



Research article

Regulation of the three-dimensional chromatin organization by transposable elements in pig spleen

Yuzhuo Li^a, Hairui Fan^{a,b}, Weiyun Qin^{a,b}, Yejun Wang^c, Shuai Chen^a, Wenbin Bao^b, Ming-an Sun^{a,d,e,f,*}^a Institute of Comparative Medicine, College of Veterinary Medicine, Yangzhou University, Yangzhou 225009, Jiangsu, China^b College of Animal Science and Technology, Yangzhou University, Yangzhou 225009, Jiangsu, China^c Youth Innovation Team of Medical Bioinformatics, Shenzhen University Health Science Center, Shenzhen 518060, China^d Joint International Research Laboratory of Important Animal Infectious Diseases and Zoonoses of Jiangsu Higher Education Institutions, Yangzhou University, Yangzhou 225009, China^e Jiangsu Co-innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonosis, Yangzhou University, Yangzhou 225009, Jiangsu, China^f Joint International Research Laboratory of Agriculture and Agri-Product Safety of Ministry of Education of China, Yangzhou University, Yangzhou 225009, Jiangsu, China

ARTICLE INFO

Keywords:

Transposable element
Endogenous retrovirus
Pig spleen
3D chromatin organization
CTCF
Hi-C

ABSTRACT

Like other mammalian species, the pig genome is abundant with transposable elements (TEs). The importance of TEs for three-dimensional (3D) chromatin organization has been observed in species like human and mouse, yet current understanding about pig TEs is absent. Here, we investigated the contribution of TEs for the 3D chromatin organization in three pig tissues, focusing on spleen which is crucial for both adaptive and innate immunity. We identified dozens of TE families overrepresented with CTCF binding sites, including LTR22_SS, LTR15_SS and LTR16_SSc which are pig-specific families of endogenous retroviruses (ERVs). Interestingly, LTR22_SS elements harbor a CTCF motif and create hundreds of CTCF binding sites that are associated with adaptive immunity. We further applied Hi-C to profile the 3D chromatin structure in spleen and found that TE-derived CTCF binding sites correlate with chromatin insulation and frequently overlap TAD borders and loop anchors. Notably, one LTR22_SS-derived CTCF binding site demarcate a TAD boundary upstream of XCL1, which is a spleen-enriched chemokine gene important for lymphocyte trafficking and inflammation. Overall, this study represents a first step toward understanding the function of TEs on 3D chromatin organization regulation in pigs and expands our understanding about the functional importance of TEs in mammals.

1. Introduction

The mammalian genomes, which are usually meters long, are packaged into three-dimensional (3D) chromatin structure in the nucleus [1]. The 3D organization enables the looping of distal enhancers to their target promoters (*i.e.* E-P loops), thus is crucial for transcriptional regulation [2]. In principle, topologically associating domains (TADs) form the basic architectural chromatin units, with the chromatin contacts preferred within instead of across TADs [3]. Recent applications of chromosome conformation capture (3C)-based techniques, particularly Hi-C, have greatly improved our understanding on 3D chromatin organization [4,5]. Increasing evidence suggest that the 3D chromatin organization is not only important for development and cell differentiation

but also associated with various diseases [6,7].

The transcription factor CCCTC-binding factor (CTCF), a highly conserved and ubiquitously expressed C2H2 zinc finger protein, is a key player for 3D chromatin organization [8]. Together with cohesion and likely through a loop extrusion mechanism, it serves as an insulator protein to create TAD boundaries and as anchors to mediate E-P loops [9,10]. While CTCF binding is largely constitutive [11], many loci also show dynamic CTCF binding during cell differentiation [12] or under stress/stimulation [13]. The biological importance of CTCF have been demonstrated by plenty of studies. For example, cell-type-specific CTCF binding is crucial for the transcriptional regulation during hematopoiesis [14]. CTCF also facilitates the acute inflammatory response in macrophages [15]. Interestingly, one IFITM3 SNP influences the risk of

* Corresponding author at: Institute of Comparative Medicine, College of Veterinary Medicine, Yangzhou University, Yangzhou 225009, Jiangsu, China.

E-mail address: mingansun@yzu.edu.cn (M.-a. Sun).

<https://doi.org/10.1016/j.csbj.2023.09.029>

Received 7 August 2023; Received in revised form 23 September 2023; Accepted 23 September 2023

Available online 25 September 2023

2001-0370/© 2023 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

severe influenza in humans by affecting CTCF boundary activity [16]. These studies highlight the importance of CTCF binding on 3D chromatin structure and transcriptional regulation.

Despite the high conservation of CTCF protein, its binding sites have undergone waves of expansions – frequently through lineage-specific transposable element (TE) insertions [17,18]. TEs are mobile DNA elements constitute a large proportion of mammalian genomes [19]. While traditionally been regarded as “junk DNA”, many TE families – particularly those belonging to endogenous retrovirus (ERV) – are known to facilitate the regulatory evolution in mammals [20–23]. Increasing evidence also suggests the importance of TEs for 3D chromatin organization by mediating CTCF binding [24]. For example, frequent binding of CTCF on TEs was observed in both human and mouse [25–28]. Comprehensive across-species comparison further suggests that TEs have facilitated species-specific expansions of CTCF binding [29]. Recent studies further revealed that TE-derived CTCF binding loci can be co-opted to mediate TAD boundaries and chromatin loops [30–32].

The domestic pig (*Sus scrofa*) is valuable for meat production, organ xenotransplantation and biomedical studies [33]. TEs are also abundant in pig genome, with ~40% of pig genome made up of TEs and 4.8% as LTRs/ERVs [34]. While plenty attentions have been paid to the risk of porcine ERVs (PERVs) for xenotransplantation [35,36], current knowledge about their regulatory function in pigs is lacking. It is also unclear how TEs may mediate the 3D chromatin organization in pigs. We and others previously analyzed the contribution of TEs (particularly ERVs) for the regulatory evolution in species like human and mouse [37,20,21, 38,39], and it is attractive to investigate their potential regulatory roles of pig TEs. The accumulation of OMICs data of pigs in recent years [40–44] provides an opportunity for the systematical characterization of the regulatory function of porcine TEs.

Here, we performed integrative analysis of public and newly generated ChIP-seq, RNA-seq and/or Hi-C data in three porcine tissues, focusing on the involvement of pig-specific ERVs for 3D chromatin organization in pig spleen. We identified dozens of TE families that are likely to mediate CTCF binding in pigs. Importantly, our data highlights the importance of pig-specific ERVs for mediating CTCF binding and TAD formation in pig spleen, which may further influence immune gene expression. Overall, this study improves our understanding about the regulatory function of porcine TEs and highlights the importance of TEs for regulating 3D chromatin organization in pigs.

2. Materials and methods

2.1. Pig spleen sample collection

Two male pigs, aged 70–80 days and weighing 80–90 kg, were selected for slaughter. The spleen was immediately collected and cut to small pieces of approximately 1 mg. After flushed with phosphate-buffered saline (PBS), the spleen samples were flash-frozen in liquid nitrogen for preservation.

2.2. In situ Hi-C experiments

Hi-C experiments were performed following a previous study [4] with slight modifications. In brief, following cross-linking with 1% formaldehyde and nuclei extraction, the chromatin was digested with Dpn II and labeled with biotinylated residue after the 5' overhangs were filled. Blunt-end ligation of crosslinked DNA fragments was performed to obtain circular molecules, and then the DNA product was purified and sonicated into DNA fragments. Hi-C library was constructed by capturing the labeled target DNA fragments with biotin labeling precipitation technique followed by size selection. Finally, the obtained Hi-C library was sequenced by BGI company as 150 bp paired-end reads on DNBSEQ platform.

