The insulator activity of tRNA-V SINEs and its implications in the transcriptional regulation of hypoxic response in zebrafish

Tesis

Entregada A La
Universidad De Chile
En Cumplimiento Parcial De Los Requisitos
Para Optar Al Grado De

Doctor en Biotecnología Molecular

Facultad De Ciencias

Por

Felipe Andrés Gajardo Escobar

Abril, 2022

Director de Tesis Dr:
Miguel Allende

FACULTAD DE CIENCIAS

UNIVERSIDAD DE CHILE

INFORME DE APROBACIÓN

TESIS DE DOCTORADO

Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de Doctorado presentada por el candidato.

Felipe Andrés Gajardo Escobar

Ha sido aprobada por la comisión de Evaluación de la tesis como requisito para optar al grado de Doctor en Biotecnología Molecular, en el examen de Defensa Privada de Tesis rendido el día .....

Director de Tesis:
Dr. Miguel Allende .....

Comisión de Evaluación de la Tesis

Dr. Marcelo Baeza .....

Dr. Alejandro Maass .....

Dr. Martín Montecinos .....

Dra. Veronica Palma .....

RESUMEN

Los elementos transponibles (ETs) son uno de los componentes más abundantes en la cromatina. A pesar de esto, su participación como factores estructurales en la organización del genoma es una materia que no ha sido estudiada a cabalidad. En este trabajo hemos usado una aproximación fundamentada en el análisis masivo de datos para predecir ETs que potencialmente participen en el establecimiento de estructuras aisladoras en el genoma de pez cebra. Nuestros resultados apuntan a retrotransposones de la superfamilia tRNA-V como ampliamente distribuidos y potencialmente capaces de actuar como aisladores en múltiples *loci*. Más aún, identificamos genes ubicados cerca de elementos tRNA-V, los cuales son diferencialmente expresados durante estrés hipóxico. Esto sugiere que elementos tRNA-V podrían estar involucrados en la regulación transcripcional de la respuesta a estrés hipóxico.

SUMMARY

Transposable elements (TEs) are a major component of chromatin. Despite this, the implications of TEs as structural players in the organization of the genome is a subject that remains poorly understood. In this work, we have used a data-drive approach to address whether SINEs participate in the establishment of insulator structures in the genome of zebrafish. Our results point to tRNA-V retrotransposons as a widespread superfamily potentially acting as insulators at multiple loci. Furthermore, we identified genes located in close proximity to tRNA-V elements, which are differentially expressed during hypoxic stress, suggesting that they may be involved in the transcriptional regulation of gene expression upon hypoxic stress.

AGRADECIMIENTOS

Datos. Fotografías abstractas fundamentadas en la organización de la materia, alimentan mi curiosidad y me quitan el sueño. Gracias por los datos. Gracias a la materia misma que les dio origen, y gracias al humano que pago en tiempo, energía y dinero su obtención. Gracias también al ingenio de los pensadores quienes concibieron las ideas que, aun hoy, siguen extendiendo los límites de la percepción humana. Entre ellos, gracias a Barbara McClintok por fijar la mirada en los "genes saltarines", y también a Alexandra Elbakyan por cuestionar y corromper el deseo humano de la propiedad en el plano del conocimiento.

En lo personal, como no agradecer a mi familia. Gracias Jose Luis Gajardo por incentivar e inspirar mi propio ingenio, y mostrarme que el mundo "funciona", tal como lo hace una bicicleta, y también a Milena Escobar del Valle, por mostrarme que aunque el mundo "funcione", aun hay algo de mágico y misterioso en aquello que nos rodea, y que a la vez nos da forma. Gracias también a mis hermanos y amigos por escuchar mis delirios científicos y alimentar la hoguera con preguntas que más de alguna vez me dejaron pensando.

Y, ¿que sería de mi sin mis amigos y compañeros en la ciencia? hay que decirlo, han tenido una paciencia inigualable con este porfiado. Gracias por las abstractas discusiones y abundante cerveza. Y por supuesto, este grato ambiente no estaría completo sin mi tutor de tesis, Miguel Allende, que ha apoyado mi desarrollo científico durante años.

Finalmente, gracias también a todos esos extraños y no tan extraños, que de una forma u otra, contribuyeron al flujo de pensamiento que llevó a las ideas expuestas en este trabajo.

INDEX

FIGURES

TABLES

INTRODUCTION

Transposable elements (TEs) are diverse and highly repetitive nucleic acid sequences present in the genome of many living organisms (Feschotte & Pritham, 2007). They owe their name to their ability to transpose from one genomic location to another (McClintock, 1950). This process is responsible for several evolutive innovations such as the development of the placenta in mammals (Emera & Wagner, 2012) and the VDJ system involved in adaptive immunity (Jones & Gellert, 2004). However, not all events of transposition produce evolutive innovations. While most of them have little to no effect on the individual phenotype, some may cause harmful consequences for the host (Arkhipova, 2018). Hence, TEs are commonly found in heterochromatin regions, silenced by epigenetic mechanisms, and only a few of them become co-opted or domesticated (Slotkin & Martienssen, 2007).

In eukaryotes, TEs can account for more than half the size of the complete genome or even more depending on the species. Indeed, they are a major component of chromatin (Wells & Feschotte, 2020). Despite their abundance, the implications of TEs as structural players in the organization of the genome is a subject that remains poorly understood. Only recently, some studies have shed light on this matter, providing evidence for the interaction between architectural proteins and sequences of TE origin.

For instance, it has been shown that SINE retrotransposons can bind the CCCTC-binding factor (CTCF), which together with cohesin are key proteins in the spatial organization of the genome, due to their implications in the formation of chromatin loops (Choudhary et al., 2020).

Generally speaking, chromatin loops occur whenever cohesin extrudes the chromatin fiber and forms a loop that gets stacked at convergently oriented CTCF binding sites. In this context, CTCF is relevant for the establishment of the loop boundaries. The resulting CTCF-cohesin-chromatin complex is called the loop-extrusion complex (LEC) and has interesting properties from a regulatory perspective (Davidson & Peters, 2021). On one hand, it may promote long-range interactions between enhancers and promoters, thus being able to influence gene expression. On the other hand, it may participate in the generation of insulator structures, which block enhancer-promoter interactions and constrains the spread of epigenetic modifications associated with silent chromatin (Chen & Lei, 2019; Yang & Corces, 2012; Brasset & Vaury, 2005). In this regard, it is interesting that some SINEs possess intrinsic insulator activity, and this activity can be modulated by factors associated with stress (Roman et al, 2011; Raab et al., 2012).

Considering these antecedents, it is paradoxical that few (Wang et al., 2015) of the methodologies that have been proposed to discover novel insulators do even consider TEs as useful factors for achieving such a goal (Srinivasan et al., 2022; Belokopytova &

Fishman, 2021). In this work, we propose a methodology for detecting novel CTCF/Cohesin-mediated insulators by exploiting their known relationship with TEs and the presence of signatures commonly associated with insulators.

Briefly, our methodology considers an initial screening for selecting TE families more widely distributed in the vicinity of genes, thus, having the potential to affect gene regulation due to their proximity with promoters and 3'-UTR regions. Then, based on occupancy data from CHIP-seq experiments, we describe these insertions in terms of their occupancy by architectural proteins and regulatory-relevant histone modifications and compare the expression levels of genes upstream and downstream of insertions, based on data from RNA-seq experiments. Finally, we consider differential expression analyses between control and stress conditions in order to identify putative insulators whose activity may be affected by stress. The final result of our workflow is a complete description of multiple TE *loci* across the genome, which are proximal to genes and whose regulation is potentially affected by stress.

As a proof of concept, we applied this methodology to the genome of zebrafish using publicly available data from normoxic and hypoxic individuals. We found that the tRNA-V superfamily is the most commonly found TE in the vicinity of genes, and moreover, our analyses revealed sets of insertions with characteristic patterns of occupancy for CTCF, Rad21 (Cohesin component), and H3K27me3, which also show

consistent differences in gene expression levels upstream and downstream of the insertion, thus, supporting insulator activities for tRNA-V elements. We further inspected some specific *loci* whose putative insulator activity was likely to be affected during hypoxic stress. Altogether, our data support that some tRNA-V elements may act as insulators in the genome of zebrafish, and highlights hundreds of *loci* where the observed changes in gene expression upon hypoxic stress may be explained by alterations in the putative insulator activity of tRNA-V elements.

Hypothesis

Some TEs participate as insulator elements in the genome of zebrafish, and alterations of their insulator activity promote transcriptional changes during hypoxic stress.

Main goal

Identify families and particular TEs with insulator activity in the genome of zebrafish highlighting those with the potential to explain transcriptional changes during hypoxic stress.

Methodological goal

Develop a methodology for detecting novel CTCF/Cohesin-mediated insulators by exploiting their known relationship with TEs and the presence of signatures commonly associated with insulators

Specific goals

1. Find families of TEs with a higher probability to occur in the vicinity of genes.

2. Characterize tRNA-V elements according to their binding with architectural proteins, factors associated with hypoxic response, and epigenetic marks.

3. Evaluate the activity of tRNA-V insertions as insulator elements in normoxic and hypoxic conditions

RESULTS

SINEs of the tRNA-V superfamily are frequently found in the vicinity of genes

In order to explore the distribution of different TE families in the vicinity of genes we first quantified the number of TEs on different genomic regions (see methods). Since we are interested in TEs with the potential to exert regulatory influence over promoters, we narrowed our screening to intergenic regions, which we defined as 1 kb upstream and downstream of genes.

As a result, we obtained matrices containing the number of fragments of TEs in several regions:

- Sense TE insertions overlapping 1kb upstream of the TSS
- Antisense TE insertions overlapping 1kb upstream of the TSS
- Sense TE insertions overlapping 1kb downstream of the end of the transcript
- Antisense TE insertions overlapping 1kb downstream of the end of the transcript
- Sense TE insertions overlapping Exons
- Antisense TE insertions overlapping Exons

In accordance with the proportions observed genome-wide for TE orders in the genome of zebrafish (Wells & Feschotte, 2020), our results show that DNA elements are the most abundant in all the regions considered, followed by LTRs, LINEs, and SINEs with the least number of fragments (Figure 1).

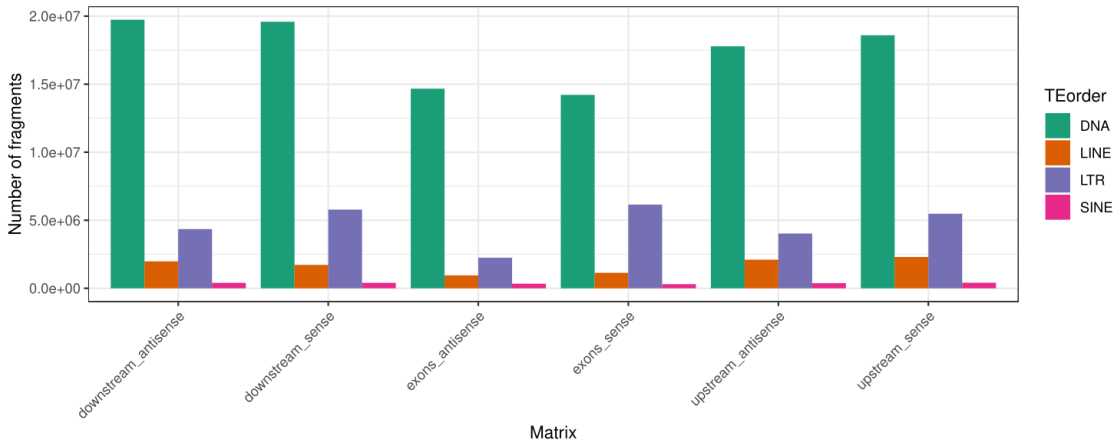


Figure 1: Abundance of TEs in gene-associated regions by order according to the generated abundance matrices.

Since some families of TEs might be not too abundant, but highly spread among many genes, we calculated the proportion of genes with at least one insertion for each of the TE families in the genome of zebrafish, thus quantifying their prevalence in intergenic regions (See methods).

We found that, despite being the least abundant, SINEs generally have a higher prevalence in both upstream and downstream regions of genes, contrasting with other

orders of TEs (Figures 2A and 3A). Notably, when visualizing these results at a more particular level in the classification of TEs, the tRNA-V superfamily appears as the one with the highest Jaccard indexes, and thus, it is the most prevalent superfamily in the vicinity of genes. For instance, according to our results in upstream regions, it is more common for a gene to have at least one tRNA-V element (~10%), rather than one from any other retrotransposon (~1%).

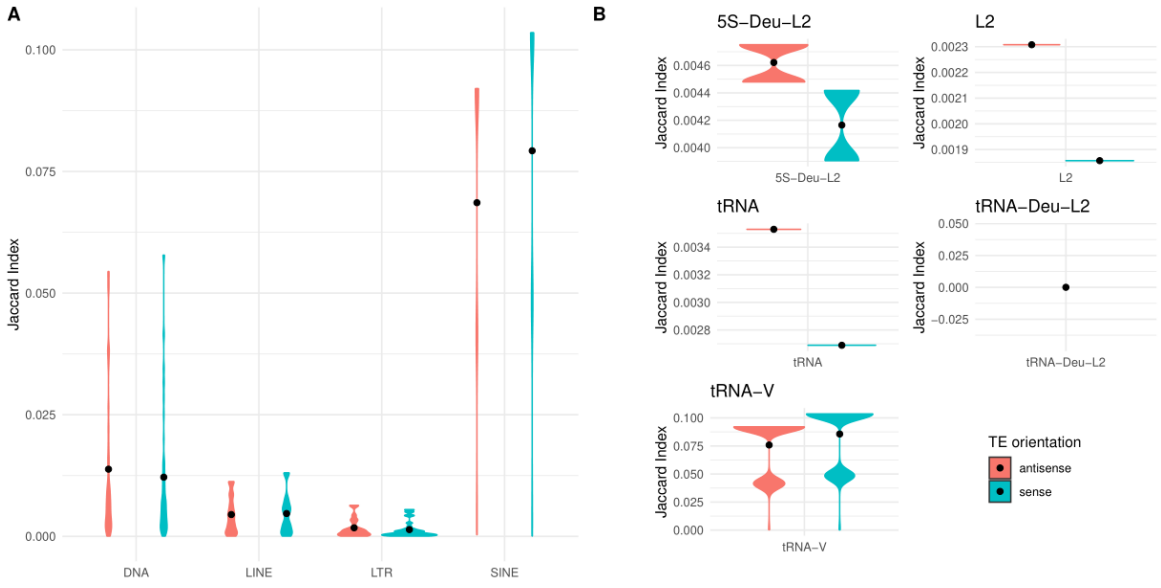


Figure 2: tRNA-V SINEs are more frequently found in upstream regions of genes in comparison to other orders of TEs. (A) The distribution of frequencies for each family of TEs in upstream regions of genes grouped by order and strand; (B) The distribution of frequencies for each family of TEs in upstream regions of genes for all SINE superfamilies by strand.

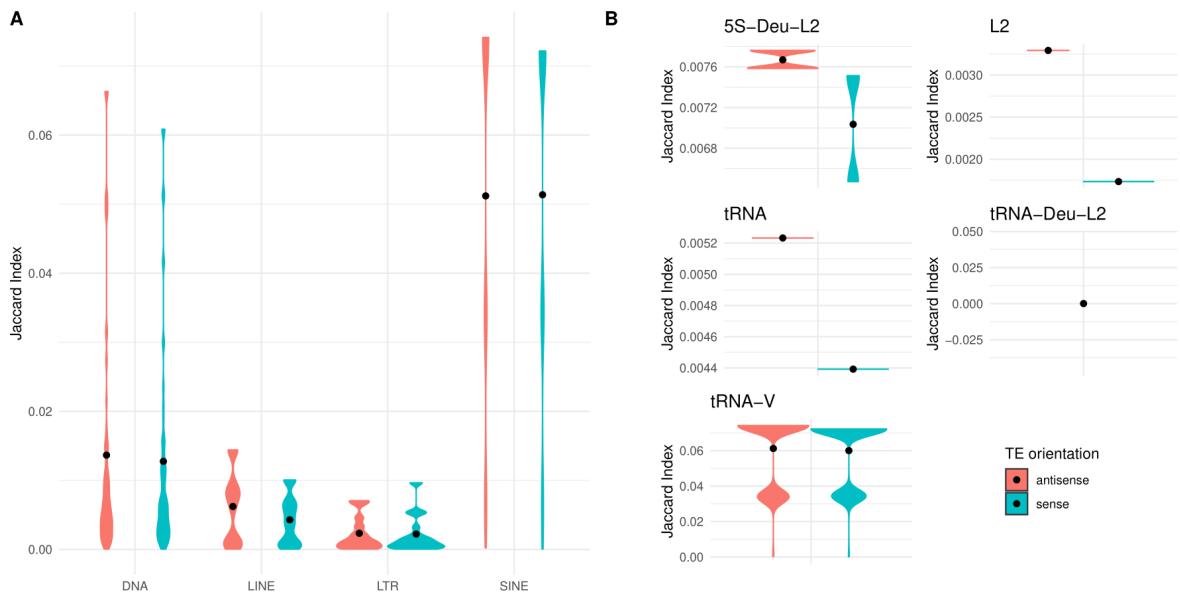


Figure 3: tRNA-V SINEs are more frequently found in downstream regions of genes in comparison to other orders of TEs. (A) The distribution of frequencies for each family of TEs in downstream regions of genes grouped by order and strand; (B) The distribution of frequencies for each family of TEs in downstream regions of genes for all SINE superfamilies by strand.

In conclusion, our results reveal an insertional bias for SINEs, and particularly for the tRNA-V superfamily, to be found close to promoter regions and 3'-UTRs, contrasting with other orders of TEs in the genome of zebrafish. This finding, together with literature linking tRNA and tRNA-like genes with insulator activities (See Discussion), prompted us to focus our analyses on tRNA-V insertions, with a particular interest in those located in the vicinity of genes.

Multiple groups of tRNA-V elements show characteristic arrangements and co-occupancy by CTCF, Rad21, and H3K27me3

To evaluate the participation of elements from the tRNA-V superfamily in the assembly of insulator structures, we characterized tRNA-V elements located in the vicinity of genes in terms of their occupancy by architectural proteins (CTCF and Rad21), transcription factors associated with hypoxic stress (HIF1-alpha), and epigenetic marks associated with reduced transcription (H3K27me3). We performed all these analyzes using publicly available CHIP-seq data of non-treated or control individuals from multiple works (See Data sources section on Methods). Briefly, we pre-processed and mapped sequencing reads to the reference genome of zebrafish, filtering ambiguously mapped reads and PCR duplicates, and explored the sequencing coverage and depth across windows of 1 kb centered at tRNA-V insertions overlapping genes. As a result, we obtained heatmap representations depicting the occupancy of each factor and epigenetic modifications mentioned above at tRNA-V insertions (Figure 4).

