Transcriptome Analysis of Sheep IVF embryos based on single Cell RNA-Seq

Di Fang
Tarim University
Na Li
Tarim University
Jianqing Zhao
Tarim University
Xue Rui
Tarim University
Nan Li
Tarim University
Qinghua Gao (gsy1997@126.com)
Tarim University

Research Article

Keywords: sheep, in vitro fertilization, embryo, transcriptome, RNA-Seq

Posted Date: December 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-375410/v1

Abstract

Objective: The purpose of the present study was to explore the transcriptome differences of sheep in IVF. Embryos were fertilized in vitro 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. The experiments aimed to promote the development of sheep embryo production technology in vitro.

Methods: The single embryos derived from sheep 16-cells and morulae were produced by IVF technology as samples 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 16-cell and morula embryos were 47355570-50855888, of which 93.17%-94.23% 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 9802-10505 sites of novel transcripts in sheep 16-cells and morulas respectively. A total of 10 genes (including novel transcripts) were expressed at the morula stage.The following screening criteria were used: | log2ratio | ≥1 and Q-value < 0.05. The differences in the 17,607 DEGs between the 16-cell and morula embryos were significant (370 upregulated and 47 downregulated). From the 16-cell stage to the morula stage, 16,343 DEGs were classified and annotated. It was found that the metabolic processes of BP, CC and MF3 focused on biological functions, such as organic synthesis and nucleic acid synthesis. KEGG analysis indicated 206 pathways that were involved in the 16-cell to morula embryo stage. Two pathways were used for enrichment namely the spliceosome and the signaling pathways regulating pluripotency of stem cells, respectively.

Conclusions: 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.

1 Introduction

Embryonic development is the result of the connection and cooperation of several genes expressed in time and space(Cao et al. 2014; Graf et al. 2014a; Jiang et al. 2014; Yan et al. 2013). 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(Morisato and Anderson 1995; Newport and Kirschner 1982). 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(Dean 2002), the bovine in the 8-16 cell stage, the pig in the 4-cell stage and the human in the 4-8 cell stage(Morgan et al. 2005). 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(Dean 2002).

In 2009, Tang et al(Tang et al. 2009) 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.(Yan et al. 2013)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.(Xue et al. 2013) 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(Tong L et al. 2019).

The RNA content of early embryos is low(Xue-Ying Y et al. 2016) and it is difficult to meet the minimum starting amount of RNA required for transcriptome sequencing. However, with the establishment and development of microtranscriptome sequencing techniques, such as Smart-seq2(L L D et al. 2014), previous studies have successfully used RNA-Seq technology to analyze the regulatory mechanism of

early embryonic development in human, pig and bovine species(Tang et al. 2009; Tong L et al. 2019; Xue et al. 2013). 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 IVF16- cell and morula 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 lay the foundation for further improvement of sheep IVF.

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 Collection of ovaries

Ovaries were collected from the Aksu slaughterhouse. Following slaughtering, the adult ewes had both their ovaries removed, which were sprayed with 75% ethanol. The samples were incubated at 38°C in the presence of sterilized normal saline [plus double antibody (green streptomycin mixture)] and transported back to the laboratory within 2-3 h. The fallopian tubes, mesentery and other excess tissues on the ovary were cut off, washed once with normal saline and preheated at 38℃ in advance. The samples were quickly washed 3 times with 75% ethanol and finally 3 times with normal saline at 37°C. The samples were placed in a 37°C water bath without passing through the ovarian beaker.

2.4 COCs collection

The egg collection fluid was preheated in a 38°C water bath 2 h in advance and approximately 10 ovaries were obtained. The water was removed on the filter paper and placed into petri dishes (60 mm) containing the egg collection fluid. Follicles that were 2-6 mm in diameter were scratched with a disposable surgical blade. A small amount of egg collection fluid was absorbed by a 5-ml syringe in order to rinse the ovary cutting site and facilitate the follicle fluid flow. The samples were left to precipitate for a few min and the eggs were visualized using a microscope. Grade A and B COCs were selected for mature culture.

2.5 COCs culture

In the super-clean platform, the mature culture droplets were transferred to a 35 mm petri dish and 20 μl mature liquid was absorbed by the liquid transfer gun. The spacing was controlled at the bottom of the petri dish in order to produce microdroplets. 3-4 ml sterilized paraffin oil was added and 50 μl mature liquid was mixed with the original droplets after the paraffin oil was paved. The selected grade A and B COCs were washed with IVM solution 3 times and transferred into pre-balanced droplets. Each droplet was cultured in a carbon dioxide incubator (5% CO_2, 38.5°C, saturated humidity) for 24 h. The diffusion of granulosa cells was observed. The loosely arranged COCs of the granulosa cells was transferred to PBS containing 0.1% hyaluronidase through an oral straw and 200 μl of the liquid was repeatedly blown to COCs. When 1-2 layers of the granulosa cells were preserved, the blowing was terminated and the samples were rapidly transferred to the IVF solution. Following three times washing in the IVF solution, the first polar body excretion was observed. The oocytes excluded by the first polar body were used for *in vitro* fertilization.

2.5.1 In vitro fertilization

The 0.5 ml diluted semen was slowly injected into the diluted semen from the bottom of the centrifuge containing 1 ml sperm capacitation solution. The diluted semen was incubated in a constant temperature incubator (38.5°C, 5% CO_2, 100% humidity) for 30 min. The supernatant was centrifuged for 5 min and discarded. This process was repeated. The mature oocytes with 1-2 layers of granulosa cells were transferred into 50 μl pre-balanced fertilized droplets following three times cleaning of the fertilized fluid, with a maximum of 15 COCs per droplet. The upstream semen was added to the fertilization droplets (the sperm density was approximately $1*10^6$ /ml). The sperm and eggs were incubated in a CO_2 incubator for 18 h (38.5°C, 5% CO_2, 100% humidity).

2.6 Removal of zona pellucida

When the embryo was transferred into 10 μl PE Streptomyces protease microdroplets with an oral straw and incubated at 50°C for 30 s at the final concentration of 5 mg PE, the gap between the zona pellucida and the blastomere was enlarged and slightly deformed. Washing of the embryo was performed in 2 ml aseptic PBS buffer (5% FBS+2% PS).The zona pellucida was gently washed with an oral straw (smaller than the diameter of the embryo) 5 times in DPBS. The 16-cell and morula embryos were collected and transferred to 10 μl lysates and liquid nitrogen was injected into the embryos within 2 min. Following freezing for several h, 2 repeats were used for each group. The 16-cells were marked as E1601 and E1602 and the morulae were marked as EM01 and EM02. The samples were transported by dry ice and sent to An Nuoda Genome Technology (Beijing) Co. Ltd.

2.7 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.

