Transcriptomic Comparative Analysis of Two Breeds of Mongolian Sheep at 16-day Embryos

Hong Su, Lu Chen, Wenrui Guo, Daqing Wang, Aolei Dou, Jie Su, Yanyan Yang, Ying Tian, Tingyi He, Caiyun Wang, Chenguang Du, Haijun Li, Xihe Li, Guifang Cao, Yongli Song, and Fuxiang Bao

1College of Veterinary Medicine, Inner Mongolia Agricultural University, China
2Animal Embryo and Developmental Engineering Key Laboratory of Higher Education Institutions of Inner Mongolia Autonomous Region, China
3Inner Mongolia Autonomous Region Key Laboratory of Basic Veterinary Medicine, China
4Institute of Animal Husbandry, Inner Mongolia Academy of Agriculture and Animal Husbandry
5College of Life Sciences, Inner Mongolia University, China

Hong Su, Lu Chen, and Wenrui Guo contributed equally to this work.

ABSTRACT

Early embryonic development determines fetal and adult development. However, little is known about the specific development of the early Mongolian sheep embryo. We aim to analyze the differentially expressed genes (DEGs) between 16-day embryos of two breeds of Mongolian sheep, namely, Hulunbuir short-tailed sheep and Ujumqin sheep, by mRNA sequencing. A total of 2152 DEGs were identified. Significantly DEGs include WNT5A, WNT2, AXIN2, BMP5, TGFB2, and SMAD9, which are involved in multiple signaling pathways, and genes involved in digestion and absorption and cellular differentiation, such as DPP4, ACE2, PLPP1, CD22, CD44, TEC, and JAK1. In the pathway analysis, the development of Hulunbuir short-tailed sheep 16-day embryo H(E16) highlighted pathways primarily associated with cellular interactions and organogenesis. These pathways are critical for transcriptional control during early embryonic development, especially organ formation. Genes that sophisticate organs and body mechanisms, and genes associated with protein and fat digestion, hematopoietic stem cell lineage and osteoblast differentiation, are enriched during the development of the Ujumqin sheep 16-day embryo U(E16). Our work provides important complementary reference data for the study of early embryonic development in sheep.

INTRODUCTION

The domestication of sheep (Ovis aries), one of the first domesticated animals, always accompanies human migration and social development (Chessa et al., 2009; Lv et al., 2015). The inner Mongolia region of China is the main breeding area for meat sheep in China, with about 17 taxa collectively known as Mongolian sheep. Among them, Ujumqin sheep and Hulunbuir short-tailed sheep are well-known and profitable breeds distributed in the Hulunbuir Grassland and Xilingol grassland, respectively (Li et al., 2018; Zhi et al., 2018). Ujumqin sheep is a multi-vertebral breed with a fat tail (Li et al., 2018). The tail of Hulunbuir short-tailed sheep is shorter than those of other breeds, and this trait is genetically stable. In addition, Hulunbuir short-tailed sheep have more intermuscular fat, which is evenly distributed, and it is a typical kind of short-tailed sheep. Due to the short-tailed phenotype, Hulunbuir short-tailed sheep have been applied to artificial and natural selection to increase meat production (Zhi et al., 2018). Li et al. (2018) showed by whole-genome resequencing (WGRS) that the multi-vertebral characteristic in Ujumqin sheep was associated with...
mutations in CAMK1D, MLLT6, FOXJ1, ADAMTS32 and GUSB. Zhu et al. (2021a) used isobaric tags for relative and absolute quantitation (iTRAQ) to reveal candidate proteins of fat deposition in Chinese native sheep with different tails and identified APOA2, GALK1, ADIPOQ and NDUFS4 genes. They also identified SPAG17, Thx15, VRTN, NPC2, BMP2, and PDGFD as the most promising candidates for tail-type traits through a genome-wide association study of genes related to tail fat deposition in Chinese sheep (Zhu et al., 2021b). Most studies on sheep tail morphology focus on tail fat deposition. For example, Xu et al. (2017) conducted a genome-wide association analysis to determine the genetic basis of fat deposition in the tail of two types of Mongolian sheep, namely, small-tailed Han sheep and large-tailed Han sheep. Moradi et al. (2012) identified candidate regions associated with fat deposition by genome-wide selection scan and found that increased homozygosity of the region on chromosomes 5 and X favored fat-tailed sheep breeds and increased homozygosity of the region on chromosome 7 favored thin-tailed sheep. Luo et al. (2021) conducted whole-genome sequencing (WGS) for three types of Mongolian sheep and found that GLIS1, LOC101117953, PDGFD and T might be associated with fat deposition in the sheep tail and tail morphology. They also suggested that GLIS1 may play a key role in germ layer differentiation in sheep embryos. Few studies focus on the formation of sheep tail vertebrae. Zhi et al. (2018) identified nine candidate genes associated with the short-tailed phenotype by WGS of Hulunbuir short-tailed sheep. Among these genes, the T gene was the strongest candidate because it was associated with vertebrae development. Guillomot et al. (2004) concluded in Staging of ovine embryos and expression of the T-box genes Brachyury and Eomesodermin around gastrulation that the expression level of the T gene was the highest in 16-day sheep embryo (E16). T gene significantly affects germ layer development. Pennimpede et al. (2012) found that in vivo knockdown of T gene resulted in skeletal defects and tail degeneration. The studies on both Ujumqin sheep and Hulunbuir short-tailed sheep are expected to improve breeding technology and increase economic benefits. We found studies on embryo development in sheep are rare. We chose Hulunbuir short-tailed sheep 16-day embryo H(E16) and Ujumqin sheep 16-day embryo U (E16) for reference transcriptome sequencing, which will help us to analyze the different developmental states of the two Mongolian sheep embryos during the same developmental period for better understanding of embryo development and scientific regulation of breeding production. It is expected that this paper can help to improve sheep welfare and increase economic efficiency.

