

RESEARCH

Open Access



# Mammary epithelial cell transcriptome reveals potential roles of lncRNAs in regulating milk synthesis pathways in Jersey and Kashmiri cattle

Peerzada Tajamul Mumtaz<sup>1,2†</sup>, Basharat Bhat<sup>3†</sup>, Eveline M. Ibeagha-Awemu<sup>4</sup>, Qamar Taban<sup>1</sup>, Mengqi Wang<sup>4</sup>, Mashooq Ahmad Dar<sup>1</sup>, Shakil Ahmad Bhat<sup>1</sup>, Nadeem Shabir<sup>1</sup>, Riaz Ahmad Shah<sup>1</sup>, Nazir A. Ganie<sup>1</sup>, Dinesh Velayutham<sup>5</sup>, Zulfqar ul Haq<sup>6</sup> and Syed Mudasir Ahmad<sup>1\*</sup>

## Abstract

**Background:** Long noncoding RNAs (lncRNAs) are now proven as essential regulatory elements, playing diverse roles in many biological processes including mammary gland development. However, little is known about their roles in the bovine lactation process.

**Results:** To identify and characterize the roles of lncRNAs in bovine lactation, high throughput RNA sequencing data from Jersey (high milk yield producer), and Kashmiri cattle (low milk yield producer) were utilized. Transcriptome data from three Kashmiri and three Jersey cattle throughout their lactation stages were utilized for differential expression analysis. At each stage (early, mid and late) three samples were taken from each breed. A total of 45 differentially expressed lncRNAs were identified between the three stages of lactation. The differentially expressed lncRNAs were found co-expressed with genes involved in the milk synthesis processes such as *GPAM*, *LPL*, and *ABCG2* indicating their potential regulatory effects on milk quality genes. KEGG pathways analysis of potential *cis* and *trans* target genes of differentially expressed lncRNAs indicated that 27 and 48 pathways were significantly enriched between the three stages of lactation in Kashmiri and Jersey respectively, including mTOR signaling, PI3K-Akt signaling, and RAP1 signaling pathways. These pathways are known to play key roles in lactation biology and mammary gland development.

**Conclusions:** Expression profiles of lncRNAs across different lactation stages in Jersey and Kashmiri cattle provide a valuable resource for the study of the regulatory mechanisms involved in the lactation process as well as facilitate understanding of the role of lncRNAs in bovine lactation biology.

**Keywords:** Cattle, Mammary epithelial cells, Lactation stages, lncRNA, RNA sequencing

## Introduction

The mammary gland is a vital organ for milk synthesis and secretion, providing essential nutrients for mammalian offspring and human nourishment. Lactation is the maternal physiological response, however, the milk yield greatly varies among cattle breeds [1, 2]. The lactation process is influenced by genetic, epigenetic, and environmental factors [3]. Elucidating regulatory mechanisms of

\*Correspondence: mudasirbio@gmail.com

<sup>†</sup>Peerzada Tajamul Mumtaz and Basharat Bhat contributed equally to this work.

<sup>1</sup> Division of Animal Biotechnology, Faculty of Veterinary Sciences and Animal Husbandry, Sher-e- Kashmir University of Agricultural Sciences and Technology – Kashmir, SKUAST-K, Shuhama, Jammu 190006, India

Full list of author information is available at the end of the article



the lactation process is not only vital for improving milk quality and production but also helps in understanding other processes related to milk synthesis. Mammary epithelial cells are the factories of milk lipids, proteins, and carbohydrates during lactation [4]. The milk-producing mammary epithelial cells (MECs) are the functional unit of mammary gland, and the proliferation of MECs is a key determinant of lactation of mammary gland [5]. Various genes and regulatory molecules are expressed at different lactation stages, which play crucial roles in the regulation of lactation [6–8]. To improve the milk production performance in cattle, deeper understanding of the molecular biology of lactating mammary glands, such as expression and regulatory mechanisms of milk related genes, including the regulatory roles of lncRNAs is essential.

Previously thought of as junk transcripts and pseudogene remnants, non-coding RNA molecules including lncRNA have emerged as essential components of cellular activity, regulating a plethora of functions within multicellular organisms [9]. lncRNAs are defined as transcripts  $\geq 200$  bp in length and without the potential to code for a protein. They are often polyadenylated and usually exhibit low expression levels and poor sequence conservation [10–13]. lncRNAs activate or repress genes at multiple levels (i.e., DNA, RNA and proteins) by acting as guide, signaling, scaffolding or decoy molecules [14]. It has been shown that lncRNAs play important roles in certain tissues as well as have species-specific expression [15], but some multifunctional lncRNAs have also been shown to display broad and conservative expression profiles [16, 17]. In mammals, lncRNAs are known to play roles in various biological processes including growth, reproduction and health [18, 19]. It is also well established that lncRNAs are involved in the development of the mammary gland and diseases like breast cancer [20–22]. In bovine, the lncRNA repertoire in several tissues including the mammary gland [23–28], milk exosomes [29], skeletal muscle/adipose tissues [30, 31] gastrointestinal tract tissues [32, 33] and sperm/testis [34, 35] has been characterized. Increasing evidence supports the notion that lncRNAs are associated with developmental, metabolic and immunological regulation, as well as adaptations and phenotypic variation of complex traits in domestic animals [26, 36, 37]. However, the role of lncRNA in the bovine milk production and mammary gland processes is less clear.

The overall objective of this study was to systematically identify the profiles of differentially expressed lncRNAs during different stages of lactation in cattle through high-throughput RNA sequencing. We hoped that by studying the relationship between differentially expressed (DE) lncRNAs and milk quality and yield related genes

(mRNAs), our study would shed light on the complex mechanisms underlying the milk production.

## Materials and methods

### Experimental animals and sampling

The experimental design was the same to that of our earlier study [38], RNA-Sequencing data were download from GSE107366. Briefly, the data from three healthy cows of each Kashmiri and Jersey breed in their 2nd/3rd parity/lactation were utilized. All animals were of same age and milk yield was also same. Animals were maintained at the University dairy farm, Mountain Livestock Research Institute, Share-Kashmir University of Agricultural Sciences and Technology-Kashmir, India. The animals were stall fed individually, offered complete feeds based on oats hay (30 parts), sorghum (30 parts) and concentrate mixture (40 parts) to meet their nutrient requirements as per ICAR, India (2013). The concentrate mixture consisted of maize (20%), wheat bran (25%), deoiled rice bran (15%), mustard oilcake (15%), cotton seed cake (16%), molasses (5%), mineral mixture (2%), salt (1%) and urea (1%). Fresh milk samples (1.5L/cow) were collected in the morning aseptically by hand milking the four quarters of the cows at 15 (D15), 90 (D90) and 250 (D250) days in milk, representing early, mid and late lactation stages, respectively. The milk production was recorded in kgs. and values are significantly comparable between the animals [38].

### Isolation of milk epithelial cells

Milk epithelial cells (MECs) were isolated from freshly collected milk following the protocol [4] with some modifications as reported in our previous study [38]. Briefly, milk sample (1.5L) was aliquoted into 250 ml centrifuge tubes, and 100 ml of 4°C diethylpyrocarbonate (DEPC) treated phosphate buffered saline (PBS) added. Samples were centrifuged at 2800 x g at 4°C for 2 min and the fat and skimmed milk were removed. The pellet and the remaining supernatant fraction (1 ml) were mixed with 800  $\mu$ l of DEPC–PBS and transferred into a 2 ml tube. After adding 200  $\mu$ l EDTA (0.5 M pH 8.0, 4°C), the samples were centrifuged at 14,000 g for 1 min at 4°C. The supernatant was discarded, and the pellets resuspended in 200  $\mu$ l cold (4°C) DEPC–PBS. Solutions containing pellets from the same cow were pulled together and the suspension centrifuged at 5100 x g for 5 min at 4°C. Thereafter, the supernatant was discarded, and the pellet resuspended in 1.25 ml cold (4°C) PBS containing 1% bovine serum albumin (Sigma, USA). For the separation of MEC from other cell types, MEC specific anti-cytokeratin peptide 18 antibodies (Clone KS-B17.2, Sigma–Aldrich, USA) coated beads (Dynabeads Pan Mouse IgG, Invitrogen) were used. The Purified MECs were checked

for possible contamination with other cell types by quantification of the expression of marker genes for various milk somatic cell types as described by Bhat et al. [38].

