Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Transcriptome Analysis of Sheep Embryos In vivo Based on Single Cell RNA-Seq

Di Fang
Tarim University
Weikun Tao
Tarim University
Jie Wang
Tarim University
Fei Huang
Tarim University
Qinghua Gao (gsy1997@126.com)
Tarim University https://orcid.org/0000-0003-3660-4095

Research Article

Keywords: sheep, embryo, transcriptome, RNA-Seq

Posted Date: January 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1199760/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Objective: The purpose of the present study was to explore the transcriptome differences of sheep embryos. Embryos were at different developmental stages in order to assess the differences of the function, classification and metabolic pathway of differentially expressed genes and to provide a theoretical basis for revealing the regulatory mechanism of sheep early embryo development.

Methods: 8-cell, 16-cell, morula and blastocysts were collected and the sequencing library was constructed by the Smart-Seq2 amplification technology. The transcriber was sequenced by Illumina HiSeqXten high-throughput sequencing technology and the effective sequences were analyzed by functional annotation and related bioinformatic analysis.

Results: The results indicated that the Clean reads of 8-cell, 16-cell, morula and early blastocysts embryos were 441698590-48957974, of which 93.71-95.29% reads were compared with the reference genome sequence of sheep; 10 were compared with alternative splicing types of transcription terminal site (TTS) and with the transcription start site (TSS), which accounted for the largest proportion in sheep embryo transcripts. There were 171170-211487 sites of single nucleotide polymorphisms (SNPs) and 6435-8023 transcripts of novel transcripts in sheep embryos respectively. We used FDR < 0.05 and Fold Change > 2 as the criteria to screen for differential genes by comparing pairwise differences at four stage during embryo development. A total of 8281 differentially expressed mRNAs were identified, including 840 in E16vsE8, 6631 in E32vsE16, 810 in BlavsE32. Using the GO enrichment analysis, we explored the function of the DEGs. No significant difference was found at E16 vs E8.At E32 vs E16,Cellular components contained 127 significance terms (P < 0.05), 92 terms were significant enriched in molecular function, And biological processes involved 338 significance terms; At Bla vs E32,Cellular components contained 7 significance terms (P < 0.05). A total of 40 significance KEGG pathway terms were enriched in E32 vs E16.

Conclusions: In this study, individual embryonic transcriptome sequencing of sheep was established for high-throughput sequencing and analysis of the transcriptomes of sheep 8-cell, 16-cell, morula, and early blastocysts. The number of differentially expressed genes was identified at different stages of sheep embryo development and the function, classification and metabolic pathway of differentially expressed genes were obtained. The current study offers substantial information on the identification of the sheep embryo transcriptome, revealing the molecular regulatory mechanism of sheep embryo development. and selects 30 key genes, and its function needs further exploration and research.

1 Introduction

Embryonic development is the result of the connection and cooperation of several genes expressed in time and space[1–4]. It is a dynamic process affected by certain parameters, such as time and space. During a specific period of embryonic development, certain differentially expressed genes are involved in different biological functions and metabolic pathways that can regulate the normal development of the embryos.

Maternal mRNA is synthesized during oogenesis and accumulates in oocytes until fertilization. Following several cleavages of mammalian eggs, maternal mRNA is specifically degraded and replaced by zygotic transcripts. This transformation is called zygotic genome activation (ZGA) and controls embryonic development. Maternal genes encode transcripts and proteins during oogenesis in lower species, such as *Drosophila melanogaster* and *Xenopuslaevis*, guiding the transformation of oocytes to fertilized eggs[5, 6]. The normal activation of ZGA is critical to embryonic development and each species has a specific zygote activation time. The mouse activation of ZGA occurs in the 1-2 cell stage[7], the bovine in the 8-16 cell stage, the pig in the 4-cell stage and the human in the 4-8 cell stage[8]. The process of transformation from the maternal to the zygotic gene is called MZT. The maternal genes are very important for coordinating cell division and activating embryonic zygotic genes. It has been speculated that hundreds of maternal genes were involved in the process of zygotic gene activation, whereas certain maternal genes that were involved in ZGA activation were verified. So far, a relatively low number of maternal genes that have been shown to be involved in ZGA activation have been found in mammals. Eight maternal effect genes, such as Mater, Hsf1, Dnmt10, Pms2, Zar1, Npm2, Stella and Zfp36l2, have been identified in mice and these genes have been studied by mouse knockout models. It was found that some of these genes were involved in early embryonic development, notably from 1-cell to 2-cell stage[7].

In 2009, Tang et al[9] initiated the first single-cell RNA sequencing (scRNA-Seq) technique to analyze the whole transcriptome of a single mouse blastocyst for the first time. In recent years, bioinformatics analytical methods, such as (PCA) or unsupervised clustering were used to analyze the composition of individual embryonic cells at different early developmental stages. The dynamic changes of pre-implantation of the embryonic cell transcriptome were revealed. The blastocysts indicated apparent heterogeneity. Through unsupervised clustering analysis, the blastocyst cells could automatically cluster into the three following groups of cells: trophoblast, ectoderm and primitive endoderm. Prior to blastocysts, no significant heterogeneity was noted between blastomeres of single embryo. For example, blastomeres in morulae expressed the different marker genes Sox2 and Id2, which could distinguish inner and outer cells by marker genes. Even in the 2-cell stage, differential expression among genes was noted between blastomeres. Yan et al.[1]used transcriptome sequencing to identify several genes out of the 2,500 genes that were upregulated during embryonic genome activation. These genes were associated with RNA transcription, translation and epigenetic regulation. Xue et al[10]. demonstrated that by injecting sperm cells from the same donor into different oocytes following fertilization, 5% of all gene transcripts were detected by paternal SNP. At each stage parental genes and the cell cycle-related gene CDCA2 were noted which indicated maternal-specific expression in both 2-cell and 4-cell stage embryos. Single cell transcriptome sequencing has become a revolutionary tool, which addresses the problem of transcriptome sequencing of limited RNA samples. This technique can sequence single mammalian somatic cells and reveal the heterogeneity of single cells[11].

The RNA content of early embryos is low and it is difficult to meet the minimum starting amount of RNA required for transcriptome sequencing[12]. However, with the establishment and development of microtranscriptome sequencing techniques, such as Smart-seq2[13], previous studies have successfully used RNA-Seq technology to analyze the regulatory mechanism of early embryonic development in human, pig and bovine species[9–11]. However, the number of studies that have examined the expression of functional genes and their potential association with the regulation of sheep embryonic development is very limited. The achievements of sheep genome sequencing and the development of high-throughput sequencing technology provide a rapid and effective

method for studying the molecular mechanism of sheep embryonic development at the taxonomic level. In the present study, the RNA library of the sheep single 8-cell,16-cell, morula and blastocysts embryos was amplified by the Smart-Seq2 method and the IlluminaHiSeqXten technique was used for high-throughput sequencing analysis in order to reveal the regulatory mechanism of sheep early embryo development and understand the genetic structure of its transcriptome, and provide useful resources and markers for future functional genomics studies.

