RESEARCH

Transcriptomic analysis reveals molecular insights into lactation dynamics in Jakhrana goat mammary gland

Mahesh Shivanand Dige[1*], Ankita Gurao[1], Lalit Pratap Singh[1,2], Meenakshi Chitkara[1,2], Manoj Kumar Singh[3], Gopal Dass[3], Arun Kumar Verma[3,4], Rakesh Kumar Pundir[1] and Ranjit Singh Kataria[1]

Abstract

Background Goat milk is gaining popularity as a superior alternative to bovine milk due to its closer resemblance to human milk. Understanding the molecular processes underlying lactation is crucial for improving milk quality and production in goats. However, the genetic mechanisms governing lactation in goats, particularly in indigenous breeds like the Jakhrana, remain largely unexplored.

Results In this study, we performed a comprehensive transcriptomic analysis of Jakhrana goat mammary glands during early and late lactation stages. We isolated milk somatic cells and conducted RNA sequencing, followed by transcript quantification and mapping against the ARS1.2 Capra hircus reference assembly. Our analysis identified differentially expressed genes (DEGs) and commonly expressed genes (CEGs) across the lactation phases. Early lactation showed enrichment of genes encoding antimicrobial peptides and lubrication proteins, while late lactation exhibited heightened expression of genes encoding major milk proteins. Additionally, DEG analysis revealed upregulation of pivotal genes, such as the ABC transporter gene *MRP4*, implicated in modulating milk composition and quality.

Conclusion Our findings provide insights into the genetic mechanisms underlying lactation dynamics in the Jakhrana goat. Understanding these mechanisms could help in improving milk production and quality in goats, benefiting both the dairy industry and consumers.

Keywords Goat milk, Lactation, Mammary glands, Transcriptomic

*Correspondence:
Mahesh Shivanand Dige
maheshdige@gmail.com; dige.mahesh@icar.gov.in
[1]Division of Animal Genetic Resources, ICAR- National Bureau of Animal Genetic Resources, Karnal, Haryana 132001, India
[2]Division of Animal Biotechnology, ICAR-National Dairy Research Institute, Karnal, Haryana 132001, India
[3]Division of Animal Genetics and Breeding, ICAR - Central Institute for Research on Goats, Makhdoom, Uttar Pradesh, India
[4]Animal Nutrition, Management and Product Technology Division, ICAR, ICAR- Central Institute for Research on Goats, Makhdoom, Uttar Pradesh, India

Dige *et al. BMC Genomics*   (2024) 25:874

Page 2 of 11

Introduction

Goats are vital milk producers in many tropical regions, playing a significant role in the nutritional source of various developing countries [1]. The increasing global recognition of goat milk's medicinal properties has been underscored by its significant health benefits for human consumption [2]. This acceptance is reflected in the rapid expansion of the goat milk market across numerous nations [3]. Studies have highlighted a plethora of advantages offered by goat milk for human health, including immune system modulation [4], enhanced mineral bioavailability [5, 6], anti-inflammatory [7] and antioxidant properties [8], as well as allergy alleviation [9] and antimicrobial [10] and anticancer effects [11]. Goat milk is esteemed for its nutritional richness, providing a range of macro- and micro-nutrients crucial for newborn growth and human health [12, 13]. Notably, the fat and protein fractions in goat milk serve as essential sources of energy and nutrients [14]. The Jakhrana goat (*Capra hircus*), a native breed from Rajasthan, India, is renowned for its adaptability and high milk production [15]. Improving milk quality, particularly for infants or those with compromised immune systems, is a key objective in goat farming [16], necessitating a comprehensive understanding of milk components and their regulatory mechanisms.

Transcriptomic analysis has emerged as a powerful tool for unraveling the molecular mechanisms underlying complex biological processes, including lactation in dairy animals. In goat farming, transcriptomic studies provide valuable insights into the genetic regulation of milk synthesis and secretion, crucial for enhancing milk yield and quality. Lactation is a dynamic process characterized by distinct phases, each associated with specific gene expression patterns and physiological changes in the mammary gland. Among the various factors influencing the milk composition of the goat, lactation stage is one of the crucial factors since from the onset of gestation to the involution the mammary gland goes through number of cycles of cell proliferation and degeneration. The phases occurring in between the gestation and involution are the effective lactation stages i.e., early, mid, and late. The extremes of the lactation cycles including the early and late is when the number of mammary alveoli starts proliferating and decline in number respectively. The early lactation phase is marked by an increased need for protection against pathogens, supported by the upregulation of genes encoding antimicrobial peptides and proteins [17]. As lactation progresses into the late phase, there is a shift towards milk production and secretion, reflected in the upregulation of genes encoding major milk proteins. While several studies have identified differentially expressed genes in various livestock species [18–26], research on the milk transcriptomic of Indian goats remains limited. To address this gap, we conducted a longitudinal study to explore the lactation dynamics of the Jakhrana goat using milk somatic cells, a non-invasive sample type. In this study, we conducted transcriptomic analysis of the mammary gland in Jakhrana goats at early and late lactation stages. Our aim was to identify differentially expressed genes and commonly expressed genes between these two phases, shedding light on the molecular mechanisms driving lactation in goat.

Methods

Milk sample procurement and milk somatic cell isolation

The Jakhrana goat breed, belonging to different lactation phases, were the subjects of the study. Three of each early lactation Jakhrana (A, B, and D) and three of the late lactation Jakhrana (H, I, and JEIKT) comprised the two groups. Early lactation referred to the first weeks of lactation, and late lactation referred to more than 10 weeks of lactation. The institute manages a nucleus flock of purebred Jakhrana animals, which are raised under semi-intensive conditions that include 6–7 h of grazing, stall feeding with dry fodder or straw, and seasonally available green fodder. Additionally, pregnant and lactating animals received concentrated feed at a rate of 500 g per day for pregnant animals and 400 g per liter of milk produced for milking animals, with water available adlibitum. The milk samples of Jakhrana goat were procured from ICAR-Central Institute for Research on Goats (Makhdoom, India). Before collecting the milk samples for somatic cell isolation, the foremilk were allowed to let down, and the mammary gland was surface disinfected using povidone iodine. The 50 mL of milk was collected from all the subject animals. Following the collection of milk samples into sterile plastic flasks, the samples were promptly shipped to the milk quality laboratory in an icebox (4 °C) for milk somatic cell isolation as per the protocol described in [16]. An initial spin at 1700×g was performed for 25 min to collect the fat, which was later removed. 45 mL of 1×PBS containing 0.5 M EDTA was added to the content and again centrifuged at 1700×g for 15 min. The supernatant was discarded, and the pellet was dispensed in 20 mL of 1×PBS containing 0.5 M EDTA. A final centrifugation at 2000×g aided in the separation of purified milk somatic cell pellets.