This revealed that tRNA-V elements are rather heterogeneous regarding their occupancy by factors and epigenetic marks. For instance, while some *loci* show a mild enrichment close to the center of the window (the upper part of the heatmaps), others are mostly depleted of it (the lower part of the heatmaps) (Figure 4).

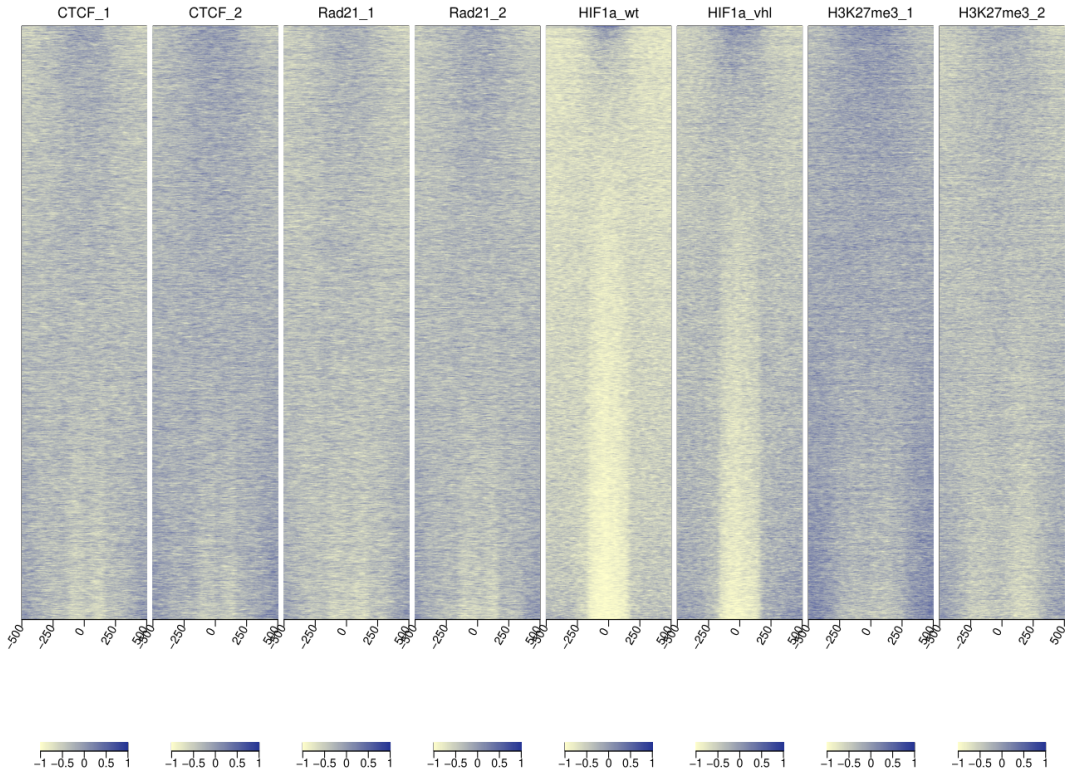


Figure 4: tRNA-V insertions are heterogeneous regarding their occupancy by factors and epigenetic marks. Heatmaps depicting the intensity of the CHIP-seq signal for architectural proteins (CTCF and Rad21), transcription factors (HIF1-alpha), and histone modifications (H3K27me3) across 1 kb windows centered at tRNA-V fragments overlapping genes.

Next, in order to explore any underlying structure in the occupancy data at tRNA-V elements, we performed hierarchical clustering using the "*ward.D2*" method in R. We obtained 27 discrete groups of tRNA-V insertions characterized for having characteristic

patterns of occupancy for CTCF, Rad21, and HIF1-alpha at different sites relative to the insertion (Figures 5 and 6). Interestingly, several of these groups (e.g. clusters 1, 6, 7, and 8) show co-occupancy between CTCF, and the subunit of cohesin, Rad21, something that has been previously associated with the formation of chromatin loops (Hansen et al., 2017).

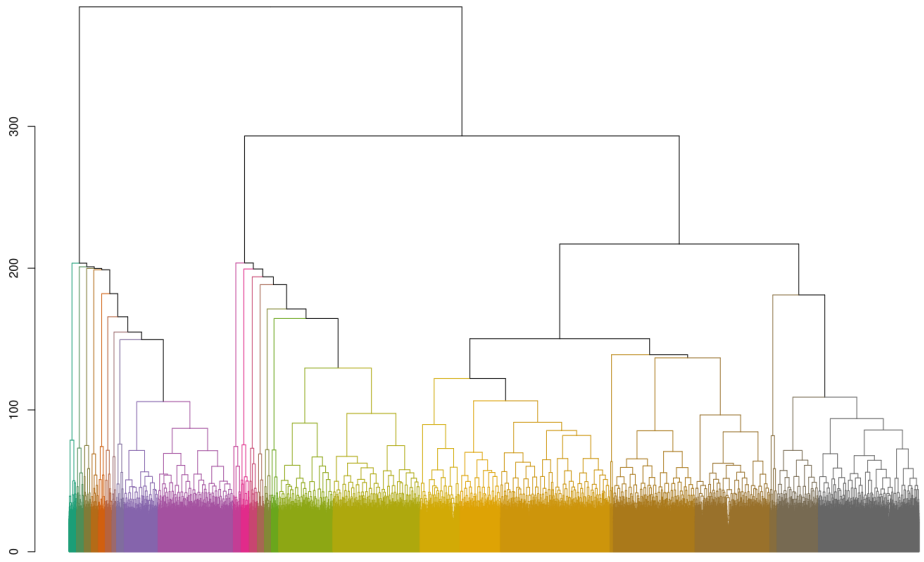


Figure 5: Cladogram depicting the relationships resulting from the hierarchical clustering performed over the CHIP-seq data.

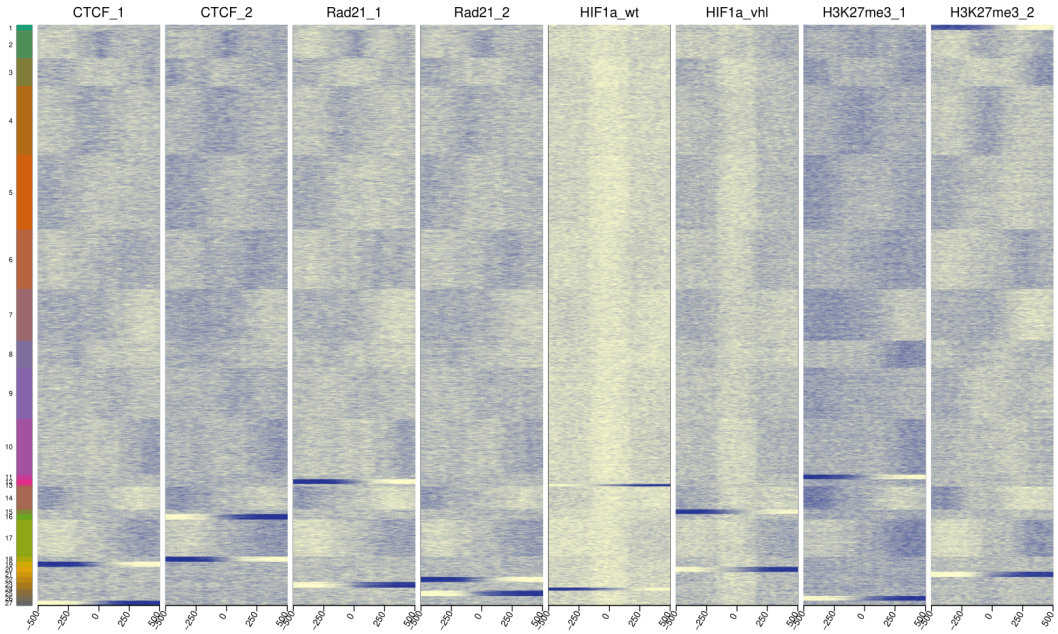


Figure 6: Groups of tRNA-V insertions are characterized by having distinctive patterns of occupancy for CTCF, Rad21, HIF1-alpha, and H3K27me3. Heatmaps depicting the intensity of the CHIP-seq signal for architectural proteins (CTCF and Rad21), transcription factors (HIF1-alpha), and histone modifications (H3K27me3) across 1 kb windows centered at tRNA-V fragments overlapping genes. Colored boxes represent the groups identified by hierarchical clustering.

In addition, we described the identified clusters in terms of their composition of families (Figure 7), fragment size, and kimura distributions (Supplementary Figures 2 and 3). This description revealed that all clusters have a similar and heterogeneous composition

of families, with HE1_DR1 being the most abundant in all of them. Similarly, all families showed distinctive fragment size and Kimura distributions in all clusters. The most notable differences were observed in low abundance tRNA-V families, such as DANA and SINE_TE.

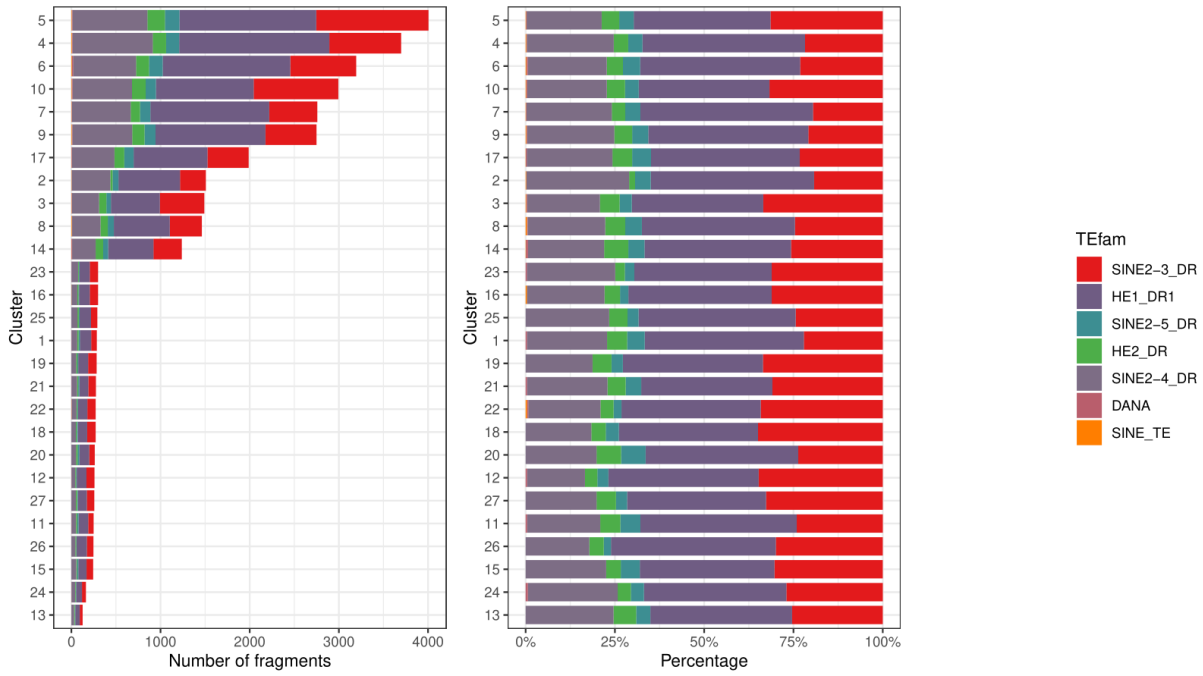


Figure 7: The identified clusters have a heterogeneous, yet similar, composition of tRNA-V families. Absolute (left) and proportional abundance (right) of fragments for each family of tRNA-V grouped by the cluster they belong to.

Finally, in order to simplify and summarize the functional implications of the identified clusters, we label them as "*H3K27me3 upstream*", "*H3K27me3 downstream*", "*H3K27me3 center*" and "*H3K27me3 not-center*", based on their pattern of occupancy for this histone modification.

Expression data support insulator activities in three groups of tRNA-V insertions

The clusters identified in the previous section were useful as they gave us the means to group tRNA-V insertions accordingly to the presence of particular epigenetic modifications and architectural proteins commonly associated with insulators. However, from a functional point of view, this information alone does not confirm an insulator activity. Hence, to evaluate the insulator activity of tRNA-V elements in a functional context, we compared the median expression of genes located upstream and downstream of tRNA-V insertions, based on RNA-seq data from normoxic and hypoxic individuals, independently.

Our results revealed that globally speaking, most clusters do not show the expected bias attributable to insulators when contrasted with the expression levels of their neighbor genes (Supplementary Figure 4). For instance, in cluster 6, the highest median

expression occurs downstream of the tRNA-V insertion, yet that region shows H3K27me3 occupancy, which is a histone modification often associated with silent chromatin. This may be due to the presence of false positives in the clusters, which could dilute the global contribution of true positive loci. Despite this, we were able to identify three clusters (5,7, and 10) whose global distributions behave in coherence with a model where the tRNA-V insertion promotes the assembly of insulator structures. Consistently, in most samples, these three clusters display their lowest median expression in the direction that overlaps regions with H3K27me3 occupancy (Figures 8-10).

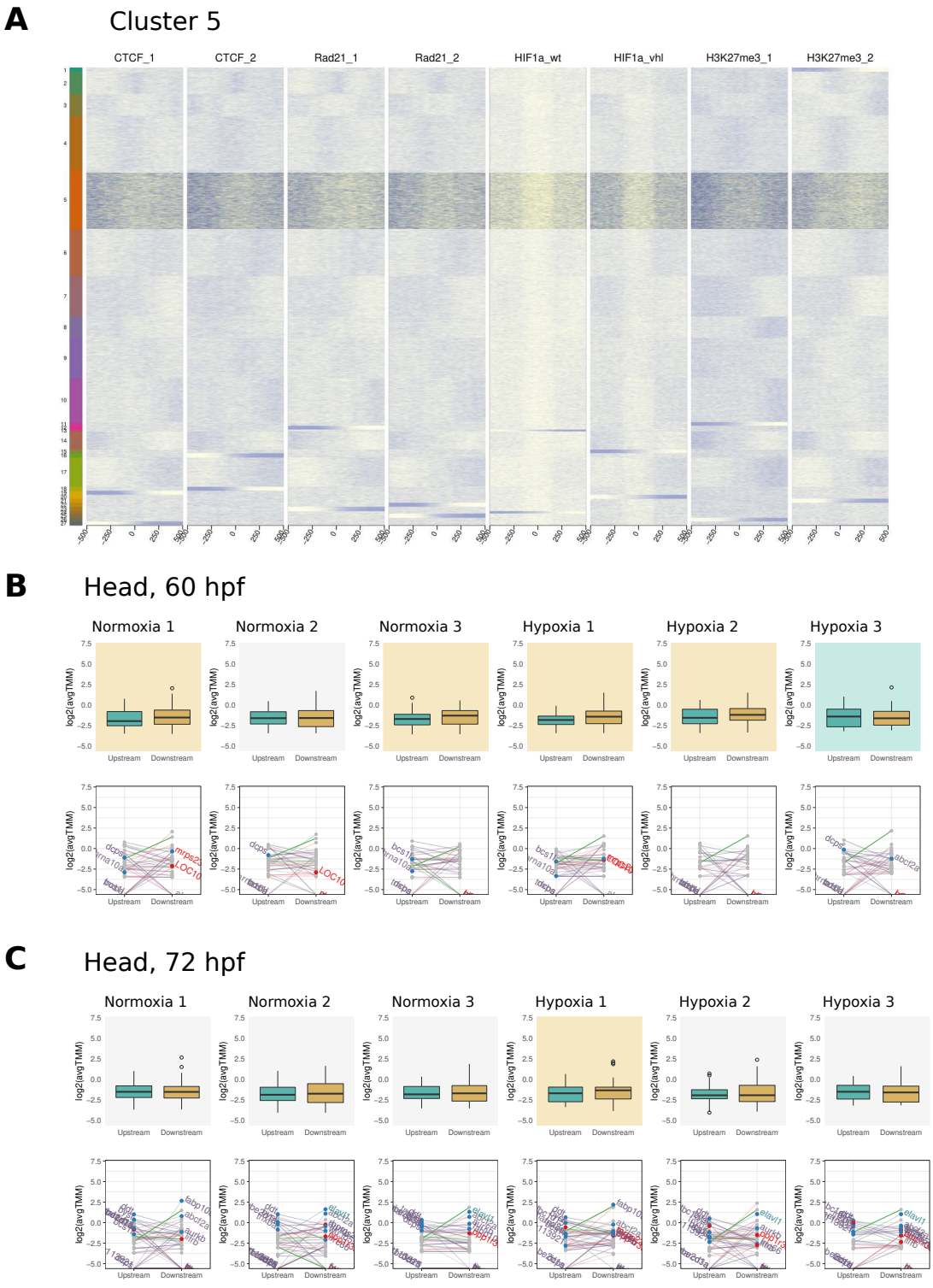


Figure 8: Gene expression levels upstream and downstream of tRNA-V insertions for cluster 5. (A) Clustered heatmap highlighting cluster 5. (B) Expression levels of genes

