Genome-wide transcriptome analysis between Small-tail Han sheep and the Surabaya fur sheep using high-throughput RNA sequencing

Xiangyang Miao and Qingmiao Luo

Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing 100193, China

Correspondence should be addressed to X Miao; Email: mxy32@sohu.com

Abstract

The Small-tail Han sheep and the Surabaya fur sheep are two local breeds in North China, which are characterized by high-fecundity and low-prolificacy breed respectively. Significant genetic differences between these two breeds have provided increasing interests in the identification and utilization of major prolificacy genes in these sheep. High prolificacy is a complex trait, and it is difficult to comprehensively identify the candidate genes related to this trait using the single molecular biology technique. To understand the molecular mechanisms of fecundity and provide more information about high prolificacy candidate genes in high- and low-fecundity sheep, we explored the utility of next-generation sequencing technology in this work. A total of 1.8 Gb sequencing reads were obtained and resulted in more than 20 000 contigs that averaged ~300 bp in length. Ten differentially expressed genes were further verified by quantitative real-time RT-PCR to confirm the reliability of RNA-seq results. Our work will provide a basis for the future research of the sheep reproduction.

Reproduction (2013) **145** 587–596

Background

High prolificacy is one of the important goals of the sheep breeding in the world. The majority of sheep species produce single lamb and a small number will produce twin lambs, which greatly affected the breeding production. Owing to the low heritability of sheep lambing number (only from 0.03 to 0.1), it was difficult to improve this feature through traditional selection. So scientists pay more attention to search the candidate genes or mutations associated with ovulation rate and multiplets. High prolificacy is a complex trait, and it is difficult to thoroughly identify the candidate genes related to this trait using the single molecular biology technique, which was affected by genes, age, season, and nutrition, and the genetic factor is the most important factor.

Small-tail Han sheep (Han sheep) and Surabaya fur sheep are two local breeds in Shandong province of North China. The former is characterized by high-fecundity with the mean litter size of 2.61 and an average lambing rate of 286.5%, while the latter is of low-prolificacy breed with an average lambing rate of 121%. Significant genetic differences between these two breeds have provided increasing interest in the identification and utilization of major prolificacy genes in these sheep.

With the development of molecular technologies and 'omics' tools, microarrays have been considered to be a standard strategy for the global analysis of gene expression. Thanks to the emergence of massively parallel cDNA sequencing technologies (Mardis 2008, Mortazavi *et al.* 2008), researchers can obtain coding sequence fragments from transcriptome data with nearly complete coverage in a reduced time and at a lower cost. Several recently developed sequencing technologies, such as the 454 Life Sciences (Roche) pyrosequencing platform, the Illumina Genome Analyzer, and Applied Biosystems SOLiD platform, can offer massively parallel production of short reads. Recent methods and experimental advances have made it possible to perform *de novo* sequence projects (Li *et al.* 2009). Many previous pioneering works have suggested that next-generation sequencing technologies provide a massive amount of useful information for the transcriptome analysis in non-model organisms (Gibbons *et al.* 2009, Wolf *et al.* 2010). As a powerful tool to explore the expression levels of thousands of genes simultaneously in cell or tissue, the microarray technology was discussed and reviewed in livestock species. For example, to gain an understanding of the genomic influence on milk quality and synthesis, Bongiorno *et al.* (2009) used sheep-specific microarray technology covering most of the species' transcriptome and representing

the first annotated microarray developed for sheep with a covering of 50% of the genome. [Peñagaricano et al. \(2012\)](#) studied the gene expression at black skin spots and white skin using microarray technology to identify the possible genes responsible for the development of this feature.

Until now, there is no high-coverage genome sequence available in the sheep, and microarray analysis of the multiplets has not been studied. In this work, to understand the molecular mechanism of fecundity and to provide more information on hyper-prolificacy candidate genes in high- and low-fecundity sheep, we explored the utility of next-generation sequencing technology for studying differential expression profiling in ovaries of the mRNAs in three groups (genotype BB Han sheep, genotype ++ Han sheep, and Surabaya fur sheep). Our work successfully detected genes that were differentially expressed between the Small-tail Han sheep and the Surabaya fur sheep and also validate our RNA-seq data by real-time RT-PCR analysis. The report showed a *de novo* assembly approach for analyzing the transcriptome data of a non-model species using short-read sequencing data.

Results

De novo assembly RNA-seq and statistics of sequencing reads

We sequenced three lanes of the cDNA library from the genotype BB Han sheep, genotype ++ Han sheep, and Surabaya fur sheep (see Materials and Methods section) respectively, and totally, about 1.8 Gb of 50 bp single-end reads (0.62, 0.6, and 0.61 Gb for the genotype BB Han sheep, genotype ++ Han sheep, and Surabaya fur sheep respectively) were obtained.

