Jiang et al. *Zool. Res.* 2024, 45(1): 138−151
https://doi.org/10.24272/j.issn.2095-8137.2023.060

ZR | Zoological Research

Article Open Access

Pig H3K4me3, H3K27ac, and gene expression profiles reveal reproductive tissue-specific activity of transposable elements

Tao Jiang[1,#], Zhi-Min Zhou[1,#], Zi-Qi Ling[1], Qing Zhang[1], Zhong-Zi Wu[1], Jia-Wen Yang[1], Si-Yu Yang[1], Bin Yang[1,*], Lu-Sheng Huang[1,*]

[1] *National Key Laboratory of Pig Genetic Improvement and Germplasm Innovation*, *Jiangxi Agricultural University*, *Nanchang*, *Jiangxi* 330045, *China*

ABSTRACT

Regulatory sequences and transposable elements (TEs) account for a large proportion of the genomic sequences of species; however, their roles in gene transcription, especially tissue-specific expression, remain largely unknown. Pigs serve as an excellent animal model for studying genomic sequence biology due to the extensive diversity among their wild and domesticated populations. Here, we conducted an integrated analysis using H3K27ac ChIP-seq, H3K4me3 ChIP-seq, and RNA-seq data from 10 different tissues of seven fetuses and eight closely related adult pigs. We aimed to annotate the regulatory elements and TEs to elucidate their associations with histone modifications and mRNA expression across different tissues and developmental stages. Based on correlation analysis between mRNA expression and H3K27ac and H3K4me3 peak activity, results indicated that H3K27ac exhibited stronger associations with gene expression than H3K4me3. Furthermore, 1.45% of TEs overlapped with either the H3K27ac or H3K4me3 peaks, with the majority displaying tissue-specific activity. Notably, a TE subfamily (LTR4C_SS), containing binding motifs for SIX1 and SIX4, showed specific enrichment in the H3K27ac peaks of the adult and fetal ovaries. RNA-seq analysis also revealed widespread expression of TEs in the exons or promoters of genes, including 4 688 TE-containing transcripts with distinct development stage-specific and tissue-specific expression. Of note, 1 967 TE-containing transcripts were enriched in the testes. We identified a long terminal repeat (LTR), MLT1F1, acting as a testis-specific alternative promoter in *SRPK2* (a cell cycle-related protein kinase) in our pig dataset. This element was also conserved in humans and mice, suggesting either an ancient integration of TEs in genes specifically expressed in the testes or parallel evolutionary patterns. Collectively, our findings demonstrate that TEs are deeply embedded in the genome and exhibit important tissue-specific biological functions, particularly in the reproductive organs.

Keywords: Transposable elements; Porcine; Histone modification; Alternative promoter; TE-containing transcript

INTRODUCTION

Transposable elements (TEs) are a class of mobile DNA sequences that constitute more than half of the DNA in many eukaryotes (Fedoroff, 2012). Based on their transposon mechanism, TEs can be divided into DNA transposons and retrotransposons, which move throughout the genome by direct cut-and-paste DNA sequence and RNA intermediates, respectively. Mammalian retrotransposons are broadly categorized into long terminal repeats (LTRs), long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs), which are repressed by different epigenetic mechanisms to prevent harmful insertion in the genome (Slotkin & Martienssen, 2007). Various studies have demonstrated that TEs provide binding sites for transcription factors (TFs) and represent an important source of regulatory elements, including promoters and enhancers (Chuong et al., 2017; Fueyo et al., 2022; Rebollo et al., 2012; Sundaram et al., 2017; Sundaram & Wysocka, 2020; Ting et al., 1992). These regulatory elements tend to be restricted to specific species or lineages (Chuong et al., 2016; Jacques et al., 2013; Kunarso et al., 2010). TEs are also expressed as a part of genes, thereby affecting gene expression and alternative splicing (Kapusta et al., 2013; Lee et al., 2022; Modzelewski et al., 2021; Senft & Macfarlan, 2021). Despite considerable progress, the tissue-specific activity and function of TEs remain largely unknown.

The pig (*Sus scrofa*) is an important agricultural animal and a valuable preclinical model for biomedical research (Lunney

Received: 08 August 2023; Accepted: 04 September 2023; Online: 05 September 2023

Foundation items: This work was supported by the National Natural Science Foundation of China (32160781)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: binyang@live.cn; LushengHuang@hotmail.com

et al., 2021; Wu et al., 2020; Yang et al., 2018; Zhong et al., 2023). The annotation of regulatory elements of the pig genome through profiling of histone modifications provides a basis for analyzing complex biological mechanisms (Zhu et al., 2022), accelerates the progress of medical research (Pan et al., 2021), and promotes the analysis of genetic mechanisms of complex porcine traits (Zhang et al., 2022b; Zhao et al., 2021). Construction of a regulatory element database will also provide a valuable resource for the development of gene editing tools, such as identifying potential safe harbors that enable stable expression of exogenous genes (Zhu et al., 2022). The Cre-loxp-mediated recombination system is a useful tool for tissue-specific gene editing that relies on tissue-specific promoters (Utomo et al., 1999). H3K27ac, which marks active enhancers and active promoters (Creyghton et al., 2010; Duttke et al., 2015), and H3K4me3, which marks active promoters (Barski et al., 2007; Bernstein et al., 2005; Pekowska et al., 2011), are two of the most relevant histone modifications associated with gene transcription. To date, however, few studies have compared the heterogeneity in H3K27ac and H3K4me3 activity or their associations with gene expression across different tissues and developmental stages, with even fewer focusing on the annotation of TEs in relation to their association with tissue-specific histone modifications and mRNA expression in pigs.

In this study, we integrated H3K4me3 ChIP-seq, H3K27ac ChIP-seq, and RNA-seq from 10 tissues of both fetal and adult pigs to annotate the regulatory sequences and TEs in the pig genome. Our results showed that H3K27ac peaks displayed a stronger association with gene expression than H3K4me3. Annotation of TEs with H3K27ac ChIP-seq, H3K4me3 ChIP-seq, and RNA-seq data across different fetal and adult tissues revealed several TE-containing transcripts and TE-derived regulatory elements associated with tissue-specific histone modifications, gene expression, alternative splicing, and alternative promoter usage, particularly in the ovaries and testes. Thus, this study provides a valuable framework for utilizing spatiotemporal H3K4me3 and H3K27ac ChIP-seq and RNA-seq data to explore the biological functions of TEs across different tissues and developmental stages.

MATERIALS AND METHODS

Ethics statement

All procedures involving animals followed the guidelines for the care and use of experimental animals approved by the State Council of the People's Republic of China. This study was approved by the Committee on Animal Biosafety of Jiangxi Agricultural University.

Sample collection

As previously described (Zhu et al., 2022), eight tissues (brain, heart, liver, lungs, muscle, small intestine (SI), ovaries, and testes) from seven fetuses (including two male and two female full-sib Large White pigs on day 75 and two male and one female Bamaxiang pigs on day 74 post-insemination) and 10 tissues (brain, heart, liver, lungs, muscle, pituitary, SI, stomach, ovaries, and testes) from eight adult pigs (two male and two female Large White pigs on day 150 and two male and two female Bamaxiang pigs on days 132–141) were collected in this study. The anatomical sites of sampled tissues were consistent in fetuses and adults. The collected tissues were frozen in liquid nitrogen immediately after slaughter and stored at −80 °C.

ChIP preparation and sequencing

ChIP was performed using the SimpleChIP® Plus Enzymatic Chromatin IP Kit (Magnetic Beads) (Cell Signaling Technology, USA). Briefly, 0.2–0.3 g of each sample was homogenized with 1 mL of phosphate-buffered saline (PBS) and cross-linked with 37% formaldehyde at room temperature for 10 min. Then, 10×glycine was added and mixed thoroughly to end the cross-linking reaction. Buffers A and B were then added for lysing. The DNA was sheared into 100–300 bp segments after sonication and enzyme digestion. ChIP was incubated with 5 μg of H3K27ac antibody (Active Motif, USA) or 3 μg of H3K4me3 antibody (Active Motif, USA) overnight to accumulate H3K27ac-marked regions or H3K4me3-marked regions, respectively. Antibody-bound DNA was collected using ChIP-grade Protein G magnetic beads, then incubated with 2 μL of 20 mg/mL proteinase K and 6 μL of 5 mol/L NaCl at 65 °C to reverse the cross-links. Immunoprecipitated DNA was then purified using a spin column. The steps of the method are binding, washing and elution. Binding of nucleic acid to a silica membrane, washing away particulates and inhibitors that are not bound to the silica membrane, and elution of the nucleic acid, with the end result being purified nucleic acid in an aqueous solution. Single-end 50 bp DNA sequencing was performed using the Illumina HiSeq 2500 platform (USA).

ChIP-seq processing

Clean reads were aligned to the pig genome (*Sus scrofa* 11.1) using Burrows-Wheeler Aligner (BWA) (Abuín et al., 2015), allowing two mismatches. Duplicate alignments were removed using SAMtools markdup (Li et al., 2009). MACS2 v.2.1.1 (Zhang et al., 2008) was used to infer peaks for each sample with the following parameter: " --q 0.01 -B --keep-dup all". Broad domains were identified using the parameter "--broad and --broad-cutoff 0.1". Peaks were merged across all samples using the "merge" function in BEDTools (Quinlan, 2014) and the number of reads in each merged peak was calculated using SAMtools bedcov (Li et al., 2009). Finally, the normalized value (peaks per million) for each merged peak was defined in R program v.3.5.1 (https://www.r-project.org/).