#### RNA extraction and library preparation

RNA extraction was accomplished with Trizol Reagent (Ambion, USA) according to the manufacturer's instructions. Absorbance (A) of RNA samples was measured at 260 and 280 nm using the spectrophotometer (Thermo Scientific, USA). The A260 was used to estimate the RNA concentration. Moreover, the quality and integrity were assessed with the Agilent 2100 Bioanalyzer (Agilent, USA). The RNA integrity number (RIN) value of the samples used for library preparation was  $\geq 8$ . IlluminaTruSeq Stranded mRNA Sample Prep kit (Illumina, USA) was used to generate cDNA libraries from 4  $\mu$ g total RNA according to the manufacturer's recommendations. Poly-T oligo-attached magnetic beads were used to purify poly-A containing mRNA and lncRNAs molecules followed by fragmentation into small pieces using divalent cations under elevated temperature. First strand cDNA was synthesized using reverse transcriptase and random primers followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. After adenylation of 3' ends of DNA fragments, hybridization was initiated by ligating Illumina PE adapter and index. cDNA fragments (200 bp) were generated and were selectively enriched to construct the final sequencing library using Illumina PCR Primer Cocktail. The sequencing was performed at SciGenome Lab (Cochin, India) using Illumina HiSeq 2500. The data was retrieved from our submitted NCBI SRA database (accession ID. SRR6324372/GSE107366).

#### Sequence data quality control, mapping and identification of lncRNAs

The raw reads were subjected to quality control using the FASTQC program v0.11.9 [39] and cleaned using cutadapt tool version 3.2 [40] and sickle tool [41]. All clean reads were mapped to the *Bos taurus* reference genome assembly ARS-UCD1.2 using HISAT2 v2.2.1 [42]. The final transcript sets were compared with known genes annotated by Ensemble Release 94. Known protein-coding transcripts and small noncoding RNA transcripts (e.g. sRNAs, tRNA, rRNA, etc.) were removed while annotated lncRNA transcripts were retained using homologous sequence similarity search with BLAST program [43, 44]. Next, the transcripts with length of < 200 bp were removed. The remaining transcripts that did not overlap with any known annotation, localized in intronic, antisense or intergenic regions were assessed for their protein coding potential using four independent algorithms, CNCI (coding-non-coding index) [45], PLEK (predictor of long non-coding RNAs and messenger

RNAs based on an improved k-mer scheme) [46], CPAT (coding potential assessment tool) [47], and Pfam (database providing alignments and hidden Markov models for protein domains) [48, 49].

CPAT recognizes coding and noncoding transcripts from a large pool of candidates using a logistic regression model built with four sequence features: open reading frame size, open reading frame coverage, Fickett TESTCODE statistics and hexamer usage bias. CPAT coding probability score ranges from 0 to 1, and the optimum cut-off for protein coding probability varies depending on the species being analyzed. In order to extract potential noncoding transcripts with a high reliability from our dataset, we selected a stringent threshold for the CPAT probability with a score < 0.02 as ncRNA [12]. The transcripts with a score below the selected thresholds were classified to possess an ambiguous coding potential. Furthermore, Pfam Scan (v1.3) was employed to identify occurrences of any known protein family domain documented in the Pfam database. CNCI software is a signature tool that effectively distinguishes between protein-coding and non-coding sequences based on their intrinsic sequence composition i.e., by profiling adjoining nucleotide triplets. The alignment-free tool, PLEK, uses a computational pipeline based on an improved k-mer scheme and a support vector machine algorithm to distinguish lncRNAs from messenger RNAs (mRNAs). This tool has > 90% accuracy when compared with other currently available tools [50].

#### Differential expression analysis

The expression levels of lncRNAs between any two stages of lactation, were measured as fragments per kilobase of exon per million fragments mapped (FPKM) using Cufflinks v2.1.1 package [51, 52]. Differentially expressed lncRNA was screened based on FDR corrected *p*-value < 0.05 and absolute  $\log_2(\text{fold change}) > 1$ .

#### Expression correlation analysis between lncRNAs and genes related to milk quality and yield traits

To identify correlation between DE lncRNA and mRNAs, Pearson correlation test was performed to calculate the co-expression coefficient between lncRNA expression data and mRNAs [53]. Two-way analysis of variance (ANOVA) was used to evaluate the statistical significance of comparisons within the lactation stages using R program v4.0.0. *P*-values (two-sided) were adjusted for FDR due to multiple testing correction [54] and FDR < 0.05 was defined as statistically significant. A lncRNA-mRNA pair was considered co-expressed if it had a significant correlation value at FDR < 0.05.

### Target gene prediction and functional analysis

For the identification of lncRNA trans-target genes, Spearman correlation was calculated between DE-lncRNAs and DE-mRNA using COR.TEST function in R [53]. Then interactions of DE-lncRNAs and DE-mRNAs with  $r$  values  $\geq 0.9$  and  $P < 0.05$  were selected as trans-target genes. The *cis* role of lncRNAs was expounded as those influencing neighboring target genes [55]. For each DE lncRNA, the nearest upstream and downstream (within 100kb) protein-coding neighbors were identified as their potential *cis*-regulatory targets. The *trans* role alludes to the impact of lncRNA on mRNA at the expression level. The expressed correlations between lncRNAs and coding genes were calculated using the Pearson method with  $p$ -value  $< 0.05$ . To understand the biological implication of the target genes we performed GO and KEGG enrichment analysis utilizing KOBAS server version 3 [56, 57]. The GO terms were categorized into biological processes (BP), cellular components (CC), and molecular functions (MF). GO terms and pathways with a  $p$ -value less than 0.05 were considered as significantly enriched.

### Quantitative real time PCR (qRT-PCR) validation of lncRNAs

To validate the repeatability and reproducibility of the RNA-Seq data, quantitative real-time PCR was performed on 10 randomly selected lncRNAs (including 5 up- and 5 down-regulated) using the same total RNA that was used to perform the RNA sequencing. Primers for qRT-PCR were designed using Primer 5.0 software (<http://www.premierbiosoft.com/primerdesign/>) and their specificity and complementarity were assessed using NCBI BLAST algorithm (Supplementary File 1). The complementary DNA (cDNA) was synthesized from 0.5  $\mu$ g of total RNA with the Revert Aid First Strand cDNA Synthesis Kit (Thermo Scientific, USA). Real-time PCR reaction mix was composed of 10  $\mu$ L SYBR Green PCR Master Mix (Roche, Germany), 0.5  $\mu$ L cDNA, 0.3  $\mu$ L (10 mM) forward and reverse primers and 9.2  $\mu$ L nuclease free water to a final volume of 20  $\mu$ L. All aliquots were then amplified by 40 cycles of denaturation at 95 °C for 5 min, annealing at 60 °C for 15 s and extension at 72 °C for 15 s. The qRT-PCRs were carried out in triplicates. The expression levels of the selected lncRNAs were normalized against two housekeeping genes, *GAPDH* and *UXT*. The relative quantification of the potential lncRNAs was determined using the  $2^{-\Delta\Delta C_t}$  method [58].

## Results

### Sequencing results and quality control

Transcriptome sequencing of 18 cDNA libraries (three at each stage of lactation per breed) generated a total of 305, 321, and 241 million raw reads for Jersey and 248, 237 and, 302 million raw reads for Kashmiri cattle at early

(D15), mid (D90) and late (D250) lactation stages, respectively. Out of these, 302 million (D15); 315 million (D90) and 217 million (D250) reads for Jersey and 235 million (D15); 235 million (D90) and 292 million (D250) reads for Kashmiri cattle passed quality control. On average, 95.8% raw reads from each library passed the quality check. Alignment of reads to the bovine reference genome ARS-UCD1.2 yielded a mean of 90.1% (Kashmiri) and 90.9% (Jersey) unique alignments to the reference genome while 9.9% (Kashmiri) and 9.1% (Jersey) reads aligned to multiple positions or did not align at all and were discarded (Table 1).

