Sequencing and Characterization of Divergent Marbling Levels in the Beef Cattle (*Longissimus dorsi* Muscle) Transcriptome

Dong Chen, Wufeng Li[1], Min Du[2], Meng Wu[3], and Binghai Cao*

College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

ABSTRACT: Marbling is an important trait regarding the quality of beef. Analysis of beef cattle transcriptome and its expression profile data are essential to extend the genetic information resources and would support further studies on beef cattle. RNA sequencing was performed in beef cattle using the Illumina High-Seq2000 platform. Approximately 251.58 million clean reads were generated from a high marbling (H) group and low marbling (L) group. Approximately 80.12% of the 19,994 bovine genes (protein coding) were detected in all samples, and 749 genes exhibited differential expression between the H and L groups based on fold change (>1.5-fold, p<0.05). Multiple gene ontology terms and biological pathways were found significantly enriched among the differentially expressed genes. The transcriptome data will facilitate future functional studies on marbling formation in beef cattle and may be applied to improve breeding programs for cattle and closely related mammals. (**Key Words:** Marbling, Beef Cattle, Transcriptome, RNA-Sequencing)

INTRODUCTION

Snow dragon beef is a high-quality beef cattle breed (crossbreed of Fuzhou Yellow cattle×Limousine×Wagyu [FLW]) with a long history of feeding and breeding aimed at obtaining abundant marbling. The quantity of marbling in a cross-section of *longissimus dorsi* (LD) muscle is considered an important trait which is associated with meat texture, tenderness, flavor, and juiciness (Hausman et al., 2009). Marbling is the primary and essential factor assessed in the quality grading of beef carcasses in the United States (USDA, 1989) and Japan (JMGA, 1988). To effectively enhance intramuscular fat deposition, it is necessary to characterize the transcriptomes associated with the divergent marbling phenotypes.

The marbling is one of the economic importance in beef. There are so many selective breedling existing which have been emphasized for many years, and carried out in large population for thousands of yeas. However, the trait of marbling is difficult to improve by traditional selection because the heritability is low and measure for marbling is only possible after slaughter. Moreover, marbling is determined polygenically. Genetics plays a key role in mediating the composition of bovine muscle tissues (Grobet et al., 1997). However, the complexity of the bovine transcriptome has not yet been fully elucidated. Novel high-throughput and deep-sequencing technologies are currently being used in genomic research, and provide new strategies for analyzing the functional complexities of transcriptomes or of the transcribed DNA of an organism, thereby generating functional genomic data. Annotation is easier for transcribed genes than for complete genomes because new sequences can be compared to conserved protein sequences and transcribed genes contain fewer repetitive elements. The holistic view of the transcriptome and its organization provided by the RNA sequencing (RNA-Seq) method have revealed many novel transcribed regions, splice isoforms and coding single nucleotide polymorphisms (Cloonan and

* Corresponding Author: B. H. Cao. Tel: +86-01062814346, Fax: +86-01062814346, E-mail: caobhchina@163.com
[1] College of Life Science, Shanxi Agriculture University, Taigu, Shanxi 030800, China.
[2] Department of Animal Sciences, Washington State University, Pullman, WA 641227, USA.
[3] Snowdragon Beef Co., Ltd. Dalian 116000, China.

Grimmond, 2008; Nagalakshmi et al., 2008). Finally, RNA-Seq generates absolute, rather than relative, gene expression measurements, thereby providing greater insight and accuracy compared with microarrays ('t Hoen et al., 2008; Marioni et al., 2008).

This study describes the first global analysis of the beef cattle transcriptome focusing on marbling formation, which was conducted using the Illumina RNA-Seq method. The objectives of this study were to associate the divergent LD muscle phenotypes with characteristic gene expression patterns and to verify the completeness and utility of the assembled transcriptome for use with several types of genome-scale analyses. In addition, the other aims of this research were to validate the RNA-Seq technology and to set up a pipeline allowing for the identification of gene expression levels between divergent marbling levels in beef cattle. For these purpose, we report a comprehensive analysis of transcriptome dynamics that aids in the elucidation of the gene expression patterns that occur during the development of marbling.