RNA isolation and library preparation for RNA-Seq

The RNA isolation was performed using the standard chloroform/isopropanol method. 200 µL chloroform was added to a well-homogenised milk somatic cell pellet, obtained from 50 mL of milk, in 1 mL of TRI Reagent™ solution. After vigorous mixing, the admixture was centrifuged at 10,000×g for 10 min to separate the aqueous phase. The aqueous phase containing the RNA fractions was treated with 500 µL of isopropanol, this will aid in

precipitating the RNA fraction when kept at room temperature for 15–20 min. To assist visualization of miniscule pellets, 30 μg of GlycoBlue™ coprecipitant was added along with the isopropanol. The precipitated fraction was centrifuged at 10,000 × g for 10 min, and the supernatant was discarded. The pellet was once washed with 70% EtOH, and the pellet was dispensed in 30 μL of nuclease-free water. The RNA was outsourced for massively parallel sequencing of the transcriptome. The library preparation was performed using QIAseq FastSelect -rRNA HMR Kit (96), Cat. No. / ID: 334,387, following the manufacturer's protocol. The samples having good RIN values (>8.0), as quantified by the Invitrogen Qubit 4 Fluorometer, were selected for further processing. RNA samples that qualified through quality inspection were used for library construction. The library was quality-checked using the Agilent TapeStation system.

Data Quality check and RNA-seq analysis
The data was quality-checked for Q30%, GC content, over-representation, and adaptor content across reads, base composition, and K-mer content using FastQC v0.12.0 [27]. The raw fastQ file was further trimmed for quality and adaptor using Trimmomatic v0.32 [28]. The clean reads were aligned against the Capra hircus index using the STAR v2.7.11a aligner, and an index was generated for the STAR using 'genomeGenerate' mode with the FASTA and GTF files of ARS1.2 (NCBI RefSeq assembly: GCF_001704415.2) assembly [29]. The default parameters were used for STAR 'genomeGenerate' and alignment. The BAM files obtained from alignment were sorted by chromosomes and coordinates, followed by a BAM file quality check. The sorted BAM files were further used for assembling the transcripts using StringTie v2.2.0, followed by merging in order to generate a non-redundant set of transcripts observed in any of the RNA-Seq samples assembled previously [30]. The merged transcript (in the -G option, reference GTF based assembly of transcripts) along with expression estimation mode (-e), accurate abundance estimations of the input transcripts were done. The estimates of abundance were converted to read counts using the prepDE.py3 script provided within the StringTie package. From the coverage values calculated by StringTie for each transcript, prepDE.py3 calculates hypothetical read counts for each transcript using the following straightforward formula: reads_per_transcript is equal to coverage*read_len / transcript_len. Meanwhile another set of analysis was done using without expression estimation mode, and merged to evaluate the transcript classes and RNA biotypes using GFFcompare [31].

The FPKM values obtained from the StringTie in ballgown mode (-B) and the matrix of the same was used to estimate the distribution of the total FPKM within sample across expressed genes (FPKM of a gene/total FPKM of sample*100), top 20 genes covering most of the FPKM distribution were term 'commonly expressed genes' or CEGs. Further, to evaluate the differentially expressed genes the pseudo-counts generated for the two groups (Jhakrana_early and Jhakrana_late) were used for differential gene expression analysis using edgeR. The comparison was done for the Jhakrana_early and Jhakrana_late groups (treatment=Jhakrana_early/Early Lactation, control=Jhakrana_late/Late Lactation). The G1-G2 would represent the difference in the transcript profile changes occurring amongst early and late lactation stages in the Jakharana breed. The CPM (count per million) cutoff for the edgeR v.3.42.4 analysis was set at 1.9 since the edgeR guidelines suggest choosing the CPM cutoff so that it corresponds to about 10–15 reads per sample (CPM cutoff=10/L; L=minimum library size or 5.2 million) [32]. The dispersion was estimated (d=0.40706, and the value was used for determining differential expression by using the exactTest function, which conducts tagwise tests using the exact negative binomial test using 'deviance' option for rejection region. The genes having logFC>2.5 and FDR<0.5 (Benjamini Hochberg) were considered upregulated, and genes having logFC>-2.5 and FDR<0.5 were considered downregulated [33] (Fig. 1).

### Functional annotation and pathway analysis
DAVID and Networkanalyst v.3.0 (https://www.networkanalyst.ca/) were used to perform ORA-based functional annotation and pathway analysis [34–36]. This analysis was performed to provide more information about the biological functions and pathways significantly enriched in up- or down-regulated genes by focusing on gene ontology (GO) terms (BP, biological process), Kyoto Encyclopedia of Genes and Genomes (KEGG), and interactome pathways (using a standard false discovery rate (FDR)<0.05). The up- and down-regulated genes from different groups were inputted together into NetworkAnalyst 3.0 (https://www.networkanalyst.ca) and the STRING interactome protein-protein interaction database was used with a confidence score cut-off of 700.

Results
Quality assessment of data
Raw reads obtained from all six samples averaged to 40 million reads whereas the average clean reads were 28.11 million reads. The percentage GC pre and post trimming were 50.94 and 47.61 respectively. The % of million reads having quality score above 30 before trimming were 85.44 and after trimming the % of million reads having quality score above PHRED 30 were 85.97. The average uniquely mapped percentage out of the total clean reads were 75.6%, and 62.7% for Jakhrana early lactation and Jakhrana late lactation respectively (Table 1).
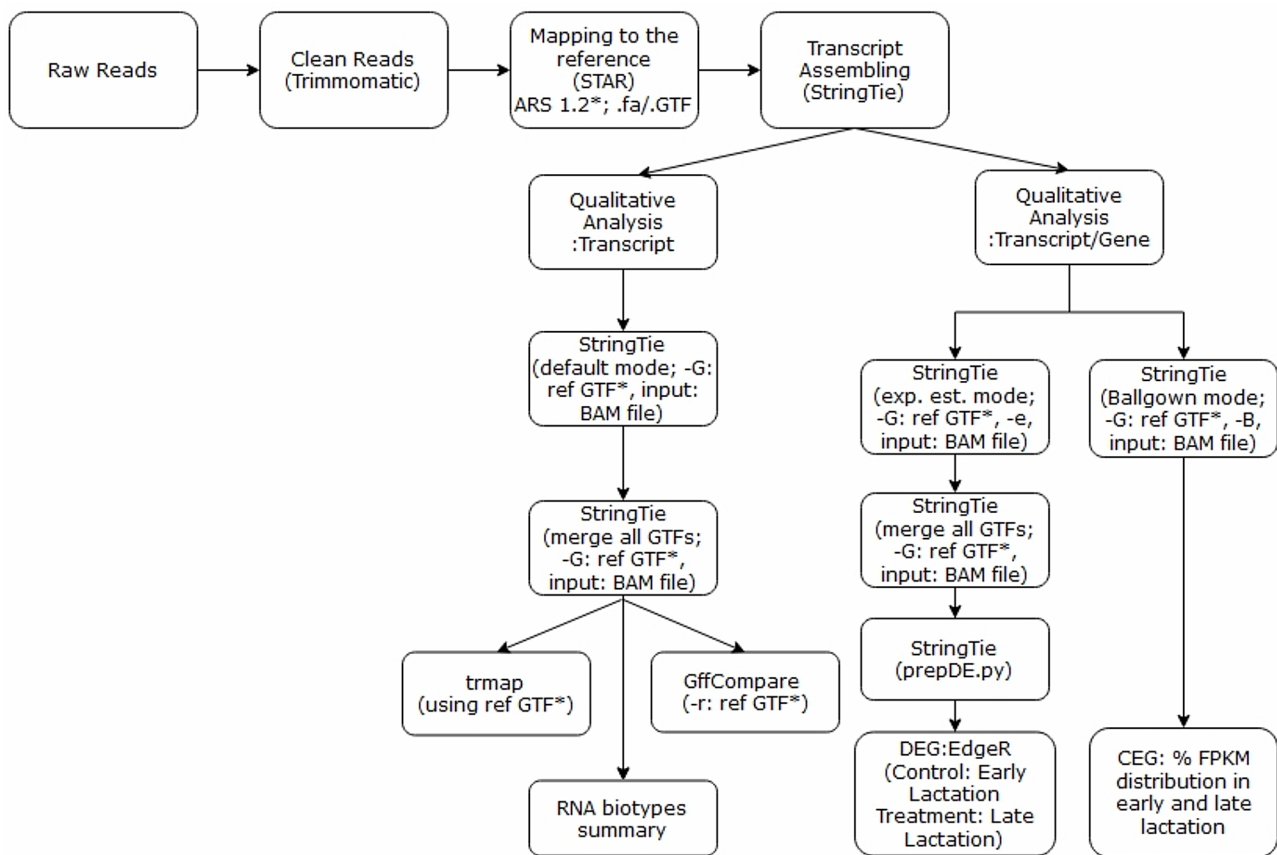
Fig. 1 The methodological pipeline used to generate the differentially expressed genes (DEGs) and commonly expressed genes (CEGs) in the present study. The reference fasta and reference GTF was derived from NCBI RefSeq assembly GCF_001704415.2 (ARS1.2)