MATERIALS AND METHODS

Samples

Hulunbuir short-tailed sheep (Fig. 1A) were from Ewenki Autonomous Banner, Hulunbuir City, Inner Mongolia Autonomous Region, China. Ujumqin sheep (Fig. 1B) were from East Ujumqin Banner, Xilinhot City, Inner Mongolia Autonomous Region, China. The two regions were about 1000 km apart. To ensure that the embryos were obtained at the same time, we transported the two types of sheep to the same ranch (Wuchuan County Breeding Sheep Ranch, Hohhot City). After one month, when the ewes and rams adapted to the local climate, we started the experiment.

Embryo sample collection

Twenty Ujumqin and 40 Hulunbuir healthy sheep, which were three years old, were treated with progesterone vaginal sponge embolization for 12 days. PMSG (330 IU/sheep) was intramuscularly injected before thrombectomy, and estrous ewes were mated after 36 hours. Embryos were collected at pregnancy 16 days (Fig. 1C, D). Embryos were washed three times using cold PBS and immediately transferred to tissue storage solution (Miltenyi, 130-100-008) with 3% penicillin-streptomycin at room temperature.

mRNA library construction and sequencing

Total RNA was isolated and purified using TRIZol reagent (Invitrogen, Carlsbad, CA, USA) following manufacturer’s procedures. RNA in each sample was
quantified, and the purity was evaluated using NanoDrop ND-1000 (NanoDrop, Wilmington, DE, USA). The RNA integrity was assessed by Bioanalyzer 2100 (Agilent, CA, USA) with RIN number >7.0 and confirmed by electrophoresis with denaturing agarose gel. Poly (A) RNA was purified from 1 μg total RNA using Dynabeads Oligo (dT) 25-61005 (Thermo Fisher, CA, USA) by two rounds of purification. Then poly (A) RNA was fragmented into small pieces using Magnesium RNA fragmentation module (NEB, cat.e6150, USA) by being kept under 94°C for 5-7 min. Then the cleaved RNA fragments were reverse-transcribed to create cDNAs by Super Script™ II Reverse Transcriptase (Invitrogen, cat. 18966649, USA), which were subsequently used to synthesize U-labeled second-stranded DNAs with E. coli DNA polymerase I (NEB, cat.m0209, USA), RNase H (NEB, cat.m0297, USA) and dUTP Solution (Thermo Fisher, cat.R0133, USA). An A-base is then added to the blunt ends of each strand, preparing them for ligation to the indexed adapters. Each adapter contains a T-base overhang for ligating the adapter to the A-tailed fragmented DNA. Single- or dual-index adapter contains a T-base overhang for ligating the adapter preparing them for ligation to the indexed adapters. Each sample was amplified with PCR under the following conditions: initial denaturation at 95°C for 3 min; eight cycles of denaturation at 98°C for 15 s, annealing at 60°C for 15 s, and extension at 72°C for 30 s; final extension at 72°C for 5 min. The average insert size for the final cDNA library was 300±50 bp. At last, we performed the 2×150bp paired-end reads/sample was kept. Moreover, an average of 50.7 million reads/sample was generated. After quality control, an average of 52.41 million reads/sample was mapped against the sheep reference genome (Ovis aries, v.3.1, Ensembl) according to manufacturer’s directions. Reverse transcription was performed using PrimeScript RT reagent kit with gDNA Eraser (RR047A, Takara). Real-time PCR was performed using the ChamQ Universal SYBR qPCR master mix on a CFX96 PCR system (Bio-rad). The primer sequences used are shown in Supplementary Table I. The mapped reads of each sample were assembled using StringTie (http://ccb.jhu.edu/software/stringtie/, version: stringtie-1.3.4d.Linux_x86_64) with default parameters (command line: ~stringtie -p 4 -G genome.gtf -o output.gtf -l sample input.bam). Then, all transcriptional activities from all samples were merged to reconstruct a comprehensive transcriptome using gff compare software (http://ccb.jhu.edu/software/stringtie/gffcompare.shtml, version:gffcompare-0.9.8.Linux_x86_64) (Pertea et al., 2015). After the final transcriptome was generated, StringTie and ballgown (http://www.bioconductor.org/packages/release/bioc/html/ballgown.html) were used to estimate the expression levels of all transcripts and expression level of mRNAs by calculating FPKM (FPKM = [total_exon_length / mapped_reads (millions) × exon_length(kB)]). The differentially expressed mRNAs with fold change>2 or fold change < 0.5 and P value<0.05 were selected by R package edgeR (https://bioconductor.org/packages/release/bioc/html/edgeR.html) or DESeq2 (http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html) (Robinson et al., 2010), and then GO ontology (Young et al., 2010) and KEGG enrichment analysis (Kanehisa et al., 2008) was performed for them.

