Song et al. *Zool. Res.* 2026, 47(4): 1059−1071
https://doi.org/10.24272/j.issn.2095-8137.2025.328
https://cstr.cn/32248.14.zr.2025.328

ZR | Zoological Research

Article Open Access

LivestockDev: A multi-omics resource for exploring cross-species comparison of livestock embryogenesis

Peng-Cheng Song[1,2,#], Yu-Chao Liang[1,#], Xiang He[1], Qi-Yan Cheng[1], Peng-Wei Hu[1], Han-Shuang Li[1], Yong Zhang[1,3], Gui-Fang Cao[1,*], Yong-Chun Zuo[1,2,*]

[1] *State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock*, *Institutes of Biomedical Sciences*, *School of Life Sciences*, *Inner Mongolia University*, *Hohhot*, *Inner Mongolia* 010021, *China*

[2] *Computer Department*, *Hohhot Vocational College*, *Hohhot*, *Inner Mongolia* 010020, *China*

[3] *College of Veterinary Medicine*, *Key Laboratory of Animal Biotechnology Ministry of Agriculture and Rural Affairs*, *Northwest A & F University*, *Yangling*, *Shaanxi* 712100, *China*

ABSTRACT

Advances in experimental and sequencing technologies have driven a surge in multi-omics data on early embryonic development in livestock species. Nevertheless, the absence of systematic data curation and standardized analytical frameworks has hindered research on early embryogenesis in livestock, limiting data integration, cross-species comparisons, and mechanistic insights. Here, we developed LivestockDev, an integrative multi-omics database dedicated to livestock embryogenesis. LivestockDev covers five major livestock species (cattle, sheep, goats, pigs, and horses) and systematically integrates multimodal datasets encompassing transcriptomics, chromatin accessibility, DNA methylation, and histone modifications across multiple embryonic stages and tissue types. Beyond providing high-quality data browsing, visualization, and download services, LivestockDev features a comprehensive suite of analytical tools specifically designed for cross-species studies of embryonic development. These tools enable fine-grained annotation of stage and tissue specific genes in early embryos, spatiotemporal comparison of gene expression across cell types and tissues, cross-species visualization of developmental gene cluster dynamics, construction of gene regulatory networks, comparative analysis of conserved protein domains, and visualization and retrieval of epigenetic modifications and single-nucleotide variants, as well as cross-species homologous gene conservation, annotation, and expression, and stage-specific gene scoring. The platform provides a modular framework and interactive web interface. It enables multi-scale, cross-species analysis of livestock embryonic data. Additionally, LivestockDev integrates over 1 594 curated publications related to livestock developmental biology, providing a valuable knowledge base for in-depth exploration of early embryogenesis. LivestockDev is the first comprehensive database focused on comparative embryonic development across multiple livestock species and is freely accessible at: http://bioinfor.imu.edu.cn/livestockDev.

Keywords: Livestock; Embryonic Development; Multi-omics Integration; Cross-species Comparison; Database

INTRODUCTION

Mammalian early embryonic development is orchestrated by precise temporal patterns and tightly regulated molecular mechanisms (Satoh, 1982; Schall et al., 2019). Investigations into the molecular genetic basis of early embryogenesis, predominantly in humans and mice, have been a longstanding focus in developmental biology (Ortega et al., 2018; Rossant & Tam, 2017; Rossant, 2018). Livestock species, including cattle, sheep, goats, pigs, and horses, are essential to animal husbandry and share conserved embryonic developmental processes, transitioning from the zygote to the morula and then to the blastocyst, followed by tissue differentiation (Bazer & Johnson, 2024). Recent studies have also focused on pluripotency factors involved in the regulation of livestock embryonic development (Wu et al., 2021). Compared to humans and mice, livestock species have less comprehensive reference genome annotations and limited sequencing data for early embryonic development, creating challenges for data integration and interpretation (Hocquette et al., 2007; Park

Received: 31 October 2025; Accepted: 09 December 2025; Online: 10 December 2025

Foundation items: This research was funded by the Hohhot City Basic Research and Applied Basic Research Project (2024GJ18), the National Natural Science Foundation of China (Grant Nos. 62571279, 62501318, and 62461046), the Group Project of Developing Inner Mongolia through Talents (2025TEL25), the China Postdoctoral Science Foundation (Grant No. 2024MD763987), and the Central Guidance Fund for Local Science and Technology Development (2024ZY0168).

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: guifangcao@126.com; yczuo@imu.edu.cn

et al., 2023).

Several specialized databases and bioinformatics platforms have been established to enhance the effective use of livestock embryonic development data. For example, EmExplorer provides insights into the temporal expression profiles of preimplantation embryonic development in five mammalian species, facilitating cross-species comparisons of functional gene sets and biological pathways. However, it is limited to transcriptome data and lacks multi-omics integration (Hu et al., 2019). FifBase is a database that curates and integrates RNA-Seq data and literature on livestock fertility. It facilitates the identification and analysis of fertility-associated genes in ten livestock species. However, it lacks epigenetic data and analyses (Li et al., 2021). LivestockExp leverages 43 710 RNA-Seq samples across 15 livestock species to analyze gene expression and alternative splicing events. However, it lacks comprehensive transcriptomic data on early embryonic development (Liu et al., 2022). GereDB establishes a gene expression regulatory network for pigs through sequence homology analysis between mice and pigs, elucidating gene regulatory relationships in pigs. However, it is limited to a single species (Huang et al., 2019). BGD provides a suite of tools for browsing, annotating, and mining cattle genomic data, advancing cattle genomics research and comparative genomics. However, its most recent data update was in 2019 (Shamimuzzaman et al., 2020). Therefore, the development of a comprehensive database integrating early embryonic development and multi-tissue expression profiles of livestock, along with a suite of analytical tools for identifying biomarkers across multiple omics layers, is essential.

Here, we developed LivestockDev, a database that integrates multimodal data on early embryonic development in livestock. This platform provides temporal gene expression profiles for pre-implantation development and extends to gene expression and epigenetic modification patterns across post-implantation stages and tissues. Researchers can use this platform to explore the dynamic regulatory mechanisms of early livestock development, including species, cell types, tissues, genomic and epigenetic modifications, gene interaction networks, gene pathways, protein sequences, three-dimensional structures, homologous genes and stage-specific gene rankings (Supplementary Table S1).

MATERIALS AND METHODS

Data processing and integration

To advance research on early embryonic development and livestock breeding, we systematically collected, processed, and integrated multi-omics data, including transcriptomics, epigenomics, proteomics, and literature data, to create LivestockDev. Data were primarily obtained from Gene Expression Omnibus (GEO) (Clough et al., 2024), Sequence Read Archive (SRA) (Katz et al., 2022), Genome Sequence Archive (GSA) (Chen et al., 2021), UniProt Knowledgebase (UniProtKB) (The UniProt Consortium, 2019), Protein Data Bank (PDB) (Berman et al., 2000), and PubMed (Sayers et al., 2025). FASTQ raw data were downloaded with IBM Aspera software v.3.11.0.5 (https://www.ibm.com/products/aspera/downloads), and Python v.3.10.1 web crawlers were used to retrieve protein sequences, three-dimensional structural information, and relevant literature by website robot policies. Homologous genes across livestock species were identified using the NCBI Orthologs dataset (downloaded from NCBI Gene, accessed in May 2024). Orthology relationships were curated by NCBI based on sequence similarity and phylogenetic methods. We retained only one-to-one orthologous pairs and standardized gene identifiers to NCBI Gene IDs. These were subsequently converted to Ensembl Gene IDs according to the reference annotation of each species. The reference genomes and corresponding gene annotation files were downloaded from Ensembl (release 104) (https://www.ensembl.org/), including *Bos taurus* (ARS-UCD1.2), *Sus scrofa* (Sscrofa11.1), *Capra hircus* (ARS1), *Ovis aries* (Oar_v3.1), and *Equus caballus* (EquCab3.0).

Bulk RNA sequencing (bulk RNA-seq) data were subjected to FastQC v.0.11.9 (Andrews, 2010) for quality control, Trim Galore v.0.6.7 (Krueger, 2012) for filtering low-quality reads, Hisat2 v.2.2.1 (Kim et al., 2019) for alignment to the reference genome, FeatureCounts v.2.0.1 (Liao et al., 2014) for quantification, Transcripts Per Million (TPM) normalization, and subsequently, differential gene expression analysis using DESeq2 v.1.34.0 (Love et al., 2014). Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) were employed for dimensionality reduction, while Gene Ontology (GO, release 2025-04, accessed May 2025; https://geneontology.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG, , release 115.1, accessed August 2025; https://www.kegg.jp/) functional annotation analyses were conducted to identify developmental stage-specific, tissue-specific, and conserved genes.

