

# Detection and Integrated Analysis of LncRNA and mRNA Relevant to Plateau Adaptation of Yak

**Jiabo Wang**

Southwest Minzu University <https://orcid.org/0000-0002-1386-0435>

**Zhixin Chai**

Southwest Minzu University

**Lei Deng**

Southwest Minzu University

**Jikun Wang**

Southwest Minzu University

**Hui Wang**

Southwest Minzu University

**You Tang**

Jilin Agricultural Science and Technology College

**Jincheng Zhong** (✉ [zhongjincheng518@sina.com](mailto:zhongjincheng518@sina.com))

**Qiumei Ji**

Tibet Academy of Agricultural and Animal Husbandry Sciences

---

## Research article

**Keywords:** Yak, lncRNA, mRNA, Plateau Adaptation, co-expression network

**Posted Date:** September 26th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.15325/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at *Reproduction in Domestic Animals* on October 5th, 2020. See the published version at <https://doi.org/10.1111/rda.13767>.

# Abstract

Background: Yaks, which are known as “The ship in the Plateau”, play an important role in the life of people on the Qinghai-Tibet Plateau. There are many potential genetic advantages for Plateau Adaptation(PA) of Yak to improve breeding and adapt the environment with low oxygen and strong ultraviolet ray. In this study, in order to filter the Different Expression (DE) genes and new long non-coding RNAs (lncRNAs), also reveal underlying co-expression and regulation network of these, we sequenced 18 samples (cerebellum and cerebrum) of Bos Taurus, Bos Grunniens × Bos Taurus and Bos Grunniens (3 individuals as replications for each species), and compared the relationship between modules and PA trait basing on the background of crossbreed genetics. Generalized linear model was used to analyze DE genes, and the different tissues as adjust variations in the model can reduce the noise of population structure. Results: On average, 50 million clean reads for each sample were obtained by the control of data quality. 12,072 pseudo lncRNAs were predicted in intersection of three software (CPC2, CNCI and CPAT). Total 4,257 significant DE transcripts were identified by using Ballgown R package ( $P < 0.01$ ), 1021 of which were protein coding genes, 14 were known lncRNAs, 661 were novel lncRNAs. By WGCNA, a co-expression network of DE mRNAs and lncRNAs with 6 modules was created to present functional relationship with PA trait. Our study provided a valuable sub-network composed by 8 hub genes (GRAMD3, MAST4, MICAL1, TLE1, ERAP1, PINK1, MRPL49 and IL34), 1 of them was known lncRNA (LOC106700748), 5 were novel lncRNAs (MSTRG247.1, MSTRG431.1, MSTRG422.1, MSTRG327.1 and MSTRG97.1) in the major module of the whole network. Conclusions: The functions of those hub genes are associated with blood pressure, reactive oxygen and metabolic process. The analysis of co-expressed genes with PA trait provides a reference way of the regulation mechanisms in PA of Yak and gets a reference method of the additive genetics expression detection between crossbreed and parents population as well.

## Background

Long non-coding RNA (lncRNA) is named as a class of non-coding RNA with more than 200 nt which participate in RNA translation or production proteins<sup>1-3</sup>. So much research indicated the expression of lncRNA plays an important role in the expression of the related mRNA in different environments<sup>4</sup>. lncRNA-mRNA association study relies on the co-expression which usually uses general linear model to explain the network. It has been showed that lncRNA-mRNA network is relevant to environment adaptation and biological response<sup>5,6</sup>.

The Yak (*Bos Grunniens*) is an only large-scale and much productive livestock in Qinghai-Tibet Plateau for supplying the living needs of Tibetan<sup>7,8</sup>. The Yaks named “The ship in the Plateau”, have a powerful adaptation to low atmospheric pressure, hypoxia concentration and high ultraviolet radiation in the Plateau<sup>8-10</sup>. It is limited by geography environments and religious belief, the genetic improvement of the productive ability of Yak slowly developed 30 years ago<sup>11,12</sup>. Recent several decades, the Zang Cattle which were cross breed by *Bos Taurus* (Cattle) and *Bos Grunniens* (Yak) have been developed a large

population for improvement of the productive ability and adaptation of the Plateau environment<sup>13-15</sup>. But the underlying gene expression, regulation and the function of the Plateau adaptation are always not clear, especially RNA expression level. Most researches were focus on SNP, SVN and other Genome Variance<sup>7,8,12,16</sup>. Based on the large gap and different between reference genome of Bos Grunniens and Bos Taurus, utilization of Zang Cattle genetics resource is an important way for detection of candidate gene and expression network in Plateau Adaptation<sup>17,18</sup>. In this study, we design a RNA-seq analysis in three breeds: Leiwuqi Yak (Bos Grunniens), Zang Cattle (Bos Grunniens × Bos Taurus) and Sanjiang Cattle (Bos Taurus) that product a gradient for the expression level of Plateau Adaptation (PA) genes. To identify Different Expression (DE) gene and predict new lncRNA, we performed RNA-seq analysis with cerebrum and cerebellum tissue of 9 animals. The network of DE mRNA and lncRNA will be conducted to veal the interactive of expression with regard to their underlying roles in Yak PA.

## Methods

### Sample Collection

Total 18 samples (3 Leiwuqi Yak cerebrum, 3 Leiwuqi Yak cerebellum, 3 Yakow cerebrum, 3 Yakow cerebellum, 3 Sanjiang Cattle cerebrum and 3 Sanjiang Cattle cerebellum) were collected in Leiwuqi (Leiwuqi Yak and Zang Yakow) and Wenchuan (Sanjiang Cattle) county, Sichuan, China. They are roughly identical health, around 4.5 years old, female. This study was approved by the Animal Ethics and Welfare Association of the Southwest Minzu University(No. 16053). As repetition, each group of 3 individuals in same breed were random selected from a herd with more than 50 individuals. The Leiwuqi Yak was set M group, the Sanjiang Cattle were set H group and the Zang Yakow were set Z group. All clean tissue samples were snap frozen in liquid nitrogen. Using TRIzol® Plus RNA Purification Kit (Invitrogen, USA), total RNA was extracted from brain tissue. Ribosomal RNA (rRNA) was removed before sequencing by Epicentre Ribo-zero™ rRNA Removal Kit.

### RNA-Seq and transcript assembly

Basing on Illumina HiSeq™ 4000 platform, we got paired-end sequencing reads of 18 tissue samples. The sequencing reads were filtered by fastp (version 0.19.8) software with default parameters. Clean reads were aligned to the Yak reference genome (GCF\_000298355.1 BosGru v2.0)<sup>19</sup> with HISAT2 (version 2.1.0) to create Sam file<sup>20</sup>. After sorting and converting to bam, we used StringTie (version 1.2.3)<sup>21</sup> to assemble transcripts for each sample. All transcripts of samples were merged together by merge function of StringTie following all sample file name and created output files for Ballgown format<sup>22</sup>. The merged file was annotated with reference genome annotation file (GCF\_000298355.1 BosGru v2.0) by gffcompare (version 0.10.8).