Data analysis

The sequence data were deposited in the sequence read archive of the NCBI databases under bioproject number GSE186602 (https://www.ncbi.nlm.nih.gov/geo/info/submitting.html, published on October 22, 2023). Cutadapt software (https://cutadapt.readthedocs.io/en/stable/version:cutadapt-1.9) was used to remove the reads that contained adaptor contamination (command line: ~cutadapt -a ADAPT1 -a ADAPT2 -o out1.fastq -p out2.fastq in 1.fastq in 2.fastq -O 5 -m 100) (Beekman et al., 2018; Wang et al., 2012). After removing the low-quality bases and undetermined bases, we used HISAT2 software (https://daehwankimlab.github.io/hisat2/, version:hisat2-2.0.4.0) to map reads to the genome (for example: Homo sapiens Ensembl v96) (command line: ~hisat2 -l 1 R1.fastq.gz -2 R1.fastq.gz -S sample_mapped.sam) (Kim et al., 2015). The reads that contained undetermined bases, we used HISAT2 software (https://daehwankimlab.github.io/hisat2/, version:hisat2-2.0.4.0) to map reads to the genome (for example: Homo sapiens Ensembl v96) (command line: hisat2 -l 1 R1.fastq.gz -2 R1.fastq.gz -S sample_mapped.sam) (Kim et al., 2015). The

RESULTS

RNA-sequecing data

From the whole transcriptome of conceptus samples (n= 3 embryos of Huulnbuir sheep at E16 using RNA fast 200 (220011, Shanghai, China) according to manufacturer’s directions. Reverse transcription was performed using PrimeScript RT reagent kit with gDNA Eraser (RR047A, Takara). Real-time PCR was performed using the ChamQ Universal SYBR qPCR master mix on a CFX96 PCR system (Bio-rad). The differentially expressed mRNAs with fold change>2 or fold change < 0.5 and P value<0.05 were selected by R package edgeR (https://bioconductor.org/packages/release/bioc/html/edgeR.html) or DESeq2 (http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html) (Robinson et al., 2010), and then GO ontology (Young et al., 2010) and KEGG enrichment analysis (Kanehisa et al., 2008) was performed for them.

qRT-PCR

RNA was isolated from Ujumqin and Huulnbuir sheep at E16 using RNA fast 200 (220011, Shanghai, China) according to manufacturer’s directions. Reverse transcription was performed using PrimeScript RT reagent kit with gDNA Eraser (RR047A, Takara). Real-time PCR was performed using the ChamQ Universal SYBR qPCR master mix on a CFX96 PCR system (Bio-rad). The differentially expressed mRNAs with fold change>2 or fold change < 0.5 and P value<0.05 were selected by R package edgeR (https://bioconductor.org/packages/release/bioc/html/edgeR.html) or DESeq2 (http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html) (Robinson et al., 2010), and then GO ontology (Young et al., 2010) and KEGG enrichment analysis (Kanehisa et al., 2008) was performed for them.
DEGs between H(E16) and U(E16)

A total of 2152 DEGs between H(E16) and U(E16) were identified, with 922 genes up-regulated in H(E16) and 1230 genes up-regulated in U(E16) (Fig. 2, Table II).

**GO functional analyses**

The GO enrichment analysis of the 992 upregulated genes in H(E16) revealed that these genes could be enriched to 877 GO terms, including 659 biological process (75.2%), 138 molecular function (15.7%) and 80 cellular component (9.1%) (Fig. 3A). The enriched GO terms highlighted several crucial factors, such as heart development, protein binding, axon guidance, outer ear morphogenesis, digestive tract morphogenesis, limb bud formation, and pharyngeal system development, for biological development.

GO annotation of the 1230 upregulated genes in U(E16) revealed that these genes could be annotated to 590 GO terms, including 403 biological process (68.3%), 121 molecular function (20.5%) and 66 cellular component (11.2%) (Fig. 3B). The enriched GO terms highlighted several crucial factors, such as extracellular space, apical plasma membrane, membrane, cell surface, platelet activation, and erythrocyte differentiation, for biological development. We used ggplot 2 to present GO enrichment analysis results in the form of bubble plots.

