



## OPEN ACCESS

## EDITED BY

Shi-Yi Chen,  
Sichuan Agricultural University, China

## REVIEWED BY

Izabela Makatowska,  
Adam Mickiewicz University, Poland  
Gonzalo Riadi,  
University of Talca, Chile

## \*CORRESPONDENCE

Shao-Hui Ye,  
✉ ysh@ynau.edu.cn

RECEIVED 12 December 2024

ACCEPTED 03 March 2025

PUBLISHED 19 March 2025

## CITATION

Li M-D, Li H-R and Ye S-H (2025) Multi-tissue transcriptomic characterization of endogenous retrovirus-derived transcripts in *Capra hircus*. *Front. Genet.* 16:1544330. doi: 10.3389/fgene.2025.1544330

## COPYRIGHT

© 2025 Li, Li and Ye. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Multi-tissue transcriptomic characterization of endogenous retrovirus-derived transcripts in *Capra hircus*

Ming-Di Li<sup>1,2</sup>, Hu-Rong Li<sup>1</sup> and Shao-Hui Ye<sup>1\*</sup>

<sup>1</sup>Department of Animal Breeding and Reproduction, College of Animal Science and Technology, Yunnan Agricultural University, Kunming, China, <sup>2</sup>Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

**Background:** Transposable elements (TEs, or transposons) are repetitive genomic sequences, accounting for half of a mammal genome. Most TEs are transcriptionally silenced, whereas some TEs, especially endogenous retroviruses (ERVs, long terminal repeat retrotransposons), are physiologically expressed in certain conditions. However, the expression pattern of TEs in those less studied species, like goat (*Capra hircus*), remains unclear. To obtain an overview of the genomic and transcriptomic features of TEs and ERVs in goat, an important farm species, we herein analyzed transcriptomes of ten *C. hircus* tissues and cells under various physiological and pathological conditions.

**Method:** Distribution of classes, families, and subfamilies of TEs in the *C. hircus* genome were systematically annotated. The expression patterns of TE-derived transcripts in multiple tissues were investigated at subfamily and location levels. Differential expression of ERV-derived reads was measured under various physiological and pathological conditions, such as embryo development and virus infection challenges. Co-expression between ERV-reads and their proximal genes was also explored to decipher the expression regulation of ERV-derived transcripts.

**Results:** There are around 800 TE subfamilies in the goat genome, accounting for 49.1% of the goat genome sequence. TE-derived reads account for 10% of the transcriptome and their abundance are comparable in various goat tissues, while expression of ERVs are variable among tissues. We further characterized expression pattern of ERV reads in various tissues. Differential expression analysis showed that ERVs are highly active in 16-cell embryos, when the genome of the zygote begins to transcribe its own genes. We also recognized numerous activated ERV reads in response to RNA virus infection in lung, spleen, caecum, and immune cells. CapAeg\_1.233:ERVK in chromosome 1 and 17 are dysregulated under endometrium development and infection conditions. They showed strong co-expression with their proximal gene *OAS1* and *TMPRSS2*, indicating the impact of activated proximal gene expression on nearby ERVs.

**Conclusion:** We generated ERV transcriptomes across goat tissues, and identified ERVs activated in response to different physiological and pathological conditions.

## KEYWORDS

transposable element, endogenous retrovirus, *Capra hircus*, goat, transcriptome

## 1 Introduction

Transposable elements (TEs, also called transposons) are mobile genetic elements consist of repetitive sequences, accounting for about half of a mammal genome (Dong et al., 2013; de Koning et al., 2011). According to the origin and mobile type, TEs could be classified into several classes and families, including DNA transposons, long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and the long terminal repeat (LTR) family, which mainly contains endogenous retroviruses (ERVs) (Lanciano and Cristofari, 2020). TEs play significant roles in shaping the size, structure and function of mammal genome. It is well recognized that the host could leverage TEs to facilitate specific biological processes. For instance, the ERV-derived envelope protein, syncytia, induces the fusion of placental trophoblast cells (Chuong, 2018), driving the evolution of placental mammals (Mi et al., 2000). Additionally, a large proportion of TEs, though being non-coding sequences, act as enhancers in host genome to regulate the expression of coding genes in various processes.

At the transcriptome level, most TEs, especially ERVs which have the potential to code for proteins, are silenced and located in heterochromatin region in the genome. When not properly silenced, they may be activated (transcribed) under certain conditions. Such abnormal TE expression may contribute to the pathogenesis of various diseases (Horvath et al., 2017). Although TEs are non-negligible in genomic analysis, they are typically ignored in transcriptomic analysis. Since TE-derived reads are highly repetitive, rendering complexities and uncertainty in attributing ambiguously aligned short reads to the exact elements, TE-associated reads are often discarded in RNA sequencing data analyses.

Because of the above technical challenges, TE-derived reads are somewhat overlooked in typical genomic and transcriptomic analyses, especially in less studied species. In the genome of *Capra hircus* (goat), the distribution and characterization of TEs are not well annotated. Whether, when, and how those TEs are expressed in goat tissues are also unclear. To gain an overview of the genomic and transcriptomic features of TEs in *C. hircus*, we herein analyzed TE- and ERV- derived transcripts at both subfamily and location levels, in a dozen bulk RNA-seq datasets of ten *C. hircus* tissues and cells under various physiological and pathological conditions. Since TEs, especially ERVs, are physiologically expressed in embryos and placenta, we initially analyzed whether any ERVs were differentially expressed during embryo development, as well as in endometrium, where the expression regulation might be regulated cooperatively. We then checked infection related datasets to investigate whether ERVs were dysregulated in response to infection, as external stress might be a source for endogenous TE activation. We generated detailed annotation files for genomic and transcriptomic analyses for goat genome, and assessed the genome-wide expression patterns of ERVs across goat tissues in various conditions, providing a reference ERV atlas for TE research in goats.

## 2 Methods

### 2.1 Study design

We initially obtained the Ensembl curated *C. hircus* genome assembly ARS1 (GCA\_001704415.1), which was well annotated for

genomic features and was widely-used, and then annotated the TEs using RepeatMasker (<http://www.repeatmasker.org>). The annotated GTF (Gene Transfer Format) file was then used for subsequent TE identification in the transcriptional analyses. RNA sequencing of goat tissues was comprehensively explored in the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/browse/>), resulting in 12 datasets from 10 *C. hircus* tissues and cells under various physiological and pathological conditions. Raw sequencing reads were processed using Tetranscripts (Jin et al., 2015) and TElocal (<https://github.com/mhammell-laboratory/TElocal>). Genomic and transcriptomic features, including TE types, genomic distributions, and expression patterns, were investigated across goat tissues (Figure 1A).

### 2.2 Data sources

The RNA-seq datasets utilized in this study were sourced from the GEO repository (accession numbers were listed in Table 1). In brief, GSE69812 includes 6 fetal skin samples from normal and hyperpigmented goats (Ren et al., 2016), while GSE164100 (Bhat et al., 2021) includes 19 skin biopsies containing secondary hair follicles from ten 24-month-old male Pashmina goats, repeatedly sampled at resting phase (telogen) and active growth phase (anagen). GSE93855 analyzed kidney, liver, and spleen of three unrelated adult female goats at high- and low-altitude (Tang et al., 2017). Since ERVs are physiologically expressed in the placenta, we analyzed expression pattern of ERVs in embryos (GSE129742) (Li et al., 2020), as well as in ovary (GSE120144) (Liu et al., 2018) and endometrium (GSE108557 and GSE184110) (Liu et al., 2021), where the gene expression might be regulated cooperatively. In addition to these tissues under physiological conditions, we also investigated infection related datasets. GSE130552 includes 12 samples from lung, spleen, and caecum of control and Peste-des-petits-ruminants virus (PPRV) infected goats at 9 days-post-infection, whereas GSE132429 analyzed monocytes and lymphocytes from PPRV infected goats (Wani et al., 2019). GSE121725 includes expression profiling of skin fibroblast cells in response to ORF virus infection, and GSE30379 analyzed mammary epithelial cells in response to *Mycoplasma agalactiae* challenge, and GSE117799 analyzed peripheral blood mononuclear cells from goats infected with *Mycobacterium avium* subsp. *Paratuberculosis* (Berry et al., 2018). To further investigate the distribution and regulation of TEs in goat, assays for transposase-accessible chromatin using sequencing (ATAC-seq) in goat tissues and cells were also screened in literature and databases. One reference ATAC-seq data from goat liver and CD4<sup>+</sup> and CD8<sup>+</sup> T cells, generated by the Functional Annotation of Animal Genomes (FAANG) project was obtained. Processed summary data, with quantified ATAC-seq peaks and chromosomal coordinates of the open regions, were downloaded from the original reference (Foissac et al., 2019), to detect TEs distributions in the context of chromatin landscape in goat tissues and cells.