2.8 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 (gb) morula stage of the 16-cell stage was obtained by calculating the expression amount of two biological repeats in the (gc) stage of the morula of the 16-cell 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(Yoav B and H. 1995), 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).

2.9 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)(Li et al. 2009). The sequenced sequences of the aligned genomes were

assembled and spliced by the Cufflinks software (v2.2.1)(Trapnell et al. 2010). 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(Roberts et al. 2011).

3 Results

3.1 Sequencing quality Evaluation and basic data Analysis

The embryo libraries of sheep 16-cells and morulas 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 four 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 four samples was estimated to 93.18-94.02%, 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-rG of the four samples basically coincided with each other, indicating that the base composition was stable and balanced and that the sequencing quality was high (Fig. 2).

3.2 Analysis of sequencing results

Following filtration of the data of the original sequence, the filtered sequence of the sheep embryos at the two developmental stages of the 16-cell and morula embryos exhibited the following range: 47355570-50855888. The obtained clean reads were compared with the reference genome by the TopHat software(Qiu et al. 2012; Trapnell et al. 2009). The results indicated 93.17-94.23% clean read alignment of the sheep reference genes in each stage and the sequence ratio of the multiple locations of the genome was 4.35-5.44%, which satisfied the requirements (Table 1). The number of transcripts and predicted new transcripts compared at each stage of development is shown in Table 2. 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 3). SNP analysis by Samtools indicated the presence of 171170-211487 loci in each developmental stage of sheep embryos from the 16-cell to the morula embryos. The novel transcripts were estimated to 9802-10505 (Table 4).

3.3 Gene expression characteristics and differentially expressed genes in early embryos

Based on the read value of the gene expression levels, the RPKM values of the genes at different stages of embryonic development were calculated by the RPKM method(Wagner et al. 2012).By comparison, 417 DEGs were identified in the 16-cell embryos compared with those of the morulae, including 47 downregulated genes and 370 upregulated genes. In order to directly reflect the difference in the multiple and significant levels of the differentially expressed genes, a volcano map was drawn (Fig. 3). Several differentially expressed genes were noted between the 16-cell and the morula embryos and the differentially expressed genes were mainly between-lg (padj) 2-5. The genes with significant differences of-lg (padj)>5 were the upregulated genes. The maternally-associated genes of sheep 16-cell and morula embryos were the following: KIT, STAT3, ZP4, BMP15, GDF9, ZP3, Pms2, Hsf1-2 and Zar1. The methylation-associated genes were the following: DMNT3B, DMNT1 and DMNT3A. The proliferation-related genes were as follows: IGF2, TRIM28 and IGF1 and the stem cell-related genes were NANOG and KLF4. The FPKM values of ATP5F1B, SNURF, UBE3A, MAPK13, JUP, ACO2, TXN2 and SARS were obtained as the logarithm of 10 and the FPKM clustering heat map (Fig. 4) was drawn by the heml software. The maternal genes exhibited the following patterns: low KIT gene expression, high BMP15, GDF9, ZP4 gene expression. Minor changes in the expression levels were noted in each period. ZP3 and STAT3 gene expression levels were slightly altered in each period, whereas the methylation-related gene expression was stable. The expression of TRIM28 was increased gradually and the expression of NANOG and KLF4 genes was increased gradually.

3.4 GO enrichment and KEGG Pathway Analysis of differentially expressed genes

GO analysis indicated that 16,343 DEGs were classified and annotated from the 16-cell stage to the morula development stage, all of which involved Biological Process (BP), Cellular Components (CC) and Molecular Function (MF). These were enriched to 88 secondary items (Fig. 5). A total of 78 second-level entries were noted in the BP classification, of which the top 10 entries involved regulation of the metabolic process, nucleobase containing compound biosynthetic process, negative regulation of cellular process, organic cyclic compound biosynthetic process, RNA metabolic process, negative regulation of biological process, positive regulation of cellular process, positive regulation of biological process, regulation of cellular process and regulation of biological process. The two secondary items enriched in CC were the spliceosomal complex and the Intracellular component. The eight secondary entries enriched in the MF classification were the transcriptional repressor activity, the RNA polymerase II proximal promoter sequence-specific DNA binding, the transcriptional repressor activity, the RNA polymerase II transcription regulatory region sequence-specific DNA binding, the transcription factor activity, the RNA polymerase II proximal promoter sequence-specific DNA binding, the transcription factor binding, the transcription factor activity, the protein binding, the transcription factor activity, the transcription factor binding, the RNA polymerase II transcription factor activity, the sequence-specific DNA binding and the transcription regulatory activity. According to the statistics of the secondary GO items, significant differences in the DEG annotation results were noted (Fig. 6). It was found that the metabolic processes of BP, CC and MF3 were focused on biological functions, such as value-added, organic synthesis and nucleic acid synthesis. KEGG analysis of differentially expressed genes during sheep embryonic development indicated 206 pathways involved in the morula developmental stage (from 16-cell to morula

stage), whereas the two main pathways with significant enrichment were the Spliceosome and the signaling pathways regulating pluripotency of stem cells, respectively.

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% (Raj A et al. 2006). 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(Gilbert et al. 2009). 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(Mortazavi et al. 2008). 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(Picelli et al. 2013).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.

4.1 Characteristics of early embryonic Development recording Group of Sheep

Following filtration of the data of the original sequencing sequence obtained from Illumina HiSeqXten sequencing, the filtered sequence of sheep embryos at the two developmental stages of 16-cell and morula was 47355570-50855888. Using the TopHat software to compare the obtained Clean reads with the reference genome, the results indicated a range of 93.17-94.23% clean reads alignment on the sheep reference gene in each stage. The sequence ratio of (multi map rate) multiple locations of the genome was 4.35-5.44%, 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 16-cells and morulae was noted, indicating that alternative splicing is common in eukaryotes and that the splicing mode of genes is constantly changing at different stages of embryonic development to regulate cell proliferation, differentiation, migration and apoptosis. 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. A total of 10 newly activated genes (including novel transcripts 2) were found in morula compared with the 16-cell embryos, which were as follows: TMEM37, PLCD1, ITGB1BP1, CCND2, PROCR, C14orf28, PRMT2, KDM3A, ENSOARG00000009515 and ENSOARG00000002677. Cyclin D (CCNDs) plays an important role in the process of the cell cycle. It is a rate-limiting factor from the G1 phase to the S phase. It mainly mediates the transmission between extracellular signal molecules and the cell mitotic cycle. By binding to cyclin-dependent kinase 4/6 (CDK4/6) and activating its active complex to phosphorylate the downstream retinoblastoma protein (RB), cyclin D plays an important role in the process of the cell cycle. Subsequently, the release of the transcription factor E2F initiates DNA replication(Coqueret 2002; Malumbres and Barbacid 2009; Zhang et al. 2014). Previous studies have found that the positive regulation of the expression of CCND2 can promote the process of the cell cycle, thus accelerating the proliferation of granulosa cells. In contrast to these observations, the downregulation of the expression of CCND2 causes an inhibition in the proliferation of granulosa cells(Summers et al. 2014; Zhang et al. 2013). Previous studies on CCNDs were mainly performed on mammalian follicles.

