Genome-wide detection and sequence conservation analysis of long non-coding RNA during hair follicle cycle of yak

Xiaolan Zhang[1], Qi Bao[1], Congjun Jia[1], Chen Li[1], Yongfang Chang[1], Xiaoyun Wu[1], Chunnian Liang[1], Pengjia Bao[1]* and Ping Yan[1]*

[1]Key Laboratory of Yak Breeding Engineering Gansu Province, Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences, Lanzhou 730050, China.

*Corresponding authors: Pengjia Bao (email:baopengjia@caas.cn), Ping Yan (email: pingyanlz@163.com)

ABSTRACT

Background: Long non coding RNA (lncRNA) as important regulator has been demonstrated playing an indispensable role in the biological process of hair follicles (HFs) growth. However, their function and expression profile in hair cycling of yak are yet unknown. It only a few functional lncRNAs have been identified, partly due to the low sequence conservation and lack of identified conserved properties in lncRNAs. Here, lncRNA-seq was employed to detect the expression profile of lncRNAs during HFs cycle of yak, and the sequence conservation between yak and cashmere goat during HFs cycle was analyzed

Result: Totally 2884 lncRNAs were identified in five phases (Jan. Mar. Jun. Aug. and Oct.) during HFs cycle of yak. We then performed differential expression analysis between three phases (Jan. Mar. and Oct.) finding that 198 differently expressed lncRNAs (DELs) were obtained in Oct-vs-Jan, 280 DELs were obtained in Jan-vs-Mar, and 340 DELs were obtained in Mar-vs-Oct. Subsequently, the nearest genes of DELs were used to explore the potential function of lncRNAs by GO and KEGG enrichment analysis, several critical pathways involved in HFs development such as Wnt signaling pathway, VEGF signaling pathway and Signaling pathways regulating pluripotency of stem cells were enriched. To further screen key lncRNAs influencing in HFs cycle, twenty-four sequence conserved DELs were obained via comparison analysis of partial DELs with previously published lncRNA-seq data of cashmere goat in HFs cycle by NCBI blast-2.9.0+, and the limited degree of sequence conservation of lncRNAs were observed.

Conclusion: Our study expanded the knowledge about lncRNAs affecting on hair cycling, and the finding related to sequence conservation properties of lncRNAs between two species at the specific trait- HFs cycle will likely provide valuable insights into the study of lncRNAs functionality and mechanism.

Keywords: Yak, Hair follicle cycling, lncRNA-seq, NCBI blast-2.9.0+

Background

Yak (Bos grunniens) is the only domesticated breed in Bos genus that can produce wools, which feature also the main embodiment that yak has been regarded as "all-purpose" livestock to provide essential living and economy resource for the inhabitants around the Qinghai-Tibetan Plateau. The hair growth cycling of yak is similar with cashmere goat also presenting a seasonal pattern under the control of the animals' endocrine systems and photoperiod that including anagen (growth), catagen (regression) and telogen (rest) [1, 2], morphological changes of hair growth take place following the stages of the cyclic process [3]. Anagen phase, an entire hair shaft from follicles were produced, which stage determines the length of hair shaft. Catagen phase, the dynamic transition from anagen to talogen, hair follicle regresses in a lower cell cycling process which caused by the increased apoptosis of epithelial cells in the bulb, outer root sheath (ORS) and outermost epithelial layer. During telogen, follicles regenerate and prepare their multipotent stem cells, receiving the signals to initiate a new cycle of hair growth [4]. Transition between the stages is the result of multiple molecular signals and their intricate interactions in skin during hair follicle development. It has been reported that several signaling pathways such as Wnt, Notch, Hedgehog and bone morphogenetic protein (BMP) are crucial for the regulation of hair follicle cycle [5]. Moreover, some hormones and molecules such as melatonin, estrogen, FGF5, VEGF, and son on have been identified influencing the development of hair cycle [6-9].

Increasing studies have been demonstrated that long non-coding RNAs (lncRNAs) play vital roles in regulating various biological such as development, cell proliferation and differentiation [10, 11]. LncRNAs regulate gene expression in a wide variety of genomic levels, including chromatin modification, transcription and post-transcriptionally, consisting of multiple binding modules to affect the gene expression on the immediate genomic vicinity (in cis) or on other genomic locations (in trans) [12-14]. Studies found that lncRNAs are crucial for maintaining pluripotency and lineage commitment [15], in dermal papilla, lncRNAs regulate the gene expression associated with HFs development and postnatal hair cycling [16]. In additionally, it was reported that lncRNAs participated in the regulation of HFs cycle in cashmere goat, sheep and Angora Rabbit as well regulating skin pigmentation in goat and bovine [17-20]. So far the researches about hair follicle growth in yak mainly focused on the expression pattern of several

genes such as *BMP2*, *TGF-β* and *HSP70* [21-23], there were no systematic molecular study involved in HFs cycle on yak was reported. The hair growth pattern of yak is similar to goat, in addition, the harsh living environment with cold and oxygen-thin air makes yak some unique traits, such as the high altitude adaptation [24], skin is the important protective barrier against cold and harsh environment, the study of lncRNAs influencing on the HFs may also broaden our knowledge on alpine adaptation of yak.

The function of lncRNAs usually were predicted by analyzing the role of their potential targets in cis or trans that colocalization or correlation. Only a small fraction of lncRNAs have been experimentally tested for their function[12, 25], such as the well studied lncRNAs Xist, HOTAIR and MALAT1 that were synchronously reported with higher sequence conversation or containing conservative domain [26-28]. In fact, contrary to coding genes, lncRNAs were mostly reported at a low level of sequence conservation and lack of identified shared properties across species [12], thus lead to the study difficulty of lncRNA function. To further understand the role of lncRNAs in various biological processes, facilitating the functional study of more lncRNAs, researches about lncRNAs in position and evolution conservation were reported, it was found that lncRNAs are conversed and plasticity in evolutionary, and the lncRNAs conserved in position are primarily connect to many developmental transcription factors [29, 30], these finding of lncRNA in conservation indicated the importance of lncRNAs in genome evolutionary and development from the view of functional commonality. However, the functional exploration of individual lncRNA need to further elaborate analysis. Recently studies demonstrated the secondary structures conservation of homologous lncRNAs among interspecies [13], which highlights the significance of structure on functionality of lncRNAs, along with the species-specific and spatiotemporally expression patterns [31]. We deemed the sequence conservation of lncRNAs between interspecies on the same specific trait is worth detecting, and the growth pattern of hair follicle between yak and cashmere goat provided us a good animal model for the sequence conservation study of lncRNAs in the same trait.