2.3. Reference genome and annotation

The reference genomes and gene annotations for human (GRCh38), mouse (GRCm37) and pig (Sscrofa11.1) were downloaded from Ensembl database [45]. Genome annotation files of GTF format were downloaded from Ensembl database [45]. Transposable element annotations for corresponding species were downloaded from UCSC Genome Browser [46].

2.4. ChIP-seq analysis

The raw reads were first trimmed with Trim Galore! v0.6.5 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), and then aligned to the corresponding reference genome (Sscrofa11.1 for pig) using Bowtie v2.3.5 [47] with default settings. Peak calling was performed using MACS2 v2.2.6 [48] with settings: `-g 2.1e9 -keep-dup all -q 0.01`. High repeatability peaks were obtained through the IDR method. BigWig files were generated using the *bamCoverage* function of deepTools v3.5.1 [49] with settings: `-normalizeUsing RPKM`.

2.5. RNA-seq analysis

Raw reads were trimmed with Trim Galore! v0.6.5, and then mapped to the reference genome (Sscrofa11.1) using STAR v2.7.3 [50]. Alignments with alignment MAPQ score < 30 were filtered using SAMtools v1.13 [51]. BigWig files were generated using the *bamCoverage* function of deepTools v3.5.1 [49] with settings: `-normalizeUsing RPKM`.

2.6. ATAC-seq analysis

Raw reads were first trimmed with Trim Galore! v0.6.5, then aligned using BWA v0.7.17 [52] with the “mem” alignment mode. Alignments with a MAPQ score < 30 were filtered using SAMtools v1.10 [51], and then PCR duplicates were marked and removed using the Picard toolkit v2.26.3 (<https://github.com/broadinstitute/picard>). Peak calling was performed using MACS2 v2.2.6 [48] with the same genomic parameters as ChIP-Seq. ATAC-Seq peaks were called in broad mark mode with a q-value cutoff of 0.05. High repeatability peaks were obtained through the IDR method. BigWig files were generated using the *bamCoverage* function of deepTools v3.5.1 [49] with settings: `-normalizeUsing RPKM`.

2.7. Hi-C analysis

The clean Hi-C reads were iteratively mapped to the pig genome with HiC-Pro v3.1.0 [53]. The *makeTagDirectory* module of HOMER [54] was used to convert the alignment file into the HOMER-style tag directory. Principal Component Analysis (PCA) was performed to infer the A/B compartments along the genome with the *runHiCpca.pl* script of HOMER [54] with settings: `-res 500000 -genome susScr11`. The TAD structure (insulation/boundaries) was defined with Arrowhead by juicer tools v1.6 [55]. The contact matrix normalized by the ICE method [56] was used to calculate the insulation score using the *matrix2insulation.pl* (<https://github.com/dekkerlab/cworld-dekker>) with the following settings: `-is 100000 -ids 60000 -nt 0.1 -ss 160000`. Chromatin loops were analyzed by HiCCUPS [5] with minor modifications. The resolution parameters were set as `“-r 5000,10000,25000 -ignore_sparsity”`.

2.8. TE analysis

TE enrichment analysis was performed following the procedure from our previous study [20]. In brief, we adopted the *fisher* function of BEDtools v2.29.2 [57] with default settings to calculate the enrichment fold and p-values of each TE families within given genomic regions (e.g. CTCF binding sites) by using Fisher’s Exact Test. To control for Family Wise Error Rate, the calculated p-values were further adjusted with

Bonferroni method.

2.9. Gene ontology analysis

Genomic distribution of CTCF peaks was performed using ChIP-seeker [58]. Gene Ontology (GO) enrichment analysis of CTCF peaks was performed by using Metascape [59] after obtaining their adjacent genes with ChIPseeker [58].

2.10. Motif analysis

Motif enrichment analysis was performed using MEME-ChIP [60]. Motif scanning on LTR22_SS consensus sequence (downloaded from the RepeatMasker website on Oct 26, 2018) was performed against CTCF motif (MA0139.1, JASPAR database) using FIMO (Grant et al., 2011) with settings: -text -thresh 1E-3.

2.11. Statistical analysis and data visualization

All statistical analyses were performed with R statistical programming language [61]. Heatmaps for ATAC-seq and ChIP-seq data were generated using deepTools v3.5.1 [49]. Hi-C data together with other 2D tracks were visualized with Juicebox [62].

3. Results

3.1. Comparison of the genome-wide binding patterns of CTCF across three pig tissues

CTCF is a key regulator for 3D chromatin organization in mammals [8]. We first compared its genome-wide binding patterns across different porcine tissues. For this purpose, we collected the ChIP-seq data of CTCF for spleen, adipose and hypothalamus from the FAANG project [41]. The matched RNA-seq, ATAC-seq and epigenomic (ChIP-seq of H3K4me3 and H3K27ac) data for pig spleen together with CTCF ChIP-seq data for human and mouse spleen were collected from the ENCODE and FAANG projects [63,41,64]. In this study, we also profiled the 3D chromatin organization of pig spleen with Hi-C. The sources of the integrated data were summarized in Table S1.

We identified 32,586, 22,821 and 27,994 CTCF peaks for adipose, hypothalamus and spleen, respectively. While many CTCF peaks are shared by all three tissues, those specific to one or two tissues are also observed (Fig. 1A,B). Specifically, 39.3% are shared by three tissues, 25.1% are shared by two tissues, and 35.7% are tissue-specific (i.e. 16.8%, 5.5% and 13.3% for adipose, hypothalamus and spleen, respectively). It confirmed the presence of shared and tissue-specific CTCF binding as reported for human previously [65]. As expected, the canonical CTCF motif is overrepresented in all groups of CTCF peaks (Fig. 1C). We further examined their genomic distribution and found that 30.2% of tissue-shared CTCF peaks occur in promoters, 28.8% in intronic regions, and 31.2% in intergenic regions (Fig. 1D). Relatively high proportions of tissue-specific CTCF peaks (particularly the adipose-specific group) occur within promoter regions (Fig. 1D). GO