MATERIALS AND METHODS

Animal harvesting and sampling

Beef cattle (FLW) were humanely harvested at Snowdragon Beef Co. Ltd., (Dalian, Liaoning, China). Live animal performance and carcass traits (69 animals) were collected (Supplementary Table 1a). The LD muscle samples were collected from sixty-nine animals immediately after exsanguinations. The samples for examining LD muscle marbling were obtained from the sixth rib. A portion of the sample was snap frozen in liquid nitrogen and used for RNA extraction, and another portion (69 animals) was used to determine the water, crude fat contents and fatty acid composition (Supplementary Table 1b). These FLW crossbreed animals were all of the same age and genetic background, and they were housed under the same conditions by Snowdragon Beef Co. Ltd. (a reputable company for the production of high marbling beef). Fat color, meat color and marbling were scored using a scale (1 = low, to 6 = high) according to the beef grading standard (Supplementary Figures 1A-C). The marbling grade was evaluated between the sixth and seventh ribs based on the official Japanese Meat Grading Standards which have been adopted by the Chinese beef cattle industry along with other meat grading standards. Based on the obtained marbling scores, two meat samples from the lowest marbling grade (labels Y1-1 and Y1-2) and two meat samples from the highest marbling grade (labels Y2-1 and Y2-2) were selected (Supplementary Figure 1D).

RNA extraction and transcriptome sequencing using

RNA sequencing technology

Muscle samples (200 g) were homogenized in liquid nitrogen, and 10 mg of each homogenized sample was used for RNA-Seq. Total RNA was isolated from these four samples using TRIzol following manufacturer's protocol (Invitrogen, Carlsbad, CA, USA). The quality of the total RNA was assessed using the RNA Nano 6000 Assay Kit with the Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). High-quality RNA (RNA integrity number >7.2) was processed using the Illumina TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA, USA) following the manufacturer's procedures. The final individual RNA-Seq libraries were constructed using the Illumina Hiseq2000 platform, which created 100 bp/paired-end (PE) sequencing reads that were validated and pooled based on their respective adapters.

Sequence alignment

The raw reads were subjected to qulity control measures, including the removals of the sequencing adapters and of all reads containing over 10% Ns, resulting in clean reads. For the data analysis, the read pairs were trimmed using Solexa QA (Massey University, Palmerston North, New Zealand, v2.0) (Cox et al., 2010), reads longer than 25 bp were retained whereas thoes in which one or a few bases presented with Q-scores of lower than 20 and thoes that were unpaired were deleted. The qualified reads were then aligned to the bovine genome (University of Maryland [UMD] Bos taurus 3.1.73 [UMD 3.1.73] produced by the University of Maryland) from the Ensembl database (http://ensembl.org/) using Tophat (v2.0.9) (Kim et al., 2013). Bowtie2 (Langmead and Salzberg, 2012) was employed to index the genome, and the maximum number of multiple hits was set to 20. The tolerance was adjusted to allow for no more than two mismatches in each alignment and zero mismatches at splice sites. Normalized gene expression was calculated as the number of fragments per kilobase of the model exon per million mapped reads using Cufflinks (Trapnell et al., 2010).

Differential gene expression and cluster analysis

The differential expression analysis was performed using the Cuffdiff program from the Cufflinks package (v 2.1.1), which tracks changes in expression with a high level of reliability (Trapnell et al., 2012). Candidate differentially expressed genes (DEGs) were identified using Fisher's exact test ($p<0.05$) and the observed fold changes (>1.5-fold) to account for multiple testing. The Cuffdiff-filtered candidate genes were clustered using a Cluster (3.0) hierarchical algorithm according to the genes and samples to assess the replication of the samples (de Hoon et al., 2004).

Gene ontology analysis

We performed a gene ontology (GO) analysis on the DEGs, which were classified into the categories of cellular component, molecular function and biological process sub-ontologies using the online server DAVID (v6.7, http://david.abcc.ncifcrf.gov/, [Huang et al., 2009]). The gene set enrichment of certain GO terms was determined based on Fisher's exact test (p<0.1).

Pathway analysis

Kyoto Encyclopedia of Genes and Genomes is the major public pathway-related database. We further performed a pathway analysis on the DEGs identified in the transcriptome using the online server DAVID (v6.7, http://david.abcc.ncifcrf.gov/, [Huang et al., 2009]). The pathway enrichment analysis identified significantly enriched metabolic pathways and signal transduction pathways among the DEGs through comparisons with the whole-genome background. The calculated p-values were corrected using Fisher's exact test (p<0.1). As a complementary approach, the candidate genes were analyzed using the Ingenuity Pathway Analysis (IPA; Ingenuity Systems, http://www.ingenuity.com).