### Identification and characterization of lncRNAs in milk derived bovine mammary epithelial

To predict lncRNA with greater certainty, four independent algorithms, CPAT, CNCI, PFAM and PLEK, were used. A total of 549 unique lncRNAs were predicted by the four independent algorithms (Supplementary File 2). The transcript length of identified lncRNAs ranged from 206 to 15,071 nucleotides, with ~44% of lncRNAs shorter than 1000 nucleotides (Fig. 1a, Supplementary File 2). Exon number of lncRNAs ranged from 2 to 43. Most lncRNAs had two exons (57.1%), followed by three exons (7.6%) and about 35.3% of lncRNAs had more than four exons (Fig. 1b). Characterization according to genomic location indicated that most lncRNAs are located in intergenic regions (459 lncRNAs), whereas 49, 13 and 8 lncRNAs are intronic, antisense or sense-overlapping lncRNAs, respectively (Fig. 1c).

To determine whether our set of identified lncRNAs possesses similar characteristics as mRNA, we compared the lncRNA expression data with mRNA expression data on the same samples. Results showed that, the lncRNAs were generally less expressed as compared to protein-coding mRNAs (Fig. 1d).

### DE lncRNAs between lactation stages in Kashmiri and Jersey cattle

A total of 10, 6 and 11 lncRNAs were DE (FDR < 0.05) between D15 vs D90, D90 vs D250, and D15 vs D250, respectively in Kashmiri cattle (Fig. 2, Supplementary File 3). Likewise, 7, 16 and 25 lncRNAs were DE (FDR < 0.05) between D15 vs D90, D90 vs D250, and D15 vs D250, respectively in Jersey cattle (Fig. 2, Supplementary File 3). Furthermore, there were 5 common lncRNAs [*XLOC\_013668* (D15 = 466.515; D90 = 93.3322; D250 = 32.1239), *XLOC\_013477* (62.1684; 306.217; 131.523), *XLOC\_021768* (1.8902; 7.13654; 8.20963), *XLOC\_000865* (1.19188; 2.83428; 2.91069), *XLOC\_003635* (0.366344; 0.883615; 0.916974)] and 12 [*XLOC\_015754* (9.73475; 1.73083; 5.26697), *XLOC\_009502* (0.415893; 0.0868364;

**Table 1** Read alignment summary

Sample No.	Total reads	QC passed reads	QC Passed %age	Aligned reads	Aligned Read %age	Unaligned Reads	Unaligned Reads %age
JERSEY-390-1	96,249,540	94,744,250	98.44%	85,869,238	90.63%	8,875,012	9.37%
JERSEY – 390-2	92,517,372	91,969,506	99.41%	84,113,665	91.46%	7,855,841	8.54%
JERSEY – 390-3	116,401,530	115,758,078	99.45%	106,605,825	92.09%	9,152,253	7.91%
JERSEY – 447-1	91,214,046	89,570,316	98.20%	79,097,456	88.31%	10,472,860	11.69%
JERSEY – 477-2	93,235,128	89,955,200	96.48%	81,828,632	90.97%	8,126,568	9.03%
JERSEY – 477-3	136,828,434	135,768,664	99.23%	125,418,364	92.38%	10,350,300	7.62%
JERSEY – 90-1	90,605,106	76,773,936	84.73%	69,093,481	90.00%	7,680,455	10.00%
JERSEY 90-2	73,121,688	69,887,565	95.57%	63,996,224	91.57%	5,891,341	9.20%
JERSEY – 90-3	77,649,922	70,905,710	91.31%	65,212,246	91.97%	5,693,464	8.03%
KASHMIRI-BL1-1	82,462,774	72,319,896	87.70%	62,308,532	86.16%	10,011,364	13.84%
KASHMIRI -BL1-2	73,887,336	73,339,212	99.26%	67,574,186	92.14%	5,765,026	7.86%
KASHMIRI -BL-1-3	92,098,428	89,792,054	97.49%	84,924,806	94.58%	4,867,248	5.42%
KASHMIRI -ML2-1	81,811,960	71,855,586	87.83%	59,079,820	82.22%	12,775,766	17.78%
KASHMIRI -ML2-2	68,425,836	66,722,732	97.51%	59,373,810	88.99%	7,348,922	11.01%
KASHMIRI -ML-2-3	87,720,824	86,711,472	98.85%	80,073,868	92.35%	6,637,604	7.65%
KASHMIRI -SL3-1	104,571,152	102,595,298	98.11%	90,594,073	88.30%	12,001,225	11.70%
KASHMIRI -SL3-2	76,974,734	76,313,042	99.14%	70,145,228	91.92%	6,167,814	8.08%
KASHMIRI -SL-3-3	120,888,074	113,697,196	94.05%	105,425,162	92.72%	8,272,034	7.28%

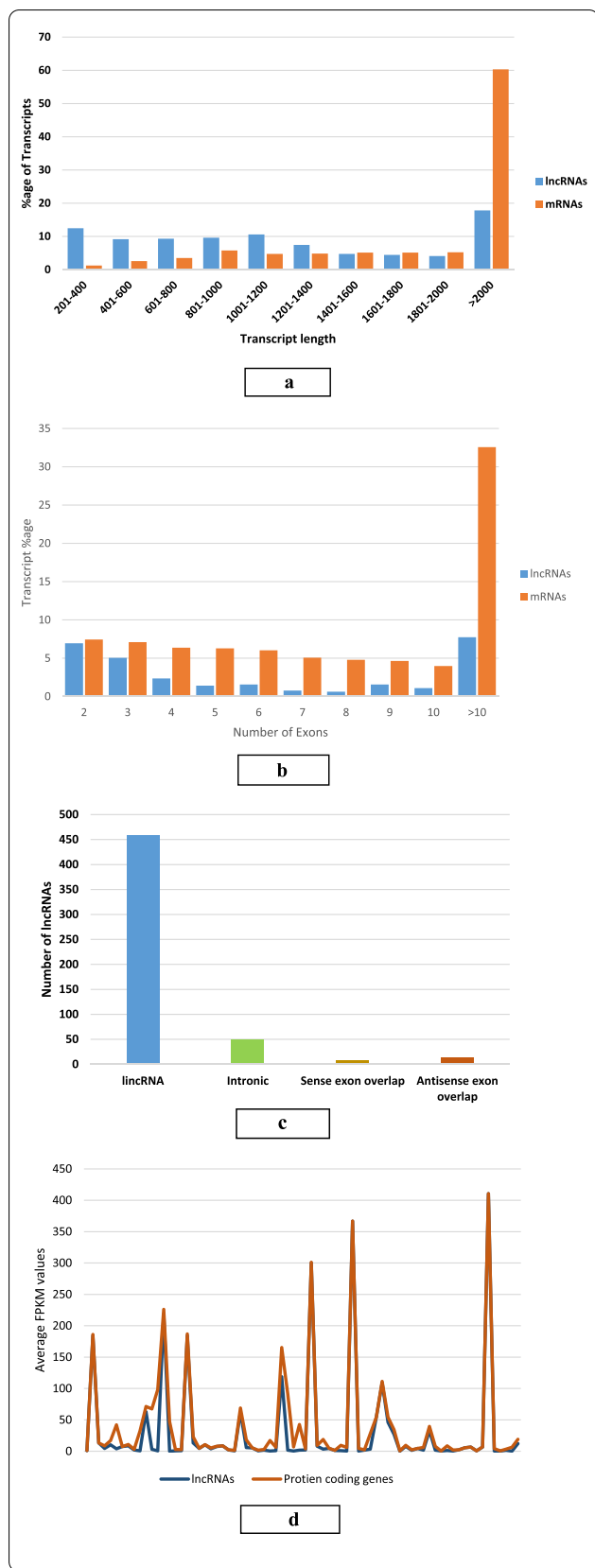
0.912991), *XLOC\_003704*(1.3756; 0.300084; 1.71051), *XLOC\_026943*(19.1323; 38.3812; 198.936), *XLOC\_002387*(0.707159; 0.492698; 2.71744), *XLOC\_002110* (0.0720901;0.336411; 3.08172), *XLOC\_005292*(0.409118; 0.770406; 15.3821), *XLOC\_015932*(1.12151; 1.52169; 7.63437), *XLOC\_025341*(0.0476; 0.08765; 0.51983), *XLOC\_020132*(96.1583; 28.32; 11.9587), *XLOC\_015281*(8.484; 1.857; 2.607), *XLOC\_007175*(0.04296; 1.4446; 3.41824), showing expression in all the 3 comparison groups of Kashmiri and Jersey cattle respectively.

