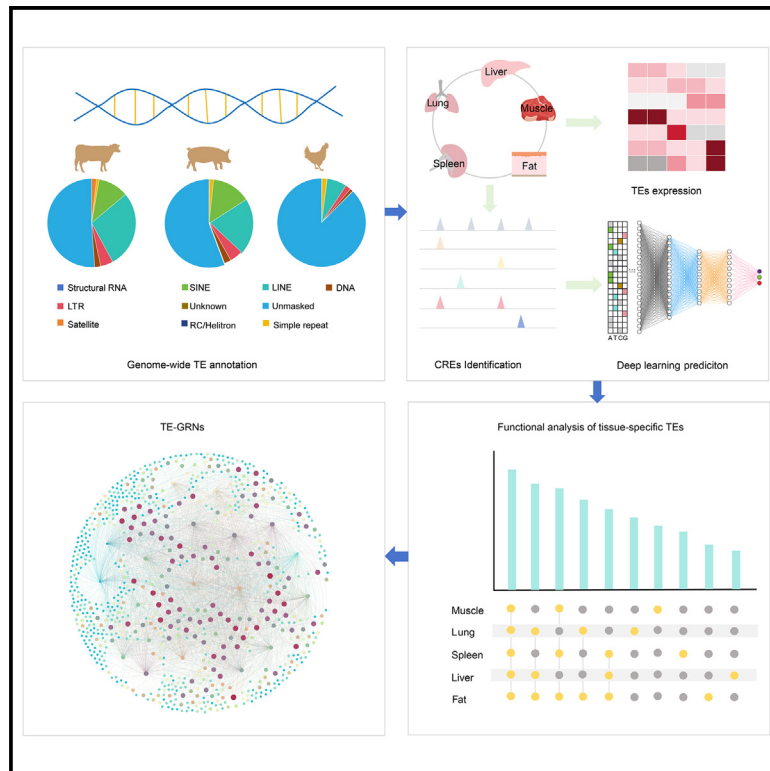


# Multi-omics analysis reveals critical *cis*-regulatory roles of transposable elements in livestock genomes

## Graphical abstract



## Authors

Chao Wang, Bowen Lei,  
Yongzhou Bao, ..., Tao Sun,  
Zhonglin Tang, Yuwen Liu

## Correspondence

liuyuwen@caas.cn

## In brief

Animals; Genomic analysis; Molecular  
Genetics; Sequence analysis

## Highlights

- Annotated and compared TE types, ages, and distributions in pig, cattle, and chicken genomes
- Integrated epigenomic data to reveal TE impacts on *cis*-regulatory elements across species
- Revealed tissue-specific TEs regulating tissue-specific biological processes
- Developed TE-mediated gene regulatory networks to reveal the genetic basis of complex traits



## Article

# Multi-omics analysis reveals critical *cis*-regulatory roles of transposable elements in livestock genomes

Chao Wang,<sup>1,2,3</sup> Bowen Lei,<sup>1,2,3</sup> Yongzhou Bao,<sup>1,3,4,5</sup> Zhen Wang,<sup>1,3,4,5</sup> Choulin Chen,<sup>1,2,3</sup> Yuanyuan Zhang,<sup>1,3,4,5</sup> Shenghua Qin,<sup>1,3</sup> Tao Sun,<sup>1,2,3</sup> Zhonglin Tang,<sup>1,2,3,6</sup> and Yuwen Liu<sup>1,2,3,6,7,\*</sup>

<sup>1</sup>Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Key Laboratory of Livestock and Poultry Multi-Omics of MARA, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, P.R. China

<sup>2</sup>Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education & Key Lab of Swine Genetics and Breeding of Ministry of Agriculture and Rural Affairs, Huazhong Agricultural University, Wuhan 430070, P.R. China

<sup>3</sup>Innovation Group of Pig Genome Design and Breeding, Research Centre for Animal Genome, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, P.R. China

<sup>4</sup>School of Life Sciences, Henan University, Kaifeng 475004, P.R. China

<sup>5</sup>Shenzhen Research Institute of Henan University, Shenzhen 518000, P.R. China

<sup>6</sup>Kunpeng Institute of Modern Agriculture at Foshan, Chinese Academy of Agricultural Sciences, Foshan 528226, P.R. China

<sup>7</sup>Lead contact

\*Correspondence: [liuyuwen@caas.cn](mailto:liuyuwen@caas.cn)

<https://doi.org/10.1016/j.isci.2025.112049>

## SUMMARY

Transposable elements (TEs) are important sources of genetic and regulatory variation, yet their functional roles in domesticated animals remain insufficiently explored. To address this gap, we comprehensively annotated TE types, ages, and distributions in the genomes of pig (*Sus scrofa*), cattle (*Bos taurus*), and chicken (*Gallus gallus*). Our analysis revealed species-specific patterns in TE abundance, amplification, and activity in modern genomes. By integrating transcriptomic and epigenomic data, we explored the impact of specific TE types on *cis*-regulatory elements (CREs) and constructed a TE expression atlas across five tissues in all three species. Our findings underscored the critical roles of tissue-specific TE expression and chromatin accessibility in regulating tissue-specific biological processes. Most notably, we developed a computational framework to uncover TE-mediated gene regulatory networks (TE-GRNs). Our findings provide valuable insights into the regulatory functions of TEs in livestock and offer a robust approach for studying TE-GRNs in diverse biological contexts.

## INTRODUCTION

Transposable elements (TEs) are a type of repetitive DNA sequence that makes up between 4% and 60% of vertebrate genomes.<sup>1</sup> Furthermore, many TEs are dynamic in that they can replicate and change position within their host genomes via transposition. TEs can be classified into two main classes based on their transposition mechanisms: DNA transposons and retrotransposons. DNA transposons primarily exploit a "cut and paste" mechanism to move from one genomic location to another, whereas retrotransposons use a "copy and paste" mechanism by being first transcribed into RNA and then reverse transcribed into DNA before being inserted into a new location in the genome.<sup>2–4</sup> Based on sequence composition, retrotransposons could be further divided into short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), and long terminal repeat (LTR) classes. TEs were initially regarded as "junk DNA" thought to merely invade the genome through their transposition ability. However, due to this replicability and

mobility, TEs are also an extensive source of mutations and genetic polymorphisms in genomic regions that potential manifest as functional.<sup>5–9</sup> Thus, by creating or deleting these important functional DNA sequences, TEs exhibit crucial roles in driving genome evolution and gene regulation.<sup>10–12</sup> For example, TEs can directly impact the coding sequences of genes<sup>13–16</sup> and provide raw material for the emergence of non-coding RNAs, such as lncRNAs,<sup>17</sup> microRNAs,<sup>18</sup> and other small RNAs.<sup>19</sup> Additionally, they can serve as foundational elements for DNA CREs, such as promoters and enhancers,<sup>20–22</sup> insulators,<sup>23</sup> and silencers.<sup>24</sup>

The diverse mechanisms by which TEs can exert influence on the genome have sparked significant interest in the development of versatile tools to study them. These tools include genome TE annotation tools, such as RepeatMasker,<sup>25</sup> REPET,<sup>26</sup> phRAIDER,<sup>27</sup> RepeatExplorer,<sup>28</sup> dnaPipeTE,<sup>29</sup> and DeepTE,<sup>30</sup> as well as TE expression quantification tools, such as Telescope,<sup>31</sup> Tetrascripts,<sup>32</sup> and SQUIRE.<sup>33</sup> Beyond annotating the genomic positions of TEs and detecting their expression levels,



the functional interpretation of TEs was greatly strengthened by functional genomics projects such as ModENCODE for model organisms<sup>34</sup> and ENCODE for humans.<sup>35</sup> These projects generated multi-omics functional genomics data that predict the *cis*-regulatory activity of DNA sequences, thus shedding light on how TEs impact the transcriptional regulatory networks by influencing CREs. For instance, Pehrsson et al. integrated data from the Roadmap Epigenomics Project to analyze the contribution of different TE classes to CREs predicted by multiple epigenomic marks across human anatomy and development.<sup>36</sup> Choudhary et al. used multi-species 3D genome data to identify TE families and subfamilies that impact lineage-specific chromatin structures during the evolution of gene regulation.<sup>37</sup> Chang et al. performed a systematic analysis of TE age and genomic distribution in zebrafish and used both bulk and single-cell transcriptomic data to explore TE expression patterns during development.<sup>38</sup> Lee et al. integrated multi-omics data to demonstrate that TEs significantly contribute to diverse tissue-specific CREs and in zebrafish. Their study revealed that TEs can drive the formation of promoters and interfere with gene transcription.<sup>39</sup> Together, these studies provide insights into the potential mechanisms by which TEs may influence complex diseases and traits.

Compared to humans and model organisms, research on TEs in livestock has received considerably less attention. Consequently, the impact of TEs on economic traits in livestock remains largely uncharted. Although recent studies in livestock functional genomics have generated multi-tissue CRE profiles in various domesticated animals,<sup>40–44</sup> the role of TEs in the formation of these CREs and their impact on gene regulatory networks (GRNs) in different tissues remains poorly understood.

In this study, we annotated TEs in the genomes of three major livestock species: pigs, cattle, and chickens. We conducted a comprehensive analysis of the genomic distribution, age, expression activity, and epigenetic regulation patterns of TEs by integrating ChIP-seq, ATAC-seq, and RNA-seq data from multiple tissues. To elucidate how TEs function as CREs to regulate biological processes, we also developed a computational framework to construct TE-GRNs. Overall, our study established a foundation for future work on analyzing how TEs contribute to the genetic basis of economically important traits in livestock.

## RESULTS

### The genomic landscape of TEs in the genomes of pigs, cattle, and chickens

We annotated TEs in three crucial livestock species using the RepeatMasker software (version 4.1.2).<sup>25</sup> In total, 3,953,666; 4,984,795; and 322,047 TEs were identified in pigs, cattle, and chickens, respectively. TEs account for approximately 9.6% of the chicken genome and a significantly higher proportion in the pig genome (43.4%) and cattle genome (47.5%) ( $p < 2.2 \times 10^{-16}$  z-test) (Figure 1A). This is consistent with previous research that demonstrated a positive correlation between the TE proportion and the genome size.<sup>45–47</sup> We further explored the genome proportion of different TE classes across the three species and found that the LINE class had the highest genome proportion in all three species, followed by the SINE class in the genomes of pigs and cattle. However, the genome proportion of the

SINE class in the chicken genome was significantly lower ( $p < 2.2 \times 10^{-16}$  z-test) (Figure 1B). Additionally, we investigated the relative abundance of each TE class among all annotated TEs in the genome. In pigs and cattle, SINE and LINE retrotransposons were found to be the most abundant, whereas in chickens, LINEs were the prevailing retrotransposons, with a considerably lower proportion of SINEs compared to pigs and cattle. Notably, in pigs and cattle, although SINE has a higher count than LINE, LINE has a higher genome proportion than SINE (Figures 1B and 1C).

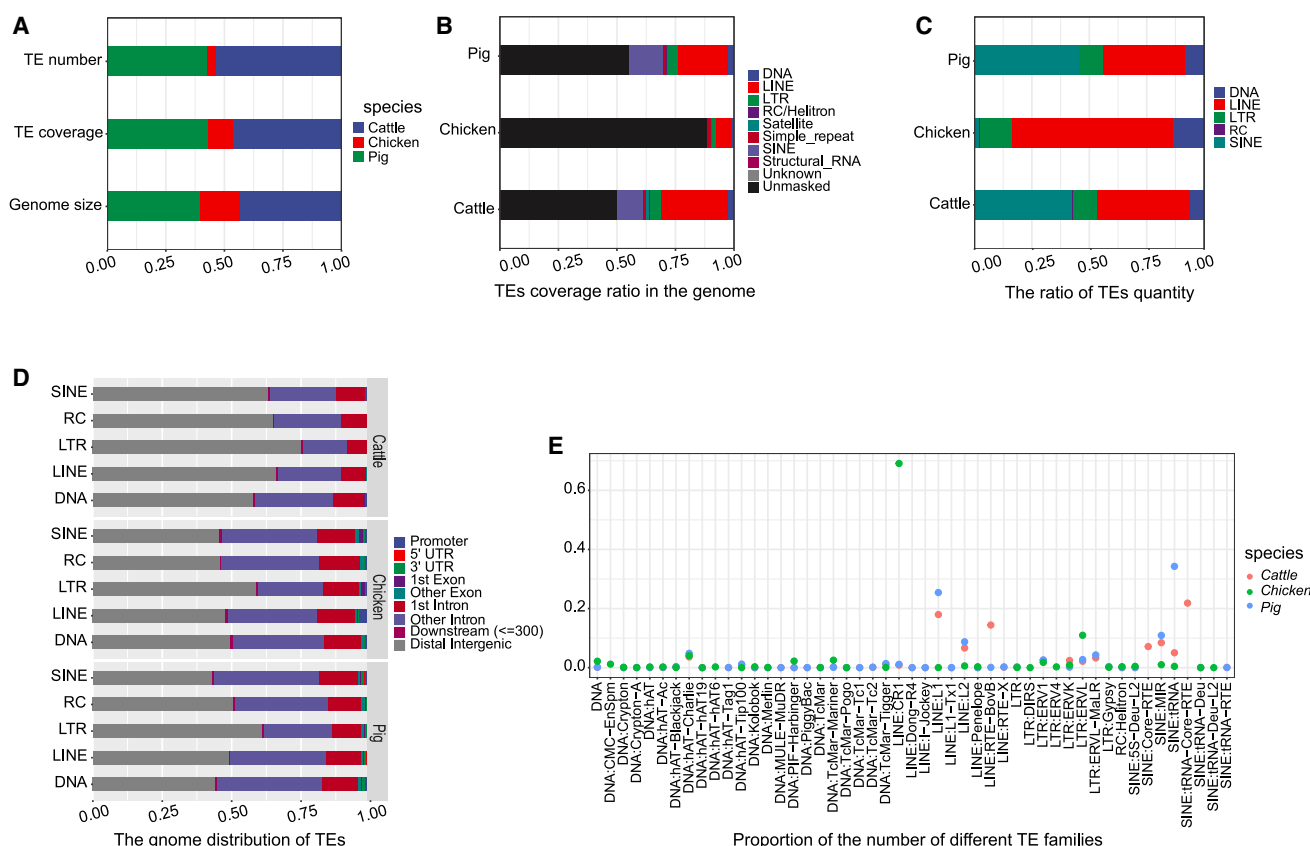
Depending on their genomic insertion sites, TEs can potentially act as CREs to regulate gene expression or directly affect protein-coding sequences of genes. We analyzed the genomic distribution of different TE classes and families and found that, in all three species, TEs tended to insert into distal intergenic and intronic regions (Figures 1D and S1A–S1C), suggesting the potential for TEs to function as enhancers.<sup>48,49</sup> Additionally, we found LTR retrotransposons exhibited a higher preference in the non-coding genomic regions compared to other types of TEs (Figure 1D).

