Transcriptome-wide gene expression profiling in oncospheres and metacestodes of *Echinococcus multilocularis*

多包条虫の虫卵から成熟包虫における遺伝子発現の推移

The United Graduate School of Veterinary Science

Yamaguchi University

FUQIANG HUANG

September 2017

Contents

List of Abbreviations

3H-T	3H-thymidine
4Wmet	4-week Immature Metacestodes
16Wmet	16-week Mature Metacestodes
AE	Alveolar Echinococcosis
AgB	Antigen B
Aonc	Activated Oncospheres
ATP	Adenosine Triphosphate
BCV	Biological Coefficient of Variation
BLAST	Basic Local Alignment Search Tool
BCV	Biological Coefficient of Variation
CE	Cystic Echinococcosis
Cmet	Metacestode Small Vesicles Cultivated *in vitro*
CPM	Counts Per Million
DC	Dendritic Cells
DEGs	Different Expression Genes
DS-cells	Dark-Stained Undifferentiated Cells
eIF4E	Eukaryotic Translation Initiation Factor 4E
ES Protein	Excretory-Secretory Protein
FDR	False Discovery Rate
FGF	Fibroblast Growth Factor
FPKM	Fragments Per Kilobase of Transcript Per Million
GAPDH	Glyceraldehyde 3-phosphate Dehydrogenase
GPI	Glycosylphosphatidylinositol
GO	Gene Ontology
HSP	Heat Shock Protein
KEGG	Kyoto Encyclopedia of Genes and Genomes
KOBAS	KEGG Orthology Based Annotation System
LCPC	Longest Contig per Component
LEL	Large Extracellular Loop
LL	Laminated Layer
LS-cells	Light-Stained Undifferentiated Cells
NGS	Next-Generation Sequence
Nonc	Non-activated Oncosphere
Met	Metacestode

Onc	Oncosphere
ORFs	Open Reading Frames
PCR	Polymerase Chain Reaction
RPKM	Reads Per Kilobase of transcript Per Million
SL-TS	Spliced-Leader Trans-Splicing
SLC10	Solute Carrier Family 10
SNP	Single-Nucleotide Polymorphism
TM Protein	Transmembrane Protein
TMG	Trimethyl-guanosine
TSP	Tetraspanin
%PF	The Total Fraction of Passing Filter Reads Assigned to an Index

List of Tables

List of Figures

General introduction

Echinococcosis is zoonotic diseases caused by infection with larvae of the genus *Echinococcus* (order Cyclophyllidea, family Taeniidae) in the intermediated hosts. Alveolar echinococcosis (AE) and cystic echinococcosis (CE) are the two main types of this disease (Eckert, 2001). Since CE and AE are important for medical and veterinary point of view, the two pathogens, *E. granulosus* (*sensu lato*) and *E. multilocularis*, are well studied. *E. granulosus* (*sensu lato*) is present in over 100 countries from all continents except Antarctica (Eckert and Deplazes, 2004) and up to eleven different strains are identified (Maule and Marks, 2006; McManus, 2009). On the other hand, *E. multilocularis* is restricted to endemic regions in the Northern Hemisphere (Davidson et al., 2012) and has a much lower global incidence. However, AE is more difficult to treat than CE (Stojkovic and Junghanss, 2012). AE is almost impossible to cure when detected at late stages of development, and is typically lethal if left untreated (Eckert and Deplazes, 2004). This makes AE to be the most serious zoonosis in the Northern Hemisphere, and it has been estimated that the global burden of AE is comparable to that of many other neglected tropical diseases such as Leishmaniasis and Trypanosomiasis, for which research efforts are much more intensive (Torgerson et al., 2010).

1. *E. multilocularis*

1.1 Life cycle and biology of *E. multilocularis*

E. multilocularis, regarded as an independent species in the 1950's (Tappe et al., 2010), is a tiny tapeworm of the fox with four or five segments, around 1.2~4.5mm in length, and completes its life cycle via transmission between two different hosts (Figure 1). The definitive host (mainly foxes and dogs) of the parasite which has a sexual and adult stage in the intestine can produce eggs (Deplazes and Eckert, 2001; Díaz et al., 2015). And the intermediate host (mainly small rodents) are infected by ingestion of eggs of the parasite (Díaz et al., 2015). After hatching of eggs, the oncospheres (the first larval stage) are released and activated in the small intestine. Activated oncospheres penetrate into the intestinal wall and are carried by blood or lymph to internal organs (mainly the liver) and cause AE (Díaz et al., 2015). Oncospheres then develop into metacestodes and can produce numerous small vesicles, which develop germinal and laminated layers. Some germinal cells initiate the production of brood capsules and protoscoleces in each vesicle (Eckert, 2001). In the susceptible intermediate host, protoscoleces start to be produced after 40 days of post infection. The definitive host acquires infection by ingesting protoscoleces and the protoscoleces establish in the small intestine of the definitive host. The head of the adult is denominated the scolex and it contains the attachment organs (suckers and rostellum with hooks). Behind the scolex, the neck region proliferates extensively and continuously generates a chain of segments (proglottids), each one developing a complete set of male and female reproductive organs.

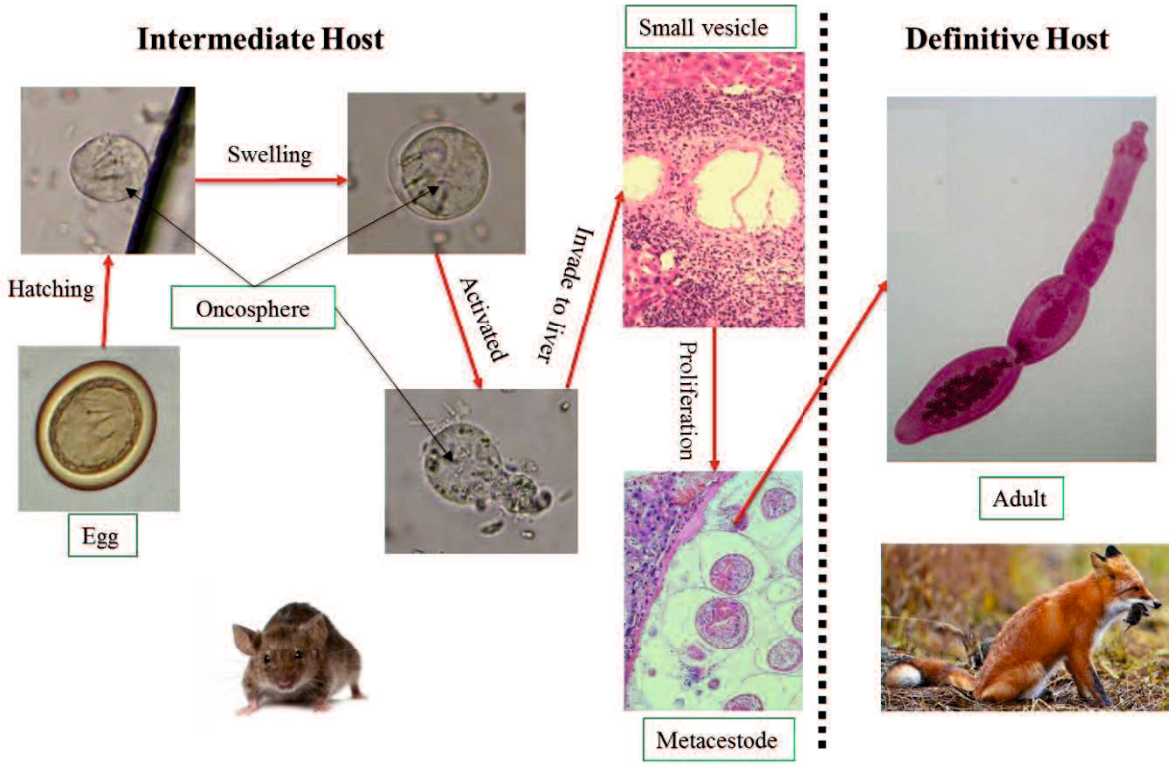


Figure 1. **Life cycle of *E. multilocularis***

1.2 Genomics of *Echinococcus* spp.

A meeting was held at the Wellcome Trust Sanger Institute that led to the still ongoing '50 helminth genomes initiative' in March 2004 (Koziol and Brehm, 2015). At this time point, there hadn't draft genome datasets of helminths, although the projects for the trematodes *Schistosoma mansoni*, *Schistosoma japonicum*, and the nematode *Brugia malayi* were in an advanced stage (Brindley et al., 2009; Koziol and Brehm, 2015). However, there were 98 Nematode and 30 Platyhelminthes draft genome datasets in the WormBase (http://parasite.wormbase.org/species.html) until 2016/10/2, and 4 datasets were *Echinococcus* spp.

Because of AE is fatal for humans and it is a well-documented laboratory model for host-parasite interplay (Gottstein and Hemphill, 2008; Brehm, 2010), study of the parasite has increased recently. And, genetic homogeneity due to inbreeding was one

aspect that greatly facilitated genome assembly and thus contributed to the high quality of the *E. multilocularis* draft genome (Tsai et al., 2013). In parallel, Next Generation Sequence (NGS) technology was used to produce draft sequences for *E. granulosus* (*sensu stricto*, G1 strain) (Tsai et al., 2013; Zheng et al., 2013) and *E. canadensis* (G7 genotype, isolated from Argentina). As for *E. granulosus* G1 strain genome, it was firstly carried out on the framework of the *E. multilocularis* (German isolates) reference genome (Tsai et al., 2013). Later, these cestodes whole genome sequencing projects were complemented by efforts of a Chinese/Australian consortium, which produced a second *E. granulosus* draft genome from a single hydatid cyst (G1 strain) using 454 and Solexa NGS (Zheng et al., 2013) and the *E.canadensis* genome is implement and almost finished. In all three completely *Echinococcus* spp. genome projects (Tsai et al., 2013; Zheng et al., 2013), gene finding and annotation was supported by extensive EST- and NGS-transcriptomic analyses of several life cycle stages such as protoscolex, metacestode and adult (Parkinson et al., 2012; Tsai et al., 2013; Zheng et al., 2013). The studied tapeworms have much smaller genomes than the related flukes (about three times) or free-living flatworms (about nine times), which is mostly due to smaller intergenic regions, smaller introns, and a lower content of repeats and mobile genetic elements in tapeworm genomes (Tsai et al., 2013). Depending on the methodology used, between 10,300 and 11,300 genes were predicted in *Echinococcus* spp. genomes (Tsai et al., 2013; Zheng et al., 2013). Furthermore, although *E. multilocularis* and *E. granulosus* have a well-developed nervous system (Brownlee et al., 1994; Fairweather et al., 1994; Camicia et al., 2013; Koziol et al., 2013), several homeobox gene families involved in neural development are missing (Tsai et al., 2013) in currently assembled genome.