Next, we performed *de novo* assembly of sequencing reads as described in the ‘Materials and Methods’ section. The bioinformatics workflow is described in [Fig. 1](#). The results of our *de novo* sequence assembly are summarized in [Table 1](#). The result showed varying amount of sequencing reads for these samples. We obtained 22 507 contigs for the genotype BB Han sheep, 20 553 contigs for the genotype ++ Han sheep, and 23 965 contigs for the Surabaya fur sheep. The total amount of sequencing reads that can contribute to the assemblies was 4.1 Gb for the genotype BB Han sheep, 3.7 Gb for the genotype ++ Han sheep, and 3.2 Gb for the Surabaya fur sheep respectively. Next, we analyzed the length distribution of the contigs. As shown in [Fig. 2](#), most of the sequences are between 150 and 500 bp, and the average length of assembly contigs is ~300 bp.

To identify the contigs that were previously uncharacterized in sheep, we downloaded all partial or complete ovine genes that were publicly available. After removing the redundant sequences, totally 2209 (from 8341) genes were obtained. In this work, we identified

Figure 1 *De novo* transcriptome assembly. This figure summarizes computational procedures for assembly, annotation, and mapping of the sequencing reads.

25 552 sheep contigs, including 25 216 new contigs for which no sequence information was previously available. This information will also provide a basis for further molecular research involving the sheep as a model organism.

Digital gene expression estimated by RNA-seq

Next, we performed expression analysis. The mapped read counts for each gene were normalized for RNA length and for the total read number in the lane according to reads per kilobase of exon model per million mapped reads (RPKM), which facilitates comparison of transcript levels between samples. As shown in [Fig. 3](#), we found significant correlation in expression among three samples. As expected, the correlation of each of the Han sheep samples with the Surabaya fur sheep ($R^2=0.69$ with the genotype BB sample and $R^2=0.76$ with the genotype ++ sample)

Table 1 Summary of sequencing reads *de novo* assembled and the alignment to bovine genes.

	BB Han sheep	++ Han sheep	Surabaya fur sheep
Total reads	6 164 788	6 008 324	6 096 940
Total mapped reads	3 950 635	3 777 991	3 246 466
De novo assembled contigs	22 507	20 553	23 965
Contigs mapped to annotated bovine genes	6304	6263	6970
Contigs N50	318	306	287
Contigs maximum length	6600	5819	3861
Contigs average length	303	295	281

Figure 2 Distribution of *de novo* assembled transcripts and their length.

was lower compared with the correlation between the two Han sheeps ($R^2 = 0.85$).

For differential gene expression measurements, DESeq package was used. The significance threshold of P value in multiple tests was set by false discovery rate (FDR). We use 'FDR < 0.01' as the threshold to judge the significance of gene expression difference. In this work, the resultant assembly contigs were aligned to the bovine genes based on the homology strategies to detect the differentially expressed genes (DEGs). Between the two

Figure 3 The correlation of gene expression between different samples. The expression levels are estimated by RPKM value.

Han sheep samples, we totally identified 2314 (1071 genes upregulated and 1243 genes downregulated) DEGs. One thousand nine hundred and twenty four (1351 genes upregulated and 573 genes downregulated)

and 1501 (1027 genes upregulated and 474 genes downregulated) DEGs were identified between the Surabaya fur sheep with the genotype BB Han sheep and genotype ++ Han sheep respectively. These results are consistent with the correlation analysis that there are few differential expression genes between the two Han sheep samples. All DEGs between each of the two Han sheep samples are listed in [Supplementary Table 1A, B and C](#), see section on [supplementary data](#) given at the end of this article, and the M-A plot is shown in [Fig. 4](#).

Gene ontology enrichment analysis

To gain insights into the biological implications, we tested for enrichment of DEGs in gene ontology (GO) terms. GO categories are organized into three groups: biological process (BP), cellular component (CC), and molecular function (MF). The three groups characterize different aspects of a gene's function and are thus examined separately in our analysis. Genes that showed a nominal significance of $P < 0.01$ were selected and tested against the background set of all genes with GO annotations. We found several GO terms significantly enriched ($FDR < 0.01$) for DEGs, among which were GO processes related to their biological functions. All enriched GO terms are shown in [Table 2](#). The results indicated that the main functional groups of DEGs between the genotype BB Han sheep and genotype ++ Han sheep in BP are multicellular organismal development and anatomical structure development, in MF are signal transducer activity, and in CC are ribonucleoprotein complex and plasma membrane part. The results showed that BP-related genes might play important roles in the difference of fecundity between BB and ++ Han sheep. The main functional groups of DEGs between the Surabaya fur sheep and genotype BB Han sheep in BP are translation-related genes, in MF are structural constituent of ribosome and receptor binding, and in CC are plasma membrane and ribonucleoprotein complex. The results showed that CC-related genes might play important roles in the difference between the Surabaya fur sheep and genotype BB Han sheep, and the main functional groups of DEGs between the Surabaya fur sheep and genotype ++ Han sheep in BP are antigen processing and presentation, in MF are receptor binding and structural constituent of ribosome-related genes, and in CC are plasma membrane and ribonucleoprotein complex. The results showed that CC-related genes might play important roles in the difference between the Surabaya fur sheep and genotype ++ Han sheep.