Quantitative polymerase chain reaction and data analysis

To confirm the detected changes in transcript levels between the H and L groups (n = 3), we quantified the expression level of 8 genes (identified via RNA-Seq) selected from the DEGs: peroxisome proliferator-activated receptor gamma, coactivator 1 beta (*PPARGC1B*), secreted phosphoprotein 2 (*SPP2*), stearoyl-CoA desaturase 5 (*SCD5*), myogenic differentiation (*MyoD*), chloride channel 4 (*CLCN4*), collagen, type I, alpha 1 (*COL1A1*), myogenin (*MyoG*), and stearoyl-CoA desaturase (*SCD*). Total RNA was extracted from Y1-1, Y1-2, Y2-1 and Y2-2 as well as from two additional samples (one from the H group and the other from the L group), which was converted to cDNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific Inc, Waltham, MA, USA) following the manufacturer's protocol. Using the Stratagene real-time quantitative polymerase chain reaction (QPCR) System, quantitative real-time PCR (RT-PCR) was performed with the SYBR Green Realtime PCR Master Mix (TOYOBO Co. Ltd., Life Science Department, Osaka, Japan). The PCR conditions were: 95°C for 5 min, followed by 40 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 50 s. The primer pairs used for the amplification reactions are shown in Table S9. All PCR efficiencies were above 98%. Relative gene expression data were obtained using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). After amplification, a melting curve (0.01°C/s) was employed to confirm product purity, and agarose gel electrophoresis was performed to confirm that only a single product of the correct size was amplified. The relative mRNA expression was normalized to that of a house keeping gene (glyceraldehyde-3-phosphate dehydrogenase).

RESULTS

High-throughput sequencing of the bovine *longissimus dorsi* muscle transcriptome

In the present study, the transcriptomic patterns of the LD muscle samples obtained from the high marbling (H) versus low marbling (L) adult cattle were compared using RNA-Seq. To this end, two rounds of linear mRNA amplification were performed to ensure that each sample produced a sufficient amount of input RNA for the analysis. Amplified RNA samples from two H and two L adult cattle produced by the same sire were sequenced with the HiSeq2000 system, resulting in approximately 13 Gb of PE raw reads averaging 100 bp in length. Table presents the overall results of the alignments of the sequencing reads to the bovine reference genome (UMD 3.1.73) and genes.

Approximately 90% of the initial reads were retained for further analysis after the filtering of low-quality reads, resulting in 56.61 million, 49.79 million, 52.60 million, and 67.04 million clean reads for the Y1-1, Y1-2, Y2-1, and Y2-2 samples, respectively. Of the total sequenced reads, more than 98% mapped to the UMD 3.1.73 reference genome for each sample. More than 87% of these clean reads uniquely mapped to specific regions of the bovine genome. Unmapped reads and reads mapped to multiple positions were excluded from further analyses (Table 1). Among the 19,994 bovine genes (protein coding) in the bovine reference genome, 16,020 were expressed in the bovine muscle samples, representing the core components of the bovine muscle transcriptome. The mean numbers of genes transcribed in the muscle transcriptomes of the L and H animals were 14,904±226.3 and 14,837±44.5 (mean± standard deviation), respectively. These transcribed genes represented 74.5% and 74.2% of the total bovine genes for the L and H animals, respectively. Strong correlations between biological replicates were also observed ($R^2 \geq 0.92$; Figure 1A) between individuals showing the same marbling grade. There were 713 genes that were only expressed in the L group, and there were 509 genes only expressed in the H group (Figure 1B).

Differentially expressed genes between the high marbling and low marbling groups

The RNA-Seq technique allows for the analysis of differential expression profiles based on transcript abundance with a high sensitivity for the identification of transcripts expressed at low levels that would otherwise be undetectable using standard microarrays (Werner, 2010). Of

Table 1. Summary of bovine genome[1] read mapping

	Low group		High group	
	Y1-1	Y1-2	Y2-1	Y2-2
Clean reads (PEs[2])/No.	63,149,080	55,541,818	58,323,610	74,561,988
Total reads after filtering (PEs)/No.	56,610,052	49,789,584	52,601,232	67,036,982
Left reads/No.	28,305,026	24,894,792	26,300,616	33,518,491
Right reads/No.	28,305,026	24,894,792	26,300,616	33,518,491
Mean read length/bp	80.67	80.35	80.98	81.05
Max length/bp	100	100	100	100
Min length/bp	25	25	25	25
Mean coverage	20.05	18.09	20.08	23.6
Standard deviation	±171.15	±199.51	±178.57	±245.34
Left mapped reads/No.	27,057,575	23,789,899	25,309,643	3,240,4620
Right mapped reads/No.	27,042,388	23,769,870	25,294,498	3,239,7139
Aligned PEs/No.	26,307,847	23,144,100	24,742,373	31,789,070
Aligned PEs with multiple alignments/No.	1,675,282	1,685,118	1,727,220	2,118,099
Aligned PEs with discordant alignments/No.	1,503,508	1,373,694	1,359,454	1,721,036
Concordant PEs alignment rate/%	87.63	87.45	88.91	89.71
PEs with matched to bovine genome sequences	27,792,116	24,415,669	25,861,768	33,012,689
PEs matched to bovine genes	24,113,839	21,789,156	22,572,732	28,965,608
PEs in intergenic regions	3,678,277	2,626,513	3,289,036	4,047,081
PEs in exon regions	22,821,023	20,869,496	21,219,845	27,427,089
PEs in intronic regions	1,292,816	919,660	1,352,887	1,538,519
FPKM[3] = 0/No.	9,184	9,591	9,493	9,392
FPKM>0/No.	15,432	1,5025	15,123	15,224
Protein_coding/No.	15,064	14,744	14,806	14,869
Bovine genes with mapped reads/%	62.69	61.04	61.44	61.85
Unchecked genes/%	37.31	38.96	38.56	38.15

FPKM, fragments Per Kilobase of exon model per Million mapped reads.

