; Yandell, M.; Feschotte, C. Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *PloS Genet.* **2013**, *9*, e1003470. [CrossRef] [PubMed]
34. Ramsay, L.; Marchetto, M.C.; Caron, M.; Chen, S.H.; Busche, S.; Kwan, T.; Pastinen, T.; Gage, F.H.; Bourque, G. Conserved expression of transposon-derived non-coding transcripts in primate stem cells. *BMC Genom.* **2017**, *18*, 214. [CrossRef] [PubMed]
35. Wang, X.; Ai, G.; Zhang, C.; Cui, L.; Wang, J.; Li, H.; Zhang, J.; Ye, Z. Expression and diversification analysis reveals transposable elements play important roles in the origin of Lycopersicon-specific lncRNAs in tomato. *New Phytol.* **2016**, *209*, 1442–1455. [CrossRef]
36. Wang, Z.X.; Schwacke, R.; Kunze, R. DNA Damage-Induced Transcription of Transposable Elements and Long Non-coding RNAs in Arabidopsis Is Rare and ATM-Dependent. *Mol. Plant* **2016**, *9*, 1142–1155. [CrossRef]
37. Zhang, X.; Xu, Y.n.; Chen, B.; Kang, L. Long noncoding RNA PAHAL modulates locust behavioural plasticity through the feedback regulation of dopamine biosynthesis. *PLoS Genet.* **2020**, *16*, e1008771. [CrossRef]
38. Melo, E.S.; Wallau, G.L. Mosquito long non-coding RNAs are enriched with Transposable Elements. *Genet Mol Biol* **2022**, *45*, e20210215. [CrossRef]
39. Graveley, B.R.; Brooks, A.N.; Carlson, J.; Duff, M.O.; Landolin, J.M.; Yang, L.; Artieri, C.G.; van Baren, M.J.; Boley, N.; Booth, B.W.; et al. The developmental transcriptome of Drosophila melanogaster. *Nature* **2011**, *471*, 473–479. [CrossRef]
40. Brown, J.B.; Boley, N.; Eisman, R.; May, G.E.; Stoiber, M.H.; Duff, M.O.; Booth, B.W.; Wen, J.; Park, S.; Suzuki, A.M.; et al. Diversity and dynamics of the Drosophila transcriptome. *Nature* **2014**, *512*, 393–399. [CrossRef]
41. Guo, B.; Li, T.; Wang, L.; Liu, F.; Chen, B. Long non-coding RNAs regulate heavy metal-induced apoptosis in embryo-derived cells. *Environ. Pollut.* **2023**, *333*, 121956. [CrossRef]
42. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, 884–890. [CrossRef] [PubMed]
43. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [CrossRef] [PubMed]
44. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.-C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [CrossRef] [PubMed]
45. Pertea, G.; Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Research* **2020**, *9*, 304. [CrossRef]
46. Kang, Y.-J.; Yang, D.-C.; Kong, L.; Hou, M.; Meng, Y.-Q.; Wei, L.; Gao, G. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **2017**, *45*, W12–W16. [CrossRef]
47. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
48. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef]
49. Smit, A.H.R.; Green, P. RepeatMasker Open-4.0. Available online: http://www.repeatmasker.org (accessed on 9 April 2019).
50. Tempel, S. Using and understanding RepeatMasker. *Methods Mol. Biol.* **2012**, *859*, 29–51. [CrossRef]
51. Jurka, J.; Kapitonov, V.V.; Pavlicek, A.; Klonowski, P.; Kohany, O.; Walichiewicz, J. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **2005**, *110*, 462–467. [CrossRef]
52. Bao, W.; Kojima, K.K.; Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **2015**, *6*, 11. [CrossRef]

53. Yu, G.C.; Wang, L.G.; He, Q.Y. ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **2015**, *31*, 2382–2383. [CrossRef] [PubMed]
54. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202. [CrossRef] [PubMed]
55. Palmer, D.; Fabris, F.; Doherty, A.; Freitas, A.A.; de Magalhães, J.P. Ageing transcriptome meta-analysis reveals similarities and differences between key mammalian tissues. *Aging* **2021**, *13*, 3313–3341. [CrossRef] [PubMed]
56. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *Bmc Bioinform.* **2008**, *9*, 559. [CrossRef]
57. Yu, G.C.; Wang, L.G.; Han, Y.Y.; He, Q.Y. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *Omics-A J. Integr. Biol.* **2012**, *16*, 284–287. [CrossRef]