Current studies have shown that PRMT2 participates in several biological processes, including immune response, inflammatory response, apoptosis and proliferation. This gene acts mainly through the synergistic action of different regulatory factors, which suggests that it can bind to different proteins and play a transcriptional regulatory role through a variety of mechanisms(Belton et al. 2012).Endothelial protein C receptor, (EPCR) is encoded by the PROCR gene, which participates in the activating protein C pathway and plays a role in cell protection, anti-inflammation and anticoagulation.

Yagisawa et al.(Pendurthi and Rao 2018) confirmed for the first time that PLCD1 had a nuclear output (NES) and nuclear input (NLS) sequence, suggesting that it could shuttle between the cytoplasm and the nucleus. During the cell cycle, the Ca^{+2} influx stimulates the increase in the formation of Ca^{+2}-dependent PLCD1-import in β1 nuclear input complexes, resulting in the accumulation of cell cycle-dependent PLCD1 in the nucleus(Yagisawa et al. 1998). The inhibition of PLCD1 activity leads to the increase of nuclear PIP2 levels and changes the transcriptional regulation and the degradation of cyclin E, thus affecting cell proliferation. In certain cases, overexpression of this protein leads to the prolongation of the G1/S and S phases. These studies suggest that PLCD1 has a specific and important function in cell cycle regulation(Cocco et al. 2006).

4.2 Activation of the embryonic genome

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(Graf et al. 2014b; Stallings et al. 2008). 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.

4.3 GO Annotation and KEGG Analysis of differentially expressed genes in embryos

The number of DEGs of the two developmental stages from the 16-cell to the morula stage was 16,343, indicating significant temporal differences in the developmental regulatory mechanism of sheep embryos at different developmental stages. GO analysis indicated that 16,343 DEGs were classified and annotated from the 16-cell stage to the morula developmental stage. All these genes involved BP, CC and MF that were enriched to 88 secondary items (Fig.5).A total of 78 second-level items were noted in the BP classification, among which the first 10 items were mainly associated with metabolic process regulation, cell process regulation and organic synthesis regulation, indicating that this period was the period of rapid embryo proliferation. Additional processes associated with the regulation of the metabolic process were the following: negative regulation of cellular process, positive regulation of cellular process, negative regulation of biological process and positive regulation of biological process. The data suggested that this mode of action could ensure the orderly and controlled proliferation. Among the two secondary items enriched by CC, the eight secondary items enriched by the spliceosomal complex and the intracellular component in MF classification were concentrated in the DNA enzyme, RNA enzyme and transcription factor components, indicating that the gene expression was high at that time and that the embryo was proliferating rapidly. From these data, we can infer that during cleavage, the embryo proliferates rapidly. Concomitantly, a mechanism exists that regulates the proliferation rate in the embryo and ensures the prevention of cell cycle block or stall and the orderly proliferation of the embryonic cells. According to the statistics of the secondary GO items, significant differences were noted in the analysis of DEG annotation (Fig.6). It was found that BP, CC and MF3 were enriched in the functions of cell proliferation, immunity, stress, signal transduction and synapse, indicating that the embryos in this period were in a period of rapid proliferation. The enrichment of pigmentation was noted. It is considered that the enrichment of this process is related to the *in vitro* culture, which provides a basis for further optimization of the *in vitro* culture system.

The two significant pathways in the KEGG analysis were the spliceosome and signaling pathways that regulated pluripotency of stem cells. The standard spliceosome is made up of five small nuclear

ribonucleoproteins (snRNPs), namely U1,U2,U4,U5and U6 and several spliceosome-associated proteins (SAPs). Among them, a few studies have been performed on six downregulated genes, such as TCEPRG1, SYF1 and PRPF19 and their functional characteristics have not been found. The discovery of the Nanog gene has recently become one of the greatest advances in maintaining the undifferentiated state of embryonic stem cells and identifying genes related to embryonic development. In the early stage of mouse embryonic development, the Nanog gene is not expressed in the early cleavage stage and trophoblast formation(Mitsui et al. 2003). The earliest expression was in the morula stage and was limited to the inner cell mass. Some *in vivo* experiments have shown that the Nanog gene is not present in undifferentiated cell mass. Instead, it is responsible for the differentiation and production of endoderm cells. The Nanog gene is a gene necessary to maintain the totipotency of inner cell mass following Oct4 initiation in the early stage of embryonic development(Yamaguchi et al. 2005). In addition, the Nanog gene is highly expressed in embryonic stem cells. When embryonic stem cells differentiate, the expression of this gene is significantly downregulated. Genetic defects in embryonic stem cells are associated with abnormal Nanog expression. This can lead to the loss of pluripotency and the initiation of differentiation. The maintenance of totipotency and self-proliferation of embryonic stem cells by the Nanog gene is mainly facilitated via binding to the regulatory region of the target genes, selectively and by inhibiting the expression of differentiation genes or promoting the expression of pluripotent genes. Therefore, the high expression of the Nanog gene is closely associated with the maintenance of totipotency of embryonic stem cells, while cell differentiation is associated with the downregulation of the Nanog gene expression.

5 Conclusions

In the present study, the transcriptome of different stages of sheep single IVF embryo development was analyzed by single cell RNA-Seq for the first time and several differential genes and related pathways were obtained. Differentially expressed genes at different stages have their own specificity in quantity, function, classification and metabolic pathway. The results indicated that the period required from the 16-cell to the morula stage was the period of rapid embryo proliferation. Its metabolic pathway focused on the synthesis and regulation of substances related to cell proliferation and it was found that there was enrichment in both positive and negative aspects of the same process involved in metabolic regulation. This in turn prevented cleavage block or stall and ensured the order of embryo proliferation. Concomitantly, the present study provided a theoretical basis for improving the technology of sheep embryo production *in vitro*. This can provide additional information on sheep gene structure and on new genes related to embryonic development.

Declarations

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).

Conflict of interest

The authors declare that they have no conflict of interest.

Ethics 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.

Consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

All data generated or analyzed during this study are included in this published article/as supplementary information files.

Code availability

Not applicable

Authors' contributions

Conceptualization: QHG; Data curation: DF, NL; Formal analysis: DF; Funding acquisition: QHG; Investigation: NL1; Methodology: DF, NL; Project administration: QHG; Resources: XR, JQZ; Visualization: DF, NL; Roles/Writing - original draft: DF, Writing - review & editing: JQZ, XR, NL1; All Authors read and approved the manuscript.

