Proceedings of the World Congress on Genetics Applied to Livestock Production, 11.

Transcriptional profiling of milk epithelial cells in Jersey and local Kashmiri cattle using RNA Sequencing

Shakil A. Bhat[1], Syed Mudasir Ahmad[*1], Basharat A. Bhat[2], Riaz A. Shah[1]

[1] Division of Biotechnology, Faculty of Veterinary Sciences and Animal Husbandry, SKUAST-Kashmir

[2] Shiv Nadar University, Delhi

* mudasirbio@gmail.com (corresponding author)

Summary

This study characterized the bovine milk transcriptome of Jersey and Local Kashmiri non-descriptive cattle at three different stages of lactation (early, mid and late lactation) and in order to explore the comparative milk transcriptome of Jersey and local Kashmiri cattle. This investigation was directed to reveal if certain unique traits are associated with the local germplasm. To this end, transcriptional profiling was conducted by Illumina RNA sequencing. Subsequently GO and pathway analysis was performed through R-package (2.34.0). It was found that casein and whey protein genes were highly expressed across three different lactation stages. In Local cattle largest differentially expressed genes (DEGs) were reported between early and mid-lactation with 1805 genes whereas, in Jersey largest DEGs were reported between early and late lactation with 3392 genes. Through GO analysis the DEGs were mainly associated with development process, cellular matrix, collagen, basement membrane, insulin stimulus and immune system process. Insulin stimulus and growth factor stimulus were significantly enriched during early lactation. In Local cattle the upregulated genes were mainly associated with immune response and immune system process. With KEGG pathway analysis it was found that P13K-AKT, JAK-STAT and MAPK pathways were significantly enriched between Kashmiri and Jersey cattle. The study provides some possible unique physico-chemical and immune properties in milk from local cattle and need to be further explored. This could lead to development of certain niche products and thereby help in conserving this unique germplasm which has been diluted through extensive cross breeding programmes.

Key words: Keywords: Transcriptome, Jersey, gene ontology, KEGG pathway analysis

Introduction

In dairy cattle, milk yield and milk components are two of the most important economic traits. An increase in the efficiency of milk synthesis both in terms of quality and quantity is a highly desirable goal for the dairy industry, which also is an ongoing challenge [1]. With the generation and development of molecular quantitative genetics, identifying genes underlying lactation process and incorporating them into genetic evaluation systems has been valuable for dairy cattle breeding programs. In the past several decades, quantitative trail loci (QTL) mapping, candidate gene analysis, and genome-wide association study (GWAS) have been the main strategies to identify causal genes or mutations for milk yield and composition in dairy cows [2–4]. Although these techniques have contributed significantly to our understanding of mechanisms on milk yield, component synthesis and metabolism, several major limitations still exist. The establishment of the *Bos taurus* genome assembly(UMD3.1 version 85), along with proteome and gene expression studies, have made it possible to estimate the number of genes involved in milk production, from mammogenesis to milk secretion. Between 6000 and 19,000 genes distributed across all 29 bovine autosomes and the X-chromosome have been reported to be differentially expressed during the lactation cycle [5]. Lactation is a dynamic physiological process characterized by an initial rapid increase in milk yield during early lactation, which peaks around 6 weeks into lactation, followed by a gradual decrease until the end of lactation [6]. To improve milk quality and milk production performance, major advances in understanding the physiology of lactating mammary glands, such as technologies and theories, have taken place in the past few decades [7]. A great number of genes are differentially expressed between different stages of lactation, and these expression alterations may play crucial roles in the regulation of lactation [8]. Lactation is an orchestrated and dynamic process aimed at providing nutrition and immunity to infants and beyond infancy [9] and has been widely investigated for years at the morphological and physiological levels but the genetics that underlie these dynamics are still to be uncovered. Recently developed “omics” technologies like transcriptomics make it possible to comprehensively and systematically study lactation at the RNA level, and to identify the most important factors or processes that may influence lactation. In order to achieve this goal, a thorough understanding of the components and the regulation of bovine milk composition is required.

Bovine milk contains a heterogeneous population of somatic cells consisting of lymphocytes, neutrophils, macrophages and exfoliated epithelial cells [10]. Mammary epithelial cells are unique in that they synthesize and secrete milk. Most of the genes expressed in the mammary gland transcriptome were present in milk epithelial cells (MEC). Compared with the mammary gland, higher numbers of genes related to milk quality and yield were uniquely expressed in MEC [11]. Even though many studies have been conducted on the physicochemical properties of cow milk and the genes expressed in bovine mammary gland [12], limited research has been published on the detailed characterization of genes expressed in milk epithelial cells. Therefore, identification and characterisation of milk quality and yield related genes expressed at different stages of lactation in MECs represent an

important step towards understanding the complex biology of milk production.

