



Identification of lncRNAs Differentially Expressed during Natural and Induced Estrus in Sheep

Bujun Mei^{1,2}, Rong Liu^{2,3}

10.18805/IJAR.B-1347

ABSTRACT

Background: The manipulation of the estrous cycle or induction of estrus is a commonly used technique in sheep industry. The goal of this study was to identify and characterize differences of non-coding RNAs (lncRNAs) expression between induced estrus and natural estrus using the BGISEQ-500 platform in 7 Mongolian sheep, which will provide insights into the regulation mechanisms of lncRNAs in different reproduction mode of sheep.

Methods: During the late spring, ovarian, pituitary, hypothalamic, pineal and uterine tissue samples were collected from four artificially induced estrus and three naturally estrus Mongolian sheep. Total RNA was extracted from the five tissues using TRIzol reagent (Invitrogen) and treated with DNase I following the manufacturer's instructions. A total of 35 sheep samples were sequenced using the BGISEQ-500 platform. Bioinformatics methods were used to analysis expression difference analysis between groups, SNP and InDel, alternative splicing, lncRNA's miRNA precursor prediction, lncRNA target gene and family prediction.

Result: 211 novel lncRNAs were systematically identified using RNA-Seq technology. Meanwhile, we found that there are diversifications of lncRNAs in induced estrus vs. nature estrus of ewes. Therefore, we predict that, under the action of exogenous hormones, many physiological processes of ewes may be affected to varying degrees through the change of lncRNA to a variety of pathways.

Key words: Ewe, Induced estrus, lncRNAs, Natural estrus, Reproductive traits, Sheep.

INTRODUCTION

Reproduction in ewes is regulated by an estrus cycle. Most sheep are seasonal estrus animals that begin gonads only in autumn and winter. The length of the estrus cycle in sheep ranges from 13 to 19 days and averages about 17 days, including four phases per cycle: proestrus, estrus, metestrus and diestrus. Estrus of sheep lasts approximately 24 to 36 hours. However, some sheep breeds are less seasonal, breeding almost year-round or in extended breeding season. Seasonal estrus is a major factor limiting sheep's fertility, due to the low reproductive rate and long interval between lambs of sheep (Malik *et al.* 2017). Therefore, in the sheep industry, estrus induction/synchronization technology is often used. Estrus and ovulation can be induced in anestrus ewes by the introduction of rams (ram effect), or by treatment with exogenous progestogen or equine chorionic gonadotropin (eCG) or using the effects of melatonin either by manipulating the photoperiod in housed sheep or using exogenous melatonin in the feed or as an implant. Inducing estrus, like estrus synchronization, is widely used in sheep production.

lncRNAs are crucial regulatory elements of gene expression in many biological processes. lncRNAs usually refer to transcripts that are not encoded by an open reading frame and are longer than 200 nt. The vital role of lncRNAs in the reproduction process is widely reported, such as germ cell specification, sex determination and gonadogenesis, sex hormone responses, meiosis, gametogenesis, placentation. lncRNAs (including Gtl2) from Dlk1 Dio3 region were positively correlated with the pluripotency of iPS cells. Wei

¹Department of Agriculture, Hetao College, Bayannur, 015000, People's Republic of China.

²Laboratory of Small Ruminant Genetics, Breeding and Reproduction, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, People's Republic of China.

³Inner Mongolia Autonomous Region Farming and Pasture Science and Technology Extension Station, Huhhot, 010000, People's Republic of China.

Corresponding Author: Bujun Mei, Department of Agriculture, Hetao College, Bayannur, 015000, People's Republic of China. Email: meibujun@163.com

How to cite this article: Mei, B. and Liu, R. (2021). Identification of lncRNAs Differentially Expressed during Natural and Induced Estrus in Sheep. Indian Journal of Animal Research. 55(12): 1421-1429. DOI: 10.18805/IJAR.B-1347.

Submitted: 12-01-2021 **Accepted:** 20-07-2021 **Online:** 21-08-2021

et al. (2019) analyzed the mechanism of Gtl2 epigenetic regulation to uncover the spatiotemporal expression patterns and changes of lncRNA Gtl2. However, few reports of them in livestock have been published. Moreover, little studies have focused on the differential expression of lncRNAs in natural and induced estrus in the sheep.

In this study, we investigated the difference of non-coding RNA expression between induced estrus and natural estrus using the BGISEQ-500 platform in 7 Mongolian sheep, which will provide insights into the regulation mechanisms of lncRNAs in different reproduction mode of sheep.

MATERIALS AND METHODS

Experimental design and sample collection

All the procedures involving animals and experiments were performed in accordance with the relevant guidelines and regulations set by the Ministry of Agriculture of the People's Republic of China. Animals were humanely sacrificed as necessary to ameliorate suffering. In this study, ovarian, pituitary, hypothalamic, pineal and uterine tissue samples were collected from four artificially induced estrus and three naturally estrus Mongolian sheep. Once all samples were obtained, they were immediately frozen in liquid nitrogen and stored in a -80°C freezer until RNA extraction.

The induction and synchronization of estrus were treated in 4 sheep during the late spring. Drugs for estrus synchronization: progesterone (P4) in the form of silicone progesterone suppositories (CRID), produced in New Zealand; pregnant mare serum gonadotropin (PMSG) and prostaglandin (PGF2 α), purchased from Ningbo Sansheng Pharmaceutical Co., Ltd. Synchronous estrus treatment scheme: P₄ (10~12 d) + PMSG (200~400 IU per ewe) + PGF2 α (0.1 mg per ewe).

