Characterization and Comparative Analysis of Transcriptional Profiles of Porcine Colostrum and Mature Milk Across Multiple Parities

Brittney N Keel (brittney.keel@ars.usda.gov)
USDA-ARS Roman L Hruska US Meat Animal Research Center https://orcid.org/0000-0002-8099-0870
Amanda K Lindholm-Perry
USDA-ARS Roman L Hruska US Meat Animal Research Center
William T Oliver
USDA-ARS Roman L Hruska US Meat Animal Research Center
James E Wells
USDA-ARS Roman L Hruska US Meat Animal Research Center
Shuna A Jones
USDA-ARS Roman L Hruska US Meat Animal Research Center
Lea A Rempel
USDA-ARS Roman L Hruska US Meat Animal Research Center

Research article

Keywords: RNA-Seq, transcriptome, milk, colostrum, total RNA, gene expression, long non-coding RNA, Lancaster method

Posted Date: October 5th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-72346/v1

Abstract

Background

The expression of genes and their regulation during lactation in swine is not well-understood. In order to gain a better understanding of genes and pathways involved in sow lactation, total RNA from colostrum and mature milk samples were sequenced from 65 sows across four parities. A stringent bioinformatic pipeline was used to identify and characterize 44,234 transcripts.

Results

The 44,234 identified transcripts included 41,875 previously annotated transcripts and 2,359 novel transcripts. Differential gene expression analysis was conducted using a generalized linear model coupled with the Lancaster method for *P*-value aggregation across transcripts. In total, 1,900 differentially expressed genes (DEG) were identified for the milk type main effect, and 373 DEG were identified for the milk type x parity interaction. Several gene ontology (GO) terms related to immune response were significant for the milk type main effect, supporting the widely-accepted hypothesis that immunoglobulins and immune cells are transferred to the neonate via colostrum.

Conclusions

This is the first study to perform global transcriptome analysis from whole milk samples in sows from multiple parities. Our results provide important information and insight into synthesis of milk proteins and innate immunity and potential targets for future improvement of swine lactation and piglet development.

Background

Colostrum and milk play a key role in survival and growth of the neonate, providing essential nutrients and antibodies [1]. Langer et al. [2] investigated differences in composition of colostrum and mature milk in several eutherian species and found that in some species colostrum contains higher concentrations of proteins than mature milk and in other species the fluids have similar composition. These differences are likely due to species-specific strategies for immunoglobulin transfer, i.e. prenatal transfer via placenta or yolk sac versus postnatal transfer via colostrum [2]. The critical importance of colostrum and milk for the newborn piglet has been well-documented [1, 3].

Piglet growth and survival are critical to the swine industry. Progeny born to primiparous sows (gilts) are born lighter, grow slower, and have higher mortality rates than those born to multiparous sows [4, 5]. It has been hypothesized that differences in lifetime performance between gilt progeny and sow progeny may be due to differences in lactation performance, specifically lower levels of immunoglobulin G (IgG) and other energetic components in the colostrum and milk of gilts. However, data from Craig et al. [6] showed no parity differences in total IgG, fat, protein, lactose, and net energy concentrations. These results suggest that the poorer performance of gilt progeny is unlikely due to insufficient nutrient levels and is more likely due to differences in colostrum and milk intake and their ability to digest and absorb each component [5].

Relatively little is known about the genetic control of lactation in swine. To date, no quantitative trait loci (QTL) for milk yield or composition have been reported. Causal genes and mutations for these traits have been identified in dairy cattle, using QTL mapping, candidate gene analysis, genome-wide association analysis (GWAS), and next-generation sequencing (NGS) technologies [7-12]. It has been shown that milk production traits are under strong epigenetic regulation [13]. Researchers have looked closely at the relationship between DNA methylation and milk production traits in dairy cattle [14]. Long non-coding RNAs (lncRNA), which are critical epigenetic regulators of cell metabolism, have recently been associated with milk proteins in Chinese Holstein cattle [15].

Recent developments in NGS technology have allowed the identification of lncRNA, RNA which are at least 200 base pairs (bp) in length and do not possess protein-coding capabilities [16]. The role of lncRNA in the transcriptome has received much attention in the last decade. They have been shown to regulate expression of genes in close proximity (*cis*-acting) as well as genes at a

distance (*trans*-acting) in the genome via different mechanisms [17]. LncRNA have also been associated with metabolic, immunological, and developmental regulation, as well as phenotypic variation in livestock [18].

A limited number of transcriptomic studies related to porcine milk production have been reported [19, 20]. These studies focused on the mammary gland transcriptome during late gestation [19] and during the transition from colostrogenesis to lactogenesis [20]. To our knowledge, there have been no studies that have reported direct sequencing of porcine whole milk samples. In fact, only six livestock RNA-sequencing (RNA-Seq) studies have reported sequencing of milk, including three in cattle [9, 21, 22], one in goat [23], one in sheep [24], and one in buffalo [25]. The emphasis of these studies was gene expression related to the lactation process, and as such, milk somatic cells were sequenced as a proxy for the mammary gland tissue [26].

In the current study, we performed total RNA-Seq on porcine colostrum and mature milk samples, from dams in parities one through four, in order to characterize both transcriptomes. We identified novel protein-coding and lncRNA transcripts and quantified expression of both known and novel transcripts. Expression profiles were compared to identify differentially expressed genes (DEG) between colostrum and mature milk across parities.

Results

High-throughput sequencing

RNA-Seq libraries were sequenced generating over 6 billion 75 bp paired-end reads, with an average of 46.2 million reads per library (Table S1). The number of reads in the colostrum libraries ranged from 22.6 to 81.8 million reads with an average of 44.4 million reads, while the number of reads in the mature milk libraries ranged from 24.2 to 97.8 million reads with an average of 48.0 million reads. After adapter removal and read trimming, the resulting high-quality reads were mapped to the Sscrofa 11.1 genome assembly with an average 99.6% read mapping rate per library. The number of reads aligning to known messenger RNA (mRNA), miscellaneous RNA (miscRNA), non-coding RNA (ncRNA), and pseudogenes in the swine genome are presented in Table S2. It was observed that most reads mapped to known mRNA, while 50.5% of colostrum reads and 44.5% of milk reads were mapped outside of annotated loci, potentially harboring novel transcripts (Figure 1).

Transcript identification and characterization

Transcripts, assembled individually for each library, were merged into a single set of 460,853 putative transcripts. This set was subjected to several filtering steps in order to remove transcriptional noise and classify transcripts (Figure 2). Transcripts identified in only one library and lowly expressed transcripts were removed, as these were considered transcriptional noise. The remaining set of transcripts was filtered to include only those with class codes '=', 'u', 'x', and 'i' (Figure S1). The transcripts with class codes 'u', 'x', and 'i' were further filtered by length, and number of exons. This set of 8,699 putative novel transcripts were then subjected to classification by open reading frame (ORF) length and protein coding potential score to complete transcript characterization. In total, 44,234 transcripts were identified in the porcine milk transcriptome, including 41,875 previously annotated transcripts as well as 2,359 novel transcripts.

Transcripts corresponded to 17,740 unique gene loci. Previously annotated transcripts corresponded to 16,515 known gene loci in the Sus scrofa genome, while unannotated protein-coding and non-coding transcripts corresponded to 971 and 571 loci, respectively. Genomic coordinates of the identified novel transcripts are given in Tables S3 and S4. Among the novel lncRNA transcripts, 224 and 171 were long intergenic non-coding RNA (lincRNA) and identified, unknown long non-coding RNA (ilncRNA), respectively, while 286 lncRNA flanked a protein-coding gene in a divergent orientation (long non-coding natural antisense transcripts; lncNAT) (Figure 3A). Using the BLAST algorithm, a total of 364 lncRNA exhibited homology with transcripts in the porcine NONCODE database, 25 lncRNA exhibited homology with non-coding transcripts in other species, and 167 lncRNA were homologous to noncoding transcripts in both swine and other species (Figure 3B; Table S5). A similar analysis identified that 1,463 of the novel mRNA transcripts were homologous to known transcripts in swine and other species (Figure 4).

Basic sequence features of the novel transcripts, including length, exon number, expression, and ORF length, are shown in Figure 5 and Table 1. Novel lncRNA were expressed at significantly lower levels than novel mRNA and known transcripts (Figure 5A). The exon number and size of the novel lncRNA and coding transcripts were notably smaller than that of known transcripts (Figure

5B,C). The ORF length of novel lncRNA was significantly shorter than ORF length in known and novel coding transcripts, while the ORF length of novel coding transcripts was longer than that of known transcripts (Figure 5D).

PCA and differential expression analysis

The principal component analysis (PCA) plot (Figure 6) showed that colostrum and mature milk transcript expression profiles seem to fall into distinct clusters, while there was no clear clustering of samples by parity. After multiple testing correction, we identified 518 differentially expressed transcripts (DET) for the milk type x parity interaction and 2,613 DET for the milk type main effect (Tables S6 and S7). Table 2 shows the classifications of DET. The DET set for the milk type main effect was comprised of 2,244 known transcripts, 286 novel coding transcripts, and 83 novel lncRNA, while the interaction DET set included 464 known transcripts and 38 and 15 novel coding transcripts and lncRNA, respectively. *P*-values of transcripts were aggregated for each gene loci to obtain DEG. A total of 1,900 DEG were identified for the milk type main effect, and 373 DEG were identified for the milk type x parity interaction (Tables S8 and S9).

Gene ontology and pathway analysis

Gene ontology (GO) analysis of the DEG indicated that genes associated with the milk type main effect were predominantly involved in binding (38.9%), catalytic activity (28.4%), transporter activity (8.5%), molecular function regulation (7.6%), and transcription regulation activity (7.3%). A total of 487 biological process, 65 molecular function, and 77 cellular component GO terms were significantly enriched in this gene set (Table S10). Additionally, 4 KEGG pathways were significantly enriched.

Like the milk type main effect genes, DEG for the milk type x parity interaction were involved in binding (37.8%), catalytic activity (27.9%), molecular transducer activity (11.7%), molecular function regulation (7.2%), and transporter activity (9.0%). Three molecular function GO terms were significantly enriched (Table S11).

Discussion

Milk production by the sow is a major limiting factor in the growth and survival of her litter. Knowledge of porcine milk composition, as well as understanding genetic factors underlying its variation, is a matter of ongoing interest. In this study, we performed the first exhaustive characterization of the porcine milk transcriptome derived from whole milk samples. The goal was to highlight differences in samples collected during early and mid-lactation and compare transcriptome profiles of dams across multiple parities.

A limited number of milk transcriptome studies in livestock have been conducted [9, 21-25], and to date, RNA-Seq of porcine milk has not been reported. Non-invasive sampling of the transcriptome of milk-producing cells via RNA secreted into porcine milk provides a powerful window into the biology of swine lactation. Lemay et al. [26] showed that RNA extracted from milk provides a better representation of RNA from mammary epithelial cells than RNA from whole mammary tissue.

