Supplementary Information

Single-cell transcriptome and metagenome profiling reveal the genetic basis of rumen functions and convergent developmental patterns in ruminants

Juan Deng, Ya-Jing Liu, Wen-Tian Wei,Qi-Xuan Huang, Li-Ping Zhao, Ling-Yun Luo, Qi Zhu, Lin Zhang, Yuan Chen, Yan-Lin Ren, Shan-Gang Jia, Yu-Luan Lin, Ji Yang, Feng-Hua Lv, Hong-Ping Zhang, Feng-E Li, Li Li, and Meng-Hua Li

Correspondence: menghua.li@cau.edu.cn (M.H. L.), lily@sicau.edu.cn (L. L.)

This PDF file includes:

1. Supplementary Methods

Single-cell suspension preparation

After preservation in the MAC tissue storage solution (Miltenyi Biotec, Bergisch Gladbach, Germany) for up to 24 h, each biopsy sample was minced into small pieces with Iris scissors and digested in 0.25% trypsin with RPMI-1640 solution [Gibco™, Thermo Fisher Scientific (China) Co., Ltd, Shanghai, China] for 10 min–30 min at 37°C. After centrifugation (300 rpm) for 5 min, the first precipitates were collected and washed with PBS containing 10% BSA. After centrifugation (300 rpm) for 5 min, secondary precipitates were digested in mixed solution with 2 mg/ml collagenase I and II (Sigma, St Louis, MO, USA) containing RPMI-1640 (Gibco™) for 30 min–1.5 h at 37°C. The duration of the two digestion steps should be adjusted according to the digestion state of the rumen samples at different timepoints. The cell suspension was then passed through a 70 μm nylon cell strainer (BD Falcon, BD Biosciences, San Jose, CA, USA) to remove tissue debris and cell aggregates. After centrifugation (200 rpm) for 3 min and then centrifuging (300 rpm) the supernatant for 5 min, the final precipitates were resuspended in PBS, and a cell suspension with viability > 80% was used. Dead cells were eliminated to increase the efficiency of sorting robust and live cells for single-cell experiments using the MACS® Dead Cell Removal Kit (Miltenyi Biotec, Germany). Details were in Supplemental Table S1.

Single-cell RNA-seq library construction and sequencing

RNA barcoding from thousands of individual cells with a set of uniquely barcoded primers was performed using the 10× Genomics single-cell RNA sequencing system (10× Genomics, Pleasanton, CA, USA). First, cells in the sorted single-cell suspension were counted and diluted to the final concentration in DMEM or

DMEM/F12 medium supplemented with 10% fetal bovine serum (FBS) (Gibco, UK) prior to analysis. Single-cell suspensions were then normalized and loaded onto a Chromium Controller instrument (10× Genomics, USA) to generate single-cell gel-bead-in-emulsions (GEMs), targeting 8,000–10,000 cells at different developmental stages. Thus, individual cells were isolated into droplets with gel beads coated with unique primers bearing 10× cell barcodes, unique molecular identifiers (UMIs) and polyA sequences. Reverse transcription reactions for barcoded full-length cDNA amplification were performed followed by emulsion breaking using the recovery agent and cDNA clean up with DynaBeads MyOne Silane Beads (Thermo Fisher Scientific, China).

Bulk cDNA was amplified on a Biometra TProfessional Thermocycler Basic Gradient with 96-well Sample Block (Montreal Biotech Inc. Germany) using the following cycling conditions: initial denaturation at 98 °C for 3 min, followed by 11 cycles of 15 s at 98 °C, 20 s at 63 °C, and 1 min at 72 °C and a final 1 min at 72 °C. cDNA libraries were prepared using the Chromium Single Cell 3' Reagent v2 Chemistry Kit. The detailed protocol consisted of the following sequential steps: (1) fragmentation, end repair and A-tailing; (2) post fragmentation, end repair, A-tailing and double-sided size selection with SPRIselect; (3) adaptor ligation; (4) post ligation cleanup with SPRIselect; (5) sample index PCR; (6) post sample index PCR and double-sided size selection with SPRIselect; (7) post library construction quality control with the Agilent Bioanalyzer High Sensitivity chip; and (8) post library construction quantification by qPCR (Tombor et al. 2021).

Histological analysis

Frozen or normal rumen tissues for histological examination were fixed in 4% paraformaldehyde (Beijing Solarbio Science & Technology Co., Ltd. Beijing, China) at 4°C for 24 h, dehydrated in 70% ethanol for 5–10 min, and incubated in pure fresh xylene solution for 30 min. Afterward, the samples were embedded in paraffin blocks and cut into 5–7 μm thin sections using a rotary Leica RM2255 microtome (Leica, Nussloch, Germany). To prevent tissue detachment from the slides during the staining process, the samples were then transferred to 3-amino-propyltriethoxysilane (APES; ZSGB-BIO, Beijing, China)-coated slides.

Hematoxylin and eosin (H&E) staining was implemented following the routine procedures described previously (García et al. 2012). In summary, the slides were deparaffinized in 100% xylene solutions for 30 min, followed by rehydration in an ethanol/dH_2O series (90%, 70% and 50%, 5 min each). Then, the slides were stained with hematoxylin solution for 7 min and rinsed twice with distilled water for 5 min. To remove excess stain, the stained tissues were rinsed with 1% (v/v) HCl-ethanol solution for 3-5 sec, followed immediately by washing with 45°C water for 5 min. After dehydration, the slides were stained with 1% eosin ethanol solution and rinsed with 100% ethanol solution for 10 min. Finally, the slides were mounted with neutral resin as the mounting medium, and brightfield photographs were taken using an optical microscope (McAudi Industrial Group Co., Ltd., China).

scRNA-Seq data preprocessing

The sequencing raw base call (BCL) files generated by Illumina sequencers were demultiplexed into FASTQ format using the “cellranger mkfastq” function. The generated FASTQ files were then aligned to the sheep reference genome

Oar_rambouillet_v1.0 (RefSeq assembly accession GCF_002742125.1) or the goat reference genome *ARS1* (RefSeq assembly accession: GCF_001704415.1) to rebuild an index for read alignment using the 'mkref' function. After alignment, the "cellranger count" subcommand with the setting '--expect-cells 8000' was used to count the gene expression reads and the feature barcoding reads from a sample per GEM well. Finally, the rates of bases with a Phred score of ≥30 (Q30) in UMIs (unique molecular identifiers), the cell barcodes and the RNA reads were over 90% (Supplemental Table S2). We implemented the filtering of low-quality cells meeting the following filtering metrics: > 200 expressed genes, > 3 UMI counts, and transcripts in less than three cells (Ge et al. 2020; Ma et al. 2020). The number of fractions of mitochondrial genes per cell is listed in Supplemental Table S2.

Detect and filter potential doublets

The above data were processed by the following he steps: (1) 'paramSweep_v3' was used to implement PCA, and the resulting PC distance matrix was used to calculate the proportion of artificial nearest neighbors (pANN); (2) 'summarizeSweep' was used to compute the bimodality coefficient across the pN-pK (pN, the number of artificial doublets; pK, the neighborhood size) parameter space; (3) 'find.pK' was used to compute and visualize the mean-variance normalized bimodality coefficient (BCmvn) score for each pK value tested in the parameter sweep, and the optimal pK values were determined for each sample based on the BCmvn score; and (4) 'doubletFinder_v3' was used to generate artificial doublets from an existing sc-RNA-seq dataset, and the number of doublets for each sample is shown in Supplemental Table S2.

Cell clustering analysis of merged data by using Seurat

The workflow was as follows: (1) the gene expression data were ln-transformed and normalized to scale the sequencing depth to 10,000 molecules per cell using the 'NormalizeData' function; (2) the top highly variable genes (n = 3,000) from the datasets were selected using the default 'vst' selection method in the 'FindVariableFeatures' function; (3) feature scaling and centering were performed by the 'ScaleData' function; (4) PCA was applied to the set of top highly variable genes using the 'RunPCA' function; (5) the 'dimensionality' of the dataset was determined via the 'ElbowPlot' function; (6) the 'FindNeighbors' function was used to calculate the Jaccard index; (7) the 'FindClusters' function with 'resolution' = 0.6 was used to cluster the cells; and (8) visualization methods such as T-distributed stochastic neighbor embedding (t-SNE) and UMAP were performed by running the 'RunUMAP' and 'RunTSNE' functions using the same PCs as input features.

Analysis of differentially expressed genes

The 'FindMarkers' function with the settings "test.use="wilcox" || logfc.threshold = 0 || min.pct=0.1" was used to find up-regulated genes (logFC > 0.25 and $P_{adj} < 0.05$) and down-regulated genes (logFC < -0.25 and $P_{adj} < 0.05$) between the timepoints or stages. Additionally, we implemented the 'FindAllMarkers' function with settings "test.use="wilcox" || logfc.threshold = 0 || min.pct=0.1" to find up-regulated genes (LogFC > 0.25 and $P_{adj} < 0.05$) and down-regulated genes (LogFC < -0.25 and $P_{adj} < 0.05$) between the cell types at each timepoint or stage.

Single-cell trajectory and RNA velocity analysis

The workflow was as follows: (1) the 'newCellDataSet' function was used to upload

the processed data into Monocle's main class; (2) the functions ‘estimateSizeFactors’ and ‘estimateDispersions’ were used to estimate the “size factors” to normalize for differences in mRNA recovered across cells, and the "dispersion" values were used for differential expression analysis; (3) the ‘detectGenes’ function, with the setting “min_expr = 0.1”, was used to determined how many copies were expressed for a particular gene and how many genes were expressed in a given cell; (4) the ‘reduceDimension’ function was used for dimensionality reduction with the settings “max_components = 2 || norm_method = 'log' || num_dim = 20 || reduction_method = 'tSNE'”; and (5) the ‘setOrderingFilter’ function was used to mark genes that were used for clustering in subsequent calls to clusterCells; (6) The ‘clusterCells’ function was used to cluster cells into a specified number of groups with an unsupervised algorithm (by default, density peak clustering); (7) The ‘orderCells’ function was used to learn a "trajectory" describing the biological process and calculate where each cell fell within that trajectory. We took a "root" state from the e45 counts to specify the start of the trajectory; and (8) to find genes whose expression patterns varied according to pseudotime, we used the ‘differentialGeneTest’ function with the setting “fullModelFormulaStr = ‘~sm.ns (Pseudotime)’”.

The loompy.connect function was implemented to connect a .loom file. Loom-annotated matrices of the 10x dataset was then loaded and analyzed in R using the velocyto.R. Next, we normalized and clustered cells using pagoda2, estimated RNA velocity using gene-relative model with k = 20 cell kNN pooling, and top/bottom 2% expression quantiles for gamma fit, Finally, we visualized the velocity vectors in the UMAP embedding with differentiable velocity vector fields.

Transcription factor (TF)-target gene regulatory network analysis

The workflow was as follows: (1) the 'geneFiltering' function with the settings "minCountsPerGene = 3 * 0.0025 * ncol(exprMat) || minSamples=ncol(exprMat)*.0025" was used to filter genes; (2) The 'runCorrelation' function was used on the input expression matrix to calculate the Spearman correlation; (3) The 'runGenie3' function was used to identify potential TF targets based on coexpression; (4) The 'runSCENIC_1_coexNetwork2modules' function was used to convert the output from GENIE3 to coexpression modules; (5) The 'runSCENIC_2_createRegulons' function with the settings "coexMethod=c("w001", "w005", "top50", "top5perTarget", "top10perTarget", "top50perTarget")" was used to perform TF-motif enrichment analysis and identify the direct targets (regulons); (6) The 'runSCENIC_3_scoreCells' function was used to score regulons on the individual cells (AUCell); and (7) The 'runSCENIC_4_aucell_binarize' function was used to binarize AUCell.

Immunofluorescence staining

For immunofluorescence staining, rumen tissues were perfused with PBS and fixed in 4% paraformaldehyde (PFA) (~pH 7.4) for 48 h. Subsequently, the tissues were processed for paraffin embedding with an automatic tissue processor and embedded in paraffin wax blocks. After washing with polylysine three times and dehydrating following established protocols in a cold graded ethanol series (70%, 80%, 90% and 100%), the paraffin-embedded rumen tissue sections at a 3.5 μm thickness were deparaffinized in dewaxing medium (Baso Biotechnology, China). The tissue sections were then exposed to 1× citrate-based antigen retrieval solution (Sangon Biotech, China), microwaved until boiling at 100°C for 5 min, allowed to stand still for 5 min,

and then boiled for 5 min. Ruminal tissue sections were blocked with Immunostaining Blocking Dilution Buffer (Sangon Biotech, China) for 1 h. After incubation using immunostaining blocking buffer for 1 h, the sections were incubated overnight at 4°C in primary antibodies diluted using the Bond primary antibody diluent (Abcam, UK). The specimen tissues were then washed twice in PBS and incubated in the secondary antibody dilution (Abcam, UK) for 1 h at 37°C. After washing off the excess primary and secondary antibodies, the sections were stained with 4’,6-diamidino-2-phenylindole (DAPI, Beijing Solarbio Science & Technology AB).