The objective of this study was to investigate the lncRNAs affecting on HFs cycle of yak and detect the potential function of lncRNAs involved. In order to obtain more critical lncRNAs that involved in HFs cycle of yak, we analyzed the sequence conservation of lncRNAs between yak and cashmere goat during hair follicle cycle. This study will contribute to a systematic knowledge about lncRNAs regulating HFs

cycle of yak and screened lncRNAs that are sequence conserved between yak and goat, which would get insights into lncRNAs in hair growth biology and lay a foundation for their further functional study.

Methods

Animals and Sample collection

Tianzhu white yaks with fine fiber production trait were used in this study. All the yaks were obtained from Tianzhu white yak propagation bases of Gansu province, China The female adult yaks similar in shape (coefficient of relationship <0.125) were selected for sample collection in different hair cycle stages that divided into five points including anagen (Jun., Aug., Oct.), catagen (Jan.) and telogen(Mar.). Skin samples of three yaks in each point were collected from the body side of each group and rapidly frozen in liquid nitrogen for further processing. The samples were collected at the same or adjacent site at each distinct stage.

All the experimental procedures involved in this study were approved by the Animal Management and Ethics Committee of the Lanzhou Institute of Animal Science and Veterinary Medicine, Chinese Academy of Agricultural Sciences (Permit No. SYXK-2016-0039).

RNA isolation, library preparation, and sequencing

Total RNA of skin tissues was extracted using Trizol reagent (Invitrogen, CA, USA) following the manufacturer's instructions, after grinding them in liquid nitrogen. RNA integrity was analyzed using RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). RNA concentration and purity was measured with the Nanodrop 2000 photometer spectrophotometer (Implen, Los Angeles, USA). Approximately 3 μg RNA was used for RNA library preparation. First, ribosomal RNA was removed using TruSeq Stranded Total RNA with Ribo-Zero Gold Kit (Illumina, CA, USA). Subsequently, interruption reagent was added to break the rRNA-depleted RNA into short ragments, the later RNA is used as a template, random hexamer primers were used to synthesize the first-strand cDNA, and then second-strand cDNA was synthesized using DNA Polymerase I and RNase H. In the reaction system of second-strand cDNA, dUTP was used instead of dTTP. Then different adapters were connected and the UNG enzyme method was used to digest the strand that containing dUTP, only retaining the first cDNA strands with different linkers. After purification of the cDNA strands, remaining overhangs were

converted into blunt ends via polymerase/exonuclease activities and adenylation of the 3' ends were performed. The sequencing adapter was ligated and the cDNA fragments of 150-200 bp in length were selected to perform PCR amplification, finally, the PCR products were purified (AMPure XP system) and RNA library quality was assessed with an Agilent 2100 Bioanalyzer system (Kapa BioSystems, MA, USA). Clustering of the index-coded samples were performed using TruSeq PE Cluster reagent (Illumina, CA, USA) on a cBot Cluster generation system according to the manufacturer's instructions. Then the cDNA libraries were subjected to standard paired-end sequencing with an Illumina Hiseq 2000 platform and 100-bp paired-end reads were generated.

Quality control and transcriptome assembly

Raw reads were trimmed by removing adapter sequences, reads containing over 10 % of ploy -N, and low-quality bases (>50 % of bases whose Phred scores were <5 %) using in-house Perl scripts. The valid bases, Phred score (Q30) and GC content were used to filter high quality clean data. Paired-end clean reads alignment to the reference genome (versionGCA_005887515.2 BosGru3.0) was performed using the hisat2 [32]. The mapped reads of each library were assembled using Cufflinks (v2.1.1) in a reference-based approach [33], and for the prediction of lncRNAs, reassemble transcripts was performed using a streaming neural network algorithm by StringTie software [34], which could accurately determine the direction of transcript chains for the data of chain-specific library.

Identification of lncRNAs

Based on the characteristics of lncRNAs, a strict four-step was used to identify lncRNAs from the assembled transcripts:(1) spliced transcripts were compared with reference transcripts using Cuff-compare v2.1.1[33] to remove known coding transcripts or known loci and the location type of the remaining transcripts were determined; (2) transcripts with length <200 bp or exon number <2 were removed; (3) the coding potential of transcripts was analyzed to remove transcripts with coding potential. The analytic software is CNCI (v2) [35], CPC (0.9-r2) [36], Pfam-scan (v1.3) [37], and PLEK [38]. (4) the transcripts obtained in step three were used to alignment the annotated lncRNAs in yak by Blastn software, repeat sequences were moved, then combined with known lncRNAs for the further quantitative analysis. FEELnc software was used to count the types of lncRNAs according to the positional relationship between lncRNA and known protein coding transcripts from four sides: direction, type, location and subtype.

Differential expression analysis of lncRNA

Fragments per kb for a million reads (FPKM) of lncRNA transcripts was calculated with Cuffdiff (v2.1.1) [39]**.** DESeq [40] was used to normalize the lncRNA counts of each skin sample (baseman value to estimate expression), transcripts with a *P*-adjust value <0.05 and fold change >2 were described as differentially expressed between any two distinct stages of hair follicle growth. Based on the expression of lncRNAs, sample similarities and expression distances between samples were analyzed via principal component analysis (PCA analysis) and sample to sample cluster analysis, respectively.

Function enrichment analysis of the nearby genes of differential lncRNAs

The nearest protein-coding gene paired with each lncRNA was searched and as the potential regulatory target gene of that lncRNA. If no gene was detected within 100 kb upstream or downstream of a lncRNA, that lncRNA was excluded in subsequent analysis. GO enrichment analysis of the nearby genes of between any two groups were analyzed with Gene Ontology database (http://geneontology.org/). KEGG pathway enrichment were analyzed with KEGG online website (http://www.genome.jp/kegg/), the top 20 enriched terms with the number of DELs greater than two were shown with R package. Protein-protein interaction network of all the nearby genes paired with DELs was analyzed with STRING database (https://string-db.org/), and further visualized with Cytoscape (http://www.cytoscape.org/).

Verification of sequencing data using Real-time quantitative PCR (RT-qPCR)

Total RNA from the skin tissue subjected to RNA sequencing were also used to verify the sequencing results by real-time quantitative PCR (RT-qPCR). The first- strand cDNA was obtained using a PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRa, Dalian, China). The RT-qPCR was performed using SYBR Premix Ex Taq II (TaKaRa, Dalian, China) on the Bio-Rad CFX96 Touch-Real Time PCR Detection System (Bio-Rad, Hercules, CA, USA). Reaction conditions were set as follows: 95 ℃ for 3min, followed by 40 cycles of 95 ℃ for 12 s and 60 ℃ for 30s. The relative expression of measured lncRNAs and mRNAs were analyzed

with the $2^{-\Delta\Delta Ct}$ method and normalized by *GAPDH*. The primers used for RT-qPCR were designed with oligo6 and listed in additional file 1.

