

# Transcriptomic analysis of the *Echinococcus granulosus* protoscolex in the encystation process

**Junjie Fan**

Chongqing Medical University

**Hongye Wu**

Chongqing Medical University

**Kai Li**

Chongqing Medical University

**Xunuo Liu**

Chongqing Medical University

**Qingqing Tan**

Chongqing Medical University

**Wenqiao Cao**

Chongqing Medical University

**Xiang Li**

Chongqing Medical University

**Peng Liao**

Chongqing Medical University

**Bo Liang**

Chongqing NO.18 Middle School

**Bin Ye** (✉ [yebin@cqmu.edu.cn](mailto:yebin@cqmu.edu.cn))

---

## Research

**Keywords:** Echinococcus granulosus, Encystation process, Differentially expressed genes, Protoscolex, RNA-seq

**Posted Date:** January 7th, 2020

**DOI:** <https://doi.org/10.21203/rs.2.20164/v1>

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

[Read Full License](#)

---

# Abstract

Back ground: Echinococcosis (CE) is a zoonosis and in humans it occurs as a result of infection by the larva of *Echinococcus granulosus*. CE seriously affects the development of animal husbandry and endangers human health. Due to the lack of in-depth understanding of the cystic fluid formation pathway, prevention and treatment of CE have been lack of innovative methods.

Result: High throughput RNA-sequencing (RNA-seq) of protoscoleces (PSCs) in the encystation process of total three biological replicates for each period on 0d, 10d, 20d, 40d and 80d were analyzed. The results demonstrated, a total of 32,401 transcripts and 14,903 genes, including numbers new genes, new transcript, stage-specific genes and differently expression genes (DEGs). Genes encoding proteins involved in several signaling pathways, such as putative G-protein coupled receptor (GPCR), tyrosine kinases and serine/threonine protein kinase were predominantly up-regulated during encystation process of PSCs. Moreover, three major antioxidant proteins of PSCs were identified, and these proteins demonstrated have a high expression level, including cytochrome c oxidase, thioredoxin glutathione, and glutathione peroxidase. Intriguingly, The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis suggested that up-regulated DEGs involved in the vasopressin-regulated water reabsorption metabolic pathway might play important roles in the protein, carbohydrate, and other substances transport.

Conclusions: The present study carried out the transcriptomic analysis of the encystation process of *E. granulosus* PCSs, which provide valuable information for the mechanism of cystic fluid formation during the encystation process. These results provide a basis and reference for further studies for investigation of the molecular mechanisms involved in PSC growth and development. Keywords: *Echinococcus granulosus*, Encystation process, Differentially expressed genes, Protoscoleces, RNA-seq

## Background

Cystic echinococcosis (CE), also called hydatid disease, is a chronic neglected zoonotic disease caused by the larvae of *Echinococcus granulosus* which endangers human health and causes huge economic losses in the animal husbandry [1]. The recent epidemiological studies showed that at least 270 million people (58% of the total population) are at risk of CE in Central Asia, including areas of Mongolia, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan, Iran, Pakistan and western China [2, 3]. The life cycle of *E. granulosus* characterized by long-term growth of larval stages (hydatid cysts) in the internal organs of humans and other intermediate hosts, especially in the liver and lungs [4]. Although benzimidazoles have been widely used to treat CE, the main disadvantages were the poor absorption and high hepatotoxicity [5, 6]. Thus, there remains great interest in developing new chemotherapies against CE.

The hydatid cysts is composed of cyst wall and contents, including cystic fluid, protoscoleces (PSCs) and brood capsules. The cyst wall mainly consists of the outer protective acellular mucin-rich laminated layer

[7] and the inner layer of germinal cells (blastoderm) [8]. The germinal layer has many cellular nuclei, which can produce groups of vesicles into the cyst lumen, these vesicles become brood cysts after cell vacuolation and then develop into the initial PSCs [9]. Secondary hydatid cysts are formed by encystation during both *in vivo* development and *in vitro* culture of the PSCs [10]. However, despite it is clear that during the process of encystation, the volume of cysts gradually increases, as well as the amount of cystic fluid, and the echinococcus cyst fluid is the internal environment for the growth and development of germinal cells and PSCs, the functional genes and the key metabolic pathways linked to this encystation process are still unknown.

Recently, the second-generation high-throughput RNA sequencing technology is mainly used to detect the expression levels of all genes or transcriptomes in a sample under specific physiological conditions, and it plays an important role in gene expression and transcriptome regulation [11]. While whole-genome sequencing, proteomic analyses, and transcriptomic investigations recently been carried out to identify the DEGs and proteins between the different life stages (adult, oncosphere, hydatid cyst wall, larval worms, and pepsin/H<sup>+</sup>-activated PSCs) of *E.granulosus* [12-15]. An unexplored issue associated with the parasites survival in its host is what is the key DEGs and main metabolic pathways that used by PSCs to produce the energy, formation cyst fluid and transport molecular needed to maintain its growth and development during the encystation process.

Here, for the first time, we reported there were 1,991 and 2,517 up- and down-regulated DEGs, respectively, during encystation process in five different stages (0d, 10d, 20d, 40d, and 80d) of PSCs by using RNA seq analysis method. Through enrichment analysis of DEGs, it was found that PSCs were mainly involved in vasopressin-regulated water reabsorption during the process of encystation. Our study identified a range of potential target genes of drugs and vaccines treatment and provided metabolism pathways that are primarily involved in the PSCs encystation process. These findings could facilitate the development of new intervention tools for the treatment and control of CE.

## Materials And Methods

### Protoscolex collection and RNA extraction

*E. granulosus* hydatid cysts were collected from a naturally infected sheep liver, which freshly obtained from a slaughterhouse in Urumqi, Xinjiang, China. The hydatid fluid was aspirated aseptically using a 50 ml syringe, and then the PSCs were collected from hydatid fluid under aseptic conditions and washed 15 times with sterile PBS. Afterward, the PSCs were digested in 1% (W/V) pepsin for 30 min at 37 °C to release PSCs from the brood capsules. During the digestion, the samples were agitated every 5 minutes and observed under an inverted microscope until the tissues were digested completely. Subsequently, the collected PSCs were washed with a solution of PBS that contained penicillin and streptomycin antibiotics (Sigma, USA), and stained with 0.4% trypan blue until the viability rate was more than 90%. PSCs culture conditions were referred to Liu et al., 2018 [16]. The PSCs were cultured in RPMI 1640 medium (Gibco, USA) (pH 7.2), with 10% fetal bovine serum and 1% penicillin-streptomycin. The average number of PSCs

in monophasic culture medium was 4000 per bottle, and the incubator conditions were 37 °C and 5% CO<sub>2</sub>. The culture medium was changed every 2 to 3 days. Different stages in culture medium were isolated based on the morphological classification described by Elissondo et al [17]. The isolated of PSCs on different stages were stored in liquid nitrogen for RNA extraction. PSCs tissues were collected from 15 batches of five different culture periods (three replicates each): EgPSC\_0d, EgPSC\_10d, EgPSC\_20d, EgPSC\_40d, and EgPSC\_80d. Total RNA was extracted from all samples by using TRIZOL-Reagent kit (Invitrogen, CA, USA). The extraction was carried out according to the instructions, and the concentration and integrity of RNA was detected by ultraviolet spectrophotometer and agarose gel electrophoresis.

### **Construction of strand-specific libraries and RNA-seq**

Conversion of mRNA from total RNA in each sample to obtain a strand-specific cDNA library was carried out according to the manufacturer's instructions, using the Illumina® TruSeq® RNA Sample Preparation Kit v2 (Illumina, CA, USA) and Oligo (dT) bead enriched PSCs mRNA. Combined different cDNA libraries according to the effective concentration and target data volume, before undertaking Illumina sequencing (Mig, Shanghai, China).

### **Data assembly and annotation**

The raw reads sequences including linker sequences, low-quality sequences, long sequences, short length sequences and ambiguous N (unknown nucleotide). Thus the software SeqPrep (<https://github.com/jstjohn/SeqPrep>) was used to perform quality control of the raw data and obtain high-quality data (clean data) before analyzing [18]. TopHat2 [19] software (<http://ccb.jhu.edu/software/tophat/index.shtml>) was subjected to sequence alignment analysis with the reference genome of *E.granulosus*. Then, total sequences were assembled and spliced using cufflinks software (<http://cole-trapnell-lab.github.io/cufflinks/>). After that, all transcripts were aligned to the public databases by BLASTn and BLASTx (E-value cut-off  $\leq 10^{-5}$ ) to provide information of transcripts expression and function, in this step, these unannotated transcripts were defined as new transcripts. The public databases include the nonredundant protein (NR) (<ftp://ftp.ncbi.nih.gov/blast/db/>); Kyoto Encyclopedia of Genes and Genomes (KEGG) (<http://www.genome.jp/kegg/>); A manually annotated, non-redundant protein sequence database (Swiss-Prot) (<http://www.uniprot.org/>); Homologous protein family (Pfam) (<http://pfam.xfam.org/>); Clusters of Orthologous Groups of proteins (COG) (<http://www.ncbi.nlm.nih.gov/COG/>) and Gene Ontology (GO) (<http://www.geneontology.org/>). The best alignment results were used to decide the direction of sequences.

### **Inter-samples correlation analysis**

RSEM (<http://deweylab.github.io/RSEM/>), as a quantitatively analyzed the expression levels of genes and transcripts software, could be used to further analyzed the DEGs between different samples, and to reveal the regulation mechanism of genes by combining sequence function information [20]. Besides, RSEM distinguished which transcripts were different subtypes of the same gene through established the

maximum likelihood abundance estimation model [21]. Venn diagram analysis was performed to display genes that specific and co-expression between samples.

## Differential expression analysis

Since there were three biological replicates per period, the raw data were statistically analyzed directly using the DESeq2 software, which based on the negative binomial distribution [22]. EgPSC\_0d was used as the control group, while other groups were the experimental groups, and a differential expression analysis of the genes from the samples was performed. Differentially expressed genes (DEGs) was threshold settings as previous studies ( $p\text{-adjust} < 0.05$  and  $|\log_2FC| > 1$ ), transcript expression folds greater than 2 per group, and p-value less than 0.05, after multiple-test corrections, were selected [23].

