Genome-wide identification and characterisation of long non-coding RNAs in two Chinese cattle breeds

Xianbo Jia, Yang He, Shi-Yi Chen, Jie Wang, Shenqiang Hu and Song-Jia Lai

ABSTRACT

Long non-coding RNAs (lncRNAs) have been essentially involved in all kinds of important biological processes, such as cell differentiation and development. In this study, we comprehensively profiled genome-wide lncRNAs of bovine skeletal muscle in two Chinese cattle breeds using RNA-seq approach. The longissimus dorsi samples were collected from three Charolais (CH) and three Tongjiang (TJ) adult cattle, respectively. Approximately 80 million reads were obtained per library, and by which we assembled 31,009 unique transcripts. We successfully identified a total of 4155 different lncRNAs in six muscle samples, consisting of 3204 intronic lncRNAs, 737 intergenic lncRNAs, 113 sense lncRNAs and 101 antisense lncRNAs. Among them, 32 lncRNAs were differentially expressed (Fold Change $>1.0$, FDR $<0.05$) between CH and TJ. Among these lncRNAs, 1744 identified bovine lncRNAs were located adjacent to 1146 protein-coding genes ($<100$ kb upstream and downstream) and functionally enriched in organ development, gene regulation and signal transduction. Our results provide a catalogue of bovine skeletal muscle-related lncRNAs, which will complement the list of bovine lncRNAs and contribute to understanding of the molecular mechanism of underpinning muscle development in cattle.

HIGHLIGHTS

- The first study to compare specialised beef cattle breed with Chinese local beef cattle breed on long non-coding RNA (lncRNA).
- The 4155 different lncRNAs in beef cattle genome were identified to complement the list of bovine lncRNAs.
- The 1146 protein-coding genes were found maybe in cis-regulatory relationships with lncRNAs to understand the molecular mechanism underpinning muscle development in cattle.

Introduction

Long non-coding RNA (lncRNA) is a kind of non-coding RNA longer than 200 nucleotides, and some lncRNAs are even longer than 100 kb (Ponting et al. 2009). Initially, lncRNA was regarded as the ‘noise’ of genomic transcription with no biological function, a by-product of RNA polymerase II transcription. Borsani et al. (1991) reported that Xist was involved in the regulation of X chromosome inactivation. Since then, increasing studies have shown that lncRNAs play important roles in many biological processes, including transcriptional regulation, epigenetic modification, cell differentiation, development, reproduction, aging, as well as in varieties of tumours and other diseases (Bonasio and Shiekhattar 2014; Meseure et al. 2015; Taylor et al. 2015; Schmitz et al. 2016; Bolha et al. 2017; Kopp and Mendell 2018; Fernandes et al. 2019; Wu et al. 2019). Thus, lncRNA has attracted extensive attention from researchers in the field of biology. lncRNA functions are mainly based on its secondary structure, which binds to proteins, can cause chromatin remodelling and affect the function of transcription factors (Fernandes et al. 2019). Also, it can bind to microRNAs at its linear level to affect the expression of messenger RNA (mRNA) indirectly (Zhao et al. 2019). In addition, it can bind to the mRNAs directly to affect their translation, splicing and degradation.
With the development of genetic array and high throughput sequencing technology, more and more lncRNAs have been discovered in human, rat and mouse (Zhao et al. 2016). Recently, lncRNAs have also been identified in pig (Yang et al. 2017; Jin et al. 2018; Kern et al. 2018), chicken (Muret et al. 2017; Kern et al. 2018; Li et al. 2019), sheep (Ren et al. 2016; Ling et al. 2017; Bush et al. 2018; Sulayman et al. 2019) and cattle (Billerey et al. 2014; Koufariotis et al. 2015; Tong et al. 2017; Bush et al. 2018; Ibeagha-Awemu et al. 2018; Kern et al. 2018; Weikard et al. 2018; Yang et al. 2018; Choi et al. 2019; Wang et al. 2019; Yue et al. 2019).

Besides, several studies thus far have investigated the bovine lncRNAs. Huang et al. (2012) predicted a total of 449 putative lncRNAs located in 405 intergenic regions from public bovine-specific ESTs. Weikard et al. (2013) identified 4848 potential lncRNAs in bovine skin samples (pigmented and non-pigmented) by deep next-generation sequencing. Billerey et al. (2014) defined a set of 584 different lincRNAs with 418 lincRNAs found in all nine muscle samples used paired-end RNA sequencing. Koufariotis et al. (2015) captured 9778 high-confidence putative lncRNAs from RNA-Seq data across 18 tissues of a single cow. Liu et al. (2017) identified 7188 bovine skeletal muscle lncRNAs by RNA-Seq. Kern et al. (2018) identified 7235 lncRNAs in cattle using RNA-seq from eight tissue types. Yang et al. (2018) detected 3746 lncRNAs in mammary glands of Holstein cows using RNA-seq. Ibeagha-Awemu et al. (2018) detected 7235 lncRNAs in cattle using RNA-seq from eight tissue types. Yang et al. (2018) detected 3746 lncRNAs in mammary glands of Holstein cows using RNA-seq. Ibeagha-Awemu et al. (2018) detected 4995, 1568 and 4243 lncRNAs in bovine mammary gland, ileum and rumen tissues using RNA-seq. Li et al. (2018) found 28,558 lncRNAs expressed in the bovine mammary gland by RNA-seq. However, the number of all recently identified bovine lncRNAs identified is rather lower than that of human and mouse (Zhao et al. 2016), suggesting that further efforts are needed to comprehensively discover all bovine lncRNAs.