RNA-seq library preparation, sequencing, and processing

Total RNA was extracted from frozen samples using TRIzol reagent (Invitrogen, USA), followed by mRNA collection using magnetic beads attached to poly-T oligos. The collected mRNA was then reverse-transcribed into cDNA. AMPure XP Beads were used for fragment enrichment and polymerase chain reaction (PCR) amplification to obtain strand-specific cDNA libraries. These libraries were then subjected to paired-end 150 (PE150) sequencing on the Illumina HiSeq 4000 platform (USA).

The STAR program (Dobin et al., 2013) was used to map the clean reads to the *Sus scrofa* 11.1 genome using Ensembl GTF (98.111) (Supplementary Table S1). The transcripts were assembled using StringTie (Pertea et al., 2015) and quantified with FeatureCounts (Liao et al., 2014). Genes with transcripts per million (TPM) values less than 1 in 90% of the samples were excluded. In total, 17 095 genes were retained for further analysis.

Predicting expression using ridge regression and peak-gene correlations

A total of 94 samples with two histone modifications

(H3K4me3 and H3K27ac) and corresponding transcriptome data were collected, including data from 63 samples from two previous studies (Supplementary Table S2) and data from 31 samples generated in this study. From Kern et al. (2021), a total of 11 samples were collected from seven tissues (adipose, cerebellum, cerebral cortex, hypothalamus, liver, lungs, and spleen) of a 6-month-old male Yorkshire breed. From Zhao et al. (2021), a total of 52 samples were collected from five tissues (muscle, liver, fat, spleen, and heart) of 6-month-old male MeiShan, Large White, Enshi Black, and Duroc breeds. A uniform ChIP-seq data analysis pipeline was applied to the 63 downloaded datasets. We retained 18 324 genes with TPM>1 in at least 10% of the samples for further analysis. The correlations between genes and peaks were assessed using Spearman correlation analysis. Genes with transcription start sites (TSSs) located within 500 kb of the peak centers were considered. For each peak-gene pair, *P*-values were corrected using the qvalue package, with peak-gene pairs with *Q*-values less than 0.01 retained.

We evaluated three distinct models, each with different predictors, including solely H3K4me3 peaks, solely H3K27ac peaks, or both peaks, to determine their capacity to predict gene expression in three peak-gene distance scenarios, i.e., association of peaks within 50 kb, 200 kb, and 500 kb of corresponding genes. For each prediction model, the whole dataset was divided into training and test data. Ridge regression was used to train a prediction model using the training dataset, after which the model was evaluated using the test data. Pearson correlation coefficients (r^2) between the predicted and measured values of gene expression were calculated as a measure of prediction accuracy.

Identification of tissue-specific promoters

Tissue specificity was defined with a Tau score threshold of 0.95 (Kryuchkova-Mostacci & Robinson-Rechavi, 2017) on the H3K27ac and H3K4me3 ChIP-seq data and RNA-seq data. Tissue-specific peaks within 1 kb upstream of the TSSs of tissue-specific genes and supported by both histone markers were defined as candidate tissue-specific promoters.

Identification of accessible TEs in multiple tissues at different developmental stages

Based on tissues and developmental stages, 84 samples of H3K27ac peaks were divided into 17 stage-tissue groups. TEs in the reference genome *Sus scrofa* 11.1 were identified using RepeatMasker (Tarailo-Graovac & Chen, 2009), as described previously (Chen et al., 2022). TEs with 50% of their sequence overlapping with at least one peak were considered peak overlapping TEs, using intersectBed, as described previously (Judd et al., 2021).

Enrichment analysis of TEs

Analysis of TE enrichment within specific stage-tissue group peaks was conducted using a Perl script (https://github.com/4ureliek/TEanalysis), as described previously (Kapusta et al., 2013), which randomized the location of TEs and maintained the distance to the nearest TSS and number of TEs on each chromosome. This procedure was bootstrapped 1 000 times. TE enrichment in a peak set was deemed significant if the ratio of observed to expected values exceeded 2.

Quantification of transcripts and TE-gene junctions

We remapped 112 RNA-seq datasets to the swine genome assembly (*Sus scrofa* 11.1) using STAR aligner v.2.7.2b (Dobin et al., 2013) with the following parameters: " --outFilterMultimapNmax 500 --alignIntronMax 1000000 --outFilterType BySJout --alignMatesGapMax 1000000". A two-pass alignment strategy was used to increase the detection sensitivity of the spliced RNA-seq reads. For first-pass alignment, the STAR index was installed with the gene annotations provided by RefSeq. The second STAR index was updated using the results from the first alignments. Quantification of TEs and transcripts was performed using TEcount (Jin et al., 2015), as described previously (Modzelewski et al., 2021). Finally, counts of exon model per million mapped reads (CPM) was adopted to normalize raw counts. Stage- and tissue-specific transcript expression analysis was performed using the Tau score, as described previously (Zhu et al., 2022).

For TE-gene junctions, all junctions were assembled and annotated using a previously described pipeline (Modzelewski et al., 2021). In brief, StringTie2 (Pertea et al., 2015) was used for transcript assembly with the following parameters: "-j 2 -s 5 -f 0.05 -c 2". The output was then merged using TACO (Niknafs et al., 2017). Splicing junctions detected in the RNA-seq dataset were then collected. The starting position of the junction was annotated using information on TEs collected from RepeatMasker (Tarailo-Graovac & Chen, 2009). Junctions with at least 10 reads were retained and normalized with transcripts per kilobase of exon model per million mapped reads. Stage- and tissue-specific TE-gene junctions were determined using the Tau score.

Motif analysis

Enrichment of TF-binding motifs in peaks was analyzed using HOMER (Heinz et al., 2010) with default parameters, and the results were verified using MEME (Bailey et al., 2015). FIMO (Grant et al., 2011) was used to predict the locations of potential TF motifs with the database of known vertebrate transcription factor motifs. TF motif heatmap visualization of binding regions was conducting using deepTools (Ramírez et al., 2014).

RESULTS

Samples and data in this study

We performed H3K4me3 and H3K27ac ChIP-seq and RNA-seq on eight tissues from four Large White and three Bamaxiang pig fetuses (prenatal days 74–75) and 10 tissues from four Large White and four Bamaxiang adult pigs (postnatal days 132–150) of both sexes (Figure 1A). In total, we successfully obtained 84 H3K27ac ChIP-seq, 42 H3K4me3 ChIP-seq, and 112 RNA-seq datasets, which passed quality control procedures for follow-up study, including 31 matched datasets for H3K4me3, H3K27ac, and RNA-seq from the same samples (Figure 1B; Supplementary Tables S1, S3, S4). Among these data, 70 H3K27ac ChIP-seq and 62 RNA-seq datasets have been reported previously (Zhu et al., 2022), while the remaining data were obtained in this study for the first time. After removing low-quality reads and duplicates, we generated an average of 23.82 million effective reads per ChIP-seq assay (Supplementary Tables S3, S4). Hierarchical clustering of samples based on the signals of the two histone modifications and gene expression showed that the samples were first grouped by sequencing assay and then by tissue type (Supplementary Figure S1A), similar to the results of Pan et al. (2021). Furthermore, the means of the fraction of reads in peaks (FRiP), normalized strand cross-correlation