[1] Genome = UMD3.1.73.

[2] For paired-end data, 1 pair of sequencing results was counted as 1 read. This pair was only counted as a uniquely mapped read if both ends were uniquely mapped.

[3] FPKM = total exon fragments/mapped reads(millions)×exon length (kb).

the 16,020 total analyzed genes, 749 (366 up-regulated in the H group and 383 upregulated in the L group) showed fold changes of greater than 1.5 (p<0.05) between the two groups (Supplementary Table 2). A cluster analysis was performed to identify the DEGs presenting with similar temporal changes between the two groups (Figure 1C).

Gene ontology analysis

To further investigate the biological processes associated with the DEGs, we performed a GO analysis by running queries for each DEG against the GO database (Huang et al., 2009), which provided information on relevant molecular functions, cellular components and biological processes. Summaries of the classifications of the DEGs into the three GO sub-ontologies (top 20) are presented in Figure 2 and Table S3. In total, 271 GO terms were assigned to the DEGs, which were further classified into the three independent sub-ontologies as follows: i) 195 (72.0%) terms corresponding to biological processes, ii) 33 (12.2%) terms corresponding to molecular functions and iii)

43 (15.9%) terms corresponding to cellular components. Several terms related to immune system processes, development and extracellular matrix (ECM) signaling figured prominently. Interestingly, the DEGs showing higher expressions in the L group were associated with immune system processes. In contrast, those DEGs presenting with higher expressions in the H group were associated with biosynthetic processes and RNA processing, for example, the positive regulation of cellular biosynthetic processes (GO:0031328), the regulation of differentially expressed RNA metabolic processes (GO:005125) and the regulation of transcription from the RNA polymerase II promoter (GO:0006357). Notably, 25 specific GO terms related to lipid metabolism (noted in Table S6) were present in the middle of the list of all DEGs between the L and H groups, indicating that a substantial fraction of the genes identified were potentially associated with the traits under study.

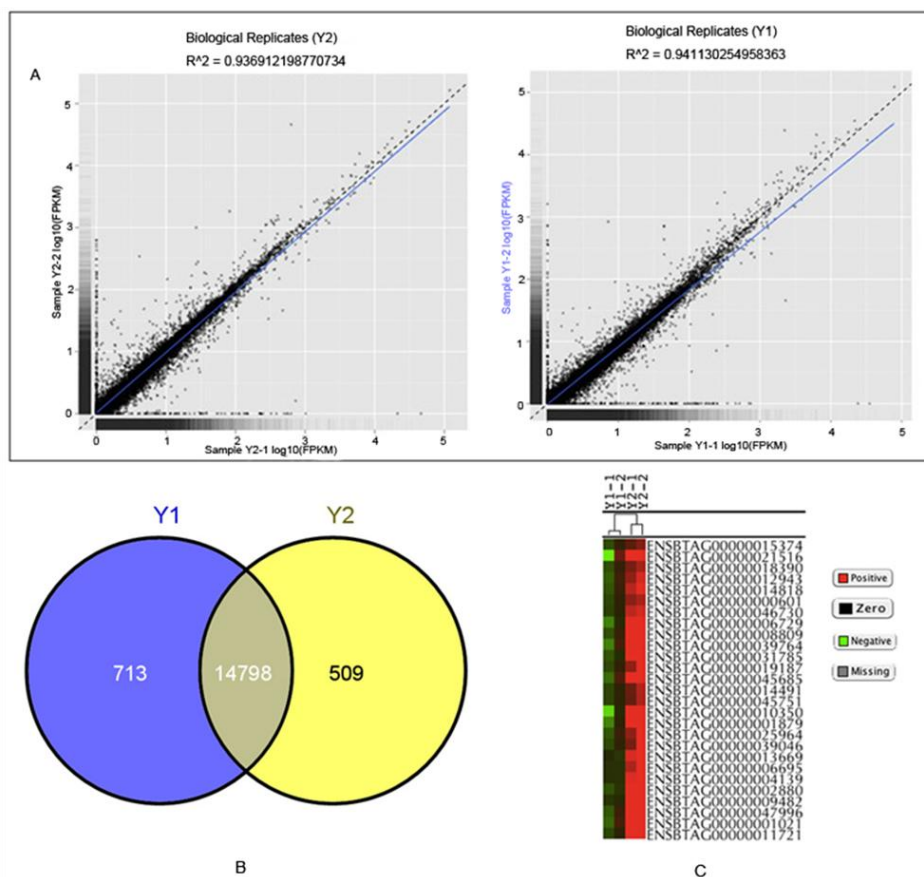**Metabolic pathways and gene networks affected by the**