Total RNA was isolated from 130 fresh whole milk samples (65 colostrum and 65 mature milk) from dams across four parities. In most milk transcriptome studies, milk is fractionated, and RNA is extracted from somatic cells, milk fat, or whey. Total RNA concentrations tend to be higher in the milk fat and somatic cells than in the whey fraction, while RNA integrity of somatic cells is higher than those of milk fat and whey [27, 28]. Low RIN values in this study (average RIN = 4.0) are likely due to the presence of small amounts of cytoplasmic material in milk fat globules [29], bacteria and small RNAs (miRNA) in the fat fraction [30], and degraded and/or free RNAs. Each milk fraction has its own place in research settings. The advantages and disadvantages of each RNA source has previously been summarized [26]. In this study, we chose to utilize whole milk samples in order to capture the broader transcriptomic signatures of porcine colostrum and milk. We were able to process the samples much more quickly than had we fractionated the milk, and our sample represents the entirety of what is being ingested by the growing piglet.

Libraries were sequenced to an average depth of 46 million reads per library. A depth of 40 million reads is considered sufficient for reliable detection of major splice isoforms for abundant and moderately abundant transcripts [31]. When generating our sequence data, we targeted a depth of 50 million reads per library. However, there was considerable variation in sequence depth across libraries. Some of this variation can be attributed to technical aspects of NGS technology, such as the stochasticity of sequencing, RNA quality, and library preparation.

A total of 44,234 transcripts were identified in this study, of which approximately 95% are annotated in the current swine genome build. Transcripts corresponded to 17,740 unique gene loci, including 16,515 known porcine genes. The number of expressed genes is comparable to those reported in similar studies in sheep [24] and goat [23]. A smaller number of expressed genes (~13,500) was reported in the buffalo milk transcriptome [25]. This discrepancy is likely to due to the swine, sheep, and goat reference genomes being more complete and of higher quality.

In general, gene expression values covered a wide range of magnitudes (Figure 7), and the gene expression profiles of colostrum and mature milk were not highly correlated (Pearson correlation coefficient 0.222). There was a large overlap (12 out of 15) in the top fifteen most abundantly expressed genes in colostrum and mature milk (Table 3). Among the top expressed genes were *CSN3, CSN2, CSN1S1, LALBA, FASN, EEF1A1, TPT1, SAA3,* and *WAP,* which have been previously identified among the top expressed genes in milk samples from other species [23-25, 32, 33].

As expected, many of the top expressed genes were related to biosynthesis of milk proteins. Expression levels of *CSN2, CSN3, CSN1S1,LALBA,* and *WAP,* which encode for the synthesis of the main milk proteins casein and whey, increased from early to mid-lactation stages. A similar gene expression pattern has been identified in a previous swine study [34], as well as in goat [23], cattle [33], and sheep [24]. High expression of the *EEF1A1* gene is also related to high levels of milk protein synthesis, as *EEF1A1* is one of the most abundant protein synthesis factors [23]. Additionally, ribosomal proteins *RPLP0* and *RPS2* were among the top expressed genes in colostrum and exhibited a slight decrease in expression during mid-lactation. These genes were also found to be highly expressed in early lactation in buffalo [25].

In addition to milk protein synthesis genes, genes associated with milk fat were among the top expressed genes, and their expression was nearly constant across lactation stages. Milk fat composition is known to influence piglet growth and development [35]. The *FABP3* gene, which is involved in the uptake and transport of fatty acids, has been linked to milk fat synthesis in cattle [36]. *FASN* is directly involved in most of the short and medium-chain fatty acids in milk [37], and *XDH* is involved in the formation of the lipid droplet in milk [38].

The 20 most significant DET for milk type and interaction are given in Tables 4 and 5, respectively. Many of the most significant DET were associated with genes involved in milk fat synthesis. Transcript *MSTRG.335722.71455*, which is associated with insulin induced gene 1 (*INSIG1*), was found to be up-regulated in mature milk samples. *INSIG1* is known to regulate the expression of sterol regulatory element-binding protein 1 (*SREBP1*; also denoted *SREBF1*), which is central to milk fat synthesis [20, 39]. Although not one of the most significant DET, transcript *MSTRG.239932.53831*, associated with *SREBP1*, was identified as a DET for the milk type x parity interaction. Transcripts *MSTRG.189750.42732* (*THRSP* gene), *MSTRG.108970.23990* (*SP1* gene), *MSTRG.283058.62377* (*ANXA7* gene) are other milk fat synthesis genes among the most significant DET. In general, expression of milk fat synthesis genes was up-regulated in mature milk samples compared to colostrum, which agreed with expression patterns observed across bovine lactation stages [39]. Our results highlight that the transition from swine colostrum to mature milk is marked by a shift from high protein contents to high fat and lactose contents [40].

In this study, DEG were identified by aggregating *P*-values across transcripts associated with each gene via the Lancaster method, rather than using gene read counts directly. Using this approach not only maintains both transcript and gene-level resolution, but also bypasses issues of different variances and directions of change across constituent transcripts. This method outperforms other gene-level methods and provides a coherent analysis between transcripts and genes [41].

One of the major aims of this study was to evaluate DEG between lactation stages across multiple parities. Progeny born to multiparous sows generally exhibit superior growth performance compared to those born to primiparous sows. However, colostrum and milk composition profiles, are highly similar across parities [6].

Glucose transport is a major precursor to lactose synthesis, which is synthesized in the Golgi vesicle of mammary secretory alveolar epithelial cells during lactation [42]. Glucose-6-phosphate transporters *SLC37A2* and *SLC37A3*, glucose cotransporter *SLC5A8*, and glucose transporter *SLC2A5* were identified as DEG for the milk type main effect. Glucose transport across the plasma membrane of mammalian cells is carried out by two distinct processes one of which involves glucose transporters from the *GLUT* gene family (encoded by *SLC2A* genes) and the other which involves glucose transporters from the *SGLT* family (encoded by *SLC5A* genes). Both the *SLC2A5* and *SLC5A8* genes were up-regulated in colostrum. Crisá et al. [23] identified significant up-regulation of

members of the *SLC2A* gene family and polysaccharide and glycosamino-glycan binding molecular function to be enriched in goat colostrum samples compared to mature milk.

Members of the *SLC35* gene family encode nucleotide sugar transporters localizing at the Golgi apparatus and/or the endoplasmic reticulum. These transporters transport nucleotide sugars pooled in the cytosol into the lumen of these organelles, where most glycoconjugate synthesis occurs [47]. Currently, the *SLC35* gene family is comprised of 31 genes which are divided into 7 subfamilies, *SLC35A* to *SLC35G* [48]. GDP-fucose transporters *SLC35C1* and *SLC35C2* were identified as DEG for the milk type main effect, with *SLC35C2* up-regulated in colostrum and *SLC35C1* down-regulated. *SLC35A5* and *SLC35G1* were identified as DEG for the milk type x parity interaction, with both genes being up-regulated in mature milk in the first three parities and up-regulated in colostrum during the fourth parity (Figure 8). Crisà et al. [23] identified 3 DEG from the *SLC35* family that were up-regulated in goat colostrum compared to mature milk, as well as enrichment of glycosaminoglycan binding molecular in colostrum. Consistent with this result, we also identified the enrichment of glycosaminoglycan binding molecular function, with 29 of the 124 porcine genes associated with the GO term being present in our milk type DEG set. These results support the hypothesis that oligosaccharide metabolism decreases over the course of lactation as oligosaccharide concentrations in milk have been shown to peak in colostrum and decrease over the remainder of lactation in human [49], goat [50], and pig [51].

Forty-two genes from the PI3K-Akt pathway were found to be differentially expressed. Although not statistically significant after FDR-correction, this pathway had a nominal *P*-value of 0.025 in the enrichment analysis. The PI3K-Akt pathway is a key signaling node for lactogenic expansion and differentiation of the luminal mammary epithelium, as numerous signaling pathways that regulate lactogenic development converge on PI3K-Akt, including the insulin-like growth factor 1 receptor (*IGF1R*), *RANKL* and *RANK*, integrins, and *PRLR*-to-*JAK2*-to-*STAT5A* pathways [52]. In general, expression of DEG in the PI3K-Akt pathway was up-regulated in colostrum. The janus kinase 1 (*JAK1*) gene, a key component of the PI3K-Akt pathway, had an FDR-corrected *P*-value of 0.02 but was filtered out of the DEG list based on the log2 fold change threshold. *JAK1* was more highly expressed in mature milk, and the difference in expression level between colostrum and mature milk was more pronounced in the later parities (Figure 9). Using a novel mammary gland-specific *JAK1* knockout mouse model, Sakamoto et al. [53] demonstrated that *JAK1* is essential to involution, the mammary gland remodeling process at weaning where the milk-producing epithelial cells are replaced with adipocytes [54]. The potential effects of involution on mammary development and milk yield during the next lactation are not well understood. Ford et al. [55] noted that mammary glands in sows that were suckled during lactation were larger than non-suckled glands at the end of the involution process, suggesting a possible beneficial effect on redevelopment during the next gestation.

Several GO terms related to immune response, particularly leukocyte activation, were significantly enriched in the DEG for the milk type main effect (Table S10). The majority of genes in these pathways were up-regulated in colostrum. This finding is consistent with the widely accepted theory that immunoglobulins and immune cells are transferred to the neonate via colostrum. In pigs, the epitheliochorial nature of the placenta prohibits transfer of maternal immune cells and immunoglobulins to the fetus, and thus, the piglet relies on the successful absorption of colostral components to acquire maternal immunity [56]. Proinflammatory cytokines play an important role in the development of the neonatal immune system by mediating the early local and systemic responses to microbial challenges [57]. A total of 37 DEG were associated with cytokine secretion, including interleukin 17 receptor C (*IL17RC*), interleukin 27 receptor A (*IL27RA*), tumor necrosis receptor superfamily members 1b and 15 (*TNFRSF1B* and *TNFRSF15*), and tumor necrosis factor superfamily member 4 (*TNSFF4*). Several other genes in these gene families were shown to be up-regulated in early porcine lactation by Palombo et al. [20].

Antimicrobial proteins naturally present in colostrum and milk can kill and inhibit a broad spectrum of bacteria [58]. Milk is also known to exert chemotactic activity on neutrophils [59], an important innate host defense against microorganisms. The chemokine superfamily encodes secreted proteins involved in immunoregulatory and inflammatory processes. The CXC chemokine ligand 16 (*CXCL16*), which encodes a chemokine antimicrobial protein [60], was up-regulated in colostrum samples (Figure 10). Palombo et al. [20] identified significant up-regulation of *CXCL2* and *CXCL10* in day 1 postpartum swine milk samples compared to colostrum. Many of the main chemokines (*CXCL2, CXCL8, CXCL9, CXCL10, CXCL11, CXCL13, CXC14*, and *CXCL16*) were expressed in our samples, and expression patterns for *CXCL10* and *CXCL2* were consistent with the findings of Palombo et al. [20]. Our results differed from previously published results in that *CXCL8* was the most abundantly expressed chemokine in both colostrum and mature milk, and *CXCL3* was not expressed in our samples. One factor contributing to this discrepancy was the use of the improved reference genome (Sscrofa 11.1), where many of the gaps and misassemblies present in the Sscrofa 10.2 genome build

were resolved and the annotation was significantly improved. These results suggest that chemokine ligands may play an important role in the transition from colostrum to mature milk in swine, likely helping prompt recruitment of neutrophils.

Conclusions

This is the first study to describe the whole milk transcriptomic profile of porcine colostrum and milk across multiple parities. Using total RNA-Seq, we identified and characterized 2,359 novel transcripts expressed in milk and detected several hundred genes that were differentially expressed between colostrum and milk and across parity. Our findings have produced several highly specialized and functional candidate genes that may contribute to postnatal development and growth of piglets, as well as lactation in the sow.