2 Material And Methods

2.1 Ethics statement

All animal experiments were conducted according to the Regulations and Guidelines for Experimental Animals established by the Ministry of Science and Technology (Beijing, China, revised in 2004).The present study was approved by the Institutional Animal Care and Use Committee of Tarim University.

2.2 Chemicals and culture media

All chemicals were purchased from Sigma-Aldrich (Oakville,ON,Canada), unless otherwise specified.

2.3 Embryo collection

The animals in this study all came from the Yingjisha County improved seed breeding center in Kashi County, Xinjiang Uigur Autonomous Region, China.We chose healthy nonpregnant 6 Hu sheep aged 2–4 years.All sheep underwent superovulation (described in Table 1), artificial insemination after estrus, and washed embryos from day 4 at 8-cell, 16-cell, morula and early blastula at day 7, and 6 embryos at each period.The 8-cell,16-cell,morula and early blastula embryos were collected and transferred to 10 μl lysates and liquid nitrogen was injected into the embryos within 2 min. Following freezing for severa l h, 2 repeats were used for each group. The 8-cell,16-cell,morula and early blastula were marked as E08,E16,E32 and blastula.The samples were transported by dry ice and sent to An Nuoda Genome Technology (Beijing) Co. Ltd.

Table 1
treatment methods of Superovulation

Time	Treatment
Day1	Vaginal indwelling CIDR suppository
Day7	Intramuscular injection of vade 3ml / animal in the morning
Day11	Follropin-v 2.9ml/animal was injected intramuscularly in the morning and evening
Day12	Follropin-v 1.6ml/animal was injected intramuscularly in the morning and evening
Day13	Follropin-v 1.0ml/animal was injected intramuscularly in the morning and evening
Day14	Follropin-v 0.5ml/animal was injected intramuscularly in the morning and evening, At the same time, the suppository was removed in the morning and PG 0.1ml was injected intramuscularly
Day15	Test situation Breeding, Intramuscular injection of LHRH-A3 25μg
Day19	In the morning and evening, 25 mg progesterone injection was injected intramuscularly
Day20	In the morning and evening, 25 mg progesterone injection was injected intramuscularly
Day21	Uterine horn embryo

2.4 RNA extraction and sequencing for library construction

The construction and sequencing of the library was completed by Annuoda Genome Technology (Beijing) Co., Ltd. The single-cell samples were collected in tubes with lysis component and ribonuclease inhibitor. Subsequently, the amplification was carried out by the Smart-Seq2 method. An Oligo-dT primer was introduced to the reverse transcription reaction for the first-strand cDNA synthesis, followed by PCR amplification to enrich the cDNA and a magbead purification step to clean up the production. The cDNA was checked by the Qubit® 3.0 Flurometer and the Agilent 2100 Bioanalyzer was used to ensure the expected production with an approximate length of 1-2 kbp. Subsequently, the cDNA was sheared randomly by ultrasonic waves for the Illumina library preparation protocol including DNA fragmentation, end repair, 3'ends A-tailing, adapter ligation, PCR amplification and library validation. Following library preparation, the PerkinElmer LabChip® GX Touch and Step OnePlus™ Real-Time PCR System were introduced for library quality inspection. The libraries that satisfied these criteria were subsequently loaded on the Illumina Hiseq platform for PE150 sequencing.

2.5 Transcriptome data analysis

In order to ensure the quality of the sequencing data, the Rawreads obtained by the IlluminaHiSeqXten sequencing were filtered and the Clean reads were obtained following removal of the joint sequence, the empty read sequence and the low quality sequence (Phred quality < 5). The reference gene and genome annotation files were downloaded from the ENSEMBL website (ftp://ftp.ensembl.org/pub/release-100/fasta/ovis_aries/dna/Ovis_aries.Oar_v3.1.dna.toplevel.fa.gz). The reference genome library was built using Bowtiev (1.0.1) and subsequently the Clean Data were compared to the reference genome through HISAT2 (v2.1.0). The gene expression was calculated by FPKM (Fragments Per Kilobase Millon mapped reads). The comprehensive expression of two biological repeats in the 8-cell,16-cell, morula and early blastula stage of they stage was obtained by calculating the expression amount of two biological repeats in the 8-cell,16-cell, morula and early blastula stage of they stage by the base mean. The differentially expressed genes were selected according to the estimation of the gene expression levels by DESeq,which was achieved by the negative binomial distribution

of the statistical methods.The P_value was corrected to control false positive according to Benjamini and Hochberg methods. The corrected P_value≤0.05 was defined as the parameter required for the differentially expressed gene. DESeq2 was used to analyze the differential expression of biological repetitive samples between the two groups compared with the treatment group and the reference group. The genes with | log2Ratio |≥1 and Q < 0.05 were selected as differentially expressed genes and the number of up- and downregulated genes was obtained. The obtained DEGs were mapped to each entry in the GO database (http://www.geneontology.org/) and their number was calculated. Following correction of the P value by the Benjamini method[14], the GO entry with Q<0.05 was the GO entry used for significant enrichment of DEGs. The signaling pathways or metabolic pathways involved in the genes were analyzed by comparing their expression with the KEGG (Kyoto encyclopedia of genes and genomes) database (http://wego.genomics.org.cn).

2.6 Analysis of alternative splicing and single nucleotide Polymorphism and Prediction of New transcripts

The RPKM tool of AS profile (1.0.4) was used to analyze the structure and expression of alternative splicing events and Samtools (1.5) was used to analyze single nucleotide polymorphism (single nucleotide polymorphism, SNP)[15]. The sequenced sequences of the aligned genomes were assembled and spliced by the Cufflinks software (v2.2.1)[16]. Following filtering of the low quality sequences (length ≤ 180 bp, Q value ≤ 10), the assembled transcript sequence was compared with the gene annotation information on the sheep genome. For example, the assembled transcript sequence was not aligned with the existing genes and was located on the genomes between the existing genes. The following conditions were satisfied: the distance from the existing annotated genes was more than 200 bp, the length was not shorter than 180 bp and the sequencing depth was not less than 2. These sequences were identified as potential new transcripts and new genes[17].

2.7 Validation by real-time quantitative PCR (RT-qPCR)

Here,we selected 5 DEGs to verify the sequencing results via RT-qPCR. Primers were designed by Primer-BLAST in NCBI,as shown in Table 2. First strand cDNA was synthesized from 1μg total RNA using the reverse transcriptase Revert Aid (Takara) following manufacturers recommendations.PCR amplification was executed in reaction volumes of 10μL that included 1μL of cDNA,0.6μL of forward and reverse primers (10μM) for each gene, 5μL of 2×S6 Universal SYBR qPCR Mix (NovaBio.China),and 3.4μL of RNA-free double-distilled H_2O. The cycling conditions were as follows: 95°C for 30 s, 40 cycles of 95°C for 5 s, and annealing temperature 60 °C for 30 s; a melt curve analysis was performed at 65°C-95°C.