Teasers lasting 60 hours were started at 12 hours after removing the vaginal sponge suppository containing 150-350 mg of progesterone for estrus induction of ewe. The control group adopted natural estrus. After the vaginal sponge suppository placed into a vagina of ewe of the induction estrus group, teasers were conducted daily in the control group and the record lasted for three estrus periods (about 60 days). When the ewe stood strongly in the back-pressure test without any movement, the vulva was swollen and had yellow transparent liquid and the sheep was judged to be in estrus.

RNA extraction, construction of sequencing libraries and sequencing

The construction of the sequencing libraries and sequencing were performed by BGI Genomics Co., Ltd (Shenzhen). Total RNA was extracted from the five tissues using TRIzol reagent (Invitrogen) and treated with DNase I following the manufacturer's instructions. The total RNA samples were preserved in the RNase-Free water to directly synthesize the First-stand cDNA. Subsequently, the First-stand cDNA was performed full-length LD-PCR amplification and the amplified double-strand cDNA (ds cDNA) was purified using the AMPure XP beads; finally, the Qubit was employed for quantitative detection of ds cDNA. Then, the Covaris system was utilized for sonication on ds cDNA, the sonicated double strand short fragments were performed end repair, addition of A tail and connection of the sequencing adapter. Later, the AMPure XP beads were used to purify the fragments and the libraries with the fragment size of about 200 bp were selected. Eventually, PCR enrichment was performed to obtain the final cDNA libraries. After the construction of the libraries, Qubit2.0 was used at first for preliminary quantification, the libraries were then diluted to 2 ng/ μ l and then Agilent2100 was adopted to detect the insert size of

the libraries. When the insert size had conformed to the expectation, the effective concentration of the library was accurately quantified according to the Q-PCR method (the effective library concentration of >2 nM), so as to guarantee the library quality. After qualified library detection, different libraries were pooled based on the effective concentration and the target data volume demand prior to BGISEQ-500 sequencing.

Analysis of transcriptomic data

In this project, a total of 35 sheep samples were sequenced using the BGISEQ-500 platform. Each sample produced an average of 12.49 Gb of data and the average genome alignment rate was 82.66%. A total of 111,293 transcript expressions were detected, of which 44,117 were new lncRNAs, 23,663 were new mRNAs, 4,195 were known lncRNAs and 39,318 were known mRNAs.

To guarantee the quality of the sequencing data, the Raw reads obtained after sequencing were carried out data filtering, as well as removal of adapter sequence, null read sequence and the low-quality sequence (Phred quality < 5), so as to obtain the clean reads. Then, the HISAT software was selected and the parameter mismatch was set at 2 (Kim *et al.* 2015), while the remaining parameters had adopted the software default parameters, so as to compare and analyze the Clean reads and the reference genome, then use StringTie to assemble (Pertea *et al.* 2015).

After obtaining a new transcript, coding capacity of the transcript was predicted to distinguish between mRNA and lncRNA. Three prediction software (CPC, txCdsPredict and CNCI) and a database (pfam) were used to make predictions (Kong *et al.* 2007). All three prediction software score the transcript's coding ability and then set a scoring threshold to distinguish between lncRNA and mRNA. If transcripts can be alignment with pfam, it is considered to be mRNA, otherwise it is lncRNA (Finn *et al.* 2016). Among the four judgment methods, at least three methods' results are consistent, then the transcript can be confirmed as mRNA or lncRNA.

Subsequently, we used Bowtie2 to align clean reads to the reference sequence and then use RSEM to calculate the expression of genes and transcripts (Li and Dewey 2011, Langmead and Salzberg 2012). The gene expression quantities were calculated using Reads per kilobase transcriptome per million mapped reads (RPKM); at the same time, gene differential expression was analyzed according to the DEGseq method for non-biological replicate samples and $|\log_2 \text{ratio}| \geq 1$ and p-value < 0.05 were used as the thresholds of gene differential expression, so as to screen the differentially expressed genes (DEGs) (Wang *et al.* 2010). Later, the significantly enriched GO entries in the obtained DEGs were calculated using the Goseq method, which were then compared with the Kyoto encyclopedia of genes and genomes (KEGG) database using the Blast2GO software (Conesa *et al.* 2005), so as to analyze the DEGs-related signaling pathways or metabolic

pathways. Afterwards, we used the software GATK for SNP and InDel testing of each sample and mapping results to specific genes. Meanwhile, variable splicing events of different samples were detected by Asprofile software.

RESULTS AND DISCUSSION

Sequencing quality evaluation and baseline data analysis

Upon detection, the micro-libraries constructed in this study had satisfied the requirement of transcriptome sequencing. The Agilent High Sensitivity DNA Kit was adopted for Agilent 2100 Bioanalyzer detection and the peak graphs of the 35 samples suggested the presence of obvious target product main peaks within the length scope of 1-2 kb, some small fragment products could be observed at below 1kb, but they had accounted for a small proportion, suggesting that the original samples had good integrity, which were eligible upon judgment and had conformed to the requirement of library construction. The Q30 percentage of the 35 samples ranged from 89.70% to 94.10%, indicating high sequencing quality and library construction quality; therefore, the sequencing data were accurate and reliable, which could be used in subsequent analysis. In the sequencing results, the base C/G proportions of the 35 samples were close to 50% and the contents were basically corresponding and coinciding, suggesting stable and balanced base composition and high sequencing quality.

