Fan et al. *Zool. Res.* 2026, 47(3): 945−959
https://doi.org/10.24272/j.issn.2095-8137.2025.393
https://cstr.cn/32248.14.zr.2025.393

ZR | Zoological Research

Article Open Access

Decoding the enhancer-driven regulatory network for lactoferrin across mammalian species

Hui-Mei Fan[1,3,#], Yu-Yi Lu[1,3,#], Zhen-Yu Wei[1,3,#], Xin-Mei Li[1], Hao Liu[1], Si-Yu Zhang[1], Wei Wang[1], Xiao-Ru Yan[1], Ya-Yi Liu[2], Nan-Ge Ma[1], Yue Zhang[1], Ya-Mei Wu[1], Yu Jiang[1,3], Yu Wang[1,3,4,*]

[1] *Key Laboratory of Animal Genetics*, *Breeding and Reproduction of Shaanxi Province*, *College of Animal Science and Technology*, *Northwest A & F University*, *Yangling*, *Shaanxi* 712100, *China*

[2] *College of Veterinary Medicine*, *Northwest A & F University*, *Yangling*, *Shaanxi* 712100, *China*

[3] *Key Laboratory of Livestock Biology*, *Northwest A & F University*, *Yangling*, *Shaanxi* 712100, *China*

[4] *National Center of Technology Innovation for Dairy*, *Hohhot*, *Inner Mongolia Autonomous Region* 010010, *China*

ABSTRACT

Lactoferrin (LTF) is a pleiotropic milk protein with essential roles in mucosal immunity, antimicrobial defense, and iron homeostasis, yet the regulatory architecture underlying mammary-enriched *LTF* expression remains incompletely understood. In this study, multi-omics integration and comparative genomics were used to resolve the regulatory landscape, transcriptional control, and evolutionary diversification of mammalian *LTF*. Cross-tissue profiling revealed that *LTF* was highly expressed in glandular tissues, predominantly in the salivary and mammary glands. Single-cell transcriptomic analysis further localized *LTF* expression mainly to secretory luminal epithelial cells in human and bovine mammary tissues. Correlation analysis of 585 Holstein cows demonstrated that LTF expression was negatively associated with somatic cell score. Multi-omics analysis identified three functional enhancers, including an upstream enhancer harboring 20 cis-expression quantitative trait loci (eQTLs) and two evolutionarily conserved ruminant-specific enhancer elements. Functional validation demonstrated that EHF and ELF5 directly targeted these enhancers and regulated *LTF* transcription. Cross-species analysis also identified a human enhancer at chr3:46488833–46489420 that contributed to the exceptionally high *LTF* expression in humans and retained 3.11-fold enhancer activity in bovine MAC-T cells. By integrating transcription factor prediction, expression correlation with *LTF*, transcription factor abundance, and binding-site enrichment, a high-resolution cross-species regulatory network was constructed, comprising three enhancer elements and 39 conserved core transcription factors expressed across mammalian mammary tissues. These findings establish an evolutionarily diversified regulatory framework for lactation-specific *LTF* expression and identify candidate regulatory elements and transcription factors that may inform strategies to improve mammary gland health and increase LTF content in dairy cattle.

Keywords: Milk-derived cells; Enhancer elements; Genetic; Epigenomics; Cross-species; Lactoferrin gene expression regulation

INTRODUCTION

Milk is a major source of high-quality protein, calcium and bioactive molecules in the human diet and plays an indispensable role in both food production and nutritional health. Lactoferrin (LTF), a multifunctional iron-binding glycoprotein abundant in mammalian milk and mucosal secretions, is among the most biologically active whey proteins due to its broad antimicrobial, antiviral, immunomodulatory, and iron-regulatory functions (Agrizzi Verediano et al., 2023; Bullen et al., 1972; Farnaud & Evans, 2003; Legrand & Mazurier, 2010; Li et al., 2019). However, LTF abundance varies markedly among mammals. Human milk contains the highest reported concentrations, ranging from 1 to 7 mg/mL and accounting for approximately 20% of total breast milk protein, whereas bovine milk contains substantially lower concentrations, ranging from 0.02 to 0.35 mg/mL (Abedi-Firoozjah et al., 2025; Oberčkal et al., 2023; Rodríguez-Franco et al., 2005). This pronounced interspecies difference has driven the use of LTF supplementation in infant formula to better approximate the functional composition of human milk, particularly for neonatal infection prevention and improved iron absorption (Ochoa et al., 2012). In livestock production, dietary LTF supplementation has also been associated with enhanced immune performance (Shao et al.,

Received: 17 October 2025; Accepted: 11 December 2025; Online: 12 December 2025

Foundation items: This study was supported by the National Key Research and Development Program of China (2021YFF1001000) and National Center of Technology Innovation for Dairy (2024-JSGG-028)

[#]Authors contributed equally to this work

*Corresponding author, E-mail: wang_yu@nwafu.edu.cn

2021) and improved gut health (Xie et al., 2021). Given these nutritional, immunological, and therapeutic properties, LTF has emerged as an increasingly prominent focus of dairy biology, neonatal nutrition, and biomedical research.

Despite substantial progress in defining LTF function, the regulatory architecture that controls mammalian *LTF* expression remains poorly characterized. Previous studies have mapped genetic determinants of milk LTF concentration, identified candidate regulatory variants, and revealed associations between *LTF* polymorphisms and dairy traits (Lopdell et al., 2024; Mao et al., 2015; O'Halloran et al., 2010; Raschia et al., 2018; Teng, 1999; Viale et al., 2017). However, these studies have not systematically clarified the conserved and lineage-divergent cis-regulatory framework responsible for *LTF* expression. Resolving this regulatory framework requires integration of cross-species genomic conservation, chromatin accessibility, enhancer activity, cell-type-specific expression, and population-level genetic variation.

To address this gap, the present study integrated multi-omics data from seven mammalian species, including cattle and goats, with single-cell transcriptomics, ATAC-seq profiling, and cis-expression quantitative trait locus (cis-eQTL) analysis in 585 cows. This approach identified three *LTF* enhancers and 39 core transcription factors, including EHF and ELF5, and enabled construction of a high-resolution regulatory network for mammalian *LTF* expression. Comparative genomics revealed human-specific and ruminant-specific regulatory divergence, and dual-luciferase reporter assays validated enhancer activity. These findings define an evolutionarily informed regulatory framework for mammalian *LTF* expression and identify candidate molecular targets for genome editing-assisted dairy breeding and the development of nutritionally improved milk products.

MATERIALS AND METHODS

Sample collection and cell harvesting

A total of 585 Chinese Holstein dairy cows with similar parity and lactation stage were selected from a large-scale commercial dairy farm. All cows were clinically healthy, showed no evidence of mastitis or major metabolic disease, and were maintained under uniform feeding and management conditions with a total mixed ration (TMR) diet. All animals had complete pedigree records and were enrolled in Dairy Herd Improvement (DHI) testing. For each cow, 100 mL of milk was collected, immediately placed on ice, and transported to the laboratory for processing. Milk cells were isolated by centrifugation at 1 200 ×*g* for 15 min at 4°C, followed by careful removal of the upper fat layer and whey fraction. The cell pellet was resuspended in prechilled phosphate-buffered saline (PBS, Gibco, USA) and washed by centrifugation at 1 200 ×g for 10 min at 4°C. The pellet was then transferred to a 1.5 mL Eppendorf (EP) tube and washed twice under the same conditions. The final cell pellet, containing mammary epithelial cells and other somatic cells, was lysed with TRIzol reagent for total RNA extraction and transcriptome sequencing, with each sample requiring more than 6G of sequencing data.

To support cross-species and multi-omics analyses, a subset of cows (n=10) was randomly selected from the 585 cattle population. Additional milk samples were collected from 10 water buffaloes, 10 goats, 10 sheep, two red deer, and three healthy lactating human participants, and mammary gland tissues were collected from 10 mice. All samples were obtained during peak lactation. Samples from cows, water buffaloes, goats, sheep, and red deer were collected at 30–60 days postpartum; human milk samples were obtained at 3–6 months postpartum; and mouse mammary gland samples were collected on lactation day 10. Samples allocated for RNA sequencing were lysed using TRIzol reagent and immediately stored in liquid nitrogen before library construction and sequencing. The remaining samples were resuspended in cell freezing medium and cryopreserved in liquid nitrogen for assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq), cleavage under targets and tagmentation (CUT&Tag), single-cell RNA sequencing (scRNA-seq), and single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq).

All animal sampling procedures involving cows, buffaloes, goats, sheep, red deer and mice were reviewed and approved by the Animal Care Committee of Northwest A&F University (Ethics Approval XN-20230529). Human milk samples were collected from lactating women who provided informed consent, and all samples were anonymized prior to analysis.

Comparative analysis of tissue-specific *LTF* expression in humans and cattle

All statistical analyses and data visualizations were performed in R (v.4.1.2). Human tissue expression data for *LTF* were obtained from the adult GTEx dataset (https://www.gtexportal.org/) (Supplementary Table S1), whereas corresponding bovine data were retrieved from the cGTEx dataset (https://cgtex.roslin.ed.ac.uk/) (Supplementary Table S2). Expression values were $\log_2$ transformed, and tissue-specific distributions of *LTF* expression were visualized as boxplots using ggplot2 (v.4.0.0) (Chen et al., 2022).

Single-cell analysis of *LTF* expression in humans and cattle

The human milk gene expression matrix was obtained from Nyquist et al. (2022) and converted into a Seurat object using CreateSeuratObject (Tan et al., 2023). Raw scRNA-seq data from bovine mammary tissue were processed with Cell Ranger and aligned to the ARS-UCD1.2 reference genome (Xu & Jia, 2021). Cells with fewer than 200 or more than 2 500 detected genes, or those with mitochondrial transcripts accounting for more than 5% of total transcripts, were excluded (Osorio & Cai, 2021). Gene expression data were normalized using NormalizeData with a scale factor of 10 000 (Lytal et al., 2020). A total of 2 000 highly variable genes were identified, and the expression matrix was scaled using ScaleData to remove unwanted sources of variation (Tsuyuzaki et al., 2020). Principal component analysis (PCA) was then performed, and the top principal components were used for unsupervised clustering. Cell clusters were identified using the FindClusters function at a resolution of 0.3. Uniform manifold approximation and projection (UMAP) was applied for dimensionality reduction and two-dimensional (2D) visualization.

Following clustering, distinct cell populations were identified according to established lineage-specific marker genes. Secretory luminal epithelial cells were characterized by high expression of *ELF5*, *KRT15*, *COBL*, and *CCL28*; hormone-responsive luminal epithelial cells by *PRLR*, *ESR1*, *ERBB4*, and *TTC6*; basal cells by *KRT17*, *KRT5*, *ACTA2*, and *TP63*; fibroblasts by *DCN*, *APOD*, *COL1A2*, and *COL1A1*; macrophages by *C1QB*, *CSF1R*, *MRC1*, and *CD163*; T cells by *PTPRC*, *CD3E*, *CD44*, and *IL7R*; lymphatic endothelial cells by *MMRN1*, *CAVIN2*, *LYVE1*, and *PROX1*; and vascular

endothelial cells by *PECAM1*, *ENG*, *FABP4*, and *RGCC*. These canonical markers were used to assign clusters to the major cell populations represented in mammary tissue. Expression of *LTF* across annotated cell populations was visualized using DotPlot (Tan et al., 2023). All single-cell analyses were performed using Seurat (v.4.0.3) in R (v.4.1.2), and figures were generated with ggplot2 (Wu et al., 2022).