Single-cell RNA sequencing (scRNA-seq) data were generated using the 10x Genomics (Zheng et al., 2017) platform and SMART-seq2 (Picelli et al., 2014) method. The preprocessing pipeline followed a similar approach to bulk RNA-seq data processing, with downstream analyses involving clustering, differential gene expression analysis, and identification of developmental genes using SCANPY v.1.7.2 (Wolf et al., 2018).

Epigenomic data, including DNA methylation and histone modifications, were processed following standard pipelines with explicit quality control and alignment settings. Raw sequencing reads were subjected to FastQC v.0.11.9 for quality control, and low-quality bases (Phred score<20) and adapter sequences were trimmed using Trim Galore v.0.6.7. Reads with a length shorter than 30 bp after trimming were discarded. For DNA methylation data, reads were aligned to the reference genome using Bismark v.0.22.3 (Krueger & Andrews, 2011) with parameters "--bowtie2 --score_min L,0,-0.2", and duplicate reads were removed with the deduplicate_bismark function. CpG methylation levels were quantified as the ratio of methylated cytosines to total coverage per site, excluding sites with coverage <5×. For histone modification data generated by chromatin immunoprecipitation sequencing (ChIP-seq), reads were aligned with HOMER v.4.11 (Heinz et al., 2010) using default parameters. Peaks were called with findPeaks under the "histone" setting, and input DNA was used as control to reduce background noise. Chromatin accessibility, methylation levels, and single nucleotide variant (SNV) information were computed using DeepTools v.3.5.1(Ramírez et al., 2014) (bamCoverage) and visualized in JBrowse v.1.16 (Buels et al., 2016) for further analysis.

Proteomic data were retrieved via UniProtKB and PDB APIs, enabling the study of protein interaction patterns through structural annotation of three-dimensional protein structures. Literature data were extracted from PubMed, comprising 1 594 articles with curated titles, abstracts, journal names, and author information, along with an integrated search

functionality for database users.

All data were integrated into LivestockDev, offering query, visualization, and download functionalities. This provides high-quality multi-omics resources to support livestock breeding and developmental research and facilitates the advancement of multimodal causal discovery and intelligent breeding tools (Supplementary Figure S1).

Batch effect correction and data normalization

To minimize batch effects arising from differences in experimental platforms, sequencing depth, and study sources, we applied batch correction during the data integration process. Expression matrices from individual datasets were first normalized within each dataset using TPM/ counts per million (CPM)zw transformation and $\log_2$-transformed where appropriate. Batch-specific variations across datasets were subsequently corrected using the ComBat algorithm (sva R package, par.prior=TRUE; https://bioconductor.org/packages/sva/), while preserving biological differences such as developmental stage or tissue-specific expression patterns.

Cross-species alignment of developmental stages

Developmental stage alignment across cattle, sheep, goat, pig, and horse was anchored to conserved morphological milestones—oocyte, zygote, 2-cell, 4-cell, 8-cell, 16-cell, morula, blastocyst, and inner cell mass (ICM)/trophectoderm (TE) segregation. This strategy avoids biases of days post-fertilization due to interspecies developmental speed variation and aligns with established practice in cross-species expression resources such as Bgee (Bastian et al., 2021).

Stage-specific gene scoring framework

To systematically evaluate the stage-specific relevance of genes during early embryonic development, we established a multi-metric scoring framework. Expression profiles were first normalized using variance stabilizing transformation and subsequently organized into a matrix in which each row corresponds to a gene and each column to a sample from a specific developmental stage. For each stage, samples from that stage were designated as the positive group and samples from all other stages as the negative group, and gene-level metrics were then computed to quantify stage-specificity.

1. Log_2 fold change ($\log_2$FC)

$$\log_2\text{FC}_g = \log_2 \frac{\overline{X}_g^{s+}}{\overline{X}_g^{s-} + \varepsilon} \tag{1}$$

The average expression of gene *g* in the positive and negative groups was denoted as the group-specific mean, and a small constant (set to 1×10^{-8}) was introduced to prevent division by zero.

2. Welch's *t*-test (*t*-statistic and *P*-value):

$$t_g^s = \frac{\overline{X}_g^{s+} - \overline{X}_g^{s-}}{\sqrt{\frac{\text{var}\left(X_g^{s+}\right)}{n_+} + \frac{\text{var}\left(X_g^{s-}\right)}{n_-}}} \tag{2}$$

An associated *P*-value was computed to test the null hypothesis that the mean expression of gene *g* did not differ between the positive and negative groups. Multiple testing correction was subsequently performed using the Benjamini–Hochberg false discovery rate procedure.

3. Cohen's d effect size:

$$d_g^s = \frac{\overline{X}_g^{s+} - \overline{X}_g^{s-}}{s_{\text{pooled}}},\ s_{\text{pooled}} = \sqrt{\frac{(n_+ - 1)\,\text{var}\left(X_g^{s+}\right) + (n_- - 1)\,\text{var}\left(X_g^{s-}\right)}{n_+ + n_- - 2}} \tag{3}$$

4. Point-biserial correlation(r_{pb}):

$$r_{\text{pb},g}^s = \text{corr}\left(X_g, y^s\right) \tag{4}$$

The indicator variable was defined as one if a sample belonged to stage s and zero otherwise, thereby quantifying the linear association between gene expression and stage membership.

5. ANOVA *F*-statistic across all stages:

$$F_g = \frac{\text{Between-group variance of}g}{\text{Within-group variance of}g} \tag{5}$$

Computed using the standard one-way ANOVA, assessing the overall stage-specific expression variance.

6. Ranking metrics: For each gene and each evaluation metric (including $\log_2$ fold change, Cohen's d, point-biserial correlation coefficient, *F*-statistic, *t*-test *P*-value, and point-biserial correlation *P*-value), genes were ranked in descending order of importance, except for *P*-values which were ranked in ascending order. The resulting ranks were then recorded for each gene–metric–stage combination.

7. Aggregated rank and normalized score: the mean rank was computed across all metrics for each gene:

$$\text{meanrank}_g^s = \frac{1}{K}\sum_{M=1}^{K} R_{g,M}^s \tag{6}$$

where K=7 is the number of metrics. The aggregated stage-specific score was then normalized to the interval [0,1]:

$$\text{aggscore}_g^s = 1 - \frac{\text{meanrank}_g^s - \min\left(\text{meanrank}^s\right)}{\max\left(\text{meanrank}^s\right) - \min\left(\text{meanrank}^s\right) + \delta} \tag{7}$$

with δ set to 10^{-12} to avoid division by zero. A higher score reflects stronger stage-specific importance.

Database implementation

LivestockDev is a distributed system developed using the Vue3 v.3.2.13 (https://vuejs.org/) framework for the front-end and the Django Rest Framework (DRF) v.3.12.4 (https://www.django-rest-framework.org/) for the back-end. MySQL v.8.0.23 (https://www.mysql.com/) is utilized as the database solution to ensure efficient data storage and retrieval. The database is designed with multiple tables organized by data type and usage requirements, each optimized for specific query needs. Indexing techniques are also implemented to enhance data retrieval efficiency, enabling rapid queries of large-scale datasets. The system employs the Nginx web server (https://nginx.org/en/) for deployment and management. Upon completion of development, the system is packaged and encrypted. The front-end, back-end, and database are deployed on separate servers to segregate data storage from computation, ensuring data security and optimizing performance. Additionally, when selecting visualization tools, we prioritized performance stability and computational speed, choosing ECharts v.5.3.3 (Li et al., 2018), Molstar v.4.11.2 (Sehnal et al., 2021), STRING v.11.5 (Szklarczyk et al., 2021), JBrowse v.1.16 (Buels et al., 2016), and BioRender (https://www.biorender.com/) to meet the demands of fluid and real-time online visualization. LivestockDev is freely accessible to the public and can be used on both desktop and mobile devices without requiring login or registration. The

platform is optimized for compatibility with multiple browsers, including Chrome (recommended), Internet Explorer, Opera, Firefox, Microsoft Edge, and Safari.

RESULTS

LivestockDev framework

LivestockDev is a multi-omics database designed for visualizing and analyzing the early development of livestock, encompassing five mammalian species (cattle, pigs, sheep, goats, and horses) and integrating 56 publicly available multi-omics datasets (Figure 1A). The platform systematically compiles and organizes data on early embryonic development in livestock, including single-cell transcriptomic, bulk transcriptomic, and epigenomic data derived from 1 819 samples, covering 29 cell types and 74 tissues and organs, meticulously classified according to the chronological progression of pre-implantation embryogenesis. Comprehensive metadata for each dataset and sample are provided, encompassing breed, cell type, tissue type, experimental methodology, and other relevant characteristics, in addition to a curated repository of 1594 peer-reviewed articles pertaining to livestock developmental biology (Figure 1B).

Three primary analytical modules are featured, comprising eight specialized tools and 22 distinct web pages. Data on the platform are centrally managed through a MySQL relational database, enabling seamless support for data analysis, visualization, and downloading, as well as robust querying capabilities. As additional public datasets are made available and user feedback is incorporated, future iterations of the platform will expand both the volume of datasets and the breadth of analytical features.