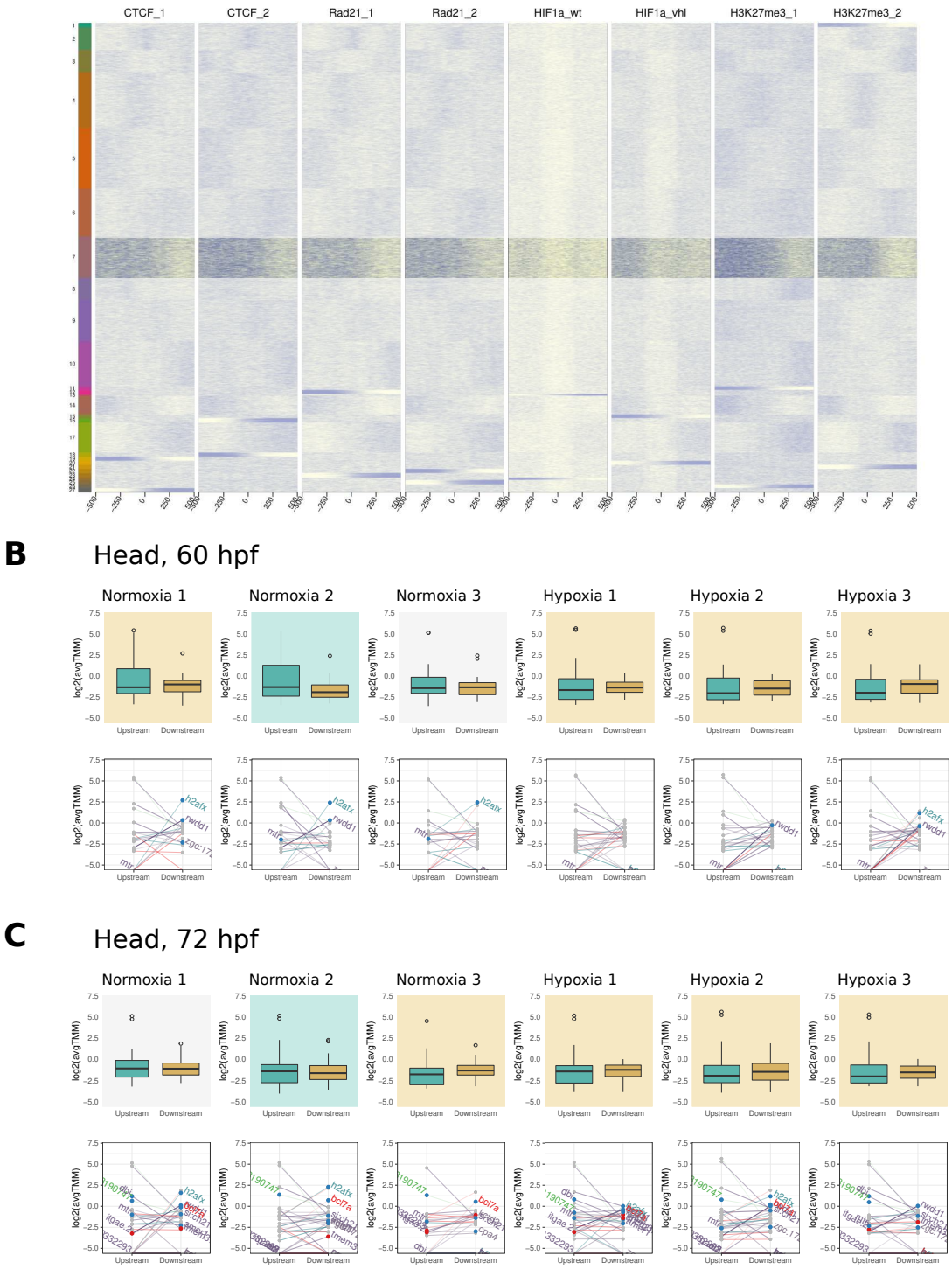


Figure 9: Gene expression levels upstream and downstream of tRNA-V insertions for cluster 7. (A) Clustered heatmap highlighting cluster 7. (B) Expression levels of genes

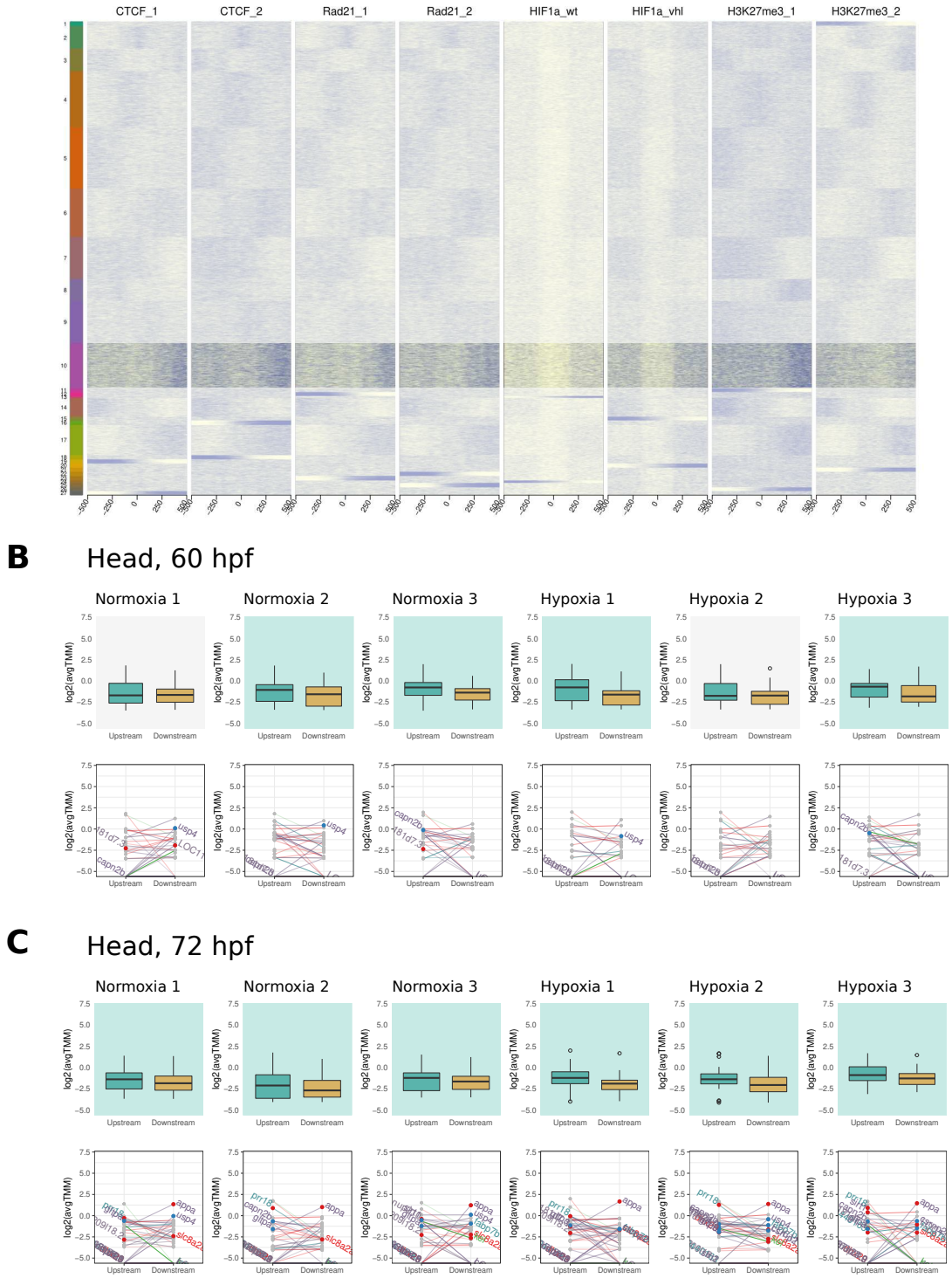


Figure 10: Gene expression levels upstream and downstream of tRNA-V insertions for cluster 10. (A) Clustered heatmap highlighting cluster 10. (B) Expression levels of

(continuation of caption for Figure 8)

upstream and downstream of tRNA-V elements belonging to cluster 5 in samples corresponding to head tissue of 60 hpf zebrafish embryos. The upper and lower pannels depict global and individual gene expression levels, respectively. Only DE genes are named. (C) Same as (B), but from head tissue of 72 hpf zebrafish embryos. Note that the background is colored based on the region with the highest median log gene expression.

(continuation of caption for Figure 9)

upstream and downstream of tRNA-V elements belonging to cluster 7 in samples corresponding to head tissue of 60 hpf zebrafish embryos. The upper and lower pannels depict global and individual gene expression levels, respectively. Only DE genes are named. (C) Same as (B), but from head tissue of 72 hpf zebrafish embryos. Note that the background is colored based on the region with the highest median log gene expression.

(continuation of caption for Figure 10)

genes upstream and downstream of tRNA-V elements belonging to cluster 10 in samples corresponding to head tissue of 60 hpf zebrafish embryos. The upper and lower pannels depict global and individual gene expression levels, respectively. Only DE genes are named. (C) Same as (B), but from head tissue of 72 hpf zebrafish embryos. Note that the background is colored based on the region with the highest median log gene expression.

The putative insulator activity of specific tRNA-V insertions may be affected by hypoxic stress

With the aim of detecting *loci* where the insulator activity of tRNA-V elements may be affected by stress, we performed differential expression analyses between normoxic and hypoxic conditions focusing on differentially expressed genes located near tRNA-V elements identified as putative insulators. Besides, we analyzed the obtained set of genes in the STRING database and explored their genomic context using the Gviz package in R. Tables 1-4 (Appendix) and Figure 11 summarizes these results for all pairs of contiguous genes close to tRNA-V elements, where at least one is DE in hypoxia.

Figure 11: Multiple DE genes with tRNA-V insertions on their vicinities have experimental evidence supporting protein interactions between them. Network of interactions for DE genes having tRNA-V elements labeled as putative insulators on their vicinities, based on data from the STRING database. Gene names colored in red are upregulated during hypoxia, while those in blue are downregulated.

DISCUSSION

Elements of the tRNA-V superfamily may act as insulators in zebrafish

In this work we have reported evidence that tRNA-V elements can act as insulators in the genome of zebrafish. Insulators are elements that allow the assembly of chromatin structures that have two characteristic features. First, they act as a chromatin barrier blocking the spreading of epigenetic marks associated with silent chromatin. Second, they are capable of blocking enhancer-promoter interactions, thus affecting gene expression. These activities are the result of complex interactions of chromatin and architectural factors, such as CTCF and cohesin (Matharu & Ahanger, 2015). Indeed, CTCF and cohesin are key components of the loop extrusion complex (LEC), a piece of molecular machinery known to participate in the formation of chromatin loops. Briefly, CTCF recruits cohesin to CTCF binding sites (Rubio et al., 2008; Bowers et al., 2009), and therefore, they tend to occupy near or overlapping positions in the genome (Nora et al., 2017). This is consistent with our results, where we identified groups of tRNA-V elements with different patterns of occupancy for several factors. Notably, most of these groups showed co-occupancy between CTCF, Rad21 (subunit of cohesin), and H3K27me3, but not with HIF1-alpha. Indeed, tRNA-V elements appear to be mostly

depleted of HIF1-alpha. This may be explained by a sterical impediment due to the presence of another structure involving CTCF and cohesin, such as the LEC.

Additionally, we found that some tRNA-V elements are located at transition sites of CTCF, Rad21, and H3K27me3 signals. H3K27me3 is an epigenetic modification associated with silent chromatin commonly found next to domain boundaries (Cuddapah et al., 2009; Cai et al., 2021). This suggests that tRNA-V elements may have chromatin barrier activity, participating in the delimitation of chromatin domains. Interestingly, in three of the groups identified in this work (clusters 5, 7 y 10), the region with the lowest median gene expression coincides with regions having H3K27me3 occupancy, supporting insulator functions for at least a population of tRNA-V elements. It must be noted, however, that in this case, we cannot discriminate if the observed tendency is the result of chromatin barrier or enhancer blocking activities. In summary, we propose a model where some tRNA-V elements may gain context-dependent insulator functions, showing signatures of chromatin barrier and enhancer-blocking activities, and thus, likely to contribute to the regulation of gene expression (Figure 12).

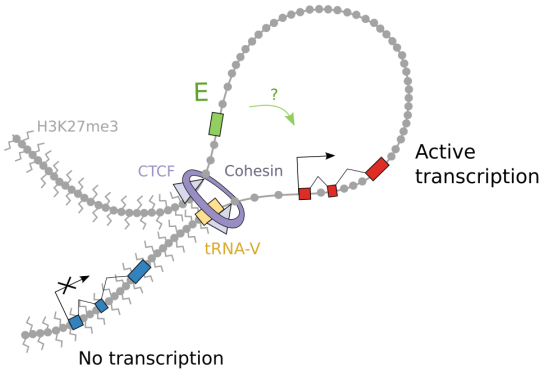


Figure 12: Model of a tRNA-V element with insulator activities.

On the other hand, our results also show a considerable proportion of tRNA-V elements which do not have the expected pattern of occupancy regarding the gene expression data. Two non-exclusive hypotheses can explain this observation. First, if the number of elements functioning as actual insulators is small, then their global contribution to median gene expression levels may become diluted. Second, we cannot reject that CTCF or Rad21 may have context-specific functions, not necessarily related to insulation activities. For example, in mammals there is an enrichment of CTCF at boundary regions, however, this represents only 15% of all CTCF binding sites in the genome (Dixon et al., 2012); Moreover, in zebrafish, despite being fundamental for chromatin

structure, CTCF is not even enriched at boundary regions (Franke et al, 2021; Pérez-Rico et al., 2020).

The insulator activity of tRNA-V elements may be associated with their tRNA-related features

This is not the first time that tRNA-related elements are found to have insulation functions. Currently, most of the evidence comes from studies on yeast and mammals. These studies remark on the involvement of the transcriptional machinery in the insulator activity of tRNA-related elements. For instance, tRNA genes, which are transcribed by RNA polymerase III, are also known as insulators in humans (Raab et al., 2012). This is not a coincidence since multiple reports have pointed to elements within the Pol III promoter as responsible for insulation (Bortle & Corces, 2012). More specifically, Pol III-mediated transcription requires the binding of TFIIIC to sites known as A and B boxes, which allow the assembly of the Pol III pre-initiation complex (Cieśla et al., 2018). Notably, TFIIIC has also been observed to exert insulator activities (Simms et al., 2008). Moreover, it has been shown that mutations in the B box impair both, the barrier function of tRNA genes, and the binding of the TFIIIC (Donze & Kamakaka, 2001).

Although we did not account for data directly supporting the involvement of the Pol III machinery in the insulator activity of tRNA-V elements, some antecedents push us to speculate about it. First, like tRNA genes, tRNA-V elements are thought to be transcribed by Pol III. This is supported by their structure. From 5' to 3' they contain, first, a tRNA-related domain with a Pol III promoter; then a highly conserved region almost identical in all tRNA-V families; and finally, a highly recombinogenic region containing a LINE-related domain relevant for SINEs retrotransposition (Ogiwara et al., 2002; Kramerov & Vassetzky, 2011). This is the general structure of tRNA-V elements, however, some families show internal redundancy of tRNA-related domains. Such is the case of the DANA family, which contains three conserved regions where two of them possess A and B boxes (Izsvák et al., 1996). This is interesting since multimerization of tRNA genes (or their promoters) has been observed to increase the strength of weak insulators in humans, chickens, and yeast (Kirkland et al., 2013; Raab et al., 2012). Notably, we found that DANA is the second tRNA-V family with the highest Kimura values, reflecting that some insertions may be subject to evolutive pressures.

Additionally, our results show that tRNA-V elements are the most widespread TEs in the vicinity of genes, with a slight bias towards upstream regions. This is consistent with reports in humans, where expressed Pol III genes are usually found near functional Pol II promoters (Moqtaderi et al., 2010). Moreover, some works have already reported insulator functions involving both Pol II and Pol III machinery (Donze & Kamakaka,

2001). For example, murine B1-X35S SINEs have been shown to require Pol II and Pol III mediated transcription in order to exert their insulator activity. Specifically, the binding of the AHR factor to B1-X35S elements triggers an exchange mechanism that recruits Pol II while releasing Pol III (Roman et al, 2011), resulting in an enhanced insulator function. Altogether, this evidence highlights the potential involvement of Pol II and Pol III machinery in the insulator activity of tRNA-V elements, however, further research will be needed in order to understand the nature of this relationship.

The insulator activity of tRNA-V may change upon stress

Few works have addressed whether insulators change their activity during stress. Only a decade ago the first report on this phenomenon was published. Tiana et al. found that an insulator element was responsible for the differential regulation of the pair of contiguous genes, *gys1,* and *ruvbl2*, during hypoxic stress (Tiana et al, 2012). These genes are in the opposite orientation and share a bidirectional ~700 bp promoter containing a hypoxia-responsive element (HRE). Interestingly, the authors found that the binding of HIF1-alpha triggers the upregulation of *gys1*, but not *ruvbl2*.

As discussed in the previous section, this is not the only example highlighting the relevance of promoter regions for insulation, regardless, it is one of the few in the subject of stress, and to our knowledge, the only one in hypoxia. A piece of

complementary evidence comes from other types of stress, for instance, studies in human cells show that heat stress induces distal changes in chromatin architecture, affecting enhancer-promoter interactions and thus gene expression. Notably, the authors of this work highlight, a "pre-wired" regulation defined by the pause-release of promoter-proximal Pol II (Vihervaara et al., 2017).

Considering these antecedents, we defined two criteria for detecting *loci* affected by the activity of insulators. We reasoned that (*i*), if there is any pair of contiguous genes where at least one of them is DE during hypoxia, and (*ii*), if there is a tRNA-V element with insulator characteristics in their vicinity, and particularly near promoter regions, then, one possibility is that the observed differential regulation may be a consequence of changes in the insulator function of the tRNA-V element. In order to test this idea, we performed a screening using publicly available RNA-seq data from hypoxic experiments in zebrafish, together with CHIP-seq data from normoxic individuals. In particular, we identified DE genes located in the vicinity of tRNA-V elements with insulator characteristics and compared these results with data of H3K27me3 (associated with silent transcription) and H3K4me3 (associated with active transcription and promoters) occupancy. This allowed us to evaluate two different models of insulator-mediated changes in gene expression. Nonetheless, it must be noted that other regulatory pathways may be exerting influence in the regulation of transcription induced by

hypoxia. For this reason, we will focus the discussion on a few *loci* with a reasonable amount of evidence to be considered bonafide insulators.

The first model accounts for situations where a gene is found up-regulated during hypoxia while located in a transcriptionally silent region in normoxic individuals (Figure 17A). Our results reveal that this is the case of the pair of contiguous genes *zyx*/*kel*, which codes for "Zyxin" and "Kell Metallo-Endopeptidase", respectively (Figures 13 and 14). On this *locus,* there is a tRNA-V element with a relatively high Kimura distance (24.47) near the *kel* promoter. Notably, *kel* is upregulated in hypoxia while the gene upstream of the insertion, *zyx*, does not report significant changes. This is in contrast with a transcriptionally silent chromatin state of *kel* in normoxic individuals, as inferred by the lack of H3K4me3 occupancy over its promoter and a slightly higher level of H3K27me3 occupancy downstream of the tRNA-V insertion.