Figure 1. Bioinformation for the 4 samples. (A) Correlations of 4 samples; (B) venn diagram showing the shared and unique genes among the high (H) and low (L) groups. The two groups of *longissimus dorsi* tissues (H and L) share over 14,000 genes. Total unique transcripts = 1,222; Y1 = 713; Y2 = 509; (C) clustering of the list of differentially expressed genes between the L and H groups.

differentially expressed genes

Different genes work in association with each other to exert their biological functions. To gain insights into the metabolic pathways and gene transactivation networks that contribute to the development of marbling in beef cattle as well as to obtain more information on the differences in expression patterns between the groups, the results from two exploration tools (IPA software and DAVID) were compared.

First, the IPA software was used to assess the pathways represented in our gene list (749 genes). A total of 9 canonical metabolic pathways and 8 signaling pathways were significantly represented in Supplementary Table 4. The complement system pathway was the most prominent with 8 members being differentially expressed (7/8 of these genes were downregulated; complement component 2 (*C2*), complement factor B (*CFB*), complement component 1, s subcomponent (*C1S*), complement component 1, q subcomponent, C chain (*C1QC*), complement component 1, q subcomponent, A chain (*C1QA*), complement component receptor 2 (*CR2*), and complement component 1, q subcomponent, B chain (*C1QB*). In addition, integrin-linked kinase (ILK) signaling (19 genes), the NRF2 (NF-E2-

related factor 2) mediated oxidative stress response (5 downregulated and 8 upregulated genes) and PKCθ (Protein Kinase C-θ) signaling in T lymphocytes (12 genes) were also present. The detected networks identified through IPA are presented in Supplementary Table 4. Several regulatory networks involved in protein synthesis, lipid metabolism, molecular transport, muscle growth and function-related pathways were amongst the most relevant networks represented in the list of DEGs. Genes related to cellular growth, proliferation and DNA replication functions were also included. Major genes networks centered on the apolipoprotein A-I (*APOA1*), abhydrolase domain containing 5 (*ABHD5*), actin, alpha, cardiac muscle 1 (*ACTC1*), sorbin and SH3 domain containing 1 (*SORBS1*), and carboxypeptidase E (*CPE*) genes.

Second, the results of the DAVID pathway analysis between the two groups are presented as 3 lists (749 overall differentially expressed, 366 upregulated in the H group, and 383 upregulated in the L group) in Supplementary Table 4. Twenty-three pathways were significantly represented among all of the DEGs. This analysis revealed three top-ranking pathways, which are listed in their respective order: complement and coagulation cascades (14
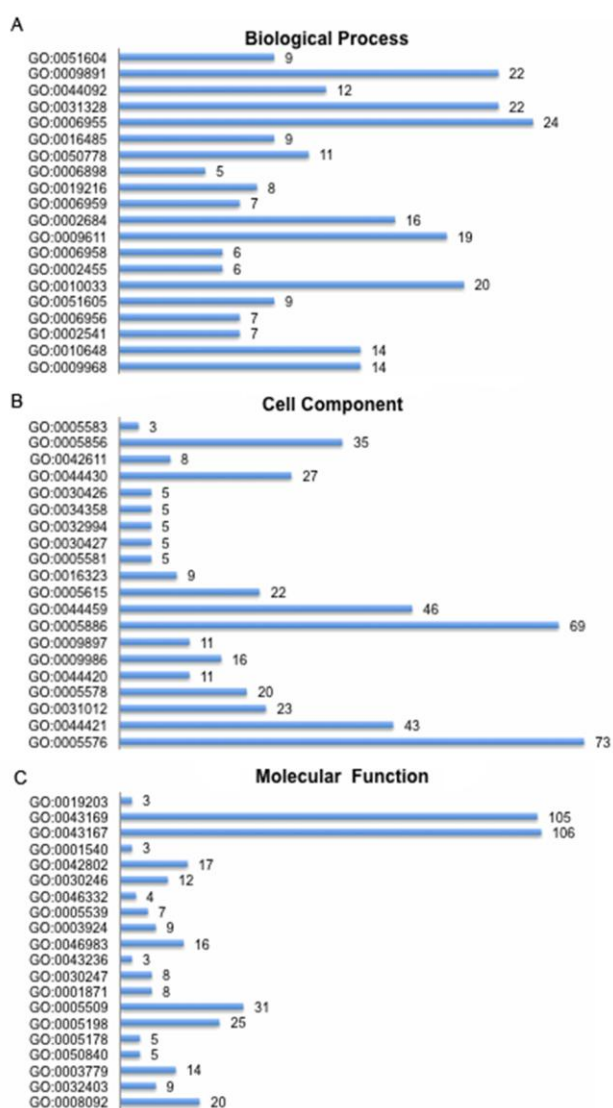
A Biological Process