## Verification of RNA sequencing data by Q-PCR

12 DEGs were selected to verify the accuracy of the Illumina His seq X (Illumina, CA, USA) sequencing data, including six up-regulated DEGs and six down-regulated DEGs, respectively. The Primer 5.0 and Oligo 7.0 software were used to design and evaluate specific primers (Table 1). According to the previous study, Actin II was employed as an endogenous reference gene [24], and these primers were synthesized by Tsingke Biotech Company (Beijing, China). ALL 15 RNA templates that were used for RNA-seq were reverse transcribed to cDNA using Prime Script™ RT reagent Kit (Takara, Dalian, China) according to the manufacturer's instructions, individually. The final cDNA products were diluted 10 times with nuclease-free water before QPCR. QPCR was performed using SYBR Premix Ex Taq GC kit (Takara, Dalian, China). The reaction system comprised 10 μl SYBR QPCR mix (2x). Forward and reverse primers (0.8 μl, 10 μM each), 4 μl diluted cDNA and added ddH<sub>2</sub>O to bring the volume to 20 μl. The reaction conditions were as follows: 95 °C, 2 min; 95 °C, 10 s; 55 °C, 10 s; 72 °C, 20 s, 40 cycles; dissolution curve conditions: 65 °C, 5s; 95 °C, 5s; 4 °C, 30s. Each gene was processed in triplicate replicates using CFX-96 Touch™ Real-Time PCR Detection System (Bio-Rad, USA) and the expression of all the selected DEGs was evaluated using 2<sup>-ΔΔCt</sup> method.

# Results

## RNA extraction and RNA-seq quality assessment

At the beginning of the *in vitro* culture, PSCs were invaginated, and the structure of the calcareous corpuscles and the apical hook could be observed under the inverted microscope (Fig. 1A). After 10 days of incubation, the morphology of the PSCs changed greatly, the apical hook and the sucker were evaginated (Fig. 1B). On day 20, the micro-cysts appeared with a thin laminated layer, and the small hook of scolex moved toward the central position of these micro-cysts, moreover, the calcareous corpuscles of the micro-cysts were observed clearly (Fig. 1C). After 40 days the micro-cysts and laminated layer were observed clearly, the apical hooks and sucker were not completely degraded (Fig. 1D). On 80 days, the

apical hook and sucker disappeared or degraded, the calcareous corpuscles decreased, and a considerable increment in the size of the cysts could be observed (Fig. 1E). Samples from three batches of PSCs specimens for five different culture periods (EgPSC\_0d, EgPSC\_10d, EgPSC\_20d, EgPSC\_40d, and EgPSC\_80d) were collected. Total RNA extraction results showed that the RNA bands were slightly degraded, with no pigment contamination, obvious protein content, sugar, or other impurities.

## Sequence assembly and annotations

Through the second-generation high throughput sequencing Illumina platform, the cDNA library for each sample was constructed and sequenced, which average generated 51,300,044 clean reads representing a total of 7,597,846,677 (7.5 Gb) nucleotides of each sample (Table 2). The GC content is one of the important characteristics of the genome base sequence, which can reflect the structure, function and evolutionary information of the gene [25]. Interestingly, the average GC content of the *E. granulosus* transcriptional level (46.44%) was similar to both its genome (42.1%) and coding regions (49.3%) [26]. Early studies estimated the *S. mekongi*, *B. malayi* and *S. mansoni* had a GC content of 34.1%, 30.5% and 35.3%, respectively [27, 28], indicated that compared with other parasite species, the GC content of *E. granulosus* was higher.

After alignment and assembled clean reads, a total of 14,903 genes and 32,401 transcripts were obtained, including 3584 new genes and 21,082 new transcripts, which had not been reported in the previously published transcriptome of the parasite. Among 32,401 transcripts, 20,511 were annotated by GO (63.3%), 13,730 by KEGG (42.38%), 17,152 by COG (52.94%), 28,188 by NR (87%), 16,511 by Swiss-Prot (51.08%), 17,874 by Pfam (55.16%). In addition, Venn diagram analysis of function annotation showed that 4,458 reference transcripts and 53 new transcripts were simultaneously annotated into all gene databases. Furthermore, some genes or transcripts could not be annotated and were defined as hypothetical protein, which suggested that the group of identified transcriptome dataset is sufficiently reliable for further characterization [29]. The length of the transcript was widely distributed, and the majority of the transcripts longer than 1,800 bp (34.65%) in length. In contrast, the transcripts length between 0 and 200 bp were the least (1.34%) (Fig. 2). The obtained transcriptome raw reads dataset has been submitted to the NCBI Short Read Archive (SRA) database under the accession number: SRP172517.

## Venn diagram analysis and differential expression analysis between samples

Based on the expression matrix, inter-sample Venn diagram analysis was performed to obtain the co-expression and specific expression genes (Fig. 3A) or transcripts (Fig. 3B) between groups. Intriguingly, although a total of 14,903 genes and 32,401 transcripts were obtained by transcriptome sequencing at different developmental stages of PSCs in the encystation process, these genes and transcripts did not always present in every period of the cysts formation process. Conversely, the genes and transcripts were specifically expressed in different stages utilizing inter-sample Venn diagram analysis, especially at EgPSC\_0d, the number of stage-specific genes and transcripts was the largest, 1079 and 1990, respectively. Secondly, the stage-specific genes and transcripts at the 20d stage were 277 and 704,

respectively. While EgPSC\_0d were the period when the original PSCs was just removed from the fresh sheep liver, and EgPSC\_20d was the initial period of cyst formation, both of which were key periods of the encystation process. In the previous transcriptome sequencing of the *E. granulosus* indicated that the presence of stage-specific genes among the cyst wall, PSCs and pepsin/H<sup>+</sup>-activated PSCs [12], which indicated that the stage-specific genes exist not only in different stages of the *E. granulosus* development but also in different periods of development.

For the DEGs analysis of each sample, EgPSC\_0d was used as the control group compared with other groups (EgPSC\_0d vs 10d, EgPSC\_0d vs 20d, EgPSC\_0d vs 40d and EgPSC\_0d vs 80d). The parameter set as P-value cutoff of < 0.05 and an FC cutoff of > 2 to identify the significant DEGs, including 1,991 up-regulated and 2,517 down-regulated DEGs, respectively. A set of gene which composed of 1,991 up-regulated DEGs was established to provide the functional classification compared with the COG database. The result showed that most of the genes (25%) were classified into the poorly characterized category, which described as function unknown. Secondly, 144 genes (7.2%) were classified as cellular processes and signaling, which described as intracellular trafficking, secretion, and vesicular transport (Additional file 1: Table S1). Given that high levels of expression often indicate a fundamental role, these represent interesting goals for further research. The finding can open the way to the further studies.

Based on Venn diagram analysis, the total of 1,991 gene were up-regulated and 2,517 were down-regulated DEGs, we found that 52 genes were consistently on the rise during the process of encystation (Fig. 4A), while 152 genes have continued to decline (Fig. 4B). Through cluster heat map analysis of 52 consistently up-regulated DEGs (Fig. 4C), showing that hypothetical protein (EGR\_10197) and (EGR\_05435) increased significantly at the late stage of encystation, and both of them were annotated as a hypothetical protein. Besides, gene1117 described as an integral component of the membrane and classified into cellular component term of the GO database. Moreover, 13 genes annotated to COG database among 52 DEGs and served many important biological functions, such as amino acid transport and metabolism, intracellular trafficking, secretion, and vesicular transport (Table 3). After analyzed the correlation coefficients between the 52 DEGs with FDR< 0.05 (Fig. 4D), a visualization network was obtained, which showed that cornifelin (EGR\_03608), platelet glycoprotein (EGR\_09887), potassium voltage-gated channel subfamily C member 3 (EGR\_05864) and others are related to the expression of many other genes.

Another interesting found was that none of the genes remained up-regulated during the encystation process when we analyzed by Venn diagram analysis of the gene sets among EgPSC\_0d vs 10d, EgPSC\_10d vs 20d, EgPSC\_20d vs 40d and EgPSC\_40d vs 80d (Additional file2: Figure S1). Only ten genes exist at most of four gene sets simultaneously, and two genes of them continued up-regulated from EgPSC\_10d to 80d, EGR\_05443, and EGR\_03870, respectively, and the other 8 genes were continuously up-regulated from EgPSC\_0d to 10d and EgPSC\_20d to 80d. In addition, of the 10 genes, except EGR\_09887 and EGR\_09469, which described as platelet glycoprotein and dynein light chain, respectively, the remain genes were described as hypothetical proteins (Table 4).

## GO and KEGG enrichment analysis of DEGs

GO was a comprehensive database that aggregates all gene-related research results from around the world and it could standardize biological terms for genes and gene products from different databases, providing a uniform definition and description of gene and protein functions [30]. To obtain a better understanding of the enriched functions of 1,991 DEGs in the process of encystation of *E. granulosus*, GO term enrichment analysis was performed. According to the GO classification of DEGs, there were 1094, 148 and 263 genes assigned to cellular component (CC), biological process (BP) and molecular function (MF) categories, respectively, of which 709, 0 and 122 genes were significantly DE (FDR value < 0.05) (Additional file 3: Table S2). The top three predominant terms for BP were endoplasmic reticulum unfolded protein response, O-glycan processing, and ion transport; for CC they were integral component of membrane, intrinsic component of membrane, and membrane part; and for MF they were phosphoric ester hydrolase activity, hydrolase activity, acting on ester bonds and beta-1,3-galactosyltransferase activity (Fig. 5).

KEGG is a large-scale database for systematic analysis of gene function, genomic information, and functional annotation. By analyzing the results using the information in this database, genes or transcripts can be classified according to the metabolic pathways or functions involved, including organismal systems (OS), metabolism (M), environmental information processing (EIP), human diseases (HD), genetic information processing (GIP) and cellular processes (CP) [25]. KEGG pathway enrichment analysis revealed that the 1,991 DEGs classified into 291 pathways (Additional file 4: Table S3) (Fig. 6). Analysis of the top 20 most significantly enriched pathways showed that the map 04962, which described as vasopressin-regulated water reabsorption metabolic pathway was the most significantly enriched pathways (FDR value < 0.05) (Fig. 7), followed by pantothenate and CoA biosynthesis (map00770) and Hippo signaling pathway (map04391).

Since the morphological changes of the PSCs were greatest in the first 10 days, we performed GO and KEGG enrichment analysis of the top ten of GO term and KEGG pathway with FDR value < 0.05 on the DEGs between EgPSC\_0d and EgPSC\_10d. After analyzed, 43 and 61 specific genes were obtained, respectively (Additional file 5: Figure S2, Additional file 6: Figure S3).

## Real-time quantitative PCR analysis

To validate the RNA-seq data and the analyses derived from them, 12 DEGs were selected to measure their relative expression level by QPCR, including six up- and six down-regulated DEGs. The inter-sample statistical analysis was performed using T-test.  $0.01 < P < 0.05$  was considered significant (\*),  $P < 0.01$  (\*\*) and  $P < 0.001$  (\*\*\*) were considered extremely significant. The QPCR data were compared with the relative expression level of the gene in the RNA-seq TPM value to confirm the reliability of the transcriptome sequencing results (Fig. 8).