The second model exposes a similar situation, but with the difference that here the discrepancy occurs at genes found downregulated during hypoxia, while located in a transcriptionally active region (Figure 17B). For instance, the pair of genes *trappc6b*/*tpx2*, which codes for "Trafficking Protein Particle Complex Subunit 6B" and "TPX2 Microtubule Nucleation Factor" respectively, are located in the vicinity of the tRNA-V element with the highest Kimura distance (40.14) in our set (Figures 15 and 16). At this locus, the *tpx2* gene is downregulated during hypoxia, while *trappc6b* does

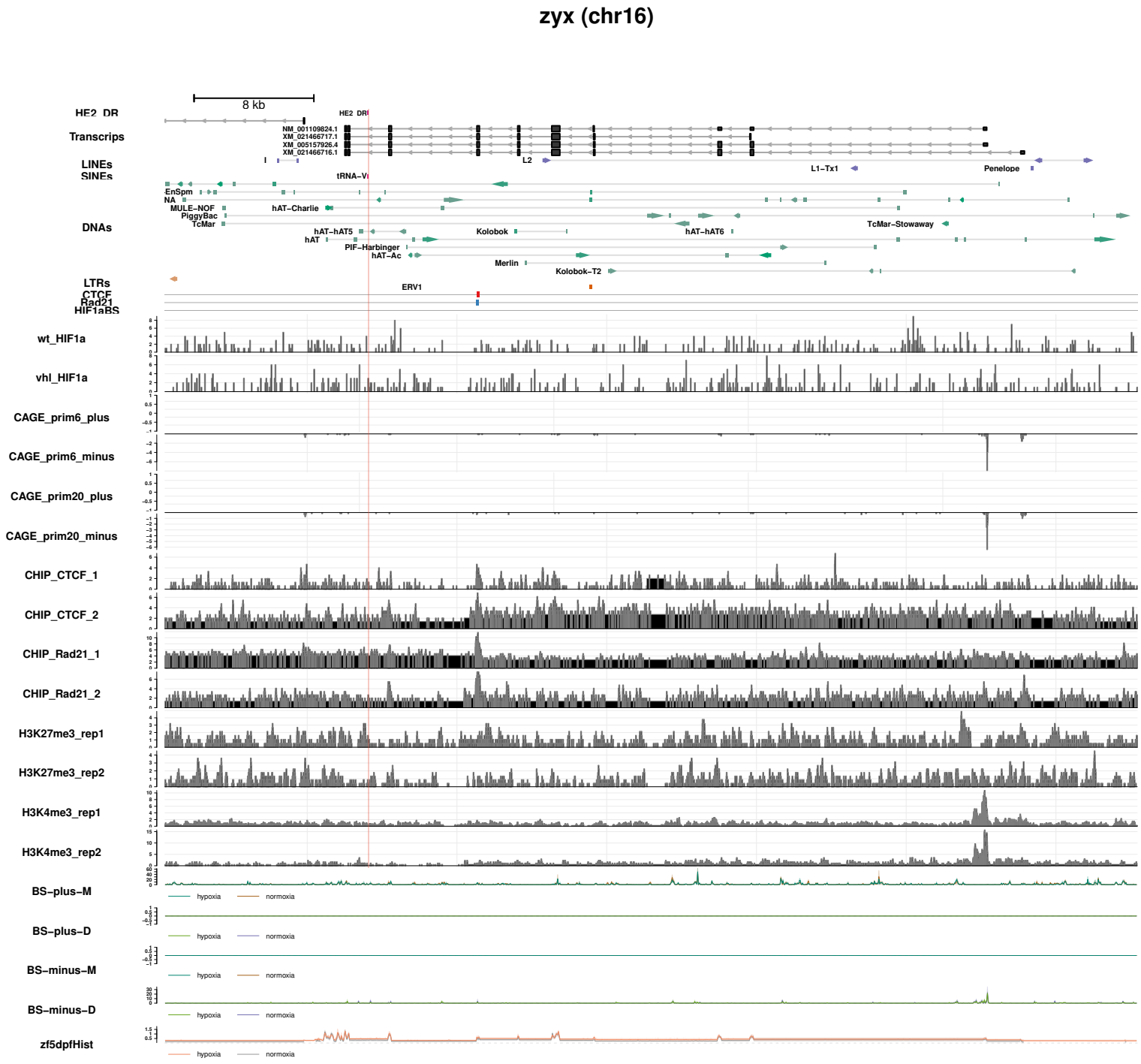


Figure 13: Genomic context of the *zyx*/*kel locus* centered at the *zyx* gene. The tRNA-V element is highlighted with a vertical red bar.

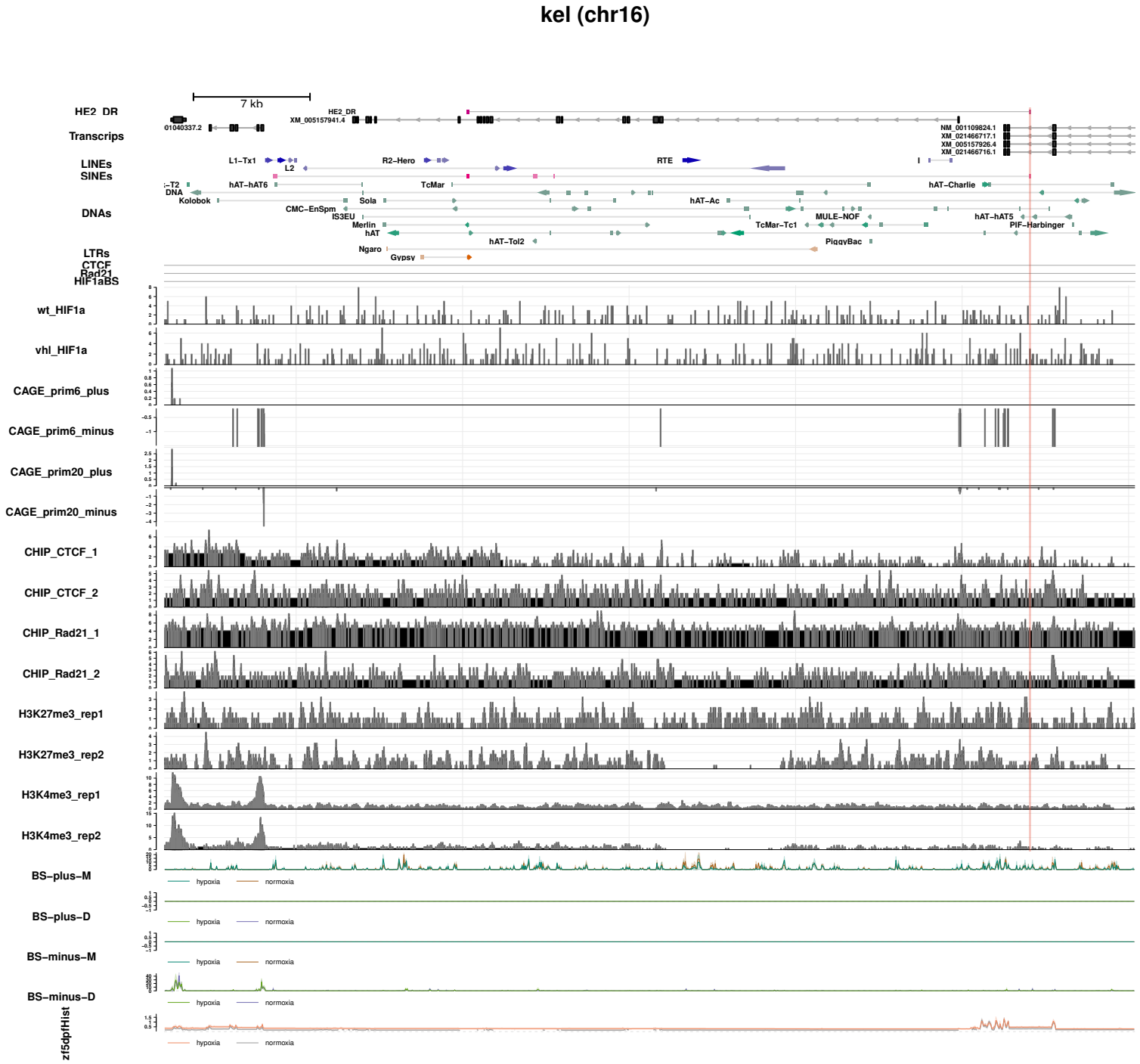


Figure 14: Genomic context of the *zyx/kel* locus centered at the *kel* gene. The tRNA-V element is highlighted with a vertical red bar.

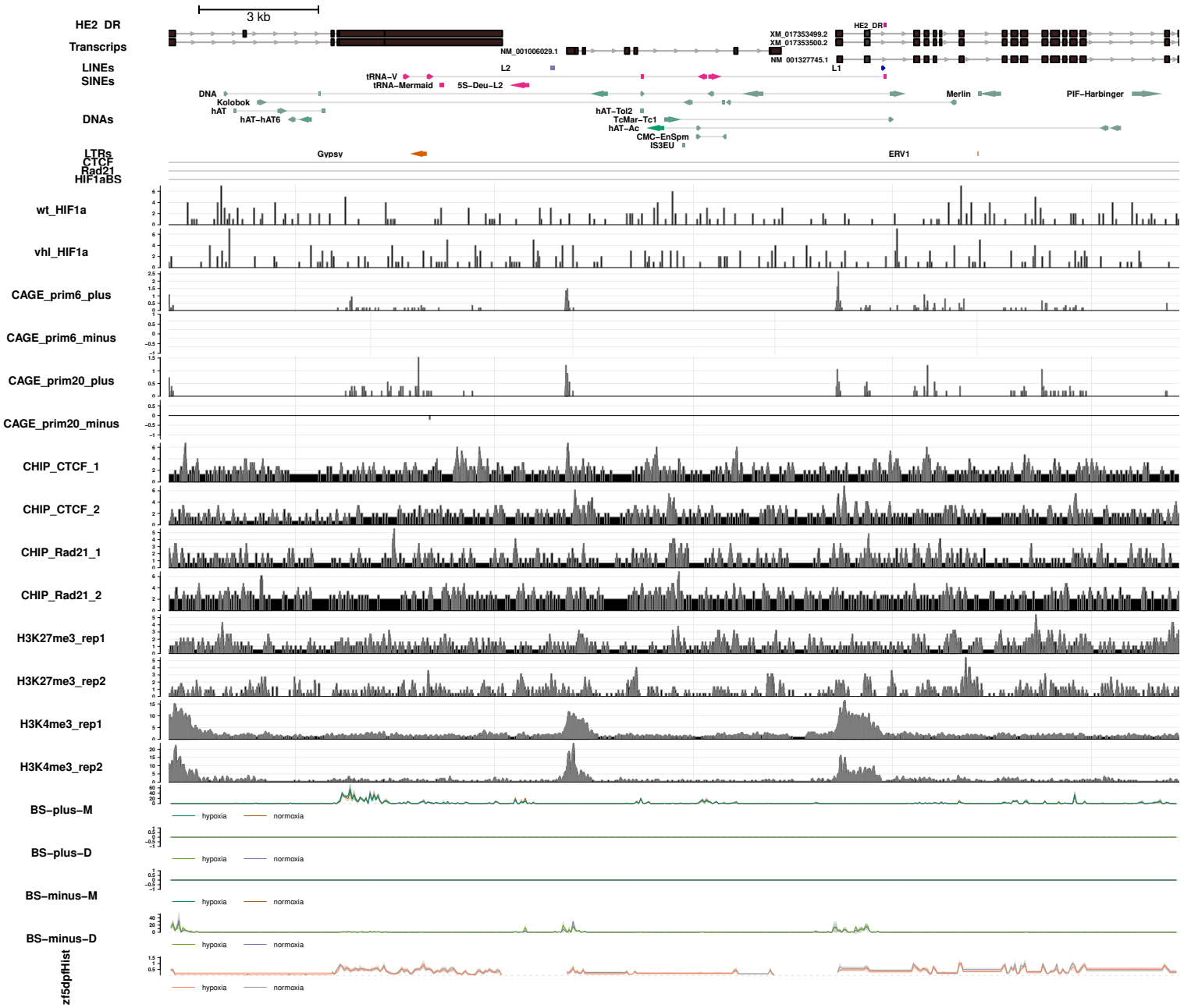


Figure 15: Genomic context of the *trappc6b*/*tpx2 locus* centered at the *trappc6b* gene. The tRNA-V element is highlighted with a vertical red bar.

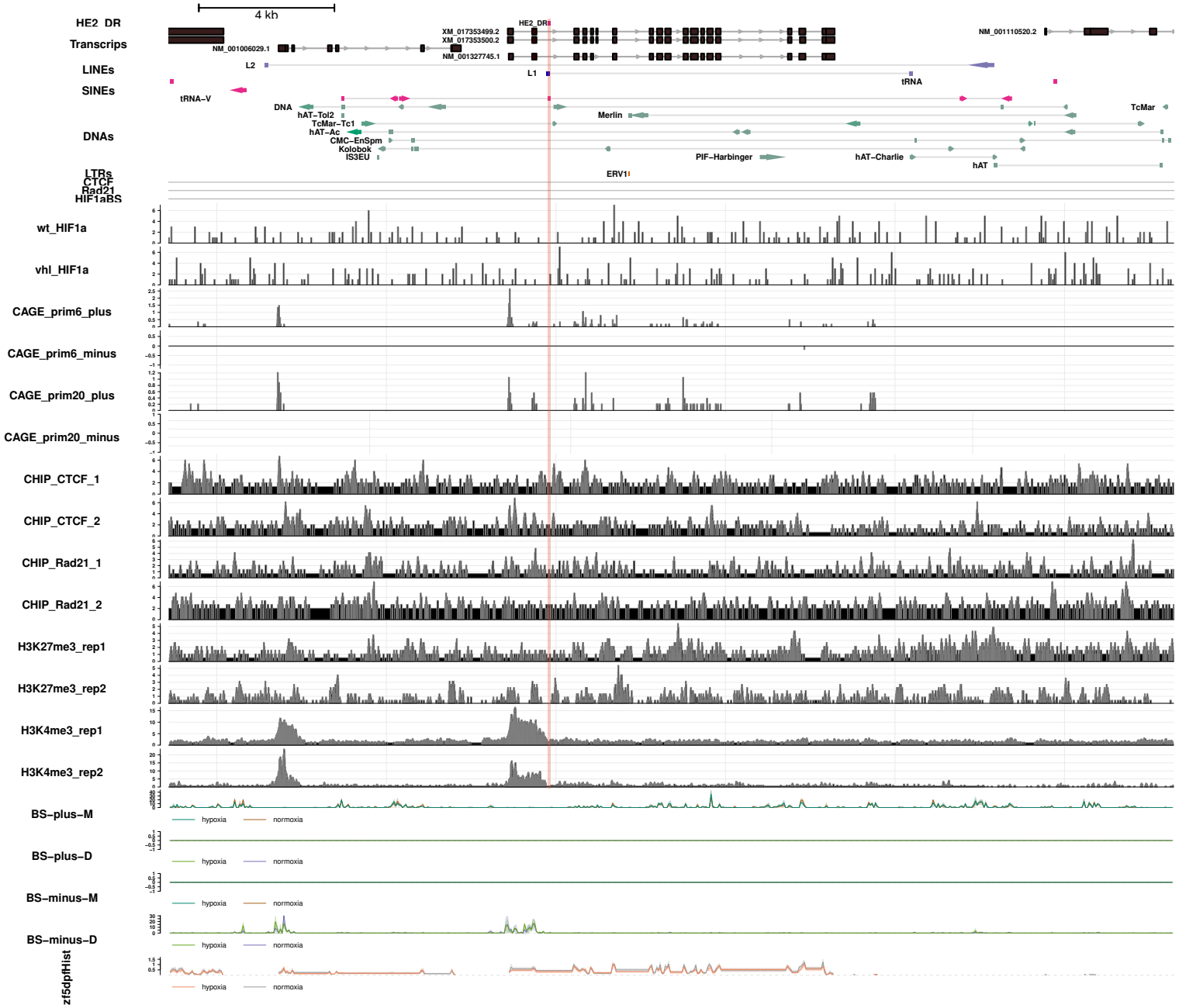


Figure 16: Genomic context of the *trappc6b*/*tpx2 locus* centered at the *tpx2* gene. The tRNA-V element is highlighted with a vertical red bar.

not show significant changes in gene expression. This contrasts with the H3K4me3 occupancy data, which suggests that both genes are transcriptionally active in normal conditions. More precisely, the H3K4me3 occupancy at the *tpx2* promoter, together with the absence of the H3K4me3 signal in the promoter of the gene immediately downstream of *tpx2,* points to the existence of an insulator element between both genes, likely at the tRNA-V insertion.

We propose that these discrepancies between the expression and histone modification occupancy data may be explained by changes in chromatin loops induced by hypoxia, which are not visible in the H3K27me3 or H3K4me3 data since their corresponding samples come from normoxic individuals. Altogether, our results suggest that the dynamic recognition (or omission) of particular tRNA-V elements as CTCF binding sites induced by stress may promote the displacement of loop boundaries, as a result of the extrusion of the chromatin fiber mediated by cohesin (Figure 17). This might "slide" a gene promoter from a heterochromatin region to a transcriptionally active domain, thus increasing its expression levels (as in the case of *kel*) (Figure 17A); Or conversely, it may "slide" the gene promoter towards a heterochromatic domain, thus decreasing its expression levels (as in the case of *tpx2*) (Figure 17B). Notably, most genes in our analysis were found to better fit the second model. Similar mechanisms have been suggested in recent years (Rowley et al., 2018), and are supported by the dynamic nature of the LEC. For instance, single-molecule imagining experiments have shown that the

residence and rebinding time is notoriously lower in CTCF than in cohesin (~1-2 min vs ~22 min, and ~1 min vs ~33 min, respectively) (Hansen et al., 2017), opening the possibility to more dynamic control of gene expression through changes in the boundaries of chromatin loops.

