



Analysis of differential expression of hair follicle tissue transcriptome in Hetian sheep undergoing different periodic changes

Xueyan Chen, SunShuang Sun and Yiming Sulaiman

College of Animal Science, Xinjiang Agricultural University, Urumqi, Xinjiang, China

ABSTRACT

Background. This study provides new information on long non-coding RNA (lncRNA) and messenger RNA (mRNA) expression profiles in the hair follicles of Hetian sheep *via* the sequencing and analysis of the transcriptome of skin hair follicles during three periods of periodicity change. This is important for improving the quality of carpet wool, providing a preliminary basis for further research on the targeting relationship of these mRNAs and their target genes, and providing a scientific basis for marker-assisted selection of Hetian sheep.

Methods. The periodic variation of anagen (P I, May, $n = 3$), catagen (p II, October, $n = 3$), and telogen (p III, January, $n = 3$) of the skin hair follicle tissue of three Hetian sheep ewes were selected. Skin samples were collected from the right mid-side of each sampled sheep at three hair follicle developmental stages. The three sheep were used for each developmental stage as biological and technical replicates for transcriptome sequencing and analysis.

Results. The statistical power of this experimental design, calculated in RNASeqPower, was 0.92. Differential expression analysis revealed 81 lncRNAs that were differentially expressed (46 up-regulated and 35 down-regulated) and 129 mRNAs that were differentially expressed (46 up-regulated and 83 down-regulated) during the PI and PII periods. Between the PI and PIII periods, a total of 144 differentially expressed lncRNAs and 693 differentially expressed mRNAs were identified. Of these, 73 lncRNAs were significantly up-regulated and 71 were significantly down-regulated, while 474 mRNAs were significantly up-regulated and 219 were down-regulated. Additionally, a total of 87 lncRNAs were found to be differentially expressed, with 40 up-regulated and 47 down-regulated, along with 39 differentially expressed mRNAs (23 up-regulated and 16 down-regulated), between the PII and PIII stages. The functional assessment revealed that the mRNA expressed in the cells is related to the membrane, cell processes, metabolism, extracellular region, and other GO items. It is enriched in thyroid hormone synthesis, choline metabolism, cancer, AMPK, Hedgehog, and other signaling pathways.

Conclusion. A total of 2,286 lncRNAs (including 965 known and 1,321 novel lncRNAs) and 20,879 mRNAs were identified. These co-expressed differentially expressed genes could be used as candidate genes for studying the periodic changes of the hair follicles in Hetian sheep.

Submitted 7 May 2024
Accepted 28 October 2024
Published 25 November 2024

Corresponding author
Yiming Sulaiman,
ysulaiman@xjau.edu.cn

Academic editor
Fernando Mata

Additional Information and
Declarations can be found on
page 15

DOI 10.7717/peerj.18542

© Copyright
2024 Chen et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Agricultural Science, Genetics, Molecular Biology, Zoology

Keywords Hetian Sheep, lncRNAs, Hair follicles, mRNA

INTRODUCTION

Hetian sheep are a local breed of short-tailed heterogeneous semi-rough sheep that have economic advantages such as resistance to drought and heat, coarse feeding, and diseases (Shi et al., 2022; Yu et al., 2011). Their white and light-colored coats have long lines and high fiber elasticity that make them a high-quality raw material for weaving carpets and jacquard blankets (Wang et al., 2021; Yu et al., 2013). However, after many years of breeding and selection, the existing Hetian sheep population has poor wool qualities such as low wool yield, less fine bottom wool, and more dry and dead wool (Zhao et al., 2019a; Zhao et al., 2019b). Wool quality and yield are mainly determined by the hair follicle (Matsuzaki & Yoshizato, 1998), which has the characteristics of periodic growth. To improve wool yield and quality, it is essential to determine the periodicity of hair follicle changes in Hetian sheep and investigate the molecular mechanisms that regulate hair follicle development.

Hair follicles undergo periodic growth, which is mainly induced by changes in external temperature (Panteleyev, Paus & Christiano, 2000). The growth cycle is consistent with the annual season and includes three stages: anagen, catagen, and telogen (Yin, 2009). For mammals such as cattle and sheep, the growing season typically occurs between April and November of each year. The catagen phase usually takes place between December to January of the following year (Wu & Lu, 2015). Wang (2018) studied 4-month-old plain and mountainous Hetian sheep and used morphological and qPCR methods to determine that KAP4.2 was highly expressed in the skin. The total follicle density growth trend of Hetian sheep was found to be evident between February to August. In the growing season (February), the growth activity of hair follicles was weak, while the growth activity of hair follicles is better and strong in the 4 to 8 months group. The hair follicle density of the mountain-type Hetian sheep between June and August show significant growth because mountain-type Hetian sheep live at an altitude of 1,900 m, and temperatures vary widely. After years of observational studies, our research team determined that mountain-type Hetian sheep hair follicle growth occurs in May, the general catagen occurs in October, and January is the resting period (Wang, 2018).

It has been reported that a number of specific genes are selectively expressed or inhibited in specific cells of the hair follicle, and a large number of regulatory factors such as miRNA and long non-coding RNAs (lncRNAs) are involved in this process (Lei et al., 2012; Shanhe et al., 2017; Tian et al., 2024). Epidermal growth factor and nerve growth factor are involved in the process of hair follicle development. The WNT, Notch, and TGF- β signaling pathways may play an important role in the growth and development of hair follicles (Zhao et al., 2019a; Zhao et al., 2019b).

lncRNAs are a class of non-coding RNAs larger than 200 nt (Dunham, Kundaje & Aldred, 2012) and the most abundant non-coding RNA in eukaryotes. Depending on the cis-acting regulatory mode, lncRNAs may affect the transcription of its surrounding genes. In trans-acting mode, lncRNAs can regulate the expression of genes distally or in other chromosomes. This combination of genes affects the translation of downstream proteins, which play biological functions and participate in various biological processes such as cell differentiation, apoptosis, growth, and development

(*Mao et al., 2021; Marchese, Raimondi & Huarte, 2017*). lncRNAs are important in regulating hair follicle development. Compared to telogen, 41 lncRNAs were upregulated and 157 lncRNAs were downregulated in anagen. Meanwhile, the effects of lncRNAs were detected with the target genes identified through cis- and trans-acting regulatory modes, suggesting that these lncRNAs have certain functions in the development and regulation of the hair follicle cycle. Fifteen significant differentially-expressed lncRNAs have been identified by analyzing the expression profile of the lncRNAs. These crucial differentially expressed lncRNAs may be involved in the initial molecular mechanism that regulates hair follicle development (*Yue et al., 2016*). *Guo (2015)* identified 6,127 lncRNAs in growing and resting cashmere using the RNA-Seq method, including 54 differentially expressed lncRNAs. These differentially expressed lncRNAs were found to regulate the periodic growth process of cashmere.

To date, no reports have been made on the screening and correlation analysis of differentially expressed lncRNAs during the periodic changes of Hetian sheep hair follicles. Additionally, the molecular mechanism behind Hetian wool growth and development, as well as the formation of high-quality carpet wool, remains unclear. This study utilized RNA-Seq technology to identify differentially expressed lncRNAs and mRNAs in the hair follicles of Hetian sheep during different periods of hair follicle periodic changes. A regulatory network was constructed, and potential biological functions of the differentially expressed lncRNAs were identified through bioinformatics analysis.

MATERIALS & METHODS

Ethics statement

The methods were performed in accordance with the guidelines of good experimental practices adopted by the Institute of Animal Husbandry. All experimental protocols were approved by the Animal Ethics Committee (AEC) of Xinjiang Agricultural University, China on April 20, 2020 (20200420).