Methods

Population and sampling

A four-breed composite line (Maternal Landrace × High-lean Landrace × Duroc × Yorkshire) maintained at U.S. Meat Animal Research Center (USMARC) for at least 18 generations was used for the collection of data in this project and have been previously described [61, 62]. Litter sizes were adjusted within 48 hours of farrowing to ensure litters were approximately equal in size but did not exceed the number of functional teats. Mammary excretion samples were collected on day of farrowing (d 0; colostrum) and again on day 10 post-farrowing (d 10; mature milk) from a total of 65 dams, 16 first parity (P1), 25 second parity (P2), 15 third parity (P3), and 9 fourth parity (P4). The power calculation for this experiment was conducted using the online RNASeqSampleSize tool ([63]; http://cqs.mc.vanderbilt.edu/shiny/RnaSeqSampleSize/). The power of using 65 animals to detect ~1,000 DEG, with maximum dispersion 0.5 and minimum fold change of 2.82, at FDR level 0.05 from 17,740 expressed genes was found to be 0.99.After sample collection, animals remained at USMARC and progressed through the breeding system according to standard operating procedures.

In most cases, no external stimulant (i.e., oxytocin) was needed to collect colostrum at time of farrowing, as farrowing stimulates endogenous oxytocin production and milk letdown activity. However, if enough colostrum could not be collected within 10-30 minutes, an intramuscular injection of oxytocin (20 IU) was administered to stimulate colostrum letdown. Teats were sprayed with iodine (5%) and ethanol (70%) and wiped clean with a chem-wipe, and then 10 mL of colostrum was collected manually from the third and fourth teat on one side of the sow. On d 10, piglets were separated from the sow for approximately one hour, and sows were given an intramuscular injection of 20 IU oxytocin to stimulate milk letdown. Teats were cleaned, and 10 mL of milk was collected manually from the third and fourth teat on one side of the sow. Fresh samples were transported to the laboratory on ice. Samples (250 µL) were aliquoted into individual lysis D matrix tubes (MP Biomedicals, LLC, Solon, OH) with 1 mL TRIzol reagent (Invitrogen, Thermo Fisher Scientific, Waltham, MA) and stored at -80°C until RNA isolation.

RNA isolation and sequencing

RNA was isolated using the FastPrep-24 5G Instrument (MP Biomedicals, LLC) with cryogenic lysis. Briefly, RNA was isolated by high-speed cellular disruption using multi-directional, simultaneous bead beating of sample material (i.e., colostrum or milk) with a cool adapter for cryogenic lysis at 6.0 m/sec for 40 sec. Lysed samples were transferred into a clean tube, and completion of isolation occurred following manufacturer's recommended protocol for TRIzol. The final RNA pellet was dried at RT for 10 min and resuspended in 30 µL water (Invitrogen UltraPure DNase/RNase-free, Thermo Fisher Scientific). RNA was quantified using a NanoDrop UV-Vis spectrophotometer (Thermo Fisher Scientific) and RNA integrity was assessed using an Agilent Bioanalyzer System (Agilent, Santa Clara, CA).

Total RNA samples extracted from colostrum or milk were prepared for RNA sequencing with the TruSeq Stranded Total RNA sample preparation kit (Illumina, San Diego, CA) following the guidelines of the manufacturer. Libraries were quantified with RT-qPCR using the NEBNext Library Quant Kit (New England Biolabs, Inc., Beverly, MA, USA) on a CFX384 thermal cycler (Bio-Rad, Hercules, CA, USA), and the size and quality of the library was evaluated with an Agilent Bioanalyzer DNA 1000 kit (Santa Clara, CA, USA). The libraries were diluted to 4nM with Illumina RSB. Libraries were paired-end sequenced with 150 cycle high output sequencing kits on an Illumina NextSeq 500 instrument.

Processing RNA-Seq data

Alignment of RNA-Seq reads was carried out as follows. First, quality of the raw paired-end sequence reads in individual fastq files was assessed using FastQC (Version 0.11.5; www.bioinformatics.babraham.ac.uk/projects/fastqc), and reads were trimmed to remove adapter sequences and low-quality bases using the Trimmomatic software (Version 0.35) [64]. The remaining reads were mapped to the Sscrofa 11.1 genome assembly (NCBI accession AEMK00000000.2) using Hisat2 (Version 2.1.0) [65] with default parameters.

Mapped transcripts were assembled for each library using Stringtie (Version 1.3.3) [66]. The NCBI Sscrofa 11.1 reference annotation (Release 106) was used to guide the assembly process. Transcripts from all samples were merged together using Stringtie merge mode to build a consensus set of transcripts.

Identification and characterization of novel transcripts

Transcript expression levels were quantified for each library using fragments per kilobase of exon per million mapped reads (FPKM) [67]. Transcripts expressed in a single sample, and transcripts with FPKM < 0.3 in all samples were removed. Gffcompare (Version 0.11.2) [68] was used to compare the list of assembled transcripts with the Sus scrofa reference annotation (NCBI release 106). Transcripts overlapping known transcript classes in the reference annotation (gffcompare class code '=') were assigned to the appropriate annotation class, while transcripts with gffcompare class codes 'x' (exonic overlap on the opposite strand), 'i' (fully contained in reference intron), and 'u' (unknown, intergenic) were considered to be potential novel transcripts.

A modified version of the discovery pipeline described in Cai et al. [15] was used to further filter transcripts and classify novel transcripts (Figure 2): (i) Filter out transcripts with short lengths (< 200 bp) and single exons; (ii) ORF obtained using TransDecoder (Version 5.5.0) [69]. Transcripts with no predicted ORF were filtered out, and transcripts ORF length ≥ 120 amino acids were considered protein-coding. (iii) Protein coding potential assessed using CPC2 (Version 2.0) [70], PLEK (Version 1.2) [71], and CNIT [72]; (iv) Transcripts were translated to amino acid sequences using the Transeq utility from EMBOSS (https://www.ebi.ac.uk/Tools/st/emboss_transeq/), and HMMER [73] was used to search for known protein domains against the Pfam database (Release 32.0) [74]. Transcripts with significant Pfam hits (E-value < 10.0) were classified as protein-coding. After steps (iii) and (iv), transcripts with significant Pfam hits, CPC2 classification "coding", PLEK score > 0, and CNIT > 0 were classified as protein-coding, and transcripts with no significant Pfam hits, CPC2 classification "noncoding", PLEK score < 0, and CNIT score < 0 were classified as non-coding. All other transcripts were discarded, as their coding potential was ambiguous.

The BLASTN algorithm from the BLAST+ package [75] was used to identify homology between (1) novel lncRNA and the NONCODE database (Version 5) and (2) novel mRNA and the NCBI Non-redundant Nucleotide database (nt; Version 5, https://ftp.ncbi.nlm.nih.gov/blast/db/). BLASTN was run with default parameters, and an E-value cutoff of 10.0 was used to define homologous sequences.

Differential expression and functional analyses

Raw read counts for the 44,234 transcripts and the 17,740 corresponding genes were normalized using DESeq2 (Version 1.26) [76]. The PCA plot, using normalized read counts, was generated using the plotPCA function from the DESeq2 package. Differential expression analysis of transcripts was performed using DESeq2 with the following generalized linear model:

$$Y = Type + Parity + Type\ x\ Parity.$$

Transcripts with FDR-adjusted *P*-value ≤ 0.01 were considered DET for the type x parity interaction. Transcripts that were not DET for the interaction term with |log2 fold change| ≥ 2 and FDR-adjusted *P*-value ≤ 0.01 were considered DET for the milk type main effect.

Transcript *P*-values were aggregated for each gene using the Lancaster method [77]

in order to generate gene-level analysis. This approach has been described in detail by Yi et al. [41]. Briefly, the Lancaster method is an extension of the Fisher method [78] for *P*-value aggregation, where under the null hypothesis that all genes have zero effect, the

test statistic $T = \sum_{i=1}^{K} \phi_{w_i}^{-1}(p_i)$ follows a chi-squared distribution with $df = \sum_{i=1}^{K} w_i$. Here, denotes the number of transcripts associated with the gene, $W_1,\ldots,W_K$ a set of weights for the transcript *P*-values $P_1,\ldots,P_K$, and $\phi_{w_i}^{-1}$ the inverse CDF of the gamma distribution with shape parameter $\alpha_i = \frac{w_i}{2}$ and scale parameter $\beta = 2.$

Here, the baseMean parameter from the DESeq2 output was used as in the Lancaster method. Aggregated *P*-values were corrected using the Benjamini-Hochberg method. Genes with FDR-adjusted *P*-value ≤ 0.01 were considered DEG. DEG for the milk type x parity interaction were removed from the milk type main effect DEG set. Log2-fold changes (log2FC; mature milk vs. colostrum) were computed for each of the genes, and genes with were filtered out of the milk type main effect DEG set.

Enrichment analysis of gene function and cellular pathways was performed for DEG using the iPathwayGuide software (Version 1910; Advaita Bio, http://advaitabio.com/ipathwayguide) with the default Mus musculus data as background. For GO analysis, an over-representation test, based on a hypergeometric distribution, was used to compute the statistical significance of observing more than the expected number of DEG. A GO term was considered statistically significant at FDR-corrected $P \leq 0.05$. Pathway over-representation analysis was performed by comparing the number of affected genes associated with a pathway between groups. Pathways were considered statistically significant at FDR-corrected $P \leq 0.05$.

Abbreviations

IgG: immunoglobin

QTL: quantitative trait loci

GWAS: genome-wide association study

NGS: next-generation sequencing

lncRNA: long non-coding RNA

bp: base pairs

RNA-Seq: RNA-sequencing

DEG: differentially expressed gene

mRNA: messenger RNA

miscRNA: miscellaneous RNA

ncRNA: non-coding RNA

ORF: open reading frame

lincRNA: long intergenic non-coding RNA

ilncRNA: identified unknown long non-coding RNA

lncNAT: long non-coding natural antisense transcripts

PCA: principal component analysis

DET: differentially expressed transcript

GO: gene ontology

USMARC: U.S. Meat Animal Research Center

NCBI: National Center for Biotechnology Information

SRA: Sequence Read Archive

FPKM: fragment per kilobase of exon per million mapped reads

Declarations

Ethics approval and consent to particpate

The USMARC Animal Care and Use Committee reviewed and approved all animal procedures. The procedures for handling pigs complied with the Guide for the Care and Use of Agricultural Animals in Agricultural Research and Teaching [79].

Consent for publication

Not applicable.

Availability of data and material

Sequence data used in this study was submitted to the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) with Accession Number PRJNA640341.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

BNK, LAR, WTO, and JEW conceived of the study and participated in its design and coordination. All authors were involved in the acquisition of data, and BNK performed data analyses. BNK drafted the manuscript, and ALP, WTO, SAJ, JEW, and LAR contributed to the writing and editing. All authors read and approved the final manuscript.

Acknowledgements

The authors wish to acknowledge Shanda Watts for her outstanding technical and laboratory assistance; the USMARC Core Laboratory for performing the sequencing; Heather Oeltjen-Bruns and Miguel Cervantes for their assistance with sample collection; Mike Judy for assistance with data organization; and the USMARC Swine Operations crew.

