Integrated analysis of long noncoding RNA and mRNA expression profiles reveals the potential role of long noncoding RNA in different bovine lactation stages

X. Zheng,[1] C. Ning,[1] P. Zhao, W. Feng, Y. Jin, L. Zhou, Y. Yu, and J. Liu[2]
National Engineering Laboratory for Animal Breeding, Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture,
College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

ABSTRACT

Long noncoding RNA (lncRNA) play a critical role in mammary development and breast cancer biology. Despite their important role in the mammary gland, little is known of the roles of lncRNA in bovine lactation, particularly regarding the molecular processes underlying it. To characterize the role of lncRNA in bovine lactation, 4 samples of Holstein cow mammary gland tissue at peak and late lactation stages were examined after biopsy. We then profiled the transcriptome of the mammary gland using RNA sequencing technology. Further, functional lncRNA-mRNA coexpression pairs were constructed to infer the function of lncRNA using a generalized linear model, followed by gene ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. More than 1,000 putative lncRNA were identified, 117 of which were differentially expressed between peak and late lactation stages. Bovine lncRNA were shorter, with fewer exon numbers, and expressed at significantly lower levels than protein-coding genes. Seventy-two differentially expressed (DE) lncRNA were coexpressed with 340 different protein-coding genes. The KEGG pathway analysis showed that target mRNA for DE lncRNA were mainly related to lipid and glucose metabolism, including the peroxisome proliferator-activated receptors and 5′ adenosine monophosphate-activated protein kinase signaling pathways. Further bioinformatics and integrative analyses revealed that 12 DE lncRNA potentially played important roles in bovine lactation. Our findings provide a valuable resource for future bovine transcriptome studies, facilitate the understanding of bovine lactation biology, and offer functional information for cattle lactation.

Key words: long noncoding RNA, bovine lactation, mammary gland, coexpression, RNA sequencing

INTRODUCTION

Transcriptional studies indicated that a large proportion of eukaryotic transcripts do not code for proteins (Kapranov et al., 2002; Cheng et al., 2005; Cabili et al., 2011). Generally, nonprotein coding RNA are defined by the size of mature transcripts, which includes small/short nonprotein coding RNA (<200 nucleotides) and long noncoding RNA (**lncRNA**; >200 nucleotides). Micro RNA (miRNA) are the most widely known small/short nonprotein coding RNA and are known to regulate the expression of targeted genes (Gindin et al., 2015; Mogilyansky et al., 2016). In the past few years research interest on lncRNA has rapidly emerged and, accordingly, an increasing number of lncRNA have been revealed in different species, including *Homo sapiens* (McHugh et al., 2015), rat (Gopalakrishnan et al., 2015), goat (Ren et al., 2016), pig (Wang et al., 2016), chicken (He et al., 2016), and bovine (Sun et al., 2016). Previous studies revealed the epigenetic roles of lncRNA (Gindin et al., 2015) in X chromosome inactivation (Sado et al., 2005; Ogawa et al., 2008) and imprinting (Mancini-Dinardo et al., 2006), as well as their regulatory action (Mo et al., 2017) at the transcriptional and posttranscriptional levels during zebrafish embryogenesis development (Pauli et al., 2012), which is associated with diseases (Gutschner and Diederichs, 2012; He et al., 2016).

Holstein cows are mainly used for milk production. Lactation is a dynamic and multifactorial complex process that involves the development of the mammary gland and the synthesis and secretion of milk (McManaman and Neville, 2003). Knowledge about lactation regulation is crucial for understanding the mechanisms of basic cellular processes involved in lactation (Gold-

man et al., 1998), which may have important implications for improving milk yield and quality and disease status but also for research purposes.

Currently, the main strategy for exploring the physiological mechanisms underlying different lactation stages focuses on the detection of the expression and function of coding genes and related biological processes (Bionaz and Loor, 2008a, 2011; Bionaz et al., 2012; Wickramasinghe et al., 2012), and some genes, such as *ASCL1*, *FABP3*, *LPIN1*, *SLC27A6*, and *AGPAT6*, were reported to be involved in cow lactation (Bionaz and Loor, 2008a). However, recent studies suggested that lncRNA are key regulators of mammary gland development and lactation processes (Askarian-Amiri et al., 2011; Standaert et al., 2014; Sandhu et al., 2016). Although bovine lncRNA have been identified from expressed sequence tag or RNA sequencing (**RNA-seq**) data (Huang et al., 2012a; Tong et al., 2017), little is known about the biological function and role of lncRNA in mammary glands during the different lactation stages. To identify lncRNA with potential functions during lactation and to examine the transcriptome landscape of mammary glands in 2 important lactation stages in Holstein cows, we performed Ribo-Zero RNA-seq analysis using the HiSeq 2500 platform (Illumina, San Diego, CA).

MATERIALS AND METHODS

Ethics Statements

All animal experiments were performed according to protocols approved by the Institutional Animal Care and Use Committee of China Agricultural University (permit no. DK1023; Beijing, People's Republic of China). All experiments were conducted in compliance with the regulations and guidelines established by that committee.

Experimental Animals

Four multiparous, healthy, and mastitis-free Chinese Holstein cows were selected from Zhonghe Dairy Farm (Taiyuan, Shanxi Province, China) for the study. Two of the cows were in their second parity and the other 2 were in their third parity when the mammary gland tissues were collected. The 305-d milk yield and fat and protein percentages in the milk of the 4 cows were 6,832 to 8,028 kg, 3.64 to 4.18%, and 3.36 to 3.41%, respectively. These cows were kept in freestall housing in the same cubicle partition of the barn during the experimental period, were all fed the same TMR diet, and had access to water ad libitum. They were milked 3 times daily in the milking parlor.

Sample Collection

Using the same 4 Chinese Holstein cows, we performed mammary gland tissue biopsies at 2 lactation stages: the peak stage (90 d postpartum) and the late stage (270 d postpartum). The biopsy experiment was performed according to previously described methods (Schmitz et al., 2004; Li et al., 2016). The selected biopsy site of the same mammary area for 4 cows was shaved and disinfected using 75% ethanol, anaesthetized with Su-Mian-Xin (846 compound anesthetic agent, 30–40 mg i.v./cow; China Agricultural University Veterinary Teaching Hospital, Beijing, China), and injected subcutaneously with 1 mL of procaine (China Agricultural University Veterinary Teaching Hospital). Then, a small incision (~1.5 cm) was made on the skin at the midpoint of a rear quarter of the mammary gland and connective tissue was blunt-dissected away using shears and tweezers, exposing the secretory tissues. The mammary gland secretory tissue biopsy (~100 mg) was then obtained and stored in liquid nitrogen until protein isolation. The suture was tied as the cannula was removed, and pressure was applied to reduce the collection of blood under the skin. Upon completion of the experiment, all 4 cows immediately received antibiotic prophylaxis and anti-inflammatory therapy. The biopsy wounds healed and the milk yield returned to prebiopsy values within 3 d. It has been demonstrated that dairy cows recover rapidly postbiopsy, with no negative effects on their health (Lima et al., 2016).

Illumina Deep Sequencing

Total RNA was isolated using the standard TRIzol (Invitrogen, Carlsbad, CA) protocol. The quality and concentration of the preparation were assessed using a NanoDrop system (Agilent Technologies, Santa Clara, CA) and on a 1% agarose gel, and the integrity was measured using an Agilent 2100 Bioanalyzer (Agilent Technologies) as previous reported (Kiewe et al., 2009).

For each sample, 3 μg of total RNA was used as input material for the RNA sample preparation. Sequencing libraries were generated using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, Ipswich, MA) according to the manufacturer's recommendations, and index codes were added to attribute the sequences to each sample. In brief, Ribo-Zero (Epicenter, Madison, WI) fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5×; NEB). First-strand cDNA was synthesized using a random hexamer primer and M-MuLV reverse transcriptase (RNase H). Second-strand cDNA synthesis

was subsequently performed using DNA polymerase I and RNase H. In the reaction buffer, deoxynucleotide triphosphates (dNTP) with dTTP were replaced with dUTP. Remaining overhangs were converted into blunt ends through exonuclease/polymerase activities. After adenylation of the 3′ ends of the DNA fragments, ligation with NEBNext Adaptor containing a hairpin loop structure was carried out to prepare for hybridization. To select cDNA fragments of preferentially 150 to 200 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, MA). Then, 3 µL of User Enzyme (NEB) was incubated with the size-selected adaptor-ligated cDNA at 37°C for 15 min, followed by 5 min at 95°C, before the PCR. The PCR was then performed with Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, Basingstoke, UK), universal PCR primers, and index (X) primer. Finally, the products were purified (AMPure XP system, Beckman Coulter) and the library quality was assessed on an Agilent Bioanalyzer 2100 system (Agilent Technologies).