The comparison of clean reads with reference genome indicated that, the total mapped of the 35 samples were greater than 82.42%, suggesting extremely successful mRNA extraction and library construction, excluding the contamination by various exogenous impurities. Multiple mapped suggested that the sequences obtained after sequencing might be compared to the similar regions in genome and its proportion should be appropriately lower than 10%. In this study, such data of the 35 samples were below 3.5%, demonstrating excellent data quality, which could provide excellent foundation for subsequent analysis. Exon number distribution of transcripts, including mRNA and lncRNA. Based on the structural features of lncRNAs and

the functional characteristics of noncoding proteins, we have established a series of rigorous screening conditions to screen lncRNAs. Wayn diagram of lncRNA for Coding ability prediction using software (CPC, txCdsPredict, CNCI) and database (pfam) is shown in Fig 1. In order to more intuitively show the number of genes in different FPKM intervals for each sample, we conducted statistics on the number of genes in three cases of FPKM (FPKM \leq 1, FPKM 1~10, FPKM \geq 10).

Expression difference analysis between groups

In this study, Bowtie2 was adopted to align clean reads to the gene reference sequence and counted the gene coverage. The differentially expressed transcripts among different samples could be obtained through differential analysis and the corresponding genes or target genes of these transcripts might exert their functions under different treatments or in different phenotypes. The DEGseq software was to analyze the gene expression levels of all samples, with union being used as the model (Fig 2). FPKM=1 was used as the threshold and that of >1 indicated up-regulation of gene expression, while that of <1 suggested gene down-regulation of gene expression. Moreover, the default screening thresholds were set as $|\log_2\text{ratio}|\geq 1$ and p-value <0.05 . The overall distribution of the differentially expressed transcripts or genes could be intuitively observed from the scatter plot and the DEGseq was utilized to analyze the differential expression of genes among the samples. It could be seen from Fig 2, compared with natural estrus group, induced estrus group was associated with 23485 DEGs, including 1480 up-regulated and 1714 down-regulated ones. In order to better understand gene function, we will annotate the assembled novel mRNA and known mRNA. The annotated databases include NR, NT, GO, KOG, KEGG, SwissProt. We use Blast or Diamond software to annotate mRNA with NT, NR, KOG, KEGG and SwissProt and use Blast2GO and NR software annotation results for GO analysis.

Gene Oncology (referred to as GO for short) is the international standard classification system for gene function. GO can be divided into three parts, including Molecular Function, Biological Process and Cellular Component. It could be seen from Fig 3 that, the differentially expressed DEGs compared between induced estrus and natural estrus group were enriched to molecular functions (such as catalytic activity, transcription activity, signal transducer activity and molecular function regulator), biological processes (such as cellular process, biological regulation, metabolic process, response to stimulus and multicellular organismal process) and cellular components of cell part, organelle, membrane and macromolecular complex.

Furthermore, we had also employed KEGG and KOG Pathway enrichment analysis on the DEGs, as displayed in Fig 4-6. The results suggested that, the differentially expressed genes between induced estrus and natural estrus

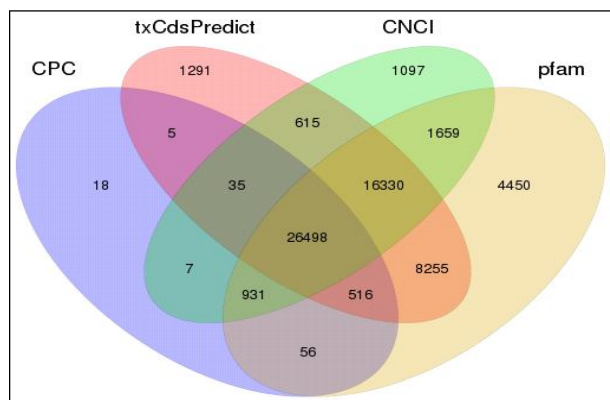


Fig 1: lncRNA coding ability prediction Venn plot.

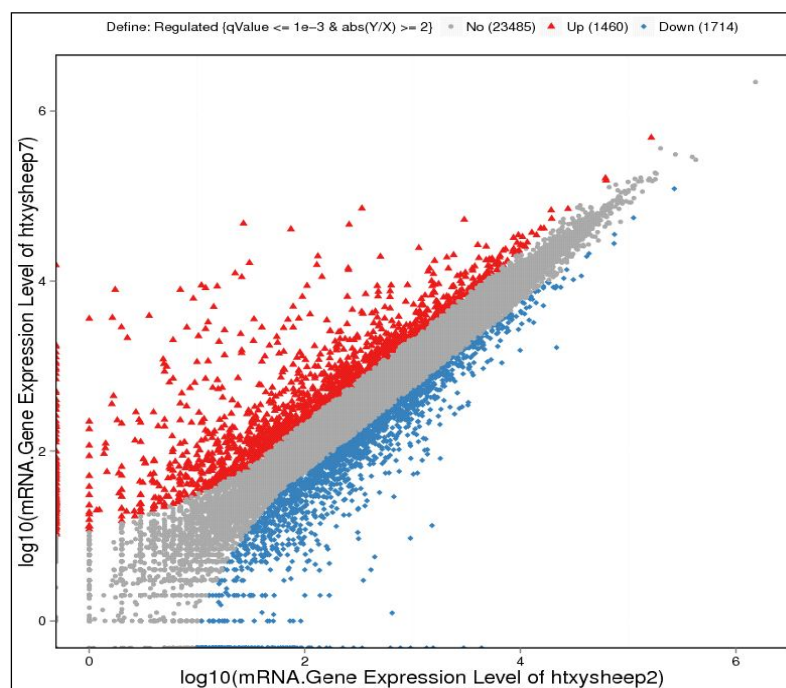


Fig 2: Scatter plot of differential genes between groups. X axis: gene logarithmic expression of a sample; Y axis: gene logarithmic expression of another sample; color: blue indicates down regulation, Orange indicates up-regulated genes, brown indicates non-significantly different genes.

