Genetics and Molecular Research

Comparative analysis of the liver tissue transcriptomes of Mongolian and Lanzhou fat-tailed sheep

X. Cheng, S.G. Zhao, Y. Yue, Z. Liu, H.W. Li and J.P. Wu

Lab of Animal Genetic, Faculty of Animal Science and Technology, Gansu Agricultural University, Lanzhou, Gansu, China

Corresponding author: J.P. Wu
E-mail: wujp@gsau.edu.cn

Genet. Mol. Res. 15 (2): gmr.15028572
Received February 19, 2016
Accepted April 15, 2016
Published May 20, 2016
DOI http://dx.doi.org/10.4238/gmr.15028572

ABSTRACT. Research on gene regulation has been made possible with the help of RNA sequencing applications such as RNA-Seq technology for high-throughput sequencing platforms. Recent studies have explored the transcriptomes from different tissues of domestic animals using RNA-Seq technology, but little research has been done to study the transcriptomes of breeds of sheep having different adipose tissue deposition mechanisms, such as Mongolian and Lanzhou fat-tailed sheep. In this study, Mongolian and Lanzhou fat-tailed sheep were selected as experimental breeds, and six libraries (three libraries per breed) were constructed. A total of 286 Mb of high-quality reads was obtained, and three-quarters of the reads were mapped to the reference genome per library. In addition, there were 16,257, 16,186, 16,254, 16,827, 16,437, and 15,761 known reference genes in the six constructed libraries (LCL1, LCL2, LCL3, MCL1, MCL2, and MCL3, respectively). Seven genes were differentially expressed: four were upregulated and three were downregulated in liver tissue between the MCL and LCL groups; 65,303, 65,442, 63,426, 76,267, 69,853, and 57,439 potential cSNPs were detected in the six libraries, respectively,

with the G/R substitution occurring most commonly. There were 24,239, 22,283, 22,457, 26,635, 27,093, and 18,700 alternate splicing (AS) events in the six libraries. Intron retention was the most common AS event, followed by alternative 3' splice sites. These results indicate that there are many differences in the liver transcriptomes of Mongolian and Lanzhou fat-tailed sheep breeds. Such results may provide fundamental information for further research on defining the sheep genome.

Key words: Sheep; Liver tissue; RNA-Seq; Transcriptome

INTRODUCTION

Sheep are economically important livestock species that provide meat, milk, fat, and wool for human consumption. Sheep enterprises often serve as important sources of financial elasticity for herders and farmers and also play important social and cultural roles (Atti et al., 2004). Under the effects of long-term natural selection, sheep have evolved into different ecotypes, such as fat-tailed and small-tailed sheep, according to their biological classification. The fat-tailed characteristic is regarded as an adaptive mechanism to the effects of a challenging nutritional environment. The "fat-tail" can provide energy during migration and in seasons when the pasture is dormant or low amounts of dry matter are available (Atti et al., 2004). In addition, the percentage of intramuscular fat in fat-tailed sheep is low; thus, fat-tailed sheep can provide a low-fat protein source. In contrast, short-tailed sheep have higher stores of intramuscular fat (Kashan et al., 2005). Therefore, understanding the mechanism of fat metabolism and deposition would be useful in improving the quality of meat and for human health.

Mongolian and Lanzhou fat-tailed sheep are two breeds native to the cooler and drier regions of western China. Lanzhou fat-tailed sheep have a pure white coat, medium-sized head, and rectangular body. The fat-tail is hypertrophic and droops toward the hock (Ding et al., 1986). Mongolian sheep are found primarily on the Mongolian plateau. This breed is adapted to areas that are extensively managed and has a short, fat tail. Its tail is generally longer than wide, and the tail is curled up in an S-shape.

RNA-Seq provides comprehensive data, which are not only useful for understanding gene structure but also provide a better understanding of the biological function of genes. In addition, data from RNA-Seq can help identify new transcripts (Hansey et al., 2012), single nucleotide polymorphisms (Esteve-Codina et al., 2011), alternative splicing events (Xiong et al., 2012), and other information. Several recent studies have used this type of next-generation sequencing platform in the liver tissue of pigs (Gunawan et al., 2013a,b; Sodhi et al., 2014), cows (Graugnard et al., 2013), rats (Tran and Huang, 2014; Wang et al., 2014), and fish (Yang et al., 2012; Lau et al., 2014; Magnanou et al., 2014).

More recently, research on sheep transcriptomes that have used next-generation sequencing platforms is very limited, and no global transcriptome analysis of sheep livers using RNA-Seq has been published. Thus, for a better understanding of fat metabolism and deposition in sheep, we analyzed liver transcriptome data from Mongolian and Lanzhou fat-tailed sheep. We characterized the sheep liver transcriptome and analyzed differentially expressed genes (DEGs), single nucleotide polymorphisms in the coding regions (cSNPs), and other information using bioinformatic analyses.

MATERIAL AND METHODS

Animals and samples

A total of nine male Mongolian and nine male Lanzhou fat-tailed sheep (2 months of age) were used for this experiment. Lambs were grouped into three replicates within each breed. All lambs were healthy and development was consistent with age. Management and feed rations were the same for the two groups. Sheep were slaughtered after 13 months following the "Guidelines on Ethical Treatment of Experimental Animals" (2006) No. 398, which was formulated by the Ministry of Science and Technology, China. Liver tissues were collected and frozen in liquid nitrogen within 10 min of slaughter. Subsequently, liver tissues were stored at -80°C until RNA isolation.