Correlation analysis of *LTF* expression and lactation traits

Raw RNA-seq reads were quality filtered using TrimQC (v.0.11.5), and the retained reads were aligned to the ARS-UCD1.2 reference genome using HISAT2 (v.2.0.1) to generate BAM files (Gill & Dhillon, 2022). Alignment files were sorted using SAMtools (v.1.22.1) (Li et al., 2009), and gene-level read counts were obtained using featureCounts. Transcript abundance was subsequently expressed as transcripts per million (TPM) (Li et al., 2023).

Outlying Dairy Herd Improvement (DHI) records were excluded using the three-standard-deviation method (3σ principle) (Li et al., 2017). Phenotypic measurements and TPM values for *LTF* were $\log_2$-transformed to improve distributional properties (Zhao et al., 2021). Associations between *LTF* expression and milk yield, milk fat percentage, milk protein percentage, and somatic cell count were evaluated separately using linear regression. Each lactation trait was modeled as the dependent variable, *LTF* expression was included as the explanatory variable, and days in milk and parity were fitted as fixed-effect covariates. Pearson correlation coefficients between *LTF* expression and individual lactation traits were also calculated and visualized using ggplot2 (v.4.0.0) in R (Hao et al., 2022).

Cows were subsequently classified into high-expression (TPM>500; n=17) and low-expression (TPM<40; n=20) groups according to *LTF* transcript abundance. Differential expression analysis was performed using gene-level count data using DESeq2 (v.1.46.0). Differentially expressed genes (DEGs) were defined by a Benjamini-Hochberg-adjusted P<0.01 and an absolute fold change>2 (Love et al., 2014). Gene identifiers were annotated using the org.Bt.eg.db package, followed by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses. Terms and pathways with an adjusted P<0.05 were considered significantly enriched (Li et al., 2020).

Identification and functional annotation of bovine *LTF* cis-eQTLs

Genotypes were inferred from the RNA-seq data and imputed against a reference panel constructed from whole-genome sequencing (WGS) data for more than 3 000 cattle (Zhang et al., 2023), including data from the 1000 Bull Genomes Project (Hayes & Daetwyler, 2019). Reference haplotypes were phased using Beagle (v.5.4) (Browning et al., 2018), and genotype imputation for the 585 lactating cows was performed using GLIMPSE2 (Rubinacci et al., 2023). Imputed variants were subjected to stringent quality control, and only sites with INFO scores greater than 0.99 were retained. Variants located within 100 kb upstream and downstream of the *LTF* boundaries were extracted for subsequent cis-eQTL mapping.

Expression values for *LTF* were normalized using the trimmed mean of M-values (TMM) method implemented in edgeR (v.3.21) (Robinson & Oshlack, 2010), followed by inverse normal transformation. Cis-eQTL mapping was performed using TensorQTL (v.1.0.3) (Taylor-Weiner et al., 2019), with the cis-regulatory interval defined as the *LTF* gene body and 100 kb flanking regions on both sides. Five genotype principal components and five expression principal components were included as covariates to account for population structure and latent expression-related confounding (Zhou et al., 2022).

Statistical significance was determined using the permutation-based multiple-testing procedure implemented in TensorQTL (Aguet et al., 2023). Adaptive permutations were first used to calculate empirical P-values for all variants within the LTF cis-window. The resulting empirical P-values were then adjusted using the Benjamini-Hochberg correction procedure, and the empirical threshold corresponding to a genome-wide false discovery rate (FDR) of 0.05 was used to derive the nominal significance threshold. Variants with a nominal P-value below this threshold were classified as significant *LTF* cis-eQTLs. Results were integrated into a comprehensive dataset, and cis-eQTL signals were explored using visualization tools (e.g., BED files), combined with epigenetic data to identify regulatory elements associated with *LTF*. Linkage disequilibrium (LD) among significant cis-eQTLs was estimated and visualized using LDBlockShow (Dong et al., 2020).

ATAC-seq library preparation and data processing

Cells were lysed on ice for 5–15 min and centrifuged at 500 ×*g* for 5 min at 4°C. After removal of the supernatant, the pellet was resuspended in 1 mL of cold Dulbecco's PBS (Gibco, USA), filtered through a 40 μm cell strainer, and transferred to a 1.5 mL EP tube. Cell counting was assessed by trypan blue staining (Thermo Fisher Scientific, USA). For each reaction, 50 000 nuclei were incubated with Tn5 transposase at 37°C for 1 h. Libraries were purified, amplified by polymerase chain reaction (PCR), quantified using a Qubit fluorometer (Thermo Fisher Scientific, USA), and evaluated with an Agilent 2100 Bioanalyzer (Agilent Technologies, USA) (USA; Li et al., 2023). Raw sequencing reads were processed using Trimmomatic (v.0.36) and FastQC (v.0.11.5), with adapters and low-quality reads removed (Dahiya et al., 2023). Clean reads were aligned to the ARS-UCD1.2 reference genome using Bowtie2 (v.2.5.4) (Carter et al., 2023). Overall alignment rates exceeded 90%, and unique mapping rates ranged from 43% to 68% across samples. BAM files were sorted and deduplicated using SAMtools (v.1.22.1) and Picard (v.3.4.0), respectively (Kim et al., 2019). Reads with mapping quality scores <30, improperly paired reads, duplicate reads, or reads aligned to the mitochondrial genome were filtered out. Accessible chromatin regions were identified using MACS2 (v.2.1.0) with a significance threshold of P=10^{-5} (Wu et al., 2023). Data quality was further evaluated using the fraction of reads in peaks (FRiP), which ranged from 49% to 66% across samples.

Cross-species identification of conserved *LTF* regulatory elements and inference of chromatin interactions

ATAC-seq peaks located within 100 kb upstream or downstream of *LTF* were identified in cattle (*Bos taurus)*, water buffaloes (*Bubalus bubalis*), goats (*Capra hircus*), sheep (*Ovis aries*), red deer (*Cervus canadensis*), humans (*Homo sapiens*), and mice (*Mus musculus*). Peaks were annotated using ChIPseeker (v.1.42.0) (Yu et al., 2015), with promoter regions defined as 3 kb upstream and downstream of the transcription start site, and chromatin accessibility profiles were visualized using Integrative Genomics Viewer (IGV) (v.2.19.1) (Chokeshaiusaha et al., 2018). Bovine (*Bos taurus*) regulatory sequences were used as references to assess cross-species sequence similarity with BLASTN (Yoon et al.,

2017). BLASTN analysis was performed using an *E*-value cutoff of 1×10^{-5}, a word size of 7, match/mismatch scores of 1/−1, and gap opening and extension penalties of 5. Evolutionary conservation across candidate regulatory regions was evaluated using PhastCons and PhyloP scores derived from multispecies whole-genome alignments. Conservation scores for cattle (*Bos taurus*, ARS-UCD1.2) were sourced from the 110-way Multiz Align track, comprising 78 ruminant and 32 outgroup species, in the Animal Omics Database of Northwest A&F University (http://animal.omics.pro/code/index.php/RGD). Human and mouse conservation scores were obtained from the 100 Vertebrates conservation track (multiz alignment of 100 species) in the UCSC Genome Browser (Benegas et al., 2023).

Cellular heterogeneity within bovine mammary tissue was resolved by unsupervised analysis of scATAC-seq data. High-quality scATAC-seq profiles were subjected to iterative latent semantic indexing (LSI) for dimensionality reduction and graph-based clustering using ArchR (v.1.0.3), which partitioned nuclei according to genome-wide chromatin accessibility patterns (Granja et al., 2021; Ranzoni et al., 2021).

Cell identities were assigned by transferring annotations from the corresponding scRNA-seq dataset to scATAC-seq clusters using promoter accessibility profiles. Within the secretory luminal epithelial cells, three chromatin-defined substates (LumSec state 1, state 2, and state 3) were distinguished based on subtle differences in chromatin accessibility patterns (Stuart et al., 2019; Xu et al., 2021). Potential regulatory connections among accessible regions were subsequently inferred through co-accessibility analysis in ArchR (Granja et al., 2021).

CUT&Tag library preparation and data processing

CUT&Tag profiling was performed using primary antibodies against histone H3 lysine 27 acetylation (H3K27ac), CCCTC-binding factor (CTCF), and signal transducer and activator of transcription 5 (STAT5). The workflow comprised cell collection, antibody incubation, transposase-mediated tagmentation, PCR amplification, and library sequencing (Chen, 2023; Zhang et al., 2025). Raw reads were quality-controlled using fastp (v.0.23.5), and clean reads were aligned to the ARS-UCD1.2 reference genome using Bowtie2 (v.2.5.4) to generate SAM files (Langdon, 2015). These files were converted to BAM format, and unmapped reads were removed using SAMtools (v.1.22.1) (Li & Durbin, 2009). BAM files were subsequently sorted and indexed using SAMtools, and used to generate coverage tracks using bamCoverage in deepTools for subsequent visualization (v.3.5.6) (Ding et al., 2022). Sample correlations matrix was generated using multiBigwigSummary. Biological replicates were merged before peak calling with MACS2 (P<0.05), and significant regions were exported in narrowPeak format. BAM files were converted to BED format using BEDTools (v.2.31.1), fragments longer than 1 000 bp were filtered out, and the remaining fragments were sorted by chromosomal position for downstream peak analysis (Quinlan & Hall, 2010).

Identification of *LTF*-associated co-expression modules by weighted gene co-expression network analysis (WGCNA)

Genes co-expressed with *LTF* were identified through WGCNA using the WGCNA R package (v.1.73) (Tan et al., 2023). Briefly, genes were filtered based on median absolute deviation (MAD) > 0.01. Sample and gene quality were checked using the goodSamplesGenes function. A soft-thresholding power of 12 was selected using the pickSoftThreshold function to approximate scale-free network topology. Network construction and module identification were performed using blockwiseModules with the following parameters: networkType= “unsigned”, minModuleSize=30, and mergeCutHeight=0.25. Genes assigned to the same co-expression module as *LTF* were designated members of the *LTF*-associated module.

Transcription factor annotation and regulatory network construction

Transcription factors were annotated using AnimalTFDB (v.4.0) (Shen et al., 2023), and enriched DNA-binding motifs were identified with MEME (Bailey et al., 2009). Predicted regulatory relationships between transcription factors and *LTF* were integrated and visualized as a network using Cytoscape (v.3.10.0) (Wu et al., 2024).

Dual-luciferase reporter assay

Dual-luciferase reporter assays were performed to evaluate enhancer activity and transcription factor-dependent regulation. Candidate enhancer sequences were first inserted into the pGL4.17[luc2/Neo] plasmid, whereas transcription factor coding sequences were cloned into the pcDNA3.1 plasmid. HEK-293T cells and bovine mammary epithelial MAC-T cells were obtained from Shanghai Sixin Biotechnology Co., Ltd. (China). Cells were seeded into 24-well plates at a density of 5×10^5 cells per well and cultured in Dulbecco’s modified Eagle medium (Gibco, USA) containing 10% fetal bovine serum (FBS, Gibco, USA). Transfection was performed using Lipofectamine™ 3000 Transfection Reagent (Thermo Fisher Scientific, CN) when HEK-293T and MAC-T cultures reached approximately 60% and 70% confluence, respectively. Before transfection, cells were starved in Opti-MEM for 1 h and then transferred to complete medium containing 5% fetal bovine serum (FBS) **()**. Each well received 0.4 μg of pGL4.17, 0.1 μg of pRL-TK (Promega, USA), and 0.4 μg of EHF-pcDNA3.1. After transfection, cells were incubated at room temperature for 10 min and then at 37°C for 8 h. The medium was subsequently replaced, and cells were cultured for an additional 48 hours. At 48 hours post-transfection, cells were lysed, and firefly and Renilla luciferase activities were measured using a TransDetect Double-Luciferase Reporter Assay Kit (TransGen Biotech, China) on a GloMax 20/20 Luminometer (Promega, USA). Firefly luciferase activity was normalized to Renilla luciferase activity for each sample (Beijing Quanshi Jin Bio Co., Ltd., China) (Wang et al., 2024).