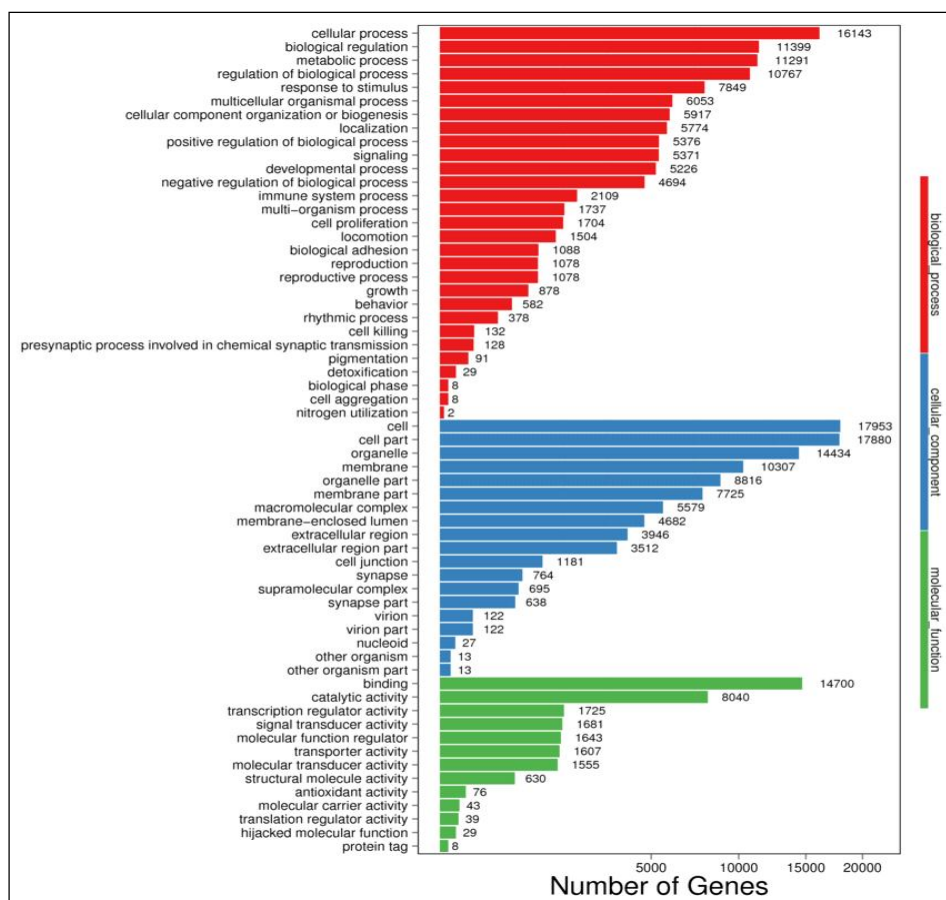


Fig 3: GO annotation statistics.

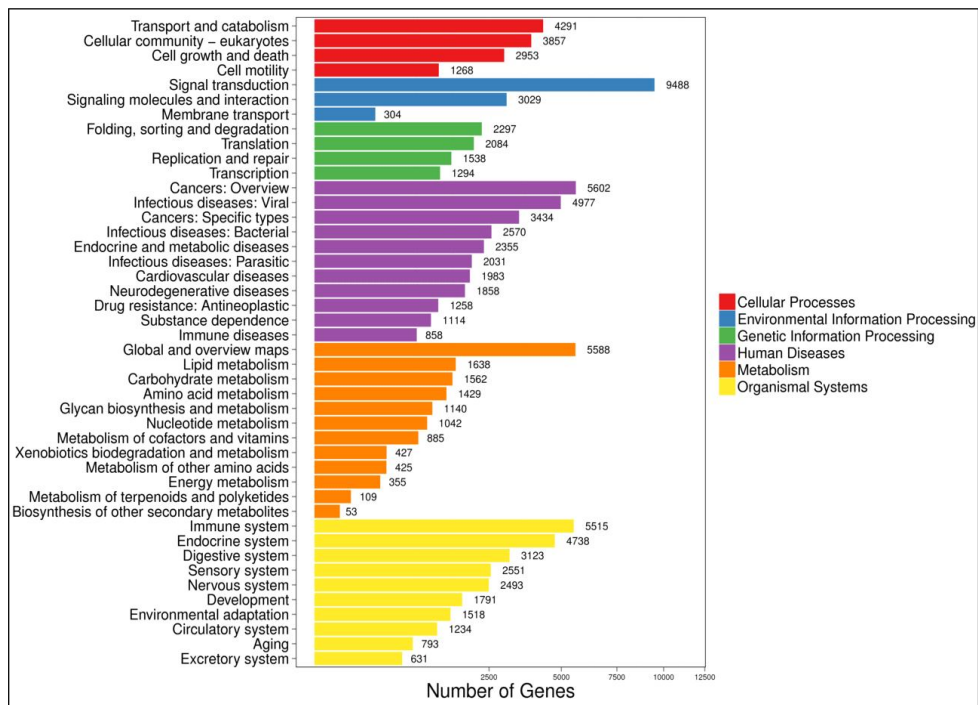


Fig 4: KEGG statistics.