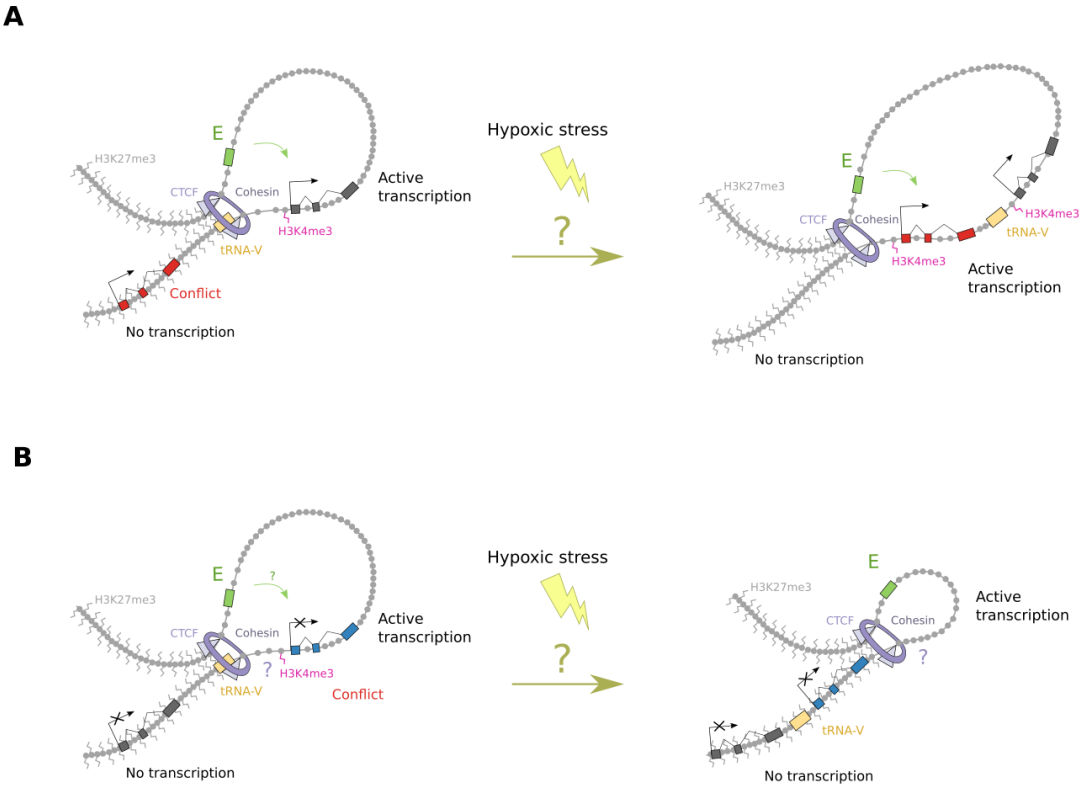


Figure 17: Implications of hypoxia in the proposed model of tRNA-V mediated insulation. (A) Model explaining upregulation of genes located on heterochromatin domains in normoxic samples. (B) Model explaining downregulation of genes located on transcriptionally active domains in normoxic samples.

Regarding the possibility of a coordinated transcriptional response to hypoxic stress, involving alterations in the putative insulator function of tRNA-V insertions, we cannot confirm or reject this hypothesis. However, some of our results appear to support this. Indeed, several downregulated genes with tRNA-V insertions on their vicinities have experimental evidence supporting protein interactions between them in other organisms (Figure 11). For instance, there are experimentally determined protein interactions involving the products of *tpx2*, *aurkb*, *h2afx*, *nup210,* and *capn2b* genes, all of which we found downregulated in hypoxia. From this subset, *aurkb* (Figure 18) is the one with the major number of direct interactions, involving the products of *tpx2*, *h2afx,* and *nup210*. Accordingly, Tpx2 is a known co-factor of the product of *aurkb*, the Aurora B Kinase (Aurkb) (Iyer et al., 2012). Another example is the pair of contiguous genes, *capn1b,* and *capn2b,* that code for the Calpain 1 and Calpain 2 large subunits, respectively (Figures 19 and 20). These proteins are known to interact with Zyx, which in turn interacts with Aurkb. Moreover, there is a tRNA-V element near the promoter of *capn1b* (Kimura distance of 14.86; Cluster 10). Notably, we found that *capn2b* is downregulated in hypoxia, but not *capn1b.* This is consistent with expression data from previous works showing that *capn2b* and *capn1b* have tissue-specific patterns of gene expression through development (Lepage et al., 2008), supporting the existence of an insulator element at this *locus*. Notably, both Aurkb and Calpain have been observed to

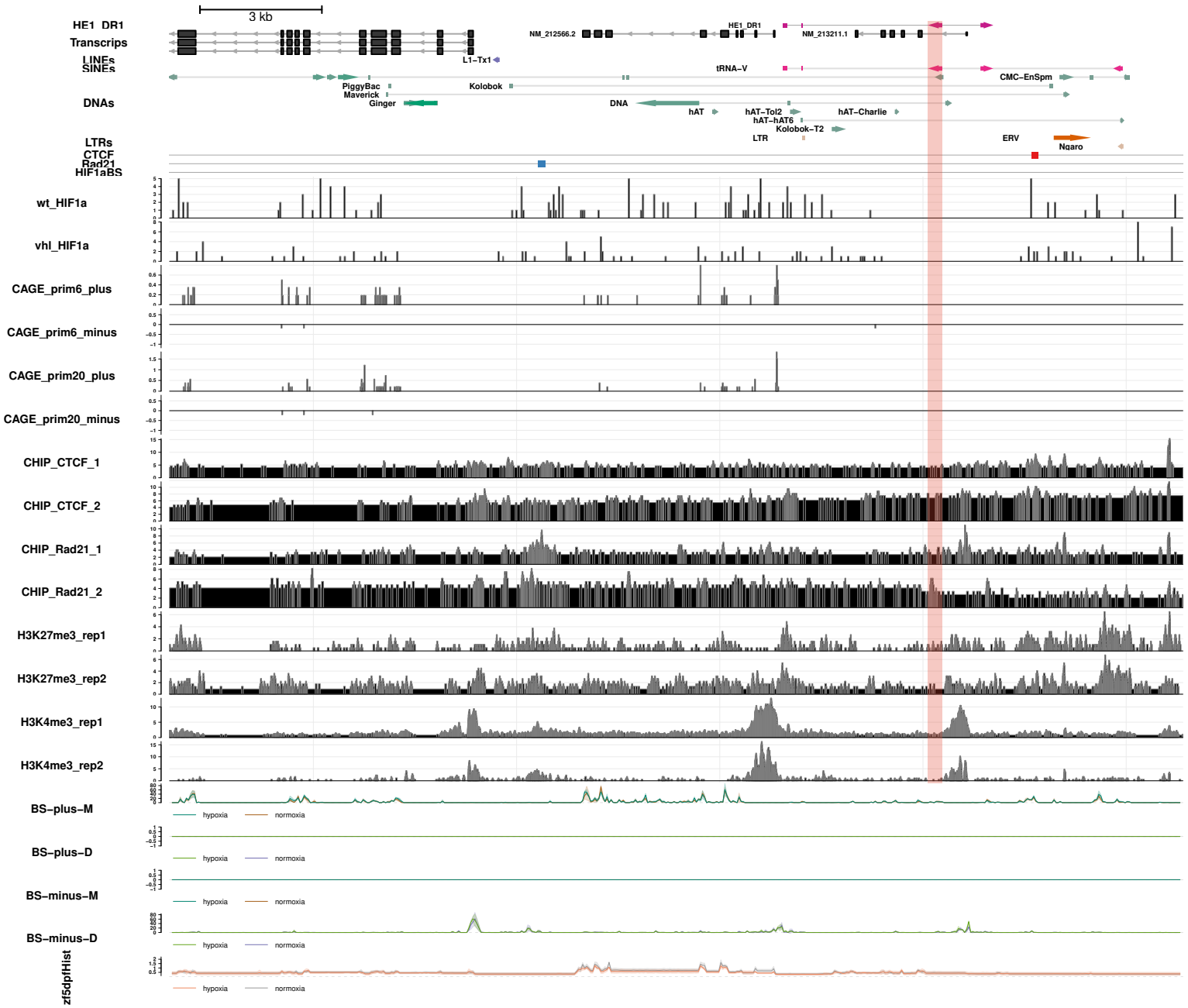


Figure 18: Genomic context of the *aurkb*/*tmem107 locus* centered at the *aurkb* gene. The tRNA-V element is highlighted with a vertical red bar.

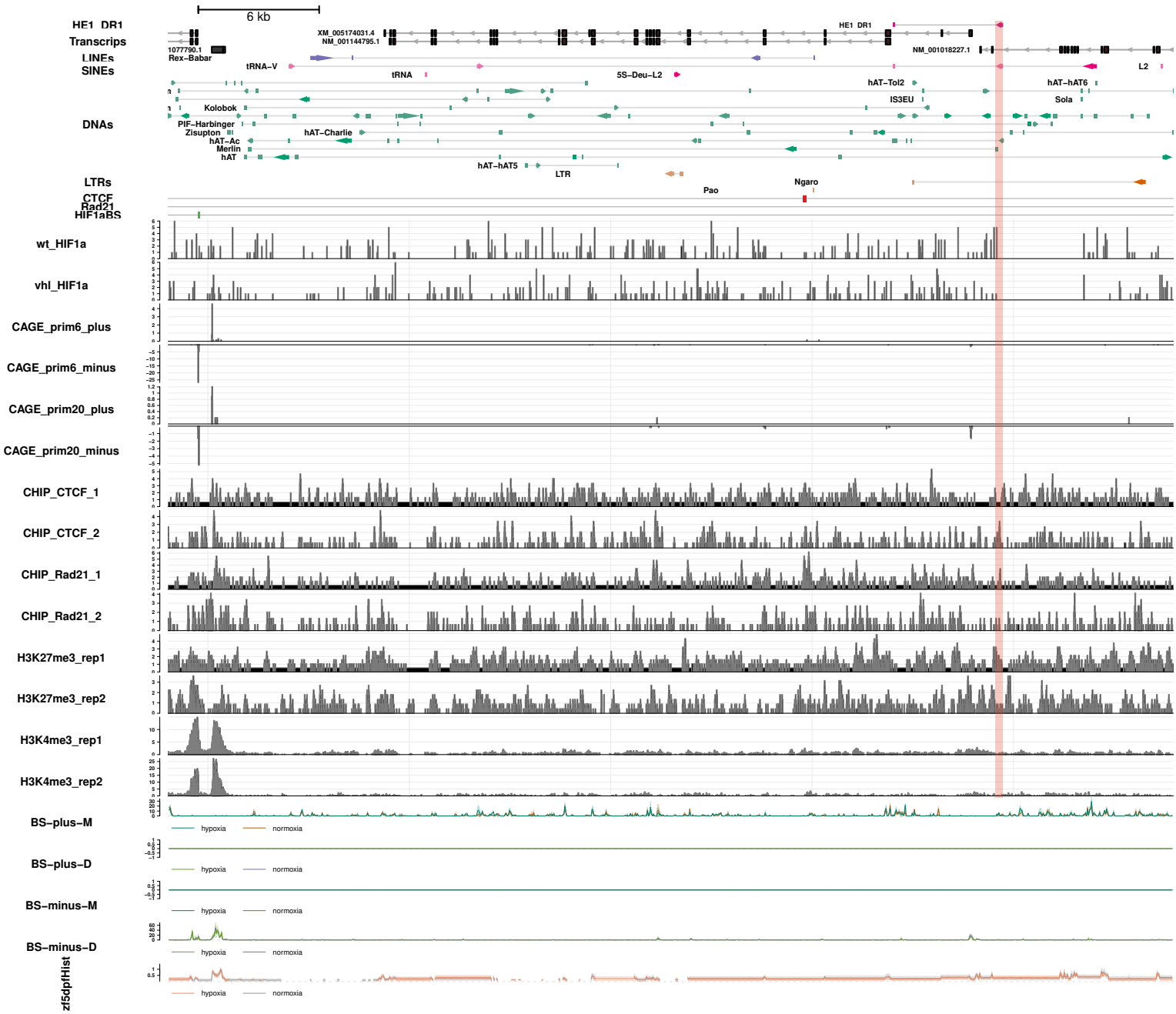


Figure 19: Genomic context of the *capn2b*/*capn1b locus* centered at the *capn1b* gene. The tRNA-V element is highlighted with a vertical red bar.

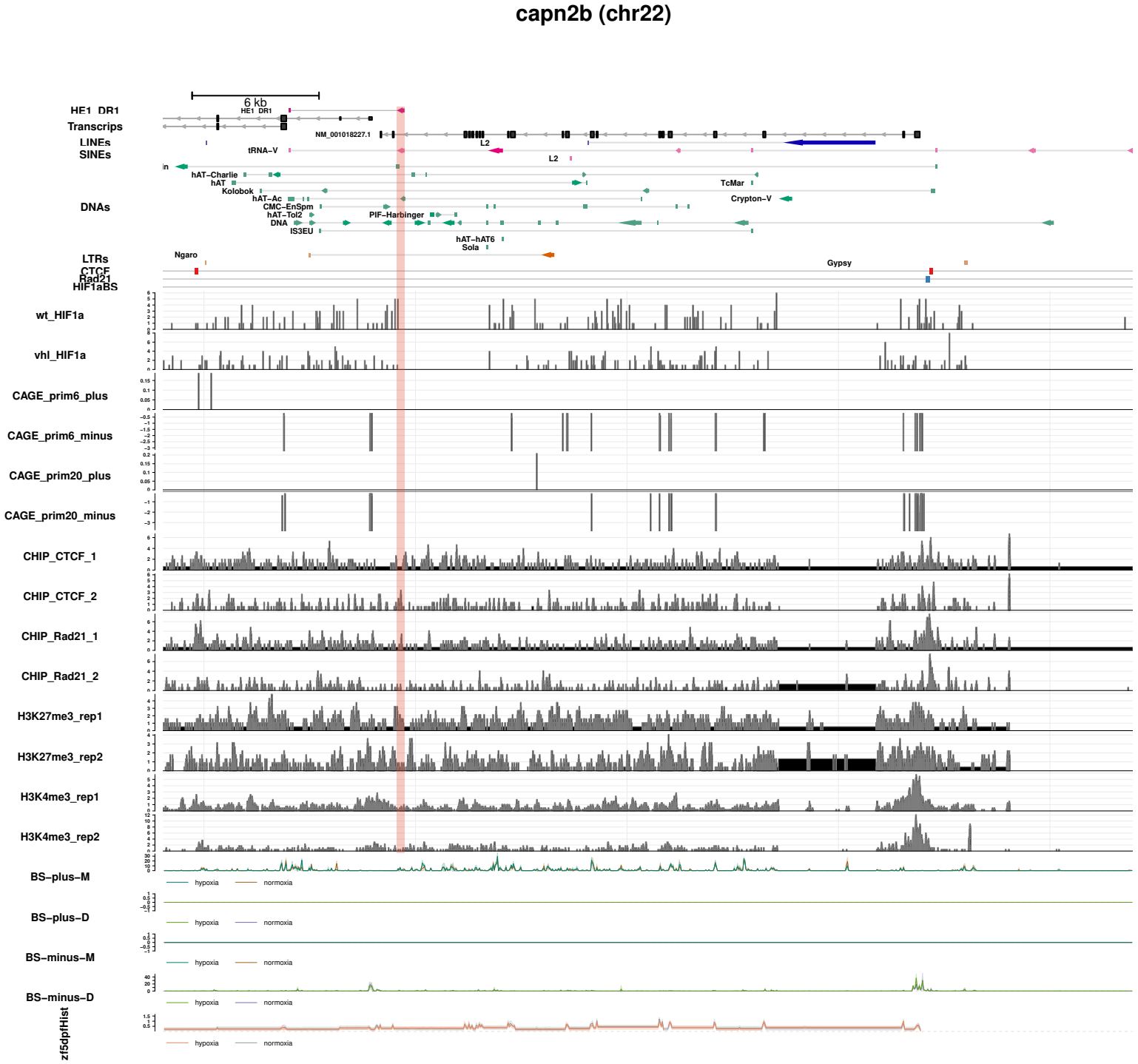


Figure 20: Genomic context of the *capn2b*/*capn1b locus* centered at the *capn2b* gene. The tRNA-V element is highlighted with a vertical red bar.

participate in Vhl-independent pathways for HIF1-alpha degradation (Biswas et al., 2020; Zho et al., 2006).

Finally, there is an interesting observation shared in several of the discussed *loci*. For *tpx2*, *capn1b, capn2b*, *tmem107*, *aurkb*, among other genes, the H3K4me3 peak visible at promoter regions gradually decreases until becoming unidentifiable from noise at the tRNA-V insertion. This supports the chromatin barrier activity of tRNA-V elements, and interestingly, coincides with the presence of H3K27me3 histone modifications, suggesting that such sites may be bivalent domains. These domains are characterized by the co-existence of both H3K4me3 and H3K27me3 histone modifications and are present in the promoters of genes expressed at low levels, such as many developmental genes. Although not fully understood, it is thought that bivalent domains silence developmental genes by keeping them "poised" for activation (Bernstein et al., 2006). In this regard, DNA methylation and the TET family of proteins are known to play a role in determining which sites may act as bivalent domains (Kong et al., 2016). DNA methylation is a modification largely known for silencing TEs, and in somatic cells, is associated with the activity of the polycomb complex (Déléris et al., 2021). Interestingly, it has been found that TET1 and PRC2 (a component of the polycomb complex) co-localize at bivalent promoters in ES cells (Neri et al., 2013). Moreover, Pol II pausing, a paradigm largely associated with responses to external stimuli (Rougvie et al., 1988) and cell fate (Gaertner et al., 2012), have been shown to be correlated with the

presence of bivalent domains (Liu et al., 2017), and interestingly, there is evidence supporting its participation in the transcriptional regulation of tRNA genes (Gerber et al., 2020).

Considering this evidence, it is tempting to speculate about the involvement of multiple epigenetic mechanisms, such as bivalency of histone modifications, Pol II pausing, DNA methylation, and chromatin architecture, in the putative insulation mediated by tRNA-V elements. However, the literature exploring the interplay of several of these paradigms with insulators is rather scarce. Further research and a more integrated framework will be required in order to shed light on the specific mechanisms involved in the potential insulator activity of tRNA-V elements.

CONCLUSIONS

We developed a methodology for detecting CTCF-mediated insulators based on the characterization of TEs biased towards gene vicinities. Using occupancy data for architectural proteins and epigenetic modifications, together with expression data from RNA-seq experiments, we identified the tRNA-V superfamily of SINEs as potentially contributing with insulators to the genome of zebrafish. Moreover, we found that the tRNA-V superfamily is more prevalent in the vicinities of genes in comparison to other superfamilies of TEs and described a discrete number of groups of tRNA-V insertions with characteristic patterns of occupancy for CTCF, Rad21, HIF1-alpha, and the H3K27me3 histone modification. Three of these groups showed biases in gene expression levels of genes upstream and downstream of tRNA-V insertions, which is consistent with the presence of the H3K27me3 repressive mark, supporting an insulator activity for a subset of tRNA-V insertions in the genome. Finally, differential expression analyses of normoxic versus hypoxic samples revealed hundreds of differentially expressed genes in the vicinities of tRNA-V elements with putative insulator activity, suggesting that this activity may be affected by hypoxic stress.