Clustering of the index-coded samples was performed on a cBot Cluster Generation system using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina HiSeq 2500 platform, and 150-bp paired-end reads were generated.

Bioinformatics Analysis

Quality Control. Raw data (raw reads) of fastq format were first cleaned by self-developed Perl scripts. In this step, clean data (clean reads) were obtained by removing adapter-containing, ploy-N-containing, and low-quality reads from the raw data. At the same time, the Q20, Q30, and GC contents of the clean data were calculated. All the downstream analyses were based on the high-quality clean data.

Mapping and Transcriptome Assembly. The clean data were aligned to the bovine UMD 3.1 reference genome by TopHat2 (Kim et al., 2013). Cufflinks software was used to assemble the transcripts (Trapnell et al., 2012). Transcripts from all samples were then merged together with Cuffmerge (Trapnell et al., 2012) to build a consensus set of transcripts across samples.

Pipeline of lncRNA Identification. The lncRNA was identified by the following steps: (1) the known protein-coding genes from the Ensembl database were removed with Cuffcompare (Trapnell et al., 2010); (2) transcripts with a single exon and a length of less than 200 bp were removed; (3) transcripts with protein-coding potency were removed with CNCI (Sun et al.,

2013); (4) transcripts that had similarity to known proteins from the UniProt database were removed by BLASTX (Altschul et al., 1997); and (5) any transcript with a maximal open reading frame of more than 100 AA was excluded.

Comparative Sequence Analysis. We compared the positions of identified lncRNA with those of previously identified bovine lncRNA that are publicly available in the Noncode 2016 database (Zhao et al., 2016). In addition, comparisons between our bovine mammary sequences and the human sequences were conducted with the BLAST software suite (Altschul et al., 1990), using thresholds of 50% for the query coverage and less than 1e-06 for the *E*-value.

Expression Analysis. The Python script htseq-count included in HTSeq (Anders et al., 2015) was used to count reads that overlapped any exon for each lncRNA and mRNA. The fragments per kilobase of transcript per million mapped reads value was used to evaluate the lncRNA and gene expression, which used the effective library size to replace the original library size (Trapnell et al., 2010). To determine differentially expressed lncRNA, edgeR (Robinson et al., 2010) software was used, applying a generalization of the paired *t*-test for paired samples. Those lncRNA with fold changes above 2.0 and a *P*-value of less than 0.05 were assigned as being differentially expressed at different lactation stages.

Coexpression Analysis. The coexpression of lncRNA and mRNA was analyzed using edgeR, applying a generalized linear model to regress mRNA expression on lncRNA expression and adjusting for lactation stage effects and baseline differences among different cows. The generalized linear model in edgeR is

$$Y_{mRNA} = \beta_{cows} + \beta_1 X_{lncRNA} + \beta_2 X_{lactation\ stage},$$

where β_{cows} is the baseline (intercept) of different cows; β_1 is the lncRNA effect on mRNA; β_2 is the effect of the lactation stage; X_{lncRNA} and $X_{lactation\ stage}$ are the corresponding design matrices; and Y_{mRNA} follows negative binomial dispersion. An lncRNA was considered to influence the expression of an mRNA (coexpression) if β_1 deviated significantly from zero. To reduce false positives, Bonferroni correction was applied to control for multiple testing, and a stringent *P*-value threshold ($P < 2.12E\text{-}09$) was applied to declare statistical significance.

Functions of lncRNA in Bovine Mammary Gland

Considering the limitation of gene ontology (**GO**) and Kyoto Encyclopedia of Genes and Genomes (**KEGG**)

ZHENG ET AL.

in annotating genes in the bovine genome, we had converted the cow refGene/Ensembl gene IDs to orthologous human Entrez Gene IDs using BioMart (http://www.ensembl.org/biomart/). To infer the function of the lncRNA, their coexpressed mRNA were subjected to GO enrichment using KOBAS 3.0 software (http://kobas.cbi.pku.edu.cn/anno_iden.php). A *P*-value of less than 0.05, as determined by Fisher's exact test, was set as the criterion for significance. To explore the relationship of differentially expressed lncRNA (**DE-lncRNA**) with milk composition and milk production traits, we integrated them with previously reported QTL mapping and genome-wide association study data (http://www.animalgenome.org/cgi-bin/QTLdb). The Panther classification system (http://www.pantherdb.org/) was used to identify the function class of the total mRNA and differentially expressed genes (**DEG**).

Analysis of the Relationship of lncRNA with miRNA

We detected the binding sites of miRNA in the lncRNA promoter region by using miRanda software, applying an empirical alignment score of 160 and minimum free energy of −20 kcal/mol (Jalali et al., 2013). The miRNA data were obtained from the miRBase database (Griffiths-Jones et al., 2006).

Verification of Gene Expression Profiles Using Quantitative PCR

For validation of the gene expression data obtained by RNA sequencing between the 2 lactation periods, quantitative PCR (**qPCR**) was performed on 13 select DEG and 5 lncRNA using the total RNA used for RNA-seq. Total RNA was purified and reverse transcribed into cDNA using the PrimeScript RT Reagent Kit with gDNA Eraser [Takara Biotechnology (Dalian) Co. Ltd., Dalian, China] according to the manufacturer's instructions. The mRNA quantities were then measured by qPCR using a LightCycler 480 Real-Time PCR System (Roche, Hercules, CA). The qPCR assays were performed in a total volume of 20 μL containing 10 μL of SYBR Green mixture, 7 μL of deionized water, 1 μL of cDNA template, and 1 μL of each primer. The thermal conditions were as follows: 95°C for 5 min, followed by 45 cycles of 95°C for 10 s, 60°C for 10 s, and 72°C for 10 s. The mRNA levels of the DEG were normalized to the geometric mean of the data obtained for the endogenous control genes *GAPDH*, *MARVELD1*, and *LRP10* that were stably expressed in bovine mammary gland tissue (Kadegowda et al., 2009; Saremi et al., 2012). The specific primer sequences used for the qPCR assays are presented in Supplemental Table S1 (https://doi.org/10.3168/jds.2018-14900). Each qPCR assay was carried out in triplicate. The relative gene expression level was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

The quantitative expression data are presented as the mean ± standard error of the mean of 3 independent experiments. In each experiment, the assay was performed in triplicate. Statistical significance of the results was tested using the paired *t*-test.

RESULTS

Overview of RNA-Seq and Identification of lncRNA and Protein-Coding Genes in the Mammary Gland of Chinese Holstein Cows

Changes in the expression of lncRNA during peak and late lactation stages were examined based on the transcriptomes of Chinese Holstein cow mammary gland samples obtained from Ribo-Zero RNA-seq. Overall, 754.76 million clean reads were obtained, with an average of 94.35 million reads (ranging from 77.86 to 115.11) for each sample. Average quality values for sequencing were 94.38% (Q20) and 87.23% (Q30; Supplemental Table S2, https://doi.org/10.3168/jds.2018-14900). Approximately 86.86% of the total reads were uniquely mapped to the UMD 3.1 reference genome (http://www.ncbi.nlm.nih.gov/genome/guide/cow/index.html). There were 3.28% nonspecifically mapped reads and 9.86% unmapped reads. Only the uniquely mapped reads were considered in further analysis. Transcripts of each sample were then assembled in Cufflinks and merged using cuffmerge software, resulting in 917,267 nonredundant transcript isoforms from 831,398 loci in peak and late lactation stages. To identify lncRNA in the bovine mammary gland, transcripts were filtered following the rigorous set of criteria listed in Materials and Methods, and 1,657 lncRNA transcripts from 1,181 candidate lncRNA loci were finally obtained.

To examine the predicted lncRNA, we compared the genomic positions of the 1,181 candidate lncRNA loci with the mapping positions of previously identified bovine lncRNA, publicly available in Noncode 2016. In total, 269 (22.8%) lncRNA identified here overlapped with previously reported bovine lncRNA genes, in agreement with previous lncRNA detection percentages reported for bovine muscle (Billerey et al., 2014). These results suggested that the bovine genome remains poorly annotated and that a large number of new transcript isoforms are still unannotated. In addition, 19.48% (230/1,181) of the predicted lncRNA obtained in this study shared local similarity with human known lncRNA (*E*-value <1e-6).