Statistical analysis

All data are expressed as mean±standard deviation (SD). Statistical analyses were performed using GraphPad Prism 8 (Chang et al., 2023). Student’s *t*-test, analysis of variance (ANOVA), Wilcoxon rank-sum test, log-rank test, and Pearson correlation analysis were applied as appropriate for the experimental design. Statistical significance was defined as P<0.05. Significance levels were denoted as follows: ns, not significant; *: P<0.05; **: P<0.01; ***: P<0.001; and ****: P<0.0001 (Zhang et al., 2025).

RESULTS

Enrichment of *LTF* expression in luminal epithelial cells of the mammary gland

Cross-tissue analysis of human GTEx and bovine cGTEx

datasets revealed pronounced tissue-dependent variation in *LTF* expression under physiological conditions. In humans, *LTF* expression was highest in the mammary gland, salivary gland, and blood compared with other tissues (Figure 1A). A similar gland-enriched pattern was observed in cattle, with predominant expression in the mammary gland, nasal mucosa, and salivary gland and substantially lower expression in nonglandular tissues, including the tibial artery, brain, spleen, and kidney (Figure 1B).

Cell-type-specific expression of *LTF* was subsequently resolved using scRNA-seq data from human milk and bovine mammary tissue. Based on UMAP clustering and established markers for mammary epithelial and immune cell populations (Nguyen et al., 2018; Nyquist et al., 2022), cells were identified as hormone-responsive luminal epithelial cells (LumHR), secretory luminal epithelial cells (LumSec), T cells, macrophages, and neutrophils. In human milk, *LTF* expression was highest in LumSec cells (Figure 1C), followed by LumHR cells, while its expression in immune cell types was almost undetectable (Figure 1C; Supplementary Figure S1A). Bovine mammary tissue also showed significantly higher *LTF* expression in LumSec cells than in other bovine cell populations (Figure 1D). Nevertheless, *LTF* abundance in all bovine cell types (including LumSec) was markedly lower than that observed in the corresponding human LumSec population. Notably, expression in bovine LumHR cells, lymphatic endothelial cells, macrophages, T cells, vascular endothelial cells, fibroblasts, and basal cells was minimal or near background levels (Figure 1D; Supplementary Figure S1B).

Overall, these findings identified a conserved cell-type-specific pattern of *LTF* expression in humans and cattle, characterized by predominant enrichment in LumSec cells. This shared localization supports a conserved role for LTF in mammary secretion during lactation.

Association of *LTF* expression with lactation traits and mammary immune status

Matched transcriptomic data and DHI records from 585 Holstein cows were analyzed to determine whether *LTF* expression was associated with lactation performance or mammary health. Linear regression models were applied to evaluate the relationships between *LTF* expression and various traits after correcting for the fixed effects of milk production days and parity. Somatic cell count was converted to somatic cell score (SCS) to reduce distributional skewness and improve conformity with model assumptions. After adjustment for these covariates, *LTF* expression showed no significant association with milk yield, milk fat percentage, or milk protein percentage (all $P>0.05$; Figure 2A). In contrast, *LTF* expression was significantly negatively associated with SCS (regression coefficient=−0.348, $P=3e-04$), indicating that higher *LTF* expression is associated with lower mammary inflammatory burden. This relationship is consistent with established immunoregulatory functions of LTF (Xu et al., 2017; Zabolewicz et al., 2014).

To examine transcriptional differences associated with extreme *LTF* expression, cows were stratified into high-expression (TPM>500, $n=17$) and low-expression (TPM<40, $n=20$) groups. Somatic cell count was significantly lower in the high-expression group ($P<0.01$), consistent with the negative correlation between *LTF* expression and SCS observed in the full cohort ($r=-0.23$; Supplementary Figure S2F). Differential expression analysis identified 196 up-regulated genes in the high-expression group, including *KRT17*, *KRT15*, *KRT8*, *KRT5*, *CALB1*, and *CALCB* (Figure 2C). These genes were enriched in pathways associated with defense against *Staphylococcus aureus* infection, estrogen signaling, and calcium reabsorption (Figure 2D; Supplementary Tables S3, S4) and showed preferential expression in basal and LumSec cell populations (Supplementary Figure S1B).

A further 1 229 genes were down-regulated in the low-expression group (Figure 2C) and were predominantly associated with immune and inflammatory pathways, including cytokine-cytokine receptor interaction and inflammatory mediator regulation (Figure 2E; Supplementary Tables S5, S6). Reduced expression of these genes, including *LRRK2*, *IL1RAP*, and *CCR3*, may lead to decreased mammary defense capacity and consequently higher SCS. The absence of significant associations between *LTF* expression and milk yield, milk fat percentage, or milk protein percentage suggests that *LTF* expression in the bovine mammary gland is more closely related to mammary immune status than to production traits. Its association with immune defense-related transcriptional programs and lower somatic cell burden supports further evaluation of *LTF* expression as a candidate molecular marker of mastitis resistance in dairy cattle.

Genetic determinants of *LTF* expression

Cis-eQTLs can alter transcription by modulating the activity of nearby regulatory elements (Brown et al., 2013). Integration of cis-eQTL mapping with ATAC-seq profiles from 585 dairy cows identified 481 cis-eQTLs within 100 kb upstream and downstream of *LTF* (Figure 3A). A prominent accessible chromatin region located 4 660 bp upstream of *LTF* (chr22:52946414–52947911) contained 20 cis-eQTLs, including four previously reported variants (Lopdell et al., 2024). The enrichment of cis-eQTLs within this accessible upstream element supports its role as a candidate regulator of LTF expression in dairy cattle. 20 cis-eQTLs were distributed across two strong LD blocks (D'=1, $r^2>0.85$). Among the 20 significant cis-eQTLs, 17 were located within two LD blocks (block 1: 5 variants; block 2: 12 variants), whereas the remaining three were isolated eQTLs not belonging to any LD block (Figure 3B). Genotype-based stratification classified the 585 cows into homozygous reference ($n=461$), heterozygous ($n=117$), and homozygous alternate ($n=7$) groups, with the homozygous reference group exhibiting significantly higher *LTF* expression (Figure 3C; Supplementary Figure S3A and Table S7). Dual-luciferase reporter assays further showed that the reference sequence produced 1.80-fold greater enhancer activity than the alternate sequence ($P<0.001$; Figure 3D). Collectively, these results indicate that variants within this enhancer alter regulatory activity and thereby contribute to variation in *LTF* expression.

Conserved ruminant regulatory elements controlling *LTF* expression

To elucidate the evolutionary conservation of cis-regulatory elements surrounding *LTF* in ruminants, ATAC-seq profiles from seven mammalian species were compared: cattle, water buffalo, goat, sheep, red deer, human and mouse. Open chromatin regions within 100 kb upstream and downstream of *LTF* were examined to identify candidate promoters and enhancers (Figure 4A). Given that enhancers are typically located in intronic or intergenic regions (Panigrahi & O'Malley, 2021), multi-species alignment using the cattle genome as a reference identified two open chromatin peaks, designated E1

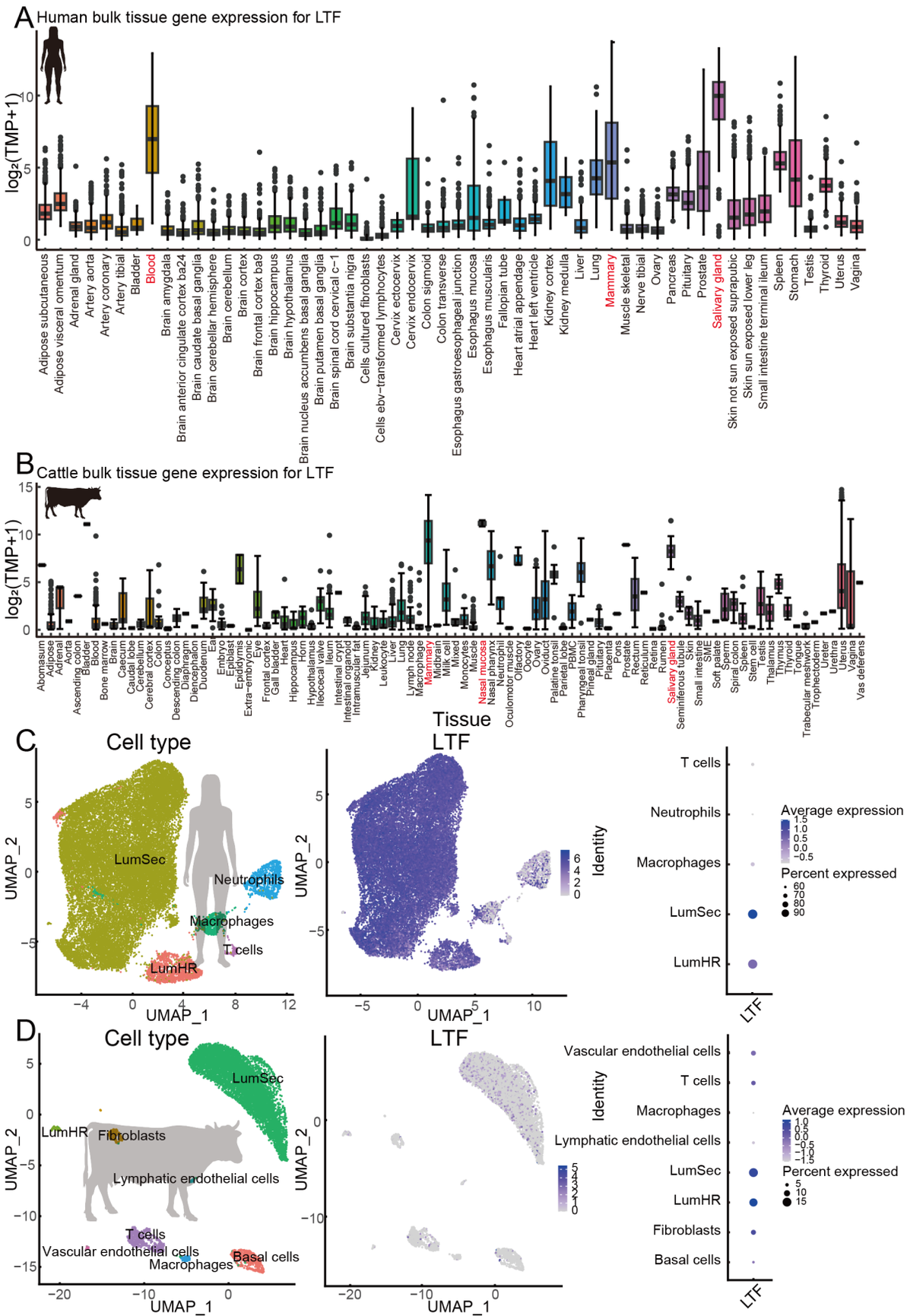

Figure 1 Tissue- and cell-type-specific expression of *LTF*

A: Distribution of *LTF* expression across adult human tissues. B: Distribution of *LTF* expression across bovine tissues. Red shading indicates tissues with the highest *LTF* expression in both species. C, D: Uniform manifold approximation and projection (UMAP) of cells from human milk and bovine mammary tissue, respectively, showing the major annotated cell populations. Heatmaps show *LTF* expression across cell clusters, with color intensity indicating expression level. Dot plots show the detection frequency and mean expression levels of *LTF* among different cell types.