To present the data more intuitively, LivestockDev offers a suite of advanced visual and analytical tools. At the cellular level, the platform includes tools for clustering cell and tissue types, conducting differential gene expression analyses, and visualizing gene expression dynamics across various cell types during early embryonic development. At the tissue level, it provides online tools for analyzing gene expression across multiple tissues and organs from different livestock species. At the genomic level, a genome browser is available to facilitate the retrieval of reference genomes for various livestock species, along with epigenetic data and SNVs. At the gene level, the platform integrates differential gene expression analysis with pathway and functional annotations, provides homologous gene search across species, and ranks stage-specific genes, thereby offering statistical insights into development-associated gene sets across different tissues and cell types. Detailed gene annotations are provided, including gene family classifications, GO annotations, subcellular localization, gene interaction networks, and structural data for three-dimensional protein conformations. At the sequence level, LivestockDev supplies coding sequences (CDS) and amino acid sequences (AA) for development-related genes. These comprehensive tools collectively support in-depth genomic and functional investigations of livestock embryonic development (Figure 1C).

Statistical insights

Based on the classification of five livestock species, some datasets share the same GEO Series (GSE) accession number. LivestockDev ultimately integrates 11 scRNA-seq, 41 bulk RNA-seq, and 16 epigenomic datasets, with some datasets spanning multiple omics types (Supplementary Table S2). Cattle, sheep, and goats have the most comprehensive data, while only three datasets are available for horses (Figure 2A). We systematically classified the species in the collected datasets, labeling those without breed information as “Unknown” (Figure 2B). LivestockDev includes 1 819 samples with multi-omics data spanning early embryonic stages and various tissues (Supplementary Table S3). Embryo samples are relatively comprehensive,

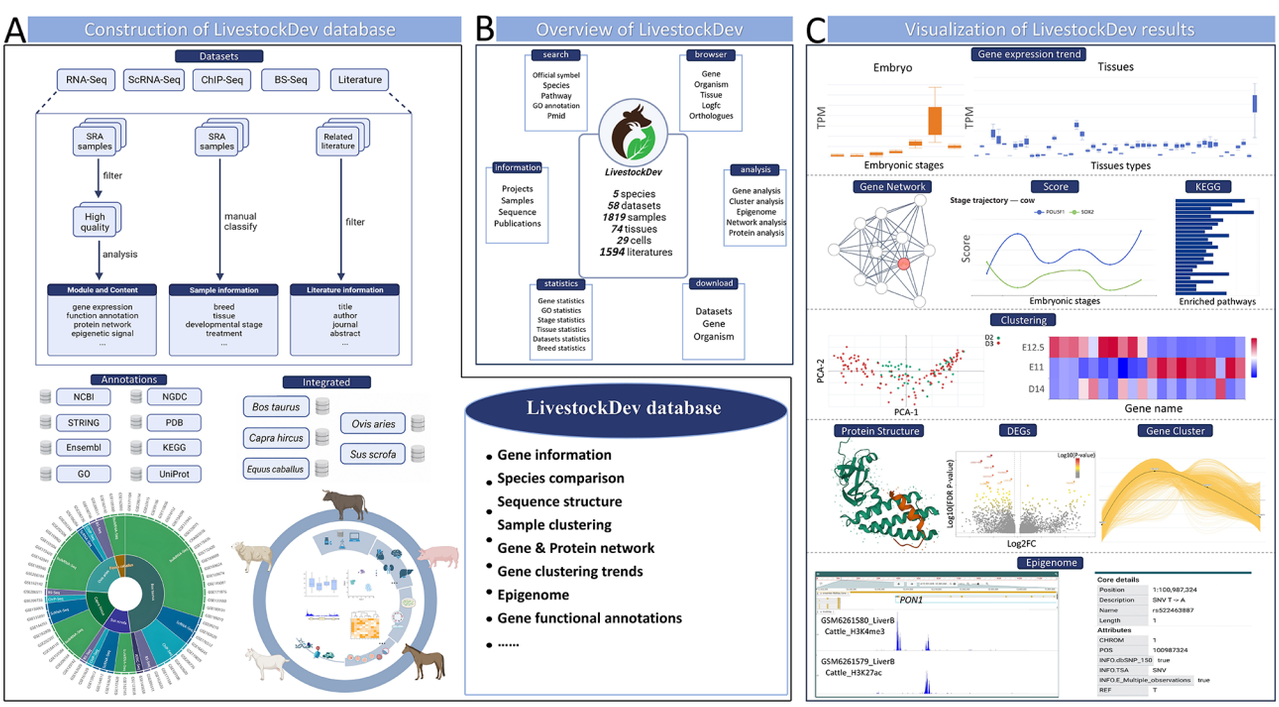

Figure 1 The LivestockDev framework

A: Multi-omics database construction. B: Overview of the multi-species integrated database. C: Modular system design, and interactive visualization tools.

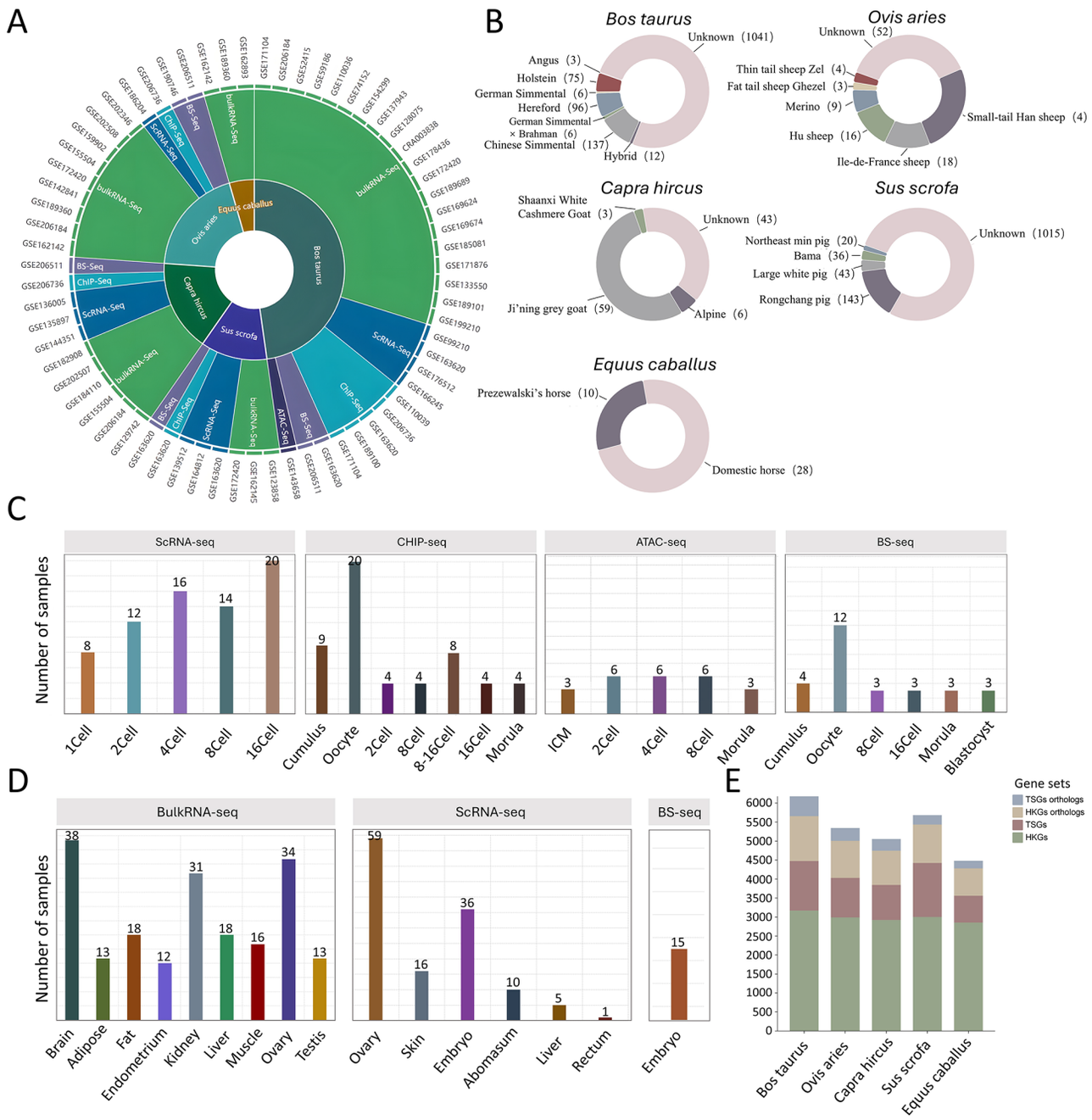

Figure 2 Summary of LivestockDev data

A: 56 datasets from different omics across multiple species. B: Breed distribution across species. C: Sample distribution across early embryonic developmental stages in different omics. D: Sample distribution across tissues in different omics. E: Proportion of HKGs, HKG orthologs, TSGs, and TSG orthologs.

representing various developmental stages and encompassing different sequencing technologies (Figure 2C). The distribution of tissue samples indicates that bulk RNA-seq data are abundant in the brain and liver, whereas epigenomic data are relatively scarce in these tissues (Figure 2D). Tissue-specific genes (TSGs) and housekeeping genes (HKGs) were identified within the development-related gene set. The proportion of orthologous genes (OGs) in each species relative to the other four livestock species was calculated using the livestock homologous gene set from the NCBI database, to evaluate the cross-species conservation of HKGs and TSGs (Figure 2E). The results show that the proportion of homologous genes among HKGs and TSGs decreases with increasing evolutionary distance. We further quantified the number of cross-species homologous genes, using cattle as the reference due to the most complete dataset. Specifically, 906 genes have no homologs in other species, 466 genes have homologs in one species, 851 genes in two species, 2 543 genes in three species, and 1 0511 genes have homologs in all four other species (Supplementary Figure S2). Annotations of these genes and their cross-species expression profiles can be explored using the Homologous Gene tool in LivestockDev.