Charolais (CH) cattle is a worldwide specialised beef cattle breed. Tongjiang (TJ) cattle is a local cattle population of Sichuan Province, China, which has long been used as work cattle. It is a small type and slowly growing cattle with high meat quality. At present, TJ cattle is in the process of herd improvement to beef. So, it is the necessary work to compare the genetic structure between the two breeds. In this study, we reported the identification and characterisation of lncRNAs in longissimus dorsi of CH cattle and TJ cattle using an Illumina HiseqX platform. Our results will provide a useful resource for understanding the function and regulation of lncRNAs in beef cattle and annotating the bovine genome, as well as contribute to better investigate skeletal muscle development in mammals.

Material and methods

Ethics statement

All animal experimental protocols were approved by Biological Studies Animal Care and Use Committee at Institute of Animal Sciences, Sichuan Agricultural University, Chengdu, China.

Animals and sample preparation

In this study, six bovine longissimus dorsi samples were taken from three CH cattle and three TJ adult cattle, respectively. The six cattle had no genetic relationship with each other for at least three generations. They were free access to water and food under natural lighting and had not suffered any diseases and taken any drugs in the previous 1 year. After slaughtering at 17 months old, samples were collected from longissimus dorsi muscle between 12th and 13th rib within 30 min. All samples were immediately frozen in liquid nitrogen and then stored at −80 °C for RNA extraction.

RNA isolation and Illumina deep sequencing

Total RNA from six samples was isolated using RNAprep pure Tissue Kit (DP431, TIANGEN, Beijing, China) following the manufacturer’s instruction. The integrity and purity of total RNA were analysed by NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA) and RNA Nano 6000 Assay Kit in a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) with RNA integrity number >8. Complementary (cDNA) libraries were constructed using approximately 3 ug of total RNA per sample. Clustering of the index-coded samples used the TruSeq PE Cluster Kit v3-cBot- HS on the cBot Cluster Generation System (Illumina, San Diego, CA). After cluster generation, the libraries were sequenced on an Illumina HiseqX platform and 150 bp paired-end reads were generated.

Transcriptome assembly

Raw reads were first subjected to quality filtering, which removed the reads containing 5% of poly(N), and low quality reads (>20% of the bases with Q ≤ 10) to generate clean reads. Then, the Q20, Q30 and GC% of the clean data were calculated. The clean
reads were mapped to the reference genome (Btau_5.0.1, NCBI) using HISAT2 tool and default parameters (Kim et al. 2015). The mapped reads from each library were assembled with StringTie software (Pertea et al. 2015), and multiple assembled transcript files were merged to produce a unique set of transcripts using the Taco software (Niknafs et al. 2017).

**Gene expression analysis**

Transcript abundance was quantified as the transcripts per million (TPM) (Li et al. 2010). The transcripts with TPM level < 1 were filtered out. The significance of the differentially expressed IncRNAs was analysed using an R package DESeq2 (Love et al. 2014) and multiple comparisons were corrected for p value with the false discovery rate (FDR). A FDR value < .05 and absolute value of log2FoldChange > 1 were used as the threshold to statistically define the significant difference.

**Identification of putative IncRNAs**

To obtain putative IncRNAs, a stringent analysis pipeline was used to filter the assembled transcripts as follows: (1) Transcripts overlapping with bovine protein-coding genes were removed. (2) Single exon transcripts were removed. (3) Transcripts less than 200 bp were removed. (4) Transcripts with coding potential score predicted by the coding potential calculator (CPC) (Kong et al. 2007) ≥ −1 were removed. After these steps, the remaining transcripts were considered as the putative IncRNAs.

**Enrichment analysis**

To predict the target genes of IncRNAs, we searched for protein-coding genes 100-kb up- and down-stream of each putative IncRNA. These target genes were subjected to functional exploration according to the Gene Ontolog (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) enrichment analysis using DAVID online analytical tools to predict their functional roles (Huang da et al. 2009). GO terms and KEGG pathways with p < .05 and Fold Enrichment > 2 were considered to be significantly enriched.