**KEGG functional analyses**

KEGG enrichment analysis of the 922 up-regulated genes in H(E16) revealed that these genes were enriched in a total of 43 signaling pathways (Fig. 4A), including axon guidance, Hippo signaling pathway, signaling pathways regulating pluripotency of stem cells, TGF-beta signaling pathway, cardiac muscle contraction, MAPK signaling pathway, Wnt signaling pathway, etc.

KEGG enrichment analysis of the 1230 upregulated genes in U(E16) revealed that these genes were enriched in 68 signaling pathways (Fig. 4B), including protein digestion and absorption, fat digestion and absorption, PPAR signaling pathway, hematopoietic cell lineage, arachidonic acid metabolism, PI3K-Akt signaling pathway, etc.

**qRT-PCR analysis**

Ten genes were randomly selected from all mRNA sequencing data for qRT-PCR reaction to verify the accuracy of mRNA sequencing results (Fig. 5). The qRT-PCR results showed that the expression of mRNAs was different, and the expression trend was consistent with that obtained by sequencing, indicating that the mRNA sequencing results were reliable.

---

**Table I. Primer sequences for qRT-PCR of sheep genes.**

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Gene ID</th>
<th>Primer (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>443005</td>
<td>F GCGCAGCTCAAGGCAGAGAAC R CAGTACTCCGACCCAGCACTACG</td>
</tr>
<tr>
<td>ANXA1</td>
<td>101123269</td>
<td>F TGCTAAGGGTGACCGGATGAGG R GCTCCTGCGTTGAAGATGGTACG</td>
</tr>
<tr>
<td>AHSG</td>
<td>443392</td>
<td>F CCAGGGGTCTGCTGGAAGGTTTG R AGGTGGACTCTCTGTTGGGAAGTACG</td>
</tr>
<tr>
<td>CDKN1A</td>
<td>100302083</td>
<td>F ACCACTGGACCCGGTTCGCTTCGTC R GATCACGGCTGCTGTTGGGAGTACG</td>
</tr>
<tr>
<td>FST</td>
<td>443323</td>
<td>F TCAAGGTGGATGATTTTCAACGG R GCTTCTCCTGCTGTTGGGAGTACG</td>
</tr>
<tr>
<td>PRNP</td>
<td>493887</td>
<td>F GTGGCTATCACTGGCTGGAAGGTTTG R GGGAGTGTCTCTCCTGTTGGGAGTACG</td>
</tr>
<tr>
<td>YAP1</td>
<td>100913160</td>
<td>F AGAGGGCTGGCCTCGCATACAAA R CCAACAGGTCGCCGGCCTACG</td>
</tr>
<tr>
<td>TBXT</td>
<td>10114280</td>
<td>F CACCAAGAGGGCAGGGAGATG R CGACTTCTGCCAGGCTGTTGCC</td>
</tr>
<tr>
<td>NR2F1</td>
<td>10121774</td>
<td>F TCTGTCGCTTTGAGTTAGTAGA R GAGGACTGATTGACGATGTAAG</td>
</tr>
<tr>
<td>HOXB9</td>
<td>10114883</td>
<td>F CGAAGGAAGCAGGACGAGGAGATG R CAGCAGCTGATTGTTTGAGGAG</td>
</tr>
<tr>
<td>SOX2</td>
<td>10110563</td>
<td>F GATCAGCTATGCTCCTCCCG GTACAGAAGTCTGTCGCTG</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** The horizontal coordinate, log2(FC), represents the fold change of expression of genes in different samples. The vertical coordinate, -log10(pVal), represents the statistical significance of the difference in gene expression. Genes down-regulated in H(E16) are described as up-regulated in U(E16).
Transcriptomic Comparative Analysis of Two Breeds of Mongolian Sheep

Fig. 3. Top 20 GO terms of H(E16) (A); U(E16) (B). The horizontal coordinate Rich Factor represents the number of differential genes in the GO/total number of genes in the GO; A larger Rich factor indicates a higher degree of GO enrichment. The vertical coordinate is GO term, i.e., GO functional annotation. In the scatter plot, the size of the dots represents the number of differential genes, and the color of the dots represents the $P$-value of the enrichment analysis, i.e., the significance of the enrichment, with a $P$-value smaller than or equal to 0.05 indicating significant enrichment. The GO enrichment analysis scatter plot is based on the significance ($P$-value) of the enrichment of the top 20 GO terms.

Table II. Top 10 upregulated genes in H(E16) and top 10 upregulated genes in U(E16).