#### Functional prediction of the roles of differentially expressed lncRNAs

To explore the functional role of DE lncRNAs, we identified their *cis* and *trans* target genes. A total of 459 (84 in Jersey and 375 in Kashmiri) potential *cis* and 8877 (6798 in Jersey and 2079 in Kashmiri) *trans* target genes were identified for 48 (in Jersey) and 27 (in Kashmiri) DE lncRNAs. To obtain an insight into the plausible associated functions of lncRNAs at different stages of lactation, the potential target genes of DE lncRNAs in three comparisons (D15 vs. D90, D90 vs. D250, and D15 vs. D250) were subjected to GO and pathway enrichment analyses. In Kashmiri cattle, target mRNAs of DE lncRNAs were enriched in 499 GO terms. Among them, 252 were categorized as *biological processes* (BP), 63 *molecular*

*functions* (MF), and 184 *cellular component* (CC) terms. Whereas in Jersey cattle, 507 enriched GO terms were found. Among them, 245 were BP, 64 were MF and 198 were CC. The top three BP GO terms for Kashmiri cattle for each pair of comparison groups are *regulation of lipid transport*, *metabolic process*, and *cholesterol homeostasis* (D15 vs D90); *cellular macromolecule metabolic process*, *lipid homeostasis* and *intracellular protein trans-membrane transport* (D90 vs D250); and *regulation of cellular metabolic process*, *secretion by cell* and *response to stress* (D15 vs D250) (Supplementary Files 4 and 5). Whereas in Jersey cattle top enriched BP terms are *response to lipid*, *DNA biosynthetic process*, and *fatty acid homeostasis* between D15 vs D90. *Regulation of macromolecule biosynthetic process*, *vesicle targeting*, and *RNA processing* are enriched terms in D90 vs D250. *Golgi vesicle transport*, *regulation of cholesterol transport* and *regulation of immune system process* are found enriched between D15 vs D250. Interestingly, several BP terms related to lactation, including *regulation of gastrulation*, *fatty acid homeostasis*, *regulation of cholesterol transport* and *golgi vesicle transport* were identified.

Results of pathways analysis indicated that 10, 11, and 19 pathways were significantly enriched in Jersey cattle for D15 vs D90, D15 vs D250 and D90 vs D250 comparisons, respectively (Supplementary Files 4 and 5). Similarly, 10, 6 and 10 pathways were significantly enriched in Kashmiri cattle for D15 vs D90, D15 vs D250 and D90 vs

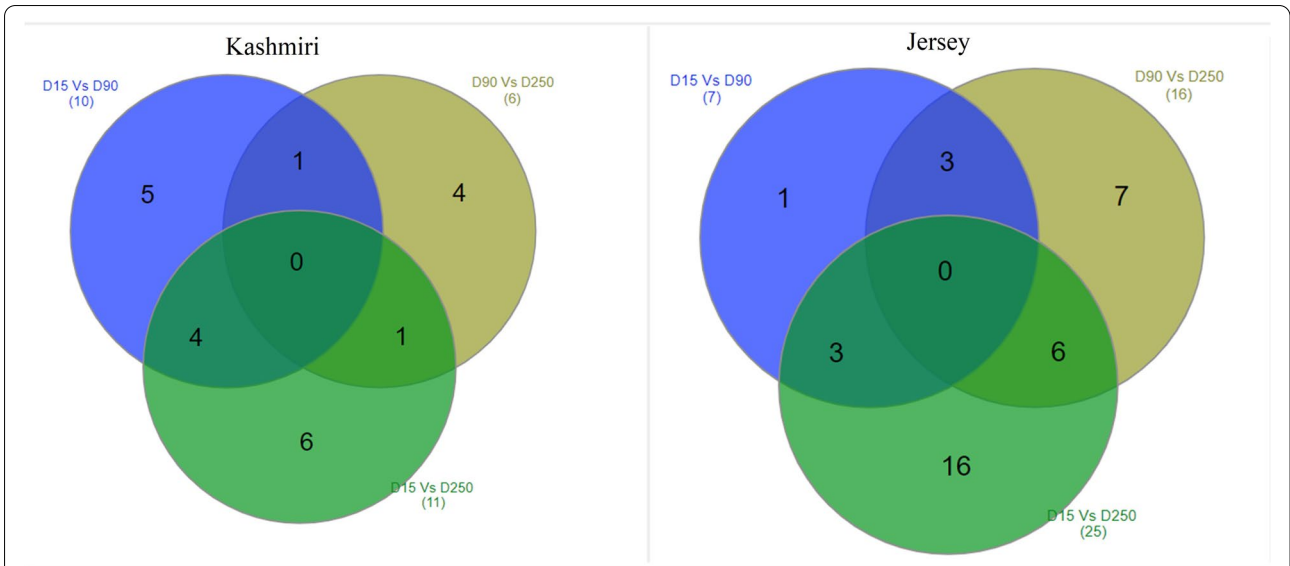


**Fig. 1** Features of lncRNAs compared with protein coding genes (mRNAs). **A** Length distribution of lncRNAs compared with mRNAs. **B** Exon number distribution of lncRNA transcripts compared with mRNAs. **C** Classification of identified lncRNAs on the basis of genomic location and orientation. **D** Comparison of average expression values of lncRNAs and mRNAs

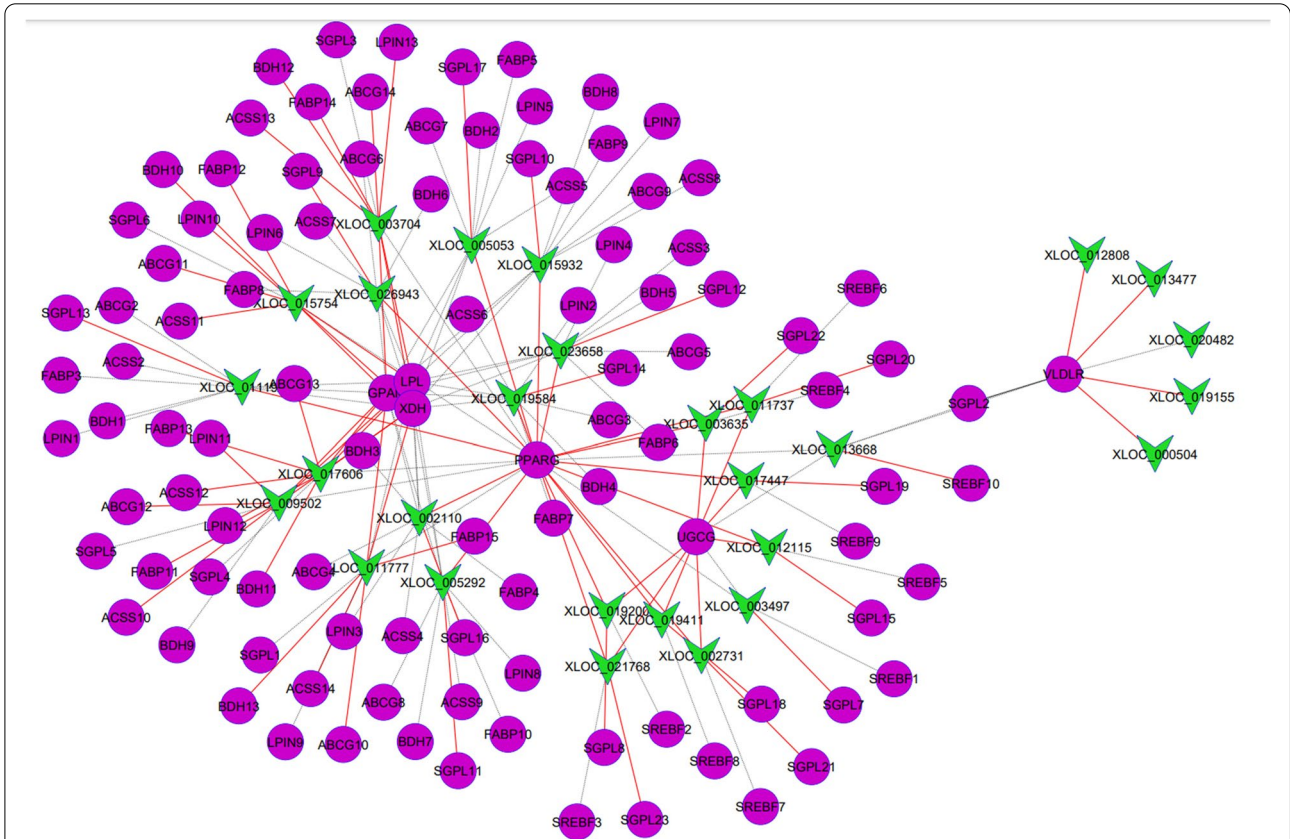
D250 comparisons, respectively. *Homologous recombination, sphingolipid metabolism and DNA replication; metabolic pathways, carbon metabolism and biosynthesis of amino acids; DNA replication, mismatch repair and TLR signaling pathway* were the top pathways for D15 vs D90, D15 vs D250 and D90 vs D250 comparison groups in Jersey cattle. While *mTOR signaling pathway, WNT signaling pathway and PI3K-Akt signaling pathway; collecting duct acid secretion, fanconi anemia pathway, N-Glycan biosynthesis and adherens junction* were the top enriched pathways for D15 vs D90, D15 vs D250 and D90 vs D250 comparison groups in Kashmiri cattle.