The 14,207 protein-coding genes identified (fragments per kilobase of transcript per million mapped reads >1) from the bovine mammary gland transcriptome data were classified into 26 RNA categories according to their function. As expected, they displayed a high diversity of functionalities (Supplemental Figure S1; https://doi.org/10.3168/jds.2018-14900), including nucleic acid binding (1,247 genes), hydrolase (964 genes), transcription factor (730 genes), and transporter (481 genes). Bionaz et al. (2012) detected the mammary gland transcriptome at 6 points relative to parturition and indicated that the mammary gland increased overall metabolism (e.g., large increase in carbohydrate, lipid metabolism during lactation). Those events were coupled with a large increase in membrane transport and hydrolase activity, which is in agreement with our results. Li et al. (2016) explored the bovine transcriptome from the mammary tissue of Chinese Holstein cows at peak lactation and the nonlactating period, and the total genes identified were similar to the results from our study.

Genomic Features of Bovine Mammary Gland lncRNA Transcripts

To characterize bovine mammary gland lncRNA, we analyzed their nucleotide composition, expression levels, and structure. We first compared the nucleotide composition among putative lncRNA, mRNA, and random intergenic regions and found that the rank of GC content in mRNA from high to low was exons, 1 kb upstream of exon, and random intergenic regions (Figure 1, middle panel). Such patterns were observed across the different regions of lncRNA (Figure 2, left panel). Our results were consistent with those of previous studies (Huang et al., 2012a; Haerty and Ponting, 2015).

To assess the distribution of bovine putative lncRNA across bovine chromosomes, we analyzed the number of putative lncRNA per millions of base pairs (Mbp) of genome within each individual chromosome. The 1,181 putative lncRNA identified here were not equally distributed across the different bovine chromosomes (Figure 2A). The maximum density of putative lncRNA was found on chromosome 25 (0.65/Mbp), whereas the minimum density was found on chromosome 24 (0.32/Mbp). We also found that the average length of bovine mammary lncRNA was about two-thirds that of protein-coding genes (mean length of 1,435 nucleotides for lncRNA vs. 2,039 nucleotides for protein-coding transcripts; Figure 2B). Moreover, lncRNA had, on average, fewer exons per transcript (~2.43) than protein-coding genes (10.03; Figure 2C). These values

are similar to those estimated for other mammalian lncRNA (Gao et al., 2017). Notably, average expression of bovine mammary lncRNA was more than 4-fold that of protein-coding genes (Figure 2D).

Differential Expression of lncRNA in Mammary Gland Tissue Between Peak and Late Lactation Stages

To investigate key lncRNA and mRNA involved in the bovine lactation process, DE-lncRNA and DEG in the bovine mammary gland were examined at peak and late lactation stages. In total, 117 DE-lncRNA were identified between these 2 lactation stages {$P < 0.05$ at a threshold fold-change value [FC(late lactation/peak lactation)] ≥ 2 or ≤ 0.5; Supplemental Table S3, https://doi.org/10.3168/jds.2018-14900; Figure 3A]. Interestingly, the 117 DE-lncRNA consisted of 6 downregulated and 111 upregulated lncRNA transcripts in late lactation samples compared with peak lactation samples.

We also found 254 DEG between the 2 lactation periods [false discovery rate <0.1 and |(fold change)| ≥ 2; Supplemental Table S4, https://doi.org/10.3168/jds.2018-14900; Figure 3B]. Among these genes, 119 (46.9%) had higher expression level in peak lactation and 135 (53.1%) had higher expression level in late
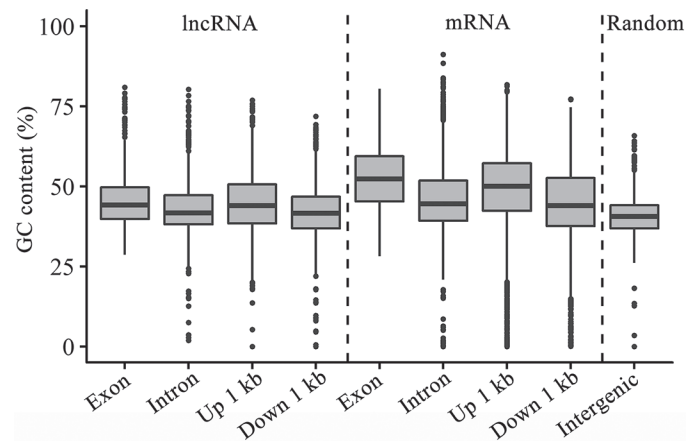

Figure 1. The comparison of GC content between predicted long noncoding RNA (lncRNA) and mRNA. The GC contents of exon, intron, 1,000 bp upstream of 5-flanking regions of lncRNA and mRNA, and random intergenic regions are plotted as box plots. To obtain random intergenic regions, a 1,000-bp window was moved randomly in regions with no protein-coding gene alignment. In total, 1,000 random intergenic regions were used for analysis. The center line represents the median; the top and bottom of the box are quartile 3 and 1, respectively; the upper whisker is the minimum value of quartile 3 + 1.5 × interquartile range (IQR) and the maximum value of data; the lower whisker is the maximum value of quartile 1 − 1.5 × IQR and the minimum value of data. Here, quartile 1 means the 25th percentile; quartile 3 means the 75th percentile; IQR means interquartile range and is equal to quartile 3 to quartile 1. A value less than the lower whisker or more than the upper whisker is represented as a black dot.

lactation. The GO analysis of DEG in peak lactation showed that these were mostly involved in cellular process, metabolic process, and biological regulation (biological process GO category) and in binding, catalytic activity, and transporter activity (molecular function GO category). On the other hand, genes upregulated in late lactation played major roles in localization, apoptosis, process system development, and cell adhesion (biological process) and in hydrolase activity, receptor activity, and transporter activity (molecular function).

Coexpression of lncRNA and mRNA in Bovine Mammary Gland Tissue

Increasing evidence has revealed that lncRNA have an important role in regulating the expression of protein-coding genes (Guo et al., 2015). Therefore, we applied bioinformatics analysis to construct lncRNA-mRNA coexpression pairs. Among the 76,259 lncRNA-mRNA pairs that were significantly coexpressed in mammary gland tissue, 2,573 (3.37%) were located in
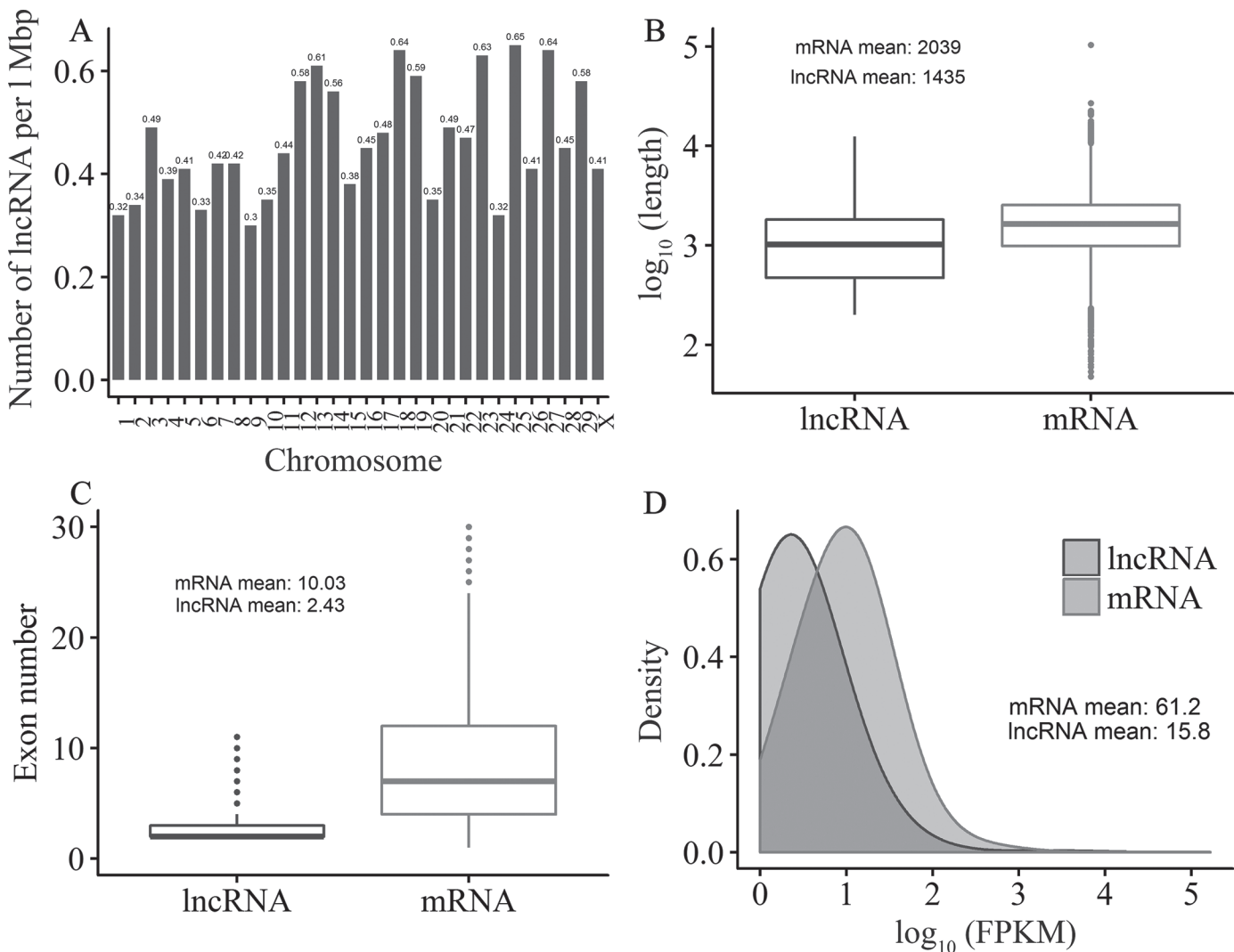