## Discussion

The parasite *E. granulosus* is a cestode tapeworm that acts as the causative agent of CE, one of the 17 neglected tropical diseases to be recently prioritized by the World Health Organization [15]. During its life cycle, hydatid cysts produce the pre-adult form, which has the ability to either differentiate into an adult worm (strobilation) in the terminal host (dog) or dedifferentiate into a secondary hydatid cyst in the intermediate host (human and livestock) [31]. Furthermore, the cysts may be developed in almost any internal organ or tissue by hematogenous dissemination, such as heart, bone and nervous system [32]. The early infection is asymptomatic, with the cyst volume increases, eventually causing physical damage result from mechanical compression of the surrounding tissues and organs [33]. The major sources of morbidity and mortality are pressure effects from cyst size, location in a sensitive organ, or cyst rupture caused by a spontaneous or external force (traumas or surgery) with subsequent anaphylaxis or dissemination of the secondary infection [34]. Thus, novel prevention and control strategies need to be developed based on basic and applied *E. granulosus* biology, especially genomic, transcriptomic and proteomic analysis. However, our understanding of key DEGs and the main metabolic pathways of PSCs in the encystation process is limited.

The previous study found that PSCs encystation generally appeared during the first week of incubation in vitro, and on day 20, some micro-cysts with a complete laminated layer were observed. Between day 38 and day 42, micro-cysts completely developed could be observed [35]. The times when we observed the above important nodes in our laboratory were 10d, 20d and 40d, respectively, and after 80 days of in vitro culture, the cysts are large enough. These phenomena was slightly different from the previous description, one possible reason for this is the different culture medium and genotype, not only in the time of bladder formation but also in the development process was slightly different [17].

Owing to the problem of isolation PSCs from the abdominal cavity of mouse is time-consuming and difficult to observe the specific encystation state [36]. Here, we provided a new insight to compare the gene expression profile of different developmental stages in the in vitro encystation process of PSCs using RNA-seq. Compared to previous hybrid-based microarrays and sanger sequence-based methods, RNA-Seq provided abundant genomic information on specimens [37]. In addition to quantifying gene expression, data obtained by RNA-Seq facilitate the identification of new genes, new transcripts, DEGs and classification of functions, but also helps to understand the different mechanisms involved in the growth and development of the specimen and important metabolic pathways [38]. Transcripts per million reads (TPM), the number of reads from a transcript per million reads, was used to measure the level of expression. Unlike the fragment per kilobase of exon model per million mapped reads (FPKM), TPM was first homogenizes the length of sequencing and then homogenizes the depth of the gene. The homogenization process of TPM made the total expression levels in different samples consistent, produced the comparison of gene expression levels more intuitive [39].

Here, we obtained a total of 32,401 transcripts and 14,903 genes, identified numbers stage-specific genes, and demonstrated that 1,991 and 2,517 genes were significantly up-regulated and down-regulated, respectively. Genes encoding proteins involved in several signaling pathways, such as putative G-protein coupled receptor (GPCR), tyrosine kinases, and serine/threonine protein kinase were predominantly

upregulated during encystation process of PSCs. These signaling pathways play major roles in key functions like movement, development, and reproduction in parasites [40]. Our findings are paralleled to the previous studies of *S. mekongi* [27] and *S. japonicum* [41]. A previous study classified more than 60 putative GPCRs of *E. granulosus* as potential anthelmintic drugs target [14]. In this study, we filtered out coding putative GPCRs genes with low expression and un-expression through transcript sequencing, narrowed down to six genes (EGR\_01295, EGR\_00873, EGR\_07838, EGR\_08773, EGR\_00585 and EGR\_06296), in particular, EGR\_00873 was significant up-regulate during EgPSC\_0d to EgPSC\_20d. In addition, EGR\_10296, and EGR\_09273, which encodes the tyrosine kinase and serine/threonine protein kinase, respectively, did not increase during the first 10 days of PSCs *in vitro* culture but showed a significant up-regulate trend at the later stage. Members of these pathways characterized here might represent novel drug targets for anti-parasitic intervention and require further study.

The vast majority of invertebrate species have shown an immune response that can inactivate and eliminate penetrating parasites, especially involves the formation of reactive oxygen species (ROS) [42]. ROS is a key trigger of the inflammatory activation of macrophages in malaria, simultaneously, it can damage proteins, carbohydrates, and DNA, which is harmful to the survival of parasites [43]. In order to neutralize the host's immune response and oxidative damage caused by oxygen free radicals, protective antioxidant proteins are produced by the body of parasites (such as schistosomiasis) [44]. In the present study, we identified three major antioxidant proteins of PSCs, including cytochrome c oxidase, thioredoxin glutathione, and glutathione peroxidase. Cytochrome c oxidase (EGR\_08027), an up-regulated DEGs, has the molecular function of binding to heme and transporting protein in the mitochondrial electron-transport chain according to the COG database annotation, consistent with previous observations that it can regulate mitochondrial oxidative metabolism of *E. granulosus* [45]. Previous reports suggested that cytochrome c oxidase dysfunctions is associated with increasing mitochondrial ROS production as well as stimulating inflammation and apoptosis [46, 47]. Hinted that in the face of external environmental pressure, cytochrome c oxidase of PSCs might play essential roles in the antioxidant defenses during the encystation process. Another well-characterized antioxidant defense is the thioredoxin glutathione system, which plays a critical role in maintaining the redox balance in the parasite, such as *S. japonicum* [48], *E. granulosus* [12], *T. brucei* [49], and *S. mansoni* [44]. In this study, we filtered out three relative proteins through differential expression analysis, including thioredoxin (EGR\_6727), thioredoxin related transmembrane protein (EGR\_01550) and thioredoxin domain containing protein (EGR\_09201). Lastly, glutathione peroxidase (EGR\_06748), which presented continuous high expression of PSCs during the encystation process, involved in clearance of cytotoxic and genotoxic compounds and protection against oxidative damage [50]. Studies have reported that glutathione peroxidase of *E. granulosus* displayed the ability to bind non-substrate molecules, particularly anthelmintic drugs [51], and it acts as an essential enzyme for the survival of the schistosomiasis in the redox environment has been actively explored as a potential drug target [52].

Here, we analyzed the key genes previously identified by genomic [26], proteomic [53] and transcriptomic [12] studies of *E. granulosus*, and found that some genes were expressed significant differences in the encystation process of PSCs. For example, Heat shock protein, which has important roles in

posttranslational modification, protein turnover, and ATP binding, was reported that coding gene EGR\_09650 was only expressed in adult worms, while EGR\_10561 was highly expressed in the hydatid cyst membrane [26]. However, in this sequencing, we found that both of them are present in the PSCs. It is worth noting that EGR\_10561 is the gene with the consistently highest expression among all genes coding heat shock proteins during the encystation process of PSCs. Cadherin, which has important roles in cell adhesion and recognition, overexpression of cadherin induced parasite aggregation in *T. vaginalis* [54].

Although more than forty genes were found encoding cadherin by sequencing, all the genes except EGR\_06182 showed low transcription expression levels. Glycosylphosphatidylinositol (GPI)-anchored wall transfer protein, despite it is important and exclusive to *E. granulosus*, the expression level of the encoding gene (EGR\_06221) is low during the PSCs encystation. One possible reason for this is the heterologous proteins is covalently anchored to the PSCs cell wall by fusing them with the anchoring domain of GPI-anchored cell wall transfer protein [55], but in this study, PSCs was cultured *in vitro* and there was no heterogeneous source protein. Furthermore, tetraspanins are transmembrane proteins previously described as potential vaccine candidates for other helminth infections and are also found in the membranes of the tegument and extracellular vesicles of *O. viverrini* [56]. In some organisms, their role is associated with virulence and pathogenesis [57]. A wide variety of molecules have been found to be released by *E. granulosus*, such as tetraspanin proteins, channel protein that transport lipids, and nucleic acids [58, 59]. Here, we identified two encoding tetraspanin genes (EGR\_11042 and EGR\_06311), showed a strikingly up-regulate trends during the encystation of PSCs, we speculated that the overexpression of genes related to membrane protein transport in PSCs might be a compensatory mechanism by these worms to adapt to the external environments to satisfy their own continuously encystation development processes.

The most highly enriched KEGG pathway involved in 1,991 up-regulated DEGs of PSCs during the encystation process was vasopressin-regulated water reabsorption metabolic pathways (map 04962), which include a total of 24 DEGs (Additional file 7: Table S4). Dynein light chain (DLC), ras-related protein, synaptobrevin-like protein and aquaporin 4 (AQP 4) was specifically upregulated in the encystation process of PSCs. The vasopressin-regulated water reabsorption metabolic pathways have been reported to play a critical role in regulating water, urea, and sodium transport to keep the water balance of body [60]. However, due to the special structure and life history of the parasite, there is no aquaporin 2, vasopressin and its corresponding V2 receptor (V2R) genes in PSCs. Previous reports indicated that DLC is involved in the process of protein, carbohydrate and other substances transport and retrograde vesicle transport [61]. Ras-related protein are involved in the process of vesicle transport after activation, parasites can secrete extracellular vesicles to achieve intercellular communication, and transfer biologically active molecules to the host cells in order to modulate host immune response [62]. Hence, vesicle transport is considered an important pathway for substances transport and information transmission between parasites and external environment [53]. In this study, we detected the presence of AQP4 in the cell membrane of PSCs, and eukaryotes usually permeate water and regulate water reabsorption by activating AQPs, such as *S. japonicum* [63] and *S. mansoni* [64]. Therefore, studying the

role of vasopressin-regulated water reabsorption metabolic pathways in PSCs can provide clues useful for the development of new drug targets. However, these results were approached in the manner of bioinformatics analysis and require further studies or investigation.

## Conclusions

In summary, through RNA-seq, we presented a transcriptomic analysis during the encystation process of PSCs. The study revealed numbers of new genes, new transcripts, DEGs, stage-specific genes and the key metabolic pathways. These data provide valuable information to understand the growth and development mechanisms of PSCs, which can be used as a base for further studies on new drugs and vaccine targets.

## Abbreviations

DEGs, differently expression genes; KEGG, Kyoto Encyclopedia of Genes and Genomes; NCBI, National Center for Biotechnology Information; PBS, phosphate buffer saline; PCR, polymerase chain reaction; Q-PCR, quantitative real-time PCR; RNA-seq, RNA sequencing.

## Declarations

### Funding

This work was supported by the National Natural Science Foundation of China (grant number: 81672045).