Table 2
QT-PCR primers

GENE	primer	Tm	Product length(bp)
GAPDH	F-GGTGATGCTGGTGCTGAG	58	184
	R-TGCTGACAATCTTGAGGGTAT		
MAP4	F-GCCAGGTTGAAGGGTTTCCA	59	275
	R-AGCAAATGTGTCGCAGGGTA		
SLC39A8	F-GTTGCTGTGTTTGGCGGATT	59	200
	R-GGCTCTGTGACAGCTGGATT		
EMC2	F-TGTCAGCAGTATGCCGAAGT	59	218
	R-ATTTGACTGGCTGCCCAACT		
VPS45	F-GCAGAACCCCAGAGTGACAG	59	209
	R-CTTTGGGGCTGAAGAGGTCA		
ELMOD2	F-TAACCCTGAGAAGGATGCCAG	59	218
	R-CAATGTCAGCCCACTGCTTG		

2.8 Statistical analysis

GAPDH were used as housekeeping genes for qPCR, and the mean expression levels of GAPDH were considered to be the expression levels of housekeeping genes. 8-cell was used to normalized the gene expression,and the relative gene expression was calculated by $2^{-\Delta\Delta CT}$ method.

3 Results

3.1 Sequencing quality Evaluation and basic data Analysis

The embryo libraries of sheep E08,E16,E32 and early blastula constructed in this study met the requirements of transcriptome sequencing. The embryos were detected by the Agilent 2100 Bioanalyzer using the Agilent High Sensitivity DNA Kit. The 8 sample peaks indicated apparent peaks of target products in the length range of 1-2 kb fragments, whereas some small fragments below 1 kb were noted. However, the proportion was small, indicating that the integrity of the original sample was optimal and that it met the requirements of database construction (Fig. 1). The percentage range of Q30 of the 8 samples was estimated to 92.22-93.03%(Fig. 2), indicating that the sequencing quality and library construction quality were high and that the sequencing data were accurate and reliable. This could meet the needs of follow-up analysis. In the sequencing results, the contents of A-T and C-G of the 8 samples basically coincided with each other, indicating that the base composition was stable and balanced and that the sequencing quality was high (Fig. 3).

3.2 Analysis of sequencing results

Following filtration of the data of the original sequence, the filtered sequence of the sheep embryos at the four developmental stages of the 8-cell, 16-cell, morula and blastocysts embryos exhibited the following range: 441698590-48957974. The obtained clean reads were compared with the reference genome by the TopHat software[18, 19]. The results indicated 93.71-95.29% clean read alignment of the sheep reference genes in each stage and the sequence ratio of the multiple locations of the genome was 5.79-15.62%, which satisfied the requirements (Table 3). The number of transcripts and predicted new transcripts compared at each stage of development is shown in Table 4. Transcript expression was presented by FPKM (Fragments per kilo-base of exon per million fragments mapped) value. The FPKM distribution of mRNAs is shown in Fig. 4.Alternative splicing analysis by ASprofile software indicated mainly 10 types of splicing events in sheep embryos: transcription terminal site(TTS),transcription start site(TSS),skipped exon(SKIP),multi-SKIP(MSKIP), alternative exon (AE), intron retention(IR) and multi-IR(MIR), among which TSS and TTS accounted for the largest proportion (Table 5). SNP analysis by Samtools indicated the presence of 171170-211487 loci in each developmental stage of sheep embryos from the 8-cell,16-cell, morula and early blastula embryos(Table 6). Count the number and proportion of unique alignment sequences (i.e. sequences that align only one position of the genome) on the three functional elements of genes (exon, intron and intergenic)(Table 7).

Table 3
RNA-Seq and mapping to the reference genome

	8-cell-1(E8C01)	8-cell-2(E8C02)	16-cell-1(E16C01)	16-cell-2(E16C02)	32-cell-1(E32C01)	32-cell-2(E32C02)	BL-cell-1(Bla01)	BL-cell-2(Bla02)
Clean Reads	41,170,258	45,850,662	41,828,458	39,256,590	41,169,756	40,471,658	46,618,120	39,995,524
Mapped Reads	38,823,560	43,234,720	39,485,132	37,016,538	38,628,316	37,924,122	44,423,045	37,756,956
Mapping Rate	0.943	0.9429	0.944	0.9429	0.9383	0.9371	0.9529	0.944
UnMapped Reads	2,346,698	2,615,942	2,343,326	2,240,052	2,541,440	2,547,536	2,195,075	2,238,568
MultiMap Reads	2,838,617	2,756,045	2,421,368	2,388,476	6,429,274	5,386,962	6,418,791	5,326,921
MultiMap Rate	0.0689	0.0601	0.0579	0.0608	0.1562	0.1331	0.1377	0.1332

Table 4
Detected transcripts and novel transcripts

Embryos	No. of transcripts	No. of novel transcripts
8-cell-1(E8C01)	12652	7394
8-cell-2(E8C02)	12857	7729
16-cell-1(E16C01)	13457	7904
16-cell-2(E16C02)	13711	8023
32-cell-1(E32C01)	13117	7303
32-cell-2(E32C02)	12746	7469
BL-cell-1(Bla01)	11996	6435
BL-cell-2(Bla02)	13008	7132

Table 5
Alternative splicing events detected during embryonic development

AS_type	8-cell-1(E8C01)	8-cell-2(E8C02)	16-cell-1(E16C01)	16-cell-2(E16C02)	32-cell-1(E32C01)	32-cell-2(E32C02)	BL-cell-1(Bla01)	BL-cell-2(Bla02)
XAE	273	329	334	373	269	349	144	287
XIR	124	86	108	128	136	158	82	100
TTS	20,674	20,616	21,350	21,284	18,969	19,106	16,628	18,024
XMIR	8	4	6	2	8	8	16	10
XMSKIP	226	268	384	262	144	214	70	160
IR	492	502	570	658	758	820	576	632
XSKIP	766	908	1,068	1,014	456	534	418	472
SKIP	4,420	4,564	5,394	5,152	2,104	2,646	1,978	2,230
AE	1,603	1,647	1,945	1,936	1,115	1,336	852	1,127
MIR	22	20	28	56	38	28	44	36
TSS	20,953	21,064	21,903	21,960	19,704	19,560	16,999	18,476
MSKIP	1,008	962	1,182	1,094	270	456	274	370

Table 6
Putative SNPs detected during embryonic development

Sample	8-cell-1(E8C01)	8-cell-2(E8C02)	16-cell-1(E16C01)	16-cell-2(E16C02)	32-cell-1(E32C01)	32-cell-2(E32C02)	BL-cell-1(Bla01)	BL-cell-2(Bla02)
SNP	137,600	115,816	148,264	147,307	250,020	190,623	168,868	176,018
InDel	10,273	8,221	10,400	10,476	19,446	15,948	13,959	13,034
Total	147,873	124,037	158,664	157,783	269,466	206,571	182,827	189,052

3.3 Gene expression characteristics and differentially expressed genes in early embryos

We used FDR < 0.05 and Fold Change > 2 as the criteria to screen for differential genes by comparing pairwise differences at four stage during embryo development. A total of 8,281 differentially expressed mRNAs were identified, including 840 in E 16vs E8, 6631 in E32 vs E16, 810 in Bla vs E32. And the number of DEGs at different stage was summarized in Fig. 5. To identify genes that play a key role in embryonic development throughout the embryonic stage, we performed Venn on genes at different stages. A total of 30 key genes were found in the intersection, generated from the Venn of DEGs Fig. 6,and the expression of 32 key genes was shown in Table 8.