References

Belton A, Gabrovsky A, Bae YK, Reeves R, Iacobuzio-Donahue C, Huso DL, Resar LM (2012) HMGA1 induces intestinal polyposis in transgenic mice and drives tumor progression and stem cell properties in colon cancer cells PloS one 7:e30034 doi:10.1371/journal.pone.0030034

Cao S et al. (2014) Specific gene-regulation networks during the pre-implantation development of the pig embryo as revealed by deep sequencing BMC genomics 15:4 doi:10.1186/1471-2164-15-4

Cocco L, Faenza I, Fiume R, Maria Billi A, Gilmour RS, Manzoli FA (2006) Phosphoinositide-specific phospholipase C (PI-PLC) beta1 and nuclear lipid-dependent signaling Biochimica et biophysica acta 1761:509-521 doi:10.1016/j.bbalip.2006.03.001

Coqueret O (2002) Linking cyclins to transcriptional control Gene 299:35-55 doi:10.1016/s0378-1119(02)01055-7

Dean J (2002) Oocyte-specific genes regulate follicle formation, fertility and early mouse development Journal of reproductive immunology 53:171-180 doi:10.1016/s0165-0378(01)00100-0

Gilbert I, Scantland S, Sylvestre EL, Gravel C, Laflamme I, Sirard MA, Robert C (2009) The dynamics of gene products fluctuation during bovine pre-hatching development Molecular reproduction and development 76:762-772 doi:10.1002/mrd.21030

Graf A, Krebs S, Heininen-Brown M, Zakhartchenko V, Blum H, Wolf E (2014a) Genome activation in bovine embryos: review of the literature and new insights from RNA sequencing experiments Animal reproduction science 149:46-58 doi:10.1016/j.anireprosci.2014.05.016

Graf A, Krebs S, Zakhartchenko V, Schwalb B, Blum H, Wolf E (2014b) Fine mapping of genome activation in bovine embryos by RNA sequencing Proceedings of the National Academy of Sciences of the United States of America 111:4139-4144 doi:10.1073/pnas.1321569111

Jiang Z et al. (2014) Transcriptional profiles of bovine in vivo pre-implantation development BMC genomics 15:756 doi:10.1186/1471-2164-15-756

L L D, R X X, Y. LW (2014) RNA-Seq analysis of yak ovary: improving yak gene structure information and mining reproduction-related genes Sci China Life Sci 2014 44:307-317

Li H et al. (2009) The Sequence Alignment/Map format and SAMtools Bioinformatics (Oxford, England) 25:2078-2079 doi:10.1093/bioinformatics/btp352

Malumbres M, Barbacid M (2009) Cell cycle, CDKs and cancer: a changing paradigm Nature reviews Cancer 9:153-166 doi:10.1038/nrc2602

Mitsui K et al. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells Cell 113:631-642 doi:10.1016/s0092-8674(03)00393-3

Morgan HD, Santos F, Green K, Dean W, Reik W (2005) Epigenetic reprogramming in mammals Human molecular genetics 14 Spec No 1:R47-58 doi:10.1093/hmg/ddi114

Morisato D, Anderson KV (1995) Signaling pathways that establish the dorsal-ventral pattern of the Drosophila embryo Annual review of genetics 29:371-399 doi:10.1146/annurev.ge.29.120195.002103

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

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 30:675-686 doi:10.1016/0092-8674(82)90272-0

Pendurthi UR, Rao LVM (2018) Endothelial cell protein C receptor-dependent signaling Current opinion in hematology 25:219-226 doi:10.1097/moh.0000000000000416

Picelli S, Björklund Å K, Faridani OR, Sagasser S, Winberg G, Sandberg R (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells Nature methods 10:1096-1098 doi:10.1038/nmeth.2639

Qiu Q et al. (2012) The yak genome and adaptation to life at high altitude Nature genetics 44:946-949 doi:10.1038/ng.2343

Raj A, Peskin C S, D. T (2006) Stochastic mRNA synthesis in mammalian cells PLoS Biol 4:e309 doi:10.1371/journal.pbio.0040309

Roberts A, Pimentel H, Trapnell C, Pachter L (2011) Identification of novel transcripts in annotated genomes using RNA-Seq Bioinformatics (Oxford, England) 27:2325-2329 doi:10.1093/bioinformatics/btr355

Stallings JD, Zeng YX, Narvaez F, Rebecchi MJ (2008) Phospholipase C-delta1 expression is linked to proliferation, DNA synthesis, and cyclin E levels The Journal of biological chemistry 283:13992-14001 doi:10.1074/jbc.M800752200

Summers AF et al. (2014) Altered theca and cumulus oocyte complex gene expression, follicular arrest and reduced fertility in cows with dominant follicle follicular fluid androgen excess PloS one 9:e110683 doi:10.1371/journal.pone.0110683

Tang F et al. (2009) mRNA-Seq whole-transcriptome analysis of a single cell Nature methods 6:377-382 doi:10.1038/nmeth.1315

Tong L, Wei X J, Y. PS (2019) Application of Single-Cell Transcriptional Sequence in Study of Reproductive Development J Int Reprod Health/Fam Plan 38:217-221

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq Bioinformatics (Oxford, England) 25:1105-1111 doi:10.1093/bioinformatics/btp120

Trapnell C et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation Nature biotechnology 28:511-515 doi:10.1038/nbt.1621

Wagner GP, Kin K, Lynch VJ (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples Theory in biosciences = Theorie in den Biowissenschaften 131:281-285 doi:10.1007/s12064-012-0162-3

Xue-Ying Y, Qin-Qin G, H. H-S (2016) Transcriptome of Bovine IVF Embryos Treated with Glutathione Acta Veterinaria Et Zootechnica Sinica 47:1363-1372 doi:10.11843/j.issn.0366-6964.2016.07.008

Xue Z et al. (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing Nature 500:593-597 doi:10.1038/nature12364

Yagisawa H et al. (1998) Replacements of single basic amino acids in the pleckstrin homology domain of phospholipase C-delta1 alter the ligand binding, phospholipase activity, and interaction with the plasma membrane The Journal of biological chemistry 273:417-424 doi:10.1074/jbc.273.1.417

Yamaguchi S, Kimura H, Tada M, Nakatsuji N, Tada T (2005) Nanog expression in mouse germ cell development Gene expression patterns : GEP 5:639-646 doi:10.1016/j.modgep.2005.03.001

Yan L et al. (2013) Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells Nature structural & molecular biology 20:1131-1139 doi:10.1038/nsmb.2660

Yoav B, H. Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing Journal of the Royal Statistical Society Series B (Methodological) 57

Zhang Q, Sakamoto K, Wagner KU (2014) D-type Cyclins are important downstream effectors of cytokine signaling that regulate the proliferation of normal and neoplastic mammary epithelial cells Molecular and cellular endocrinology 382:583-592 doi:10.1016/j.mce.2013.03.016