Sample collection and RNA extraction

The research samples came from the Kunlun sheep farm in Yutian County, Xinjiang, China ($39^{\circ}29'$ to $81^{\circ}9'$ to $82^{\circ}51'E$, $35^{\circ}14'$ to $39^{\circ}29'N$). Three female Hetian sheep aged 12 months under the same feeding and management conditions from the same flock were randomly selected. The feeding and management conditions were relatively the same for all the animals. Skin samples were collected from the right mid-side of each sampled sheep at the three hair follicle developmental stages (anagen, catagen, and telogen). The three sheep were used for each developmental stage as biological and technical replicates. After hair shearing and alcohol disinfection, approximately one cm^2 of skin samples were collected from the same shoulder region of each sheep at different stages of hair follicle growth period, starting with anagen (P1, May, $n = 3$), then catagen (P2, October, $n = 3$), and telogen (P3, January of the following year, $n = 3$). Three pieces of back skin tissue from each sheep were taken without analgesia according to experimental protocols (20200420). The skin tissues were rinsed in ice-cold DEPC-treated water, cut into small pieces, submerged in RNAlater (ABI, Foster City, CA, USA), and frozen at -70°C until RNA and total protein

extraction. After each sampling, the ewes were well cared for in the later stage. The three sampled sheep were alive after sampling. All experimental procedures with sheep used in the present study were given prior approval by the Animal Ethics Committee (AEC) of Xinjiang Agricultural University under contract (20200420).

High-throughput sequencing and analysis of the transcriptome

A total of nine libraries in the three time periods mentioned above were constructed using Illumina Xten double terminal sequencing at Shanghai Jingzhou Gene Technology Co., Ltd. The quality of the raw sequencing data was evaluated, and the sequences were filtered using fastp (Chen *et al.*, 2018). The raw data were uploaded to the NCBI database (<https://www.ncbi.nlm.nih.gov/sra/PRJNA1095361>). Hisat2 (Kim *et al.*, 2019) was used to align the genome sequence, and the reference genome was Oar_V3.1.91 (https://useast.ensembl.org/Ovis_aries_texel/Info/Index; ftp://ftp.ensembl.org/pub/release-91/fasta/ovis_aries/dna/Ovis_aries.Oar_v3.1.dna.toplevel.fa.gz). StringTie (Pertea *et al.*, 2016) was used to compare the fragments within each gene segment after alignment. The trimmed mean of M values (TMM) algorithm was then used for normalization, and finally, the FPKM value of each gene was calculated. Gffcompare (Pertea & Pertea, 2020) was used to compare the results obtained by StringTie splicing with the genome.

In the first application, the fragments within each gene segment after comparison were counted using StringTie, normalized using the TMM algorithm, and the FPKM value for each gene was calculated. The results were stored in an Excel file and the table gave the FPKM value for each gene or transcript in addition to the number of reads aligned to the gene (shown in the [Supplementary Files](#)). For sequences sequenced to specific genes in the genome, the software edgeR was used for statistical quantification, and the number of sequenced reads corresponding to each gene and homogenizing between samples was calculated, a process known as gene quantification. On this basis, the differential genes obtained from the comparison of the two groups were calculated based on the information of the experimental groups. This includes gene quantification and transcript quantification, and both coding transcripts (mRNA) and non-coding transcripts (lncRNA). For each sample, the expression of each gene is calculated using FPKM as the count quantity, which is a value that can be used for comparison between different samples. Count quantity represents the specific number of reads compared to each gene. It calculates the sequence alignment for each transcript using the transcript as a reference (Ioanna *et al.*, 2018; Griffith-Jones *et al.*, 2006).

Screening of differentially expressed lncRNAs and mRNA

The I, II, and III set samples represent anagen, regression, and telogen, respectively. PI, PII, and PIII represent the averages of three samples in the same period (01180635 I, 00617149 I, and 01180656 I, respectively) and were divided into three groups for comparison: PII *vs.* PI represented the control group, PII represented the treatment group, and PI was the reference group; PIII *vs.* PI represented the control group, PIII represented the treatment group, PI represented the reference group; and PIII *vs.* PII represented the control group, PIII represented the treatment group, and PII represented the reference group. Differential

expression between sample treatment and reference groups was analyzed using EdgeR (Robinson, McCarthy & Smyth, 2010). The screening criteria were Q-value (corrected P -value) < 0.05 , and the expression value was upregulated two-fold (fold-change ≥ 2) or downregulated 2-fold (fold-change ≤ 0.5).

Prediction of the differentially expressed lncRNA target genes

The interaction between the lncRNAs and mRNAs was identified by both cis- and trans-acting regulatory modes. The cis-acting target genes of lncRNAs were predicted by searching the mRNA upstream and downstream of lncRNAs within the range of 10 kb on the genome as the target gene of lncRNAs. The trans-target genes were predicted based on the principle of sequence complementation matching. The complementary mRNAs were compared with the lncRNAs using BLAST, and the thermodynamic parameter values of the lncRNA and mRNA complementary pairing were calculated using RNAplex software (Tafer & Hofacker, 2008). The results above the software threshold were selected as the target genes of the lncRNAs. The differentially expressed mRNA was isolated from the target genes of the differentially expressed lncRNAs. GO and KOBAS 2.0 (<https://bio.tools/kobas>) were used for annotation of differential genes (Chen et al., 2011). The part of differential mRNA in the target gene of differential lncRNAs was isolated, and the relationship of interaction was determined according to the target relationship of lncRNAs-mRNA. The Cystoscope software (Shannon et al., 2003) was used for visual analysis. The miRanda tools (<http://mirtoolsgallery.tech/mirtoolsgallery/node/1055>) were used for the co-expression analyses, and the first 400 interactions with a correlation coefficient greater than 0.7 and $P < 0.05$ were selected, while those with a correlation coefficient less than 400 were subjected to the actual conditions.

In order to avoid false positive results, the enrichment analysis result P was corrected by multiple tests (false discovery rate, FDR). When $FDR \leq 0.05$, GO terms and pathways were defined as significant enrichment of candidate genes.

Quantitative real-time polymerase chain reaction

In order to verify the accuracy of transcriptome sequencing, 10 of the differentially expressed lncRNAs and 10 of the differentially expressed mRNAs were randomly selected, with large differential expression and high CPM values between PIII and PI periods detected for relative expression analysis using qRT-PCR. RT-qPCR was performed in 96-well plates (Thermo Fisher Scientific, Waltham, MA, USA) on the StepOnePlus system (Applied Biosystems, Waltham, MA, USA). Primer sequences and characteristics are shown in Table S3. The reaction mix was performed using 0.3 μ l of FastStart Universal SYBR Green Master (Thermo Fisher Scientific), 1 μ l of 10 μ M primer mix, 1 μ l of a diluted 1:10 cDNA, and water for a final volume of 20 μ l. Cycling conditions were 95 °C for 1 minute, and 40 cycles of 95 °C for 5 s, and 60 °C for 40 s. All RT-qPCR experiments were performed using three biological and three technical replicates. The upstream primers for qRT-PCR were designed according to the mature sequences and were quantified in real time using the SYBR Green dye method and the expression of Actin as the internal reference gene. Three replicates were set up for each sample. The results were used to calculate the relative

Table 1 Clean statistics of raw reads.

Sample	Total reads	Clean reads	Clean ratio	no rRNA	rRNA ratio	no rRNA pair
S00617149I	90,982,528	89,864,784	98.77%	89,759,340	0.12%	89,759,340
S00617149II	96,468,902	95,843,912	99.35%	90,462,322	5.61%	90,462,322
S00617149III	86,368,040	85,588,616	99.10%	84,823,360	0.89%	84,823,360
S01180635I	82,175,600	81,447,962	99.11%	79,434,522	2.47%	79,434,522
S01180635II	84,978,642	83,951,322	98.79%	83,198,370	0.90%	83,198,370
S01180635III	86,625,880	85,687,400	98.92%	85,629,238	0.07%	85,629,238
S01180656I	91,730,370	90,951,556	99.15%	87,969,762	3.28%	87,969,762
S01180656II	86,157,106	85,329,020	99.04%	84,280,136	1.23%	84,280,136
S01180656III	82,772,778	81,732,942	98.74%	80,996,134	0.90%	80,996,134

Table 2 Cross-reference the information with all relevant databases.