In the current study MECs were selected as RNA source based on previous study that has shown MECs as a representative source of mammary gland tissue [13]. Moreover, MECs provide a more accessible method compared with invasive approaches, such as mammary gland biopsies. Herein, we report for the first time a complete dataset detailing the MEC milk transcriptome from local Kashmiri cattle (a poor performing non-descript cow) and Jersey cattle at three different stages of lactation using RNA-seq. The differential gene expression (DGE) was studied at three time points/stages transition (day 15), peak (day 90) and late (day 250) lactation. A global analysis was conducted first on the bovine milk transcriptome by studying the highly expressed genes in each stage of lactation and genes with statistically significant expression between the stages. Then a detailed and integrated analysis was conducted on the expression profiles of candidate genes for milk quality and yield traits which could lead to improved selection of dairy cattle while providing new insights into complex milk biology. The study was also aimed to compare the milk transcriptome of two breeds with different genetic structure and different dairy characteristics. An investigation of local Kashmiri cattle can reveal if specific good production or milk properties are available and thereby create a new use for them. This could lead to development of niche products and thereby help in conserving them through proper breeding plans.

Materials and methods

Animals and sampling

Three healthy local Kashmiri and Jersey cows in their 2nd/3rd lactation at the University dairy farm, MLRI-Srinagar were selected for the study. The animals were kept in free stall housing, fed with total mixed ration and had access to water *ad libitum*. Fresh milk samples were collected by hand milking the four quarters of the cows at 15th, 90th and 250th day of lactation, with three biological replicates per stage of lactation with a total of 9 samples per group. The milk samples were immediately transported to laboratory in ice cool containers.

Isolation of milk epithelial cells

The milk epithelial cells (MECs) were isolated from the whole fresh milk following the protocol of Boutinaud *et al*. (2008) with some modifications. Immediately after collection of milk samples (1.5 litres of milk), aliquots of 125 ml were distributed into 250-ml centrifuge tubes, and 100 ml of 4^{o}C diethylpyrocarbonate (DEPC) treated phosphate buffered saline (PBS) buffer prepared from distilled water were added to each milk sample. The samples were defatted by 20 min centrifugation at 2800g at 4^{o}C. Afterwards, the fat layer was discarded, and the skim milk was carefully removed. The pellet and the remaining supernatant fraction (1 ml) were mixed with 800µl of 4^{o}C DEPC–PBS and transferred into a 2 ml tube. After adding 200 µl EDTA (0.5 M pH 8.0, 4^{o}C), the samples were centrifuged at 14,000g for 1 min at 4^{o}C. The supernatants were discarded, and the pellets were resuspended in 200 µl 4^{o}C DEPC–PBS. After merging the resuspended pellets, the suspension was centrifuged at 5100g for 5 min at 4^{o}C. Thereafter, the supernatant was discarded, and the

pellet was finally resuspended in 1.25 ml 4°C PBS containing 1% bovine serum albumin (BSA, Sigma). For the separation of immune cells and MEC, mammary epithelial cell specific anti-cytokeratin peptide 18 antibody (Clone KS-B17.2, Sigma–Aldrich) coated beads (Dynabeads Pan Mouse IgG, Invitrogen) were used. The separation was performed as described by Boutinaud *et al.,*(14). Briefly, 25 µl of Dynabeads were transferred to a 1.5-mltube and washed twice with 1% BSA–PBS to remove the preservative. The Dynabeads were resuspended in 1 ml 1% BSA–PBS and transferred to a 1.5-ml tube containing 3 µl of KRT18 antibodies. The suspension was incubated for 30 min at 4°C on a Sample Mixer. Then, the tube was placed in the magnetic particle concentrator (Dyna Mag™ 5, Invitrogen) for 30 s. After another washing step and aspiration of the supernatant containing unbound antibodies, the antibody-coated Dynabeads were resuspended in 250 µl 1% BSA–PBS. These 250 µl were added to the 1.25 ml cell suspension and incubated for 1 h at 4°C on the Sample Mixer. Finally, the specifically bound cells were collected by magnetic incubation for 1 min. The bead bound cell pellet was washed and immediately used for RNA extraction.

RNA extraction and sequencing

The extracted MECs were washed by PBS and were subjected to RNA extraction by Trizol method (*Ambion*-USA) according to the manufacturer's instructions. RNA was quantified by spectrophotometer (ThermoFisher, USA) and the quality and integrity was assessed by bioanalyzer (Agilent, USA). Illumina TruSeq stranded mRNA sample preparation kit was used to generate cDNA libraries according to the manufacturer's recommendations. Approximately ~4ug of total RNA were used to prepare the RNA-seq library using the TruSeq RNA Sample Prep Kits (Illumina, USA). The poly-A containing mRNA molecules were purified using poly-T oligo-attached magnetic beads. Following purification, the mRNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were used to synthesize first strand cDNA using reverse transcriptase and random primers followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. After adenylating of 3' ends of DNA fragments, hybridisation was initiated by ligating Illumina PE adapter and index. cDNA fragments (200bp) were generated and were selectively enriched to construct the final sequencing library using Illumina PCR Primer Cocktail. An Illumina Hiseq 2500 platform was used to sequence the libraries.