**Expression correlation network of lncRNAs with candidate genes related to milk quality and yield traits in Jersey and Kashmiri cattle**

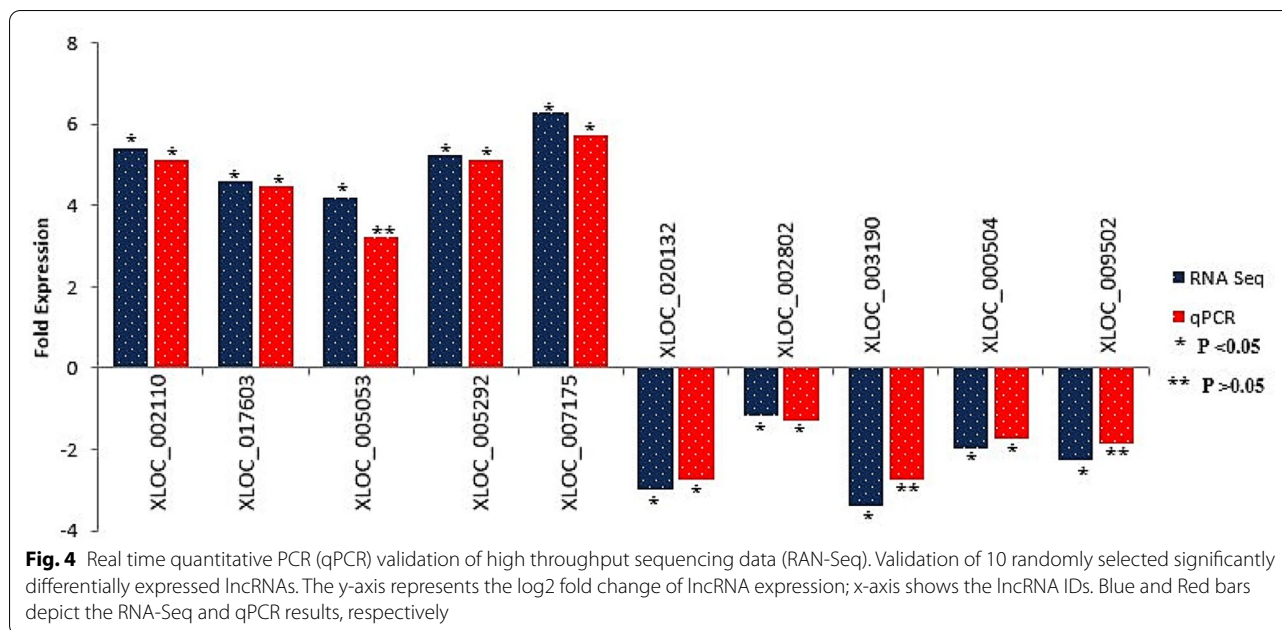
To further explore the potential regulatory mechanism of the DE lncRNAs, their relationship with protein-coding genes with roles in milk quality and yield traits was explored through correlation analysis. The results of lncRNA and coding genes correlation analysis are shown in Supplementary File 6. 3768 (Jersey) and 4048 (Kashmiri) significant correlations (FDR<0.05) were found between lncRNAs and their potential target mRNAs. Furthermore, 360 significant correlations were found between DE lncRNAs and 46 candidate genes for milk quality and yield traits (179 and 189 in Jersey and Kashmiri cattle, respectively) (Fig. 3). Among the 179 correlations in Jersey, 104 were positive and 65 were negative correlations while of the 189 correlations found in Kashmiri cattle, 117 were negative and 72 were positive correlations. Interestingly, we found that 13 lncRNAs correlated with *LALBA* in Jersey cattle while no lncRNA correlated with this gene in Kashmiri. In addition, *XLOC\_011777* correlated positively with *GPAM* and *ABCG2* genes in Jersey while the correlation of *XLOC\_011777* was negative with both genes in Kashmiri cattle. Interestingly, the 5 commonly expressing lncRNAs in Kashmiri cattle were showing correlation with *UGCG, VLDLR, SREBF1, PPARG, SGPL1* where as the 12 commonly expressing lncRNAs were showing expression correlation with *GPAM, ABCG2, SCAP, FASN, SPTLC1, UGCG, BDH1, LALBA, SLC2A8, LPL, CSNK1, NOS2, and MFGE8*. All these genes are promising candidate genes and central for milk protein and fat synthesis regulation hence the yield.



**Fig. 2** Significantly differentially expressed lncRNAs between D15 vs D90, D15 vs D250 and D90 vs D250 in Kashmiri and Jersey cattle



**Fig. 3** lncRNAs and milk quality and yield related genes co-expression network. The green triangles represent differentially expressed lncRNAs while pink spheres represent candidate genes for milk quality and yield traits. Red lines represent positive correlations while black dotted lines represent negative correlations



#### Verification of lncRNA expression profiles using qRT-PCR

To confirm the RNA-Seq data, the expression of 10 randomly selected lncRNAs (5 up- and 5 down-regulated) was examined by qRT-PCR. The expression patterns of lncRNAs showed a similar trend between the methods of RNA-Seq and qRT-PCR (Fig. 4). Pearson correlation coefficient between RNA-Seq data and qRT-PCR data was 0.97, indicating that the RNA-Seq data was highly correlated with that of the qRT-PCR.

#### Discussion

Progressively, studies are unraveling the roles of non-coding RNA molecules in biological processes, but their roles are still classical in the area. The mammary epithelial cells are the primary cells involved in synthesis and secretion of milk and proteins in it. Lactation process is influenced by various molecule, including lncRNAs. Until now, very little is known about the bovine lncRNA transcriptome in mammary epithelial cells. In this study, we conducted a preliminary investigation of lncRNA expression profiles in mammary epithelial cells of Jersey and Kashmiri cows to assess the potential of lncRNAs as regulators of milk quality and yield during the three stages of lactation curve. Consequently, the present work provides an important resource of lncRNA repertoire in mammary epithelial cells for future studies.

The identification and characterization of lncRNAs, particularly in the mammary gland, is limited compared with lncRNAs in humans and other model organisms [59–61]. In bovine mammary gland, the main focus has

been on genes and miRNAs rather than on lncRNAs [62–64]. Identification of breed-specific, milk related lncRNAs along with their expression pattern during the different stages of lactation could provide an insight into their contribution to the inter-breed variation in milk yield and could be targeted to increase the milk yield and consequently improve the profitability of the farm system in future. In addition, knowledge of these special genetic elements between breeds shall be useful to develop accurate genomic prediction equations that can operate effectively across breeds. In the present study, we identified 549 putative lncRNAs with high confidence across three lactation stages in two bovine breeds with differing milk production ability. We classified predicted lncRNAs into categories based on their genomic location to understand their functional spectrum as the relative position between a lncRNA and its neighboring protein-coding genes is a key determinant of their regulatory relationship [65]. To our knowledge, this is the first report to systematically identify lncRNAs from RNA-seq data during different stages of the bovine lactation. The identified lncRNAs have fewer exons, shorter transcript lengths, and lower expression levels in comparison with known protein-coding transcripts, which are in agreement with earlier report [66].