Sample.ID	Clean.Reads	miRBase matureTotal.A.R	Rfam Total.A.R	GenomeTotal.A.R	Genome Total.A.R.P	Uniq. Reads	Uniq.A.R	Uniq.A.R.P
S0617149I	24,734,835	9,190,105	19,088,781	21,241,263	85.88%	1,261,644	1,073,735	85.11%
S0617149II	19,905,899	7,732,352	15,797,163	17,385,377	87.34%	1,047,837	924,456	88.23%
S0617149III	20,835,097	8,425,210	17,151,450	18,176,324	87.24%	566,931	375,684	66.27%
S1180635I	18,071,465	5,987,699	14,767,751	15,853,567	87.73%	574,681	464,704	80.86%
S1180635II	23,247,853	9,133,227	19,268,227	20,326,162	87.43%	576,101	380,432	66.04%
S1180635III	21,133,345	8,680,199	17,706,242	18,554,892	87.80%	487,585	324,435	66.54%
S1180656I	19,437,266	5,773,534	12,501,434	14,892,836	76.62%	1,092,439	589,932	54.00%
S1180656II	22,195,423	8,453,494	18,120,710	19,395,202	87.38%	790,493	608,143	76.93%
S1180656III	23,481,915	10,189,875	20,041,164	21,070,478	89.73%	676,323	428,483	63.35%

expression of target genes using the $2^{-\Delta\Delta Ct}$ method, where $\Delta\Delta Ct = Ct$ target gene— Ct internal reference gene, and $\Delta\Delta Ct = \Delta\Delta Ct$ experimental group— $\Delta\Delta Ct$ control group with 01180635 I as the control group. All the results were compared with the sequencing results, presented in Microsoft Excel 2016, and were uploaded as a separate zip file.

RESULTS

Original data and quality control

After removing the raw sequencing data and passing through the filtering process, a total of 21,499,205 bp high quality data were obtained on average for each sample, the quality of the data after filtering the raw data was more than 86%, and the Q30 was more than 90% for all samples (Table 1 and Tables S1 and S2). The number of reads grows exponentially with respect to the coverage of the genome and then converges to a certain ratio of saturation.

Sequence alignment

The statistical power of this experimental design, calculated in RNASeqPower, was 0.92. Clean reads were aligned to the miRBase (<https://www.mirbase.org/>), Rfam (<http://rfam.org/>), and Genome databases using Hisat2, and the alignment results are shown in Table 2.

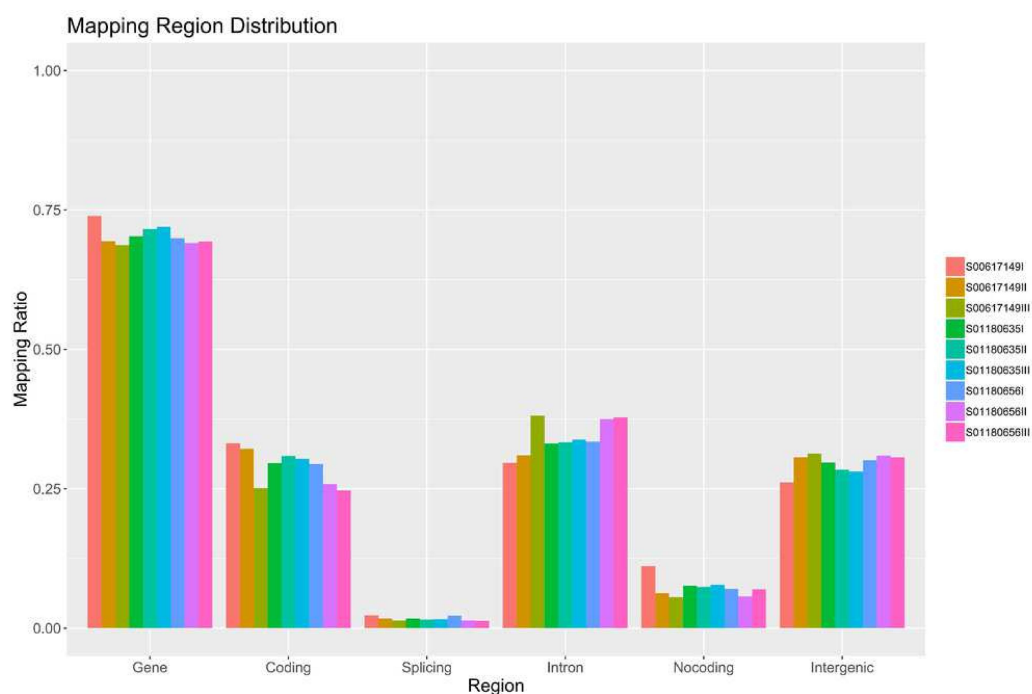


Figure 1 Functional region distribution of the sequences on the genome.

Full-size DOI: [10.7717/peerj.18542/fig-1](https://doi.org/10.7717/peerj.18542/fig-1)

8,173,966 reads were aligned to the miRBase database for each sample on average, and the average comparable reads for each sample were 37.90%. A total of 18,544,011 reads were aligned to the Genome database for each sample on average, and the average comparable reads for each sample were 86.35%, with 73.09% having only one unique alignment position on the reference sequence. A total of 17,160,325 reads were aligned to the Rfam database for each sample on average, and the average number of comparable reads for each sample was 79.8%. According to the alignment results, reads of different lengths were extracted for statistics, which could visually show the composition of the data under different lengths. [Figure 1](#) illustrates the specific locations on the genome where the sequences obtained by sequencing were compared to the genome, *i.e.*, their distribution in different functional regions of the genome.

By examining the expression correlation among the samples, we found that the square of the Pearson correlation coefficient (R^2) was greater than 0.93. The results of gene quantification are shown in [Figs. 2](#) and [3](#) and indicate that the expression patterns among the samples were very similar, with high test repeatability and low variation, which allowed for the next test analysis. Principal component analysis (PCA) is designed to assess a specific grouping of samples and determine whether the distribution of samples in the results is consistent with the experimental design groupings, and was used to demonstrate the relationship and variation among samples ([Fig. 4](#)).

From the alignment results of the three databases, the small RNA fragments were classified and annotated, and their number was counted (uploaded as [Supplemental Files](#)).

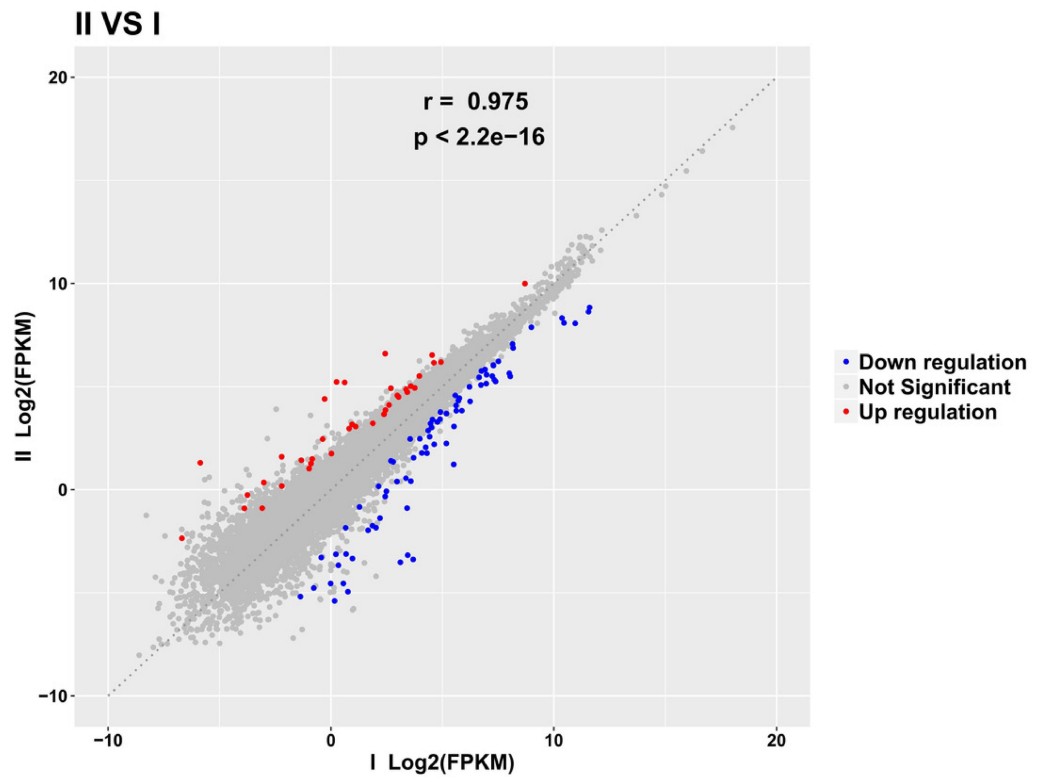


Figure 2 The mRNA scatter diagram showing the differential expression of PII vs. PI.