## Identification of lncRNAs

All transcripts in merged transcript annotation file were identified using four filters to predict potential lncRNAs. First, transcripts shorter than or equal 200 nt were removed; Second, transcripts with class code “u”( unknown intergenic), “x”( exonic overlap on the opposite strand), “i”( fully contained within a reference intron), “j”(multi-exon with at least one junction match), “o”(other same strand overlap with reference exons) were kept; Third, transcripts with low expression levels ( $FPKM \leq 1$ ) were filtered out; And fourth, transcripts more than 1 exon were retained<sup>1,23</sup>. The fasta format data was extracted by using getfasta function of bedtools from reference genome<sup>24</sup>. Whole remaining transcripts were considered as new transcripts. Then we used three methods to predict the potential protein coding ability of these new transcripts. The first method is CPC2<sup>25</sup>, which is python function for predicting potential protein coding ability with fasta data of new transcripts. The second method is CNCI<sup>26</sup>, which is used SVM (Support Vector Machine) model for prediction. The last method is CPAT<sup>27</sup>, which is trained with Hexamer to fit the logistic regression model. The intersection of these three predicted results was considered as new lncRNA for downstream analyses.

## Different Expression of mRNAs and lncRNAs

New lncRNAs, known mRNAs and known lncRNAs were captured from reference genome following annotation file and identification of new lncRNAs. Their expressions were calculated in Fragment PerKilobase Million (FPKM) using Ballgown, which is integrated tool for statistical analysis of assembled transcriptomes. Because of absolutely opposite adaptation for Plateau between Yak and Sanjiang Cattle, we filtered Different Expression (DE) of all lncRNA and known mRNA between Yak group and Sanjiang Cattle group. The species was as covariate factor and the different tissues (cerebellum and cerebrum) were set as adjust variate factor in the model. The Yakow was used in expression trend and association module analyses. Thresholds were set  $|\log_2\text{FoldChange}| \geq 2$  and adjusted P value  $< 0.05$  for significant differentially expressed gene. All FPKMs of different tissues were used to calculate replication relationship. Heatmap was drawn with all FPKMs of different tissues to show the repeatability validated.

Gene ontology enrichment analysis of DE mRNA were performed in R package “clusterProfiler” with database “org.Bt.eg.db”<sup>28,29</sup>.

## Co-expression network between DE RNA and PA

All significant DE known lncRNAs, novel lncRNAs and mRNAs between Yak (M) and Sanjiang Cattle (H) were filtered by Ballgown. All expression levels of these significant gene and lncRNA among the whole three population (Yak, Yakow and Sanjiang Cattle) were constructed to create unsigned co-expression network using the Weighted Gene Co-expression Network Analysis (WGCNA) R package<sup>30</sup>. The soft power

was set 1 to 30 lops to optimize association model for co-expression network and the threshold was 60%. Among all genes and lncRNAs pairwise Pearson's correlations were calculated to create an adjacency matrix. The Topological Overlap Measure (TOM) was estimated by the optimum model using adjacency matrix. Modules of co-expression genes and lncRNAs were identified by using the dynamic tree cut algorithm. Each module was labelled individual color.

In addition, the trait which related to Plateau Adaptation was set 1, 0,-1 for Yak, Yakow and Sanjiang Cattle. Comparing the correlation between the trait and modules which were created by our co-expression network, we selected best correlated module with the PA trait. Heatmap was drawn with the module eigengenes using the first principal component to capture the variation in genes and lncRNA expression. All genes in each module were used to visualize the network relationship figure with the Cytoscape<sup>31</sup>, the nodes were names of whole genes and lncRNAs, the edges were the distance based on the expression levels between each genes and lncRNAs. Major hub genes in the most significant module with PA trait were selected to be forced on with correlated known lncRNA and novel lncRNA. We used hub genes network to reveal the regulation between hub genes and lncRNA.

A summary pipeline of such experimental workflow, bioinformatics statistical software is presented in Fig. 1.

## Results

### RNA Sequencing

Total 18 samples were sequenced in this study. After quality control and adaptor removal, an average of 50 million clean reads for each sample were obtained (Supplementary Table 1). Using Hisat2, each sample was totally mapped an average of 91.57% mapping ratio to the reference genome (GCF\_000298355.1 BosGru v2.0). 84.22% of these were unique mapped reads and other 2.98% were multiple mapped reads.

Total 101,835 transcripts were assembled in the merged assembled result file of 18 samples, of which there were 32,667 transcripts expressed in all 18 samples. All transcript expression levels of 18 samples were presented (Fig. S1) with  $\log_2(\text{FPKM}+1)$ . We also presented the distribution of transcript count per known gene (Fig. S2). There were 3,755 genes with only one transcript and the ratio of total gene was 24.16%, which indicated most of gene expressed with multiple transcripts. We designed a smoothed color density representation (Fig. S3) to validate repeatability and compare the different expression level between Cerebrum and Cerebellum in each sample. The correlated level (from weak to strong) was indicated by white, blue, cyan, yellow and red.

### Identification of lncRNAs

To identify new novel lncRNAs, we further filtered the length, exon number, expression level, and class code of the new transcripts. New transcripts were identified with class codes “i”, “u”, “x”, “j” and “o” from the whole 100,807 transcripts. Next, we removed transcripts which were shorter than 200 nt, and possessed only one exon and with a FPKM  $\leq 1$ . At last, there were 55,542 new transcripts remained for prediction protein coding with three software (CPC2, CNCI and CPAT). CPC2 predicted 21,642 pseudo lncRNAs, CNCI reported 16,819 pseudo lncRNAs and CPAT indicated 36,759 pseudo lncRNAs. Of these, 12,072 pseudo lncRNAs were in intersection of three software (Fig 2A), which were considered as novel lncRNAs. Based on huge different between Yak and Cattle reference, our novel lncRNAs were not matched into NONCODE dataset. Most of new lncRNAs were unknown intergenic (64.17%), 19.75% of new lncRNAs were multi-exon match, 10.32% of new lncRNAs were fully contained with intron, 3.98% of new lncRNAs were overlap with opposite and 1.78% of the new lncRNAs were overlap with exons (Fig. 2B). Comparing the Transcripts Per Million (TPM) distribution of novel lncRNAs and known lncRNAs was presented in Fig. 2C. The most of novel lncRNAs and known lncRNAs were low expressed.