### 2.3 Genomic annotation of TEs

Genomic annotation of TEs was conducted according to literature (Lanciano and Cristofari, 2020; Tarailo-Graovac and

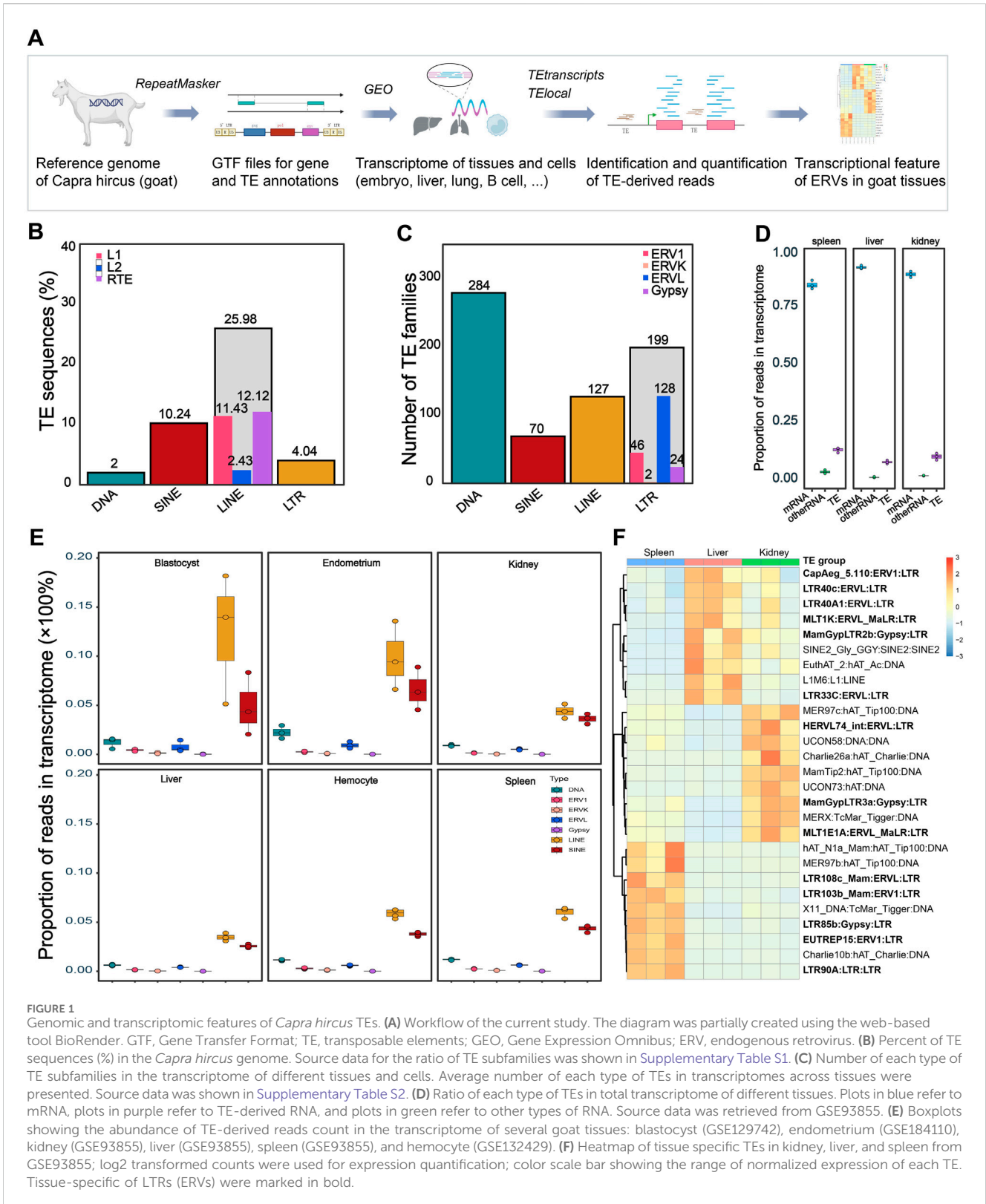


FIGURE 1

Genomic and transcriptomic features of *Capra hircus* TEs. (A) Workflow of the current study. The diagram was partially created using the web-based tool BioRender. GTF, Gene Transfer Format; TE, transposable elements; GEO, Gene Expression Omnibus; ERV, endogenous retrovirus. (B) Percent of TE sequences (%) in the *Capra hircus* genome. Source data for the ratio of TE subfamilies was shown in Supplementary Table S1. (C) Number of each type of TE subfamilies in the transcriptome of different tissues and cells. Average number of each type of TEs in transcriptomes across tissues were presented. Source data was shown in Supplementary Table S2. (D) Ratio of each type of TEs in total transcriptome of different tissues. Plots in blue refer to mRNA, plots in purple refer to TE-derived RNA, and plots in green refer to other types of RNA. Source data was retrieved from GSE93855. (E) Boxplots showing the abundance of TE-derived reads count in the transcriptome of several goat tissues: blastocyst (GSE129742), endometrium (GSE184110), kidney (GSE93855), liver (GSE93855), spleen (GSE93855), and hemocyte (GSE132429). (F) Heatmap of tissue specific TEs in kidney, liver, and spleen from GSE93855; log<sub>2</sub> transformed counts were used for expression quantification; color scale bar showing the range of normalized expression of each TE. Tissue-specific of LTRs (ERVs) were marked in bold.

Chen, 2009; Dong et al., 2015; Bourque et al., 2018). The Ensembl curated assembly ARS1 was used for genome mapping and annotation. The reference sequence (FASTA file) was accessed at [https://ftp.ensembl.org/pub/release-112/fasta/capra\\_hircus/dna/Capra\\_hircus.ARS1.dna.toplevel.fa.gz](https://ftp.ensembl.org/pub/release-112/fasta/capra_hircus/dna/Capra_hircus.ARS1.dna.toplevel.fa.gz), and the genomic

feature annotation GTF file was accessed at [https://ftp.ensembl.org/pub/release-112/gtf/capra\\_hircus/Capra\\_hircus.ARS1.112.gtf.gz](https://ftp.ensembl.org/pub/release-112/gtf/capra_hircus/Capra_hircus.ARS1.112.gtf.gz). The TEs of the *C. hircus* genome were annotated using the RepeatMasker (version 4.1.6) software (Tarailo-Graovac and Chen, 2009), with default blast mode *rmblastn* (version 2.14.0+). *De novo*

TABLE 1 Characteristics of analyzed RNA-seq datasets from various goat tissues.

Condition	Tissue/cell	Sample description	Sample size	Mapping rate	Data source	SRA accession
Physiological conditions	Skin	Fetal skin sampled from the normal and hyperpigmented goats	6	75.70%	GSE69812	SRP059421
		Skin biopsies containing secondary hair follicles from 24-months old male Pashmina goats, repeatedly sampled at resting phase (telogen) and active growth phase (anagen)	19	77.7%	GSE164100	SRP299900
	Liver, spleen, kidney	kidney, liver, and spleen of three unrelated adult female goats at high- and low-altitude	12	89.03%	GSE93855	SRP055702
	Oocyte and embryo	Matured oocytes and <i>in vivo</i> developed embryos from the 2-cell to the blastocyst stages	21	76.80%	GSE129742	SRP192394
	Ovary	Ovary at follicular and luteal phases	10	85.70%	GSE120144	SRP163266
	Endometrial epithelium	Peri-implantation endometrial epithelium at 6-day,16-day, and19-day of pregnancy	9	75.70%	GSE108557	SRP127612
	Endometrial epithelial cell	Endometrial epithelial cells with or without IFNT treatment	6	85.70%	GSE184110	SRP337048
Pathological conditions	Lung, spleen and caecum	Lung, spleen, and caecum of control and Peste-des-petits-ruminants virus (PPRV) infected goats at 9 days post infection	12	77.10%	GSE130552	SRP194369
	Monocytes and lymphocytes	Monocytes and lymphocytes from Peste-des-petits-ruminants virus (PPRV) infected goats	16	75.20%	GSE132429	SRP200956
	Skin fibroblast cells	Skin fibroblast cells in response to Orf virus infection	6	79.50%	GSE121725	SRP166752
	Mammary epithelial cells	Mammary epithelial cells to <i>Mycoplasma agalactiae</i> challenge at different time-points (3 h, 12 h, and 24 h) post infection	12	80.30%	GSE30379	SRP007396
	Peripheral blood mononuclear cells	PBMC from goats that were vaccinated and infected with <i>Mycobacterium avium</i> subsp. Paratuberculosis	15	86.30%	GSE117799	SRP155536

Note, raw RNA-seq data retrieved from Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/browse/>).

TE annotations in goat were previously conducted by Dong et al. (2013), where they performed RepeatMasker annotation, together with RepeatModeller- and LTR\_FINDER-based *de novo* repeat annotations. Their annotations were integrated into the RepeatMasker libraries of current RepeatMasker version, we thus used “-species *C. hircus*” to call the priori annotations as the reference for our current genome assembly. To improve the coverage of TE annotation, the repeat library *FamDB* (CONS-Dfam\_with RBRM\_3.8) was also included for *rmbblastn*. The RepeatMasker generated tables were then parsed to filter out repeats like rRNA, scRNA, snRNA, srpRNA, and tRNA. The `makeTEgtf.pl` script ([http://labshare.cshl.edu/shares/mhammellab/www-data/TEtranscripts/TE\\_GTF/makeTEgtf.pl.gz](http://labshare.cshl.edu/shares/mhammellab/www-data/TEtranscripts/TE_GTF/makeTEgtf.pl.gz)) was used to reformat RepeatMasker tables into GTF file for subsequent analysis. Each TE and ERVs in the table were given a unique identifier, with genomic location, element name, subfamily and class information extracted from the table and were included in the GTF file. *Bedtools (-intersect)* was used to define the genomic location of ERVs in intergenic, intronic, and exonic regions.