Spliced-leader trans-splicing (SL-TS) is an mRNA maturation process, similar to intron splicing, which has been shown to occur in some parasites but not in theirs hosts (mammals) and are suggested to be exploitable for drug development for helminthiases (Liu et al., 2009). It is shown that the spliced leader, which is donated by a small RNA, is encoded elsewhere on the genome of *E. multilocularis* by genes that are tandemly arrayed (Brehm et al., 2000a). The transcripts of trans-spliced genes thus all harbor an identical exon at their 5′ end which, in the case of *E. multilocularis*, is 36 bases long and contains a trimethyl-guanosine (TMG) cap, which differs from the 7-methyl-guanosine (7mG) cap present at the 5′ end of usual (non-trans-spliced) mRNAs (Brehm et al., 2000a; Lasda and Blumenthal, 2011). The function of trans-splicing remains unclear. But, certain hypotheses argue that it is an adaptive process for coordinated gene regulation or translational control, others argue the spliced-leader genes might be selfish DNA components that 'hijack' essential nuclear genes until they cannot be eliminated from the genome, leading to an expansion of Spliced-leader genes in tandem arrays (Blaxter and Liu, 1996; Blumenthal, 2004). It is revealed that 13% genes in the *E. multilocularis* genome are trans-spliced (Tsai et al., 2013). Among the trans-spliced genes, some of them are involved in essential cellular processes such as transcriptional and translational control, splicing or replication (Brehm et al., 2000a; Tsai et al., 2013). Hence, if trans-splicing or the translation of trans-spliced messages could be inhibited, this would surely result in lethal effects for all parasite cells (Koziol and Brehm, 2015). It is difficult for design drugs that directed against the splicing process itself to against cestodes, since trans-splicing is carried out by the canonical spliceosome, the components of which are highly conserved between cestodes and mammals. However, as already suggested previously (Liu et al., 2009), the eukaryotic translation initiation

factor 4E (eIF4E), which initiates translation by binding to the mRNA cap structure, could be a promising target. In mammals, eIF4E only recognizes 7mG cap structures, whereas in *Schistosoma*, the trans-splicing eIF4E recognizes both 7mG and TMG caps (Liu et al., 2009). The *E. multilocularis* genome contains one single copy gene for eIF4E and it is reasonable to assume that this factor also recognizes both types of caps (Koziol and Brehm, 2015). Hence, structural differences in the cap-binding structures of parasite and host eIF4E, which are likely to exist (Liu et al., 2009), could possibly be exploited for the development of small molecule compounds that target trans-splicing in cestodes (Liu et al., 2009).

Hence, with the establishment of the *in vitro* cultivation system (Brehm and Spiliotis, 2008) and the availability of the genome datasets, *E. multilocularis* is a promising model to study host-parasite interface. To facilitate an effective and wide use as a model, deep understanding of *E. multilocularis* biology especially at the molecular level and gene expression at different life-cycle stages is necessary and imperative for drugs design and vaccination development.

1.3 Structure of oncospheres and metacestodes of *Echinococcus* spp.

Numbers of studies have been undertaken to describe the structure of taeniid (Platyhelminthes: Cestoda: Taeniidae) eggs and/or oncospheres (Jabbar et al., 2010; Swiderski, 1983), including several research on *Echinococcus* spp.(Heath and Smyth, 1970; Heath and Lawrence, 1976; Swiderski, 1983; Harris et al., 1989; Holcman et al., 1994; Swiderski et al., 2016). Studies describe the structure of oncosphere of *Echinococcus* spp. mainly based on sections taken at random, and results show that mature *Echinococcus* spp. eggs have a thick embryophore and its ultrastructure shows it

is made of thick elongated blocks united by electron-lucid cement and the oncosphere membrane is a thin cytoplasmic layer surrounding the oncosphere. Furthermore, study by Swiderski (2016) described oncospheral hook morphogenesis for *E. multilocularis* and show that the blade and base gradually protrude outside the oncoblast plasma membrane during hook growth.

During early metacestode development in *E. multilocularis*, two types of undifferentiated cells were described at the ultrastructural level: “light stained undifferentiated cells” (LS-cells) and “dark-stained undifferentiated cells” (DS-cells), both of which were found in mitosis (Sakamoto and Sugimura, 1970). LS-cells were only found during the earliest stages of the oncosphere to metacestode metamorphosis, and were proposed to give rise to the DS-cells. DS-cells accumulate during the formation of brood capsules and protoscoleces, and it was proposed that DS-cells differentiate into several cell types such as tegumental cells, muscle cells and glycogen storing cells (Sakamoto and Sugimura, 1970). Ultrastructural studies of *E. granulosus* metacestodes also described other differentiated cell types, such as calcareous corpuscle cells and excretory cells (cells of the excretory tubules and flame cells) (Lascano et al., 1975; Smith and Richards, 1993).

2. Transcriptome and RNA-seq

A transcriptome is the full range of messenger RNA molecules expressed by an organism. The term "transcriptome" can also be used to describe the array of mRNA transcripts produced in a particular cell, tissue type or life stages. In contrast with the genome, which is characterized by its stability, the transcriptome actively changes

(Velculescu et al., 1997). And it is believed that organisms maintain their stability against external and internal changes by regulating gene expression at certain time point. Characterization of all the RNAs transcribed therefore helps us understand profound bio-processes (Velculescu et al., 1997).

2.1 Transcriptome: an entire dynamic RNA profile

The term transcriptome was first used in yeast to describe the genes that were being expressed and their expression levels at distinct cell phases (Velculescu et al., 1997; Zheng, 2012). It is proved that alternatively splice is a common phenomenon in eukaryote, for example, transcripts from genes with multiple exons compose up to 95% and more than seven alternative splicing events occur in every multi-exon gene in human tissue (Pan et al., 2008). Transcriptome analysis provides a wealth of bio-information as it records direct information of the transcription of an organism's genome. Transcriptome analysis allows one to analyze such expression patterns as alternative splicing and find new exons, especially ones of a small size, or genes or single-nucleotide polymorphism (SNP) (Sultan et al., 2008). In particular, it is quite useful for annotating genomes of organisms that are poorly understood (Yassour et al., 2009). Moreover, transcriptome analysis is a powerful tool to be able to compare the up or down regulation of gene expression in cells or tissues exposed to different conditions and stages. A study comparing matured metacestodes and adult of *E. multilocularis* revealed that there are more than 1,000 transcripts are significant different expressed (Tsai et al., 2013). This analysis gives us valuable data to search the target genes for diagnostic and vaccination for echinococcosis.

2.2 RNA Sequencing (RNA-Seq): a deep high-throughput technology for transcriptional characterization

RNA-Seq, also called whole transcriptome shotgun sequencing (Morin et al., 2008), using NGS to reveal the presence and quantity of RNA in a biological sample at a given moment in time (Wang et al., 2009; Chu and Corey, 2012). Different from Automated Sanger Sequencing (first generation), NGS technologies have several different approaches, including 454, Illumina and SOLiD, et al., but the most popular one is the Illumina platform.

Chapter 1. RNA sequencing of oncospheres and metacestodes of *E. multilocularis*

Abstract

In the present study, seven samples at stage of non-activated oncospheres (Nonc), activated oncospheres (Aonc), 4-week metacestodes *in vivo* (4Wmet), 16-week metacestodes *in vivo* (16Wmet) and *in vitro* cultivated metacestodes (Cmet) of *E. multilocularis* were collected with the aim of measuring dynamics expressed RNA transcripts that occur during parasite development. The single-end (s4Wmet and s16Wmet) and pair-end (pNonc, pAonc, p4Wmet, pCmet) sequencing by Illumina's Genome Analyzer platform and other bioinformatics analyses was used for RNA sequencing approach, respectively. The result show that 700 million clean reads with > 90% of all bases having Phred (Q) scores above 30, and most of *de novo* assembled contigs can matched to the reference genome of *E. multilocularis* which indicated that all sequenced reads of this seven samples and the assembled contigs of pair-end sequence samples are reliable.

Key word: Oncospheres, Metacestodes, Reads, Contigs, Transcriptome

1. Introduction

For the development of diagnostic antigen or vaccine targets of AE, gene expression data of oncospheres and early larval stages metacestodes were needed. At present, little gene expression data has been published for oncospheres and early larval stages. Thus, experiments on identifying antigens for use in immunodiagnostic assays is a crucial

point in the improvement of the diagnostic tool and must be based on the developmental stage of the parasite.

Now, the genome database of *E. multilocularis* has been recently published, since we want to check the trans-spliced transcripts in *E. multilocularis* genome, all the reads of pair-end sequenced samples will *de novo* assembled by Trinity software (Grabherr et al., 2011) and for check the *de novo* assemble is reliable or not, the BLASTN algorithm (Altschul et al., 1997) was used to Blast to *E. multilocularis* reference genome (German isolates).

2. Materials and Methods

2.1 Ethics statement

This study was carried out in strict accordance with the recommendations set out in the Guidelines for Animal Experimentation of the Japanese Association for Laboratory Animal Science, and the protocol for the animal experiments was approved by the ethics committee of the Hokkaido Institute of Public Health (Permission number: K25-02).

2.2 Preparation of parasite samples

E. multilocularis isolated in Hokkaido (Nemuro strain) was routinely maintained through a dog–cotton rat life cycle at the Hokkaido Institute of Public Health (Sapporo, Japan). Dogs were orally administered 5×10^5 *E. multilocularis* protoscoleces and the infection was terminated 35–77 days post infection by administering two tablets of Droncit® (Kouguchi et al., 2016).

2.2.1 Non-activated oncospheres (Nonc)

Feces were collected from experimentally infected dogs at 35 days post-infection. Eggs were isolated from feces by filtering by mesh, natural sedimentation and flotation with sugar solution. The isolated eggs were treated with 3% sodium hypochlorite for 20min for removal of the embryophore and sterilization. Non-activated oncospheres were collected at two times for biological replicates: September 2013 (sample, Nonc1) and December 2013 (sample, Nonc2).

2.2.2 Activated oncospheres (Aonc)

Techniques for activation of non-activated oncospheres were as previously described (Holcman et al., 1994; Santivanez et al., 2010). Briefly, non-activated oncospheres were activated with 1% pancreatin (Nacalai Tesque, Inc.), 1% hog bile extract (MP Biomedicals, LLC) and 0.2% Na_2CO_3 in RPMI 1640 (Gibco) at 38°C for 20min, and then cultivated in RPMI 1640 with 10% fetal calf serum (Gibco) at 38°C for 24h.

2.2.3 4-week immature metacestodes (4Wmet)

The DBA/2 mice were sacrificed after four weeks post oral infections with eggs and small lesions with early stage larvae were collected from the livers. The collected larvae were examined as 4-week metacestodes miniature vesicles (4Wmet) and have no protoscoleces, brood capsules and calcareous corpuscles.

2.2.4 16-week mature metacestodes (16Wmet)

The DBA/2 mice were sacrificed after 16 weeks post oral infections with eggs and lesions with larvae were collected from the livers. The collected larvae were examined

as 16-week metacestodes with protoscoleces, brood capsules and calcareous corpuscles (16Wmet).

2.2.5 Metacestodes small vesicles cultivated *in vitro* (Cmet)

In vitro cultivation of *E. multilocularis* was carried out as described previously (Spiliotis et al., 2004; Brehm and Spiliotis, 2008) . In short, cyst masses of metacestodes from intraperitoneal passage DBA/2 mice at 16 weeks were cut into small pieces and cultivated in DMEM (Gibco) with 10% fetal calf serum (Gibco) at 37°C. Miniature cysts were grown to small vesicles (2-4 mm in diameter) in several weeks but were harvested before the formation of brood capsules and protoscoleces.

2.3 Extraction of total RNA

Total RNA was extracted with protocols from Onc (2 samples), Aonc, 4Wmet (2 samples), 16Wmet and Cmet using Trizol (Invitrogen, cat.no.15596-026) and RNase-Free DNase Set (QIAGEN, cat.no.79254). Briefly, parasites were homogenized in 1 ml Trizol using mortar by adding liquid nitrogen. Then the total RNA was extract from Trizol, which contain the samples by the protocol for Trizol RNA isolation. The extracted RNA from each sample was eluted into nuclease-free water. To get rid of contaminating genomic DNA, the recovered RNA was purified by RNase-Free DNase Set (QIAGEN, cat.no.79254) according to the manual and suspended into RNase-free water. The concentration was determined by Nanodrop (Thermo Scientific).