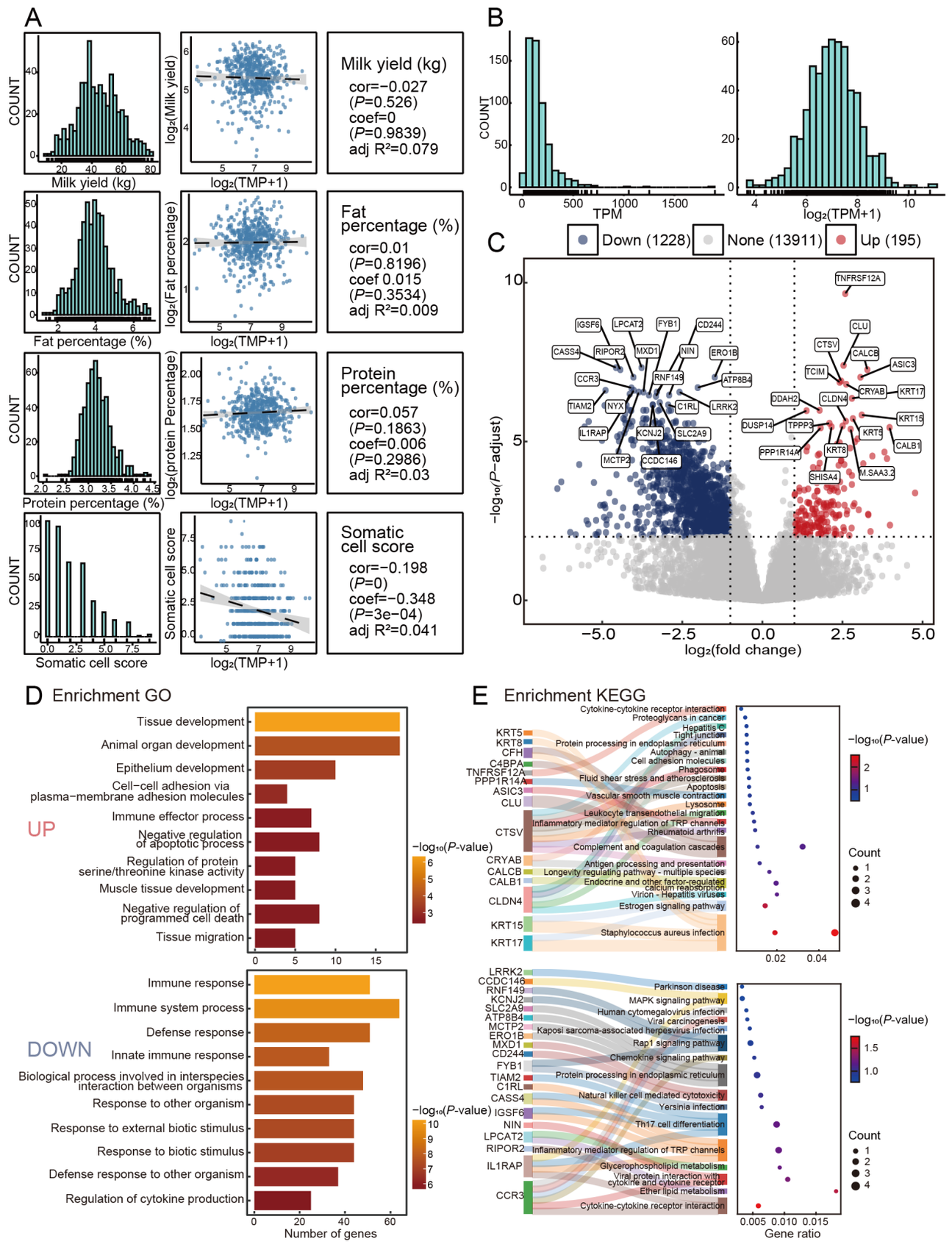

Figure 2 Potential factors influencing *LTF* expression

A: Linear associations between *LTF* expression and milk yield, milk fat percentage, milk protein percentage, and SCS after adjusting for days in milk and parity. Scatter points represent raw data, and shaded areas indicate 95% confidence intervals of the regression line. Regression coefficients and corresponding *P* values are shown. B: Distribution of raw *LTF* expression values across samples before and after log2 transformation. C: Volcano plot showing differentially expressed genes between cows with high and low *LTF* expression. “Up” and “down” indicate the top 20 significantly up-regulated and downregulated genes, respectively, whereas “none” denotes genes without significant differential expression. D, E: GO and KEGG enrichment analyses of up-regulated and down-regulated genes, respectively.

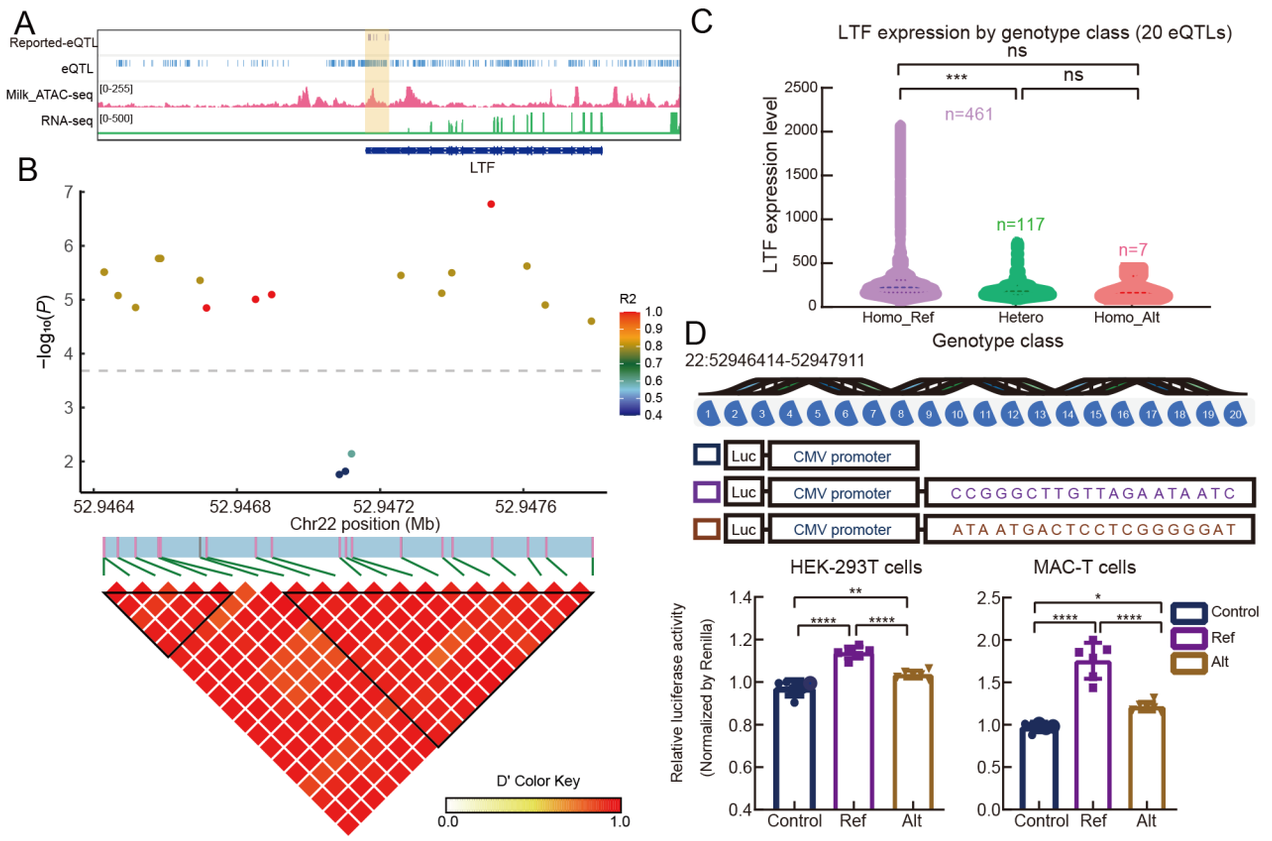

Figure 3 Identification of an upstream cis-regulatory enhancer containing LD blocks associated with bovine *LTF* expression

A: Distribution of cis-eQTLs within 100 kb upstream and downstream of *LTF*, including six previously reported variants. Bovine ATAC-seq and RNA-seq profiles are displayed in Integrative Genomics Viewer (IGV), with yellow highlighted regions marking identified regulatory elements. B: Manhattan plot of 20 cis-eQTLs in the upstream enhancer region. X-axis indicates genomic positions, y-axis shows −log-transformed *P* values, and color intensity reflects linkage disequilibrium (LD) strength (r^2). Corresponding heatmap shows pairwise D' values, where each square denotes LD level between two cis-eQTLs. C: Violin plots showing *LTF* expression according to genotype. Midline indicates median, upper/lower lines represent Q3 and Q1 respectively, with genotypes classified as homozygous reference (Homo_Ref), heterozygous (Hetero), and homozygous alternate (Homo_Alt). D: Dual-luciferase assay using enhancer elements containing 20 cis-eQTLs (vector backbone: pGL4.17 with a CMV promoter). HEK-293T, human embryonic kidney cells; MAC-T: bovine mammary epithelial cell line; Ref, reference; Alt, alternate.

(22:52981819–52982561; 742 bp), located within a downstream intronic region, and E2 (22:52988484–52989210; 726 bp), located within an intergenic region. These elements were positioned 29.2 kb and 35.9 kb from *LTF*, respectively, and were conserved across cattle, water buffalo, goats, sheep, and red deer but were not conserved in humans or mice (Figure 4A). Because regulatory function is often associated with sequence conservation (Pennacchio & Visel, 2010), the two regions were examined further by comparative sequence analysis. Both elements showed >90% sequence identity and 100% coverage in water buffalo, goats, sheep, and red deer relative to the bovine reference (Supplementary Table S8). Multispecies alignment in the UCSC Genome Browser and elevated PhastCons and PhyloP scores provided additional evidence of ruminant-specific conservation (Figure 4B), indicating that these regions likely play critical roles in regulating LTF transcription in these species.

Chromatin accessibility and regulatory connectivity were further examined by scATAC-seq in lactating goat mammary tissue (Figure 4C). Chromatin accessibility at E1 and E2 was significantly enriched in secretory luminal epithelial cells, including LumSec states 1, 2, and 3. Co-accessibility analysis linked E1 and E2 to one another, to the *LTF* promoter, and to additional upstream regulatory regions associated with *LTF*, supporting coordinated regulation of *LTF* expression.