RNA isolation, library construction, and Illumina sequencing

Total RNA was extracted using an RNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer protocol. RNA was quantified and checked for purity and integrity using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The RNA integrity number (RIN) of each RNA was >7.0 and, the purity (28S/18S) was >1.0. Total RNA from all liver samples within each replicate in each breed was pooled prior to library preparation for these six experimental groups in equimolar quantities.

Six pooled RNA samples (LCL1, LCL2, LCL3, MCL1, MCL2, and MCL3) were used for cDNA library preparation. Each pooled sample was subjected to DNase I digestion, and mRNA was extracted using Oligo (dT) magnetic beads and fragmented. After cDNA synthesis using fragmented mRNA as a template, the purified cDNA fragments were repaired and connected with sequencing adapters. Next, suitable fragments were purified and amplified by agarose gel electrophoresis and using an ABI StepOnePlus™ Real-Time PCR System (ABI, Carlsbad CA, USA), respectively. Last, the constructed libraries were sequenced on an Illumina HiSeq™ 2000 (Illumina, Inc., San Diego, CA, USA) at the Wuhan Genomics Institute (BGI, Shenzhen, China).

Reference sequence and mapping

After sequencing, quality control was performed as follows: Poor quality reads and sequencing adaptors were deleted. Then, the SOAPaligner/SOAP2 software (Li et al., 2009) was used to map reads, using the reference version of the sheep genome, with the advantage of faster matching and greater accuracy.

Detection of the gene expression level and DEGs

After mapping to the reference genome, the expression level of each gene was calculated based on the RPKM value (reads per kilobase of transcript per million mapped reads) (Mortazavi et al., 2008), using the following formula:

$$RPKM = \frac{10^6 C}{NL/10^3} \quad \text{(Equation 1)}$$

Provided RPKM is the expression of the gene, C is the number of reads uniquely matched to the gene, N is the number of total reads uniquely matched to the reference genome, and L is the base number of the coding region of the gene. The RPKM eliminates the effect of gene length and differences in sequencing, so that the gene expression level can be directly compared between different samples.

Differential expression analysis can identify genes that are differentially expressed between different samples. DEG analyses were performed by contrasting the differences in gene expression levels between different samples. The correction of false positive and negative errors was performed using the false discovery rate (FDR) (Benjamini and Hochberg, 1995). The P value threshold was determined by controlling the FDR value. After obtaining the FDR value of the difference test, the ratio of expression between different samples was calculated on the basis of gene expression (the RPKM value) (Benjamini and Yekutieli, 2001; Wang et al., 2010). The smaller the FDR value, the greater the fold-change value, which indicates that the gene expression difference is more significant. In this analysis, a gene where the FDR was £0.001 and the absolute value of the Log_2 ratio was ³1 was considered to be a DEG.

GO classification analysis

Gene Ontology (GO), an international standard for gene function classification, provides a dynamically updated, controlled vocabulary to describe genes and the properties of gene products in organisms (Ashburner et al., 2000). The three GO categories are molecular function, cellular components, and biological processes. GO analysis was performed using Gene Ontology database (http://geneontology.org/).

Identification of novel transcript units

The discovery of novel transcript units and the identification of shear in different transcript units are important characteristics in which RNA-Seq is superior to gene chip analysis. In order to discover novel transcribed regions, assembled transcript units were compared to annotated transcript units from reference genomes. Transcript units with lengths ³180 bp and sequencing depths ³2, which had a distance of at least 200 bp away from annotated genes, were identified as novel transcript units.

Analysis of cSNPs

cSNPs can be detected by comparison with a reference genome using the SOAPsnp software, which is part of the SOAP (Short Oligonucleotide Analysis Package) software (Li et al., 2009). The consensus sequence was obtained if clean reads mapped to the reference genome. On the basis of the consensus sequence and sequencing quality scores for each base, the program calculates the probability of the potential genotypes at each site. Then, posterior probability is calculated by Bayes' theorem, and the genotype with the highest posterior probability may be inferred.

Identification of alternative splicing

The genes of all eukaryotes are discontinuous, with exons interspersed between

introns. The initial transcription product must be modified and processed to allow for protein synthesis. Splicing modifies precursor mRNA by removing introns and joining exons. Alternative splicing (AS) makes it possible for a gene to generate different mRNA transcript units that are (probably) translated into distinguishable proteins. AS occurs in a number of ways. Many new types of AS have been discovered in humans (Wang et al., 2008), rice (Zhang et al., 2010), and other species (Filichkin et al., 2010). There are seven types of AS: exon skipping (ES); intron retention (IR); alternative 5' splice site (A5SS); alternative 3' splice site (A3SS); alternative first exon; alternative last exon; and mutually exclusive exon (Zhang et al., 2010). The last three AS event types had high false positive rates. Therefore, we only analyzed four types of AS events (i.e., ES, IR, A5SS, and A3SS). AS can be identified using the TopHat software (Trapnell et al., 2009).

RESULTS

Illumina HiSeq™ 2000 sequencing and read mapping

The liver transcriptomes (LCL1, LCL2, LCL3, MCL1, MCL2, and MCL3) of two phenotypically extreme sheep breeds (based on tail fat deposition) were sequenced, and six groups of paired-end reads were obtained. All reads were 91 bp. After removing the sequencing adaptors and poor quality reads, the SOAPaligner/SOAP2 (Li et al., 2009) software was used to map the reads against the reference sheep genome assembly. The results of clean reads were obtained (Tables 1 and 2). Observed percentages of reads mapped to the *Ovis aries* genome per library were 75.49, 77.17, 77.04, 76.82, 77.24, and 77.20%, respectively, and approximately 50% of the reads for each library could be mapped to the reference genes. Approximately 50% of reads for each library were identical to the *O. aries* genome, but only about 40% of reads were identical to the reference genes. Furthermore, >60% of reads had unique locations within the genome, and nearly 50% were matched to these unique locations in the reference genes. The results were almost equivalent for the two breeds.