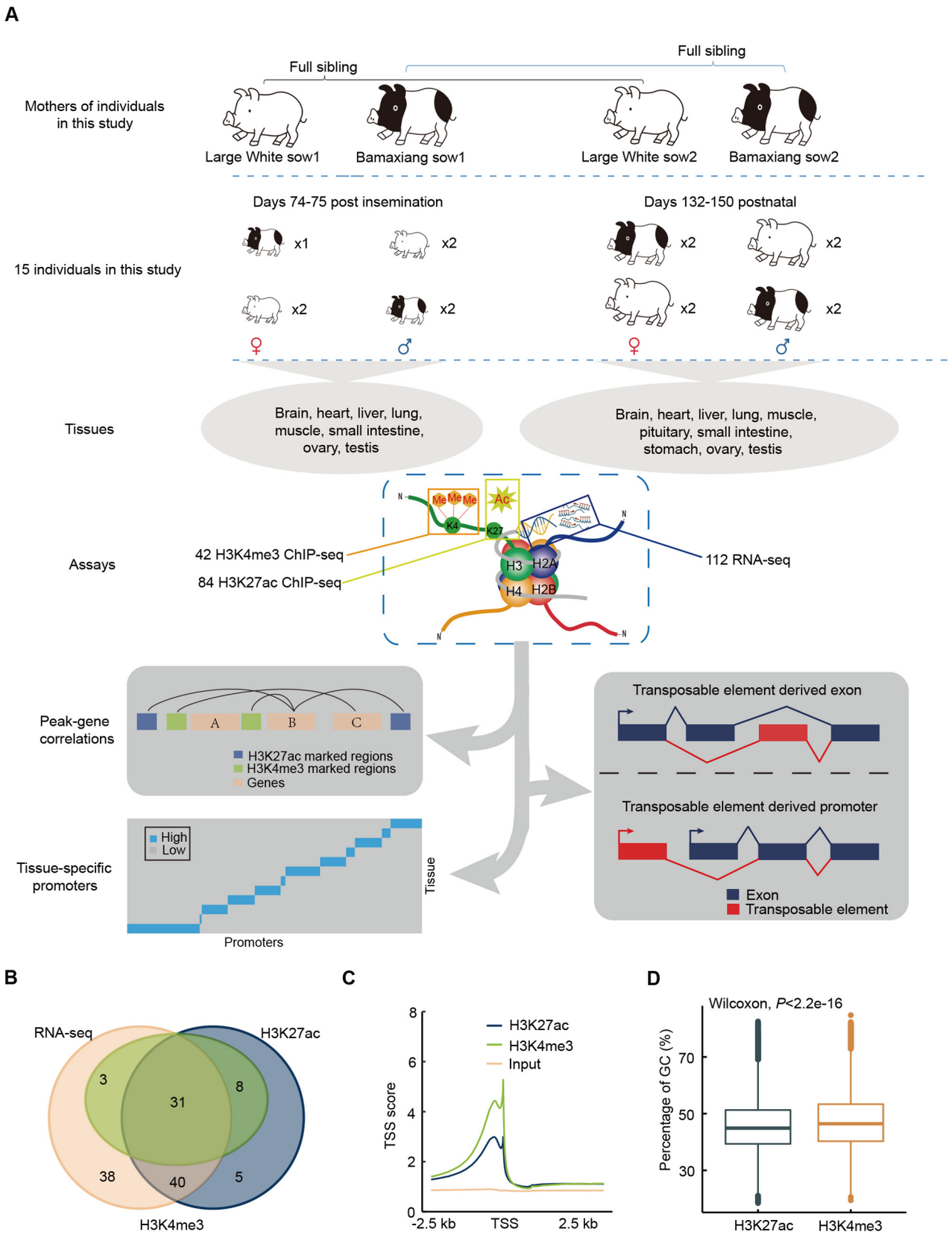

Figure 1 Scheme of samples and experiments in this study

A: Experimental design of study. B: Venn diagram showing overlap of samples by the three assays. Different colors indicate different assays, pink for RNA-seq, blue for H3K4me3 ChIP-seq, and green for H3K27ac ChIP-seq. C: Average read coverages of two histone modifications at 2 500 bp upstream and 2 500 bp downstream regions of genes. D: Box plot showing GC contents in H3K27ac and H3K4me3 peaks. Statistical significance of comparisons between H3K27ac and H3K4me3 peaks was calculated using wilcox.test in the R program.

coefficient (NSC), and relative strand cross-correlation coefficient (RSC) were 0.20, 1.06, and 2.98, respectively (Supplementary Tables S3, S4), meeting all data standards from the ENCODE guidelines and thus supporting the quality of the ChIP-seq data. We identified an average of 24 957 (range: 11 368 to 59 338) H3K4me3 and 38 525 (range: 10 506 to 81 255) H3K27ac peaks per sample (Supplementary Figure S1B, C). After combining peaks across different samples, we obtained 115 181 and 271 376 merged H3K4me3 and H3K27ac peaks, respectively, further enriching the regulatory element catalog for pigs obtained in previous study (Zhu et al., 2022). Of the 115 181 merged H3K4me3 peaks, 98 041 (85.12%) overlapped with at least one of the merged H3K27ac peaks, indicating that H3K27ac tagged most of the active promoter regions exhibiting H3K4me3 activity (Howe et al., 2017; Roth et al., 2001). H3K4me3 peaks exhibited a higher degree of enrichment around TSSs and higher GC content compared to the H3K27ac peaks (Figure 1C, D), indicating that H3K4me3 is more likely to occur at TSSs and be subjected to DNA methylation than H3K27ac (Hughes et al., 2020).

Associations of H3K27ac and H3K4me3 with gene expression

We next compared the H3K4me3 and H3K27ac peaks in terms of their associations with gene expression. To enhance the power of peak-gene association analysis, we downloaded 63 publicly available H3K27ac and H3K4me3 ChIP-seq and RNA-seq datasets (Supplementary Table S2). The data were analyzed using the same analytical pipelines, assembling a dataset of 94 samples with matched H3K27ac and H3K4me3 ChIP-seq and RNA-seq data. In total, 333 867 H3K27ac and 126 651 H3K4me3 merged peaks were identified in the 94 samples. For each gene, we compared the predictive performance of three models (H3K27ac model, H3K4me3 model, and two-modifications model) for histone peaks located within 50 kb, 200 kb, and 500 kb of the TSSs of the corresponding genes. Notably, the H3K27ac model showed higher prediction accuracy of gene expression than the H3K4me3 model in all peak-gene distance scenarios, suggesting a stronger gene regulatory role of H3K27ac than H3K4me3 (Figure 2A). The predictive power of the H3K27ac model was comparable to that of the two-modification model, which may reflect functional redundancy of H3K27ac and H3K4me3 in their regulatory role in gene expression (Karlić et al., 2010).

The integrated data also enabled the identification of potential target genes of the two histone modification markers. We calculated the correlations between the two histone markers and genes within 500 kb and identified 77 989 H3K4me3-gene and 438 615 H3K27ac-gene correlations (*Q*-values<0.01), providing a valuable resource for exploring the links between regulatory elements and genes across pig tissues. For both histone markers, positive peak-gene correlations were enriched around the TSSs of the corresponding genes, whereas negative correlations were under-represented in the TSS regions (Figure 2B; Supplementary Figure S2A). Results showed that 16 939 out of 29 408 H3K4me3 peaks showed stronger associations with distant rather than adjacent genes, e.g., the peak (chr4:90 642 880–90 645 761) showed the highest correlation with *PEA15*, which was located 369 kb from the peak (Supplementary Figure S2B, C), suggesting that some promoters may play enhancer roles, in agreement with observations from Hi-C data (Delaneau et al., 2019).

Identification of tissue-specific promoters

Identification of tissue-specific promoters will facilitate tissue-specific genome editing of genes using the Cre/loxP recombinase system (Utomo et al., 1999). Therefore, we identified H3K27ac and H3K4me3 peaks linked to tissue-specific gene expression. After applying strict filtration procedures (see Methods for more details), we identified 230 tissue-specific promoters simultaneously supported by the tissue-specific activity of H3K27ac, H3K4me3, and mRNA expression (Figure 2C, D; Supplementary Tables S5, S6). Among these genes, we identified the heart-specific gene *TBX20*, a gene required for normal heart function (Kirk et al., 2007; Qian et al., 2008), and kidney-specific gene *PAX2*, which is associated with renal maldevelopment (Fletcher et al., 2005). These tissue-specific promoters are promising candidates for developing tissue-specific editing tools, e.g., Cre/loxP recombinase system.

Enrichment analysis of TEs in H3K27ac peaks

Once considered as "junk DNA" (Biémont, 2010), TEs constitute a rich source of *cis*-regulatory elements involved in the regulation of eukaryotic gene expression via binding to TFs (Bourque et al., 2018). At present, however, the regulatory functions of TEs in pigs remain largely unexplored. First, we investigated the proportion of each TE class in certain functional genomic regions (Supplementary Figure S3A). Results showed that all types of TEs were strongly enriched in intronic and intergenic regions but not in TSSs or exons and were depleted in the two histone modification regions. Despite this, 1.45% (56 763/3 925 013) of the TEs in the pig genome overlapped with H3K27ac or H3K4me3 peaks. To investigate the heterogeneity of the epigenetic states of TEs across stage-tissue groups, we explored the TEs overlapping with the H3K27ac peaks in the 84 H3K27ac ChIP-seq data samples, given that the H3K27ac peaks tagged most of the H3K4me3 peaks (85.12%), and more samples were assayed for H3K27ac ChIP-seq. Results showed that 28–7 031 H3K27ac peaks overlapped with at least one TE in the different stage-tissue groups (Figure 3A). Peaks overlapping with TEs had lower intensities than other peaks across the stage-tissue groups (Supplementary Figure S3B), supporting the notion that TEs tend to be transcriptionally repressed (Colonna Romano & Fanti, 2022). Additionally, over half of the TE-overlapping peaks were active in only a single stage-tissue group (Supplementary Figure S3C), suggesting stage- and tissue-specific regulatory elements (Supplementary Figure S3D). Further analysis was conducted on various TE subfamilies to assess their enrichment in H3K27ac peaks across the different stage-tissue groups using a previously described pipeline (Judd et al., 2021). Three unique TEs, namely, LTR_ERVK_LTR4C_SS, LTR_ERV1_ERV1-2B_SSc-LTR, and LTR_ERVK_LTR6B_SS, were identified as significantly enriched in H3K27ac peaks in at least one stage-tissue group (Figure 3B).

LTR4_SS, with more than 50 observed bound copies, showed greater enrichment in both adult and fetal ovaries than in other tissues (Figure 3B, C). We further performed TF binding motif enrichment analysis on the 34 peaks overlapping the LTR4C_SS element in the ovaries using MEME (Bailey et al., 2015) and HOMER (Heinz et al., 2010). Analyses revealed enrichment of the binding motifs of SIX4 and SIX1,