MATERIALS AND METHODS

Quantification of TEs in the vicinity of genes

Although it is assumed that TEs insert randomly all over the genome, there is abundant evidence that supports the contrary (Bourque et al, 2018). Instead, TEs have insertional biases toward specific regions depending on the order and family they belong to, and also in regard to their relationship with other TEs. Hence, instead of assuming a random distribution of insertions, we quantified the number of TE fragments for every TE family in the vicinity of genes. To accomplish this, we built a series of functions for building what we refer to as TE abundance matrices, which store count data from fragments of TEs overlapping specific genomic regions in a TE family-specific manner. These functions were mostly written in R and made use of the GenomicRanges and GenomicFeatures libraries from Bioconductor, among others.

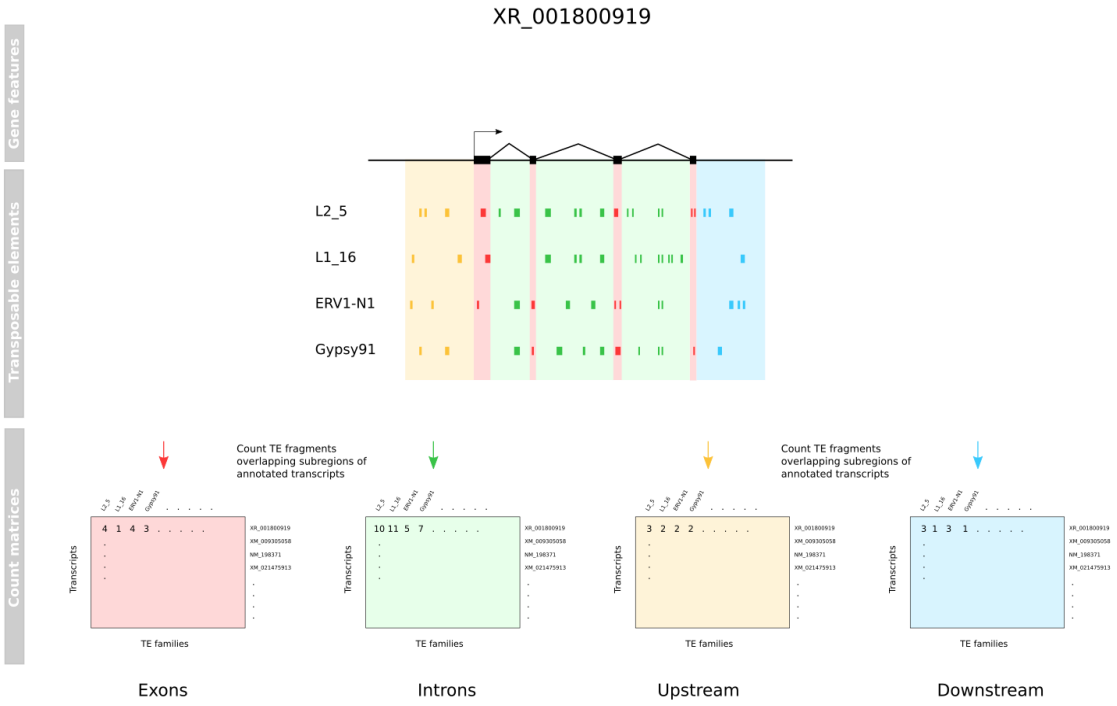


Figure 21: Schematic genomic context of a representative gene and its associated abundance matrices. Colored boxes represent the position of TEs overlapping specific regions on the gene. (yellow: upstream flank; red: exons; green: introns; blue: downstream flank).

Figure 21 shows a diagram representing the strategy used to quantify TEs in specific regions. Briefly, we intersected TE annotations with the ranges corresponding to our region of interest and counted their occurrence at the TE family level. For example, the region highlighted in yellow represents the upstream region of genes and only fragments of TEs in this range would be included in the quantification. The same is applicable to any other region, such as downstream of genes, exons, and introns. Finally, the

family-specific count of fragments is stored in a matrix where rows are the regions of interest and columns families of TEs.

In addition, this strategy can generate matrices that account separately for fragments in sense or antisense orientation with respect to the associated gene. This allows us to assess for deviations from what would be expected by chance, that is, no significant bias in the number of sense or antisense fragments. Deviations from this might be indicative of selection pressure or other unknown factors since theoretically TEs are supposed to behave according to the neutral theory of evolution.

Altogether, these matrices are the starting material to assess more directed questions about the influence of TEs in the expression of neighbor genes.

Calculation of the proportion of genes containing TEs using the Jaccard/Tanimoto coefficient

Jaccard/Tanimoto coefficient is a widely used parameter to assess the similarity between two binary vectors. It is calculated by dividing the intersection of the vectors by their union.

$$J(A,B) = A \cap B / A \cup B$$

A special case in the calculation of the Jaccard/Tanimoto coefficient occurs when both vectors have the same length and one of them is filled completely with 1s (this vector is known as intercept). In this case, as every position of the intercept has a 1, the denominator equals the total number of elements, and hence the coefficient represents a proportion. We have followed this rationale to calculate the proportion of genes with at least one TE of any given family inserted in its vicinity.

For this sake, we generated binary matrices from the TE abundance matrices obtained previously, and also an intercept with the size of the number of genes in the zebrafish genome. In this sense, we have taken a gene-centered approach such that the regions for which the prevalence of TEs is being evaluated are indexed relative to their closest or overlapping gene. Finally, it is important to note that what we refer to here as vicinity is limited to regions of 1kb upstream and downstream of genes.

Sequencing reads workflow

In order to facilitate the processing of sequencing reads from different studies we developed a workflow using Snakemake.

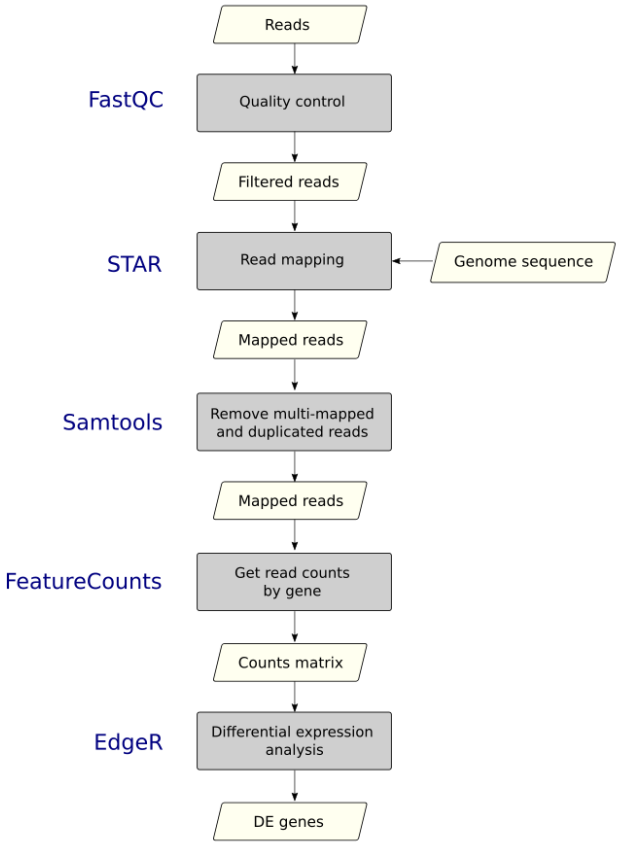


Figure 22: Diagram of the workflow used for processing sequencing reads.

Briefly, it first maps sequencing reads to a reference genome using STAR. The resulting BAM files are then filtered by removing poor quality alignments and PCR duplicates, in order to minimize ambiguously mapped reads. Finally, featureCounts is used to get read counts and build a matrix, which is used as input for a differential expression analysis with the EdgeR package.

Identification of groups of tRNA-V elements from CHIP-seq data

In order to evaluate any degree of enrichment of CTCF, Rad21, HIF1-alpha, and H3K27me3 at tRNA-V elements, we processed sequencing reads from publicly available CHIP-seq experiments using the Snakemake workflow described in the previous section. We obtained the read counts in every 50 bp window on the GRCz11/danRer11 version of the zebrafish genome. Then, we normalized these counts correcting by the normalization factors calculated by edgeR, and used this data to generate heatmaps displaying a 1,000 bp window centered at tRNA-V elements using the Genomation package on a custom R script.

On the other hand, we performed hierarchical clustering (using the "ward.D2" method and a tree height of 100) which allowed us to describe 27 discrete groups of tRNA-V elements with characteristic patterns of CTCF, Rad21, HIF1-alpha, and H3K27me3 occupancy. As before, we generated heatmaps using the Genomation R package, but now highlighting the identified groups.

Particularly, H3K27me3 is an epigenetic mark associated with facultative heterochromatin, which has been shown to repress gene expression in a mechanism involving chromatin interactions (Cai et al, 2021). Considering this, we classified the identified groups into one of four categories. (1) "H3K27me3 downstream", when there are H3K27me3 marks only downstream of the tRNA-V element; (2) "H3K27me3

upstream", when there are H3K27me3 marks only upstream of the tRNA-V element; (3) "H3K27me3 center" when there are H3K27me3 over the tRNA-V element, but no upstream or downstream; and finally (4) "H3K27me3 not-center" when there are H3K27me3 marks upstream and downstream, but not over the tRNA-V element.

Gene expression workflow

In order to facilitate the processing of data from multiple hypoxic and normoxic experiments we used Snakemake and R for developing a workflow for estimating absolute gene expression and performing differential expression analysis.

For estimating absolute gene expression we compared the expression levels of pairs of contiguous genes separated by tRNA-V elements. For this sake, we normalized reads counts using the normalization factors generated by edgeR and estimated absolute gene expression levels considering only reads mapping to exons. Then, we globally compared gene expression levels upstream and downstream of tRNA-V elements with putative insulator activity.

For the differential expression analysis, we used the edgeR package. More specifically, we built specific model matrices for each sample including metadata associated with the sample itself and relative to the platform used for sequencing. This allowed us to correct

for unwanted variation in our differential expression analysis. Then, we tested for differential expression and considered only those genes with a false discovery rate (FDR) lower than 0.05.

Finally, we selected DE genes located no more than 5 kb apart from a tRNA-V element and explored their functions and possible interactions using the STRING database.

Data sources

We recollected and analyzed public CHIP-seq data for two architectural proteins: CTCF and Rad21; the hypoxia-inducible factor 1 alpha (HIF1-alpha); and two histone modifications: H3K27me3 which is associated with heterochromatin regions, and H3K4me3, which is associated with active transcription and promoters. After checking for reads quality, we processed these datasets using the Snakemake workflow described in the "Sequencing reads workflow" section.

We performed the expression analyses using publicly available RNA-seq data from two different works where zebrafish embryos were exposed to hypoxic stress using different experimental designs. On one hand, Long *et* al. generated RNA-seq data from whole 5 dpf larvae exposed to hypoxic stress, cold stress, and the respective untreated control. On the other, Milash *et* al. generated RNA-seq data from heads of larvae at 24, 36, 48,

60, and 72 hpf collected after exposure to acute hypoxic for 12 hrs. and untreated larvae as well (Figures 23 and 24). We processed these reads using the workflows described in the “Sequencing reads workflow” and "Gene expression workflow" sections.

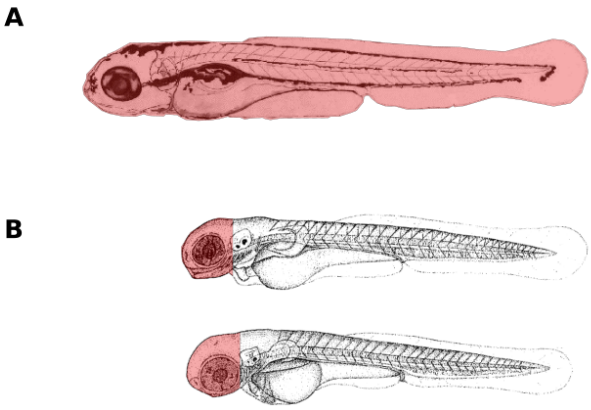


Figure 23: Schematic representation of the source of RNA-seq data used in this work. (A) Whole-body of 120 hpf individuals, from Long et al; (B) Pooled heads of 60 and 72 hpf individuals, separately.

Figure 24: Experimental designs of the datasets used in this work

We selected the best-behaved samples considering the number of differentially expressed genes detected, multidimensional scaling (MDS) plots, P-value, and FDR distributions. The selected samples were 60 hpf, 72hpf (Milash et al), and 120 hpf (Long et al). Finally, we used data from other sources in our genome context visualizations. These are detailed in Table 5 (Appendix).

REFERENCES

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., … & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome biology*, *19*(1), 1-12.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proceedings of the National Academy of Sciences, 36(6), 344-355. https://drive.google.com/open?id=1cMetJSYYSWWJOhsRXaXBOI8C0O1AiQ5s

Long, Y., Yan, J., Song, G., Li, X., Li, X., Li, Q., & Cui, Z. (2015). Transcriptional events co-regulated by hypoxia and cold stresses in Zebrafish larvae. *BMC genomics*, *16*(1), 1-15.

Shang, H., Li, Q., Feng, G., & Cui, Z. (2011). Identification and characterization of alternative promoters, transcripts and protein isoforms of zebrafish R2 gene. *PloS one*, *6*(8), e24089.

Aanes, H., Østrup, O., Andersen, I. S., Moen, L. F., Mathavan, S., Collas, P., & Alestrom, P. (2013). Differential transcript isoform usage pre-and post-zygotic genome activation in zebrafish. *BMC genomics*, *14*(1), 1-15.

Batut, P., Dobin, A., Plessy, C., Carninci, P., & Gingeras, T. R. (2013). High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome research*, *23*(1), 169-180.

Persson, J., Steglich, B., Smialowska, A., Boyd, M., Bornholdt, J., Andersson, R., … & Ekwall, K. (2016). Regulating retrotransposon activity through the use of alternative transcription start sites. *EMBO reports*, *17*(5), 753-768.

Miao, B., Fu, S., Lyu, C., Gontarz, P., Wang, T., & Zhang, B. (2020). Tissue-specific usage of transposable element-derived promoters in mouse development. *Genome biology*, *21*(1), 1-25.

Villanueva-Cañas, J. L., 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 research*, *47*(13), 6842-6857.

Reyes, A., & Huber, W. (2018). Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic acids research*, *46*(2), 582-592.

Li, B. J., Zhu, Z. X., Qin, H., Meng, Z. N., Lin, H. R., & Xia, J. H. (2020). Genome-wide characterization of alternative splicing events and their responses to cold stress in tilapia. *Frontiers in genetics*, *11*, 244.

Horváth, V., Merenciano, M., & González, J. (2017). Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends in Genetics*, *33*(11), 832-841.

Lanciano, S., & Cristofari, G. (2020). Measuring and interpreting transposable element expression. Nature Reviews Genetics, 21(12), 721-736.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1), 139-140.

Tanave, A., Imai, Y., & Koide, T. (2019). Nested retrotransposition in the East Asian mouse genome causes the classical nonagouti mutation. *Communications biology*, *2*(1), 1-11.

Wagner, A. (2005). Energy constraints on the evolution of gene expression. *Molecular biology and evolution*, *22*(6), 1365-1374.

Kapusta, A., Kronenberg, Z., Lynch, V. J., Zhuo, X., Ramsay, L., Bourque, G., ... & Feschotte, C. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet*, *9*(4), e1003470.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21.

Kazachenka, A., Bertozzi, T. M., Sjoberg-Herrera, M. K., Walker, N., Gardner, J., Gunning, R., ... & Ferguson-Smith, A. C. (2018). Identification, characterization, and heritability of murine metastable epialleles: implications for non-genetic inheritance. *Cell*, *175*(5), 1259-1271.

Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Gonçalves, Â., Kutter, C., ... & Odom, D. T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*, *148*(1-2), 335-348.

Pérez-Rico, Y. A., Barillot, E., & Shkumatava, A. (2020). Demarcation of Topologically Associating Domains Is Uncoupled from Enriched CTCF Binding in Developing Zebrafish. *Iscience*, *23*(5), 101046.

Lefevre, P., Witham, J., Lacroix, C. E., Cockerill, P. N., & Bonifer, C. (2008). The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. Molecular cell, 32(1), 129-139.

Magbanua, J. P., Runneburger, E., Russell, S., & White, R. (2015). A variably occupied CTCF binding site in the ultrabithorax gene in the Drosophila bithorax complex. *Molecular and cellular biology*, *35*(1), 318-330.

He, J., Fu, X., Zhang, M., He, F., Li, W., Abdul, M. M., ... & Hutchins, A. P. (2019). Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nature communications*, *10*(1), 1-13.

Kang, J., Lienhard, M., Pastor, W. A., Chawla, A., Novotny, M., Tsagaratou, A., ... & Rao, A. (2015). Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. *Proceedings of the National Academy of Sciences*, *112*(31), E4236-E4245.

Tafani, M., Sansone, L., Limana, F., Arcangeli, T., De Santis, E., Polese, M., ... & Russo, M. A. (2016). The interplay of reactive oxygen species, hypoxia, inflammation, and sirtuins in cancer initiation and progression. *Oxidative medicine and cellular longevity*, *2016*.

Coulter, J. B., O'Driscoll, C. M., & Bressler, J. P. (2013). Hydroquinone increases 5-hydroxymethylcytosine formation through ten eleven translocation 1 (TET1) 5-methylcytosine dioxygenase. *Journal of biological chemistry*, *288*(40), 28792-28800.

Kietzmann, T., Petry, A., Shvetsova, A., Gerhold, J. M., & Görlach, A. (2017). The epigenetic landscape related to reactive oxygen species formation in the cardiovascular system. *British journal of pharmacology*, *174*(12), 1533-1554.

Meng, H., Chen, G., Gao, H. M., Song, X., Shi, Y., & Cao, L. (2014). The emerging nexus of active DNA demethylation and mitochondrial oxidative metabolism in post-mitotic neurons. *International journal of molecular sciences*, *15*(12), 22604-22625.

Tiana, M., Villar, D., Perez-Guijarro, E., Gomez-Maldonado, L., Molto, E., Fernandez-Minan, A., ... & Del Peso, L. (2012). A role for insulator elements in the regulation of gene expression response to hypoxia. *Nucleic acids research*, *40*(5), 1916-1927.

Merkenschlager, M., & Odom, D. T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. *Cell*, *152*(6), 1285-1297.

King, M. R., Matzat, L. H., Dale, R. K., Lim, S. J., & Lei, E. P. (2014). The RNA-binding protein Rumpelstiltskin antagonizes gypsy chromatin insulator function in a tissue-specific manner. *Journal of cell science*, *127*(13), 2956-2966.