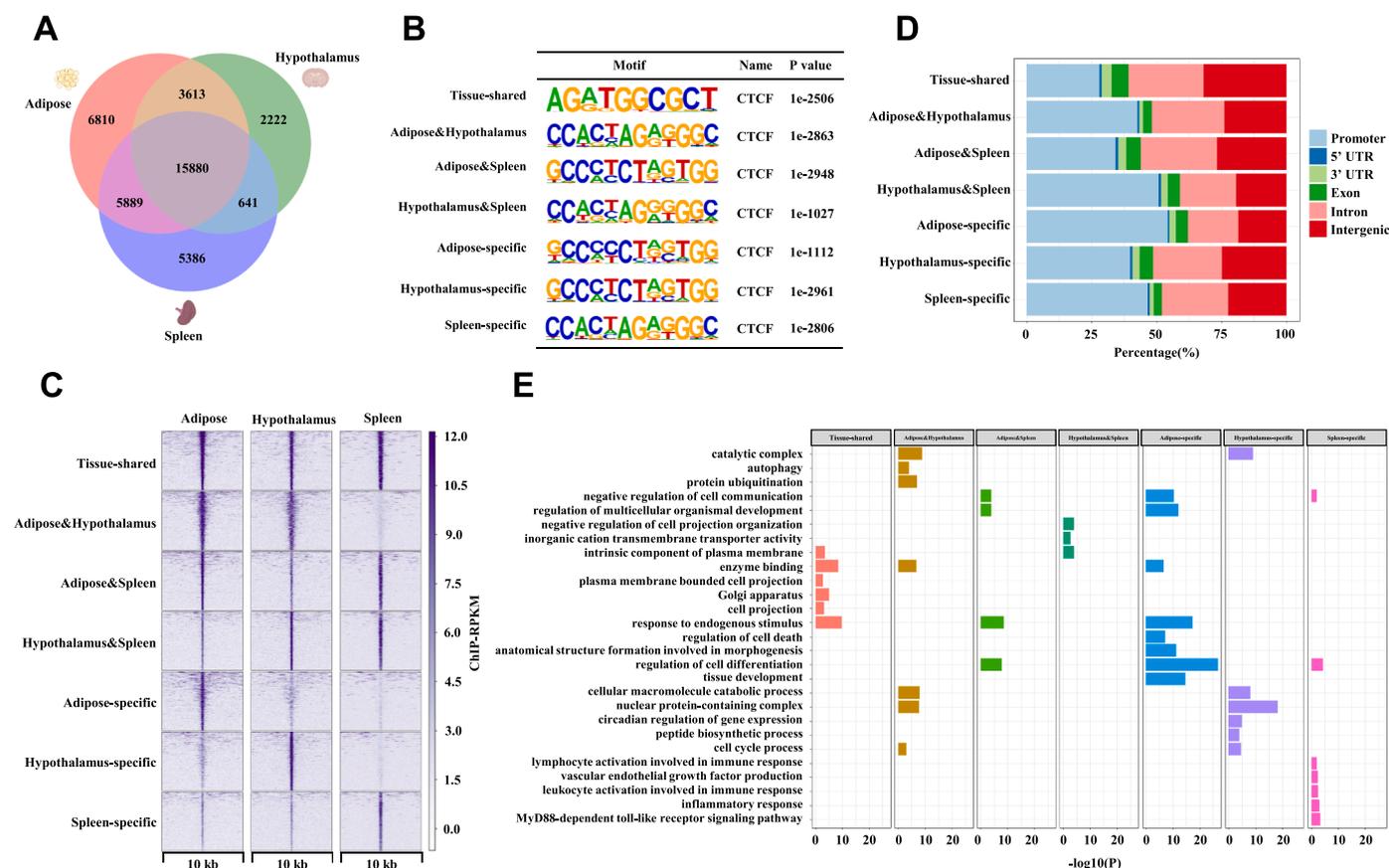


Fig. 1. Characterization of the genome-wide binding of CTCF across three pig tissues. A. Venn diagram shows the overlap of CTCF peaks of adipose, hypothalamus and spleen. B. Heatmap shows the binding pattern of CTCF in different groups of CTCF peaks in adipose, hypothalamus and spleen. C. Enrichment of the canonical CTCF motif in different groups of CTCF peaks. D. Genomic distribution of different groups of CTCF peaks. E. Representative GO terms enriched for different groups of CTCF peaks.

enrichment analysis confirmed the association of each group of CTCF peaks with the properties of corresponding tissues (Fig. 1E, Table S2). For example, spleen-specific CTCF peaks are associated with immune response, while adipose-specific peaks are associated with regulation of cell differentiation and development. Together, these results characterize the patterns and potential functional relevance of CTCF binding across different pig tissues.

3.2. Thousands of CTCF binding sites are derived from conserved or pig-specific TE families

Multiple studies revealed the contribution of TEs for the 3D chromatin organization in human and mouse [24], yet current knowledge on the function of pig TEs is lacking. Here, we identified dozens of TE families that are overrepresented in the CTCF peaks for the three pig tissues (Fig. 2A, Table S3). Notably, the degree of TE enrichment is significantly higher in spleen relative to the other two tissues (Fig. 2B), which is further confirmed by the comparison across different groups of shared or tissue-specific CTCF peaks (Fig. S1). Most of the significantly enriched TE families belong to LTRs/ERVs and DNA transposons (Fig. 2C, Table S3). The significantly enriched TE families include those that are known to create CTCF binding sites in human and mouse, such as MER20, MER91B, LTR41 and LTR50 [26,30,66,67]. Interestingly, three pig-specific ERV families including LTR22_SS, LTR15_SS and

LTR16_SS are also among the top enriched (Fig. 2D, Table S3). These results suggest the involvement of both conserved and lineage-specific TEs for mediating CTCF binding in pigs.

To compare the CTCF binding on TEs across species, we determined the significantly enriched TE families in the CTCF peaks for human and mouse spleen based on the overlaps of TE elements and CTCF peaks (Fig. S2, Table S3). The enriched TE families were compared across species (Fig. 2E, S2), and impressively, the degree of enrichment for multiple pig-specific ERV families (particularly LTR22_SS) is remarkably higher than the TE families previously reported to mediate 3D chromatin organization, such as MER20 which is a DNA transposon family shared by placental mammals [26,30,66]. Specifically, 9.9% (n = 221) of the 2228 LTR22_SS elements are bound by CTCF in pig spleen, which is much more frequent than MER20 elements with only 1.8% (252/13,982) being CTCF-bound. Of note, CTCF binding on some LTR22_SS elements show tissue-specificity, such as demonstrated by the representative loci flanking *OAZ2*, *CD48* and *IL17A*, respectively (Fig. S3). CTCF binding is centered on LTR22_SS elements, likely by recognizing the canonical CTCF motif which is significantly enriched in these elements (Fig. 2F). Moreover, the LTR22_SS consensus also harbors two canonical CTCF motifs (Fig. S4), indicating that LTR22_SS elements have inherited motifs to bind CTCF since inserted into pig genome. Together, these results indicate that distinct conserved and lineage-specific TEs may have the potential to create CTCF binding sites in pigs.