To further understand the contribution of different TE families to genome composition, we analyzed the proportion of each TE family in the total annotated TEs. We found that each species has species-specific TE families that substantially outnumbered other TE families. For instance, in chickens, the LINE/CR1 family exhibits a notably higher proportion compared to other retrotransposon families. In pigs, the SINE/tRNA family displays a particularly elevated proportion. On the other hand, in cattle, the families of SINE/Core-RTE, SINE/tRNA-Core-RTE, and LINE/RTE\_BovB show relatively higher proportions (Figure 1E). This result indicates that a TE family may have different functional importance and thus were subject to species-specific selection during evolution.

Overall, we successfully annotated TEs in the genomes of three major livestock species, namely chicken, pig, and cattle, and compared the similarities and differences in their genomic distribution and abundance, providing a solid foundation for further research on how TEs function in the livestock genomes.

### Amplification patterns of TEs differ among different livestock species

The abundance and genomic distribution of TEs are closely linked to the evolutionary dynamics of TEs. Therefore, we investigated the transposition rate of TE classes at different time points during species evolution in pigs, cattle, and chickens. Our findings revealed that each species exhibited distinct bursts of TE amplification. Although LINEs and SINEs were more abundant than other TEs in pigs and cattle, their burst patterns differed between the two species. Pigs experienced two bursts, whereas cattle had three (Figures 2A and 2B). In chickens, both SINEs and LINEs had only one amplification burst, but LINEs maintained high levels of amplification while SINEs lost their ability to amplify in the modern chicken genome (Figure 2C). Moreover, DNA TEs had one burst in pigs and cattle but two in chickens. On the other hand, LTRs displayed only one burst in all three species (Figures 2A–2C). We further investigated the transposition rate of TE families within each class in the three species. Our analyses showed that, within a given TE class,



**Figure 1. TE annotation and genomic landscape across pigs, cattle, and chickens**

(A) TE quantity, genomic coverage, and genome size proportion distribution in the three species.  
(B) Comparison of genomic coverage of different TE classes in the three species.  
(C) Proportion of different TE classes among all annotated TE classes in the three species.  
(D) Genomic distribution of different TE classes in the three species.  
(E) Proportion of different TE families among all annotated TE families in the three species.

distinct TE families exhibited varying patterns of amplification (Figure 2D), as evidenced by differences in the timing and amplitude of burst. Accordingly, only some TE families are still active in the modern genomes of pigs, cows, and chickens. In the SINE class, Core-RTE and tRNA families are still active in cattle and pigs, respectively. Within the LINE class, RTE-BovB and L1 families in cattle, L1 family in pigs, and CR1 family in chickens are still capable of transposition. In the LTR class, ERVK and ERV1 families of cattle and ERVK and ERVL families of chickens still exhibit transposition activity, whereas all LTR families of pigs have lost their activity almost entirely. Notably, all DNA families have lost transposition activity in the modern genomes of all three species.

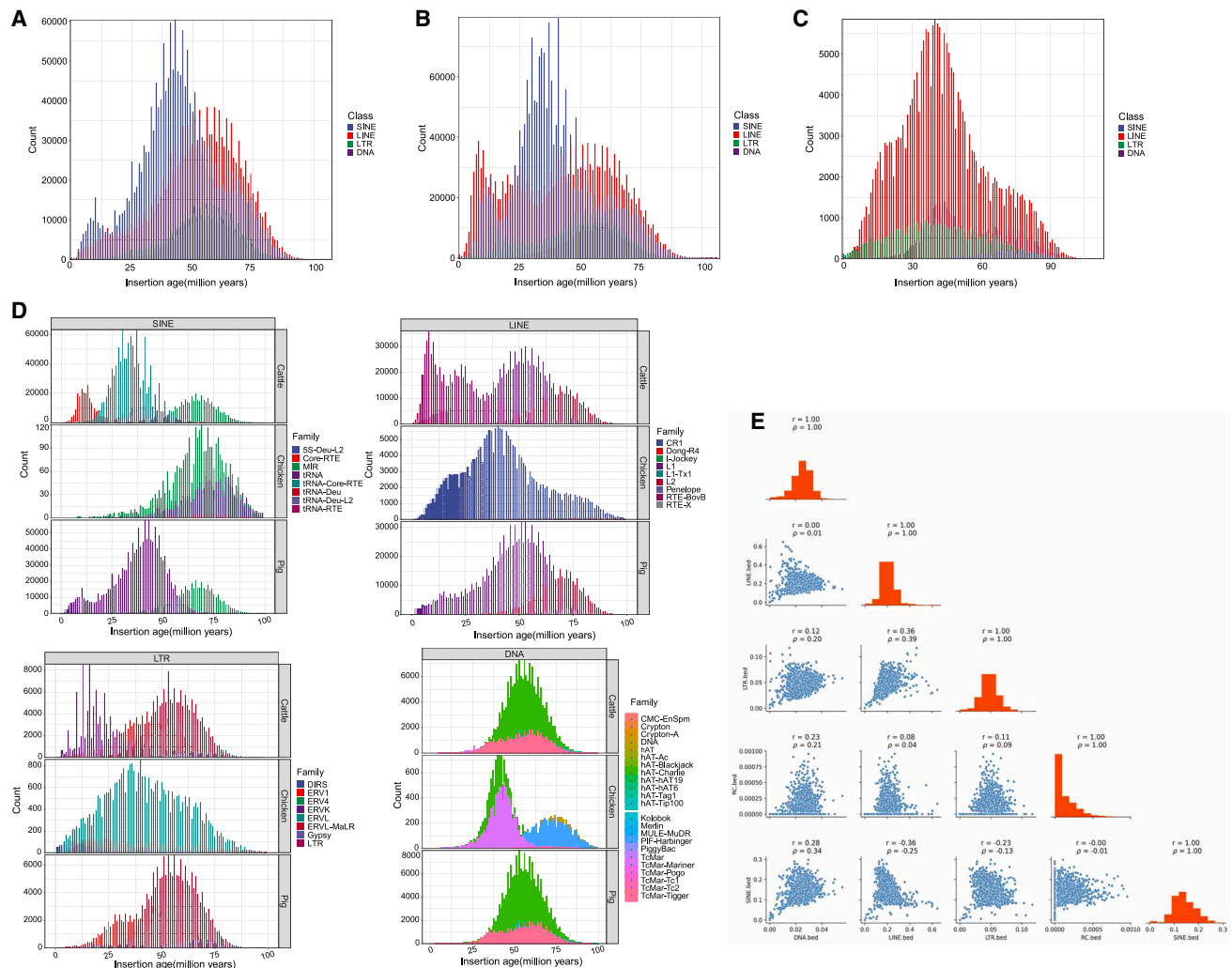
To explore the genomic distribution relationship of different TE classes, we adopted the approach used in a previous study, which calculated the density of each TE class (as genome sequence fraction) in non-overlapping 2-Mb windows along the genome and then analyzed the pairwise correlation between different TE classes.<sup>38</sup> Using this method, we revealed the correlation of genomic distribution between TE pairs of interest. Specifically, we found that LTR and LINE densities are positively correlated in all three species (Figures 2E, S1D, and S1E). In

addition, there are species-specific correlation patterns between TE classes. For example, in pigs and cattle, SINE density is negatively correlated with LINE density (Figures 2E and S1D). In chickens, LTR density exhibits a high negative correlation with SINE density, whereas LTR density is positively correlated with LINE and DNA densities, respectively (Figure S1E).

Overall, we systematically analyzed the amplification patterns and correlation of genomic density distribution for different TE classes in the three species, which may shed light on the significant events that these species have experienced throughout their evolutionary history.

### TEs play an important role in regulating tissue-specific chromatin accessibility

Open chromatin regions (OCRs) are exposed in intracellular environment, vulnerable to transcription factor (TF) binding and associated with *cis*-regulatory activity affecting gene expression. Using ATAC-seq data from five production-trait-related tissues (muscle, fat, lung, spleen, and liver) in the three livestock species, we explored the chromatin accessibility of different TE classes. The SINE class showed enrichment for open

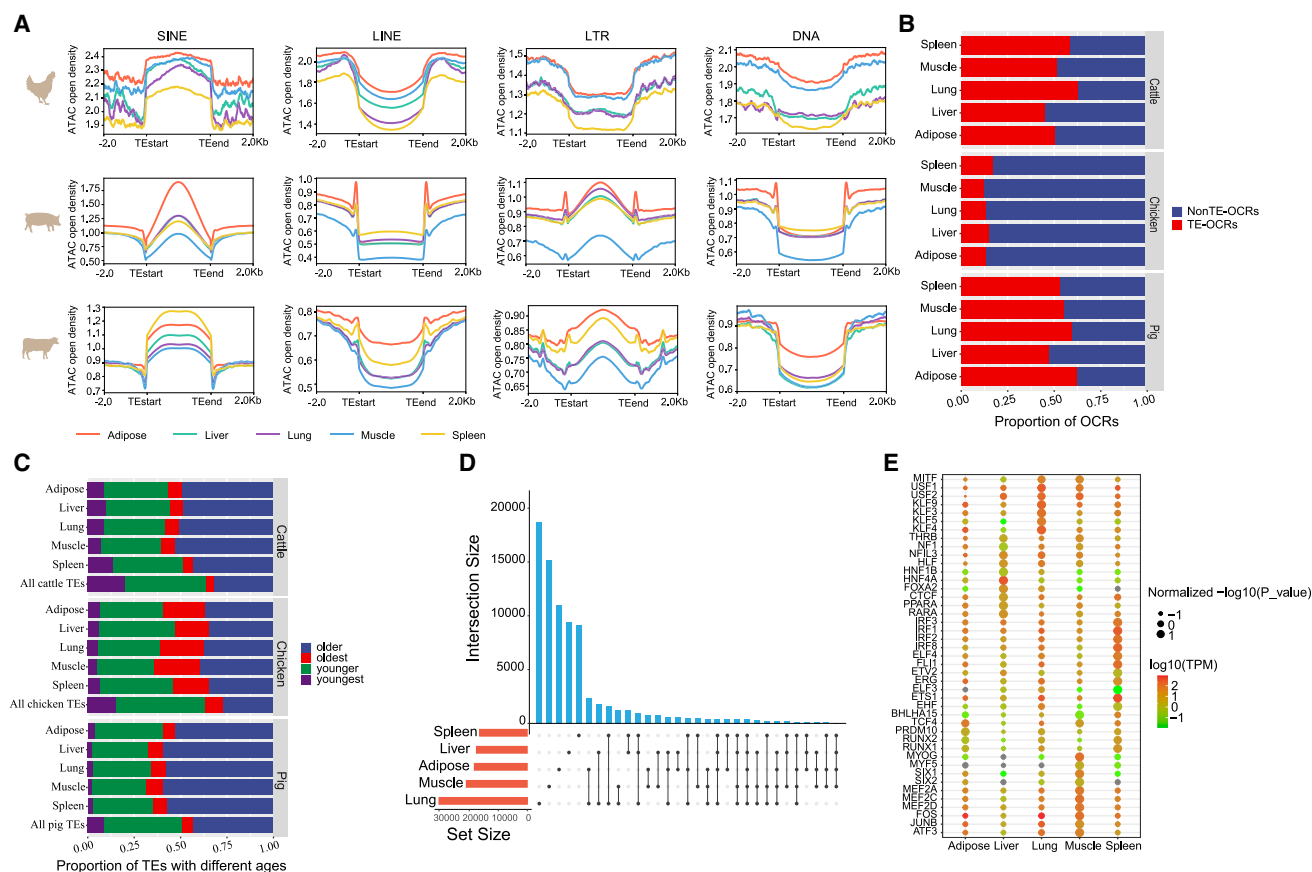


chromatin signals in every tissue of the three species, whereas the LINE and DNA classes exhibited depletion. An intriguing observation is the depletion of chromatin-accessible signals in LTR across all tissues in chickens. In contrast, pigs and cattle exhibit relatively high levels of enrichment of chromatin-accessible signals attributed to LTR (Figure 3A). These results suggest that SINEs have the potential to act as CREs in all three species, whereas the *cis*-regulatory potential of LTRs is specific to pigs and cattle.