Table 7
Distribution statistics of unique alignment sequences in reference genome region

#Library	8-cell-1(E8C01)	8-cell-2(E8C02)	16-cell-1(E16C01)	16-cell-2(E16C02)	32-cell-1(E32C01)	32-cell-2(E32C02)	BL-cell-
Exon	6,707,141(31.97%)	7,752,870(34.97%)	6,647,002(31.45%)	6,185,164(30.87%)	5,306,854(24.42%)	5,777,827(29.81%)	6,959,2
Intron	4,399,041(20.97%)	4,035,432(18.20%)	4,477,297(21.18%)	4,094,111(20.43%)	6,541,309(30.10%)	5,017,548(25.88%)	5,509,3
Intergenic	9,874,822(47.07%)	10,378,951(46.82%)	10,011,928(47.37%)	9,757,694(48.70%)	9,886,647(45.49%)	8,589,507(44.31%)	12,431,

3.4 GO enrichment and KEGG Pathway Analysis of differentially expressed genes

Using the GO enrichment analysis, we explored the function of the DEGs. The top 10 enrichment terms in the three sections (Cellular Component, Molecular Function, Biological Process) were displayed in Fig. 7A-E. No significant difference was found at E16 vs E8.At E32 vs E16,Cellular components contained 127 significance terms (P < 0.05), such as intracellular part,cytoplasmic part,intracellular organelle part,intracellular organelle,organelle part, 92 terms were significant enriched in molecular function,the top 5 terms included for example, catalytic activity,heterocyclic compound binding,nucleoside phosphate binding,nucleotide binding,organic cyclic compound binding. And biological processes involved 338 significance terms; the top 5 terms included cellular metabolic process,metabolic process,organic substance metabolic process,primary metabolic process,nitrogen compound metabolic process.At Bla vs E32,Cellular components contained 7 significance terms (P<0.05),such as intracellular part,endoplasmic reticulum part,endoplasmic reticulum membrane,proteasome core complex, alpha-subunit complex,Ndc80 complex,And biological processes involved 1 significance terms.

The KEGG enrichment analysis of the DEGs was shown in Fig. 7F.A total of 40 significance KEGG pathway terms were enriched in E32vsE16, for instance, Oxidative phosphorylation,Huntington disease,Biosynthesis of amino acids,Peroxisome.

3.5 Validation of candidate genes

To reveal the key genes associated with embryonic development, we screened several genes with higher expression levels among the 30 key differentially expressed genes, including MAP4,SLC39A8,EMC2,VPS45,ELMOD2.Then,the candidate genes were verified via RT-qPCR (Fig. 8A-E). Notably, similar results were reported as those obtained through sequencing, which confirmed the reliability of the sequencing data.

4 Discussion

Mammalian single somatic cell RNA has a concentration range of 10-30 pg, of which rRNA accounts for approximately 85% of the total RNA content. Approximately 15-20% of the rRNA content is as low as 1-5%[20]. The basic requirements for the total RNA used in general transcriptome sequencing are as follows: concentration of RNA samples ≥ 400 ng/µl, total amount ≥ 20 µg, purity of A260/280 between 1.8-2.4, A260/230 > 1.8, optimal integrity (RIN 7.0-8.0) and RNA28S/8S > 1.0mm. However, the total RNA concentration of each embryo from the mammalian 2-cell to the blastocyst stage is only 200-2,000 pg[21]. Therefore, the trace RNA cannot meet the basic requirements of the construction of the transcriptome sequencing library and of high-throughput sequencing. With the continuous development of a new generation of high-throughput sequencing technology, RNA-Seq can be used to sequence the total transcriptional activity of any biological growth and developmental stage without designing probes in advance. Moreover, it can accurately detect gene expression under various conditions and discover several unknown molecular regulatory mechanisms[22]. In the present study, RNA-Seq technology was used for the first time to reveal the mechanism of sheep early embryo development from the perspective of single embryo transcriptome, which provides a new method for improving sheep embryo production *in vitro*. It also provides a theoretical basis for further improving sheep gene structure information and identification of new genes related to embryonic development. Therefore, Smart-Seq2 amplification technology was used to enrich the samples and construct a sequencing library[23].The Illumina Hi SeqXten high-throughput sequencing technology was used for transcriptome sequencing. The results of sequencing quality evaluation and data analysis indicated that the sequencing quality and library construction quality were high and that the sequencing data were accurate and reliable.

Following filtration of the data of the original sequence, the filtered sequence of the sheep embryos at the four developmental stages of the 8-cell, 16-cell, morula and blastocysts embryos exhibited the following range: 441698590-48957974. Using the TopHat software to compare the obtained Clean reads with the reference genome, the results indicated a range of 93.71-95.29% clean reads alignment on the sheep reference gene in each stage. The sequence ratio of (multi map rate) multiple locations of the genome was 5.79-15.62%, which satisfied the requirements. The correlation between biological repeats was 90.86% and the repeatability between the samples was optimal. In the present study, a large number of alternative splicing in sheep 8-cell, 16-cell, morula and blastocysts,Alternative splicing of transcripts is prevalent in mammalian cells and could make major differences for maintenance of cell identity and function[24, 25]. Certain genes are expressed at specific stages of embryonic development and play key regulatory roles at specific stages. It was found that the large number of novel transcripts was due to the lack of the Y chromosome sequence in the reference genome of the sheep.

Embryonic development is the result of the expression of genetic information according to a certain time, space and order. This result (genetic program) is derived from the genetic program of development. The development of early embryos belongs to the process of maternal regulation, indicating that it is regulated by a large number of mRNAs and proteins synthesized during oogenesis and maturation. Maternal mRNA plays an important physiological role in the early stage of embryonic development. Following developmental progression, maternal mRNA and protein are gradually degraded, while embryonic genome activation (embryonic genome activation, EGA) is initiated and developmental transition from maternal regulation to embryonic regulation (maternal to-embryonic transition, MET) is facilitated. Different species exert different stages of MET. Mouse MET occurs in the 2-cell stage, human and pig occur in the 4-cell to the 8-cell stage and the bovine embryo genome mainly transcribes from the 8-cell to the 16-cell stage.