Validation of RNA-seq results by real-time RT-PCR

To validate our RNA-seq data, real-time RT-PCR analysis was performed on ten selected differentially expressed transcripts ([Table 1](#)). These genes belong to divergent

Figure 4 M-A plot of differential expressed genes. The red dots represent genes with significant gene expression differences and black dots represent genes without significant gene expression differences.

functional categories or pathways. For instance, four genes (latent transforming growth factor β (TGF β)-binding protein 1, TGF β 1-induced transcript 1 protein, TGF β inducible early growth response protein 2 and latent TGF β -binding protein 4) were associated with TGF β . Three genes (Sigma non-opioid intracellular receptor 1, scavenger receptor class B member 1, and *NOTCH2*) were grouped as receptor. In addition, two

Table 2 Enriched gene ontology (GO) terms for all differential expression genes: 'CC', 'MF', and 'BP' represent cellular component, molecular function, and biological process respectively.

	GO term	Description	DEGs	All genes	P value	FDR
BB Han sheep vs ++ Han sheep						
CC	GO:0005578	Proteinaceous extracellular matrix	44	67	3.72×10^{-4}	1.16×10^{-2}
	GO:0031012	Extracellular matrix	51	78	1.53×10^{-4}	9.55×10^{-3}
	GO:0044420	Extracellular matrix part	23	32	1.59×10^{-3}	2.31×10^{-2}
	GO:0044459	Plasma membrane part	101	182	1.67×10^{-3}	2.31×10^{-2}
	GO:0005615	Extracellular space	37	54	3.05×10^{-4}	1.16×10^{-2}
	GO:0005840	Ribosome	92	135	1.83×10^{-8}	2.29×10^{-6}
	GO:0030529	Ribonucleoprotein complex	122	219	4.66×10^{-4}	1.16×10^{-2}
	GO:0022626	Cytosolic ribosome	11	12	9.65×10^{-4}	2.01×10^{-2}
	GO:0016323	Basolateral plasma membrane	29	42	1.13×10^{-3}	2.02×10^{-2}
	GO:0004871	Signal transducer activity	191	164	1.83×10^{-4}	3.48×10^{-2}
MF	GO:0030247	Polysaccharide binding	18	4	3.85×10^{-4}	4.88×10^{-2}
	GO:0003735	Structural constituent of ribosome	87	35	1.34×10^{-9}	5.10×10^{-7}
BP	GO:0006412	Translation	134	82	1.01×10^{-7}	9.74×10^{-5}
	GO:0007155	Cell adhesion	101	71	1.04×10^{-4}	2.49×10^{-2}
	GO:0007275	Multicellular organismal development	272	244	5.49×10^{-5}	2.41×10^{-2}
	GO:0048856	Anatomical structure development	249	221	7.53×10^{-5}	2.41×10^{-2}
Surabaya fur sheep vs BB Han sheep						
CC	GO:0005886	Plasma membrane	154	314	1.33×10^{-5}	5.61×10^{-4}
	GO:0044459	Plasma membrane part	89	182	1.02×10^{-3}	2.15×10^{-2}
	GO:0005840	Ribosome	93	912	8.74×10^{-16}	1.10×10^{-13}
	GO:0030529	Ribonucleoprotein complex	120	209	2.45×10^{-9}	1.54×10^{-7}
	GO:0022626	Cytosolic ribosome	11	12	1.65×10^{-4}	5.20×10^{-3}
	GO:0005615	Extracellular space	32	51	2.24×10^{-4}	5.65×10^{-3}
MF	GO:0004857	Enzyme inhibitor activity	32	52	3.74×10^{-4}	3.50×10^{-2}
	GO:0003735	Structural constituent of ribosome	87	118	8.01×10^{-16}	3.00×10^{-13}
	GO:0005102	Receptor binding	75	144	2.38×10^{-4}	2.96×10^{-2}
	GO:0005179	Hormone activity	14	17	2.07×10^{-4}	2.96×10^{-2}
BP	GO:0006412	Translation	118	209	1.30×10^{-8}	1.21×10^{-5}
Surabaya fur sheep vs ++ Han sheep						
CC	GO:0005886	Plasma membrane	124	327	1.72×10^{-4}	5.34×10^{-3}
	GO:0042612	MHC class I protein complex	6	7	3.04×10^{-3}	3.38×10^{-2}
	GO:0044459	Plasma membrane part	71	184	2.49×10^{-3}	3.38×10^{-2}
	GO:0005578	Proteinaceous extracellular matrix	32	65	3.93×10^{-4}	9.76×10^{-3}
	GO:0031012	Extracellular matrix	38	76	7.54×10^{-5}	3.11×10^{-3}
	GO:0015630	Microtubule cytoskeleton	40	94	2.97×10^{-3}	3.38×10^{-2}
	GO:0005783	Endoplasmic reticulum	64	167	4.79×10^{-3}	4.57×10^{-2}
	GO:0030016	Myofibril	8	11	3.27×10^{-3}	3.38×10^{-2}
	GO:0005840	Ribosome	65	129	1.58×10^{-7}	1.96×10^{-5}
	GO:0030529	Ribonucleoprotein complex	91	214	9.53×10^{-6}	5.91×10^{-4}
	GO:0042383	Sarcolemma	11	17	2.28×10^{-3}	3.38×10^{-2}
	GO:0022626	Cytosolic ribosome	9	12	1.24×10^{-3}	2.56×10^{-2}
	GO:0033179	Proton-transporting V-type ATPase, V0 domain	5	5	2.00×10^{-3}	3.38×10^{-2}
	GO:0005102	Receptor binding	67	150	2.22×10^{-5}	4.21×10^{-3}
MF	GO:0003735	Structural constituent of ribosome	60	118	3.18×10^{-7}	1.20×10^{-4}
BP	GO:0019882	Antigen processing and presentation	18	26	2.29×10^{-5}	2.16×10^{-2}