Integration of milk RNA-seq data from five ruminant species with transcription factor motif prediction identified 20 highly expressed transcription factors with predicted conserved binding sites in E1 and E2 (Supplementary Figure S4A, B). Among these candidates, conserved EHF-binding motifs were identified in both enhancers (Figure 6C), supporting a potential regulatory role for EHF. Dual-luciferase reporter assays confirmed higher reporter activity relative to the control, with E1 producing a greater increase than E2 (1.15-fold vs. 1.08-fold, $P<0.001$). Co-transfection with an EHF expression vector further enhanced reporter activity for E1 and E2 by 1.12-fold and 1.06-fold, respectively (both $P<0.01$; Figure 4D). Together, these findings identify E1 and E2 as conserved ruminant enhancers that are accessible in secretory luminal epithelial cells and responsive to EHF, supporting coordinated regulation of *LTF* transcription during lactation.

Transcriptional basis of exceptionally high *LTF* expression in humans

Cross-species expression profiling demonstrated that *LTF* expression in humans (TPM=12 248.5) was substantially higher than that in 11 other mammalian species (Figure 5A), paralleling the markedly greater concentration of LTF protein in human milk. This concordance suggests that species-specific transcriptional regulation contributes to the exceptionally high abundance of *LTF* in human milk.

To elucidate the regulatory elements associated with high

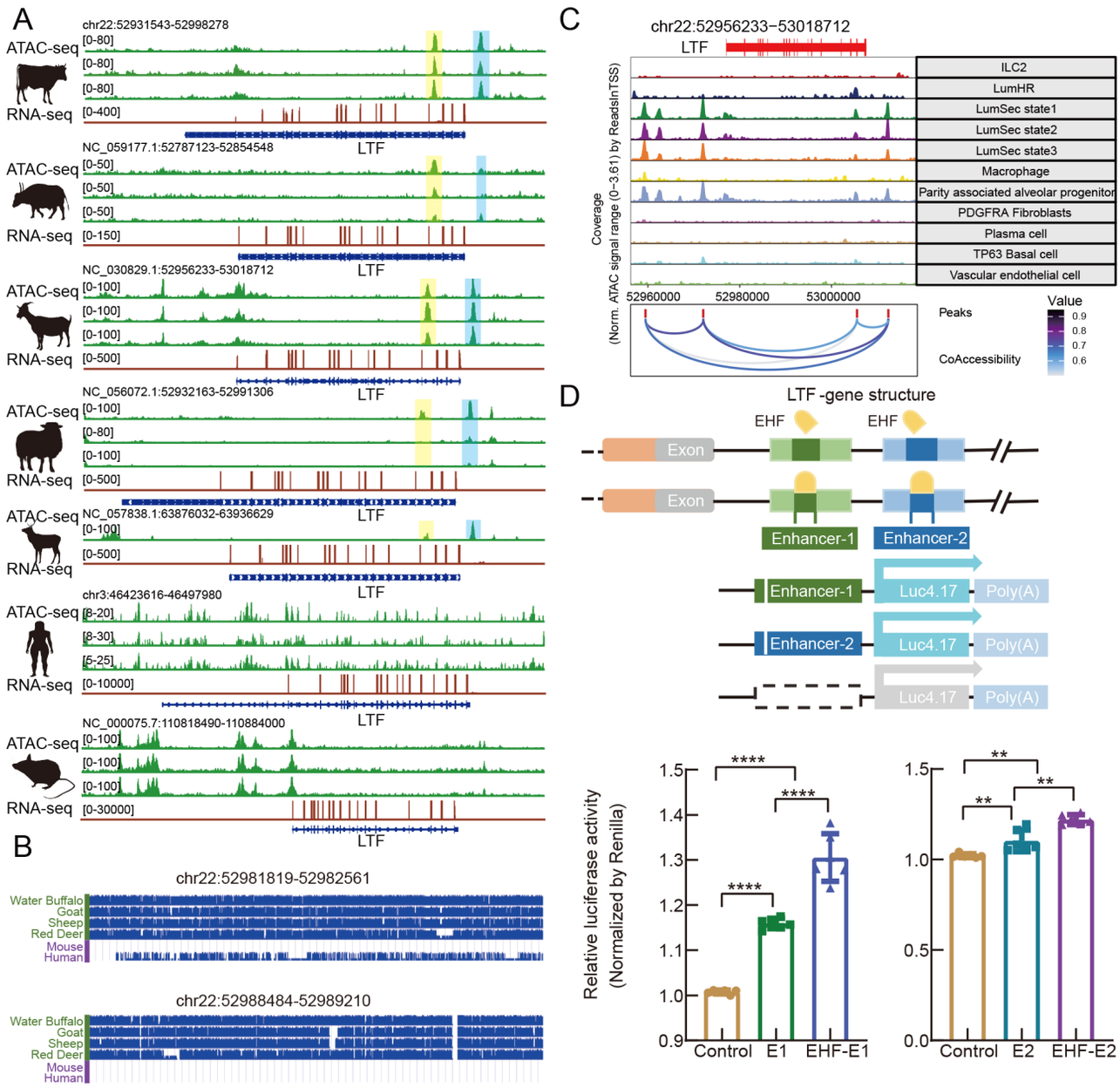

Figure 4 EHF drives *LTF* expression by two conserved regulatory elements in ruminants

A: Comparative analysis of *LTF* loci in cattle, water buffalo, goat, sheep, red deer, human, and mouse. ATAC profiles from three samples and *LTF* expression from one sample are shown for each species. Yellow and blue shaded regions indicate the conserved enhancer elements E1 and E2, respectively, and dark blue icons represent *LTF* transcripts in each species. B: Sequence conservation of E1 and E2 across ruminants. C: Co-accessibility analysis of scATAC-seq data from lactating goat mammary tissue. Arcs connect the *LTF* promoter, the upstream enhancer, E1, and E2, with darker shading indicating stronger co-accessibility. D: Dual-luciferase reporter assay of E1 and E2 and activation by EHF. E1 and E2 were cloned into pGL4.17, and EHF was expressed from pcDNA3.1. Experimental groups comprised control, E1 alone, E2 alone, EHF+E1 co-transfection (EHF+E1), and EHF+E2 co-transfection (EHF+E2). MAC-T, bovine mammary epithelial cells.

LTF expression in humans, accessible chromatin regions within 100 kb upstream and downstream of the human LTF locus were examined using ATAC-seq data from human milk cells. Integration with mouse epigenomic profiles, including ATAC-seq and ChIP-seq data for H3K27ac, H3K4me1, H3K4me3, STAT5, glucocorticoid receptor (GR), mediator complex subunit 1 (MED1), and RNA polymerase II, identified a candidate enhancer at chr3:46488833-46489420 with a positionally corresponding regulatory region in the mouse genome (Figure 5B). However, despite positional conservation, sequence alignment revealed limited conservation, with 88.89% identity across only part of the region and the predicted ELF5-binding motif incompletely conserved in the corresponding mouse, cattle, and goat sequences (Figure 5C), indicating lineage-specific divergence within this enhancer region.

Dual-luciferase reporter assays were used to evaluate the activity of the human enhancer in HEK-293T and bovine MAC-T cells. In HEK-293T cells, the enhancer increased activity by 1.25-fold relative to the control (P<0.001), with ELF5 co-expression producing a further 1.26-fold increase (P<0.001). In MAC-T cells, the enhancer increased reporter activity by 3.11-fold, whereas ELF5 co-transfection induced a 1.22-fold increase (P<0.001; Figure 5D). Thus, the human enhancer remained functional in bovine mammary epithelial cells and responded to ELF5 despite limited sequence conservation

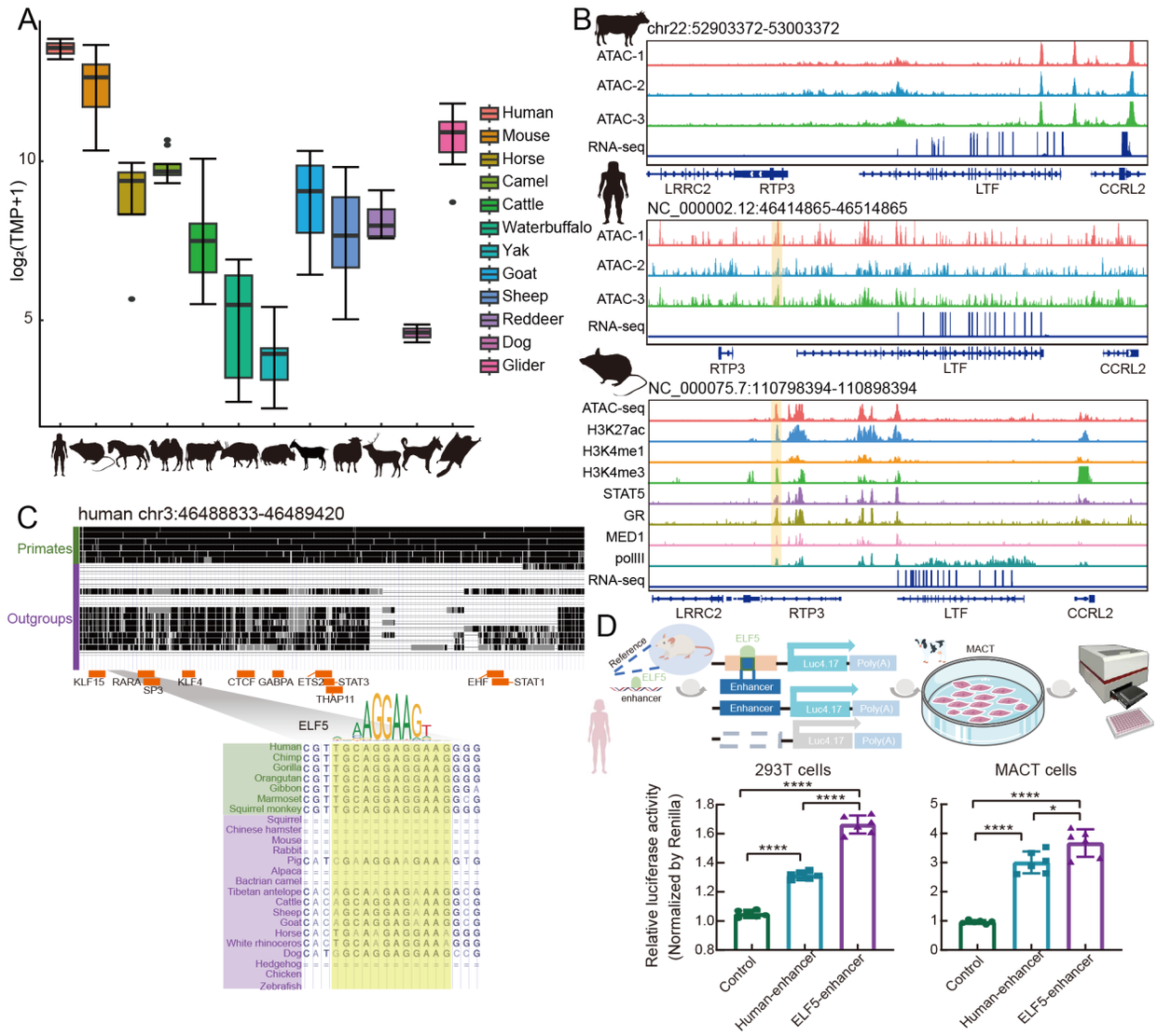

Figure 5 Identification and functional characterization of a human *LTF* enhancer

A: Distribution of *LTF* expression across 12 mammalian species. Higher median values indicate greater transcript abundance. B: ATAC-seq profiles across regions located within 100 kb upstream and downstream of the bovine and human *LTF* loci, alongside mouse ATAC-seq and ChIP-seq data. Yellow shaded regions indicate candidate regulatory elements specifically conserved between mice and humans but absent in bovines. C: Predicted transcription factors and corresponding core binding motifs within the candidate enhancer. D: Dual-luciferase reporter analysis of the candidate human enhancer identified by integration with mouse epigenomic data. Enhancer was cloned into pGL4.17, and ELF5 was expressed from pcDNA3.1. Reporter activity was measured in HEK-293T and MAC-T cells transfected with vector control, human enhancer alone, or human enhancer together with ELF5. HEK-293T, human embryonic kidney cells; MAC-T, bovine mammary epithelial cells.

across species. These findings identify a species-divergent cis-regulatory element that may contribute to exceptionally high *LTF* expression in humans and provide a candidate sequence for future regulatory element humanization studies.