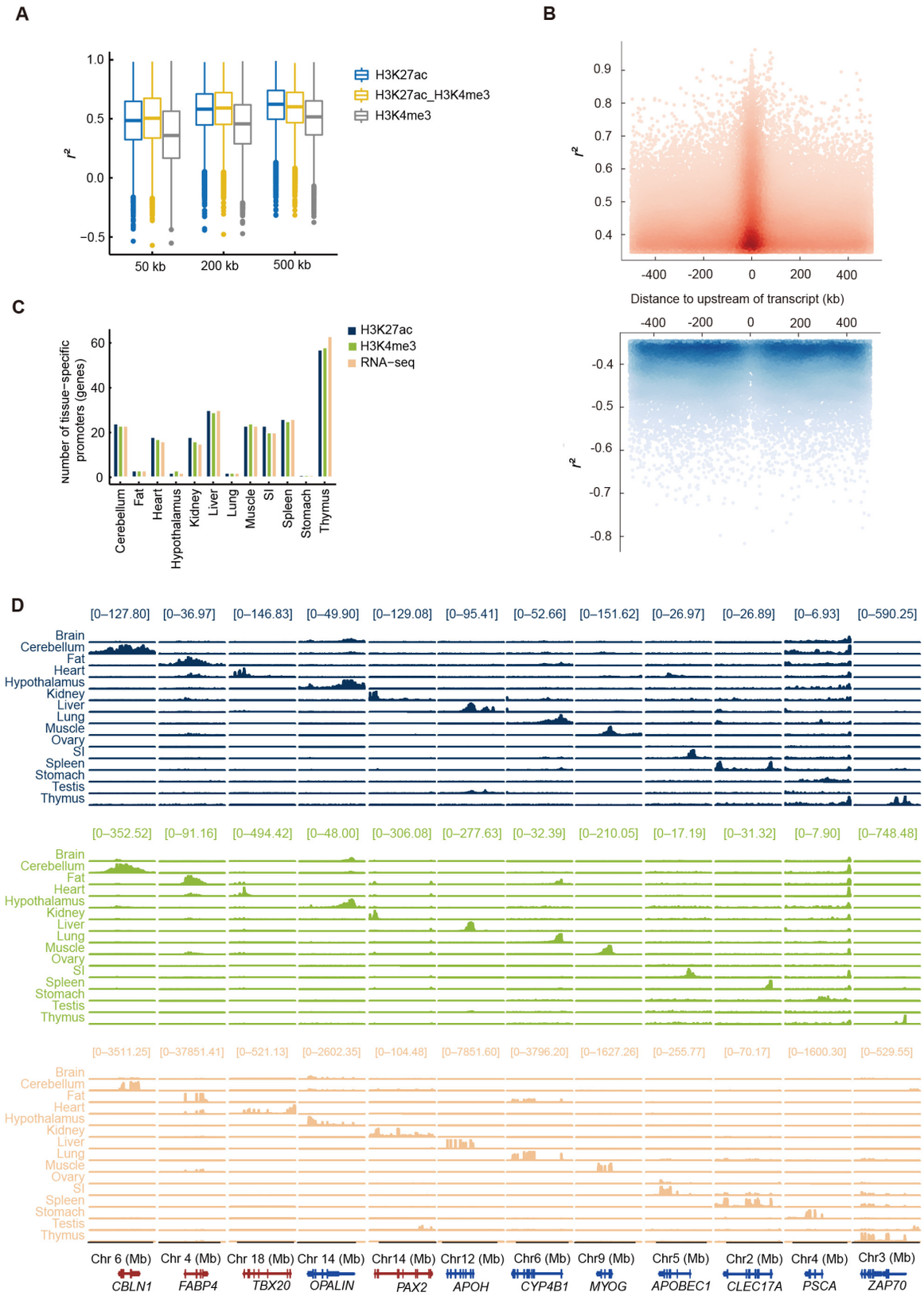

Figure 2 Histone modification levels are predictive for gene expression

A: Pearson correlations between predicted and measured gene expression levels by different models; X-axis corresponds to distance between center of peaks to TSSs of genes. B: Distribution of distance for significant positive peak-gene pairs and negative peak-gene pairs from H3K4me3 are shown in upper and lower panels, respectively. Dark color indicates clustering of peak-gene pairs at this distance. C: Number of tissue-specific peaks (genes) in different tissues. D: Representative examples of tissue-specific promoters. Each track represents average number of base-pairs per million mapped reads in 20 bp bins across all samples in corresponding tissue. Scale of track was consistent for each assay. Different colors indicate different types of assays as described in C.

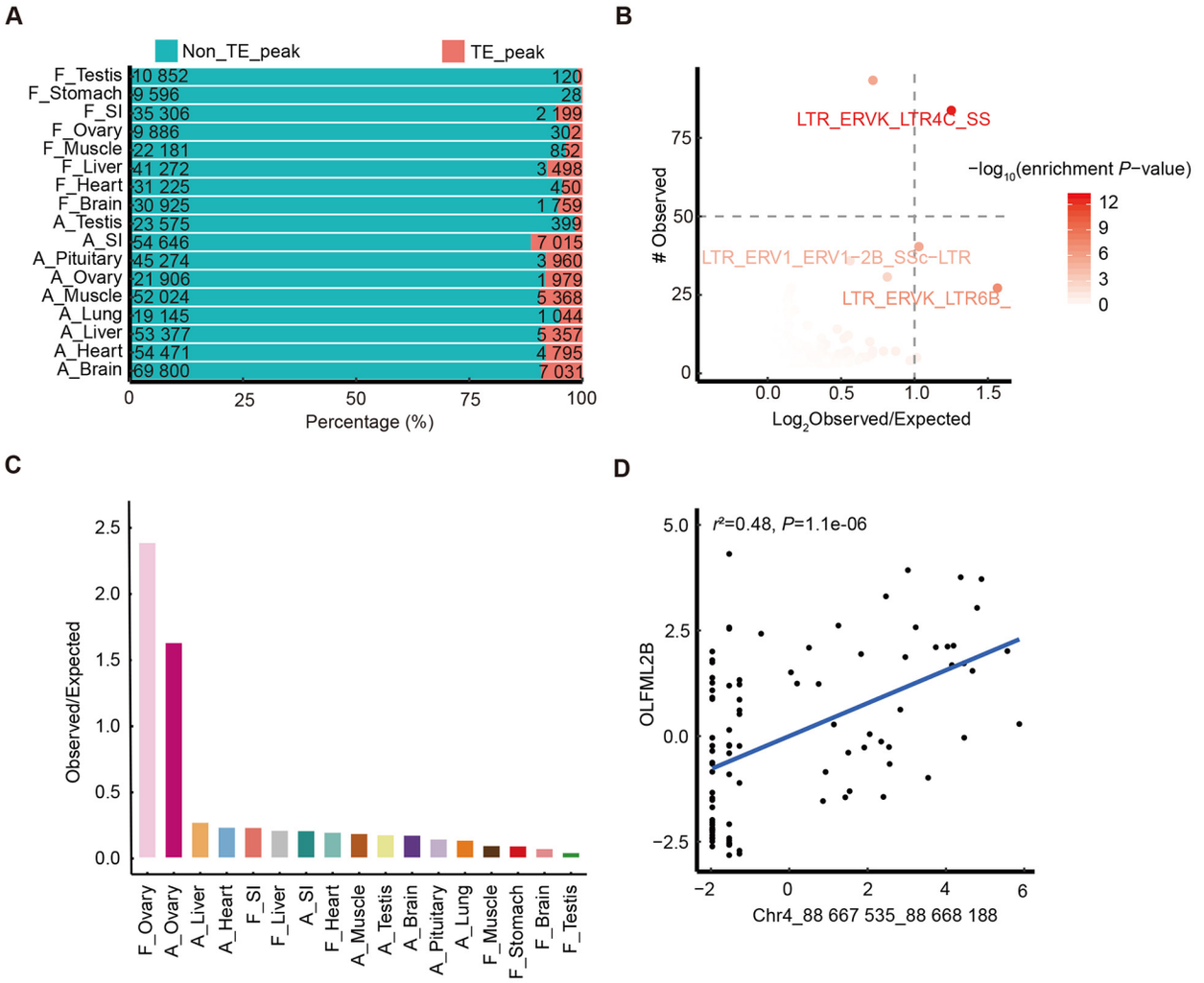

Figure 3 LTR4C_SS elements are highly active peaks enriched in ovaries

A: Bar plot showing numbers and fractions of peaks in stage-tissue groups that overlapped with TEs. B: All TEs detected in ChIP-seq peaks, plotted by number of elements observed compared to ratio of observed to expected occurrences for that particular TE. Expected values were calculated by bootstrapping 1 000 times. Dashed line denotes cutoff of >50 elements observed to filter false positives. C: Ratio of observed to expected occurrences of LTR4C_SS, as in B, for each stage-tissue group. D: Correlation between peak (chr4_88 667 535_88 668 188), which overlapped with LTR4C_SS and target gene (*OLFML2B*). Each dot represents peak intensity and gene expression from each sample.

further supported by the higher ChIP signatures around the binding motifs of SIX1 and SIX4 (Supplementary Figure S3E, F). *SIX1* is implicated in the advance of late-stage ovarian carcinoma by resisting TRAIL-mediated apoptosis (Behbakht et al., 2007), and *SIX4* is known to play an important role in maintaining the resting state of primary follicles and in restricting ovarian growth (Zhang et al., 2016). Overall, these results suggest that LTR4C_SS provides the binding motifs of SIX4 and SIX1, which may be critical for ovarian function in pigs. The overlapping LTR4C_SS peaks had an average distance of 118.87 kb from their closest genes (Supplementary Figure S3G), indicating that the LTR4C_SS element primarily acts as an enhancer. We identified four genes (*OLFML2B*, *SNORA11G*, *NFKBIZ*, and *ENSSSCG00000006350*) associated with the LTR4C_SS overlapping peaks (Figure 3D; Supplementary Figure S3H–J). Among these, *OLFML2B* (*P*=1.1e-6) participates in cell growth, maintenance, cell cycle regulation, apoptosis, and cell communication (Liu et al., 2019). *SNORA11G* (*P*=4.8e-6) encodes a snoRNA class essential for rRNA stability, processing, and fidelity of ribosome assembly and translation (Holley & Topkara, 2011). Dysfunction of the transcription factor *NFKBIZ* (*P*=1.4e-4) is associated with premature ovarian failure-18 (Safran et al., 2010).

Expression of TEs across tissues and developmental stages

Increasing evidence suggests that the expression of TEs in tissues is involved in placental development, immune system, brain development, and other biological functions (Bourque et al., 2018; Cornelis et al., 2017; Joly-Lopez & Bureau, 2018; Naville et al., 2016). Thus, we quantified both TEs and TE-gene junctions based on multiply mapped reads from the RNA-seq data using TEcount and TEtranscript (Jin et al., 2015), respectively (Supplementary Figure S4A). Results showed that TEs accounted for 0.12% to 0.83% of the RNA-seq reads in each stage-tissue group (Supplementary Figure S4B). We identified 8 068 transcribed TE-gene junctions in at least one of the tissues (Supplementary Figure S4C). The majority (89.71%) of the involved TEs were retrotransposons (Supplementary Figure S4C). Additionally, except for the testes, these genes showed higher expression in fetal tissues

than in adult tissues (Supplementary Figure S4C). Most TEs (84.02%) were expressed as parts of exons, while 20.56% of TEs were located in coding sequences, indicating potential in creating novel protein sequences (Supplementary Figure S4C).