Conservation analysis of lncRNAs between yak and cashmere goat during hair follicle cycle

Sequence conservation analysis of lncRNAs between yak and goat during HFs cycle was performed using NCBI blast-2.9.0+ with default parameters. The lncRNA data of goat during HFs cycle was downloaded from a previous paper [19, 41]. The sequence conservation properties were detailed analysis using dottup website, WebLogo3 (http://weblogo.threeplusone.com/), Clone Manager bioinformatics softwares etc.

Results

Identification and differential expression analysis of lncRNAs in yak skin

A total of 748.61M raw reads were produced from fifteen skin samples in five phase groups of yak using the Illumina HiSeq 2000 Platform in which 733.79 M reads were clean (additional file 2), accounting for 98% of raw reads. After screening with rigorous criteria and analyzing the coding potential with the software CNCI, CPC, Pfam and PLEK, a total of 2884lncRNAs from all samples were identified (Figure 1A). Of those transcripts, 70.1% (2023) were novel lncRNAs, the average length of novel lncRNAs is 1152 bp. Furthermore, our data showed that approximately 40% of the novel lncRNAs were antisense lncRNA and 60% were sense lncRNA, besides, approximately 90% of the novel lncRNAs with 2-3 number of exon (Figure 1B). These data of lncRNAs presented here were similar to previously described lncRNAs sequence feature [42].

The expression levels of lncRNA were analyzed using Cuffdiff v2.1.1, differentially expressed lncRNAs (DELs) of every two distinct groups were screened with the threshold was set as Fold Change ≥ 2 and P-adjust value <0.05, numbers of all the DELs between every comparison groups and the common and specific DELs were vasulized and showed in Figure 1C. According to the result, we selected Oct in anagen with Jan (catagen) and Mar (telogen) for further functional analysis, because of both Mar-vs-Oct and Jan-vs-Oct with more number of DELs, while the number of DELs was relative minor in Mar-vs-Jun and Aug-vs-Oct, Jun and Aug were preliminary excluded so as to simplify subsequent analysis. In the differential expression analysis of the three groups (Jan., Mar. and Oct.), 280 DELs were obtained in Jan-vs-Mar, 340 and 198 DELs were obtained in Mar-vs-Oct and Jan-vs-Oct, respectively, 22 of those DELs were shared in the three comparisson groups (Figure 1D). PCA was used to analysis the sample similarities, a more concentrated distribution of samples in the same group could be clearly observed, in additional, grouped

together between samples was presented by sample to sample cluster analysis (Figure 1E), these data indicated our samples with high reliability.

GO and KEGG enrichment analysis

To investigate the potential function of DELs between the three stage comprision groups, the nearest protein-coding genes paired with DELs were used to perform GO and KEGG enrichment analysis. The top 30 GO terms (top 10 terms in each GO category) were showed in Figure 2A. The result showed that the GO terms mainly enriched in Keratinization, keratin filament, carbohydrate metabolic process and thiamine transport which involved in hair formation or nutrient metabolism. In Mar-vs-Oct, protein ubiquitination was multiple enriched, as a result, it was reported that ubiquitin-mediated proteolysis pathway is important for distinguishing the SFs (secondary hair follicles) and PFs (primary hair follicles) in cashmere goats [43, 44], which suggestting that the regulation of protein ubiquitination was related to the transition between PFs and SFs from Mar to Oct in our present study.

KEGG enrichment analysis was performed and the top 20 KEGG pathways with the number of DELs greater than two were showed in Figure 2B. We found the VEGF signaling pathway, Wnt signaling pathway and Signaling pathways regulating pluripotency of stem cells which pathways previously have been reported to play indispensable roles during HFs cycle were enriched [45-47]. In addition, Hormone and Nutrition metabolism signaling pathway (eg. Estrogen signaling pathway, Vitamin digestion and absorption and Glycine, serine and threonine metabolism), Immune response signaling pathway (eg. Antigen processing and presentation, Natural killer cell mediated cytotoxicity) were enriched, which maybe also essential for the hair follicles growth [6, 48]. We observed the pathways emerged among the distinct stages coincidentally corresponded with the biological process of hair cycling seems to orchestrate a dynamic genesis of HFs cycle in molecular level, for example, in Jan-vs-Mar the only up-regulated pathway is Vitamin digestion and absorption may due to the relative resting sate in telogen (Mar.) and reserve energy for preparing the next HFs cycle (Figure S1). What's more, HIF-1 signaling pathway involved in perception of hypoxia was enriched, which may relate to the high altitudes environment with frigid and hypoxia of yak living. Protein-protein interaction network of coding genes paired with all the DELs of this three distinct stages were analyzed using STRING Database (https://string-db.org), the result showed that *ITCH* (E3 ubiquitin-protein ligase Itchy homolog), *PSMD1*, *FBXL19* and *HSPA9* primarily

involved in protein ubiquitination (queried from UniProt database) were key nodes of the DELs paired genes (Figure 2C), ubiquitin-mediated proteolysis pathway was important for distinguishing SFs and PFs in cashmere goats [43, 44], and it was recently reported that ubiquitylation process associated with Hedgehog signaling a crucial signaling that affecting on hair follicle regeneration [49-51].

Comparative analysis of DELs in yak with Cashmere goat lncRNA data

To further screen the DELs that are more likely to affect HFs cycle of yak, 93 differently expressed protein coding genes (DEGs) paired with 110 DELs were obtained by combining analysis with our previously mRNA sequencing data in yak HFs cycle, the detailed information of these DELs and DEGs were presented in additional file 3. We found that the expression tendency and abundance of most DELs were consistent with its paired coding gene, although a few pairs such as TCONS_00055063 with its paired gene NOS3 has an opposite expression trendency. Six lncRNAs and four LncRNA-paired protein coding genes which differently expressed at least any two distinct stages were selected to verify the transcriptome sequencing data by RT-qPCR. LncRNAs were selected based on the expression aboundance and potential function in hair development of the paired protein coding gene. The RT-qPCR result was showed in Figure 3A.