### Different Expression Analysis

To identify DE genes between Yak (M) and Sanjiang Cattle (H), we used  $|\log_2\text{FoldChange}|$  of FPKM to filter transcripts. Total 4,257 significant DE transcripts were identified by using Ballgown R package ( $P < 0.01$ ) with species as covariate factor and tissue as adjust variate factor, 867 of DE transcripts were upregulated and 1,710 were downregulated (Fig. 3A). Of these, 1,021 were protein coding genes, 14 were known lncRNAs, 661 were novel lncRNAs from our prediction. Next, we analyzed significant DE transcripts between Cerebrum and Cerebellum in Yak and Cattle group. There were 2,342 significant DE transcripts ( $P < 0.01$ ), of these 173 were upregulated and 573 were downregulated (Fig. S4), 778 were protein coding genes, 10 were known lncRNAs, 350 were novel lncRNAs from our prediction. Based on significant DE genes, we performed Gene ontology enrichment analysis and used R package “clusterProfiler” with database “org.Bt.eg.db”. The significant DE genes were focused on top 5 GO terms (Fig. 3B), transcription coactivator activity, hormone receptor binding, nuclear hormone receptor binding and snoRNA binding with more than 5 genes on each term.

### Co-expression network

Association co-expression network of whole DE genes and lncRNAs (Additional file FPKM\_allDE\_genes\_lncRNA.csv) was constructed by WGCNA to reveal underlying regulatory function and cluster associated expression modules. The R square of filter criteria was set 0.6 to optimize and select soft power for creating prediction model (Fig. S5). The associated expression modules were clustered into 6 groups (the grey module was mismatching module), and each group (exclude grey module) contained more than 50 genes. The hierarchical relationship between the genes and lncRNAs in modules was drew at Fig. 4A. There were total 944 genes in the largest module with turquoise color. To compare the relationship between each module and PA trait, we put the PA trait into network as an individual module. We found our largest module was strongest relationship with PA trait (Fig. 4B). The PA trait was designed

as threshold trait with 18 samples (see Method). And all modules were clustered into two parties, the one contained turquoise modules and PA trait, the other contained blue, yellow, brown and green modules. For all DE genes and lncRNAs in the co-expression network, the 6 modules according to the topological overlap matrix (TOM) were chosen to visualize the correlation of clusters (Fig. 5A, Additional file TOM\_allDE\_genes\_lncRNA.csv) by Cytoscape. In this network, the fifth nodes group containing the least genes and lncRNAs was grey module which was clustered mismatching genes and lncRNAs, The other five nodes groups correspond with five modules. The edges showed the functional regulatory relationship between genes and lncRNAs. The coding genes, novel lncRNAs and known lncRNAs were marked with red, green and blue (Fig. 5AB).

There were total 979 edges (genes and lncRNAs) in the top significant terms of the first related module. We selected top 10 hub genes (more than 950 edges), 5 known lncRNAs (more than 750 edges) and 5 novel lncRNAs (more than 480 edges) to describe the underlying regulation functional network focusing on hub genes (Fig. 6). A sub-network with GRAMD3 (GRAM domain containing 2B), MAST4 (Microtubule Associated Serine/Threonine 4), RBFox2 (RNA binding FOX-1 homolog 2), MICAL1 (Microtubule Associated Monooxygenase, Calponin And LIM Domain containing 1), TLE1 (TLE family member 1), SFRP4 (Secreted Frizzled Related Protein 4), ERAP1 (Endoplasmic Reticulum Amino Peptidase 1), PINK1 (PTEN Induced Kinase 1), MRPL49 (Mitochondrial Ribosomal Protein L49) and IL34 (Inter Leukin 34) plays a role in cell structure GO terms, such as nucleus, protein binding, cytoplasm and cytosol. All of novel lncRNAs participated in the sub-network, but only LOC106700748 played more effect than other known lncRNAs. Although known gene was with highest edges in whole network, RBFox2 did not participated in the sub-network of hub genes.

## Discussion

In this study, we performed 18 samples of total RNAs from Yak, *Bos grunniens* × *Bos taurus*, Sanjiang Cattle sequencing to identify DE genes, related lncRNAs and describing the co-expression network. The PA trait of Yak is influenced by thousands of genes and lncRNAs<sup>7,19</sup>. Most of these are specific from plain Cattle. Yak and Cattle were estimated to have been diverged approximately 4.9 million years ago, therefore the F1 crossbreed Yakow between Yak and Cattle presented male sterile which could be restored gradually by female Yakow backcross to Yak after several generations<sup>14</sup>. Most of researches about PA between Yak and Cattle have focused on specific DNA or physiological structure, this study is the first time for detection DE genes and lncRNA about PA with Yak, Yakow and Cattle. Consequently, our work provides an important expression regulation network between hub genes and lncRNAs in PA.

### The Genetics Characteristics of Yak, Yakow and Cattle

The method of filtered DE genes was based on expression levels of different environments under same or similar genetic background. In most of researches about traits of different animal species, especially the Cattle or Yak, 3 repetitive individuals were used to reduce the variance in each group. The gaps among each phenotype group were considered to relate to genetic variance and different expression level of

genes. In this study, we performed DE genes and lncRNAs analysis in three species (Yak, Yakow and Cattle). The Yakow shared genetics background with Yak and Cattle, and the phenotype (PA) is similar to Yak<sup>13</sup>, so we just compared the different expression levels of Yak and Cattle. We assumed the candidate gene of PA coming from Yak, therefore, we used Bos Grunniens reference genome to be mapped all RNA-seq reads, making sure that detected genes and the genetics background in each group were same. The reads unique mapping ratio (Table S1) showed the ratio of Yak was the highest, of Yakow was the second and of Cattle was the least, which also indicated the genetics background relationship between each species and reference genome.

## DE Analysis and Co-expression Network

All expression of mRNA surpassed lncRNA, which was the same result as many researches. After filtered DE genes and lncRNA, we just got 14 known lncRNAs, which may depend on the reference genome with 425 known lncRNAs annotation quality. Few researches about lncRNAs have perplexed the development of expression-regulation network in Yak. Our results of novel lncRNAs will prove more information for detection and identification of lncRNAs.

Most of DE genes enriched on Binding and Activity terms, which indicated the function of DE genes were not only single pathway. That was a complex expression-regulation network between genes and other RNA to influence the adaptation of organism. Many previous studies explored the complex relationship between genes and lncRNAs. When we created a predicted functional module to cluster DE genes and lncRNAs, we used indicator (2,1,0) as PA trait in Yak, Yakow and Cattle. It based on the prior observed phenotype from previous researches, which helped us to figure out the relationship between co-expression module and PA trait<sup>32,33,34</sup>. Although PA trait of Yak does not show the twice effect with Cattle, this gradient of phenotype suggested the trend of PA in Yak, Yakow and Cattle.