Out of the 8 068 transcribed TE-gene junctions, 4 688 stage- and tissue-specific TE-gene junctions were identified with Tau scores greater than 0.85 (Supplementary Figure S4D), involving 483 TE subfamilies, including 123 LINE (25.47%), 37 SINE (7.67%), 191 LTR (39.54%), and 132 DNA transposon (27.33%) subfamilies. Among the 4 688 TE-gene junctions, 3 456 specifically influenced the exons of 2 135 genes. Notably, 57% of these genes (1 220/2 135) showed stage- and tissue-specific expression, including 10 in the fetal SI and 503 in the adult testes (Figure 4A). These genes are involved in pathways associated with the biology of the corresponding tissues (Figure 4A), such as cardiac muscle contraction in the adult heart and carboxylic acid catabolic processes in the adult liver. These findings support the notion that TEs can impact protein coding sequences and that a considerable proportion of TEs expressed in a tissue-specific manner may play important roles in tissue-specific functions.

Notably, more than 41% (1 967/4 688) of the stage-tissue specific TE-exon junctions were found in the adult testes, consistent with the observation that genes expressed in the testes more frequently experience TE insertion events (Bhalla, 2020; Reik et al., 2001). Several corresponding genes, including *ADAM30*, *ADAM32*, and *ZPBP*, are known to be involved in testicular biological functions, while other genes have not previously been shown to act in the testes. For example, *CSNK1A1* displayed high expression across all stage-tissue groups at the gene level (Figure 4C), while the TE (PRE1e)-exon 9 junction in *CSNK1A1* showed adult testis-specific expression (Figure 4B, D), indicating the potential role of PRE1e in the transcription of *CSNK1A1* in adult testes. Interestingly, other casein kinases, such as *CSNK1G2*, are known to suppress necroptosis-promoted testicular aging (Li et al., 2020). Similar patterns were also observed in other stage-tissue groups. For example, MamSINE1 serves as an adult brain-specific expressed exon for *ERMN*, LTR16B2 serves as an adult liver-specific expressed exon for *LECT2*, L1MB3 serves as a fetal brain-specific expressed exon for *STMN4*, MamRTE1 serves as an adult muscle-specific expressed exon for *TMOD4*, L1ME4c serves as a fetal ovary-specific expressed exon for *FIGLA*, and Tigger19a serves as an adult SI-specific expressed exon for *MS4A12* (Supplementary Figure S5A–F).

Stage tissue-specific alternative promoters derived from TEs

Furthermore, 717 junctions were identified where the corresponding TEs functioned as alternative promoters for 600 genes, among which 351 showed stage- and tissue-specific expression (Supplementary Figure S6A). Further screening of these TEs using the two ChIP-seq datasets led to the identification of 97 TE-derived promoters that overlapped with both histone modification regions (Figure 5B; Supplementary Table S7). This group included the promoter-overlapping

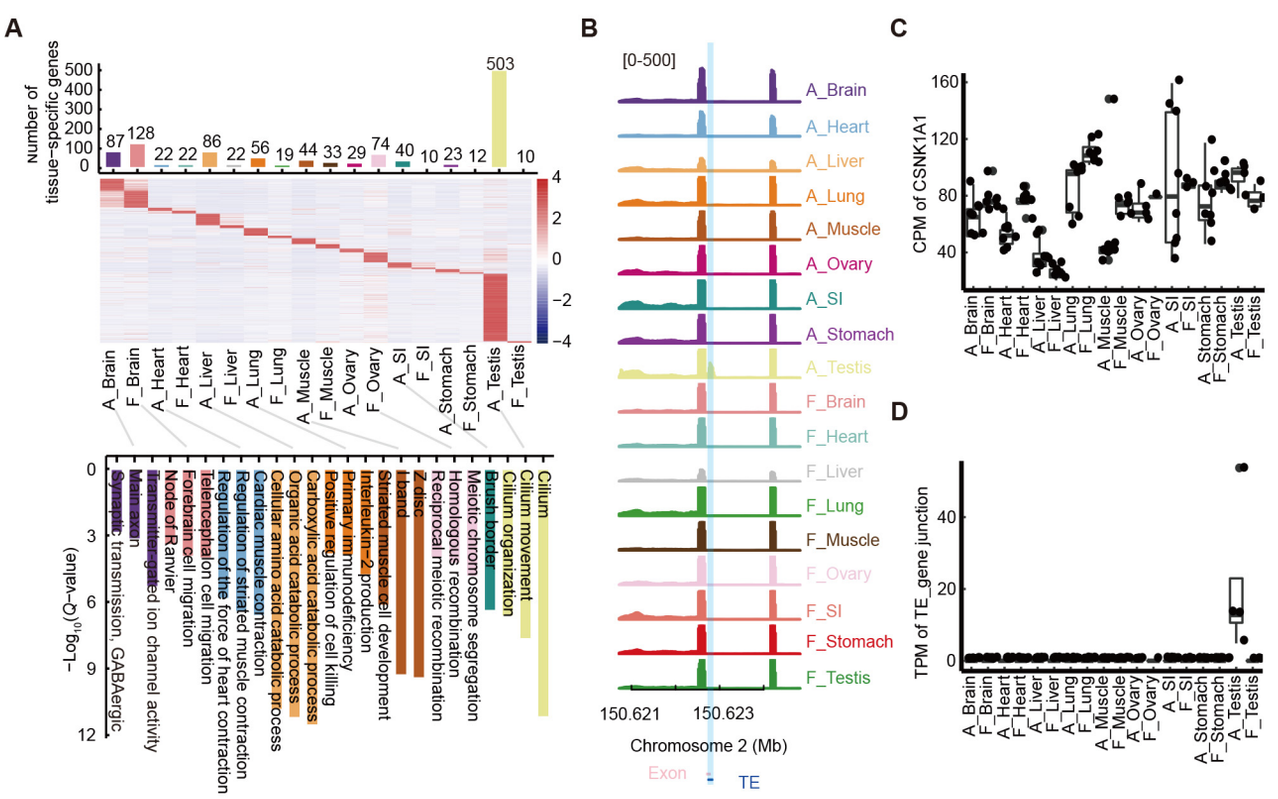

Figure 4 Expression of TE-containing genes

A: Heatmap showing stage-tissue specific gene expression of 3 456 TE-gene junctions, where TEs intersect with exons of corresponding genes. Bar plot above heatmap represents number of stage-tissue specific TE-containing genes. Bar plot below heatmap represents enriched GO terms and KEGG pathways in stage-tissue specific TE-containing genes. B: Tracks of TE-exon expression across stage-tissue groups; Y-axis corresponds to read depths of RNA-seq data. C: Expression of *CSNK1A1* across stage-tissue groups. D: Expression of PRE1e-*CSNK1A1* junction across stage-tissue groups.

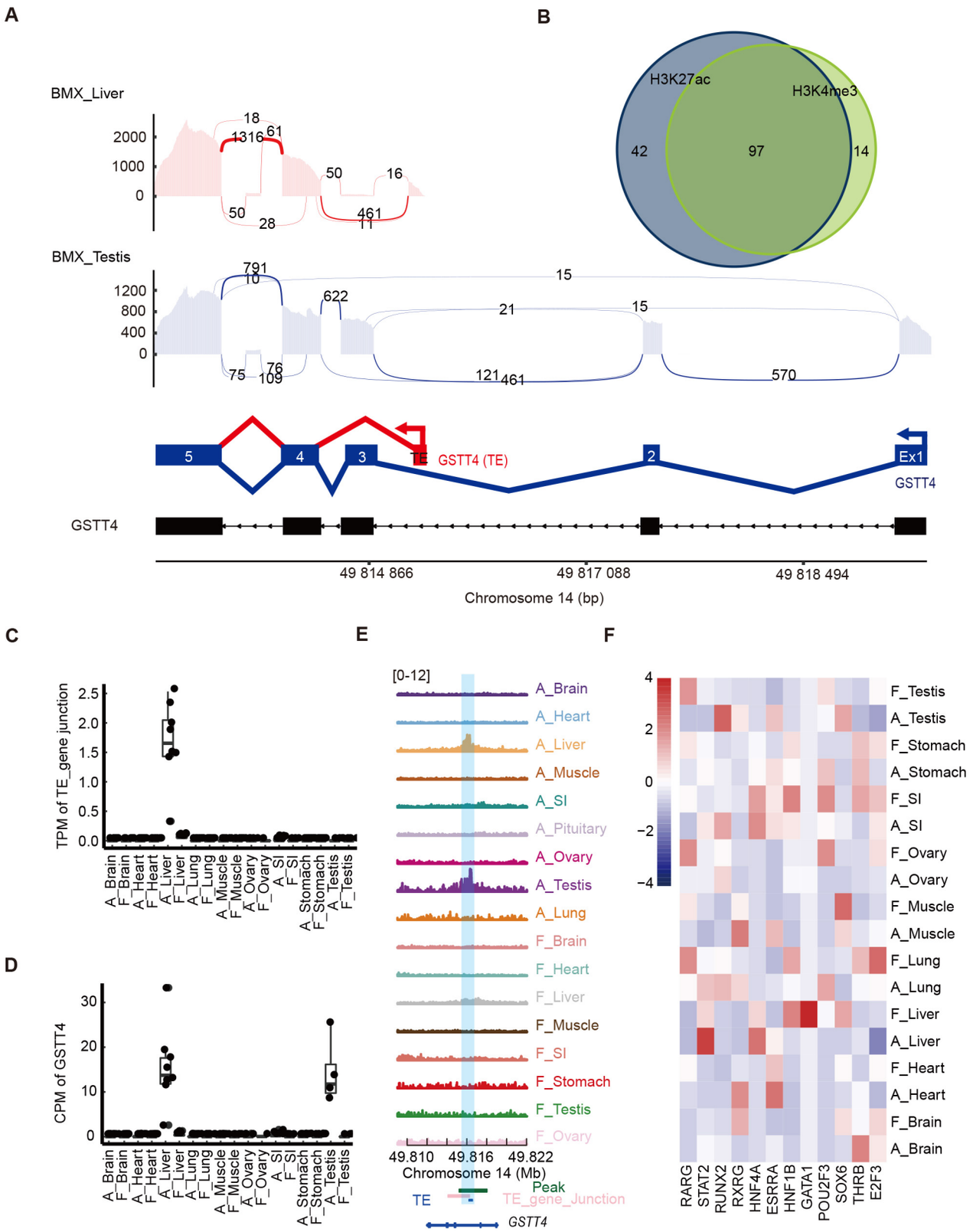

Figure 5 TE-associated tissue-specific alternative promoter usage of *GSTT4* gene