LncRNA has been reported to be tissue specific and with lower sequence conservation, in order to explore sequence conservation level between yak and cashmere goat during HFs cycle and obtain more reliable DELs that regulating hair follicle growth, NCBI blast-2.9.0+ was used to research the conserved lncRNAs of these 110 DELs with the lncRNA data of cashmere goat in HFs cycle [19]. Blast result showed that twenty-four DELs among these 110 DELs were queried with sequence conservation with cashmere goat lncRNA data in differ degree. The twenty-four DELs and its paired twenty-three DEGs were showed in Figure 3B with heat map. In addition, we comparative analysis of the 93 DEGs of yak with the differentially expressed genes between anagen and telogen of cashmere goat [19] found that 10 genes among these DEGs differently expressed in both yak and goat during HFs cycle and the expression tendency in cyclic process was consistent (Figure 3C), the genes *GPRC5D*, *FER*, *EMX2* and *ERAP2* which paired lncRNAs were simultaneously conserved in sequence between yak and cashmere goat, *GPRC5D*, *EMX2* was reported to be associated with hair follicle morphogenesis [52, 53]. These results verified our

sequencing data and further screened some potentially important DELs and their paired genes affecting on HFs cycle by comparative analysis of lncRNA-seq data between yak and cashmere goat.

Sequence conservation analysis of lncRNAs between yak and cashmere goat

In the above section, we investigate the sequence conservation of the 110 DELs paired with DEGs of yak with cashmere goat lncRNA data during HFs cycle by NCBI blast-2.9.0+. We then detailed analysis the sequence conservation properties of these aligned lncRNAs. In the goat lncRNAs data(532 sequences) [19], more than half lncRNA sequences(303) were aligned with those twenty-four DELs of yak, length of the alignments distributed between 28-3765 bp and all the matched percent is over 80% (additional file 4). The length distribution of the matched sequence in cashmere goat lncRNA data was counted and showed in Figure 4A, indicating that largely of the matched sequence distributed between 101-200 bp, accounting for nearly 60% of all the alignments [19] (Figure 4A, additional file 4). Furthermore, the matching regions of every aligned lncRNA in goat data were analyzed, we found that most of the queried lncRNAs in yak with numerous alignments in goat lncRNA data while the matching regions on one lncRNA of yak were constrained (additional file 4-5), partial queries that aligned more than ten lncRNA sequences in goat data were presented in table 1, for example, the matching region of TCONS_00027937 with high sequence conservation which aligned 36 lncRNAs sequences of goat while all of these aligned sequences roughly matched to the region of 43-160 of TCONS_00027937. Then we randomly selected TCONS_00008989 and TCONS_00020227 the first two lncRNAs listed, and TCONS_00016111 which with longest matching sequence to 3765 bp to perform further detailed analysis for their matching regions.

For TCONS_00008989, all the alignments of goat data were mapped to the roughly region of 243-442 of TCONS_00008989, we then manually performed a local multiple sequence alignment at the matching region of TCONS_00008989 with other aligned eleven different sequences in goat, the result showed that the matched region between 315-356 about 40 bp of TCONS_00008989 was highly conserved among ten of these aligned sequences (Figure S2), the sequence logo of this highly conserved region was presented using WebLogo3 (http://weblogo.threeplusone.com/) (Figure 4B). We speculated this highly conserved sequence may be an essential binding module for dictating the function of TCONS_00008989.

TCONS_00020227 is a intronic lncRNA, which was aligned to 31 lncRNA sequences of goat data mainly matched to four roughly regions they were 2506-3121, 4251-4452, 4723-4898 and 4886-5053 within

297 TCONS_00020227 (additional file 4, table 1). Comparing to TCONS_00008989, the matching regions in
298 TCONS_00020227 were relative various, then we detected whether there are repeats on TCONS_00020227
299 by analysis the Dot Plot of TCONS_00020227 with dottup website, two pairs of repeats on
300 TCONS_00020227 were appeared (Figure 4C) and the detailed regions of the repeats were presented by
301 self pair-wise alignment using Clone Manager software (Figure 4C) which result showed that the latter
302 repeats was coincident with the matching regions of 4723-4898 and 4886-5053 by NCBI blast-2.9.0+
303 (Figure 4C and table 1), the slightly difference in sequence value may be leaded by distinct analysis method,
304 but the roughly range was comparable. Another repeats appeared in Dot Plot of TCONS_00020227could
305 also be observed via Clone Manager that is 4252-4449 (equivalent to 4251-4452 by NCBI blast-2.9.0+)
306 which repeated with the range of 3387-3563 although with a lower percent match, it could also implicate
307 the conservative of another mainly matched region 4251-4452 in a way. This result showed that the
308 matching regions repeatedly existed in TCONS_00020227 and further indicated the importance of this
309 matched sequences and which may be a common property of lncRNAs in conservation module.

310 Finally, we analyzed the sequence conservation of TCONS_00016111 which with longest matching
311 sequence to 3765 bp. Due to TCONS_00016111 with a longer matching region, we explore the sequence
312 conservation of TCONS_00016111 by Nucleotide BLAST in Nucleotide collection databases with default
313 parameters. Blast result showed that lncRNA TCONS_00016111 were mapped to multiple specie genomes
314 including Ovis canadensis canadensis, whales, Bos Taurus, Capra hircus and so on, most of these
315 alignments were ncRNAs (Figure S3). Interestingly, the distance tree of the selected 100 sequences in
316 defaulted first page of BLAST result showed that TCONS_00016111 with Ovis canadensis canadensis and
317 whales were closer at distance tree (Figure 4D), suggesting there may be related between Ovis canadensis
318 Canadensis, whales and yak in some traits in evolution. What's more, we found that two mainly matching
319 regions 2881-3029 and 3191-3365 of TCONS_0001611 searched by NCBI blast-2.9.0+ between yak and
320 goat lncRNA data were emerged in nearly all of the mapped sequences in the first page by Nucleotide
321 BLAST (Figure 4E), more importantly, these two regions largely repeated in the genomes of Ovis
322 canadensis Canadensis (bighorn sheep) and Bos mutus (yak) and in different chromosomes, the number of
323 repeats from hundreds to more than one thousand, an example of the searched results was showed in Figure
324 4F. Why this two regions simultaneously repeated in genome in abundance, then we found the two regions

were partly reverse-complementary so that the range of 2881-3365 could form a stable secondary structure (Figure S4), which may be one of the conserved secondary structure of TCONS_0001611, consistenting with the reported conservation properties of lncRNAs that most lncRNA are conserved in secondary structure [13]. In addition, miR-34a in bovine, Pan troglodytes, Gorilla gorilla, and the pre-miR-34a and pri-miR-34a in human were found within the alignments of TCONS_00016111(Figure 4E), implicating that TCONS_00016111 also act as a miRNA precursor.