58. Ramirez, F.; Duendar, F.; Diehl, S.; Gruening, B.A.; Manke, T. deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **2014**, *42*, W187–W191. [CrossRef]
59. Zhang, Y.; Liu, T.; Meyer, C.A.; Eeckhoute, J.; Johnson, D.S.; Bernstein, B.E.; Nussbaum, C.; Myers, R.M.; Brown, M.; Li, W.; et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **2008**, *9*, R137. [CrossRef]
60. Quinlan, A.R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinform.* **2014**, *47*, 11.12.1–11.12.34. [CrossRef]
61. Meng, J.; Cui, X.; Rao, M.K.; Chen, Y.; Huang, Y. Exome-based analysis for RNA epigenome sequencing data. *Bioinformatics* **2013**, *29*, 1565–1567. [CrossRef]
62. Liu, Q.; Gregory, R.I. RNAmod: An integrated system for the annotation of mRNA modifications. *Nucleic Acids Res.* **2019**, *47*, W548–W555. [CrossRef]
63. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]
64. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]
65. Zhang, P.; He, D.; Xu, Y.; Hou, J.; Pan, B.-F.; Wang, Y.; Liu, T.; Davis, C.M.; Ehli, E.A.; Tan, L.; et al. Genome-wide identification and differential analysis of translational initiation. *Nat. Commun.* **2017**, *8*, 1749. [CrossRef] [PubMed]
66. Lauria, F.; Tebaldi, T.; Bernabo, P.; Groen, E.J.N.; Gillingwater, T.H.; Viero, G. riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Comput. Biol.* **2018**, *14*, e1006169. [CrossRef] [PubMed]
67. Ji, Z. RibORF: Identifying Genome-Wide Translated Open Reading Frames Using Ribosome Profiling. *Curr. Protoc. Mol. Biol.* **2018**, *124*, e67. [CrossRef]
68. Gregory, R.I.; Sliz, P.; Shvarts, T.; Liu, Q. RiboToolkit: An integrated platform for analysis and annotation of ribosome profiling data to decode mRNA translation at codon resolution. *Nucleic Acids Res.* **2020**, *48*, W218–W229. [CrossRef]
69. Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **2016**, *11*, 2301–2319. [CrossRef]
70. Li, K.; Vaudel, M.; Zhang, B.; Ren, Y.; Wen, B. PDV: An integrative proteomics data viewer. *Bioinformatics* **2019**, *35*, 1249–1251. [CrossRef]
71. Luo, H.; Bu, D.; Sun, L.; Chen, R.; Zhao, Y. De novo approach to classify protein-coding and noncoding transcripts based on sequence composition. *Methods Mol. Biol.* **2014**, *1182*, 203–207. [CrossRef]
72. Anver, S.; Sumit, A.F.; Sun, X.M.; Hatimy, A.; Thalassinos, K.; Marguerat, S.; Alic, N.; Bahler, J. Ageing-associated long non-coding RNA extends lifespan and reduces translation in non-dividing cells. *EMBO Rep* **2024**, *25*, 4921–4949. [CrossRef]
73. Huang, Q.; Li, J.; Qi, Y.; He, X.; Shen, C.; Wang, C.; Wang, X.; Xia, Q.; Zhang, Y.; Pan, Z.; et al. Copper overload exacerbates testicular aging mediated by lncRNA:CR43306 deficiency through ferroptosis in Drosophila. *Redox Biol.* **2024**, *76*, 103315. [CrossRef] [PubMed]
74. Perez-Lluch, S.; Klein, C.C.; Breschi, A.; Ruiz-Romero, M.; Abad, A.; Palumbo, E.; Bekish, L.; Arnan, C.; Guigo, R. bsAS, an antisense long non-coding RNA, essential for correct wing development through regulation of blistered/DSRF isoform usage. *PLoS Genet* **2020**, *16*, e1009245. [CrossRef] [PubMed]
75. Camilleri-Robles, C.; Amador, R.; Tiebe, M.; Teleman, A.A.; Serras, F.; Guigó, R.; Corominas, M. Long non-coding RNAs involved in Drosophila development and regeneration. *NAR Genom. Bioinform.* **2024**, *6*, lqae091. [CrossRef] [PubMed]
76. Wang, C.; Duan, Y.; Duan, G.; Wang, Q.; Zhang, K.; Deng, X.; Qian, B.; Gu, J.; Ma, Z.; Zhang, S.; et al. Stress Induces Dynamic, Cytotoxicity-Antagonizing TDP-43 Nuclear Bodies via Paraspeckle LncRNA NEAT1-Mediated Liquid-Liquid Phase Separation. *Mol. Cell* **2020**, *79*, 443–458.e7. [CrossRef]
77. Dayal Aggarwal, D.; Mishra, P.; Yadav, G.; Mitra, S.; Patel, Y.; Singh, M.; Sahu, R.K.; Sharma, V. Decoding the connection between lncRNA and obesity: Perspective from humans and Drosophila. *Heliyon* **2024**, *10*, e35327. [CrossRef]