DNA extraction, assessment of DNA integrity, concentration, and quality

DNA was extracted following well-established protocols involving repeated bead-beating plus column filtration (Yu and Morrison 2004). DNA integrity and concentration were assessed by electrophoresis on 1% agarose gels, and DNA quality was determined using a Qubit® 2.0 Fluorometer (ThermoFisher Scientific, MA, USA).

Rumen cells with *Prevotella copri* RNA sequencing

Total RNA was extracted from cultured cells using TRIzol (Invitrogen, Carlsbad, CA, USA) following the manual instruction. RNA quality and integrity were examined and evaluated by 1% agarose gels, NanoPhotometer® spectrophotometer (IMPLEN, CA, USA), and RNA Nano 6000 Assay Kit in Bioanalyzer 2100 system (Agilent Technologies, CA, USA). Then, 1 μg RNA was used for library preparation using the NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) as recommended by the manufacturer, with the index codes added to adaptors for multiplexing samples. Libraries were sequenced on an Illumina NovaSeq platform

and 150 bp paired-end reads were generated.

2. Supplementary Figures

Supplemental Figure S1

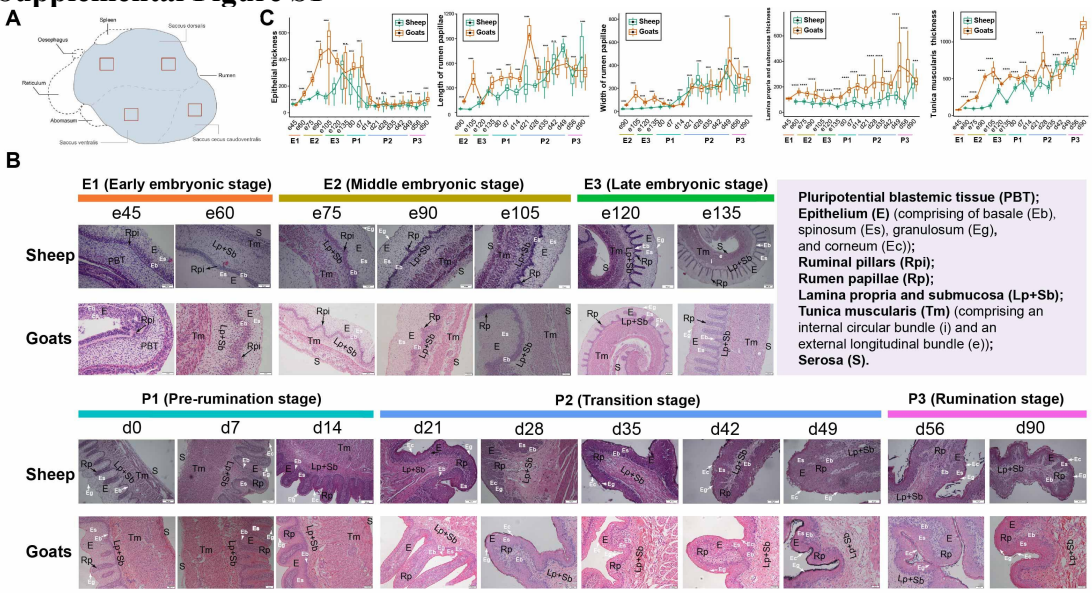


Fig. S1 (A) Schematic diagram of rumen structure. The red boxes indicate the sampling sites of the rumen tissue. **(B)** Histomorphometric measurements of rumen tissues of sheep and goats (embryonic days 45–135 and 0–90 postpartum); hematoxylin-eosin (HE) staining; bar: 50 μm, 100 μm and 200 μm. E: epithelium; Rp: rumen papillae; Rpi, ruminal pillars; Lp+Sb: lamina propria and submucosal tissue; Tm: tunica muscularis; S: serosa. **(C)** Morphometric analysis of rumen tissues in sheep and goats during embryonic and postnatal development (μm). The box plots show the five-number summary of a set of data, including the minimum value, 25% quantile (lower), median, 75% quantile (upper), and maximum value. Black dots represent the mean values of each variable and are connected by the polylines. Two hundred measurements have been taken for each variable such as epithelial thickness, length of rumen papillae, width of rumen papillale, lamina propria and submucosa thickness, and tunica muscularis thickness. n.s., not significant; *, $P < 0.05$; **, $P < 0.01$, ***, $P < 0.001$; ****, $P < 0.0001$.

Supplemental Figure S2

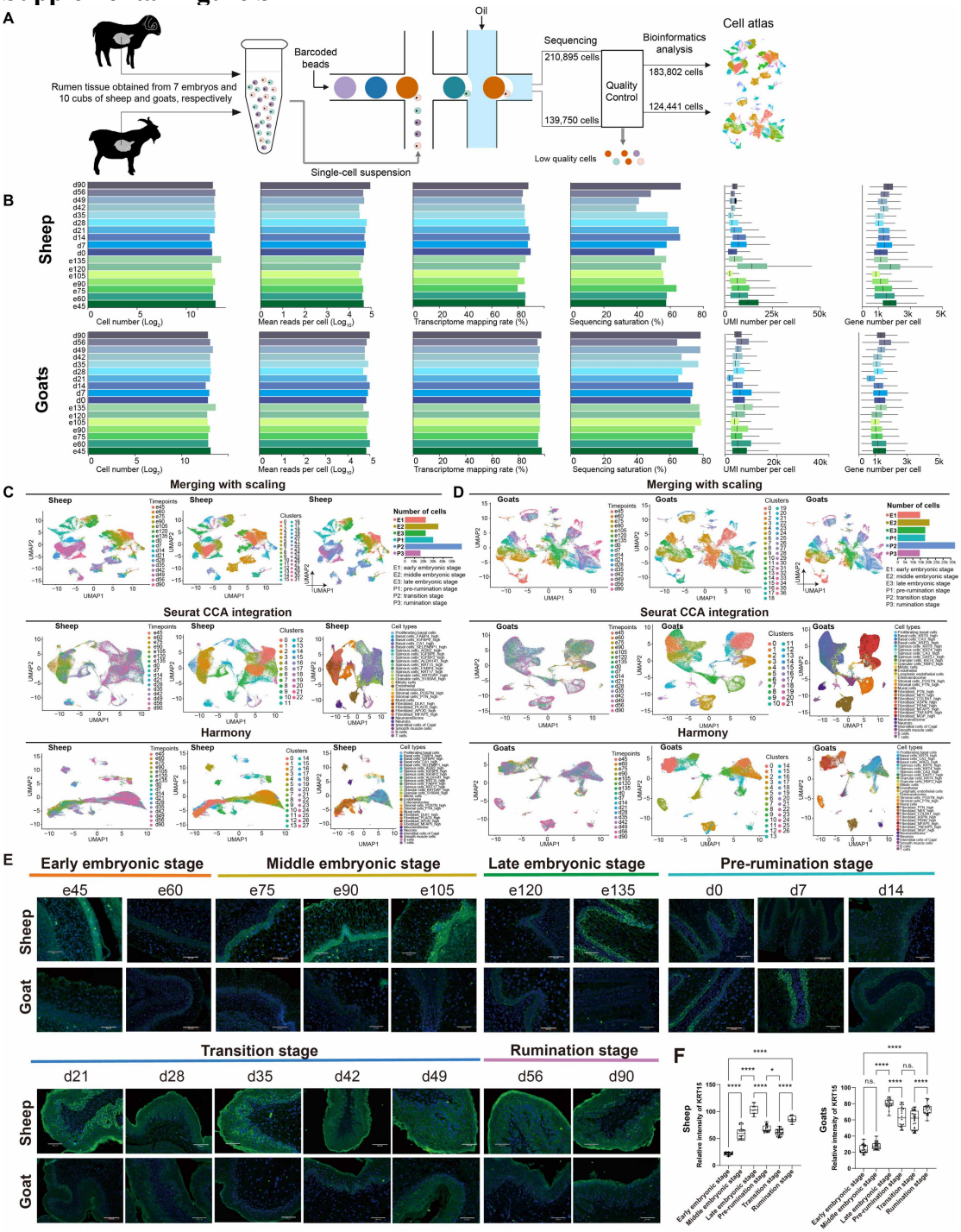


Fig. S2 (A) Flowchart of the scRNA-seq data analysis. **(B)** Cell number, mean reads per cell, transcriptome mapping rate, sequencing saturation, gene number and unique molecular identifier (UMI) number per cell for the scRNA-seq data at the seventeen timepoints of sheep and goat rumen. **(C, D)** UMAP plots, showing the cell clusters at the seventeen timepoints or six stages in sheep (C) and goats (D), were performed after batch correction by the merging with scaling, CCA integration in Seurat and harmony, respectively. The right panels shown the number of cells in each stage. **(E)** Immunohistochemical localization of selected key marker proteins encoded by KRT15 in the epithelial cells of rumen at different developmental stages. Experiments were repeated for 2–3 rumen slices per timepoint. Scale bars: 20 mm. **(F)** Box plots of relative immunofluorescence intensity of *KRT15* gene. Ten fields were randomly selected from each section at each developmental stage to measure the fluorescence intensity. The data are shown as means ± SEM (the standard error of the mean). *, $P < 0.05$; **, $P < 0.01$, ***, $P < 0.001$. n.s., not significant.

Supplemental Figure S3

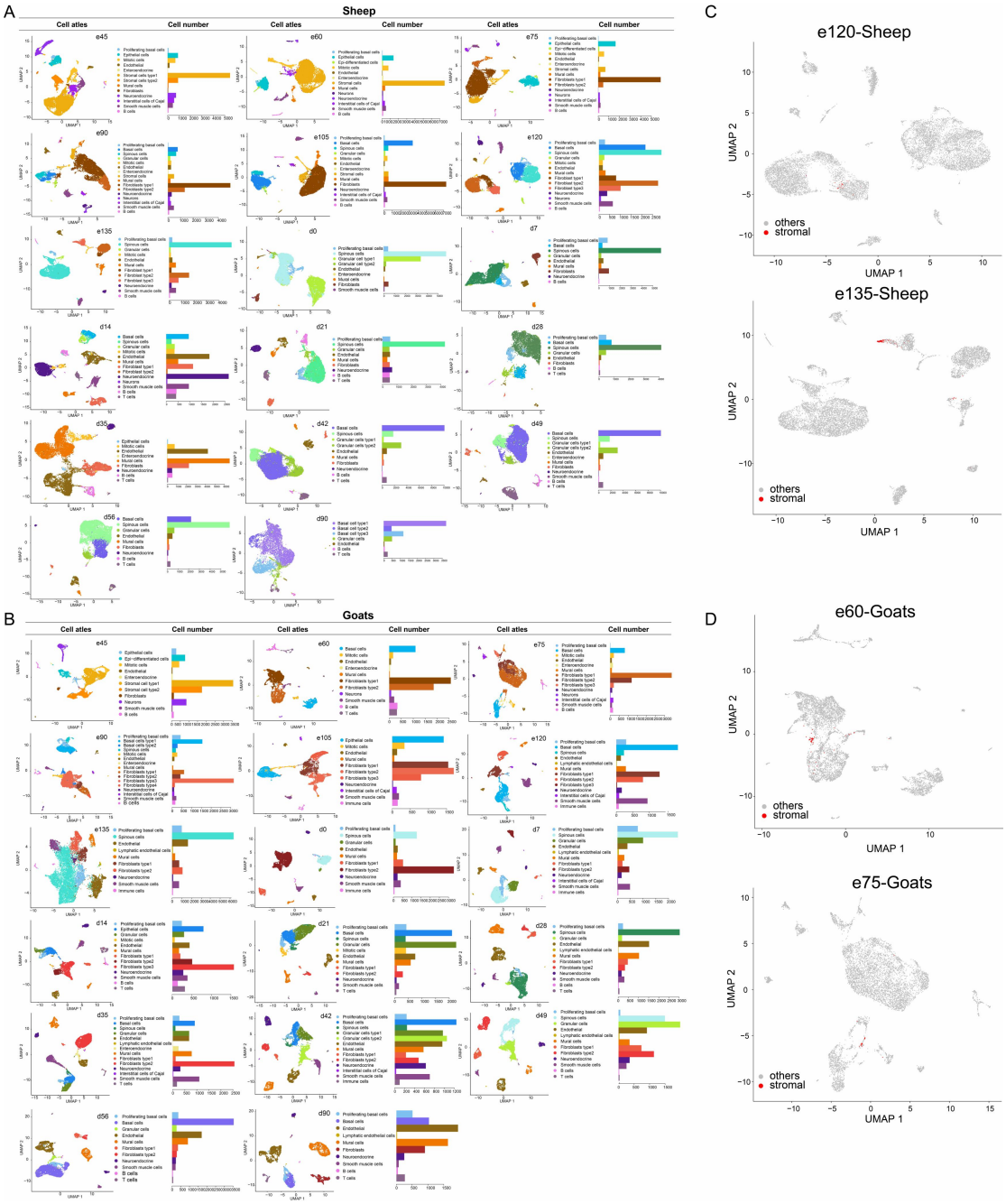


Fig. S3 (A, B) UMAP plots showing the cell types and the cell number of the types at the seventeen timepoints in sheep (A) and goats (B). **(C, D)** UMAP plots of individual stages in sheep (C) and goats (D), the cells marked red color points represent the stromal cells identified by the method of merging with scaling.