Table 1 The summary of raw fastQ and the clean or trimmed fastQ files along with mapping statistics

| Group Name	Sample Name	Read Strand	Raw fastQ files			Clean fastQ files			Alignment summary
			% GC	M Seqs	Q30%	% GC	M Seqs	Q30%	Uniquely aligned (M reads)
Early lactation	A	F	48	58.1	87.14	48	50.4	86.94	38.4
		R	49	58.1	86.02	48	50.4	88.41	
	B	F	52	26	87.79	52	22.4	87.75	16.1
		R	53	26	81.86	53	22.4	82.63	
	D	F	47	33.8	88.21	47	30.2	87.9	23.4
		R	47	33.8	87.45	48	30.2	89.72	
Late lactation	H	F	49	28.2	90.71	49	25	90.769	18.9
		R	50	28.2	83.11	50	25	84.13	
	I	F	56	31.5	91.89	56	28.3	92.03	16.3
		R	59	31.5	67.45	60	28.3	66.75	
	JEIKT	F	53	39.1	88.07	54	32	87.61	18.3
		R	54	39.1	85.58	55	32	87.11	

%Dups refers to the percentage of duplicated reads, %GC indicates the percentage GC content, M Seqs refers to the total number of reads per sample in millions, and Q30% refers to the percentage of reads having PHRED quality score above 30. The forward read strand and the reverse read strand has been represented as F and R respectively.

Transcript analysis

Utilizing StringTie for assembly, we identified a total of 166,154 transcripts from the merged GTFs of all six samples (Table 2). The qualitative analysis enabled the classification of transcripts based on their code and RNA biotypes. Among these transcripts, 49,285 were identified as complete transcripts. The majority of transcripts were categorized as mRNA (80%), with the remaining

Table 2 The classification of transcripts based on their assembly to the reference GTF

| Transcript classification codes	Description	No of transcript(s)
x	Exonic overlap on the opposite strand (like 'o' or 'e' but on the opposite strand	2
u	None of all (unknown, intergenic)	78,565
p	Possible polymerase run-on (no actual overlap)	1192
o	Other same strand overlap with ref exons	303
n	Retained intron(s), not all introns matched/covered	1870
m	Retained intron(s), all introns matched or retained	2418
k	Containment of reference (reverse containment)	1203
j	Multi-exon with at least one junction match	5888
i	Fully contained within a reference intron	25,244
e	Single exon transfrag partially covering an intron, possible pre-mRNA fragment	1
c	Contained in reference (intron compatible)	183
=	Complete, exact match of intron chain	49,285
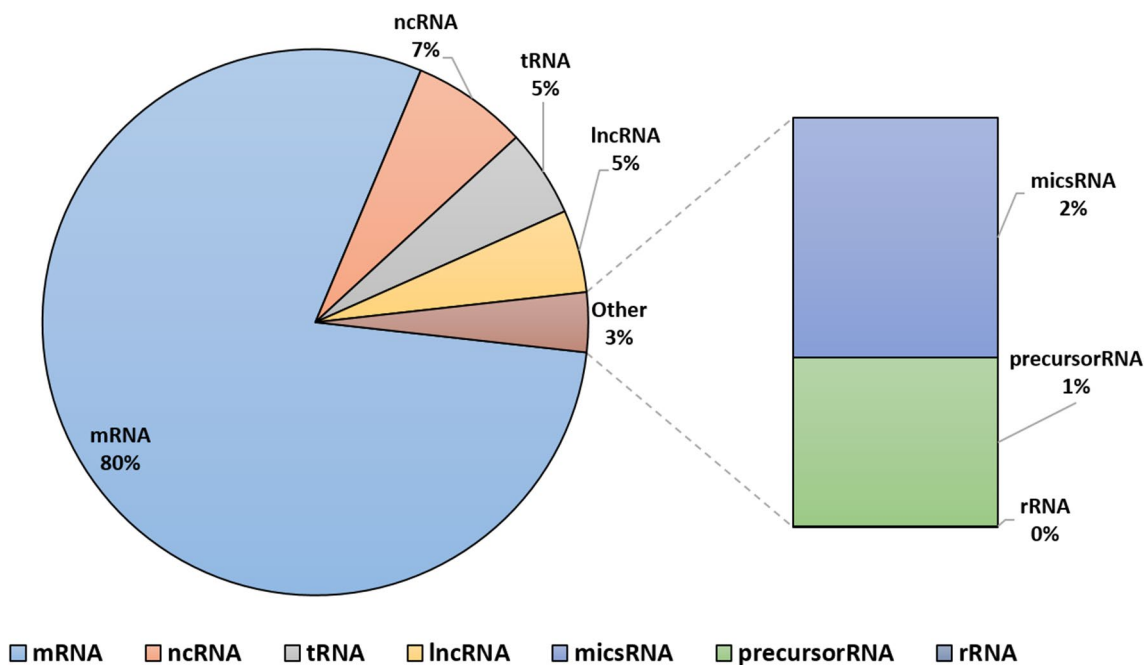
Fig. 2 The RNA biotypes classification of the transcripts matching exact to the intron chain of reference GTF, represented as '=' code in Table 2

belonging to ncRNA, tRNA, lncRNA, micsRNA, precursorRNA, and rRNA categories (Fig. 2).