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene name</th>
<th>Description</th>
<th>log2 (FC)</th>
<th>pVal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H(E16)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSOARG00000002632</td>
<td>ENSOARG00000002632</td>
<td>Integrin subunit alpha 7</td>
<td>6.05</td>
<td>6.9x10^{-22}</td>
</tr>
<tr>
<td>ENSOARG00000011427</td>
<td>ITGA7</td>
<td>Coiled-coil domain containing 66</td>
<td>2.85</td>
<td>1.6x10^{-11}</td>
</tr>
<tr>
<td>ENSOARG00000015275</td>
<td>CCDC66</td>
<td></td>
<td>3.08</td>
<td>2.5x10^{-10}</td>
</tr>
<tr>
<td>ENSOARG00000026938</td>
<td>ENSOARG00000026938</td>
<td></td>
<td>13.09</td>
<td>1.5x10^{-9}</td>
</tr>
<tr>
<td>ENSOARG00000069666</td>
<td>UNC5B</td>
<td>Unc-45 myosin chaperone B</td>
<td>2.73</td>
<td>6.0x10^{-9}</td>
</tr>
<tr>
<td>ENSOARG0000004153</td>
<td>NKK2-5</td>
<td>NK2 homeobox 5</td>
<td>3.05</td>
<td>8.5x10^{-9}</td>
</tr>
<tr>
<td>ENSOARG00000008405</td>
<td>CASQ1</td>
<td>Calsequestrin 1</td>
<td>2.54</td>
<td>4.6x10^{-4}</td>
</tr>
<tr>
<td>ENSOARG00000017490</td>
<td>THBS4</td>
<td></td>
<td>2.22</td>
<td>1x10^{-7}</td>
</tr>
<tr>
<td>ENSOARG00000009361</td>
<td>ENSOARG00000009361</td>
<td></td>
<td>3.20</td>
<td>1.2x10^{-7}</td>
</tr>
<tr>
<td>ENSOARG00000012682</td>
<td>HRC</td>
<td>Histidine rich calcium binding protein</td>
<td>2.24</td>
<td>2.1x10^{-7}</td>
</tr>
<tr>
<td><strong>U(E16)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSOARG00000011224</td>
<td>ENSOARG00000011224</td>
<td>Meprin A subunit alpha</td>
<td>-13.20</td>
<td>1.8x10^{-17}</td>
</tr>
<tr>
<td>ENSOARG00000011165</td>
<td>HEMGN</td>
<td>Hemogen</td>
<td>-5.02</td>
<td>2.4x10^{-14}</td>
</tr>
<tr>
<td>ENSOARG00000001867</td>
<td>ENSOARG00000001867</td>
<td>Feline leukemia virus subgroup C receptor-related protein 2-like</td>
<td>-13.47</td>
<td>3.6x10^{-13}</td>
</tr>
<tr>
<td>ENSOARG00000010944</td>
<td>ENSOARG00000010944</td>
<td>Hemoglobin subunit alpha-1/2</td>
<td>-4.04</td>
<td>5.8x10^{-13}</td>
</tr>
<tr>
<td>ENSOARG00000006087</td>
<td>ENSOARG00000006087</td>
<td>Pancreatic trypsin inhibitor</td>
<td>-3.84</td>
<td>6X10^{-13}</td>
</tr>
<tr>
<td>ENSOARG00000015872</td>
<td>ENSOARG00000015872</td>
<td>Ubiquitin-conjugating enzyme E2 D3</td>
<td>-2.89</td>
<td>1.3X10^{-12}</td>
</tr>
<tr>
<td>ENSOARG0000003018</td>
<td>S100G</td>
<td></td>
<td>-5.54</td>
<td>2X10^{-12}</td>
</tr>
<tr>
<td>ENSOARG00000010469</td>
<td>ATP6V0A4</td>
<td>ATPase H+ transporting V0 subunit a4</td>
<td>-5.66</td>
<td>6.3X10^{-12}</td>
</tr>
<tr>
<td>ENSOARG00000026828</td>
<td>ENSOARG00000026828</td>
<td></td>
<td>-16.27</td>
<td>7.7X10^{-12}</td>
</tr>
<tr>
<td>ENSOARG00000007169</td>
<td>KLF4</td>
<td>Kruppel like factor 4</td>
<td>-3.79</td>
<td>1.9X10^{-4}</td>
</tr>
</tbody>
</table>
**DISCUSSION**