Figure 2. Genomic features of identified long noncoding RNA (lncRNA). (A) The lncRNA are not equally distributed on different bovine chromosomes. (B) The transcript length of lncRNA was shorter compared with protein-coding genes. (C) The lncRNA have fewer exons than protein-coding genes. (D) LncRNA expressed at lower levels than protein-coding genes. FPKM = fragments per kilobase of transcript per million mapped reads. The center line represents the median; the top and bottom of the box are quartile 3 and 1, respectively; the upper whisker is the minimum value of quartile $3 + 1.5 \times$ interquartile range (IQR) and the maximum value of data; the lower whisker is the maximum value of quartile $1 - 1.5 \times$ IQR and the minimum value of data. Here, quartile 1 means the 25th percentile; quartile 3 means the 75th percentile; IQR means interquartile range and is equal to quartile 3 to quartile 1. A value less than the lower whisker or more than the upper whisker is represented as a black dot. Color version available online.

Table 1. Top 5 long noncoding RNA (lncRNA) and mRNA with the highest numbers of associations in coexpression analysis

lncRNA ID	lncRNA chromosome	No of. mRNA	mRNA ID	mRNA chromosome	No. of. lncRNA
XLOC_000752	1	251	ENSBTAG00000040279	19	361
XLOC_274111	18	249	ENSBTAG00000020582	25	318
XLOC_412271	23	244	ENSBTAG00000016554	1	308
XLOC_226575	16	228	ENSBTAG00000015525	13	300
XLOC_667393	7	224	ENSBTAG00000017593	23	300

the same chromosome (*cis*-acting), whereas the remaining (96.63%) were located on different chromosomes (*trans*-acting). This suggested that most of the lncRNA *trans*-regulate mRNA expression. Such regulation pattern was also observed in a human cancer-related study (Wu et al., 2016).

The 76,259 lncRNA and mRNA coexpression pairs consisted of 906 unique lncRNA and 354 unique mRNA. Among the 906 lncRNA, 72 were differentially expressed between peak and late lactation and showed coexpression with 340 different mRNA (Supplemental Table S5; https://doi.org/10.3168/jds.2018-14900). Fourteen lncRNA (19.4%) showed differential coexpression with a single mRNA, whereas 58 (80.56%) lncRNA showed differential coexpression with at least 2 mRNA. The top 5 co-expressed lncRNA and mRNA are shown in Table 1.

Functional Characterization of the Identified lncRNA

To better understand the function of the lncRNA identified in this study, 3 aspects were taken into account: (1): annotation of their coexpressed mRNA; (2) comparison of their positions with cattle QTL; and (3) identification of miRNA binding sites in the lncRNA promoter region. To infer the biological function of the identified lncRNA, their coexpressed mRNA were annotated according to GO and KEGG databases. The 340 mRNA showing differential coexpression with lncRNA were significantly enriched in glycolysis and gluconeogenesis ($P = 6.25\text{E-}06$), peroxisome proliferator-activated receptors (**PPAR**; $P = 1.11\text{E-}05$), and 5′ AMP-activated protein kinase (**AMPK**) signaling pathways ($P = 1.20\text{E-}05$). The most relevant KEGG pathways in the lactation process are shown
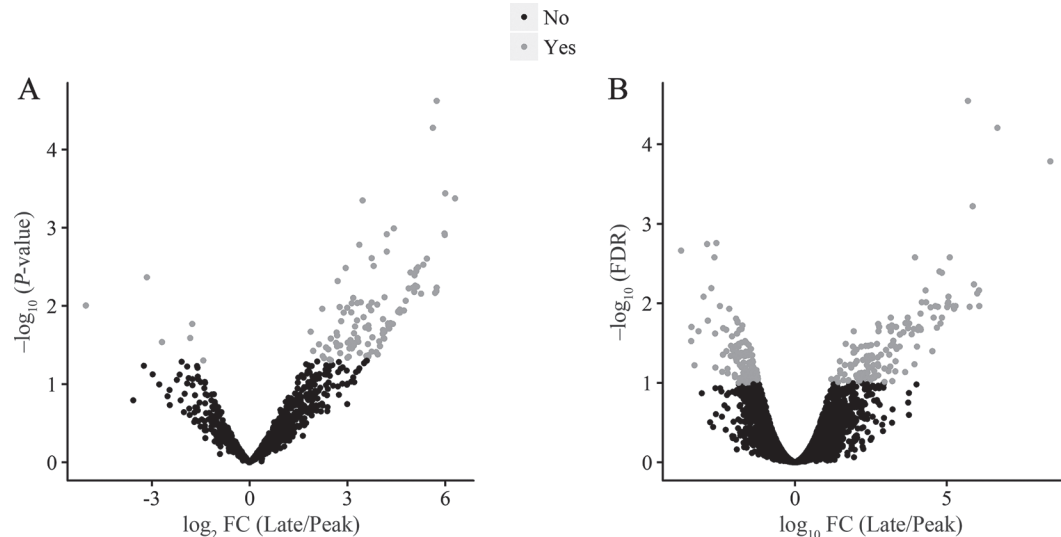

Figure 3. (A) The differentially expressed long noncoding RNA (lncRNA) in mammary tissues between peak and late lactation stages. The x-axis shows $\log_2$ of the fold-change (FC) differences in expression, and the y-axis displays the $-\log_{10}$ P-value. The gray (red) dots represent the significantly differentially expressed lncRNA ($P < 0.05$); the black dots represent the transcripts whose expression levels did not reach statistical significance ($P > 0.05$). (B) The differentially expressed protein-coding genes in bovine mammary tissues between peak and late lactation stages. Each point in the figure represents an mRNA. The x-axis shows $\log_2$ of the FC differences in expression, and the y-axis displays the $-\log_{10}$ false discovery rate (FDR)–adjusted P-value. The gray (red) dots represent the significantly differentially expressed genes (FDR <0.05); the black dots represent the genes whose expression levels did not reach statistical significance (FDR >0.05). Color version available online.

in Supplemental Table S6 (https://doi.org/10.3168/jds .2018-14900), and the overrepresented pathways in 2 lactation stages are shown in Figure 4. Notably, KEGG data indicated that the coexpressed mRNA for lncRNA XLOC_000752, XLOC_306924, XLOC_274111, XLOC_517858, XLOC_518578, XLOC_555176, and XLOC_626085 were enriched in metabolic pathways such as AMPK, mechanistic target of rapamycin, PPAR signaling pathway, and fatty acid metabolic pathway.

We also compared the locations of lncRNA with that of known bovine QTL (http://www.animalgenome .org/QTLdb/, release 33; Hu et al., 2016). The cattle QTL database contains 99,652 QTL, representing 574 different traits. Most (1,022) lncRNA were located in 357 QTL, including milking, production, reproduction, health traits, meat, carcass, and exterior traits. Furthermore, 267 (26.13%) lncRNA were located in previously reported milk-related QTL regions (Supplemental Table S7; https://doi.org/10.3168/jds.2018 -14900), and 5 of these lncRNA overlapped with the coexpressed mRNA for lncRNA mentioned above (i.e., XLOC_000752, XLOC_306924, XLOC_274111, XLOC_518578, and XLOC_626085).

Emerging evidence suggests that lncRNA might be directly regulated by miRNA (Cornett and Lutz, 2014), thereby indirectly regulating gene expression by competing with shared miRNA. The relationship between the 17 lncRNA transcripts and 24 miRNA associated with milk traits is shown in Supplemental Table S8 (https://doi.org/10.3168/jds.2018-14900).