2.4 Library construction and sequencing

The mRNA was extracted using the Illumina mRNA-Seq Sample Preparation Kit

according to manufacturer instructions. Briefly, total RNA was subjected to poly (A) selection using Sera-Mag Magnetic Oligo-dT beads. Poly (A+) RNA was partially degraded by incubating in fragmentation buffer at 94°C for 5 min. The first-strand cDNA was synthesized using random primers and SuperScript II (Invitrogen), and the second-strand cDNA was synthesized using RNaseH and DNA pol I (Illumina). Illumina GA sequencing adaptors were ligated to the cDNA ends. Double-stranded cDNA was size-fractionated by 6% polyacrylamide gel electrophoresis, and the band of 200 bp cDNA was recovered and amplified using Phusion DNA Polymerase (Finnzymes) in 15 cycles by PCR. Finally, >50 bp single- or pair- end read RNA-seq tags were generated (Figure 1-1) using the Genome Analyzer IIx (Illumina, San Diego, CA, USA) following methods in the User Guide.

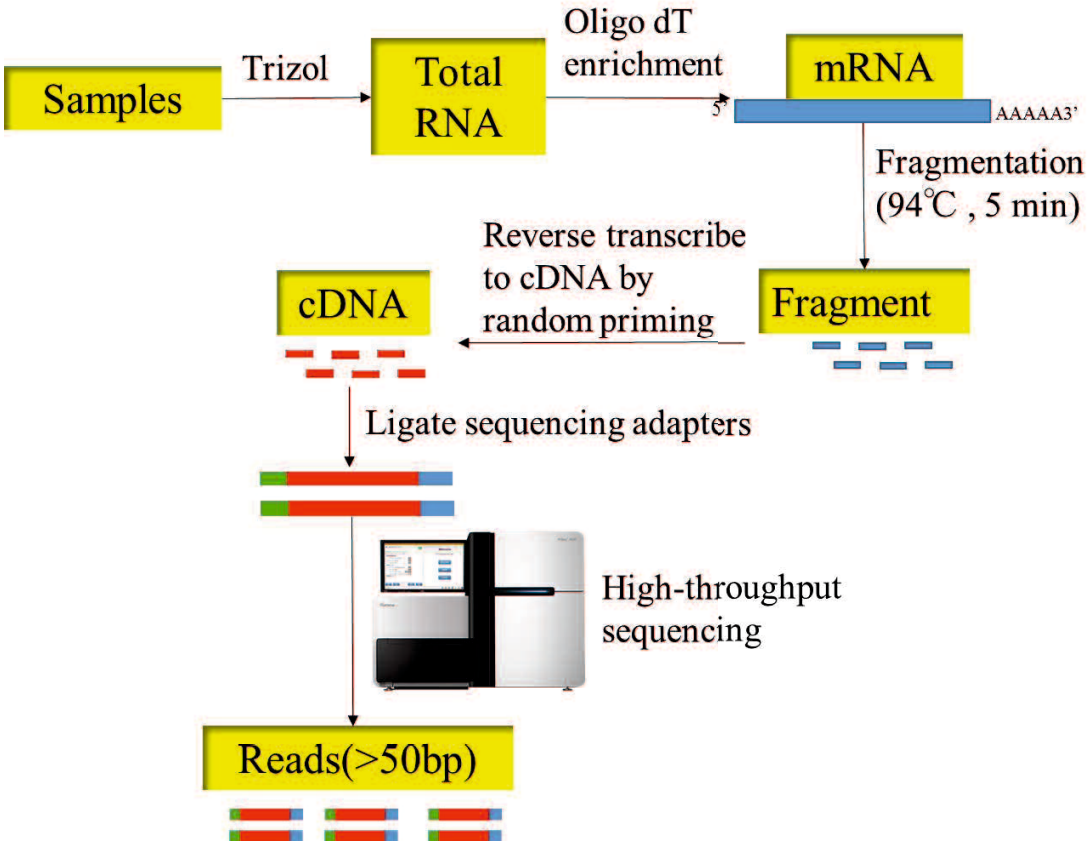


Figure 1-1. Flowchart of *E. multilocularis* RNA-Seq. technology.

2.5 Sequence data quality control

To assess the quality of each lane, reads obtained from each life-cycle stages were filtered by Perl script using the following criteria: 1) trim adapter; 2) remove Illumina-filtered reads; 3) remove reads with no-call bases (ex: AATC "N" ATGATAG); and 4) remove mouse-mapped reads.

2.6 *De novo* assembly for 100bp pair-end reads

As for Spliced-leader and trans-splicing analysis in *E. multilocularis* and it recommends using pair-end reads for *de novo* assembly by Trinity software (Grabherr et al., 2011), we conduct *de novo* transcriptome assembly using 100 par-end reads from oncospheres and metacestodes of *E. multilocularis*. Filtered reads were extracted using SAMtools (Li et al., 2009) and used for subsequent assemblies. *De novo* transcriptome assemblies of these filtered reads were performed using the Trinity software (Grabherr et al., 2011). The command line used for assembly was Trinity.pl–seq Type fq–JM 10G–left reads-1.fq–right reads-2.fq–min_contig_length 200. The final output from Trinity was a large number of assembled FASTA sequences. After assembly, TransDecoder (http://transdecoder.sf.net) was used to identify open reading frames (ORFs) with complete coding sequences. To remove possible sources of contamination from assemblies of individual samples, the list of contigs from individual samples was used as a query for a BLASTN (Altschul et al., 1997)(2.2.31 release of NCBI-BLAST+) search against *E. multilocularis* reference genome (German isolates) that deposited in WormBase ParaSite (http://parasite.wormbase.org/species.html).

3. Result

3.1 RNA-Seq sequencing data Analysis

pNonc1, pNonc2, pAonc, p4Wmet and pCmet (Figure 1-2) were used for pair-end sequencing (p) and s4Wmet and s16Wmet (Figure 1-3) were used for single-end sequencing (s). And it was show that the clean reads which originated from pair-end sequencing were almost twice for single-end sequencing at the stages of metacestodes, *in vivo* (Table 1-1) which indicated that the initial mRNA extract from different 4Wmet samples were most equal. The quality of obtained reads was excellent with more than 90% reads having a quality score at Q30 (error probability of 0.001) or higher (Table 1-1).

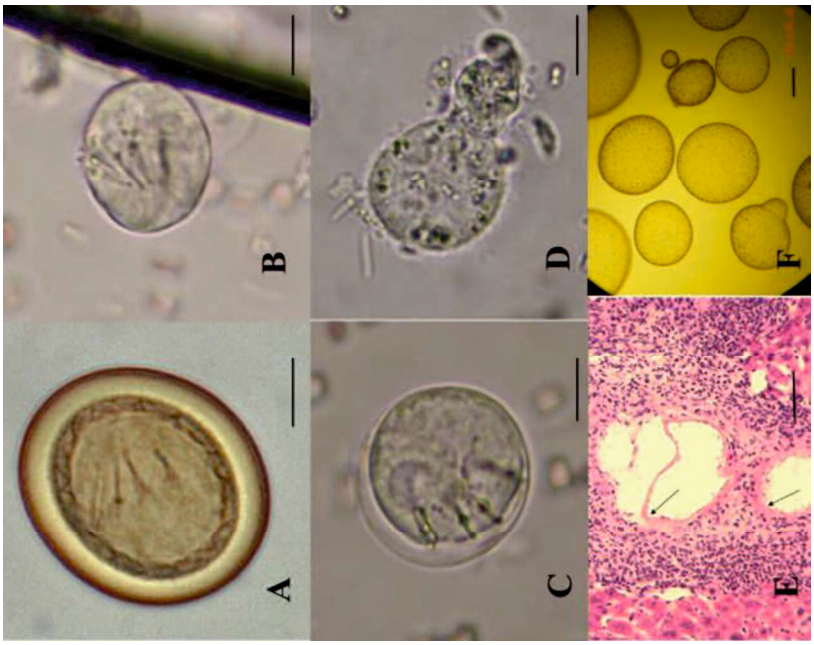


Figure 1-2. Morphology of different life cycle stages of *E. multilocularis.* A: Egg; B: Non-activated oncospheres (Nonc); C: Activating oncospheres with the hooks dispersive and body swelling; D: Activated oncospheres (Aonc) with the hooks aggregation in the smaller lobe after 24 hours activation; E: 4-week metacestodes, *in vivo* (4Wmet); F: Metacestodes, cultivated *in vitro* (Cmet); Bar: 10μm (A-D), 100μm (E) and1mm (F); Arrowhead: Miniature vesicles.

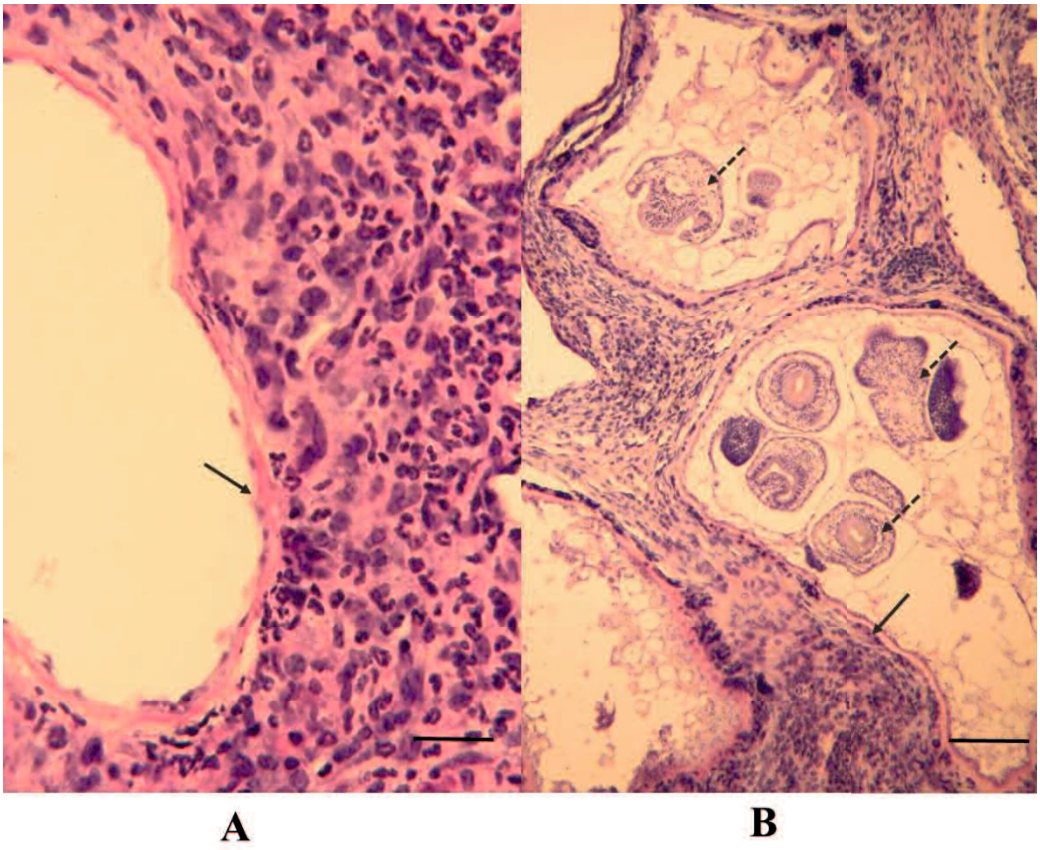


Figure 1-3. Morphology of metacestode samples of *E. multilocularis* for single-end sequencing. **A:** 4-week immature metacestodes (s4Wmet) in the liver of a mouse; **B:** 16-week mature metacestodes (s16Wmet) in the liver of a mouse; Bar: 25μm (A), 100μm (B); Arrow: germinal (nucleated) inner layer; Virtual arrow: protoscolex.