Data analysis

Quality control for paired-end reads

Raw data (raw reads) of fastq format were first processed using in-house perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapters, reads containing ploy-N and low quality reads from raw data. At the same time, Q30 (the proportion of bases with a phred base quality score greater than 30, i.e., the proportion of read bases whose error rate is less than 0.1%) and GC content of the clean data were calculated. All the downstream analyses were based on the clean data.

Reads mapping on the bovine reference genome and gene expression analysis

The bovine genome UMD3.1 (ftp://ftp.ensembl.org/pub/release-85/gtf/bos_taurus/Bos_taurus.UMD3.1.85.gtf.gz) was utilized as the reference genome for the assembly. Index of the reference genome was built using Bowtie v2.0.6 [15] and paired-end clean reads were aligned to the reference genome using TopHat v2.1.1[16] (http://tophat.cbcb.umd.edu/). Also, a data-base of splice junctions were generated by TopHat based on the gene model annotation files[17] (ftp://ftp.ensembl.org/pub/release-77/gtf/bos_taurus). HTSeq v0.6.1 was used to count the reads numbers mapped to each gene [18]. Transcript abundances were estimated as reads per kilobase of exon model per million mapped reads (FPKM), which was calculated based on the length of the gene and reads count mapped to this gene. A gene was defined as expressed if it was detected above 0.01 RPKM in any given sample [19].

Differential expression analysis

The DEGs and transcript analysis across the three stages (Early, Peak and Late lactation) were performed using Cuffdiff package [20]. For Cuffdiff, the commonly used fragments per kilo-base of transcript per million mapped fragments (FPKM) value [21] in pair-end sequencing experiments incorporated two normalization steps; i.e., the number of fragments was normalized by the transcript's length and the total yield of the fragments to ensure accurate quantification of the gene's expression. TopHat's read alignments were assembled by Cufflinks and then the differentially expressed genes and transcripts across different stages were detected and quantified by Cuffdiff, which is included in the Cufflinks package, using a rigorous sophisticated statistical analysis [20]. The differentially expressed genes, their corresponding attributes, fold changes (in log2 scale), p -values, and false discovery rate corrected values were reported in the output files from Cuffdiff.

Gene ontology (GO) and KEGG analysis of DEGs

GO and pathway enrichment analysis of DEGs was implemented in the GO stats R package (2.34.0) [22], in which gene length bias was corrected. GO terms and KEGG pathways (http://www.genome.jp/kegg/) with P-value less than 0.05 were considered significantly enriched by DEGs.

Results

Sequencing and mapping of the bovine mammary epithelial cell transcriptome

Sequencing of 18 libraries of milk epithelial cells at three different stages of lactation in Kashmiri cattle and Jersey cows generated a total of 1.65 billion reads (range 68,43 - 136,83 million reads/library). Out of this number, 1.47 billion reads (95.82%) passed quality control and were aligned to the bovine genome UMD3.1. Alignment of the sequence reads against the bovine genome UMD3.1yielded 82.22%-94.8% of uniquely aligned reads across different samples and there were 5.42%-17.78% unmapped or multi-position matched reads and they were excluded from further analyses. Furthermore, using a pairwise approach, three cows per group were used to eliminate the background noise of individual-specific transcription, enabling acquisition of more relevant data from the two groups. According to the FPKM value, mapped genes were divided into a low expression group (< 10 FPKM), a moderate expression group (≥ 10 FPKM to 500 FPKM) and a high expression group (≥ 500 FPKM)

(**Table 1**). For both breeds, the lactation time point with the highest number of expressed genes was 2rd stage (36931genes in Jersey and 31887 genes in local cattle).

Table1: *RNA-seq gene expression distribution*

	Jersey			Kashmiri		
	1st	2nd	3rd	1st	2nd	3rd
Total number of expressed genes	31850	36931	36702	24621	31887	38866
Highly expressed genes(>500 FPKM)	219	245	339	203	180	150
Medium expressed genes(10-500 FPKM)	4935	4817	5496	4970	4723	4713
Lowly expressed genes(<10 FPKM)	26696	31869	30867	19448	26984	34003

Top most expressed genes in each stage of lactation in two breeds

The number of genes with the highest FPKM values (≥ 500) in each breed and lactation time are shown in Table1. The results show that in bovine milk, the top-10 genes have very high expression values. For both breeds, Jersey and local cattle, the ten highest expressed genes at each time point accumulate approximately 70% of the total gene FPKM reads, which means that a small number of genes contribute to a large fraction of the total RNA extracted from milk epithelial cells (**Table 2**). A total of 13 genes, ranging between means of 2494.05 to 245241.9 FPKM, are encompassed in the top-10 highest expressed genes.