**Validation by qRT-PCR**

Six IncRNAs were randomly chosen for qRT-PCR validation, whose primers were designed using Primer3 online tool (Table 1). The total RNAs for qRNA-PCR were converted to cDNA using the PrimeScript RT reagent Kit (TaKaRa, Dalian, China) following the manufacturer’s instructions. The PCR was performed according to the SYBR Premix Ex Taq™ II instructions (TaKaRa) on a CFX96 Real-Time PCR System (Bio-Rad, Hercules, CA). The PCR cycling condition involved 2 min incubation at 95 °C, followed by 39 times of 95 °C for 10 s, Tm for 10 s (Tm indicated in Table 1). The relative gene expression levels were calculated based on GAPDH gene expression using the $2^{-\Delta\Delta C_{t}}$ method.

**Results**

**Overview of RNA sequencing and transcriptome assembly**

To comprehensively and accurately identify IncRNAs expressed in the bovine longissimus dorsi muscle, we constructed six cDNA libraries (CH1-3, TJ1-3) from longissimus dorsi muscle of CH cattle and TJ cattle in this study. A total of 489,465,128 raw reads were generated in all six libraries with a length of 150 bases, resulting in a total of 73.42 gigabases. The result of RNA quality was shown in Table 2. After filtering out adaptors, poly-N, low quality and rRNA reads, a total of 488,266,748 clean reads were remained. The percentage of clean reads in each library ranged from 99.48% to 99.94%. The reads were then aligned into bovine Btau_5.0.1 reference genome using the HISAT2 software. Approximately 94.74% (94.32%~95.14%) of the total clean reads were mapped to the bovine reference genome and 66.66% (65.61%~68.19%) of the reads were uniquely mapped to specific regions of the bovine genome. The mapped sequences in each library were reconstructed and assembled using the

| Table 1. Primers used in the qPCR analysis. |
|--------------------------------------------|
| **Gene name** | **Forward primer** | **Reverse primer** | **Product size, bp** | **Tm, °C** |
| MSTRG.5226.1 | ttctggacctatgagaggag | cggcaagacagggagag | 187 | 60.1 |
| MSTRG.11027.1 | gcacaacttcactctca | cttaccacttcaacctctcc | 231 | 60.1 |
| MSTRG.2792.1 | ctgctgcgtctctctcttc | ggcgtctctctctctctctct | 264 | 59.8 |
| MSTRG.3551.1 | tggcttcttgtgccttcttc | gaggagaggagaggagctt | 297 | 60.5 |
| MSTRG.6194.7 | aagtctctctctctctctct | gtcctctctctctctctctct | 257 | 60.0 |
| MSTRG.8585.2 | ttgaagacagagactggctga | atagaagaccgatggagatgg | 288 | 59.8 |
| GAPDH | cgaccactttctgatgccat | tctcagtgtggtggagatgg | 288 | 59.8 |
StringTie and a total of 31,009 unique assembled transcripts were obtained.

**Identification of putative lncRNAs in bovine skeletal muscle**

To identify lncRNAs, we developed a stringent filtering pipeline to discard transcripts with one or more of the following: homology to known genes, lengths < 200 bp, single exon and CPC score $\leq -1$. A total of 4155 lncRNA candidates were identified by our pipeline (Supplementary Table S1), including 3204 intronic lncRNAs, 737 intergenic lncRNAs, 113 sense lncRNAs and 101 antisense lncRNAs in the six libraries (Figure 1(a)). The 4155 lncRNAs were distributed, but not equally, over all chromosomes (Figure 1(b)), with the highest density on chromosome X (212 lncRNAs), whereas the lowest density was on chromosome Y (1 lncRNA). There were also 19 lncRNAs located in Bos taurus unplaced genomic scaffolds. The size of 4155 lncRNAs ranged from 200 to 8199 bp with an average length of 2114 bp (Figure 1(c)). Approximately 80.51% of lncRNAs were shorter than 2800 bp. Their average number of exon was 3.24 with a range of 2–16 (Figure 1(d)). Most of these lncRNAs contained 2–4 exons, accounting for 87.32% of the 4155 lncRNAs.

**Identification of differentially expressed lncRNAs**

Expression levels of lncRNA transcripts were estimated by Transcripts Per Kilobase Million (TPM) using Kallisto software. A total of 263 lncRNAs were expressed specifically in longissimus dorsi of CH cattle, while 180 lncRNAs expressed specifically in that of TJ cattle (Figure 2). A total of 3712 lncRNAs were expressed in the longissimus dorsi muscle of both breeds, of which 2999 lncRNAs were expressed in longissimus dorsi muscle of all individuals. The R package DESeq was used to get differential expression levels of lncRNAs between the two breeds. Differential expression analysis (CH vs. TJ) suggested that 20 lncRNAs were identified up-regulation and 12 lncRNAs were identified down-regulation in CH cattle (Supplementary Table S1). Principal components analysis (PCA) of differentially expressed lncRNAs showed distinct clustering of CH cattle and TJ cattle (Figure 2).