Overview of LivestockDev database

LivestockDev boasts an intuitive and well-structured homepage with a navigation bar, species search, global search, database resource list, and an overview of the database (Figure 3A). It also offers five functional modules for analyzing early embryonic development in livestock, along

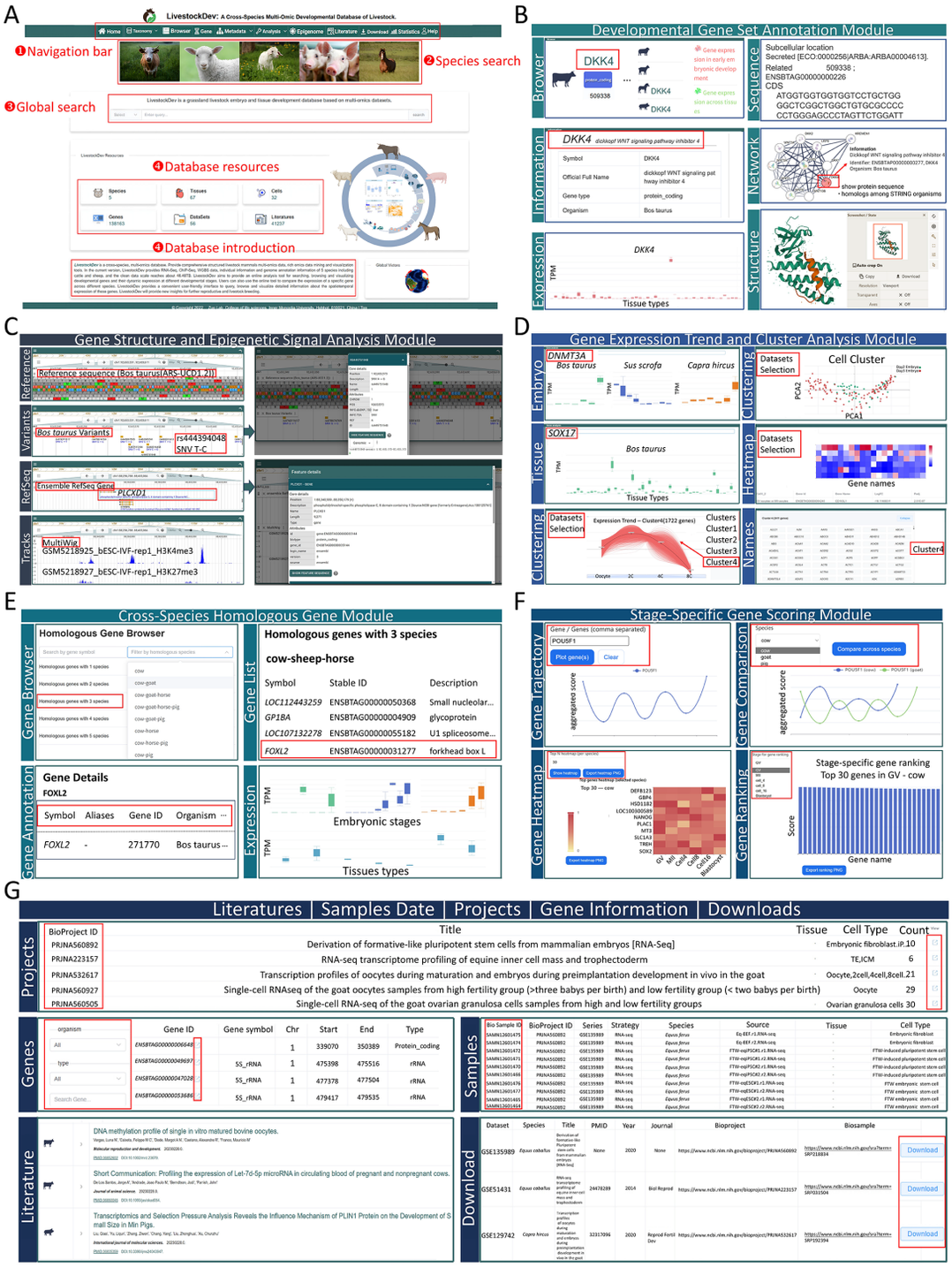

Figure 3 Overview of LivestockDev Database

A: The homepage integrates a navigation bar, species-specific entry, global search, and summaries of resources, including species, tissues, cells, genes, datasets, and literature, providing a structured entry point for multi-omics exploration. B: The annotation module combines gene information, coding and amino acid sequences, tissue-specific expression profiles, protein–protein interaction networks, and 3D structures. In the interaction network, nodes represent genes and edges indicate functional associations, with edge thickness reflecting interaction strength. C: The genome browser displays reference sequences, SNVs, gene models, and epigenetic signal tracks in a unified interface. Users can navigate genomic regions across chromosomes, zoom into nucleotide resolution, and link gene features with epigenetic modifications to analyze regulatory mechanisms. Variants are highlighted with colored markers, and epigenetic tracks are shown as signal peaks aligned to genomic loci. D: The expression module visualizes transcriptional dynamics across embryonic stages and tissues in livestock species. Integrated clustering approaches (PCA and UMAP) identify differentially expressed genes, and genes with similar expression patterns are grouped into clusters. In heatmaps, red and blue denote up- and down-regulated expression, respectively. E: The homology module enables filtering of homologous genes across species, provides detailed annotations, and compares expression dynamics. Conserved genes are listed in tabular form, with stable IDs and functional descriptions for clarity. F: The stage-specific scoring module quantifies gene importance through developmental trajectories, cross-species comparisons, heatmaps, and stage-specific rankings. G: The statistics and download module provides metadata for projects, samples, and annotated genes, together with curated literature and direct download services.

with various statistics on the collected data and a dedicated download page.

Developmental gene set annotation module: This module identifies and annotates development-related gene sets across various tissues and cell types in livestock based on differential gene expression and pathway analyses. It integrates data from multiple authoritative databases, offering extensive functional data for in-depth analysis. In addition to providing coding and amino acid sequence information, the module explores gene functions in tissue and subcellular localization, interaction networks, biological processes, molecular functions, and 3D protein structures. For instance, users can explore the cattle developmental gene *DKK4* through multiple functional views. In the "Browser Gene" panel (top left), users can quickly locate *DKK4* and related genes. The "Information" panel (middle left) provides detailed biological annotations, including gene type, family, homologs across species, and subcellular localization. The "Specific Sequence" panel (top right) displays CDS and amino acid sequences retrieved from UniProt, while the "Dynamic Expression" panel (bottom left) shows tissue- and cell-specific expression patterns of *DKK4*. Furthermore, the "P–P Interaction" panel (middle right) constructs gene regulatory networks, visualizing interactions between *DKK4* and other developmental regulators or metabolites. Finally, the "Protein Secondary Structure" panel (bottom right) renders the 3D structure of the *DKK4* protein directly on the webpage. Together, these features highlight how the module integrates multi-level information to facilitate functional exploration of developmental genes. (Figure 3B).