In total 27 and 48 lncRNAs were differentially expressed in pairwise comparisons in Kashmiri and Jersey, respectively. These lncRNAs may have specific biological roles in bovine mammary gland during the lactation cycle. As it has been demonstrated that lncRNAs are key players in tissue physiology and organogenesis



[27, 67–70]. We also identified 5 and 12 DE-lncRNAs expressing in all the 3 lactation stages of Kashmiri and Jersey cattle respectively. These DE-lncRNAs were correlated with a network of genes essential for coordinating milk synthesis and secretion. Compared with Kashmiri cattle, the mammary tissue transcriptome of Jersey cattle had a completely different rank of expressing lncRNAs in terms of abundance. Recently, a study on the bovine mammary gland identified 36 lincRNAs (long intergenic ncRNAs) located in milk related quantitative trait loci (QTL), suggesting their association with milk quality and production [28]. Another study characterized lncRNAs in mammary gland tissues of cows at mid lactation and identified lncRNAs with potential roles in mammary gland functions [27]. The identification of lncRNAs associated with the development of mammary gland and lactation will contribute in selecting decisions to further improve productivity and healthy breeding policies of cattle.

It is well established that an array of genes is involved, directly or indirectly with growth and development of the mammary gland as well as initiation and maintenance of the lactation cycle [71, 72]. lncRNAs can regulate the expression of their neighboring genes (known as in *cis*) as well as distant genes (known as in *trans*) [73]. In this study, bioinformatics analysis of putative lncRNAs target genes suggest roles in pathways such as *MAPK*, *PI3K-Akt*, *NF-kappa B*, *mTOR*, T cell receptor signaling, Hedgehog (SHH) signaling pathway, Glucagon signaling pathway, *AMPK* signaling pathway, Insulin signaling pathway, and Toll-like receptor signaling pathways. Key roles for these pathways in mammary gland development and lactation have been reported [74–77]. The SHH signaling is an essential pathway and is involved in mammary gland development [78]. Moreover, the Glucagon regulates the mammary gland development and lactation through activating GPCR [79]. *AMPK* plays an essential role in cellular energy sensing [80] and *mTOR* activation [81]. This signaling pathway is found involved in regulating the effect of glucose supply and utilization in the lactating mammary gland [82]. Insulin signaling pathway is an important regulator of milk synthesis and secretion in the lactating animals [83]. Also, lncRNAs like *XLOC\_011777*, *XLOC\_019584*, *XLOC\_011194*, and *XLOC\_003497*, were found to interact with some lactation-related candidate genes like *LALBA*, *LPL*, *GPAM*, *SREB1* and *LIPIN1* genes. These genes have an important role in the biosynthesis of milk fat, protein, and lactose [84, 85]. In human, cow and mouse lactation cycles, *LALBA* expression levels were found to be similar to that of other milk protein genes (146, 68 and 96% expression relative to  $\beta$ -casein) and levels of expression of both these milk protein genes increased dramatically with onset of

lactation [86]. During lactation lipoprotein lipase (LPL) is elevated in mammary tissue and depressed in adipose tissue to redirect lipids for incorporation into milk fat [87]. While activation of *SREB1*, together with *THRSP* and *ESRRA* via the concomitant decrease in progesterone concentration and increase prolactin signaling, is most likely central for milk lipid synthesis regulation in the human mammary gland [88].

Comprehensive analysis of lncRNAs and mRNAs expression profiles gives a better understanding of the biological functions of DE lncRNAs. Therefore, we performed correlation analysis of the DE lncRNAs and protein coding genes, and identified significant correlations between the transcripts in both breeds. Narrowing down the analysis to candidate genes related to milk quality and yield traits, we found that 34 lncRNAs correlated with candidate genes for milk quality and yield traits suggesting potential roles in lactation. Interestingly, most of the lncRNAs correlated positively with candidate genes related to milk quality and yield traits in Jersey compared to Kashmiri cattle where they were mostly negatively correlated, which could be one of the mechanisms responsible for the differential milking performance between the two breeds. Importantly, some of the described lncRNAs might target mRNAs, which have important roles in the mammary gland throughout the lactation cycle. For example, lncRNA *XLOC\_002110* is expressed at D90 (peak lactation) stage in Jersey only and shows correlation with *SLC2A8* gene. *SLC2A8* is a member of the transporter superfamily with predominant roles in the active transport of glucose across the plasma membrane [89]. Higher expression of *SLC2A8* has been reported during mid-lactation (D90) in Jersey cows (high yielding) [39], suggesting important roles during the peak stage of lactation. Glucose uptake by mammary epithelial cells is an important step in milk synthesis during lactation, and hence directly influences the milk yield [90]. Therefore, *XLOC\_002110* might be involved in the lactation process through regulation of the expression of *SLC2A8*. Based on these results, it is clear that target genes of the putative lncRNAs like *XLOC\_002110*, *XLOC\_019584*, *XLOC\_011194*, *XLOC\_005773* and *XLOC\_003497* could be involved in bovine mammary gland development and their roles need to be confirmed. Based on our results and data from recent studies, lncRNAs as core regulatory elements have significant roles in the physiology of the lactation cycle and in the development of the mammary gland in cattle.

## Conclusion

In this study, 549 putative lncRNA transcripts were found at three stages of lactation in Jersey and Kashmiri cattle breeds. A total of 27 and 48 lncRNAs were DE between at least one comparison pair in Kashmiri and Jersey

cattle, respectively. Target genes of DE lncRNAs were enriched in pathways (*MAPK*, *PI3K-Akt*, *NF-kappa B*, *mTOR*, T cell receptor signaling and Toll-like receptor signaling pathways) with roles in lactation and mammary gland development. Expression correlation analysis reveals that lncRNAs like *XLOC\_002110*, *XLOC\_019584*, *XLOC\_011194*, *XLOC\_005773* and *XLOC\_003497* might be important regulators of candidate genes for milk quality and yield traits in cattle. Compared to Kashmiri cattle a strong correlation between DE lncRNAs and milk candidate genes was found in Jersey, which could explain the differences in milking performance between the two breeds. This study mapped the expression profiles of lncRNAs across lactation stages and their relationships with candidate genes related to milk quality and yield traits in Jersey and Kashmiri cattle. This study therefore provides a valuable resource for the study of lncRNA roles in lactation biology. However, better understanding of the molecular mechanisms involving lncRNA functions in milk synthesis of the animals is of prime importance.

#### Abbreviations

lncRNAs: Long non-coding Ribonucleic acids; MEC: Mammary epithelial cells; DE: Differentially expressed; GPAM: Glycerol-3-phosphate acyltransferase 1, mitochondrial; LPL: Lipoprotein lipase; ABCG2: ATP-binding cassette superfamily G member 2; RNA: Ribonucleic acid; DNA: Deoxyribonucleic acid; CNCI: Coding-non-coding index; PLEK: Predictor of long non-coding RNAs and messenger RNAs based on an improved k-mer scheme; CPAT: Coding potential assessment tool; rRNA: Ribosomal Ribonucleic acid; pfam: Database providing alignments and hidden Markov models for protein domains; mRNA: Messenger Ribonucleic acid; FPKM: Fragments per kilobase of exon per million fragments mapped; CSN2: Casein 2; cDNA: Complementary DNA; DNase: Deoxyribonuclease; dNTP: Deoxyribonucleotide triphosphate; ds: Double-stranded; EDTA: Ethylenediaminetetraacetic acid; Fig: Figure; DE: Differentially expressed; KEGG: Kyoto Encyclopedia of Genes and Genomes; lncRNA: Long non-coding RNA; M: Million; MAPK: Mitogen-activated protein kinase; mRNA: Messenger RNA; miRNA: MicroRNA; NCBI: National Center for Biotechnology Information; NFkB: Nuclear factor-kappa B; RNA-seq: RNA sequencing.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08406-x>.

**Additional file 1.**

**Additional file 2.**

**Additional file 3.**

**Additional file 4.**

**Additional file 5.**

**Additional file 6.**

#### Acknowledgments

The authors would like to thank the staff of the University Dairy Farm of Mountain Livestock Research Institute-Manasbal, Srinagar-Kashmir for care and management of the experimental animals.