Supplemental Figure S4

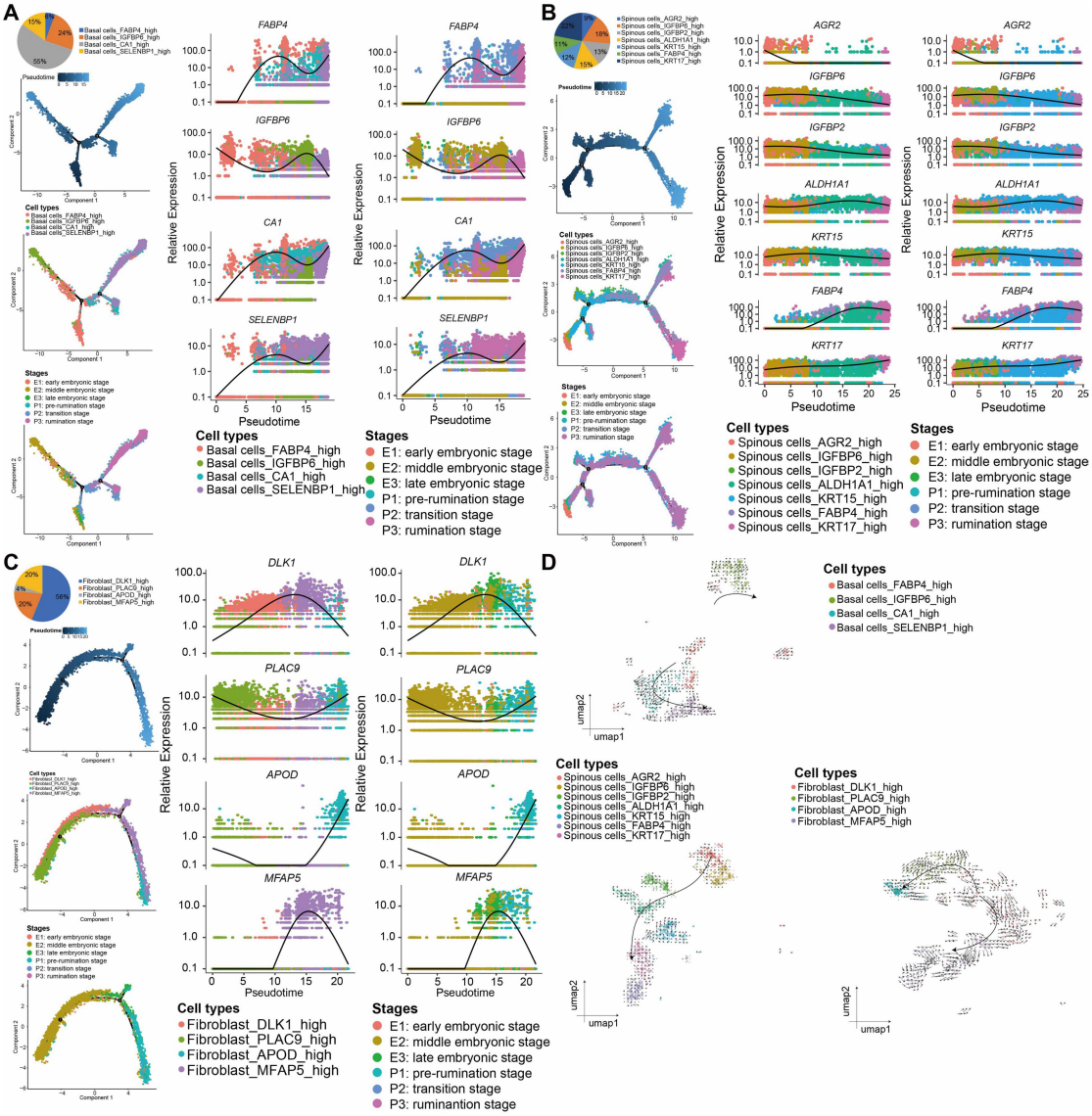


Fig. S4 Pseudotime analyses of cell subtypes in the rumen tissues of sheep.
Pie chart showing the relative percentages of cell subtypes in basel cells, spinous cells and fibroblasts, and pseudotime trajectory analysis of the differentiation of cell subtype cells and the expressions of marker genes in basel cells, spinous cells and fibroblasts, respectively. Cells are colored based on pseudotime, cell subtypes and developmental stages.

Supplemental Figure S5

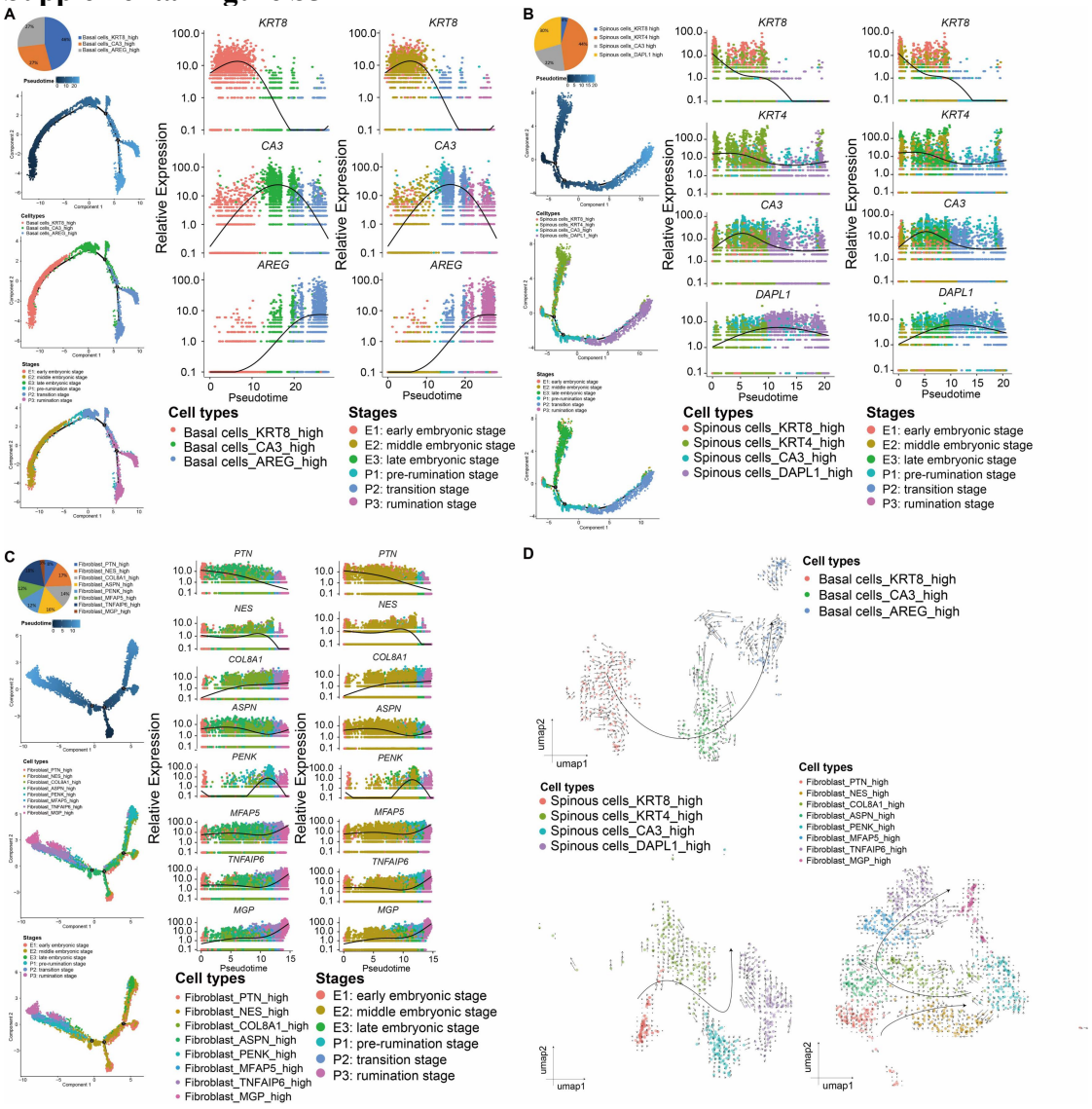


Fig. S5 Pseudotime analyses of cell subtypes in the rumen tissues of goats.

Pie chart showing the relative percentages of cell subtypes in basel cells, spinous cells and fibroblasts, and pseudotime trajectory analysis of the differentiation of cell subtype cells and the expressions of marker genes in basel cells, spinous cells and fibroblasts, respectively. Cells are colored based on pseudotime, cell subtypes and developmental stages.

Supplemental Figure S6

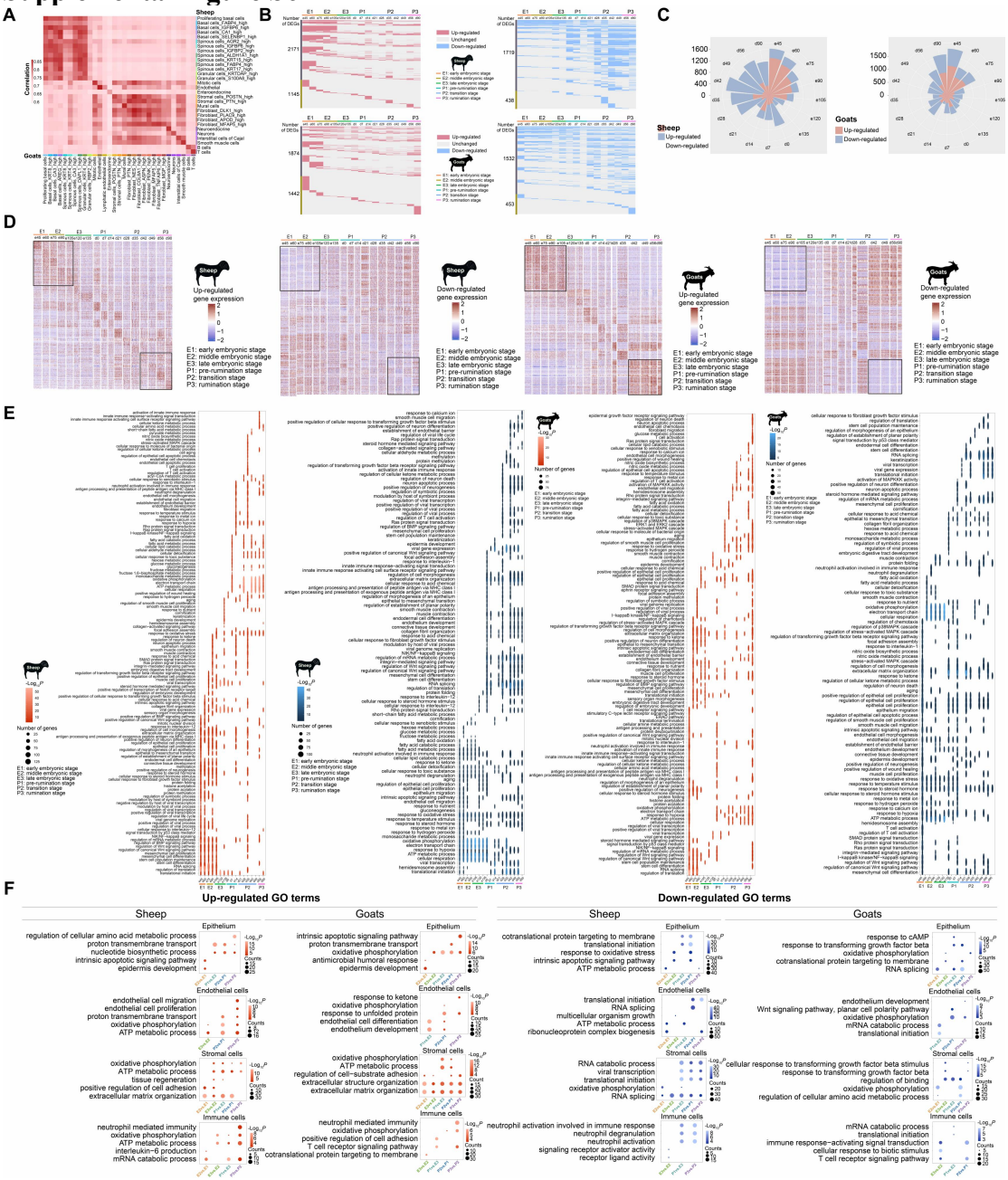


Fig. S6 Differentially expressed genes (DEGs) at the 17 developmental timepoints in sheep and goat rumen tissue. (A) Spearman correlation of gene expressions in the rumen cell types between sheep and goats. **(B)** Heatmaps showing the number of DEGs among the 17 time-points in sheep and goats. The upper rows denote the DEGs shared by at least two time-points and the lower rows are DEGs specific for individual time-points. **(C)** Rose diagrams showing the numbers of DEGs at each time-point in sheep and goat. **(D)** Heatmap plots showing the up- and down-regulated of top 50 DEGs in sheep and goat rumen tissues at the seventeen timepoints. **(E)** Representative and important gene ontology (GO) terms based on the up- and down-regulated DEGs in the sheep and goat rumen tissues at the seventeen timepoints the ($P < 0.05$). **(F)** Diagram showing the significantly enriched Gene Ontology (GO) terms for the up- and down-regulated genes ($P_{adj} < 0.05$, $|logFC| > 0.25$) as revealed by the pairwise comparisons of different cell types between different stages in sheep and goat rumen.