For ethical reasons, many questions about the human embryo cannot be directly investigated, and no animal model can truly reproduce human embryonic development. Pregnant sheep have been widely used to study maternal-fetal interactions, and a thorough understanding of the developmental events of the sheep embryo is necessary. In addition, studies of early embryonic development in sheep can help us to better regulate sheep gestation and reproduction, which will also contribute to economic development. *Sanchez et al.* (2021) showed that in 6-day sheep embryos produced in vivo, miRNAs regulate genes associated with the cell cycle and differentiation of stem cells and embryonic cells. The retrieved miRNAs also regulate the expression of genes involved in cellular signaling system; 30, Regulation of actin cytoskeleton; 31, Human cytomegalovirus infection; 32, Phenylalanine metabolism; 33, Platinum drug resistance; 34, Mineral absorption; 35, NF-kappa B signaling pathway; 36, Human immunodeficiency virus 1 infection; 37, Fructose and mannose metabolism; 38, Riboflavin metabolism; 39, Drug metabolism - cytochrome P450; 40, Glutathione metabolism; 41, PI3K-Akt signaling pathway; 42, Adipocytokine signaling pathway; 43, Glioma; 44, Hepatitis B; 45, Transcriptional misregulation in cancer; 46, Glycolysis/ Gluconeogenesis; 47, AMPK signaling pathway; 48, Prolactin signaling pathway; 49, Tuberculosis; 50, Toll-like receptor signaling pathway; 51, Rheumatoid arthritis; 52, Porphyrin and chlorophyll metabolism; 53, Steroid hormone biosynthesis; 54, C-type lectin receptor signaling pathway; 55, Apoptosis; 56, Drug metabolism other enzymes; 57, Amyloidosis; 58, Jak-STAT signaling pathway; 59, Hepatitis C; 60, Endocrine resistance; 61, T cell receptor signaling pathway; 62, Bacterial invasion of epithelial cells; 63, Leukocyte transendothelial migration; 64, Viral carcinogenesis; 65, Amphetamine addiction; 66, Melanoma; 67, Natural killer cell mediated cytotoxicity; 68, Metabolism of xenobiotics by cytochrome P450. Green nodes represent specific pathways.
signaling pathways, such as MAPK, Wnt, TGF-beta, p53, and Toll-like receptors (Sanchez et al., 2021). These results are partially consistent with our data. By KEGG enrichment analysis of up-regulated genes in H(E16), we found that these genes could be enriched in signaling pathways regulating pluripotency of stem cells, TGF-beta signaling pathway, MAPK signaling pathway, and Wnt signaling pathway. In addition, we found that upregulated genes in H(E16) could also be enriched in axon guidance, Hippo signaling pathway, and cardiac muscle contraction, suggesting that the neurological and cardiac organ development of H(E16) is more prominent than that of 6-day sheep embryo.

GO analysis showed that H(E16) had more developing organs, such as heart, nerves, ear, digestive tract, limb buds, and pharynx, compared to U(E16) at the same age, which was mainly undergoing some physiological processes related to cell growth and hematopoietic cell differentiation. Based on the KEGG enrichment analysis of H(E16) and U(E16), we found that upregulated genes in U(E16) were enriched in protein digestion and absorption, fat digestion and absorption, PPAR signaling pathway, hematopoietic cell lineage, and PI3K-Akt signaling pathway, suggesting that U(E16) may have completed development of some organs while H(E16) was still in the state of organ development. This finding is consistent with the result of GO enrichment analysis, indicating that U(E16) developed faster than H(E16), which is a surprising finding. Clark et al. (2017) provided a sheep gene expression profiling dataset that extended the RNA-seq dataset to include data of adult sheep tissues such as the liver, spleen, ovary, testis, kidney, muscle, thymus, and left ventricle, and to include data during early development (blastocyst), early development (day 23), early development (day 35), and maternal development (days 23, 35 and 100 days). Their work provides references for sheep research. However, until now, there have been no relevant data for 16-day sheep, and our study fills this research gap.

Above all, for H(E16), we focus more on signal pathways related to organogenesis, such as axon guidance, Hippo signaling pathway, signaling pathways regulating pluripotency of stem cells, TGF-beta signaling pathway, cardiac muscle contraction, MAPK signaling pathway, and Wnt signaling pathway. For U(E16), we focus more on protein digestion and absorption, fat digestion and absorption, Hematopoietic cell lineage, osteoclast differentiation, etc.

H(E16)