genes were implicated in hormone regulation (StAR protein, mitochondrial, and folliculin) and one was assigned to metalloproteinase (ADAM metalloproteinase with thrombospondin type 1 motif 7). For these genes, the expression fold changes ($2^{-\Delta\Delta Ct}$) were also measured by real-time RT-PCR (Table 3). Their expression kinetics from the real-time RT-PCR results was consistent with the RNA-seq analysis.

Discussion

It is well known that mutations that increase ovulation rate and affect reproductive performance have been discovered in the bone morphogenetic protein 15 gene

(*BMP15*; breeds showing this mutation include the Inverdale, Hanna, Belclare, Cambridge, and Laucune; Shimasaki *et al.* 1999), its receptor gene (*BMPRI1B*; the Booroola breed), and growth differentiation factor 9 gene (*GDF9*; the Cambridge and Belclare breeds) (Paradis *et al.* 2009). In Han sheep, point mutations of the *BMP15* gene B2 (CfwdarwT), the *BMPRI1B* gene FecB (A746G), and the *GDF9* gene G3 (G477A) and one novel single-nucleotide mutation (G729T) have been detected by PCR-SSCP or PCR-RFLP. FecB is a single mutation gene that results in a marked increase in ovulation rate and prolificacy in Han sheep; however, the mechanisms whereby this mutation alters ovarian function in Han sheep are still unclear, and the absence of the FecB mutations in the other prolific breeds does not preclude

Table 3 Validation of selected RNA-seq-based gene expression by real-time RT-PCR analysis.

Gene id	Transcript id	Description	RNA-seq			Real-time RT-PCR		
			BB/+ +	BB/S	+ +/S	BB/S	+ +/S	Primer sequence
ENSBTAG000000033345	ENSBTAT000000047426	Steroidogenic acute regulatory protein, mitochondrial (source: UniProtKB/Swiss-Prot accession no.: Q28918)	-2.21	-	-	2.40	5.33	F: TCCATGTGCGTG CTGGCTGG R: GGAGCACCATG CAGGTGGGG
ENSBTAG000000015804	ENSBTAT000000020990	Sigma non-opioid intracellular receptor 1 (source: UniProtKB/Swiss-Prot accession no.: Q58DH7)	-2.45	-1.63	0.86	0.66	0.85	F: GCCCGTGCACC ACTGTCTCC R: GCTCTTCGGCA CCGCTCTGG
ENSBTAG000000014269	ENSBTAT000000016321	Scavenger receptor class B member 1 (source: UniProtKB/Swiss-Prot accession no.: O18824)	-2.63	-1.52	1.16	0.78	1.17	F: TGGCCTCCATCC TAACG R: TCGAACCAATT AGCGG
ENSBTAG000000019839	ENSBTAT000000026435	Latent transforming growth factor β (TGF β)-binding protein 1 (source: RefSeq peptide accession no.: NP_001096561)	2.48	1.22	1.38	2.07	1.06	F: GGAAGGACCCT TCTACATT R: TTGTGACAATAC GGCTGG
ENSBTAG000000002391	ENSBTAT000000003094	TGF β 1-induced transcript 1 protein (source: UniProtKB/Swiss-Prot accession no.: Q3MHZ4)	0.69	1.52	-	1.97	2.68	F: CAGGTCTTTCGA GGTGG R: CTTCACTGCAC CTTCTG
ENSBTAG000000008010	ENSBTAT000000010532	Follistatin (source: UniProtKB/Swiss-Prot accession no.: Q3B7L5)	-2.09	-	-	1.82	0.89	F: CAGGCAAAGGA GGTGTG R: TATTGAAGCGG AGCAGTT
ENSBTAG000000046218	ENSBTAT000000066246	TGF β -inducible early growth response protein 2 uncharacterized protein (source: UniProtKB/TrEMBL accession no.: A3KE41)	0.75	-	-	1.90	1.19	F: CCAGGTGTCTCC GGGGCTTG R: TGCCGGCTGAG CATGTGGTG
ENSBTAG000000004757	ENSBTAT000000065855	Latent TGF β -binding protein 4 (source: HGNC symbol accession no.: 6717)	-0.55	-0.54	1.22	2.37	1.39	F: GTCGGTGTGGT ACAGAIG R: GCAGACGTGAA CGAGTG
ENSBTAG000000001396	ENSBTAT000000038814	ADAM metalloproteinase with thrombospondin type 1 motif, 7 (source: HGNC symbol accession no.: 223)	-0.82	-	2.52	2.55	1.67	F: CTCACAGTGGC TGTCCGGG R: CCGCCAACTTCC TGGCCCTG
ENSBTAG000000007909	ENSBTAT000000010400	Notch2 (source: HGNC symbol accession no.: 7882)	-1.17	-2.15	-0.93	1.31	0.98	F: TCTAGACGGCCT AACACC R: CTCTAGGGAATC GACGGG