Table 2: *Highly expressed genes at three different stages of lactation in two different breeds*

Genes	Kashmiri			Genes	Jersey		
	1st Stage (FPKM)	2nd Stage (FPKM)	3rd Stage (FPKM)		1st Stage (FPKM)	2nd Stage (FPKM)	3rd Stage (FPKM)
CSN1S1	151546	128456.2	194567.1	CSN1S1	164564.5	156743.9	234598.2
CSN1S2	71398.8	89764.6	125657	CSN1S2	150945	167512.3	182657.2
CSN3	161987.2	159574.6	192456.5	CSN3	181457.6	150675.1	231241.7
CSN2	211651.8	185657.23	227861.2	CSN2	243561.9	200165	245241.9
LGB	147876.25	98365.7	148641.1	LGB	137645.7	100387	140034.5
LALBA	34224.31	37937.7	52945.3	LALBA	41669.1	40597.8	45782.2
RPLP1	21400.1	15712.7	7041	RPLP1	9184.13	3006.93	3774.83
RPS28	17835.6	9313.4	3436.38	RPS28	8888.89	-	-
RPS20/snoU54	14922.6	6510.56		RPS20/snoU54	8756.15	-	-
RPLPO	11925.8	-	3133.07	RPLPO	5337.15	-	-
RPS12	-	-	-	CCL14	-	7509.43	10501.5
B2M	-	7468.17	2723.17	B2M	-	10721	14357.5
				RPS12	-	3089.68	2494.05

'-'Indicate that the FPKM values are below threshold values

Differentially expressed genes across different stages of lactation

For three different lactation stages in Kashmiri and Jersey cattle, a total of 1805, 608 and 1080 genes were differentially expressed between 1st and 2nd, 2nd and 3rd and 1st and 3rd stages respectively. Likewise, a total of 1174, 978 and 3392 genes were differentially expressed across different stages in Jersey cattle (**Table 3**). The largest number of differentially expressed genes are observed between 1st and 2nd stage in Local cattle and between 1st and 3rd stage in Jersey cattle.

A total of five DEGs with highest fold change were selected from each stage for further analysis. **Table 4** shows the list of these five genes in Local and Jersey cattle with their respective FPKM values across different time points of lactation.

Table 3: *Number of significantly differentially expressed genes across different stages of lactation*

Breed	Stage of lactation	upregulated	downregulated	Total no. of DEGs (q value≤0.05)
Kashmiri cattle	1st vs 2nd	580	1225	1805
	2nd vs 3rd	403	205	608
	1st vs 3rd	624	456	1080
Jersey	1st vs 2nd	538	636	1174
	2nd vs 3rd	910	68	978
	1st vs 3rd	2672	720	3392

Table 4: *Five Differentially expressed genes with highest fold change between different stages of lactation*

Kashmiri cattle											
Stage 1st vs 2nd				Stage 2nd vs 3rd				Stage 1st vs 3rd			
Gene	Stage1 (FPKM value)	Stage 2 (FPKM value)	Fold change	Gene	Stage1 (FPKM value)	Stage 2 (FPKM value)	Fold change	Gene	Stage1 (FPKM value)	Stage 2 (FPKM value)	Fold change
SNORD50	7047.23	3.28723	-11.066	B3GNT6	0.004071	0.532565	7.03145	PEAR1	0.023992	3.08885	7.00836
TMEM232	1051.25	0.505671	-11.216	PEAR1	0.056574	3.08885	5.77078	DMP1	0.05297	6.48591	6.93599
ATP6V0D2	0.074882	4.92161	6.03836	TMEM232	0.115735	4.01734	5.11735	SLC18B1	0.00407	0.443622	6.76822
CPM	0.13847	4.89707	5.14427	LMO7	0.418231	14.4041	5.10603	SNORD50	7047.23	1.65661	-12.0546
SPDEF	34.5312	0.415322	-6.37753	ME1	0.048525	1.40554	4.85625	TMEM232	1051.25	0.55980	-10.8749
Jersey cattle											
Stage 1st vs 2nd				Stage 2nd vs 3rd				Stage 1st vs 3rd			
Gene	Stage1 (FPKM value)	Stage 2 (FPKM value)	Fold change	Gene	Stage1 (FPKM value)	Stage 2 (FPKM value)	Fold change	Gene	Stage1 (FPKM value)	Stage 2 (FPKM value)	Fold change
TMSB4X	0.12323	6.5137	5.72403	SLC27A6	0.066849	6.06855	6.5043	CD69	0.347745	76.2555	7.77667
IL1A	3.90345	136.758	5.13073	SLC25A21	0.052236	2.99908	5.84334	P2RY14	0.401521	56.9604	7.14834
SPATA3	7.38234	0.078409	-6.55691	CCDC13	0.063189	3.04106	5.58875	ND6	3.06231	290.037	6.56547
IRX1	40.9304	0.490715	-6.38214	MAPK4	0.030915	1.48648	5.58747	KMO	1.20894	97.2595	6.33002
SLC27A6	5.29831	0.066849	-6.38214	HHATL	0.035794	1.63308	5.51174	STRA8	0.94622	63.9683	6.0704