Table 1-1. Overview of the RNA-seq. Result.

Samples	Yield (Mbases)	% PF	Raw Reads	Clean Reads	% of >= Q30 Bases (%PF)	Mean Quality Score (%PF)
pNonc1	13,251	93.34	141,966,744	131,761,968	91.66	35.74
pNonc2	13,870	94.45	146,847,812	136,531,516	93.53	36.39
pAonc	12,735	94.76	134,400,788	126,184,658	93.48	36.24
p4Wmet	12,430	93.77	132,558,666	121,105,597	91.36	35.79
s4Wmet	9,102	91.89	72,304,466	66,440,573	87.75	34.27
s16Wmet	8,174	93.78	74,498,150	69,707,919	93.31	36.13
pCmet	10,051	93.57	107,407,454	98,702,084	90.28	35.38

Note1: pNonc1: Non-activated oncosphere 1 (pair-end sequencing); pNonc2: Non-activated oncosphere2 (pair-end sequencing); pAonc: Activated oncosphere (pair-end sequencing); p4Wmet: 4-week metacestode (*in vivo*, pair-end sequencing); s4Wmet: 4-week metacestode (*in vivo*, single-end sequencing); s16Wmet: 16-week metacestode (*in vivo*, single-end sequencing); pCmet: Metacestode (*in vitro*, pair-end sequencing)

Note2: % Reads Identified (%PF): The total fraction of passing filter reads assigned to an index

Note3: Q10 means 1 in 10bases is mistake; Q20 means 1 in 100 base is mistake; Q30 means 1 in 1000 base is mistake

3.2 *De novo* assembly for 100bp pair-end reads

Illumina sequencing of four *E. multilocularis* developmental stage yielded 100 bp length paired-end clean reads. Reads were then generated contigs with average lengths are 1,848bp (pNonc1), 1,163bp (pNonc2), 1,748bp (pAonc), 772bp (p4Wmet) and 1,330bp (pCmet). The length of N50 of pNonc1 and pAonc were longer and the ratio of completed ORFs predicted by TransDecoder was also higher at pNonc1 and pAonc. Furthermore, the length of N75 of p4Wmet is shortest (Table 1-2). The highest ratio of BLASTN matched contigs with E-value between 0 to 1e-150 (Figure 1-4) against the *E. multilocularis* reference genome (German isolates) in pNonc1 and pAonc indicated that most of the contigs were assembled as full-length sequences in pNonc1 and pAonc. The

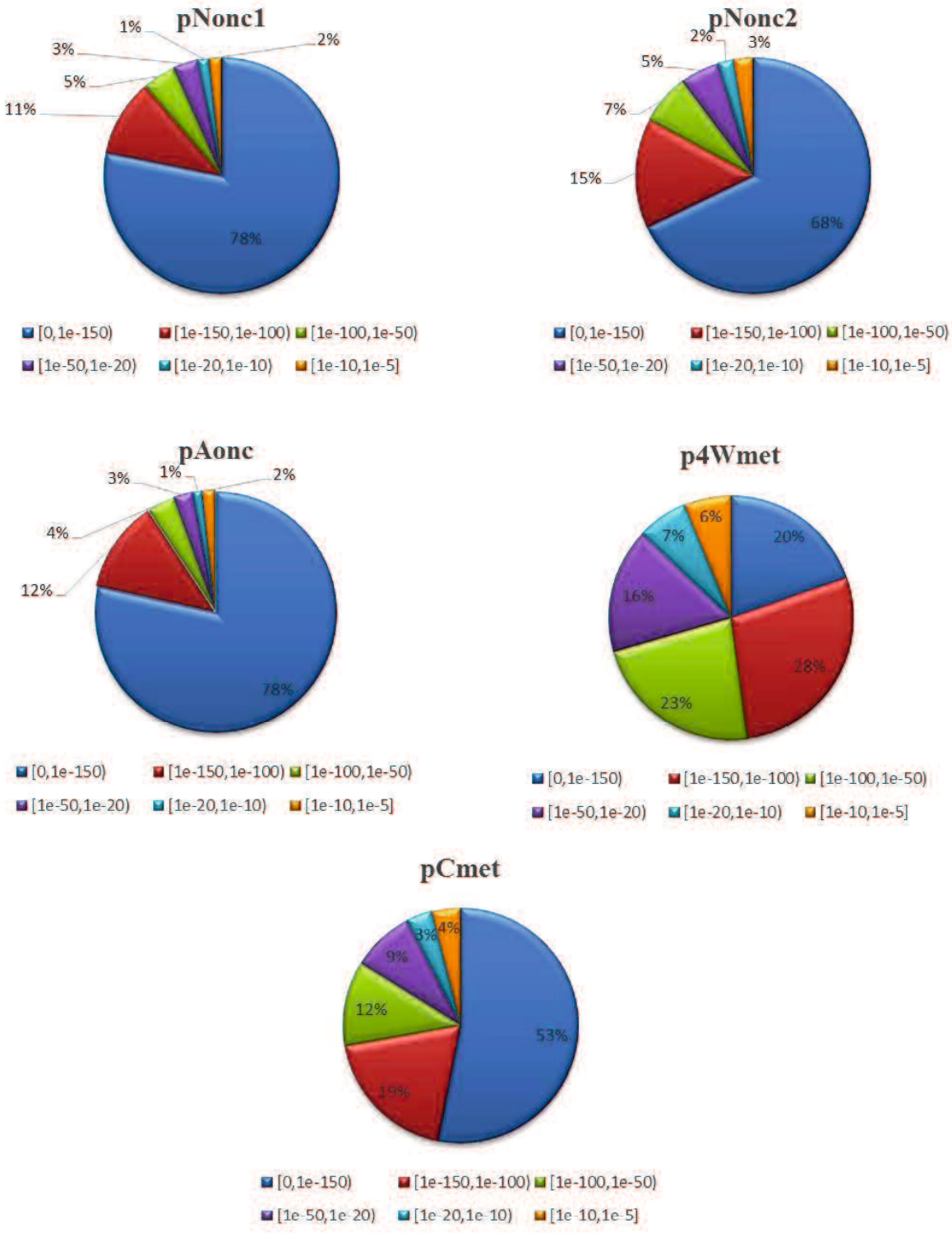
low ratio of BLASTN matched contigs with E-value between 0 to 1e-150 (Figure 1-4) but a little higher ratio of similarity of 100% against the *E. multilocularis* reference genome (German isolates) in p4Wmet means that the matched region length of contigs are short. After removing contaminating sequences, the GC% content of p4Wmet and pCmet changed from 58.72% and 52.50% to 50.44% and 49.04%, respectively, but only about 1% variation was found in non-activated and activated oncospheres (Tables 1-2 and 1-3). However, the GC content of the contaminating sequences filtered Cmet transcriptome was near to reference transcriptome of *E. multilocularis* (Table 1-3). This result indicated that most contigs in this study were correctly assembled.

Table 1-2. Metrics for *E. multilocularis* transcriptome assembly and predicted peptides.

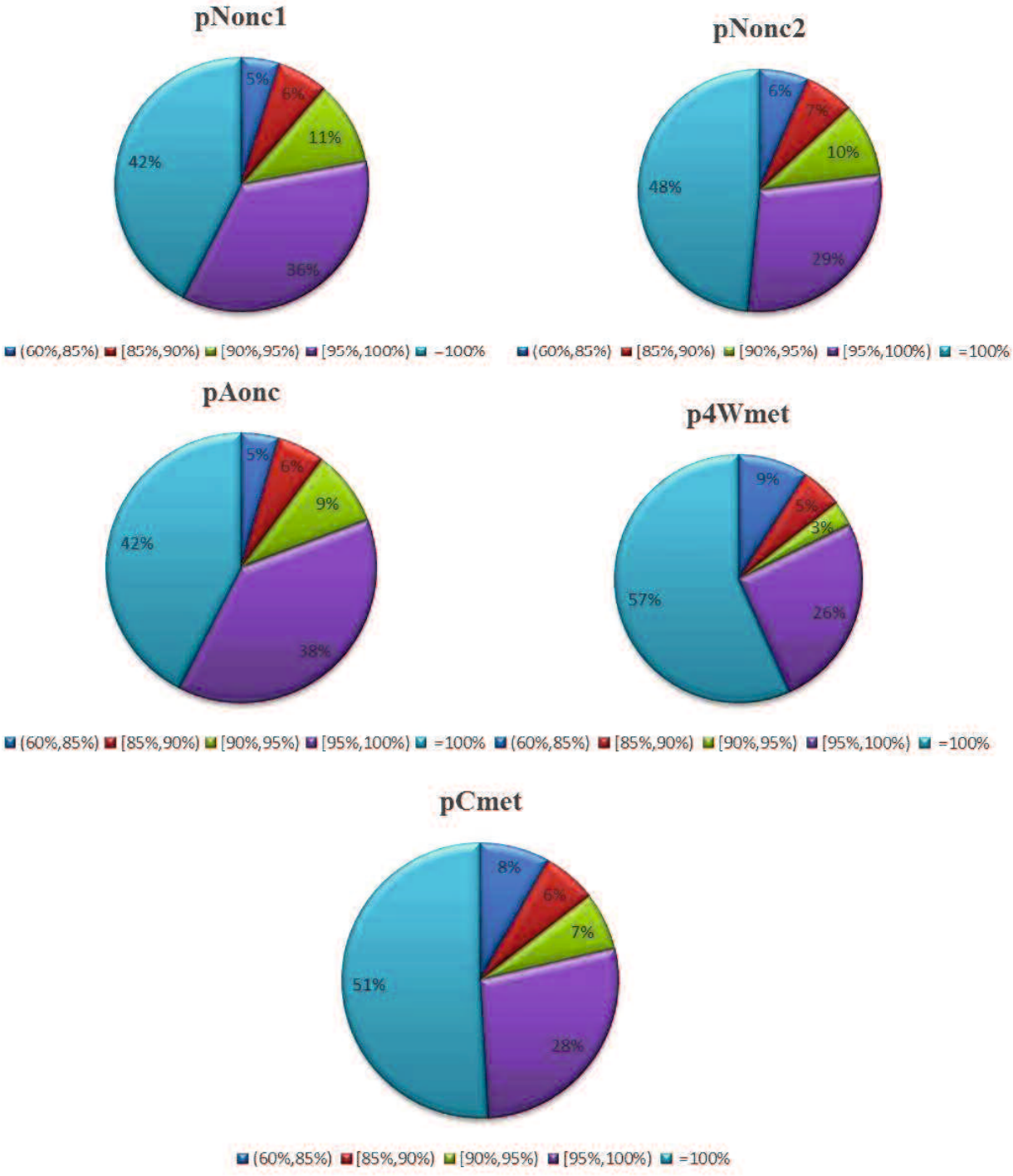
		pNonc1	pNonc2	pAonc	p4Wmet	pCmet
Reads	Raw reads	141,966,744	146,847,812	134,400,788	132,558,666	107,407,454
	Clean reads	131,761,968	136,531,516	126,184,658	231,874	98,702,084
	Phred score >30	91.66	93.53	93.48	91.36	90.28
Assembled contigs	#Contigs	214,410	109,275	182,883	24,949	49,255
	#Components	70,199	52,792	76,479	21,685	28,826
	Maximum length (bp)	17,379	13,450	19,110	11,414	30,022
	Minimum length (bp)	201	201	201	201	201
	Average length (bp)	18,48	1,163	1,748	772	1,330
	Median length (bp)	1,341	759	1,088	474	687
	N25	4,692	3,001	5,106	2,108	4,652
	N50	3,038	1,904	3,189	1,204	2,544
	N75	1,509	1,060	1,840	587	1,263
	GC%	47.24	48.51	46.73	58.72	52.50
Predicted peptides*	Total ORFs	123,154	60,660	103,533	19,428	28,936
	Total LCPC ORFs	25,558	25,927	26,622	17,009	14,882
	complete ORFs	70,605	23,898	59,257	22,88	12,147
	5prime partial ORFs	13,943	9,023	10,557	3,739	6,510
	3prime partial ORFs	25,690	12,055	17,325	3,467	4,305
	internal ORFs	12,916	15,684	16,394	9,934	5,974

*ORF predicted by TransDecoder.