Gene structure and epigenetic signal analysis module: The module is centered around the development and visualization of a genome browser, providing an interactive platform for exploring epigenomic data. At the top of the browser, chromosome numbers and genomic coordinates are displayed, along with an integrated chromosome-switching function that allows users to navigate genomic regions across different chromosomes. The browser also supports zoom-in and zoom-out functionalities for detailed inspection. Chromosomal coordinates are entered in the standard "chromosome:start …end" format. For instance, "1:77 570 552…77 746 671" specifies the genomic interval from position 77 570 552 to 77 746 671 on chromosome 1. Below the chromosome display, the browser presents nucleotide sequences and gene annotation data. When the visualization window spans fewer than 1 000 base pairs (bp), the underlying nucleotide sequence is revealed and linked to known genomic variation sites. Users can click on specific SNVs to access detailed mutation information. Gene annotation features include gene transcripts, intronic and exonic structures, and transcriptional orientation, with particular emphasis on aligning epigenetic modification signals to genomic loci. Clicking on a gene opens a detailed annotation panel, displaying gene-specific details such as transcript isoforms, CDS, and corresponding amino acid sequences. The browser arranges epigenomic datasets in an organized manner, with sample labels positioned on the far left. Labels follow the standardized format "dataset_tissue/cell_modification" to facilitate dataset identification. For example, "GSM5218925_bESC-IVF-rep1_H3K4me3" denotes an H3K4me3 modification dataset derived from bovine embryonic stem cells (bESCs) obtained via in vitro fertilization (IVF), in GSM5218925 (Figure 3C).

Expression trend and cluster analysis module: This module provides a comprehensive visualization of gene expression dynamics during early embryonic development in three livestock species: cattle, pigs, and goats, which facilitates in-depth exploration of transcriptional patterns across developmental stages via an integrated gene retrieval system. It incorporates well-curated early embryonic development gene sets from the literature, enabling users to refine their analyses by selecting genes of interest via the "Gene List" interface. The developmental stages are systematically classified based on spatiotemporal attributes, encompassing oocytes, zygotes, 2-cell, 4-cell, 8-cell, 16-cell, morulae, blastocysts, inner cell mass, and trophectoderm. Beyond embryonic development, the module supports transcriptomic profiling across diverse tissues and organs in multiple livestock species, allowing comparative analyses of gene expression patterns across phylogenetic and anatomical contexts. Unlike the precisely orchestrated temporal transitions in early embryogenesis, gene expression in tissues and organs is predominantly shaped by spatial heterogeneity. To enhance the resolution of tissue-specific expression patterns, tissues are meticulously categorized, such as subdividing the cattle kidney into cortex and medulla, to delineate the transcriptional landscapes of developmental genes. Additionally, the module integrates a robust clustering framework for cell and tissue-level analyses, leveraging PCA for linear dimensionality reduction and UMAP for nonlinear dimensionality reduction in single-cell data. These approaches generate high-resolution clustering maps that distinguish distinct cellular or tissue subpopulations within developmentally relevant datasets. Based on clustering outcomes, differentially expressed gene sets are systematically extracted from various cell and tissue groups for downstream analysis. Furthermore, the module prioritizes the top 20 highly differentially expressed genes between sample groups, visualizing their expression fluctuations through clustering, with fold changes quantified by $\log_2$FC. Genes exhibiting similar expression trends were clustered and visualized (Figure 3D).

Cross-species homologous gene conservation, annotation, and expression module: This module enables users to explore homologous genes across multiple livestock species, providing a comprehensive framework for conservation analysis, functional annotation, and expression comparison. Users can quickly locate target genes through the Homologous Gene Browser (top left), which allows filtering by the number and combination of conserved species. For example, *FOXL2* is identified as a homologous gene shared by cow, sheep, and horse in the Homologous Gene List panel (top right). Detailed biological information, such as gene identifiers, descriptions, orthologous relationships, and statistical parameters, is displayed in the Gene Annotation Details panel (bottom left). Furthermore, the Cross-Species Gene Expression panel (bottom right) enables visualization of *FOXL2* expression dynamics across embryonic stages and tissues in different species, facilitating comparative analysis of gene regulation during development. Together, these features make the module a powerful resource for investigating cross-species conservation and functional divergence of homologous genes (Figure 3E).

Stage-specific gene scoring module: This module provides a systematic framework for quantifying and visualizing the stage-specific importance of genes during early embryonic development. The Developmental Gene Trajectory panel (top left) displays the temporal scoring trajectory of a user-defined

gene across developmental stages, with species selection and gene search functionalities. The Multi-species Gene Comparison panel (top right) enables direct comparison of scoring trajectories for the same gene across multiple species, thereby facilitating cross-species analyses of developmental regulation. The Top Genes Heatmap panel (bottom left) allows users to define the number of top-ranked genes, generating heatmaps that capture stage-specific scoring patterns within a selected species. Complementarily, the Stage-specific Gene Ranking panel (bottom right) provides ranked bar plots of the top 30 genes for a chosen developmental stage and species, enabling rapid identification of key regulators. Collectively, these features establish the module as a powerful tool for uncovering dynamic and stage-specific gene regulation across species (Figure 3F).

Comprehensive statistics and download services for genes, projects, samples, and literature: To enhance data accessibility and analytical utility, the database provides open access to comprehensive project- and sample-level metadata associated with all collected datasets. The “Projects” page systematically presents the BioProject ID, project title, number of associated samples, as well as corresponding tissue and cell type information for each dataset. It also supports page-level navigation, allowing users to access a detailed “DatasetInfo” subpage by clicking the “View” button for further exploration of project-specific content. In parallel, the “Samples” page offers fine-grained metadata for each individual sample, including BioSample ID, sequencing series, and library strategy, enabling precise retrieval and filtering. At the gene level, the “Gene” page integrates and displays all currently annotated genes from five domesticated mammalian species, including Gene ID, Gene Symbol, Gene Type, and functional descriptions. This module is equipped with flexible query options, allowing users to refine their search by selecting specific organisms and gene types through dropdown menus or by directly inputting gene names into the search bar for targeted retrieval. In addition to biological data, the database also compiles an extensive collection of literature, encompassing 1 594 peer-reviewed publications related to developmental biology in five livestock species. The “Literature” page organizes references by species and provides key metadata such as article title, authors, journal name, publication date, PMID, and DOI. Users can access the full entry on PubMed via clickable PMID links and view article abstracts through an expandable interface. Notably, all curated and processed datasets are freely available for download, and users can access the dedicated “Download” section via the main navigation bar to retrieve raw data, metadata files, and related documentation efficiently (Figure 3G).

Characterization and comparative analysis of *TBX3* in early embryonic development across Livestock species

The *TBX3* gene plays a crucial role in early embryonic development across various livestock species, participating in key processes such as zygotic genome activation (Li et al., 2025), cell differentiation (Esmailpour & Huang, 2012), and lineage segregation (Wei et al., 2017). We performed annotation of the *TBX3* gene using LivestockDev, where essential information, including the gene symbol, gene type, and associated GO terms, is displayed (Figure 4A). Additionally, the protein structure of *TBX3* was analyzed and illustrated (Figure 4B). A previous study reported elevated expression of *TBX3* and *FGFR4* in late-stage cattle preimplantation embryos (Wei et al., 2017). Using the newly developed visualization tool on our platform, we validated these expression patterns in cattle and further identified analogous trends in porcine and caprine embryos (Figure 4C). Interestingly, *TBX3* transcription was first detected at the 8-cell stage across all three species, indicating this phase may mark a critical developmental transition associated with pluripotency activation. Gene interaction network of *TBX3* demonstrated direct/indirect connectivity with core pluripotency regulators (*SOX2*, *KLF4*, *NANOG*, *ESRRB*; Figure 4D). Through our platform, we delineated temporal expression profiles of these factors during embryogenesis. *KLF4* and *ESRRB* exhibited pronounced upregulation in late preimplantation phases (16-cell to morula stages), corroborating their proposed roles in pluripotency initiation and maintenance (Lavagi et al., 2018). Notably, interspecies divergence emerged in specific gene regulatory networks. While *BMP4*—a pivotal determinant of cell fate operating synergistically with the transcription cofactor *POUSFI* (Wang et al., 2012)—was ubiquitously upregulated during early embryogenesis across species, cattle *BMP4* uniquely maintained persistent expression through the morula stage (Figure 4F). This interspecies variability implies evolutionary divergence in developmental regulation mechanisms. Leveraging our integrated epigenomic mapping pipeline, we discovered a potential enhancer in the cattle *TBX3* locus (Chr17: 60 086 087-60 091 678, intron 2) specifically active in IVF-derived embryonic stem cells. This region exhibited pronounced enrichment of active histone marks (H3K4me3, H3K27ac; Figure 4E), suggesting its role in increasing the transcriptional level of *TBX3* during embryogenesis.