## 2.4 Transcriptomic identification of TEs and ERVs at subfamily and location levels

The quality of raw sequencing data was assessed using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Raw reads were trimmed and adapters were removed using trimomatic-0.39-2 (Bolger et al., 2014). Trimmed reads were aligned to the *C. hircus* genome using STAR-2.7.4a (Dobin et al., 2013). Coding-gene and TE -derived reads were quantified using TEtranscripts (v2.2.3) and TELocal (v1.1.1) with their default parameters (Jin et al., 2015; Kabiljo et al., 2022). The aligned BAM file, together with two annotation GTF files for genes and TEs was used as the input data for TEtranscripts to identify TE subfamilies, while TELocal was used for single location TE and ERV identification. TE-derived transcripts were annotated according to the TE GTF file generated from the RepeatMasker. Because of the repetitive nature of TEs, there might be frequent multi-mapping and overlapping TEs, which may lead to bias to subsequent differential expression analysis. The following two steps were considered to minimize such impact. Firstly, RepeatMasker handles ambiguity of

overlapping TEs or multi-mapping by scoring sequence features and context. Distal overlapping TEs were identified and reported separately in the output file, while overlapping TEs in proximal were fused. Then, TE transcripts used equal weighting and expectation maximization strategies to avoid bias for differential expression analysis. If a read was mapped to multiple TEs, the read was equally weighted for each TE, to avoid bias by single mapping. The software use expectation maximization to estimate TE expression from multi-mapping reads, to ensure accurate quantification of TE expression. The strategy for such reads mapping and quantification are the same for all samples, resulting in comparable count matrix for subsequent differential expression analysis.

## 2.5 Differential expression analysis

Following the generation of a count table for gene and TE transcripts, differential expression analysis of genes and TEs was performed using R package DESeq2 (v1.20.0) with default parameters (Love et al., 2014). Normalized count, defined as counts divided by sample-specific size factors, determined by median ratios method of normalization, was used for differential expression analysis, correlation analysis, and visualization. Differentially expressed TEs were identified at the subfamily level. Considering the huge number of TEs at the single location level, we focused on differentially expressed ERV reads, which are highly involved in various physiological and pathological conditions at the location level. Since most of the genomic ERVs have no single read mapped to the annotated region, ERVs with raw count >0 were arbitrarily defined as expressed and were subjected to subsequent differential expression analysis. A gene, TE, and ERV-derived transcript with false discovery rate (FDR-adjusted *P*-value) less than 0.005 was defined as significantly differentially expressed. The ERVs and nearby genes at the location level were visualized using the Integrative Genomics Viewer (IGV) (version 2.18.2) with the read alignments (BAM) file. Since there is possibility that the proposed expression of candidate TE-derived transcript might be the by-product of the host gene expression, those TEs annotated to be located in coding regions (exonic, 5'UTR, and 3' UTR in Figure 2C) were excluded in the differential expression analysis. Volcano plots showing differential expression of TEs and ERVs were made using the R package *ggplot2* (version 3.5.1). Heatmaps were created using the R package *pheatmap* (version 1.0.12). Correlations between ERV expression level and coding gene levels were measured by Spearman's rho correlation.

## 3 Results

### 3.1 Genomic and transcriptomic features of TEs

Distribution of classes, families, and subfamilies of TEs in the *C. hircus* genome was systematically annotated. Their expression pattern in multiple tissues under different physiological and pathological conditions was investigated at TE subfamily and location levels (Figure 1A). In the genomic context, 49.1% of the

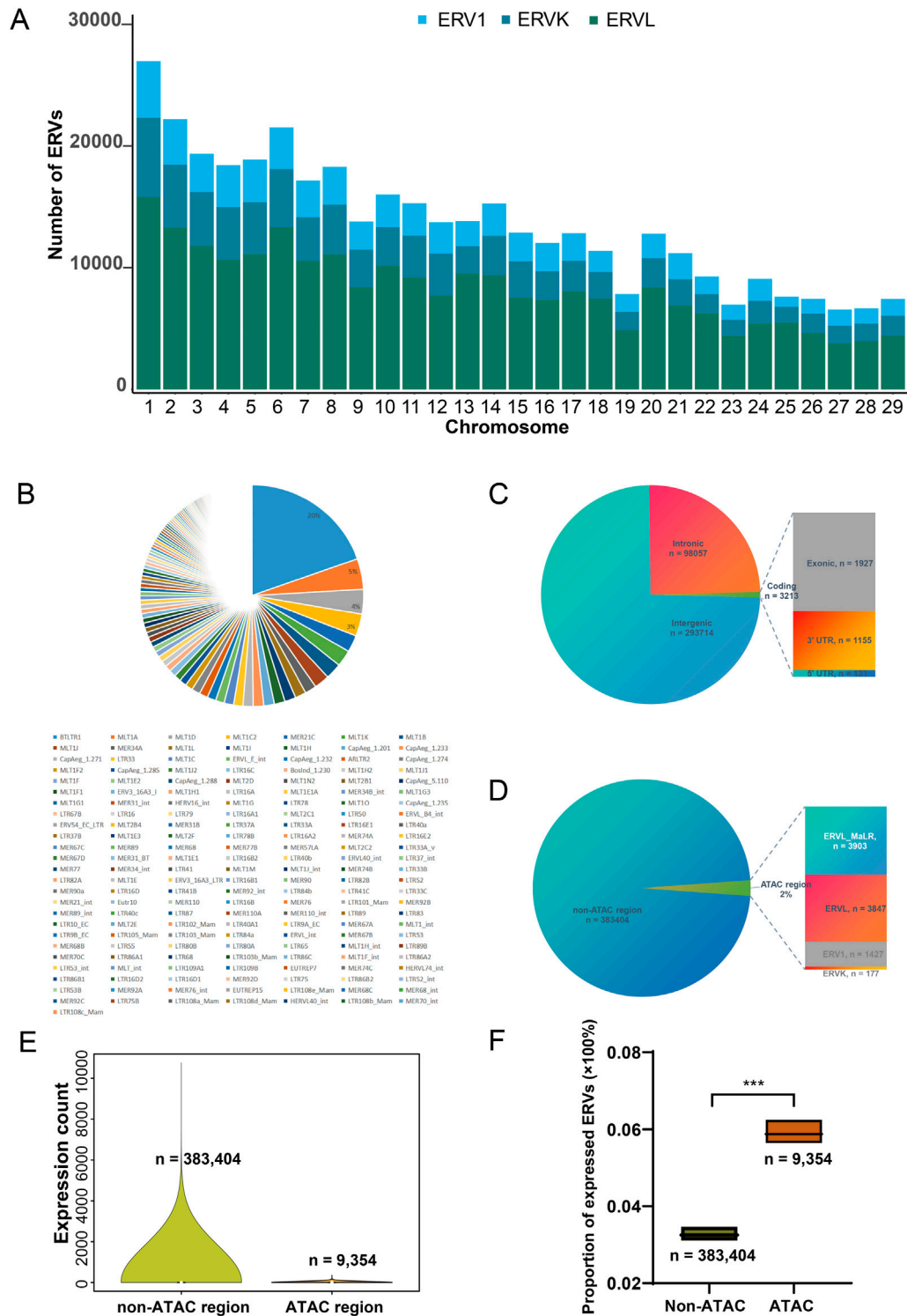
goat genome is composed by TE sequences, with 25.98% are LINES, 10.24% SINEs, 3.98% LTRs (ERVs) and 1.97% are DNA transposons (Figure 1B; Supplementary Table S1). These TE sequences could be divided into about 749 (average number of the TE subfamilies in transcriptome of analyzed tissues, range from 622 to 794) subfamilies, of which the majority are DNA transposons and LTR transposons in the transcriptome (Figure 1C; Supplementary Table S2).

Though they account for near half the genome sequence, TE-derived reads only account around 10% of the transcriptome in various tissues (Figure 1D). Consistent with the genomic sequence ratio (Figure 1B), abundance of SINEs and LINES in the transcriptome is the highest among the ~700 TE subfamilies (Figure 1E; Supplementary Table S2). And the SINEs and LINES derived transcripts showed a high abundance in the transcriptome of blastocyst and endometrium, partially due to a high level of transcriptomic variations at both subfamily (Supplementary Figure S1) and location (Supplementary Figure S2) levels within the group. In spite of the sample heterogeneity, SINEs and LINES derived transcripts each account for around 5% on average, while DNA TEs and LTR TEs each account for 1%, of the transcriptome across the tissues (Figures 1D, E). The comparable expression abundance of each TE family in different tissues and cells suggested constitutive expression of TEs in the transcriptome. Indeed, the constitutive expression of TEs is robust in physiological conditions like skins sampled from the normal and hyperpigmented goats, since few dysregulated TEs were observed in such condition (Supplementary Table S3, GSE69812). Notably, in pathological conditions, such as tissues infected with the Peste-des-petits-ruminants virus (PPRV), a large portion of the TE subfamilies are dysregulated with a stringent cutoff (FDR-adjusted *P*-value <0.005) for significantly differential expression: 165 (21.5%) in lung, 155 (20.2%) in spleen, and 654 (85.3%) in caecum were altered under PPRV challenge (Supplementary Table S3, GSE130552). This observation indicated extensive dysregulation and active involvement of TE expression in pathological conditions like external infection.