Table 1. Summary of reads mapped to the *Ovis aries* genome.

	LCL1	LCL2	LCL3	MCL1	MCL2	MCL3
Total reads	46,686,582	49,032,008	48,983,064	47,400,744	47,268,062	47,206,272
Total bp	4,201,792,380	4,412,880,720	4,408,475,760	4,266,066,960	4,254,125,580	4,248,564,480
Total mapped reads	35,244,498 (75.49%)	37,836,252 (77.17%)	37,735,830 (77.04%)	36,413,318 (76.82%)	36,511,068 (77.24%)	36,443,428 (77.20%)
Perfect match	24,400,691 (52.26%)	26,009,702 (53.05%)	26,460,534 (54.02%)	25,657,739 (54.13%)	26,133,571 (55.29%)	24,121,577 (51.10%)
≤5-bp mismatch	10,843,807 (23.23%)	11,826,550 (24.12%)	11,275,296 (23.02%)	10,755,579 (22.69%)	10,377,497 (21.95%)	12,321,851 (26.10%)
Unique match	30,961,864 (66.32%)	32,373,010 (66.02%)	32,531,150 (66.41%)	31,198,952 (65.82%)	31,425,101 (66.84%)	31,279,814 (66.26%)
Multi-position match	4,282,634 (9.17%)	5,463,242 (11.14%)	5,204,680 (10.63%)	5,214,366 (11.00%)	5,085,967 (10.76%)	5,163,614 (10.94%)
Total unmapped reads	11,442,084 (24.51%)	11,195,756 (22.83%)	11,247,234 (22.96%)	10,987,426 (23.18%)	10,756,994 (22.76%)	10,762,844 (22.80%)

Table 2. Summary of reads mapped to the reference genes.

	LCL1	LCL2	LCL3	MCL1	MCL2	MCL3
Total reads	46,686,582	49,032,008	48,983,064	47,400,744	47,268,062	47,206,272
Total bp	4,201,792,380	4,412,880,720	4,408,475,760	4,266,066,960	4,254,125,580	4,248,564,480
Total mapped reads	27,969,522 (59.91%)	27,796,936 (56.89%)	27,474,265 (56.09%)	26,082,404 (55.03%)	26,676,817 (56.44%)	25,901,174 (54.87%)
Perfect match	20,434,616 (43.77%)	19,858,978 (40.50%)	20,074,242 (40.98%)	19,095,296 (40.28%)	19,802,583 (41.89%)	17,769,328 (37.64%%)
≤5-bp mismatch	7,534,906 (16.14%)	7,937,958 (16.19%)	7,400,023 (15.11%)	6,987,108 (14.74%)	6,874,234 (14.54%)	8,131,846 (17.23%)
Unique match	23,942,017 (51.28%)	23,980,845 (48.91%)	23,732,130 (48.45%)	22,607,239 (47.69%)	22,643,637 (47.90%)	22,278,963 (47.19%)
Multi-position match	4,027,505 (8.63%)	3,816,091 (7.78%)	3,742,135 (7.64%)	3,475,165 (7.33%)	4,033,180 (8.53%)	3,622,211 (7.67%)
Total unmapped reads	18,717,060 (40.09%)	21,235,072 (43.31%)	21,508,799 (43.39%)	21,318,340 (44.97%)	20,591,245 (43.56%)	21,305,098 (45.13%)

Analysis of gene expression

The transcriptomic analysis of liver tissue from the two breeds of sheep revealed extensive gene expression. In the six libraries, there were 16,257, 16,186, and 16,254 known reference genes in Lanzhou fat-tailed sheep, whereas there were 16,827, 16,437, and 15,761 known reference genes in Mongolian sheep (**S1-S6 Tables**). The percentage of genes expressed at more than 1000 RPKM was <1%, approximately 4% of genes were expressed at 100-1000 RPKM, and 95% of genes were expressed at <100 RPKM (Table 3). The maximum RPKM values were 21,476.89, 34,736.43, 28,940.24, 36,236.47, 37,483.01, and 28,813.44 for the six libraries (i.e., LCL1, LCL2, LCL3, MCL1, MCL2, and MCL3), respectively. There were similar read distributions in the six libraries (Figure 1). Of the reference genes, 38, 34, and 34% had 90-100% coverage in Lanzhou fat-tailed sheep, and 39, 38, and 29% had 90-100% coverage in Mongolian sheep.