Distribution of differentially expressed genes between local and Jersey cattle

In the comparison between local and Jersey cattle for DEGs between respective stages, a total of 1592, 2064 and 2641 DEGs were reported between different stages. It was found that the number of differentially expressed genes increased between 1st to 3rd stage of lactation (**Table 5**). More number of genes were found to be upregulated in Jersey in comparison to Local

cattle. To simplify the comparative analysis between these two genotypes only 5 genes with highest fold change were selected for further analysis. **Table 6** shows the detailed information regarding these top 5 genes at each lactation stage.

Table 5: *DEGs between local cattle and jersey*

Kashmiri cattle vs Jersey	**upregulated**	**downregulated**	**Total no. of DEGs (q value≤0.05)**
1^{st} vs 1^{st}	1082	510	1592
2^{nd} vs 2^{nd}	1212	852	2064
3^{rd} vs 3^{rd}	2384	257	2641

Table 6: *Five Differentially expressed genes with highest fold change between local and jersey cattle*

Stage	Genes	FPKM values (local)	FPKM values (Jersey)	Fold change
local 1^{st} stage vs Jersey 1^{st} stage	SLC18B1	0.00407	3.65084	9.80905
	bta-mir-223	0.314812	96.6193	8.26168
	CD-207	0.077504	13.1987	7.411
	SNORD50	7047.23	1.19075	-12.531
	TMEM237	1051.25	1.05273	-9.96376
local 2^{nd} stage vs Jersey 2^{nd} stage	NLRP12	0.071833	23.1786	8.33392
	DMXL2	0.617948	189.453	8.26014
	MGLL	0.025106	5.32724	7.72919
	STEAP4	0.813946	136.958	7.39458
	GPR84	3.96817	643.733	7.3185
local 3^{rd} stage vs Jersey 3^{rd} stage	NLRP12	0.209232	100.401	8.90646
	bta-mir-223	0.064803	21.3033	8.3608
	PTX3	4.66395	1505.93	8.33489
	GPR84	4.53768	1287.36	8.14824
	STEAP4	1.64204	391.072	7.8958

Gene ontology and KEGG pathway analysis of DEGs

Multiple pathways and GO terms including biological processes, cellular components and molecular functions were significantly enriched for DEGs across lactation between and within the breed. Multiple significant GO terms are related to developmental process, extracellular matrix, collagen and basement membrane. The important pathways identified were metabolic pathways, chemokine pathway, osteoclast and JAK-STAT pathway across lactation in Jersey cows. In addition to above pathways Fructose and mannose metabolic pathway, amino sugars and nucleotide pathways and Pyruvate metabolism pathways were significantly enriched in local cattle. In the comparative analysis between Jersey and Local cattle metabolic pathway P13K-AKT signalling pathway, JAK-STAT pathway and MAPK pathway were significantly enriched.

Discussion

In this study, we investigated the whole milk transcriptome profile of the bovine milk epithelial cells in Jersey and local Kashmiri indigenous cattle through high throughput RNA sequencing which has many advantages over traditional cDNA microarray technologies (23).

On the basis of RIN values of extracted RNA samples from both experimental subjects, sequencing was done at three stages of lactation with three biological replicates. We sequenced the transcripts deeply and performed the analysis through cuffdiff programme. The average percentage of uniquely mapped reads for each sample was substantially higher than that obtained in another study of bovine milk transcriptome analysis, where approximately 65% of the total uniquely mapped to the Btau 4.0 reference genome (24). From the global view, the transcriptome results revealed that the majority of genes fall under low expression group having FPKM<10 across lactation for both the breeds. The average gene expression levels reported here are in agreement with previous reports on the mammary gland transcriptome of the other ruminant species (24,25). However, maximum gene expression at late lactation in local cattle could be the reason for short lactation period compared to jersey cattle. As these genes were mainly associated with involution of mammary gland and apoptosis of epithelial cells.

As the lactation cycle is dynamic in nature (6) and is characterised by different milk phenome at different stages. The initial stage (stage 1st) is associated with highest fat and protein percentage whereas, mid lactation (stage 2nd) is characterised by highest yield. The yield further decreases in late lactation with increase in fat and protein percentage. The DGE profile across lactation was found to follow the lactation cycle dynamics (6). Highest number of DEGs were found between stage 1st and 2nd in local and between 1st and 3rd in Jersey cattle. This may be due to longer lactation length in Jersey cow. It was also found that highest number of genes were expressed in stage 1st in both breeds. The higher protein and fat content during initial stages of lactation could be a possible reason for such expression pattern (26).