To investigate the impact of TEs on OCRs in different tissues, we analyzed the number of OCRs overlapping TE insertions in each tissue. We defined OCRs with TE insertions as TE-driven OCRs (TE-OCRs). In pigs and cattle, TE-OCRs accounted for around 50% of all OCRs, whereas in chickens, they constituted

only approximately 15% of OCRs (Figure 3B). These findings suggest that the degree to which TEs contributed to creating CREs substantially varies between chickens and the other two species. To further explore the impact of different TE classes on OCRs in different tissues, we analyzed the proportion of the four major TE classes within TE-OCRs. Across tissues, we found consistent proportions of all four TE classes in TE-OCRs within each species. When compared to all annotated TEs, our analysis revealed a significant depletion of LINES and an overrepresentation of LTRs within TE-OCRs across all tissues in pigs and cattle ( $p < 2.2 \times 10^{-16}$  z-test) (Figure S4A; Table S1). In chickens, except for the spleen, LTRs showed a significant depletion within TE-OCRs, whereas SINEs exhibited a significant enrichment across all tissues ( $p < 2.2 \times 10^{-16}$  z-test) (Figure S4A; Table S1). These





**Figure 3. The effect of TEs on OCRs in various tissues of pigs, cattle, and chickens**

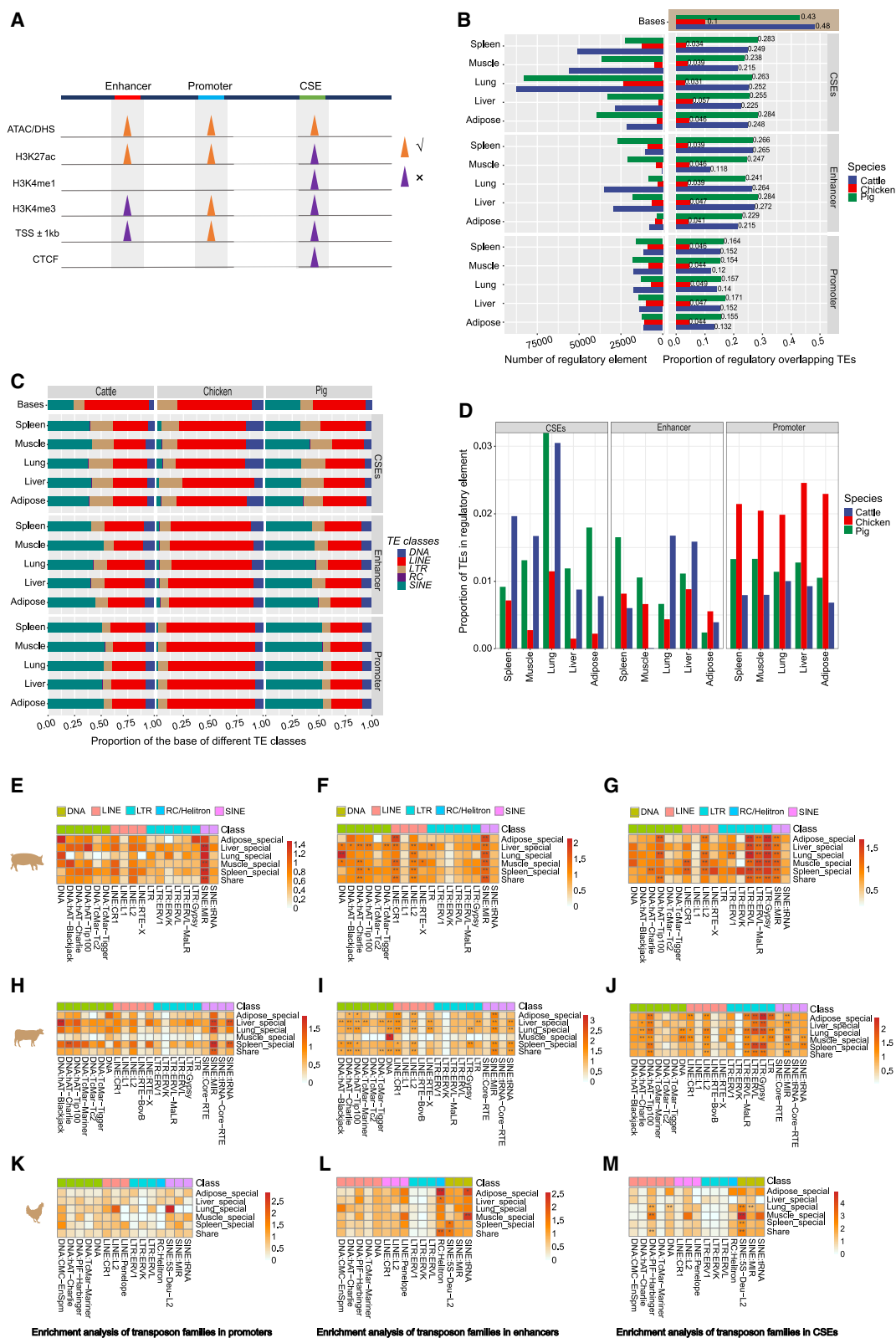
(A) The reads density distributions of chromatin accessibility of four major TE classes across different tissues in three species.  
 (B) The distribution of the proportion of TE-driven and non-TE-driven OCRs across different tissues in three species.  
 (C) The age distribution of OCRs-residing TEs in different tissues across three species.  
 (D) The bar plot shows the number of chromatin-accessible tissue-specific or shared TEs in pig across different tissues. The red bar represents the total number of TEs with chromatin accessibility features in that tissue, and the blue bar represents the number of TEs with chromatin accessibility features shared among specific tissues.  
 (E) Bubble plot showing significant enrichment of TF motifs in chromatin-specific accessible TEs in each tissue of pig, along with the expression levels of the corresponding TFs in that tissue. Select the top 20 significantly enriched TFs (Q value < 0.01) in each tissue and display only the TFs that are expressed in that tissue. The  $-\log_{10}(p\text{ value})$  values were scaled across tissues using the `scale()` function in R program.

findings suggest that the contribution of different TE classes to functional CREs is conserved across tissues but varied across species.

Furthermore, we classified TE into four distinct types based on their estimated age: youngest (0–25 MYa), younger (25–50 MYa), older (50–75 MYa), and oldest (75–100 MYa). This categorization allows for exploring the enrichment patterns of TEs of varying ages within TE-OCRs across diverse tissues. Using all annotated TEs as background, we found a significant enrichment of old TEs in TE-OCRs across different tissues in the three species, while young TEs were depleted ( $p < 0.01$ , z-test) (Figure 3C; Table S2). This finding indicates that the preference for older TEs in forming CREs is not limited to a specific species but is, in fact, a cross-species phenomenon. Genomes may predominantly preserve TE-OCRs exhibiting functional and evolutionary advantages, specifically those contributing to enhanced adaptability. Consequently, it is plausible that the oldest TE-OCRs conferred

evolutionary advantages, leading to their preservation despite their long presence in the genome.

Moreover, we were interested in understanding the extent to which TEs contribute to tissue-specific CREs, as these CREs are typically involved in tissue-specific biological processes.<sup>41</sup> Across all the three species, we observed a clear pattern that the majority of TEs within TE-OCRs exhibit tissue-specific chromatin accessibility (Figures 3D, S2A, and S2B). We hypothesized that these tissue-specific TEs contributed to the formation of tissue-specific CREs, which recruit binding of TFs that play crucial roles in maintaining tissue-specific biological processes. To test this hypothesis, we performed motif enrichment analysis for tissue-specific TEs. After selecting the top 20 enriched TF motifs in tissue-specific TEs, we only retained TFs that were significantly enriched (Q value < 0.01) and expressed in the corresponding tissue. Surprisingly, we found that tissue-specific TEs in all three species were significantly enriched for DNA-binding motifs of



(legend on next page)

TFs that are closely relevant to tissue development and highly expressed in the corresponding tissues (Figures 3E, S2C, and S2D). This result highlights the important role in regulating tissue-related biological processes of tissue-specific TEs with tissue-specific chromatin accessibility.

### TEs affect the epigenetic states of CREs

In order to comprehensively investigate the impact of TEs on specific types of CREs, we classified CREs into promoters, enhancers, and candidate silencer elements (CSEs) based on their epigenetic states (detailed in the “STAR methods” section, Figure 4A). Based on this classification strategy, we successfully constructed an atlas of three types of CREs across five tissues in the three species (Figure 4B left panel; Table S3). To ensure that the identified CSEs can truly be classified as a specific type of element with a unique pattern of nucleotide compositions, we used deep learning to test whether CSEs could be distinguished from other sequences solely by their DNA sequences. The model we built achieved good predictive ability for CSEs, as demonstrated by high area under the curve (AUC) values in chickens (0.759), pigs (0.850), and cattle (0.857), respectively (Figures S3A–S3C). As expected, the model could also well predict enhancers and promoters. The performance of the model in predicting chicken CREs exhibits a slight decrease, potentially due to the limited amount of training data available for chickens.

To investigate the extent to which CREs were constructed on TEs, we analyzed the proportion of TE bases within CREs across five tissues in the three species (Figure 4B, right panel). With the TE base coverage in the genome as the background, we observed a depletion of TEs in CREs across different tissues in the three species. This pattern could be explained by the defense mechanisms of host genome against transposon insertion, which prevent the insertion of TEs into important functional DNA regions, such as CREs, as a way to minimize deleterious impact on the host's fitness.<sup>50</sup> Across tissues within a species, the TE base proportion of each CRE type present is similar. However, there were variations in the TE base proportions among different CRE types within a species. In pigs and cattle, enhancers and silencers showed a higher proportion of TEs compared to promoters. In chickens, the TE base proportion of CREs was notably smaller, aligning with the lower overall genome proportion of TEs in the chicken genome (Figure 4B, right panel). To further investigate how CREs were constructed on different TE classes, we investigated the contributions of different classes of TEs to CREs (Figure 4C). Using the base proportions of different TE classes in the genome as the background, we found notable variations in the preference of different TE classes within CREs in

pigs, chickens, and cattle. Specifically, in pigs and cattle, LINEs exhibited depletion across all types of CREs, whereas SINEs displayed enrichment within CREs, except for pig CSEs. Additionally, both pig and cattle CSEs exhibited enrichment of LTRs. In chickens, DNA elements were enriched in CSEs, whereas LINEs were enriched in enhancers and promoters (Figure 4C).

We also examined the percentage of TEs that overlapped with promoters, enhancers, and silencers (Figure 4D). Our analysis revealed that the proportion of TEs overlapping enhancers and silencers varied considerably across tissues in all three species, whereas the proportion of TEs overlapping promoters remained relatively stable. Notably, in chickens, the proportion of TEs overlapping enhancers and silencers was relatively low compared to that of pigs and cattle. However, the proportion of TEs overlapping promoters in chickens is higher. These findings suggest that TEs in chickens may have a greater propensity to directly engage in mediating promoter function. Moreover, we observed that the proportion of different TE classes overlapping each CRE type were stable across tissues within each species. However, notable differences were observed in the CRE-overlapping proportion of some TE classes across species (Figure S4B).

To further explore the impact of different TE families on different types of CREs, we categorized the TEs by family and the CREs by tissue-specificity. Subsequently, we utilized permutation tests to evaluate the enrichment of different TE families in CREs specifically active in various tissues. In all three species, we found that fewer TE families were enriched in tissue-specific promoters than in enhancers and CSEs (Figures 4E–4M). Particularly, the chicken promoters did not enrich for any type of TE families (Figure 4K). These results suggest that TEs made less contribution to promoters than to other CREs. Additionally, in pigs and cattle, we found that LTR families were more enriched in CSEs than in enhancers of different tissues (Figures 4F, 4G, 4I, and 4J), indicating that LTRs generally have a greater impact on CSEs than enhancers. An important observation is that we found a significant enrichment of LTR families in the CSEs of various tissues in pigs and cattle but not in chickens (Figures 4G, 4J, and 4M). This finding provides a plausible explanation for the distinct chromatin accessibility patterns observed in LTRs among pig, cattle, and chicken tissues. Notably, LTRs exhibit a considerably higher level of accessibility in pig and cattle tissues, while demonstrating comparatively lower accessibility in chicken tissues (Figure 3A). To further investigate why LTRs have a greater impact on CSEs activity in pig and cattle tissues compared to chickens, we performed a motif enrichment analysis of LTR DNA sequences within the CSEs of the three species by

### Figure 4. The contribution of TEs to different CREs across different tissues and three species

- (A) Diagram of CREs categorization.  $\sqrt{\phantom{x}}$  indicates the presence of the epigenetic mark, and  $\times$  indicates the absence of the epigenetic mark.  
 (B) Number of three types of regulatory elements identified in different tissues of three species (left panel). The base proportion of CREs within TEs across five tissues in three species (right panel). The beige block shows the coverage of TE bases in the genome of three species.  
 (C) The distribution of base proportion of different TE classes within specific regulatory elements across different tissues of three species. Base proportion distribution of different TE classes in all annotated TEs of the genome as background.  
 (D) The proportion of TEs with specific regulatory element states among all annotated TEs in the genome across different tissues of three species.  
 (E–M) TE family enrichment in different types of CREs across the three species. (\*) FDR < 0.05 and (\*\*) FDR < 0.01 after correcting for multiple testing using the Bonferroni method.



HOMER software.<sup>51</sup> After selecting the top 20 significantly enriched TFs ( $Q < 0.01$ ) that also have orthologues in that species, we identified four TFs (MITF, USF1, CLOCK, MNT) that are shared between pigs and cattle, whereas no TFs are shared between chickens and either pigs or cattle (Figure S4C). Importantly, among the four TFs enriched in LTRs of CSEs in pigs and cattle, USF1 has been reported to display transcriptional repressor activity.<sup>52,53</sup> This finding proposed possible mechanism that LTRs might affect the activity of CSEs specifically in cattle and pigs by recruiting repressive TFs.

Altogether, our analyses offer a theoretical framework for assessing the significance of TEs in species-specific elements and CREs. This framework facilitates further investigation into how TEs reconfigure transcriptional regulatory networks in diverse tissues and contribute to the development of complex traits.

### Tissue-specific expression of TEs are involved in tissue-related biological processes

TEs not only function as CREs but they can also regulate gene expression in *trans* through their transcription. In order to investigate the potential impact of TE expression on biological processes, we downloaded transcriptome data from five tissues of chickens, pigs, and cattle, each with two biological replicates. We then utilized SQUIRE software to construct a TE expression atlas across tissues within each species.<sup>33</sup> Our study defined a TE as transcriptionally detectable in a tissue when its average FPKM in that tissue is greater than 1. Interestingly, we found that although there were significant differences in the number of detectable TEs among each species and across different tissues, the cross-tissue fluctuation pattern in the number of detectable TEs was similar among pigs, cattle, and chickens, with the spleen having the highest detectable TEs and muscle the lowest (Figure 5A). Subsequently, we analyzed the proportion of detectable TE classes in different tissues (Tables S4A–S4C). Compared to all annotated TEs, SINEs were preferentially expressed in each tissue of the three species. However, LINES still comprised the largest proportion of detectable chicken TEs due to their predominant presence in the chicken genome (Figure 5B).