Zhang Q et al. (2013) MicroRNA-181a suppresses mouse granulosa cell proliferation by targeting activin receptor IIA PloS one 8:e59667 doi:10.1371/journal.pone.0059667

Tables

Table 1 RNA-Seq and mapping to the reference genome

	16-cell-1 E16	16-cell-2 E16	Morula-1 EM	Morula-2 EM
Clean Reads	67,939,390	58,488,860	53,327,656	53,723,168
Mapped Reads	64,038,390	55,115,234	49,685,950	50,358,084
Mapping Rate	0.9426	0.9423	0.9317	0.9374
UnMapped Reads	3,901,000	3,373,626	3,641,706	3,365,084
MultiMap Reads	3,181,456	2,783,301	2,318,186	2,924,575
MultiMap Rate	0.0468	0.0476	0.0435	0.0544

Table 2 Detected transcripts and novel transcripts

Embryos	No. of transcripts	No. of novel transcripts
16-cell-1 E16	14850	10020
16-cell-2 E16	14848	10505
Morula-1 EM	14013	9802
Morula-2 EM	14261	10107

Table 3 Alternative splicing events detected during embryonic development

AS_type	16-cell-1 E16	16-cell-2 E16	Morula-1 EM	Morula-2 EM
TSS	27,179	27,452	23,972	25,903
TTS	25,390	25,677	23,184	24,490
SKIP	5,028	6,148	5,770	7,464
AE	1,956	2,438	2,217	2,662
MSKIP	1,018	1,286	2,082	1,976
XSKIP	1,116	1,472	1,206	1,530
IR	584	778	876	806
XMSKIP	606	708	758	788
XAE	794	702	751	598
XIR	156	218	192	234
MIR	16	24	48	28
XMIR	10	6	12	2

Table 4 Putative SNPs detected during embryonic development

Sample	16-cell-1 E16	16-cell-2 E16	Morula-1 EM	Morula-2 EM
SNP	199,252	185,622	159,986	185,523
InDel	12,235	11,558	11,184	12,468
Total	211,487	197,180	171,170	197,991

Figures

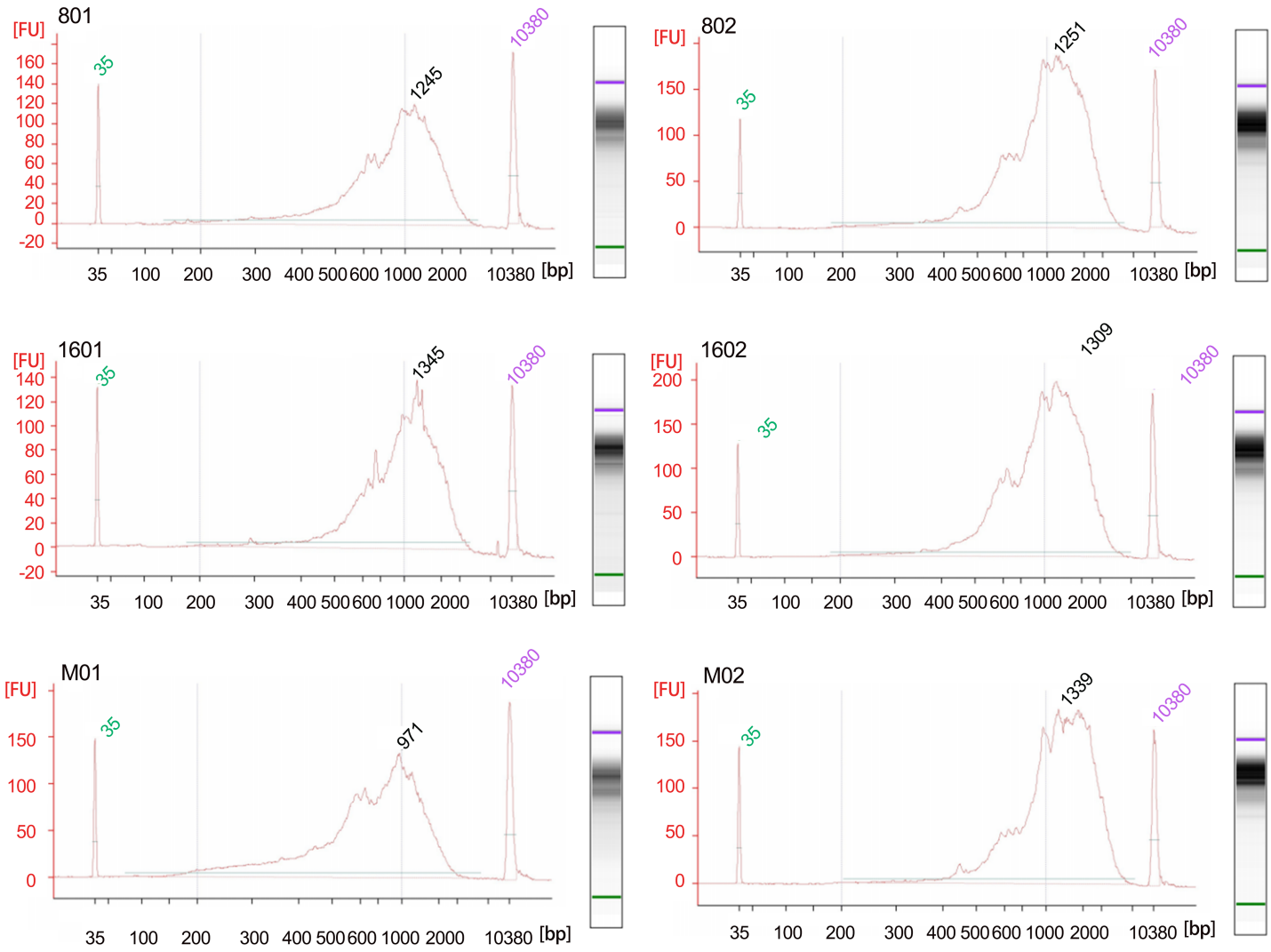

Figure 1

peak figure of Agilent 2100 for cDNA integrity detection by Smart-Seq2 PCR amplification Note: Figure 1601, 1602:16- cell cDNA peak diagram; M01 and M02: the cDNA peak diagram of morula