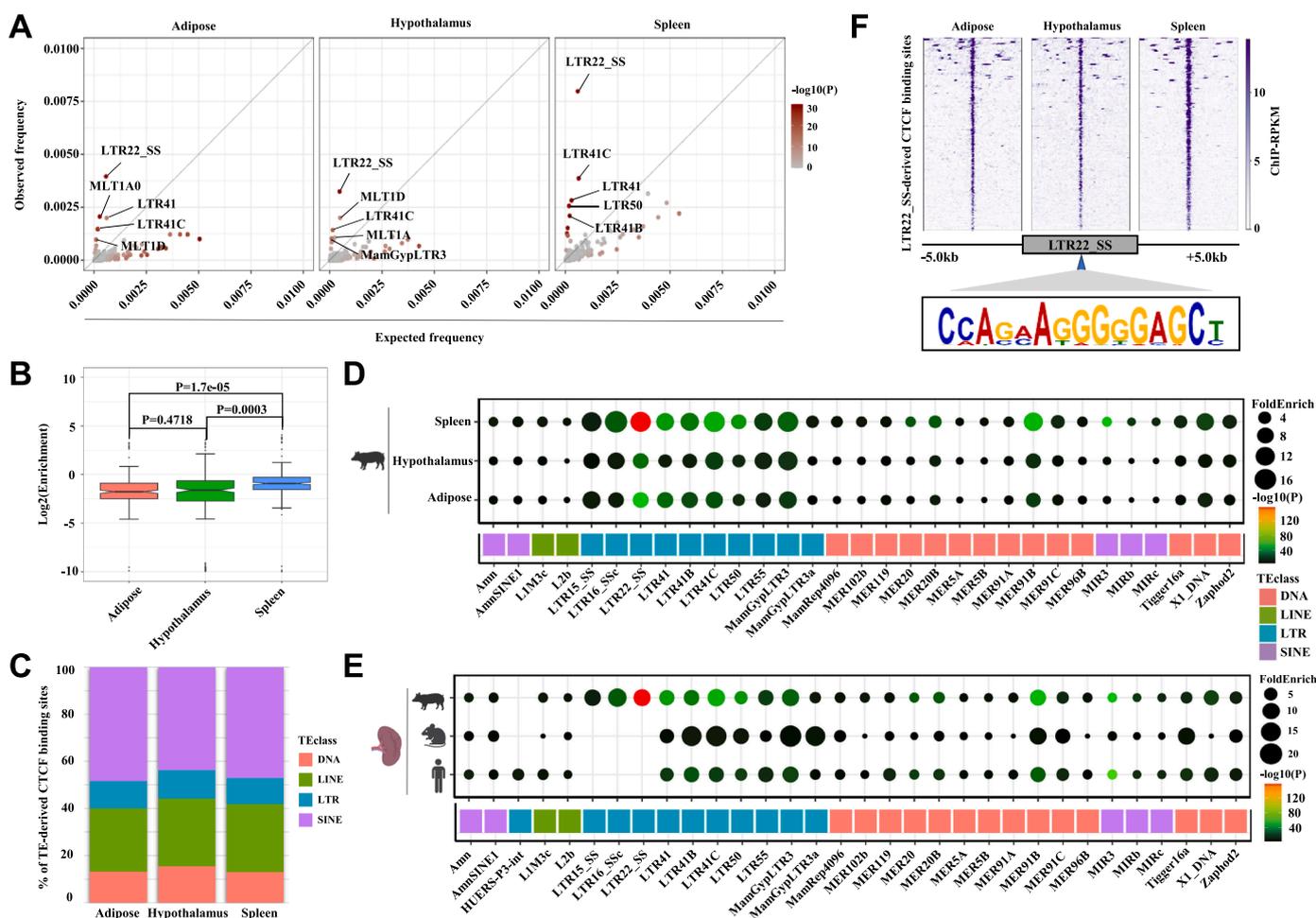


Fig. 2. Distinct conserved and lineage-specific TE families are overrepresented in the CTCF binding sites in pigs. A. Scatter plots show the enrichment of different TE families in the CTCF peaks in three pig tissues. The y-axis indicates the observed fraction of CTCF peaks that overlap each TE family, and the x-axis is the expected fraction of peaks that overlap each TE family. These values are calculated based on the results generated with the *fisher* function of BEDtools. B. Comparison of the TE enrichment in the CTCF peaks across three pig tissues. P-values calculated by Student’s t-test are indicated. C. Top enriched TE families in the CTCF peaks for human, mouse and pig tissues. D. Binding of CTCF on LTR22_SS elements based on ChIP-seq data. The presence of the CTCF motif in LTR22_SS consensus sequence is also indicated below.

3.3. TE-derived CTCF binding sites are highly accessible and enriched with active histone marks

To characterize the epigenetic status of TE-derived CTCF binding sites, we integrated the ATAC-seq and histone modification data - focusing on spleen since CTCF binding on TEs is more frequent in this tissue. Spleen is a tissue important for both innate and adaptive immunity [68]. We first compared the CTCF binding sites derived from major TE classes (*i.e.* LTR, LINE, SINE, DNA) or non-TEs. Over 80.0% of TE-derived CTCF binding sites overlap ATAC-seq peaks which is even higher than non-TE-derived ones (Fig. 3A), yet the degree of accessibility on TE-derived CTCF binding sites is relatively low (Fig. S5). The two active histone marks H3K27ac and H3K4me3 are also enriched on CTCF binding sites (Fig. S5), which agrees with previous observation in human and mouse [65,69]. Interestingly, the occupancy of these histone marks on TE-derived CTCF binding sites is lower than those not overlapping with TEs (Fig. 3A,B, S5). For example, 12.0% of LTR/ERV-derived CTCF binding sites are marked with H3K27ac, which is much lower than the 41.2% for non-TE-derived ones (Fig. 3A). These results suggest that TE-derived CTCF binding sites are highly accessible but rarely serve as active *cis*-elements like promoters and enhancers.

Given that multiple ERV families such as LTR22_SS are among the top enriched, we further compared the CTCF binding sites derived from different ERV families and observed substantial differences regarding CTCF binding intensity, chromatin accessibility and histone modifications (Fig. 3C). For example, relative to pig-specific ERV families like LTR22_SS, LTR15_SS and LTR16_SSc, CTCF binding sites derived from ancestral ERV families like LTR41, LTR41B/C, LTR50 and MamGypLTR3 are more accessible and have strong intensity of CTCF binding. Interestingly, those with strongest CTCF binding don't always have

active histone marks, given that H3K27ac is not enriched on MamGypLTR3-derived CTCF binding sites. We further demonstrate that for LTR22-derived CTCF binding sites, only 6.8% (*n* = 15) and 4.1% (*n* = 9) are marked with H3K27ac and H3K4me3, respectively (Fig. 3D). Interestingly, LTR22_SS-derived CTCF binding sites are highly associated with adaptive immune system and adaptive immune response (Fig. 3E), matching the function of spleen. Together, these results characterized the epigenetic patterns of TE-derived CTCF binding sites and indicate the potential involvement of ERV-derived CTCF binding sites for immune-related regulation in spleen.

3.4. Profiling of the 3D chromatin organization in pig spleen with Hi-C experiment

Hi-C is a powerful technique for studying 3D chromatin organization [4]. To profile the 3D chromatin organization of pig spleen, we conducted *in situ* Hi-C experiments with two biological replicates (Table S1). A total of 430 million valid paired contacts were obtained, which generated a chromatin interaction map at desirable resolution (Fig. 4A). The replicates are pooled for analysis after validating their high consistency (Fig. S6). Principal component analysis (PCA) identified active A compartments and inactive B compartments, which comprise 49.9% and 50.1% of the whole genome, respectively (Fig. 4B). As expected, A compartments are more accessible, enriched with the active histone marks H3K27ac and H3K4me3, and have higher transcription level (Fig. 4A). Notably, A compartments also show stronger enrichment of CTCF binding (Fig. 4A), probably due to their higher accessibility. Using the pooled data, we identified 1927 TADs and 3410 loops, respectively (Table S4, S6). Closer inspection of the HOXA locus revealed the typical TAD structures of this region, as well as a loop on this locus (Fig. 4C).