References

1. Theil PK, Lauridsen C, Quesnel H. Neonatal piglet survival: impact of sow nutrition around parturition on fetal glycogen deposition and production and composition of colostrum and transient milk. 2014;8:1021-1030.
2. Langer P. Differences in the composition of colostrum and milk in eutherians reflect differences in immunoglobulin transfer. Mammal. 2009;90(2):332-339.
3. Quesnel H, Farmer C, Theil PK. Colostrum and milk production. Chapter 8. In Farmer C (ed.) The gestating and lactating sow. Wageningen Academic Publishers, Wageningen, the Netherlands, 2015.
4. Carney-Hinkle EE, Tran H, Bundy JW, Moreno R, Miller PS, Burkey TE. Effect of dam parity on litter performance, transfer of passive immunity, and progeny microbial ecology. Anim. Sci. 2013;91:2885-2893.
5. Craig JR, Collins CL, Bunter KL, Cottrell JJ, Dunshea FR, Pluske JR. Poorer lifetime growth performance in gilt progeny compared with sow progeny is largely due to weight differences at birth and reduced growth in the preweaning period, and is not improved by progeny segregation after weaning. Anim. Sci. 2017;95:4904-4916.

6. Craig JR, Dunshea FR, Cottrell JJ, Wijesiriwardana UA, Pluske JR. Primiparous and multiparous sows have largely similar colostrum and milk composition profiles throughout lactation. Animals (Basel.). 2019;9(2):35.
7. Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, et al. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics. 1995;139:907-920.
8. 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;21:592-598.
9. Wickramasinghe S, Hua S, Rincon G, Islas-Trejo A, German JB, Lebrilla CB, Medrano JF. Transcriptome profiling of bovine milk oligosaccharide metabolism genes using RNA-sequencing. PLoS ONE. 2011;6(4):e18895.
10. Littlejohn MD, Tiplady K, Fink TA, Lehnert K, Lopdell T, Johnson T, et al. Sequence-based association analysis reveals an MGST1 eQTL with pleiotropic effects on bovine milk composition. Nat. Sci. Rep. 2016;6:25376.
11. Gao Y, Jiang J, Yang S, Hou Y, Liu GE, Zhang S. CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. BMC Genomics. 2017;18:265.
12. Van Den Berg I, Hayes BJ, Chamberlain AJ, Goddard ME. Overlap between eQTL and QTL associated with production traits and fertility in dairy cattle. BMC Genomics. 2019;20:291.
13. Singh K, Molenaar AJ, Swanson KM, Gudex B, Arias JA, Erdman RA, et al. Epigenetics: a possible role in acute and transgenerational regulation of dairy cow milk production. Animal. 2012;6:375-381.
14. Thompson RP, Nilsson E, Skinner MK. Environmental epigenetics and epigenetic inheritance in domestic farm animals. Anim. Rep. Sci. 2020;106316.
15. Cai W, Li C, Liu S, Zhou C, Yin H, Song J, et al. Genome wide identification of novel long non-coding RNAs and their potential associations with milk proteins in Chinese Holstein cattle. Front. Genet. 2018;9:281.
16. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. 2012;81:145-166.
17. Yan P, Luo S, Lu JY, Shen X. Cis- and trans-acting lncRNAs in pluripotency and reprogramming. Curr. Opin. Genet. Devel. 2017;46:170-178.
18. Weikard R, Demasius W, Kuehn C. Mining long noncoding RNA in livestock. Anim. Genet. 2017;48(1):3-18.
19. Zhao W, Shahzad K, Jiang M, Graugnard DE, Rodriguez-Zas SL, Luo J, et al. Bioinformatics and gene network analyses of the swine mammary gland transcriptome during late gestation. Biol. Insights 2013;7:193-216.
20. Palombo V, Loor JJ, D'Andrea M, Vailati-Riboni M, Shahzad K, Krogh U, Theil PK. Transcriptional profiling of swine mammary gland during the transition from colostrogenesis to lactogenesis using RNA sequencing. BMC Genomics. 2018;19:322.
21. Medrano JF, Rincon G, Islas-Trejo A. Comparative analysis of bovine milk and mammary gland transcriptome using RNA-Seq. In 9th World Congress in Genetics Applied to Livestock Production, Leipzig, Germany, 2010, 852.
22. Wickramasinghe S, Rincon G, Islas-Trejo A, Medrano JF. Transcriptional profiling of bovine milk using RNA sequencing. BMC Genomics. 2012;13:45.
23. Crisà A, Ferrè F, Chillemi G, Moioli B. RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. BMC Vet. Res. 2016;12:264.
24. 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. Nat. Sci. Rep. 2015;5:18399.
25. Arora R, Sharma A, Sharma U, Girdhar Y, Kaur M, Kapoor P, et al. Buffalo milk transcriptome: a comparative analysis of early, mid and late lactation. Nat. Sci. Rep. 2019;9:5993.
26. Lemay DG, Hovey RC, Hartono SR, Hinde K, Smilowitz JT, Ventimiglia F, et al. Sequencing the transcriptome of milk production: milk trumps mammary tissue. BMC Genomics. 2013;14:872.
27. Alsaweed M, Hepworth AR, Lefèvre C, Hartmann PE, Geddes DT, Hassiotou F. Human milk microRNA and total RNA differ depending on milk fractionation. J. Cell. Biochem. 2015;116:2397-2407.
28. Li R, Dudemaine PL, Zhao X, Lei C, Ibeagha-Awemu EM. Comparative analysis of the miRNome of bovine milk fat, whey and cells. PLoS One. 2016;11(4):e0154129.
29. Huston GE, Patton S. Factors related to the formation of cytoplasmic crescents on milk fat globules. J. Dairy Sci. 1990;73:2061-2066.

30. Munch EM, Harris RA, Mohammad M, Benham AL, Pejerrey SM, Showalter L, et al. Transcriptome profiling of microRNA by next-gen deep sequencing reveals known and novel miRNA species in the lipid fraction of human breast milk. PLoS One. 2013;8(2):e50564.
31. Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods. 2008;5:1-8.
32. Lemay DG, Ballard OA, Hughes MA, Morrow AL, Horseman ND, Nommsen-Rivers L. RNA Sequencing of the human milk fat layer transcriptome reveals distinct gene expression profiles at three stages of lactation. PLoS One. 2013;8:e67531.
33. Canovas A, Rincon G, Bevilacqua C, Islas-Trejo A, Brenaut P, Hovey RC, et al. Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-Sequencing. Nat. Sci. Rep. 2014;4:5297.
34. Su Z, Dong X, Zhang B, Zeng Y, Fu Y, Yu J, et al. Gene expression profiling in porcine mammary gland during lactation and identification of breed- and developmental-stage-specific genes. Science in China Series C. 2006;49:26-36.
35. Cordero G, Isabel B, Morales J, Menoyo D, Piñero C, Daza A, Lopez-Bote CJ. Conjugated linoleic acid (CLA) during last week of gestation and lactation alters colostrum and milk fat composition and performance of reproductive sows. Feed Sci. Tech. 2011;168:232-240.
36. Bionaz M, Loor JJ. ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. J. Nutr. 2008;139:1019-1024.
37. Suburu J, Shi L, Wu JJ, Wang S, Samuel M, Thomas MK et al. Fatty acid synthase is required for mammary gland development and milk production during lactation. Am. Physiol-Endoc. M. 2014;306:E1132-E1143.
38. Tansey JT, Sztalryd C, Hlavin EM, Kimmel AR, Londos C. The central role of perilipin a in lipid metabolism and adipocyte lipolysis. IUBMB Life 2004;56:379-385.
39. Bionaz M, Loor JJ. Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics. 2008;9:366.
40. Klobasa F, Werhahn E, Butler JE. Composition of sow milk during lactation. J. Anim. Sci. 1987;64(5):1458-1466.
41. Yi L, Pimentel H, Bray NL, Pachter L. Gene-level differential analysis at transcript-level resolution. Genome Biol. 2018;19:53.
42. Zhao FQ. Biology of glucose transport in the mammary gland. J. Mammary Gland Biol. Neoplasia 2014;19(1):3-17.
43. Newburg DS. Glycobiology of human milk. Biochemistry (Mosc.) 2013;78:771-785.
44. Fong B, Ma K, McJarrow P. Quantification of bovine milk oligosaccharides using liquid chromatography-selected reaction monitoring-mass spectrometry. J. Agric. Food Chem. 2011;59:9788-9795.
45. Claps S, Di Napoli MA, Sepe L, Caputo AR, Rufrano D, Di Trana A, et al. Sialyloligosaccharides content in colostrum and milk of two goat breeds. Small Rumin. Res. 2014;121:116-119.
46. Salcedo J, Frese SA, Mills DA, Barile D. Characterization of porcine milk oligosaccharides during early lactation and their relation to the fecal microbiome. J. Dairy Sci. 2016;99(10):7733-7743.
47. Ishida N, Kawakita. Molecular physiology and pathology of the nucleotide sugar transporter family (SLC35). Eur. J. Physiol. 2004;447:768-775.
48. Song Z. Roles of the nucleotide sugar transporters (SLC35 family) in health and disease. Mol. Aspects Med. 2013;34(2-3):590-600.
49. Xu G, Davis JCC, Goonatilleke E, Smilowitz JT, German JB, Lebrilla CB. Absolute quantitation of human milk oligosaccharides reveals phenotypic variations during lactation. J Nutr. 2017;147(1):117-124.
50. Claps S, Di Napoli MA, Sepe L, Caputo AR, Rufrano D, Di Trana A, et al. Sialyloligosaccharides content in colostrum and milk of two goat breeds. Small Rumin. Res. 2014;121:116-119.
51. Tao N, Ochomicky KL, German JB, Donovan SM, Lebrilla CB. Structural determination and daily variations of porcine milk oligosaccharides. J. Agric. Food Chem. 2010;58:4653-4659.
52. Williams MM, Vaught DB, Joly MM, Hicks DJ, Sanchez V, Owens P, et al. ErbB3 drives mammary epithelial survival and differentiation during pregnancy and lactation. Breast Cancer Res. 2017;19(1):105.
53. Sakamoto K, Wehde BL, Yoo KH, Kim T, Rajbhandari N, Shin HY, et al. Janus kinase 1 is essential for inflammatory signaling and mammary gland remodeling. Mol. Cell. Biol. 2016;36(11):1673-1690.