Normalisation of gene expression by library size and gene length was done to determine the most expressed genes at each lactation stage. This normalisation approach facilitates the comparison of genes within a sample (27). The results show that in bovine milk top 10 genes have very high FPKM values. These highly expressed genes in both breeds contribute approximately more than 65% of the total gene expression levels, indicating that smaller number of genes contribute to a large fraction of total RNA extracted from MECs. In our study it was found that caseins are included in the top expressed gene category in contrast to findings of Wickramsinghe et al. (24) in Holstein cows. Such observation could be due to higher milk protein content in Jersey cattle (28).

The genes with highest expression levels in terms of FPKM values are CSN1S1, CSN1S2, CSN3, CSN2, LGB, and LALBA, which encode four casein and two whey proteins. These proteins constitute about 3-4% of total milk composition (29). Casein and whey protein expression remained almost constant throughout the lactation in both the breeds. However, the expression was higher in Jersey cattle as compared to Local cows. It is likely that these genes with great effects have been fixed through long term genetic selection. Apart from major casein and whey proteins there are other genes which are highly expressed in two breeds: RPLP1, RPS28, RPLPO and B2M. The RPLP1 gene encodes a ribosomal phosphoprotein that is a component of the 60S subunit. This protein, which is a functional equivalent of the *E. coli* L7/L12 ribosomal protein, belongs to the L12P family of ribosomal

proteins. It plays an important role in the elongation step of protein synthesis. RPLP1 deletion in primary mouse embryonic fibroblasts didn't change the global protein expression but did change the expression patterns of specific protein subsets involved in protein folding and the unfolded protein response, cell death, protein transport and signal transduction, among others [30]. Deficiency of the RPS28 and RPLPO proteins causes cell death through reactive oxygen species (ROS) accumulation and MAPK1/ERK2 signalling pathway activation [31]. It has been observed that RPS28 is present throughout the lactation in Local cattle in comparison to Jersey cows where it is present in early lactation only. On the other hand, RPLPO is present in early and late lactation in Local cattle and in early lactation in Jersey cows. This indicates the high antioxidant activity of milk from local cattle as compared to Jersey cattle. B2M encodes the beta-2-microglobulin protein, an integral component of the Fc receptor heterodimer, which is involved in transferring Immunoglobulin G (IgG) from serum into milk across mammary epithelial cells (32). It was highly expressed in Jersey milk in mid and late lactation, whereas it was totally absent during the initial stages of lactation in both breeds. Some B2M haplotypes have been reported to be related to higher concentrations of IgG1 in bovine milk [33]. Increasing IgG levels in milk could become important as IgG enhanced dairy products are in demand by consumers to obtain protective immunity [34]. Chemokine (C-C motif) ligand 14 (CCL14) is a small cytokine belonging to the CC chemokine family. It is exclusively expressed in Jersey cows. CCL14 is involved in Cellular calcium homeostasis, immune response, positive regulation of cell proliferation [35].

In differential gene expression studies we found five genes showing highest fold change across different stages of lactation. In Local cattle it was found that only three genes SNORD50, PEAR1 and TMEM232 were common to the top five genes across different stages. SNORD50 is non-coding gene and shows the highest fold change and its regulatory function is not known but should be explored for its possible role milk synthesis. PEAR1 (Platelet endothelial aggregation receptor 1) has been found to be co-expressed with SYK, GLUP1 and CRIM1. SYK have a role in angiogenesis whereas GLUP1 causes apoptosis of cells. This indicates the role of PEAR1 in the development and involution of mammary gland. Another gene TMEM232 was co-expressed with SLC25A46 and NAA10 (N-alpha-acetyltransferase 10). SLC25A46 is involved in solute carrier across mitochondrial membrane and hence may play a role in milk production. NAA10 was found to regulate the mTOR pathway which regulates the milk protein synthesis [36].

Large number of GO terms were significantly enriched by DEGs across different stages of lactation. These terms were involved in mammary gland development, protein and lipid metabolism process, signal transduction, cell cycle/death, differentiation and immune functions. These terms have been associated different mammary gland physiological functions [37,38]. The early stages of lactation were mainly enriched by insulin stimulus (GO: 0032868) and growth factor stimulus (GO: 0070848) terms, which play a crucial role in lactogenesis through prolactin signalling [39].

In the comparison between Jersey and Local Kashmiri cattle GO enrichment terms for annotated DEGs were categorized into different functional groups. The most significantly GO terms within this dataset were immune response (GO: 0006955), and immune system