### Acknowledgments

We are grateful to the editor and anonymous reviewers for their helpful comments and constructive suggestions.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and material

RNA-Seq reads have been deposited in the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA) database under the accession number: SRP172517.

### Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

JJF and BY conceived and designed the study. HYW, XNL and KL provided samples. JJF and QQT performed the experiments and analyzed the transcriptomic data. JJF drafted the manuscript. BL, PL, WQC and XL helped in study design, study implementation and manuscript revision. All authors read and approved the final manuscript.

## References

1. Zhang W, Zhang Z, Wu W, Shi B, Li J, Zhou X, et al. Epidemiology and control of echinococcosis in central Asia, with particular reference to the People's Republic of China. *Acta Trop*. 2015;141:235–43.
2. Siyadatpanah A, Anvari D, Emami Zeydi A, Hosseini SA, Daryani A, Sarvi S, et al. A systematic review and meta-analysis of the genetic characterization of human echinococcosis in Iran, an endemic country. *Epidemiol Health*. 2019;41:e2019024.
3. Mahmoudi S, Mamishi S, Banar M, Pourakbari B, Keshavarz H. Epidemiology of echinococcosis in Iran: a systematic review and meta-analysis. *BMC Infect Dis*. 2019;19:929.
4. Filippou D, Tselepis D, Filippou G, Papadopoulos V. Advances in liver echinococcosis: diagnosis and treatment. *Clin Gastroenterol Hepatol*. 2007;5:152–9.
5. Xing G, Zhang H, Liu C, Guo Z, Yang X, Wang Z, et al. Sodium arsenite augments sensitivity of *Echinococcus granulosus* protoscoleces to albendazole. *Exp Parasitol*. 2019;200:55–60.
6. El-On J. Benzimidazole treatment of cystic echinococcosis. *Acta Trop*. 2003;85:243–52.
7. Thompson RCA. Biology and systematics of *Echinococcus*. *Adv Parasitol*. 2017;95:65–109.
8. Albani CM, Elissondo MC, Cumino AC, Chisari A, Denegri GM. Primary cell culture of *Echinococcus granulosus* developed from the cystic germinal layer: biological and functional characterization. *Int J Parasitol*. 2010;40:1269–75.
9. Díaz A, Casaravilla C, Allen JE, Sim RB, Ferreira AM. Understanding the laminated layer of larval *Echinococcus* II: immunology. *Trends Parasitol*. 2011;27:264–73.
10. Díaz A, Casaravilla C, Irigoín F, Lin G, Previato JO, Ferreira F. Understanding the laminated layer of larval *Echinococcus* I: Structure. *Trends Parasitol*. 2011;27:204–13.
11. Ronza P, Robledo D, Bermúdez R, Losada AP, Pardo BG, Sitjà-Bobadilla A, et al. RNA-seq analysis of early enteromyxosis in turbot (*Scophthalmus maximus*): new insights into parasite invasion and immune evasion strategies. *Int J Parasitol*. 2016;46:507–17.
12. Parkinson J, Wasmuth JD, Salinas G, Bizarro C V, Sanford C, Berriman M, et al. A transcriptomic analysis of *Echinococcus granulosus* larval stages: implications for parasite biology and host adaptation. *PLoS Negl Trop Dis*. 2012;6:e1897.
13. Liu S, Zhou X, Hao L, Piao X, Hou N, Chen Q. Genome-wide transcriptome analysis reveals extensive alternative splicing events in the protoscoleces of *Echinococcus granulosus* and *Echinococcus*

- multilocularis*. Front Microbiol. 2017;8:1–14.
14. Tsai IJ, Zarowiecki M, Holroyd N, Garcarrubio A, Sanchez-Flores A, Brooks KL, et al. The genomes of four tapeworm species reveal adaptations to parasitism. Nature. 2013;496:57–63.
  15. Debarba JA, Monteiro KM, Moura H, Barr JR, Ferreira HB, Zaha A. Identification of newly synthesized proteins by *Echinococcus granulosus* protoscoleces upon induction of strobilation. PLoS Negl Trop Dis. 2015;9:1–18.
  16. Liu C, Yin J, Xue J, Tao Y, Hu W, Zhang H. *In vitro* effects of amino alcohols on *Echinococcus granulosus*. Acta Trop. 2018;182:285–90.
  17. Elissondo MC, Dopchiz MC, Brasesco M, Denegri G. *Echinococcus granulosus*: first report of microcysts formation from protoscoleces of cattle origin using the *in vitro* vesicular culture technique. parasit. 2004;11:415–8.
  18. Li H, Li H, Durbin R, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25:1754–60.
  19. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.
  20. Zhu Y, Cao X, Zhang X, Chen Q, Wen L, Wang P. DNA methylation-mediated klotho silencing is an independent prognostic biomarker of head and neck squamous carcinoma. Cancer Manag Res. 2019;11:1383–90.
  21. Lateef A, Prabhudas SK, Natarajan P. RNA sequencing and de novo assembly of *Solanum trilobatum* leaf transcriptome to identify putative transcripts for major metabolic pathways. Sci Rep. 2018;8:15375.
  22. Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: An R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. 2010;26:136–8.
  23. He JJ, Ma J, Elsheikha HM, Song HQ, Huang SY, Zhu XQ. Transcriptomic analysis of mouse liver reveals a potential hepato-enteric pathogenic mechanism in acute *Toxoplasma gondii* infection. Parasites and Vectors. 2016;9:1–13.
  24. Fen W, Xunuo L, Hongye W, Kai L, Junjie F, Bin YE. Selecting of optimal reference genes for detecting transcript expression profiles of candidate aquaporins genes of *Echinococcus granulosus* during the development of hydatid cysts from protoscoleces. J Third Mil Med Univ. 2018;40:1533–41.
  25. Huang Y, Xiong JL, Gao XC, Sun XH. Transcriptome analysis of the Chinese giant salamander (*Andrias davidianus*) using RNA-sequencing. Genomics Data. 2017;14:126–31.
  26. Zheng H, Zhang W, Zhang L, Zhang Z, Li J, Lu G, et al. The genome of the hydatid tapeworm *Echinococcus granulosus*. Nat Genet. 2013;45:1168–75.
  27. Phuphisut O, Ajawatanawong P, Limpanont Y, Reamtong O, Nuamtanong S, Ampawong S, et al. Transcriptomic analysis of male and female *Schistosoma mekongi* adult worms. Parasit Vectors. 2018;11:504.

28. Almeida GT, Amaral MS, Beckedorff FCF, Kitajima JP, DeMarco R, Verjovski-Almeida S. Exploring the *Schistosoma mansoni* adult male transcriptome using RNA-seq. *Exp Parasitol*. 2012;132:22–31.
29. Li WH, Zhang NZ, Yue L, Yang Y, Li L, Yan HB, et al. Transcriptomic analysis of the larva *Taenia multiceps*. *Res Vet Sci*. 2017;115:407–11.
30. Acharya S, Saha S, Pradhan P. Novel symmetry-based gene-gene dissimilarity measures utilizing Gene Ontology: application in gene clustering. *Gene*. 2018;679:341–51.
31. Guo B, Zhang Z, Zheng X, Guo Y, Guo G, Zhao L, et al. Prevalence and molecular characterization of *Echinococcus granulosus sensu stricto* in Northern Xinjiang, China. *Korean J Parasitol*. 2019;57:153–9.
32. Mandal S, Deb Mandal M. Human cystic echinococcosis: epidemiologic, zoonotic, clinical, diagnostic and therapeutic aspects. *Asian Pac J Trop Med*. 2012;5:253–60.
33. Daneshpour S, Kefayat A, Mofid M, Rostami Rad S, Yousofi Darani H. Effect of hydatid cyst fluid antigens on induction of apoptosis on breast cancer cells. *Adv Biomed Res*. 2019;8:27.
34. Mandal S, Deb Mandal M. Human cystic echinococcosis: epidemiologic, zoonotic, clinical, diagnostic and therapeutic aspects. *Asian Pac J Trop Med*. 2012;5:253–60.
35. Dezaki ES, Yaghoubi MM, Spiliotis M. Comparison of ex vivo harvested and in vitro cultured materials from *Echinococcus granulosus* by measuring expression levels of five genes putatively involved in the development and maturation of adult worms. *Parasitol Res*. 2016;115:4405–16.
36. Albani CM, Cumino AC, Elissondo MC, Denegri GM. Development of a cell line from *Echinococcus granulosus* germinal layer. *Acta Trop*. 2013;128:124–9.
37. Lippuner C, Ramakrishnan C, Basso WU, Schmid MW, Okoniewski M, Smith NC, et al. RNA-Seq analysis during the life cycle of *Cryptosporidium parvum* reveals significant differential gene expression between proliferating stages in the intestine and infectious sporozoites. *Int J Parasitol*. 2018;48:413–22.
38. Nie H, Dong S, Li D, Zheng M, Jiang L, Li X, et al. RNA-Seq analysis of differentially expressed genes in the grand jackknife clam *Solen grandis* under aerial exposure. *Comp Biochem Physiol-Part D Genomics Proteomics*. 2018;28:54–62.
39. Alcolea PJ, Alonso A, Baugh L, Paisie C, Ramasamy G, Sekar A, et al. RNA-seq analysis reveals differences in transcript abundance between cultured and sand fly-derived *Leishmania infantum* promastigotes. *Parasitol Int*. 2018;67:476–80.
40. Camicia F, Celentano AM, Johns ME, Chan JD, Maldonado L, Vaca H, et al. Unique pharmacological properties of serotonergic G-protein coupled receptors from cestodes. *PLoS Negl Trop Dis*. 2018;12:1–27.
41. de Saram PSR, Ressurreição M, Davies AJ, Rollinson D, Emery AM, Walker AJ. Functional mapping of protein kinase A reveals its importance in adult *Schistosoma mansoni* motor activity. *PLoS Negl Trop Dis*. 2013;7:e1988.
42. Vorontsova YL, Slepneva IA, Yurlova NI, Ponomareva NM, Glupov V V. The effect of trematode infection on the markers of oxidative stress in the offspring of the freshwater snail *Lymnaea*