54. Watson CJ. Key stages in mammary gland development – Involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ. Breast Cancer Res. 2006;8:203.
55. Ford JA, Kim SW, Rodriguez-Zas SL, Hurley WL. Quantification of mammary gland tissue size and composition changes after weaning in sows. Anim. Sci. 2003;81(10):2583-2589.
56. Bandrick M, Ariza-Nieto C, Baidoo SK, Molitor TW. Colostral antibody-mediated and cell-mediated immunity contributes to innate and antigen-specific immunity in piglets. Dev. Comp. Immunol. 2014;43(1):114-120.
57. Nguyen TV, Yuan L, Azevedo MSP, Jeong K, Gonzalez A-M, Saif LJ. Transfer of maternal cytokines to suckling piglets: in vivo and in vitro models with implications for immunomodulation of neonatal immunity. Vet. Immunol. Immunopathol. 2007;117:236.
58. Hurley WL, Theil PK. Perspectives on immunoglobulins in colostrum and milk. Nutrients. 2011;3:442-474.
59. Rainard P, Riollet C, Berthon P, Cunha P, Fromageau A, Rossignol C, et al. The chemokine CXCL3 is responsible for the constitutive chemotactic activity of bovine milk for neutrophils. Mol. Immunol. 2008;45:4020-4027.
60. Tohyama M, Sayama K, Komatsuzawa H, Hanakawa Y, Shirakata Y, Dai X, et al. CXCL16 is a novel mediator of the innate immunity of epidermal keratinocytes. Int. Immunol. 2007;19(9):1095-1102.
61. Rempel LA, Freking BA, Miles JR, Nonneman DJ, Rohrer GA, Schneider JF, Vallet JL. Association of porcine heparanase and hyaluronidase 1 and 2 with reproductive and production traits in Landrace-Duroc-Yorkshire population. Genet. 2011;2:20.
62. Rempel LA, Vallet JL, Lents CA, Nonneman DJ. Measurements of body composition during late gestation and lactation in first and second parity sows and its relationship to piglet production and post-weaning reproductive performance. Livestock Sci. 2015;178:289-295.
63. Zhao S, Li CI, Guo Y, Sheng Q, Shyr Y. RNASeqSampleSize: real data based sample size estimation for RNA sequencing. BMC Bioinformatics. 2018;19(1):1-8.
64. Bolger AM, Lohse M, Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114-2120.
65. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Meth. 2015;12:357-360.
66. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Biotech. 2015;33:290-295.
67. Lee S, Seo CH, Lim B, Yang JO, Oh J, Kim M, et al. Accurate quantification of transcriptome from RNA-Seq data by effective length normalization. Nucleic Acids Res. 2011;39(2):e9.
68. Pertea M, Kim D, Pertea G, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie, and Ballgown. Nat. Protoc. 2016;11(9):1650-1667.
69. Hass BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction for RNA-Seq: reference generation and analysis with Trinity. Protoc. 2013;8(8):1494.
70. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, Gao G. CPC2: a fast and accurate potential calculator based on sequence intrinsic features. Nucleic Acids Res. 2017;45(W1):W12-16.
71. 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;15:311.
72. Guo JC, Fang SS, Wu Y, Zhang JH, Chen Y, Liu J, et al. CNIT: a fast and accurate web tool for identifying protein-coding and long non-coding transcripts based on intrinsic sequence composition. Nucleic Acids Res. 2019;47(W1):W516-522.
73. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:W29-37.
74. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44(D1):D279-285.
75. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
76. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;12(12):550.
77. Lancaster H. The combination of probabilities: an application of orthonormal functions. Austral. J. Statist. 1961;3:20-33.

78. Fisher RA. Statistical methods for research workers. 1932. Edinburgh: Oliver and Boyd.
79. McGlone, J. Guide for Care and Use of Agricultural Animals in Agricultural Research and Teaching. Fed. Animal Science Society, Savoy, IL. 2010.

Tables

Table 1. Median characteristics of expressed transcripts.

	Novel lncRNA	Novel Coding	Known Transcripts
Expression[1]	0.1392*†	0.2262	0.2262
Length[2]	15.7*	14.1*	29.6
Number Exons	2*	2*	10
ORF Length[3]	321*†	363	339

[1] Measured in $\log_{10}$(FPKM+1).

[2] Measured in kbp.

[3] Measured in bp.

* Left-tailed Wilcoxon rank-sum *P*-value < 0.05 compared to known transcripts.

† Left-tailed Wilcoxon rank-sum *P*-value < 0.05 compared to novel coding.

Table 2. Classifications of DET for milk type main effect and milk type x parity interaction.

	Milk Type	Milk Type x Parity
Known Transcripts	2,244	464
Novel Coding Transcripts	286	38
Novel lncRNA	83	15
Total	2,613	517

Table 3. Top expressed genes in porcine colostrum and mature milk.

Gene Symbol	Description	Colostrum*	Mature Milk*
LOC102723301	RNA, 7SL, cytoplasmic 1	19.10 (1)	20.20 (3)
KLHDC2	Kelch domain containing 2	18.81 (2)	19.85 (4)
CSN3	Casein kappa	18.47 (3)	21.40 (2)
LOC100737553	Peptidyl-prolyl cis-trans isomerase A pseudogene	18.26 (4)	21.96 (1)
EEF1A1	Eukaryotic translation elongation factor 1 alpha 1	18.13 (5)	17.28 (10)
CSN1S1	Casein alpha s1	17.67 (6)	19.85 (5)
XDH	Xanthine dehydrogenase	17.17 (7)	17.41 (9)
TPT1	Tumor protein, translationally-controlled 1	16.95 (8)	16.57 (13)
FASN	Fatty acid synthase	16.79 (9)	16.71 (12)
CSN2	Casein beta	16.72 (10)	18.18 (7)
LOC102163816	Uncharacterized	16.58 (11)	16.53 (15)
FABP3	Fatty acid binding protein 3	16.46 (12)	16.71 (11)
RPLP0	Ribosomal protein lateral stalk subunit P0	16.29 (13)	14.95
EEF2	Eukaryotic translation elongation factor 2	16.24 (14)	15.15
RPS2	Ribsomal protein S2	15.98 (15)	14.37
LALBA	Lactalbumin alpha	14.98	18.61 (6)
SAA3	Serum amyloid A-3 protein	14.61	17.71 (8)
WAP	Whey acidic protein	13.24	16.56 (14)

*Average normalized gene expression value across samples. Number in parenthesis is ranking in the top fifteen expressed genes.

Table 4. Twenty most significant DET associated with milk type.

Transcript	Gene	Gene Description
MSTRG.34152.5080*	ALDH1A1	Aldehyde dehydrogenase 1 family member A1
MSTRG.204265.45537*	CCDC71L	Coiled-coil domain containing 71 like
MSTRG.150583.14*	MSTRG.150583	Novel coding RNA
MSTRG.255032.57048*	LOC110256328	ncRNA (uncharacterized)
MSTRG.282439.62156**	TET1	Tet methylcytosine dioxygenase 1
MSTRG.356301.74734*	TSPAN6	Tetraspanin 6
MSTRG.332729.70598**	LOC100513767	Colostrum trypsin inhibitor-like
MSTRG.241000.54008**	KCNH8	Potassium voltage-gated channel subfamily H member 8
MSTRG.250391.56376*	LOC110256292	ncRNA (uncharacterized)
MSTRG.18578.2786**	SHC4	SHC adaptor protein 4
MSTRG.161346.37587**	PSME1	Proteasome activator subunit 1
MSTRG.96359.19955*	C4H8orf46	Vexin
MSTRG.335722.71455*	INSIG1	Insulin induced gene 1
MSTRG.59760.11972*	LOC106508433	ncRNA (uncharacterized)
MSTRG.161229.37414*	HECTD1	HECT domain E3 ubiquitin protein ligase 1
MSTRG.189750.42732*	THRSP	Thyroid hormone responsive
MSTRG.137708.32180**	RSPO1	R-spondin 1
MSTRG.256226.57149*	SAMD7	Sterile alpha motif domain containing 7
MSTRG.145605.33246*	CTH	Cystathionine gamma-lyase
MSTRG.233871.51381**	TMEM106A	Transmembrane protein 106A

* Indicates up-regulation in mature milk.

** Indicates down-regulation in mature milk.

Table 5. Twenty most significant DET associated with milk type x parity interaction.

Transcript	Gene	Gene Description
MSTRG.108970.23990	SP1	Sp1 transcription factor
MSTRG.103400.8310	LOC100156563	Elongation factor 1-alpha 1-like
MSTRG.231019.50113	GPS1	G protein pathway suppressor 1
MSTRG.284779.62750	ADIRF	Adipogenesis regulatory factor
MSTRG.144177.33094	LOC100511460	RING finger and transmembrane domain-containing protein 1
MSTRG.230974.50069	C12H17orf62	Cytochrome b-245 chaperone 1
MSTRG.122057.26907	ACSF3	Acyl-CoA synthetase family member 3
MSTRG.336830.71885	LOC102162486	Putative RNA-binding protein Luc7-like 2
MSTRG.152860.35484	LOC110261493	Sperm acrosome membrane-associated protein 4-like
MSTRG.24606.3702	ZNF532	Zinc finger protein 532
MSTRG.166401.38675	LOC100157061	ncRNA (uncharacterized)
MSTRG.287208.63174	SLC35G1	Solute carrier family 35 member G1
MSTRG.283058.62377	ANXA7	Annexin A7
MSTRG.123610.27185	WWOX	WW domain containing oxidoreductase
MSTRG.122076.26887	CPNE7	Copine 7
MSTRG.113056.24989	TMTC1	Transmembrane O-mannosyltransferase targeting cadherins 1
MSTRG.41262.6306	LRRC8A	Leucine rich repeat containing 8 VRAC subunit A
MSTRG.264927.58828	PROS1	Protein S
MSTRG.234423.51759	LASP1	LIM and SH3 protein 1
MSTRG.132550.30513	PERM1	PPARGC1 and ESRR induced regulator, muscle 1

Figures

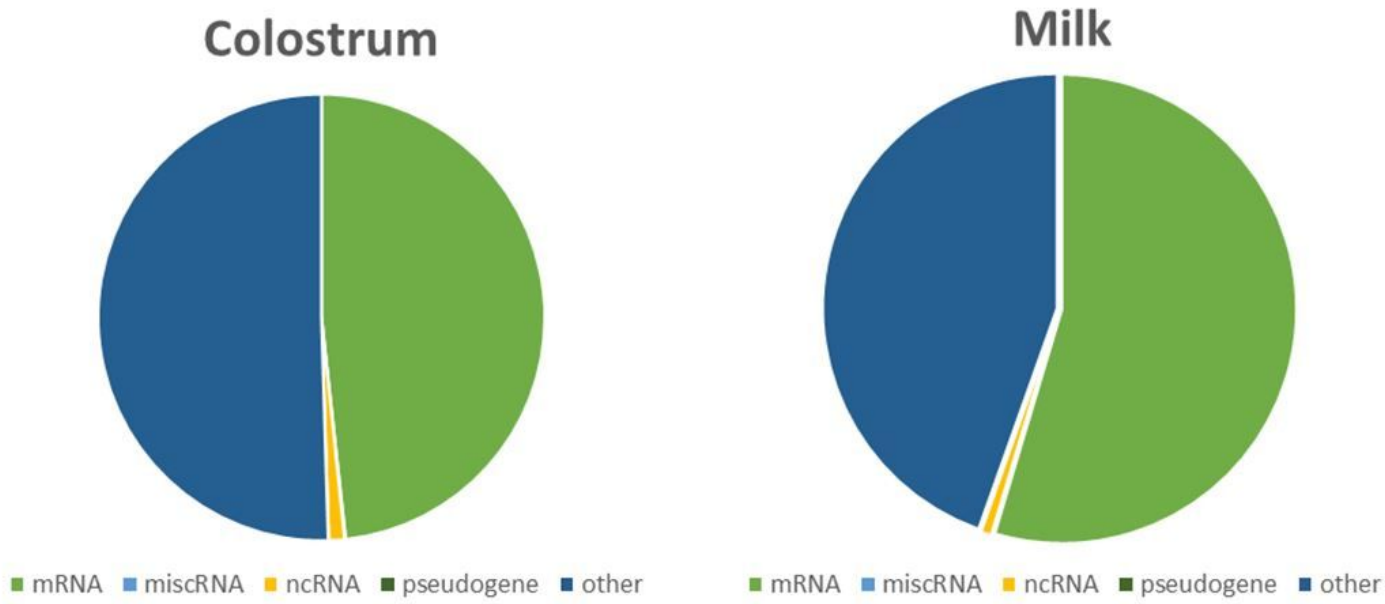

Figure 1

Distribution of reads aligning to the Sus scrofa 11.1 genome.