As TE expression exhibited tissue specificity in different species by K-means clustering (Figures S5A–S5F), we further identified TEs detectable in a tissue-specific manner in the three species (Figure 5C). Depending on their genomic insertion sites, TEs in the non-coding regions can potentially act as CREs to regulate gene expression. We performed enrichment analysis of tissue-specific TEs and non-tissue-specific TEs in the non-coding region of different species. In pigs and cattle, tissue-specific TEs were significantly enriched in the non-coding region compared to non-tissue-specific TEs, whereas this difference was not significant in chickens (Figure 5D). In addition, we found that non-tissue-specific TEs in pigs and cattle, but not in chickens, were more enriched in highly conserved DNA sequences compared to tissue-specific TEs (Figure 5E). These findings suggest that, compared to those of chickens, TEs in the genomes of pigs and cattle made a more prominent contribution to tissue-specific CREs. In general, CREs specifically active in a certain tissue are known to play a crucial role in biological processes related to that

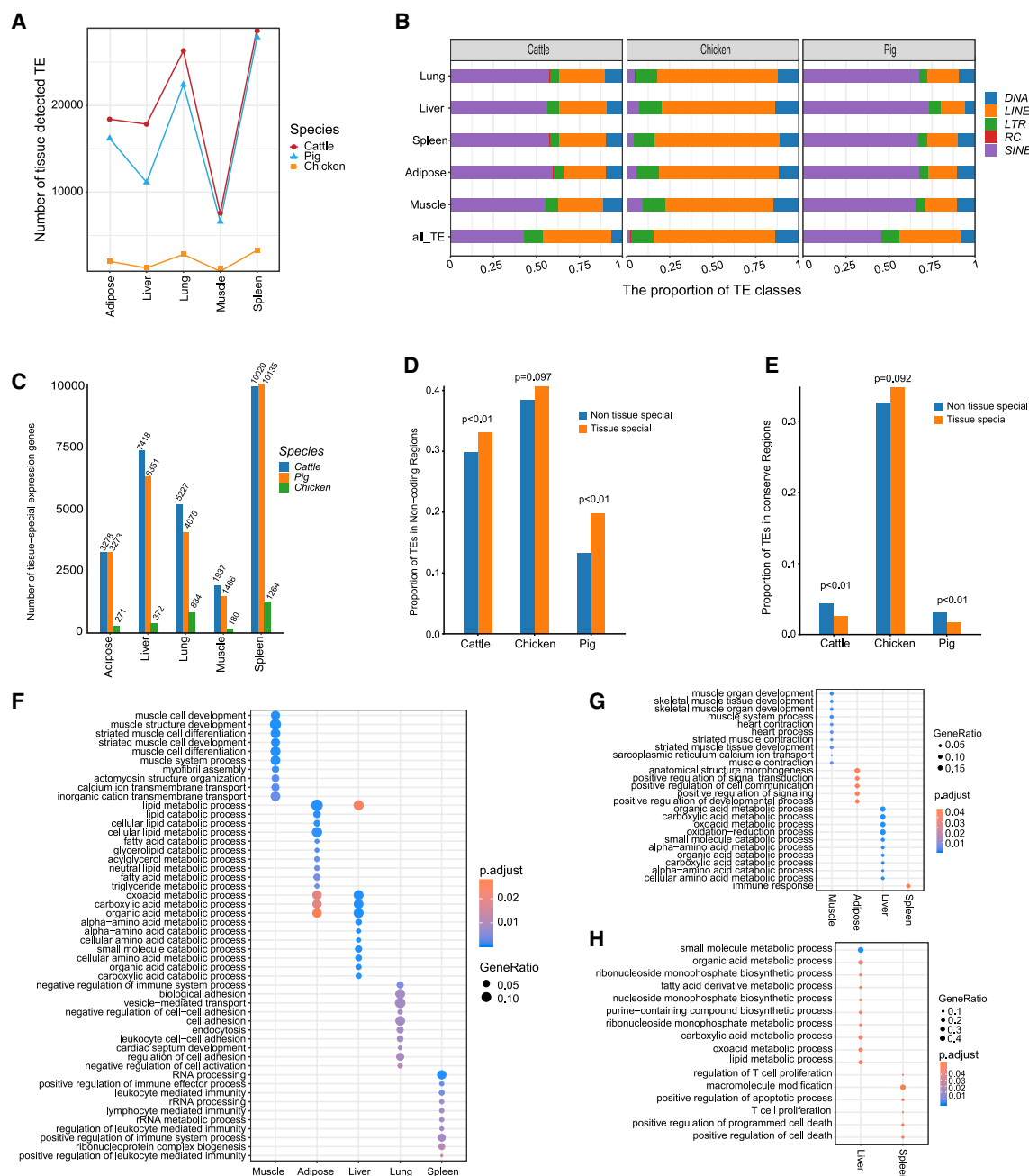
tissue. In light of this knowledge, we investigated whether tissue-specific TEs also revealed similar regulatory capacities. As TEs are known to have the potential to regulate the expression of their adjacent genes,<sup>54,55</sup> we performed gene ontology (GO) analysis on the neighboring genes of tissue-specific TEs (Tables S5A–S5C). We found that in pigs, neighboring genes of tissue-specific TEs in muscle, fat, liver, lung, and spleen were significantly enriched in tissue-relevant biological processes, respectively (Figure 5F). Cattle had a similar enrichment pattern except for the lung (Figure 5G). However, such an enrichment pattern was only observed in the liver and spleen in chickens (Figure 5H).

Overall, our results suggest that TE expression exhibits tissue specificity in different species and that the tissue-specific expression of TEs implies regulatory potential in maintaining tissue-specific biological functions, especially in mammals.

### Using tissue-specific TEs to dissect the GRNs underlying tissue-related biological processes

TE expression is regulated by chromatin modifications and TF binding.<sup>56</sup> Our study has revealed the biological significance of TEs that have tissue-specific expression or tissue-specific chromatin accessibility. Consequently, we proceeded to identify hub TEs in a tissue, which we defined as those exhibiting both tissue-specific expression and chromatin accessibility. To further study how tissue-specific TEs participate in tissue-relevant biological processes, we developed a computational framework to construct TE-GRNs. In contrast to conventional undirected GRNs, we leveraged these hub TEs to construct a directed GRN that encompasses transcriptional regulatory signaling from upstream TFs to TEs, then to downstream target genes. The specific strategy, illustrated in Figure 6A, involves not only identifying TE-upstream TFs and TE-downstream target genes using a co-expression approach but also requires DNA sequence motif scanning to identify upstream TFs and examine TE-neighboring genes for potential downstream target genes.

Since the liver is an important metabolic organ and substantially affects many economic traits of domesticated animals,<sup>57–59</sup> we used it as an example to demonstrate the construction of TE-GRNs underlying liver-relevant biological processes in pigs, cattle, and chickens (Figures 6B–6D; Tables S6A–S6C). Analyzing TE-GRNs within the liver across the three species yielded insights into the shared and species-specific TFs upstream of hub TEs. For example, we identified four TE-upstream TFs (LHX2, NR5A2, HNF4A, and GATA4) that were common to all three species. Additionally, we discovered two TFs (ETV2 and SP5) specific to pigs and four TFs (ELF5, HAND2, PAX8, and CUX2) specific to cattle. Chickens exhibited two species-specific TFs (FOXA1 and TBX6) (Figure 6E). Notably, among the identified upstream TFs associated with these hub TEs in the three species, we observed a significant number of TFs that exhibited a remarkable enrichment of motifs within TEs with liver-specific chromatin accessibility, including HNF4A, the FOXA gene family, and PPARG (Figures 3E, S2C, and S2D). HNF4A,<sup>60,61</sup> GATA4,<sup>62,63</sup> FOXA gene family,<sup>64,65</sup> and PPARG<sup>66</sup> have been reported to play important regulatory roles in liver-relevant biological processes. These findings support the authenticity and reliability of our approach in identifying key TFs acting upstream of TEs. To further confirm that the liver



**Figure 5. TEs specifically expressed in tissues are involved in tissue-related biological processes**

(A) Dynamic changes in the number of detectable expressed TEs in different tissues of three species.

(B) Comparison of the proportion of different classes of TEs among the detectable expressed TEs in each tissue (with the proportion of different classes of annotated TEs in the genome as background).

(C) Number of tissue-specific expressed TEs in different tissues of three species.

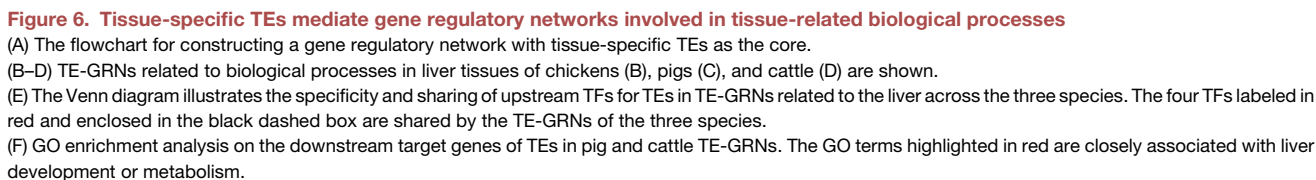
(D) The proportion of tissue-specific TEs and non-tissue-specific TEs in the non-coding region;  $p$  value is determined by Fisher's exact test.

(E) The proportion of tissue-specific TEs and non-tissue-specific TEs in the conserved element;  $p$  value is determined by Fisher's exact test.

(F–H) GO enrichment analysis of tissue-specific expressed TEs in pigs (F), cattle (G), and chickens (H) by extracting neighboring genes of tissue-specific expressed TEs. Only select the top 10 significantly enriched GO terms for tissue-specific expressed TEs for display.

TE-GRNs are involved in important biological processes of the liver, we conducted GO analysis of the genes in these networks of the three species, respectively. Genes in TE-GRNs of pigs and

cattle are significantly enriched in metabolic pathways, and genes in pig TE-GRNs are also enriched in the biological pathways of liver development (Figure 6F). However, genes in chicken



TE-GRN did not enrich in liver-related biological processes. This could potentially be attributed to the small number of genes in chicken TE-GRN and a correspondingly lower statistical power for enrichment analysis. Nevertheless, among the TE-downstream target genes in chicken TE-GRN, several genes are closely related to metabolism, including PAQR9, which can influence hepatic ketogenesis and fatty acid oxidation by regulating the protein stability of PPARA.<sup>67</sup> Additionally, PIT54 is highly expressed in the chicken liver and may be closely associated with liver physiology, function, and development.<sup>68</sup>

In summary, we presented a computational framework to decipher TE-GRNs underlying liver-related biological processes. Our strategy not only facilitates future understanding of the genetic basis of liver-related economic traits, such as feed conversion rate and lipid metabolism, but also holds great promises for dissecting the molecular underpinnings of other important economic traits by adopting a TE-centric perspective.

## DISCUSSION

TEs are crucial component of the genome that can influence diseases or phenotypes by modulating gene regulatory networks. With the growing recognition of the functional significance of TEs, an expanding body of research has embraced the integration of multi-omics data. This integration has facilitated investigations into the regulatory role of TEs across a diverse array of biological processes, not only in humans but also in model animals. Nevertheless, the livestock genetics community has been lagging behind in generating such data, leading to a shortage of research that comprehensively explore the regulatory landscape of TEs in domesticated animals.

In our research, we constructed comprehensive maps of TEs in pigs, cattle, and chickens and compared their evolutionary dynamics as well as contributions to different types of CREs. We systematically analyzed the expression of TEs across tissues, revealing the importance of tissue-specific TEs in maintaining tissue-specific biological processes. Given the important role of TEs in rewiring transcriptional regulatory networks, we presented a computational framework to construct TE-GRNs based on tissue-specific TEs. As a proof of concept, we built TE-GRNs in the liver tissues of pigs, cattle, and chickens, identifying TE-upstream TFs and TE-downstream target genes that gain multiple lines of literature support. This study not only explores the TE regulatory landscape across multiple tissues in three major domesticated animals but also presents a computational framework to study how TEs influence molecular pathways underlying complex traits and diseases.

The genome annotation of TEs is essential for the study of any TE-related biological questions. Following the TE annotation efforts in several previous studies,<sup>36–39</sup> we used RepeatMacker software to annotate TEs in the reference genomes of pigs, cattle, and chickens, respectively. Our analysis showed that TEs in pigs and cattle accounted for 43.4% and 47.5% of their respective genomes, whereas in chickens, TEs made up only 9.6% of the genome. Given that propagation of TE copies during evolution have expanded the genome size, previous studies have suggested a positive correlation between the level of TE proportion and the size of a genome.<sup>45,46</sup> Consistently, this correlation holds

true for TE proportions across the genomes of pigs, cattle, and chickens. Furthermore, we have identified highly abundant species-specific TE families, such as SINE/tRNA in pigs; LINE/CR1 in chickens; and SINE/Core-RTE, SINE/tRNA-Core-RTE, and LINE/RTE\_BovB in cattle. Our analysis also confirms several previous studies, which showed that Chicken Repeat 1 (CR1) was the most abundant repeat family in chickens,<sup>69</sup> SINE/tRNA the most abundant repeat family in pigs,<sup>70</sup> and non-LTR LINE RTE (BovB) and BovB-derived SINEs the cattle-specific repeats.<sup>71</sup>

In this study, we analyzed the age distribution of all annotated TEs in three species based on the single-nucleotide mutation rate in each species' genome. By comparing the burst patterns of different TE classes and families during evolution, we identified TE types that still exhibit transposition activity in the modern genome of specific species. For instance, SINE/tRNA and LINE/L1 have undergone high levels of amplification during pig evolution and remain active in the modern pig genome. Notably, a recent burst in SINE/tRNA was observed, consistent with previous findings that SINE/tRNA is currently the most active transposon type in the modern pig genome.<sup>72–74</sup> In this work, we also analyzed the relationship between the genomic distribution density of different classes of TEs in the three species and found that the correlation levels of genomic distribution density between different TEs have species-specific and shared pattern. For example, LTR and LINE densities are positively correlated in all three species. SINE density is negatively correlated with LINE densities only in pigs and cattle, which is consistent with previous reports on mammals.<sup>38,75</sup> Additionally, SINE density shows a stronger negative correlation with LTR densities, whereas LINE density is positively correlated with LTR and DNA densities in chickens. These findings suggest that the modes of cooperation among different classes of TEs in maintaining genome stability may vary across species.