**Prediction of lncRNA functions**

A number of studies have indicated that lncRNAs can act in cis to regulate nearby coding genes (Marchese et al. 2017). We predicted the potential targets of lncRNAs upon applying a cut-off of 100 kb upstream and downstream of the 4155 lncRNAs. A total of 1146 protein-coding genes were identified as the nearest neighbours of the 1744 lncRNAs. Then we performed gene ontology analysis of these 1146 protein-coding genes to explore the lncRNAs functions using DAVID Bioinformatics Resources. The analysis was performed on biological process (GOTERM_BP_DIRECT), cellular component (GOTERM_CC_DIRECT), molecular function (GOTERM_MF_DIRECT) and KEGG_PATHWAY using the functional annotation tool with a corrected Fisher exact $p$ value $<.05$ and folder enrichment $>2$. GO functional annotation showed that 239 genes were functionally enriched in 38 GO terms, including 23 biological processes, 7 cellular component and 8 molecular function (Table 3). These terms were mainly associated with organ development, gene regulation and signal transduction, such as skeletal muscle cell differentiation (GO:0035914), heart development (GO:0007507), chondrocyte differentiation (GO:0002062), positive regulation of MAP kinase activity (GO:0043406), positive regulation of cell migration (GO:0030335), positive regulation of vascular permeability (GO:0043117), post-synaptic membrane (GO:0045211), voltage-gated potassium channel complex (GO:0008076) and cAMP binding (GO:0005057). For pathways, 52 genes were enriched in

### Table 2. Summary of reads mapping to the bovine transcriptomes.

| Sample | CH1 | CH2 | CH3 | TJ1 | TJ2 | TJ3 |
|--------|-----|-----|-----|-----|-----|-----|
| Number of raw data | 82,371,264 | 81,802,544 | 80,607,132 | 82,070,242 | 81,279,218 | 81,334,728 |
| Number of raw base (Gb) | 12.36 | 12.27 | 12.09 | 12.31 | 12.19 | 12.20 |
| % Clean data Q20 | 97.45% | 97.43% | 97.23% | 97.67% | 97.30% | 97.19% |
| % Clean data Q30 | 93.80% | 93.72% | 93.37% | 94.26% | 93.47% | 93.25% |
| % GC Content | 50.71% | 49.72% | 49.20% | 49.40% | 49.30% | 48.70% |
| Number of total Clean Reads | 81,942,408 | 81,707,456 | 80,541,884 | 81,856,618 | 80,933,636 | 81,284,746 |
| % total clean reads | 99.48% | 99.88% | 99.92% | 99.74% | 99.57% | 99.94% |
| Number of mapped reads | 77,677,753 | 77,738,137 | 76,483,186 | 77,508,940 | 76,522,548 | 76,667,123 |
| % mapped reads | 94.80% | 95.14% | 94.96% | 94.69% | 94.55% | 94.32% |
| Number of uniquely mapped reads | 54,295,039 | 55,699,972 | 54,921,510 | 54,131,781 | 53,092,465 | 53,330,921 |
| % uniquely mapped reads | 66.26% | 68.17% | 68.19% | 66.13% | 65.60% | 65.61% |

CH: Charolais; TJ: Tongjiang.
5 KEGG pathways, including gap junction (bta04540), insulin signalling pathway (bta04910), prostate cancer (bta05215), insulin resistance (bta04931) and glutamatergic synapse (bta04724).

Experimental validation of putative lncRNAs

Six differently expressed lncRNAs were randomly selected for validation by qPCR. The results shown that these six lncRNAs were expressed in all samples.
## Table 3. The GO term and KEGG functional annotation of genes nearby IncRNAs.