- stagnalis*. Parasitol Res. 2019;118:3561–4.
43. Ty MC, Zuniga M, Götz A, Kayal S, Sahu PK, Mohanty A, et al. Malaria inflammation by xanthine oxidase-produced reactive oxygen species. EMBO Mol Med. 2019;11:e9903.
44. Sayed AA, Williams DL. Biochemical characterization of 2-cys peroxiredoxins from *Schistosoma mansoni*. J Biol Chem. 2004;279:26159–66.
45. Cancela M, Paes JA, Moura H, Barr JR, Zaha A, Ferreira HB. Unraveling oxidative stress response in the cestode parasite *Echinococcus granulosus*. Sci Rep. 2019;9:15876.
46. Zalewska A, Ziembicka D, Żendzian-Piotrowska M, Maciejczyk M. The Impact of high-fat diet on mitochondrial function, free radical production, and nitrosative stress in the salivary glands of wistar rats. Oxid Med Cell Longev. 2019;2019:1–15.
47. Ozcan C, LI Z, Kim G, Jeevanandam V, Uriel N. Molecular mechanism of the association between atrial fibrillation and heart failure includes energy metabolic dysregulation due to mitochondrial dysfunction. J Card Fail. 2019;25:911–20.
48. LI Y, LI P, PENG Y, WU Q, HUANG F, LIU X, et al. Expression, characterization and crystal structure of thioredoxin from *Schistosoma japonicum*. Parasitology. 2015;142:1044–52.
49. Wilkinson SR, Horn D, Prathalingam SR, Kelly JM. RNA interference identifies two hydroperoxide metabolizing enzymes that are essential to the bloodstream form of the *African trypanosome*. J Biol Chem. 2003;278:31640–6.
50. Bombaça ACS, Brunoro GVF, Dias-Lopes G, Ennes-Vidal V, Carvalho PC, Perales J, et al. Glycolytic profile shift and antioxidant triggering in symbiont-free and H<sub>2</sub>O<sub>2</sub>-resistant *Strigomonas culicis*. Free Radic Biol Med. 2019;ppi:S0891-5849.
51. Lopez-Gonzalez V, La-Rocca S, Arbildi P, Fernandez V. Characterization of catalytic and non-catalytic activities of EgGST2-3, a heterodimeric glutathione transferase from *Echinococcus granulosus*. Acta Trop. 2018;180:69–75.
52. Gaba S, Jamal S, Drug Discovery Consortium OS, Scaria V. Cheminformatics models for inhibitors of *Schistosoma mansoni* thioredoxin glutathione reductase. Sci World J. 2014;2014:1–9.
53. Santos GB do., Monteiro KM, da Silva ED, Battistella ME, Ferreira HB, Zaha A. Excretory/secretory products in the *Echinococcus granulosus* metacestode: is the intermediate host complacent with infection caused by the larval form of the parasite? Int J Parasitol. 2016;46:843–56.
54. Chen YP, Riestra AM, Rai AK, Johnson PJ. A novel cadherin-like protein mediates adherence to and killing of host cells by the parasite trichomonas vaginalis. MBio. 2019;10:1–15.
55. Inokuma K, Kurono H, den Haan R, van Zyl WH, Hasunuma T, Kondo A. Novel strategy for anchorage position control of GPI-attached proteins in the yeast cell wall using different GPI-anchoring domains. Metab Eng. 2020;57:110–7.
56. Phung LT, Chaiyadet S, Hongsrichan N, Sotillo J, Dieu HDT, Tran CQ, et al. Recombinant *Opisthorchis viverrini* tetraspanin expressed in *Pichia pastoris* as a potential vaccine candidate for opisthorchiasis. Parasitol Res. 2019;118:3419–27.

57. Tomii K, Santos HJ, Nozaki T. Genome-wide analysis of known and potential tetraspanins in *Entamoeba histolytica*. *Genes (Basel)*. 2019;10:885.
58. Coakley G, Maizels RM, Buck AH. Exosomes and other extracellular vesicles: the new communicators in parasite infections. *Trends Parasitol*. 2015;31:477–89.
59. Siles-Lucas M, Sánchez-Ovejero C, González-Sánchez M, González E, Falcón-Pérez JM, Boufana B, et al. Isolation and characterization of exosomes derived from fertile sheep hydatid cysts. *Vet Parasitol*. 2017;236:22–33.
60. Park E-J, Kwon T-H. A Minireview on vasopressin-regulated aquaporin-2 in kidney collecting duct cells. *Electrolytes Blood Press*. 2015;13:1.
61. Horgan CP, Hanscom SR, Jolly RS, Futter CE, McCaffrey MW. Rab11-FIP3 binds dynein light intermediate chain 2 and its overexpression fragments the *Golgi complex*. *Biochem Biophys Res Commun*. 2010;394:387–92.
62. Fox AR, Maistriaux LC, Chaumont F. Toward understanding of the high number of plant aquaporin isoforms and multiple regulation mechanisms. *Plant Sci*. 2017;264:179–87.
63. Huang Y, Li W, Lu W, Xiong C, Yang Y, Yan H, et al. Cloning and in vitro characterization of a *Schistosoma japonicum* aquaglyceroporin that functions in osmoregulation. *Sci Rep*. 2016;6:1–8.
64. Faghiri Z, Camargo SMR, Huggel K, Forster IC, Ndegwa D, Verrey F, et al. The tegument of the human parasitic worm *Schistosoma mansoni* as an excretory organ: the surface aquaporin SmaQP is a lactate transporter. *PLoS One*. 2010;5:e10451.

## Tables

### Table 1

Primers used in quantitative PCR

Gene id	Gene name	Gene database prediction	Forward Primer (FP) and Reverse Primer (RP)	Product Size (bp)
Up-regulated DEGs				
Gene7722	EGR_03608	cornifelin	FP: AGTCTGCTTCCATTGACGACCA RP: CCATCGCTGTCGCCTTCGGTA	149
Gene11293	EGR_00384	Thioredoxin-related transmembrane protein	FP: TGCAAATGTACGCTACTACCC  RP: ACGTAAAGTAGGCTCAAGCAA	88
Gene10136	EGR_01086	Vesicle transport through interaction with t-SNAREs 1A	FP: AGATAGCAGCAACTACCCGTTT  RP: TTCACATCAGCAGGCATGTCC	91
Gene5023	EGR_06269	Minor histocompatibility H13	FP: TATTACCGTTTTCCGTCGAGA RP: GCGGCAAATATCTTAAACACC	96
Gene1575	EGR_09743	Collagen, type III, alpha	FP: AAAGGACTCTACTGGACCTCG RP: GAAGTCCACTGATCGCACGTT	84
Gene3068	EGR_08242	Signal transducing adapter molecule	FP: TGATGCCCAAATGCTACACC RP: CGAAATCGAATAGAGCACGAAC	140
Down-regulated DEGs				
Gene885	EGR_10426	Transcription elongation factor	FP: TGATTGAGCCACTACCGAAGCA RP: AGTCCTCCAGGCATATTCCAC	123
Gene2683	EGR_08627	Fatty acid-binding protein	FP: AAACCCGTGCTTACCTTCG RP: CATCCGTGGTGTTCATCG	145
Gene1583	EGR_09751	heat shock protein	FP: TCCAGGCCGTATTACTGAGCA RP: ATGACTTTAGTGCCACCGT	96
Gene2093	EGR_09229	LIM and SH3 domain protein	FP: GCCAGTGTGACGTAGTCCTC RP: CGTTTGTGTGTTCTTCGCCAA	107
Gene10528	EGR_00894	hypothetical protein	FP: AGTCAATTCAAGTAATCGTGCTG RP: TTCTCGCTTGATTTTCGCCACT	80
Gene7715	EGR_03601	BoIA-like protein 3	FP: TTCTTAATGTCTGCCGCCAT RP: AGTAGAGCCTCCTTATAGTCC	104
Control primer				
Gene2256	EGR_09068	Actin II	FP: ATGGTTGGTATGGGACAAAAGG RP: TTCGTCACAATACCGTGCTC	118

**Table 2**

Sequencing data statistics

Sample	Raw reads	Raw bases	Clean reads	Clean bases	Error Rate (%)	Q20 (%)	Q30 (%)	GC Content(%)
EgPSC_0d_1	60,847,206	9,187,928,106	60,032,580	8,950,587,025	0.0274	97.19	91.78	47.07
EgPSC_0d_2	65,486,882	9,888,519,182	64,718,518	9,648,903,196	0.0267	97.47	92.42	47.43
EgPSC_0d_3	44,838,914	6,725,837,100	44,709,544	6,679,256,433	0.012	98.31	95.24	46.25
EgPSC_10d_1	50,879,460	7,682,798,460	49,912,092	7,433,785,290	0.0232	98.79	95.88	41.48
EgPSC_10d_2	45,451,476	6,863,172,876	44,642,728	6,648,774,193	0.0234	98.73	95.67	43.67
EgPSC_10d_3	51,627,418	7,795,740,118	50,541,444	7,440,873,654	0.0234	98.72	95.73	45.1
EgPSC_20d_1	62,302,076	9,407,613,476	46,103,180	6,618,060,319	0.0231	98.78	96.06	48.29
EgPSC_20d_2	56,237,888	8,491,921,088	45,097,156	6,377,338,828	0.0232	98.76	96.02	48.01
EgPSC_20d_3	54,920,334	8,292,970,434	52,009,700	7,464,895,752	0.0233	98.77	95.88	45.28
EgPSC_40d_1	59,219,662	8,942,168,962	58,640,210	8,770,998,130	0.0235	98.66	95.6	47.32
EgPSC_40d_2	52,049,106	7,859,415,006	51,576,654	7,738,872,659	0.0237	98.61	95.46	48.24
EgPSC_40d_3	54,113,540	8,171,144,540	53,665,904	8,042,017,580	0.0238	98.58	95.35	47.31
EgPSC_80d_1	49,360,298	7,453,404,998	48,784,352	7,308,163,088	0.0239	98.52	95.23	47.27
EgPSC_80d_2	49,876,008	7,531,277,208	49,296,316	7,387,535,278	0.024	98.49	95.13	46.91
EgPSC_80d_3	50,275,646	7,591,622,546	49,770,288	7,457,638,729	0.0241	98.45	95.03	46.99
Average	53,832,394	8,125,702,273	51,300,044	7,597,846,677	0.0232	98.46	95.09	46.44

**Table 3**

COG classification statistics of the consistently up-regulated differential expressed genes

Category	Functional description	Gene id	Gene name	Gene description
Information storage and processing	RNA processing and modification	gene9743	EGR_01647	Transcription elongation regulator
		gene5441	EGR_05864	Potassium voltage-gated channel subfamily C member 3
metabolism	Amino acid transport and metabolism	gene8350	EGR_02924	Ornithine aminotransferase
Information storage and processing	Translation, ribosomal structure and biogenesis	gene2299	EGR_09013	Epsin-1
		gene11020	EGR_00111	NEDD8-conjugating enzyme UBE2F
poorly characterized	Function unknown	gene7950	EGR_03318	Synaptic vesicle membrane protein VAT-1-like protein
		gene10887	EGR_00659	WD repeat-containing protein
		gene1175	EGR_10139	5'-tyrosyl-DNA phosphodiesterase
		gene7701	EGR_03587	Josephin-2
		gene7722	EGR_03608	Cornifelin
		gene246	EGR_11073	Glycosyltransferase-like protein LARGE2
Cellular processes and signaling	Intracellular trafficking, secretion, and vesicular transport	gene1459	EGR_09846	1,5-anhydro-D-fructose reductase
		gene1432	EGR_09887	Platelet glycoprotein