Construction of a cross-species regulatory network for *LTF* expression

Integration of genetic, epigenomic, and functional data identified three candidate enhancers within 100 kb upstream and downstream of *LTF* (Figure 6A). The upstream enhancer at chr22:52946414–52947911 contained multiple cis-eQTLs and coincident epigenetic signals, linking local genetic variation to regulation of *LTF* expression. The two additional enhancers, E1 at chr22:52981819–52982561 and E2 at chr22:52988484–52989210, were located downstream of *LTF* and showed strong conservation across ruminants, supporting conserved roles in the regulation of *LTF* expression in these species. Dual-luciferase assays confirmed enhancer activity for all three elements.

Motif analysis identified 125, 27, and 135 predicted transcription factor-binding motifs within the upstream enhancer, E1, and E2, respectively. Six transcription factors (EHF, IRF2, IRF9, KLF5, RARA, and STAT2) were predicted to bind all three enhancers (Figure 6B). Because sequence-specific recognition of cis-regulatory motifs is central to transcriptional control (Smith & Matthews, 2016), the core binding motifs within each enhancer were further examined (Figure 6C). Variants within the upstream enhancer overlapped a core ELF5-binding motif, supporting potential allele-dependent regulation of *LTF* expression. E1 and E2 both contained predicted EHF-binding motifs but differed in the surrounding regulatory sequences, indicating distinct regulatory properties within a conserved ruminant enhancer system.

Predicted transcription factors associated with the three

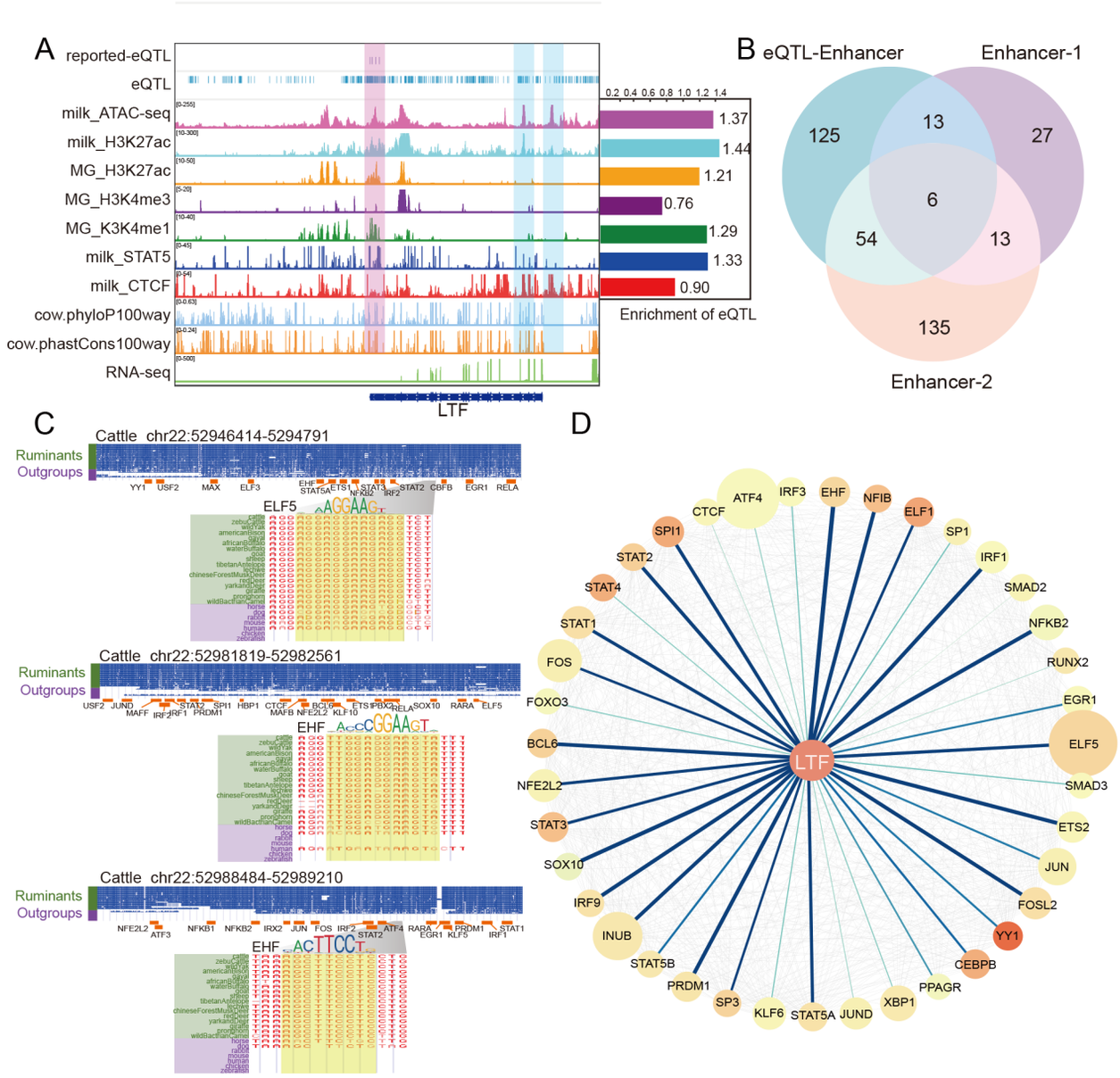

Figure 6 Construction of a regulatory network for *LTF* expression

A: Integrated view of genetic and epigenomic features within 100 kb upstream and downstream of *LTF*, including cis-eQTLs, ATAC-seq profiles, CUT&Tag profiles for H3K27ac, CTCF, and STAT5, and ChIP-seq profiles for H3K27ac, H3K4me3, and H3K4me1. Rectangles indicate enrichment of cis-eQTLs within accessible chromatin regions; values above and below 0 denote positive and negative associations, respectively. B: Venn diagram showing transcription factors predicted to bind each of the three regulatory elements and factors shared among the elements. C: Core transcription factor-binding motifs identified within the three regulatory elements. D: Correlations between *LTF* expression and transcription factor genes conserved across species. Line thickness and color intensity represent correlation strength, with thicker and darker lines indicating stronger correlations. Node size reflects gene expression levels, and node color indicates enrichment of the corresponding binding motif.

enhancer elements were subsequently integrated with *LTF*-associated factors identified through WGCNA. A cross-species regulatory network for *LTF* was constructed by combining correlations between transcription gene expression and *LTF* expression, transcription factor abundance, and enrichment of predicted transcription factor-binding motifs. This analysis identified a conserved set of highly expressed transcription factors closely associated with *LTF* expression, including ELF5, YY1, EHF, and ATF4 (Figure 6D).

DISCUSSION

This study explored the regulatory architecture and evolutionary diversification of mammalian *LTF* expression through integrated multi-omics analyses and cross-species comparisons. Human GTEx and bovine cGTEx datasets revealed that *LTF* expression was strongly enriched in glandular tissues, particularly in the mammary and salivary glands, consistent with established roles of *LTF* in iron homeostasis and mucosal immune defense (Jańczuk et al., 2022; Plate & Wiseman, 2017; Ramírez-Rico et al., 2023; Tsou et al., 2017; Xia et al., 2025). Single-cell transcriptomic analysis further identified LumSec cells as the predominant source of *LTF* expression in mammary tissue, providing cell-level insight into the functional specialization of *LTF* within the epithelial compartment responsible for milk protein synthesis and secretion (The GTEx Consortium, 2020).

Systematic analysis of 585 Holstein cows further linked *LTF*

expression to mammary immune status. Notably, *LTF* expression was significantly negatively associated with SCS (P<0.01), consistent with a potential contribution to innate immune protection in the mammary gland (Angelopoulou et al., 2024; Harmon, 1994). In cows with high *LTF* expression (TPM>500, n=17), differential expression analysis revealed significant enrichment of pathways related to immune defense, epithelial barrier function, keratinocyte activation, and calcium signaling, linking elevated *LTF* expression to a transcriptional program associated with mammary immune protection (Qi et al., 2021; Shen et al., 2025).

At the cis-regulatory level, integration of cis-eQTL mapping with ATAC-seq identified an accessible enhancer 4 660 bp upstream of bovine *LTF* that contained 20 cis-eQTL variants. These variants were distributed across two strong LD blocks (D'=1, r^2>0.85), including a 12 variant block that may constitute a major regulatory interval. Dual-luciferase reporter assays showed greater enhancer activity for the reference sequence than for the alternate sequence, providing new perspectives on the genetic basis of cattle *LTF* expression. Interestingly, only seven cows carried the homozygous alternate genotype, potentially reflecting the elimination of unfavorable genotypes under natural selection pressure, although further validation of the functional effects of these variants in larger, genetically diverse populations is required. Consistent with previous studies of lactoferrin regulation in humans and other species (Haraksingh & Snyder, 2013; Martyn et al., 2025), these findings provide the first systematic characterization of the fine-scale regulatory mechanisms governing bovine *LTF* expression.

Compared with the established regulatory paradigms of other major milk protein genes, the present findings revealed a more complex regulatory organization of *LTF*. For instance, expression of bovine α-lactalbumin (*LALBA*) and β-lactoglobulin (*BLG*) is strongly influenced by lactogenic hormones, including prolactin, through transcription factors such as STAT5 acting at proximal promoter regions (Lemay et al., 2007; Watson, 2006). In contrast, bovine *LTF* regulation involved an upstream enhancer harboring 20 cis-eQTLs, together with two ruminant-specific conserved downstream enhancers, E1 and E2. This combination of allele-sensitive and lineage-conserved regulatory elements suggests a more distributed cis-regulatory architecture than that described for *LALBA* and *BLG*. Such regulatory complexity may contribute to the pronounced interspecies variation in *LTF* expression and to differences in lactoferrin abundance between humans and ruminants.

E1 and E2 showed remarkable sequence conservation (>90% identity) across cattle, water buffaloes, goats, sheep, and red deer but were absent in humans and mice, aligning with the concept of lineage-specific divergence in mammary regulatory programs (Zemke et al., 2023) and the evolutionary diversification of lactation (Capuco & Akers, 2009; Hayssen, 1993; Oftedal, 2020). In lactating goat mammary tissue, scATAC-seq revealed preferential accessibility of both enhancers in LumSec cells. Co-accessibility analysis further linked E1 and E2 to the *LTF* promoter and to additional regulatory regions, supporting coordinated cis-regulatory activity without establishing direct chromatin looping (Ito et al., 2022; Monfils & Barakat, 2021; Ramasamy et al., 2023). Integration of RNA-seq data from five ruminant species identified 20 transcription factors that were highly expressed in milk-derived cells, including EHF, for which highly conserved predicted binding motifs were present in both enhancers and which is known to regulate epithelial gene expression (Fossum et al., 2017; Monfils & Barakat, 2021; Reehorst et al., 2021). Dual-luciferase reporter assays confirmed activity for both enhancers, with E1 producing a greater increase than E2 relative to the control and EHF co-transfection further enhancing reporter activity (P<0.001). These findings provide a molecular basis for understanding evolutionary adaptation of milk protein expression in ruminants.