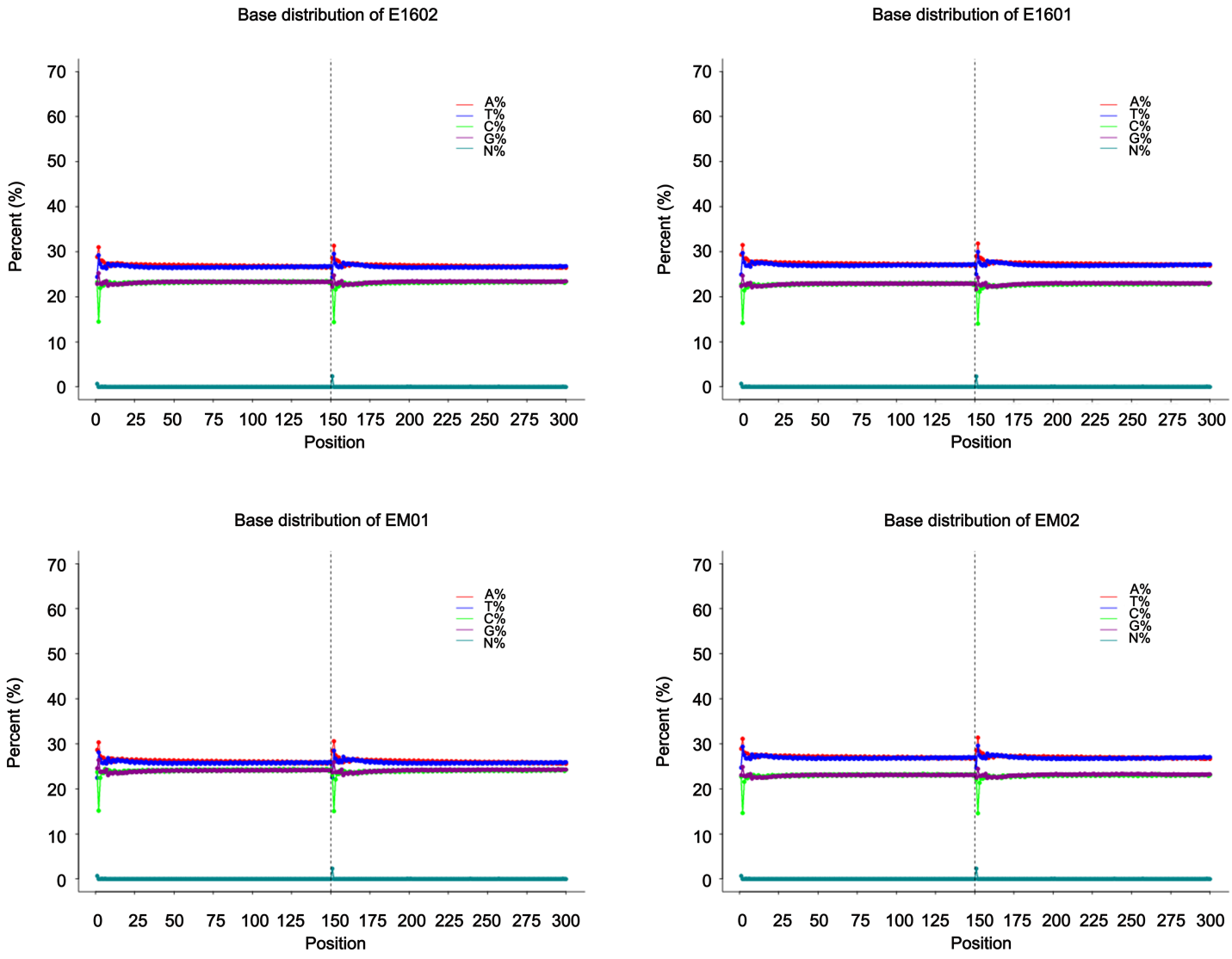

Figure 2

Content of A,T,G,C for reads of 16-cell（E1601、E1602） and morula（EM01、EM02）

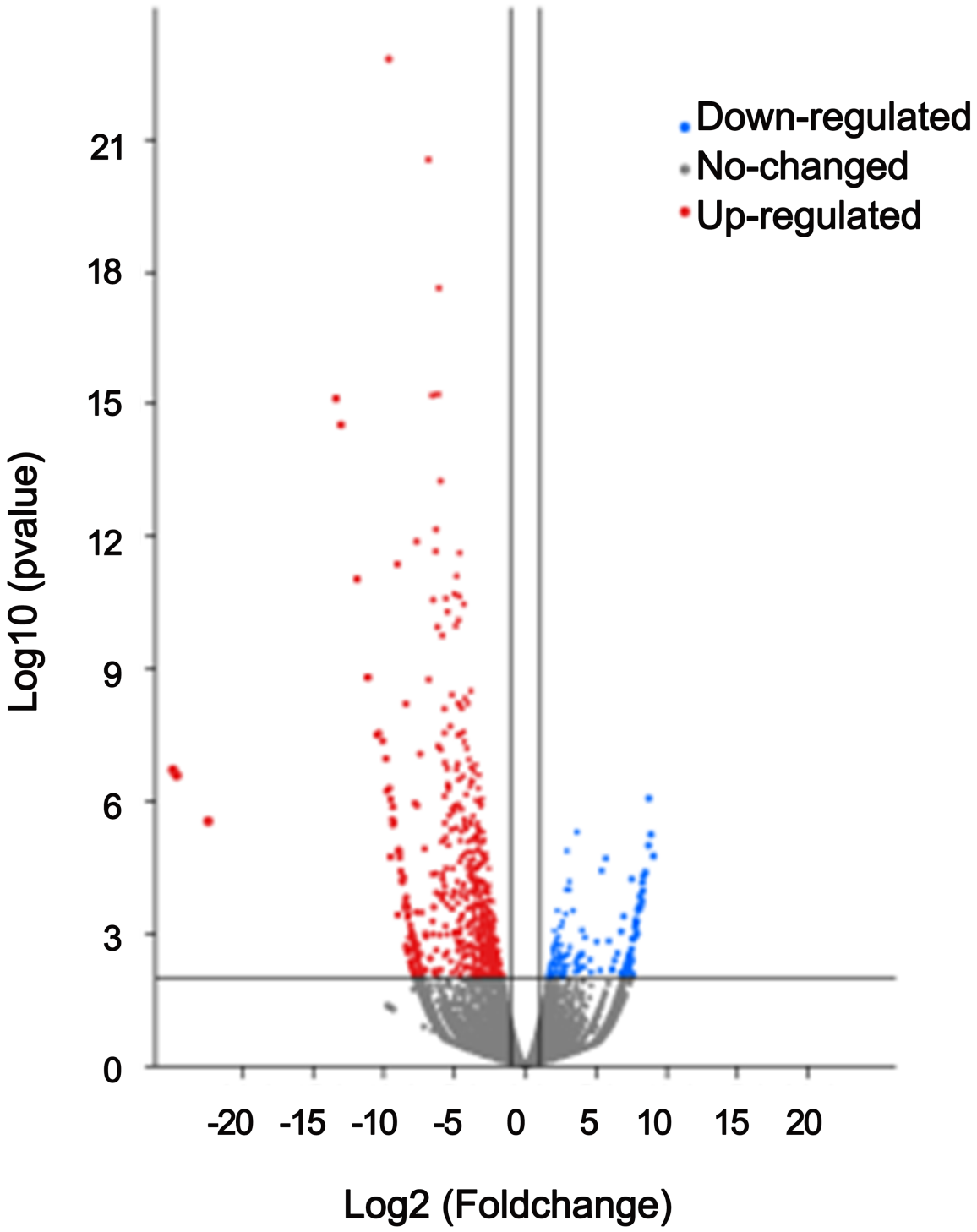

Figure 3

volcanic diagram of sheep IVF 16-cell and morula differentially expressed genes (16-cell VS morula abscissa shows the changes in expression multiples in different experimental groups/samples, and the ordinate shows the statistically significant changes in expression levels. Different colors represent different classifications).