GO term	value
GO:0051604	9
GO:0009891	22
GO:0044092	12
GO:0031328	22
GO:0006955	24
GO:0016485	9
GO:0050778	11
GO:0006898	5
GO:0019216	8
GO:0006959	7
GO:0002684	16
GO:0009611	19
GO:0006958	6
GO:0002455	6
GO:0010033	20
GO:0051605	9
GO:0006956	7
GO:0002541	7
GO:0010648	14
GO:0009968	14

B Cell Component

GO term	value
GO:0005583	3
GO:0005856	35
GO:0042611	8
GO:0044430	27
GO:0030426	5
GO:0034358	5
GO:0032994	5
GO:0030427	5
GO:0005581	5
GO:0016323	9
GO:0005615	22
GO:0044459	46
GO:0005886	69
GO:0009897	11
GO:0009986	16
GO:0044420	11
GO:0005578	20
GO:0031012	23
GO:0044421	43
GO:0005576	73

C Molecular Function

GO term	value
GO:0019203	3
GO:0043169	105
GO:0043167	106
GO:0001540	3
GO:0042802	17
GO:0030246	12
GO:0046332	4
GO:0005539	7
GO:0003924	9
GO:0046983	16
GO:0043236	3
GO:0030247	8
GO:0001871	8
GO:0005509	31
GO:0005198	25
GO:0005178	5
GO:0050840	5
GO:0003779	14
GO:0032403	9
GO:0008092	20

Figure 2. Top 20 functional gene ontology annotations for the differentially expressed genes (DEGs). Bar graphs (A), (B), and (C) show the three independent gene ontology (GO) information categories of biological processes, molecular functions and cellular components, respectively. In each bar graph, the abscissa represents the number of DEGs, and the ordinate shows the ID numbers of the GO terms. All GO categories listed exhibit enrichment with $p<0.05$.

downregulated genes in the H group), systemic lupus erythematosus (14/16 downregulated genes in the H group) and ECM-receptor interactions (7 downregulated genes, including collagen, type IV, alpha 6 (*COL4A6*), tenascin C (*TNC*), secreted phosphoprotein 1 (*SPP1*), collagen, type I, alpha 2 (*COL1A2*), *COL1A1*, CD44 molecule (*CD44*), and ribosomal protein L26 (*RPL26*), and 5 upregulated genes, including integrin, beta 6 (*ITGB6*), syndecan 4 (*SDC4*), agrin (*AGRN*), thrombospondin 4 (*THBS4*), *BT.3034*. Similar to the IPA results, the highest impact factor was found for the complement system cascades followed by 14-

3-3-mediated signaling, these findings were in contrast with the DAVID pathway results. Among these findings were in contrast with the DAVID pathway results, which revealed several pathways that were similar to those identified using the IPA tool but with a different ranking of relevance. Interestingly, certain significant pathways (the complement and coagulation cascades, ECM-receptor interactions and the mitogen-activated protein kinase (MAPK) signaling pathway) were very relevant to the lipid metabolism or muscle and/or adipose tissue differentiation. In addition, 5/7 PPAR signaling genes (solute carrier family 27 (fatty acid transporter), member 6 (*SLC27A6*), carnitine palmitoyltransferase 2 (*CPT2*), *SORBS1*, phosphoenolpyruvate carboxykinase 1 (*PCK1*) and *SLC27A1*) and 4 adipocytokine signaling pathway genes (*OB*, *PRKAG2*, *PCK1*, and *PPARGC1A*) were upregulated in the H group, which are significant to the development of marbling.

Gene expression confirmation via quantitative polymerase chain reaction

The differential expressions of 8 genes were confirmed through RT-PCR. Three genes with relevant functions in lipid metabolism (*SCD5*, *SCD*, and *PPARGC1B*), 2 with a relevant functions in myogenesis (*MYOD1* and *MYOG*), one related to fibrogenesis (*COL1A1*), one showing high expression levels in Y2 (*CLCN4*) and one showing high expression in Y1 (*SPP2*), were subjected to quantification by RT-PCR. The fold-change ratios between the H and L groups were consistent with the RNA-Seq data for all 8 genes (Figure 3).