In addition to pathological challenge, regulation of TEs expression might also be variable across tissues (Figure 1F). In the 27 tissue-specific TEs (Figure 1F), half of them were LTR (which defines ERV) derived reads, indicating a variable nature of ERVs among the TE transcripts. Moreover, many ERVs have the potential to code for proteins, and are more active than other types of TEs in various processes. We thus focused on ERVs in subsequent analyses.

### 3.2 Genomic and transcriptomic features of ERVs

There are three major families of ERVs in the goat genome, ERV1, ERVK, and ERVL. These ERVs are located in all goat chromosomes, with the number of ERVs increasing along with chromosomal length (Figure 2A). Among 176 ERV subfamilies, the most abundant ERVs are BTLTR1, MLT1A, MLT1D, and MLT1C2 at the location level (Figure 2B). In the 392,758 non-redundant ERV insertions, most are located in intergenic regions, and 0.8% (*n* = 3213 ERV locations/insertions) of them located in coding regions (Figure 2C). We further investigated whether these



**FIGURE 2** Genomic features of ERVs in the goat genome. **(A)** Number of ERVs in all goat chromosomes. **(B)** Pie chart representing the proportions of ERV families in the genome. **(C)** Distribution of ERV insertions in genomic contexts. UTR, untranslated region; Coding, regions from 5′UTR, exons to 3′UTR. Note that some TE elements were annotated to be both intronic and located in coding region due to alternative splicing, leading to double counting; and the number of non-redundant ERVs are 392,758. **(D)** Ratio of ERV insertions in open chromatin regions. ATAC region, those genomic fragments revealed to be open chromatin by transposase-accessible chromatin using sequencing (ATAC-seq) of liver and immune cells from goat (Foissac et al., 2019). **(E)** Distribution of expression levels in liver for ERVs located in open chromatin region. Shown expression levels were measured by total count of mapped reads. **(F)** Proportion of expressed ERVs (raw count > 0) within and outside of the ATAC-region (region with ATAC peaks) from normal goat liver and T cells (Foissac et al., 2019). Expression profile of liver from GSE93855 was used.

ERVs are located in heterochromatin region or accessible chromatin region. The ATAC-seq data from normal goat liver and T cells, a reference of chromatin accessibility in ruminants (Foissac et al., 2019), was used to assess accessible chromatin regions. Based on the released open chromatin positions in the ATAC-seq data, we found that 2% ( $n = 9,354$  ERV locations/insertions) of the genomic ERVs located in open chromatin region in goat liver or immune cells (Figure 2D). Those ERVs located outside of the open chromatin region show more various expression in goat tissues, due to the large number of such ERVs (Figure 2E). Since most of the genomic ERVs have no mapped read (raw count = 0) in the transcriptome, we thus measured the proportion of expressed ERVs (raw count > 0) within and without the ATAC-region (region with ATAC peaks). The proportion of expressed ERVs located in the ATAC-region was significantly higher than that of non-ATAC regions (0.058% vs 0.033%,  $P < 0.001$ , Figure 2F). However, more than 90% of the 392,758 ERVs remain silenced in analyzed tissues under physiological conditions (Figure 2F), consistent to the well-established knowledge that most ERVs are silenced in the genome. We then moved on to investigate where and how ERVs may become dysregulated in goat tissues.

### 3.3 Dysregulated ERVs in the reproduction system

Consistent with the robustness of TEs expression in physiological conditions, no ERVs were significantly altered during skin pigmentation (Supplementary Table S4, GSE69812). Since ERVs play essential roles during embryo development, we investigated whether any ERVs are differentially expressed in goat embryos under different developmental stages. Differentially expressed ERVs were also explored in the ovary and the endometrium (Supplementary Table S4), where the expression regulation might be cooperated with the embryo to ensure success reproduction. We found that among the 12,711 expressed ERVs, most dysregulated ERVs are activated during embryo development, and reached peak at 16-cell embryo (Figures 3A–C). In particular, there were 642 ERVs upregulated ( $\log_2$ Fold Change [ $\log_2FC$ ] > 0, FDR-adjusted  $P$ -value [ $P_{adj}$ ] < 0.005) and 57 ERVs downregulated ( $\log_2FC < 0$ ,  $P_{adj} < 0.005$ ) in 16-cell embryo compared with 8-cell embryo (Figures 3A, B). For instance, BosInd\_1.230:ERV1:LTR (chr1:12676110-12676292,  $\log_2FC = 13.265$ ,  $P_{adj} = 4.867 \times 10^{-22}$ ), CapAeg\_1.232:ERV1:LTR (chr1:143876497-143876779,  $\log_2FC = 14.649$ ,  $P_{adj} = 1.135 \times 10^{-20}$ ), and BosInd\_1.230:ERV1:LTR (chr18:60592269-60592640,  $\log_2FC = 11.199$ ,  $P_{adj} = 1.848 \times 10^{-18}$ ) are the most highly activated ERVs in 16-cell embryos (Figure 3A; Supplementary Table S4, GSE129742), when the genome of the zygote begins to transcribe its own genes. During this stage, the genome is of high accessibility, rendering the extensive expression of ERVs as well. In the late embryo stage, the blastocyst, only one ERV CapAeg\_1.233:ERVK:LTR (chr1:6646491-6648140,  $\log_2FC = 7.603$ ,  $P_{adj} = 0.0022$ ) was upregulated compared with morula.

In peri-implantation endometrial epithelium, there were only 1,024 ERVs expressed, among which there were 32 upregulated and 6 downregulated ERVs at 16 days post-implantation compared to 6 days endometrial epithelium (Figure 3A; Supplementary Table S4,

GSE108557). None of these ERVs are overlapped with those dysregulated in embryos (Figure 3B), partially due to limited number of analyzed ERVs in endometrium. Since IFNT signaling is essential for implantation, there were a dozen ERVs upregulated in endometrial epithelial cells with IFNT treatment, which were consistent in 16 days post-implantation endometrial epithelium compared to 6 days (Figures 3C, D; Supplementary Table S4, GSE184110). In addition to endometrium, some ERVs (BosInd\_1.230:ERV1 and LTR89B:ERVL) were also activated in skin during rapid growth of the hair follicles (Figures 3A, B), indicating the importance of cell proliferation state for ERV expression.

### 3.4 Dysregulated ERVs in response to infections

In addition to the cell proliferation-responsive ERVs, we recognized numerous activated ERVs in response to virus infection in the lung, spleen, caecum, and immune cells (Figures 4A, B; Supplementary Table S4). In goats infected by the Peste-des-petits-ruminants virus (PPRV, a single strand RNA virus), there were more upregulated ERVs in caecum and B cells (Figures 4A, B). 327 out of 4,542 ERVs were upregulated and 164 downregulated in caecum under PPRV infection, with 20 ERVs consistently upregulated in all three tissues (Supplementary Table S4). The most significantly upregulated were MER74A:ERVL:LTR (chr17:9732009-9732446), CapAeg\_5.110:ERV1:LTR (chr1:141232530-141232858), and MER34A:ERV1:LTR (chr26:4,0321713-4,0321937). Intriguingly, MLT1E1A:ERVL\_MaLR:LTR (chr11:29358543-29359075) was significantly upregulated in caecum ( $\log_2FC = 3.012$ ,  $P_{adj} < 1.0 \times 10^{-14}$ ) but downregulated in lung ( $\log_2FC = -1.673$ ,  $P_{adj} = 5.222 \times 10^{-11}$ ). Nine ERVs were consistently dysregulated in the PPRV-infected tissues and immune cells (Figures 4C, D), for instance, CapAeg\_5.110:ERV1 and MER21C:ERVL were upregulated in all tissues and immune cells. Notably, we observed opposite direction of ERV activation between monocyte and lymphocytes, namely, those ERVs upregulated in PPRV-infected lymphocytes and tissues were downregulated in PPRV-infected monocyte (Figures 4C, D). This observation was in accordance with different anti-virus roles of lymphocytes and monocytes.

Intriguingly, DNA virus, like Orf virus (ORFV), and non-viral infection (e.g., *Mycoplasma agalactiae* and *M. avium* subsp. *Paratuberculosis*) activate less ERVs in the transcriptome (Supplementary Table S4, GSE121725, GSE30379, and GSE117799), indicating that those RNA-virus-derived endogenous ERVs may be hitchhiking on the infection processes of external RNA virus infection.