Supplemental Figure S7

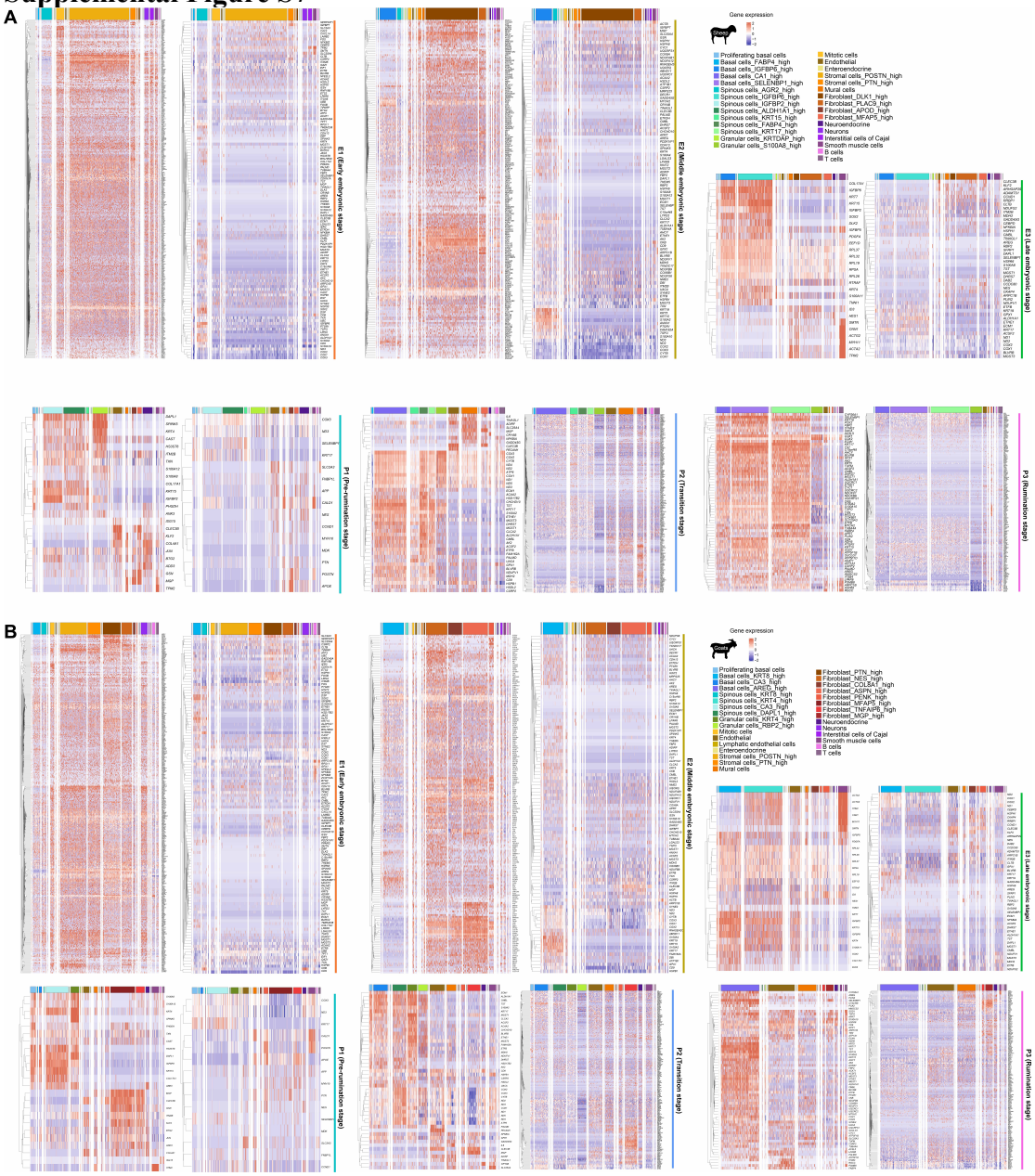


Fig. S7 Expression patterns of common differentially expressed genes (DEGs) between sheep and goats in different cell types. (A, B) Heatmap plots showing the common up- and down-regulated DEGs across the different developmental stages in sheep (A) and goats (B).

Supplemental Figure S8

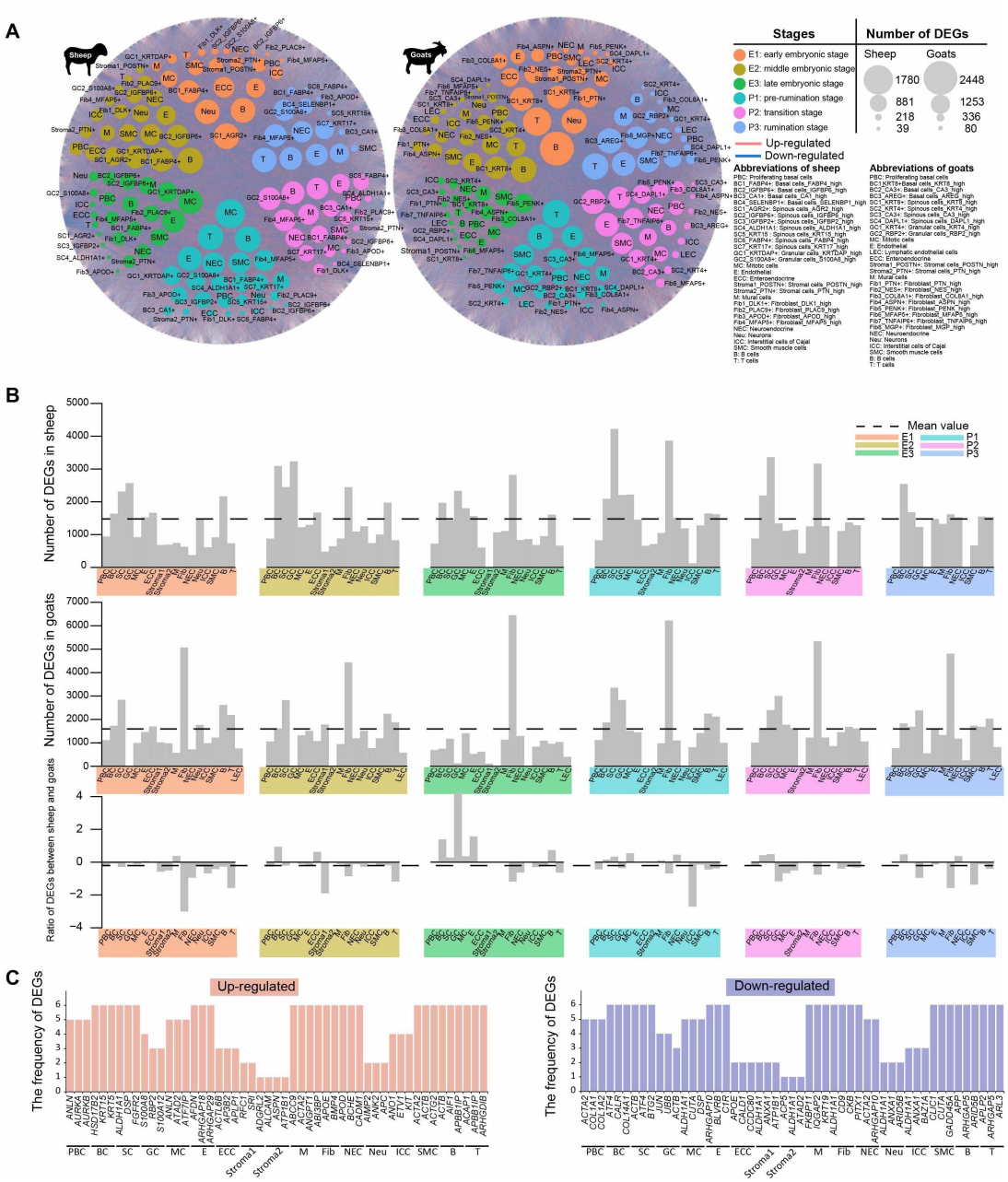


Fig. S8 Up-regulated and down-regulated differentially expressed genes (DEGs) indifferent developmental stages and different cell types of sheep and goats. (A) Network plots showing the number of DEGs ($P_{adj} < 0.05$, $|logFC| > 0.25$) in each cell type at the six major developmental stages. The internal nodes denote cell types. The gray circular edge denotes the collections of DEGs. Each cell type is connected with its DEGs by the internal lines of the network. **(B)** Bar plots showing the numbers of DEGs, and the ratio of DEGs between sheep and goats for each cell type. **(C)** Bar plots showing frequencies of the top three common up- and down-regulated DEGs in the six developmental stages.

Supplemental Figure S9

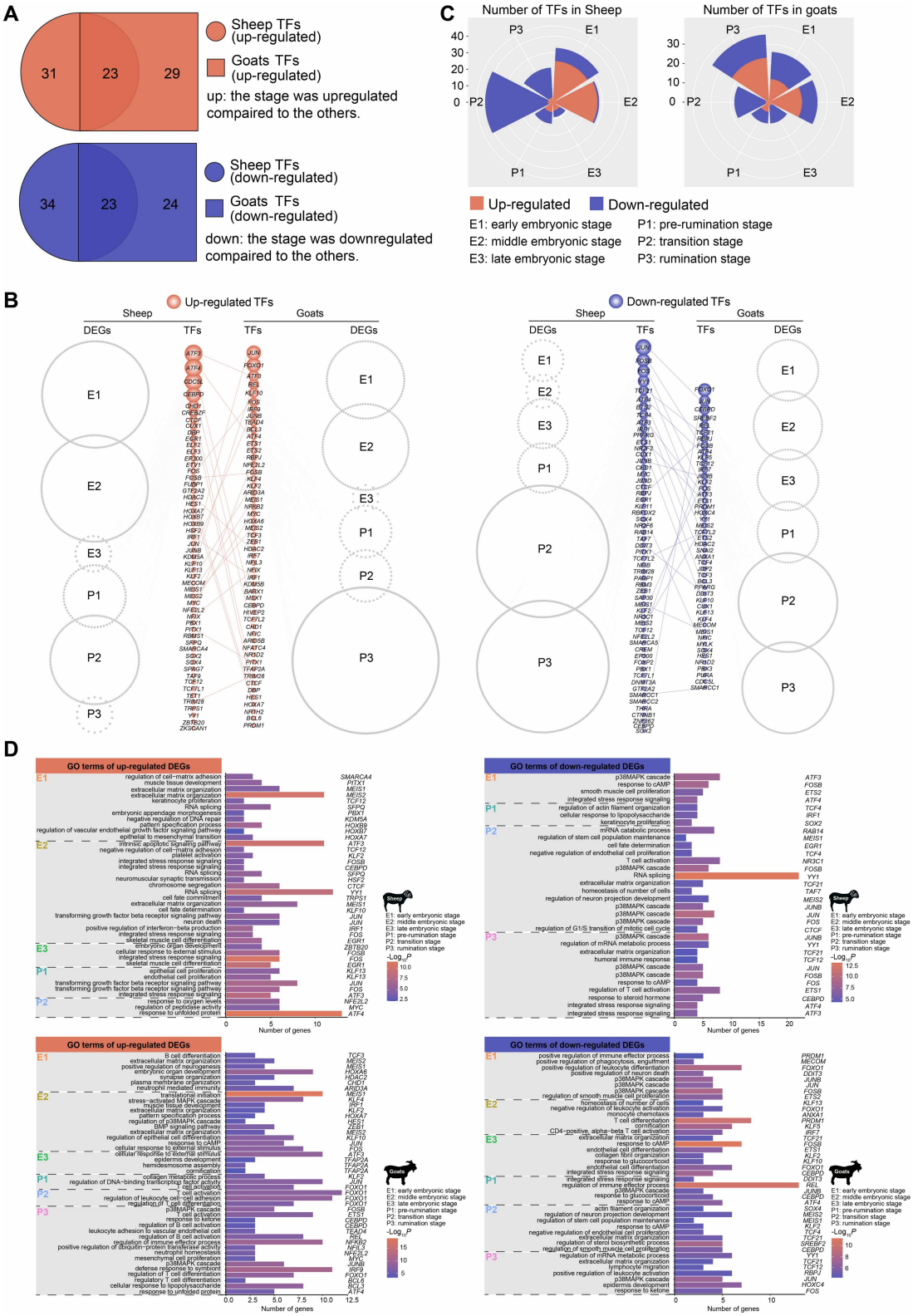


Fig. S9 Changes in core regulatory transcription factors (TFs) during the rumen development. (A) Venn diagrams showing the common and species-specific up- and down-regulated DEGs for the TFs in sheep and goat. The up- or down-regulated DEGs was defined by comparing the expressions of a gene at one stage with its expressions of all the other stages. **(B)** Network visualization of potential up-regulated and down-regulated TFs. The colored node sizes are proportional to the number of associated DEGs for the significant TFs. The connecting line in the middle denotes common TFs between sheep and goat. The circles of grey dots showed the relevant up-regulated and down-regulated DEGs for the TFs in the six major developmental stages. **(C)** Rose diagrams showing the numbers of TFs in the six major developmental stages. **(D)** TF target gene enrichment analysis at the six developmental stages between sheep and goats.