A: Assembled transcripts of *GSTT4* genes in adult BMX liver and adult BMX testis. Top: Sashimi-plot visualization of splicing-junctions from L1MC5-derived promoter. Bottom: Schematic representation of different exon usage. For TE-derived promoter, first exon originated from the TE-derived TSS and was spliced to a fourth exon and again to the fifth exon, skipping the canonical first to third exons. B: Venn diagram showing overlap between TEs involved in stage-tissue specific TE-gene junctions with H3K27ac and H3K4me3 peaks. C: Expression levels of *GSTT4* in all stage-tissue groups. D: Expression levels of L1MC5:GSTT4 junction in all stage-tissue groups. E: H3K27ac tracks in TE-derived *GSTT4* alternative promoter region. F: Heatmap of expression levels of TFs with potential binding motif in L1MC5. Color of each block represents scale value of gene expression level.

junction L1MC5:GSTT4, where the *GSTT4* gene, which was mainly expressed in the adult liver and testes (Figure 5D), is involved in glutathione metabolic processes. The L1MC5 element appeared to facilitate an alternative promoter for *GSTT4* in the liver, resulting in liver-specific expression of exons 4–5 (Figure 5A, C). In comparison, *GSTT4* expressed exons 1–5 in the testes, with low expression in the other tissues (Figure 5A; Supplementary Figure S6B). The liver-specific expression of the L1MC5-GSTT4 junction was associated with liver-specific activity of the two histone markers, while the high activity of the two modifications in the testes was unexpected (Figure 5E). Among all inferred TFs with potential TF-binding motifs in L1MC5 (chr14:49 816 220–49 816 558), *STAT2* showed significantly higher expression in the adult liver than in the other tissues (Figure 5F). *STAT2* is a transcriptional activator that responds to cytokines and growth factors and may drive the liver-specific expression of the L1MC5-GSTT4 junction-containing transcripts. Similar patterns were also observed in the adult testes. For example, MIR serves as an adult testis-specific promoter causing testis-specific expression of the *DLGAP5* transcript (Supplementary Figure S7A–C).

Cross-species conservation of tissue-specific expression of TE-gene junctions

We next examined the cross-species conservation of tissue-specific expression of TE-gene junctions. Using the same analytical pipeline, 8 560 and 3 957 stage- and tissue-specific TE-gene junctions were identified by analyzing published RNA-seq data from humans and mice, respectively (Supplementary Table S8). Notably, 31 genes that displayed conserved tissue-specific expression of TE-gene junctions were identified in the three species (Supplementary Tables S9–S11), among which 29 showed testis-specific expression. For example, the retrotransposon MLT1F1 is reported to drive testis-specific promoter usage of *SRPK2* (SRSF Protein Kinase 2, a gene associated with spermatogenesis and alternative splicing (Giannakouros et al., 2011)) in the three species (Figure 6A–C). Our results indicated that the insertion of MLT1F1 into the genome may precede the divergence of the three species. The MLT1F1 sequences in *SRPK2* from the three species showed 39%–51% sequence identity (Supplementary Figure S8). To identify potential TFs driving conserved alternative promoter use in the three species, FIMO (Grant et al., 2011) was used to predict the locations of TF-binding motifs in the MLT1F1 sequences, revealing a core "GTCATAAAA" sequence shared by all three species, predicted to bind to 14 TFs (Supplementary Figure S8). Among these TFs, HOX proteins are crucial for testicular and epididymal physiological functions and can affect male fertility (Ferguson & Agoulnik, 2013; Topaloğlu et al., 2021; Xian et al., 2015). Species-conserved TE-derived alternative promoters were not limited to the same subfamily of TEs, i.e., different TE subfamilies acted on the same gene across the three species. For example, the MIR element serves as a testis-specific promoter for *SUGP2*, which mediates the alternative splicing regulatory network during spermatogenesis in the testes (Zhan et al., 2021). The subfamilies of MIR were MIRb, MIR, and MIR1_Amn in pigs, mice, and humans, respectively (Supplementary Figure S9).

DISCUSSION

In this study, we performed a comprehensive analysis of H3K4me3 and H3K27ac ChIP-seq and RNA-seq data from 10 tissues of seven fetuses and eight adult pigs of both sexes, then annotated the regulatory sequences of TEs in the pig genome. In total, we identified 115 181 H3K4me3 peaks and 271 376 H3K27ac peaks, further enriching the repository of regulatory elements for pigs. We examined the predictive efficiency of different combinations of histone modifications on gene expression using ridge regression models and found that the activity of H3K27ac exhibited higher associations with gene expression than that of H3K4me3. Additionally, we identified ~230 tissue-specific promoters that may be particularly valuable for tissue-specific gene editing. Together, these results provide novel resources and new insights into the regulation of gene expression.

The colonization of TEs poses a considerable threat to genome integrity (Ardeljan et al., 2017), which may increase the risk of genomic instability (Ayarpadikannan & Kim, 2014). TE fragments embedded in the genome may also have diverse functional consequences (Fueyo et al., 2022). For example, Chang et al. (2022) found that TEs are sources of new protein-coding exons in zebrafish, while Lee et al. (2022) demonstrated that TEs contribute to exons and promoters of genes in zebrafish. Thus, we performed enrichment analysis of TEs using H3K27ac ChIP-seq. A notable finding was the identification of the TE LTR4C_SS, which was significantly enriched in ovarian tissue and harbored binding sites for SIX1 and SIX4 motifs, suggesting involvement in gonadal development.

We then focused on the expression of TEs across stage-tissue groups using RNA-seq to reveal temporal- and spatial-specific patterns of TE expression. In total, we identified 8 068 TE-gene junctions, with the corresponding genes expressed at higher levels in the fetuses than in the adults, except in the testes, reflecting imperfect epigenetic regulatory mechanisms in the fetuses. We also identified 4 688 stage- and tissue-specific TE-gene junctions, most of which were expressed in the adult testes. Remarkably, 3 456 of the 4 688 TE-gene junctions contained TE-derived exons. The corresponding genes (1 220/2 150) displayed stage- and tissue-specific patterns, with 503 genes specifically expressed in the adult testes and enriched in pathways relevant to their corresponding tissues. Notably, PRE1e, located within a *CSNK1A1* exon, demonstrated adult testis-specific expression, while the overall expression of *CSNK1A1* remained consistent across all stage-tissue groups. This gene phosphorylates many proteins and participates in Wnt signaling. Several factors could explain the preferential activation of TEs in adult testes: (1) TEs replicate in germline cells, allowing transportation of novel insertions to the next generation (Cosby et al., 2019); (2) males produce four meiotic products per meiotic division and thus have a higher tolerance for TE insertion (Bhalla, 2020); (3) subtle and brief temperature increases during spermatogenesis can increase TE activity and mobility (Kurhanewicz et al., 2020); (4) chromatin remodeling during spermatogenesis ensures developmental reprogramming, providing favorable conditions for TE activation and mobility (Castañeda et al., 2011); and (5) newly emerged genes tend to exhibit testis-specific expression before becoming more widespread, aligning with the "out of testis" hypothesis (Kaessmann, 2010). Thus, it would be interesting to investigate TE activity in the context of developmental transitions and differentiation trajectories of cells in single-cell data of testes in future work (Zhang et al.,