Discussion

The development of hair follicle cycle is the characteristic of hair growth in mammals, including anagen, catagen and telogen [54]. Transition between the stages of hair cycle were controlled by the unique regenerating system of follicular epithelial and mesenchymal cells with the interaction of their multiple molecular signals [55]. LncRNAs as key regulatory molecules have been found to play an important role during HF cycle, and which were especially studied a lot in fur-producing animals due to the economic benefit of animal fiber and observable changes in hair growth phenotype [17, 56, 57]. The hair growth pattern of yak is similar to cashmere goat presenting a seasonal phenomenon under the control of photoperiod. In this study, we firstly investigated the expression profile of lncRNAs during HFs cycle and analyzed the potential function of differently expressed lncRNAs of yak. In HFs cycle, anagen is the longest stage that more than half of a whole period (one year) [47] and there no available division of HFs cycle in yak was reported to data, in order to confirm the optimum phase during anagen that distinguish from other stages, the skin samples of yak in five phases that including three phases of anagen (Jun., Aug. and Oct.), catagen (Jan.) and telogen (Mar.) during HFs cycle were collected by referencing the period of cashmere goat and the phenotypic changes of yak [47]. Thus, samples in total five phases were employed to perform lncRNA sequencing, after preliminary analysis the differently expressed lncRNA data between each comprision groups, October was selected as the optimum phase in anagen with Jan.(catagen) and Mar.(telogen) for further function analysis. Previous studies reported that the stage division of HFs cycle in shaibei cashmere goat in north china as follows: anagen is from Apr. to Oct., catagen is from Nov. to Jan., and telogen is from Feb. to Mar.. The climatic environment of yak in our present study is closer with shaibei cashmere goat in this previously reported, DELs between Jan., Mar. and Oct. in our sequencing date were relative more than other comparison groups and Oct. with more changes comparing Jun. and Aug. in

anagen, which is consistent with previous study reported that the radio of secondary to primary follicles (S/P) in Oct. is the highest in cashmere goat [47]. After confirm the optimum three stages of HFs cycle in yak, we performed functional analysis of the DELs of these three distinct groups.

LncRNAs could affect the gene expression on the immediate genomic vicinity (in cis) [12, 14]. In currency study, the nearest gene within 100kb of lncRNA was searched and used to predict the potential function of DELs by GO and KEGG enrichment analysis, the result showed that in the transition from Aatagen to Catagen (Oct-vs-Jan), the primarily enriched GO terms of cell communication and metabolism (eg. Cell surface receptor signaling, carbohydrate metabolic process and protein O-linked mannosylation) were down-regulated, the transform of cell itself was increased (eg. Keratinization and mitotic cell cycle); in the transition from Catagen to telogen (Jan-vs-Mar), the function of nutrition metabolism (eg. Thiamine transport, response to glucose and regulation of insulin secretion) was mainly enriched, which may relate to the preparing for follicle reset; Insterestingly, the function of protein ubiquitination was enriched in the transition from telogen to Aatagen (Mar-vs-Oct), previous studies reported that ubiquitin-mediated proteolysis pathway is important for distinguishing the SFs (secondary hair follicles) and PFs (primary hair follicles) of cashmere goats [43, 44], which is coincident to our data that Oct. is the phase with thriving SFs and the growth of PFs maybe vigorous in Mar. [47]. Signaling pathways involved in hair follicle morphogenesis were important for maintain the normal periodical cycle of hair growth [5], KEGG enrichment analysis found that the Wnt signaling was enriched in Jan-vs-Mar, which was widely reported playing a curial role in HFs development [45, 58]. VEGF signaling pathway was enriched both in Jan-va-Mar and Mar-vs-Oct, *VEGF* (vascular endothelial growth factor) was reported that regulation of the HFs cycle by inducing the angiogenesis of dermal papilla [9]. HIF-1 signaling pathway were enriched in Mar-vs-Oct, which may associate with the hypoxia adaptability of yak [59]. And from a holistic perspective, the pathways emerged among the distinct stages seem to orchestrate a dynamic genesis of HFs cycle in a molecular level that correspond with the biological process of hair follicle development.

There were a lot of differently expressed lncRNAs during HFs cycle in various species were examed following the rapid development of genome sequencing technology, but the identified functional lncRNAs are rarely, partly due to the lower level of sequence conservation of lncRNA across species [12]. In order to obtain more reliable DELs affecting on hair follicle growth, the DELs with paired differently expressed

nearby gene were screened and further analyzed the sequence conversation of these lncRNAs between yak and goat during HFs cycle. Twenty-four DELs were found with a sequence conservation between cashmere goat and yak lncRNAs data in differ degree. Sequence conversation was an important factor to reveal the function of lncRNAs, the well-studied lncRNAs such as Xist, HOTAIR and MALAT1 were reported with higher sequence conversation or conserved domain [25-27], and it was identified that a highly conserved sequence element of lncRNA is essential for its function [60] although lack of sequence conservation does not directly imply a lack of function [61]. Previous studies have reported that lncRNAs were specific expressed in tissue and species and with poor conservation at sequence level [11, 62], in our present study, Yak and Goat are close in evolutionary distance which all belong to the Bovidae family of ruminants, and the seasonal cyclic development of hair follicles of yak is similar to that of goats, including anagen, catagen and telogen, which just provide us a well molding to study the sequence conservation of lncRNAs at the same trait in different species. NCBI blast-2.9.0+ result showed that twenty-four lncRNAs among 110 DELs of yak were aligned with cashmere goat lncRNA data, accounting for 20% of total queries (additional file 4) and the matched sequence of ~200 bp length with > 80% sequence similarity is the largest among the length distribution of matching regions, which is consistent with the sequence conservation feature of lncRNAs between human and zebrafish [63]. Many well-studied lncRNAs exhibit well-conserved RNA secondary structures [13], we then analyzed the sequence conservation properties of these aligned lncRNAs, we found that largely of the queried conservation lncRNAs were aligned to multiple lncRNA sequences in goat data, while the matching regions in queried lncRNAs of yak are constrained, these sequences probably contain elements essential for their activity [64-66]. In currently study, we selected three lncRNAs for detailed analysis, among this three lncRNAs, TCONS_00008989 aligned to fourteen lncRNA sequences of goat data and all of these aligned sequence nearly matched to one region that about 200 bp of TCONS_00008989 and approximately 40 bp was highly conversed which may be the key recognition sequence of TCONS_00008989. Another two selected lncRNAs TCONS_00020227 and TCONS_00016111 were all aligned to nearly thirty lncRNA sequences in goat data and their number of matched regions are about four and five regions respectively. Repeat matching regions of TCONS_00020227 was observed which probaly a commonality of many conserved lncRNAs. Because of TCONS_00016111was aligned to a longer sequence with a cashmere goat lncRNA, Online Blast was performed and found that the molecular evolution tree of TCONS_00016111 was differ with the species

tree of yak (http://asia.ensembl.org/info/about/speciestree.html), the distance of this molecular evolution tree of yak was closer with Ovis Canadensis Canadensis and whales while bovine in species tree. And notably, there are two matched regions that could form a stable seconary structure in TCONS_00016111were largely repeated from the number of hundreds to over one thousand in the genome of Ovis Canadensis Canadensis and yak (Figure S4; Figure 4F), which implicating TCONS_00016111 may link with some trait among these species in evolution and the repeatly emerged sequence may act as an important module to recognize RNA, DNA or protein for the activity of their genomic loci [13, 60], meanwhile, these data also indicated the efficiency of the aligned conservation elements by NCBI blast-2.9.0+.