KIT, STAT3, ZP4, BMP15, GDF9, ZP3, Pms2, Hsf1-2 and Zar1 were confirmed to be maternal genes[26, 27]. The expression of these maternal genes from the 16-cell to the morula sheep embryos did not change significantly. The expression levels of the zygotic genome-related genes were not high. However, the marker gene NANOG of EGA was initially expressed in the morula stage. Therefore, it is concluded that sheep EGA may occur from the 16-cell to the morula stage. Methylation-related genes are stably and highly expressed, suggesting that the methylation of embryos is being reconstructed, which is consistent with previous studies.

The number of DEGs in 840 in E16vsE8, 6631 in E32vsE16, 810 in BlavsE32. And the number of DEGs at different stage was summarized in Fig. 5. indicating significant temporal differences in the developmental regulatory mechanism of sheep embryos at different developmental stages.There was 6631 differential genes from 16-cell development to the morula period, mainly due to the substantial activation of zygotic genes during the ZGA period of sheep embryos, as a result of substantial consumption of maternal material.We identified 30 key genes, and the analysis found that these genes were functional enriched in the cell cycle, RNA, DNA synthesis regulation, and verified the developmental characteristics of energy metabolism during this period.We identified 30 key genes, and the analysis found that these genes were functional enriched in the cell cycle, RNA, DNA synthesis regulation, and verified the developmental characteristics of energy metabolism during this period.But its functionality will need to be further explored.

GO analysis indicated that No significant difference was found at E16 vs E8.At E32 vs E16,Cellular components contained 127 significance terms ($P<0.05$). And biological processes involved 338 significance terms.At Bla vs E32,Cellular components contained 7 significance terms.The GO and KEGG enrichment analyses found that the main differences were concentrated in the 16-cell to morula period, and the difference types were mainly cell metabolism and compound synthesis, indicating that after zygotic genome activation, the zygotic genome began to regulate embryonic development.In bovine, comparative transcriptome analysis demonstrated that lineage differentiation at the blastocyst stage is associated with marked differences in expression of genes involved in various biological processes, such as metabolism, endocytosis, paracrine signaling, and cellular architecture[28].

5 Conclusions

In this study, a single cell transcriptome sequencing method for single sheep embryos was established. The transcriptome of sheep 8-cell,16-cell, morula and early blastocyst were sequenced and analyzed by single cell RNA-Seq for the first time. The number of differentially expressed genes at different stages of sheep embryonic development was identified, and the function, classification and metabolic pathway of differentially expressed genes were obtained. The results showed that there were the most differences in differential genes and metabolic pathways from 16 cells to morula stage, and the metabolic pathways were mainly related to cell proliferation, energy metabolism, cell cycle and so on. It also showed that during this period, the zygotic genome began to play a role in regulating the development of embryos. This study provided a lot of information for the identification of sheep embryo transcriptome, revealed the molecular regulation mechanism of sheep embryo development, improved the technology of in vitro embryo culture, and screened out 30 key genes, whose functions need to be further explored.

Declarations

Authors Contributions

Conceptualization: Qinghua Gao

Data curation: Di Fang

Formal analysis: Di Fang

Funding acquisition: Qinghua Gao

Investigation: Weikun Tao

Methodology:Di Fang, Weikun Tao

Project administration: Qinghua Gao

Resources:Fei Huang,Jie Wang

Software: Di Fang

Visualization:Di Fang, Jie Wang

Roles/Writing - original draft: Di Fang

Writing - review & editing:Di Fang

All Authors read and approved the manuscript.

Acknowledgements

The authors would like to thank Ruoyang Hu for his assistance in data generation.

Funding

This work was supported by National Natural Science Foundation of China "Study on the Molecular Mechanism of Bovine SRY Gene Promoting the Proliferation and Differentiation of Male Cleavage Embryos" (31660657); XPCC's young and middle-aged technological innovation leading talent team construction project (2015BC001); Graduate Scientific Research Innovation Project (TDBSCX201906).

Compliance with ethical standards

Conflict of interest The authors declare that they have no confict of interest.

Ethical approval All animal experiments were conducted according to the Regulations and Guidelines for Experimental Animals established by the Ministry of Science and Technology (Beijing, China, revised in 2004).The present study was approved by the Institutional Animal Care and Use Committee of Tarim University..

References

1. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J, Huang J, Li M, Wu X, Wen L, Lao K, Li R, Qiao J, Tang F (2013) Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. NAT STRUCT MOL BIOL 20(9):1131–1139. doi: 10.1038/nsmb.2660
2. Cao S, Han J, Wu J, Li Q, Liu S, Zhang W, Pei Y, Ruan X, Liu Z, Wang X, Lim B, Li N (2014) Specific gene-regulation networks during the pre-implantation development of the pig embryo as revealed by deep sequencing. BMC Genomics 15(1):4. doi: 10.1186/1471-2164-15-4
3. Graf A, Krebs S, Heininen-Brown M, Zakhartchenko V, Blum H, Wolf E (2014) Genome activation in bovine embryos: Review of the literature and new insights from RNA sequencing experiments. ANIM REPROD SCI 149(1–2):46–58. doi: 10.1016/j.anireprosci.2014.05.016