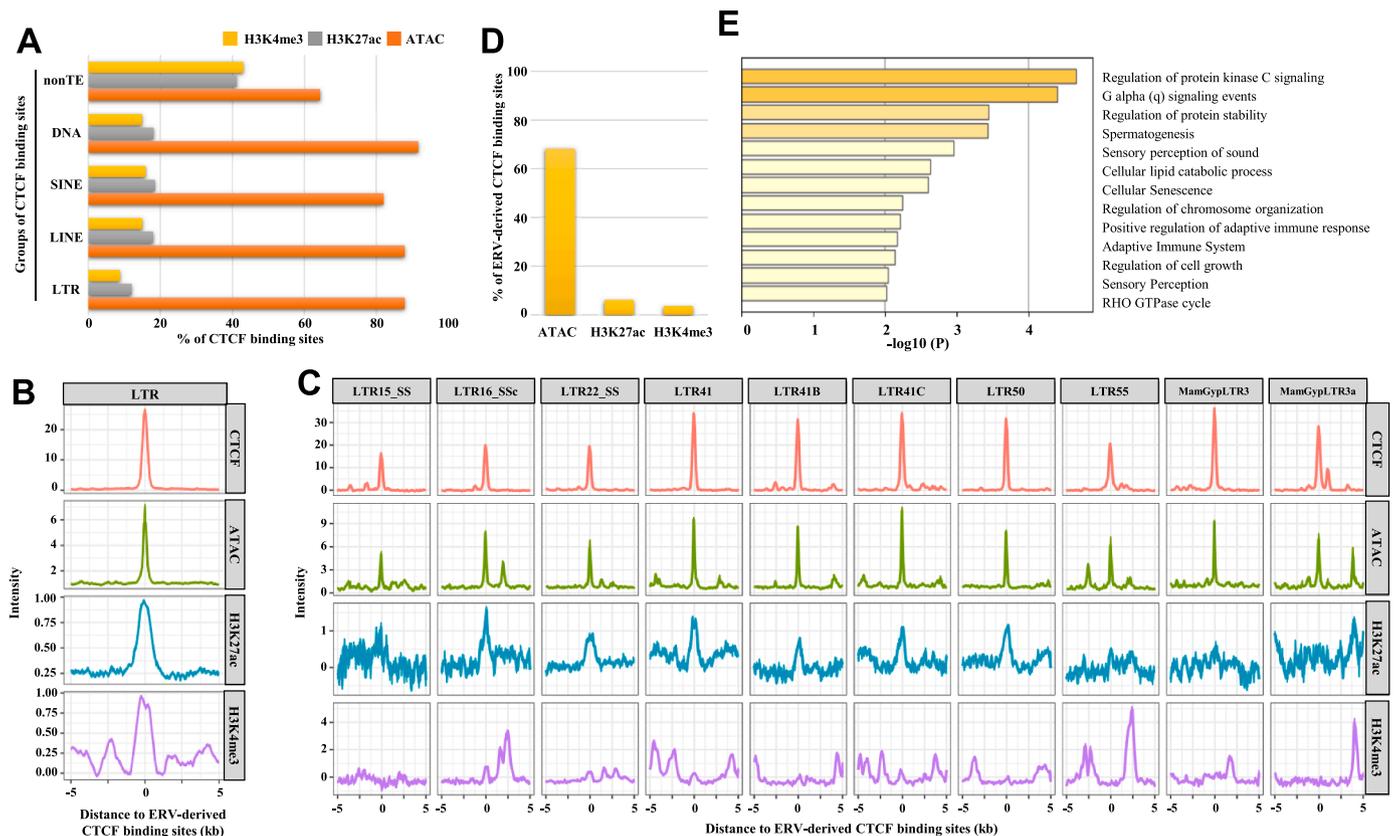


Fig. 3. Association of TE-derived CTCF binding sites with chromatin accessibility and histone marks. **A.** Bar plots show the frequency of ATAC-seq, K4me3 and K27ac occupancy on CTCF peaks derived from LTR/ERV, LINE, SINE, DNA or not. **B.** Intensity of CTCF, ATAC-seq, H3K4me3 and H3K27ac on the ERV-derived CTCF binding sites. **C.** Similar to B, but for CTCF binding sites derived from each significantly enriched ERV family. **D.** Overlap of LTR22_SS-derived CTCF binding sites to the peaks for ATAC-seq, H3K4me3 and H3K27ac, respectively. **E.** GO enrichment results for LTR22_SS-derived CTCF binding sites.

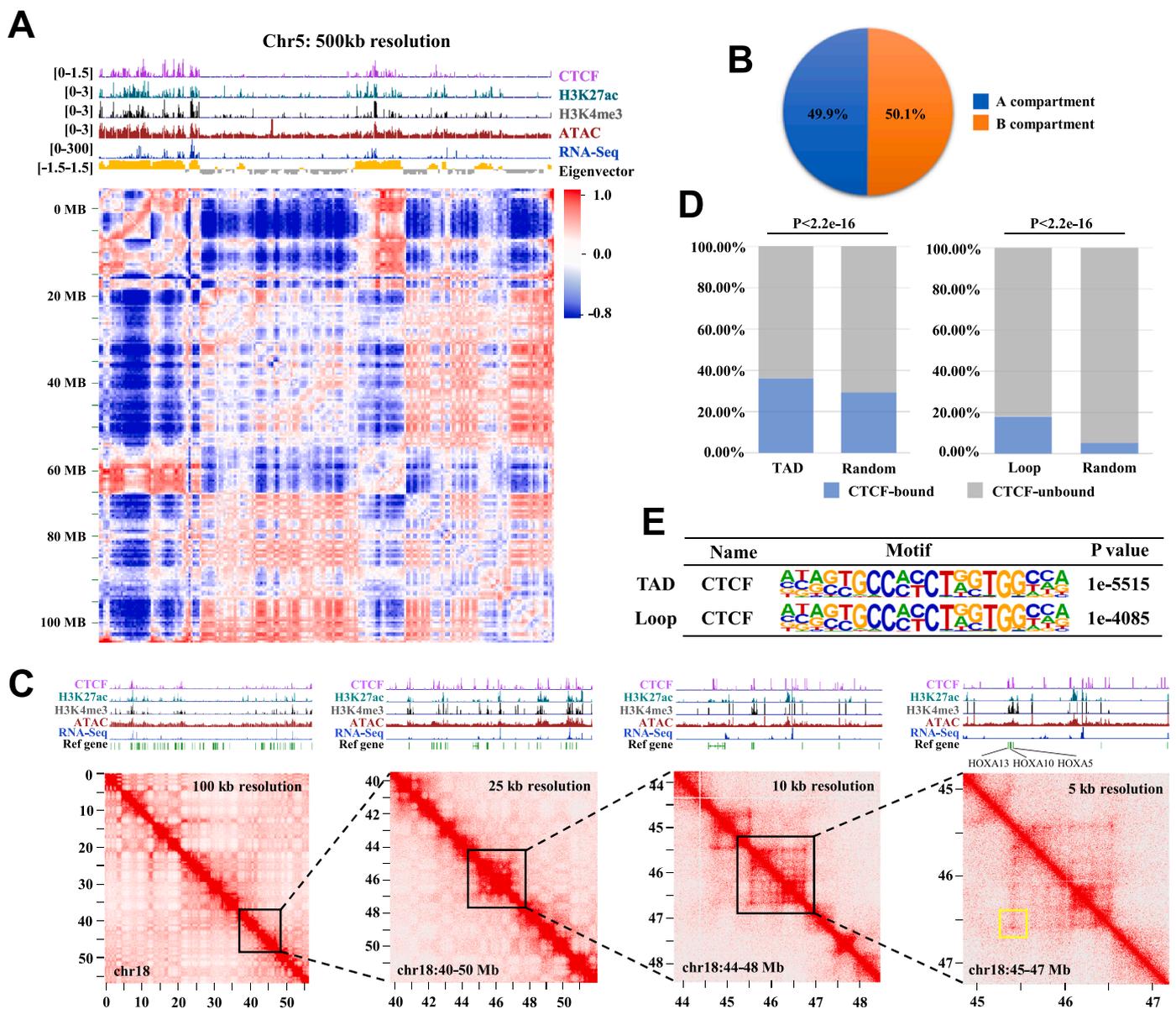


Fig. 4. Profiling of the 3D chromatin organization in pig spleen by *in situ* Hi-C. **A.** Correlation map at 500 kb resolution for chromosome 5 in pig spleen, with the tracks for ATAC-seq, ChIP-seq (CTCF, H3K27ac, H3K4me3) and RNA-seq shown alongside at top. **B.** Pie to show the proportions of compartment A and B, respectively. **C.** TAD structure surrounding the HOXA cluster, with resolutions of 100 kb, 25 kb, 10 kb and 5 kb, respectively. At 5 kb resolution, a representative loop is highlighted in yellow square. **D.** The percentages of TAD boundaries and loop anchors bound by CTCF. Randomly shuffled regions are used as control. P-values calculated by using Fisher's Exact Test are indicated. **E.** Enrichment of the canonical CTCF motif in TAD boundaries and loop anchors.

Both the TAD boundaries and loop anchors frequently overlap CTCF peaks and are enriched with the canonical CTCF motif (Fig. 4D,E), agreeing with the crucial role of CTCF for 3D chromatin organization. The generated *in situ* Hi-C data enable the profiling of the 3D chromatin structure in pig spleen for the first time.