Stringtie assembling and FPKM distribution
After mapping the reads with StringTie, we successfully mapped 25,485 genes and 49,556 transcripts to the reference Capra hircus genome using the -e and -B options, respectively, for estimating expression and Ballgown compatibility. Among these, 14,068 genes had a single transcript, while 9,256 genes had more than one transcript (Fig. 3A). The average transcript length across all samples was approximately 3,165 (Fig. 3B). The total FPKM values were 459,531.45, 739,544.01, 356,543.83, 548,791.64, 702,267.91, and 351,482.39 for the three Jakhrana early lactation and three Jakhrana late lactation, respectively (Fig. 3C; Supplementary Fig. S1). The top 20 genes contributing the highest percentage of FPKM across all libraries were *CSN2* (*n*=2), *CSN3*, *CSN1S2* (*n*=2), *PAEP*, *CSN1S1* (*n*=4), *GLYCAM1*, *LALBA*, *SRGN*, *LOC102169149*, *TMSB10*, *FTH1*, *SDS*, *TXNIP*, *LOC102184299*, and *S100A8* (Fig. 4A). Among these, *FTH1*, *GLYCAM1*, *LOC102169149*, *LOC102184299*, *S100A8*, *TMSB10*, *LALBA*, *TXNIP*, and *SDS* showed higher FPKM distribution in the early stages compared to the late stages. *CSN2*, *CSN1S2*, and *PAEP* had higher FPKM distribution during the late lactation stage (Fig. 4B). Most of the commonly expressed genes (CEGs) were enriched as milk proteins except for *GLYCAM1* and *SRGN* (Supplementary Fig. S2).
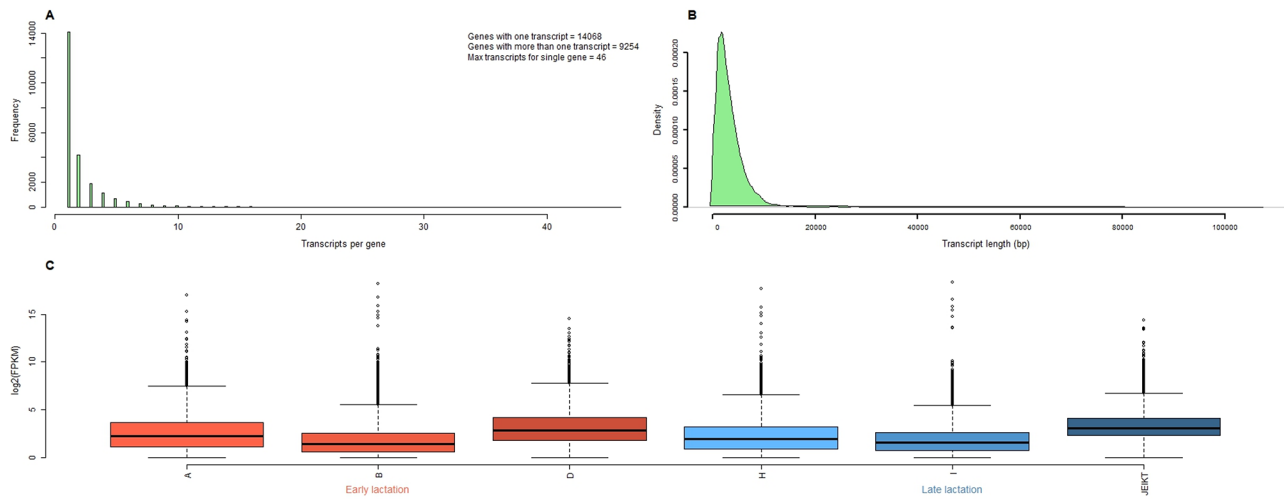
Fig. 3 **A**. Histogram presenting the distribution of transcript count per gene. **B**. The density histogram presenting the distribution of transcript length in the dataset. **C**. The box and whisker plot presenting the distribution of FPKM values across all 6 samples. The FPKM has been represented as log2 transformed values on the Y-axis
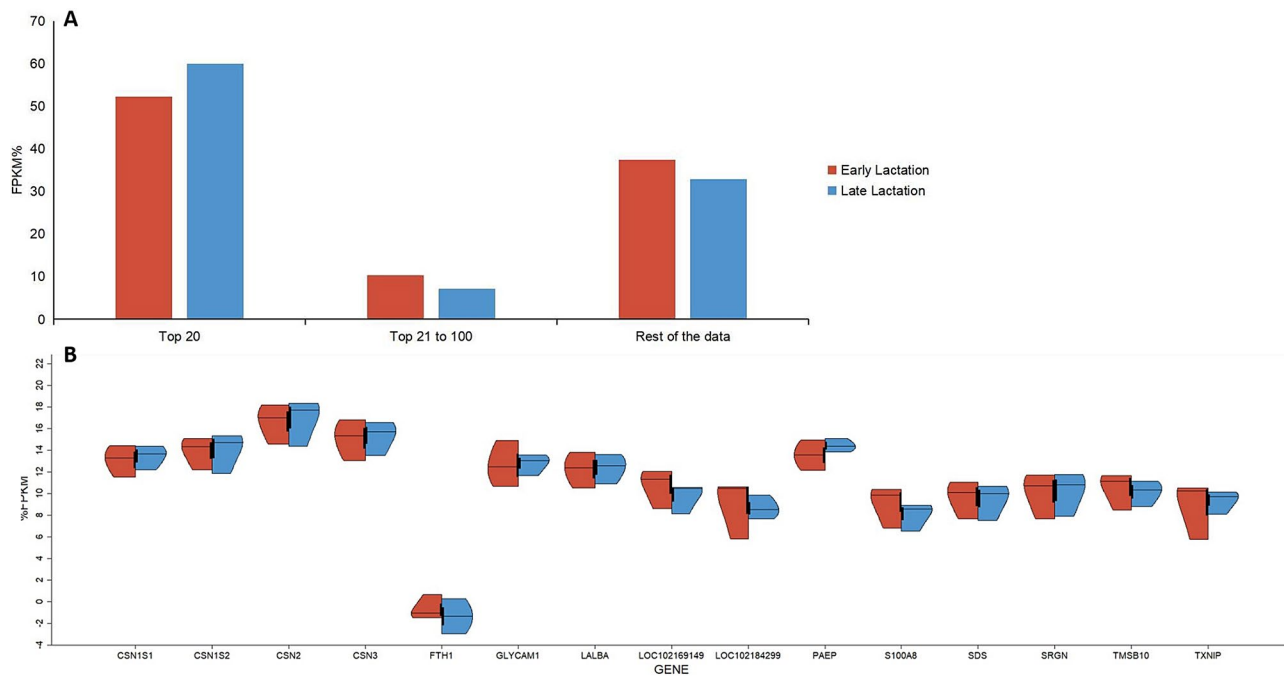