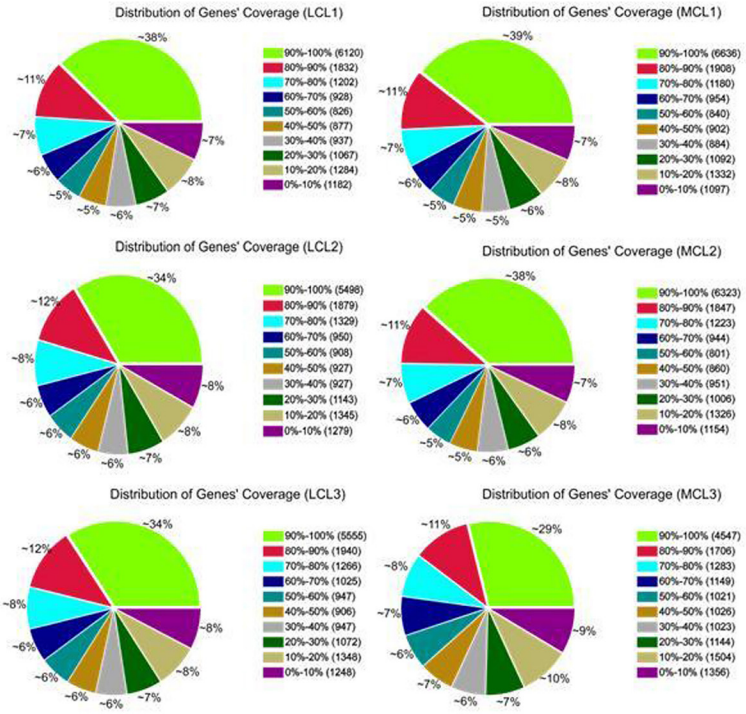

Figure 1. Distribution of gene coverage.

Table 3. Gene expression level as determined by RPKM.

RPKM	Gene number of LCL1	Gene number of LCL2	Gene number of LCL3	Gene number of MCL1	Gene number of MCL2	Gene number of MCL3
0-100	15,512 (95.42%)	15,438 (95.38%)	15,477 (95.22%)	16,030 (95.26%)	15,693 (95.47%)	14,926 (94.70%)
100-1,000	641 (3.94%)	643 (3.97%)	673 (4.14%)	701 (4.17%)	652 (3.97%)	711 (4.51%)
1,000-10,000	95 (0.58%)	94 (0.58%)	94 (0.58%)	86 (0.51%)	81 (0.49%)	115 (0.73%)
>10,000	9 (0.06%)	11 (0.07%)	10 (0.06%)	10 (0.06%)	11 (0.07%)	9 (0.06%)
Total	16257	16186	16254	16827	16437	15761

GO classification analysis

All genes identified from sheep liver tissue were annotated using a GO classification analysis and were classed into 3 categories, including cellular components, biological processes, and molecular functions, which were related to 55 types of biological functions. Under the cellular component category, large numbers of genes were categorized as cells (10,644 genes), cell parts (10,644 genes), organelles (8249 genes), or membranes (5282 genes). Under the biological process category, cellular processes (9791 genes), metabolic processes (7679 genes), and biological regulation (5195 genes) represented the greatest proportion of genes. Under the molecular function category, binding (10,067 genes), catalytic activity (5110 genes), and molecular transducer activity (1151 genes) were the most abundant subcategories (Figure 2).

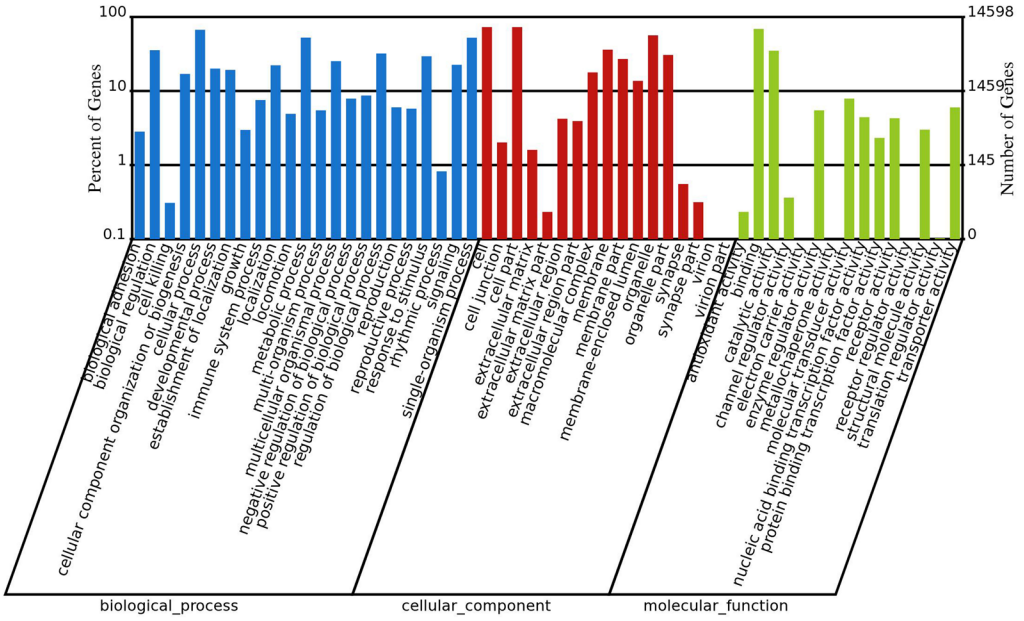

Figure 2. GO functional classification of Mongolian and Lanzhou fat-tailed sheep.

Analysis of DEGs

The DEG analysis was performed by comparing differences in the expression of genes between different samples. Using the evaluation criteria, seven genes were found to be differentially expressed, with four genes being upregulated and three genes downregulated.

Identification of novel transcript units

After matching reads to the sheep reference genome using the SOAPaligner/SOAP2 software, we obtained 1148, 1341, and 1260 novel transcript units for Lanzhou fat-tailed sheep,

and 1475, 1225, and 1331 novel transcript units for Mongolian sheep. In the six libraries, about 5% of the novel transcript units had only one exon (5.66, 5.07, 5.00, 3.67, 6.37, and 4.81%, respectively), and the longest novel transcript units contained 29, 29, 25, 29, 23, and 21 exons, respectively. In the six libraries, the number of novel transcript units on the + strands of genes was only slightly higher than that on the - strands (Table 4).