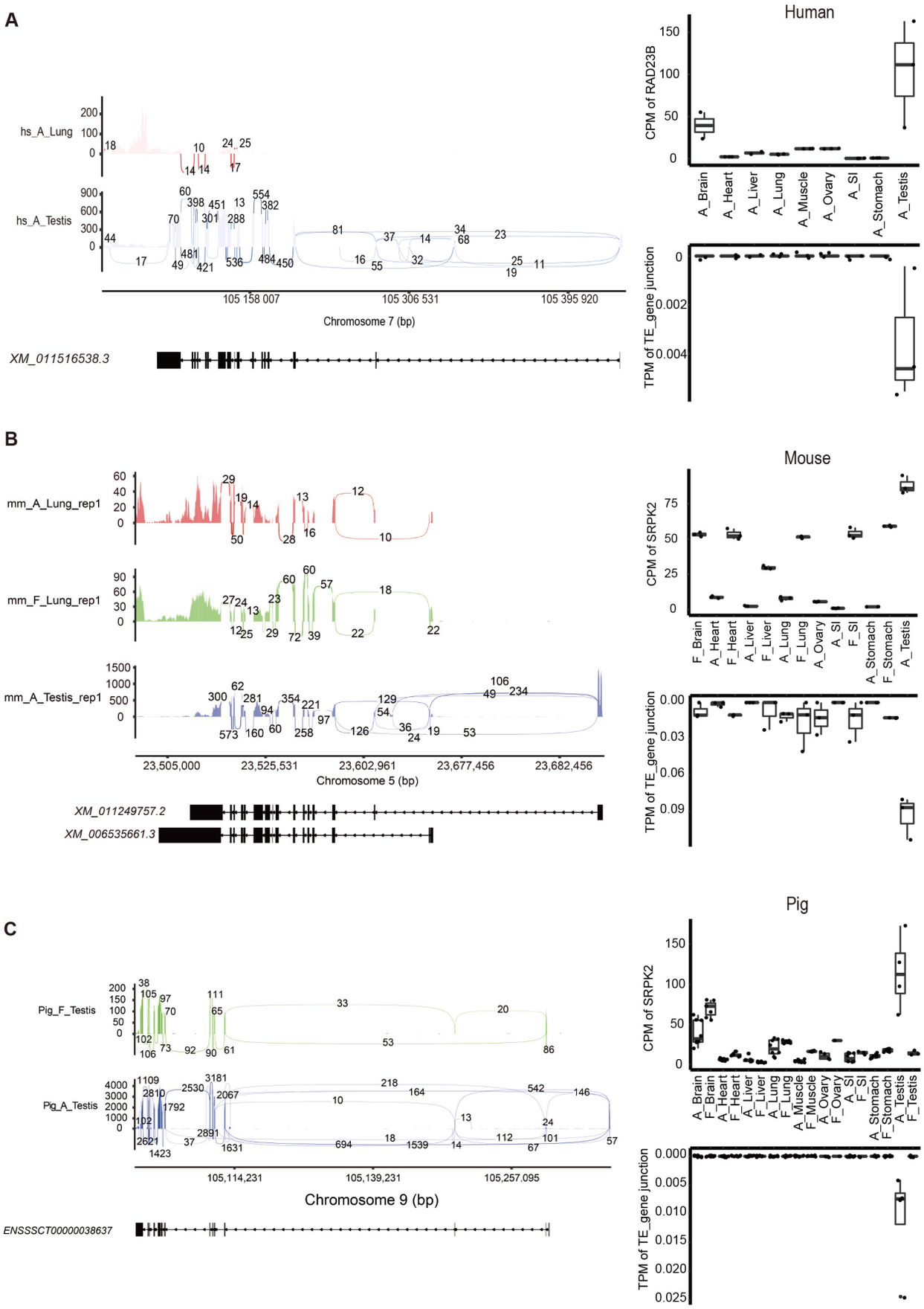

Figure 6 Retrotransposon promoters yield tissue-specific expression of *SRPK2* isoform across pigs, humans, and mice

A–C: Left, Sashimi plot showing RNA-seq reads spanning MLT1F1-SRPK2 junction in testis and lung from three species. MLT1F1 was located in promoter region of *SRPK2* in three species. Right, expression of *SRPK2* and MLT1F1-SRPK2 junction across stage-tissue groups in three species.

2022a).

Multiple lines of evidence suggest that specifically expressed TEs present in the genome can be recruited as promoter elements to reprogram host gene expression (Gardner et al., 2019; Miao et al., 2020). Similarly, we identified 351 stage- and tissue-specific TE-containing genes, with corresponding TEs functioning as alternative promoters. Most of these TE-derived promoter-driven stage- and tissue-specific genes were also expressed in the adult testes. To ensure the authenticity of these promoters, we screened 97 TE-derived alternative promoters supported by both H3K27ac and H3K4me3 peaks. For example, MIR acts as an adult testis-specific promoter, leading to testis-specific expression of the TE-containing *DLGAP5* transcript, a critical gene in cell cycle regulation with testis expression patterns (Liu et al., 2018). We also identified a TE-derived promoter in the adult liver and canonical promoters in the adult testes, both regulating the *GSTT4* gene, which is involved in glutathione metabolism. The observed enrichment of TF-binding motifs indicated that STAT2, an adult liver-specific gene, may also activate this promoter. This supports previous research showing that the same gene can transcribe different transcripts in different tissues depending on specific promoters (Lee et al., 2022). These findings further underscore the extensive recruitment of TEs in the regulation and expression of host genes (Chuong et al., 2017).

Further analysis was conducted to explore the conserved functional roles of TEs among pigs, humans, and mice. This analysis identified 31 homologous genes across these species, demonstrating TE involvement in their expression. Notably, MLT1F1 was found to regulate the expression of *SRPK2* in the adult testes. *SRPK2*, which was highly expressed in the testes, is associated with spermatogenesis (Giannakouros et al., 2011). These results suggest that the insertion of MLT1F1 into the genome likely occurred before the evolutionary divergence of humans, mice, and pigs, evolving into a crucial functional sequence across these species.

DATA AVAILABILITY

All epigenomic sequence data used in this study were submitted to the China National GeneBank DataBase (CNGBdb) with accession code: CNP0001696 (https://db.cngb.org/search/project/CNP0001696/), Genome Sequence Archive (GSA) database (https://ngdc.cncb.ac.cn/gsa/) under accession number CRA013785, Science Data Bank (doi:10.57760/sciencedb.j00139.00086) and NCBI under BioProjectID PRJNA1049515.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

L.S.H. designed and supervised this study and revised the manuscript. B.Y. wrote and revised the manuscript and supervised the data analysis; T.J. and Z.M.Z. performed the experiments, analyzed the data, and wrote the manuscript; Z.Q.L., Q.Z., Z.Z.W., J.W.Y., and S.Y.Y. performed the experiments. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We are grateful to colleagues from the State Key Laboratory of Swine Genetic Improvement and Production Technology, Jiangxi Agricultural University, for sample collection.

REFERENCES

Abuín JM, Pichel JC, Pena TF, et al. 2015. BigBWA: approaching the Burrows-Wheeler aligner to Big Data technologies. *Bioinformatics*, **31**(24): 4003−4005.

Ardeljan D, Taylor MS, Ting DT, et al. 2017. The human long interspersed element-1 retrotransposon: an emerging biomarker of neoplasia. *Clinical Chemistry*, **63**(4): 816−822.

Ayarpadikannan S, Kim HS. 2014. The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. *Genomics & Informatics*, **12**(3): 98−104.

Bailey TL, Johnson J, Grant CE, et al. 2015. The MEME suite. *Nucleic Acids Research*, **43**(W1): W39−W49.

Barski A, Cuddapah S, Cui KR, et al. 2007. High-resolution profiling of histone methylations in the human genome. *Cell*, **129**(4): 823−837.

Behbakht K, Qamar L, Aldridge CS, et al. 2007. Six1 overexpression in ovarian carcinoma causes resistance to TRAIL-mediated apoptosis and is associated with poor survival. *Cancer Research*, **67**(7): 3036−3042.

Bernstein BE, Kamal M, Lindblad-Toh K, et al. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, **120**(2): 169−181.

Bhalla N. 2020. Meiosis: is spermatogenesis stress an opportunity for evolutionary innovation?. *Current Biology*, **30**(24): R1471−R1473.

Biémont C. 2010. A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics*, **186**(4): 1085−1093.

Bourque G, Burns KH, Gehring M, et al. 2018. Ten things you should know about transposable elements. *Genome Biology*, **19**(1): 199.

Castañeda J, Genzor P, Bortvin A. 2011. piRNAs, transposon silencing, and germline genome integrity. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **714**(1-2): 95−104.

Chang NC, Rovira Q, Wells J, et al. 2022. Zebrafish transposable elements show extensive diversification in age, genomic distribution, and developmental expression. *Genome Research*, **32**(7): 1408−1423.

Chen JQ, Zhang MP, Tong XK, et al. 2022. Scan of the endogenous retrovirus sequences across the swine genome and survey of their copy number variation and sequence diversity among various Chinese and Western pig breeds. *Zoological Research*, **43**(3): 423−441.

Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science*, **351**(6277): 1083−1087.

Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics*, **18**(2): 71−86.

Colonna Romano N, Fanti L. 2022. Transposable elements: major players in shaping genomic and evolutionary patterns. *Cells*, **11**(6): 1048.

Cornelis G, Funk M, Vernochet C, et al. 2017. An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental *Mabuya* lizard. *Proceedings of the National Academy of Sciences of the United States of America*, **114**(51): E10991−E11000.

Cosby RL, Chang NC, Feschotte C. 2019. Host-transposon interactions: conflict, cooperation, and cooption. *Genes & Development*, **33**(17-18): 1098−1116.

Creyghton MP, Cheng AW, Welstead GG, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(50): 21931−21936.

Delaneau O, Zazhytska M, Borel C, et al. 2019. Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science*, **364**(6439): eaat8266.