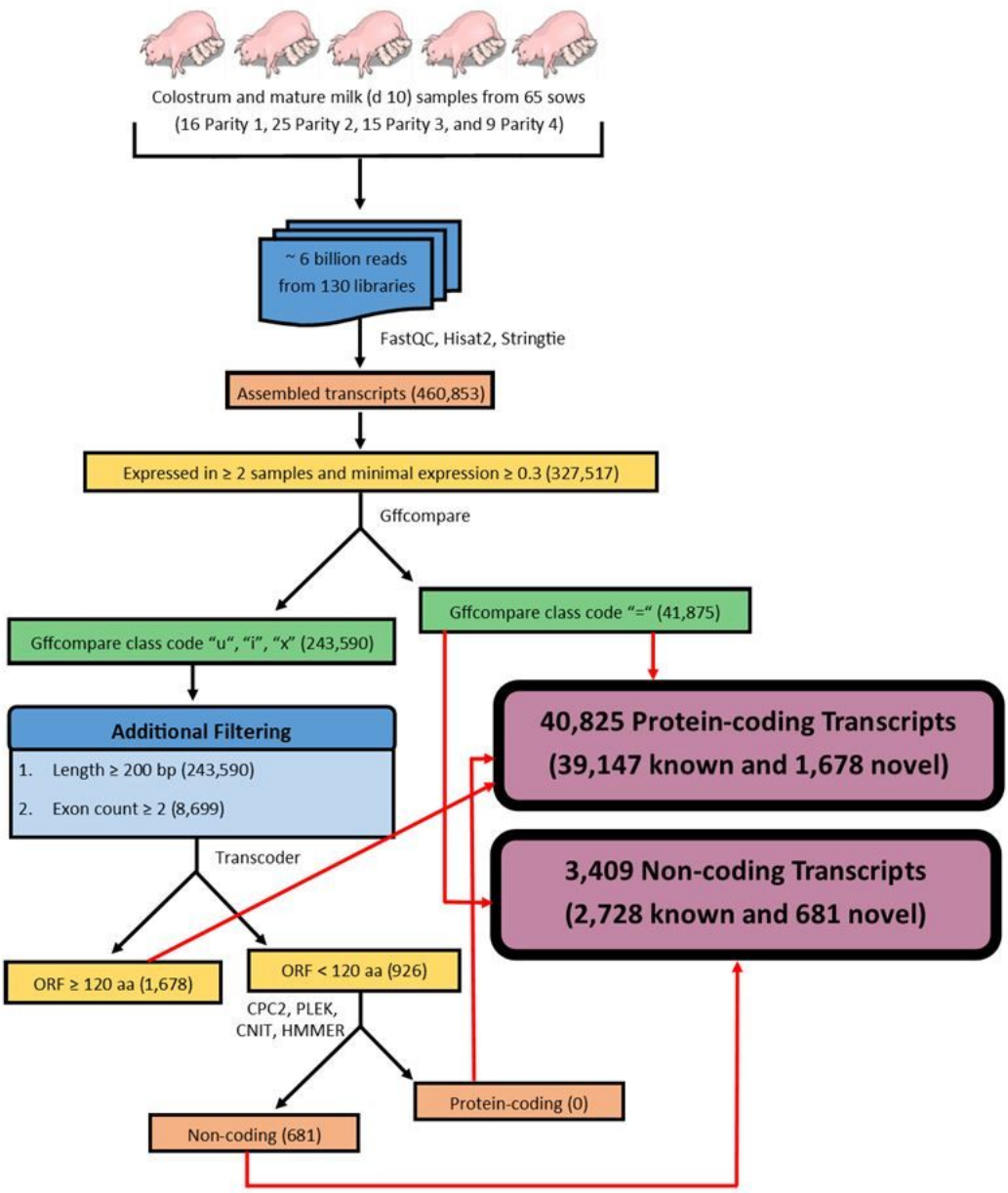

Figure 2

Computational pipeline used to determine novel transcripts from RNA-Seq data.

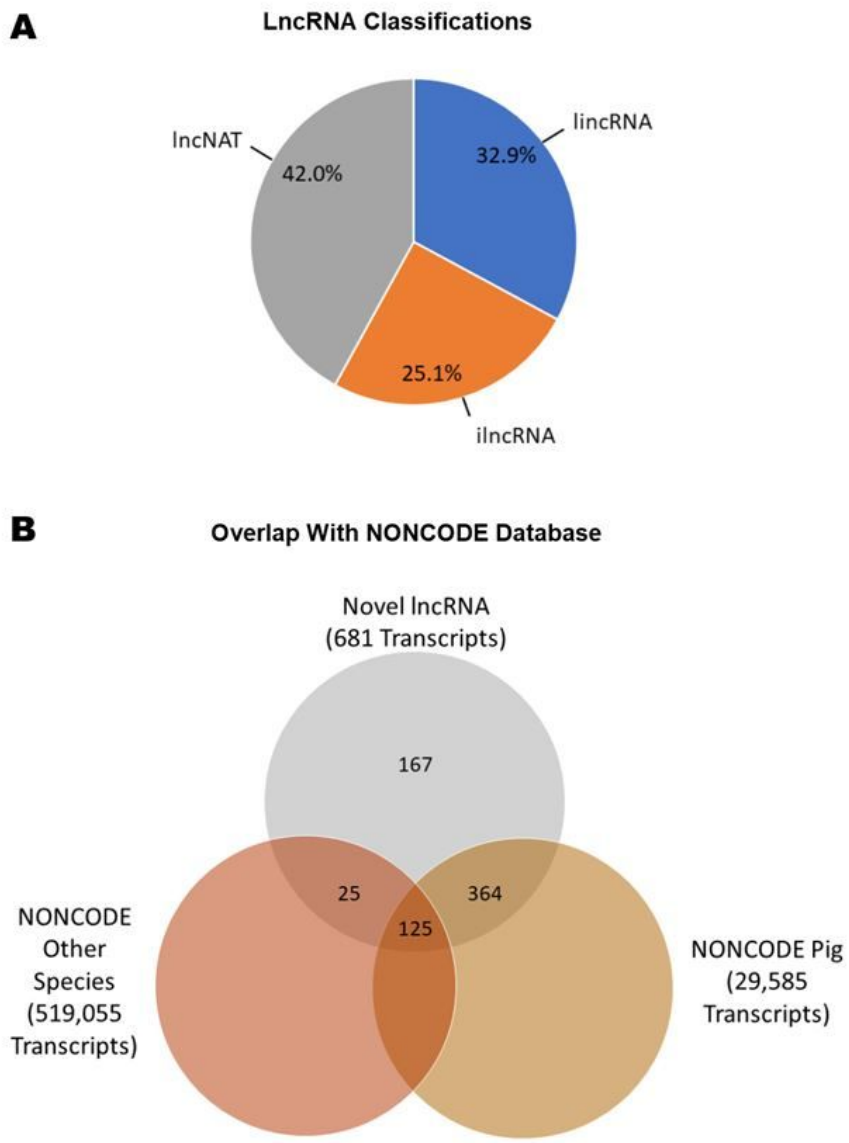

Figure 3

Classification of novel lncRNA.

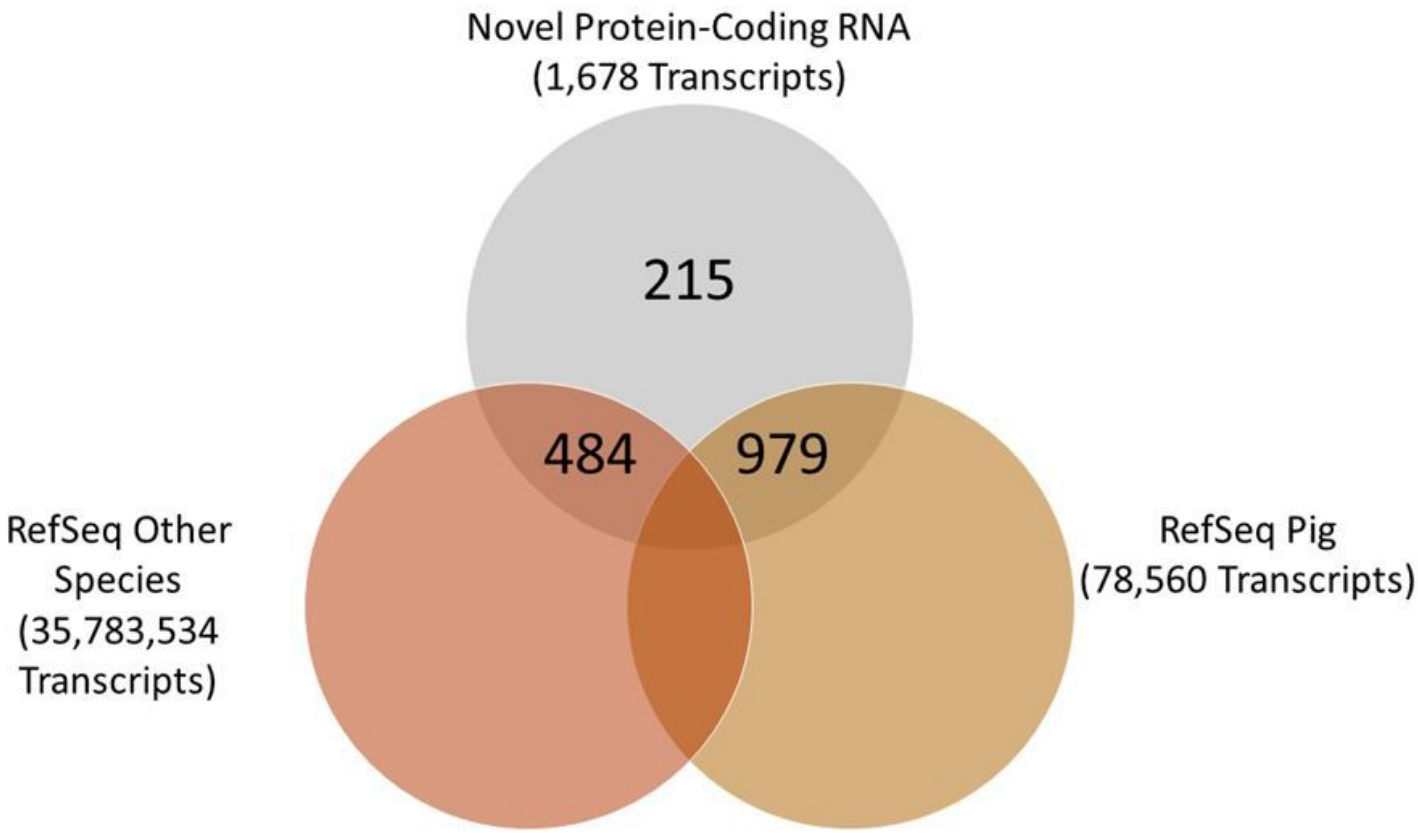

Figure 4

Overlap of novel protein-coding transcripts with RefSeq database.