In the analysis of gene expression trends within Cluster 4 in goats, we identified the *TBX3* gene, whose expression progressively increased in parallel with the advancement of embryonic development (Figure 4G), corroborating the findings presented in Figure 4C. This finding strongly suggests that *TBX3* plays a pivotal role in orchestrating gene regulatory networks during early embryonic development. Furthermore, we discovered a cohort of genes exhibiting expression trends strikingly similar to those of *TBX3* (Figure 4H). The parallel expression profiles of these genes imply that they may function in concert with *TBX3* in driving essential processes during embryonic development. Further analysis of these genes will be crucial in elucidating the molecular mechanisms underlying *TBX3*-mediated regulatory interactions and its role in embryonic development. LivestockDev provides comprehensive visualization of gene expression dynamics throughout embryonic development and facilitates in-depth investigation of differential *TBX3* gene expression across tissues and organs. Our findings reveal that *TBX3* exhibits elevated expression levels in the lung, anterior prostate, prostate, and thyroid tissues of cattle (Figure 4I). In pigs, *TBX3* also showed high expression level in the small intestine and lung (Supplementary Figure S3), and the consistently high expression level in lung across both species suggests a conserved role of *TBX3* in regulating pulmonary function. Previous research has demonstrated that *TBX3* plays a pivotal role in regulating lung development in mice (Douglas & Papaioannou, 2013). This suggests that *TBX3* may have similar physiological and regulatory functions in the lung and other tissues with high expression level in cattle. Our results align with previous studies, supporting the hypothesis that *TBX3* high expression in these organs is linked to its critical regulatory roles in physiological processes.

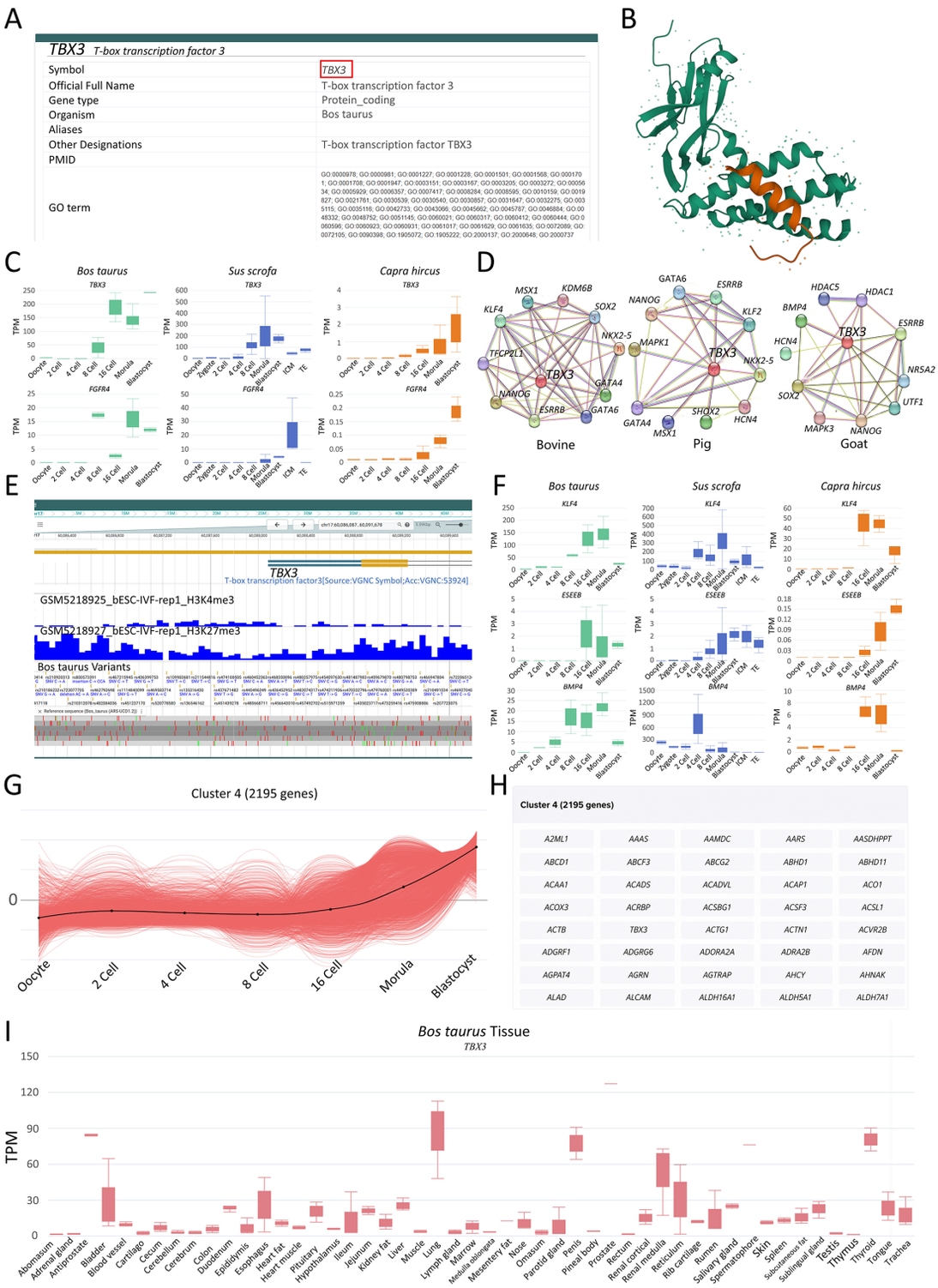

Figure 4 Functional Analysis of Key Genes in Early Embryonic Development

A: Detailed annotation of the *TBX3* gene, showing the gene symbol, type, and associated GO terms. B: Structural visualization of the TBX3 protein, with orange highlighting predicted secondary elements within the 3D conformation. C: Expression patterns of *TBX3* and *FGFR4* during early embryonic development in cattle, pigs, and goats. Boxplots use distinct colors for each species (green: cattle, blue: pig, orange: goat), and y-axis values represent transcript levels (TPM). D: Gene interaction networks of *TBX3* in cattle, pigs, and goats. Nodes represent genes and edges indicate functional or regulatory associations, with connection density reflecting interaction confidence. *TBX3* shows direct or indirect links to pluripotency regulators. E: Epigenetic mapping at the cattle *TBX3* locus. Blue tracks denote H3K4me3 and H3K27ac histone modification peaks in IVF-derived embryonic stem cells, overlapping an intronic enhancer region (Chr17:60 086 087–60 091 678). F: Expression patterns of *KLF4*, *ESRRB*, and *BMP4* across developmental stages in cattle, pigs, and goats. Boxplot colors follow the same species coding as in panel C. While *BMP4* is upregulated early in all species, cattle uniquely maintain high expression through the morula stage. G: Gene cluster analysis in goats, showing Cluster 4 (red lines) with a general upward trajectory of expression during development. *TBX3* belongs to this cluster, supporting its progressive activation during embryogenesis. H: Identification of genes with expression patterns similar to *TBX3* across species, displayed in tabular format. I: Tissue-specific expression of *TBX3* in cattle, shown as red boxplots for individual tissues.

Collectively, this work establishes a multidimensional framework integrating transcriptional dynamics, epigenetic landscapes, and interspecies variability during livestock embryogenesis. The developed platform and datasets provide critical resources for advancing comparative embryology and assisted reproductive technologies.

Stage-specific and Cross-species dynamics of *SOX2* and *NANOG*

SOX2 and *NANOG* are core pluripotency factors that play indispensable roles in early embryogenesis by regulating cell fate specification and sustaining stem cell identity (Aguila et al., 2022). To investigate their cross-species and stage-specific dynamics, we systematically analyzed their temporal expression patterns using our integrated platform.

In cattle, *NANOG* expression exhibited a sharp surge at the 8-cell stage, decreased at the 16-cell stage, and subsequently increased again to reach its maximum level at the blastocyst stage. In contrast, *SOX2* showed an earlier peak at the 4-cell stage, progressively declined toward the morula stage, and was reactivated during blastocyst formation. In goat embryos, *NANOG* expression initiated earlier, with peak levels detected at the 16-cell stage, whereas *SOX2* displayed maximal expression at the morula stage. This late-cleavage stage activation pattern is consistent with the established roles of both factors in maintaining the inner cell mass and pluripotent lineage (Aguila et al., 2022; Kim et al., 2025). Within each species, the trajectories of *SOX2* and *NANOG* displayed coordinated yet distinct activation dynamics, underscoring their complementary contributions to preimplantation development (Figure 5A). To further quantify these stage-specific dynamics, we employed our gene scoring tool. The scoring curves corroborated the temporal expression peaks observed in Figure 5A: *NANOG* attained its maximum score at the 16-cell stage, whereas *SOX2* scored highest at the morula stage, consistent with their respective expression trajectories. These results indicate that *NANOG* activation precedes that of *SOX2* by one developmental stage. Genes with the highest dynamic scores were significantly enriched for established pluripotency regulators, with *SOX2* and *NANOG* prominently highlighted. Although the overall trends in the scoring profiles closely mirrored the raw expression patterns, subtle discrepancies were observed, such as minor shifts in the precise stage of peak scoring relative to expression curves. Such differences may reflect stage-specific regulatory mechanisms, potentially involving post-transcriptional regulation or sensitivity of the scoring algorithm, while the general concordance between score and expression underscores the robustness of the approach (Figure 5B).