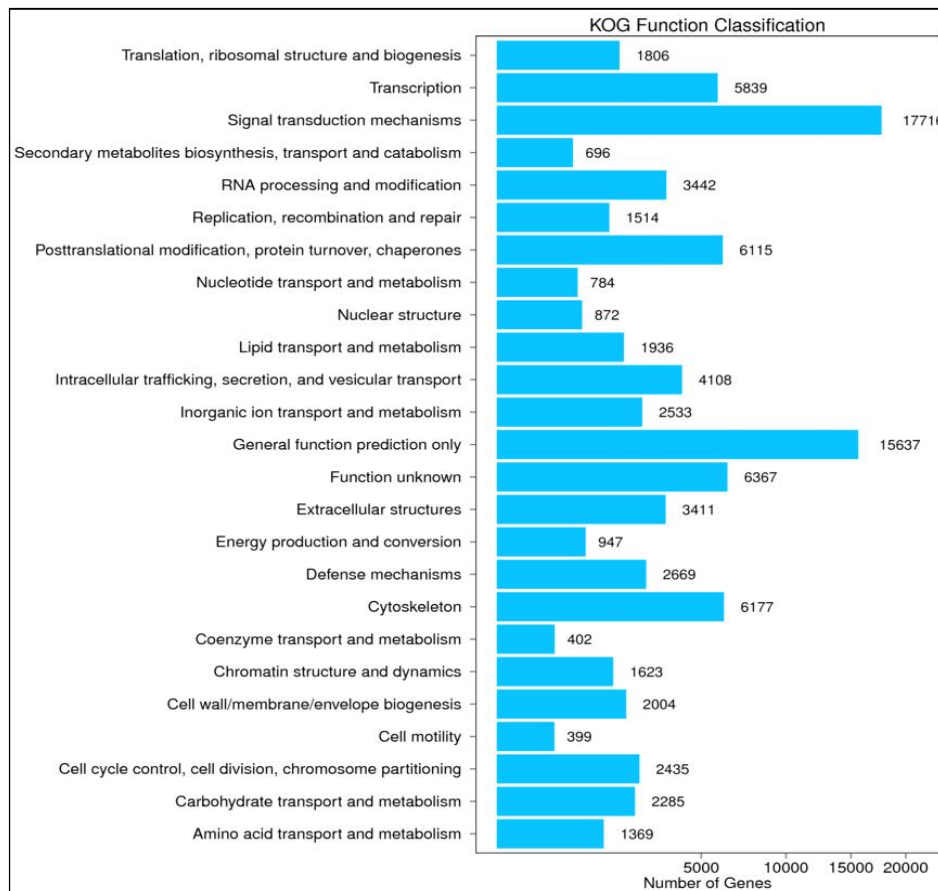


Fig 5: KOG statistics.

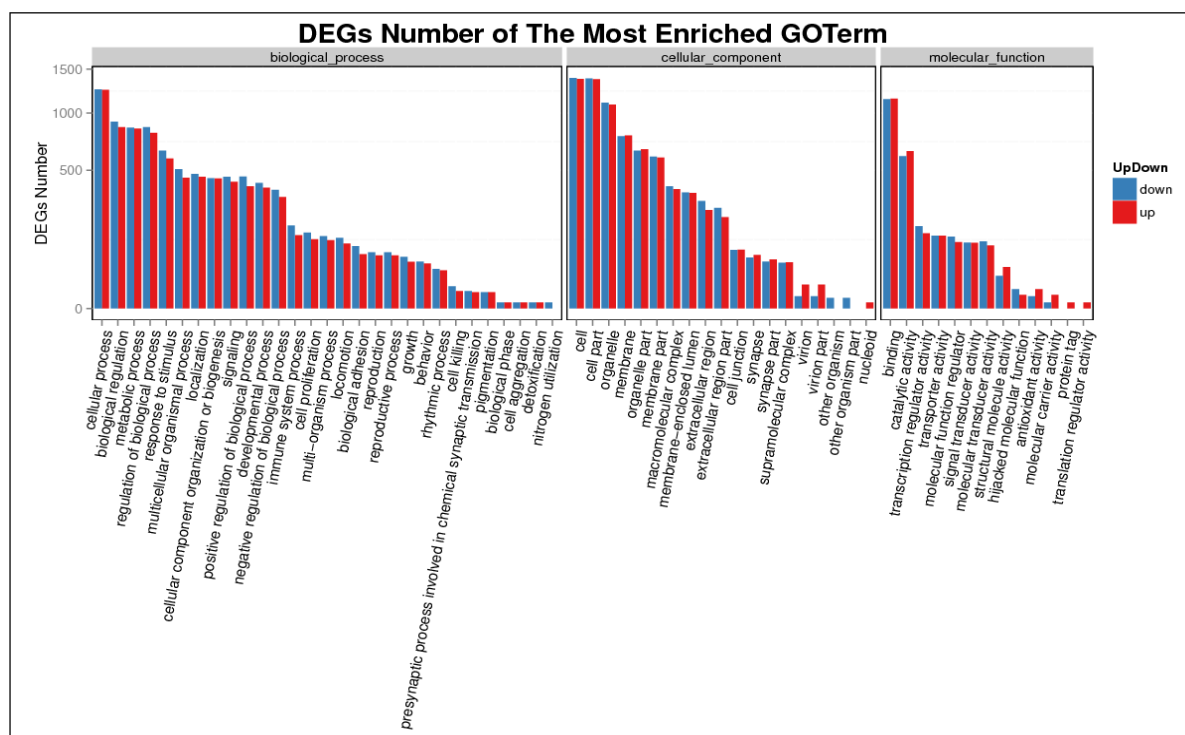


Fig 6: Differential genes between groups: up- and down-regulated GO statistics.

group were enriched to the cellular processes (such as transport and catabolism, cellular community-eukaryotes, cell growth and death, cell motility), environmental information processing (signal transduction, signaling molecules and interaction, membrane transport), genetic information processing (folding, sorting and degradation, translation, replication and repair, transcription), metabolism (global and overview maps, lipid metabolism, carbohydrate metabolism, amino acid metabolism, glycan biosynthesis and metabolism), organismal systems (immune system, endocrine system, digestive system and sensory system). In addition, we had also done KEGG Pathway enrichment analysis on the DEGs between all samples.