Construction of a Putative Metabolic Network Regulated by lncRNA-mRNA Coexpression Pairs

A possible metabolic network underlying lactation processes in peak and late lactation stages in the bovine mammary was constructed. It comprised 26 genes involved in 4 pathways—namely, glycolysis and gluconeogenesis, insulin, PPAR, and AMPK signaling pathways—and these 26 genes were regulated by 18 lncRNA (Figure 5). Among the 18 lncRNA, 12 were associated with milk traits based on the cattle QTL database (Table 2). Six of the 26 target genes were differentially expressed between peak and late lactation stages, including lipoprotein lipase (*LPL*), stearoyl-CoA desaturase (*SCD*), aldehyde dehydrogenase 1 family member A3 (*ALDH1A3*), phosphoenolpyruvate carboxykinase 2 (*PCK2*), PPARG coactivator 1α (*PPARGC1A*), and 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4 (*PFKFB4*; Table 2).

Validation of RNA-Seq Results by qPCR

To confirm the reliability of RNA-seq results, we performed qPCR to validate the expression of DE-lncRNA and DEG. Five DE-lncRNA and 13 DEG were randomly selected for qPCR validation based on the same samples used for RNA-seq. Results indicated that patterns found for RNA-seq data were in excellent agreement with that found for qPCR data (Supplemental Figure S2; https://doi.org/10.3168/jds.2018-14900), confirming the reliability of RNA-seq data.

DISCUSSION

In the present study, we used a longitudinal biopsies design to investigate whole-transcriptome changes in the lactating mammary gland via high-throughput RNA sequencing. By adopting stringent lncRNA identification criteria, more than 1,000 candidate lncRNA were identified in the Chinese Holstein cow mammary gland during peak and late lactation stages. The lncRNA identified here shared similar genomic features with those reported for other mammalian genomes. We found 117 DE-lncRNA and 254 protein-coding genes and constructed lncRNA-mRNA coexpression pairs by using a generalized linear model. Functional enrichment analysis of coexpressed mRNA showed that these were mostly related to lipid and glucose metabolism pathways, including PPAR and AMPK signaling pathways. Integrated interpretation of differential gene expression indicated that 12 lncRNA were the most promising candidate lncRNA genes affecting milk traits. To our knowledge, our study is the first to systematically identify lncRNA from RNA-seq data obtained for the
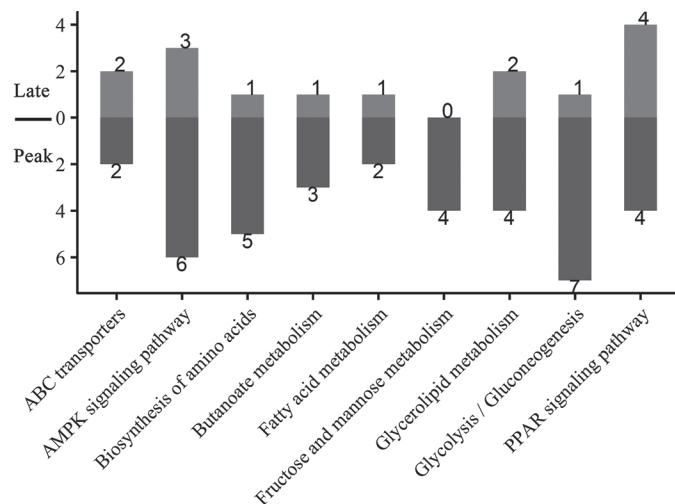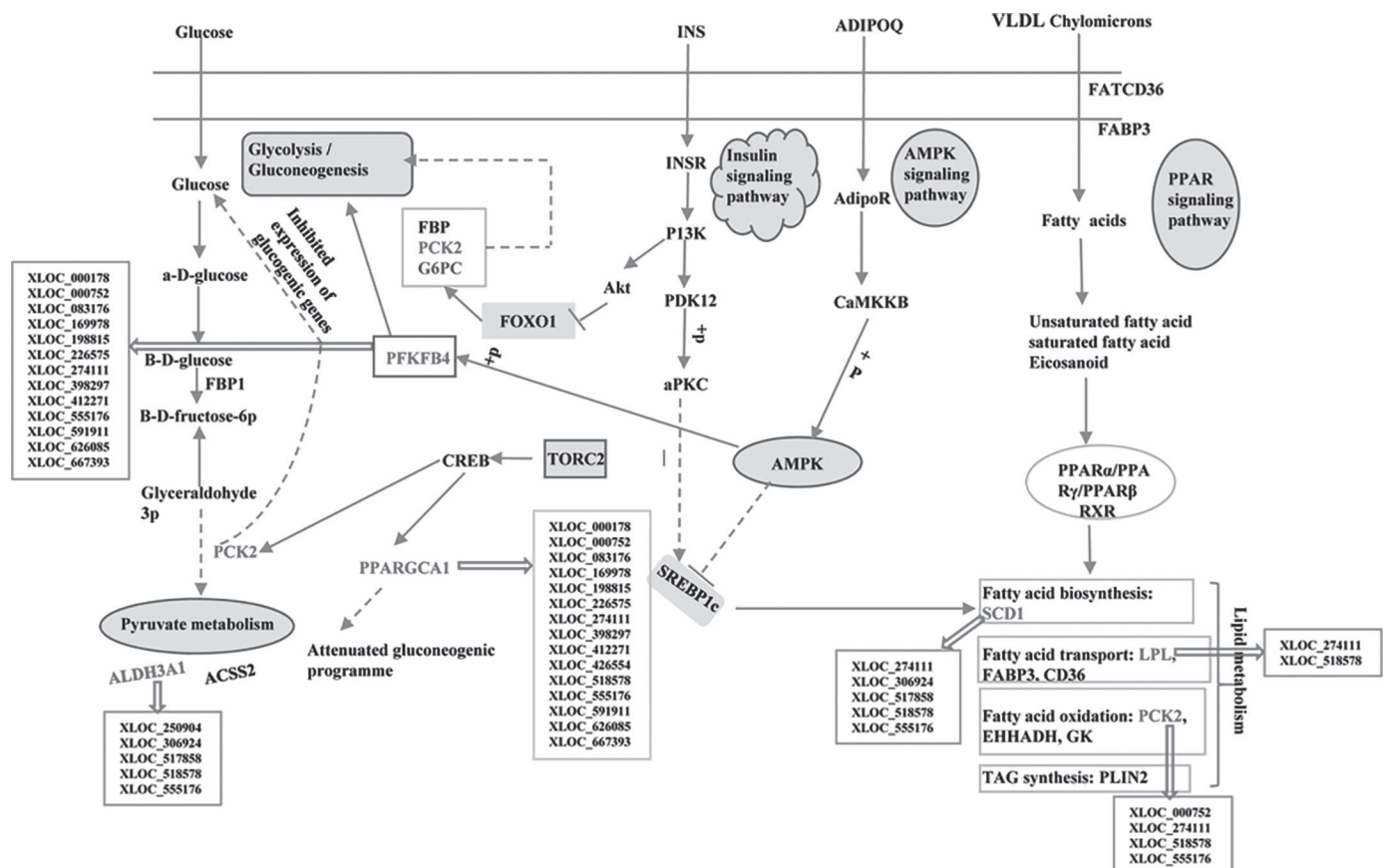

Figure 4. Overrepresented pathways of differentially expressed genes. Gray (red) = late lactation; black (blue) = peak lactation. ABC = ATP-binding cassette; AMPK = AMP-activated protein kinase; PPAR = peroxisome proliferator-activated receptors. Color version available online.

Table 2. Detailed information about 12 candidate long noncoding RNA (lncRNA) in bovine mammary tissue potentially affecting milking performance

lncRNA ID	Target gene symbol	Trait[1]	References
XLOC_000178	*AGPAT1, BDH1, PFKFB4, PPARGC1A*	PY	Daetwyler et al. (2008)
XLOC_169978	*PPARGC1A, BDH1, AGPAT1, PFKFB4, PPARGC1A*	PY, FY	Ashwell et al. (2004), Harder et al. (2006), Lund et al. (2008)
XLOC_198815	*AGPAT1, BDH1, PFKFB4, PPARGC1A*	PP	Schopen et al. (2009)
XLOC_226575	*AGPAT1, BDH1, PFKFB4, PPARGC1A, TSTA3*	MY, PY	Lillehammer et al. (2007)
XLOC_274111	*AGPAT1, LPL, PCK2, BDH1, PFKFB4, PPARGC1A, SCD, TSTA3*	FY, MY, PY	Olsen et al. (2002), Bennewitz et al. (2004), Harder et al. (2006)
XLOC_306924	*SCD, ALDH1A3, BDH1*	FY, PY	Zhang et al. (1998), Lillehammer et al. (2007), Bouwman et al. (2011)
XLOC_398297	*PFKFB4, BDH1, PPARGC1A*	FY, PP	Ashwell et al. (2004), Harder et al. (2006)
XLOC_412271	*AGPAT1, PFKFB4, BDH1, PPARGC1A, TSTA3*	PP	Bennewitz et al. (2003)
XLOC_518578	*ALDH1A3, PCK2, LPL, BDH1, PPARGC1A, SCD*	FP, MY, PY	Heyen et al. (1999), Viitala et al. (2003), Daetwyler et al. (2008)
XLOC_000752	*BDH1, PCK2, PFKFB4, PPARGC1A, TSTA3*	Milk iron content	Buitenhuis et al. (2015)
XLOC_626085	*AGPAT1, BDH1, PFKFB4, PPARGC1A, TSTA3*	Milk protein content	Huang et al. (2012b)
XLOC_250904	*ALDH1A3*	Milk casein percentage	Schopen et al. (2009)

[1]PY = protein yield; FY = fat yield; PP = protein percentage; MY = milk yield; FP = fat percentage.