Supplemental Figure S10

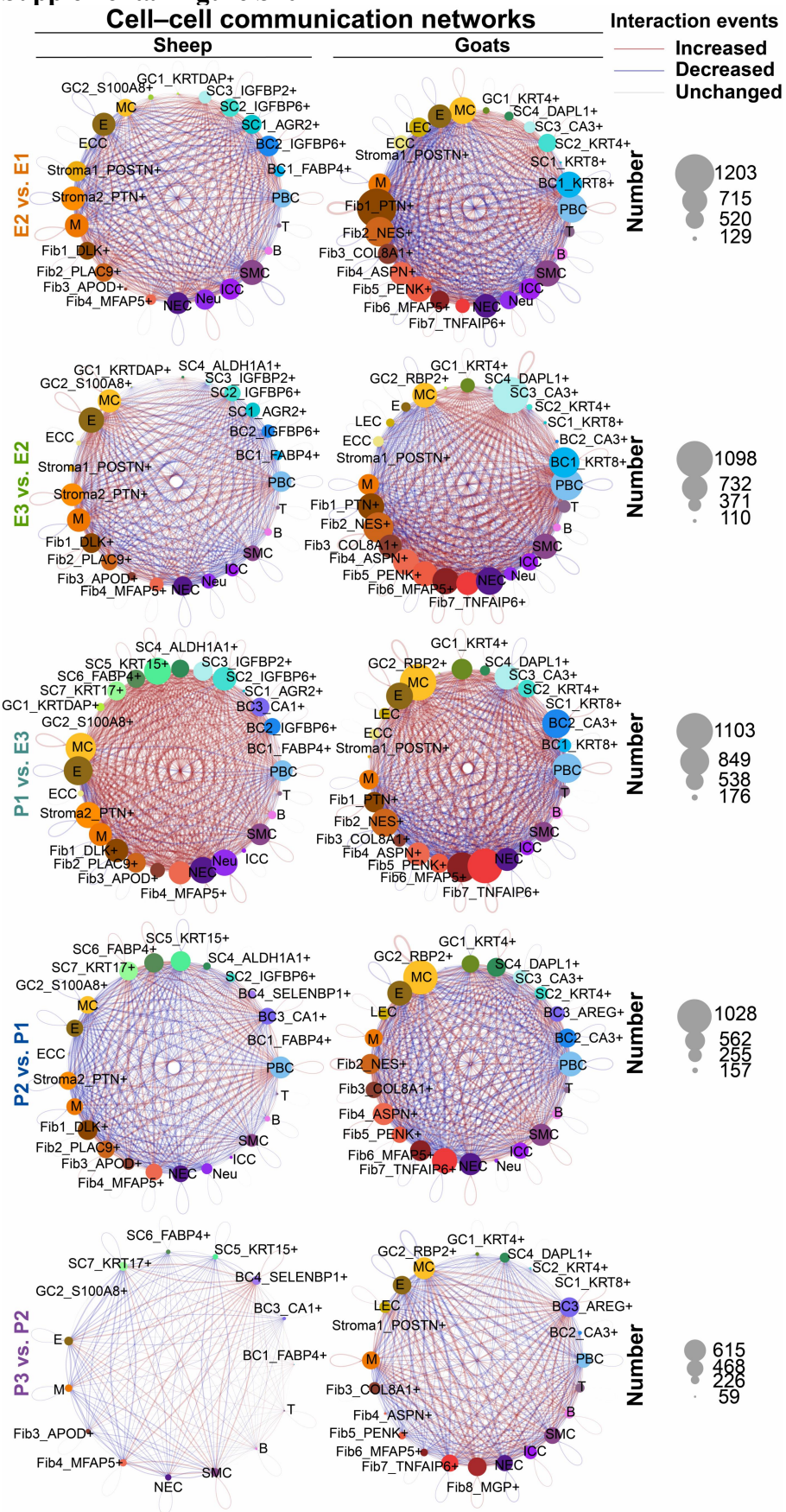


Fig. S10 Network plots showing the changes in LR interaction events. The thickness of the lines is proportional to the number of LR interaction events. Red lines represent an increase in the number of LR events relative to the previous stage, and blue lines represent a decrease. The abbreviation of the cell names is shown in Fig. 1.

Supplemental Figure S11

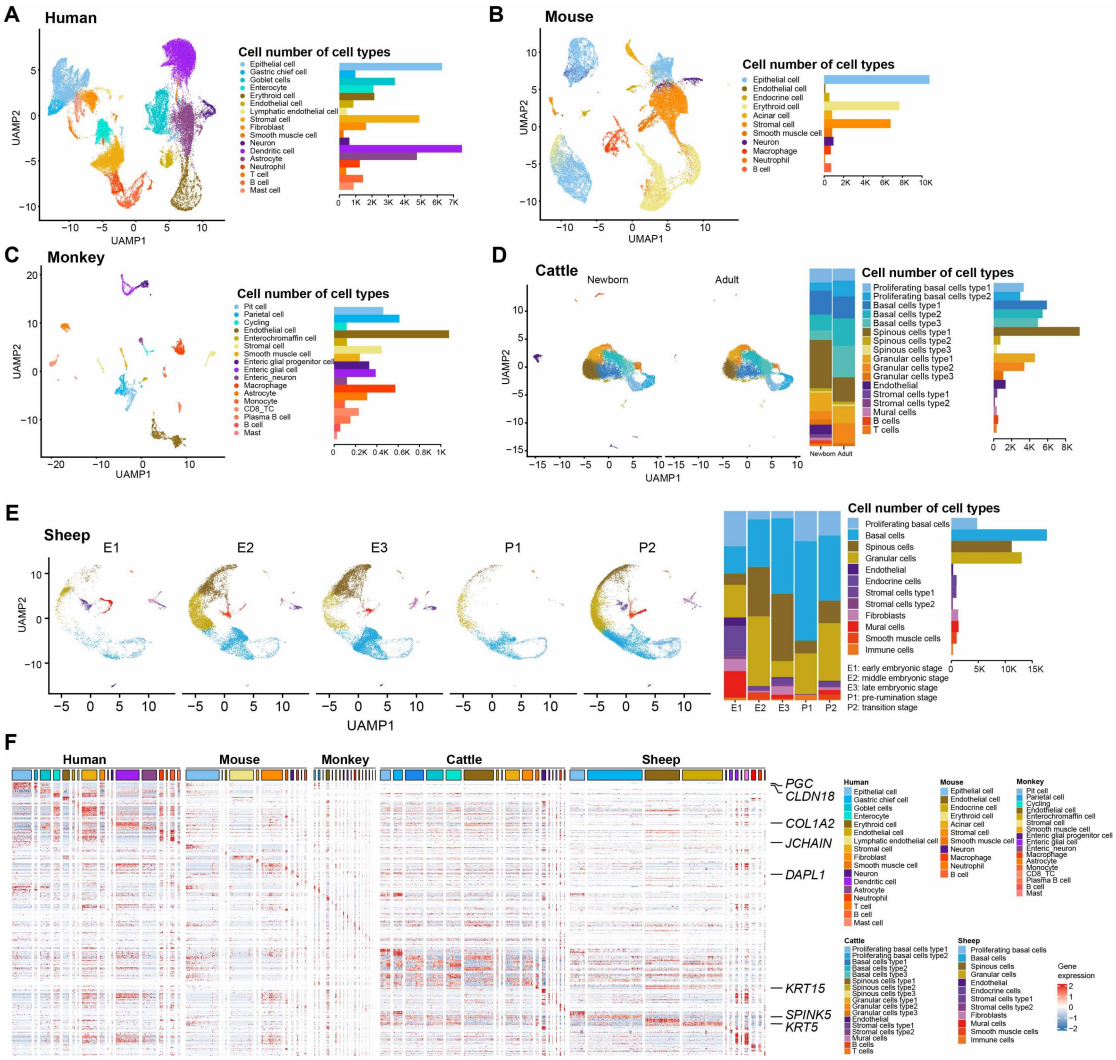


Fig. S11 Cross-species rumen and stomach single cell transcriptome atlases. **(A-E)** UMAP plot showing the single-cell atlas of stomach or rumen in human, mouse, monkey, cattle, and sheep. Dots with colors represent different cell types. Bar plots show the number of cells profiled for each type after quality control. **(F)** Heatmap showing the cross-species comparisons of the top 50 marker genes for all the cell types in the human, mouse, monkey and cattle.