| Ontology | Term | Description | Folder E. | p Value |
|----------|------|-------------|-----------|---------|
| BP       | GO:0021987 | Cerebral cortex development | 4.39 | .00 |
|          | GO:0035914 | Skeletal muscle cell differentiation | 3.52 | .01 |
|          | GO:0043406 | positive regulation of MAP kinase activity | 3.96 | .01 |
|          | GO:0050852 | T cell receptor signalling pathway | 3.85 | .01 |
|          | GO:0048565 | Digestive tract development | 4.24 | .01 |
|          | GO:0006783 | Heme biosynthetic process | 6.59 | .02 |
|          | GO:0042733 | Embryonic digit morphogenesis | 3.22 | .02 |
|          | GO:0007507 | Heart development | 2.18 | .02 |
|          | GO:0046513 | Ceramide biosynthetic process | 4.49 | .02 |
|          | GO:0021766 | Hippocampus development | 3.60 | .02 |
|          | GO:0002062 | Chondrocyte differentiation | 3.60 | .02 |
|          | GO:0007156 | Homophilic cell adhesion via plasma membrane adhesion molecules | 2.24 | .02 |
|          | GO:0030335 | positive regulation of cell migration | 2.02 | .03 |
|          | GO:0048041 | Focal adhesion assembly | 5.65 | .03 |
|          | GO:0007411 | Axon guidance | 2.25 | .03 |
|          | GO:0035690 | Cellular response to drug | 3.21 | .04 |
|          | GO:0007286 | spermatid development | 2.51 | .04 |
|          | GO:0035019 | somatic stem cell population maintenance | 3.80 | .04 |
|          | GO:0007157 | Heterophilic cell–cell adhesion via plasma membrane cell adhesion molecules | 3.04 | .05 |
|          | GO:0043117 | Positive regulation of vascular permeability | 8.48 | .05 |
|          | GO:0032836 | Glomerular basement membrane development | 8.48 | .05 |
|          | GO:0030814 | Regulation of cAMP metabolic process | 8.48 | .05 |
|          | GO:0006198 | cAMP catabolic process | 8.48 | .05 |
|          | GO:0005911 | Cell–cell junction | 2.72 | .00 |
|          | GO:0042629 | Mast cell granule | 7.12 | .02 |
|          | GO:0045211 | Postsynaptic membrane | 2.13 | .03 |
|          | GO:0042734 | Presynaptic membrane | 3.76 | .04 |
|          | GO:0031256 | Leading edge membrane | 8.39 | .05 |
|          | GO:0000164 | Protein phosphatase type 1 complex | 8.39 | .05 |
|          | GO:0008076 | Voltage-gated potassium channel complex | 2.41 | .05 |
|          | GO:0005057 | Receptor signalling protein activity | 8.01 | .00 |
|          | GO:0030552 | cAMP binding | 5.27 | .01 |
|          | GO:0003823 | Antigen binding | 4.77 | .02 |
|          | GO:0004879 | RNA polymerase II transcription factor activity, ligand-activated sequence-specific DNA binding | 4.17 | .03 |
|          | GO:0008603 | cAMP-dependent protein kinase regulator activity | 10.01 | .03 |
|          | GO:0005201 | Extracellular matrix structural constituent | 3.25 | .04 |
|          | GO:0005546 | Phosphatidylinositol-4,5-bisphosphate binding | 3.16 | .04 |
|          | GO:004115 | 3′,5′-cyclic-AMP phosphodiesterase activity | 8.58 | .04 |
| KP       | bta04540 | Gap junction | 3.03 | .00 |
|          | bta04910 | Insulin signalling pathway | 2.16 | .02 |
|          | bta05215 | Prostate cancer | 2.57 | .02 |
|          | bta04931 | Insulin resistance | 2.21 | .04 |
|          | bta04724 | Glutamatergic synapse | 2.17 | .04 |

BP: biological process; CC: cellular component; MF: molecular function; KP: KEGG pathway; Folder E.: folder enrichment.

![Figure 3](image-url)  // Validation of six differentially expressed lncRNAs by qPCR. Data are the mean ± SEM. *p<.05; **p<.01.
And their expression patterns were consistent with the expression levels of the RNA-seq. This result suggests that most of our putative lncRNAs would be truly expressed in vivo (Figure 3).

Discussion

lncRNA rapidly became a hot spot in the field of genetic research, since it was demonstrated as important regulators of gene expression in a variety of biological processes. An increasing number of lncRNAs have been identification and characterised in multiple species with the developed next-generation sequencing technology, especially with the advent of high-throughput RNA-seq methods (Jathar et al. 2017; Cao et al. 2019). The latest NONCODE database source contains 548,640 lncRNA transcripts, which points to 172,216 lncRNAs from the human genome, 131,697 lncRNAs from the mouse genome and only 23,515 lncRNAs from the bovine genome (Zhao et al. 2016). The number of lncRNAs in the bovine genome seems to be much underestimated. In previous studies, the EST and RNA-seq data were both used to identify lncRNAs in cattle. For example, 21,951 lncRNAs and 449 putative lncRNAs were identified from bovine ESTs data (Huang et al. 2012). On the other hand, several studies reported that thousands of lncRNAs were detected in diverse bovine tissues by RNA-seq data (Kern et al. 2018; Choi et al. 2019; Wang et al. 2019). In this study, we identified a total of 4155 bovine lncRNAs based on transcriptomic data from six RNA-seq datasets in longissimus dorsi muscle of CH and TJ cattle. This result increases in the number of bovine lncRNAs and further improves the current database of information on bovine lncRNAs.