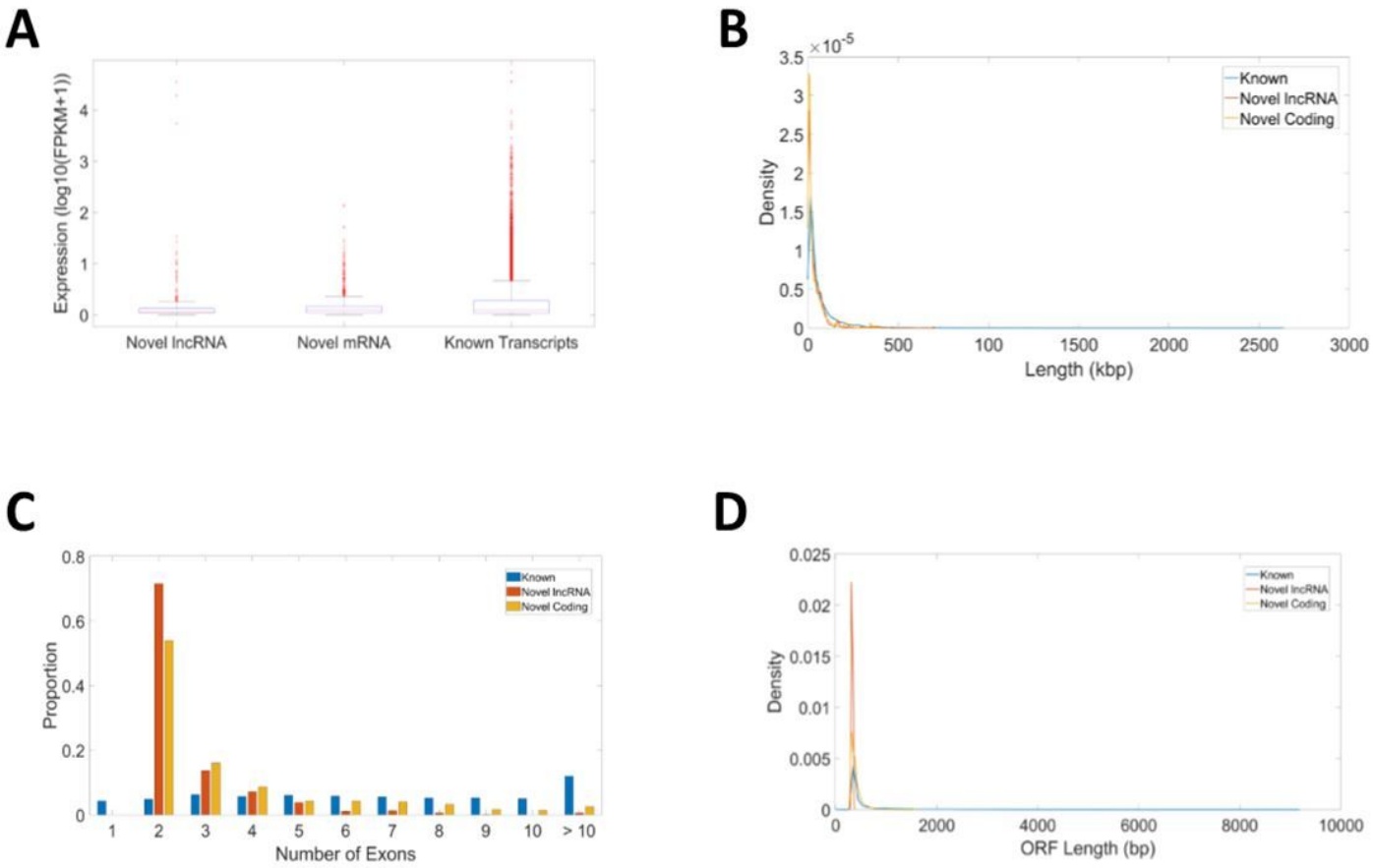

Figure 5

Basic features of transcripts. (A) Expression level of transcripts. (B) Length distribution of transcripts. (C) Number of exons for transcripts. (D) ORF length distribution of transcripts.

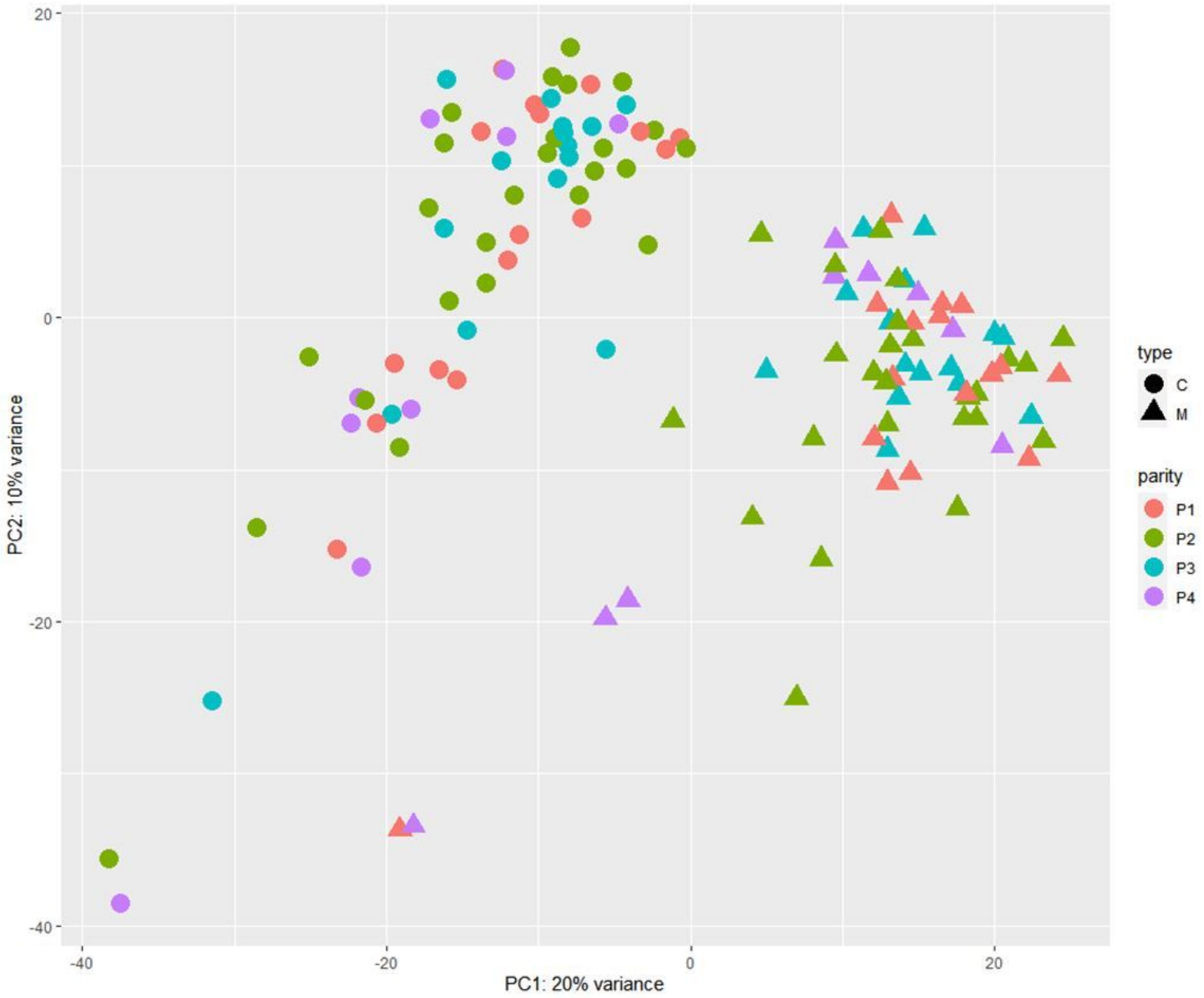

Figure 6

PCA plot of colostrum (C) and mature milk (M) transcripts from dams in parities 1-4.

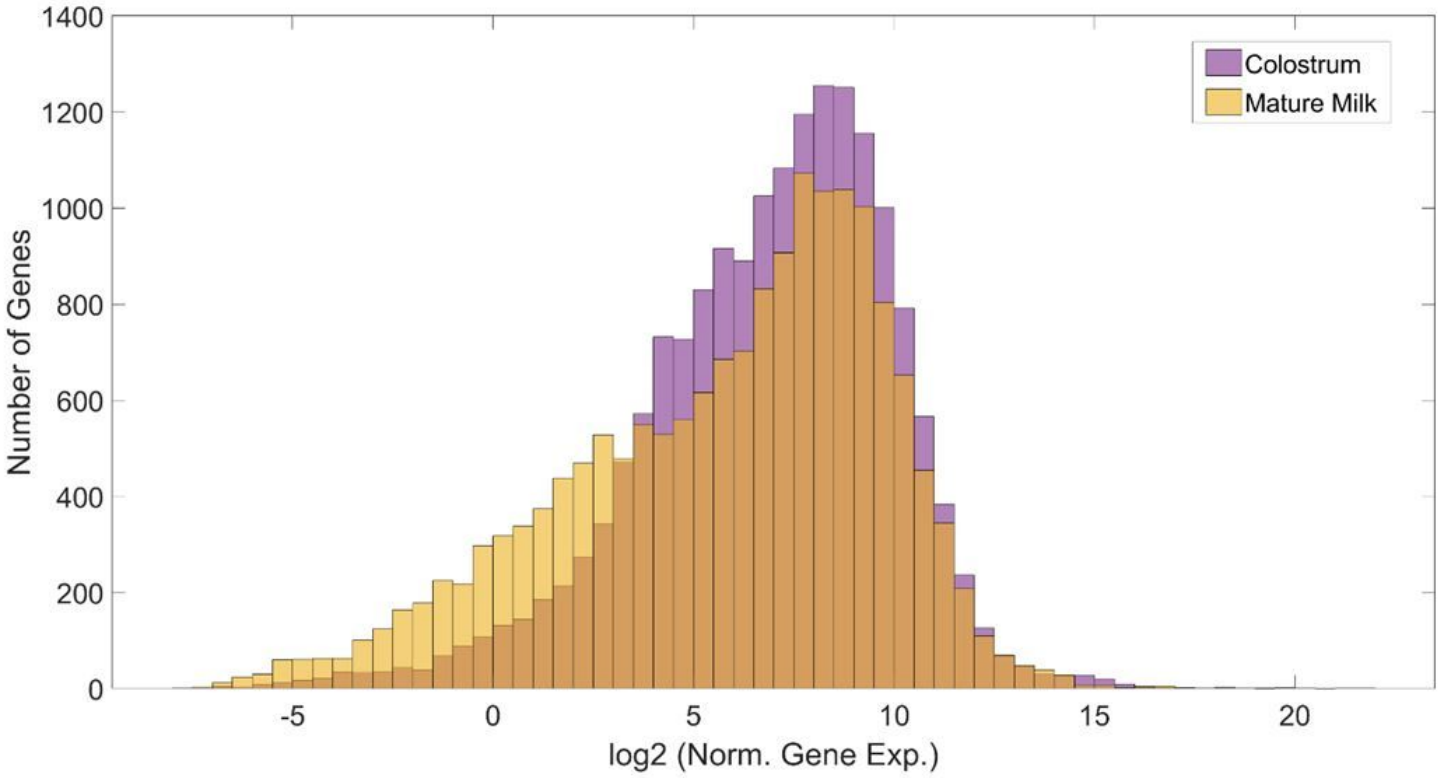

Figure 7

Plot of gene expression distribution for colostrum and mature milk samples. Values are averaged across samples in each group.