### 3.5 Regulatory mechanism of dysregulated ERVs by proximal genes

We then investigated how and why those ERVs are dysregulated under certain conditions. ERVs altered in various conditions (indicated by Figures 3C, 4C) were subjected to subsequent analysis (Supplementary Table S5). Among the commonly altered 35 ERVs, five out of the 14 differentially expressed ERVs shared by 16 days-post-implantation

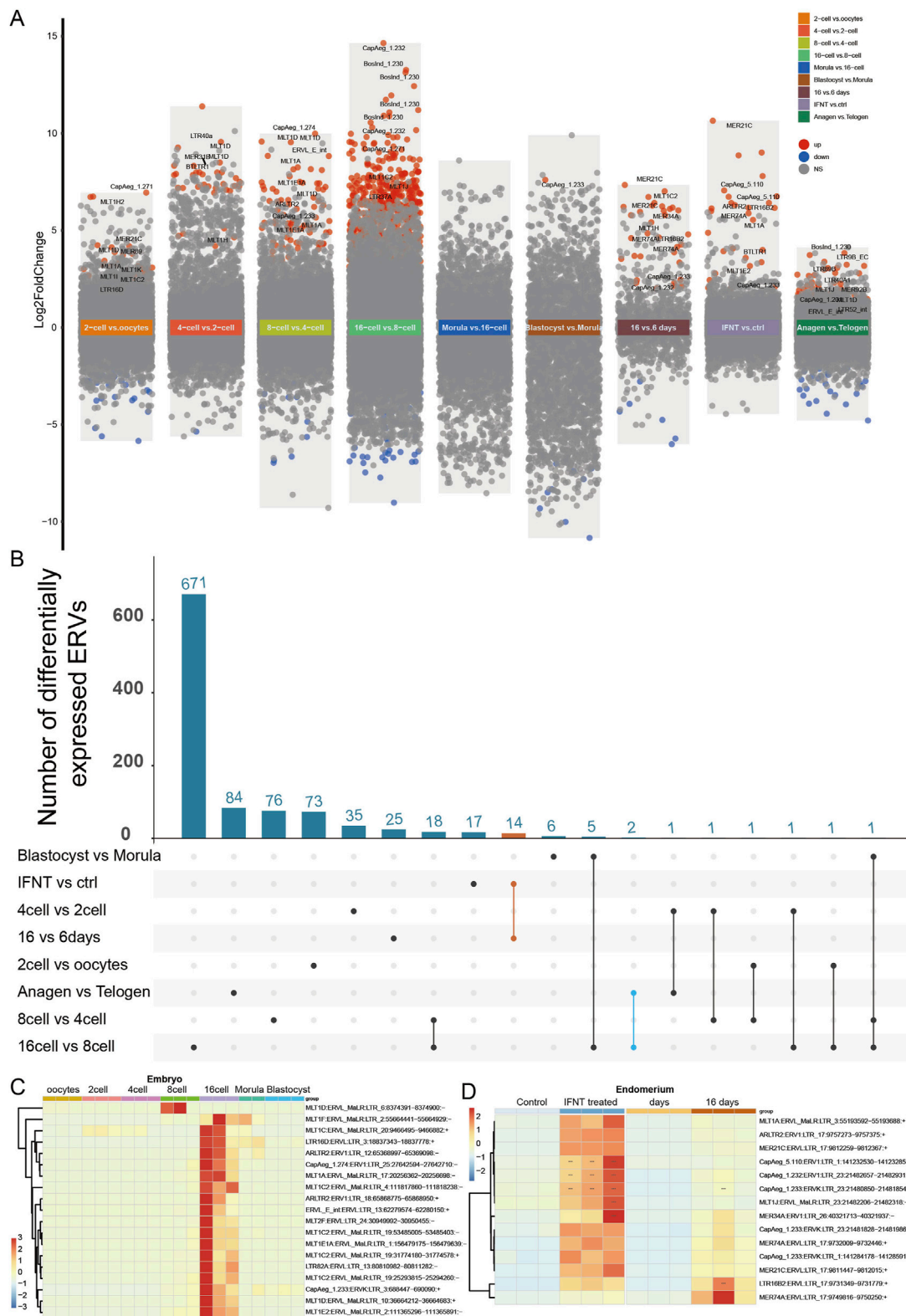
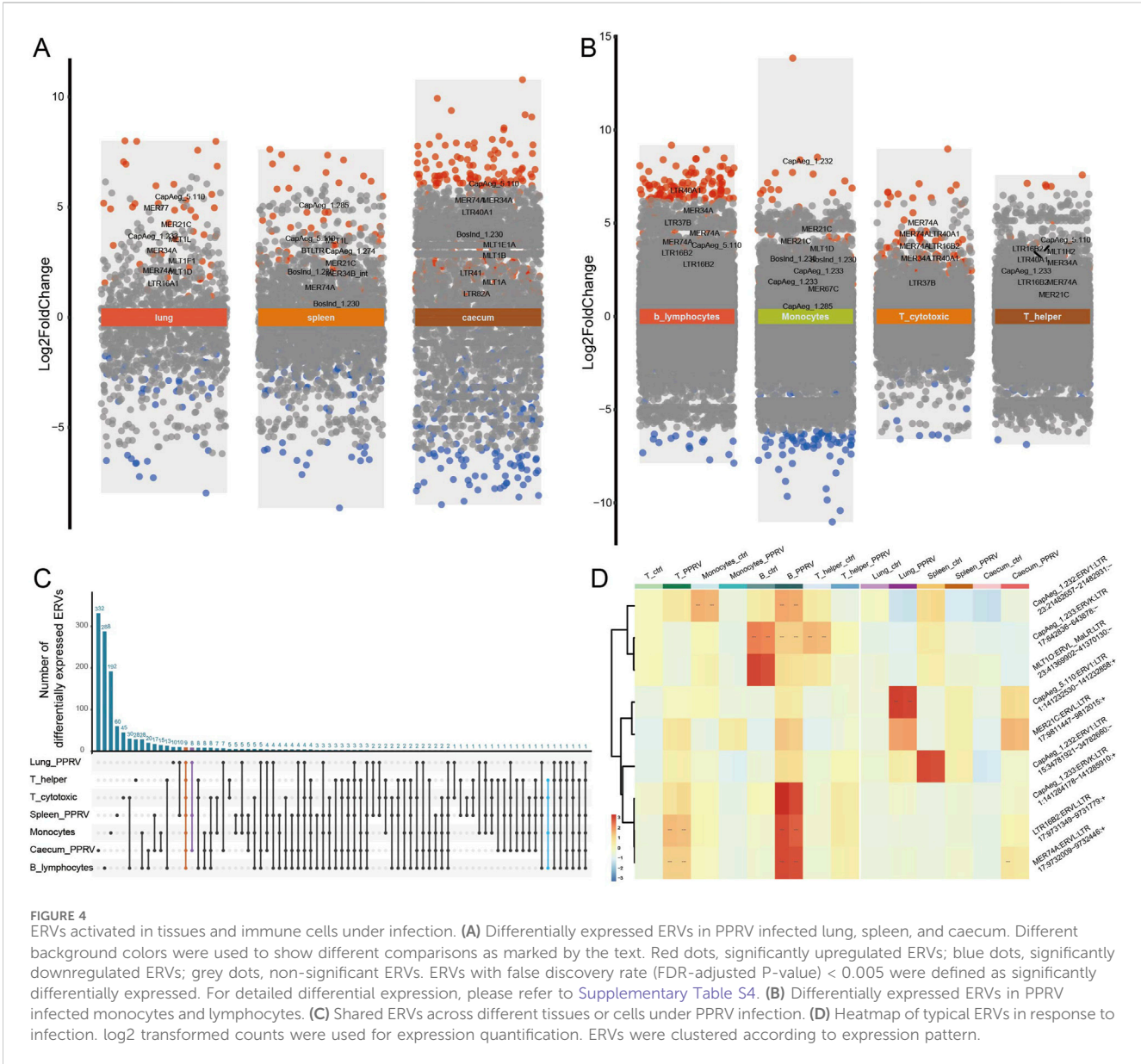


FIGURE 3

ERVs activated in embryo and peri-implantation endometrial epithelium. (A) Differentially expressed ERVs in embryos, endometrium, and skin at different stages. Different background colors were used to show different comparisons as marked by the text. Red dots, significantly upregulated ERVs; blue dots, significantly downregulated ERVs; grey dots, non-significant ERVs; vs. versus. ERVs with false discovery rate (FDR-adjusted P-value) < 0.005 were defined as significantly differentially expressed. "16 vs. 6 days", ERVs from endometrial epithelium at 16 days peri-implantation compared to 6 days; "INFT vs ctrl", INFT treated endometrial epithelium compared to control group; "Anagen vs Telogen", ERVs from skin biopsies sampled at active growth phase (anagen) compared to resting phase (telogen). For detailed differential expression, please refer to [Supplementary Table S4](#). (B) Number of activated ERVs shared by different comparisons. Orange line, activated ERVs shared by different endometrium (listed in [Supplementary Table S5](#)); blue (Continued)