**Table 4**

Statistics on the expression of 10 consistently up-regulated genes

Gene ID	Gene Name	Gene Description	TPM of different stages				
			0d	10d	20d	40d	80d
gene7373	EGR_03870	hypothetical protein	1178.71	230.68	527.33	2007.29	4039.22
gene5859	EGR_05443	hypothetical protein	11.67	10	23.96	199.76	395.56
gene2035	EGR_09260	hypothetical protein	0.14	2.24	0.85	4.85	10.39
gene5066	EGR_06312	hypothetical protein	1.71	8.72	2.01	78.61	154.78
gene9360	EGR_01986	hypothetical protein	44.24	100.54	20.87	55.48	251.26
gene1117	EGR_10197	hypothetical protein	5.91	53.94	13.79	907.96	1823.58
gene9359	EGR_01985	hypothetical protein	40.05	71.03	14.37	46.97	190.86
gene8053	EGR_03148	hypothetical protein	63.06	122.69	38.63	115.20	275.06
gene1432	EGR_09887	Platelet glycoprotein	0.56	3.95	5.30	20.32	48.29
gene1839	EGR_09469	Dynein light chain 2	501.76	1340.03	248.19	1400.56	3242.24

## Figures

Fig 1. A



Fig 1. D

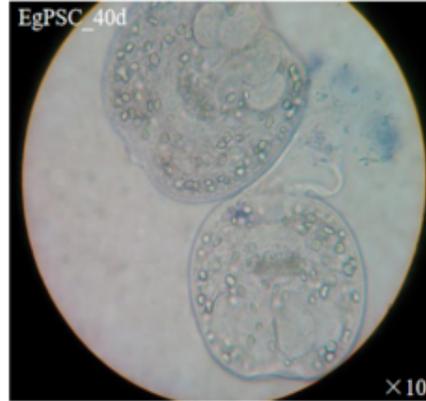


Fig 1. B

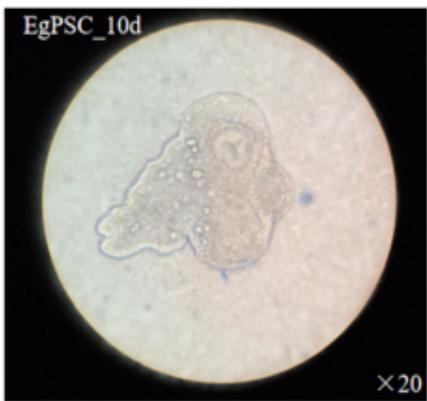


Fig 1. E

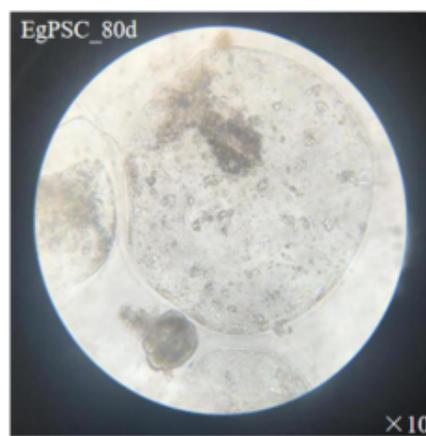
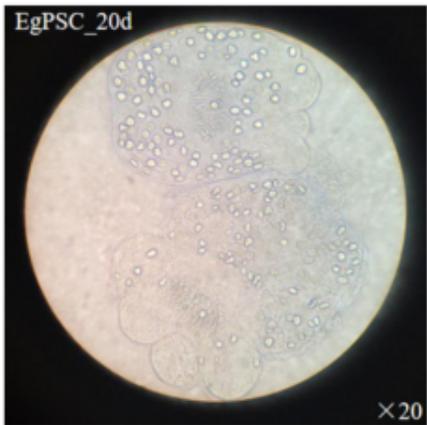


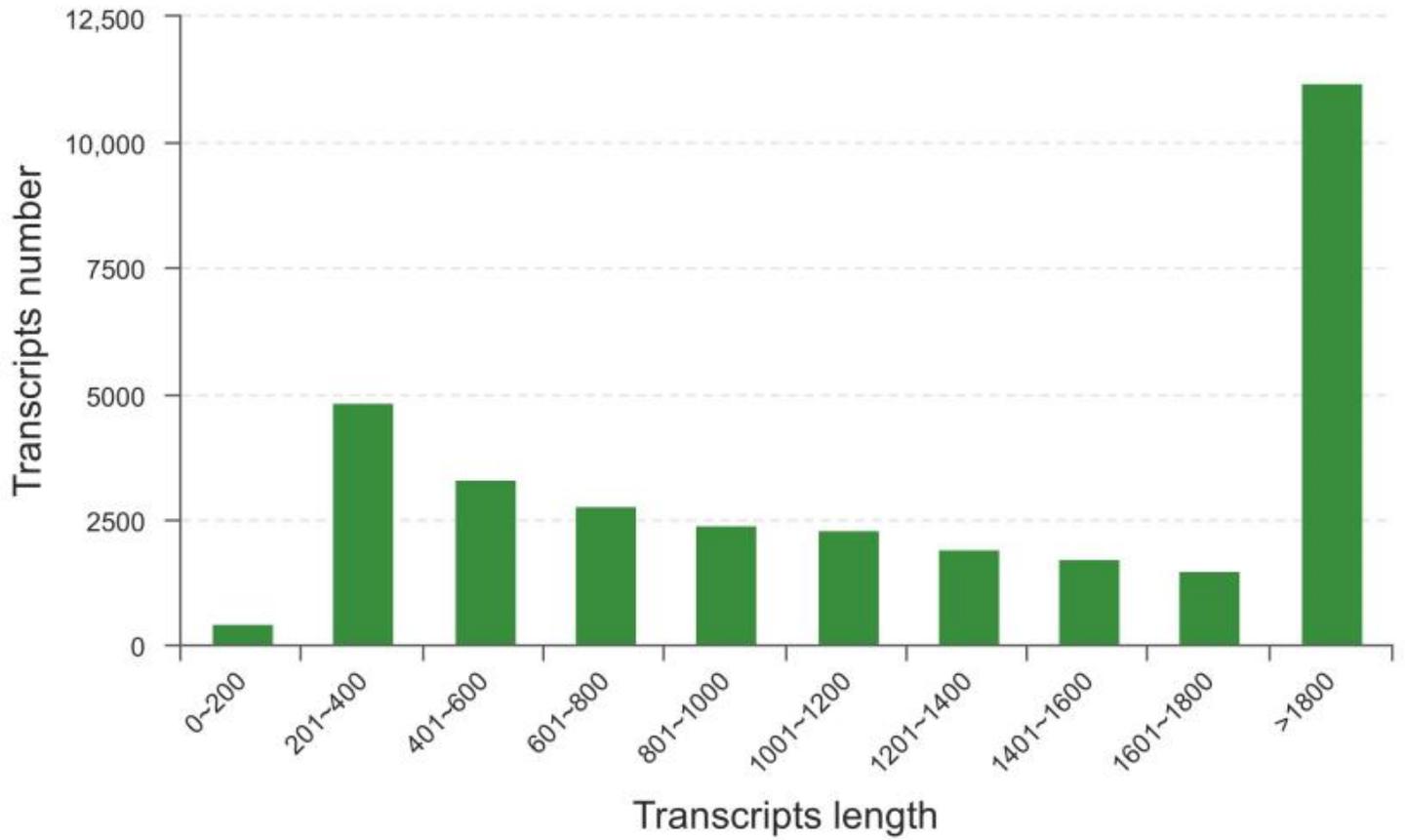
Fig 1. C



## Figure 1

Protoscoleces were cultured in vitro at different developmental stages. (A) The 0th day of culture (EgPSC\_0d); (B) The 10th day of the complete gestation period (EgPSC\_10d); (C) The 20th day of the complete gestation period (EgPSC\_20d); (D) The 40th day of the complete gestation period (EgPSC\_40d); (E) The 80th day of the vesicle enlargement period (EgPSC\_80d).

## Length distribution of Transcripts



**Figure 2**

Assembled transcript length distribution of *E. granulosus* protoscoleces.

Fig 3. A

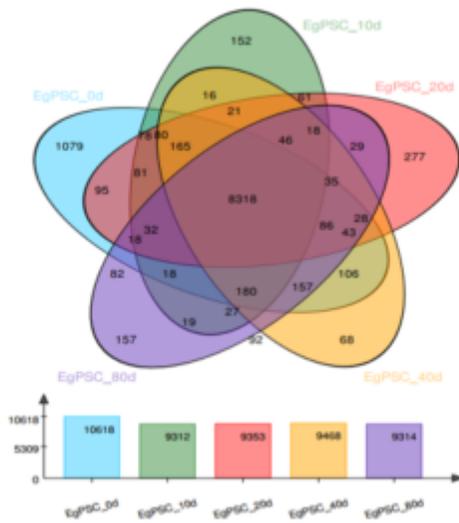


Fig 3. B

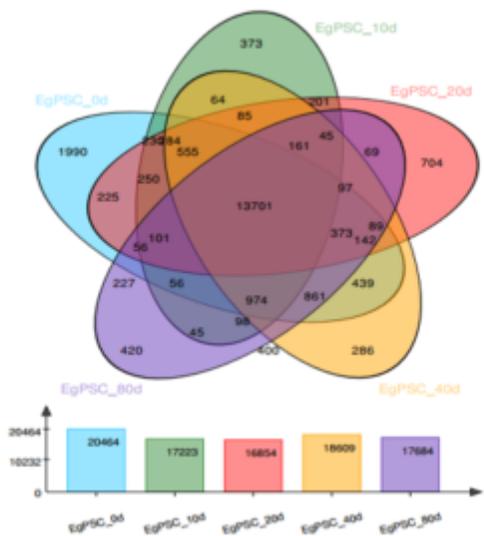


Figure 3

Venn diagram shows the commonality and difference of different samples. Note: Circles of different colors represent a set of genes (A) and transcripts (B) expressed in a sample, where the values represent the number of common and unique expressed genes/transcripts between a two or three group sample, and the sum of all numbers inside the circle represents the group. The sum of the number of expressed genes, the cross-region of the circle represents the number of shared genes of each group of expressed genes/transcripts.