Full-size DOI: 10.7717/peerj.18542/fig-2

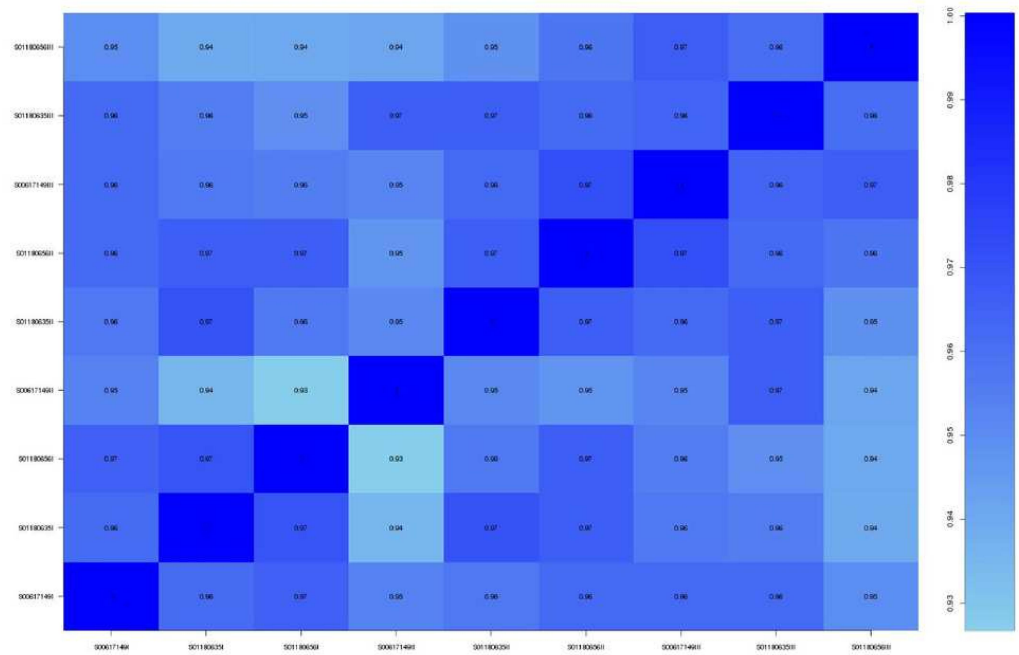


Figure 3 Correlation test between samples.

Full-size DOI: 10.7717/peerj.18542/fig-3

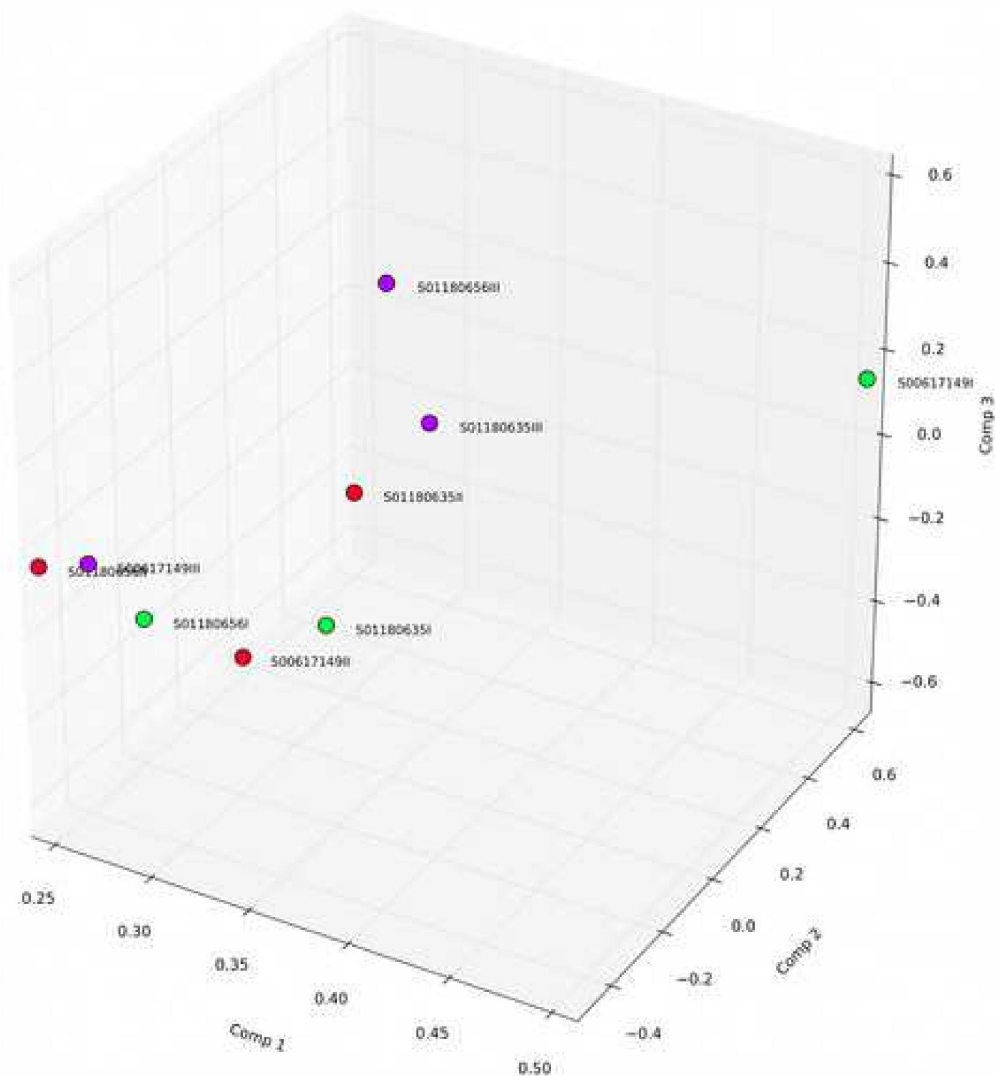


Figure 4 Principal component analysis.

Full-size  DOI: [10.7717/peerj.18542/fig-4](https://doi.org/10.7717/peerj.18542/fig-4)

The proportion of comparison between all samples and the genome was found to be more than 98.83%, more than 98.30% of all samples had comparison pairs on the reference sequence, and the unique comparison rate on the reference sequence was 0.07–0.53%. As shown in Fig. 1, the sequences were mainly distributed in the alignment gene, coding, splice site, intron, non-coding, and intervening regions. The reads had the highest proportion in the gene region and the lowest proportion in the splicing site.