In the current study, we also systematically investigated the impact of TEs on epigenetic marks associated with CREs. Our findings showed that TEs have a greater contribution to the OCRs in pigs and cattle than in chickens (Figure 3B). This difference might imply the differential contributions of TEs to CREs between mammals and avians. Previous studies have demonstrated that TEs can act as CREs to regulate gene expression by carrying specific TF-binding motifs.<sup>39,76–80</sup> Similarly to He et al.,<sup>81</sup> we found that tissue-specific TEs enriched in pigs, cattle, and chickens contain numerous TF binding motifs related to tissue growth and development, suggesting that tissue-specific TEs are crucial for maintaining tissue-related biological processes. A genome contains diverse types of CREs that regulate gene expression through different molecular mechanisms. To further explore the similarities and differences in the effects of TEs on the various types of CREs in the three species, we identified promoters, enhancers, and CSEs in pigs, cattle, and chickens and built a multitask deep learning model that showed good performance in classifying the three CRE types by DNA sequences. We systematically analyzed the differences in the extent to which TEs contributed to CREs in the three species, which enhanced our understanding of the differential potential of TEs to influence CREs in different species.

In addition to their primary role in transposition, TEs exhibit a multifaceted impact on biological systems, influencing various

aspects such as gene expression, chromatin accessibility, activation of cellular signaling pathways (including the interferon response), RNA interference responses, and even contributing to aging processes or antiviral activities.<sup>82</sup> Genes expressed in a tissue-specific manner are intricately linked to the tissue-relevant biological processes. Although there are several sources available for tissue-specific gene expression, including HPA,<sup>83</sup> TIGER,<sup>84</sup> GTEx,<sup>85</sup> and TisGDB,<sup>86</sup> research on tissue-specific TE expression is still limited. In this study, we not only constructed TE expression maps for five tissues in the pigs, cattle, and chickens but also identified TEs with tissue-specific expression patterns. We showed that these tissue-specific TEs are intimately associated with tissue-related biological processes. Therefore, our study provides a valuable reference for future investigations on how TEs mediate epigenetic regulation in the evolution and domestication of pigs, cattle, and chickens.

To further dissect how TEs regulate various biological processes, we developed a computational framework to construct TE-GRNs. In the liver tissue, the TE-GRNs consisted of TE-downstream target genes that are enriched in liver-related biological processes, as well as TE-upstream TFs that include liver-related pioneering TFs, which supports the reliability of our network construction approach. While our TE-GRN is a promising advancement in understanding TE biology, there is still considerable room for improvement in its accuracy and comprehensiveness. With the increasing availability of different types of multi-omics data in various biological contexts, identifying TE-downstream target genes can be improved by chromatin looping information from HiC data, and identifying TE-upstream TFs can be facilitated by high-throughput protein-DNA interaction data, such as ChIP-Seq and CUT&Tag. With the increasing generation of scRNA-seq and scATAC-seq data across diverse biological contexts and the ongoing development of quantitative tools for assessing single-cell TE expression,<sup>55,81</sup> the TE-GRN construction approach could be adopted to studying how TEs contributed to transcriptional regulatory networks at cell-type-specific level. This will enable more powerful analyses of the TE-mediated phenotype variations that exclusively manifest within certain cell types. Overall, by revealing the molecular pathways involved in tissue-related biological processes, the development of TE-GRNs represents a valuable resource for future investigations into the molecular basis of phenotypic variation, from a TE-centric perspective.

In this study, we comprehensively analyze TEs in three major domesticated animals: pigs, cattle, and chickens, exploring their potential roles in genomic evolution, epigenetic mechanisms, and gene expression regulation. In addition, we present a computational framework for constructing TE-GRNs, providing a better understanding of the impact of TEs on complex traits and diseases. Our research significantly contributes to the understanding of the importance of TEs and their potential roles in shaping the evolutionary trajectory and elucidating complex traits of domesticated animals.

### Limitations of the study

Although we have constructed a universal TE-GRN to elucidate complex biological processes through multi-omics integration, the application of TE-GRN has certain limitations. For instance,

due to the source and quality of the omics data, we only used five tissues to identify tissue-specific expression and accessibility of TEs. This limited number of tissue types may introduce some bias in identifying tissue specificity. Additionally, the omics data used in this study for the three species were obtained from specific developmental stages, which might not capture the dynamic nature of development in constructing regulatory networks. Furthermore, methods that rely on histone markers to identify regulatory elements can result in false positives. These markers are only indicators of regulatory elements and do not establish causality. High-throughput experimental technologies, such as STARR-seq, that directly measure the activity of *cis*-regulatory elements will offer a more precise assessment of their regulatory capacity, potentially enhancing the accuracy of regulatory network construction.<sup>87,88</sup> In conclusion, as multi-omics data for livestock and poultry continue to accumulate, we believe that our methodological framework will achieve better performance and broader applicability.

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Yuwen Liu ([liuyuwen@caas.cn](mailto:liuyuwen@caas.cn)).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

- The all raw sequencing data utilized in this study are available in the Gene Expression Omnibus (GEO) under accession number GSE158430.
- The original code, TE and gene expression matrix data, as well as model training data have been deposited at Zenodo and are publicly available as of the date of publication. Accession numbers are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

### ACKNOWLEDGMENTS

This work was supported by the China National Key R&D Program during the 14th Five-year Plan Period (2021YFF1200500 to Y.W.L.), the National Natural Science Foundation of China (32070595 and 32372858 to Y.W.L.), and the Shenzhen Innovation and Entrepreneurship Plan—Major special project of science and technology (KJZD20230923115003006 to Y.W.L.).

### AUTHOR CONTRIBUTIONS

Conceptualization, C.W. and Y.W.L.; methodology, C.W.; investigation, C.W., B.W.L., Z.W., Y.Z.B., Y.Y.Z., and S.Q.; writing—original draft, C.W.; writing—review & editing, C.W. and Y.W.L.; funding acquisition, Y.W.L.; supervision, Y.W.L. and Z.L.T.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)



# METHOD DETAILS

- Data sources
- TE annotation
- TE genomic landscape analysis
- Analyzing the impact of TEs on OCRs
- Identification and deep learning classification of different DNA CREs
- Contribution of TEs to different DNA CREs
- Enrichment analysis of TE families in different DNA CREs
- Quantification of TE loci expression and identification of tissue-specific TE expression
- Enrichment analysis of tissue-specific expressed TEs
- Construction of TE-GRNs

# QUANTIFICATION AND STATISTICAL ANALYSIS

# ADDITIONAL RESOURCES

# SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2025.112049>.

Received: January 14, 2024

Revised: July 10, 2024

Accepted: February 13, 2025

Published: February 18, 2025

# REFERENCES

1. Sotero-Caio, C.G., Platt, R.N., Suh, A., and Ray, D.A. (2017). Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* 9, 161–177. <https://doi.org/10.1093/gbe/evw264>.
2. Thomas, J., and Pritham, E.J. (2015). Helitrons, the eukaryotic rolling-circle transposable elements. *Microbiol. Spectr.* 3, 891–924. <https://doi.org/10.1128/microbiolspec.MDNA3-0049-2014>.
3. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. <https://doi.org/10.1038/35057062>.
4. Finnegan, D.J. (1989). Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5, 103–107. [https://doi.org/10.1016/0168-9525\(89\)90039-5](https://doi.org/10.1016/0168-9525(89)90039-5).
5. Eickbush, T.H., and Furano, A.V. (2002). Fruit flies and humans respond differently to retrotransposons. *Curr. Opin. Genet. Dev.* 12, 669–674. [https://doi.org/10.1016/s0959-437x\(02\)00359-3](https://doi.org/10.1016/s0959-437x(02)00359-3).
6. Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., and Mager, D.L. (2006). Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet.* 2, e2. <https://doi.org/10.1371/journal.pgen.0020002>.
7. Stewart, C., Kural, D., Strömberg, M.P., Walker, J.A., Konkel, M.K., Stütz, A.M., Urban, A.E., Grubert, F., Lam, H.Y.K., Lee, W.P., et al. (2011). A comprehensive map of mobile element insertion polymorphisms in humans. *PLoS Genet.* 7, e1002236. <https://doi.org/10.1371/journal.pgen.1002236>.
8. Zhao, P., Gu, L., Gao, Y., Pan, Z., Liu, L., Li, X., Zhou, H., Yu, D., Han, X., Qian, L., et al. (2023). Young sines in pig genomes impact gene regulation, genetic diversity, and complex traits. *Commun. Biol.* 6, 894. <https://doi.org/10.1038/s42003-023-05234-x>.
9. Zhao, P., Peng, C., Fang, L., Wang, Z., and Liu, G.E. (2023). Taming transposable elements in livestock and poultry: a review of their roles and applications. *Genet. Sel. Evol.* 55, 50. <https://doi.org/10.1186/s12711-023-00821-2>.
10. Biemont, C. (2010). A brief history of the status of transposable elements: from junk dna to major players in evolution. *Genetics* 186, 1085–1093. <https://doi.org/10.1534/genetics.110.124180>.
11. Biemont, C., and Vieira, C. (2006). Genetics: junk dna as an evolutionary force. *Nature* 443, 521–524. <https://doi.org/10.1038/443521a>.
12. Volff, J.N. (2006). Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* 28, 913–922. <https://doi.org/10.1002/bies.20452>.
13. Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405. <https://doi.org/10.1038/nrg2337>.
14. Lev-Maor, G., Ram, O., Kim, E., Sela, N., Goren, A., Levanon, E.Y., and Ast, G. (2008). Intronic alus influence alternative splicing. *PLoS Genet.* 4, e1000204. <https://doi.org/10.1371/journal.pgen.1000204>.
15. Xing, J., Wang, H., Belancio, V.P., Cordaux, R., Deininger, P.L., and Batzer, M.A. (2006). Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc. Natl. Acad. Sci. USA* 103, 17608–17613. <https://doi.org/10.1073/pnas.0603224103>.
16. Schmitz, J., and Brosius, J. (2011). Exonization of transposed elements: a challenge and opportunity for evolution. *Biochimie* 93, 1928–1934. <https://doi.org/10.1016/j.biochi.2011.07.014>.
17. Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., and Feschotte, C. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding mas. *PLoS Genet.* 9, e1003470. <https://doi.org/10.1371/journal.pgen.1003470>.
18. Piriyaopongsa, J., Mariño-Ramírez, L., and Jordan, I.K. (2007). Origin and evolution of human micromas from transposable elements. *Genetics* 176, 1323–1337. <https://doi.org/10.1534/genetics.107.072553>.
19. Mccue, A.D., and Slotkin, R.K. (2012). Transposable element small rnas as regulators of gene expression. *Trends Genet.* 28, 616–623. <https://doi.org/10.1016/j.tig.2012.09.001>.
20. Chuong, E.B., Rumi, M.A.K., Soares, M.J., and Baker, J.C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* 45, 325–329. <https://doi.org/10.1038/ng.2553>.
21. Jacques, P.É., Jeyakani, J., and Bourque, G. (2013). The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* 9, e1003504. <https://doi.org/10.1371/journal.pgen.1003504>.
22. Trizzino, M., Park, Y., Holsbach-Beltrame, M., Aracena, K., Mika, K., Caliskan, M., Perry, G.H., Lynch, V.J., and Brown, C.D. (2017). Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res.* 27, 1623–1633. <https://doi.org/10.1101/gr.218149.116>.
23. Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G.D., Marshall, A., Flícek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and ctcf binding in multiple mammalian lineages. *Cell* 148, 335–348. <https://doi.org/10.1016/j.cell.2011.11.058>.
24. Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430, 471–476. <https://doi.org/10.1038/nature02651>.
25. Tarailo-Graovac, M., and Chen, N. (2009). Using repeatmasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* 4, 4–10. <https://doi.org/10.1002/0471250953.bi0410s25>.
26. Quesneville, H., Nouaud, D., and Anxolabéhère, D. (2003). Detection of new transposable element families in drosophila melanogaster and anopheles gambiae genomes. *J. Mol. Evol.* 57, S50–S59. <https://doi.org/10.1007/s00239-003-0007-2>.
27. Schaeffer, C.E., Figueroa, N.D., Liu, X., and Karro, J.E. (2016). Phraider: pattern-hunter based rapid ab initio detection of elementary repeats. *Bioinformatics* 32, i209–i215. <https://doi.org/10.1093/bioinformatics/btw258>.
28. Novak, P., Neumann, P., Pech, J., Steinhaisl, J., and Macas, J. (2013). Repeatexplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence

reads. *Bioinformatics* 29, 792–793. <https://doi.org/10.1093/bioinformatics/btt054>.