#### Authors' contributions

SMA: conception of study design, supervised the experiments and data analysis, and provided input on data interpretation. PTM: conducted the

experiment, and drafted the article, BB: performed data analyses and wrote manuscript, EMI: provided intellectual content, proofreading and data interpretation. DV, NS, RAS and NAG: provided intellectual content. ZH, QT, MAD, SAB: provided input on data interpretation. All authors revised and approved the final version of the manuscript.

#### Funding

The Department of Biotechnology, Ministry of Science and Technology, Government of India supported the study under grant BT/PR114 08/AAQ/1/591/2014. We would like to acknowledge the Computational Biology team from NAHEP SKUAST-Kashmir for their support and end-to-end data analysis.

#### Availability of data and materials

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ncbi.nlm.nih.gov/sra/SRR6324372>

#### Declarations

##### Ethics approval and consent to participate

Animal care and use procedures were approved by the Institutional Animal Ethics Committee of Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir (SKUAST-K). All animal procedures were conducted in strict accordance with the rules and guidelines outlined by the SKUAST-K Animal Welfare Committee.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

##### Author details

<sup>1</sup>Division of Animal Biotechnology, Faculty of Veterinary Sciences and Animal Husbandry, Sher-e-Kashmir University of Agricultural Sciences and Technology – Kashmir, SKUAST-K, Shuhama, Jammu 190006, India. <sup>2</sup>Department of Biochemistry, School of Life Sciences Jaipur National University, Jaipur, India. <sup>3</sup>Division of Animal Breeding and Genetics, Faculty of Veterinary Sciences and Animal Husbandry, SKUAST-K, Shuhama, Jammu, India. <sup>4</sup>Agriculture and Agri-Food Canada, Sherbrooke Research and Development Centre, Sherbrooke, Quebec, Canada. <sup>5</sup>AgriGenom laboratory Pvt. Limited, Cochin, India. <sup>6</sup>Division of Livestock Production and Management, SKUAST-K, Srinagar, India.

Received: 9 March 2021 Accepted: 15 February 2022

Published online: 04 March 2022

#### References

1. Eva M, Strucken YC, Laurenson SM, Gudrun BA. Go with the flow-biology and genetics of the lactation cycle. *Front Genet.* 2015;118:6.
2. Reece RP. The physiology of Milk production. *J Dairy Sci.* 1956;39(6):726–34.
3. Ngoc duy D, Ibeagha-Awemu EM. Non-coding RNA roles in ruminant mammary gland development and lactation. In book: *Current Topics in Lactation* (ed). 2017. <https://doi.org/10.5772/67194>.
4. Boutinaud M, Ben CM, H., Delamaire E., Guinard-Flament J. Milking and feed, restriction regulate transcripts of mammary epithelial cells purified from milk. *J Dairy Sci.* 2008;91:988–98.
5. Capuco AV, Wood DL, Baldwin R, Mcleod K, Paape MJ. Mammary cell number, proliferation, and apoptosis during a bovine lactation: relation to milk production and effect of bST. *J Dairy Sci.* 2001;84:2177–87.
6. Do DN, Li R, Dudemaine PL, Ibeagha-Awemu EM. MicroRNA roles in signaling during lactation: an insight from differential expression, time course and pathway analyses of deep sequence data. *Sci Rep.* 2017;7:1–19.
7. Peng J, Zhao JS, Shen YF, Mao HG, Xu NY. MicroRNA expression profiling of lactating mammary gland in divergent phenotype swine breeds. *Int J Mol Sci.* 2015;16:1448–65.
8. Zhang C, et al. Deep RNA sequencing reveals that microRNAs play a key role in lactation in rats. *J Nutr.* 2014;144:1142–9.

9. Castillo J, Stueve TR, Marconett CN. Intersecting transcriptomic profiling technologies and long non-coding RNA function in lung adenocarcinoma: discovery, mechanisms, and therapeutic applications. *Oncotarget*. 2017;8:81538.
10. Guttman M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009;458:223–7.
11. Kapranov P, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*. 2007;316:1484–8.
12. Derrien T, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22:1775–89.
13. Pang KC, Frith MC, Mattick JS. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet*. 2006;22:1–5.
14. Bhat S, et al. Long non-coding RNAs: mechanism of action and functional utility. *Non-coding RNA Res*. 2016;43:50:1.
15. Ballarino M, et al. Novel long noncoding RNAs (lncRNAs) in myogenesis: a miR-31 overlapping lncRNA transcript controls myoblast differentiation. *Mol Cell Biol*. 2015;35:728–36.
16. Li P, et al. A liver-enriched long non-coding RNA, lncLSTR, regulates systemic lipid metabolism in mice. *Cell Metab*. 2015;21:455–67.
17. Canovas A, Rincon G, Islas-Trejo A, Wickramasinghe S, Medrano JF. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mamm Genome*. 2010;21:592–8.
18. Dhanoa JK, Sethi RS, Verma R, Arora JS, Mukhopadhyay CS. Long non-coding RNA: its evolutionary relics and biological implications in mammals: a review. *J Anim Sci Technol*. 2018;60:25.
19. Hansji H, Leung EY, Baguley BC, Finlay GJ, Askarian-Amiri ME. Keeping abreast with long non-coding RNAs in mammary gland development and breast cancer. *Front Genet*. 2014;5:379.
20. Huang J, et al. Long non-coding RNA UCA1 promotes breast tumor growth by suppression of p27 (Kip1). *Cell Death Dis*. 2015;5:1008.
21. Wang H, et al. A novel long non-coding RNA regulates the immune response in MAC T cells and contributes to bovine mastitis. *FEBS J*. 2019;286:1780–95.
22. Zhang Y, et al. Long intergenic non-coding RNA expression signature in human breast cancer. *Sci Rep*. 2016;6:37821.
23. Zheng X, et al. Integrated analysis of long noncoding RNA and mRNA expression profiles reveals the potential role of long noncoding RNA in different bovine lactation stages. *J Dairy Sci*. 2018;101:1–13.
24. Yang B, et al. Transcriptome sequencing to detect the potential role of long non-coding RNAs in bovine mammary gland during the dry and lactation period. *BMC Genomics*. 2018;19:605.
25. Ibeagha-Awemu EM, Li R, Dudemaine PL, Do DN, Bissonnette N. Transcriptome analysis of long non-coding RNA in the bovine mammary gland following dietary supplementation with linseed oil and safflower oil. *Int J Mol Sci*. 2018;19:3610.
26. Tong C, et al. Identification and characterization of long intergenic non-coding RNAs in bovine mammary glands. *BMC Genomics*. 2017;18:468.
27. Koufariotis LT, Chen YPP, Chamberlain A, Vander Jagt C, Hayes BJ. A catalogue of novel bovine long noncoding RNA across 18 tissues. *PLoS One*. 2015;10.
28. Cai W, et al. Genome wide identification of novel long non-coding RNAs and their potential associations with milk proteins in Chinese Holstein cows. *Front Genet*. 2018;9:281.
29. Karlsson O, et al. Detection of long non-coding RNAs in human breast-milk extracellular vesicles: implications for early child development. *Epigenetics*. 2016;11:721–9.
30. Choi JY, Shin D, Lee HJ, Oh JD. Comparison of long noncoding RNA between muscles and adipose tissues in Hanwoo beef cattle. *Anim Cells and Syst*. 2019;23:50–8.
31. Kern C, et al. Genome-wide identification of tissue specific long non-coding RNA in three farm animal species. *BMC Genomics*. 2018;19:684.
32. Ibeagha-Awemu EM, Do DN, Dudemaine PL, Fomenky BE, Bissonnette N. Integration of lncRNA and mRNA transcriptome analyses reveals genes and pathways potentially involved in calf intestinal growth and development during the early weeks of life. *Genes (Basel)*. 2018;9:142.
33. Weikard R, et al. Long noncoding RNAs are associated with metabolic and cellular processes in the jejunum mucosa of pre-weaning calves in response to different diets. *Oncotarget*. 2018;9:21052–69.
34. Wang X, et al. Integrated analysis of mRNAs and long noncoding RNAs in the semen from Holstein bulls with high and low sperm motility. *Sci Rep*. 2019;9:2092.
35. Gao Y, et al. Analysis of long non-coding RNA and mRNA expression profiling in immature and mature bovine (*Bostaurus*) testes. *Front Genet*. 2019;10:646.
36. Qu Z, Adelson DL. Bovine ncRNAs are abundant, primarily intergenic, conserved and associated with regulatory genes. *PLoS One*. 2012;7:e42638. 10.1371.
37. Weikard R, Hadlich F, Kuehn C. Identification of novel transcripts and noncoding RNAs in bovine skin by deep next generation sequencing. *BMC Genomics*. 2013;14:789.
38. Bhat SA, Ahmad SM, Ibeagha-Awemu EM, Bhat BA, Dar MA, Mumtaz PT, et al. Comparative transcriptome analysis of mammary epithelial cells at different stages of lactation reveals wide differences in gene expression and pathways regulating milk synthesis between Jersey and Kashmiri cattle. *PLoS One*. 2019;14:2.
39. Andrews S. FastQC. Qual Control Tool High Throughput Seq Data. 2010;370.
40. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17.
41. Joshi NA, Fass JN. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). Software. 2011.
42. Kim D, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907–15.
43. Johnson M, et al. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008;36.
44. Bhat B, et al. TM-aligner: multiple sequence alignment tool for transmembrane proteins with reduced time and improved accuracy. *Sci Rep*. 2017;7.
45. Sun L, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. 2013;41.
46. Li A, Zhang J, Zhou Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC bioinformatics*. 2014;15.
47. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res*. 2013;41.
48. Finn RD, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2013;42.
49. Bhat B, et al. Comparative transcriptome analysis reveals the genetic basis of coat color variation in pashmina goat. *Sci Rep*. 2019;9(1):1–9.
50. Li L, et al. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol*. 2014;15.
51. Trapnell C, et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28:511–5.
52. Ravasi T, et al. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res*. 2006;16:11–9.
53. Kassambara A, Kassambara MA. "Package 'ggpubr'" *R package version 0.1.6*. 2020.
54. Cuevas A, Febrero M, Fraiman R. An anova test for functional data. *Computat Stat Data Anal* 47n. 2004;1:111–22.
55. Guil S, Esteller M. Cis-acting noncoding RNAs: friends and foes. *Nat Struct Mol Biol*. 2012;19:1068–75.
56. Bhat BA, et al. "Biological Networks: Tools, Methods, and Analysis." *Essentials of Bioinformatics, Volume I*. Cham: Springer; 2019. p. 255–86.
57. Xie C, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res*. 2011;39:W316–22.
58. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>ΔΔCT method. *Methods*. 2001;25:402–8.
59. Cabili MN, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25:1915–27.
60. Lv J, et al. Identification of 4438 novel lincRNAs involved in mouse pre-implantation embryonic development. *Mol Genet Genomic Med*. 2015;290:685–97.
61. Pauli A, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res*. 2012;22:577–91.