3.5. TE-derived CTCF binding sites create hundreds of TAD boundaries and chromatin loops in pig spleen

Taking advantage of the newly generated Hi-C data, we further inspected the association of TE families with the TAD boundaries and chromatin loops in pig spleen. We identified 21 and 13 TE families that are enriched within the CTCF binding sites from TAD boundaries and chromatin loops, respectively (Fig. 5A,B, Table S6, S7). These TE families are highly consistent with those enriched in CTCF binding sites (Fig. 5C). Thirteen TE families, including the pig-specific LTR22_SS and

LTR16_SSc family ERVs, are simultaneously enriched in all CTCF binding sites and those within TAD boundaries and loop anchors (Fig. 5D). Notably, all these thirteen TE families belong to LTRs/ERVs, SINEs or DNA transposons, while none is from LINES (Fig. 5D, Table S7, S8). It confirms the previous report that LINES are under-represented at TAD boundaries, probably due to the length constrain [70].

To learn more about the insulation function of TEs, we further compared the insulation score for the CTCF binding sites derived from TEs or non-TEs. We demonstrate that the insulation score of the CTCF binding sites derived from each TE classes are comparable to those that don't overlap with TEs (Fig. 5E). Inspection of the CTCF binding sites derived from eight top TE families confirmed the insulation function for all of them (Fig. 5F). These results support the functional importance of TE-derived CTCF binding sites for chromatin insulation. As an example, we identified one LTR22_SS-derived CTCF binding site (is located 50 kb downstream of the gene) on the TAD boundary upstream of XCL1

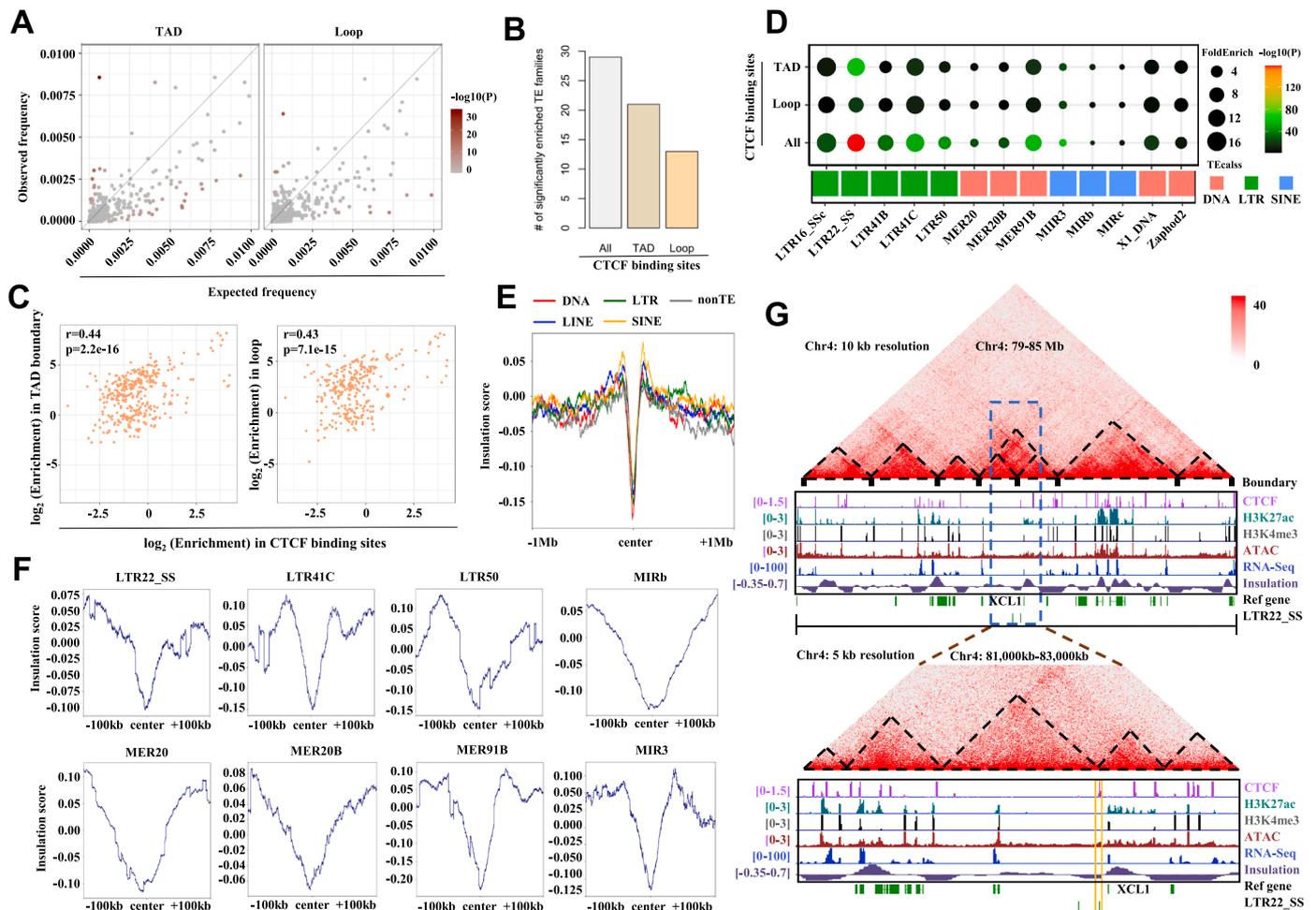


Fig. 5. Association of TE-derived CTCF binding sites to TAD boundaries and chromatin loops in pig spleen. **A.** Scatter plots to show the TE enrichment in the CTCF binding sites within TAD boundaries and chromatin loops, respectively. **B.** Bar plot shows the number of significantly enriched TE families in all CTCF binding sites or those within TAD boundaries or loop anchors. **C.** Comparison of the fold enrichment for different TE families in all CTCF binding sites or those within TAD boundaries and loop anchors, respectively. Pearson's r and p -values are indicated. **D.** Enrichment of the TE families that are enriched in all CTCF binding sites or those within TAD boundaries and loop anchors, respectively. **E.** Insulation score flanking the CTCF binding sites derived from major classes of TEs or not. **F.** Similar to **E**, but for CTCF binding sites derived from representative TE families. **G.** TAD structure together with ATAC-seq, ChIP-seq (CTCF, H3K27ac, H3K4me3), RNA-seq and insulation score tracks on the XCL1 locus. The LTR22_SS-derived CTCF binding sites on the TAD boundary upstream of XCL1 is highlighted in orange rectangle.

(Fig. 5G). XCL1, also known as lymphotactin, is a chemokine produced mainly by CD8⁺ T and natural killer cells and has important role in lymphocyte trafficking and inflammation [71]. Together, these results suggest the importance of TE-derived CTCF binding sites for mediating the 3D chromatin organization in pig spleen.

4. Discussion

The regulatory function of TEs in mammals has gained increasing attentions in recent years [23,24]. Like other mammals, the genome of pig is also abundant with TEs, with its 44.9% made up of TEs. However, most previous studies focusing on porcine ERVs for xenotransplantation [35], and studies on their regulatory function in pigs just began to emerge. For example, a few studies suggest that pig TEs may regulate immune gene expression either by acting as enhancers or through “viral mimicry” [72,73]. Apart from that, current understanding about the function of TEs in pigs remains obscure. In this study, we investigated the contribution of TEs, particularly lineage-specific ERV families, for the 3D chromatin organization in pigs. This study is initiated by the comparison of CTCF binding across three pig tissues, and then, we focused on spleen which is a tissue crucial for adaptive and innate immunity [68,74].