29. Goubert, C. (2023). Assembly-free detection and quantification of transposable elements with dnapipe. *Methods Mol. Biol.* 2607, 25–43. [https://doi.org/10.1007/978-1-0716-2883-6\\_2](https://doi.org/10.1007/978-1-0716-2883-6_2).
30. Yan, H., Bombarely, A., and Li, S. (2020). Deepte: a computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics* 36, 4269–4275. <https://doi.org/10.1093/bioinformatics/btaa519>.
31. Bendall, M.L., de Mulder, M., Iñiguez, L.P., Lecanda-Sánchez, A., Pérez-Losada, M., Ostrowski, M.A., Jones, R.B., Mulder, L.C.F., Reyes-Terán, G., Crandall, K.A., et al. (2019). Telescope: characterization of the retro-transcriptome by accurate estimation of transposable element expression. *PLoS Comput. Biol.* 15, e1006453. <https://doi.org/10.1371/journal.pcbi.1006453>.
32. Jin, Y., Tam, O.H., Paniagua, E., and Hammell, M. (2015). Tetrascripts: a package for including transposable elements in differential expression analysis of rna-seq datasets. *Bioinformatics* 31, 3593–3599. <https://doi.org/10.1093/bioinformatics/btv422>.
33. Yang, W.R., Ardeljan, D., Pacyna, C.N., Payer, L.M., and Burns, K.H. (2019). Squire reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Res.* 47, e27. <https://doi.org/10.1093/nar/gky1301>.
34. Mouse ENCODE Consortium; Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., and Kaul, R. (2012). An encyclopedia of mouse dna elements (mouse encode). *Genome Biol.* 13, 418. <https://doi.org/10.1186/gb-2012-13-8-418>.
35. ENCODE Project Consortium (2012). An integrated encyclopedia of dna elements in the human genome. *Nature* 489, 57–74. <https://doi.org/10.1038/nature11247>.
36. Pehrsson, E.C., Choudhary, M.N.K., Sundaram, V., and Wang, T. (2019). The epigenomic landscape of transposable elements across normal human development and anatomy. *Nat. Commun.* 10, 5640. <https://doi.org/10.1038/s41467-019-13555-x>.
37. Choudhary, M.N.K., Quaid, K., Xing, X., Schmidt, H., and Wang, T. (2023). Widespread contribution of transposable elements to the rewiring of mammalian 3d genomes. *Nat. Commun.* 14, 634. <https://doi.org/10.1038/s41467-023-36364-9>.
38. Chang, N.C., Rovira, Q., Wells, J., Feschotte, C., and Vaquerizas, J.M. (2022). Zebrafish transposable elements show extensive diversification in age, genomic distribution, and developmental expression. *Genome Res.* 32, 1408–1423. <https://doi.org/10.1101/gr.275655.121>.
39. Lee, H.J., Hou, Y., Maeng, J.H., Shah, N.M., Chen, Y., Lawson, H.A., Yang, H., Yue, F., and Wang, T. (2022). Epigenomic analysis reveals prevalent contribution of transposable elements to cis-regulatory elements, tissue-specific expression, and alternative promoters in zebrafish. *Genome Res.* 32, 1424–1436. <https://doi.org/10.1101/gr.276052.121>.
40. Zhao, Y., Hou, Y., Xu, Y., Luan, Y., Zhou, H., Qi, X., Hu, M., Wang, D., Wang, Z., Fu, Y., et al. (2021). A compendium and comparative epigenomics analysis of cis-regulatory elements in the pig genome. *Nat. Commun.* 12, 2217. <https://doi.org/10.1038/s41467-021-22448-x>.
41. Kern, C., Wang, Y., Xu, X., Pan, Z., Halstead, M., Chanthavixay, G., Saelao, P., Waters, S., Xiang, R., Chamberlain, A., et al. (2021). Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nat. Commun.* 12, 1821. <https://doi.org/10.1038/s41467-021-22100-8>.
42. Zhu, Y., Zhou, Z., Huang, T., Zhang, Z., Li, W., Ling, Z., Jiang, T., Yang, J., Yang, S., Xiao, Y., et al. (2022). Mapping and analysis of a spatiotemporal h3k27ac and gene expression spectrum in pigs. *Sci. China Life Sci.* 65, 1517–1534. <https://doi.org/10.1007/s11427-021-2034-5>.
43. Fang, L., Liu, S., Liu, M., Kang, X., Lin, S., Li, B., Connor, E.E., Baldwin, R.L., 6th, Tenesa, A., Ma, L., et al. (2019). Functional annotation of the cattle genome through systematic discovery and characterization of chromatin states and butyrate-induced variations. *BMC Biol.* 17, 68. <https://doi.org/10.1186/s12915-019-0687-8>.
44. Pan, Z., Wang, Y., Wang, M., Wang, Y., Zhu, X., Gu, S., Zhong, C., An, L., Shan, M., Damas, J., et al. (2023). An atlas of regulatory elements in chicken: a resource for chicken genetics and genomics. *Sci. Adv.* 9, eade1204. <https://doi.org/10.1126/sciadv.ade1204>.
45. Gao, B., Shen, D., Xue, S., Chen, C., Cui, H., and Song, C. (2016). The contribution of transposable elements to size variations between four teleost genomes. *Mobile DNA* 7, 4. <https://doi.org/10.1186/s13100-016-0059-7>.
46. Talla, V., Suh, A., Kalsoom, F., Dinca, V., Vila, R., Friberg, M., Wiklund, C., and Backström, N. (2017). Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (leptidea) butterflies. *Genome Biol. Evol.* 9, 2491–2505. <https://doi.org/10.1093/gbe/evx163>.
47. Shao, F., Han, M., and Peng, Z. (2019). Evolution and diversity of transposable elements in fish genomes. *Sci. Rep.* 9, 15399. <https://doi.org/10.1038/s41598-019-51888-1>.
48. Du, A.Y., Zhuo, X., Sundaram, V., Jensen, N.O., Chaudhari, H.G., Saccone, N.L., Cohen, B.A., and Wang, T. (2022). Functional characterization of enhancer activity during a long terminal repeat's evolution. *Genome Res.* 32, 1840–1851. <https://doi.org/10.1101/gr.276863.122>.
49. Todd, C.D., Deniz, Ö., Taylor, D., and Branco, M.R. (2019). Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. *Elife* 8, e44344. <https://doi.org/10.7554/eLife.44344>.
50. Arkhipova, I.R. (2018). Neutral theory, transposable elements, and eukaryotic genome evolution. *Mol. Biol. Evol.* 35, 1332–1337. <https://doi.org/10.1093/molbev/msy083>.
51. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Mol. Cell* 38, 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>.
52. Hadjiagapiou, C., Borthakur, A., Dahdal, R.Y., Gill, R.K., Malakooti, J., Ramaswamy, K., and Dudeja, P.K. (2005). Role of usf1 and usf2 as potential repressor proteins for human intestinal monocarboxylate transporter 1 promoter. *Am. J. Physiol. Gastrointest. Liver Physiol.* 288, G1118–G1126. <https://doi.org/10.1152/ajpgi.00312.2004>.
53. Doni Jayavelu, N., Jajodia, A., Mishra, A., and Hawkins, R.D. (2020). Candidate silencer elements for the human and mouse genomes. *Nat. Commun.* 11, 1061. <https://doi.org/10.1038/s41467-020-14853-5>.
54. Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18, 71–86. <https://doi.org/10.1038/nrg.2016.139>.
55. Rodríguez-Quiroz, R., and Valdebenito-Maturana, B. (2022). Solote for improved analysis of transposable elements in single-cell rna-seq data using locus-specific expression. *Commun. Biol.* 5, 1063. <https://doi.org/10.1038/s42003-022-04020-5>.
56. He, J., Fu, X., Zhang, M., He, F., Li, W., Abdul, M.M., Zhou, J., Sun, L., Chang, C., Li, Y., et al. (2019). Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat. Commun.* 10, 34. <https://doi.org/10.1038/s41467-018-08006-y>.
57. Zhao, Y., Hou, Y., Liu, F., Liu, A., Jing, L., Zhao, C., Luan, Y., Miao, Y., Zhao, S., and Li, X. (2016). Transcriptome analysis reveals that vitamin a metabolism in the liver affects feed efficiency in pigs. *G3 (Bethesda)* 6, 3615–3624. <https://doi.org/10.1534/g3.116.032839>.
58. Horodyska, J., Hamill, R.M., Reyer, H., Trakooljul, N., Lawlor, P.G., McCormack, U.M., and Wimmers, K. (2019). Rna-seq of liver from pigs divergent in feed efficiency highlights shifts in macronutrient metabolism,

- hepatic growth and immune response. *Front. Genet.* 10, 117. <https://doi.org/10.3389/fgene.2019.00117>.
59. Luan, Y., Zhang, L., Hu, M., Xu, Y., Hou, Y., Li, X., Zhao, S., Zhao, Y., and Li, C. (2019). Identification and conservation analysis of cis-regulatory elements in pig liver. *Genes* 10, 348. <https://doi.org/10.3390/genes10050348>.
  60. Huang, K.W., Reebye, V., Czyst, K., Ciriello, S., Dorman, S., Reccia, I., Lai, H.S., Peng, L., Kostomitsopoulos, N., Nicholls, J., et al. (2020). Liver activation of hepatocellular nuclear factor-4alpha by small activating rna rescues dyslipidemia and improves metabolic profile. *Mol. Ther. Nucleic Acids* 19, 361–370. <https://doi.org/10.1016/j.omtn.2019.10.044>.
  61. Torres-Padilla, M.E., Fougère-Deschatrette, C., and Weiss, M.C. (2001). Expression of hnf4alpha isoforms in mouse liver development is regulated by sequential promoter usage and constitutive 3' end splicing. *Mech. Dev.* 109, 183–193. [https://doi.org/10.1016/s0925-4773\(01\)00521-4](https://doi.org/10.1016/s0925-4773(01)00521-4).
  62. Geraud, C., Koch, P.S., Zierow, J., Klapproth, K., Busch, K., Olsavszky, V., Leibing, T., Demory, A., Ulbrich, F., Diett, M., et al. (2017). Gata4-dependent organ-specific endothelial differentiation controls liver development and embryonic hematopoiesis. *J. Clin. Invest.* 127, 1099–1114. <https://doi.org/10.1172/JCI90086>.
  63. Borok, M.J., Papaioannou, V.E., and Sussel, L. (2016). Unique functions of gata4 in mouse liver induction and heart development. *Dev. Biol.* 410, 213–222. <https://doi.org/10.1016/j.ydbio.2015.12.007>.
  64. Heslop, J.A., and Duncan, S.A. (2020). Foxa factors: the chromatin key and doorstop essential for liver development and function. *Genes Dev.* 34, 1003–1004. <https://doi.org/10.1101/gad.340570.120>.
  65. Horisawa, K., Udono, M., Ueno, K., Ohkawa, Y., Nagasaki, M., Sekiya, S., and Suzuki, A. (2020). The dynamics of transcriptional activation by hepatic reprogramming factors. *Mol. Cell* 79, 660–676. <https://doi.org/10.1016/j.molcel.2020.07.012>.
  66. Kasano-Camones, C.I., Takizawa, M., Ohshima, N., Saito, C., Iwasaki, W., Nakagawa, Y., Fujitani, Y., Yoshida, R., Saito, Y., Izumi, T., et al. (2023). Pparalpha activation partially drives nafld development in liver-specific hnf4a-null mice. *J. Biochem.* 173, 393–411. <https://doi.org/10.1093/jb/mvad005>.
  67. Lin, Y., Chen, L., You, X., Li, Z., Li, C., and Chen, Y. (2021). Pparalpha regulates hepatic ketogenesis and fatty acid oxidation during fasting by modulating protein stability of pparalpha. *Mol. Metabol.* 53, 101331. <https://doi.org/10.1016/j.molmet.2021.101331>.
  68. Ahn, J., Woodfint, R.M., Lee, J., Wu, H., Ma, J., Suh, Y., Hwang, S., Cressman, M., and Lee, K. (2019). Comparative identification, nutritional, and physiological regulation of chicken liver-enriched genes. *Poult. Sci.* 98, 3007–3013. <https://doi.org/10.3382/ps/pez057>.
  69. Liu, G.E., Jiang, L., Tian, F., Zhu, B., and Song, J. (2009). Calibration of mutation rates reveals diverse subfamily structure of galliform cr1 repeats. *Genome Biol. Evol.* 1, 119–130. <https://doi.org/10.1093/gbe/evp014>.
  70. Fang, X., Mou, Y., Huang, Z., Li, Y., Han, L., Zhang, Y., Feng, Y., Chen, Y., Jiang, X., Zhao, W., et al. (2012). The sequence and analysis of a chinese pig genome. *GigaScience* 1, 16. <https://doi.org/10.1186/2047-217X-1-16>.
  71. Adelson, D.L., Raison, J.M., and Edgar, R.C. (2009). Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. *Proc. Natl. Acad. Sci. USA* 106, 12855–12860. <https://doi.org/10.1073/pnas.0901282106>.
  72. Zhao, P., Du, H., Jiang, L., Zheng, X., Feng, W., Diao, C., Zhou, L., Liu, G.E., Zhang, H., Chamba, Y., et al. (2020). Pre-1 revealed previous unknown introgression events in eurasian boars during the middle pleistocene. *Genome Biol. Evol.* 12, 1751–1764. <https://doi.org/10.1093/gbe/evaa142>.
  73. Zhou, R., Yao, W., Xie, C., Zhang, L., Pei, Y., Li, H., Feng, Z., Yang, Y., and Li, K. (2020). Developmental stage-specific a-to-i editing pattern in the postnatal pineal gland of pigs (sus scrofa). *J. Anim. Sci. Biotechnol.* 11, 90. <https://doi.org/10.1186/s40104-020-00495-6>.
  74. Chen, C., D'Alessandro, E., Murani, E., Zheng, Y., Giosa, D., Yang, N., Wang, X., Gao, B., Li, K., Wimmers, K., and Song, C. (2021). Sine jumping contributes to large-scale polymorphisms in the pig genomes. *Mobile DNA* 12, 17. <https://doi.org/10.1186/s13100-021-00246-y>.
  75. Medstrand, P., van de Lagemaat, L.N., and Mager, D.L. (2002). Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res.* 12, 1483–1495. <https://doi.org/10.1101/gr.388902>.
  76. Patoori, S., Barnada, S.M., Large, C., Murray, J.I., and Trizzino, M. (2022). Young transposable elements rewired gene regulatory networks in human and chimpanzee hippocampal intermediate progenitors. *Development* 149, dev200413. <https://doi.org/10.1242/dev.200413>.
  77. Frank, J.A., and Feschotte, C. (2017). Co-option of endogenous viral sequences for host cell function. *Curr. Opin. Virol.* 25, 81–89. <https://doi.org/10.1016/j.coviro.2017.07.021>.
  78. Ellison, C.E., and Bachtrög, D. (2015). Non-allelic gene conversion enables rapid evolutionary change at multiple regulatory sites encoded by transposable elements. *Elife* 4, e05899. <https://doi.org/10.7554/eLife.05899>.
  79. Judd, J., Sanderson, H., and Feschotte, C. (2021). Evolution of mouse circadian enhancers from transposable elements. *Genome Biol.* 22, 193. <https://doi.org/10.1186/s13059-021-02409-9>.
  80. Trizzino, M., Kapusta, A., and Brown, C.D. (2018). Transposable elements generate regulatory novelty in a tissue-specific fashion. *BMC Genom.* 19, 468. <https://doi.org/10.1186/s12864-018-4850-3>.
  81. He, J., Babarinde, I.A., Sun, L., Xu, S., Chen, R., Shi, J., Wei, Y., Li, Y., Ma, G., Zhuang, Q., et al. (2021). Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scet. *Nat. Commun.* 12, 1456. <https://doi.org/10.1038/s41467-021-21808-x>.
  82. Lanciano, S., and Cristofari, G. (2020). Measuring and interpreting transposable element expression. *Nat. Rev. Genet.* 21, 721–736. <https://doi.org/10.1038/s41576-020-0251-y>.
  83. Uhlen, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419. <https://doi.org/10.1126/science.1260419>.
  84. Liu, X., Yu, X., Zack, D.J., Zhu, H., and Qian, J. (2008). Tiger: a database for tissue-specific gene expression and regulation. *BMC Bioinf.* 9, 271. <https://doi.org/10.1186/1471-2105-9-271>.
  85. Carithers, L.J., and Moore, H.M. (2015). The genotype-tissue expression (gtex) project. *Biopreserv. Biobanking* 13, 307–308. <https://doi.org/10.1089/bio.2015.29031.hmm>.
  86. Kim, P., Park, A., Han, G., Sun, H., Jia, P., and Zhao, Z. (2018). Tissgdb: tissue-specific gene database in cancer. *Nucleic Acids Res.* 46, D1031–D1038. <https://doi.org/10.1093/nar/gkx850>.
  87. Wang, C., Chen, C., Lei, B., Qin, S., Zhang, Y., Li, K., Zhang, S., and Liu, Y. (2024). Constructing erna-mediated gene regulatory networks to explore the genetic basis of muscle and fat-relevant traits in pigs. *Genet. Sel. Evol.* 56, 28. <https://doi.org/10.1186/s12711-024-00897-4>.
  88. Zhang, S., Wang, C., Qin, S., Chen, C., Bao, Y., Zhang, Y., Xu, L., Liu, Q., Zhao, Y., Li, K., et al. (2024). Analyzing super-enhancer temporal dynamics reveals potential critical enhancers and their gene regulatory networks underlying skeletal muscle development. *Genome Res.* 34, 2190–2202. <https://doi.org/10.1101/gr.278344.123>.
  89. Yu, G., Wang, L.G., and He, Q.Y. (2015). Chipseeker: an r/bioconductor package for chip peak annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383. <https://doi.org/10.1093/bioinformatics/btv145>.
  90. Ito, K., and Murphy, D. (2013). Application of ggplot2 to pharmacometric graphics. *CPT-PHARMACOMET. SYST. PHARMACOL.* 2, e79. <https://doi.org/10.1038/psp.2013.56>.

91. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
92. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of ngs alignment formats. *Bioinformatics* 31, 2032–2034. <https://doi.org/10.1093/bioinformatics/btv098>.
93. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). Deeptools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–W191. <https://doi.org/10.1093/nar/gku365>.
94. Quinlan, A.R., and Hall, I.M. (2010). Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
95. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). Upsetr: an r package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
96. Sachs, M.C. (2017). Plotroc: a tool for plotting roc curves. *J. Stat. Software* 79, 2. <https://doi.org/10.18637/jss.v079.c02>.
97. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. <https://doi.org/10.1093/bioinformatics/btw313>.
98. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). Clusterprofiler: an r package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. <https://doi.org/10.1089/omi.2011.0118>.
99. Kim, D., Langmead, B., and Salzberg, S.L. (2015). Hisat: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. <https://doi.org/10.1038/nmeth.3317>.
100. Liao, Y., Smyth, G.K., and Shi, W. (2014). Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
101. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. <https://doi.org/10.1101/gr.1239303>.
102. Lee, B.T., Barber, G.P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J.N., Hinrichs, A.S., Lee, C.M., et al. (2022). The ucsc genome browser database: 2022 update. *Nucleic Acids Res.* 50, D1115–D1122. <https://doi.org/10.1093/nar/gkab959>.
103. Howe, K.L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., Bhai, J., et al. (2021). Ensembl 2021. *Nucleic Acids Res.* 49, D884–D891. <https://doi.org/10.1093/nar/gkaa942>.
104. Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. <https://doi.org/10.1007/BF01731581>.
105. Kumar, S., and Subramanian, S. (2002). Mutation rates in mammalian genomes. *Proc. Natl. Acad. Sci. USA* 99, 803–808. <https://doi.org/10.1073/pnas.022629899>.
106. Nam, K., Mugal, C., Nabholz, B., Schielzeth, H., Wolf, J.B.W., Backström, N., Künstner, A., Balakrishnan, C.N., Heger, A., Ponting, C.P., et al. (2010). Molecular evolution of genes in avian genomes. *Genome Biol.* 11, R68. <https://doi.org/10.1186/gb-2010-11-6-r68>.
107. Miao, B., Fu, S., Lyu, C., Gontarz, P., Wang, T., and Zhang, B. (2020). Tissue-specific usage of transposable element-derived promoters in mouse development. *Genome Biol.* 21, 255. <https://doi.org/10.1186/s13059-020-02164-3>.
108. de Almeida, B.P., Reiter, F., Pagani, M., and Stark, A. (2022). Deepstarr predicts enhancer activity from dna sequence and enables the de novo design of synthetic enhancers. *Nat. Genet.* 54, 613–624. <https://doi.org/10.1038/s41588-022-01048-5>.
109. Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21, 650–659. <https://doi.org/10.1093/bioinformatics/bti042>.
110. Valdebenito-Maturana, B., Torres, F., Carrasco, M., and Tapia, J.C. (2021). Differential regulation of transposable elements (tes) during the murine submandibular gland development. *Mobile DNA* 12, 23. <https://doi.org/10.1186/s13100-021-00251-1>.
111. Karakulah, G., Arslan, N., Yandim, C., and Suner, A. (2019). Tefectr: an r package for studying the potential effects of transposable elements on gene expression with linear regression model. *PeerJ* 7, e8192. <https://doi.org/10.7717/peerj.8192>.



## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological samples</b>		
The multi-tissue epigenomic and transcriptomic datasets for pigs, cattle, and chickens	Kern et al., 2021 <sup>41</sup>	GEO accession: GSE158430
<b>Deposited data</b>		
TE and gene expression matrices	Zenodo	<a href="https://zenodo.org/records/14638981">https://zenodo.org/records/14638981</a>
<b>Software and algorithms</b>		
RepeatMasker v4.1.2	Tarailo-Graovac et al., 2009 <sup>25</sup>	<a href="https://www.repeatmasker.org/RepeatMasker/">https://www.repeatmasker.org/RepeatMasker/</a>
SQUIRE v0.9.9.9a-beta	Yang et al., 2019 <sup>33</sup>	<a href="https://github.com/wyang17/SQUIRE/releases">https://github.com/wyang17/SQUIRE/releases</a>
R (4.0.5)	R core team	<a href="https://www.R-project.org/">https://www.R-project.org/</a>
R package ChIPseeker (v1.32.0)	Yu et al., 2015 <sup>89</sup>	<a href="https://www.bioconductor.org/packages/release/bioc/html/ChIPseeker.html">https://www.bioconductor.org/packages/release/bioc/html/ChIPseeker.html</a>
R package ggplot2 (v3.5.1)	Ito et al., 2013 <sup>90</sup>	<a href="https://cran.r-project.org/web/packages/ggplot2/index.html">https://cran.r-project.org/web/packages/ggplot2/index.html</a>
Homer	Heinz et al., 2010 <sup>51</sup>	<a href="https://anaconda.org/bioconda/homer">https://anaconda.org/bioconda/homer</a>
Trim Galore (v0.6.10)	N/A	<a href="https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/">https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/</a>
Bowtie2 (v2.5.1)	Langmead et al., 2012 <sup>91</sup>	<a href="https://github.com/BenLangmead/bowtie2/releases">https://github.com/BenLangmead/bowtie2/releases</a>
Sambamba (v1.0.0)	Tarasov et al., 2015 <sup>92</sup>	<a href="https://lomereiter.github.io/sambamba/">https://lomereiter.github.io/sambamba/</a>
deepTools (v3.5.2)	Ramirez et al., 2014 <sup>93</sup>	<a href="https://deeptools.readthedocs.io/en/develop/content/installation.html">https://deeptools.readthedocs.io/en/develop/content/installation.html</a>
BEDTools (v2.31.0)	Quinlan et al., 2010 <sup>94</sup>	<a href="https://github.com/arq5x/bedtools2/releases">https://github.com/arq5x/bedtools2/releases</a>
R package UpSetR (v1.4.0)	Conway et al., 2017 <sup>95</sup>	<a href="https://github.com/hms-dbmi/UpSetR/releases/tag/v1.4.0">https://github.com/hms-dbmi/UpSetR/releases/tag/v1.4.0</a>
Keras (v2.9.0)	keras team	<a href="https://github.com/keras-team/keras">https://github.com/keras-team/keras</a>
TensorFlow (v2.9.1)	N/A	<a href="https://github.com/tensorflow/tensorflow/releases">https://github.com/tensorflow/tensorflow/releases</a>
R package plotROC (v2.3.1)	Sachs et al., 2017 <sup>96</sup>	<a href="https://cran.r-project.org/web/packages/plotROC/index.html">https://cran.r-project.org/web/packages/plotROC/index.html</a>
R package ComplexHeatmap (v2.6.2)	Gu et al., 2016 <sup>97</sup>	<a href="https://github.com/jokergoo/ComplexHeatmap">https://github.com/jokergoo/ComplexHeatmap</a>
R package clusterProfiler (v3.14.3)	Yu et al., 2012 <sup>98</sup>	<a href="https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html">https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html</a>
Hisat2 (v2.2.1)	Kim et al., 2015 <sup>99</sup>	<a href="https://daehwankimlab.github.io/hisat2/">https://daehwankimlab.github.io/hisat2/</a>
featureCounts (v2.0.2)	Liao et al., 2014 <sup>100</sup>	<a href="https://subread.sourceforge.net/featureCounts.html">https://subread.sourceforge.net/featureCounts.html</a>
Cytoscape	Shannon et al., 2003 <sup>101</sup>	<a href="https://cytoscape.org/">https://cytoscape.org/</a>
The original code for data analysis	Zenodo	<a href="https://zenodo.org/records/14638981">https://zenodo.org/records/14638981</a>

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This study utilized publicly available data from pigs, cattle, and chickens released in the Zhou et al. study (GSE158430).<sup>41</sup> The chickens were male F1 crosses from highly inbred Lines 6 and 7, euthanized at 20 weeks of age by exposure to CO<sub>2</sub> gas. Two castrated male Yorkshire littermate pigs were humanely slaughtered at 6 months using electrocution, following standard procedures in slaughterhouses. The cattle were intact male Line 1 Herefords, slaughtered at 14 months using captive bolt under USDA inspection.



All animal procedures were conducted in accordance with Protocol for Animal Care and Use #18464, approved by the Institutional Animal Care and Use Committee (IACUC).

## METHOD DETAILS

### Data sources

In this study, all data were sourced from public databases. To explore the regulatory functions of TEs, we obtained ATAC-seq, Chip-seq, and RNA-seq data for liver, spleen, lungs, muscles, and adipose tissues of pigs, cattle, and chickens, as reported by Zhou et al. (GSE158430).<sup>41</sup> The reference genome sequences (bosTau9.fa, galGal6.fa, susScr11.fa) for these species were downloaded from the UCSC database for TE annotation.<sup>102</sup> To ensure compatibility and coherence between TE annotations and epigenetic data, the reference genome versions employed for TE annotation were aligned with those in Zhou et al.'s published article for the epigenetic data analysis. Additionally, for a conservative analysis of TEs, we obtained the information on genome conserved element regions specific to each species (gerp\_constrained\_elements.bos\_taurus.bb, gerp\_constrained\_elements.gallus\_gallus.bb, gerp\_constrained\_elements.sus\_scrofa.bb) from the Ensembl database.<sup>103</sup>

### TE annotation

TE annotation was performed using RepeatMasker<sup>25</sup> software (version 4.1.2) with the rmblastn engine, incorporating the Dfam\_Consensus-20181026 and Repbase-20181026 libraries. The following parameters were set: -species, -a, -s, -gff, -cutoff 225. In the resulting annotation files, we excluded nested TEs, unknown TEs, and repetitive sequences categorized as Simple\_repeat, Satellite, Low\_complexity, and Unknown. Only the precisely identified TEs were retained for subsequent research analysis. To calculate TE divergence and evaluate the coverage of various TE families within the genome, we employed the calcDivergence-FromAlign.pl and createRepeatLandscape.pl scripts, respectively.

TE insertion age is calculated by adjusting the mutation rate of transposon members from RepeatMasker. First, the mutation rate provided by RepeatMasker (d) is converted to a proportion (D) by dividing by 100. Then, using the Jukes-Cantor formula,<sup>104</sup> the average substitution rate (K) is calculated as follows:

$$X = -\frac{300}{4} \times \ln\left(1 - D \times \frac{4}{300}\right)$$

Finally, the insertion age (T) is obtained by dividing the average substitution rate by twice the mutation rate per year (2r).<sup>104</sup>

$$T = -\frac{K}{2r}$$

It is important to note that our study used different values for the rate of nucleotide substitution per site per year (r) depending on the species. Specifically, we used a consistent single nucleotide mutation rate of  $r = 2.2 \times 10^{-9}$  substitutions/site/year across mammalian genomes<sup>105</sup> and a mutation rate of  $1.91 \times 10^{-9}$  substitutions per site per year in avian genomes.<sup>106</sup>

### TE genomic landscape analysis

In this study, the genomic coverage of different TE families was analyzed using the createRepeatLandscape.pl script in RepeatMasker. TE coverage across the genome was quantified as the proportion of total TE bases relative to the genome size. The differences in TE coverage among the three species were compared using the prop.test function in R (version 4.0.5) to perform a Z-test. The relative abundance of each TE class or family was determined by dividing the count of each TE category by the total annotated TE count. To explore the distribution patterns of different TE classes across the genomes of three species, we downloaded the genome annotation files for each species from the Ensembl database (Sus\_scrofa.Sscrofa11.1.100.chr.gff3, Bos\_taurus.ARS-UCD1.2.104.gff3, Gallus\_gallus.GRCg6a.104.chr.gff3). We then used the annotatePeak function from the R package ChIPseeker (version 1.32.0)<sup>90</sup> to analyze the distribution proportions of various TE classes across different genomic regions. Additionally, to investigate the amplification patterns of different TE classes or families during genome evolution, we utilized R program (version 4.0.5) to analyze the count of each class or family at different time points based on TE insertion age. All visualizations for the analyses described above were performed using the R package ggplot2 (version 3.5.1).<sup>90</sup> In exploring the relationship between the genomic distribution of different TE classes, we followed a similar approach as previous studies.<sup>38</sup> We first divided the genome into non-overlapping 2-Mb windows and then calculated the density of each TE class (as a fraction of the genome sequence) within each 2-Mb window. Subsequently, we employed the corr() method from Python's pandas library (version 2.0.3) to analyze the pairwise correlation between the different TE classes.