62. Scheel TK, et al. Global mapping of miRNA-target interactions in cattle (*Bostaurus*). *Sci Rep*. 2017;7:1–13.
63. Lai YC, et al. Bovine milk transcriptome analysis reveals microRNAs and RNU2 involved in mastitis. *FEBS J*. 2019.
64. Li Z, Liu H, Jin X, Lo L, Liu J. Expression profiles of microRNAs from lactating and non-lactating bovine mammary glands and identification of miRNA related to lactation. *BMC Genomics*. 2012;13:731.
65. Laurent GS, Wahlestedt C, Kapranov P. The landscape of long noncoding RNA classification. *Trends Genet*. 2015;31:239–51.
66. Long Y, Wang X, Youmans DT, Cech TR. How do lncRNAs regulate transcription?. *Sci Adv*. 2017;3.
67. Weikard R, Demasius W, Kuehn C. Mining long noncoding RNA in livestock. *Anim Genet*. 2017;48:3–18.
68. Salviano-Silva A, Lobo-Alves SC, Almeida RCD, Malheiros D, Petzl-Erler ML. Besides pathology: Long non-coding RNA in cell and tissue homeostasis. *Non-coding RNA*. 2018;4:3.
69. Zhao W, et al. Systematic identification and characterization of long intergenic non-coding RNAs in fetal porcine skeletal muscle development. *Sci Rep*. 2015;5:8957.
70. Tang Z, et al. Comprehensive analysis of long non-coding RNAs highlights their spatio-temporal expression patterns and evolutionary conservation in *Sus scrofa*. *Sci Rep*. 2017;7:43166.
71. Dai WT, Zou YX, White RR, Liu JX, Liu HY. Transcriptomic profiles of the bovine mammary gland during lactation and the dry period. *Funct Integr Genomics*. 2018;18:125–40.
72. Williams MM, et al. ErbB3 drives mammary epithelial survival and differentiation during pregnancy and lactation. *Breast Cancer Res*. 2017;19:105.
73. Villegas VE, Zaphiropoulos PG. Neighboring gene regulation by antisense long non-coding RNAs. *Int J Mol Sci*. 2015;16:3251–66.
74. Schmidt JW, et al. Stat5 regulates the PI3-kinase/Akt1 pathway during mammary gland development and tumorigenesis. *Mol Cell Biol*. 2014;34:1363–77.
75. Rädler PD, Wehde BL, Wagner KU. Crosstalk between STAT5 activation and PI3K/AKT functions in normal and transformed mammary epithelial cells. *Mol Cell Endocrinol*. 2017;451:31–9.
76. Zhang R, et al. Th-POK regulates mammary gland lactation through mTOR-SREBP pathway. *PLoS Genet*. 2018;14.
77. Fata JE, et al. The osteoclast differentiation factor osteoprotegerin-ligand is essential for mammary gland development. *Cell*. 2000;103:41–50.
78. Hui M, Cazet A, Nair R, Watkins DN, O'Toole SA, Swarbrick A. The hedgehog signalling pathway in breast development, carcinogenesis and cancer therapy. *Breast Cancer Res*. 2013;15:1–14.
79. Andreck ER, Mori S, Rempel RE, Chang JT, Nevins JR. Patterns of cell signaling pathway activation that characterize mammary development. *Development*. 2008;135:2403–13.
80. Hardie DG, Ross FA, Hawley SA. AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nat Rev Mol Cell Biol*. 2012;13:251–62.
81. Inoki K, Zhu T, Guan KL. TSC2 mediates cellular energy response to control cell growth and survival. *Cell*. 2003;115:577–90.
82. Bionaz M, Hurley W, Loor J. Milk protein synthesis in the lactating mammary gland: insights from transcriptomics analyses. *Milk protein*. 2012;11:285–324.
83. Neville MC, Webb P, Ramanathan P, Mannino MP, Pecorini C, Monks J, et al. The insulin receptor plays an important role in secretory differentiation in the mammary gland. *Am J Physiol-Endocrinol Metab*. 2013;305:1103–14.
84. Osorio JS, Lohakare J, Bionaz M. Biosynthesis of milk fat, protein, and lactose: roles of transcriptional and posttranscriptional regulation. *Physiol Genomics*. 2016;48(4):231–56.
85. Bionaz M, Loor JJ. Gene networks driving bovine milk fat synthesis during the lactation cycle. *BMC Genomics*. 2008;9:366.
86. Sharp JA, Lefèvre C, Nicholas KR. Lack of functional alpha-lactalbumin prevents involution in cape fur seals and identifies the protein as an apoptotic milk factor in mammary gland involution. *BMC Biol*. 2008;6:1–15.
87. Jensen DR, Bessesen DH, Etienne J, Eckel RH, Neville MC. Distribution and source of lipoprotein lipase in mouse mammary gland. *J Lipid Res*. 1991;32:733–42.
88. Mohammad MA, Haymond MW. Regulation of lipid synthesis genes and milk fat production in human mammary epithelial cells during secretory activation. *Am J Physiol Endocrinol Metab*. 2013;305:E700–16.
89. Olson AL, Pessin JE. Structure, function, and regulation of the mammalian facilitative glucose transporter gene family. *Annu Rev Nutr*. 1996;16:235–56.
90. Liu H, Zhao K, Liu J. Effects of glucose availability on expression of the key genes involved in synthesis of milk fat, lactose and glucose metabolism in bovine mammary epithelial cells. *PLoS One*. 2013;8.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