Table 4. Summary of novel transcript units.

+/-	LCL1	LCL2	LCL3	MCL1	MCL2	MCL3
+	600	719	656	770	648	683
-	548	622	604	705	577	648
Total	1148	1341	1260	1475	1225	1331

Analysis of cSNPs

After matching reads to the sheep reference genome, we located candidate SNPs in the coding regions. There were 65,303, 65,442, 63,426, 76,267, 69,853, and 57,439 potential cSNPs in LCL1, LCL2, LCL3, MCL1, MCL2, and MCL3, respectively. The most common variations were G/R substitutions followed by C/Y and T/Y substitutions (Table 5).

Table 5. Summary of SNPs in the coding region.

Allele variation	LCL1	LCL2	LCL3	MCL1	MCL2	MCL3
A/G	5039	4225	4505	4923	5631	3886
T/C	5007	4415	4435	4778	5626	3824
C/T	3433	2869	2993	3247	4028	2675
G/A	3290	2778	2880	3211	3909	2616
G/C	952	861	835	924	1058	739
C/G	980	823	850	964	1079	765
C/A	973	836	876	1022	1155	837
G/T	990	823	871	954	1115	809
A/C	817	704	729	830	970	682
T/G	906	747	802	826	971	665
A/T	624	517	547	601	699	455
T/A	612	535	532	615	664	482
C/Y	8382	9127	8645	10,684	8634	8015
G/R	8580	9181	8763	11,000	8806	7911
A/R	6822	7468	6871	8688	7036	6331
T/Y	6880	7537	6847	8908	7166	6420

Identification of alternative splicing

AS is a type of regulated process during gene expression and can generate different mRNA transcript units for a gene. AS frequently occurs in eukaryotes and can increase the biological diversity of a protein, which is encoded by a gene. Four known types of alternative splicing were detected in this study, including ES, IR, A5SS, and A3SS. Splicing in the six libraries accounted for approximately 40% of AS. There were 24,239, 22,283, 22,457, 26,635, 27,093, and 18,700 AS events in the six libraries, respectively. The distribution of AS events is shown in Tables 6 and 7. Of the different types of AS, IR occurred more often than other types; more than 33% of all AS events in the six libraries (38.33, 41.63, 41.22, 39.77, 37.96,

and 41.04%, respectively) were IR. ES occurred at the lowest frequency (18.91, 16.01, 16.60, 16.71, 17.56, and 16.64%, respectively) (Figure 3). In the six libraries, the number of ES, A5SS, and A3SS occurring on the + strands of genes was more than the number occurring on the - strands (Tables 6 and 7).

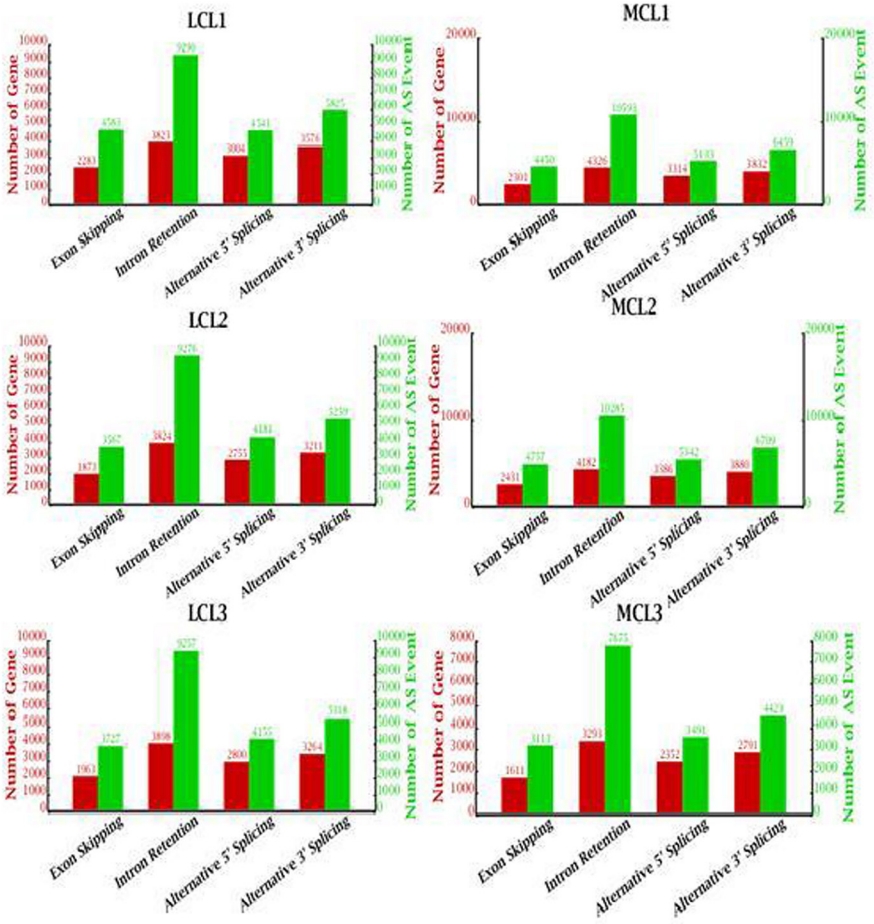

Figure 3. Summary of alternative splicing. This figure shows the number of alternative splices, including the four types in the six samples. The red bars and coordinates depict the number of genes, and the green bars and coordinates depict the number of events.