Compared with ruminants, human samples showed exceptionally high *LTF* expression (TPM=12 248.5), suggesting that humans may have evolved unique transcriptional regulatory mechanisms to support the critical functions of LTF in immune defense and iron homeostasis (Cai et al., 2017; Platenburg et al., 1994). High lactoferrin concentrations in human milk may be particularly important during early life, when immune maturation remains incomplete and nutritional demands associated with rapid development are substantial. Milk-derived lactoferrin contributes to passive antimicrobial protection, establishment of the infant gut microbiome, and regulation of iron availability during brain development (Cai et al., 2018; Garrido et al., 2013). Our results indicated that human-specific regulatory elements remained active in MAC-T cells, demonstrating cross-species functionality in a bovine mammary epithelial context. Comparative analysis revealed that while a positionally conserved element existed in mice, sequence conservation was markedly reduced, possibly due to loss or rearrangement during murine evolution (Amaral et al., 2018; Phan et al., 2025), reflecting species-specific evolution of regulatory elements. Integration of multi-omics data further delineated a cross-species regulatory network for *LTF*, identifying three key enhancers and a conserved set of associated transcription factors. Differential recognition of these elements by factors such as ELF5 and EHF supports species-specific organization of *LTF* transcriptional control. These findings not only refine current understanding of mammalian *LTF* regulation but also provide a theoretical basis for the development of mammary bioreactors incorporating humanized regulatory elements.

CONCLUSIONS

Integrated multi-omics and cross-species analyses defined the regulatory architecture and evolutionary characteristics of mammalian *LTF* expression. The study identified conserved ruminant enhancers associated with mammary epithelial regulation, characterized a lineage-divergent human enhancer that retained activity in bovine mammary epithelial cells, and constructed a cross-species regulatory network comprising three enhancers and 39 conserved core transcription factors, including ELF5 and EHF. These findings not only deepen our understanding of *LTF* transcriptional regulation but also provide candidate regulatory targets for the development of humanized regulatory element-based mammary bioreactors. Although reporter assays established enhancer activity *in vitro*, gene-edited animal models will be required to verify the *in vivo* functions of these regulatory elements and investigate network dynamics during mammary gland development and lactation. This research establishes a novel paradigm for studying milk protein gene evolution and lays a theoretical foundation for developing nutritionally enhanced dairy products.

DATA AVAILABILITY

Gene expression data from multiple human tissues were obtained from the GTEx database (https://gtexportal.org/). Bovine genome annotation data

were sourced from the Ensembl database (ARS-UCD1.2). All other relevant data are available within the article and its supplementary information files, or from the corresponding author upon reasonable request (https://www.gtexportal.org/, https://cgtex.roslin.ed.ac.uk/, https://duos.broadinstitute.org/ ENA BrowserGEO Accession viewer.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

H.M.F. formal analysis, investigation, resources, visualization, writing original draft, writing review and editing. Y.Y.L. formal analysis, methodology, resources, visualization. Z.Y.W. resources. X.M.L. resources, visualization. H.L. resources, validation. S.Y.Z. resources. W.W. resources. X.R.Y. resources. Y.Y.L. resources. N.G.M. resources. Y.Z. resources. Y.M.W. resources. Y.J. resources, supervision. Y.W. funding acquisition, investigation, resources, supervision, writing review and editing. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank the High-Performance Computing Center (HPC) of Northwest A&F University (NWAFU) for providing computing resources.

REFERENCES

Abedi-Firoozjah R, Behnam B, Balouch Zehi Z, et al. 2025. Detection and quantification of lactoferrin: innovations, applications, and challenges. *Food Chemistry*, **466**: 142204.

Agrizzi Verediano T, Agarwal N, Juste Contin Gomes M, et al. 2023. Effects of dietary fiber on intestinal iron absorption, and physiological status: a systematic review of *in vivo* and clinical studies. *Critical Reviews in Food Science and Nutrition*, **63**(27): 9017–9032.

Aguet F, Alasoo K, Li YI, et al. 2023. Molecular quantitative trait loci. *Nature Reviews Methods Primers*, **3**(1): 4.

Amaral PP, Leonardi T, Han N, et al. 2018. Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. *Genome Biology*, **19**(1): 32.

Angelopoulou A, Harris HMB, Warda AK, et al. 2024. Somatic cell count as an indicator of subclinical mastitis and increased inflammatory response in asymptomatic lactating women. *Microbiology Spectrum*, **12**(10): e04051–23.

Bailey TL, Boden M, Buske FA, et al. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, **37**(S2): W202–W208.

Benegas G, Batra SS, Song YS. 2023. DNA language models are powerful predictors of genome-wide variant effects. *Proceedings of the National Academy of Sciences of the United States of America*, **120**(44): e2311219120.

Brown CD, Mangravite LM, Engelhardt BE. 2013. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genetics*, **10**(7): e1004546.

Browning BL, Zhou Y, Browning SR. 2018. A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, **103**(3): 338–348.

Bullen JJ, Rogers HJ, Leigh L. 1972. Iron-binding proteins in milk and resistance to escherichia coli infection in infants. *Proceedings of the Royal Society of Medicine*, **65**(12): 1086.

Cai CX, Eck P, Friel JK. 2017. Gene expression profiles suggest iron transport pathway in the lactating human epithelial cell. *Journal of Pediatric Gastroenterology and Nutrition*, **64**(3): 460–464.

Cai XK, Duan YF, Li Y, et al. 2018. Lactoferrin level in breast milk: a study of 248 samples from eight regions in China. *Food & Function*, **9**(8): 4216–4222.

Capuco AV, Akers RM. 2009. The origin and evolution of lactation. *Journal of Biology*, **8**(4): 37.

Carter JL, Stevens H, Ridge PG, et al. 2023. Short sequence aligner benchmarking for chromatin research. *International Journal of Molecular Sciences*, **24**(18): 14074.

Chang JJ, Wu H, Wu J, et al. 2023. Constructing a novel mitochondrial-related gene signature for evaluating the tumor immune microenvironment and predicting survival in stomach adenocarcinoma. *Journal of Translational Medicine*, **21**(1): 191.

Chen GM, Luo DQ, Zhong N, et al. 2022. GPC2 is a potential diagnostic, immunological, and prognostic biomarker in pan-cancer. *Frontiers in Immunology*, **13**: 857308.

Chen SF. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*, **2**(2): e107.

Chokeshaiusaha K, Puthier D, Nguyen C, et al. 2018. A demonstration of the H3 trimethylation ChiP-seq analysis of galline follicular mesenchymal cells and male germ cells. *Asian-Australasian Journal of Animal Sciences*, **31**(6): 791–797.

Dahiya S, Katiyar A, Rai S, et al. 2023. Ceftriaxone-resistant salmonella typhi isolated from paediatric patients in north India: insights into genetic profiles and antibiotic resistance mechanisms. *Indian Journal of Medical Microbiology*, **46**: 100448.

Ding Y, Cai KW, Liu LQ, et al. 2022. mHapTk: a comprehensive toolkit for the analysis of DNA methylation haplotypes. *Bioinformatics*, **38**(22): 5141–5143.

Dong SS, He WM, Ji JJ, et al. 2020. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings in Bioinformatics*, **22**(4): bbaa227.

Farnaud S, Evans RW. 2003. Lactoferrin-a multifunctional protein with antimicrobial properties. *Molecular Immunology*, **40**(7): 395–405.

Fossum SL, Mutolo MJ, Tugores A, et al. 2017. Ets homologous factor (EHF) has critical roles in epithelial dysfunction in airway disease. *Journal of Biological Chemistry*, **292**(26): 10938–10949.

Garrido D, Dallas DC, Mills DA. 2013. Consumption of human milk glycoconjugates by infant-associated bifidobacteria: mechanisms and implications. *Microbiology*, **159**(Pt 4): 649–664.

Gill N, Dhillon B. 2022. RNA-seq data analysis for differential expression. *In*: Coleman J. Fusarium Wilt: Methods and Protocols. New York: Springer, 45–54.

Granja JM, Corces MR, Pierce SE, et al. 2021. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature Genetics*, **53**(3): 403–411.

Hao QH, Peng W, Wang J, et al. 2022. The correlation between internet addiction and interpersonal relationship among teenagers and college students based on Pearson's correlation coefficient: a systematic review and meta-analysis. *Frontiers in Psychiatry*, **13**: 818494.

Haraksingh RR, Snyder MP. 2013. Impacts of variation in the human genome on gene regulation. *Journal of Molecular Biology*, **425**(21): 3970–3977.

Harmon RJ. 1994. Physiology of mastitis and factors affecting somatic cell counts. *Journal of Dairy Science*, **77**(7): 2103–2112.

Hayes BJ, Daetwyler HD. 2019. 1000 bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. *Annual Review of Animal Biosciences*, **7**: 89–102.

Hayssen V. 1993. Empirical and theoretical constraints on the evolution of lactation. *Journal of Dairy Science*, **76**(10): 3213–3233.

Ito S, Das ND, Umehara T, et al. 2022. Factors and mechanisms that influence chromatin-mediated enhancer-promoter interactions and transcriptional regulation. *Cancers*, **14**(21): 5404.

Jańczuk A, Brodziak A, Czernecki T, et al. 2022. Lactoferrin-the health-promoting properties and contemporary application with genetic aspects.