Screening of the differentially expressed lncRNAs and mRNAs

According to the screening threshold, a total of 81 differentially expressed lncRNAs were identified in the PI and PII stages (Fig. 5), including 46 upregulated lncRNAs and 35 downregulated lncRNAs. In PI and PIII, 144 differentially expressed lncRNAs

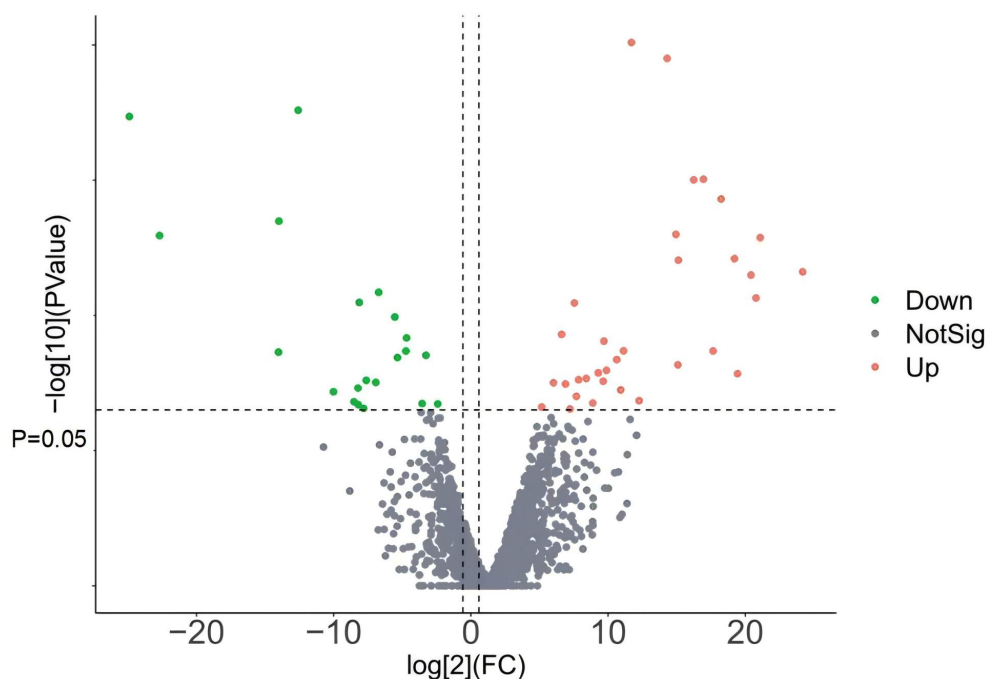


Figure 5 LncRNAs volcano of P II vs P I.

Full-size DOI: 10.7717/peerj.18542/fig-5

were identified, including 73 significantly upregulated lncRNAs and 71 significantly downregulated lncRNAs. In P II and P III, 87 differentially expressed lncRNAs were identified, including 40 upregulated and 47 downregulated lncRNAs.

A total of 129 differentially expressed mRNAs were identified in the P I and P II stages, including 46 upregulated and 83 downregulated mRNAs. In the P I and P III stages, 693 differentially expressed mRNAs were identified, including 474 significantly upregulated mRNAs and 219 significantly downregulated mRNAs. Additionally, 39 differentially expressed mRNAs were identified at the P II and P III stage, including 23 upregulated and 16 downregulated mRNAs. The KIF7 gene from the KIF family, KRT84, KRT15, and KRTAP11-1 from the KRT family, as well as HOXB5 from the HOX family, are associated with hair characteristics.

Prediction of the target genes of the lncRNAs

The prediction of cis targets for the identified 1,509 cis-acting lncRNAs resulted in 1,199 potential target genes. Additionally, trans targets were predicted for 627 trans-acting lncRNAs, resulting in 1,994 potential target genes. Previous studies have shown that many of the predicted target genes, such as KRT84, KRT25, and KRT15 of the KRT family; HOXA1, HOXB2, HOXD3 of the HOX family; and BMP3 of the BMP family are related to hair traits. Additionally, 958 target genes were predicted to differentially express lncRNAs at P I and P II, while 1,076 target genes were predicted to differentially express lncRNAs at P I and P III. Furthermore, 795 target genes were predicted to be differentially expressed by the lncRNAs at P II and P III.

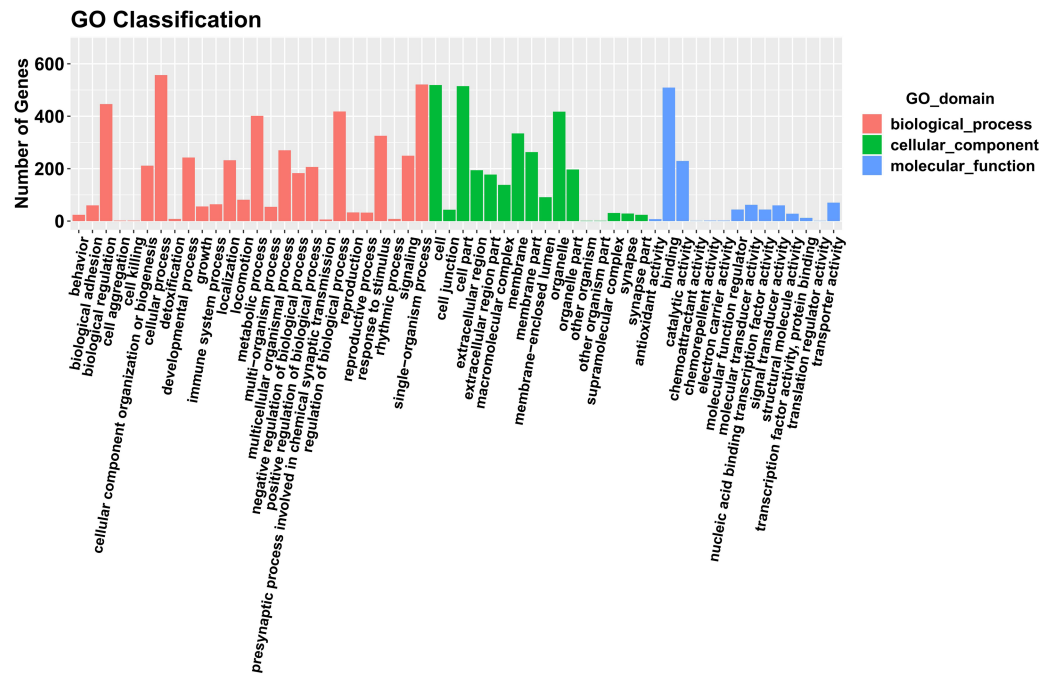


Figure 6 The GO enrichment diagram of the combined analyses of the lncRNA target genes and mRNA with differential expression of PIII vs. PI. Genes involved in biological processes are shown in red. Genes contained in the cellular component are shown in green. Genes included in the molecular function are shown in blue.

Full-size DOI: 10.7717/peerj.18542/fig-6

Enrichment analysis of GO and KEGG

Our analysis predicted target genes and differentially expressed mRNAs, which were mainly annotated in the cell, membrane part, cellular process, metabolic process, extracellular region, and other GO items. Bar charts in Fig. 6 represent significantly enriched GO items in each period of comparison. The KEGG enrichment analysis revealed significant enrichment in thyroid hormone synthesis, choline metabolism in cancer, AMPK signaling pathway, and the Hedgehog signaling pathway. Figure 7 displays the KEGG function distribution of PIII vs. PI target gene and mRNA combined analyses.

Differentially expressed lncRNAs and mRNA co-expression network

Figure 8 displays the co-expression network of the lncRNAs-mRNA between the comparison groups. Red dots represent differentially expressed lncRNAs and blue dots represent differentially expressed mRNAs, with larger dots indicating more interactions they are involved in. This suggests that Hetian sheep miRNAs have multiple targets. Common genes were found in all three comparison periods, including differentially expressed ENSOARG0000007090 and MSTRG.25673.2 in lncRNAs, as well as differentially expressed CYP2F1 and MTERF4. These genes played a crucial role in altering the hair follicles of the sheep during the three periods and could be considered as potential candidate genes for further analysis. Additionally, genes with large dots and dense interaction lines in the network can also be analyzed as candidate genes.