SNP and InDel analysis

Single Nucleotide Polymorphisms (SNPs) refer to changes in DNA sequence caused by changes in single nucleotides (A, T, C, or G), resulting in diverse genomes among species or individuals, including the base replacement and substitution *etc.* After comparing with the reference genome, we used the software GATK to perform SNP detection on each sample and mapped them to specific genes. We detected approximately 195,601 to 2,505,134 SNPs and 13,016 to 191,097 InDel in the 35 samples. The most frequent SNPs were A>G, C>T, G>A and T>C. The most frequent length of InDel was one nucleotide.

Alternative splicing analysis

We use Asprofile software to detect alternative splicing events in different samples. We detected 11 types of Alternative exon ends (AE), Intron retention (IR), Multi-IR

(MIR), Multi-exon SKIP (MSKIP), Skipped exon (SKIP), Alternative 5' first exon (TSS), Alternative 3' last exon (TTS), Approximate AE (XAE), Approximate IR (XIR), Approximate MSKIP (XMSKIP), Approximate SKIP (XSKIP). The top two AS events were TSS and TTS AS events, which accounted for more than 82% of AS events in each library (Fig 7).

lncRNA's miRNA precursor prediction

We used Blast to align lncRNA to miRBase to find potential miRNA precursors. lncRNA with miRNA precursor alignment coverage greater than 90% was selected (Griffiths-Jones *et al.* 2006). Our study found that the number of novel lncRNAs align to miRNA precursors is 211. Number of positions aligned to miRNA precursor (a lncRNA may have different positions to match miRNA precursor) is 2,942.

lncRNA target gene prediction and family analysis

The function of lncRNA is mainly achieved by acting on the target gene through cis or trans. The basic principle of Cis action target gene prediction believes that the function of lncRNA is related to the protein coding genes whose coordinates are close, so the mRNA adjacent to lncRNA is selected as its target gene. However, trans regulation does not depend on the positional relationship. We predict it by calculating the binding energy. If there is an overlap between lncRNA and the target gene, we would carefully classify the situation of overlap (Bu *et al.* 2012), which help us understand the details of cis regulation and Overlap classification statistics is shown in Fig 8.

We use software INFERNAL to align lncRNA to Rfam database, so as to annotate the family of lncRNA

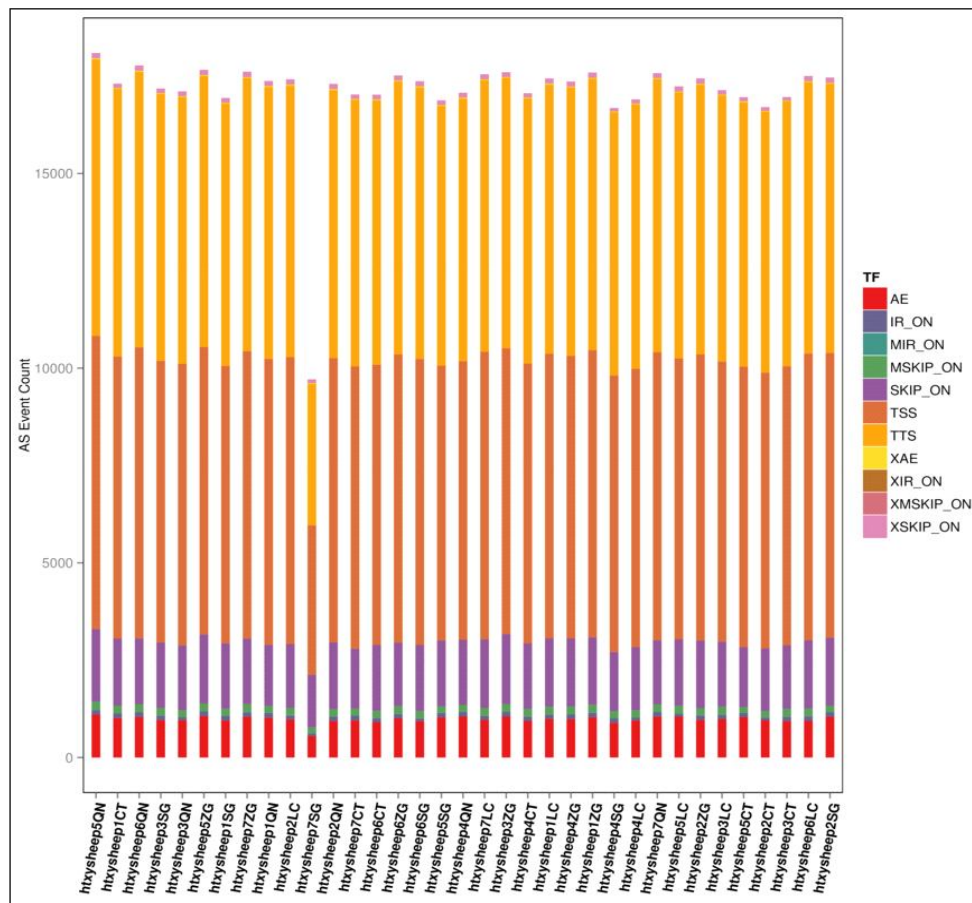


Fig 7: Alternative splicing statistics.