DISCUSSION

Similar to the microarray method, RNA-Seq permits the identification and classification of genes such as those associated with muscle growth (Reecy et al., 2006) and adipogenesis [Wang et al., 2005; Wang et al., 2009]) from transcript data. RNA-Seq is a useful approach to transcriptome analyses, that is quickly superseding microarray-based methods for the measurements of gene expression levels. RNA-Seq technologies enabled the sequencing and annotation of a reference transcriptome, which was used to perform comprehensive gene expression analyses associated with the development of marbling in beef cattle in the current study. However, the new level of detail offered by RNA-Seq raises novel statistical and computational challenges, which limit the widespread adoption of whole-genome transcriptional profiling for the detection of genes that are differentially expressed among different experiments.

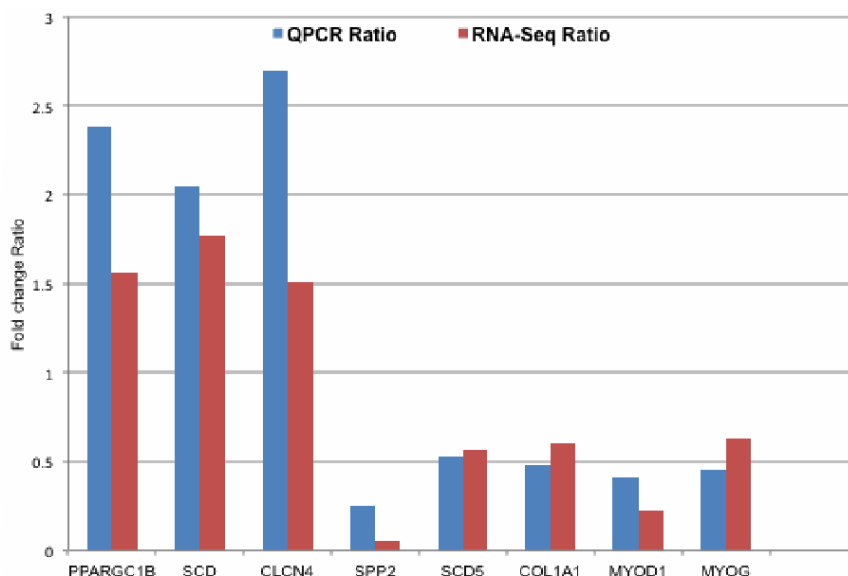In the present study, we generated 13 Gb of raw PE sequence reads averaging 100 bp in length by the Illumina

Figure 3. Comparisons of the RNA sequencing and quantitative polymerase chain reaction expression ratios (low and high groups) for the selected genes.

sequencing of mRNA obtained from two groups of cattle that were phenotypically divergent with regard to their LD muscle fat content marbling. The alignments of all of the sequencing reads to the bovine UMD 3.1.73 reference genome yielded >200 million reads that were successfully mapped. Only approximately 3.9 million reads (approximately 1.7%) did not align to the UMD 3.1.73 reference genome. The high sequence coverage (>20) of the reference genome in addition to the low levels of reference and sequencing errors that were detected verifies the accuracy of our assembly. By comparing our transcriptome assembly with sequences from other species with well-characterized genomes, we obtained identities approximately 80.12% of the entire transcriptome, corresponding to 16,020 unique genes, thereby illustrating the sensitivity of RNA-Seq for transcript discovery, even for genes expressed at low levels (Wang et al., 2009). We also studied gene expression differences between two groups showing different marbling levels (1, LD muscle with high marbling; 2, LD muscle with low marbling). ultimately, 749 differentially expressed reference sequences were selected as putative candidates that play roles in marbling formation in LD muscle. Although this study was small, our results revealed significant differences in the expression of a wide variety of genes, which may be assessed in future studies as potential members of signaling pathways associated with marbling in beef cattle.

With the rapid increase in the use of high-throughput sequencing, sufficient transcriptomic data sets are now available, particularly for mammals. A comparison of the transcriptomes of different beef cattle breeds (FLW crossbreed cattle were analyzed in the present study, and Qinchuan beef cattle (He and Liu, 2013) and Holstein cows (Driver et al., 2012) have been analyzed in a similar study) revealed similarities in the GO distributions of the genes within each sub-ontology between these breeds. For the FLW breed, immune- and lipid-related processes were among the most highly represented GO terms in the biological process sub-ontology. However, for the Qinchuan and Holstein breeds, developmental- and structure- related processes were the most prominently represented terms in the biological process sub-ontology, such as cell adhesion and biological adhesion. Although a few differences existed within each category, the general organization and the main terms identified in our study were similar to these reported by the other studies (Driver et al., 2012; He and Liu, 2013). Pathway-based analyses allow for the better understanding of the biological functions of genes and identify significantly enriched metabolic pathways and signal transduction pathways associated with DEGs. In our study, we also detected some lipid metabolism- related pathways and networks, such as the PPAR signaling and adipocytokine signaling pathways, in addition to commonly enriched pathways. The results obtained from our pathway analysis were similar to those reported by a previous study involving the feed of propionate to developing Angus cattle, in which the lipid metabolism, small-molecule biochemistry, carbohydrate metabolism and molecular transport pathway were also observed (Baldwin et al., 2012). These results indicate that similar species are likely to exhibit the enrichments of varying metabolic process and biological process GO categories that depent upon physiological, environmental and time differences.