Fig 4. A

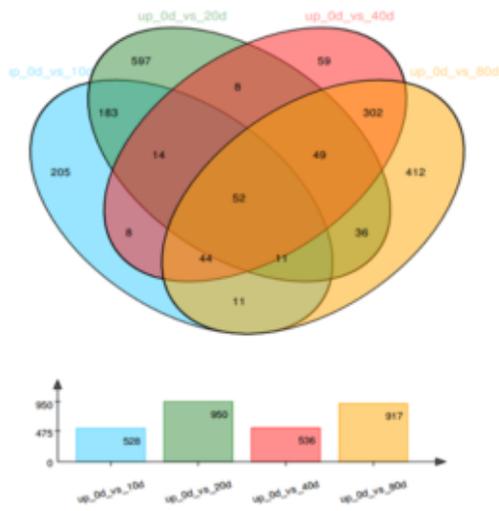


Fig 4. B

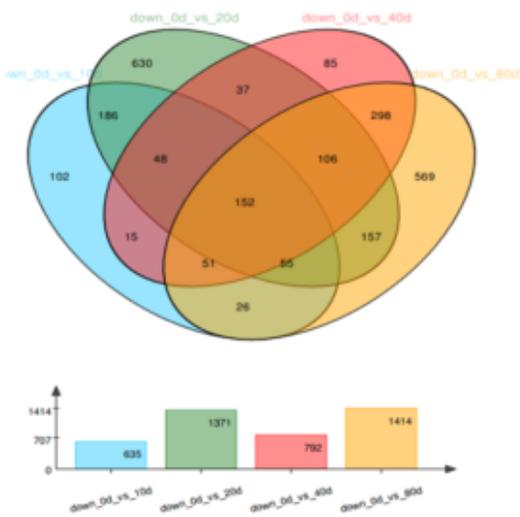


Fig 4. C

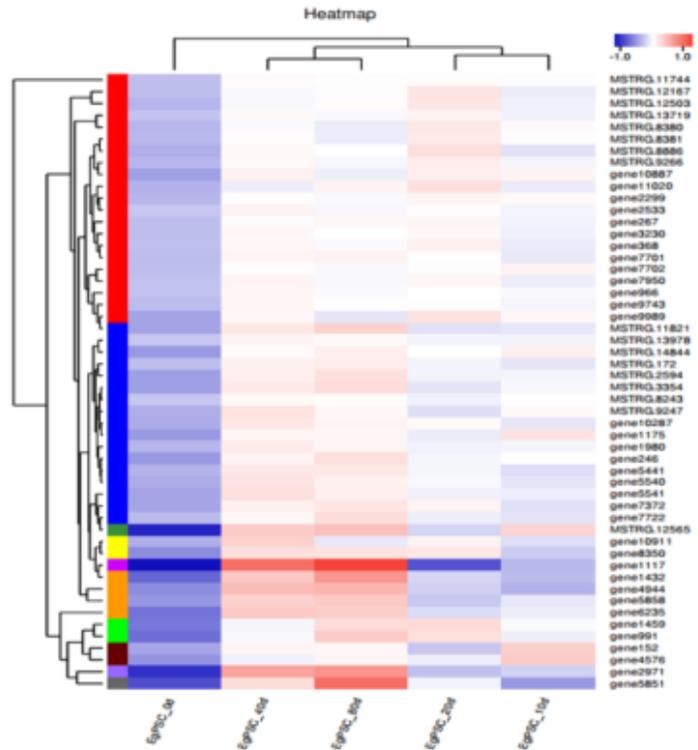


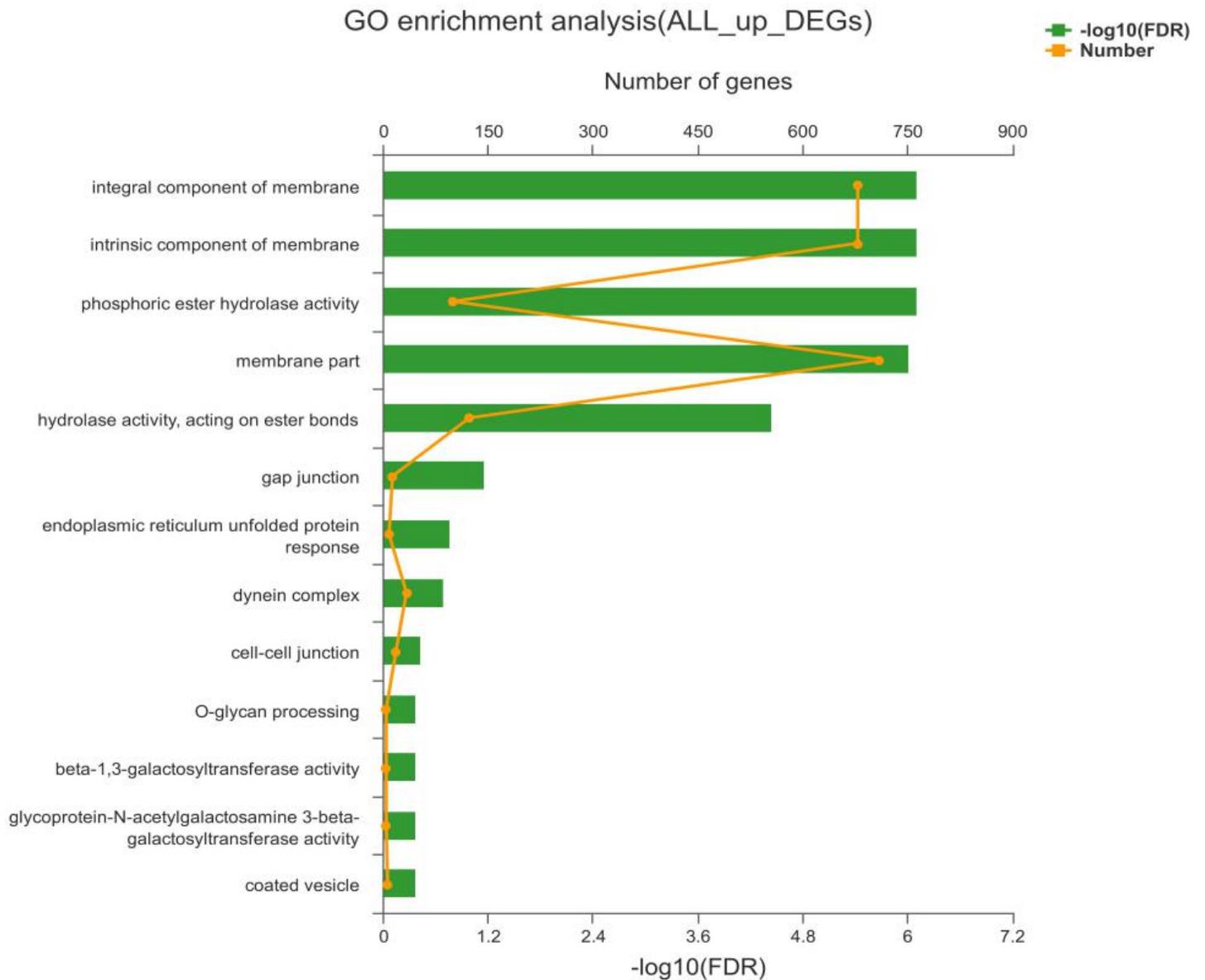
Fig 4. D



Figure 4

Transcription profile analysis of significantly differentially expressed genes. Venn diagram showing the total number of differential expressed genes in each group in the large circles (EgPSC\_0d as the control group), including significantly up-regulated DEGs (A) and down-regulated DEGs (B), respectively, and the number of genes expressed in both groups are shown in overlapping portion of the circle; (C). Hierarchical clustering map of 52 up-regulated DEGs showing the overall TPM. log<sub>10</sub> (TPM+1) values are scaled and

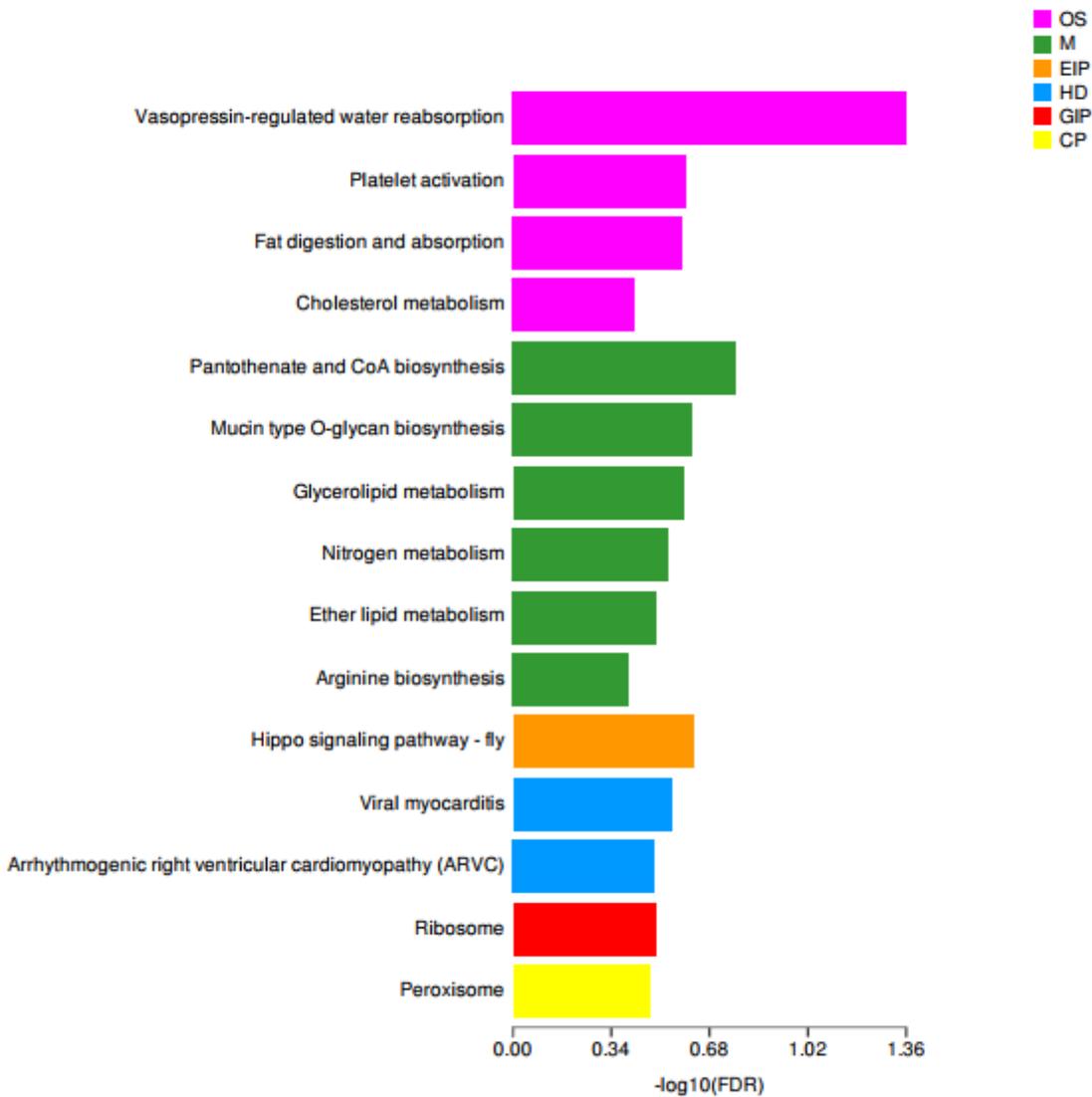
clustered, with red representing up-regulated genes and blue representing down-regulated genes. The upper part is the tree diagram of the sample cluster, and the name of the sample is below. The closer the two-sample branches are, the closer the expression pattern of all genes in the two samples is, means the closer the change trend of gene expression quantity is; (D). Visualization of inter-gene expression correlation of 52 up-regulated DEGs, each node represents a gene, and the line between nodes represents the correlation between gene expressions, the larger the node, the more the gene is correlated with other genes.