N25, N50, N75: The contig length that using equal or longer contigs produces 25%, 50%, 75% the bases of the genome. The N25, N50 , N75 size is computed by sorting all contigs from largest to smallest and by determining the minimum set of contigs whose sizes total 25%, 50%, 75% of the entire genome.



A. E-value distribution



B. Similarity distribution

Figure1-4. Characteristics of similarity search of *de novo* contigs against *E. multilocularis* reference genome. A: E-value distribution of BLAST hits for each transcript with a cutoff E-value of 10−5; **B:** Similarity distribution of the top BLAST hit for each transcript.

Table 1-3. Summary of *de novo* assembled data of *E. multilocularis* after the contamination filtered

	pNonc1	pNonc2	pAonc	p4Wmet	pCmet	Reference Transcriptome (German isolates)
#Transcripts	192,861	86,068	158,820	9,114	41,150	10,669
#components	50,220	30,550	52,874	8,590	24,382	—
Maximum length (bp)	17,379	13,450	19,110	11,414	30,022	33,585
Minimum length (bp)	201	201	201	201	201	33
Average length (bp)	1,983	1,332	1,959	405	1,127	1,504
Median length (bp)	1,527	996	1,404	291	594	1,071
N25	4,779	3,167	5,167	801	3,682	3,684
N50	3,124	2,065	3,272	415	2,067	2,199
N75	1,924	1,266	1,948	273	1,042	1,275
GC%	46.75	47.19	46.40	50.44	49.04	49.89
Total length	382,396,537	114,639,338	311,205,473	3,691,924	46,372,592	15,326,092

4. Discussion

Most of the *Taenia* immunizing antigens (Johnson et al., 1989; Harrison et al., 1996; Lightowlers et al., 1996a; Lightowlers et al., 1996b; Flisser et al., 2004; Gonzalez et al., 2005; Gauci et al., 2008), which prove to be effective to protect livestock and humans, were all cloned from the infective stage within the parasites' egg (oncospheres). With the parasite genome project implement and the development of the sequence technique, there are several transcriptome datasets available for *Echinococcus* spp. (Parkinson et al., 2012; Tsai et al., 2013; Zheng et al., 2013; Pan et al., 2014;) . And for *E. multilocularis*, transcriptome datasets for mature metacestodes cultivated *in vitro* and immature /mature adult are available, but it is very dangerous to prepare activated oncospheres for this parasite, there haven't public transcriptome datasets for

non-activated oncosphere, activated oncosphere and metacestode that developed post oral infections with oncospheres till four weeks and 16 weeks. In the present study, we would like to analysis transcriptome of *E. multilocularis* from its larval, especially the activated oncospheres using NGS technology. We get more than 700 million clean reads (contain the host sequenced reads) from all the sequenced samples. It is already observed that larval tissue in the liver of 1-3 weeks post oral infections in DBA/2 mice were very small. In the present study, the lesions were identified in the livers and lesions with the parasite (4Wmet) were separated and extracted after four weeks post oral infections of egg of the parasite. The extracted samples contained more host tissue than the parasites. Thus, the low ratio of BLASTN matched contigs with E-value between 0 to 1e-150 against *E. multilocularis* reference genome (German isolates) but high ratio of similarity of 4Wmet were main caused by the contamination of the host RNA and didn't the problem from the quality of reads. And the hypothesis also proved by the significantly decreased when filtering the mouse-mapped reads (Form 131,761,968 clean reads reduced to 231,874 clean reads). Moreover, most of *de novo* assembled contigs could be matched to the reference genome of *E. multilocularis*, which indicated that all sequenced reads of the seven samples and the assembled contigs of pair-end sequencing samples were reliable.

Chapter 2. Different gene expression and function annotation in oncospheres and metacestodes of *E. multilocularis*

Abstract

In order to get the different expressed genes of different life-cycle stages of *E. multilocularis*, the reference genome of *E. multilocularis* (German Isolates) was used as reference to align the sequenced reads. In the present study, the mapped data show that there were 1,300 DEGs in oncospheres versus metacestodes, from which there were752 DEGs are up-regulation when oncospheres transform to metacestodes and 84 DEGs in Aonc versus Nonc. Furthermore, all of these DEGs were up-regulation when Nonc transform to Aonc. In addition, for DEGs in oncospheres versus metacestodes, amyloid beta A4 protein, EG95, some diagnostic antigen GP50, major egg antigen (HSP20) and Tetraspanin 3 (TSP3) were highly expressed in Onc, however, Antigen B subunits (EmAgB8/1, 2,3 and 4), Tetraspanin 5, 6 (TSP5 and TSP6) and tegumental protein were highly expressed in metacestodes. Strikingly, 97% (938/968) of the predicted trans-splicing genes are expressed at the stages of oncospheres and metacestodes, though 20% (2,177/10,669) genes in the reference transcriptome were almost no expression. Moreover, the 769 and 1980 predicted ES and TM proteins of the *E. multilocularis* revealed an enrichment of 'extracellular region' and 'transmembrane transporter activity' at gene ontology level, respectively. The protease analysis showed that there were 257 proteases and 55 proteases inhibitor. And most of proteinases have relatively higher expression levels in 16 Wmet, which indicated these proteinases might

play a more important role in regulating host immune response during the chronic stage of echinococcosis. In contrast, proteases inhibitor, especially Kunitz-type protease inhibitors, were highly expression in oncospheres which suggesting some proteases inhibitor might play an important role to block the proteolytic attack in the host alimentary tract.

This study demonstrated that, the genes expression levels in *E. multilocularis* were change in the transformation and the development. Genes that are highly expressed in non-activated/activated oncospheres, immature/mature metacestode could be explored as novel candidates for diagnostic antigens and vaccine targets.

Key words: Genome, Transcriptome, Oncosphere, Metacestode, Different Expression

1. Introduction

Hosts of *E. multilocularis* produce immune responses to reject and/or limit the growth of the parasite. The parasite can also produce molecules to avoid these immune attacks (Zhang et al., 2008). With immune responses to larval *Echinococcus* spp. infections divided into “establishment” and “established metacestode” phases. The parasite is thought to be more susceptible to immune attack during the “establishment” phase (Siracusano et al., 2011).

DBA/2 mice are thought to be highly susceptibility to AE based on mature protoscolex formation and subsequent active growth of larval parasites in 4 inbred strains of mice (Matsumoto et al., 2010). Differential expression of stage-specific molecules in *in vivo* and *in vitro* 4-week metacestodes has been clearly demonstrated in

this parasite, suggesting that differently expressed molecules may play an important role in the process of *E. multilocularis* infection and modulation of the immune response (Tsai et al., 2013). Moreover, the specific IgG and IgM levels in DBA/2 mice against crude antigens became positive at 4 or 9 weeks post-infection and continued to increase until 16 weeks post-infection (Matsumoto et al., 2010) suggesting that metabolism of the parasite and host responses vary during different growth periods of metacestodes. However, gene expression profile data of metacestodes based on experimental infection through oral ingestion of parasite eggs (termed primary AE) remains lacking.

The *E. multilocularis* reference genome (German Isolates) was sequenced by the Parasite Genomics group at the Wellcome Trust Sanger Institute in collaboration with Klaus Brehm. The initial version of the genome was described in Tsai et al. at 2013, which 9% of the sequence is contained in 9 chromosome scaffolds that have only 23 gaps and one chromosome was complete from telomere to telomere. The gene models have since been subject to iterative improvement. Until 2016/10/26, the total scaffold length was about 115 MB and the longest one was about 20.1 MB and has 10,663 coding genes and 10,669 transcripts. The developing reference genome of *E. multilocularis* makes it possible to predict the gene expression level accurately. In addition, the available software, such as Blast2GO (Götz et al., 2008), InterproScan (Jones et al., 2014) and annotated database, like KEGG (Kanehisa et al., 2015), RefSeq (Pruitt et al., 2007) and Uniprot (Wu et al., 2006), MEROPS (Rawlings et al., 2015) make it easy to prediction the function of DEGs.

2. Materials and Methods

2.1 Mapping and quantification statistics

Clean RNA-Seq reads larger than 50 were mapped to *E. multilocularis* reference genome (January 2016) using TopHat with default parameters (Trapnell et al., 2009). Then, the mapped read number for each gene was counted by htseq-count (Anders et al., 2014) which was integrated with Galaxy software (https://usegalaxy.org/) using the default parameter, and then transformed to counts per million (CPM) and Reads Per Kilobase of exon model per Million mapped reads (RPKM). To validate NGS data, nine genes common to the pNonc1 and pCmet and six antigen candidates from s4Wmet and s16Wmet were selected for real-time PCR analysis, respectively. The primers employed for amplification of 15 genes of *E. multilocularis* and *GAPDH* (EmuJ_000254600, internal control) were designed by OligoArchitect™ (http://www.oligoarchitect.com) and are shown in Table 2-1. The real-time PCR was performed using Applied Biosystems 7300 Real-time PCR System with SYBR-Green detection (SYBR Premix, TaKaRa) according to the manufacturer's instructions. Each reaction was run in triplicate, after which the average threshold cycle (Ct) was calculated per sample and the relative expression of genes was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Table 2-1. Primers for real-time PCR