lncRNAs can be either single-exonic or multi-exonic, while it is a challenge to distinguish true lncRNAs from the abundant one exon, low expression fragments and expression noise assembled from RNA-Seq data (Cabili et al. 2011; Derrien et al. 2012). Thus, we only chose multi-exonic lncRNAs for further analysis, as was done in other studies (Tong et al. 2017; Kern et al. 2018; Yang et al. 2018; Yue et al. 2019), and depends on a highly stringent filtering pipeline to minimise the number of false-positive lncRNAs. Consequently, 4155 non-protein-coding transcripts were considered as highly reliable lncRNAs, which was less than previous studies due to our stringent criteria (Billerey et al. 2014; Liu et al. 2017; Choi et al. 2019; Yue et al. 2019). These lncRNAs have different expression distribution in different chromosomes. The vast majority of putative lncRNAs were intronic or intergenic in our study, which confirmed that intragenic and intergenic regions were the major sources of non-coding RNAs (Mattick and Makunin 2006; Xu, Bai, et al. 2017). lncRNAs are known to be shorter transcript lengths and fewer exon numbers than mRNAs (Wang et al. 2018; Sulayman et al. 2019). Approximately 80.51% of our 4155 lncRNAs was shorter than 2800 bp and 87.32% of them had 2–4 exons, which was in agreement with previous studies in cattle (Billerey et al. 2014; Liu et al. 2017; Choi et al. 2019; Yue et al. 2019). These investigations provide valuable information about the characteristics of lncRNAs in bovine.

Although thousands of lncRNAs have been identified and analysed in bovine muscle, several lncRNAs have been reported to play important roles in bovine myoblast proliferation and differentiation (Sun et al. 2016; Jin et al. 2017; Xu, Ji, et al. 2017; Yue et al. 2019). Muscle mass and growth rate are two economically important traits in beef production. The differences in muscle mass and growth between CH and TJ cattle provide a good model for studying the mechanisms that regulate muscle development (Chen et al. 2008). In this study, 32 lncRNAs were differentially expressed between CH and TJ cattle and may be involved in muscle development. It may provide a better understanding of the dynamic gene regulation of muscle mass and growth differences between the two breeds and lays important groundwork for further research regarding the regulatory role of lncRNA in muscle development.

Emerging evidence suggests that lncRNAs often as cis-acting element are actively involved in various developmental and physiological processes. For example, the bovine lncRNAs TCONS_00041733 regulate differentially expression coding gene EFNA1, which is involved in vascular development, inflammatory response and cell proliferation and angiogenesis as well as sperm morphology and membrane functionality (Wang et al. 2019). In the cis prediction, we sought out 4155 lncRNAs that neighboured 1146 protein-coding genes (<100 kb upstream and downstream). These neighbouring genes were significantly functionally enriched in 38 GO terms (23 biological process, 7 cellular component and 8 molecular function) and 5 KEGG pathways, which were mainly associated with organ development, gene regulation and signal transduction. Eight of the cis target protein-coding genes enriched in GO term associated with skeletal muscle cell differentiation (BCL9L, HIVEP3, MEF2D, RB1, MAFF, EMD, BTG2 and ASB2), indicating the important roles of lncRNAs in skeletal muscle development. It was consistent with previous studies that lncRNAs regulated nearby protein-coding genes to involve in muscle development and diseases (Zhan et al. 2016; Liu et al. 2017; Yue et al. 2019). For
example, the most recent study reported that the bovine IncKBTBD10 affect its proximity gene KBTBD10 to involve in myogenesis (Chen et al. 2019).

Conclusions

In this study, 4155 IncRNAs were identified in six bovine longissimus dorsi muscle tissues by RNA-seq technology. The major characteristics of IncRNAs were similar to previous studies, including unequal distribution on chromosomes, relatively shorter ORF and fewer exons. Thirty-two different expression IncRNAs were found out between CH cattle and TJ cattle. The identified IncRNAs may act in cis to target 1146 protein-coding genes, which were mainly associated with organ development, gene regulation and signal transduction. This study provides a useful database of bovine muscle IncRNAs for future investigations of their regulatory roles in bovine muscle development.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was financially supported by the Earmarked Fund for Supported by Sichuan Science and Technology Programme [Grant No: 2016NYZ0050].

ORCID

Jie Wang http://orcid.org/0000-0003-0269-0677

References

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

Bolha L, Ravnik-Glavac M, Glavac D. 2017. Long noncoding RNAs as biomarkers in cancer. Dis Markers. 2017:7243968.

Bonasio R, Shiekhattar R. 2014. Regulation of transcription by long noncoding RNAs. Annu Rev Genet. 48:433–455.

Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D, Capra V, Grompe M, Pizzuti A, Muzny D, Lawrence C, et al. 1991. Characterization of a murine gene expressed from the inactive X chromosome. Nature. 351:325–329.

Bush SJ, Muriuki C, McCulloch MEB, Farquhar IL, Clark EL, Hume DA. 2018. Cross-species inference of long non-coding RNAs greatly expands the ruminant transcriptome. Genet Sel Evol. 50:20.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25:1915–1927.

Cao M, Zhao J, Hu G. 2019. Genome-wide methods for investigating long noncoding RNAs. Biomed Pharmacother. 111:395–401.