4. Jiang Z, Sun J, Dong H, Luo O, Zheng X, Obergfell C, Tang Y, Bi J, Neill RO, Ruan Y, Chen J, Tian X (2014) Transcriptional profiles of bovine in vivo pre-implantation development. BMC Genomics 15(1):756. doi: 10.1186/1471-2164-15-756
5. Newport J, Kirschner M (1982) A major developmental transition in early xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell 30(3):675–686. doi: 10.1016/0092-8674(82)90272-0
6. Morisalo D, Anderson KV, THAT ESTABLISH THE DORSAL-VENTRAL PATTERN, OF THEDROSOPHILA EMBRYO (1995)ANNU REV GENET, 29(1),371–399. doi: 10.1146/annurev.ge.29.120195.002103
7. J Dean (2002) Oocyte-specific genes regulate follicle formation, fertility and early mouse development. J REPROD IMMUNOL 53(1–2):171–180. doi: 10.1016/s0165-0378(01)00100-0
8. Morgan HD, Santos F, Green K, Dean W, Reik W (2005) Epigenetic reprogramming in mammals. HUM MOL GENET 14(suppl1):R47–R58. doi: 10.1093/hmg/ddi114
9. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, MA Surani (2009) mRNA-Seq whole-transcriptome analysis of a single cell. NAT METHODS 6(5):377–382. doi: 10.1038/nmeth.1315
10. Xue Z, Huang K, Cai C, Cai L, Jiang C, Feng Y, Liu Z, Zeng Q, Cheng L, Sun YE, Liu J, Horvath S, Fan G (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature 500(7464):593–597. doi: 10.1038/nature12364
11. Tong LI, Jia-wei XU, SUN Ying-pu (2019) Application of Single-Cell Transcriptional Sequence in Study of Reproductive Development. J Int Reprod Health/Fam Plan 38(03):217–221
12. Xue-ying YU (2016) GUOQin-qin,HAO Hai-sheng,SUN Wei-jun,ZHAO Xue-ming,ZHU Hua-bin,YANG Ling,DU Wei-hua. Transcriptome of Bovine IVF Embryos Treated with Glutathione.Xu mu shou yi xue bao, 47(7), 1363-1372. doi: 10.11843/j.issn.0366-6964.2016.07.008
13. LD L, XX R, WY L, X T, ZJ T, ZX D, W Y, L J (2014) RNA-Seq analysis of yak ovary: improving yak gene structure information and mining reproduction-related genes. SCI CHINA LIFE SCI 44(03):307–317
14. Yoav B, Yosef H (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.Journal of the Royal Statistical Society. Series B (Methodological), 57(1)
15. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079. doi: 10.1093/bioinformatics/btp352
16. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. NAT BIOTECHNOL 28(5):511–515. doi: 10.1038/nbt.1621
17. Roberts A, Pimentel H, Trapnell C, Pachter L (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics 27(17):2325–2329. doi: 10.1093/bioinformatics/btr355
18. Q QIU, HA LEWIN GZHANGLAUVILBCAPITANUJMA, K WANG XQIANYLANGRZHOULWANG, T JXIA, MA, S LIAO SPAN, XU LU HHOU, YE YWANGXZANG, YIN HMA, Z JZHANG, WANG, W QIAN, ZHANG Y, T YONEZAWA DZHANG, ZHANG MHASEGAWAYZHONGWLIUY, J ZHUANGSZHANGRLONG, JA WANGHYANGJWANG, DN LENSTRA, YI COOPER, Z WUJWANGPSHIJLIU, C YE, Q CAO, DM LARKIN (2012) The yak genome and adaptation to life at high altitude. NAT GENET 44(8):946–949. doi: 10.1038/ng.2343
19. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25(9):1105–1111. doi: 10.1093/bioinformatics/btp120
20. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA Synthesis in Mammalian Cells. PLOS BIOL 4(10):e309. doi: 10.1371/journal.pbio.0040309
21. Gilbert I, Scantland S, Sylvestre E, Gravel C, Laflamme I, Sirard M, Robert C (2009) The dynamics of gene products fluctuation during bovine pre-hatching development. MOL REPROD DEV 76(8):762–772. doi: 10.1002/mrd.21030
22. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. NAT METHODS 5(7):621–628. doi: 10.1038/nmeth.1226
23. Picelli S, Björklund ÃK, Faridani OR, Sagasser S, Winberg G, Sandberg R (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. NAT METHODS 10(11):1096–1098. doi: 10.1038/nmeth.2639
24. Codina-Fauteux V, Beaudoin M, Lalonde S, Lo KS, Lettre G (2018) PHACTR1 splicing isoforms and eQTLs in atherosclerosis-relevant human cells. BMC MED GENET 19(1):97. doi: 10.1186/s12881-018-0616-7
25. Park J, Jang H, Shin S, Cho H, Choi J, Kim N, Lee H, Do K, Song K, Cho B (2015) Molecular Analysis of Alternative Transcripts of the Equine Cordon-Bleu WH2 Repeat Protein-Like 1 (*COBLL1*) Gene. ASIAN AUSTRAL J ANIM 28(6):870–875. doi: 10.5713/ajas.14.0722
26. Graf A, Krebs S, Heininen-Brown M, Zakhartchenko V, Blum H, Wolf E (2014) Genome activation in bovine embryos: review of the literature and new insights from RNA sequencing experiments. ANIM REPROD SCI 149(1–2):46–58. doi: 10.1016/j.anireprosci.2014.05.016
27. Stallings JD, Zeng YX, Narvaez F, Rebecchi MJ (2008) Phospholipase C-δ1 Expression Is Linked to Proliferation, DNA Synthesis, and Cyclin E Levels. J BIOL CHEM 283(20):13992–14001. doi: 10.1074/jbc.M800752200
28. Ozawa M, Sakatani M, Yao J, Shanker S, Yu F, Yamashita R, Wakabayashi S, Nakai K, Dobbs KB, Sudano MJ, Farmerie WG, PJ Hansen (2012) Global gene expression of the inner cell mass and trophectoderm of the bovine blastocyst. BMC DEV BIOL 12(1):33. doi: 10.1186/1471-213X-12-33

Tables

Due to technical limitations, table 8 is only available as a download in the Supplemental Files section.

Figures

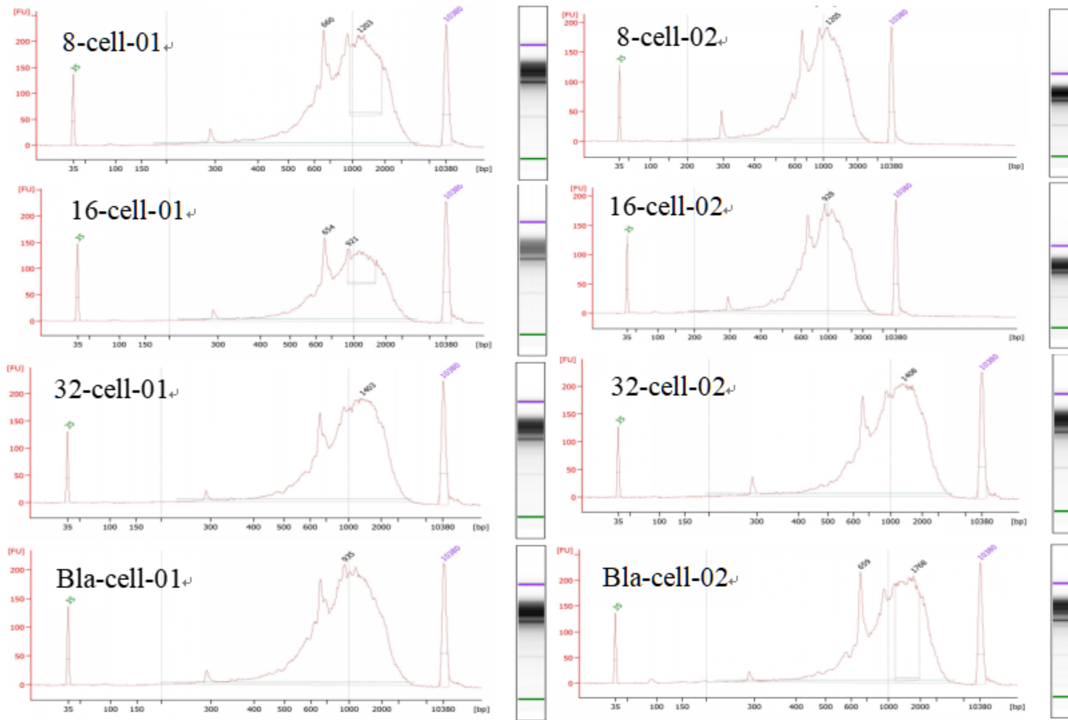

Figure 1

peak figure of Agilent 2100 for cDNA integrity detection by Smart-Seq2 PCR amplification (Note: Figure 8-cell-01 and 8-cell-02:8-cell cDNA peak diagram; 16-cell-01 and 16-cell-02:16-cell cDNA peak diagram;32-cell-01 and 32-cell-02:Morula cDNA peak diagram;Bla-cell-01 and Bla-cell-02:early blastocyst cDNA peak diagram;)