In the sub-network, there were 10 hub genes, 5 novel lncRNAs and 5 know lncRNAs. Of these hub genes, ERAP1 is a key gene playing the major function of regulated blood pressure in mammal animal<sup>35</sup>. The stress of oxygen levels in the endoplasmic reticulum will drop down, when the ERAP1 is released, and blood pressure is lowered to maintain balance<sup>36</sup>. IL34, GRAMP3, PINK1 and MRPL49 were identified as being associated with regulation of reactive oxygen species and metabolic process<sup>37-42</sup>. The genes RBFOX2, MICAL1, TLE1, SFRP4 and MAST4 were identified in protein binding, ATP binding and magnesium ion binding, they also played function in protein kinase activity and threonine kinase activity<sup>43-49</sup>. All hub genes contained more than 950 edges from whole co-expression network. These evidence suggests that the network of expression and regulation in PA trait is complex and meta-pathway biological processes. We selected top 10 enriched known lncRNA and novel lncRNA to present the co-expression relationship between genes and lncRNAs. The expressions of MSTRG.247.1, MSTRG.431.1, MSTRG.422.1, MSTRG.327.1 and MSTRG.97.1 are correlated with IL34, ERAP1, PINK1, MICAL1, TLE1 and GRAMD3. LOC106700748 is correlated with ERAP1, TLE1, MSTRG.431.1, MSTRG.422.1 and MSTRG.327.1.

# Conclusions

The result had presented that the PA trait was regulated by many genes and lncRNAs, which were composed in an underlying functional network. The function of those hub genes was associated with blood pressure, reactive oxygen and metabolic process. The analysis of co-expressed genes with PA trait provides a reference way of the regulation mechanisms in PA of Yak and reference method of the additive genetics expression detection between crossbreed and parents population.

# Declarations

## Ethics approval and consent to participate

The protocol for Animal Care and Use, was approved by the Animal Ethics and Welfare Association of the Southwest Minzu University (No. 16053) and experiments were performed according to the regulations and guidelines established by this committee.

## Consent to publish

Not applicable

## Availability of data and materials

Whole raw RNA sequence data of 2 tissues of each 9 animals used in this study were submitted to the National Center for Biotechnology Information Sequence Read Archive (SRA) with Accession Number GSE132452 .

## Competing interests

The authors declare that they have no competing interests.

## Funding

This work is supported by the Program of Chinese National Beef Cattle and Yak Industrial Technology System (CARS-37) and Sichuan Science and Technology Program (Grant No.2019YJ0257).

## Authors' Contributions

JBW: Processed and analyzed RNA sequencing data, implemented software, method and drafted the manuscript; LD: Processed qPCR; ZXC: Participated in discussions regarding data analyses; JKW: Participated in discussions regarding interpretation of results; HW: Performed animal phenotyping and analyses for animal selection, collected samples; YT: Participated in discussions regarding interpretation of results; JCZ: Designed experiments, participated in discussions regarding data analyses. All authors made significant contributions editing the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We thank Xing Huang , Jun Hao and Wenbo Ji (all from Southwest Minzu University) for sample collection.

## Authors' Information

1. Key Laboratory of Qinghai-Tibetan Plateau Animal Genetic Resource Reservation and Utilization (Southwest Minzu University), Ministry of Education—Chengdu, Sichuan, 610041; 2. Qinghai-Tibetan Plateau Animal Genetic Resource Reservation and Utilization Key Laboratory of Sichuan Province, Chengdu, Sichuan, 610041; 3. JiLin Agricultural Science and Technology College, Changchun, Jilin, 132101;

## References

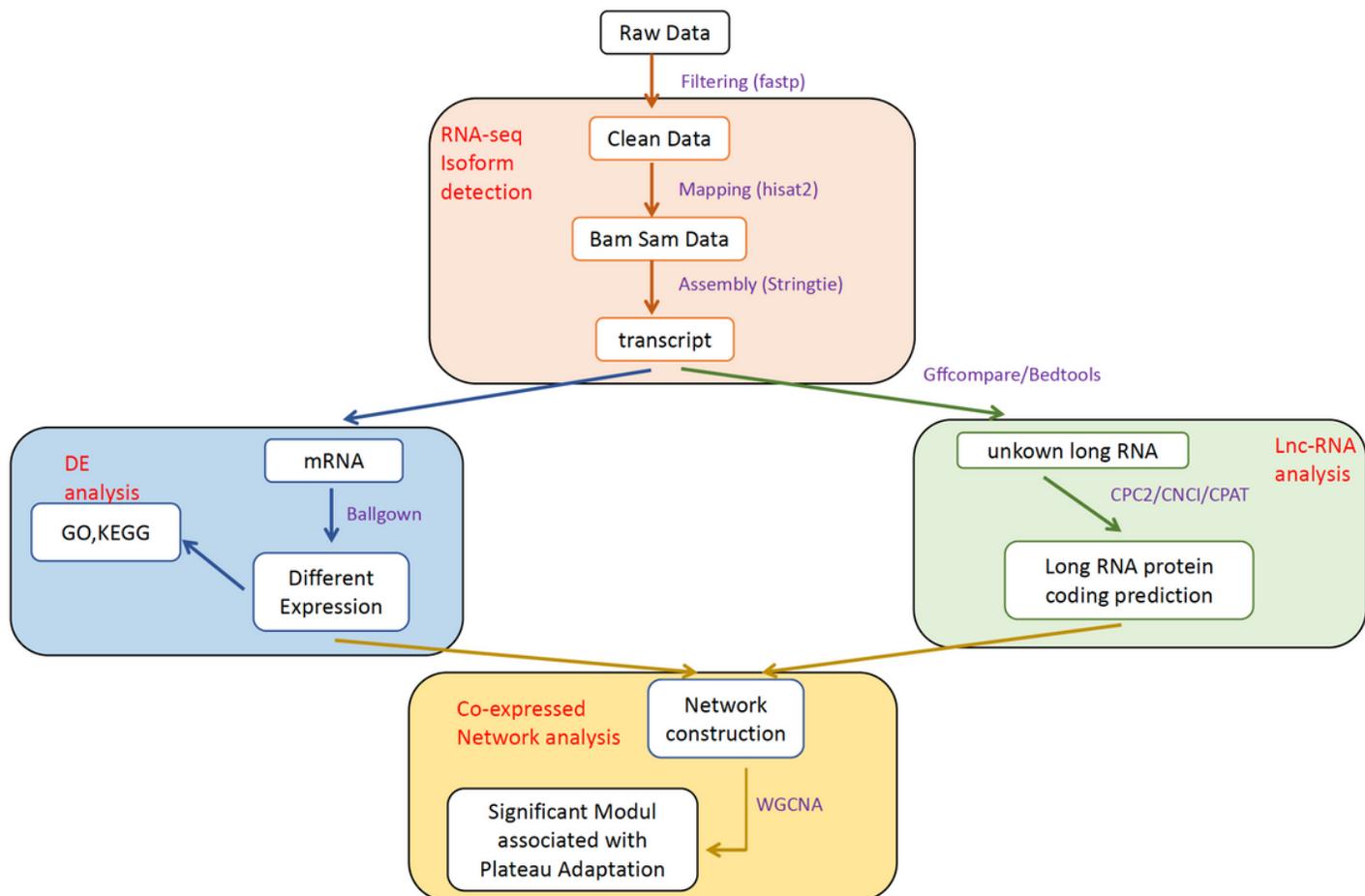
1. Cloutier, S. C., Wang, S., Ma, W. K., Petell, C. J. & Tran, E. J. Long Noncoding RNAs Promote Transcriptional Poising of Inducible Genes. *PLoS Biol.* **11**, 32–34 (2013).
2. Cusanelli, E. & Chartrand, P. Telomeric repeat-containing RNA TERRA: a noncoding RNA connecting telomere biology to genome integrity. *Front. Genet.* **6**, 1–9 (2015).
3. Beck, Z. T. *et al.* Regulation of glucose-dependent gene expression by the RNA helicase Dbp2 in *Saccharomyces cerevisiae*. *Genetics* **198**, 1001–1014 (2014).
4. Riaz, N., Wolden, S. L., Gelblum, D. Y. & Eric, J. HHS Public Access. **118**, 6072–6078 (2016).
5. Huang, L. *et al.* Comparative transcriptome sequencing of tolerant rice introgression line and its parents in response to drought stress. *BMC Genomics* **15**, 1–16 (2014).
6. Salleh, S. M., Mazzoni, G., Løvendahl, P. & Kadarmideen, H. N. Gene co-expression networks from RNA sequencing of dairy cattle identifies genes and pathways affecting feed efficiency. *BMC Bioinformatics* **19**, 1–15 (2018).
7. Qiu, Q. *et al.* Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nat. Commun.* **6**, 1–7 (2015).
8. Hu, Q. *et al.* The Yak genome database: an integrative database for studying yak biology and high-altitude adaption. *BMC Genomics* **13**, 1 (2012).
9. Zhang, X. *et al.* Genome-wide patterns of copy number variation in the Chinese yak genome. *BMC Genomics* **17**, 1–12 (2016).
10. ZHONG Jincheng, *et al.* Mitochondrial complete genome sequencing and phylogenetic research on wild yak. *Acta Ecol. Sin.* (2014). doi:10.5846/stxb201310142481