Chen M, Li X, Zhang X, Li Y, Zhang J, Liu M, Zhang L, Ding X, Liu X, Guo H. 2019. A novel long non-coding RNA, IncKBTBD10, involved in bovine skeletal muscle myogenesis. In Vitro Cell Dev Biol Anim. 55:25–35.

Chen SY, Liu YP, Wang W, Gao CZ, Yao YG, Lai SJ. 2008. Dissecting the matrilineal components of Tongjiang cattle from southwest China. Biochem Genet. 46:206–215.

Choi JY, Shin D, Lee HJ, Oh JD. 2019. Comparison of long noncoding RNA between muscles and adipose tissues in Hanwoo beef cattle. Anim Cells Syst (Seoul). 23:50–58.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 22:1775–1789.

Fernandes JCR, Acuna SM, Aoki JI, Floeter-Winter LM, Muxel SM. 2019. Long non-coding RNAs in the regulation of gene expression: physiology and disease. ncRNA. 5:17.

Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 4:44–57.

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

Ibeagha-Awemu EM, Do DN, Dudemaine PL, Fomenky BE, Bissonnette N. 2018. Integration of IncRNA and mRNA transcriptome analyses reveals genes and pathways potentially involved in calf Intestinal growth and development during the early weeks of life. Genes (Basel). 9:142.

Ibeagha-Awemu EM, Li R, Dudemaine PL, Do DN, Bissonnette N. 2018. Transcriptome analysis of long non-coding RNA in the bovine mammary gland following dietary supplementation with linseed oil and safflower oil. Int J Mol Sci. 19:3610.

Jathar S, Kumar V, Srivastava J, Tripathi V. 2017. Technological developments in IncRNA biology. Adv Exp Med Biol. 1008:283–323.

Jin L, Hu SL, Tu T, Huang ZQ, Tang QZ, Ma JD, Wang X, Li XW, Zhou X, Shuai SR, et al. 2018. Global long noncoding RNA and mRNA expression changes between prenatal and neonatal lung tissue in pigs. Genes (Basel). 9:443.

Jin CF, Li Y, Ding XB, Li X, Zhang LL, Liu XF, Guo H. 2017. Inc133b, a novel, long non-coding RNA, regulates bovine skeletal muscle satellite cell proliferation and differentiation by mediating miR-133b. Gene. 630:35–43.

Kern C, Wang Y, Chitwood J, Korf I, Delany M, Cheng H, Medrano JF, Van Eenennaam AL, Ernst C, Ross P, et al. 2018. Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species. BMC Genomics. 19:684.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 12:357–360.

Kong L, Zhang Y, Ye QZ, Liu XQ, Zhao SQ, Wei L, Gao G. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 35:W345–W349.

Kopp F, Mendell JT. 2018. Functional classification and experimental dissection of long noncoding RNAs. Cell. 172:393–407.
Koufariotis LT, Chen YP, Chamberlain A, Vander Jagt C, Hayes BJ. 2015. A catalogue of novel bovine long noncoding RNA across 18 tissues. PLoS One. 10:e0141225.

Li D, Li F, Jiang K, Zhang M, Han R, Jiang R, Li Z, Tian Y, Yan F, Kang X, et al. 2019. Integrative analysis of long noncoding RNA and mRNA reveals candidate IncRNAs responsible for meat quality at different physiological stages in Gushi chicken. PLoS One. 14:e0215006.

Li Q, Qiao J, Zhang Z, Shang X, Chu Z, Fu Y, Chu M. 2018. Identification and analysis of differentially expressed long non-coding RNAs of Chinese Holstein cattle responses to heat stress. Anim Biotechnol. 31:9–16.

Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. 2010. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics. 26:493–500.

Ling YH, Xu LN, Zhu L, Sui MH, Zheng Q, Li WY, Liu Y, Fang FG, Zhang XR. 2017. Identification and analysis of differentially expressed long non-coding RNAs between multiparous and uniparous goat (Capra hircus) ovaries. PLoS One. 12:e0183163.

Liu XF, Ding XB, Li X, Jin CF, Yue YW, Li GP, Guo H. 2017. An atlas and analysis of bovine skeletal muscle long noncoding RNAs. Anim Genet. 48:278–286.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.

Marchese FP, Raimondi I, Huarte M. 2017. The multidimensional mechanisms of long non-coding RNA function. Genome Biol. 18:206.

Mattick JS, Makunin IV. 2006. Non-coding RNA. Hum Mol Genet. 15:R17–29.

Meseure D, Alsibai KD, Nicolas A, Bieche I, Morillon A. 2015. Long noncoding RNAs as new architects in cancer epigenetics, prognostic biomarkers, and potential therapeutic targets. Biomed Res Int. 2015:320214.