Fig. 4 **A**. the barplot represent the percentage of total FPKM in the JK_early lactation and JK_late lactation group represented by proportion of genes indicated in the X-axis. **B**. The split violin plot represent the log2transformed average FPKM values of CSN1S1, CSN1S2, CSN2, CSN3, FTH1, GLYCAM1, LALBA, LOC102160149, LOC120184209, PAEP, S100A8, SDS, SRGN, TMSB10, and TXNP during the early and late lactation in Jakhrana goat

Differential gene expression analysis

The psuedogene count generated from StringTie were considered for the analysis of genes that are overall differentially expressed (DEG) based on various FDR cutoffs (0.5 and 0.05). The commonly expressed genes essentially neither form the most differentially expressed genes nor they were statistically significant (Supplementary Table 1),The multidimensional scaling plot (MDS) generated using edgeR couldn't segregate two groups of Jakhrana

goat by lactation stages, the transitional stage between the lactation phase may have diluted the clustering (Fig. 5A). To establish a consistent cutoff for analysis, the library sizes of the six samples were considered, resulting in a tags/genes reduction from 25,485 to approximately 12,773 using a counts per million (CPM) cutoff of 1.9 (Fig. 5B). The mean dispersion across all genes was calculated to be 0.40706, with a gene-wise biological coefficient of variation of 0.638 (Fig. 5C). Differential
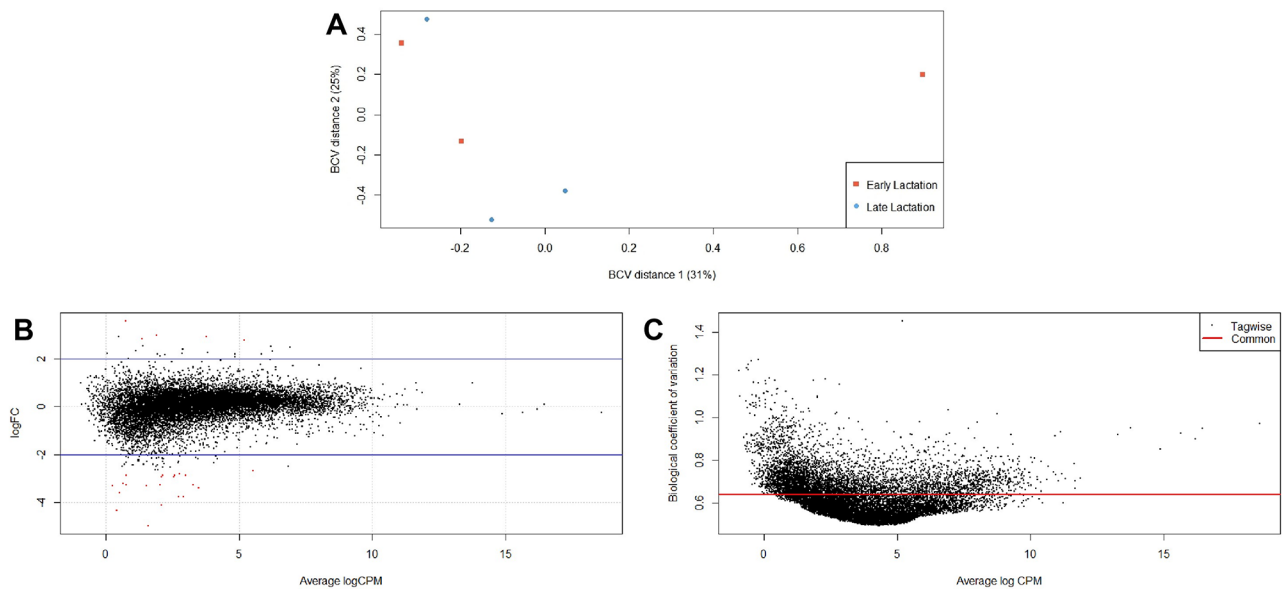
Fig. 5 A. The multidimensional scaling plot representing the clustering of late lactation (JK_late) and early lactation (JK_early) samples. **B**. the MA plot showing the relationship between logCPM (count per million) and logFC (fold-change) across the tags/genes (black dots). The blue solid represent the logFC cut-off (+/- 2.5). The red dot represent the significantly expressed genes/tags at FDR 0.5
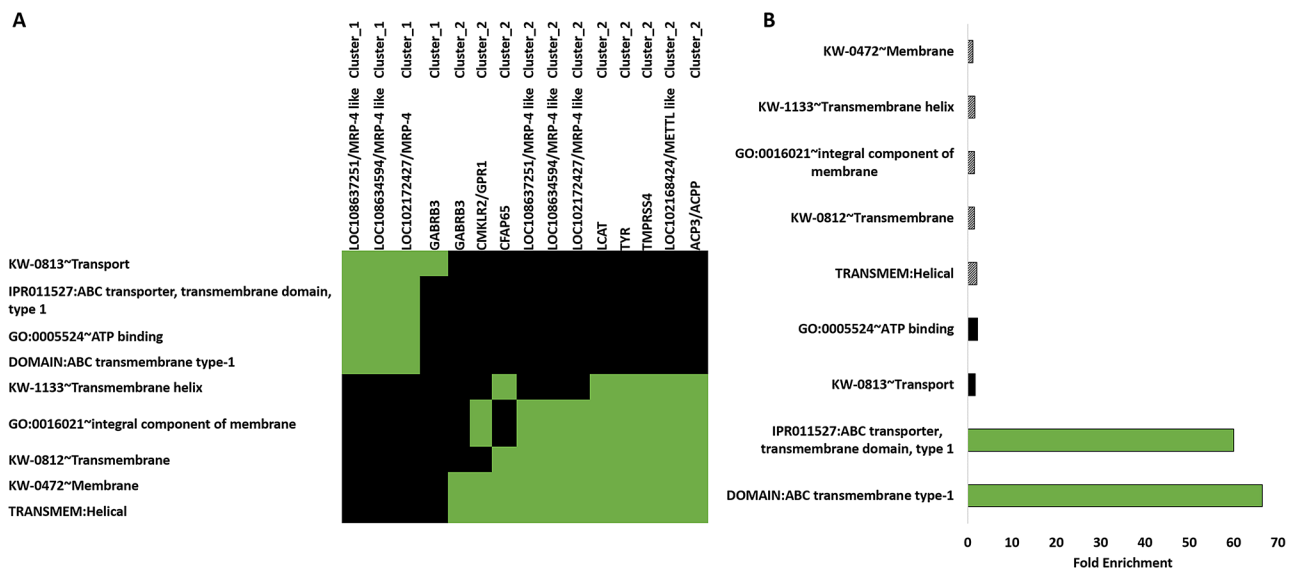
Fig. 6 A. Clusters identified using DAVID functional annotation. Two clusters were identified using the DAVID functional annotation clustering tool. **B**. Bar graph representing the DEGs enriched for the terms at various fold enrichments. The significantly enriched terms have been represented in green bars (Bonferroni FDR < 0.05)

expression analysis was conducted based on these parameters, using the dispersion value and exact test deviance goodness of fit statistics, with JK Late lactation considered as the control and JK early lactation as the treatment group. The analysis identified a total of 5 DEGs that were downregulated (logFC < -2.5; FDR<0.05), 7 DEGs that were downregulated (logFC < -2.5; FDR<0.1), 10 downregulated and 1 upregulated (logFC>+2.5 & logFC < -2.5; FDR<0.2), and 21 downregulated and 5 upregulated (logFC>+2.5 & logFC < -2.5; FDR<0.5) genes (Supplementary Table S2).