Supplemental Figure S12

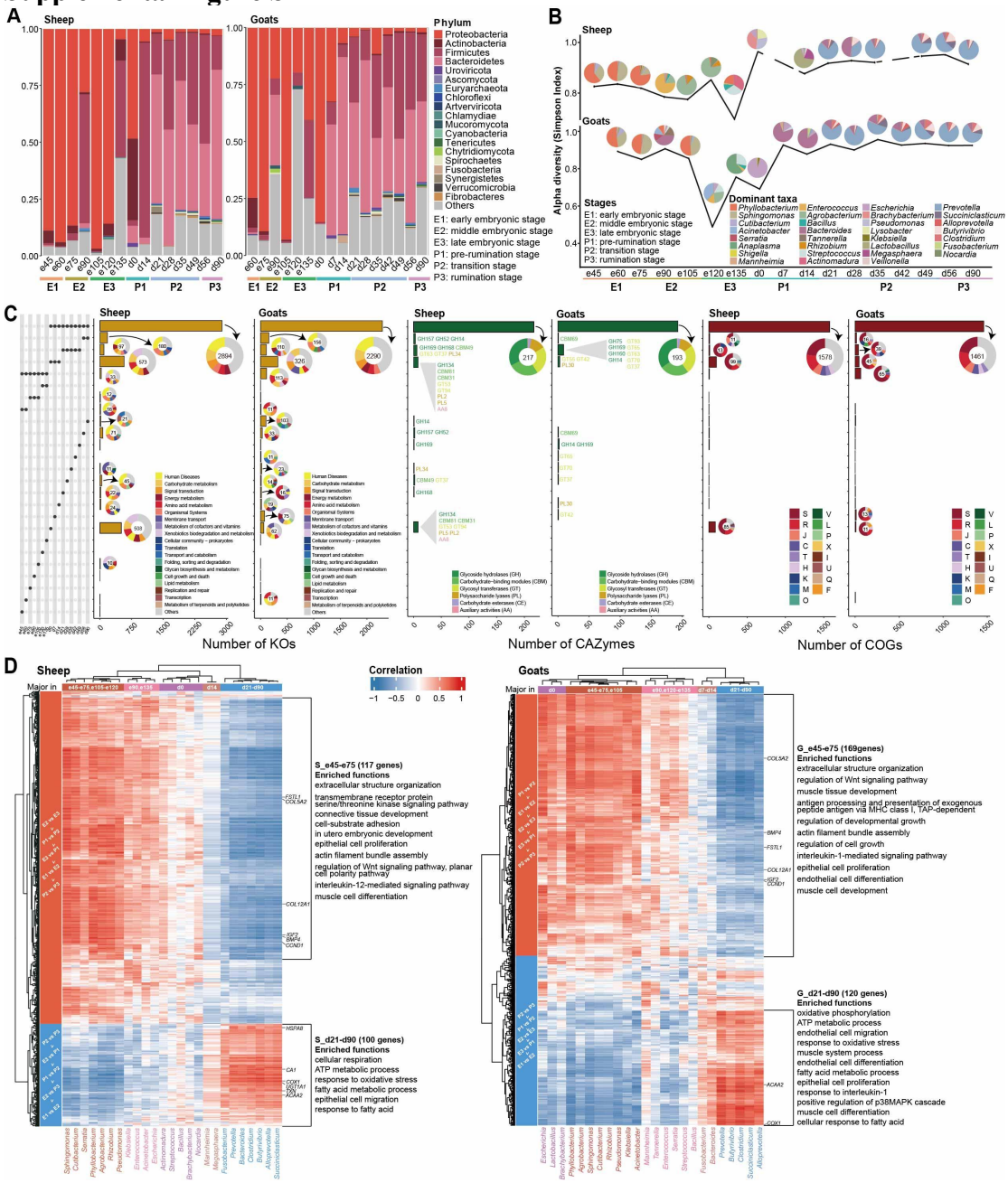


Fig. S12 Taxonomic and functional annotation of ruminal microbiota and associations among the bacterial composition and differentially expressed genes between adjacent stages in sheep and goats. (A) Relative abundances of the ruminal microbial phyla across the rumen developmental stages in sheep and goat. **(B)** Dynamics of microbial composition during the rumen developmental stages in sheep and goats. Pie charts show the proportions of the most abundant microbial genera, which collected the top three abundant microbial genera at each time point of the rumen development. The lines represent the alpha-diversity by Simpson index at the species level. **(C)** Comparison of the functional items (KOs in the left, CAZymes in the middle and COGs in the right) among the rumen developmental stages in the microbiome of sheep and goat. The left panel shows sets included in the intersection and independent sites, and the right bar or pie charts show the categories of the functional items in these sets. The COGs categories are following: C, Energy production and conversion; F, Nucleotide transport and metabolism; H, Coenzyme transport and metabolism; I, Lipid transport and metabolism; J, Translation, ribosomal structure and biogenesis; K, Transcription; L, Replication, recombination and repair; M, Cell wall/membrane/envelope biogenesis; O, Posttranslational modification, protein turnover, chaperones; P, Inorganic ion transport and metabolism; Q, Secondary metabolites biosynthesis, transport and catabolism; R, General function prediction only; S, Function unknown; T, Signal transduction mechanisms; U, Intracellular trafficking, secretion, and vesicular transport; V, Defense mechanisms; X, Mobilome: prophages, transposons. **(D)** Heatmap of correlation between the expressions of host differential expressed genes (DEGs) between adjacent stages and the relative abundances of major rumen bacterial genera in sheep and goat. Based on the dominance of bacterial genera in one specific stage, five clusters were generated for 25 and 24 bacterial genera for sheep and goat respectively. A cluster, which was dominant in e45-e75, consisted of *Sphingomonas*, *Cutibacterium*, *Phyllobacterium*, *Agrobacterium*, *Rhizobium*, and *Pseudomonas*, and was positively correlates with the expression of the development- and immune-related genes (S_e45-e75 for sheep and G_e45-e75 for goat). In S_e45-e75 and G_e45-e75, 117 genes and 169 genes were assigned with related GO terms, separately. We overlayed the genes of S_e45-e75 and G_e45-e75, and got 67 overlapped genes. Another cluster, which was dominant in d21-d90, consisted of *Prevotella*, *Butyrivibrio*, *Clostridium*, *Succiniclasticum*, and *Alloprevotella*, and was correlated with the expression of the energy metabolism-, cell migration-, immune- and fatty acid metabolic process-related genes (S_d21-d90 for sheep and G_d21-d90 for goat). In S_d21-d90 and G_d21-d90, 100 genes and 120 genes were assigned with these related GO terms, separately. We got 36 overlapped genes between S_d21-d90 and G_d21-d90.

Supplemental Figure S13

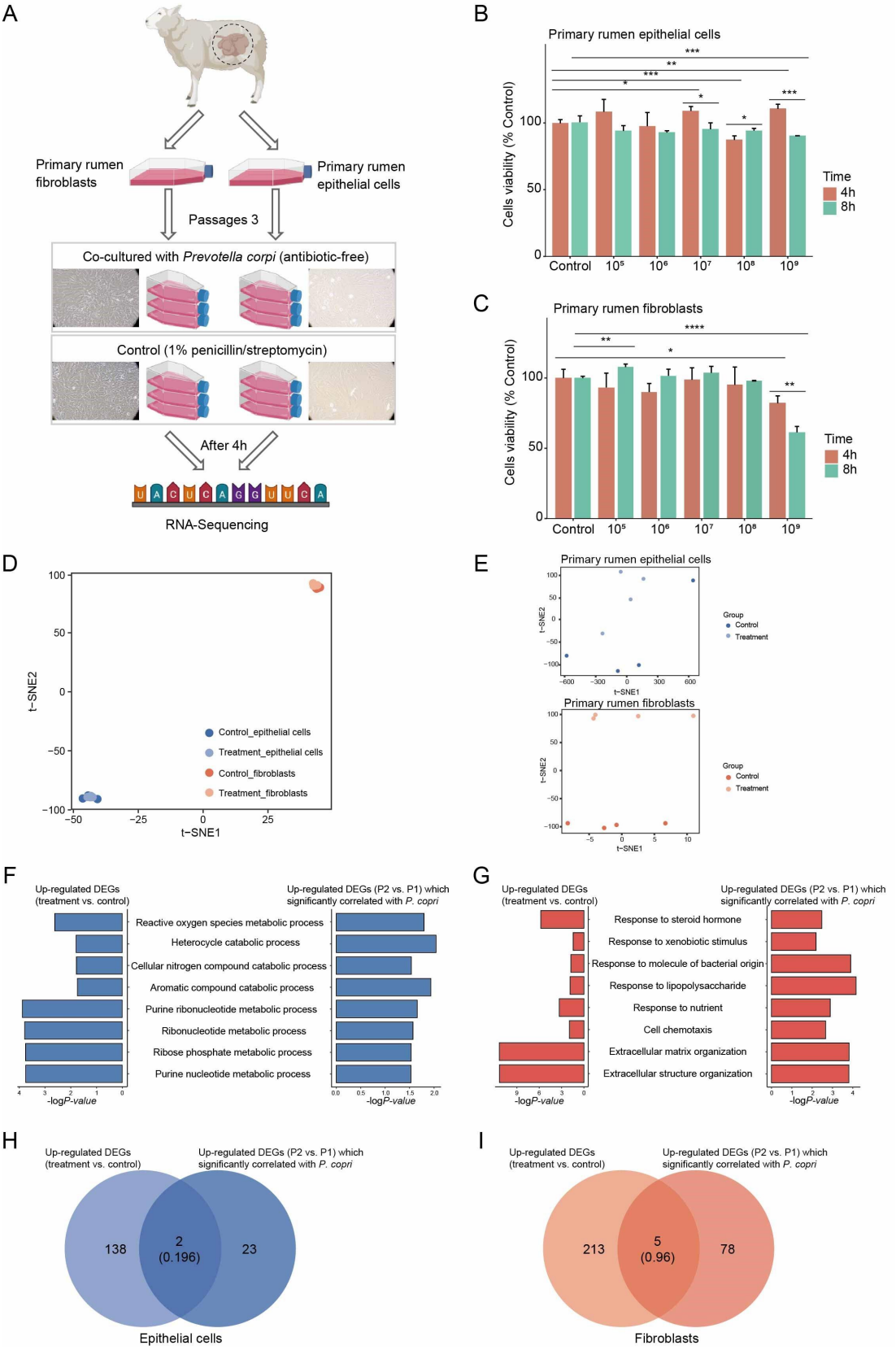
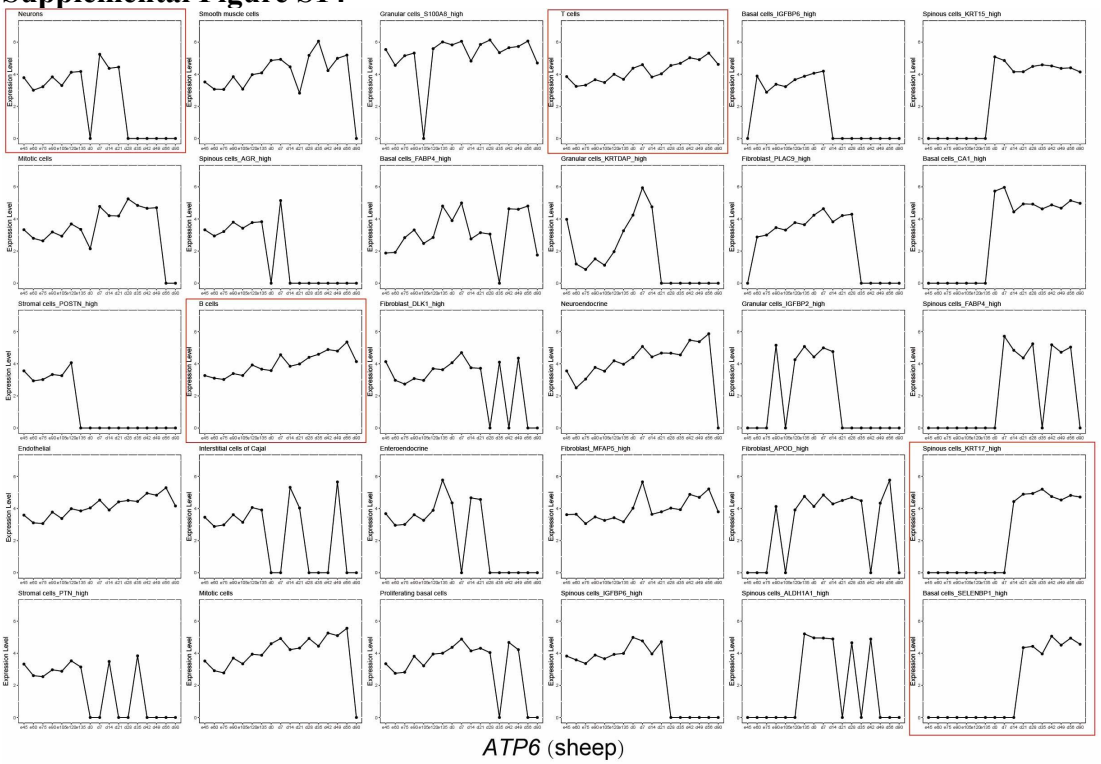
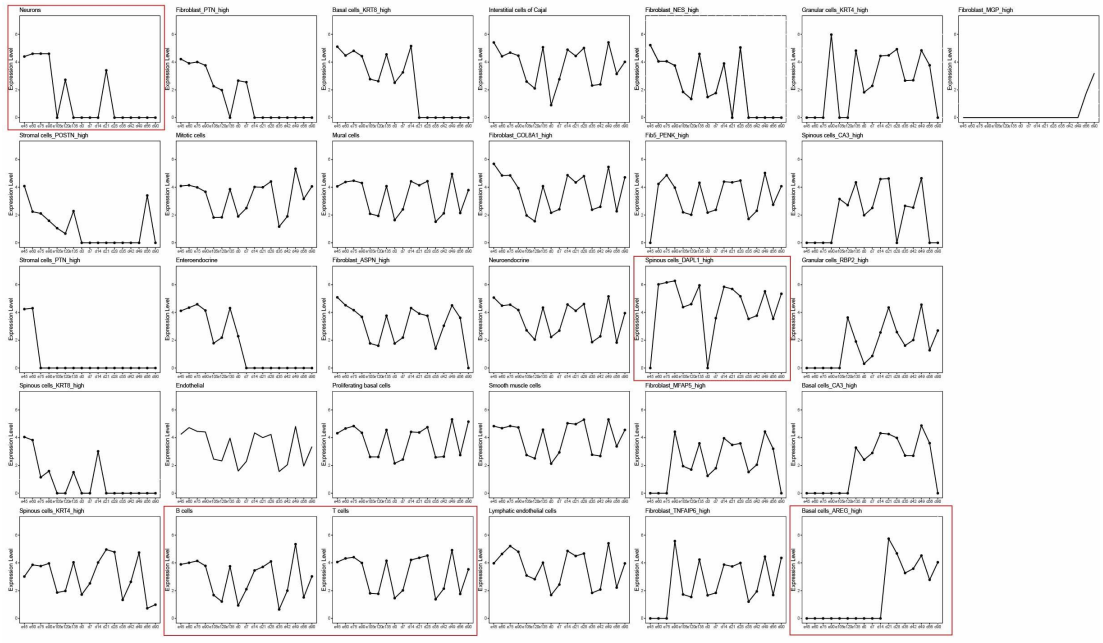


Fig. S13 Transcriptomic analysis of co-cultured rumen cells with *Prevotella copri*. **(A)** Overview of experiment of co-culture microbial with cells in *vitro*. **(B, C)** Cells viability of primary rumen epithelial cells (B) and primary rumen fibroblasts (C) in different concentration and duration for co-culture of cells with *P. copri* were evaluated by CCK-8 assay. *P value* were determined by t-test. ****: $P < 0.0001$, ***: $P < 0.001$, **: $P < 0.01$, *: $P < 0.05$. **(D)** Plot of t-SNE of all samples based on gene expression. **(E)** Plot of t-SNE of primary rumen epithelial cells and primary rumen fibroblasts based on gene expression. **(F, G)** In in rumen epithelial cells (F) or fibroblasts (G), the partial common GO terms in up-regulated DEGs (treatment vs. control) and some up-regulated DEGs (P2 vs. P1; significantly correlated with *Prevotella copri*). Only GO terms with "pvalueCutoff = 0.05" and "minGSSize = 3" were considered. **(H, I)** Venn diagram of overlap between up-regulated DEGs in primary rumen epithelial cells (H) or primary rumen fibroblasts (I) with up-regulated DEGs (P2 vs. P1) which significantly ($P < 0.05$) correlated with *Prevotella copri*. And comparison between the overlap and the overlap expected by chance.