process (GO: 0002376) which were associated to upregulated genes in Local cattle. In the cellular component GO terms category, the most significant ones were related to extracellular matrix (GO: 0005576) and extracellular region part (GO: 0044421). The GO: 0005576 annotation is associated with upregulated genes in Jersey cattle, whereas GO: 0044421 is associated with upregulated genes in Local cattle. Based on the functional GO analysis, gene expression in Local cattle is shown to be largely enriched by genes involved in immune and stress responses (GO: 0006955, GO: 0006950). This indicates the high level of immune response of Local cattle and it needs further investigations to fully explore the comparative immune system strength in Local germplasm and to exploit it further through proper breeding programmes. The GO term related to endopeptidase activity (GO: 0052548), which is associated with upregulated gene in Local cattle. This may be of special interest as endopeptidase activity has special effects on phsico-chemical properties of milk and hence on dairy products which may be of special use to local population. The KEGG pathway analysis showed that P13K-AKT, MAPK and JAK-STAT pathways are significantly enriched during lactation in Jersey and Local cattle. P13K-AKT regulates the mammary epithelial cell differentiation through prolactin action. The mammary differentiation due to P13K-AKT activation results in autocrine prolactin secretion which in turn activates JAK-STAT pathway [41]. Lemay et al [5] observed that P13K-AKT pathway to be highly significant during lactation in mouse mammary gland. The genes involved in this pathway are highly expressed in Jersey during early lactation, indicating the enhanced cell growth and proliferation during early lactation as compared to late lactation stages. The JAK-STAT pathway is important during lactation and the key genes like SOCS3, STAT2, STAT3, STA6, JAK2, JAK3 showed high expression pattern in Jersey cattle. These genes have possible roles in mammary gland development and lactation [40]. The MAPK pathway is another signalling pathway which plays a critical role in mammary epithelial cell development and enhances milk production by modulating alveolar cell proliferation and branching [41]. This pathway was highly enriched in upregulated genes in Jersey Cattle.

Conclusion

This study presents the first comprehensive overview on the dynamic expression profile of the bovine milk transcriptome. A comparative analysis of the milk transcriptome of two cattle breeds also provided us with a valuable insight into the expression profile differences which could serve to identify the key genes responsible for different phenotypic characters between the two breeds. It was found that casein and whey genes are highly expressed and constitute more than 65% of the total transcript abundance. RPLPO and RPS28 were found to be more highly expressed in local cattle milk indicating its high possible antioxidant properties. Also immune response and immune system process GO terms were significantly expressed in Local cattle. The KEGG pathway analysis revealed that the JAK-STAT, MAPK and P13K-AKT pathways have been significantly enriched by upregulated genes in Jersey cows. The study provides some possible unique physico-chemical and immune properties in milk from local cattle which need to be further explored. This could lead to the development of certain niche products and thereby help in conserving this unique germplasm which has

been diluted through extensive cross breeding programmes.

References

1. Bionaz M. et al. Old and new stories: revelations from functional analysis of the bovine mammary transcriptome during the lactation cycle. PLoS One 7, e33268 (2012).
2. Georges M. et al. . Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139, 907–20 (1995).
3. Andersson L. Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. Genetica 136, 341–9 (2009).
4. Schennink A., Bovenhuis H., Leon-Kloosterziel K. M., van Arendonk J. A. & Visker M. H. Effect of polymorphisms in the FASN, OLR1, PPARGC1A, PRL and STAT5A genes on bovine milk-fat composition. Anim Genet 40, 909–16 (2009).
5. Lemay et al., 2009. The bovine lactation genome: insights into the evolution of mammalian milk. *Genome Biology.* Volume 10, Issue 4, Article R43
6. Stanton T. L., Jones L. R., Everett R. W., Kachman S. D. (1992). Estimating milk, fat, and protein lactation curves with a test day model. J. Dairy Sci. 75, 1691–1700.
7. Suárez-Vega A, Gutiérrez-Gil B, Klopp C, Robert-Granie C, Tosser-Klopp G, Arranz JJ. Characterization and comparative analysis of the milk transcriptome in two dairy sheep breeds using RNA sequencing. Sci Rep. 2015;5:18399.
8. Li, Z.; Liu, H.Y.; Jin, X.L.; Lo, L.J.; Liu, J.X. Expression profiles of microRNAs from lactating and non-lactating bovine mammary glands and identification of miRNA related to lactation. BMC Genom. 2012, 13, 731.
9. Strucken EM, Laurenson YCSM and Brockmann GA (2015) Go with the flow—biology and genetics of the lactation cycle. Front. Genet.
10. Pan Q., Shai O., Lee L. J., Frey B. J. & Blencowe B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40, 1413–5 (2008).
11. Boutinaud M, Jammes H. Potential uses of milk epithelial cells: a review. Reprod Nutr Dev. 2002;42:133–147.
12. Li C, Cai W, Zhou C, Yin H, Zhang Z, Loor JJ, Sun D, Zhang Q, Liu J, Zhang S. RNA-Seq reveals 10 novel promising candidate genes affecting milk protein concentration in the Chinese Holstein population. Sci Rep. 2016.
13. Finucane K, McFadden T, Bond J, Kennelly J, Zhao F-Q. Onset of lactation in the bovine mammary gland: gene expression profiling indicates a strong inhibition of gene expression in cell proliferation. Funct Integ Genomics. 2008;8:251–264.
14. Canovas, A. et al. Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-Sequencing. Sci. Rep. 4, 5297 (2014).