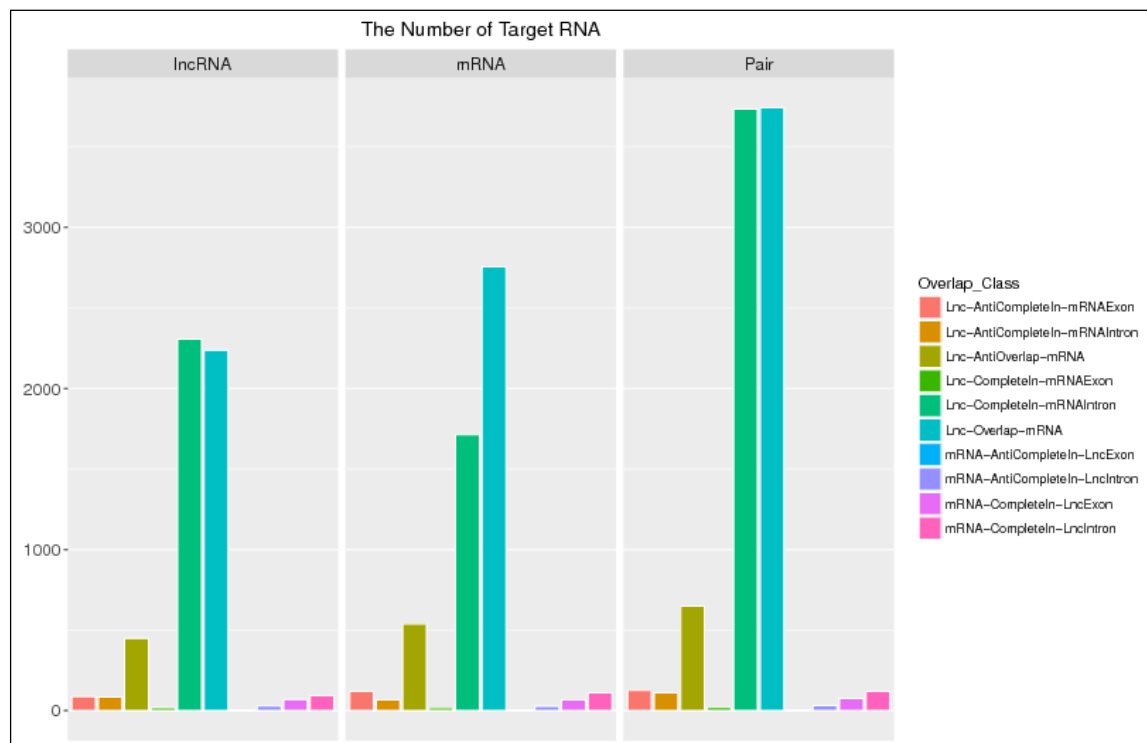


Fig 8: Overlap classification of lncRNA target gene prediction.

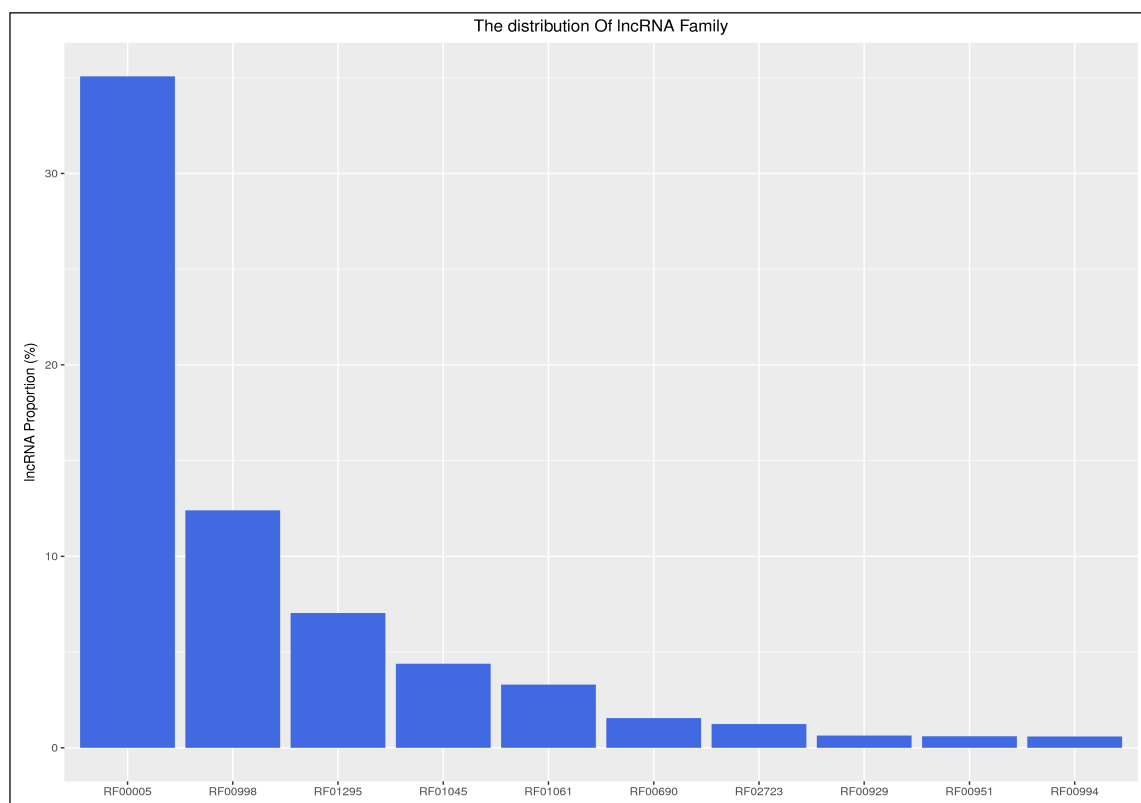


Fig 9: lncRNA family analysis statistics.