411 **Supplemental Figure S14**



ATP6 (sheep)



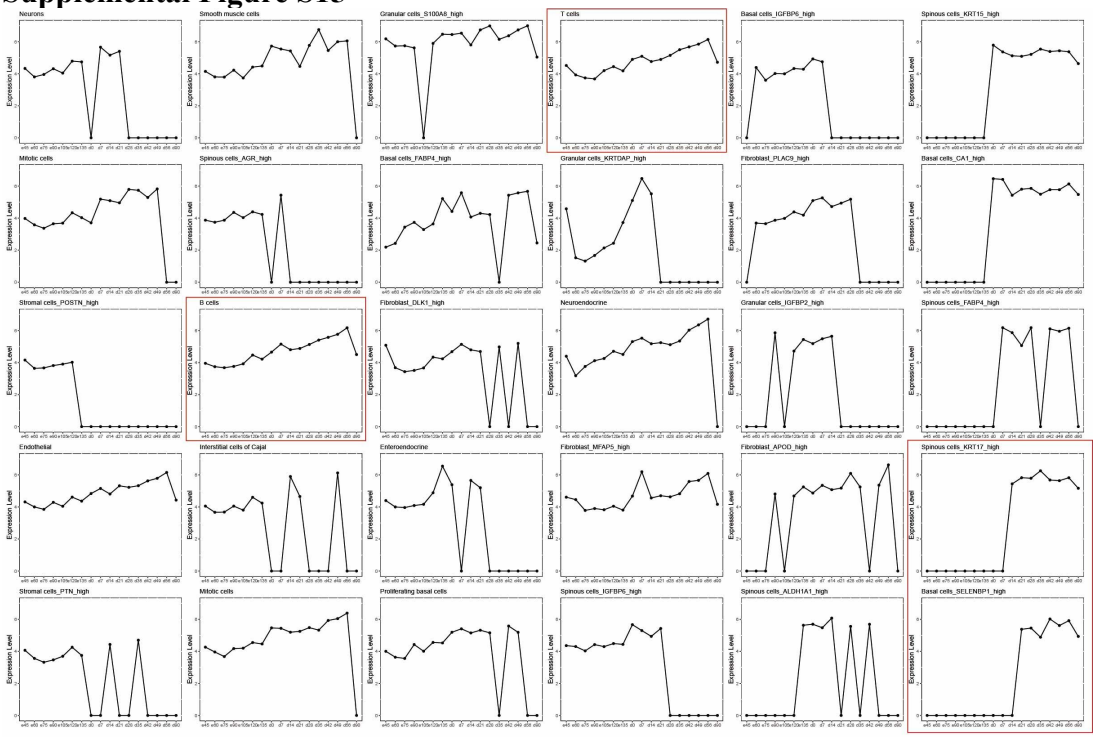
412 *ATP6* (goats)

413 **Fig. S14** The expression level of *ATP6* in all types of cells in sheep and goats.

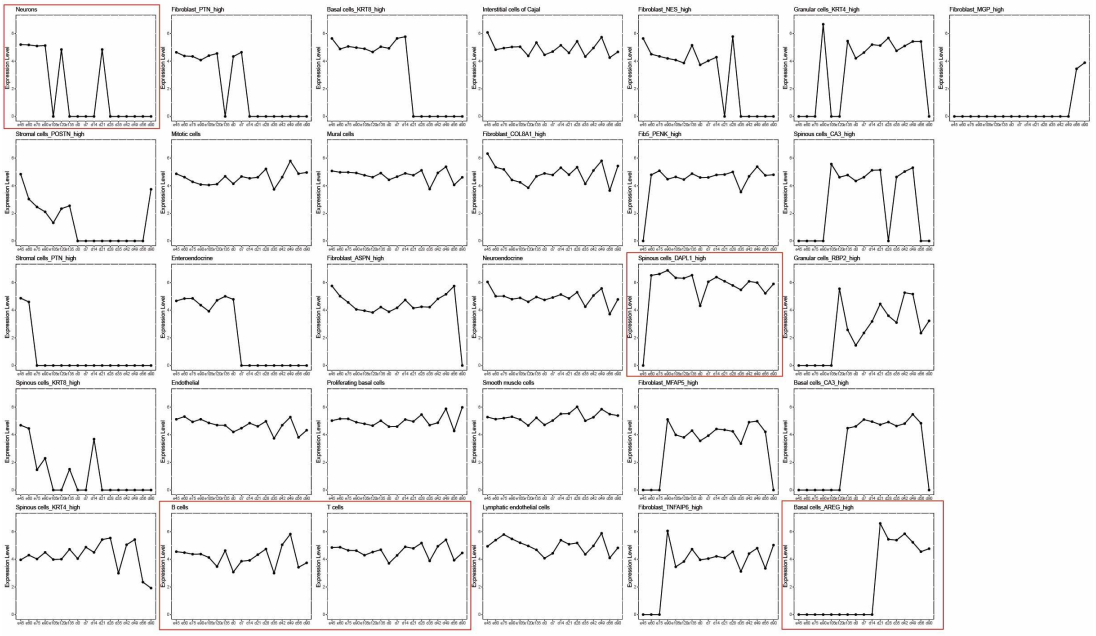
414

415

Supplemental Figure S15



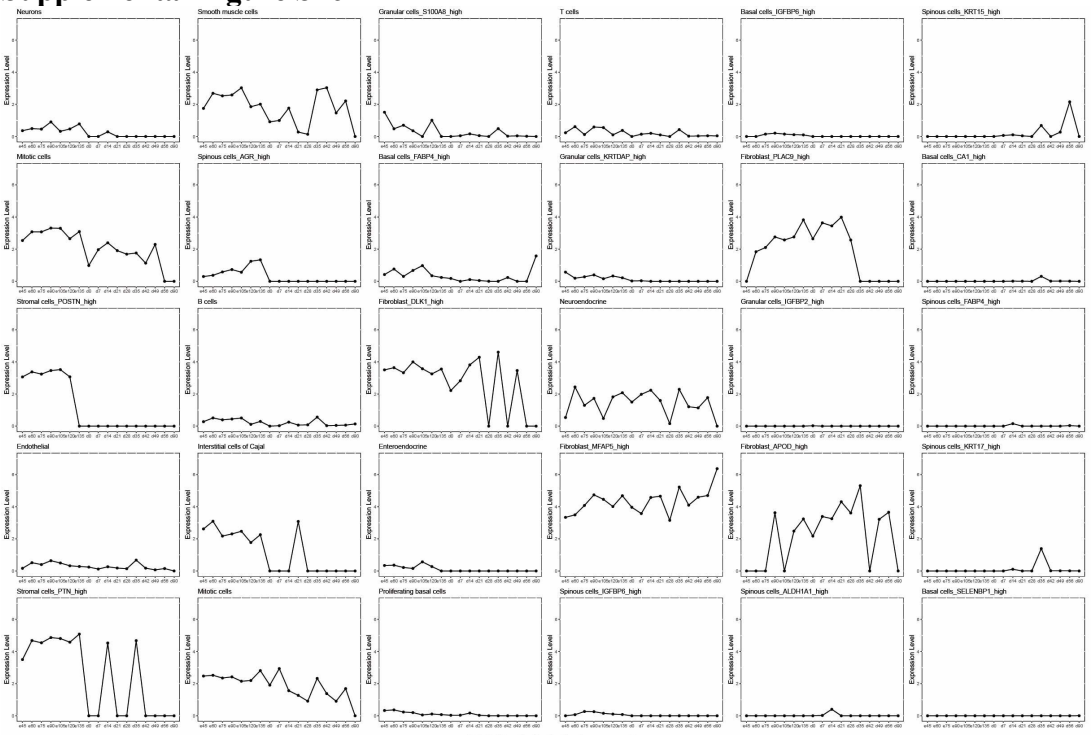
CYTB (sheep)



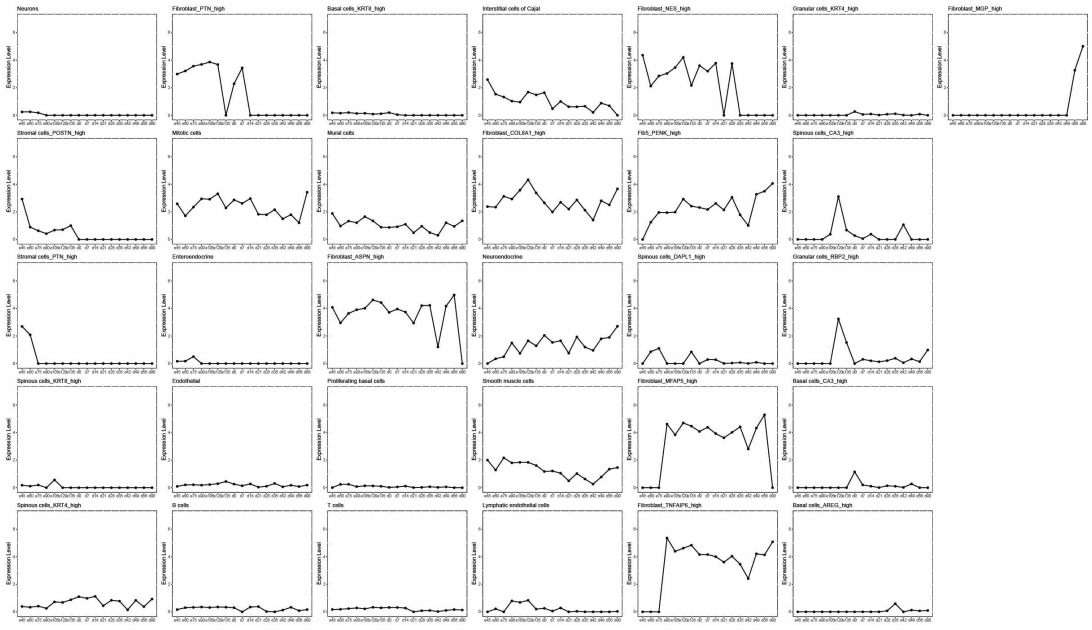
CYTB (goats)

Fig. S15 The expression level of *CYTB* in all types of cells in sheep and goats.

Supplemental Figure S16



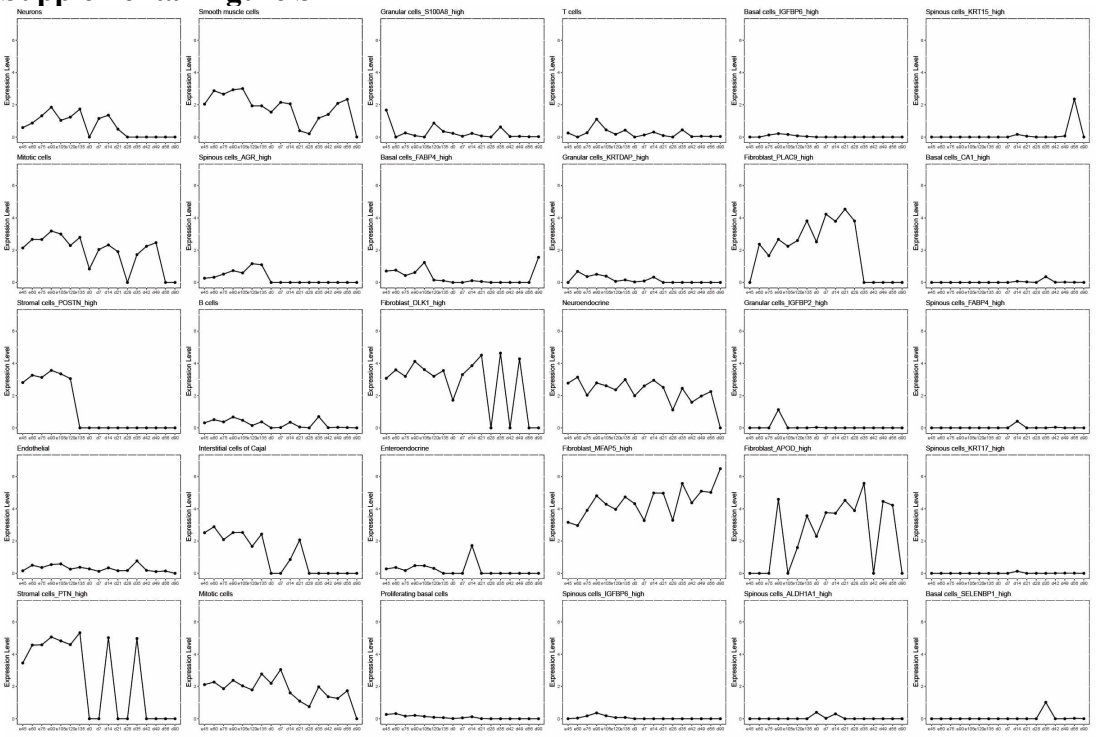
COL1A1 (sheep)



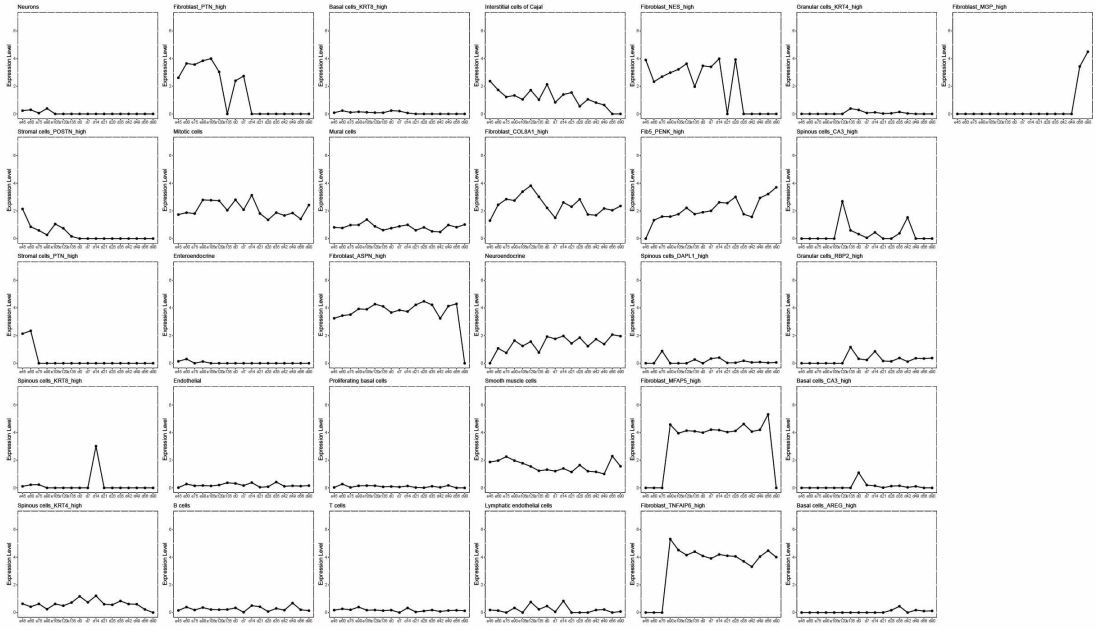
COL1A1 (goats)

Fig. S16 The expression level of *COL1A1* in all types of cells in goats and sheep.