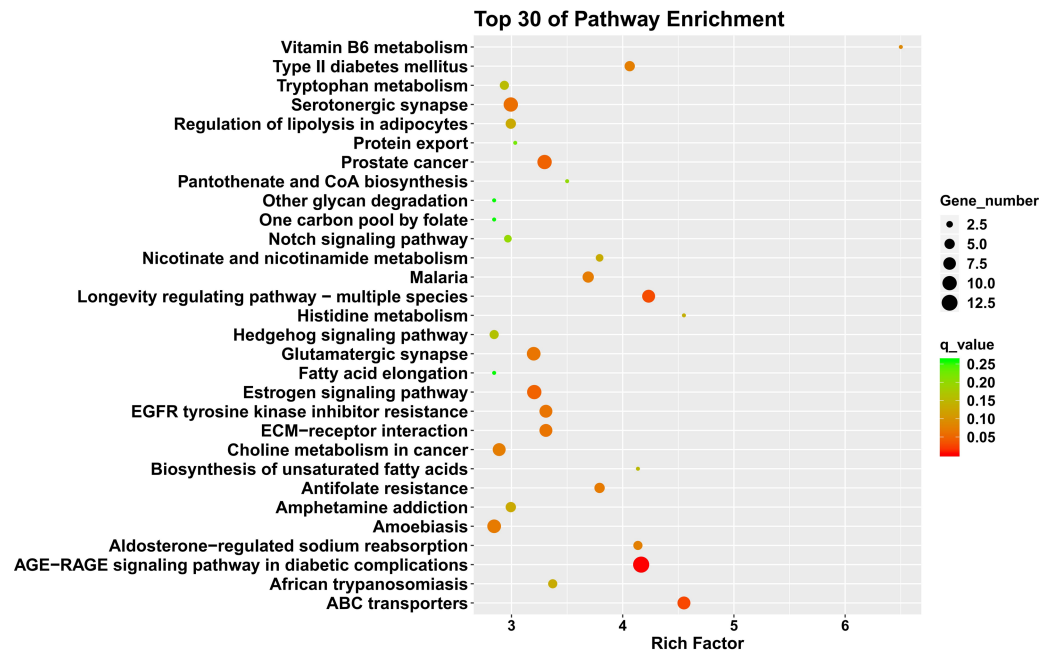


Figure 7 KEGG function distribution of the PIII vs. PI target gene and mRNA combined analysis. The magnitude of the qvalue of the color response. Going from green to red indicates an increasingly significant degree of enrichment.

Full-size DOI: 10.7717/peerj.18542/fig-7

qRT-PCR verification

The sequencing results were compared and graphically represented using Excel (uploaded as [Supplemental Files](#)) and the results are shown in [Figs. 9](#) and [10](#). The expression trend was consistent with the RNA-Seq results, indicating the accuracy and reliability of the genome sequencing results.

DISCUSSION

In recent years, there have been significant developments in high-throughput sequencing and genome research. The RNA-Seq technology has been widely used to discover and identify new lncRNAs in different species (*Consortium SM-I, 2014*). RNA-Seq technology has been used by researchers to identify lncRNAs in developing chicken muscle (*Li et al., 2012*). However, there have been few studies on lncRNA in Hetian sheep, a native breed of Xinjiang. To investigate the variation in transcriptome expression and skin tissue during the periodic changes of the hair follicle in Hetian sheep, we utilized high-throughput sequencing to analyze the differential expression of lncRNAs and mRNAs across three distinct periods. Additionally, we constructed a network regulation diagram. This research aims to provide a theoretical basis for the molecular mechanism of hair follicle development in Hetian sheep and offer a reference for molecular-assisted breeding.

Upon comparing the transcripts and differential expressions of lncRNAs and mRNAs, we discovered that the number of mRNAs exceeded that of lncRNAs. Additionally, lncRNAs were found to be tissue-specific and expressed at a lower level than the protein-coding

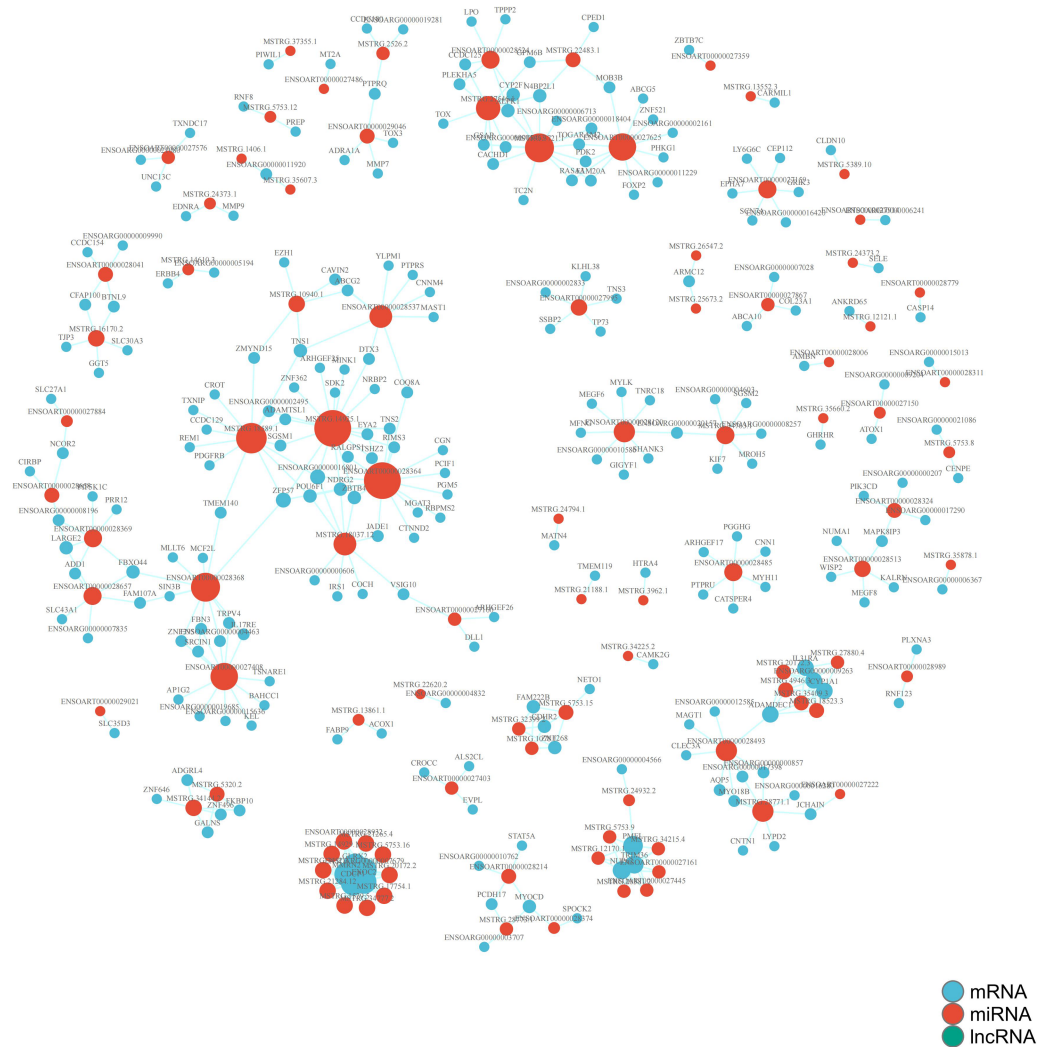


Figure 8 The lncRNA-mRNA co-expression network between the PIII vs PI comparison groups. Red dots represent differentially expressed lncRNAs and blue dots represent differentially expressed mRNAs, with larger dots indicating more interactions they are involved in.

Full-size [DOI: 10.7717/peerj.18542/fig-8](https://doi.org/10.7717/peerj.18542/fig-8)

genes (Mercer *et al.*, 2008; Derrien *et al.*, 2012; Kutter *et al.*, 2012). The study identified 2,286 candidate lncRNA transcripts and 20,879 mRNA transcripts from three periods of periodic hair follicle changes. The expression levels of lncRNAs were lower than those of protein-coding genes. Additionally, the genes were found to express differently during different periods. The group comparing anagen and telogen showed the most differentially expressed genes, indicating tissue- and time-specific gene expression.