(Nawrocki *et al.* 2009, Burge *et al.* 2013). Rfam is a database containing information on various ncRNA families, including conserved regions of RNA secondary structure, mRNA cis-acting elements and other RNA elements. It divides ncRNAs into different families based on their common ancestors at the evolutionary. The predicted statistical graph of the lncRNA family is as shown in Fig 9.

CONCLUSION

In recent years, high-throughput sequencing has been performed using animal reproductive tissues, such as endometrium, ovaries, placenta, follicles and granulosa cells, to identify the genes and genetic loci that affect reproduction traits. lncRNAs play key roles in many important physiological and pathological processes. At present, several functional lncRNAs have been identified in prediction and regulation of lncRNAs in sheep. However, the role of lncRNAs in the ovarian, pituitary, hypothalamic, pineal and uterine tissue of sheep between natural and induced estrus processes is unclear. In our study, 211 novel lncRNAs were systematically identified using RNA-Seq technology. Many studies have shown that lncRNAs exert their regulatory functions through a series of complex and elaborate regulatory processes. In this study, we found that there are large number diversifications of lncRNAs in induced estrus vs. nature estrus of ewes. Bioinformatics analysis showed that inducing estrus in sheep changes the expression level of lncRNA in many biological processes, such as catalytic activity, metabolic process, cellular

community *etc.* (Chu *et al.* 2017). We found that lncRNAs would regulate the expression of adjacent genes by cis or trans mechanism of action, through 10 ways of overlap between lncRNA and target gene. Under the action of exogenous hormones, various forms of alternative splicing would appear in the gene expression process. In order to further research the functions of these differentially expressed lncRNAs in induced estrus of sheep, we analyzed the target genes of these differentially expressed lncRNAs. We found that lncRNA, such as LTCONS_00000214, LTCONS_00000450 and LTCONS_00000466 *etc.*, target mRNAs. Meantime, the expression levels of the two lncRNAs and mRNAs were significantly difference in the induced and nature estrus of sheep and the trends of lncRNAs expression and their cis-target genes were consistent. Therefore, we predict that, under the action of exogenous hormones, many physiological processes of ewes may be affected to varying degrees through the change of lncRNA to a variety of pathways.

This study implies that there may be a few differences between the mechanism of natural estrus and artificially induced estrus. The influence of exogenous hormones on the physiological state of the ewe during artificially induced estrus still exist many issues that deserve to be studied and further explored.

ACKNOWLEDGEMENT

We are thankful to Bayannaoer Academy of Agriculture and Animal Husbandry for their assistance in the sampling.

Funding

This work was supported by Natural Science Foundation of China (No. 31760660), Natural Science Foundation of Inner Mongolia Autonomous Region (No. 2019MS03092).

Competing interests

The authors declare that there are no competing interests associated with the manuscript.

Author contribution

Bujun Mei designed the study and analyzed the RNA-seq data and drafted the manuscript. Rong Liu joined the research and improved the manuscript.

REFERENCES

- Bu, D., Yu, K., Sun, S., Xie, C., Skogerbo, G., Miao, R., Xiao, H., Liao, Q., Luo, H., Zhao, G., Zhao, H., Liu, Z., Liu, C., Chen, R. and Zhao, Y. (2012). NONCODE v3.0: Integrative annotation of long noncoding RNAs. *Nucleic Acids Res.* 40(Database issue): D210-215.
- Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., Eddy, S.R., Gardner, P.P. and Bateman, A. (2013). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* 41(Database issue): D226-232.
- Chu, Q., Zhou, B., Xu, F., Chen, R., Shen, C., Liang, T., Li, Y. and Schinckel, A.P. (2017). Genome-wide differential mRNA expression profiles in follicles of two breeds and at two stages of estrus cycle of gilts. *Sci. Rep.* 7(1): 5052.
- Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M. and Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 21(18): 3674-3676.
- Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J. and Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* 44(D1): D279-285.
- Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34(Database issue): D140-144.
- Kim, D., Langmead, B. and Salzberg, S.L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nat Methods.* 12(4): 357-360.
- Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L. and Gao, G. (2007). CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35 (Web Server issue): W345-349.
- Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 9(4): 357-359.
- Li, B. and Dewey, C.N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 12: 323.
- Malik, Z.S., Dalal, D.S., Patil, C.S. and Dahiya, S.P. (2017). Genetic studies on growth, reproduction and wool production traits in Harnali sheep. *Indian Journal of Animal Research.* 51(5): 813-816.
- Nawrocki, E.P., Kolbe, D.L. and Eddy, S.R. (2009). Infernal 1.0: Inference of RNA alignments. *Bioinformatics.* 25(10): 1335-1337.
- Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33(3): 290-295.
- Wang, L., Feng, Z., Wang, X., Wang, X. and Zhang, X. (2010). DEGseq: An R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics.* 26(1): 136-138.
- Wei, Y., Lang, J., Zhang, Q., Yang, C.R., Zhao, Z. A., Zhang, Y., Du, Y. and Sun, Y. (2019). DNA methylation analysis and editing in single mammalian oocytes. *Proceedings of the National Academy of Sciences of the United States of America.* 116(20): 9883-9892.