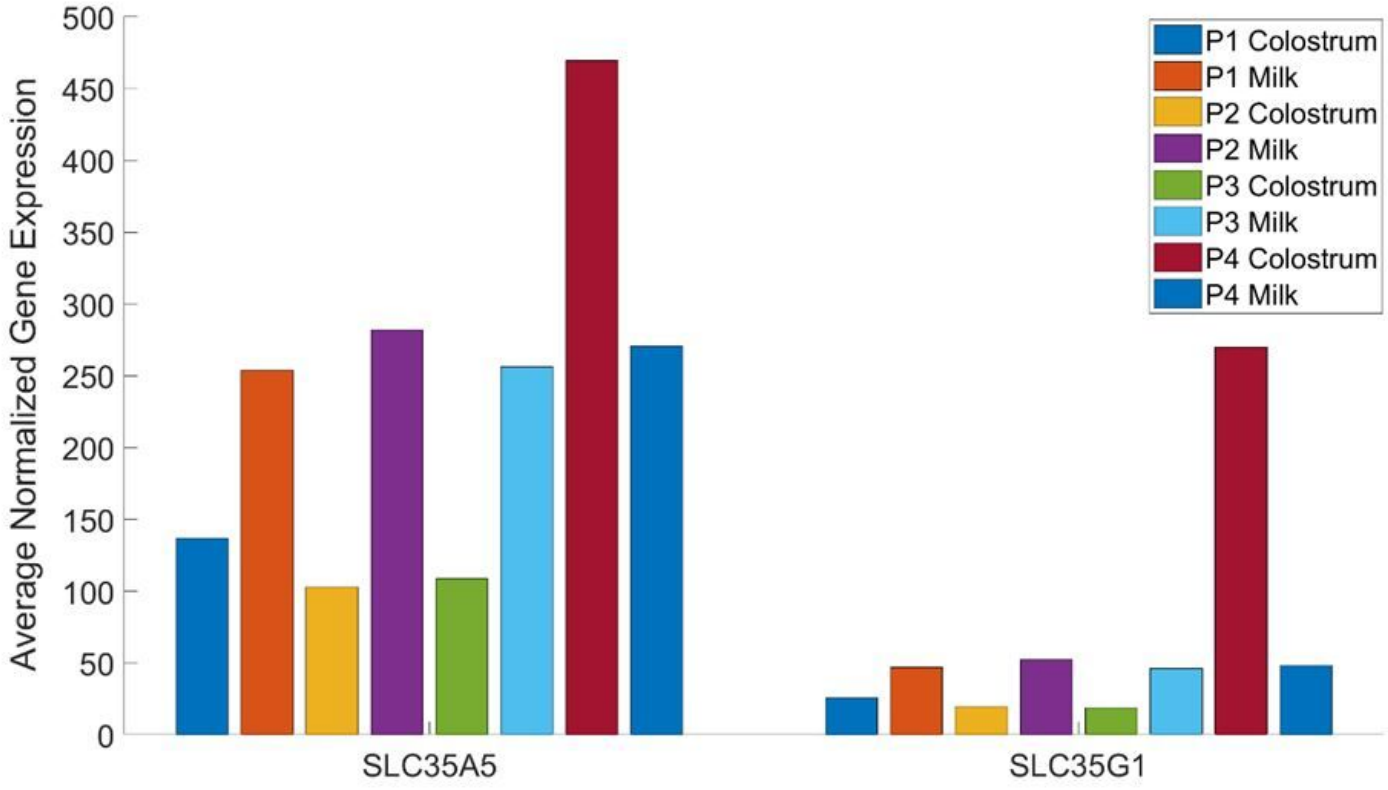

Figure 8

Average gene expression of DEG in the SLC35 family that were significant for the milk x parity interaction.

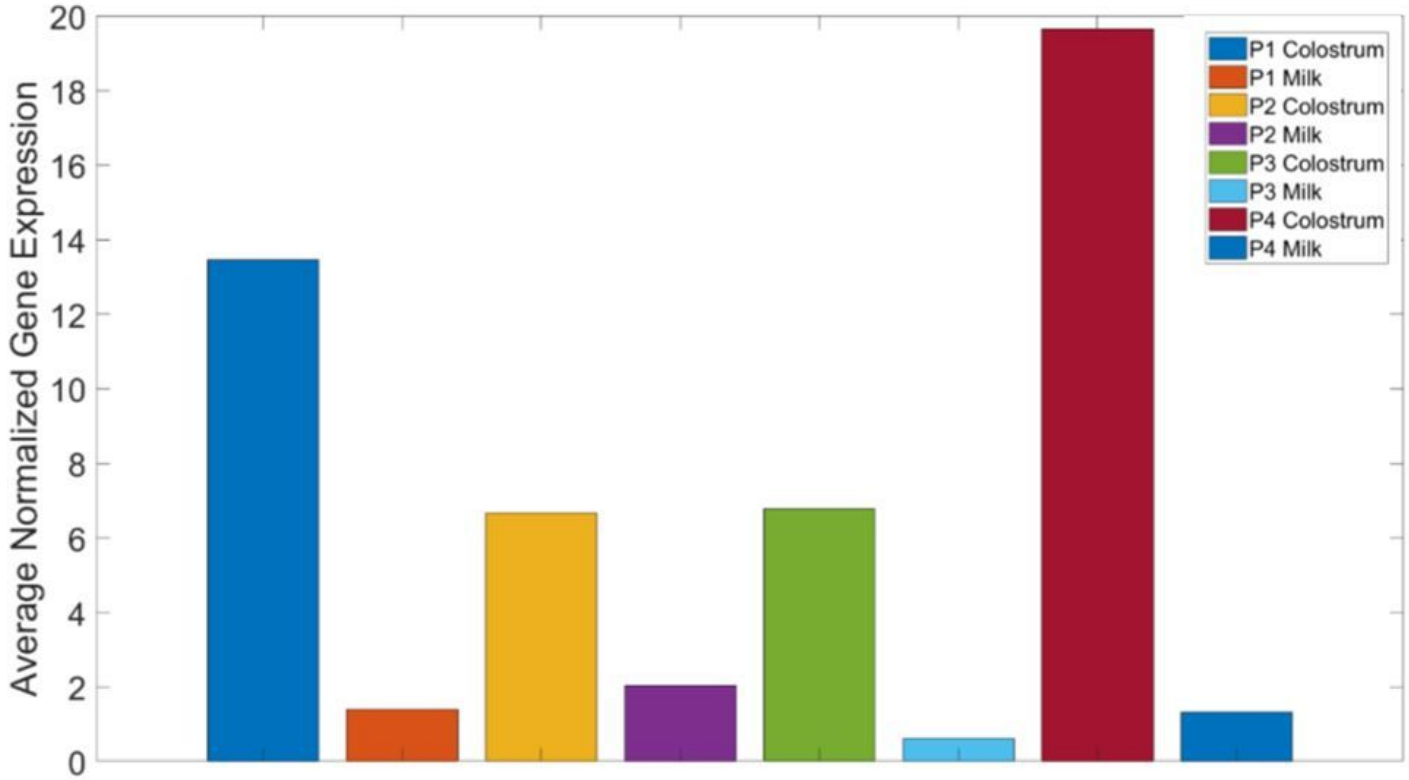

Figure 9

Average gene expression values of JAK1 gene.

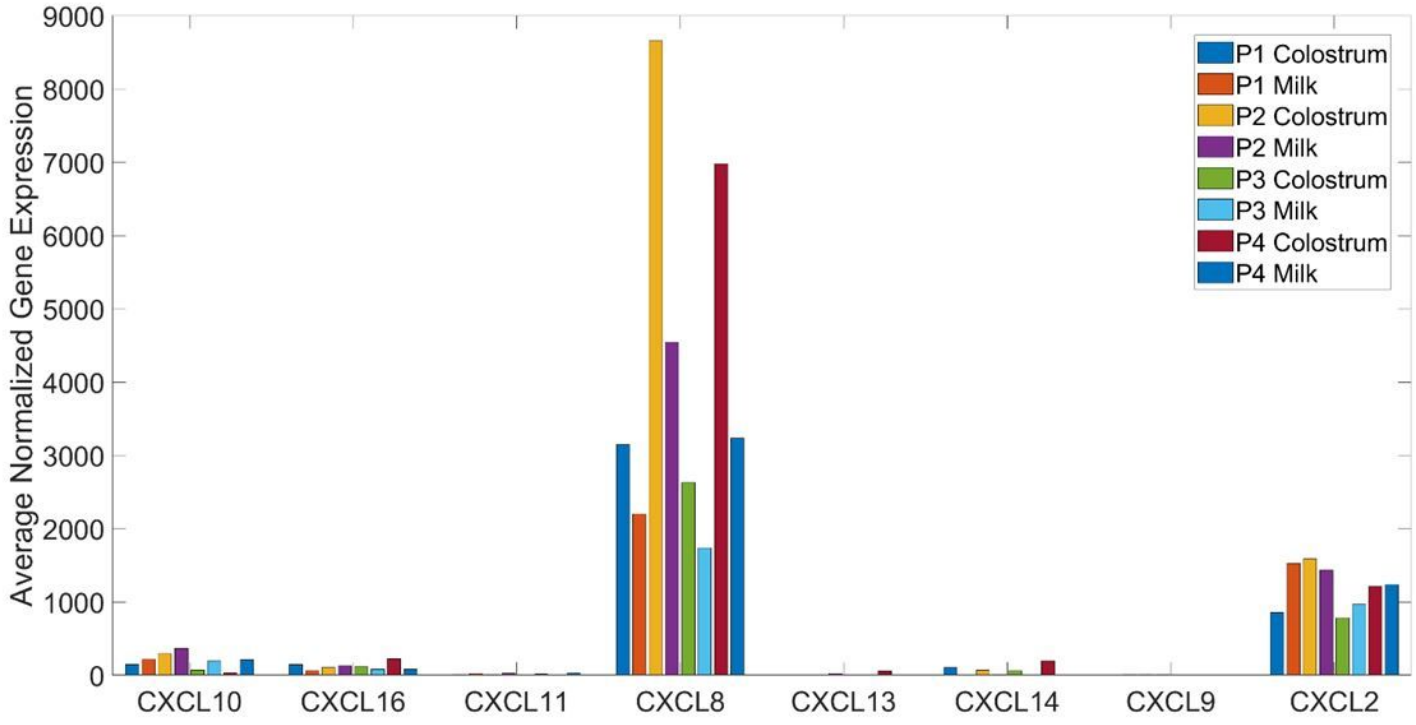

Figure 10

Average gene expression values of genes in chemokine superfamily (CXC).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS11.xlsx
- TableS10.xlsx
- TableS9.xlsx
- TableS8.xlsx
- TableS7.xlsx
- TableS6.xlsx
- TableS5.xlsx
- TableS4.xlsx
- TableS3.xlsx
- TableS2.xlsx
- TableS1.xlsx
- FigureS1.docx