When comparing the expression of differentially expressed genes at each period, we found that there were more differentially expressed genes in the telogen and anagen phases than in the other two comparison groups. This suggests that more differentially expressed genes are involved in the transition from telogen to anagen. Compared to the anagen phase, the number of genes involved in telogen increased, and the number of upregulated genes

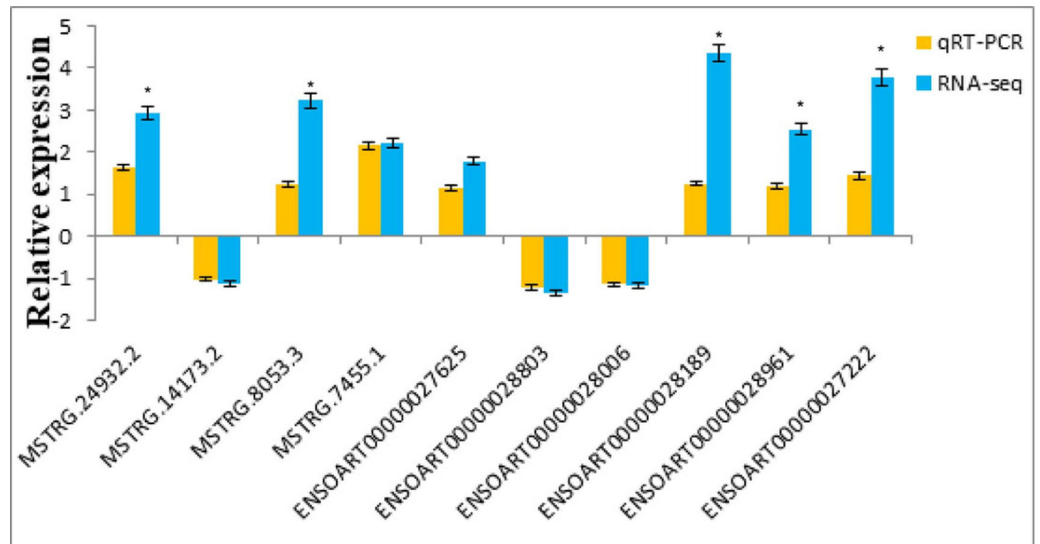


Figure 9 Comparison of the qRT-PCR and RNA-Seq of lncRNAs.

Full-size [DOI: 10.7717/peerj.18542/fig-9](https://doi.org/10.7717/peerj.18542/fig-9)

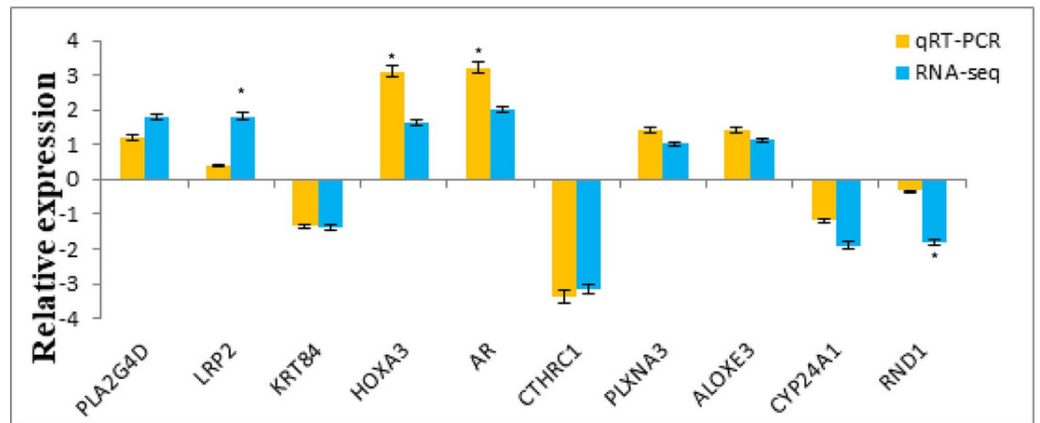


Figure 10 Comparison of the mRNA qRT-PCR and RNA-seq.

Full-size [DOI: 10.7717/peerj.18542/fig-10](https://doi.org/10.7717/peerj.18542/fig-10)

was also consistent. *Li et al. (2014)* found that among the genes with different expression levels, most of the genes related to the cell cycle were upregulated in telogen. *Guo (2015)* discovered that among the 55 differentially expressed lncRNAs in the anagen and telogen of cashmere, 32 lncRNAs were upregulated and 23 lncRNAs were downregulated during telogen. This finding is consistent with previous research.

Research has confirmed that the Hedgehog signaling pathway's sonic hedgehog can regulate the anagen phase of hair follicles and the transition from telogen to anagen (*Yang et al., 2024*). The transmission of signals from the cell surface to the MAPK signaling pathway in the nucleus is crucial for the periodic changes of the hair follicles (*Drew et al., 2007*). The co-expression analysis of differentially expressed lncRNAs and mRNA revealed

significant enrichment in the AMPK and Hedgehog signaling pathways, which is consistent with previous studies. The studies suggest that lncRNAs may regulate downstream target mRNA expression through the pathway during periodic changes of hair follicles. Further research should verify the relationship between differentially expressed lncRNAs and mRNAs using the dual luciferase-reporter gene system and conduct genome verification on the mRNAs. Finally, this study selected the key candidate genes for Hetian wool traits to provide molecular-assisted markers for breeding Hetian sheep. The study provides a preliminary basis for more in-depth investigation of the relationship between lncRNAs and targeted mRNA in Hetian sheep. However, further analysis and verification are necessary to determine the specific regulatory mechanism.

CONCLUSION

This study identified 965 known lncRNAs and predicted 1,321 new lncRNAs through comparative analysis of transcriptional libraries obtained from sequencing Hetian sheep hair follicles at different stages of hair follicle growth period during anagen, regression, and telogen. Additionally, 20,879 mRNAs were identified, further enriching the sheep lncRNA and mRNA information database. The study found differential expression of the two at different periods. Additionally, bioinformatics tools were used to predict the network structure and function of key targets and pathways related to the Hetian wool traits of these sheep. This provides a new perspective for future studies.

ACKNOWLEDGEMENTS

We would like to thank all the participants of the study. We would also like to thank LEXIS for language editing.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Xinjiang Natural Science Foundation Project “Preliminary analysis of Key microRNA regulatory network for carpet Wool Traits in Hetian Sheep” (2017D01A33) and the “Xinjiang Agriculture Research System”. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Xinjiang Natural Science Foundation Project “Preliminary analysis of Key microRNA regulatory network for carpet Wool Traits in Hetian Sheep”: (2017D01A33).
“Xinjiang Agriculture Research System”.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Xueyan Chen conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- SunShuang Sun conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Yiming Sulaiman conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Animal Ethics

The following information was supplied relating to ethical approvals (*i.e.*, approving body and any reference numbers):

This study was authorized by the Animal Ethics Committee (AEC) of Xinjiang Agricultural University, China (20200420).

Data Availability

The following information was supplied regarding data availability:

The sequence reads are available at GenBank: [PRJNA1095361](https://www.ncbi.nlm.nih.gov/nuclseq/PRJNA1095361).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.18542#supplemental-information>.

REFERENCES

- Chen X, Xizeng M, Jiaju H, Yang D, Jianmin W, Shan D, Lei K, Ge G, Chuan-Yun L, Liping W. 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research* 39(Web Server issue):W316–W322 DOI 10.1093/nar/gkr483.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34(17):i884–i890 DOI 10.1093/bioinformatics/bty560.
- Consortium SM-I. 2014. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the sequencing quality control consortium. *Nature Biotechnology* 32:903–914 DOI 10.1038/nbt.2957.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Guigó R. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research* 22(9):1775–1789 DOI 10.1101/gr.132159.111.
- Drew CF, Lin CM, Jiang TX, Blunt G, Mou C, Chuong CM, Headon DJ. 2007. The Edar subfamily in feather placode formation. *Developmental Biology* 305(1):232–245 DOI 10.1016/j.ydbio.2007.02.011.
- Dunham I, Kundaje A, Aldred SF. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74 DOI 10.1038/nature11247.