RNA-seq data are shown Log2 ratio, and positive and negative values of Log2 ratio are either up- or downregulated genes in the three pair-comparisons. No significant fold changes are indicated as '-'. Values of real-time RT-PCR analysis are fold changes of RNA transcripts, which were calculated by the $2^{-\Delta\Delta C_T}$ method (Livak & Schmittgen 2001) with actin as an internal control.

major gene effects on prolificacy in sheep. There might be unidentified different genes present in Han sheep that change the gene expression patterns of *BMPR1B*, *BMP15*, and *GDF9* and related genes or pathways affecting ovulation rate. So the whole ovaries of Han sheep (genotype BB and genotype ++) and Surabaya fur ewes (considered as the low-fecundity control) were used to screen DEGs or pathways by next-generation sequencing technology.

BMP/Smad signaling pathway components including the BMPs (BMP1, BMP2, BMP4, BMP6, BMP7, and BMP15; Paradis *et al.* 2009, Canty-Laird *et al.* 2010, Sugiura *et al.* 2010, Otsuka *et al.* 2011, Paulini & Melo 2011), the BMP receptors (BMPR1A, BMPR1B, and BMPR2) (9), intracellular transducers (Smad1, Smad5, and Smad4), and related molecules (GDF9 and TGF- β RI) (Knight & Glister 2006, Feary *et al.* 2007, Paradis *et al.* 2009) are now known to have effects on reproduction. Furthermore, different modes of cross talk between the BMP signaling pathway and the MAPK, phosphatidylinositol-3 kinase (PI3K)/Akt, Wnt, Hedgehog, and Notch pathways have also been investigated. And it is reported that the SLIT/ROBO pathway has functional roles in pivotal processes in ovary development by inhibiting cell migration and promoting apoptosis and subsequently determined fertility in adult life.

In this study, we successfully demonstrated that the next-generation sequencing technology could be a useful way for the gene expression profiling in the non-model species that is lacking genomic information. Next-generation sequencing method provides a low-cost, labor-saving, and rapid ways of transcriptome sequencing and characterization. Recently, *de novo* transcriptome analyses, that is, *de novo* assembly of short reads from mRNA without genome reference, have emerged. Several studies have reported the transcriptome sequencing of various non-model species using next-generation sequencing technologies (Gibbons *et al.* 2009, Wolf *et al.* 2010). Although many studies are based on the long-read sequence data using 454 pyrosequencing or used a hybrid approach, the *de novo* assembly of transcriptome employing short reads (Illumina or SOLiD) has also received extensive attention because of its relatively low cost. In this study, we presented a *de novo* assembly approach for the transcriptome for the non-model species using only short-read sequence data and showed a strategy for identifying sequences with different expression levels between different sheep samples (the genotype BB Han sheep, genotype ++ Han sheep, and Surabaya fur sheep). The strategy of differential gene expression analysis described here can be potentially employed in any species. Although currently we can only speculate on the functions of the candidate genes for the comparison of different sheep, further detailed analyses of our transcriptome data could contribute to the development of molecular studies and may

accelerate functional genomic studies of their biological implications. Therefore, we concluded that next-generation sequencing technologies will also provide new insights into the biological study of phenotypic variation in wild organisms.

Changes in gene expression between Small-tail Han sheep (Han sheep) and Surabaya fur sheep were detected by RNA-seq analysis. GO enrichment analysis results suggested that BP-related genes might play important roles in the difference of fecundity between BB and ++ Han sheep, and CC-related genes might play important roles in the difference between the Surabaya fur sheep and genotype BB or ++ Han sheep. Genes confirmed as differentially regulated in hyper-prolificacy included those involved in plasma membrane, ribonucleoprotein complex, and translation-related genes, which implied that there are some other genes associated with high prolificacy besides *FecB* gene. The real-time RT-PCR results showed that TGF β -associated genes were up-regulated in BB and ++ genotype, which implied that TGF β played important roles in high prolificacy of Small-tail Han sheep. BMPs belong to TGF β superfamily, and Bonnet *et al.* (2011) have proved that BMP15 was expressed in sheep granulosa cells and oocytes during early follicular development by laser capture microdissection and which played important role in follicular development and ovarian function (Otsuka *et al.* 2011, Paulini & Melo 2011). In our study, latent TGF β -binding protein 1 was highly upregulated in Small-tail Han sheep, especially in BB genotype, implying that it might play critical roles for high prolificacy. The function of this gene might be involved in the assembly, secretion, and targeting of TGF β 1 to its stored and/or activated sites and might play critical roles in controlling and directing the activity of TGF β 1. The results suggested that other members of TGF β superfamily were also involved in the high prolificacy besides BMP15.