Figure 5. Putative metabolic map for lipid and lactose in the mammary gland. The genes whose expression changed more than 2-fold are shown in gray (pink). The hollow (red) arrow showed the target relationship between long noncoding RNA (lncRNA) and their predicted target genes. Color version available online.

bovine mammary tissue gland at different lactation stages.

The lncRNA identified in the present study were of high quality. To reduce false positives, we adopted a highly stringent filtering pipeline to remove transcripts evidencing protein-coding potential. As expected, lncRNA identified here were similar to those identified in previous studies, with fewer exon numbers, shorter transcript lengths, lower GC content, and lower expression levels than known protein-coding genes (Huang et al., 2012a). In addition, the number of putative lncRNA detected in the present study was higher than that reported in previous cow mammary gland studies (Tong et al., 2017). More important, the use of longitudinal biopsies allowed the direct observation of changes in gene expression in the mammary gland tissue of identical animals between peak and late lactation stages. This avoided the influence of different genetic backgrounds on DE-lncRNA and DEG detection. Finally, the validation of RNA-seq data by qPCR revealed that both results were consistent. Thus, the data set obtained in the present study may provide a valuable source for the bovine NONCODE database.

An effective method for the construction of lncRNA-mRNA coexpression pairs is presented here. We used a generalized linear model under the negative binomial framework to identify lncRNA-mRNA coexpression pairs. The negative binomial model can account for gene-specific variability from both biological and technical sources (Lund et al., 2012). Most previous studies explored the potential relationship of lncRNA and mRNA by correlation (Guo et al., 2015) and basic local alignment search tool (BLAST; Lin et al., 2014) procedures. Although both these approaches and our generalized linear model are based on expression correlation, lncRNA-mRNA coexpression analysis indicated the lncRNA regulation of mRNA; however, coexpression results from correlation and BLAST do not always indicate such regulation. Using a generalized linear model provides directionality and eliminates the lactation stage effect, which might reduce false-positive findings. Functional studies in humans have used similar models. Wu et al. (2016) used this method to identify lncRNA-mRNA coexpression pairs associated with breast cancer, and Guo et al. (2015) investigated the expression patterns of lncRNA and mRNA in ovarian cancer malignant progression. Thus, using this model might help improve lncRNA annotation.

Important metabolic and molecular changes between peak and late lactation stages in Chinese Holstein cows were revealed in the present study. The KEGG pathway analysis for the DE-lncRNA targeting mRNA indicated that lipid and glucose metabolic pathways were the most changed between peak and late lactation stages and that PPAR and AMPK signaling pathways were related to lactation. The AMPK pathway had not been reported in the mammary gland during the lactation stage. In addition, the overrepresented pathways differed between the 2 lactation stages. For example, steroid hormone biosynthesis, glycerolipid metabolism, and biosynthesis of AA were overrepresented in peak lactation, and this was in agreement with a previous study in the ovine mammary gland (Paten et al., 2015). There were more overrepresented metabolic pathways at the peak lactation stage than at the late lactation stage, reflecting the heavier request of energy and metabolites at peak lactation than at late lactation. Previous studies indicated limited growth of the mammary gland after early lactation (Capuco et al., 2001; Boutinaud et al., 2004), justifying the enhanced activity of secretory cells to increase milk yield in peak lactation (Boutinaud et al., 2004). Therefore, high activity in pathways related to lipid and glucose metabolism fulfills the enormous energy requirement for mammary cell proliferation or milk yield and secretion (Davies et al., 2006). These findings will improve knowledge on the bovine lactation process.

Twelve candidate lncRNA affecting bovine milk performance were found by bioinformatics analyses. For example, lnc-XLOC_274111 showed coexpression relationships with multiple protein-coding genes, including *FABP3* and *FABP4*, which code for 2 highly abundant fatty acid binding proteins in the bovine mammary gland (Bionaz and Loor, 2008b). These proteins transport endothelial long-chain fatty acids to the endoplasmic reticulum for the synthesis of triglycerides that are eventually incorporated into lipid droplets for secretion (Bionaz and Loor, 2008b). Target mRNA for XLOC_000752 were significantly enriched in the PPAR, AMPK, and insulin signaling pathways and in the biosynthesis of AA, carbon metabolism, and glycerolipid metabolism pathways. The PPAR, mechanistic target of rapamycin, and AMPK signaling pathways mediate lipid metabolism through several lipid metabolism–related transcription factors (Chen et al., 2013). Based on previous reports and on our KEGG pathway results, lnc-XLOC_274111 is likely to be involved in milk fat metabolism and mammary gland function during bovine lactation. However, further validation in functional experiments is required to verify this hypothesis. Nevertheless, these findings will be valuable for further studies on the bovine lactation mechanism.

The high number of putative lncRNA located within known QTL regions, particularly in chromosomal regions harboring QTL for milk traits, indicates that the collection of lncRNA found in the mammary tis-

sue transcriptome might provide useful information for elucidating their role in the mechanisms underlying the genetic variability of milk traits (van Binsbergen et al., 2012). Moreover, because the functional loci reported in a genome-wide association study may fall into non-coding regions (Michailidou et al., 2013), genetic variants mapping to lncRNA may play an important role in regulating gene expression levels via lncRNA-mRNA coexpression networks. For instance, we found that an SNP locus (rs29020019) was associated with milk fat percentage and was mapped to the DE-lncRNA XLOC_170433. This lncRNA showed coexpression with gene *TGM3*, which has been shown to be differentially expressed in Chinese Holstein cows with high and low milk production (Bai et al., 2016). Therefore, the findings of the present study provide further insight to functionally characterize genetic loci and SNP variants and could be useful for post-genome-wide association studies.

CONCLUSIONS

Overall, the 117 lncRNA and 254 protein-coding genes that were differentially expressed in the bovine mammary gland between peak and late lactation stages, the lncRNA-mRNA coexpression network pairs involved in lipid and glucose metabolism pathways, and the 12 potential lncRNA regulating milk traits provide a comprehensive set of lncRNA related to bovine lactation and will facilitate understanding this biological process and the molecular mechanisms underlying it. The data set obtained here also provides a valuable resource for the bovine lncRNA database and contributes to bovine functional research (https://figshare.com/articles/Integrated_analysis_of_lncRNA_and_mRNA_expression_profiles_reveals_the_potential_role_of_lncRNAs_in_different_bovine_lactation_stages/6633494).

ACKNOWLEDGMENTS

REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Anders, S., P. T. Pyl, and W. Huber. 2015. HTSeq—A Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169.

Ashwell, M. S., D. W. Heyen, T. S. Sonstegard, C. P. Van Tassell, Y. Da, P. M. VanRaden, M. Ron, J. I. Weller, and H. A. Lewin. 2004. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. J. Dairy Sci. 87:468–475.

Askarian-Amiri, M. E., J. Crawford, J. D. French, C. E. Smart, M. A. Smith, M. B. Clark, K. Ru, T. R. Mercer, E. R. Thompson, S. R. Lakhani, A. C. Vargas, I. G. Campbell, M. A. Brown, M. E. Dinger, and J. S. Mattick. 2011. SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. RNA 17:878–891.

Bai, X., Z. Zheng, B. Liu, X. Ji, Y. Bai, and W. Zhang. 2016. Whole blood transcriptional profiling comparison between different milk yield of Chinese Holstein cows using RNA-seq data. BMC Genomics 17(Suppl. 7):512.

Bennewitz, J., N. Reinsch, C. Grohs, H. Leveziel, A. Malafosse, H. Thomsen, N. Xu, C. Looft, C. Kuhn, G. A. Brockmann, M. Schwerin, C. Weimann, S. Hiendleder, G. Erhardt, I. Medjugorac, I. Russ, M. Forster, B. Brenig, F. Reinhardt, R. Reents, G. Averdunk, J. Blumel, D. Boichard, and E. Kalm. 2003. Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. Genet. Sel. Evol. 35:319–338.