Conclusion

Our current study presented the first data on the lncRNA of yak during HFs cycle by lncRNA-seq, the function of the identified differently expressed lncRNAs were indirectly predicted via enrichment analysis the function of their nearest mRNA. In addition, several limited degree of sequence conserved lncRNAs between yak and cashmere goat in HFs cycle were searched by NCBI blast-2.9.0+ and the sequence conservation properties were analyzed, the conserved sequence elements may play an important role in regulation of hair cycling, our finding will provide a valuable insight into the study of lncRNAs functionality and mechanism in the process of HFs cycle.

Abbreviations

LncRNAs: Long non-coding RNAs; HFs: Hair follicles; DELs: Differentially expressed lncRNAs; DEGs: Differentially expressed genes; KEGG: Kyoto encyclopedia of genes and genomes.

Acknowledgements

Not applicable.

Funding

This work was supported by the Agricultural Science and Technology Innovation Program [CAAS-ASTIP-2014-LIHPS-01]; the National Beef Cattle Industry Technology & System [CARS-37]; and the Herbivorous Livestock Industry and Technical System of Gansu Province [GARS08].

Availability of data and materials

The data supporting the findings of this study are available within its supplementary information.

Authors' contributions

XZ analyzed the data and wrote the article. QB and CJ contributed significantly to bioinformatics analysis, visualized the data. CL1 and YC extracted the RNA and performed the RT-qPCR. XW and CL2 collected the skin samples. PB and PY conceived the study, designed the experiment, interpreted the results and revised the article. All authors read and approved the manuscript.

Ethics approval

All procedures involving animals were performed according to the guidelines of the China Council on Animal Care and the Ministry of Agriculture of the People's Republic of China and approved by the Animal Ethics Committee of Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Figure legends

Figure 1. long noncoding RNAs (lncRNAs) identification and comparative analysis of differential expression in hair growth cycling of yak. A. Screening of the candidate lncRNAs in skin transcriptome. Four tools (CPC, PLEK, CNCI, and PFAM) were employed to analyze the coding potential of lncRNAs. Those lncRNAs shared by the four analytical tools were designated as candidate lncRNAs and used for subsequent analysis. B. The length distribution, classification and exon number distribution of novel lncRNA. C. Statistic of differently expressed lncRNA in every two comparison groups (left), and common and specific differently expressed lncRNAs between different comparison groups (right). D. Venn diagrams of differently expressed lncRNA in three comparative groups including Jan-vs-Mar, Jan-vs-Oct

and Mar-vs-Oct, the lncRNAs shared with the three groups indicated differently expressed in every comparison groups. E. PCA and sample to sample cluster analysis between samples in Jan., Mar. and Oct..

Figure 2. GO and KEGG enrichment analysis of the nearest genes of differently expressed lncRNAs. A. GO analysis of the nearest genes of differentially expressed lncRNAs, the top 30 GO terms (top 10 terms in each GO category) were shown. B. KEGG enrichment analysis of the nearest genes of differently expressed lncRNAs, the top 20 pathways and with the number of genes greater than two were illustrated as bubble plot. The size of the bubble indicates gene number and the color indicates pvalue. C. Protein to protein interactive network of all the nearest genes of differently expressed lncRNAs visualized by Cytoscape. The color indicates node degree.

Figure 3. Verification of sequencing data by RT-qPCR and screening of sequence conserved DELs between yak and Cashmere goat. A. The expression level of differently expressed lncRNAs and several paired differently expressed mRNA validated by RT-qPCR, the relative expression of measured lncRNAs and mRNAs were analyzed with the $2^{-\Delta\Delta Ct}$ method and normalized by *GAPDH*. Data were presented as means ± SEM (n = 3), the same letter means no significant difference, different letters mean significant difference. B. Cluster analysis of sequence conserved DELs between yak and cashmere goat and their paired differently expressed mRNAs. C. Heat map of differently expressed mRNAs with paired DELs during HFs cycle shared in yak and cashmere goat.

Figure 4. Conservation properties analysis of sequence conserved lncRNAs between yak and cashmere goat. A. Length distribution of the matching sequence of aligned cashmere goat lncRNA in matched sequences of yak using NCBI blast-2.9.0+. B. Sequence logo of the highly conserved sequences between TCONS_00008989 and its aligned ten cashmere goat lncRNA by WebLogo3. C. Dot Plot of TCONS_00020227and the detailed ranges of the repeats on TCONS_00020227 were analyzed by dottup website and Clone Manager software, respectively. The boxes with the same color indicate the same region of TCONS_00020227 and the red boxes and green boxes represented two pairs of repeats on TCONS_00020227. D. The Blast Tree View of TCONS_00016111 produced using BLAST pairwise alignments with Neighbor Joining algorithm. Query_55251 presented TCONS_00016111and other terms on the tree were the first seven sequences of TCONS_00016111 BLAST result. E. Graphic Summary of

489 selected 100 sequences in defaulted first page of TCONS_00016111 BLAST result. Three mainly matching
490 regions of TCONS_00016111were marked in green box. F. Typical alignments example of the two key
491 matching regions about 2881-3029 and 3191-3365 in the chromosome of Ovis canadensis Canadensis and
492 Bos mutus (yak), the number of matches in each of these two chromosomes was highlighted using red box.
493