Our data indicated that plenty of genes were differentially regulated in different sheep. As shown (Supplementary Table 1), most components of 40S (e.g. S10–S21) and 60S (e.g. L5–L15) ribosomal complex are differentially expressed, suggesting that protein synthesis activity could be a critical determinant for fecundity and prolificacy. It has been suggested that male fertility is affected by mitochondrial haplotypes (Yee *et al.* 2013). Among those DEGs, ATP synthase subunits, NADH dehydrogenase, and Diablo homolog are linked to mitochondrial and cellular metabolism (Wathes *et al.* 2012). Different signaling pathway molecules are also differentially expressed, such as Wnt-2b and β -catenin in regulating Wnt pathway (Mosimann *et al.* 2009), MAPK and MAPKK in MAPK pathway (Raabe & Rapp 2002), and GSPM1 in G-protein signaling (Blumer *et al.* 2008). In addition, proteins required for transport are also differentially expressed, such as Zinc transporter ZIP13, intraflagellar transport 122 homolog, long-chain fatty

acid transport protein 1, and protein transport protein Sec61 subunits.

Conclusions

In this study, with the use of next-generation RNA-seq technology, we have characterized the DEGs in Small-tail Han sheep and the Surabaya fur sheep in a genome-wide level. Our data highlight that multiple genes, such as latent TGF β -binding protein 1, are potentially critical for regulating sheep fecundity and prolificacy. Together, besides providing the key knowledge related to the high prolificacy of Small-tail Han sheep, our study will also be served as a very useful genetic resource for other relevant research community. Nonetheless, further studies are still needed to verify the physiological functions of candidate genes for high prolificacy through genetic engineering.

Materials and Methods

Samples

Totally, 108 Han ewes from the fine nucleus herd, bred by Ao-Te sheep breeding farm in Qingdao, Shandong province of China, were used for blood sampling to identify the FecB mutation in the *BMPT1B* gene. Among these ewes, three adult Han ewes with genotype BB and genotype ++ were regarded as two groups in high-fecundity Han sheep. Meanwhile, three adult Surabaya fur ewes (aged from 2.5 to 3 years) were killed for ovary sampling as the control low-fecundity group in the same condition. All animals were raised under the condition of free access to water and food in natural lighting. All experimental procedures were performed according to authorization granted by the Chinese Ministry of Agriculture.

All experimental ewes were treated with intravaginal sponges (40 mg; Chronogest, Intervet, Federal District, México) impregnated with fluorogesterone acetate during 10 days. And after 10 days, sponges were removed and pregnant mare serum gonadotropin (Ningbo Hormone Co., Ningbo, China) was injected i.m. at doses of 400 IU each sheep to synchronize estrus (Quintero-Elisea *et al.* 2011). After estrus was diagnosed, all ewes have reached spontaneous estrus after an estrous cycle and then animals were killed between 24 and 36 h after spontaneous estrus was detected. Whole ovaries were excised and samples were collected with better ovulation points on the surfaces of the ovaries. All samples were immediately snap-frozen in liquid nitrogen and stored at -70°C for total RNA extraction.

RNA extraction and RNA-seq

Total RNA was extracted from ovaries in three groups (BB Han ewes, ++ Han ewes, and Surabaya fur ewes) using TRIzol (Invitrogen, Inc.) according to the manufacturer's instruction. The extracted RNA was treated with RNase-free DNase I (Ambion, Inc., Austin, TX, USA) to remove any potential genomic DNA contamination.

After extracting the total RNA from the samples, mRNA was enriched using the oligo(dT) magnetic beads. Adding the fragmentation buffer, the mRNA was interrupted to short fragments (about 200 bp) and then the first-strand cDNA was synthesized by random hexamer primer using the mRNA fragments as templates. Buffer, dNTPs, RNase H, and DNA polymerase I were added to synthesize the second strand. The double-stranded cDNA was purified with QiaQuick PCR extraction kit and washed with elution buffer for end repair and single nucleotide A (adenine) addition. Finally, sequencing adaptors were ligated to the fragments. The required fragments were purified by agarose gel electrophoresis and enriched by PCR amplification. The library products were ready for sequencing analysis via Illumina HiSeq 2000.

Bioinformatics analysis

We applied a series of *de novo* assembly pipeline to these sequencing reads. *De novo* assembly was carried out by Inchworm nested in Trinity Software packages (Grabherr *et al.* 2011). It utilizes the K-mer graph method to assemble Illumina RNA-seq reads. Although it prefers strand-specific RNA-seq reads, Inchworm can also deal with the nonstrand-specific RNA-seq reads generated from the RNA-seq experiments. Low-coverage artifacts or redundancies from different tissues were removed by CD-HIT (Li & Godzik 2006) with an identity threshold of 95%. The detailed work flow is described in Fig. 1. To find new transcripts, we downloaded sheep genes from NCBI database and then compared all these transcripts to these sequences using BLASTN.

Expression profiling

The mapped read counts for each gene were normalized for RNA length and for the total read number in the lane according to RPKM. The RPKM method is able to eliminate the influence of different gene length and sequencing discrepancy on the calculation of gene expression. Therefore, the calculated gene expression can be directly used for comparing the difference of gene expression among samples (Mortazavi *et al.* 2008). For differential gene expression analysis with count data, the DESeq package in R was used (Anders & Huber 2010). We used FDR to determine the significance threshold of *P* value in multiple tests in the experiment.