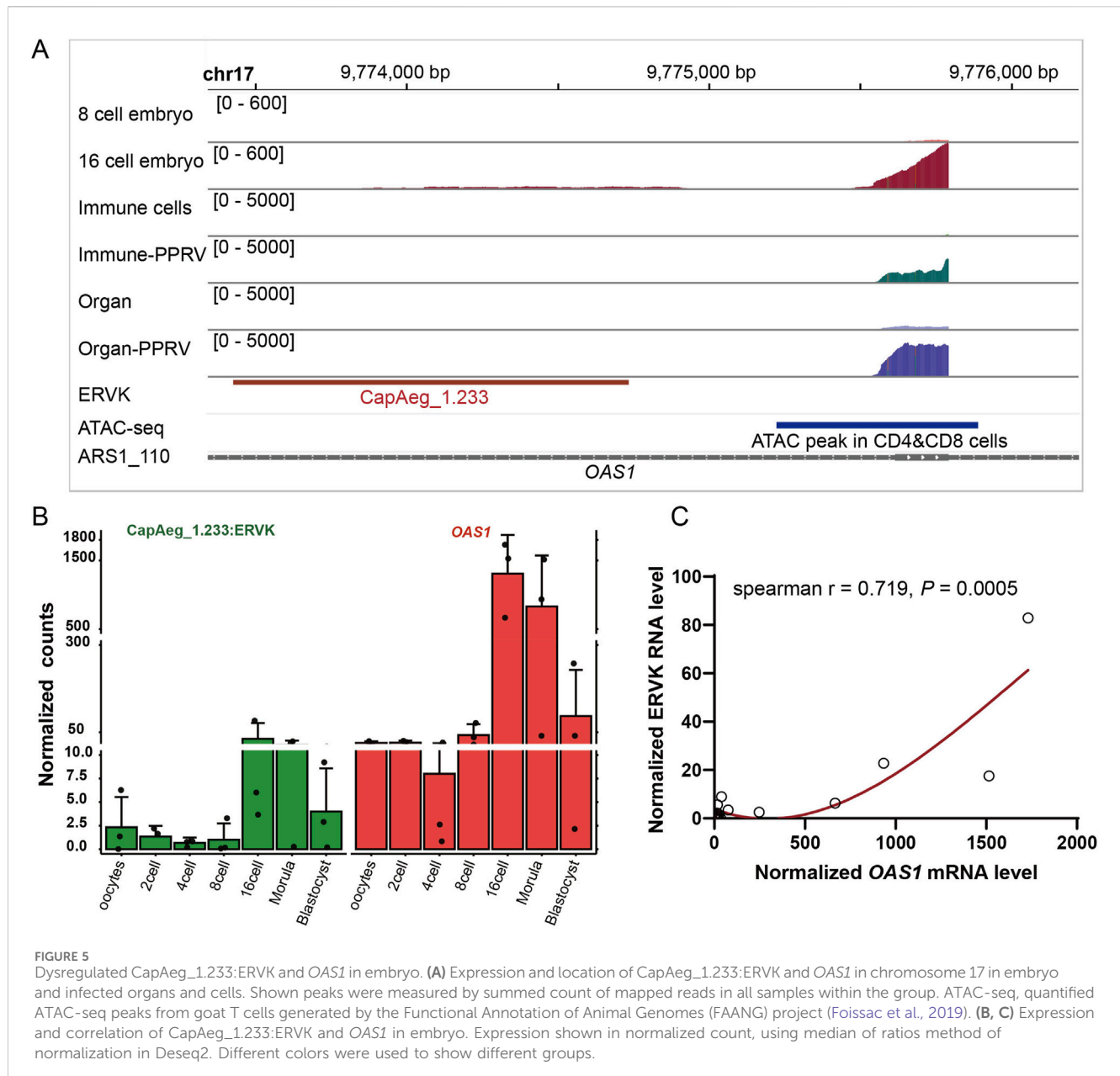


**FIGURE 3 (Continued)**  
 line, activated ERVs shared by endometrial epithelium and embryos. **(C)** Heatmap of typical ERVs altered in 16-cell and 8-cell embryos. **(D)** Heatmap of dysregulated ERVs shared by different endometrium [N = 14, orange line in **(B)**]. log<sub>2</sub> transformed counts were used for expression quantification. ERVs were clustered according to expression pattern.



endometrial epithelium and IFNT treated endometrial epithelial cells located around an important interferon-induced gene, *OAS1*. And in nine differentially expressed ERVs in PPRV-infected tissues and cells, there were three located within *OAS1*. Though *OAS1* activated in both embryo development and virus infection, the proximally activated ERVs are different, indicating specific physiological functions of distinct ERVs. LTR16B2:ERVL and MER74A:ERVL were significantly upregulated under infection (Supplementary Figures S3A, B),

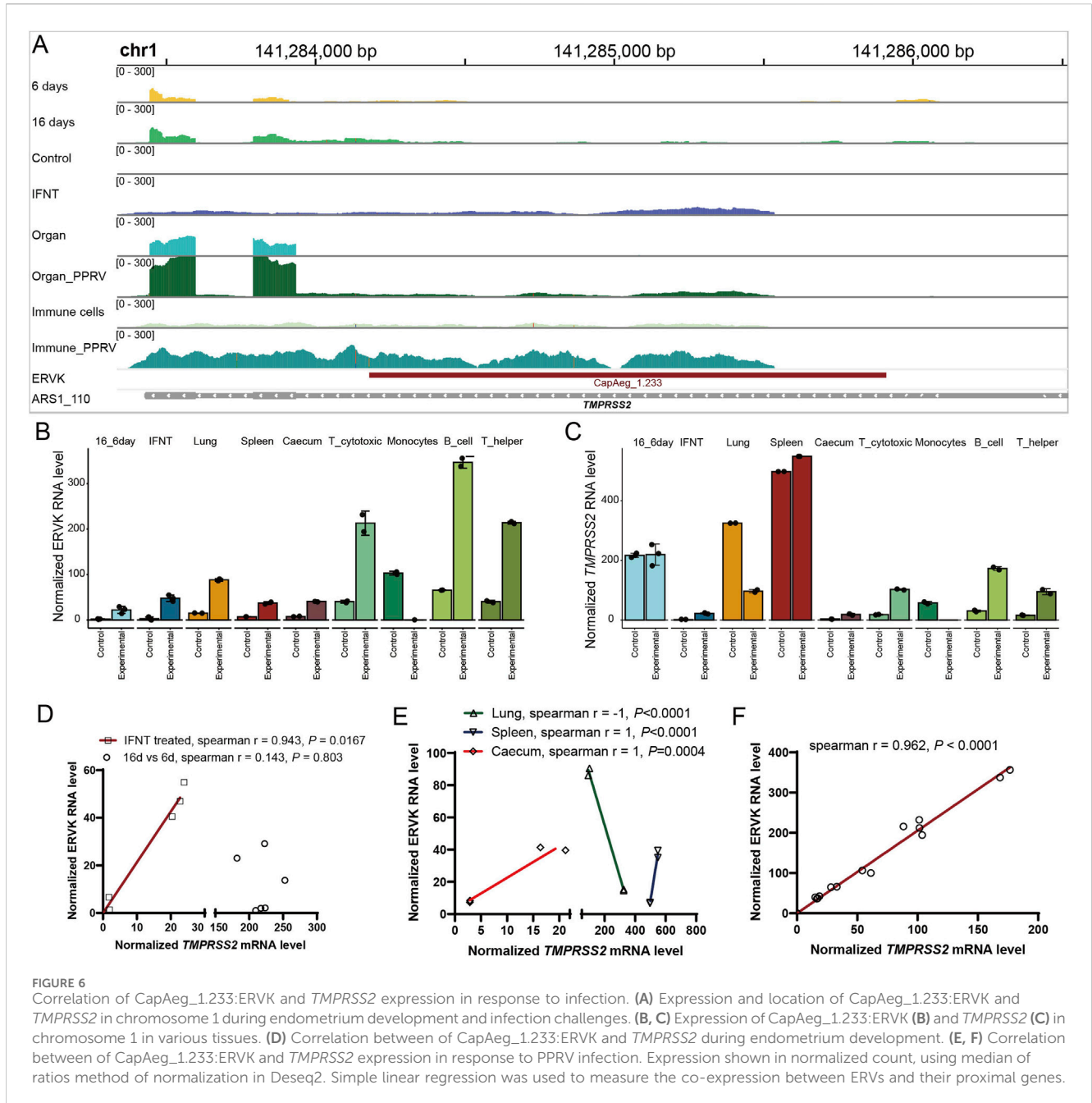
and showed positive correlation with the expression of *OAS1* (Supplementary Figures S3B, C), whereas CapAeg\_1.233:ERVK was specifically upregulated in embryos (Figure 5A) and also showed positive correlation with *OAS1* (Figures 5B, C). Notably, these ERVs are located in the intronic region of *OAS1*, rendering the possibility that the proposed expression of candidate TE-transcript might be the by-product of the host gene expression. Indeed, LTR16B2 and MER74A are such cases, since the reads distributed equally within and outside of the TE along the intron



(Supplementary Figure S3A). And CapAeg\_1.233:ERVK was reasonably expressed independently, with peaks exactly in TE region and no reads outside of TEs were observed along the intron (Figure 5A). We thus conclude that though ERVs around OAS1 were proposed to be frequently altered in various conditions, only CapAeg\_1.233:ERVK showed robust and specific expression during embryo development. Whether infection may induce the expression of MER74A and LTR16B2, independent of infection-induced OAS1, remain to be determined. Because of the incomplete of the annotation of goat genome, we cannot rule out whether there are an unknown infection-inducible OAS1 isoform or ERV-derived lncRNA.