References

1. Waters WKBNM. Fleece growth in Australian cashmere goats. III. The seasonal patterns of cashmere and hair growth, and association with growth hormone, prolactin and thyroxine in blood. Crop Pasture Sci. 1993;44(5):1035-1050.
2. Mcdonald B, Hoey W, Hopkins P. Cyclical fleece growth in cashmere goats. Aust J Agr Res. 1987;38(3):597.
3. Baker RE, Murray PJ. Understanding hair follicle cycling: a systems approach. Curr Opin Genet Dev. 2012;22(6):607-612.
4. Alonso L, Fuchs E. The hair cycle. J Cell Sci. 2006;119(Pt 3):391-393.
5. Rishikaysh P, Dev K, Diaz D, Qureshi W, Filip S, Mokry J. Signaling Involved in Hair Follicle Morphogenesis and Development. Int J Mol Sci. 2014;15(1):1647-1670.
6. Oh HS, Smart RC. An estrogen receptor pathway regulates the telogen-anagen hair follicle transition and influences epidermal cell proliferation. P Natl Acad Sci. 1996;93(22):12525-12530.
7. Ibraheem M, Galbraith H, Scaife J, Ewen S. Growth of secondary hair follicles of the Cashmere goat in vitro and their response to prolactin and melatonin. J Anat. 1994;185(Pt 1):135-142.
8. Foitzik K, Lindner G, Mueller-Roever S, Maurer M, Botchkareva N, Botchkarev V, Handjiski B. Control of murine hair follicle regression(catagen) by TGF-beta1 in vivo. FASEB J. 2000;14(5):752-760.
9. Castexrizzi N, Lachgar S, Charveron M, Gall Y. Implication of VEGF, steroid hormones and neuropeptides in hair follicle cell responses. Ann Dermatol Vener. 2002;129(5 Pt 2):783-786.
10. Guttman M, Rinn JL. Modular regulatory principles of large non- coding RNAs. Nature. 2012;482(7385):p.339-346.
11. Sarropoulos I, Marin R, Cardoso-Moreira M, Kaessmann H. Developmental dynamics of lncRNAs across mammalian organs and species. Nature. 2019;571(7766):510-514.
12. Amaral PP, Leonardi T, Han N, Viré E, Gascoigne DK, Arias-Carrasco R, Büscher M, Pandolfini L, Zhang A, Pluchino S. Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. Genome Biol. 2018;19(1):32.
13. Nitsche A, Stadler PF. Evolutionary clues in lncRNAs. Wiley Interdiscip Rev RNA. 2017;8(1):e1376.
14. Roberts TC, Morris KV, Weinberg MS. Perspectives on the mechanism of transcriptional regulation by long non-coding RNAs. Epigenetics. 2014;9(1):13-20.
15. Mitchell G, Julie D, Bryce W C, Manuel G, Jennifer K G, Glen M, Geneva Y, Anne Bergstrom L, Robert A, Laurakay B et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011;477(7364):295-300.
16. Lin CM, Liu Y, Huang K, Chen X-c, Cai B-z, Li H-h, Yuan Y-p, Zhang H, Li Y. Long noncoding RNA expression in dermal papilla cells contributes to hairy gene regulation. Biochem Bioph Res Co. 2014;453(3):508-514.
17. Zhao B, Chen Y, Hu S, Yang N, Wang M, Liu M, Li J, Xiao Y, Wu X. Systematic Analysis of Non-coding RNAs Involved in the Angora Rabbit (Oryctolagus cuniculus) Hair Follicle Cycle by RNA Sequencing. Front Gene. 2019;10(407):eCollection 2019.
18. Sulayman A, Tian K, Huang X, Tian Y, Xu X, Fu X, Zhao B, Wu W, Wang D, Tulafu AYaH. Genome-wide identification and characterization of long non-coding RNAs expressed during sheep fetal and postnatal hair follicle development. Sci Rep-UK. 2019;9(1):8501.
19. Wang S, Wei G, Zhixin L, Yang G, Beilei J, Lei Q, Zhiying Z, Xin W. Integrated analysis of coding genes and non-coding RNAs during hair follicle cycle of cashmere goat (Capra hircus). BMC Genomics. 2017;18(1):767.
20. Weikard R, Hadlich F, Kuehn C. Identification of novel transcripts and noncoding RNAs in bovine skin by deep next generation sequencing. BMC Genomics. 2013;14:798.
21. Song L-L, Cui Y, Yu S-J, Liu P-G, Liu J, Yang X, He J-F, Zhang Q. Expression Characteristics of BMP2, BMPR-IA and Noggin in Different Stages of Hair Follicle in Yak Skin. Gen Comp Endocr. 2018;260:18-24.

22. Song L-L, Cui Y, Xiao L, Yu S, He J. DHT and E2 Synthesis-Related Proteins and Receptors Expression in Male Yak Skin During Different Hair Follicle Stages. Gen Comp Endocr. 2020;286:113245.
23. Song L-L, Cui Y, Yu S-J, Liu P-G, He J-F. TGF-beta and HSP70 profiles during transformation of yak hair follicles from the anagen to catagen stage. J Cell Physiol. 2019; 234(9):15638-15646.
24. Jia C, Wang H, Li C, Wu X, Zan L, Ding X, Guo X, Bao P, Pei J, Chu M et al. Genome-wide detection of copy number variations in polled yak using the Illumina BovineHD BeadChip. BMC Genomics. 2019;20(1):376.
25. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai M-C, Hung T, Argani P, Rinn JL et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010;464(7291):1071-1076.
26. Gutschner T, Hämmerle M, Diederichs S. MALAT1 —a paradigm for long noncoding RNA function in cancer. J Mol Med. 2013;91(7):791-801.
27. Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF. The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell. 1992;71(3):527-542.
28. Jeon Y, Lee JT: YY1 Tethers Xist RNA to the Inactive X Nucleation Center. Cell. 2011;146(1):119-133.
29. Nitsche A, Rose D, Fasold M, Reiche K, Stadler PF. Comparison of splice sites reveals that long noncoding RNAs are evolutionarily well conserved. RNA. 2015;21(5):801-812.
30. Johnsson P, Lipovich L, Grandér D, Morris KV. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochim Biophys Acta. 2014;1840(3):1063-1071.
31. Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. Cell Rep. 2015;11(7):1110-1122.
32. Kim D, Langmead B, Salzberg SL. HISAT. a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357-360.
33. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562-578.
34. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290-295.
35. Liang S, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Yi Z. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res. 2013;41(17):e166.
36. Lei K, Yong Z, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Ge G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007;35(Web Server issue):W345–W349.
37. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J. Pfam: the protein families database. Nucleic Acids Res. 2014;42(D1):D222-D230.
38. Li A, Zhang J, Zhou Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. BMC Bioinformatics. 2014;15(1):311.
39. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511-515.
40. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
41. Grove EA, Tole S, Limon J, Yip L, Ragsdale CW. The hem of the embryonic cerebral cortex is defined by the expression of multiple Wnt genes and is compromised in Gli3-deficient mice. Development. 1998;125(12):2315-2325.