GO enrichment analysis

GO is an international standardized gene functional classification system that offers a dynamic updated controlled vocabulary and a strictly defined concept to comprehensively describe properties of genes and their products in any organism. GO terms and annotations were obtained in our MySQL database, which was downloaded from the December 2011 release (Smedley *et al.* 2009). Hypergeometric test was applied to map all DEGs in GO terms and search significantly enriched GO terms in DEGs comparing to the genome background. In this work, this analysis was performed using FatiGO Software (Al-Shahrour *et al.* 2004).

The calculating formula is:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

where N is the number of all genes with GO annotation; n is the number of DEGs in N ; M is the number of all genes that are annotated to the certain GO terms; and m is the number of DEGs in M . The calculated P value goes through Bonferroni's correction, taking corrected P value ≤ 0.01 as a threshold. GO terms fulfilling this condition are defined as significantly enriched GO terms in DEGs. This kind of analysis is able to recognize the main biological functions that DEGs exercise.

Quantitative real-time PCR

To validate the array data, the expression of ten interest genes were confirmed using quantitative real-time PCR (qRT-PCR). Total RNA (0.5 μ g) was used to synthesize first-strand cDNA using PrimerScript RT reagent Kit (TaKaRa Biotech, Dalian, China; code: DRR037A). Each cDNA sample was diluted ten times in ddH₂O, and 1 μ l of this dilution was used as a template for qRT-PCR. The qRT-PCR reactions were performed in a 20 μ l volume containing 10 μ l 2 $\times$ SYBR Green Master Mix (TIANGEN Biotech, Dalian, China; code: FP204), 50 ng cDNA, and 400 nM of forward and reverse primers in a Roche HOLD CYCLE LightCycler 480 II. The amplification conditions were 95 °C for 15 min of initial stage, followed by 40 cycles of 95 °C for 10 s and 60 °C for 30 s. β -Actin was used as the internal control (F-primer: TCCGCAAAGACCTCTACG, R-primer: CCGAG-TACTTGCCTCA). The gene primers were listed in Table 3. The comparative Ct method was used to calculate the relative gene expression level across the samples. The relative expression level of each gene in one sample (Δ Ct) was calculated as follows: Ct target gene – Ct β -actin. The relative expression of each gene in two different samples ($\Delta\Delta$ Ct) was calculated as follows: Δ Ct (sample 1) – Δ Ct (sample 2). Values of real-time RT-PCR analysis are fold changes of RNA transcripts, which were calculated by the $2^{-\Delta\Delta$ Ct method (Livak & Schmittgen 2001).

Supplementary data

This is linked to the online version of the paper at <http://dx.doi.org/10.1530/REP-12-0507>.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This work was supported by a grant from The Major Science and Technology Project of New Variety Breeding of Genetically Modified Organisms (nos 2009ZX08008-004B and 2008ZX08008-003), the National High Technology Research

Development Program of China (863 Program no. 2008AA10Z140), the National Natural Science Foundation of China (no. 30571339), and the Innovation Research Foundation of CAAS (no. 2004-CAAS-1).