The next-generation sequencing approach used in the

present study allowed for the assembly and annotation of a reliable, comprehensive reference transcriptome. The transcriptome sequencing was performed to evaluate the divengent marbling phenotypes of FLW crossbreed cattle using the RNA-Seq method. This study revealed a large number of genes with known and unknown functions, thus greatly expanding a gene expression profile assoicated with the development of marbling in these cattle. This study also represents a first step toward an improved understanding of the functions of these genes and provides comprehensive and novel insights to facilitate for future research for beef quality improvement in beef cattle.

ACKNOWLEDGMENTS

REFERENCES

Baldwin, R. L. V., R. W. Li, C. Li, J. M. Thomson, and B. J. Bequette. 2012. Characterization of the longissimus lumborum transcriptome response to adding propionate to the diet of growing Angus beef steers. Physiol. Genomics 44:543-550.

Cloonan, N. and S. M. Grimmond. 2008. Transcriptome content and dynamics at single-nucleotide resolution. Genome Biol. 9: 234.

Cox, M. P., D. A. Peterson, and P. J. Biggs. 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11:485.

de Hoon, M. J. L., S. Imoto, J. Nolan, and S. Miyano. 2004. Open source clustering software. Bioinformatics 20:1453-1454.

Driver, A. M., F. Peñagaricano, W. Huang, K. R. Ahmad, K. S. Hackbart, M. C. Wiltbank, and H. Khatib. 2012. RNA-Seq analysis uncovers transcriptomic variations between morphologically similar *in vivo-* and *in vitro*-derived bovine blastocysts. BMC Genomics 13:118.

Grobet, L., L. J. R. Martin, D. Poncelet, D. Pirottin, B. Brouwers, J. Riquet, A. Schoeberlein, S. Dunner, F. Menissier, J. Massabanda, R. Fries, R. Hanset, and M. Georges. 1997. A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. Nat. Genet. 17:71-74.

Hausman, G. J., M. V. Dodson, K. Ajuwon, M. Azain, K. M. Barnes, L. L. Guan, Z. Jiang, S. P. Poulos, R. D. Sainz, S. Smith et al. 2009. Board-invited review: the biology and regulation of preadipocytes and adipocytes in meat animals. J. Anim. Sci. 87:1218-1246.

He, H. and X. Liu. 2013. Characterization of transcriptional complexity during longissimus muscle development in bovines using high-throughput sequencing. PloS one 8(6):e64356.

Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4:44-57.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14:R36.

Langmead, B. and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357-359.

Livak, K. J. and T. D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25:402-408.

Marioni, J. C., C. E. Mason, S. M. Mane, M. Stephens, and Y. Gilad. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 18:1509-1517.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344-1349.

Reecy, J., D. M. Spurlock, and C. H. Stahl. 2006. Gene expression profiling: insights into skeletal muscle growth and development. J. Anim. Sci. 84:E150-E154.

't Hoen, P. A. C., Y. Ariyurek, H. H. Thygesen, E. Vreugdenhil, R. H. A. M. Vossen, R. X. de Menezes, J. M. Boer, G-J. B. van Ommen, and J. T. den Dunnen. 2008. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucl. Acids Res. 36:e141.

Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7: 562-578.

Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, and L. Pachter. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28:511-515.

Wang, Y. H., N. I. Bower, A. Reverter, S. H. Tan, N. De Jager, R. Wang, S. M. McWilliam, L. M. Cafe, P. L. Greenwood, and S. A. Lehnert. 2009. Gene expression patterns during intramuscular fat development in cattle. J. Anim. Sci. 87:119-130.

Wang, Y. H., K. A. Byrne, A. Reverter, G. S. Harper, M. Taniguchi, S. M. McWilliam, H. Mannen, K. Oyama, and S. A. Lehnert. 2005. Transcriptional profiling of skeletal muscle tissue from two breeds of cattle. Mamm. Genome 16:201-210.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: A revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57-63.

Werner, T. 2010. Next generation sequencing in functional genomics. Brief. Bioinform. 11:499-511.