11. Wu, J. The distributions of Chinese yak breeds in response to climate change over the past 50 years. *Anim. Sci. J.* **87**, 947–958 (2016).
12. Zhang, Y. *et al.* Characterization of the complete mitochondrial genome sequence of golden wild yak and revealed its phylogenetic relationship with 9 yak subspecies. *Mitochondrial DNA Part B Resour.* **4**, 660–661 (2019).
13. Na, R. S. *et al.* Complete mitochondrial genome of the Yakow (*Bos primigenius taurus* × *Bos grunniens*) in China. *Mitochondrial DNA Part A DNA Mapping, Seq. Anal.* **27**, 3826–3827 (2016).
14. Xi, D. *et al.* Isolation and characteristics of the melanocortin 1 receptor gene (MC1R) in the Chinese yakow (*Bos grunniens* × *Bos taurus*). *Gene* (2012). doi:10.1016/j.gene.2012.02.041
15. Xi, D. *et al.* Nucleotide sequence and polymorphism of MHC class II DQB exon 2 alleles in Chinese yakow (*Bos grunniens* × *Bos taurus*). *Int. J. Immunogenet.* (2014). doi:10.1111/iji.12109
16. Medugorac, I. *et al.* Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nat. Genet.* **49**, 470–475 (2017).
17. Wu, D. D. *et al.* Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat. Ecol. Evol.* **2**, 1139–1145 (2018).
18. Ma, Z. J. *et al.* Y-STR INRA189 polymorphisms in Chinese yak breeds. *Genet. Mol. Res.* **14**, 18859–18862 (2015).
19. Qiu, Q. *et al.* The yak genome and adaptation to life at high altitude. *Nat. Genet.* **44**, 946–949 (2012).
20. Kim, D., Langmead, B. & Salzberg, S. L. Hisat2. *Nat. Methods* (2015). doi:10.1038/nmeth.3317
21. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* (2015). doi:10.1038/nbt.3122
22. Frazee, A. C. *et al.* Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat. Biotechnol.* (2015). doi:10.1038/nbt.3172
23. Fan, H. *et al.* A Novel lncRNA Regulates the Toll-Like Receptor Signaling Pathway and Related Immune Function by Stabilizing FOS mRNA as a Competitive Endogenous RNA. *Front. Immunol.* **10**, 1–22 (2019).
24. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btq033
25. Kang, Y. J. *et al.* CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx428

26. Sun, L. *et al.* Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* (2013). doi:10.1093/nar/gkt646
27. Wang, L. *et al.* CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res.* (2013). doi:10.1093/nar/gkt006
28. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–7 (2012).
29. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2016). doi:10.1017/CBO9781107415324.004
30. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
31. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003). doi:10.1101/gr.1239303
32. Shao, B. *et al.* Morphological adaptations of yak (*Bos grunniens*) tongue to the foraging environment of the Qinghai-Tibetan Plateau. *J. Anim. Sci.* (2010). doi:10.2527/jas.2009-2398
33. Wang, H. *et al.* Comparison of nitrogen metabolism in yak (*Bos grunniens*) and indigenous cattle (*Bos taurus*) on the Qinghai-Tibetan Plateau. *Asian-Australasian J. Anim. Sci.* (2011). doi:10.5713/ajas.2011.10350
34. Wang, Z. *et al.* Phylogeographical analyses of domestic and wild yaks based on mitochondrial DNA: New data and reappraisal. *J. Biogeogr.* (2010). doi:10.1111/j.1365-2699.2010.02379.x
35. Hisatsune, C. *et al.* ERp44 Exerts Redox-Dependent Control of Blood Pressure at the ER. *Mol. Cell* **58**, 1015–1027 (2015).
36. Goto, Y., Ogawa, K., Hattori, A. & Tsujimoto, M. Secretion of endoplasmic reticulum aminopeptidase 1 is involved in the activation of macrophages induced by lipopolysaccharide and interferon- $\gamma$ . *J. Biol. Chem.* **286**, 21906–21914 (2011).
37. Wang, Y. *et al.* IL-34 is a tissue-restricted ligand of CSF1R required for the development of Langerhans cells and microglia. *Nat. Immunol.* (2012). doi:10.1038/ni.2360
38. Radaeva, O. & Simbirtsev, A. Investigating a correlation between the levels of peripheral blood cytokines and the risk for cardiovascular complications in patients with stage II essential hypertension. *Bull. Russ. State Med. Univ.* (2019). doi:10.24075/brsmu.2019.006
39. Beveridge, T. J. Structures of gram-negative cell walls and their derived membrane vesicles. *Journal of Bacteriology* (1999).