Functional annotation of differentially expressed genes
The functional enrichment using DAVID for genes 21 downregulated and 5 upregulated (logFC>+2.5 & logFC< -2.5; FDR<0.5) could enrich the terms in two prominent annotation clusters (Fig. 6A). Cluster one with an enrichment Score: 1.747 and cluster two with an enrichment Score: 0.941 comprised 4 terms (DOMAIN: ABC

transmembrane type-1, IPR011527: ABC transporter/transmembrane domain, type 1, KW-0813 ~ Transport, and GO: 0005524 ~ ATP binding) and 5 terms (TRANSMEM: Helical, KW-0812 ~ Transmembrane, GO: 0016021 ~ integral component of membrane, KW-1133 ~ Transmembrane helix, and KW-0472 ~ Membrane). At Bonferroni p-value correction < 0.1, DOMAIN: ABC transmembrane type-1 and IPR011527: ABC transporter, transmembrane domain, type 1 were significantly enriched (Sig indicated as green bar). The genes associated with these terms i.e., multidrug resistance-associated protein 4-like (*LOC108637251*), multidrug resistance-associated protein 4-like (*LOC108634594*), and multidrug resistance-associated protein 4 (*LOC102172427*) were significantly downregulated (logFC< -2.5; FDR < 0.5) in JK_early group (Fig. 6B).

Discussion

The mammary gland has a dynamic and highly adaptive molecular profile enabling varying composition of milk, strategic defense against pathogen, and degree of apoptosis during the stages of lactation [23, 37]. Indigenous goat breeds in arid and semi-arid zones of developing countries, such as India, have adapted to local agro-climatic conditions and are valued for their medicinal milk properties, yet the molecular profiles of these goats during different lactation stages remain poorly understood [38]. Among the various factors influencing the milk composition of the goat, lactation stage is one of the crucial factor since from the onset of gestation to the involution the mammary gland goes through number of cycles of cell proliferation and degeneration [39]. The extremes of the lactation cycles including the early and late is when the number of mammary alveoli starts proliferating and decline in number respectively. One such case has been studied in the present report, the Jakhrana goats in late and early lactation stages were profiled for transcriptome derived from milk somatic cells. The study presents a detailed transcriptome profile of Jakhrana goat's milk somatic cells during early and late lactation phase, emphasizing the FPKM based gene abundance and the log fold change in the early lactation group compared with late lactation.

The comparative analysis of commonly expressed genes based on FPKM in early and late lactation stages in Jakhrana goats revealed distinct trends. Among the top commonly expressed genes (*CSN2, CSN3, CSN1S2, PAEP, CSN1S1, GLYCAM1, LALBA, SRGN, LOC102169149, TMSB10, FTH1, SDS, TXNIP, LOC102184299*, and *S100A8*), six were prominent milk protein encoding genes (*PAEP, LALBA, CSN3, CSN1S1, CSN1S2*, and *CSN2*), consistent with previous findings in goat and cattle regarding the importance of these genes in milk production [16, 40]. Notably, the CSN2, CSN1S2,

and other casein genes exhibited higher expression levels during late lactation and late gestation compared to the dry period [41]. However, some studies suggest that the mRNA abundance of these caseins is actually greater in the early lactation stage [42].

All these milk protein genes exhibited higher FPKM values in the late lactation stage compared to early lactation, except for *LALBA*, which showed relatively constant expression across both stages. This stability in expression aligns with findings in buffaloes, where *LALBA* expression remained consistent across all lactation stages [18]. On the other hand, *FTH1, GLYCAM1, LOC102169149* (*S100-A12*), *LOC102184299* (*Serpin B3*), *S100A8, SDS, TMSB10*, and *TXNIP* demonstrated higher expression levels during early lactation. *FTH1* (Ferritin heavy chain 1), which encodes a major intracellular iron-storage protein, has been previously reported to have high FPKM during late lactation stages in bovines [16]. This finding contrasts with our study, although *FTH1* is also a highly prioritized protein expressed during the colostrum stage [40, 43]. Similarly, *GLYCAM1*, a gene in the mucin family encoding a milk fat globule glycoprotein, showed higher expression during early lactation in caprine milk, consistent with previous studies in bovine milk [16, 44]. *GLYCAM1* serves as an antimicrobial and lubricant for the neonatal intestinal tract during early lactation, regulated primarily by prolactin [45, 46]. Two S100 family protein encoding genes (*S100A8* and *S100A12*) were abundantly expressed in early lactation stage in caprine milk in the present study, this Ca2+-binding proteins act as antimicrobial peptides during an innate immune response [47]. This antimicrobial protein has also been reported earlier in milk from the Indian goat breeds [48]. This contradicts the results reported in Holstein Frisian, where the late lactation stages have been reported to express higher levels of *S100A8* and *S100A12* [49]. Likewise, *LOC102184299* (Serpin B3-like) is another antimicrobial peptide was over-abundant in early lactation stages. The FPKM-based trend suggests that milk proteins were abundant in late lactation stages, whereas genes encoding antimicrobial and lubrication-related proteins were abundant in early lactation stages, protecting neonate kids against common pathogens.

Differential gene expression analysis between early and late lactation stages in Jakhrana goats revealed several key findings. In the early lactation phase, several genes were significantly downregulated, including *MUC21, MPRSS4, NCCRP1, MUC2*, and *KRT78*. Among these, *MUC21* is involved in mucin production, suggesting a decrease in mucin levels during early lactation. This is intriguing, as mucins play a crucial role in mucosal protection and immune response in the mammary gland. The downregulation of these genes may indicate a shift in the mammary gland function towards milk production

and secretion. Furthermore, the study identified three major ABCC transporter genes, *LOC102172427* (*MRP4*), *LOC108634594* (*MRP4-like*), and *LOC108637251* (*MRP4-like*), that were downregulated in early lactation. These transporters are part of the ATP-binding cassette (ABC) transporter family, responsible for the efflux of metabolites, xenobiotics, and lipids into mammary epithelial cells [50]. The downregulation of these transporters suggests a modification in the composition of milk during early lactation, possibly to meet the changing nutritional needs of the neonates. It is noteworthy that human mammary epithelial cells possess MRP1, MRP2, and MRP5, with MRP4 reported in mammary gland tissues [51–53]. This indicates a conserved mechanism of drug resistance and xenobiotic efflux in mammary epithelial cells across species. Previous studies have reported the presence of an effective blood-milk barrier for bile acids, likely involving the MRP4 transporter, as indicated by [52]. The role of MRP4 and MRP4-like transporters in imparting drug resistance during lactation warrants further investigation, particularly in understanding the expression magnitude of MRP4 efflux transporter in mammary epithelial cells across indigenous and exotic goat breeds.