Ideraabdullah, F. Y., Thorvaldsen, J. L., Myers, J. A., & Bartolomei, M. S. (2014). Tissue-specific insulator function at H19/Igf2 revealed by deletions at the imprinting control region. *Human molecular genetics*, *23*(23), 6246-6259.

Lyu, X., Rowley, M. J., & Corces, V. G. (2018). Architectural proteins and pluripotency factors cooperate to orchestrate the transcriptional response of hESCs to temperature stress. *Molecular cell*, *71*(6), 940-955.

Ray, J., Munn, P. R., Vihervaara, A., Lewis, J. J., Ozer, A., Danko, C. G., & Lis, J. T. (2019). Chromatin conformation remains stable upon extensive transcriptional changes driven by heat shock. *Proceedings of the National Academy of Sciences*, *116*(39), 19431-19439.

Sanders, J. T., Freeman, T. F., Xu, Y., Golloshi, R., Stallard, M. A., Hill, A. M., ... & McCord, R. P. (2020). Radiation-induced DNA damage and repair effects on 3D genome organization. *Nature communications*, *11*(1), 1-14.

Greenald, D., Jeyakani, J., Pelster, B., Sealy, I., Mathavan, S., & van Eeden, F. J. (2015). Genome-wide mapping of Hif-1α binding sites in zebrafish. *BMC genomics*, *16*(1), 1-17.

Ragsdale, A., Ortega-Recalde, O., Dutoit, L., Besson, A. A., Chia, J. H., King, T., ... & Johnson, S. L. (2020). Paternal hypoxia exposure primes offspring for increased hypoxia resistance. *bioRxiv*

Milash, B., Gao, J., Stevenson, T. J., Son, J. H., Dahl, T., & Bonkowsky, J. L. (2016). Temporal Dysynchrony in brain connectivity gene expression following hypoxia. BMC genomics, 17(1), 1-14.

Meier, M., Grant, J., Dowdle, A., Thomas, A., Gerton, J., Collas, P., ... & Horsfield, J. A. (2018). Cohesin facilitates zygotic genome activation in zebrafish. *Development*, *145*(1).

Feschotte, C., & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.*, *41*, 331-368.

Emera, D., & Wagner, G. P. (2012). Transposable element recruitments in the mammalian placenta: impacts and mechanisms. *Briefings in functional genomics*, *11*(4), 267-276.

Jones, J. M., & Gellert, M. (2004). The taming of a transposon: V (D) J recombination and the immune system. *Immunological reviews*, *200*(1), 233-248.

Arkhipova, I. R. (2018). Neutral theory, transposable elements, and eukaryotic genome evolution. *Molecular biology and evolution*, *35*(6), 1332-1337.

Slotkin, R. K., & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature reviews genetics*, *8*(4), 272-285.

Wells, J. N., & Feschotte, C. (2020). A field guide to eukaryotic transposable elements. *Annual review of genetics*, *54*, 539-561.

Choudhary, M. N., Friedman, R. Z., Wang, J. T., Jang, H. S., Zhuo, X., & Wang, T. (2020). Co-opted transposons help perpetuate conserved higher-order chromosomal structures. *Genome biology*, *21*(1), 1-14.

Davidson, I. F., & Peters, J. M. (2021). Genome folding through loop extrusion by SMC complexes. *Nature Reviews Molecular Cell Biology*, *22*(7), 445-464.

Chen, D., & Lei, E. P. (2019). Function and regulation of chromatin insulators in dynamic genome organization. *Current opinion in cell biology*, *58*, 61-68.

Brasset, E., & Vaury, C. (2005). Insulators are fundamental components of the eukaryotic genomes. *Heredity*, *94*(6), 571-576.

Yang, J., & Corces, V. G. (2012). Insulators, long-range interactions, and genome function. Current opinion in genetics & development, 22(2), 86-92.

Román, A. C., González-Rico, F. J., Moltó, E., Hernando, H., Neto, A., Vicente-Garcia, C., ... & Fernández-Salguero, P. M. (2011). Dioxin receptor and SLUG transcription factors regulate the insulator activity of B1 SINE retrotransposons via an RNA polymerase switch. Genome research, 21(3), 422-432.

Raab, J. R., Chiu, J., Zhu, J., Katzman, S., Kurukuti, S., Wade, P. A., ... & Kamakaka, R. T. (2012). Human tRNA genes function as chromatin insulators. The EMBO journal, 31(2), 330-350.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proceedings of the National Academy of Sciences, 36(6), 344-355.

Moqtaderi, Z., Wang, J., Raha, D., White, R. J., Snyder, M., Weng, Z., & Struhl, K. (2010). Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. *Nature structural & molecular biology*, *17*(5), 635-640.

Ogiwara, I., Miya, M., Ohshima, K., & Okada, N. (2002). V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. *Genome research*, *12*(2), 316-324.

Raab, J. R., Chiu, J., Zhu, J., Katzman, S., Kurukuti, S., Wade, P. A., ... & Kamakaka, R. T. (2012). Human tRNA genes function as chromatin insulators. The EMBO journal, 31(2), 330-350.

Izsvák, Z., Ivics, Z., Garcia-Estefania, D., Fahrenkrug, S. C., & Hackett, P. B. (1996). DANA elements: a family of composite, tRNA-derived short interspersed DNA elements associated with mutational activities in zebrafish (Danio rerio). *Proceedings of the National Academy of Sciences*, *93*(3), 1077-1081.

Donze, D., & Kamakaka, R. T. (2001). RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae. *The EMBO journal*, *20*(3), 520-531.

Noma, K. I., Cam, H. P., Maraia, R. J., & Grewal, S. I. (2006). A role for TFIIIC transcription factor complex in genome organization. *Cell*, *125*(5), 859-872.

Simms, T. A., Dugas, S. L., Gremillion, J. C., Ibos, M. E., Dandurand, M. N., Toliver, T. T., ... & Donze, D. (2008). TFIIIC binding sites function as both heterochromatin barriers and chromatin insulators in Saccharomyces cerevisiae. *Eukaryotic cell*, *7*(12), 2078-2086.

Kirkland, J. G., Raab, J. R., & Kamakaka, R. T. (2013). TFIIIC bound DNA elements in nuclear organization and insulation. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1829*(3-4), 418-424.

Kramerov, D. A., & Vassetzky, N. S. (2011). Origin and evolution of SINEs in eukaryotic genomes. *Heredity*, *107*(6), 487-495.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., ... & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature, 485(7398), 376-380.

Matharu, N. K., & Ahanger, S. H. (2015). Chromatin insulators and topological domains: adding new dimensions to 3D genome architecture. *Genes*, *6*(3), 790-811.

Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J., Uebersohn, A., Abdennur, N., ... & Bruneau, B. G. (2017). Targeted degradation of CTCF decouples local insulation of chromosome domains from higher-order genomic compartmentalization. *bioRxiv*, 095802.

Bowers, S. R., Mirabella, F., Calero-Nieto, F. J., Valeaux, S., Hadjur, S., Baxter, E. W., ... & Cockerill, P. N. (2009). A conserved insulator that recruits CTCF and cohesin exists between the closely related but divergently regulated interleukin-3 and granulocyte-macrophage colony-stimulating factor genes. *Molecular and cellular biology*, *29*(7), 1682-1693.

Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., ... & Krumm, A. (2008). CTCF physically links cohesin to chromatin. *Proceedings of the National Academy of Sciences*, *105*(24), 8309-8314.

Cai, Y., Zhang, Y., Loh, Y. P., Tng, J. Q., Lim, M. C., Cao, Z., ... & Fullwood, M. J. (2021). H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. *Nature communications*, *12*(1), 1-22.

Cuddapah, S., Jothi, R., Schones, D. E., Roh, T. Y., Cui, K., & Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome research*, *19*(1), 24-32.

Franke, M., la Calle-Mustienes, D., Neto, A., Almuedo-Castillo, M., Irastorza-Azcarate, I., Acemel, R. D., ... & Gómez-Skarmeta, J. L. (2021). CTCF knockout in zebrafish induces alterations in regulatory landscapes and developmental gene expression. *Nature communications*, *12*(1), 1-19.

Kang, N., Duan, L., Tang, L., Liu, S., Li, C., Li, Y., ... & He, W. (2008). Identification and characterization of a novel thymus aging related protein Rwdd1. *Cellular & molecular immunology*, *5*(4), 279-285.

Patterson, A. J., Xiao, D., Xiong, F., Dixon, B., & Zhang, L. (2012). Hypoxia-derived oxidative stress mediates epigenetic repression of PKCε gene in foetal rat hearts. *Cardiovascular research*, *93*(2), 302-310.

McGarry, T., Biniecka, M., Veale, D. J., & Fearon, U. (2018). Hypoxia, oxidative stress and inflammation. *Free Radical Biology and Medicine*, *125*, 15-24.

Zhou, J., Kohl, R., Herr, B., Frank, R., & Brune, B. (2006). Calpain mediates a von hippel-lindau protein–independent destruction of hypoxia-inducible factor-1α. *Molecular*

biology of the cell, *17*(4), 1549-1558.

Cai, Y., Zhang, Y., Loh, Y. P., Tng, J. Q., Lim, M. C., Cao, Z., ... & Fullwood, M. J. (2021). H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. *Nature communications*, *12*(1), 1-22.

Vihervaara, A., Mahat, D. B., Guertin, M. J., Chu, T., Danko, C. G., Lis, J. T., & Sistonen, L. (2017). Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nature communications*, *8*(1), 1-16.

Iyer, J., & Tsai, M. Y. (2012). A novel role for TPX2 as a scaffold and co-activator protein of the Chromosomal Passenger Complex. *Cellular signalling*, *24*(8), 1677-1689.

Biswas, K., Sarkar, S., Said, N., Brautigan, D. L., & Larner, J. M. (2020). Aurora B Kinase Promotes CHIP-Dependent Degradation of HIF1α in Prostate Cancer Cells. *Molecular cancer therapeutics*, *19*(4), 1008-1017.

Rougeot, J., Chrispijn, N. D., Aben, M., Elurbe, D. M., Andralojc, K. M., Murphy, P. J., ... & Kamminga, L. M. (2019). Maintenance of spatial gene expression by Polycomb-mediated repression after formation of a vertebrate body plan. *Development*, *146*(19), dev178590.

Lepage, S. E., & Bruce, A. E. (2008). Characterization and comparative expression of zebrafish calpain system genes during early development. *Developmental dynamics: an official publication of the American Association of Anatomists*, *237*(3), 819-829.

Liu, J., Wu, X., Zhang, H., Pfeifer, G. P., & Lu, Q. (2017). Dynamics of RNA polymerase II pausing and bivalent histone H3 methylation during neuronal differentiation in brain development. *Cell reports*, *20*(6), 1307-1318.

Gaertner, B., Johnston, J., Chen, K., Wallaschek, N., Paulson, A., Garruss, A. S., ... & Zeitlinger, J. (2012). Poised RNA polymerase II changes over developmental time and prepares genes for future expression. *Cell reports*, *2*(6), 1670-1683.

Rougvie, A. E., & Lis, J. T. (1988). The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell*, *54*(6), 795-804.

Van Bortle, K., & Corces, V. G. (2012). tDNA insulators and the emerging role of TFIIIC in genome organization. *Transcription*, *3*(6), 277-284.

Cieśla, M., Skowronek, E., & Boguta, M. (2018). Function of TFIIIC, RNA polymerase III initiation factor, in activation and repression of tRNA gene transcription. *Nucleic acids research*, *46*(18), 9444-9455.

Wang, J., Vicente-García, C., Seruggia, D., Moltó, E., Fernandez-Miñán, A., Neto, A., ... & Jordan, I. K. (2015). MIR retrotransposon sequences provide insulators to the human genome. *Proceedings of the National Academy of Sciences*, *112*(32), E4428-E4437.

Srinivasan, S. S., Gong, Y., Xu, S., Hwang, A., Xu, M., Girgenti, M. J., & Zhang, J. (2022). InsuLock: A Weakly Supervised Learning Approach for Accurate Insulator Prediction, and Variant Impact Quantification. *Genes*, *13*(4), 621.

Belokopytova, P., & Fishman, V. (2021). Predicting genome architecture: challenges and solutions. *Frontiers in genetics*, 1776.

Rowley, M. J., & Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nature Reviews Genetics*, *19*(12), 789-800.

Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., & Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *elife*, *6*, e25776.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., ... & Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, *125*(2), 315-326.

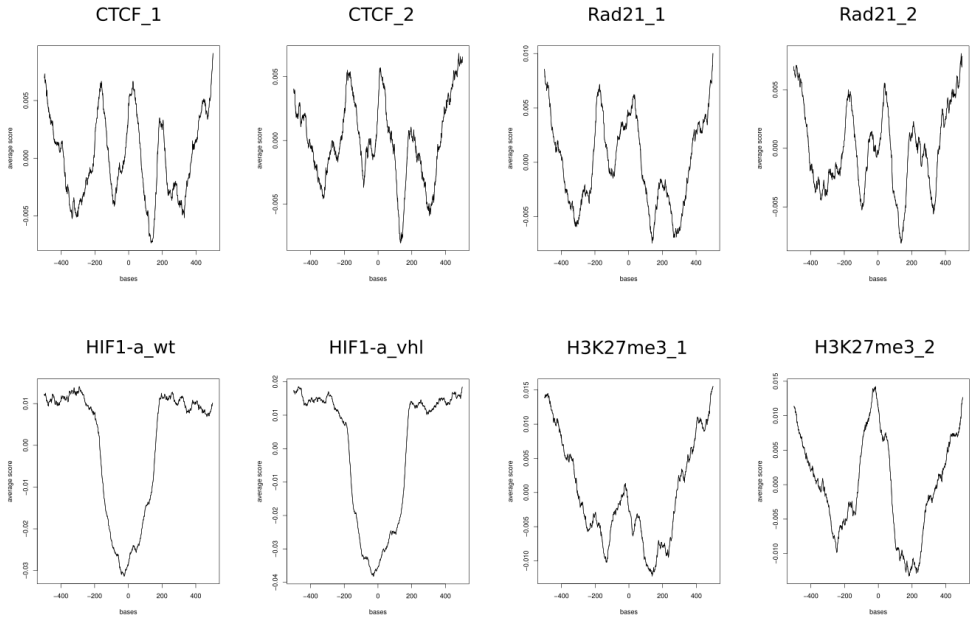
Kong, L., Tan, L., Lv, R., Shi, Z., Xiong, L., Wu, F., ... & Shi, Y. G. (2016). A primary role of TET proteins in establishment and maintenance of De Novo bivalency at CpG islands. *Nucleic acids research*, *44*(18), 8682-8692.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., ... & Oliviero, S. (2013). Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome biology*, *14*(8), 1-13.

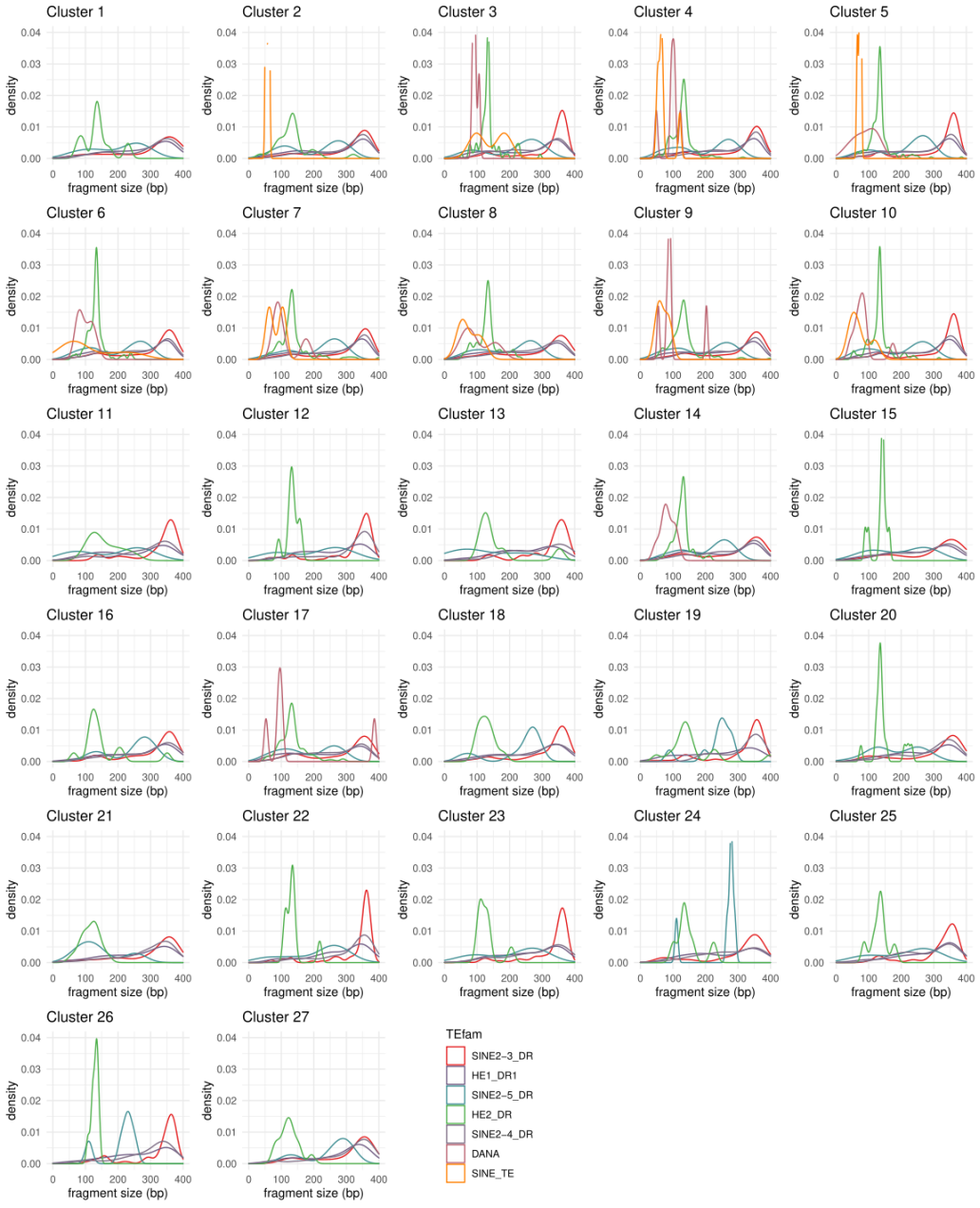
Déléris, A., Berger, F., & Duharcourt, S. (2021). Role of Polycomb in the control of transposable elements. *Trends in Genetics*, *37*(10), 882-889.