Description	Gene ID	Forward primers	Reserve primers
Major egg antigen	EmuJ_000212700	CGAAGGGTAATAAGGTGTA	TTGTAGAACTCACGATGT
Na+:K+ ATPase alpha	EmuJ_000342600	CTTCATCCACATTATCACT	CAGTAGTAGCCAAGGATA
Tetraspanin	EmuJ_000355500	CGAAGGTGATGCTGAAGA	TCCGACCACAATGAAGAC
Tegumental protein	EmuJ_000372400	CGAAGTGCTCAAGTCTGA	GCTAGAGTCGGCATTGTA
FABP2	EmuJ_000550000	AACTTCGTAGTCACTGAT	AGTCATCTCCTTGAACTT
Tetraspanin 5	EmuJ_001077100	TTCTTCTTCAATGCCATT	TACCTCCAGACTTGTTAG
Amyloid beta A4 protein	EmuJ_001136900	TTCAATGCTACATCAGGTAAT	CGCCTACATTCCTTCTTAG
GP50	EmuJ_000681200	AGCAACAACCTCTTCTTC	AGTCTTCATAGTATAAGCCAAT
ETS transcription factor	EmuJ_000770300	AACATGAGTGAGGAGAAT	CGTAGAACTTGTAGACATC
EG95	EmuJ_000368620	TTCTCGGATGGACAACTC	CCTCTCACTGCTTCTACA
AgB2	EmuJ_000381100	CTCTTGGCAATGACCTAACT	TAACATACTTCTTCAGCACCTC
AgB1	EmuJ_000381200	AAATGCTTGGCGAAATGA	CCTTAACATCTGGAACACTT
AgB3	EmuJ_000381500	GGTGATGTTGATGAAGTG	TTGGAAGAAGTCCTTGAT
AgB4	EmuJ_000381400	TCTTGTTCTCGTGGCTTT	TCGCATTATGAGGCACTT
MUC-1	EmuJ_000742900	TACTATGCTGAAGAGGAT	GGAGGTGAATAGATGAAG
MUC-2	EmuJ_000408200	TAGACAACCATCCACAACT	ATCGTAGAAGTCGCTGTT
GAPDH	EmuJ_000254600	CTTCCAACTCTGTCAATG	GCTGTCAATAACCAACTT

2.2 Differentially expressed gene analysis

For differential gene expression and related analyses, gene expression is rarely considered at the level of raw counts since libraries sequenced at a greater depth will result in higher counts. Rather, it is common practice to transform raw counts onto a scale that account for such library size differences. Popular transformations include counts per million (CPM), $\log_2$-counts per million (log-CPM), reads per kilobase of transcript per million (RPKM), and fragments per kilobase of transcript per million (FPKM) (Law et al., 2016). edgeR is often used when detect differential expression genes which designed for the analysis of replicated count-based expression data and raw

counts are converted to CPM and log-CPM values using the ‘cpm’ function (Robinson et al., 2010). It is a better choice if there are repeat when use edgeR software to detect the different expressed genes between two group according to the manual of the software (Robinson et al., 2010). But, it is dangerous to prepare activated oncospheres samples, so there was only one sample prepared for activated oncospheres. For getting more reliable result for DEGs detection, I divided the pair-compared groups depend on the biological development stages of the parasite and the read count cluster result (Figure 2-1). Firstly, I compared DEGs between oncospheres and metacestodes. And then, for deeply understand the DEGs, oncospheres were divided into non-activated and activated oncospheres and metacestodes were divided to immature metacestodes *in vivo*, mature metacestodes *in vivo*, and cultivated metacestodes. Non-activated oncospheres and immature metacestodes which have repeat were as standard to establish the dispersion when did differently expression gene analysis. So, DEGs were detected using edgeR software with $p<0.01$ and false discovery rate (FDR) smaller than 0.05 by the following groups:

Oncospheres (Nonc1, Nonc2, Aonc) vs. Metacestodes (4Wmet1, 4Wmet2, Cmet, 16Wmet);

Non-activated oncospheres (pNonc1, pNonc2) vs. Activated oncospheres (pAonc);

4-weeks metacestodes (p4Wmet, s4Wmet) vs. cultivated metacestodes, *in vitro* (pCmet);

4-weeks metacestodes (pWmet, s4Wmet) vs. 16-weeks metacestodes (s16Wmet).

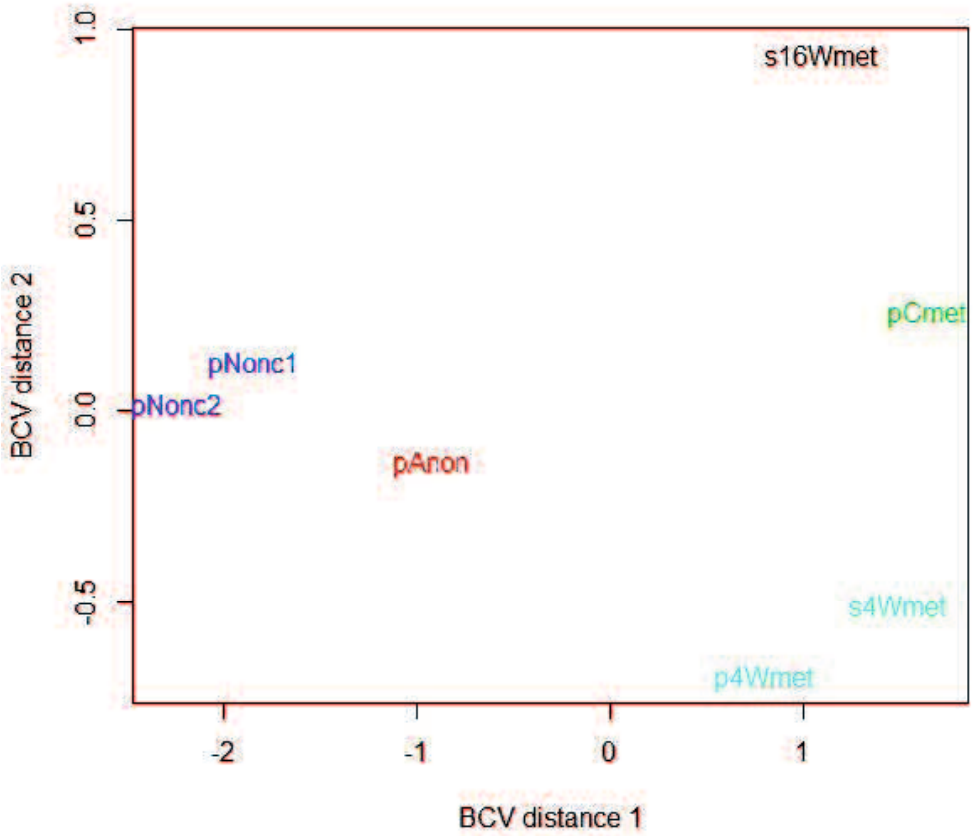


Figure 2-1. CPM value Plot of *E. multilocularis* samples. Samples from the same stage cluster together in the plot, while samples from different stages form separate clusters. This indicates that the gene expression differences of difference stages are larger than those within stage

2.3 *In silico* excretory-secretory (ES) and transmembrane (TM) proteins prediction

Because experimental identification of ES and TM proteins is time-consuming and expensive, the prediction of ES and TM proteins from sequenced genomes is a novel alternative strategy used to priorities the experimental study of new therapeutic and immunodiagnostic targets for human parasitic diseases, even though *in silico* prediction result are influenced by parameter setting, prediction model, et al., and the false positive prediction for amino sequences may exists, especially for the prediction of subcellular localization of proteins (Wang et al., 2015a). 10,669 amino sequences predicted from

the reference transcriptome of *E. multilocularis* were downloaded from WormBase ParaSite (http://parasite.wormbase.org/Echinococcus_multilocularis_prjeb122/Info/Index/) (November 2, 2016). *In silico* prediction of ES proteins and TM proteins were carried out according to the protocol described previously (Garg and Ranganathan, 2011). Briefly, the ES proteins homologues were utilizing the following four tools: SignalP (version 4.1) (Petersen et al., 2011) for classical secreted proteins; SecretomeP (version 2.0) (Bendtsen et al., 2004) for non-classical proteins; TMHMM (version 2.0) (Sonnhammer et al., 1998) for trimming transmembrane proteins; TargetP (version 1.1) (Emanuelsson et al., 2007) for trimming mitochondrial proteins. The proteins predicted to contain only one TM domain, further TM prediction was performed by the Phobius algorithm (Käll et al., 2007) to help discriminate hydrophobic helices of TM topologies from those of signal peptides, in which only the proteins confirmed by Phobius were considered as TM proteins. The predicted proteins with no transmembrane helices were thought to be ES proteins.

2.4 Protease analysis

Putative homologues of known proteases of the 10,669 amino acid sequences in *E. multilocularis* reference transcriptome were identified using the complete set of core protease sequences from the MEROPS (release 10.0) database (Rawlings et al., 2015). They consist of a non-redundant library of the catalytic unit of a protease and exclude all other functional units, such as domains of Ca^{2+}-binding and ATP-binding. These core sequences were used to avoid false positive identification of proteases due to high sequence identity in its non-catalytic parts. Core sequences were compared to the

10,669 amino acid sequences in *E. multilocularis* reference transcriptome. The MEROPS batch BLAST (Rawlings and Morton, 2008) comparisons were carried out using the 10,669 proteins as the queries, and the MEROPS peptidases as the subject database.

2.5 Spliced-leader and trans-splicing analysis

The *de novo* assembled *E. multilocularis* contigs of each stages were alignment with the spliced leader sequences (Brehm et al., 2000a; Tsai et al., 2013) using BLAST (Altschul et al., 1997) (parameters word-size: 12, E-value: 1E-10). The identified contigs were filtered according to BLAST alignment features: an alignment length of at least 12 nucleotides, one or no mismatches between query and subject and presence of the 3' terminal SL sequence ATG. If the contig was reverse-complementary to the SL sequence, the sequence was reverse complemented and the corresponding quality entry string reversed. Then the 10,669 transcripts sequences of *E. multilocularis* reference transcriptome would BLASTN (Altschul et al., 1997) to contigs which contain spliced-leader and the result showed E-value smaller then 1E-25, and identity binger than 95% would retained. And this sequences from the reference transcriptome were assigned to spliced-leader contain transcripts.

2.6 Functional annotations

Sequences of *E. multilocularis* reference transcriptome were used as queries against the National Center for Biotechnology Information non-redundant database *Platyhelminthes* Section (*Taxonomy* ID: 6157) using BLASTX (Altschul et al., 1997) with an e-value threshold of $1e^{-5}$. The BLASTX output, generated in xml format, was

used for Blast2GO analysis to annotate the transcripts with Gene Ontology (GO) terms describing biological processes, molecular functions, and cellular components (Götz et al., 2008). The e-value filter for GO annotation was $1e^{-8}$. Proper GO terms were generated using Blast2GO mapping process (Götz et al., 2008). And then, GOslim, which is integrated in the Blast2GO software, was used to slim the annotation. A sequence description was also generated from Blast2GO, based upon NR database *Platyhelminthes* Section (*Taxonomy ID*: 6157) according to e-value and identity to BLASTed genes. KOBAS (KEGG Orthology Based Annotation System, v2.0) was used to identify biochemical pathways and genes coding amino sequences were compared to amino sequences of *Schistosoma mansoni* which was the only available flatworm database in KOBAS software (Xie et al., 2011). The Interpro function domain annotations were extracted directly from the GFF3 file of the parasite that deposited at WormBase ParaSite.

2.7 Gene Oncology (GO) term enrichment analysis

GO enrichment analysis of the stages significant highly expressed genes were performed using the Fisher's Exact Test function in Blast2GO (Götz et al., 2008) with the FDR cut-off value =0.05.

3. Result

3.1 Mapping reads to the *E. multilocularis* genome

In the present study, seven RNA-Seq libraries were constructed, of which five for pair-end sequencing and two for single-end sequencing. As for pair-end sequencing reads, there are 78.4%, 80.2%, 82.7%, 0.4% and 68.9% reads were mapped to the *E.*

multilocularis reference genome (German isolates) from pNonc1, pNonc2, pAonc, p4Wmet and pCmet, respectively. In addition, there are 1.4% and 9.0% single-end sequenced reads were mapped the same genome from s4Wmet and s16Wmet, respectively (Table 2-2). In the following analysis, the reads with alignment quality less than -10 were counted by htseq-count software. In order to validate the expression profiles, 9 (Figure 2-2A) and 6 (Figure 2-2AB) genes of *E. multilocularis* for pair-end and single-end sequencing were selected for quantitative RT-PCR analysis using the same RNA samples as for RNA-Seq. The real-time PCR results confirmed the result obtained from deep sequencing analysis and showed similar trends of up- or down-regulated genes.