Table 6. Summary of alternative splicing in Lanzhou fat-tailed sheep.

AS type	LCL1			LCL2			LCL3		
	+	-	Total	+	-	Total	+	-	Total
ES	2881	1702	4583	2210	1357	3567	2309	1418	3727
IR	2960	6330	9290	2782	6494	9276	2816	6441	9257
A5SS	2311	2230	4541	2146	2035	4181	2098	2057	4155
A3SS	2975	2850	5825	2717	2542	5259	2727	2591	5318

Table 7. Summary of alternative splicing in Mongolian sheep.

AS type	MCL1			MCL2			MCL3		
	+	-	Total	+	-	Total	+	-	Total
ES	2,744	1,706	4,450	2,995	1,762	4,757	1,938	1,173	3,111
IR	3,618	6,975	10,593	3,624	6,661	10,285	1,893	5,782	7,675
A5SS	2,606	2,527	5,133	2,741	2,601	5,342	1,783	1,708	3,491
A3SS	3,381	3,078	6,459	3,551	3,158	6,709	2,300	2,123	4,423

DISCUSSION

New advances in high-throughput technologies have allowed for the large-scale analysis of genomic data, providing new opportunities for the characterization of transcriptome architectures. Due to the advantages of high-throughput technologies, the reliability, accuracy, and comprehensive data from RNA-Seq can not only help us understand gene structure, but also provide a better understanding of the biological function of genes (Wang et al., 2009). In addition, efficient Illumina high-throughput sequencing can help us discover new transcripts (Hansey et al., 2012), single nucleotide polymorphisms (Esteve-Codina et al., 2011), AS sites (Xiong et al., 2012), and other information. In this study, the transcriptomes of Mongolian and Lanzhou fat-tailed sheep were compared. Though they are both native to the cooler and drier regions of western China, they have many different traits and belong to two different ecotypes (small tailed and fat-tailed sheep, respectively), according to their biological classification.

High-quality reads of 286 Mb were obtained from an RNA-Seq analysis of the six libraries, with three-quarters of reads matching the *O. aries* genome, and approximately 50% of reads being identical to the reference gene. In similar studies of pigs (Chen et al., 2011; Esteve-Codina et al., 2011) and cattle (Huang et al., 2012), in which 61.6-77.8% of reads were matched to the reference genome, our results were comparable for sheep. The percentage of reads mapped to the reference genes were lower than that mapped to the *O. aries* genome, a result potentially due to the GC content, type of cells, and other reasons (McIntyre et al., 2011; Sendler et al., 2011). The percentage of reads mapped to the reference genes in the LCL libraries (59.91, 56.89, and 56.09%) were slightly higher than those for the MCL libraries (55.03, 56.44, and 54.87%), but the percentage of reads mapped to the *O. aries* genome were not similar. The percentage of total reads not mapped to the reference genes were high (40.09, 43.31, 43.39, 44.97, 43.56, and 45.13%, respectively), which suggests that further research would be desirable to determine the reasons for these high percentages.

Characterizing gene expression has long been of interest to researchers, not only with regard to the genes expressed in certain tissues but also with respect to the level of expression. Computational approaches need to be developed to analyze differential expression levels in cell lineages and tissues (Huang et al., 2012; Li et al., 2014). Although measuring the concentration of mRNA is still a useful method, in order to analyze gene expression, gene coverage should be calculated. There were 38, 34, and 34% of the reference genes with 90-100% coverage in the LCL libraries, and 39, 38, and 29% with similar coverage in the MCL libraries. In this study, about 50% of genes annotated in the sheep genome were covered by more than 80% of the sequenced reads, demonstrating that RNA-Seq was more sensitive in finding transcripts, even for genes with low expression patterns, which is in agreement with Wang et al. (2009). Although several recent studies have identified DEGs in the liver tissue of pigs (Gunawan et al., 2013a,b; Sodhi et al., 2014), cows (Graugnard et al., 2013), rats (Tran

and Huang, 2014; Wang et al., 2014) and fish (Yang et al., 2012; Lau et al., 2014; Magnanou et al., 2014) when using next-generation sequencing platforms, very few studies have clarified transcriptomic analyses regarding the liver tissue of sheep. In this study, seven genes were found to be differentially expressed (three were upregulated and four downregulated) in a DEG analysis comparing the liver tissues of the LCL and MCL libraries. These DEGs may serve as the focus for a future study to better understand the complexities of lipid deposition and metabolism.

Identification of cSNPs using RNA-Seq is limited to transcribed regions, since this method can only be used to discover SNPs in coding regions. An analysis of sequencing data from other studies identified numerous cSNPs in eukaryotes (Cánovas et al., 2010; Blanca et al., 2011; Nie et al., 2011). In this study, there were 65,303, 65,442, 63,426, 76,267, 69,853, and 57,439 potential cSNPs detected using sequencing reads from the six libraries, respectively. The six libraries had similar base variations, and the most common variations were G/R and C/Y substitutions, followed by T/Y. The number of SNPs on chromosomes 1, 2, and 3 were higher than other chromosomes, probably because of the longer lengths of these chromosomes.