**Figure 5**

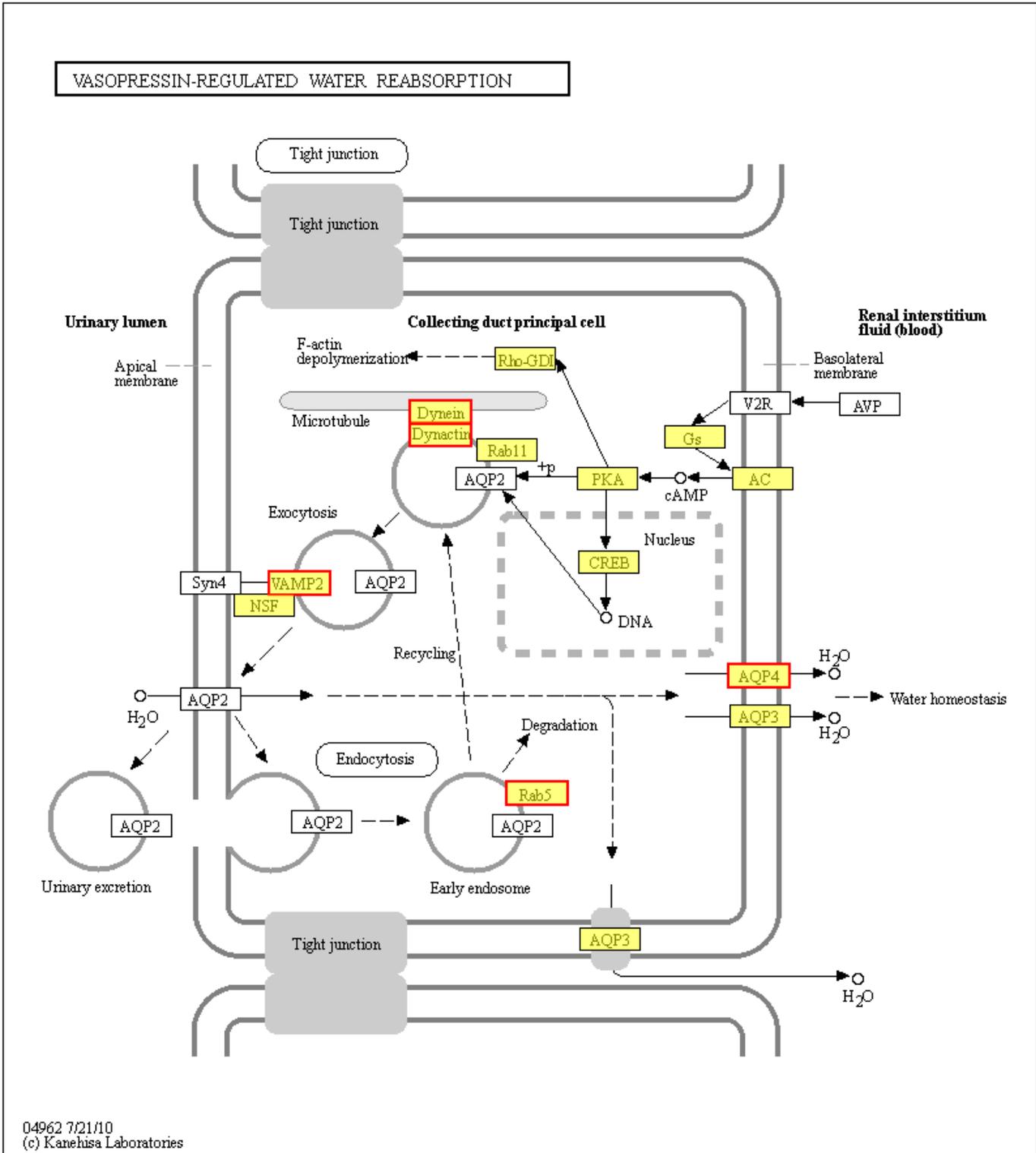
Gene Ontology (GO) transcript classification statistics. The ordinate represents the GO term, and the abscissa above represents the number of genes compared to the GO term, which corresponding to the different points on the broken line. The lower abscissa represents the significance level of enrichment, corresponding to the height of the column. Among them, the smaller FDR is, the larger the  $-\log_{10}(\text{FDR})$  value, and the more significant the GO term is.

## KEGG enrichment analysis(ALL\_up\_DEGs)



**Figure 6**

Statistical classification of transcriptome pathways. Kyoto Encyclopedia of Genes and Genomes (KEGG) assignment of DEGs in the protoscolex transcriptome of *E. granulosus*, including organismal systems, metabolism, environmental information processing, human diseases, genetic information processing, and cellular processes .



**Figure 7**

Vasopressin-regulated water reabsorption. All the products with color background in the figure belong to KEGG annotation results of this sequencing, in which the yellow background represents the reference gene, and the red border indicates that the genes in the analyzed gene set are annotated.

Fig 8. A

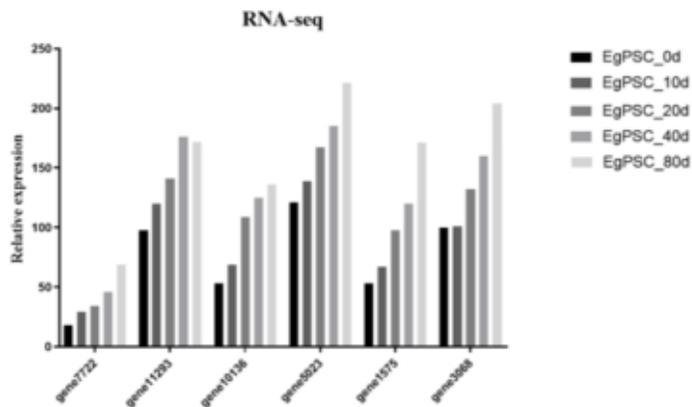


Fig 8. C

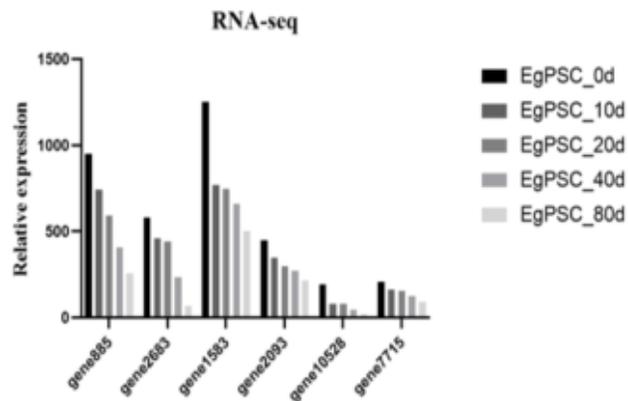


Fig 8. B

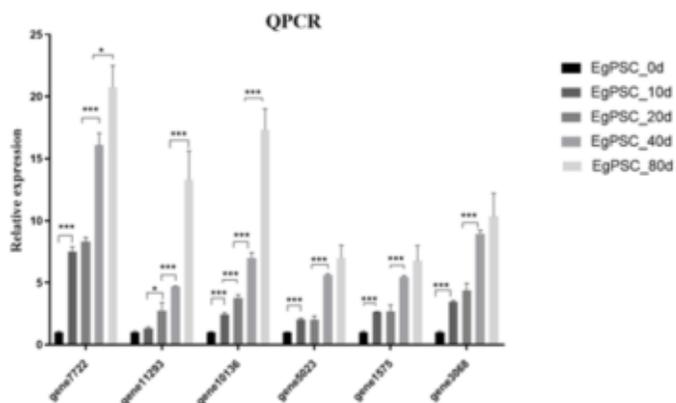


Fig 8. D

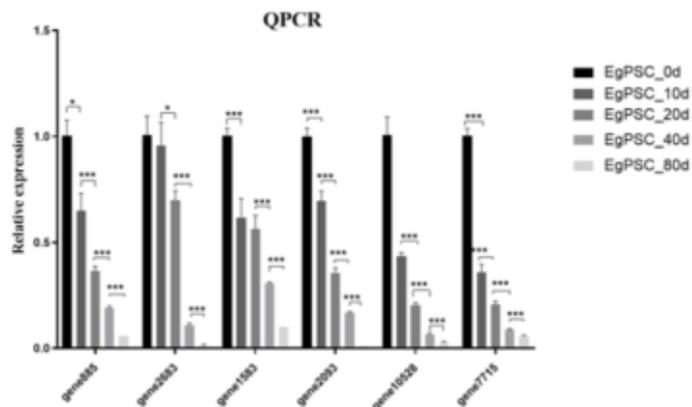


Figure 8

Results of RNA-Seq verification by QPCR. The internal control gene was selected Actin II gene.

## Supplementary Files

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

- [Additionalfile7TableS4.doc](#)
- [Additionalfile4TableS3.doc](#)
- [Additionalfile2figureS1.doc](#)
- [Additionalfile5FigureS2.doc](#)
- [Additionalfile3TableS2.doc](#)
- [Additionalfile6FigureS3.doc](#)
- [Additionalfile1TableS1.doc](#)