Table 2-2. Summary of alignment statistics in different life-cycle stages.

Samples	Aligned pairs (Pair-end)	Mapped reads (Single-end)	Alignment rate
pNonc1	971,213	—	78.4%
pNonc2	1,032,382	—	80.2%
pAonc	932,949	—	82.7%
p4Wmet	265,767	—	0.4%
pCmet	603,046	—	68.9%
s4Wmet	—	926,796	1.4%
s16Wmet	—	6,276,152	9.0%

Note: The statistics data of each sample was origin from Tophat align summary; "—": No data.

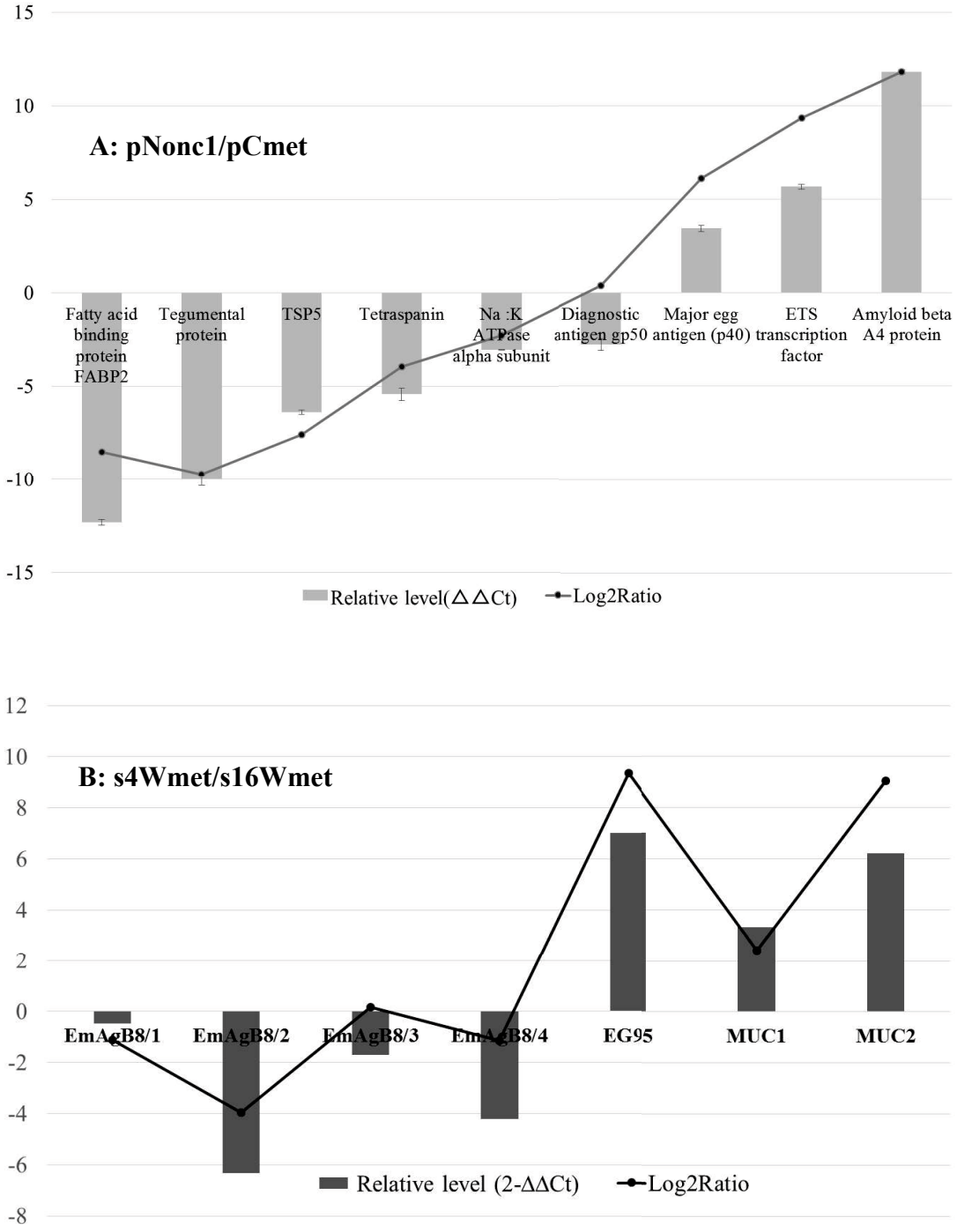


Figure 2-2. Correlation of fold-changes between RNA-seq. and real time PCR. The y-axis indicates the value of relative expression level ($2^{-\Delta\Delta Ct}$) by real-time PCR and log2Ratio of pNonc1/pCmet (**A**, pair-end) and s4Wmet/s16Wmet (**B**, single-end) by Next-generation sequencing. A: pair-end; B: single-end; GAPDH as the internal control.

3.2 Differentially expressed gene analysis

For DEGs analysis, there were 1,300 DEGs in oncospheres versus metacestodes, from which 752 were up-regulation when oncospheres transform to metacestodes (Figure 2-3). In addition, there were 84 DEGs in Aonc versus Nonc and all of these DEGs were up-regulation when Nonc transform to Aonc (Figure 2-3). Moreover, there were 82 DEGs in 4Wmet versus Cmet, of which 34 DEGs were up-regulation in 4Wmet (Figure 2-3). At last, there were 194 DEGs in 4Wmet versus 16Wmet (Figure 2-3), and 135 DEGs up-regulation in 16Wmet (Figure 2-3). As for those DEGs, most genes (1,169) were significant high expressed when oncospheres versus metacestodes (Figure 2-4), and one gene was detected to be DEGs among compared stages, and there were less DEGs when compared 4Wmet to Cmet than compared to16Wmet (Figure 2-4). For DEGs in oncospheres versus metacestodes, amyloid beta A4 protein (EmuJ_001136900.1), EG95 (EmuJ_000710400.1), some diagnostic antigen GP50 (EmuJ_000295100.1, EmuJ_000032300.1, EmuJ_000261100.1), major egg antigen (EmuJ_000212700.1), Tetraspanin 3 (EmuJ_001077400.1) were significant highly expressed in oncospheres, however, Antigen B subunits (EmAgB8/1, 2, 3 and 4), Tetraspanin 5, 6 (EmuJ_001077100.1, EmuJ_001021300.1) and tegumental protein (EmuJ_001001400.1) were significant highly expressed in metacestodes (Figure 2-5). As for Nonc versus Aonc, EG95 (EmuJ_000368620.1) and GP50 (EmuJ_000261100.1) and purine nucleoside phosphorylase (EmuJ_000635800.1) were highly expressed in Aonc (Figure 2-6), for 4Wmet versus Cmet, it was shown that Actin cytoplasmic A3 (EmuJ_000406900.1, EmuJ_000407200.1) were significant highly expressed in Cmet, however glioma pathogenesis protein 1 (EmuJ_000290500.1) and LRRP1 (EmuJ_002194800.1) are highly expressed in 4Wmet (Figure 2-7). And collagen alpha

iv chain (EmuJ_000140000.1) and EG19 antigen (EmuJ_000342900.1) were highly expressed in 16Wmet, however GP50 (EmuJ_001120900.1) was highly expressed 4Wmet (Figure 2-8).

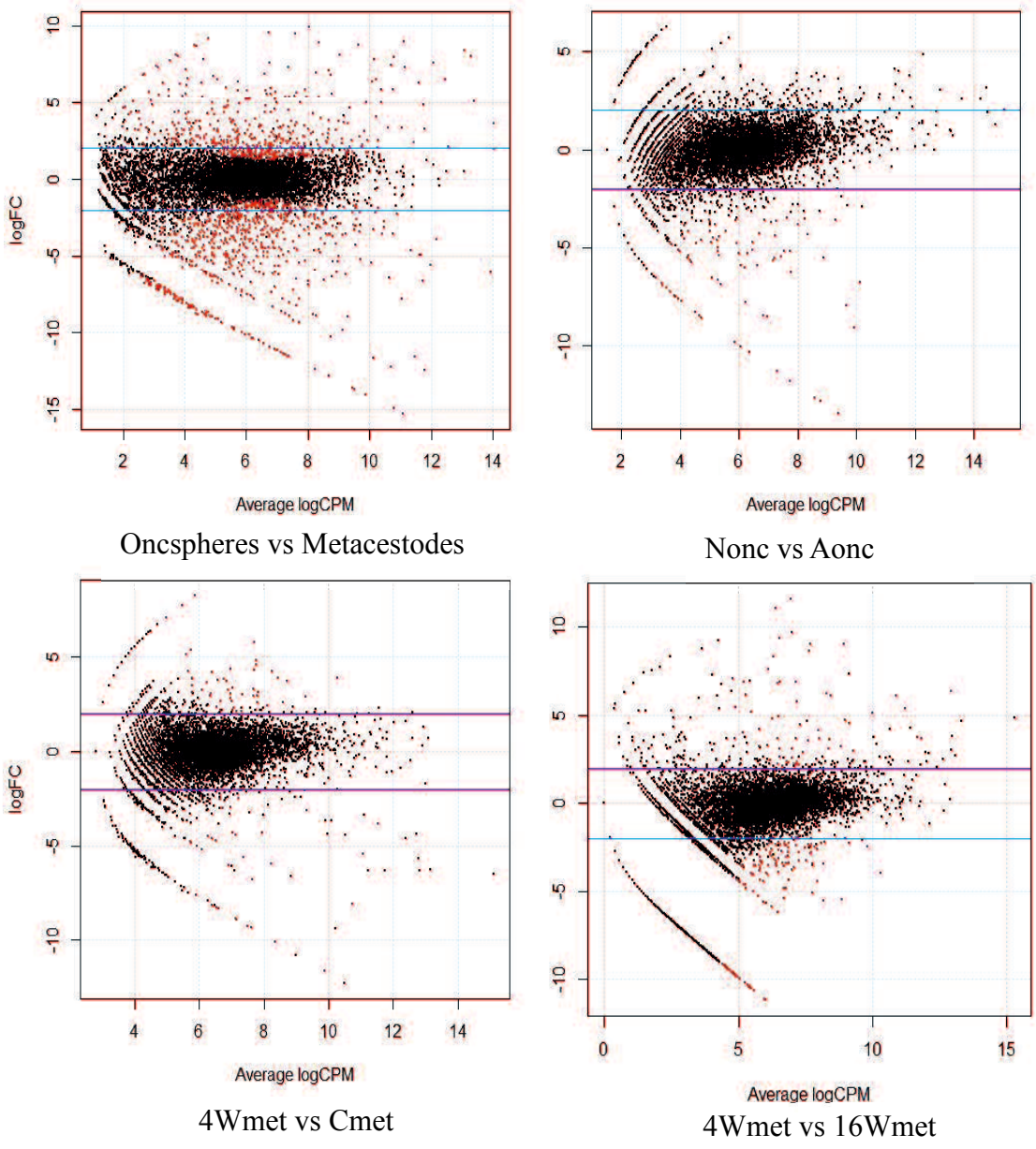


Figure 2-3. Analyses of differentially expressed genes (DEGs) among Nonc, Aonc, 4Wmet, Cmet and 16Wmet. Note： Red points means significant differently expression genes.