In our Gene Clustering tool, *NANOG* was identified in bovine cluster 2 and *SOX2* in bovine cluster 1, while in goat, *NANOG* and *SOX2* were found in clusters 3 and 4, respectively. The expression trajectories of these clusters mirrored the single-gene patterns shown in Figure 5A. For instance, bovine *NANOG* exhibited a rise–fall–rise profile across preimplantation stages, while caprine *NANOG* peaked at the 16-cell stage. Such concordance indicates that the clustering framework reliably captures stage-specific transcriptional dynamics, and the platform allows direct exploration of these gene sets (Figure 5C). In cattle, *SOX2* and *NANOG* are predominantly enriched in reproductive and lymphoid tissues, with particularly high levels in the epididymis, testis, and thymus, consistent with their established roles in germ cell development and immune-associated processes (Lenz et al., 2022). Importantly, *SOX2* demonstrates a wider spectrum of tissue activity than *NANOG*, with pronounced enrichment in the bladder and anterior prostate, whereas *NANOG* expression remains largely confined to the reproductive/immune axis. In addition, *SOX2* exhibits substantially elevated transcriptional abundance relative to *NANOG* across multiple tissues, indicative of a more pervasive activation profile. Collectively, these patterns underscore gene- and tissue-specific modes of regulatory control, with *SOX2* exerting a broader functional influence in cattle (Figure 5D).

In summary, *SOX2* and *NANOG* display coordinated yet distinct stage-specific dynamics during early embryogenesis, with *NANOG* generally preceding *SOX2*. While these temporal patterns are largely conserved across cattle and goat, species-specific shifts highlight nuanced regulatory control. Gene clustering and scoring analyses corroborate these dynamics and emphasize enrichment of key pluripotency regulators. Tissue-specific profiling further reveals a broader functional spectrum for *SOX2* compared with *NANOG*, underscoring differential roles beyond pluripotency. Collectively, these results provide an integrated, cross-species perspective on the spatiotemporal regulation of core pluripotency factors in livestock.

DISCUSSION

We developed LivestockDev by integrating big data analytics, cloud computing, and bioinformatics, creating a web-based platform for integrating and analyzing multi-omics data on early embryonic development in livestock. This platform enables the standardized bioinformatics analysis of 56 datasets (1 819 samples) from five major livestock species—cattle, goat, sheep, pig, and horse—by integrating multi-omics data from various sequencing technologies using Python and R. To address challenges in large-scale data storage and management, we implemented a MySQL relational database, archiving 10 embryonic developmental stages, 19 cell types, and 74 tissue or organ types. Additionally, functional annotations from authoritative databases, including GO, KEGG, and Ensembl, were manually curated, culminating in a structured database with 17.76 Tb of total storage capacity. Although LivestockDev integrates multi-omics data from five major livestock species, the number of available horse datasets remains relatively limited. This small sample size may reduce the robustness of intra-species analyses and weaken the reliability of cross-species comparisons involving horses. Nevertheless, we included horse data to provide a preliminary reference for comparative studies, and we anticipate that the coverage of horse datasets will be substantially expanded as more resources become publicly available in future updates of LivestockDev. Built on a three-tier system architecture (frontend, backend, and database layers), LivestockDev offers a suite of multi-level analytical tools:

1. Cell-level analysis—clustering, differential gene expression, and early embryonic development profiling.
2. Tissue-level analysis—multi-tissue developmental gene expression analysis.
3. Genome-level analysis—a genome browser for livestock reference genomes, epigenetic signals, and SNV sites.
4. Gene-level analysis—classification of development-related gene sets with functional annotation.
5. Sequence-level analysis—access to CDS and AA

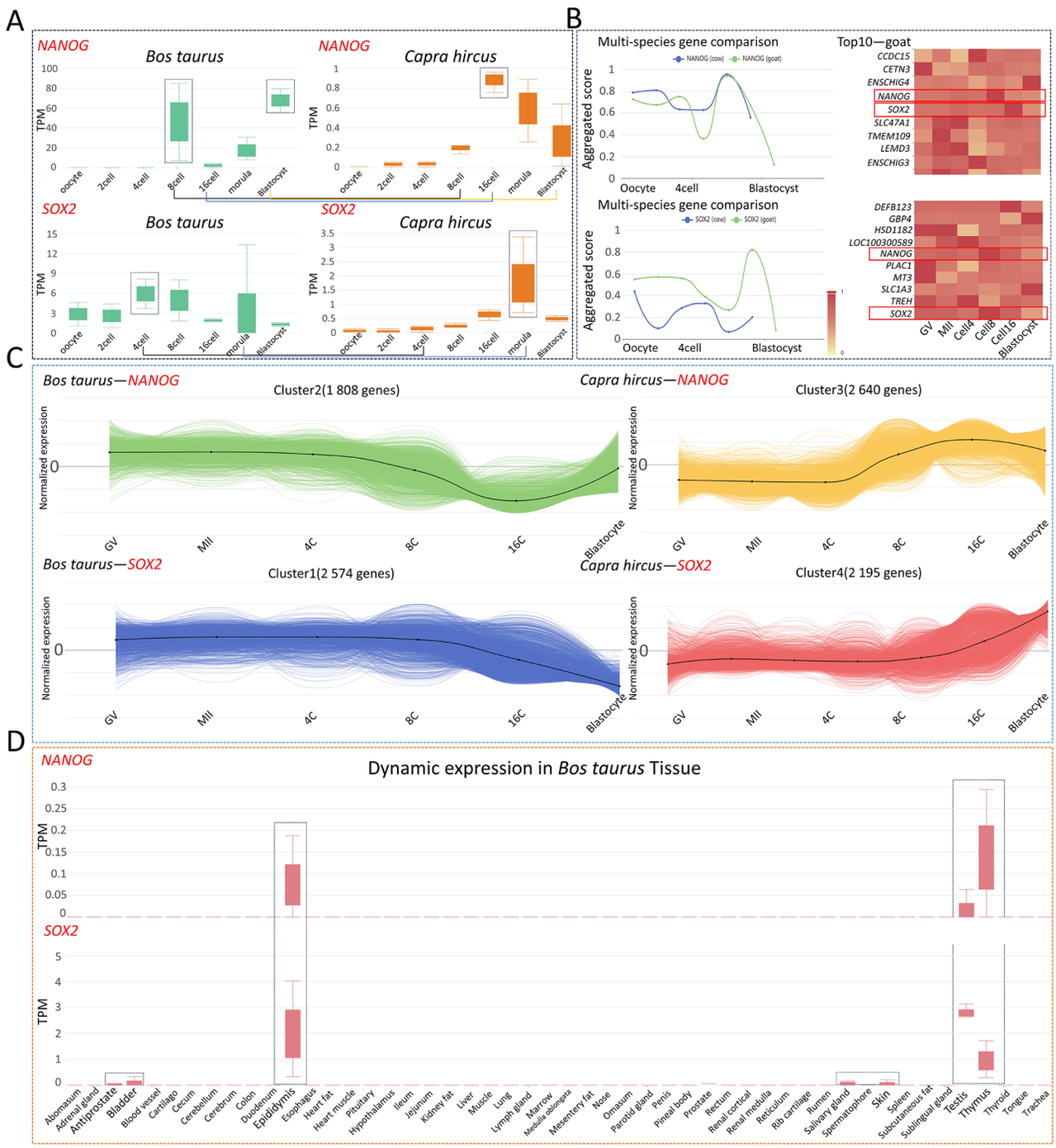

Figure 5 Comparative characterization of *SOX2* and *NANOG* across species, developmental stages, and tissues

A: Temporal expression trajectories of *SOX2* and *NANOG* during preimplantation development in cattle (green boxplots) and goat (orange boxplots). *NANOG* activation occurs earlier than *SOX2*, highlighting their coordinated yet distinct temporal dynamics. B: Gene scoring curves recapitulate stage-specific peaks of *NANOG* (blue: cow, green: goat) and *SOX2* (blue: cow, green: goat), confirming the developmental timing of their activation. Heatmaps display the top 10 ranked genes in goat (top) and cow (bottom), with red indicating higher scores and yellow lower scores, highlighting enrichment of pluripotency-associated regulators including *NANOG* and *SOX2*. C: Gene clustering analysis identifies stage-specific transcriptional clusters corresponding to *NANOG* and *SOX2*. In bovine embryos, *NANOG* belongs to Cluster 2 (green curves) and *SOX2* to Cluster 1 (blue curves), while in goat embryos they are assigned to Clusters 3 (yellow) and 4 (red), respectively. Each thin curve represents an individual gene trajectory, with the bold black line denoting the cluster trend. D: Tissue-specific expression of *SOX2* and *NANOG* in cattle, shown as red boxplots for individual tissues. *SOX2* is more broadly expressed, with high levels in epididymis, testis, thymus, bladder, and anterior prostate, whereas *NANOG* expression is largely confined to reproductive and immune tissues.

sequences of development-related genes.

In summary, LivestockDev integrates high-throughput sequencing data related to early embryonic development, multi-tissue development, cloning, and assisted reproduction in livestock. By explicitly comparing with existing databases, LivestockDev highlights its innovations in multi-omics integration and cross-species analysis. By enabling comprehensive cross-species comparisons, this platform provides a robust data-driven foundation for understanding livestock growth, development, and reproduction, ultimately

offering valuable theoretical insights for breeding improvement.