78. Chen, J.; Huang, Y.; Qi, G. LncRNA-IRAR-mediated regulation of insulin receptor transcripts in Drosophila melanogaster during nutritional stress. *Insect Mol. Biol.* **2022**, *31*, 261–272. [CrossRef]
79. Valanne, S.; Salminen, T.S.; Jarvela-Stolting, M.; Vesala, L.; Ramet, M. Immune-inducible non-coding RNA molecule lincRNA-IBIN connects immunity and metabolism in Drosophila melanogaster. *PLoS Pathog* **2019**, *15*, e1007504. [CrossRef]
80. Zhang, L.; Xu, W.; Gao, X.; Li, W.; Qi, S.; Guo, D.; Ajayi, O.E.; Ding, S.W.; Wu, Q. lncRNA Sensing of a Viral Suppressor of RNAi Activates Non-canonical Innate Immune Signaling in Drosophila. *Cell Host Microbe* **2020**, *27*, 115–128.e8. [CrossRef]

81. Kaur, R.; McGarry, A.; Shropshire, J.D.; Leigh, B.A.; Bordenstein, S.R. Prophage proteins alter long noncoding RNA and DNA of developing sperm to induce a paternal-effect lethality. *Science* **2024**, *383*, 1111–1117. [CrossRef]
82. Becker, P.B.; Thomae, A.W.; Villa, R.; Krause, S.; Schauer, T.; Müller, M. Two-step mechanism for selective incorporation of lncRNA into a chromatin modifier. *Nucleic Acids Res.* **2020**, *48*, 7483–7501. [CrossRef]
83. Patraquim, P.; Magny, E.G.; Pueyo, J.I.; Platero, A.I.; Couso, J.P. Translation and natural selection of micropeptides from long non-canonical RNAs. *Nat. Commun.* **2022**, *13*, 6515. [CrossRef] [PubMed]
84. Wang, D.; Qu, Z.P.; Yang, L.; Zhang, Q.Z.; Liu, Z.H.; Do, T.; Adelson, D.L.; Wang, Z.Y.; Searle, I.; Zhu, J.K. Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants. *Plant J.* **2017**, *90*, 133–146. [CrossRef] [PubMed]
85. Zhao, T.; Tao, X.; Feng, S.; Wang, L.; Hong, H.; Ma, W.; Shang, G.; Guo, S.; He, Y.; Zhou, B.; et al. LncRNAs in polyploid cotton interspecific hybrids are derived from transposon neofunctionalization. *Genome Biol* **2018**, *19*, 195. [CrossRef] [PubMed]
86. Li, W.; Chen, Y.; Wang, Y.; Zhao, J.; Wang, Y. Gypsy retrotransposon-derived maize lncRNA GARR2 modulates gibberellin response. *Plant J.* **2022**, *110*, 1433–1446. [CrossRef]
87. Chen, X.; Wang, C.; He, B.; Wan, Z.; Zhao, Y.; Hu, F.; Lv, Y. Transcriptome Profiling of Transposon-Derived Long Non-coding RNAs Response to Hormone in Strawberry Fruit Development. *Front. Plant Sci.* **2022**, *13*, 915569. [CrossRef]
88. Zhu, Y.; Chen, L.; Hong, X.; Shi, H.; Li, X. Revealing the novel complexity of plant long non-coding RNA by strand-specific and whole transcriptome sequencing for evolutionarily representative plant species. *BMC Genom.* **2022**, *23*, 381. [CrossRef]
89. Cerqueira de Araujo, A.; Huguet, E.; Herniou, E.A.; Drezen, J.M.; Josse, T. Transposable element repression using piRNAs, and its relevance to endogenous viral elements (EVEs) and immunity in insects. *Curr. Opin. Insect Sci.* **2022**, *50*, 100876. [CrossRef]
90. Kwapisz, M.; Morillon, A. Subtelomeric Transcription and its Regulation. *J. Mol. Biol.* **2020**, *432*, 4199–4219. [CrossRef]
91. Günther, V.; Lindert, U.; Schaffner, W. The taste of heavy metals: Gene regulation by MTF-1. *Biochim. Et Biophys. Acta (BBA)—Mol. Cell Res.* **2012**, *1823*, 1416–1425. [CrossRef]
92. Navarro, J.A.; Schneuwly, S. Copper and Zinc Homeostasis: Lessons from Drosophila melanogaster. *Front. Genet.* **2017**, *8*, 223. [CrossRef]
93. Slobodian, M.R.; Petahtegoose, J.D.; Wallis, A.L.; Levesque, D.C.; Merritt, T.J.S. The Effects of Essential and Non-Essential Metal Toxicity in the Drosophila melanogaster Insect Model: A Review. *Toxics* **2021**, *9*, 269. [CrossRef] [PubMed]
94. Kokotovic, T.; Lenartowicz, E.M.; Langeslag, M.; Ciotu, C.I.; Fell, C.W.; Scaramuzza, A.; Fischer, M.J.M.; Kress, M.; Penninger, J.M.; Nagy, V. Transcription factor mesenchyme homeobox protein 2 (MEOX2) modulates nociceptor function. *FEBS J.* **2022**, *289*, 3457–3476. [CrossRef] [PubMed]
95. Johnson, R.; Guigo, R. The RIDL hypothesis: Transposable elements as functional domains of long noncoding RNAs. *RNA* **2014**, *20*, 959–976. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.