### Analyzing the impact of TEs on OCRs

To compare the differences in chromatin accessibility signal enrichment of various classes of TEs across different tissues (Adipose, Liver, Lung, Muscle, Spleen) in three species, we downloaded the raw ATAC-seq data for these tissues and generated BigWig files for profile plotting of the accessibility signals of different TEs. The analysis proceeded as follows: first, the raw ATAC-seq data underwent quality control using Trim Galore (version 0.6.10; [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) with the

following parameters: `-q 20 -phred33 -stringency 4 -length 20 -e 0.1`. Next, the clean reads were aligned to the reference genome using Bowtie2 (version 2.5.1)<sup>92</sup> software. PCR duplicates were then removed using Sambamba (version 1.0.0)<sup>93</sup> software, and the resulting BAM files were converted to BigWig files using the `bamCoverage` script from deepTools<sup>93</sup> (version 3.5.2) with the following parameters: `-extendReads -normalizeUsing RPKM`. Finally, chromatin accessibility signals for different TE regions and their flanking sequences ( $\pm 2$  kb) were generated using the `computeMatrix` script in deepTools (version 3.5.2), and the signal intensity was visualized using the `plotProfile` script.

Additionally, to explore the impact of TEs on OCRs of these tissues across the three species, we obtained peak files from both ATAC-seq and DNase-seq for each tissue, ensuring the inclusion of two biological replicates per tissue. We then utilized the `intersect` and `merge` functions of BEDTools<sup>94</sup> (version 2.31.0) to identify consistent OCRs that are representative of each tissue.

To examine the direct influence of TEs on OCRs, we used the `intersect` command in BEDTools (version 2.31.0) to calculate the proportion of OCRs affected by TE insertions, designating these as TE-driven OCRs. TEs were then categorized into age groups based on their estimated genomic insertion ages: youngest TEs (0–25 million years (Mya)), younger TEs (25–50 Mya), older TEs (50–75 Mya), and oldest TEs (75–100 Mya). We assessed the enrichment of various TE classes and age groups within TE-driven OCRs across tissues in each species, using the proportional distribution of all genome-annotated TEs as a background reference. Enrichment was statistically tested with the `prop.test` function in R program (version 4.0.5) to perform Z-tests. Finally, the proportional distribution of TE classes and age groups within TE-driven OCRs for each tissue was visualized using R package ggplot2 (version 3.5.1).

To identify TEs with chromatin accessibility in specific tissues, we utilized the `intersectBed` function of BEDTools (version 2.31.0) with parameters of `-f 0.5` and `-F 0.2` to determine overlaps between TEs and OCR peaks in each tissue.<sup>107</sup> We considered a TE to be associated with chromatin accessibility if it overlapped with OCR peaks by at least 50%, and OCR peaks overlapped with TEs by at least 20%. Shared and tissue-specific TE accessibility was visualized using the R package UpSetR (version 1.4.0),<sup>95</sup> based on the TEs identified with chromatin accessibility in each tissue. Tissue-specific accessible TEs were identified using the `setdiff` function in R program (version 4.0.5). To identify enriched TF motifs within tissue-specific accessible TEs, we performed motif enrichment analysis using the `findMotifsGenome.pl` script from HOMER software. This analysis was carried out on tissue-specific TEs for each tissue, employing the parameters `-len 8,10,12` and `-size -100,100`. We selected the top 20 significantly enriched TFs expressed in the corresponding tissues for each species. The results were visualized using bubble plots generated with the R package ggplot2 (version 3.5.1).

### Identification and deep learning classification of different DNA CREs

To investigate the effects of different TE classes and families on various types of DNA CREs, including promoters, enhancers, and potential silencers, we further obtained peak files for CTCF, H3K4me1, H3K27ac, and H3K4me3 from adipose, liver, lung, muscle, and spleen tissues of pigs, cattle, and chickens. These files facilitate the identification of regulatory elements across these tissues. To obtain representative peak regions for each tissue, we merged the peak regions from the two biological replicates using the `intersect` and `merge` commands from BEDTools (version 2.31.0). The methods for identifying different types of DNA CREs are outlined as follows: Enhancers were defined as H3K27ac peaks located outside known coding gene transcription start sites (TSS)  $\pm 1$  kb and without overlap with H3K4me3 peaks. Promoter regions were characterized as H3K27ac peaks overlapping with known coding gene TSS  $\pm 1$  kb or with H3K4me3 peaks. Putative silencer regions were identified based on a previously described method.<sup>53</sup> In this approach, OCRs overlapping with CTCF, H3K4me1, H3K27ac, H3K4me3, and TSS were excluded, and the remaining OCRs were classified as CSEs.

To investigate whether the CSEs identified by this method possess unique sequence characteristics distinct from enhancers and promoters, we constructed a multitask deep learning classification model based on Almeida's Deepstarr regression model.<sup>108</sup> This model was used to classify and predict promoters, enhancers, and CSEs across different species. First, for the model input sequences, we extracted DNA sequence information within  $\pm 1$  kb of the center point of each regulatory element and converted it into a one-hot encoding format to serve as input data for the deep learning model. The output labels for the sequences were also converted into a one-hot encoding format based on the different types of regulatory elements. The architecture of model is based on the Basset CNN, consisting of four 1D convolutional layers with the following filters and sizes: 246 (size = 7), 60 (size = 3), 60 (size = 5), and 120 (size = 3). Each convolutional layer is followed by batch normalization, a ReLU activation function, and max-pooling with a pool size of 2. Following the convolutional layers, there are two fully connected layers, each with 256 neurons, followed by batch normalization, a ReLU activation, and dropout with a fraction of 0.4. The final layer maps to both element outputs. Hyperparameters were manually adjusted to achieve the best performance on the validation set. The model was implemented and trained using Keras (v.2.9.0) with TensorFlow (v.2.9.1) as the backend, employing the Adam optimizer (learning rate = 0.0001) and binary\_crossentropy as the loss function. Training was conducted with a batch size of 128, and early stopping was employed with a patience of 20 epochs. Model training, hyperparameter tuning, and performance evaluation were conducted on different sets. Data preprocessing involved shuffling and splitting, with 15% reserved for testing and the remaining data split into 80% for training and 20% for validation (Tables S7A–S7C). To evaluate the performance of each regulatory element type classification accuracy, we used the R package plotROC (version 2.3.1)<sup>97</sup> to calculate and visualize the AUC results based on the predicted and true values of the test set.

### Contribution of TEs to different DNA CREs

To examine the contribution of TEs to different types of DNA CREs, we calculated the proportion of TE bases across various element types. We utilized the "intersect" command in BEDTools (version 2.31.0) to extract TE sequences from specific DNA CREs, and then

quantified the proportion of TE bases relative to the total bases within each regulatory element using R program (version 4.0.5), with the genomic coverage of all annotated TEs in the species serving as a reference.

To investigate the contribution of different classes of TEs to the composition of CREs, we quantified the proportion of TE bases for each TE classes within the CREs. In this part of the analysis, the proportion of TE base for each type of TE in all annotated TEs was used as the background.

In addition, to investigate the contribution of TE insertions to different types of CREs, we utilized the ‘intersect’ command in BEDTools (version 2.31.0) to calculate the proportion of annotated TEs overlapping with specific regulatory element across tissues in three species. To further focus on the ability of different classes of TEs to integrate into different CREs, we explored the proportion of different classes of TEs based on those TEs overlapping with specific regulatory element. The above statistical results were visualized using the ggplot2 R package (version 3.5.1) in R language.

### Enrichment analysis of TE families in different DNA CREs

To examine the enrichment of TE families in different CREs across different tissues of the three species, we utilized a permutation test. This involved randomly simulating size-matched regulatory element regions in the genome 1,000 times to establish a background for calculating the fold enrichment and significance of various TE families within each regulatory element type. *p*-values for significance were corrected using the Bonferroni method. Before conducting the enrichment analysis, we filtered out low-abundance TE families since they may have low statistical power due to their low copy number in certain species. We retained TE families with more than 3000 copies in pigs and cattle and more than 1000 copies in chickens for analysis. The enrichment analysis results were visualized using the R package pheatmap (version 1.0.12).

### Quantification of TE loci expression and identification of tissue-specific TE expression

To detect the expression activity of TEs in the liver, spleen, lung, muscle, and adipose tissue of three species, we obtained transcriptome data for these tissues, including two biological replicates for each. First, we conducted quality control on the transcriptome data of each sample using Trim Galore (version 0.6.10) software with the following parameters: -q 20 -phred33 -stringency 4 -length 20 -e 0.1. Subsequently, based on the annotation information of TEs in the species genomes and RNA-seq clean data, we used SQUIRE software (version 0.9.9.9a-beta) to detect the transcriptional activity of TEs in each sample. To minimize the impact of low-expression TEs on downstream analysis, we defined TEs with an average FPKM  $\geq 1$  in the two biological replicates of a tissue as detectable TEs in that tissue. By integrating the expression levels of detectable TEs in the five tissues and using the R package ComplexHeatmap (version 2.6.2)<sup>97</sup> for plotting K-means heatmaps. To identify tissue-specifically expressed TEs, based on the matrix formed by the expression values of detectable TEs in each sample, after data normalization, we calculated the tissue-specificity index (TSI) of each TE using the following formula<sup>109</sup>:

$$\tau = \frac{\sum_{i=1}^n (1 - x_i)}{N - 1}$$

Tissue-specifically expressed TEs were identified as those with a delta value greater than 0.9. TEs with a normalized expression value of 1 within a tissue were considered specifically expressed in that tissue.

### Enrichment analysis of tissue-specific expressed TEs

To explore whether specific TE classes are enriched or depleted among the expressed TEs in each tissue, we compared the proportions of different TE classes in the detectable TEs to the proportions of all annotated TEs. This enabled us to identify the proportions of different TE classes within the expressed TEs and to observe any enrichment or depletion of certain TE classes in the expressed TEs of each tissue compared to the background.

To investigate the enrichment intensity of tissue-specifically expressed TEs compared to non-tissue-specifically expressed TEs in genomic coding regions and conserved elements, we obtained genomic conserved element for three species from Ensembl database. We defined coding regions based on genome annotation files, specifically considering the gene body regions of protein-coding genes as coding regions. Non-tissue-specifically expressed TEs were used as controls, and we counted the numbers of both tissue-specific and non-specific TEs within the coding regions and conserved elements. We constructed contingency tables and conducted Fisher’s statistical test using the fisher.test function in R. The results were then visualized using the R package ggplot2 (version 3.5.1).

To examine the biological processes associated with tissue-specific TE expression, we utilized the R package ChIPseeker (version 1.32.0) to annotate the genomic information of tissue-specific TEs and extract their neighboring genes, regarded as potential target genes. We then performed GO enrichment analysis on these target genes using the R package clusterProfiler (version 3.14.3).<sup>98</sup> The significantly enriched GO terms reflect the biological processes in which the tissue-specific expressed TEs are involved. Finally, we selected the top 10 significantly enriched GO terms in each tissue for result visualization using R package ggplot2 (version 3.5.1).

### Construction of TE-GRNs

Tissue-specific TEs are intricately linked to the biological processes of their respective tissues. To investigate the gene regulatory networks they mediate, we first identified these hub TEs by examining those that overlap with tissue-specific OCRs and exhibit tissue-specific expression patterns. Following this, we further characterized the upstream TFs and downstream target genes regulated

by these hub TEs. For identifying the downstream target genes of hub TEs, we adopted a strategy inspired by previous studies<sup>55,110,111</sup> that suggest the transcriptional activity of TEs can affect the expression of neighboring genes and exhibit co-expression patterns with them. To perform co-expression analysis between TEs and genes, we quantified the expression of genes in each sample as follows: We first aligned the transcriptome data to the genome with Hisat2 software (version 2.2.1)<sup>100</sup> with the `-rna-strandness RF` option. Subsequently, we quantified the gene reads using featureCounts software (version 2.0.2)<sup>101</sup> with the `-s 2` parameter. Finally, the expression level of each gene were normalized to TPM values. Next, we performed genomic annotation of hub TEs using the R package ChIPseeker (version 1.32.0) to identify their neighboring genes. Subsequently, we assessed the linear regression relationship between TE and neighboring gene expression levels in different tissues using the `lm` function in R program (version 4.0.5). Statistically significant TE-gene pairs (model  $p$  value  $< 0.05$ ) were retained, and the Spearman correlation between TE-gene pairs was calculated using the R base function `cor.test` to determine positive regulatory relationships ( $R_s \geq 0.3$ ). To screen for the upstream TFs of hub TEs, we employed a multi-criteria screening strategy. Firstly, we used the BEDtools `getfasta` command to accurately extract TE DNA sequences from the genome based on their strand information. Next, we downloaded the custom motif matrices files for TFs from <http://homer.ucsd.edu/homer/custom.motifs> and utilized the `scanMotifGenomeWide.pl` script in HOMER software to scan for motifs information within TE sequences, enabling us to identify potential upstream TFs associated with these TEs. Subsequently, we further used the `cor.test` function in R to calculate the correlation between TF-TE pairs (method = "spearman", adjust = "fdr"). TF-TE pairs with  $R_s \geq 0.3$  and  $\text{adjust}, p < 0.05$  were considered to have a regulatory relationship. Additionally, by identifying the co-localization of hub TEs and CREs, we were able to infer the CREs mediated by hub TEs. Overall, we determined the regulatory relationships between TF binding to TEs, resulting in chromatin opening and the regulation of target genes. Finally, we visualized these regulatory relationships in liver tissue using the Cytoscape<sup>101</sup> software.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Details of the statistical tests applied, including the statistical methods and significance levels, are provided in the main text or relevant figure legends. All statistical analyses were conducted using the R software.

## ADDITIONAL RESOURCES

This study does not include additional resources.