Foods, **12**(1): 70.
Kim D, Paggi JM, Park C, et al. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**(8): 907–915.
Langdon WB. 2015. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *BioData Mining*, **8**(1): 1.
Legrand D, Mazurier J. 2010. A critical review of the roles of host lactoferrin in immunity. *Biometals*, **23**(3): 365–376.
Lemay DG, Neville MC, Rudolph MC, et al. 2007. Gene regulatory networks in lactation: identification of global principles using bioinformatics. *BMC Systems Biology*, **1**: 56.
Li H, Durbin R. 2009. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, **25**(14): 1754–1760.
Li H, Handsaker B, Wysoker A, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16): 2078–2079.
Li HY, Li P, Yang HG, et al. 2019. Investigation and comparison of the anti-tumor activities of lactoferrin, α-lactalbumin, and β-lactoglobulin in A549, Ht29, HepG2, and MDA231-LM2 tumor models. *Journal of Dairy Science*, **102**(11): 9586–9597.
Li JS, Chen ZC, Bai YB, et al. 2023. Integration of ATAC-seq and RNA-seq analysis to identify key genes in the longissimus dorsi muscle development of the Tianzhu white yak. *International Journal of Molecular Sciences*, **25**(1): 158.
Li Q, Liang RB, Li Y, et al. 2020. Identification of candidate genes for milk production traits by RNA sequencing on bovine liver at different lactation stages. *BMC Genetics*, **21**(1): 72.
Li X, Kim Y, Tsang EK, et al. 2017. The impact of rare variation on gene expression across tissues. *Nature*, **550**(7675): 239–243.
Lopdell TJ, Trevarton AJ, Moody J, et al. 2024. A common regulatory haplotype doubles lactoferrin concentration in milk. *Genetics Selection Evolution*, **56**(1): 22.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**(12): 550.
Lytal N, Ran D, An LL. 2020. Normalization methods on single-cell RNA-seq data: an empirical survey. *Frontiers in Genetics*, **11**: 41.
Mao YJ, Zhu XR, Xing SY, et al. 2015. Polymorphisms in the promoter region of the bovine lactoferrin gene influence milk somatic cell score and milk production traits in Chinese Holstein cows. *Research in Veterinary Science*, **103**: 107–112.
Martyn GE, Montgomery MT, Jones H, et al. 2025. Rewriting regulatory DNA to dissect and reprogram gene expression. *Cell*, **188**(12): 3349–3366. e23.
Monfils K, Barakat TS. 2021. Models behind the mystery of establishing enhancer-promoter interactions. *European Journal of Cell Biology*, **100**(5-6): 151170.
Nguyen QH, Pervolarakis N, Blake K, et al. 2018. Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nature Communications*, **9**(1): 2028.
Nyquist SK, Gao P, Haining TKJ, et al. 2022. Cellular and transcriptional diversity over the course of human lactation. *Proceedings of the National Academy of Sciences of the United States of America*, **119**(15): e2121720119.
O'Halloran F, Berry DP, Bahar B, et al. 2010. Polymorphisms in the bovine lactoferrin promoter are associated with reproductive performance and somatic cell count. *Journal of Dairy Science*, **93**(3): 1253–1259.
Oberčkal J, Liaqat H, Matijašić BB, et al. 2023. Quantification of lactoferrin in human milk using monolithic cation exchange HPLC. *Journal of Chromatography B*, **1214**: 123548.
Ochoa TJ, Pezo A, Cruz K, et al. 2012. Clinical studies of lactoferrin in children. *Biochemistry and Cell Biology*, **90**(3): 457–467.
Oftedal OT. 2020. The evolution of lactation in mammalian species. *In*: Proceedings of the 94th Nestlé Nutrition Institute Workshop. Lausanne: Karger, 1–10.
Osorio D, Cai JJ. 2021. Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA-sequencing data quality control. *Bioinformatics*, **37**(7): 963–967.
Panigrahi A, O'Malley BW. 2021. Mechanisms of enhancer action: the known and the unknown. *Genome Biology*, **22**(1): 108.
Pennacchio LA, Visel A. 2010. Limits of sequence and functional conservation. *Nature Genetics*, **42**(7): 557–558.
Phan MHQ, Zehnder T, Puntieri F, et al. 2025. Conservation of regulatory elements with highly diverged sequences across large evolutionary distances. *Nature Genetics*, **57**(6): 1524–1534.
Plate L, Wiseman RL. 2017. Regulating secretory proteostasis through the unfolded protein response: from function to therapy. *Trends in Cell Biology*, **27**(10): 722–737.
Platenburg GJ, Kootwijk EPA, Kooiman PM, et al. 1994. Expression of human lactoferrin in milk of transgenic mice. *Transgenic Research*, **3**(2): 99–108.
Qi YF, Yang Y, Zhang Y, et al. 2021. Down regulation of lactotransferrin enhanced radio-sensitivity of nasopharyngeal carcinoma. *Computational Biology and Chemistry*, **90**: 107426.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**(6): 841–842.
Ramasamy S, Aljahani A, Karpinska MA, et al. 2023. The mediator complex regulates enhancer-promoter interactions. *Nature Structural & Molecular Biology*, **30**(7): 991–1000.
Ramírez-Rico G, Martinez-Castillo M, Cárdenas-Zúñiga R, et al. 2023. *Acanthamoeba castellanii* genotype T4: inhibition of proteases activity and cytopathic effect by bovine apo-lactoferrin. *Microorganisms*, **11**(3): 708.
Ranzoni AM, Tangherloni A, Berest I, et al. 2021. Integrative single-cell RNA-seq and ATAC-seq analysis of human developmental hematopoiesis. *Cell Stem Cell*, **28**(3): 472–487. e7.
Raschia MA, Nani JP, Maizon DO, et al. 2018. Single nucleotide polymorphisms in candidate genes associated with milk yield in Argentinean Holstein and Holstein x jersey cows. *Journal of Animal Science and Technology*, **60**: 31.
Reehorst CM, Nightingale R, Luk IY, et al. 2021. EHF is essential for epidermal and colonic epithelial homeostasis, and suppresses *Apc*-initiated colonic tumorigenesis. *Development*, **148**(12): dev199542.
Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, **11**(3): R25.
Rodríguez-Franco DA, Vázquez-Moreno L, Ramos-Clamont Montfort G, et al. 2005. Antimicrobial mechanisms and potential clinical application of lactoferrin. *Revista Latinoamericana De Microbiologia*, **47**(3-4): 102–111.
Rubinacci S, Hofmeister RJ, Sousa da Mota B, et al. 2023. Imputation of low-coverage sequencing data from 150, 119 UK biobank genomes. *Nature Genetics*, **55**(7): 1088–1090.
Shao YX, Zhang X, Zhang HW, et al. 2021. Effects of dietary supplementation of bovine lactoferricin on rumen microbiota, lactation, and health in dairy goats. *Frontiers in Nutrition*, **8**: 722303.
Shen SM, Wu YL, Shao ZW, et al. 2025. LTF as a potential predictive biomarker for durable benefit from first-line chemo-immunotherapy in small cell lung cancer. *Cancer Science*, **116**(6): 1522–1536.
Shen WK, Chen SY, Gan ZQ, et al. 2023. AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. *Nucleic Acids Research*, **51**(D1): D39–D45.
Smith NC, Matthews JM. 2016. Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. *Current Opinion in Structural Biology*, **38**: 68–74.
Stuart T, Butler A, Hoffman P, et al. 2019. Comprehensive integration of

single-cell data. *Cell*, **177**(7): 1888–1902. e21.

Tan ZY, Chen XR, Zuo JM, et al. 2023. Comprehensive analysis of scRNA-seq and bulk RNA-seq reveals dynamic changes in the tumor immune microenvironment of bladder cancer and establishes a prognostic model. *Journal of Translational Medicine*, **21**(1): 223.

Taylor-Weiner A, Aguet F, Haradhvala NJ, et al. 2019. Scaling computational genomics to millions of individuals with GPUs. *Genome Biology*, **20**(1): 228.

Teng CT. 1999. Regulation of lactoferrin gene expression by estrogen and epidermal growth factor: molecular mechanism. *Cell Biochemistry and Biophysics*, **31**(1): 49–64.

The GTEx Consortium. 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**(6509): 1318–1330.

Tsou YA, Tung YT, Wu TF, et al. 2017. Lactoferrin interacts with SPLUNC1 to attenuate lipopolysaccharide-induced inflammation of human nasal epithelial cells via down-regulated MEK1/2-MAPK signaling. *Biochemistry and Cell Biology*, **95**(3): 394–399.

Tsuyuzaki K, Sato H, Sato K, et al. 2020. Benchmarking principal component analysis for large-scale single-cell RNA-sequencing. *Genome Biology*, **21**(1): 9.

Viale E, Tiezzi F, Maretto F, et al. 2017. Association of candidate gene polymorphisms with milk technological traits, yield, composition, and somatic cell score in Italian Holstein-Friesian sires. *Journal of Dairy Science*, **100**(9): 7271–7281.

Wang N, Miao XD, Lu WX, et al. 2024. RUNX3 exerts tumor-suppressive role through inhibiting EXOSC4 expression. *Functional & Integrative Genomics*, **24**(3): 103.

Wang W, Wei ZY, Li ZH, et al. 2024b. Integrating genome- and transcriptome-wide association studies to uncover the host-microbiome interactions in bovine rumen methanogenesis. *iMeta*, **3**(5): e234.

Watson CJ. 2006. Key stages in mammary gland development - involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ. *Breast Cancer Research*, **8**(2): 203.

Wu HJ, Gonzalez Villalobos R, Yao X, et al. 2022. Mapping the single-cell transcriptomic response of murine diabetic kidney disease to therapies. *Cell Metabolism*, **34**(7): 1064–1078. e6.

Wu JL, Li YW, Feng DL, et al. 2023. Integrated analysis of ATAC-seq and RNA-seq reveals the transcriptional regulation network in SLE. *International Immunopharmacology*, **116**: 109803.

Wu WQ, Jiang YN, Xing DH, et al. 2024. The epigenetic regulators EP300/CREBBP represent promising therapeutic targets in MLL-rearranged acute myeloid leukemia. *Cell Death Discovery*, **10**(1): 206.

Xia B, Liang JW, Lu YL, et al. 2025. Lactoferrin influences atherosclerotic progression by modulating macrophagic AMPK/mTOR signaling-dependent autophagy. *Scientific Reports*, **15**(1): 10585.

Xie WC, Song LY, Wang XY, et al. 2021. A bovine lactoferricin-lactoferrampin-encoding *Lactobacillus reuteri* CO21 regulates the intestinal mucosal immunity and enhances the protection of piglets against enterotoxigenic *Escherichia coli* K88 challenge. *Gut Microbes*, **13**(1): 1956281.

Xu H, Jia JP. 2021. Single-cell RNA sequencing of peripheral blood reveals immune cell signatures in Alzheimer's disease. *Frontiers in Immunology*, **12**: 645666.

Xu HY, Huang WQ, Hou QC, et al. 2017. The effects of probiotics administration on the milk production, milk components and fecal bacteria microbiota of dairy cows. *Science Bulletin*, **62**(11): 767–774.

Xu K, Zhang WW, Wang C, et al. 2021. Integrative analyses of scRNA-seq and scATAC-seq reveal CXCL14 as a key regulator of lymph node metastasis in breast cancer. *Human Molecular Genetics*, **30**(5): 370–380.

Yoon SH, Ha SM, Lim J, et al. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek*, **110**(10): 1281–1286.

Yu GC, Wang LG, He QY. 2015. ChIPseeker: an R/Bioconductor package for ChiP peak annotation, comparison and visualization. *Bioinformatics*, **31**(14): 2382–2383.

Zabolewicz T, Barcewicz M, Brym P, et al. 2014. Association of polymorphism within *LTF* gene promoter with lactoferrin concentration in milk of Holstein cows. *Polish Journal of Veterinary Sciences*, **17**(4): 633–641.

Zemke NR, Armand EJ, Wang WL, et al. 2023. Conserved and divergent gene regulatory programs of the mammalian neocortex. *Nature*, **624**(7991): 390–402.

Zhang ZB, Wang A, Hu HH, et al. 2023. The efficient phasing and imputation pipeline of low-coverage whole genome sequencing data using a high-quality and publicly available reference panel in cattle. *Animal Research and One Health*, **1**(1): 4–16

Zhang ZY, Gao YN, Qian YY, et al. 2025. The Lyn/RUVBL1 complex promotes colorectal cancer liver metastasis by regulating arachidonic acid metabolism through chromatin remodeling. *Advanced Science*, **12**(5): 2406562.

Zhao YD, Li MC, Konaté MM, et al. 2021. TPM, FPKM, or normalized counts? A comparative study of quantification measures for the analysis of RNA-seq data from the NCI patient-derived models repository. *Journal of Translational Medicine*, **19**(1): 269.

Zhou HJ, Li L, Li YM, et al. 2022. PCA outperforms popular hidden variable inference methods for molecular QTL mapping. *Genome Biology*, **23**(1): 210.