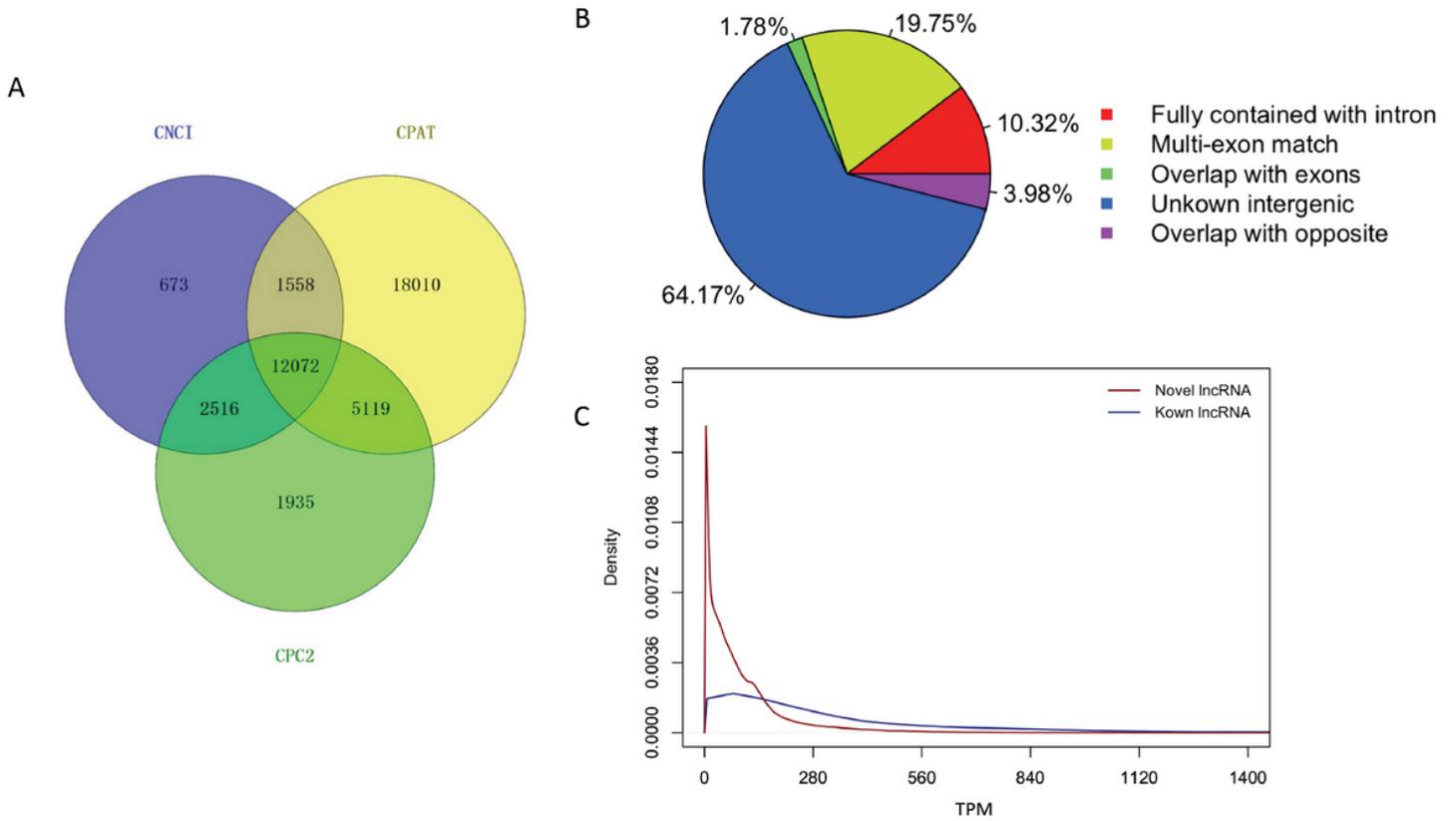
40. Xiao, B. *et al.* Reactive oxygen species trigger Parkin/PINK1 pathway–dependent mitophagy by inducing mitochondrial recruitment of Parkin. *J. Biol. Chem.* **292**, 16697–16708 (2017).
41. Geisler, S. *et al.* PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *Nat. Cell Biol.* (2010). doi:10.1038/ncb2012
42. Jung, S. J., Seo, Y., Lee, K. C., Lee, D. & Roe, J. H. Essential function of Aco2, a fusion protein of aconitase and mitochondrial ribosomal protein bL21, in mitochondrial translation in fission yeast. *FEBS Lett.* (2015). doi:10.1016/j.febslet.2015.02.015
43. Gehman, L. T. *et al.* The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. *Genes Dev.* (2012). doi:10.1101/gad.182477.111
44. Braeutigam, C. *et al.* The RNA-binding protein Rbfox2: An essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion. *Oncogene* (2014). doi:10.1038/onc.2013.50
45. Lee, B. C. *et al.* MsrB1 and MICALs Regulate Actin Assembly and Macrophage Function via Reversible Stereoselective Methionine Oxidation. *Mol. Cell* (2013). doi:10.1016/j.molcel.2013.06.019
46. Dastidar, S. G., Narayanan, S., Stifani, S. & D’Mello, S. R. Transducin-like enhancer of split-1 (TLE1) combines with Forkhead box protein G1 (FoxG1) to promote neuronal survival. *J. Biol. Chem.* (2012). doi:10.1074/jbc.M111.328336
47. Imai, Y. *et al.* TLE, the human homolog of Groucho, interacts with AML1 and acts as a repressor of AML1-induced transactivation. *Biochem. Biophys. Res. Commun.* (1998). doi:10.1006/bbrc.1998.9705
48. Muley, A. *et al.* Secreted frizzled-related protein 4: An angiogenesis inhibitor. *Am. J. Pathol.* (2010). doi:10.2353/ajpath.2010.090465
49. Ikram, M. A. K. *et al.* Novel genetic loci associated with hippocampal volume. *Nat. Commun.* (2016). doi:10.1038/ncomms13624