- Griffith-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. 2006.** miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research* **34(Database issue):**D140–D144 DOI [10.1093/nar/gkj112](https://doi.org/10.1093/nar/gkj112).
- Guo Y. 2015.** Screening of periodic growth specific lncRNAs in cashmere. Master's degree, Northwest A & F University, Yangling, China (in Chinese).
- Ioanna K, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, Bateman A, Petrov AI. 2018.** Non-coding RNA analysis using the Rfam database. *Current Protocols in Bioinformatics* **62(1):**e51 DOI [10.1002/cpbi.51](https://doi.org/10.1002/cpbi.51).
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019.** Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37(8):**907–915 DOI [10.1038/s41587-019-0201-4](https://doi.org/10.1038/s41587-019-0201-4).
- Kutter C, Watt S, Stefflova K, Wilson MD, Goncalves A, Ponting CP, Odom DT, Marques AC. 2012.** Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLOS Genetics* **8(7):**e1002841 DOI [10.1371/journal.pgen.1002841](https://doi.org/10.1371/journal.pgen.1002841).
- Lei S, Zhihua Z, Bailey TL, Perkins AC, Tallack MR, Zhao X, Hui L. 2012.** Prediction of novel long non-coding RNAs based on RNA-Seq data of mouse Klf1 knockout study. *BMC Bioinformatics* **13:**331 DOI [10.1186/1471-2105-13-331](https://doi.org/10.1186/1471-2105-13-331).
- Li T, Wang S, Wu R, Zhou X, Zhu D, Zhang Y. 2012.** Identification of long non-protein coding RNAs in chicken skeletal muscle using next generation sequencing. *Genomics* **99(5):**292–298 DOI [10.1016/j.ygeno.2012.02.003](https://doi.org/10.1016/j.ygeno.2012.02.003).
- Li H, Yu B, Li J, Su L, Liu B. 2014.** Overexpression of lncRNA H19 enhances carcinogenesis and metastasis of gastric cancer. *Oncotarget* **5(8):**2318–2329 DOI [10.18632/oncotarget.1913](https://doi.org/10.18632/oncotarget.1913).
- Mao J, Zhou Y, Lu L, Zhang P, Ren R, Wang Y, Wang J. 2021.** Identifying a serum exosomal-associated lncRNA/circRNA-miRNA-mRNA network in coronary heart disease. *Cardiology Research and Practice* **2021:**6682183 DOI [10.1155/2021/6682183](https://doi.org/10.1155/2021/6682183).
- Marchese FP, Raimondi I, Huarte M. 2017.** The multidimensional mechanisms of long noncoding RNA function. *Genome Biology* **18(1):**206 DOI [10.1186/s13059-017-1348-2](https://doi.org/10.1186/s13059-017-1348-2).
- Matsuzaki T, Yoshizato K. 1998.** Role of hair papilla cells on induction and regeneration processes of hair follicles. *Wound Repair and Regeneration* **6(6):**524–530.
- Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. 2008.** Specific expression of long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America* **105(2):**716–721 DOI [10.1073/pnas.0706729105](https://doi.org/10.1073/pnas.0706729105).
- Panteleyev AA, Paus R, Christiano AM. 2000.** Patterns of hairless (hr) gene expression in mouse hair follicle morphogenesis and cycling. *American Journal of Pathology* **157:**1071–1079.
- Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016.** Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* **11(9):**1650–1667 DOI [10.1038/nprot.2016.095](https://doi.org/10.1038/nprot.2016.095).
- Pertea G, Pertea M. 2020.** GFF utilities: GffRead and GffCompare. *F1000 Research* **9:**ISCB Comm J-304 DOI [10.12688/f1000research.23297.2](https://doi.org/10.12688/f1000research.23297.2).

- Robinson MD, McCarthy DJ, Smyth GK. 2010.** edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–140 DOI 10.1093/bioinformatics/btp616.
- Shanhe W, Wei G, Zhixin L, Yang G, Beilei J, Lei Q, Zhiying Z, Xin W. 2017.** Integrated analysis of coding genes and non-coding RNAs during hair follicle cycle of cashmere goat (*Capra hircus*). *BMC Genomics* 18:767 DOI 10.1186/s12864-017-4145-0.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003.** Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13(11):2498–2504 DOI 10.1101/gr.1239303.
- Shi R, Li S, Liu P, Guo L, Gong S, Wan Y. 2022.** Effects of testosterone on skin structure and factors related to androgen conversion and binding in Hetian sheep. *Tropical Animal Health and Production* 54(4):218 DOI 10.1007/s11250-022-03216-5.
- Tafer H, Hofacker IL. 2008.** RNAplex: a fast tool for RNA–RNA interaction search. *Bioinformatics* 24(22):2657–2663 DOI 10.1093/bioinformatics/btn193.
- Tian D, Pei Q, Jiang H, et al. 2024.** Comprehensive analysis of the expression profiles of mRNA, lncRNA, circRNA, and miRNA in primary hair follicles of coarse sheep fetal skin. *BMC Genomics* 25(1):574.
- Wang L. 2018.** Wool follicle morphological analysis and their KAPs and KIFs gene significance difference on transcription level between plain-type and mountain-type Hetian sheep. Master's thesis, Tarim University, Alaer, China (in Chinese).
- Wang HY, Li SW, Wu TH, Wu ZH, Guo JX. 2021.** The effect of androgen on wool follicles and keratin production in Hetian sheep. *Brazilian Journal of Biology* 81(3):526–536 DOI 10.1590/1519-6984.224056.
- Wu H, Lu Q. 2015.** Epigenetic modulation of hair follicle stem cells. *Epigenetics and Dermatology* 2015:75–84 DOI 10.1016/b978-0-12-800957-4.00005-9.
- Yang H, Zhu M, Wang M, Zhou H, Zheng J, Qiu L, Fan W, Yang J, Yu Q, Yang Y, Zhang W. 2024.** Genome-wide comparative analysis reveals selection signatures for reproduction traits in prolific Suffolk sheep. *Frontiers in Genetics* 15:1404031 DOI 10.3389/fgene.2024.1404031.
- Yin ZJ. 2009.** Histological study on sheep hair follicle structure. *Agricultural Technology and Equipment* 11:48–49 (in Chinese) DOI 10.3969/j.issn.1673-887X.2009.11.031.
- Yu J, Dong H, Ji Y, Kong Y, Gu L. 2011.** Current situation and protection and utilization measures of Hetian sheep. *Herbivorous Livestock* 02:7–9.
- Yu J, Tian K, Xu X, Dong H, Chen T, Rou Z, Yuan X, Deng S. 2013.** Genetic diversity in MTDNAD-LOOP region of Hetian sheep. *Chinese Herbivore Science* 33(02):19–22.
- Yue Y, Guo T, Yuan C, Liu J, Guo J, Feng R, Niu C, Sun X, Yang B. 2016.** Integrated analysis of the roles of long noncoding RNA and coding RNA expression in sheep (*Ovis aries*) skin during initiation of secondary hair follicle. *PLOS ONE* 11(6):e0156890 DOI 10.1371/journal.pone.0156890.
- Zhao B, Chen Y, Yang N, Chen Q, Bao Z, Liu M, Hu S, Li J, Wu X. 2019a.** miR8 regulates skin and hair follicle development through Wnt/ β -atenin signaling

pathway by targeting SFRP2. *Journal of Cellular Physiology* **234(11)**:20329–20341

[DOI 10.1002/jcp.28633](https://doi.org/10.1002/jcp.28633).

Zhao X, Huang L, Li B, Li Z, Yi M. 2019b. Genetic polymorphism of KRT31 and KRT85 genes and their association with wool traits in Hetian sheep. *Herbivorous Livestock* **04**:1–9.