Dobin A, Davis CA, Schlesinger F, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**(1): 15−21.
Duttke SHC, Lacadie SA, Ibrahim MM, et al. 2015. Human promoters are intrinsically directional. *Molecular Cell*, **57**(4): 674−684.
Fedoroff NV. 2012. Presidential address. Transposable elements, epigenetics, and genome evolution. *Science*, **338**(6108): 758−767.
Ferguson L, Agoulnik AI. 2013. Testicular cancer and cryptorchidism. *Frontiers in Endocrinology*, **4**: 32.
Fletcher J, Hu M, Berman Y, et al. 2005. Multicystic dysplastic kidney and variable phenotype in a family with a novel deletion mutation of *PAX2*. *Journal of the American Society of Nephrology*, **16**(9): 2754–2761.
Fueyo R, Judd J, Feschotte C, et al. 2022. Roles of transposable elements in the regulation of mammalian transcription. *Nature Reviews Molecular Cell Biology*, **23**(7): 481−497.
Gardner EJ, Prigmore E, Gallone G, et al. 2019. Contribution of retrotransposition to developmental disorders. *Nature Communications*, **10**(1): 4630.
Giannakouros T, Nikolakaki E, Mylonis I, et al. 2011. Serine-arginine protein kinases: a small protein kinase family with a large cellular presence. *The FEBS Journal*, **278**(4): 570−586.
Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**(7): 1017−1018.
Heinz S, Benner C, Spann N, et al. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, **38**(4): 576−589.
Holley CL, Topkara VK. 2011. An introduction to small non-coding RNAs: miRNA and snoRNA. *Cardiovascular Drugs and Therapy*, **25**(2): 151−159.
Howe FS, Fischl H, Murray SC, et al. 2017. Is H3K4me3 instructive for transcription activation?. *BioEssays*, **39**(1): 1−12.
Hughes AL, Kelley JR, Klose RJ. 2020. Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation. *Biochimica et Biophysica Acta* (*BBA*) - *Gene Regulatory Mechanisms*, **1863**(8): 194567.
Jacques PÉ, Jeyakani J, Bourque G. 2013. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genetics*, **9**(5): e1003504.
Jin Y, Tam OH, Paniagua E, et al. 2015. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics*, **31**(22): 3593−3599.
Joly-Lopez Z, Bureau TE. 2018. Exaptation of transposable element coding sequences. *Current Opinion in Genetics & Development*, **49**: 34−42.
Judd J, Sanderson H, Feschotte C. 2021. Evolution of mouse circadian enhancers from transposable elements. *Genome Biology*, **22**(1): 193.
Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Research*, **20**(10): 1313−1326.
Kapusta A, Kronenberg Z, Lynch VJ, et al. 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genetics*, **9**(4): e1003470.
Karlić R, Chung HR, Lasserre J, et al. 2010. Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(7): 2926−2931.
Kern C, Wang Y, Xu XQ, et al. 2021. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nature Communications*, **12**(1): 1821.
Kirk EP, Sunde M, Costa MW, et al. 2007. Mutations in cardiac T-box factor gene *TBX20* are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. *The American Journal of Human Genetics*, **81**(2): 280−291.
Kryuchkova-Mostacci N, Robinson-Rechavi M. 2017. A benchmark of gene expression tissue-specificity metrics. *Briefings in Bioinformatics*, **18**(2): 205−214.
Kunarso G, Chia NY, Jeyakani J, et al. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics*, **42**(7): 631−634.
Kurhanewicz NA, Dinwiddie D, Bush ZD, et al. 2020. Elevated temperatures cause transposon-associated DNA damage in *C. elegans* spermatocytes. *Current Biology*, **30**(24): 5007−5017.e4.
Lee HJ, Hou YR, Maeng JH, et al. 2022. Epigenomic analysis reveals prevalent contribution of transposable elements to *cis*-regulatory elements, tissue-specific expression, and alternative promoters in zebrafish. *Genome Research*, **32**(7): 1424−1436.
Li DR, Ai YW, Guo J, et al. 2020. Casein kinase 1G2 suppresses necroptosis-promoted testis aging by inhibiting receptor-interacting kinase 3. *eLife*, **9**: e61564.
Li H, Handsaker B, Wysoker A, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16): 2078−2079.
Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**(7): 923−930.
Liu JX, Liu Z, Zhang XZ, et al. 2019. Bioinformatic exploration of OLFML2B overexpression in gastric cancer base on multiple analyzing tools. *BMC Cancer*, **19**(1): 227.
Liu M, Qiu YL, Jin T, et al. 2018. Meta-analysis of microarray datasets identify several chromosome segregation-related cancer/testis genes potentially contributing to anaplastic thyroid carcinoma. *PeerJ*, **6**: e5822.
Lunney JK, Van Goor A, Walker KE, et al. 2021. Importance of the pig as a human biomedical model. *Science Translational Medicine*, **13**(621): eabd5758.
Miao BP, Fu SH, Lyu C, et al. 2020. Tissue-specific usage of transposable element-derived promoters in mouse development. *Genome Biology*, **21**(1): 255.
Modzelewski AJ, Shao WQ, Chen JQ, et al. 2021. A mouse-specific retrotransposon drives a conserved *Cdk2ap1* isoform essential for development. *Cell*, **184**(22): 5541−5558.e22.
Naville M, Warren IA, Haftek-Terreau Z, et al. 2016. Not so bad after all: retroviruses and long terminal repeat retrotransposons as a source of new genes in vertebrates. *Clinical Microbiology and Infection*, **22**(4): 312−323.
Niknafs YS, Pandian B, Iyer HK, et al. 2017. TACO produces robust multisample transcriptome assemblies from RNA-seq. *Nature Methods*, **14**(1): 68−70.
Pan ZY, Yao YL, Yin HW, et al. 2021. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nature Communications*, **12**(1): 5848.
Pekowska A, Benoukraf T, Zacarias-Cabeza J, et al. 2011. H3K4 tri-methylation provides an epigenetic signature of active enhancers. *The EMBO Journal*, **30**(20): 4198−4210.
Pertea M, Pertea GM, Antonescu CM, et al. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, **33**(3): 290−295.
Qian L, Mohapatra B, Akasaka T, et al. 2008. Transcription factor neuromancer/TBX20 is required for cardiac function in *Drosophila* with implications for human heart disease. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(50): 19833−19838.
Quinlan AR. 2014. BEDTools: the swiss-army tool for genome feature analysis. *Current Protocols Bioinformatics*, **47**: 11.12.1−11.12.34.
Ramírez F, Dündar F, Diehl S, et al. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*, **42**(W1): W187−W191.
Rebollo R, Romanish MT, Mager DL. 2012. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annual Review of Genetics*, **46**: 21−42.
Reik W, Dean W, Walter J. 2001. Epigenetic reprogramming in mammalian

development. *Science*, **293**(5532): 1089−1093.

Roth SY, Denu JM, Allis CD. 2001. Histone acetyltransferases. *Annual Review of Biochemistry*, **70**: 81−120.

Safran M, Dalah I, Alexander J, et al. 2010. GeneCards Version 3: the human gene integrator. *Database*, **2010**: baq020.

Senft AD, Macfarlan TS. 2021. Transposable elements shape the evolution of mammalian development. *Nature Reviews Genetics*, **22**(11): 691−711.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, **8**(4): 272−285.

Sundaram V, Choudhary MNK, Pehrsson E, et al. 2017. Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nature Communications*, **8**(1): 14550.

Sundaram V, Wysocka J. 2020. Transposable elements as a potent source of diverse *cis*-regulatory sequences in mammalian genomes. *Philosophical Transactions of the Royal Society B:Biological Sciences*, **375**(1795): 20190347.

Tarailo-Graovac M, Chen NS. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*,doi: 10.1002/0471250953.bi0410s25.

Ting CN, Rosenberg MP, Snow CM, et al. 1992. Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes & Development*, **6**(8): 1457−1465.

Topaloğlu U, Akbalik ME, Sağsöz H. 2021. Immunolocalization of some HOX proteins in immature and mature feline testes. *Anatomia, Histologia, Embryologia*, **50**(4): 726−735.

Utomo ARH, Nikitin AY, Lee WH. 1999. Temporal, spatial, and cell type-specific control of Cre-mediated DNA recombination in transgenic mice. *Nature Biotechnology*, **17**(11): 1091−1096.

Wu YQ, Zhao H, Li YJ, et al. 2020. Genome-wide identification of imprinted genes in pigs and their different imprinting status compared with other mammals. *Zoological Research*, **41**(6): 721−725.

Xian H, Xian Y, Liu LL, et al. 2015. Expression of β-nerve growth factor and homeobox A10 in experimental cryptorchidism treated with exogenous nerve growth factor. *Molecular Medicine Reports*, **11**(4): 2875−2881.

Yang Y, Adeola AC, Xie HB, et al. 2018. Genomic and transcriptomic analyses reveal selection of genes for puberty in Bama Xiang pigs. *Zoological Research*, **39**(6): 424−430.

Zhan JF, Li JB, Wu YR, et al. 2021. Chromatin-associated protein sugp2 involved in mRNA alternative splicing during mouse spermatogenesis. *Frontiers in Veterinary Science*, **8**: 754021.

Zhang LK, Ma HD, Guo M, et al. 2022a. Dynamic transcriptional atlas of male germ cells during porcine puberty. *Zoological Research*, **43**(4): 600−603.

Zhang Q, Li J, Zhang YF, et al. 2022b. Whole-genome sequence-based association study for immune cells in an eight-breed pig heterogeneous population. *Journal of Genetics and Genomics*, **49**(11): 1068−1071.

Zhang Y, Liu T, Meyer CA, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, **9**(9): R137.

Zhang Y, Zhao B, Roy S, et al. 2016. microRNA-309 targets the Homeobox gene *SIX4* and controls ovarian development in the mosquito *Aedes aegypti*. *Proceedings of the National Academy of Sciences of the United States of America*, **113**(33): E4828–E4836.

Zhao YX, Hou Y, Xu YY, et al. 2021. A compendium and comparative epigenomics analysis of *cis*-regulatory elements in the pig genome. *Nature Communications*, **12**(1): 2217.

Zhong LP, Zheng M, Huang YZ, et al. 2023. An atlas of expression quantitative trait loci of microRNAs in *longissimus* muscle of eight-way crossbred pigs. *Journal of Genetics and Genomics*, **50**(6): 398−409.

Zhu YL, Zhou ZM, Huang T, et al. 2022. Mapping and analysis of a spatiotemporal H3K27ac and gene expression spectrum in pigs. *Science China Life Sciences*, **65**(8): 1517−1534.