Muret K, Klopp C, Wucher V, Esquerre D, Legeai F, Lecerf F, Desert C, Boutin M, Jehl F, Acloque H, et al. 2017. Long noncoding RNA repertoire in chicken liver and adipose tissue. Genet Sel Evol. 49:6.

Niknafs YS, Pandian B, Iyer HK, Chinnaiyan AM, Iyer MK. 2017. TACO produces robust multisample transcriptome assemblies from RNA-seq. Nat Methods. 14:68–70.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 33:290–295.

Ponting CP, Oliver PL, Reik W. 2009. Evolution and functions of long noncoding RNAs. Cell. 136:629–641.

RenHX, Wang GF, Chen L, Jiang J, Liu LJ, Li NF, Zhao JH, Sun XY, Zhou P. 2016. Genome-wide analysis of long noncoding RNAs at early stage of skin pigmentation in goats (Capra hircus). BMC Genomics. 17:67.

Schmitz SU, Grote P, Herrmann BG. 2016. Mechanisms of long noncoding RNA function in development and disease. Cell Mol Life Sci. 73:2491–2509.

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

Sulayman A, Tian K, Huang X, Tian Y, Xu X, Fu X, Zhao B, Wu W, Wang D, Yasin A, et al. 2019. Genome-wide identification and characterization of long non-coding RNAs expressed during sheep fetal and postnatal hair follicle development. Sci Rep. 9:8501.

Taylor DH, Chu ET, Spektor R, Soloway PD. 2015. Long noncoding RNA regulation of reproduction and development. Mol Reprod Dev. 82:932–956.

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

Wang GZ, Du K, Hu SQ, Chen SY, Jia XB, Cai MC, Shi Y, Wang J, Lai SJ. 2018. Genome-wide identification and characterization of long non-coding RNAs during postnatal development of rabbit adipose tissue. Lipids Health Dis. 17:271.

Wang X, Yang C, Guo F, Zhang Y, Ju Z, Jiang Q, Zhao X, Liu Y, Zhao H, Wang J, et al. 2019. Integrated analysis of mRNAs and long noncoding RNAs in the semen from Holstein bulls with high and low sperm motility. Sci Rep. 9:2092.

Weikard R, Hadlich F, Hammon HM, Frieten D, Gerbert C, Koch C, Dusel G, Kuehn C. 2018. Long noncoding RNAs are associated with metabolic and cellular processes in the jejunum mucosa of pre-weaning calves in response to different diets. Oncotarget. 9:21052–21069.

Weikard R, Hadlich F, Kuehn C. 2013. Identification of novel transcripts and noncoding RNAs in bovine skin by deep next generation sequencing. BMC Genomics. 14:789.

Wu Z, Gao S, Zhao X, Chen J, Keyvanfar K, Feng X, Kajigaya S, Young NS. 2019. Long noncoding RNAs of single hematopoietic stem and progenitor cells in healthy and dysplastic human bone marrow. Haematologica. 104:894–906.

Xu J, Bai J, Zhang X, Lv Y, Gong Y, Liu L, Zhao H, Yu F, Ping Y, Zhang G, et al. 2017. A comprehensive overview of IncRNA annotation resources. Brief Bioinform. 18:236–249.

Xu X, Ji S, Li W, Yi B, Li H, Zhang H, Ma W. 2017. LncRNA H19 promotes the differentiation of bovine skeletal muscle satellite cells by suppressing Sirt1/FoxO1. Cell Mol Biol Lett. 22:10.

Yang B, Jiao B, Ge W, Zhang X, Wang S, Zhao H, Wang X. 2018. Transcriptome sequencing to detect the potential role of long non-coding RNAs in bovine mammary gland during the dry and lactation period. BMC Genomics. 19:605.

Yang Y, Zhou R, Zhu S, Li X, Li H, Yu H, Li K. 2017. Systematic identification and molecular characteristics of long noncoding RNAs in pig tissues. Biomed Res Int. 2017:6152582.

Yue B, Li H, Liu M, Wu J, Li M, Lei C, Huang B, Chen H. 2019. Characterization of IncRNA-miRNA-miRNA network to reveal potential functional ceRNAs in bovine skeletal muscle. Front Genet. 10:91.

Zhan S, Dong Y, Zhao W, Guo J, Zhong T, Wang L, Li L, Zhang H. 2016. Genome-wide identification and characterization of long non-coding RNAs in developmental skeletal muscle of fetal goat. BMC Genomics. 17:666.

Zhao Y, Chen M, Lian D, Li Y, Li Y, Wang J, Deng S, Yu K, Lian Z. 2019. Non-coding RNA regulates the myogenesis of skeletal muscle satellite cells, injury repair and diseases. Cells. 8:988.

Zhao Y, Li H, Fang S, Kang Y, Wu W, Hao Y, Li Z, Bu D, Sun N, Zhang MQ, et al. 2016. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. Nucleic Acids Res. 44:D203–D208.d