The downregulation of mucins during early lactation, as observed in this study, is intriguing, given their role in antimicrobial defense and lubrication. This finding contrasts with the overabundance of GLYCAM1, which also plays a role in antimicrobial and lubricating effects. This discrepancy may suggest a complex regulatory mechanism involving multiple factors in milk composition during lactation. Moreover, the upregulation of the *LCAT* gene during early lactation is interesting, as it encodes lecithin-cholesterol acyltransferase, which metabolizes cholesterol to cholesteryl ester. This finding suggests a potential role of LCAT in milk lipid metabolism and may have implications for milk yield regulation. In Sarda sheep, specific SNPs in the *LCAT* gene have been linked to milk yield [54]. On the other hand, LCAT activity has been observed to decrease shortly after parturition in cattle [55].

Conclusion

This study delved into the differential gene expression patterns between early and late lactation phases in Jakhrana goats. Our findings shed light on the abundance of genes encoding antimicrobial peptides and proteins during early lactation, which likely play an important role in protecting neonates from invasive pathogens. Key genes such as *FTH1, GLYCAM1, LOC102169149 (S100-A12), LOC102184299 (Serpin B3), S100A8, SDS, TMSB10,* and *TXNIP* were identified as potential contributors to this protective mechanism. Conversely, late lactation stages exhibited an upregulation of genes encoding major milk proteins, indicative of a shift towards milk production and secretion. Our analysis also revealed the upregulation of the ABC transporter gene *MRP4* during late lactation, highlighting its importance in xenobiotic efflux. These findings provide valuable insights into the molecular mechanisms underlying lactation in Jakhrana goats.

Declarations

Ethics approval
All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. Ethical approval was not required for the study as no experiment was performed on the animals. Milk samples were collected from animals milked for commercial purpose.

Consent for publication
The manuscript has consent of all the authors for its publication in current format.

Competing interests
The authors declare no competing interests.

References
1. Devendra C, Goats. Challenges for increased productivity and improved livelihoods. Outlook Agric. 1999;4:215–26.
2. Pal M, Dudhrejiya TP, Pinto S. Goat milk products and their significance. Beverage Food World. 2017;44:21–5.
3. Liang JB, Devendra C. Expanding the contribution of dairy goats in efficient and sustainable production systems. Anim Prod Sci. 2014;54:1198–203.
4. Li Z, Wang X, Deng X, Song J, Yang T, Liao Y, Gong G, Huang L, Lu Y, Wang Z. High-sensitivity qualitative and quantitative analysis of human, bovine and

Dige *et al. BMC Genomics*   (2024) 25:874

Page 10 of 11

goat milk glycosphingolipids using HILIC-MS/MS with internal standards. Carbohydr Polym. 2023;312:120795.

5. Alférez MJM, López Aliaga I, Barrionuevo M, Campos MS. Effect of dietary inclusion of goat milk on the bioavailability of zinc and selenium in rats. J Dairy Res. 2003;70:181–7.

6. Barrionuevo M, López Aliaga I, Alférez MJM, Mesa E, Nestáres T, Campos MS. Beneficial effect of goat milk on bioavailability of copper, zinc and selenium in rats. J Physiol Biochem. 2003;59:111–8.

7. Lara-Villoslada F, Debras E, Nieto A, Concha A, Gálvez J, López-Huertas E, Boza J, Obled C, Xaus J. Oligosaccharides isolated from goat milk reduce intestinal inflammation in a rat model of dextran sodium sulfate-induced colitis. Clin Nutr. 2006;25:477–88.

8. Ma Y, Li J, Huang Y, Liu X, Dou N, Zhang X, Hou J, Ma J. Physicochemical stability and in vitro digestibility of goat milk affected by freeze-thaw cycles. Food Chem. 2023;404:134646.

9. Park YW, Haenlein GFW. Therapeutic, hypo-allergenic and bioactive potentials of goat milk, and manifestations of food allergy. In: Park YW, Haenlein GFW, Wendorff WL, editors Handbook of milk of non- bovine mammals. 2017; 151–179.

10. Niaz B, Saeed F, Ahmed A, Imran M, Maan AA, Khan MKI, Tufail T, Anjum FM, Hussain S, Suleria HAR. Lactoferrin (LF): a natural antimicrobial protein. Int J Food Prop. 2019;22:1626–41.

11. Dhasmana S, Das S, Shrivastava S. Potential nutraceuticals from the casein fraction of goat's milk. J Food Biochem. 2022;46(6):e13982.

12. Casado B, Affolter M, Kussmann M. OMICS-rooted studies of milk proteins, oligosaccharides and lipids. J Proteom. 2009;196–208. https://doi.org/10.1016/j.jprot.2009.09.018.

13. German JB, Dillard CJ, Ward RE. Bioactive components in milk. Current opinion in Clinical Nutrition and Metabolic Care. 2002; 5: 653–8. https://doi.org/10.1097/01.mco.0000038808.16540.7f

14. Shi H, Zhu J, Luo J, Cao W, Shi H, Yao D, Li J, Sun Y, Xu H, Yu K, Loor JJ. Genes regulating lipid and protein metabolism are highly expressed in mammary gland of lactating dairy goats. Funct Integr Genom. 2015;15:309–21. https://doi.org/10.1007/s10142-014-0420-1.

15. Mandal A, Roy R, Bhusan S, Rout PK, Sharma MC. Environmental effects on production traits of Jakhrana goat. Indian J Anim Sci. 2010;80(11):1141.

16. Wickramasinghe S, Rincon G, Islas-Trejo A, Medrano JF. Transcriptional profiling of bovine milk using RNA sequencing. BMC Genomics. 2012;13:45. https://doi.org/10.1186/1471-2164-13-45.

17. Gomes RD, Anaya K, Galdino AB, Oliveira JP, Gama MA, Medeiros CA, Gavioli EC, Porto ALF, Rangel AH. Bovine colostrum: a source of bioactive compounds for prevention and treatment of gastrointestinal disorders. NFS J. 2021;25:1–11. https://doi.org/10.1016/j.nfs.2021.10.001.

18. Arora R, Sharma A, Sharma U, Girdhar Y, Kaur M, Kapoor P, Ahlawat S, Vijh RK. Buffalo milk transcriptome: a comparative analysis of early, mid and late lactation. Sci Rep. 2019;9(1):5993.

19. Hao Z, Zhou H, Hickford JG, Gong H, Wang J, Hu J, Liu X, Li S, Zhao M, Luo Y. Transcriptome profile analysis of mammary gland tissue from two breeds of lactating sheep. Genes. 2019;10(10):781.

20. Yao H, Dou Z, Zhao Z, Liang X, Yue H, Ma W, Su Z, Wang Y, Hao Z, Yan H, Wu Z, Wang L, Chen G, Yang J. Transcriptome analysis of the bactrian camel (*Camelus bactrianus*) reveals candidate genes affecting milk production traits. BMC Genomics. 2023;24(1):660. https://doi.org/10.1186/s12864-023-09703-9.