AS is an important regulated process during gene expression, and it can generate different mRNA transcript units for a gene. AS frequently occurs in eukaryotes. In the two sheep breeds, we found similar AS rates, with 35.69, 34.03, 34.50, 38.33, 37.46, and 30.54% of the reference genes in the six libraries undergoing 24,239, 22,283, 22,457, 26,635, 27,093, and 18,700 AS events. This is much lower than that reported for humans (86.0%) (Wang et al., 2008), but is higher than that reported for pigs (18.0%) (Chen et al., 2011), sheep (25.28-26.02%) (Zhang et al., 2013), and rice (33.0%) (Zhang et al., 2010). The most common AS identified in our research was IR, in both Mongolian and Lanzhou fat-tailed sheep. Our results were similar to those for rice (Zhang et al., 2010), but different from those for pigs (Chen et al., 2011) and Dorper and small-tailed Han sheep (Zhang et al., 2013), in which A3SS was identified as the most common AS. Our results also differed from those of humans (Sultan et al., 2008), in which ES was the primary type of AS. This phenomenon may result from differences between eukaryotes in the mechanisms of alternative splicing. In addition, the secondary structure of the precursor mRNA transcript also regulates splicing by bringing splicing elements together (Reid et al., 2009; Warf and Berglund, 2010). More studies should be conducted to determine gene regulation in the two breeds, which shed light on the complexities of lipid metabolism, and may provide useful references in addressing issues of affecting humans, including obesity and diabetes.

Based on the high-quality reads, we characterized the sheep liver transcriptome and analyzed DEGs, cSNPs, and other information by bioinformatics analysis. Our results indicated that there may be several differences in the liver transcriptomes between the two breeds, despite the fact that both breeds are native to the cooler and drier regions of western China. Our sequencing data and analyses will help guide future research.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by a special fund for the China Agriculture Research System

(program code: #CARS-40-09B), the National Natural Science Foundation (program code: #31460592), and a special fund for State Administration of Foreign Experts Affairs P.R. China (program codes: #GDT20156200087 and #GDT20156200087).

REFERENCES

Ashburner M, Ball CA, Blake JA, Botstein D, et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25: 25-29. http://dx.doi.org/10.1038/75556

Atti N, Bocquier F and Khaldi G (2004). Performance of the fat-tailed Barbarine sheep in its environment: adaptive capacity to alternation of underfeeding and re-feeding periods. *Anim. Res.* 53: 165-176. http://dx.doi.org/10.1051/animres:2004012

Benjamini Y and Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J.R. Stat. Soc.* 57: 289-300.

Benjamini Y and Yekutieli D (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29: 1165-1181.

Blanca J, Cañizares J, Roig C, Ziarsolo P, et al. (2011). Transcriptome characterization and high throughput SSRs and SNPs discovery in Cucurbita pepo (Cucurbitaceae). *BMC Genomics* 12: 104-119. http://dx.doi.org/10.1186/1471-2164-12-104

Cánovas A, Rincon G, Islas-Trejo A, Wickramasinghe S, et al. (2010). SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mamm. Genome* 21: 592-598. http://dx.doi.org/10.1007/s00335-010-9297-z

Chen C, Ai H, Ren J, Li W, et al. (2011). A global view of porcine transcriptome in three tissues from a full-sib pair with extreme phenotypes in growth and fat deposition by paired-end RNA sequencing. *BMC Genomics* 12: 448-464. http://dx.doi.org/10.1186/1471-2164-12-448

Ding SD, Wu XH and Li XM (1986). Survey report of Lanzhou-fat-tailed sheep. *Chin. J. Anim. Sci.* 1: 20-21.

Esteve-Codina A, Kofler R, Palmieri N, Bussotti G, et al. (2011). Exploring the gonad transcriptome of two extreme male pigs with RNA-seq. *BMC Genomics* 12: 552-566. http://dx.doi.org/10.1186/1471-2164-12-552

Filichkin SA, Priest HD, Givan SA, Shen R, et al. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis* thaliana. *Genome Res.* 20: 45-58. http://dx.doi.org/10.1101/gr.093302.109

Graugnard DE, Moyes KM, Trevisi E, Khan MJ, et al. (2013). Liver lipid content and inflammometabolic indices in peripartal dairy cows are altered in response to prepartal energy intake and postpartal intramammary inflammatory challenge. *J. Dairy Sci.* 96: 918-935. http://dx.doi.org/10.3168/jds.2012-5676

Gunawan A, Sahadevan S, Neuhoff C, Große-Brinkhaus C, et al. (2013a). RNA deep sequencing reveals novel candidate genes and polymorphisms in boar testis and liver tissues with divergent androstenone levels. *PLoS One* 8: e63259. http://dx.doi.org/10.1371/journal.pone.0063259

Gunawan A, Sahadevan S, Cinar MU, Neuhoff C, et al. (2013b). Identification of the novel candidate genes and variants in boar liver tissues with divergent skatole levels using RNA deep sequencing. *PLoS One* 8: e72298. http://dx.doi.org/10.1371/journal.pone.0072298

Hansey CN, Vaillancourt B, Sekhon RS, de Leon N, et al. (2012). Maize (*Zea mays* L.) genome diversity as revealed by RNA-sequencing. *PLoS One* 7: e33071. http://dx.doi.org/10.1371/journal.pone.0033071

Huang W, Nadeem A, Zhang B, Babar M, et al. (2012). Characterization and comparison of the leukocyte transcriptomes of three cattle breeds. *PLoS One* 7: e30244. http://dx.doi.org/10.1371/journal.pone.0030244

Kashan NEJ, Azar GHM, Afzalzadeh A and Salehi A (2005). Growth performance and carcass quality of fattening lambs from fat-tailed and tailed sheep breeds. *Small Rumin. Res.* 60: 267-271. http://dx.doi.org/10.1016/j.smallrumres.2005.01.001