## Figures



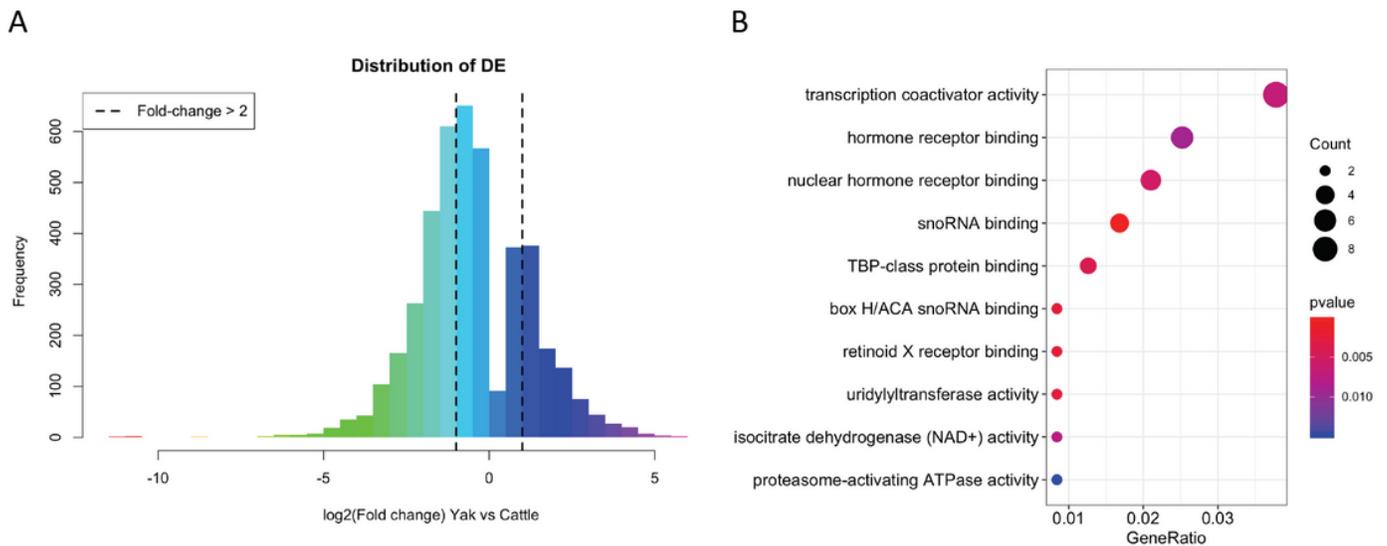
**Figure 1**

The experimental design and the pipeline of mRNA-lncRNA co-expressed network. An overview of the identification for Different Expression gene and lncRNA was divided into four parties for different analysis. We used whole DE gene, known lncRNA and novel lncRNA to create co-expressed network with WGCNA.



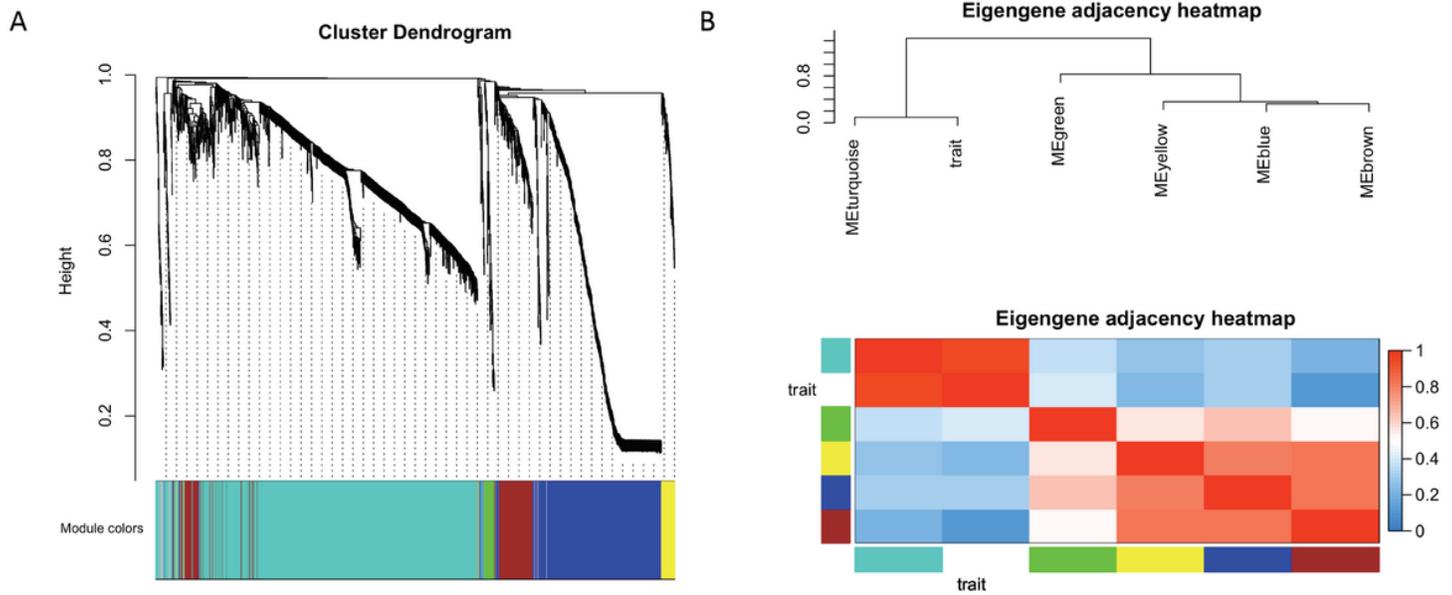
**Figure 2**

Genomic features and classification of novel lncRNA. A the intersection of predictive long noncoding RNA by three methods (CNCI, CPC2 and CPAT). B classification of novel lncRNA according to “class code” showing the relationship between a transcript and the closest reference transcript. C the TPM distribution of Novel lncRNA and known lncRNA.