21. Li Q, Liang R, Li Y, Gao Y, Li Q, Sun D, Li J. Identification of candidate genes for milk production traits by RNA sequencing on bovine liver at different lactation stages. BMC Genet. 2020;21:1–12.

22. Michailidou S, Gelasakis A, Banos G, Arsenos G, Argiriou A. Comparative transcriptome analysis of milk somatic cells during lactation between two intensively reared dairy sheep breeds. Front Genet. 2021;12:700489.

23. Xuan R, Chao T, Zhao X, Wang A, Chu Y, Li Q, Zhao Y, Ji Z, Wang J. Transcriptome profiling of the nonlactating mammary glands of dairy goats reveals the molecular genetic mechanism of mammary cell remodeling. J Dairy Sci. 2022;105(6):5238–60.

24. Yang Y, Zheng N, Yang J, Bu D, Wang J, Ma L, Sun P. Animal species milk identification by comparison of two-dimensional gel map profile and mass spectrometry approach. Int Dairy J. 2014;35(1):15–20.

25. Crisà A, Ferrè F, Chillemi G, Moioli B. RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. BMC Vet Res. 2016;12(1):264.

26. Zorc M, Dolinar M, Dovč P. A single-cell transcriptome of bovine milk somatic cells. Genes (Basel). 2024;15(3):349.

27. Andrews S, FastQC. A quality control tool for high throughput sequence data. Retrieved from www.bioinformatics.babraham.ac.uk/projects/fastqc/.2010

28. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; (15): 2114–20.

29. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.

30. Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. F1000Res. ISCB Comm. 2020; J-304. https://doi.org/10.12688/f1000research.23297.2

31. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290–5.

32. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.

33. Benjamini Y, Hochberg Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. J Educational Behav Stat. 2000;25(1):60–83.

34. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022;50(W1):W216–21.

35. Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nat Protoc. 2015;10(6):823–44.

36. Zhou G, Soufan O, Ewald J, Hancock RE, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. 2019;47(W1):W234–41.

37. Xuan R, Wang J, Zhao X, Li Q, Wang Y, Du S, Duan Q, Guo Y, Ji Z, Chao T. Transcriptome analysis of goat mammary gland tissue reveals the adaptive strategies and molecular mechanisms of lactation and involution. Int J Mol Sci. 2022;23(22):14424.

38. Singh G, Sharma RB, Kumar A, Chauhan A. Effect of stages of lactation on goat milk composition under field and farm rearing condition. Adv Anim Veterinary Sci. 2014;2(5):287–91.

39. Paape MJ, Poutrel B, Contreras A, Marco JC, Capuco AV. Milk somatic cells and lactation in small ruminants. J Dairy Sci. 2001;84:E237–44.

40. Crisa A, Ferre F, Chillemi G, Moioli B. RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. BMC Vet Res. 2016;264. https://doi.org/10.1186/s12917-016-0881-7.

41. Xuan R, Wang J, Zhao X, Li Q, Wang Y, Du S, Duan Q, Guo Y, Ji Z, Chao T. Transcriptome Analysis of Goat Mammary Gland Tissue Reveals the Adaptive Strategies and Molecular Mechanisms of Lactation and Involution. Int J Mol Sci. 2022 Nov 20;23(22):14424. https://doi.org/10.3390/ijms232214424.

42. Sharma A, Shandilya UK, Sodhi M, Jatav P, Mohanty A, Jain P, Verma P, Kataria RS, Kumari P, Mukesh M. Milk-derived mammary epithelial cells as non-invasive source to define stage-specific abundance of milk protein and fat synthesis transcripts in native Sahiwal cows and Murrah buffaloes. 3 Biotech. 2019;9(3):106. https://doi.org/10.1007/s13205-019-1642-7.

43. Lemay DG, Hovey RC, Hartono SR, Hinde K, Smilowitz JT, Ventimiglia F, Schmidt KA, Lee JW, Islas-Trejo A, Silva PI, Korf I. Sequencing the transcriptome of milk production: milk trumps mammary tissue. BMC Genomics. 2013; 14: 1–17.

44. Le Provost F, Cassy S, Hayes H, Martin P. Structure and expression of Goat GLYCAM1 gene: Lactogenic-Dependent expression in Ruminant Mammary Gland and Interspecies Conservation of the proximal promoter. Gene. 2003;313:83–9.

45. Bhat MI, Sowmya K, Kapila S, Kapila R. Potential probiotic Lactobacillus rhamnosus (MTCC-5897) inhibits Escherichia coli impaired intestinal barrier function by modulating the host tight junction gene response. Probiotics Antimicrob Proteins. 2020;12:1149–60.

46. Hou Z, Bailey JP, Vomachka AJ, Matsuda M, Lockefeer JA, Horseman ND. Glycosylation-dependent cell adhesion molecule 1 (GlyCAM 1) is induced by prolactin and suppressed by progesterone in mammary epithelium. Endocrinology. 2000;141(11):4278–83.

47. Singh P, Ali SA. Multifunctional role of S100 protein family in the Immune System: an update. Cells. 2022;11(15):2274. https://doi.org/10.3390/cells11152274.

48. Verma M, Dige MS, Gautam D, De S, Rout PK. Functional milk proteome analysis of genetically diverse goats from different agro climatic regions. J Proteom. 2020;227:103916. https://doi.org/10.1016/j.jprot.2020.103916.

49. Han Z, Fan Y, Yang Z, Loor JJ, Yang Y. Mammary transcriptome profile during peak and late lactation reveals differentially expression genes related to inflammation and immunity in Chinese Holstein. Animals. 2020;10(3):510.

50. Farke C, Meyer HHD, Bruckmaier RM, Albrecht C. Differential expression of ABC transporters and their regulatory genes during lactation and dry period in bovine mammary tissue. J Dairy Res. 2008;75(4):406–14.

51. Alcorn J, Lu X, Moscow JA, McNamara PJ. Transporter gene expression in lactating and nonlactating human mammary epithelial cells using real-time reverse transcription-polymerase chain reaction. J Pharmacol Exp Ther. 2002;303(2):487–96.

52. Blazquez AM, Macias RI, Cives-Losada C, de la Iglesia A, Marin JJ, Monte MJ. Lactation during cholestasis: role of ABC proteins in bile acid traffic across the mammary gland. Sci Rep. 2017;7(1):7475.

53. Empey PE. Xenobiotic Transporters in Lactating Mammary Epithelial Cells: Predictions for Drug Accumulation in Breast Milk. 2007; Ph.D. dissertation, University of Kentucky.

54. Dettori ML, Petretto E, Mingioni P, Vacca GM, Pazzola M. Variability of genes involved in lipid metabolism and their effect on milk yield, composition and coagulation traits in Sarda sheep. Sci E Tec Latt Casearia2022; 57–65.

55. Kessler EC, Bruckmaier RM, Gross JJ. Milk production during the colostral period is not related to the later lactational performance in dairy cows. J Dairy Sci. 2014;97(4):2186–92.

Publisher's note