In addition to the OAS1 region, CapAeg\_1.233:ERVK:LTR (chr1:141284178-141285910) proximal to another important immune gene, *TMPRSS2*, was altered during endometrium

development and under infection (Supplementary Table S5; Figure 6). It shows positive correlation with the expression of *TMPRSS2* in IFNT-treated endometrium (Figure 6D), PPRV-infected caecum, spleen (Figure 6E), and immune cells (Figure 6F). Strangely, *TMPRSS2* was downregulated in PPRV-infected lung (Figure 6C), leading to unexplainable negative correlation between CapAeg\_1.233:ERVK:LTR and *TMPRSS2* in lung (Figure 6E). Some other infection-responsive genes like *IFI144*, *MX1*, *MX2*, *IFIT3*, and *HLA-DOA* were also observed proximal to those altered ERVs (Supplementary Table S5), and showed positive co-expressions with their proximal ERVs, wherever they located (Supplementary Figure S4). We therefore propose that dysregulated proximal genes, which are active in response to the respective conditions, contribute to the dysregulation of nearby ERVs.



## 4 Discussion

TE-derived transcripts account for a non-negligible proportion of a mammalian transcriptome (Bourque et al., 2018; Dopkins and Nixon, 2024). However, most standard expression analyses ignored such reads due to the lack of tools that allow easy inclusion of TE-derived reads (Lanciano and Cristofari, 2020). In goat, the transcriptomic feature of TEs is missed, limiting the full understanding of the goat genome. In this study, we took advantage of Tetranscripts and TElocal in analyses of series of transcriptomes of goat tissues. Since Tetranscripts is highly dependent on the quality of the genomic annotation, and this is problematic for less studied species like the goat, we manually curated a GTF file for TE annotation in goat. Then we

recognized the abundance of TE-derived reads in transcriptomes of goat tissues and investigated the expression pattern of ERV-derived transcripts in various tissues and conditions. We found that TEs are constitutively expressed in the transcriptome of tissues and cells, accounting for 10% of the transcriptome. ERVs are actively altered in some conditions, especially during embryo development and in response to infection (Supplementary Tables S3, S4). Specially, we showed that ERVs on chromosome 17 respond to different physiological and pathological conditions, and are co-expressed with their proximal coding genes. These results may benefit goat-based genomic and transcriptomic research.

TEs in the goat genome have been investigated when the reference genome was released (Dong et al., 2013; Dong et al., 2015; Belay et al., 2024). The ratio of TEs, especially LTRs, is

consistent with our re-analyses using another genome assembly, supporting the reliability of the TE annotation. However, the detailed genomic features of TEs and ERVs were unclear in the goat (Chang et al., 2024), while they are well-studied in sheep and other farm animals (Klymiuk et al., 2003; Baillie et al., 2004; Arnaud et al., 2008; Garcia-Etxebarria et al., 2014). It is therefore worthwhile to conduct a deep annotation of TEs at the family and location levels for the goat genome. We herein curated the first GTF file for TE annotation in goat, which is publicly available for further validation and use.

Being an essential part of the genome sequences, TEs play multiple roles in the evolution, structure and function of mammal genome (Lanciano and Cristofari, 2020; Horvath et al., 2017; Bourque et al., 2018), such as in expression regulation (Lanciano and Cristofari, 2020; Ivancevic et al., 2024; Yu et al., 2023; Branco and Frost, 2023). Moreover, the abundance of TEs is constitutive in the transcriptome, though most TE sequences are silenced. Among those expressed TEs, ERVs gained most attention, since ERV-derived genes drive the evolution of placental mammals (Chuong, 2018; Mi et al., 2000; Bourque et al., 2018; Haig, 2012), and even function in other non-mammals like birds (Chen et al., 2022). Indeed, we found that in goat embryo, a thousand of ERVs are dysregulated during embryo development, especially in the 16-cell stage, when the zygote's genome is activated (Deng et al., 2019). During embryo development, the cell proliferates quickly, and the chromatin is highly open, rendering the concomitant expression of ERVs. Similarly, the transcription of the host cell becomes active when challenged by infection, ERVs are also highly activated. Notably, the interferon-induced gene *OAS1*, upregulated in both development and infection, resulting in the activation of its nearby ERVs. Yet it is unclear why *OAS1* is upregulated in embryo development and what's the impact of ERV activation in these progresses. In particular, these endogenous retroviruses become active when there are exogenous retroviruses, or RNA virus (PPRV in this study) infections, while DNA virus or bacterial infection induce few ERV expression. Surely there are interactions between the contemporary retroviruses and their endogenized ancestors (Kyriakou and Magiorkinis, 2023). This is supported by a recent human study showing the activity of certain ERVs in the colon of HIV reservoirs (Dopkins et al., 2024a). These observations suggested that active transcription event of the host might be hijacked by ERVs (Asimi et al., 2022; Grow, 2022), which may lead to subsequent impact on the host cells (Dopkins and Nixon, 2024; Ivancevic et al., 2024; Wang et al., 2024; Yu et al., 2024; Guo et al., 2024; Dopkins et al., 2024b; da Silva et al., 2024). Notably, such TEs should be interpreted with caution, since there is possibility that the proposed expression of candidate TE-transcript might be the by-product of the host gene expression. This bias should be particularly checked when the TEs were in exonic and intronic regions. Moreover, there might be overlapping TEs due to their repetitive nature. Though we did not conduct experimental validation since the study covers various types of tissues and conditions, we resolved ambiguity using a series of strategies. Further experimental validations are undoubtedly warranted for the above issues.

In summary, in this study we curated a GTF file for TE annotation and generated the first TE-derived transcriptomes across goat tissues. The expression pattern of ERV-derived transcripts in various tissues and conditions was comprehensively explored. These results may benefit goat-based genomic and transcriptomic researches. It may also enhance the understanding and treatment of infection threats for goat farming. The annotation of TEs might be biased, since the reference genome keeps updating due to the rapid development of sequencing techniques. Annotating these structural variations and repetitive elements using more recent genome assemblies, or using full-length sequencing data, followed by experimental validations, will undoubtedly improve future TE-related research.

## Data availability statement

The data presented in the study are available in the GEO repository, accession numbers listed in Table 1. All data generated or analyzed during this study are included in this published article and its Supplementary Material. Codes and GTF files for TE annotation at the family and location levels are publicly accessible through Figshare at [https://figshare.com/articles/dataset/GTF\\_files\\_for\\_annotating\\_transposable\\_elements\\_in\\_i\\_Capra\\_hircus\\_i\\_goat\\_genome/27898515](https://figshare.com/articles/dataset/GTF_files_for_annotating_transposable_elements_in_i_Capra_hircus_i_goat_genome/27898515) (DOI: 10.6084/m9.figshare.27898515). These files can also be accessed through the following links: [https://drive.google.com/file/d/12ClX4cFKJ5us8aq0I-xDVFDydyRTpc/view?usp=drive\\_link](https://drive.google.com/file/d/12ClX4cFKJ5us8aq0I-xDVFDydyRTpc/view?usp=drive_link) for TE family annotation, and [https://drive.google.com/file/d/1NpE\\_5eZOOAdcsUaJuks446YzqnaBjk2a/view?usp=drive\\_link](https://drive.google.com/file/d/1NpE_5eZOOAdcsUaJuks446YzqnaBjk2a/view?usp=drive_link) for TE location annotation.

## Ethics statement

Ethical approval was not required for the study involving animals in accordance with the local legislation and institutional requirements because this study uses publicly datasets, no bench experiments were conducted on animals.

## Author contributions

M-DL: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Resources, Software, Visualization, Writing—original draft, Writing—review and editing. H-RL: Investigation, Validation, Writing—review and editing. S-HY: Conceptualization, Funding acquisition, Project administration, Supervision, Writing—review and editing, Data curation.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work is supported by the Ministry of Agriculture and Rural of the People's Republic of China (19221953), and the Modern Agro-industry Technology Research System (CARS-39-01). The funder has no role in the conceptualization, design, data

collection, analysis, decision to publish, or preparation of the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## References