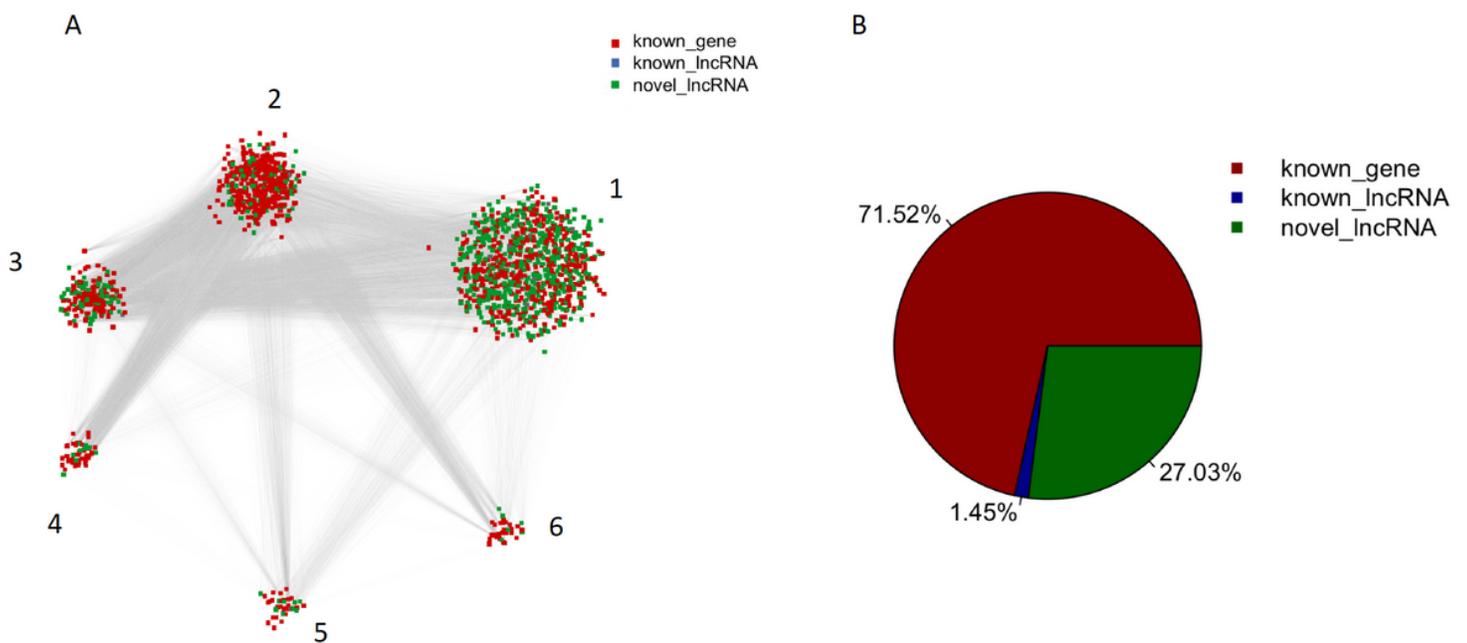
**Figure 3**

The screening and enrichment of Different Expression gene in Yak and Cattle group. A the DE gene were filtered between Yak and Cattle with 2 times Fold change. B top 10 GO terms were enriched with whole DE genes.



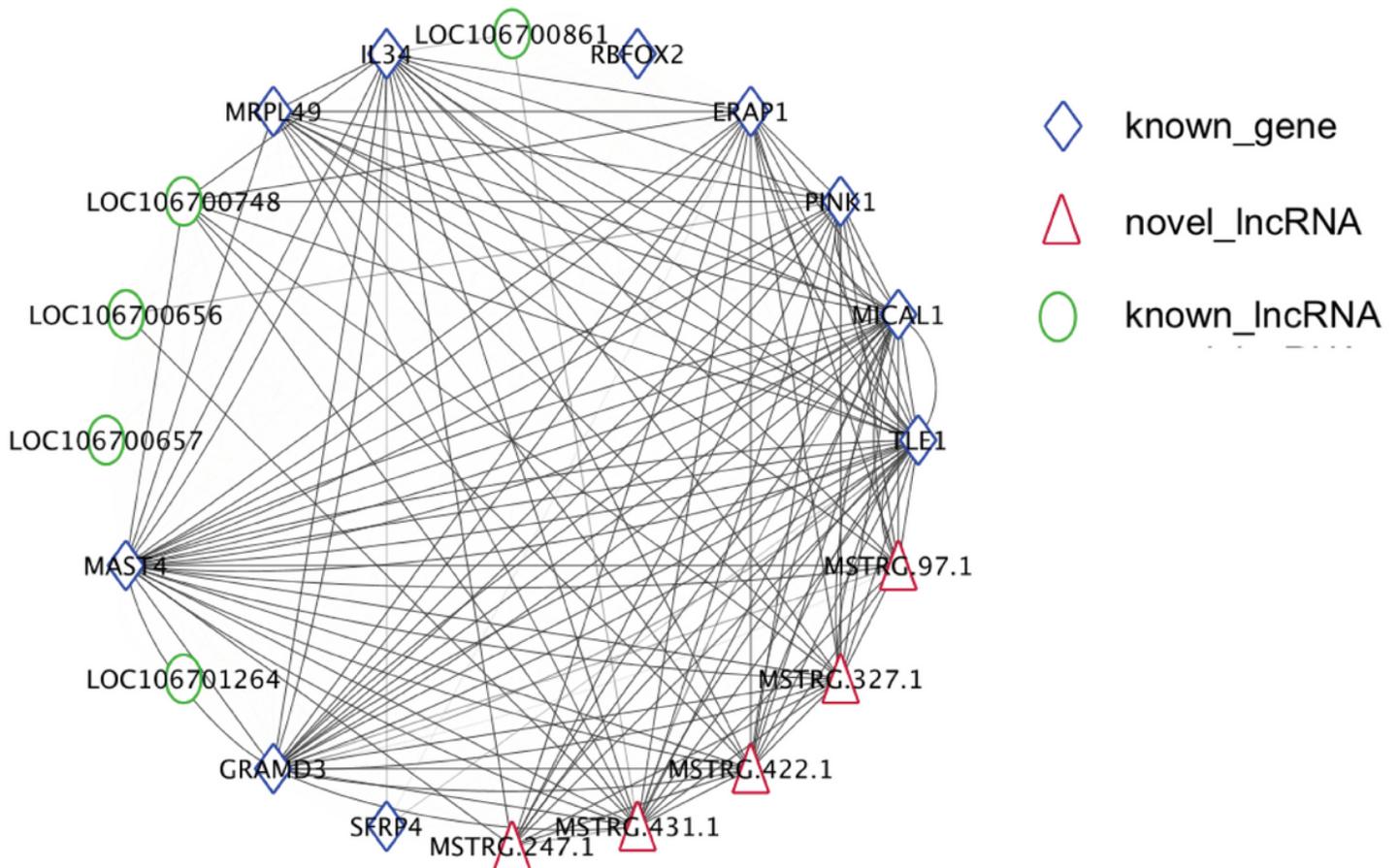
**Figure 4**

Optimum modules with whole coding gene, known lncRNA and novel lncRNA. A total five modules were created with the expression of whole coding gene, known lncRNA and novel lncRNA. B the relationship between PA trait and five modules.



**Figure 5**

Co-expression network in modules with whole coding gene, known lncRNA and novel lncRNA. A co-expression network visualization: nodes show coding gene (red), known lncRNA (blue) and novel lncRNA (green). B Classification of RNA in whole co-expression network.



**Figure 6**

Sub-network with most enrichment genes and lncRNAs. The top 10 enrichment hub genes (Blue Diamond) in whole network, top 5 known lncRNAs (Green Circle) and top 5 novel lncRNAs (Red Triangle) were selected to build sub-network. The more enrichment line indicated more related regulation of expression.

## Supplementary Files

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

- [FPKMallDEgeneslncRNA.csv](#)
- [F.S2.pdf](#)

- GOturquoisegenes.csv
- SupplementrayFigureandTable.docx
- F.S3.pdf
- F.S5.pdf
- F.S1.pdf
- F.S4.pdf