As the nervous system develops, newly differentiated neurons need to extend their axons toward synaptic targets in order to form functional neural circuits (Stoeckli, 2018). During embryonic development, various kinds of guidance information influence the direction of axon extension. A highly dynamic and sensitive structure located at the end of the axon the growth cone is responsible for interpreting guidance information. The binding of specific guidance molecules to receptors at the growth cone activates intracellular signaling pathways, and the combined action of several extracellular signals regulates the activity of the cytoskeleton, allowing axons to extend in the correct direction (de Wit and Verhaagen, 2003). In our study, upregulated genes in H(E16) were enriched in axon guidance. These genes included members of the semaphorin family that can influence processes such as axon bundling and branching and synapse formation, namely, SEMA3A, SEMA3G, SEMA4F and SEMA6C and included plxins in the semaphorin receptor family, namely PLXNA2, PLXNA3 and PLXNB1. In the axon guidance pathway, the Eph receptor and ligand families are involved in the regulation of embryonic neural stem cells (Moskowitz and Lo, 2003). Our data suggested that EPHB4, EPHA4, EPHB3, EPHB2 and EPHA7 were enriched. WNT5A was also enriched in the axon guidance pathway. The involvement of WNT5A, as an important ligand in the Wnt signaling pathway, in axon attraction connects the axon guidance pathway and the Wnt signaling pathway (Salinas, 2012). The proteins secreted by the Wnt family bind to different receptors and mediate a variety of signaling pathways to regulate the biological functions of cells, including cell polarization, cell proliferation and differentiation (Wodarz and Nusse, 1998; van Amerongen and Nusse, 2009). In our study, WNT2, AXIN2, WIF1 and PORCN (Boonekamp et al., 2021; Onizuka et al., 2012; Tribulo et al., 2018; Barrott et al., 2011) involved in the canonical Wnt signaling pathway and membrane receptors WNT5A and ROR2 (Yamada et al., 2010) involved in the non-canonical Wnt signaling pathway (PCP) were enriched. In fact, the WNT5A, WNT2, and AXIN2 were also enriched in the Hippo signaling pathway, which is an evolutionarily conserved signaling pathway that controls the size of organs of species from flies to humans (Zhao et al., 2011). We also identified enriched DLG4, the repressor in the Hippo signaling pathway, and enriched transcription factors ID1 and SOX2. The Hippo signaling pathway, as an important pathway regulating organogenesis during embryonic development, is not only associated with the Wnt signaling pathway but also interacts with the TGF-beta signaling pathway. BMP5 and TGFβ2, as important ligands in the TGF-beta signaling pathway, affect osteogenic differentiation and apoptosis through the BMP signaling pathway and canonical TGF-beta signaling pathway, respectively (Hata and Chen,
In our study, enriched transcriptional factor SMAD9 in the canonical TGF-beta signaling pathway (Hata and Chen, 2016) and the BMP inhibitor, i.e., CHRD (Hyvönen, 2003) were identified. In addition, the TGF-beta signaling pathway activated the MAPK-ERK signaling pathway and phosphorylated ERK upregulates the expression of ubiquitin ligase SMURF1 to inhibit the transportation of SMADS protein in the BMP signaling pathway into the nucleus, thereby affecting osteogenic differentiation (Sun et al., 2018). Our results showed that some factors of the CACN ligand family in the canonical MAPK signaling pathway, such as CACNA1A, CACNA1C, CACNB2, CACNB1 and CACNA1D and partial receptors and ligands of the MAPK-RAS pathway, such as FGFR8, KIT, FGFR1, VEGFC, and NTRK1 (Faiivre et al., 2006) were enriched. The upregulation of ERK was not found. In the analysis of the above signaling pathways, we were surprised to find that upregulated genes in H(E16) were not only enriched in axon guidance, Hippo signaling pathway, TGF-beta signaling pathway, MAPK signaling pathway and Wnt signaling pathway but also enriched in signaling pathways regulating pluripotency of stem cells, and cardiac muscle contraction, such as WNT2, SMAD9, SOX2, AXIN2, WNT5A, ID1, FGFR1, CACNA1C, CACNB2, CACNA1D, and CACNB1. Despite previous reports on Hulunbuir short-tailed sheep (Zhi et al., 2018) and studies on sheep embryos (Sanchez et al., 2021), the enrichment of these genes that we reported demonstrates that H(E16) is at the stage of organogenesis, which is a novel finding.

**U(E16)**

Protein is an essential dietary component for the nutritional balance of the human body as it is for sheep, which are ruminants. In U(E16), some of the upregulated genes are significantly enriched in protein digestion and absorption. Although most of these genes remain unexplored, there are still some genes that are well known, such as DPP4, also known as CD26. It is a type II transmembrane protein released from the cell membrane by a non-classical secretion mechanism. This exopeptidase selectively degrades a variety of substrates, including insulinotropic hormones, growth factors and cytokines (Nargis and Chakrabarti, 2018). XPNPEP2, a member of the aminopeptidase family, can perform amino acid excision and hydrolyze proteins to amino acids (Duan et al., 2005). MME is a member of the matrix metalloproteinase (MMP) family, also known as MMP-12. Unlike other MMP members, MME breaks down plasminogen to produce angiotatin, which has an inhibitory effect on vascular endothelial cell proliferation (Gorrin-Rivas et al., 2000). ACE2 is an important amino acid transport factor (Khodadoost et al., 2020) and regulates SLC7A7, SLC7A8, SLC7A9 and SLC15A1 of the solute carrier family (SLC) for the transmembrane transport of many important endogenous and exogenous substances, including nutrients and drugs (Peitzsch et al., 2014).

The digestion of lipids begins in the mouth, where the process of being broken down to triglycerides starts under the action of lingual lipase secreted by glands in the tongue. Lipid emulsion enters the duodenum as fine lipid droplets, which then mix with the bile and pancreatic juice and undergo significant changes in chemical and physical forms. Emulsification continues in the duodenum and is accompanied by hydrolysis and micellization, which set conditions for absorption through the intestinal wall (Iqbal and Hussain, 2009). In our study, some genes associated with fat digestion and absorption were significantly enriched. PLPP1 is a lipid phosphatase (Tang et al., 2015). MTTP is an endoplasmic reticulum resident protein (Wetterau et al., 1997) that primarily transfers triglycerides to promote optimal folding of nascent APOB and shuttles among other lipid classes such as cholesteryl esters, free cholesterol, phospholipids, ceramides, and sphingolipids to further promote lipoprotein formation (Iqbal et al., 2015; Peng et al., 2021). Fatty acid-binding protein FABP1 is a liver-specific FABP that plays an important role in intracellular lipid metabolism in the liver (Pi et al., 2019). A previous study has shown that FABP1 overexpression significantly increases fatty acid uptake by hepatocytes (Wu et al., 2016). APOB is a large amphiphilic glycoprotein that plays a central role in lipoprotein metabolism. Two forms of APOB (APOB-48 and APOB-100) are produced from the APOB gene through a unique post-transcriptional editing process. APOB-48 is necessary for chylomicron production in the small intestine, and APOB-100 is a ligand for LDL receptor-mediated endocytosis of LDL particles (Whitfield et al., 2004).