Lau K, Lai KP, Bao JY, Zhang N, et al. (2014). Identification and expression profiling of microRNAs in the brain, liver and gonads of marine medaka (*Oryzias melastigma*) and in response to hypoxia. *PLoS One* 9: e110698. http://dx.doi.org/10.1371/journal.pone.0110698

Li R, Yu C, Li Y, Lam TW, et al. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25: 1966-1967. http://dx.doi.org/10.1093/bioinformatics/btp336

Li HD, Menon R, Omenn GS and Guan Y (2014). The emerging era of genomic data integration for analyzing splice isoform function. *Trends Genet.* 30: 340-347. http://dx.doi.org/10.1016/j.tig.2014.05.005

Magnanou E, Klopp C, Noirot C, Besseau L, et al. (2014). Generation and characterization of the sea bass *Dicentrarchus labrax* brain and liver transcriptomes. *Gene* 544: 56-66. http://dx.doi.org/10.1016/j.gene.2014.04.032

McIntyre LM, Lopiano KK, Morse AM, Amin V, et al. (2011). RNA-seq: technical variability and sampling. *BMC Genomics* 12: 293. http://dx.doi.org/10.1186/1471-2164-12-293

Mortazavi A, Williams BA, McCue K, Schaeffer L, et al. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5: 621-628. http://dx.doi.org/10.1038/nmeth.1226

Nie Q, Fang M, Jia X, Zhang W, et al. (2011). Analysis of muscle and ovary transcriptome of *Sus scrofa*: assembly, annotation and marker discovery. *DNA Res.* 18: 343-351. http://dx.doi.org/10.1093/dnares/dsr021

Reid DC, Chang BL, Gunderson SI, Alpert L, et al. (2009). Next-generation SELEX identifies sequence and structural determinants of splicing factor binding in human pre-mRNA sequence. *RNA* 15: 2385-2397. http://dx.doi.org/10.1261/rna.1821809

Sendler E, Johnson GD and Krawetz SA (2011). Local and global factors affecting RNA sequencing analysis. *Anal. Biochem.* 419: 317-322. http://dx.doi.org/10.1016/j.ab.2011.08.013

Sodhi SS, Song KD, Ghosh M, Sharma N, et al. (2014). Comparative transcriptomic analysis by RNA-seq to discern differential expression of genes in liver and muscle tissues of adult Berkshire and Jeju Native Pig. *Gene* 546: 233-242. http://dx.doi.org/10.1016/j.gene.2014.06.005

Sultan M, Schulz MH, Richard H, Magen A, et al. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321: 956-960. http://dx.doi.org/10.1126/science.1160342

Tran NT and Huang CH (2014). Gene expression and gene ontology enrichment analysis for H3K4me3 and H3K4me1 in mouse liver and mouse embryonic stem cell using ChIP-Seq and RNA-Seq. *Gene Regul. Syst. Bio.* 8: 33-43. http://dx.doi.org/10.4137/GRSB.S13612

Trapnell C, Pachter L and Salzberg SL (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105-1111. http://dx.doi.org/10.1093/bioinformatics/btp120

Wang ET, Sandberg R, Luo S, Khrebtukova I, et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470-476. http://dx.doi.org/10.1038/nature07509

Wang Z, Gerstein M and Snyder M (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10: 57-63. http://dx.doi.org/10.1038/nrg2484

Wang L, Feng Z, Wang X, Wang X, et al. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26: 136-138. http://dx.doi.org/10.1093/bioinformatics/btp612

Wang C, Gong B, Bushel PR, Thierry-Mieg J, et al. (2014). The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* 32: 926-932. http://dx.doi.org/10.1038/nbt.3001

Warf MB and Berglund JA (2010). Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* 35: 169-178. http://dx.doi.org/10.1016/j.tibs.2009.10.004

Xiong J, Lu X, Zhou Z, Chang Y, et al. (2012). Transcriptome analysis of the model protozoan, *Tetrahymena thermophila*, using Deep RNA sequencing. *PLoS One* 7: e30630. http://dx.doi.org/10.1371/journal.pone.0030630

Yang D, Liu Q, Yang M, Wu H, et al. (2012). RNA-seq liver transcriptome analysis reveals an activated MHC-I pathway and an inhibited MHC-II pathway at the early stage of vaccine immunization in zebrafish. *BMC Genomics* 13: 319. http://dx.doi.org/10.1186/1471-2164-13-319

Zhang G, Guo G, Hu X, Zhang Y, et al. (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* 20: 646-654. http://dx.doi.org/10.1101/gr.100677.109

Zhang C, Wang G, Wang J, Ji Z, et al. (2013). Characterization and comparative analyses of muscle transcriptomes in Dorper and small-tailed Han sheep using RNA-Seq technique. *PLoS One* 8: e72686. http://dx.doi.org/10.1371/journal.pone.0072686

Supplementary material

S1 Table. Expression statistics for reference genes in the transcriptome of Lanzhou fat-tailed sheep (LCL1).

S2 Table. Expression statistics for reference genes in the transcriptome of Lanzhou fat-tailed sheep (LCL2).

S3 Table. Expression statistics for reference genes in the transcriptome of Lanzhou fat-tailed sheep (LCL3).

S4 Table. Expression statistics for reference genes in the transcriptome of Mongolian sheep (MCL1).

S5 Table. Expression statistics for reference genes in the transcriptome of Mongolian sheep (MCL2).

S6 Table. Expression statistics for reference genes in the transcriptome of Mongolian sheep (MCL3).