Through the comparison across adipose, hypothalamus and spleen,

we found that only about one third of CTCF peaks are shared by all three tissues. Given that only three tissues are compared, we expect that the proportion of common CTCF peaks should be even less if more tissues are included. Thus, tissue-specific CTCF binding is pervasive. This is consistent with previous studies which compared CTCF binding across mouse tissues [65,75]. Importantly, GO analysis demonstrates that tissue-specific CTCF binding events have functional relevance to the corresponding tissue, indicating that tissue-specific binding of CTCF may regulate the transcription of tissue-specific genes in pig. Mechanistically, the differential binding of CTCF is probably facilitated by other tissue-specific regulators, such as transcription factors crucial for each tissue. However, the exact factors that drive the tissue-specific binding of CTCF in these pig tissues remain unclear.

Regarding the links of TEs to the 3D chromatin organization in pigs, we achieved two major findings. First, we identified dozens of TE families that are significantly enriched within CTCF binding sites. While previous studies reported the links between TEs and CTCF binding in species like human and mouse [24], we uncovered three pig-specific ERV families, including LTR22_SS, LTR16_Ssc and LTR15_SS, that create many CTCF binding sites. Second, we found that while TEs are enriched in the CTCF binding sites for all tissues, the degree of enrichment is remarkably higher in spleen. It is not surprising given that the immune system evolves relatively fast due to the host-pathogen conflict

[76]. Indeed, TEs have been well-recognized to facilitate both innate and adaptive immunity evolution by creating lineage-specific cis-regulatory elements [77,20,22,78]. We speculate that some immune-related transcription factors may facilitate spleen-specific binding of CTCF on TEs. Support this speculation, LTR22_SS consensus sequence harbors the motifs for multiple immune-related TFs (e.g. STAT1/2, IRF7/8, NFYA/B/C).

After revealing the links between TEs and CTCF binding, we further performed Hi-C experiments to characterize the contribution of TEs to 3D chromatin structure in pig spleen. Typical chromatin structures, including A/B compartments, TADs and loops, were identified in pig spleen. We also confirmed the enrichment of active histone marks in A compartments as well as the enrichment of CTCF binding in TAD boundaries and loop anchors. Regarding TEs, we confirmed that they are overrepresented in both TAD boundaries and loop anchors. Importantly, the CTCF binding sites derived from TEs, including the pig-specific LTR22_SS ERV family, show comparable insulation score to those don't overlap TEs. Interestingly, one LTR22_SS-derived CTCF binding site demarcates a TAD border adjacent to XCL1, which is an immune gene highly expressed in spleen. These data suggest the importance of TEs for mediating 3D chromatin organization in pig spleen. We expect that Hi-C experiment with higher resolution and CRISPR/Cas9 engineering of TE-derived CTCF binding sites would further improve the understanding about the regulation of 3D chromatin organization and immune gene expression in pigs.

5. Conclusions

In summary, this study represents the first step toward understanding the function of TEs on 3D chromatin organization regulation in pigs. We identified dozens of TE families, including pig-specific ERV families, that are likely to mediate CTCF binding in pigs. Importantly, our data highlights the importance of pig-specific ERVs for mediating CTCF binding and TAD formation in pig spleen, which may further influence immune gene expression. Overall, this study improves our understanding about the regulatory function of porcine TEs and expands current knowledge about the functional importance of TEs in mammals.

Ethics approval and consent to participate

The research was conducted according to the guidelines of the Institutional Animal Care and Utilization Committee (IACUC) of the Animal Experimental Ethics Committee of Yangzhou University (permit number: SYXK (SU) IACUC 2012–0029).

Author contributions

MS conceived the project. YL and YW performed bioinformatic analyses. HF, WQ and SC performed experiments. YL, HF, WB and MS wrote the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the grants from the National Natural Science Foundation of China (32270584 & 31900422 to MS), the China Postdoctoral Foundation (2023M732990 to SC), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (23KJB180028 to SC), the 111 Project D18007, and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

CRediT authorship contribution statement

MS conceived the project. YL, YW and LZ performed bioinformatic analyses. HF, WQ and SC performed experiments. YL, HF, WB and MS

wrote the manuscript. All authors read and approved the final manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study utilized the computational resources of Yangzhou University College of Veterinary Medicine High-Performance Computing cluster.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.csbj.2023.09.029](https://doi.org/10.1016/j.csbj.2023.09.029).

References

- [1] Bonev B, Cavalli G. Organization and function of the 3D genome. *Nat Rev Genet* 2016;17:661–78.
- [2] Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* 2019;20:437–55.
- [3] Szabo Q, Bantignies F, Cavalli G. Principles of genome folding into topologically associating domains. *Sci Adv* 2019;5:eaaw1668.
- [4] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Rogoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009;326:289–93.
- [5] Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;159:1665–80.
- [6] Krijger PH, de Laat W. Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol* 2016;17:771–82.
- [7] Zheng H, Xie W. The role of 3D genome organization in development and cell differentiation. *Nat Rev Mol Cell Biol* 2019;20:535–50.
- [8] Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* 2014;15:234–46.
- [9] Davidson IF, Barth R, Zaczek M, van der Torre J, Tang W, Nagasaka K, Janissen R, Kersemakers J, Wutz G, Dekker C, et al. CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature* 2023.
- [10] Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Włodarczyk J, Rusczycki B, et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* 2015;163:1611–27.
- [11] Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 2007;128:1231–45.
- [12] Qi Q, Cheng L, Tang X, He Y, Li Y, Yee T, Shrestha D, Feng R, Xu P, Zhou X, et al. Dynamic CTCF binding directly mediates interactions among cis-regulatory elements essential for hematopoiesis. *Blood* 2021;137:1327–39.
- [13] Ottaviani D, Lever E, Mao S, Christova R, Ogunkolade BW, Jones TA, Szary J, Aarum J, Mumin MA, Pieri CA, et al. CTCF binds to sites in the major histocompatibility complex that are rapidly reconfigured in response to interferon-gamma. *Nucleic Acids Res* 2012;40:5262–70.
- [14] Ortabozkoyun H, Huang PY, Cho H, Narendra V, LeRoy G, Gonzalez-Buendia E, Skok JA, Tsigirgos A, Mazzoni EO, Reinberg D. CRISPR and biochemical screens identify MAZ as a cofactor in CTCF-mediated insulation at Hox clusters. *Nat Genet* 2022;54:202–12.
- [15] Stik G, Vidal E, Barrero M, Cuartero S, Vila-Casadesus M, Mendieta-Esteban J, Tian TV, Choi J, Berenguer C, Abad A, et al. CTCF is dispensable for immune cell transdifferentiation but facilitates an acute inflammatory response. *Nat Genet* 2020;52:655–61.
- [16] Allen EK, Randolph AG, Bhargale T, Dogra P, Ohlson M, Oshansky CM, Zamora AE, Shannon JP, Finkelstein D, Dressen A, et al. SNP-mediated disruption of CTCF binding at the IFITM3 promoter is associated with risk of severe influenza in humans. *Nat Med* 2017;23:975–83.
- [17] Merckenschlager M, Odom DT. CTCF and cohesin: linking gene regulatory elements with their targets. *Cell* 2013;152:1285–97.
- [18] Vietri Rudan M, Hadjur S. Genetic tailors: CTCF and cohesin shape the genome during evolution. *Trends Genet* 2015;31:651–60.
- [19] Wells JN, Feschotte C. A field guide to eukaryotic transposable elements. *Annu Rev Genet* 2020;54:539–61.
- [20] Du C, Jiang J, Li Y, Yu M, Jin J, Chen S, Fan H, Macfarlan TS, Cao B, Sun MA. Regulation of endogenous retrovirus-derived regulatory elements by GATA2/3 and MSX2 in human trophoblast stem cells. *Genome Res* 2023;33:197–207.