In the future, the LivestockDev multi-omics database for early embryonic development in livestock will continue to expand, with plans to incorporate new omics data types to enrich its content and functionality. Building on the systematic collection of multi-omics data from various tissues and organs, LivestockDev aims to construct a comprehensive gene expression atlas across tissues, developmental stages, and species. This will enable the identification of tissue-specific and shared genes, dynamically and stably expressed genes across development, as well as differentially expressed gene groups and co-expression modules across species. As high-quality early embryonic expression data from sheep and horses are collected and integrated, the database will further enhance its gene expression visualization tools for early embryonic development in livestock. Additionally, with the supplementation of multi-tissue gene expression data from more species, particularly from goats and horses, as exemplified by the published single-cell transcriptomic atlas of the goat testis (Yu et al., 2021), LivestockDev will strengthen its capability to analyze multi-tissue developmental processes in livestock. This study will also explore potential species-specific transcription factors and their roles in development, while further investigating the stage-specificity of gene alternative splicing events. In forthcoming studies, we will apply machine learning to identify stage-specific genes during early embryogenesis. Processed expression matrices from different developmental stages will serve as inputs, and models such as regularized regression, tree-based methods, and neural networks will be trained to capture discriminative gene features. Feature importance analyses (SHAP values) will then highlight candidate stage-specific genes, which will be validated by stability tests and functional enrichment analyses. This approach aims to uncover novel regulatory genes and provide interpretable results for further biological validation.

DATA AVAILABILITY

The processed datasets used in this work are available at the LivestockDev database website: http://bioinfor.imu.edu.cn/livestockDev.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

P.C.S. and Y.C.L. contributed to both the front-end and back-end development of the database and performed data analysis. Q.Y.C. and X.H. provided critical revisions to the manuscript, enhancing its clarity, coherence, and overall quality. H.S.L. and P.W.H. handled preliminary data processing. Y.Z., G.F.C., and Y.C.Z designed the research framework, secured funding, supervised the entire project, and ensured coordination among all team members. All authors read and approved the final version of the manuscript.

REFERENCES

Aguila L, Osycka-Salut C, Treulen F, et al. 2022. Pluripotent core in bovine embryos: a review. *Animals*, **12**(8): 1010.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.

Bastian FB, Roux J, Niknejad A, et al. 2021. The Bgee suite: integrated curated expression atlas and comparative transcriptomics in animals. *Nucleic Acids Research*, **49**(D1): D831–D847.

Bazer FW, Johnson GA. 2024. Early embryonic development in agriculturally important species. *Animals*, **14**(13): 1882.

Berman HM, Westbrook J, Feng ZK, et al. 2000. The protein data bank. *Nucleic Acids Research*, **28**(1): 235–242.

Buels R, Yao E, Diesh CM, et al. 2016. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biology*, **17**: 66.

Chen T, Chen X, Zhang S, et al. 2021. The Genome Sequence Archive Family: toward explosive data growth and diverse data types. *Genomics, Proteomics & Bioinformatics*, **19**(4): 578–583.

Clough E, Barrett T, Wilhite SE, et al. 2024. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Research*, **52**(D1): D138–D144.

Douglas NC, Papaioannou VE. 2013. The T-box transcription factors TBX2 and TBX3 in mammary gland development and breast cancer. *Journal of Mammary Gland Biology and Neoplasia*, **18**(2): 143–147.

Esmailpour T, Huang TS. 2012. TBX3 promotes human embryonic stem cell proliferation and neuroepithelial differentiation in a differentiation stage-dependent manner. *Stem Cells*, **30**(10): 2152–2163.

Heinz S, Benner C, Spann N, et al. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, **38**(4): 576–589.

Hocquette J F, Lehnert S, Barendse W, et al. 2007. Recent advances in cattle functional genomics and their application to beef quality. *Animal*, **1**(1): 159–173.

Hu BS, Zheng L, Long CS, et al. 2019. EmExplorer: a database for exploring time activation of gene expression in mammalian embryos. *Open Biology*, **9**(6): 190054.

Huang TH, Huang XL, Shi BM, et al. 2019. GEREDB: gene expression regulation database curated by mining abstracts from literature. *Journal of Bioinformatics and Computational Biology*, **17**(4): 1950024.

Katz K, Shutov O, Lapoint R, et al. 2022. The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Research*, **50**(D1): D387–D390.

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.

Kim TG, Choe YH, Kim SH, et al. 2025. Increased apoptosis in late-developing *in vitro* fertilized bovine blastocysts decreases successful pregnancy. *Animal Bioscience*, **38**(3): 454–465.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**(11): 1571–1572.

Krueger F. 2012. Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries.

Lavagi I, Krebs S, Simmet K, et al. 2018. Single-cell RNA sequencing reveals developmental heterogeneity of blastomeres during major genome activation in bovine embryos. *Scientific Reports*, **8**(1): 4071.

Lenz J, Konecna P, Tichy F, et al. 2022. Unique expression patterns of the embryonal stem cell marker SOX2 and hormone receptors suggest the existence of a subpopulation of epithelial stem/progenitor cells in porcine and bovine endometrium. *Veterinary Medicine and Science*, **8**(4): 1489–1501.

Li DQ, Mei HH, Shen Y, et al. 2018. ECharts: a declarative framework for rapid construction of web-based visualization. *Visual Informatics*, **2**(2): 136–146.

Li H, Hou JY, Chen ZY, et al. 2021. FifBase: a comprehensive fertility-associated indicators factor database for domestic animals. *Briefings in Bioinformatics*, **22**(5): bbaa432.

Li XH, Lee SH, Kim JD, et al. 2025. TBX3 is essential for zygotic genome activation and embryonic development in pigs. *Microscopy and*

Microanalysis, **31**(1): ozae123.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**(7): 923–930.

Liu JD, Lang K, Tan SX, et al. 2022. A web-based database server using 43, 710 public RNA-seq samples for the analysis of gene expression and alternative splicing in livestock animals. *BMC Genomics*, **23**(1): 706.

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.

Ortega NM, Winblad N, Reyes AP, et al. 2018. Functional genetics of early human development. *Current Opinion in Genetics & Development*, **52**: 1–6.

Park S, Lee J, Kim J, et al. 2023. Benchmark study for evaluating the quality of reference genomes and gene annotations in 114 species. *Frontiers in Veterinary Science*, **10**: 1128570.

Picelli S, Faridani OR, Björklund ÅK, et al. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols*, **9**(1): 171–181.

Ramírez F, Dündar F, Diehl S, et al. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*, **42**(W1): W187–W191.

Rossant J. 2018. Genetic control of early cell lineages in the mammalian embryo. *Annual Review of Genetics*, **52**: 185–201.

Rossant J, Tam PPL. 2017. New insights into early human development: lessons for stem cell derivation and differentiation. *Cell Stem Cell*, **20**(1): 18–28.

Satoh N. 1982. Timing mechanisms in early embryonic development. *Differentiation*, **22**(1-3): 156–163.

Sayers EW, Beck J, Bolton EE, et al. 2025. Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Research*, **53**(D1): D20–D29.

Schall PZ, Ruebel ML, Midic U, et al. 2019. Temporal patterns of gene regulation and upstream regulators contributing to major developmental transitions during Rhesus macaque preimplantation development. *Molecular Human Reproduction*, **25**(3): 111–123.

Sehnal D, Bittrich S, Deshpande M, et al. 2021. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Research*, **49**(W1): W431–W437.

Shamimuzzaman M, Le Tourneau JJ, Unni DR, et al. 2020. Bovine Genome Database: new annotation tools for a new reference genome. *Nucleic Acids Research*, **48**(D1): D676–D681.

Szklarczyk D, Gable AL, Nastou KC, et al. 2021. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, **49**(D1): D605–D612.

The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, **47**(D1): D506–D515.

Wang Z, Oron E, Nelson B, et al. 2012. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell*, **10**(4): 440–454.

Wei QQ, Zhong L, Zhang SP, et al. 2017. Bovine lineage specification revealed by single-cell gene expression analysis from zygote to blastocyst. *Biology of Reproduction*, **97**(1): 5–17.

Wolf FA, Angerer P, Theis FJ. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, **19**(1): 15.

Wu XL, Zhu ZS, Xiao X, et al. 2021. LIN28A inhibits DUSP family phosphatases and activates MAPK signaling pathway to maintain pluripotency in porcine induced pluripotent stem cells. *Zoological Research*, **42**(3): 377–388.

Yu XW, Li TT, Du XM, et al. 2021. Single-cell RNA sequencing reveals atlas of dairy goat testis cells. *Zoological Research*, **42**(4): 401–405.

Zheng GXY, Terry JM, Belgrader P, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, **8**: 14049.