Bennewitz, J., N. Reinsch, V. Guiard, S. Fritz, H. Thomsen, C. Looft, C. Kuhn, M. Schwerin, C. Weimann, G. Erhardt, F. Reinhardt, R. Reents, D. Boichard, and E. Kalm. 2004. Multiple quantitative trait loci mapping with cofactors and application of alternative variants of the false discovery rate in an enlarged granddaughter design. Genetics 168:1019–1027.

Billerey, C., M. Boussaha, D. Esquerre, E. Rebours, A. Djari, C. Meersseman, C. Klopp, D. Gautheret, and D. Rocha. 2014. Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. BMC Genomics 15:499.

Bionaz, M., and J. J. Loor. 2008a. 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. 138:1019–1024.

Bionaz, M., and J. J. Loor. 2008b. Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics 9:366.

Bionaz, M., and J. J. Loor. 2011. Gene networks driving bovine mammary protein synthesis during the lactation cycle. Bioinform. Biol. Insights 5:83–98.

Bionaz, M., K. Periasamy, S. L. Rodriguez-Zas, R. E. Everts, H. A. Lewin, W. L. Hurley, and J. J. Loor. 2012. Old and new stories: Revelations from functional analysis of the bovine mammary transcriptome during the lactation cycle. PLoS One 7:e33268.

Boutinaud, M., J. Guinard-Flamenta, and H. Jammes. 2004. The number and activity of mammary epithelial cells, determining factors for milk production. Reprod. Nutr. Dev. 44:499–508.

Bouwman, A. C., H. Bovenhuis, M. H. Visker, and J. A. van Arendonk. 2011. Genome-wide association of milk fatty acids in Dutch dairy cattle. BMC Genet. 12:43.

Buitenhuis, B., N. A. Poulsen, L. B. Larsen, and J. Sehested. 2015. Estimation of genetic parameters and detection of quantitative trait loci for minerals in Danish Holstein and Danish Jersey milk. BMC Genet. 16:52.

Cabili, M. N., C. Trapnell, L. Goff, M. Koziol, B. Tazon-Vega, A. Regev, and J. L. Rinn. 2011. Integrative annotation of human large

intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25:1915–1927.

Capuco, A. V., D. L. Wood, R. Baldwin, K. McLeod, and M. J. Paape. 2001. Mammary cell number, proliferation, and apoptosis during a bovine lactation: Relation to milk production and effect of bST. J. Dairy Sci. 84:2177–2187.

Chen, H., L. Zhang, X. W. Li, X. B. Li, G. Q. Sun, X. Yuan, L. C. Lei, J. X. Liu, L. H. Yin, Q. H. Deng, J. G. Wang, Z. X. Liu, W. T. Yang, Z. Wang, H. Zhang, and G. W. Liu. 2013. Adiponectin activates the AMPK signaling pathway to regulate lipid metabolism in bovine hepatocytes. J. Steroid Biochem. Mol. Biol. 138:445–454.

Cheng, J., P. Kapranov, J. Drenkow, S. Dike, S. Brubaker, S. Patel, J. Long, D. Stern, H. Tammana, G. Helt, V. Sementchenko, A. Piccolboni, S. Bekiranov, D. K. Bailey, M. Ganesh, S. Ghosh, I. Bell, D. S. Gerhard, and T. R. Gingeras. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 308:1149–1154.

Cornett, A. L., and C. S. Lutz. 2014. Regulation of COX-2 expression by miR-146a in lung cancer cells. RNA 20:1419–1430.

Daetwyler, H. D., F. S. Schenkel, M. Sargolzaei, and J. A. Robinson. 2008. A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. J. Dairy Sci. 91:3225–3236.

Davies, C. R., J. S. Morris, M. R. Griffiths, M. J. Page, A. Pitt, T. Stein, and B. A. Gusterson. 2006. Proteomic analysis of the mouse mammary gland is a powerful tool to identify novel proteins that are differentially expressed during mammary development. Proteomics 6:5694–5704.

Gao, X., J. Ye, C. Yang, K. Zhang, X. Li, L. Luo, J. Ding, Y. Li, H. Cao, Y. Ling, X. Zhang, Y. Liu, F. Fang, and Y. Zhang. 2017. Screening and evaluating of long noncoding RNAs in the puberty of goats. BMC Genomics 18:164.

Gindin, Y., Y. Jiang, P. Francis, R. L. Walker, O. D. Abaan, Y. J. Zhu, and P. S. Meltzer. 2015. miR-23a impairs bone differentiation in osteosarcoma via down-regulation of GJA1. Front. Genet. 6:233.

Goldman, A. S., S. Chheda, and R. Garofalo. 1998. Evolution of immunologic functions of the mammary gland and the postnatal development of immunity. Pediatr. Res. 43:155–162.

Gopalakrishnan, K., S. Kumarasamy, B. Mell, and B. Joe. 2015. Genome-wide identification of long noncoding RNAs in rat models of cardiovascular and renal disease. Hypertension 65:200–210.

Griffiths-Jones, S., R. J. Grocock, S. van Dongen, A. Bateman, and A. J. Enright. 2006. miRBase: MicroRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 34:D140–D144.

Guo, Q., Y. Cheng, T. Liang, Y. He, C. Ren, L. Sun, and G. Zhang. 2015. Comprehensive analysis of lncRNA-mRNA co-expression patterns identifies immune-associated lncRNA biomarkers in ovarian cancer malignant progression. Sci. Rep. 5:17683.

Gutschner, T., and S. Diederichs. 2012. The hallmarks of cancer: A long non-coding RNA point of view. RNA Biol. 9:703–719.

Haerty, W., and C. P. Ponting. 2015. Unexpected selection to retain high GC content and splicing enhancers within exons of multiexonic lncRNA loci. RNA 21:333–346.

Harder, B., J. Bennewitz, N. Reinsch, G. Thaller, H. Thomsen, C. Kuhn, M. Schwerin, G. Erhardt, M. Forster, F. Reinhardt, and E. Kalm. 2006. Mapping of quantitative trait loci for lactation persistency traits in German Holstein dairy cattle. J. Anim. Breed. Genet. 123:89–96.

He, Y., Y. Ding, F. Zhan, H. Zhang, B. Han, G. Hu, K. Zhao, N. Yang, Y. Yu, L. Mao, and J. Song. 2016. Corrigendum: The conservation and signatures of lincRNAs in Marek's disease of chicken. Sci. Rep. 6:19422.

Heyen, D. W., J. I. Weller, M. Ron, M. Band, J. E. Beever, E. Feldmesser, Y. Da, G. R. Wiggans, P. M. VanRaden, and H. A. Lewin. 1999. A genome scan for QTL influencing milk production and health traits in dairy cattle. Physiol. Genomics 1:165–175.

Hu, Z. L., C. A. Park, and J. M. Reecy. 2016. Developmental progress and current status of the animal QTLdb. Nucleic Acids Res. 44(D1):D827–D833.

Huang, W., N. Long, and H. Khatib. 2012a. Genome-wide identification and initial characterization of bovine long non-coding RNAs from EST data. Anim. Genet. 43:674–682.

Huang, W., F. Peñagaricano, K. Ahmad, J. Lucey, K. Weigel, and H. Khatib. 2012b. Association between milk protein gene variants and protein composition traits in dairy cattle. J. Dairy Sci. 95:440–449.

Jalali, S., D. Bhartiya, M. K. Lalwani, S. Sivasubbu, and V. Scaria. 2013. Systematic transcriptome wide analysis of lncRNA-miRNA interactions. PLoS One 8:e53823.

Kadegowda, A. K., M. Bionaz, B. Thering, L. S. Piperova, R. A. Erdman, and J. J. Loor. 2009. Identification of internal control genes for quantitative polymerase chain reaction in mammary tissue of lactating cows receiving lipid supplements. J. Dairy Sci. 92:2007–2019.

Kapranov, P., S. E. Cawley, J. Drenkow, S. Bekiranov, R. L. Strausberg, S. P. Fodor, and T. R. Gingeras. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. Science 296:916–919.

Kiewe, P., S. Gueller, M. Komor, A. Stroux, E. Thiel, and W. K. Hofmann. 2009. Prediction of qualitative outcome of oligonucleotide microarray hybridization by measurement of RNA integrity using the 2100 Bioanalyzer capillary electrophoresis system. Ann. Hematol. 88:1177–1183.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg. 2013. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14:R36.

Li, C., W. Cai, C. Zhou, H. Yin, Z. Zhang, J. J. Loor, D. Sun, Q. Zhang, J. Liu, and S. Zhang. 2016. RNA-Seq reveals 10 novel promising candidate genes affecting milk protein concentration in the Chinese Holstein population. Sci. Rep. 6:26813.