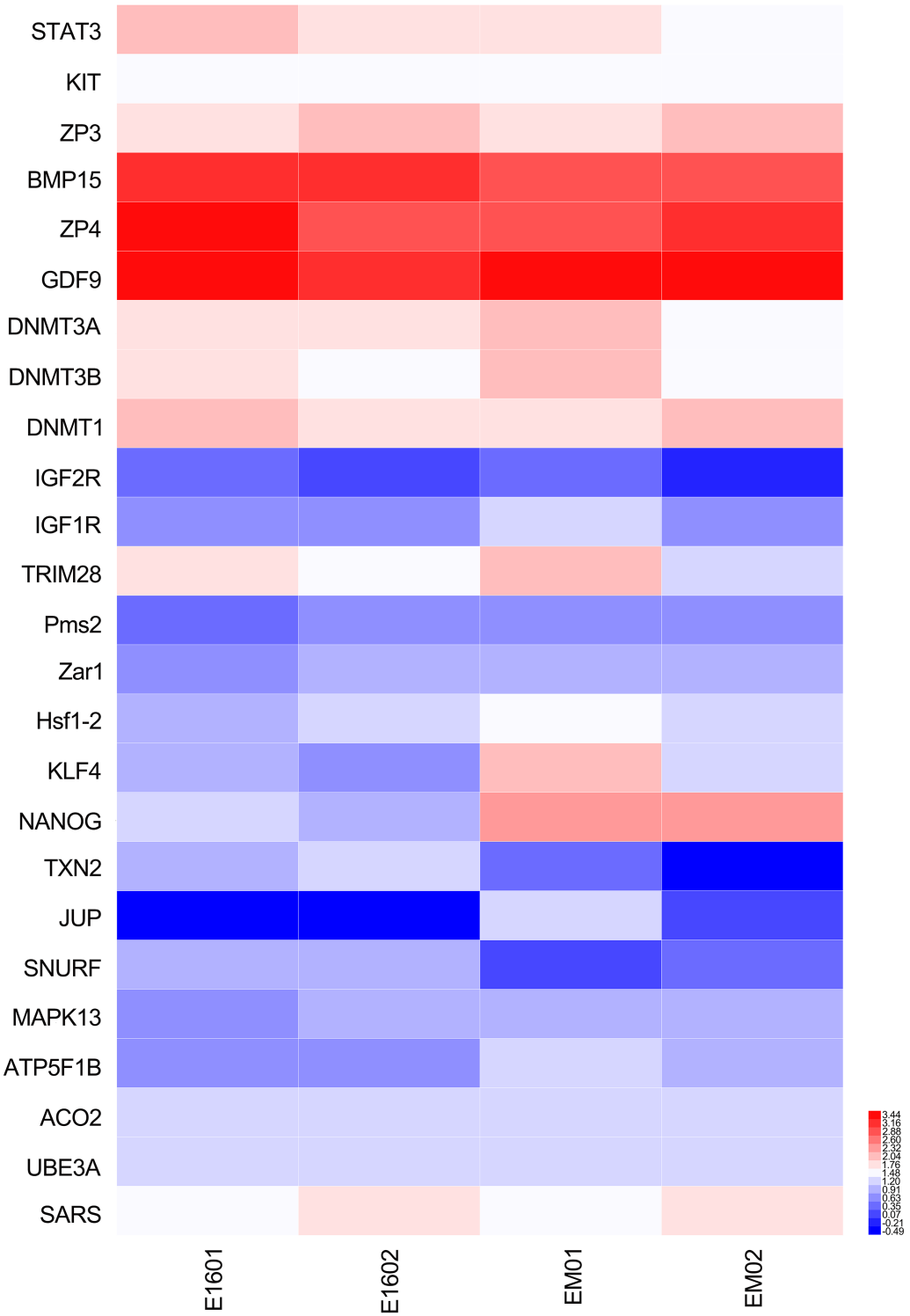

Figure 4

FPKM heat accumulation pattern of some genes in different stages of ovine cleavage (The maternally-associated genes : KIT, STAT3, ZP4, BMP15, GDF9, ZP3, Pms2, Hsf1-2 and Zar1. The methylation-associated genes were the following: DMNT3B, DMNT1 and DMNT3A. The proliferation-related genes were as follows: IGF2, TRIM28 and IGF1 and the stem cell-related genes were NANOG and KLF4. The

FPKM values of ATP5F1B, SNURF, UBE3A, MAPK13, JUP, ACO2, TXN2 and SARS were obtained as the logarithm of 10 and the FPKM clustering heat map was drawn by the heml software).