15. Boutinaud M., Ben Chedly M. H., Delamaire E., Guinard-Flament J. (2008). Milking and feed, restriction regulate transcripts of mammary epithelial cells purified from milk. *J. Dairy Sci.* 91 988–998.
16. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10, R25 (2009).
17. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–11 (2009).
18. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7, 562–78 (2012).
19. Anders, S., Pyl, P. T. & Huber, W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–9 (2015).
20. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5, 621–8 (2008)
21. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010, 28(5):511–515.
22. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008, 5(7):621–628.
23. Falcon S. & Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics23, 257–8 (2007).
24. Nagalakshmi U, Waern K, Snyder M (2010) RNA-Seq: a method for comprehensive transcriptome analysis. Curr Protoc Mol Biol Chapter 4 Unit 4.11.1-13.
25. Wickramasinghe, S., Rincon, G., Islas-Trejo, A. & Medrano, J. F. Transcriptional profiling of bovine milk using RNA sequencing. BMC Genomics 13, 45-2164-13-45 (2012).
26. Lin, J. et al. Transcriptome analysis of the mammary gland from GH transgenic goats during involution. Gene 565, 228–234 (2015)
27. Gao Y, Lin X, Shi K, Yan Z, Wang Z (2013) Bovine Mammary Gene Expression Profiling during the Onset of Lactation. PLoS ONE 8(8): e70393.
28. Dillies, M. A. et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform 14, 671–683 (2013).
29. Swank, V. A., et al. "Jersey calf performance in response to high-protein, high-fat liquid feeds with varied fatty acid profiles: Blood metabolites and liver gene expression." *Journal of dairy science* 96.6 (2013): 3845-3856.
30. Pulina, G. & Nudda, A. Milk Production in Dairy Sheep Feeding and Nutrition (ed. Poulina, G.) Ch. 1, 1–12 (CABI Publishing, 2004).
31. Perucho L, Artero-Castro A, Guerrero S, Ramón y Cajal S, LLeonart ME, Wang Z-Q (2014) RPLP1, a Crucial Ribosomal Protein for Embryonic Development of the Nervous System. PLoS ONE 9(6): e99956.
32. Ana Artero-Castro, Mileidys Perez-Alea, Andrea Feliciano, Jose A Leal, Mónica Genestar, Josep Castellvi, Vicente Peg, Santiago Ramón y Cajal & Matilde E LLeonart. Disruption of the ribosomal P complex leads to stress-induced autophagy.

Autophagy Vol. 11 , Iss. 9,2015
33. Anderson, C. L. et al. Perspective– FcRn transports albumin: relevance to immunology and medicine. Trends Immunol. 27, 343–348 (2006).
34. Zhao, S. et al. Association of polymorphisms of beta-2-microglobulin gene (β 2m) with milk IgG1 content in Chinese Holstein dairy cows. Livestock Science 143, 289–292 (2012).
35. Hurley, W. L. & Theil, P. K. Perspectives on immunoglobulins in colostrum and milk. Nutrients 3, 442–474 (2011)
36. Paola Piantoni, Ping Wang, James K. Drackley,Walter L. Hurley,and Juan J. Loor. Expression of Metabolic, Tissue Remodeling, Oxidative Stress, and Inflammatory Pathways in Mammary Tissue During Involution in Lactating Dairy Cows. Bioinform Biol Insights. 2010; 4: 85–97.
37. Wang M, Xu B, Wang H, Bu D, Wang J, Loor J-J (2014) Effects of Arginine Concentration on the *In Vitro* Expression of Casein and mTOR Pathway Related Genes in Mammary Epithelial Cells from Dairy Cattle. PLoS ONE 9(5): e95985.
38. Suárez-Vega, A., Gutiérrez-Gil, B., Klopp, C., Robert-Granie, C., Tosser-Klopp, G., Arranz, J.J.Characterization and comparative analysis of the milk transcriptome in two dairy sheep breeds using RNA sequencing. *Sci. Rep.* 2015;5:18399
39. Maller, O., Martinson, H. & Schedin, P. Extracellular matrix composition reveals complex and dynamic stromal-epithelial interactions in the mammary gland. J. Mammary Gland Biol. Neoplasia 15, 301–318 (2010).
40. Macias, H. & Hinck, L. Mammary gland development. Wiley Interdiscip. Rev. Dev. Biol. 1, 533–557 (2012).
41. Yamaji D., Kang K., Robinson G. W., Hennighausen L. (2013). Sequential activation of genetic programs in mouse mammary epithelium during pregnancy depends on STAT5A/B concentration.*Nucleic Acids Res.* 41 1622–1636.
42. Fata, Jimmie E., et al. "The MAPK ERK-1, 2 pathway integrates distinct and antagonistic signals from TGFα and FGF7 in morphogenesis of mouse mammary epithelium." *Developmental biology* 306.1 (2007): 193-207.