- Arnaud, F., Varela, M., Spencer, T. E., and Palmarini, M. (2008). Coevolution of endogenous Betaretroviruses of sheep and their host. *Cell. Mol. Life Sci.* 65 (21), 3422–3432. doi:10.1007/s00018-008-8500-9
- Asimi, V., Sampath Kumar, A., Niskanen, H., Riemenschneider, C., Hetzel, S., Naderi, J., et al. (2022). Hijacking of transcriptional condensates by endogenous retroviruses. *Nat. Genet.* 54 (8), 1238–1247. doi:10.1038/s41588-022-01132-w
- Baillie, G. J., van de Lagemaat, L. N., Baust, C., and Mager, D. L. (2004). Multiple groups of endogenous betaretroviruses in mice, rats, and other mammals. *J. Virology* 78 (11), 5784–5798. doi:10.1128/JVI.78.11.5784-5798.2004
- Belay, S., Belay, G., Nigussie, H., Jian-Lin, H., Tijjani, A., Ahbara, A. M., et al. (2024). Whole-genome resource sequences of 57 indigenous Ethiopian goats. *Sci. Data* 11 (1), 139. doi:10.1038/s41597-024-02973-2
- Berry, A., Wu, C. W., Venturino, A. J., and Talaat, A. M. (2018). Biomarkers for early stages of john's disease infection and immunization in goats. *Front. Microbiol.* 9, 2284. doi:10.3389/fmicb.2018.02284
- Bhat, B., Yaseen, M., Singh, A., Ahmad, S. M., and Ganai, N. A. (2021). Identification of potential key genes and pathways associated with the Pashmina fiber initiation using RNA-Seq and integrated bioinformatics analysis. *Sci. Rep.* 11 (1), 1766. doi:10.1038/s41598-021-81471-6
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi:10.1093/bioinformatics/btu170
- Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., et al. (2018). Ten things you should know about transposable elements. *Genome Biol.* 19, 199. doi:10.1186/s13059-018-1577-z
- Branco, M. R., and Frost, J. M. (2023). Endogenous retroviruses control human placental gene expression. *Nat. Struct. and Mol. Biol.* 30 (4), 415–416. doi:10.1038/s41594-023-00965-1
- Chang, L. L., Zheng, Y., Li, S., Niu, X., Huang, S., Long, Q., et al. (2024). Identification of genomic characteristics and selective signals in Guizhou black goat. *BMC Genomics* 25 (1), 164. doi:10.1186/s12864-023-09954-6
- Chen, J. Q., Zhang, M. P., Tong, X. K., Li, J. Q., Zhang, Z., Huang, F., 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 Res.* 43 (3), 423–441. doi:10.24272/j.issn.2095-8137.2021.379
- Chuong, E. B. (2018). The placenta goes viral: retroviruses control gene expression in pregnancy. *PLoS Biol.* 16 (10), e3000028. doi:10.1371/journal.pbio.3000028
- da Silva, A. L., Guedes, B. L. M., Santos, S. N., Correa, G. F., Nardy, A., Nali, L. H. d. S., et al. (2024). Beyond pathogens: the intriguing genetic legacy of endogenous retroviruses in host physiology. *Front. Cell. Infect. Microbiol.* 14, 1379962. doi:10.3389/fcimb.2024.1379962
- de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7 (12), e1002384. doi:10.1371/journal.pgen.1002384
- Deng, R. Z., Han, C., Zhao, L., Zhang, Q., Yan, B., Cheng, R., et al. (2019). Identification and characterization of ERV transcripts in goat embryos. *Reproduction* 157 (1), 115–126. doi:10.1530/REP-18-0336
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15–21. doi:10.1093/bioinformatics/bts635
- Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., et al. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* 31 (2), 135–141. doi:10.1038/nbt.2478
- Dong, Y., Zhang, X., Xie, M., Arefnezhad, B., Wang, Z., Wang, W., et al. (2015). Reference genome of wild goat (*capra aegagrus*) and sequencing of goat breeds provide insight into genic basis of goat domestication. *BMC Genomics* 16, 431. doi:10.1186/s12864-015-1606-1
- Dopkins, N., Fei, T., Michael, S., Liotta, N., Guo, K., Mickens, K. L., et al. (2024a). Endogenous retroelement expression in the gut microenvironment of people living with HIV-1. *Ebiomedicine*, 103. doi:10.1016/j.ebiom.2024.105133
- Dopkins, N., and Nixon, D. F. (2024). Activation of human endogenous retroviruses and its physiological consequences. *Nat. Rev. Mol. Cell Biol.* 25 (3), 212–222. doi:10.1038/s41580-023-00674-z
- Dopkins, N., Singh, B., Michael, S., Zhang, P., Marston, J. L., Fei, T., et al. (2024b). Ribosomal profiling of human endogenous retroviruses in healthy tissues. *BMC Genomics* 25 (1), 5. doi:10.1186/s12864-023-09909-x
- Foissac, S., Djebali, S., Munyard, K., Vialaneix, N., Rau, A., Muret, K., et al. (2019). Multi-species annotation of transcriptome and chromatin structure in domesticated animals. *BMC Biol.* 17 (1), 108. doi:10.1186/s12915-019-0726-5
- Garcia-Etxebarria, K., Sistiaga-Poveda, M., and Jugo, B. M. (2014). Endogenous retroviruses in domestic animals. *Curr. Genomics* 15 (4), 256–265. doi:10.2174/1389202915666140520003503
- Grow, E. J. (2022). Endogenous retroviruses steer transcriptional condensates away from pluripotency. *Nat. Genet.* 54 (8), 1068–1069. doi:10.1038/s41588-022-01111-1
- Guo, X. F., Zhao, Y., and You, F. P. (2024). Identification and characterization of endogenous retroviruses upon SARS-CoV-2 infection. *Front. Immunol.* 15, 1294020. doi:10.3389/fimmu.2024.1294020
- Haig, D. (2012). Retroviruses and the placenta. *Curr. Biol.* 22 (15), R609–R613. doi:10.1016/j.cub.2012.06.002
- Horvath, V., Merenciano, M., and Gonzalez, J. (2017). Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends Genet.* 33 (11), 832–841. doi:10.1016/j.tig.2017.08.007
- Ivancevic, A., Simpson, D. M., Joyner, O. M., Bagby, S. M., Nguyen, L. L., Bitler, B. G., et al. (2024). Endogenous retroviruses mediate transcriptional rewiring in response to oncogenic signaling in colorectal cancer. *Sci. Adv.* 10 (29), eado1218. doi:10.1126/sciadv.ado1218
- Jin, Y., Tam, O. H., Paniagua, E., and Hammell, M. (2015). TETranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* 31 (22), 3593–3599. doi:10.1093/bioinformatics/btv422
- Kabiljo, R., Bowles, H., Marriott, H., Jones, A. R., Bouton, C. R., Dobson, R. J. B., et al. (2022). RetroSnake: a modular pipeline to detect human endogenous retroviruses in genome sequencing data. *IScience* 25 (11), 105289. doi:10.1016/j.isci.2022.105289
- Klymiuk, N., Müller, M., Brem, G., and Aigner, B. (2003). Characterization of endogenous retroviruses in sheep. *J. Virology* 77 (20), 11268–11273. doi:10.1128/jvi.77.20.11268-11273.2003

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1544330/full#supplementary-material>

- Kyriakou, E., and Magiorkinis, G. (2023). Interplay between endogenous and exogenous human retroviruses. *Trends Microbiol.* 31 (9), 933–946. doi:10.1016/j.tim.2023.03.008
- Lanciano, S., and Cristofari, G. (2020). Measuring and interpreting transposable element expression. *Nat. Rev. Genet.* 21 (12), 721–736. doi:10.1038/s41576-020-0251-y
- Li, Y., Sun, J., Ling, Y., Ming, H., Chen, Z., Fang, F., et al. (2020). Transcription profiles of oocytes during maturation and embryos during preimplantation development *in vivo* in the goat. *Reprod. Fertil. Dev.* 32 (7), 714–725. doi:10.1071/RD19391
- Liu, H., Wang, C., Li, Z., Shang, C., Zhang, X., Zhang, R., et al. (2021). Transcriptomic analysis of STAT1/3 in the goat endometrium during embryo implantation. *Front. Vet. Sci.* 8, 757759. doi:10.3389/fvets.2021.757759
- Liu, Y., Qi, B., Xie, J., Wu, X., Ling, Y., Cao, X., et al. (2018). Filtered reproductive long non-coding RNAs by genome-wide analyses of goat ovary at different estrus periods. *BMC Genomics* 19 (1), 866. doi:10.1186/s12864-018-5268-7
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8
- Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., et al. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403 (6771), 785–789. doi:10.1038/35001608
- Ren, H., Wang, G., Chen, L., Jiang, J., Liu, L., Li, N., et al. (2016). Genome-wide analysis of long non-coding RNAs at early stage of skin pigmentation in goats (*Capra hircus*). *BMC Genomics* 17, 67. doi:10.1186/s12864-016-2365-3
- Tang, Q., Gu, Y., Zhou, X., Jin, L., Guan, J., Liu, R., et al. (2017). Comparative transcriptomics of 5 high-altitude vertebrates and their low-altitude relatives. *Gigascience* 6 (12), 1–9. doi:10.1093/gigascience/gix105
- Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.* Chapter 4, p. 4.10.1-4.10.14. doi:10.1002/0471250953.bi0410s25
- Wang, J. C., Lu, X., Zhang, W., and Liu, G. H. (2024). Endogenous retroviruses in development and health. *Trends Microbiol.* 32 (4), 342–354. doi:10.1016/j.tim.2023.09.006
- Wani, S. A., Sahu, A. R., Khan, R. I. N., Pandey, A., et al. (2019). Contrasting gene expression profiles of monocytes and lymphocytes from peste-des-petits-ruminants virus infected goats. *Front. Immunol.* 10: 1463. doi:10.3389/fimmu.2019.01463
- Yu, J. D., Qiu, P., Ai, J., Liu, B., Han, G. Z., Zhu, F., et al. (2024). Endogenous retrovirus activation: potential for immunology and clinical applications. *Natl. Sci. Rev.* 11 (4), nwae034. doi:10.1093/nsr/nwae034
- Yu, M., Hu, X., Pan, Z., Du, C., Jiang, J., Zheng, W., et al. (2023). Endogenous retrovirus-derived enhancers confer the transcriptional regulation of human trophoblast syncytialization. *Nucleic Acids Res.* 51 (10), 4745–4759. doi:10.1093/nar/gkad109