42. Zhang S, Qin C, Cao G, Xin W, Feng C, Zhang W. Systematic Analysis of Long Noncoding RNAs in the Senescence-accelerated Mouse Prone 8 Brain Using RNA Sequencing. Mol Ther Nucleic Acids. 2016;5(8):e343.
43. Ji XY, Jian-xun W, Bin L, Zhu-qing Z, Shao-yin F, Mekuriaw TG, Xue B, Yong-sheng B, Heng L, Wen-guang Z. Comparative Transcriptome Analysis Reveals that a Ubiquitin-Mediated Proteolysis Pathway Is Important for Primary and Secondary Hair Follicle Development in Cashmere Goats. Plos One. 2016;11(10):e0156124.
44. Huntzicker EG, Oro AE. Controlling Hair Follicle Signaling Pathways through Polyubiquitination. J Invest Dermatol. 2008;128(5):1081-1087.
45. Tsai S-Y, Sennett R, Rezza A, Clavel C, Grisanti L, Zemla R, Najam S, Rendl M. Wnt/β-catenin signaling in dermal condensates is required for hair follicle formation. Dev Biol. 2014; 385(2):179-188.
46. Yamamoto N, Tanigaki K, Han H, Hiai H, Honjo T. Notch/RBP-J Signaling Regulates Epidermis/Hair Fate Determination of Hair Follicular Stem Cells. Curr Biol. 2003;13(4):333-338.
47. Chao L, Yan L, Guangxian Z, Ye G, Sen M, Yulin C, Jiuzhou S, Xiaolong W. Whole-genome bisulfite sequencing of goat skins identifies signatures associated with hair cycling. BMC Genomics. 2018;19(1):638.
48. Nyberg KG, Machado CA, Notes A. Comparative Expression Dynamics of Intergenic Long Noncoding RNAs in the Genus Drosophila. Genome Biol Evol. 2016;8(6):1839-1858.
49. Infante P, Severini LL, Bernardi F, Bufalieri F, Marcotullio LD. Targeting Hedgehog Signalling through the Ubiquitylation Process: The Multiple Roles of the HECT-E3 Ligase Itch. Cells. 2019;8(2):98.
50. Suen W-J, Li S-T, Yang L-T. Hes1 regulates anagen initiation and hair follicle regeneration through modulation of hedgehog signaling. Stem Cells. 2020;38(2):301-314.
51. Massa F, Tammaro R, Prado MA, Cesana M, Lee BH. The deubiquitinating enzyme USP14 controls ciliogenesis and hedgehog signalling. Hum Mol Genet. 2018;28(5):764–777.
52. Gao Y, Wang X, Yan H, Zeng J, Ma S, Niu Y, Zhou G, Jiang Y, Chen Y. Comparative Transcriptome Analysis of Fetal Skin Reveals Key Genes Related to Hair Follicle Morphogenesis in Cashmere Goats. Plos One. 2016;11(3):e0151118.
53. Jacobo A, Dasgupta A, Erzberger A, Siletti K, Hudspeth AJ. Notch-Mediated Determination of Hair-Bundle Polarity in Mechanosensory Hair Cells of the Zebrafish Lateral Line. Curr Biol. 2019;29(21):3579-3587.
54. Stenn KS, Paus R. Controls of Hair Follicle Cycling. Physiol Rev. 2001;81(1):449-494.
55. Sennett R, Rendl M. Mesenchymal–epithelial interactions during hair follicle morphogenesis and cycling. Semin Cell Dev Biol. 2012;23(8):917-927.
56. Jiao Q, Yin RH, Zhao SJ, Wang ZY, Zhu YB, Wang W, Zheng YY, Yin XB, Guo D, Wang SQ et al. Identification and molecular analysis of a lncRNA-HOTAIR transcript from secondary hair follicle of cashmere goat reveal integrated regulatory network with the expression regulated potentially by its promoter methylation. Gene. 2019;688:182-192.
57. Wei G, Shan-He W, Bing S, Yue-Lang Z, Wei S, Hasan K. Melatonin promotes Cashmere goat (Capra hircus) secondary hair follicle growth: a view from integrated analysis of long non-coding and coding RNAs. Cell Cycle. 2018;17(10):1255-1267.
58. Mater DV, Kolligs FT, Dlugosz AA, Fearon ER. Transient Activation of Beta -Catenin Signaling in Cutaneous Keratinocytes Is Sufficient to Trigger the Active Growth Phase of the Hair Cycle in Mice. Genes Dev. 2003;17(10):1219-1224.
59. Ivan M, Kondo K, Yang H, Kim W, Valiando J, Ohh M, Salic A, Asara JM, Lane WS, Kaelin WG. HIFalpha Targeted for VHL-Mediated Destruction by Proline Hydroxylation: Implications for O2 Sensing. Science. 2001;292(5516):464-468.
60. Marín-Béjar O, Mas AM, González J, Martinez D, Athie A, Morales X, Galduroz M, Raimondi I, Grossi E, Guo S. The human lncRNA LINC-PINT inhibits tumor cell invasion through a highly conserved sequence element. Genome Biol. 2017;18(1):202.
61. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010;465(7301):1033-1038.

62. Zheng H, Li J, Liu D, Li H, Samudrala R, Yu J, Wong GK-S, Wang J, Zhang J. Mouse transcriptome: Neutral evolution of 'non-coding' complementary DNAs. Nature. 2004;431(7010):1-757.
63. Lin N, Chang KY, Li Z, Gates K, Rana TM. An Evolutionarily Conserved Long Noncoding RNA TUNA Controls Pluripotency and Neural Lineage Commitment. Mol Cell. 2014;53(6):1005-1019.
64. Ulitsky I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. Nat Rev Genet. 2016;17(10):601.
65. Chen J, Shishkin AA, Zhu X, Kadri S, Maza I, Guttman M, Hanna JH, Regev A, Garber M. Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. Genome Biol. 2016;17(1):19.
66. Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Gruetzner F, Kaessmann H. The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature. 2014;505(7485):635-640,a611.

Table 1 information of lncRNAs aligned with numerous lncRNA sequences of cashmere goat

Query	Length	Aligned lncRNA numbers	Partner_gene	Regions of matching sequence
TCONS_00008989	1549	14	GPRC5D	243-442
TCONS_00020227	5053	31	TMEM231	2506-3121,4251-4452,4723-4898,4886-5053
TCONS_00058749	2278	25	ASAH2	172-359, 1673-1863
TCONS_00055139	2913	13	BACH1	920-1463
TCONS_00033093	2364	11	EMX2	1929-2091
TCONS_00018581	229	15	FER	1-229
TCONS_00016111	4400	26	GPR157	690-4400(2),103-279(2),277-689(1),2881-3029,3191-3365
TCONS_00027937	6622	36	ISPD	43-160
TCONS_00000176	839	16	LIPG	1-133,131-267
TCONS_00016755	1962	15	LOC102274179	923-1962
TCONS_00027432	1635	36	OAF	581-1397(3),1058-1400,1-171
TCONS_00031781	223	15	COLEC12	1-223
TCONS_00058073	3188	11	LOC102267365	2867-3059

Note: the number in parentheses refer to the aligned numbers in that region (relative less), remaining aligned numbers are assigned to other matching regions.