Supplemental Figure S17



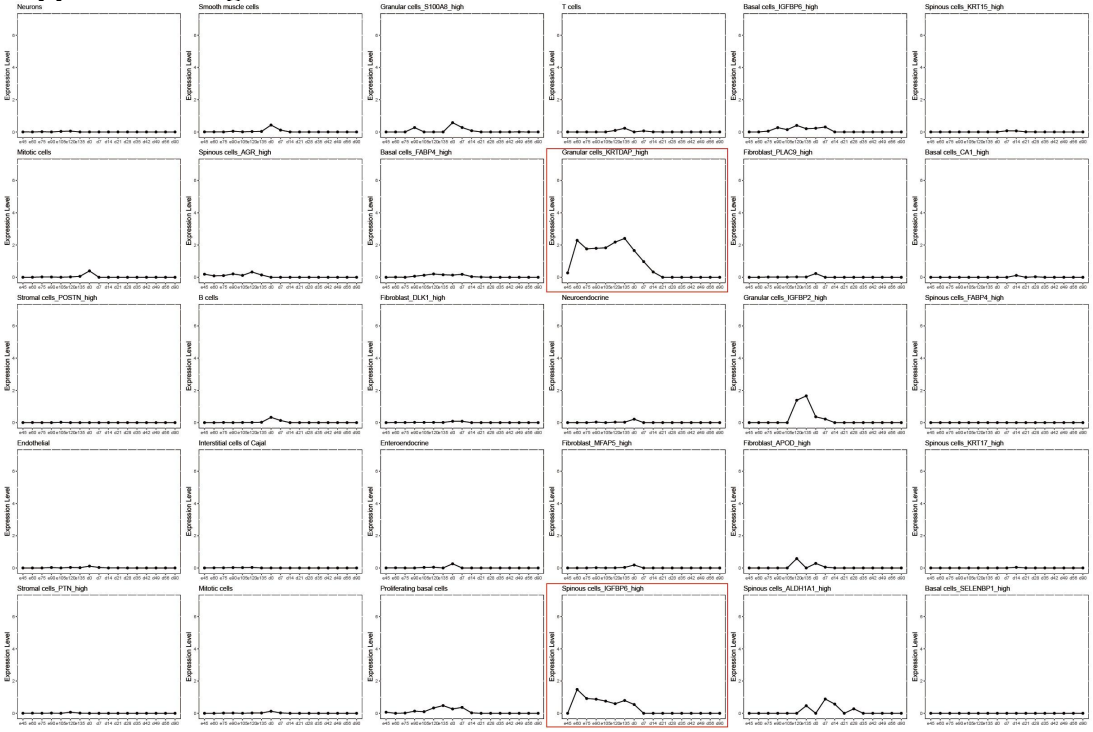
COL1A2 (sheep)



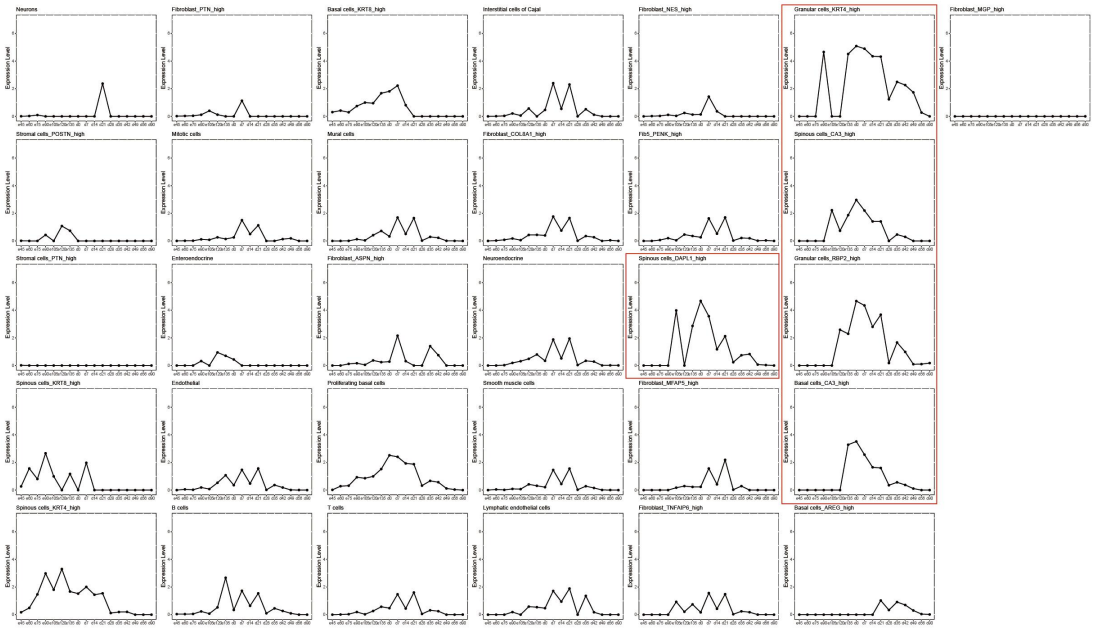
COL1A2 (goats)

Fig. S17 The expression level of *COL1A2* in all types of cells in sheep and goats.

Supplemental Figure S18



KRT4 (sheep)



KRT4 (goats)

Fig. S18 The expression level of *KRT4* in all types of cells in sheep and goats.

3. Supplemental Notes

Supplemental Note S1 Prior to embryonic day 45 (e45), the rumen typically consisted of three layers: the epithelium (E), pluripotential blastemic tissue (PBT) and serosa (S). Ruminal pillars (Rpi) became visible at approximately e45 and later (Supplemental Fig. S1B), which falls within the time range (e39–e46) reported previously (Ortega 1973; Franco et al. 1992; García et al. 2012). Ruminal papillae (Rp) started to appear at e90 (Supplemental Fig. S1B), slightly later than previously reported (e76; García et al. 2012). The mature rumen walls developed into four layers, including an internal epithelium (E), a middle layer of lamina propria and submucosal tissue (Lp+Sb), a tunica muscularis (Tm) and an external layer or serosa (S; Supplemental Fig. S1B).

Supplemental Note S2 The histological changes in rumen tissues could result from the keratinization of the epithelium after birth and the development of smooth muscle contraction for rumination (Luginbuhl 1983). In the late embryonic and pre-rumination stages, the rumens of sheep and goats were underdeveloped with poorly developed papillae but without a high degree of keratinization, a characteristic of the mature organ (Supplemental Fig. S1B; Gilliland et al. 1962). After 21 days postpartum, the length and width of the Rp increased significantly (Supplemental Fig. S1C). Meanwhile, the epithelium became stratified and keratinized, and could be divided into four different layers (i.e., the keratin, granular, spinous and basal layers; Supplemental Fig. S1B). The Sb was composed of loose connective tissue without glands, and the Tm consisted of two layers, namely, an internal circular bundle (i) and an external longitudinal bundle (e; Supplemental Fig. S1B). At most of the timepoints, the Rp of goats were taller and wider than those of sheep (Supplemental Fig. S1C). In

general, the dimensions (Rp, Lp+Sp and Tm) increased as the rumen matured, which helped increase rumination (Supplemental Fig. S1C). The epithelium became substantially thinner after birth until the transition stage but became slightly thicker after d56 of the rumination stage (Supplemental Fig. S1C).

Supplemental Note S3 We aimed to combine the single cells of rumen tissues of sheep and goats, and obtain a unified dataset that faithfully retained all sources of variability such as developmental stages and cell composition heterogeneity, while accounting for technical biases. We tested three different scRNA-seq data integration methods [e.g., the merging with scaling, cca integration in Seurat, and harmony]. We found that the method used here, i.e., the merging with scaling, is the best, which can detect all the different cell types while retain the variability in both species (Supplemental Fig. S2C,D). We projected the major cell types identified by "merging with scaling" to the post-batch correction graphs (Supplemental Fig. S2C,D). We found that major cell types and cell type composition comparison were mostly identified, indicating accuracy of the "merging with scaling" method applied in the analysis. However, the CCA integration in Seurat and the Harmony, two methods for post-batch correction, were unable to clearly identify cell subtypes. Thus, we retained the UMAP graphs by the merging with scaling in the main text.

Supplemental Note S4 In addition to cell types, we detected the cell subtypes specific at particular stages by manual annotation. We conducted the dimensionality reduction analysis at each time point, and annotated the cell subtypes. For the cell subtypes and proportions, we observed consistency with the integrated dataset (Supplemental Fig. S3; Fig. 2C,D). In the merged data of sheep, the stromal1 was

found at embryonic stage e45 (cell number, 5148), e60 (7540), e75 (149), e90 (59) and e105 (78); stromal2 was found at seven timepoints of embryonic stage e45 (cell number, 892), e60 (213), e75 (467), e90 (492), e105 (175), e120(21) and e135 (97), and after birth d14 (14), d35 (4). In the merged data of goats, the stromal1 was found at embryonic stage e45 (cell number, 3726), e60 (55), e75 (34), e90 (12), e105 (6), e120 (8) and e135 (4); stromal2 was found at embryonic stage e45 (1810) and e60 (13) (Supplemental Tables S5, S6). However, in the annotation of individual stages, stromal cells were only identified at e45 to e105 in sheep, and at e45 in goats. In sheep and goats, we selected two timepoints in each species (sheep, e120 and e135; goats, e60 and e75). Stromal cells were not identified at individual timepoint (Supplemental Fig. S3), but identified in the merged data (Fig. 2 C,D). We projected the stromal cells identified by “merging with scaling” to the individual graphs at the same timepoints (Supplemental Fig. S3). The "stromal cells" with a very small number of cells (less than 100) were not annotated because of mixing with fibroblasts with similar gene expressions.

References

Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine J-C, Geurts P, Aerts J. 2017. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**: 1083–1086. doi:10.1038/nmeth.4463

Franco A, Regodon S, Robina A, Redondo E. 1992. Histomorphometric analysis of the rumen of sheep during development. *Amer J Vet Res* 53: 1209–1217

García A, Masot J, Franco A, Gázquez A, Redondo E. 2012. Histomorphometric and immunohistochemical study of the goat rumen during prenatal development. *Anat Rec (Hoboken)* 295: 776–785. doi:10.1002/ar.22431

Ge W, Tan SJ, Wang SH, Li L, Sun XF, Shen W, Wang X. 2020. Single-cell Transcriptome Profiling reveals Dermal and Epithelial cell fate decisions during Embryonic Hair Follicle Development. *Theranostics* 10: 7581–7598. doi:10.7150/thno.44306

Gilliland R, Bush L, Friend J. 1962. Relation of ration composition to rumen development in early-weaned dairy calves with observations on ruminal parakeratosis. *J Dairy Sci* 45: 1211–1217. doi:10.3168/jds.S0022-0302(62)89598-8

Ma S, Sun S, Geng L, Song M, Wang W, Ye Y, Ji Q, Zou Z, Wang S, He X. 2020a. Caloric restriction reprograms the single-cell transcriptional landscape of Rattus norvegicus aging. *Cell* 180: 984–1001. e1022. doi:10.1016/j.cell.2020.02.008

Ortega SDR. 1973. Desarrollo prenatal del estómago de la oveja. *Doctoral thesis*.

Tombor LS, John D, Glaser SF, Luxán G, Forte E, Furtado M, Rosenthal N, Baumgarten N, Schulz MH, Wittig J et al. Single cell sequencing reveals endothelial plasticity with transient m esenchymal activation after myocardial infarction. *Nat Commun* 12: 681. doi:10.1038/s41467-021-20905-1

Yu Z, Morrison M. 2004. Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* **36**: 808–812. doi:10.2144/04365ST04