Gerber, A., Ito, K., Chu, C. S., & Roeder, R. G. (2020). Gene-specific control of tRNA expression by RNA polymerase II. *Molecular cell*, *78*(4), 765-778.

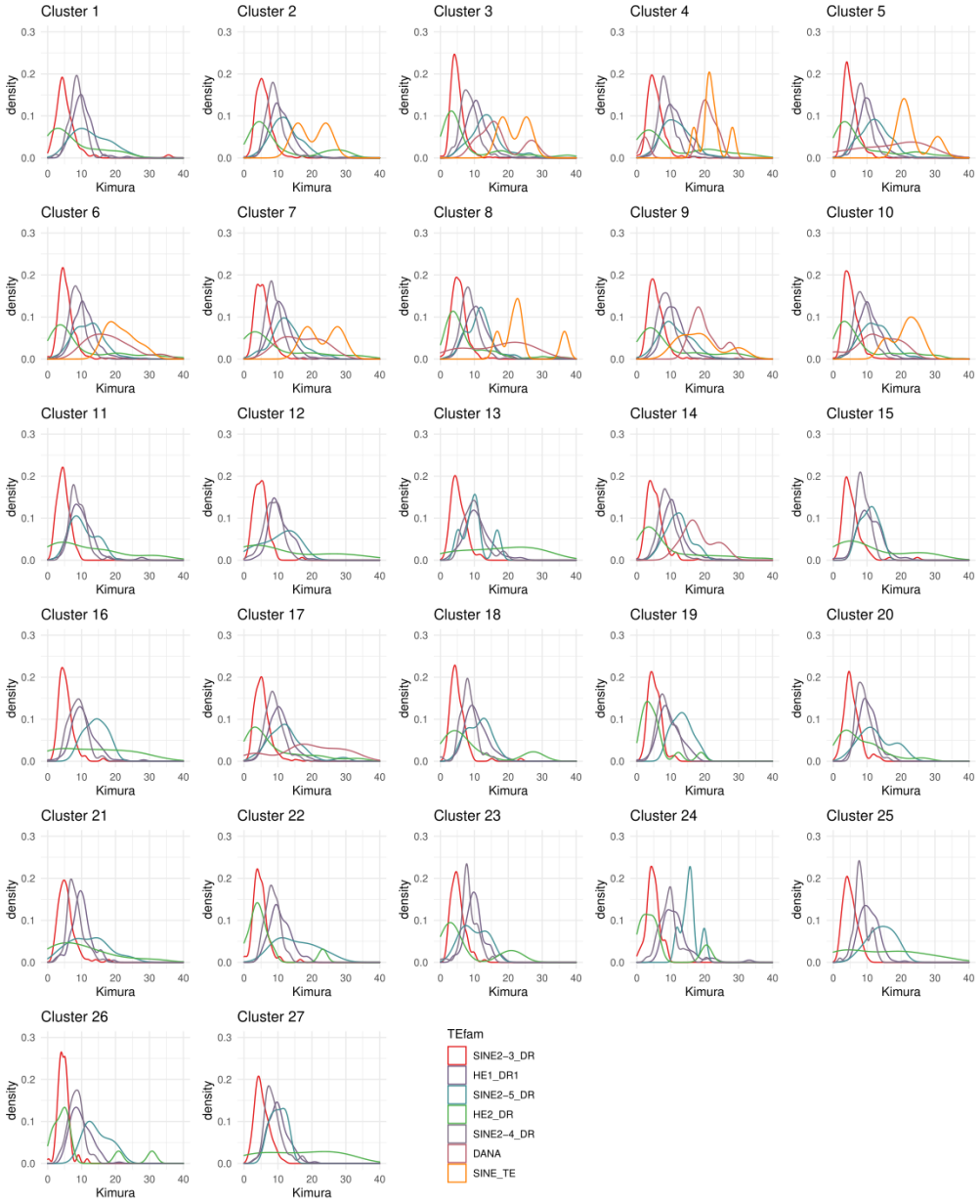
APPENDIX



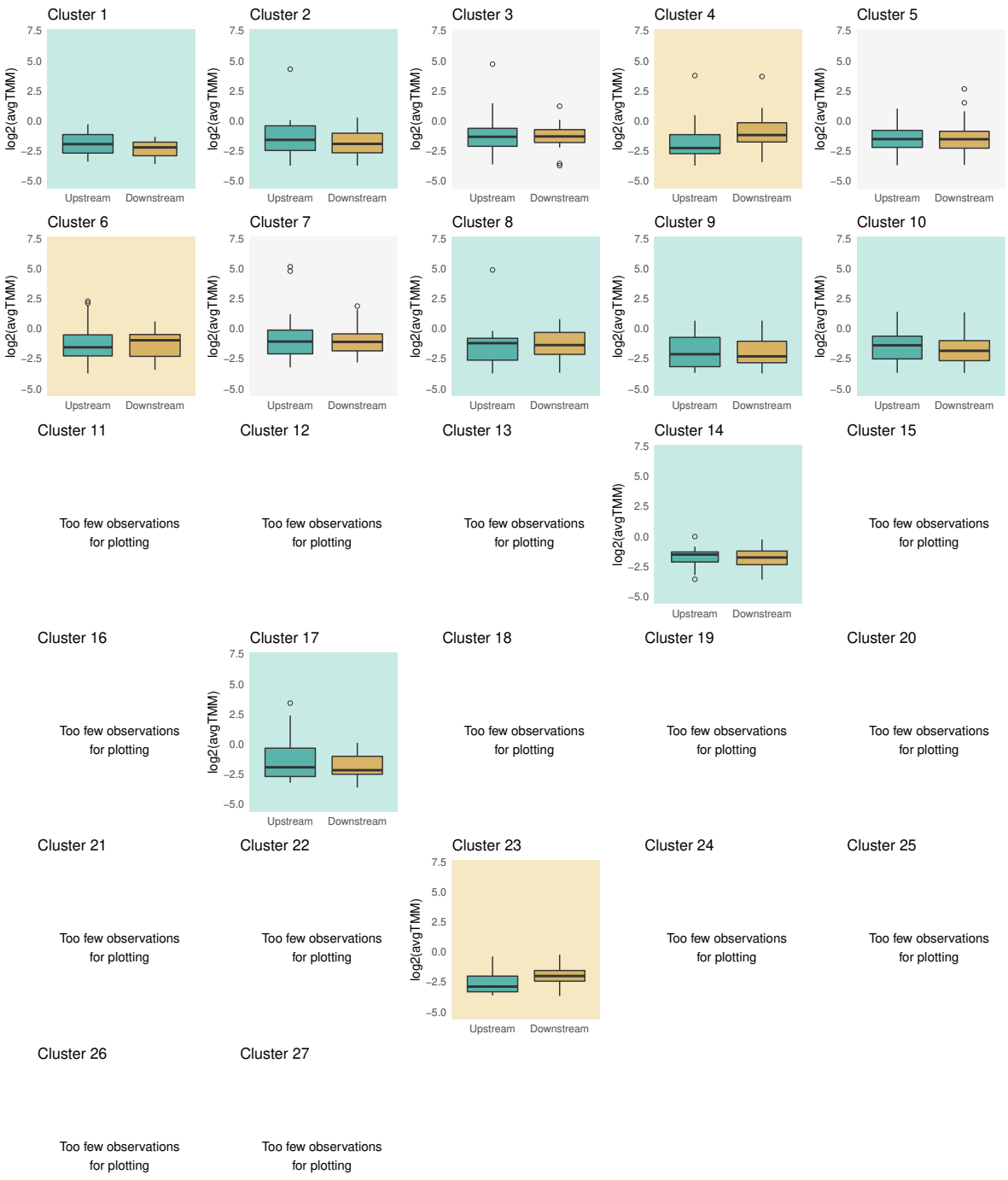
Supplementary Figure 1: Density plots showing the CHIP-seq coverage for a 1 kb window centered at tRNA-V elements overlapping genes.



Supplementary Figure 2: Fragment size distribution for all tRNA-V families by cluster



Supplementary Figure 3: Kimura two-parameter distribution for all tRNA-V families by cluster



Supplementary Figure 4: Global trends in gene expression levels upstream and downstream of tRNA-V insertions for each cluster in a representative sample. "Normoxia 1" sample is shown. The background is colored based on the region with the highest median log gene expression.

Table 1: Upregulated genes in heterochromatin domains (I). Only insertions whose downstream genes are upregulated, and from clusters labeled as "*H3K27me3 downstream"* are shown. Genes in bolt were found DE in both 60 and 72 hpf samples; others were found DE only at 72 hpf.

SINE family	Gene upstream	Gene downstream	Regulation upstream	Regulation downstream	Cluster	Kimura distance
SINE2-4_DR	igsf21a	klhdc7a	Not DE	Upregulated	2	5.6
HE1_DR1	***slc51a***	***pdk3a***	Downregulated	Upregulated	8	8.26
HE1_DR1	**matn1**	**sesn2**	Not DE	Upregulated	17	9
HE1_DR1	cyyr1	appa	Not DE	Upregulated	10	10.28
SINE2-3_DR	nccrp1	slc8a2a	Not DE	Upregulated	10	8.14
HE1_DR1	inpp5ka	tekt1	Not DE	Upregulated	17	6.23
HE1_DR1	**slc45a3**	**si:ch211-117k10.3**	Not DE	Upregulated	2	10.25
SINE2-3_DR	adgre5b.2	adgre5b.3	Not DE	Upregulated	17	6.19
HE2_DR	zyx	kel	Not DE	Upregulated	10	24.47
SINE2-4_DR	**si:ch211-181d7.3**	**LOC110437747**	Upregulated	Upregulated	10	7.5
SINE2-4_DR	**tmem120a**	**pora**	Not DE	Upregulated	8	9.79
SINE2-4_DR	pnkd	catip	Not DE	Upregulated	8	5.94

Table 2: Upregulated genes in heterochromatin domains (II). Only insertions whose upstream genes are upregulated, and from clusters labeled as "*H3K27me3 upstream"* are shown. Genes in bolt were found DE in both 60 and 72 hpf samples; others were found DE only at 72 hpf.

SINE family	Gene upstream	Gene downstream	Regulation upstream	Regulation downstream	Cluster	Kimura distance
HE1_DR1	tmem176	si:dkey-9i23.14	Upregulated	Not DE	14	10.24
SINE2-3_DR	LOC101882086	LOC110439800	Upregulated	Upregulated	14	3.22
HE1_DR1	LOC100332293	polr3e	Upregulated	Not DE	7	12.58
HE1_DR1	***cry1ab***	***zgc:153031***	Upregulated	Downregulated	14	9.2
HE1_DR1	tbc1d15	tph2	Upregulated	Not DE	5	12.28
HE1_DR1	itgae.2	cul5a	Upregulated	Not DE	7	10.34

Table 3: Downregulated genes upstream of heterochromatin domains. Only insertions whose upstream genes are downregulated, and from clusters labeled as "*H3K27me3 downstream*" are shown. Genes in bolt were found DE in both 60 and 72 hpf samples; others were found DE only at 72 hpf.

SINE family	Gene upstream	Gene downstream	Regulation upstream	Regulation downstream	Cluster	Kimura distance
HE1_DR1	*slc51a*	*pdk3a*	Downregulated	Upregulated	8	8.26
HE1_DR1	cyp2x9	cyp2x8	Downregulated	Not DE	8	10.01
SINE2-4_DR	**bbs2**	**b3gnt9**	Downregulated	Not DE	2	7.52
SINE2-4_DR	cad	dnajc5ga	Downregulated	Not DE	17	4.09
SINE2-3_DR	dhx29	mcidas	Downregulated	Not DE	10	6.45
SINE2-3_DR	msrb1b	meiob	Downregulated	Not DE	8	6.38
SINE2-3_DR	**grk1b**	**arhgef15**	Downregulated	Downregulated	8	6.6
HE1_DR1	si:ch211-197h24.9	LOC103908936	Downregulated	Not DE	10	12.24
HE1_DR1	**capn2b**	**capn1b**	Downregulated	Not DE	10	14.86
SINE2-4_DR	**atp5g3a**	**atf2**	Downregulated	Not DE	8	10.86
HE1_DR1	gltpa	trpv4	Downregulated	Not DE	10	6.72
HE1_DR1	**manea**	**fut9a**	Downregulated	Not DE	8	11.92
SINE2-4_DR	nup210	LOC108181449	Downregulated	Not DE	10	5.55
HE2_DR	ints8	qrfp	Downregulated	Not DE	8	3.96

Table 4: Downregulated genes downstream of heterochromatin domains. Only insertions whose downstream genes are downregulated, and from clusters labeled as "*H3K27me3 upstream*" are shown. Genes in bolt were found DE in both 60 and 72 hpf samples; others were found DE only at 72 hpf.

SINE family	Gene upstream	Gene downstream	Regulation upstream	Regulation downstream	Cluster	Kimura distance
SINE2-4_DR	**smarcd3a**	**abcf2a**	Downregulated	Downregulated	5	7.5
HE1_DR1	sgcb	lrrc66	Not DE	Downregulated	5	18.08
HE1_DR1	myom3	fabp10a	Not DE	Downregulated	5	6.21
HE1_DR1	**LOC110437808**	**rwdd1**	Not DE	Downregulated	7	25.89
HE2_DR	trappc6b	tpx2	Not DE	Downregulated	9	40.14
SINE2-5_DR	tspan37	elavl1	Not DE	Downregulated	5	9.03
HE1_DR1	rpl10a	si:ch211-163l21.11	Not DE	Downregulated	7	11.49
SINE2-3_DR	**nos2b**	**mrps23**	Not DE	Downregulated	5	5.08
HE1_DR1	larp1b	pgrmc2	Not DE	Downregulated	9	10.53
HE1_DR1	asb12a	amer1	Not DE	Downregulated	7	11.52
HE1_DR1	*cry1ab*	*zgc:153031*	Upregulated	Downregulated	14	9.2
HE1_DR1	tmem107	aurkb	Not DE	Downregulated	5	8.9
HE1_DR1	si:ch1073-416d2.3	ddx24	Not DE	Downregulated	14	9.59
SINE2-3_DR	wdr66	bcl7a	Not DE	Downregulated	7	2.12
SINE2-4_DR	cpa2	cpa4	Not DE	Downregulated	7	7.4
SINE2-4_DR	**si:ch73-40i7.5**	**zgc:172302**	Not DE	Downregulated	7	6.45
HE1_DR1	nxnl2	wu:fb59d01	Not DE	Downregulated	5	3.39

SINE2-5_DR	**c2cd2l**	**h2afx**	Not DE	Downregulated	7	9.47

Table 5: Source for all the data used in this work

Track	Description	Source
Transcripts	Annotation of all transcript isoforms	danRer11 UCSC annotation
TSS data		
CAGE_prim6_plus	Sequencing coverage at plus-strand from CAGE-seq experiments in wild-type individuals at prim6 developmental stage (~25hpf).	Nepal et al., 2013
CAGE_prim6_minus	Sequencing coverage at minus strand from CAGE-seq experiments in wild-type individuals at prim6 developmental stage (~25hpf).	Nepal et al., 2013
CAGE_prim20_plus	Sequencing coverage at plus-strand from CAGE-seq experiments in wild-type individuals at prim20 developmental stage (~33hpf).	Nepal et al., 2013
CAGE_prim20_minus	Sequencing coverage at minus-strand from CAGE-seq experiments in wild-type individuals at prim20 developmental stage (~33hpf).	Nepal et al., 2013
Transposable elements		
LINEs	Annotation of all fragments of TEs from the LINE order grouped by superfamily	Custom annotation using RepeatMasker and RepBase
SINEs	Annotation of all fragments of TEs from the SINE order grouped by superfamily	Custom annotation using RepeatMasker and RepBase
DNAs	Annotation of all fragments of TEs from the DNA order grouped by superfamily	Custom annotation using RepeatMasker and RepBase
LTRs	Annotation of all fragments of TEs from the LTR order grouped by superfamily	Custom annotation using RepeatMasker and RepBase
TE family	Annotation of all fragments from an specific TE family	Custom annotation using RepeatMasker and RepBase
Transcription factors		

CTCF	CTCF binding sites and coverage according to CHIP-seq data from embryos at 10 hpf	Meier et al., 2018
Rad21	Rad21 binding sites and coverage according to CHIP-seq data from embryos at 10 hpf	Meier et al., 2018
HIF1-alpha	HIF-1a binding sites and coverage according to CHIP-seq data in wild-type and vhl knock-out individuals at 4 dpf	Greenald et al, 2015
Epigenetic modifications		
H3K4me3	CHIP-seq experiments for H3K4me3 in 24 hpf embryos (Whole body)	Rougeot et al., 2019
H3K27me3	CHIP-seq experiments for H3K27me3 in 24 hpf embryos (Whole body)	Rougeot et al., 2019
Expression data		
zf24hpf	RNA-seq experiments of individuals exposed to hypoxic stress in comparison to non-treated individuals at 24hpf (Head)	Milash et al., 2016
zf36hpf	RNA-seq experiments of individuals exposed to hypoxic stress in comparison to non-treated individuals at 36hpf (Head)	Milash et al., 2016
zf48hpf	RNA-seq experiments of individuals exposed to hypoxic stress in comparison to non-treated individuals at 48hpf (Head)	Milash et al., 2016
zf60hpf	RNA-seq experiments of individuals exposed to hypoxic stress in comparison to non-treated individuals at 60hpf (Head)	Milash et al., 2016
zf72hpf	RNA-seq experiments of individuals exposed to hypoxic stress in comparison to non-treated individuals at 72hpf (Head)	Milash et al., 2016
zf5dpf	RNA-seq experiments of individuals exposed to hypoxic stress in comparison to non-treated individuals at 120hpf (Whole body)	Long et al., 2015
Methylation data		
BS-plus-M	Density of methylated sites in the plus strand from the sperm of male individuals exposed to hypoxic conditioning in comparison to non-treated individuals.	Ragsdale et al., 2020

BS-plus-D	Density of demethylated sites in the plus strand from the sperm of male individuals exposed to hypoxic conditioning in comparison to non-treated individuals.	Ragsdale et al., 2020
BS-minus-M	Density of methylated sites in the minus strand from the sperm of male individuals exposed to hypoxic conditioning in comparison to non-treated individuals.	Ragsdale et al., 2020
BS-minus-D	Density of demethylated sites in the minus strand from the sperm of male individuals exposed to hypoxic conditioning in comparison to non-treated individuals.	Ragsdale et al., 2020