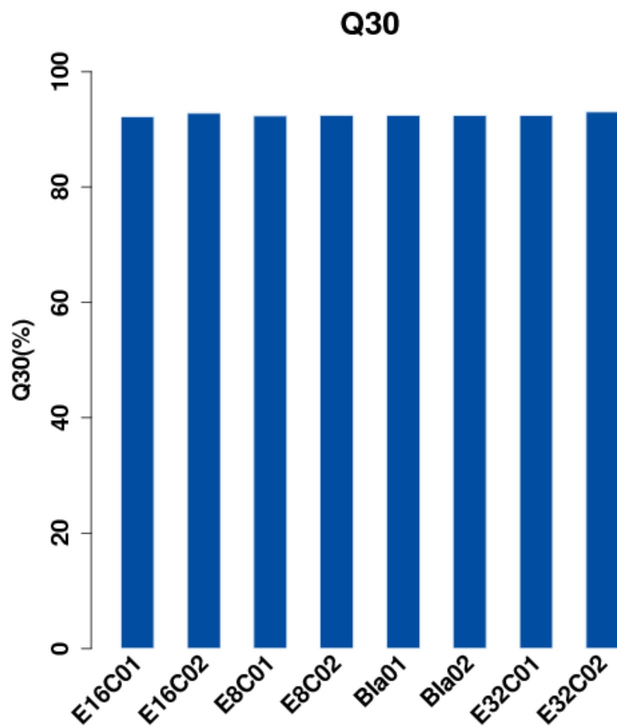

Figure 2

Base quality level of Q30 reactive sequencing. *(The proportion of the bases with sequencing quality value was greater than 30 (error rate less than 0.1%) in the total RawReads.)*

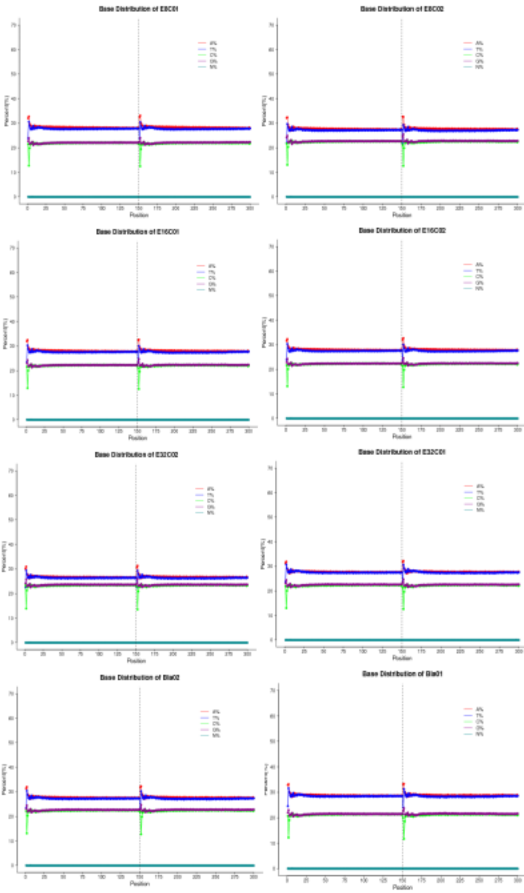

Figure 3

Content of A,T,G,C for reads of 8-cell（E8C01,E8C02）,16-cell（E1601,E1602） Morula（E32C01,E32C02）and early blastocyst（Bla01,Bla02）

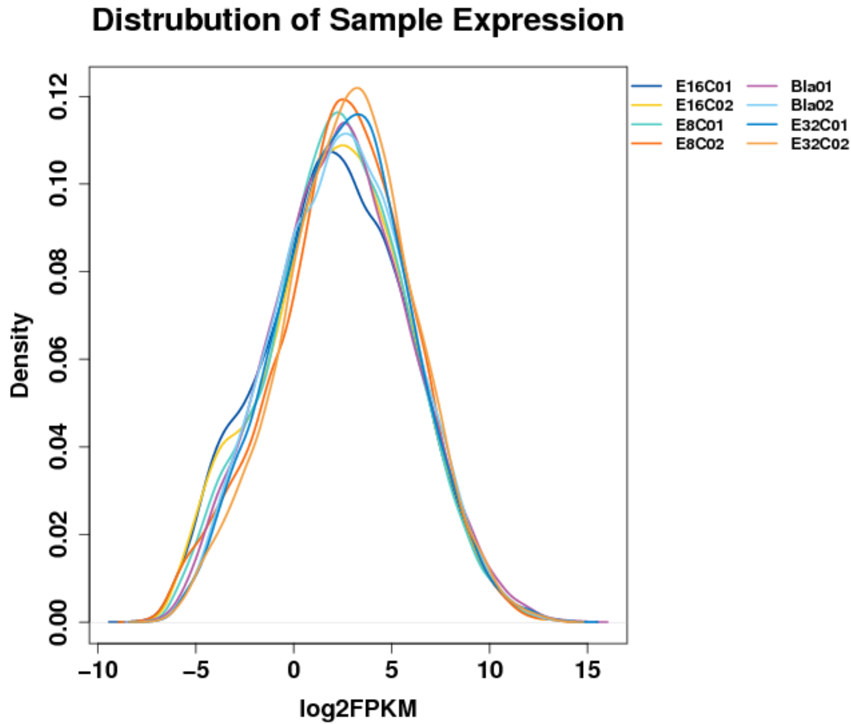

Figure 4

The density distribution of mRNAs was according to log10 (FPKM);

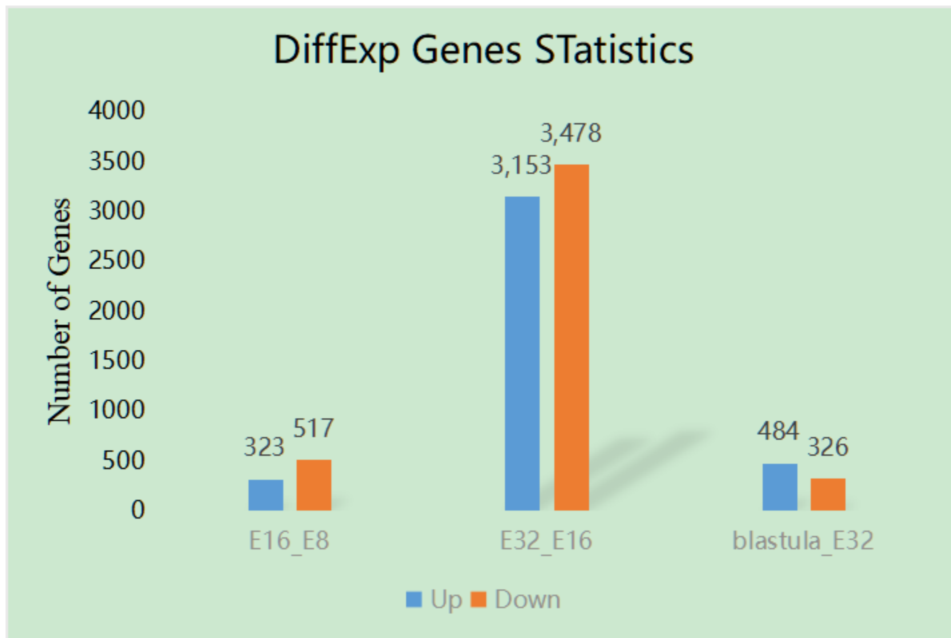

Figure 5

Differential gene statistics at different time points;