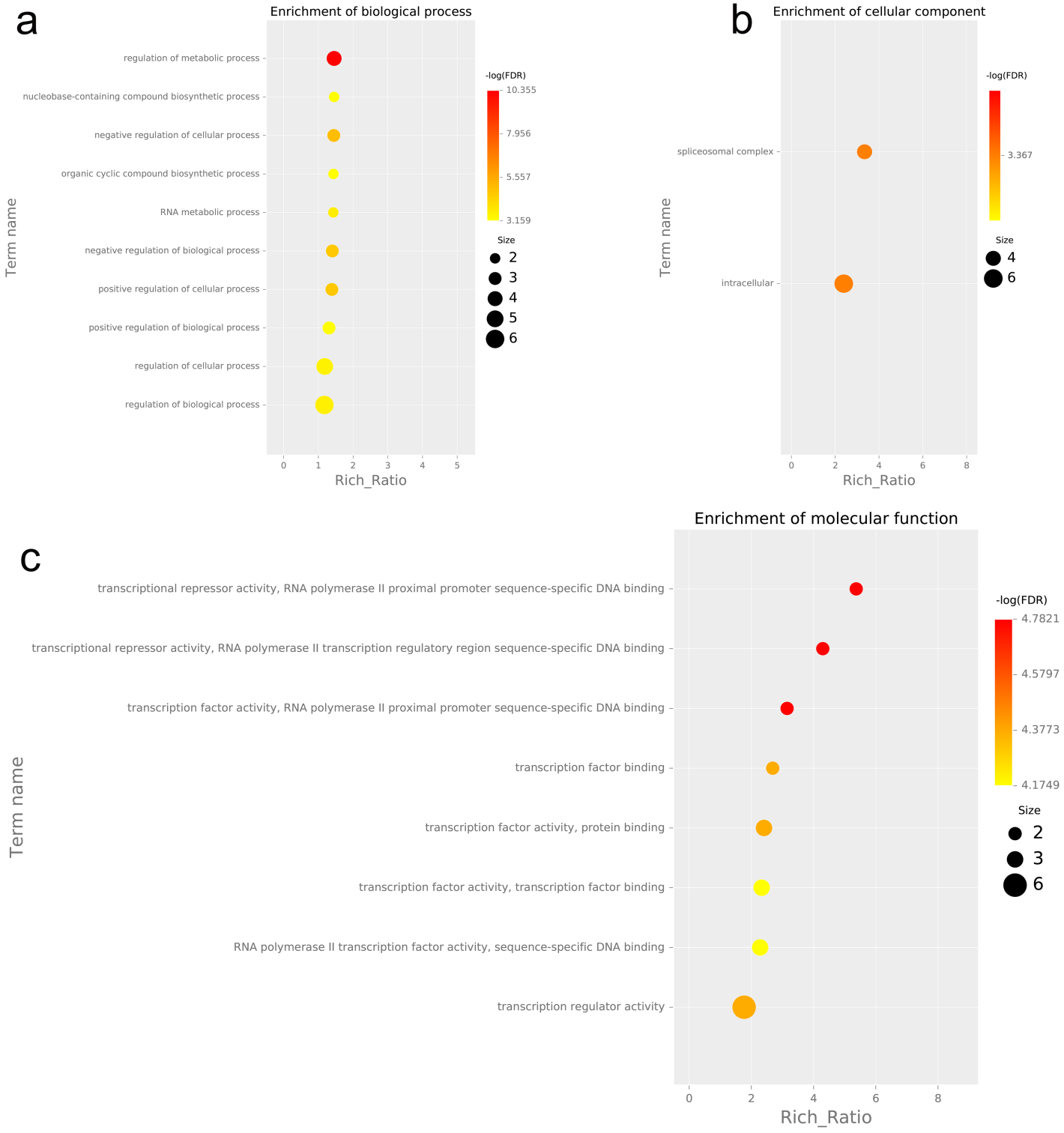

Figure 5

Gene Ontology classification of the DEGs during embryonic development

(figure a:16-cell VS morula BP enrichment;Figure b: 16-cell VS morula CC enrichment;c:16-cell VS morula MF enrichment;The larger the bubble, the darker the color, and the higher the enrichment rate.)

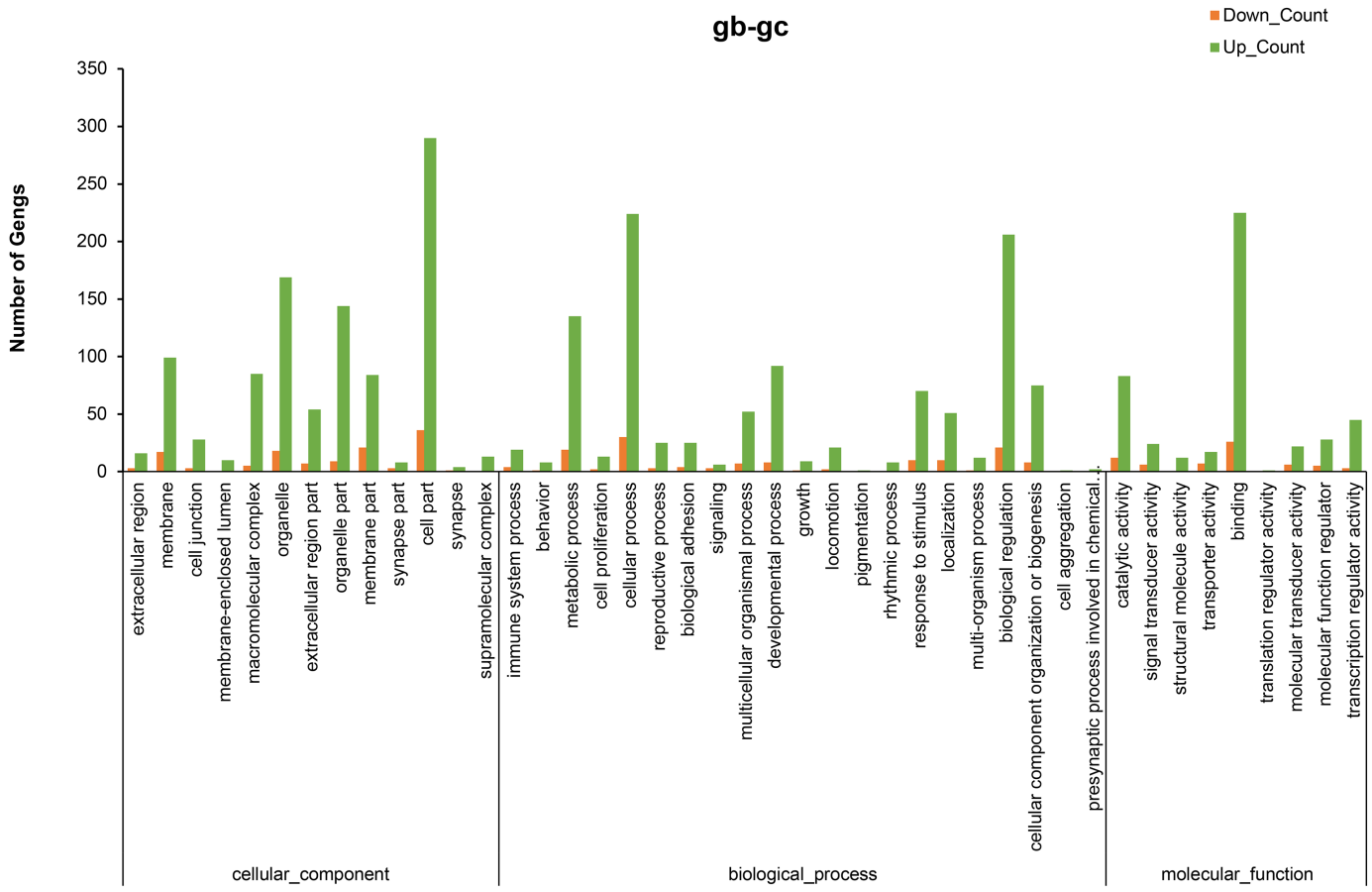

Figure 6

Gene Ontology classification of the Significant DEGs during embryonic development