- [21] Frost JM, Amante SM, Okae H, Jones EM, Ashley B, Lewis RM, Cleal JK, Caley MP, Arima T, Maffucci T, et al. Regulation of human trophoblast gene expression by endogenous retroviruses. *Nat Struct Mol Biol* 2023.
- [22] Kelly CJ, Chitko-McKown CG, Chuong EB. Ruminant-specific retrotransposons shape regulatory evolution of bovine immunity. *Genome Res* 2022.
- [23] Senft AD, Macfarlan TS. Transposable elements shape the evolution of mammalian development. *Nat Rev Genet* 2021;22:691–711.
- [24] Lawson HA, Liang Y, Wang T. Transposable elements in mammalian chromatin organization. *Nat Rev Genet* 2023.
- [25] Azazi D, Mudge JM, Odom DT, Flicek P. Functional signatures of evolutionarily young CTCF binding sites. *BMC Biol* 2020;18:132.
- [26] Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* 2008;18:1752–62.
- [27] Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* 2010;42:631–4.
- [28] Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res* 2014;24:1963–76.
- [29] Schmidt D, Schwale PC, Wilson MD, Ballester B, Goncalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 2012;148:335–48.
- [30] Choudhary MN, Friedman RZ, Wang JT, Jang HS, Zhuo X, Wang T. Co-opted transposons help perpetuate conserved higher-order chromosomal structures. *Genome Biol* 2020;21:16.
- [31] Choudhary MN, Quaid K, Xing X, Schmidt H, Wang T. Widespread contribution of transposable elements to the rewiring of mammalian 3D genomes. *Nat Commun* 2023;14:634.
- [32] Diehl AG, Ouyang N, Boyle AP. Transposable elements contribute to cell and species-specific chromatin looping and gene regulation in mammalian genomes. *Nat Commun* 2020;11:1796.
- [33] Lunney JK, Van Goor A, Walker KE, Hailstock T, Franklin J, Dai C. Importance of the pig as a human biomedical model. *Sci Transl Med* 2021;13:eabd5758.
- [34] Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens HJ, et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 2012;491:393–8.
- [35] Denner J. The origin of porcine endogenous retroviruses (PERVs). *Arch Virol* 2021;166:1007–13.
- [36] Yue Y, Xu W, Kan Y, Zhao HY, Zhou Y, Song X, Wu J, Xiong J, Goswami D, Yang M, et al. Extensive germline genome engineering in pigs. *Nat Biomed Eng* 2021;5:134–43.
- [37] Chuong EB, Rumi MA, Soares MJ, Baker JC. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* 2013;45:325–9.
- [38] Miao B, Fu S, Lyu C, Gontarz P, Wang T, Zhang B. Tissue-specific usage of transposable element-derived promoters in mouse development. *Genome Biol* 2020;21:255.
- [39] Sun MA, Wolf G, Wang Y, Senft AD, Ralls S, Jin J, Dunn-Fletcher CE, Muglia LJ, Macfarlan TS. Endogenous retroviruses drive lineage-specific regulatory evolution across primate and rodent placentae. *Mol Biol Evol* 2021;38:4992–5004.
- [40] Jin L, Tang Q, Hu S, Chen Z, Zhou X, Zeng B, Wang Y, He M, Li Y, Gui L, et al. A pig BodyMap transcriptome reveals diverse tissue physiologies and evolutionary dynamics of transcription. *Nat Commun* 2021;12:3715.
- [41] Kern C, Wang Y, Xu X, Pan Z, Halstead M, Chanthavixay G, Saelao P, Waters S, Xiang R, Chamberlain A, et al. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nat Commun* 2021;12:1821.
- [42] Li D, He M, Tang Q, Tian S, Zhang J, Li Y, Wang D, Jin L, Ning C, Zhu W, et al. Comparative 3D genome architecture in vertebrates. *BMC Biol* 2022;20:99.
- [43] Pan Z, Yao Y, Yin H, Cai Z, Wang Y, Bai L, Kern C, Halstead M, Chanthavixay G, Trakooljul N, et al. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nat Commun* 2021;12:5848.
- [44] Zhao Y, Hou Y, Xu Y, Luan Y, Zhou H, Qi X, Hu M, Wang D, Wang Z, Fu Y, et al. A compendium and comparative epigenomics analysis of cis-regulatory elements in the pig genome. *Nat Commun* 2021;12:2217.
- [45] Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, et al. Ensembl 2020. *Nucleic Acids Res* 2020;48:D682–8.
- [46] Lee CM, Barber GP, Casper J, Clawson H, Diekhans M, Gonzalez JN, Hinrichs AS, Lee BT, Nassar LR, Powell CC, et al. UCSC genome browser enters 20th year. *Nucleic Acids Res* 2020;48:D756–d761.
- [47] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.
- [48] Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9:R137.
- [49] Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014;42:W187–91.
- [50] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
- [51] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Genome Project Data Processing, S.. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
- [52] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- [53] Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* 2015;16:259.
- [54] Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576–89.
- [55] Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* 2016;3:95–8.
- [56] Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, Dekker J, Mirny LA. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* 2012;9:999–1003.
- [57] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2.
- [58] Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 2015;31:2382–3.
- [59] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;10:1523.
- [60] Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 2011;27:1696–7.
- [61] Team, R.C. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- [62] Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst* 2016;3:99–101.
- [63] Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
- [64] Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 2014;515:355–64.
- [65] Chen H, Tian Y, Shu W, Bo X, Wang S. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS One* 2012;7:e41374.
- [66] Lynch VJ, Leclerc RD, May G, Wagner GP. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* 2011;43:1154–9.
- [67] Schwale PC, Ward MC, Cain CE, Faure AJ, Gilad Y, Odom DT, Flicek P. Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes. *Genome Biol* 2013;14:R148.
- [68] Bronte V, Pittet MJ. The spleen in local and systemic regulation of immunity. *Immunity* 2013;39:806–18.
- [69] Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 2011;480:490–5.
- [70] Kentepozidou E, Aitken SJ, Feig C, Stefflova K, Ibarra-Soria X, Odom DT, Roller M, Flicek P. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol* 2020;21:5.
- [71] Lei Y, Takahama Y. XCL1 and XCR1 in the immune system. *Microbes Infect* 2012;14:262–7.
- [72] Wang X, Chen Z, Murani E, D'Alessandro E, An Y, Chen C, Li K, Galeano G, Wimmers K, Song C. A 192 bp ERV fragment insertion in the first intron of porcine TLR6 may act as an enhancer associated with the increased expressions of TLR6 and TLR1. *Mob DNA* 2021;12:20.
- [73] Zhang L, Jin J, Qin W, Jiang J, Bao W, Sun MA. Inhibition of EZH2 causes retrotransposon derepression and immune activation in porcine lung alveolar macrophages. *Int J Mol Sci* 2023;24.
- [74] Lewis SM, Williams A, Eisenbarth SC. Structure and function of the immune system in the spleen. *Sci Immunol* 2019;4.
- [75] Kubo N, Ishii H, Xiong X, Bianco S, Meitinger F, Hu R, Hocker JD, Conte M, Gorkin D, Yu M, et al. Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat Struct Mol Biol* 2021;28:152–61.
- [76] Litman GW, Cooper MD. Why study the evolution of immunity? *Nat Immunol* 2007;8:547–8.
- [77] Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 2016;351:1083–7.
- [78] Ye M, Goudot C, Hoyler T, Lemoine B, Amigorena S, Zueva E. Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers. *Proc Natl Acad Sci USA* 2020;117:7905–16.