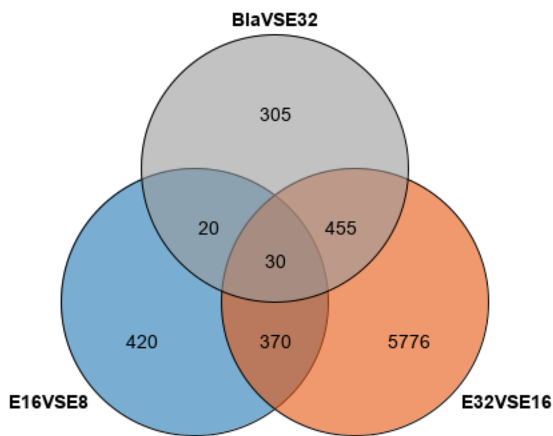

Figure 6

The Venn plot of DEGs

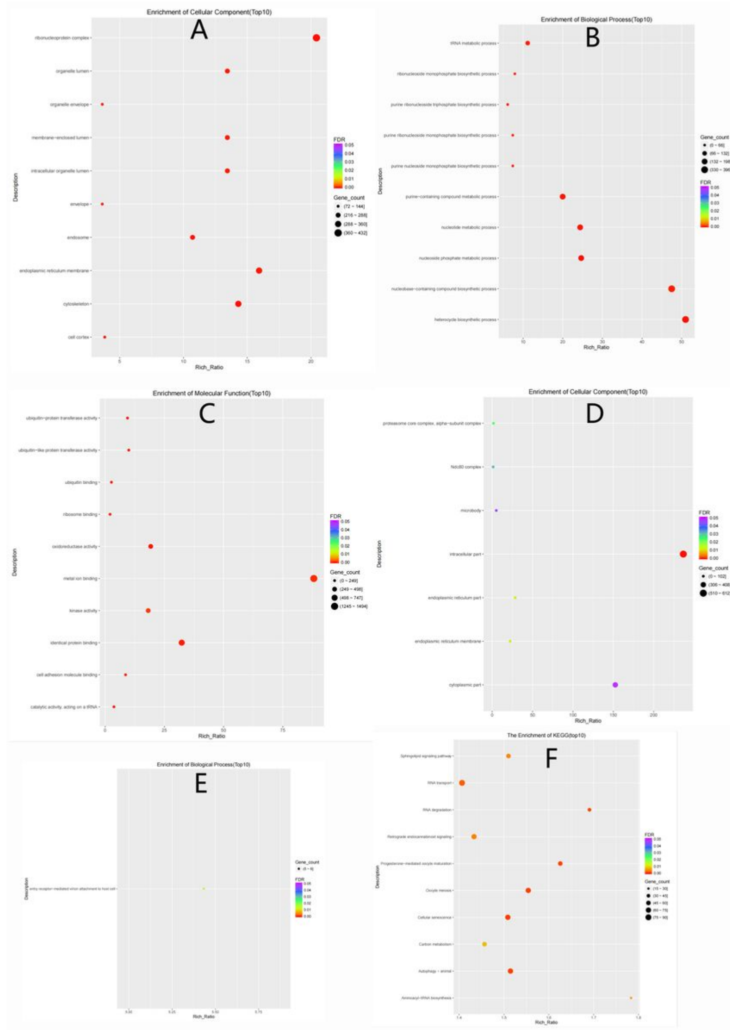

Figure 7

Gene Ontology classification of the DEGs during embryonic development

(A:Morula VS 16 CC enrichment TOP10;B:Morula VS 16 BP enrichment TOP10;C:Morula VS 16 MF enrichment TOP10;D:Bla VS Morula CC enrichment TOP10;E:Bla VS Morula BP enrichment TOP10;F:Morula VS 16 KEGG enrichment TOP10;The larger the bubble, the darker the color, and the higher the enrichment rate.)

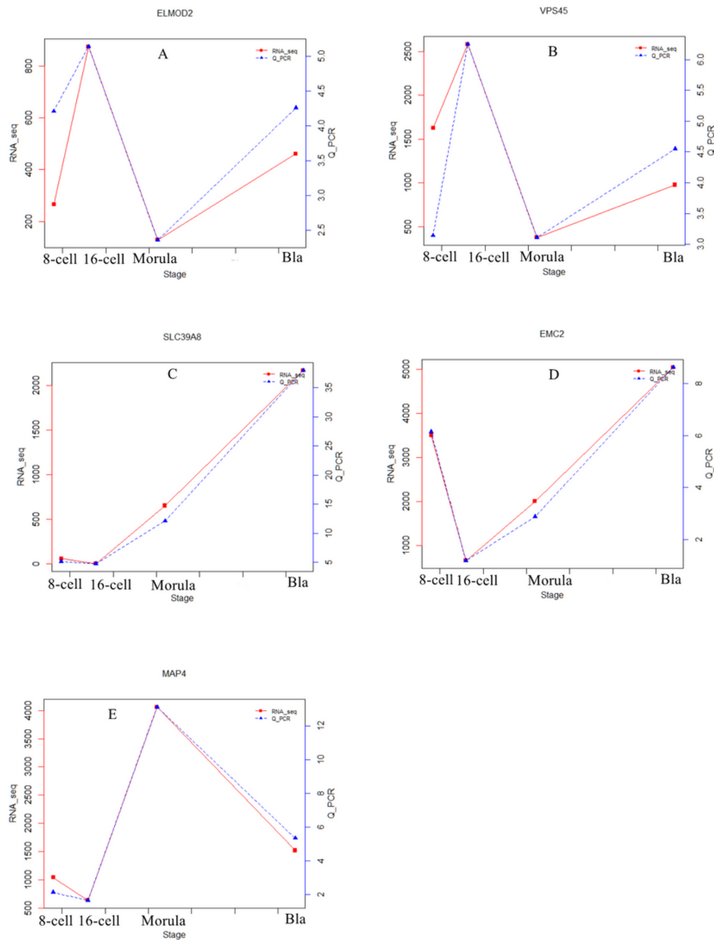

Figure 8

The validation of candidate genes.(A. ELMOD2; B.VPS45; C. SLC39A8; D. EMC2; E. MAP4. Blue means Q-PCR, red means RNAseq GAPDH were used as the reference gene for Q-PCR, RNA-seq relative expression was represent by FPKM)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table8.docx