Blood cells are derived from hematopoietic stem cells (HSCs), which can self-renew or differentiate into multi-lineage committed progenitors, such as common lymphoid progenitors (CLPs) or common myeloid progenitors (CMPs). CLP produces leukocytes or the lymphoid lineage of leukocytes natural killer (NK) cells and T and B lymphocytes. CMP produces the bone marrow lineage, which includes the remaining leukocytes, erythrocytes, and megakaryocytes that produce platelets important for clotting. Cells undergoing these differentiation processes express surface markers corresponding to specific particular stages and lineages. Thus, the cell stage is determined by the specific expression patterns of these genes (Pouzoëlles et al., 2016). MME (CD10) is a marker of the early stage of the differentiation of HSC into B cells.
CD22 is expressed during B cell formation and on the surface of the formed B cell 54,55. A recent study suggests that MME is also a pro-adipogenic, specific cell surface marker for adipose stem cells (ASCs) (Chakraborty et al., 2021). IL1R2 (CD121) and CSF1R (CD115) are marker genes for the early stage of neutrophil formation (Maekawa et al., 2019). CSF1R(CD115) and CD14 are markers for the early stage of macrophage formation (Auffray et al., 2009). IL6R (CD126) is a marker in the middle and late stages of platelet formation. ITGA1, ITGA2 and ITGA6 are expressed on the membrane of platelets (Kobayashi et al., 1994). CD44 is specifically expressed in the early phase of T cell formation and is also a surface marker of erythrocytes (An and Chen, 2018; Föger et al., 2000).

Osteoclasts are multinucleated cells derived from the hematopoietic monocyte-macrophage lineage and are responsible for bone resorption. Mice lacking FOS (encoding C-FOS) develop osteosclerosis due to a block to early differentiation in the osteoblast lineage. C-FOS is a component of dimeric transcription factor-activating protein-1 (AP-1), consisting mainly of FOS (C-FOS, FOSB, FOSL1) and JUN proteins (Matsuo et al., 2000). CSF1R is abundantly expressed in the early stage of osteoclast differentiation and is used to track osteoclast precursors (Yahara et al., 2020). The high expression level of APC5 implies the formation of osteoclasts as it is the marker for osteoclasts (Muise et al., 2016). Osteoclasts can be formed through the indirect action of JAK1 and STAT3 in inflammatory cells and other helper cells (Sims, 2020). Tyrosine kinase TEC can mediate osteoclast formation through a complex of LCP2 and BLNK (Shinohara et al., 2008).

CONCLUSIONS

This study compares the transcriptional changes of E16 of two types of Mongolian sheep, namely, Hulunbuir short-tailed and Ujumqin sheep. Upregulated genes in H (E16) associated with axon guidance, Hippo signaling pathway, signaling pathways regulating pluripotency of stem cells, TGF-beta signaling pathway, cardiac muscle contraction, MAPK signaling pathway, and Wnt signaling pathway prove transcriptional regulation related to cellular interactions and organogenesis mainly occurs in early prenatal development. The biological processes, including protein digestion and absorption, fat digestion and absorption, Hematopoietic cell lineage, and osteoclast differentiation, are predominantly evidenced by upregulated genes in U(E16), reflecting the more complex development of organs and body structures during the U(E16) phase.

ACKNOWLEDGMENTS

The author are grateful to Wuchuan County Breeding Sheep Ranch Hohhot City, China for the support in completing this study.

Funding

This project is supported by the National Natural Science Foundation of China (No. 31860689); Inner Mongolia Autonomous Region Science and Technology Plan of China (No. 2019GG241) (No. 2020ZD0007); Inner Mongolia Autonomous Region Science and Technology Achievements Transformation Project (No.2020CG0078); Key Technology Project of Inner Mongolia Autonomous Region (No. 2021GG0062).

IRB approval

The study and animals experiment in the study were reviewed by the Ethics Committee of Inner Mongolia Agricultural University.

Ethics statement

All sheep experimental procedures and protocols in this study were approved and authorized by the animal care and use committee of Inner Mongolia Agricultural University (License NO. SYXK, Inner Mongolia, 2016-0017).

Statement of conflict of interest

The authors have declared no conflict of interest.

REFERENCES