Lillehammer, M., M. Arnyasi, S. Lien, H. G. Olsen, E. Sehested, J. Odegard, and T. H. Meuwissen. 2007. A genome scan for quantitative trait locus by environment interactions for production traits. J. Dairy Sci. 90:3482–3489.

Lima, J. A., J. R. Ruas, A. C. Vasconcelos, B. F. Silper, A. M. Lana, V. A. Gheller, H. M. Saturnino, R. B. Reis, and S. G. Coelho. 2016. Effects of bovine mammary gland biopsy and increased milking frequency on post-procedure udder health, histology, and milk yield. Animal 10:838–846.

Lin, X. C., Y. Zhu, W. B. Chen, L. W. Lin, D. H. Chen, J. R. Huang, K. Pan, Y. Lin, B. T. Wu, Y. Dai, and Z. G. Tu. 2014. Integrated analysis of long non-coding RNAs and mRNA expression profiles reveals the potential role of lncRNAs in gastric cancer pathogenesis. Int. J. Oncol. 45:619–628.

Livak, K. J., and T. D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T) (-Delta Delta C) method. Methods 25:402–408.

Lund, M. S., P. Sorensen, P. Madsen, and F. Jaffrezic. 2008. Detection and modelling of time-dependent QTL in animal populations. Genet. Sel. Evol. 40:177–194.

Lund, S. P., D. Nettleton, D. J. McCarthy, and G. K. Smyth. 2012. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. Stat. Appl. Genet. Mol. 11:1–44.

Mancini-Dinardo, D., S. J. Steele, J. M. Levorse, R. S. Ingram, and S. M. Tilghman. 2006. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev. 20:1268–1282.

McHugh, C. A., C. K. Chen, A. Chow, C. F. Surka, C. Tran, P. McDonel, A. Pandya-Jones, M. Blanco, C. Burghard, A. Moradian, M. J. Sweredoski, A. A. Shishkin, J. Su, E. S. Lander, S. Hess, K. Plath, and M. Guttman. 2015. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature 521:232–236.

McManaman, J. L., and M. C. Neville. 2003. Mammary physiology and milk secretion. Adv. Drug Deliv. Rev. 55:629–641.

Michailidou, K., P. Hall, A. Gonzalez-Neira, M. Ghoussaini, J. Dennis, R. L. Milne, M. K. Schmidt, J. Chang-Claude, S. E. Bojesen, and M. K. Bolla. 2013. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat. Genet. 45:353–361.

Mo, X. B., L. F. Wu, X. W. Zhu, W. Xia, L. Wang, P. He, P. F. Bing, X. Lu, Y. H. Zhang, F. Y. Deng, and S. F. Lei. 2017. Identification and evaluation of lncRNA and mRNA integrative modules in human peripheral blood mononuclear cells. Epigenomics 9:943–954.

Mogilyansky, E., P. Clark, K. Quann, H. Zhou, E. Londin, Y. Jing, and I. Rigoutsos. 2016. Post-transcriptional regulation of BRCA2 through interactions with miR-19a and miR-19b. Front. Genet. 7:143.

Ogawa, Y., B. K. Sun, and J. T. Lee. 2008. Intersection of the RNA interference and X-inactivation pathways. Science 320:1336–1341.

Olsen, H. G., L. Gomez-Raya, D. I. Vage, I. Olsaker, H. Klungland, M. Svendsen, T. Adnoy, A. Sabry, G. Klemetsdal, N. Schulman, W. Kramer, G. Thaller, K. Ronningen, and S. Lien. 2002. A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle. J. Dairy Sci. 85:3124–3130.

Paten, A. M., E. J. Duncan, S. J. Pain, S. W. Peterson, P. R. Kenyon, H. T. Blair, and P. K. Dearden. 2015. Functional development of the adult ovine mammary gland—Insights from gene expression profiling. BMC Genomics 16:748.

Pauli, A., E. Valen, M. F. Lin, M. Garber, N. L. Vastenhouw, J. Z. Levin, L. Fan, A. Sandelin, J. L. Rinn, A. Regev, and A. F. Schier. 2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Res. 22:577–591.

Ren, H., G. Wang, L. Chen, J. Jiang, L. Liu, N. Li, J. Zhao, X. Sun, and P. Zhou. 2016. Genome-wide analysis of long non-coding RNAs at early stage of skin pigmentation in goats (Capra hircus). BMC Genomics 17:67.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

Sado, T., Y. Hoki, and H. Sasaki. 2005. Tsix silences Xist through modification of chromatin structure. Dev. Cell 9:159–165.

Sandhu, G. K., M. J. G. Milevskiy, W. Wilson, A. M. Shewan, and M. A. Brown. 2016. Non-coding RNAs in mammary gland development and disease. Adv. Exp. Med. Biol. 886:121–153.

Saremi, B., H. Sauerwein, S. Danicke, and M. Mielenz. 2012. Technical note: Identification of reference genes for gene expression studies in different bovine tissues focusing on different fat depots. J. Dairy Sci. 95:3131–3138.

Schmitz, S., M. W. Pfaffl, H. H. D. Meyer, and R. M. Bruckmaier. 2004. Short-term changes of mRNA expression of various inflammatory factors and milk proteins in mammary tissue during LPS-induced mastitis. Domest. Anim. Endocrinol. 26:111–126.

Schopen, G. C., P. D. Koks, J. A. van Arendonk, H. Bovenhuis, and M. H. Visker. 2009. Whole genome scan to detect quantitative trait loci for bovine milk protein composition. Anim. Genet. 40:524–537.

Standaert, L., C. Adriaens, E. Radaelli, A. Van Keymeulen, C. Blanpain, T. Hirose, S. Nakagawa, and J. C. Marine. 2014. The long noncoding RNA Neat1 is required for mammary gland development and lactation. RNA 20:1844–1849.

Sun, L., H. Luo, D. Bu, G. Zhao, K. Yu, C. Zhang, Y. Liu, R. Chen, and Y. Zhao. 2013. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res. 41:e166.

Sun, X., M. Li, Y. Sun, H. Cai, X. Lan, Y. Huang, Y. Bai, X. Qi, and H. Chen. 2016. The developmental transcriptome sequencing of bovine skeletal muscle reveals a long noncoding RNA, lncMD, promotes muscle differentiation by sponging miR-125b. Biochim. Biophys. Acta 1863:2835–2845.

Tong, C., Q. Chen, L. Zhao, J. Ma, E. M. Ibeagha-Awemu, and X. Zhao. 2017. Identification and characterization of long intergenic noncoding RNAs in bovine mammary glands. BMC Genomics 18:468.

Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7:562–578.

Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. Van Baren, S. L. Salzberg, B. J. Wold, and L. Pachter. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28:511–515.

van Binsbergen, R., R. F. Veerkamp, and M. P. Calus. 2012. Makeup of the genetic correlation between milk production traits using genome-wide single nucleotide polymorphism information. J. Dairy Sci. 95:2132–2143.

Viitala, S. M., N. F. Schulman, D. J. de Koning, K. Elo, R. Kinos, A. Virta, J. Virta, A. Maki-Tanila, and J. H. Vilkki. 2003. Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. J. Dairy Sci. 86:1828–1836.

Wang, Y., S. Xue, X. Liu, H. Liu, T. Hu, X. Qiu, J. Zhang, and M. Lei. 2016. Analyses of long non-coding RNA and mRNA profiling using RNA sequencing during the pre-implantation phases in pig endometrium. Sci. Rep. 6:20238.

Wickramasinghe, S., G. Rincon, A. Islas-Trejo, and J. F. Medrano. 2012. Transcriptional profiling of bovine milk using RNA sequencing. BMC Genomics 13:45.

Wu, W., E. K. Wagner, Y. Hao, X. Rao, H. Dai, J. Han, J. Chen, A. M. V. Storniolo, Y. Liu, and C. He. 2016. Tissue-specific co-expression of long non-coding and coding RNAs associated with breast cancer. Sci. Rep. 6:32731.

Zhang, Q., D. Boichard, I. Hoeschele, C. Ernst, A. Eggen, B. Murkve, M. Pfister-Genskow, L. A. Witte, F. E. Grignola, P. Uimari, G. Thaller, and M. D. Bishop. 1998. Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics 149:1959–1973.

Zhao, Y., H. Li, S. Fang, Y. Kang, W. Wu, Y. Hao, Z. Li, D. Bu, N. Sun, M. Q. Zhang, and R. Chen. 2016. NONCODE 2016: An informative and valuable data source of long non-coding RNAs. Nucleic Acids Res. 44(D1):D203–D208.