References

- Al-Shahrour F, Diaz-Uriarte R & Dopazo J 2004 FatiGO: a web tool for finding significant associations of gene ontology terms with groups of genes. *Bioinformatics* **20** 578–580. (doi:10.1093/bioinformatics/btg455)
- Anders S & Huber W 2010 Differential expression analysis for sequence count data. *Genome Biology* **11** R106. (doi:10.1186/gb-2010-11-10-r106)
- Blumer JB, Lord K, Saunders TL, Pacchioni A, Black C, Lazartigues E, Varner KJ, Gettys TW & Lanier SM 2008 Activator of G protein signaling 3 null mice: I. Unexpected alterations in metabolic and cardiovascular function. *Endocrinology* **149** 3842–3849. (doi:10.1210/en.2008-0050)
- Bongiorno S, Chillemi G, Prosperini G, Bueno S, Valentini A & Pariset L 2009 A tool for sheep product quality: custom microarrays from public databases. *Nutrients* **1** 235–250. (doi:10.3390/nu1020235)
- Bonnet A, Bevilacqua C, Benne F, Bodin L, Cotinot C, Liaubet L, Sancristobal M, Sarry J, Terenina E, Martin P *et al.* 2011 Transcriptome profiling of sheep granulosa cells and oocytes during early follicular development obtained by laser capture microdissection. *BMC Genomics* **12** 417–431. (doi:10.1186/1471-2164-12-417)
- Canty-Laird E, Carré GA, Mandon-Pépin B, Kadler KE & Fabre S 2010 First evidence of bone morphogenetic protein 1 expression and activity in sheep ovarian follicles. *Biology of Reproduction* **83** 138–146. (doi:10.1095/biolreprod.109.082115)
- Feary ES, Juengel JL, Smith P, French MC, O'Connell AR, Lawrence SB, Galloway SM, Davis GH & McNatty KP 2007 Patterns of expression of messenger RNAs encoding GDF9, BMP15, TGFBR1, BMPR1B, and BMPR2 during follicular development and characterization of ovarian follicular populations in ewes carrying the Woodlands FecX2W mutation. *Biology of Reproduction* **77** 990–998. (doi:10.1095/biolreprod.107.062752)
- Gibbons JG, Janson EM, Hittinger CT, Johnston M, Abbot P & Rokas A 2009 Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics. *Molecular Biology and Evolution* **26** 2731–2744. (doi:10.1093/molbev/msp188)
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al.* 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29** 644–652. (doi:10.1038/nbt.1883)
- Knight PG & Glistler C 2006 TGF- β superfamily members and ovarian follicle development. *Reproduction* **132** 191–206. (doi:10.1530/rep.1.01074)
- Li W & Godzik A 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22** 1658–1659. (doi:10.1093/bioinformatics/btl158)
- Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y *et al.* 2009 The sequence and *de novo* assembly of the giant panda genome. *Nature* **463** 311–317. (doi:10.1038/nature08696)
- Livak KJ & Schmittgen TD 2001 Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta$ Ct method. *Methods* **25** 402–408. (doi:10.1006/meth.2001.1262)
- Mardis ER 2008 The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24** 133–141. (doi:10.1016/j.tig.2007.12.007)
- Mortazavi A, Williams BA, McCue K, Schaeffer L & Wold B 2008 Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5** 621–628. (doi:10.1038/nmeth.1226)
- Mosimann C, Hausmann G & Basler K 2009 β -Catenin hits chromatin: regulation of Wnt target gene activation. *Nature Reviews. Molecular Cell Biology* **10** 276–286. (doi:10.1038/nrm2654)
- Otsuka F, McTavish KJ & Shimasaki S 2011 Integral role of GDF-9 and BMP-15 in ovarian function. *Molecular Reproduction and Development* **78** 9–21. (doi:10.1002/mrd.21265)
- Paradis F, Novak S, Murdoch GK, Dyck MK, Dixon WT & Foxcroft GR 2009 Temporal regulation of BMP2, BMP6, BMP15, GDF9, BMPR1A,

- BMPR1B, BMPR2 and TGFBR1 mRNA expression in the oocyte, granulosa and theca cells of developing preovulatory follicles in the pig. *Reproduction* **138** 115–129. (doi:10.1530/REP-08-0538)
- Paulini F & Melo EO** 2011 The role of oocyte-secreted factors GDF9 and BMP15 in follicular development and oogenesis. *Reproduction in Domestic Animals* **46** 354–361. (doi:10.1111/j.1439-0531.2010.01739.x)
- Peñagaricano F, Zorrilla P, Naya H, Robello C & Urioste JI** 2012 Gene expression analysis identifies new candidate genes associated with the development of black skin spots in Corriedale sheep. *Journal of Applied Genetics* **53** 99–106. (doi:10.1007/s13353-011-0066-9)
- Quintero-Elisea JA, Macias-Cruz U, Alvarez-Valenzuela FD, Correa-Calderon A, Gonzalez-Reyna A, Lucero-Magaña FA, Soto-Navarro SA & Avendaño-Reyes L** 2011 The effects of time and dose of pregnant mare serum gonadotropin (PMSG) on reproductive efficiency in hair sheep ewes. *Tropical Animal Health and Production* **43** 1567–1573. (doi:10.1007/s11250-011-9843-z)
- Raabe T & Rapp UR** 2002 KSR – a regulator and scaffold protein of the MAPK pathway. *Science's STKE* **136** 28. (doi:10.1126/stke.2002.136.pe28)
- Shimasaki S, Zachow RJ, Li D, Kim H, Iemura S, Ueno N, Sampath K, Chang RJ & Erickson GF** 1999 A functional bone morphogenetic protein system in the ovary. *PNAS* **96** 7282–7287. (doi:10.1073/pnas.96.13.7282)
- Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G & Kasprzyk A** 2009 BioMart-biological queries made easy. *BMC Genomics* **10** 22. (doi:10.1186/1471-2164-10-22)
- Sugiura K, Su YQ, Li Q, Wigglesworth K, Matzuk MM & Eppig JJ** 2010 Estrogen promotes the development of mouse cumulus cells in coordination with oocyte-derived GDF9 and BMP15. *Molecular Endocrinology* **24** 2303–2314. (doi:10.1210/me.2010-0260)
- Wathes DC, Clempson AM & Pollott GE** 2012 Associations between lipid metabolism and fertility in the dairy cow. *Reproduction, Fertility, and Development* **25** 48–61. (doi:10.1071/RD12272)
- Wolf JB, Bayer T, Haubold B, Schilhabel M, Rosenstiel P & Tautz D** 2010 Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Molecular Ecology* **19** (Suppl 1) 162–175. (doi:10.1111/j.1365-294X.2009.04471.x)
- Yee WK, Sutton KL & Dowling DK** 2013 *In vivo* male fertility is affected by naturally occurring mitochondrial haplotypes. *Current Biology* **3** R55–R56. (doi:10.1016/j.cub.2012.12.002)

Received 18 December 2012

First decision 11 January 2013

Revised manuscript received 12 March 2013

Accepted 10 April 2013