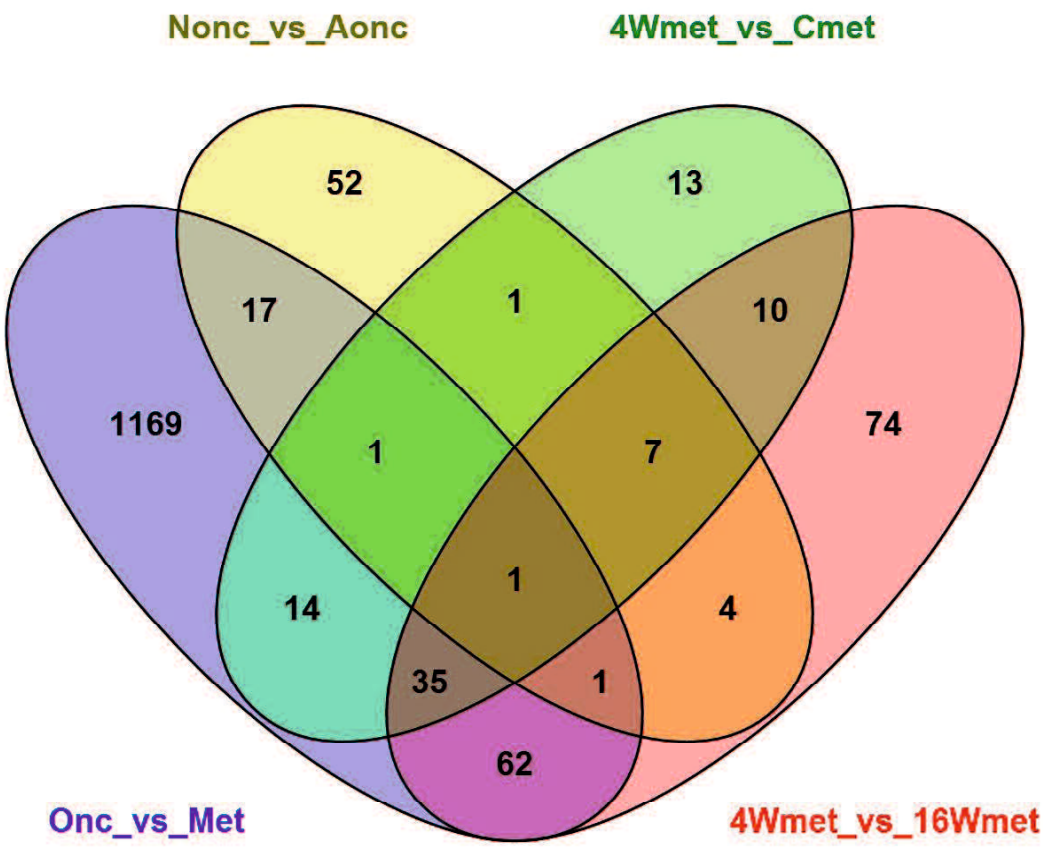


Figure 2-4. Transcriptome analysis of different expression genes in different stages.

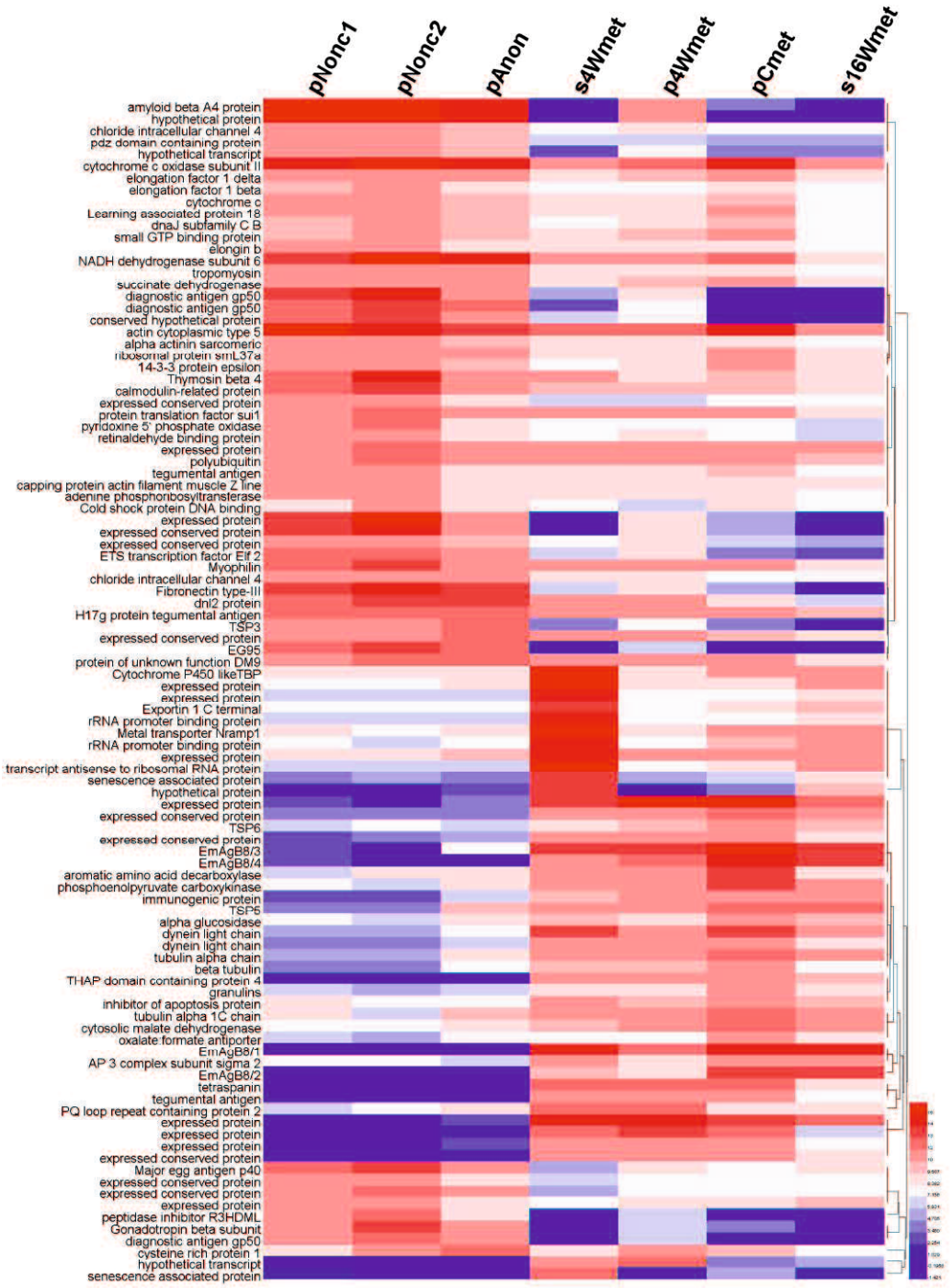


Figure 2-5. Heatmap of log-RPKM values for top 100 DEGs in oncospheres versus metacestodes. Expressions across each gene (or row) have been scaled so that mean expression is zero and standard deviation is one. Samples with relatively high expression within a gene are marked in red, samples with relatively low expression are marked in blue. Lighter shades and white represent genes with intermediate expression levels. Samples and genes have been reordered by the method of hierarchical clustering. A dendrogram is shown for the clustering of samples. **Note:** DEGs are arranged by the average log-RPKM value among all sequenced samples.

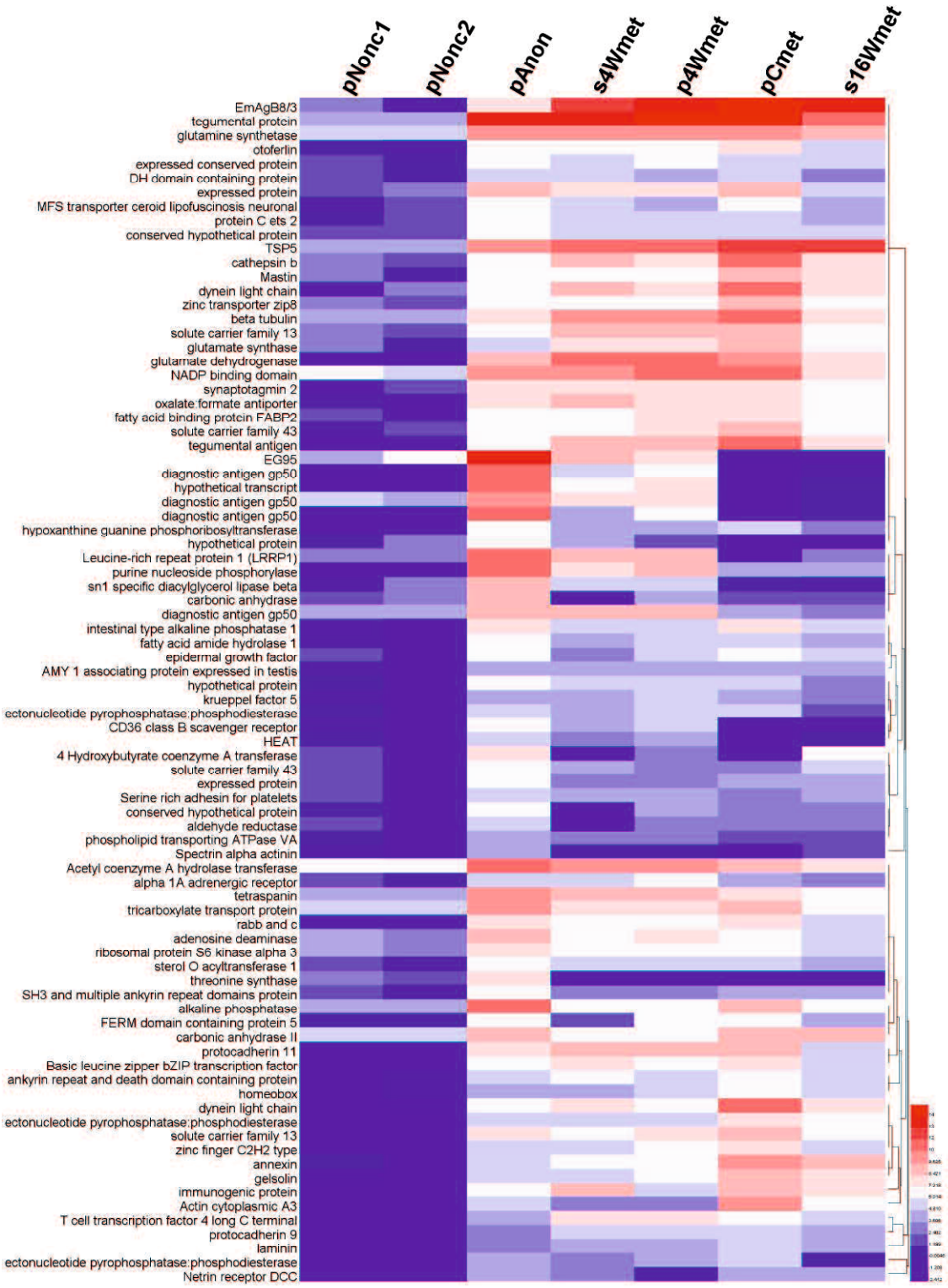


Figure 2-6. Heatmap of log-RPKM values for all DEGs in Nonc versus Aonc.

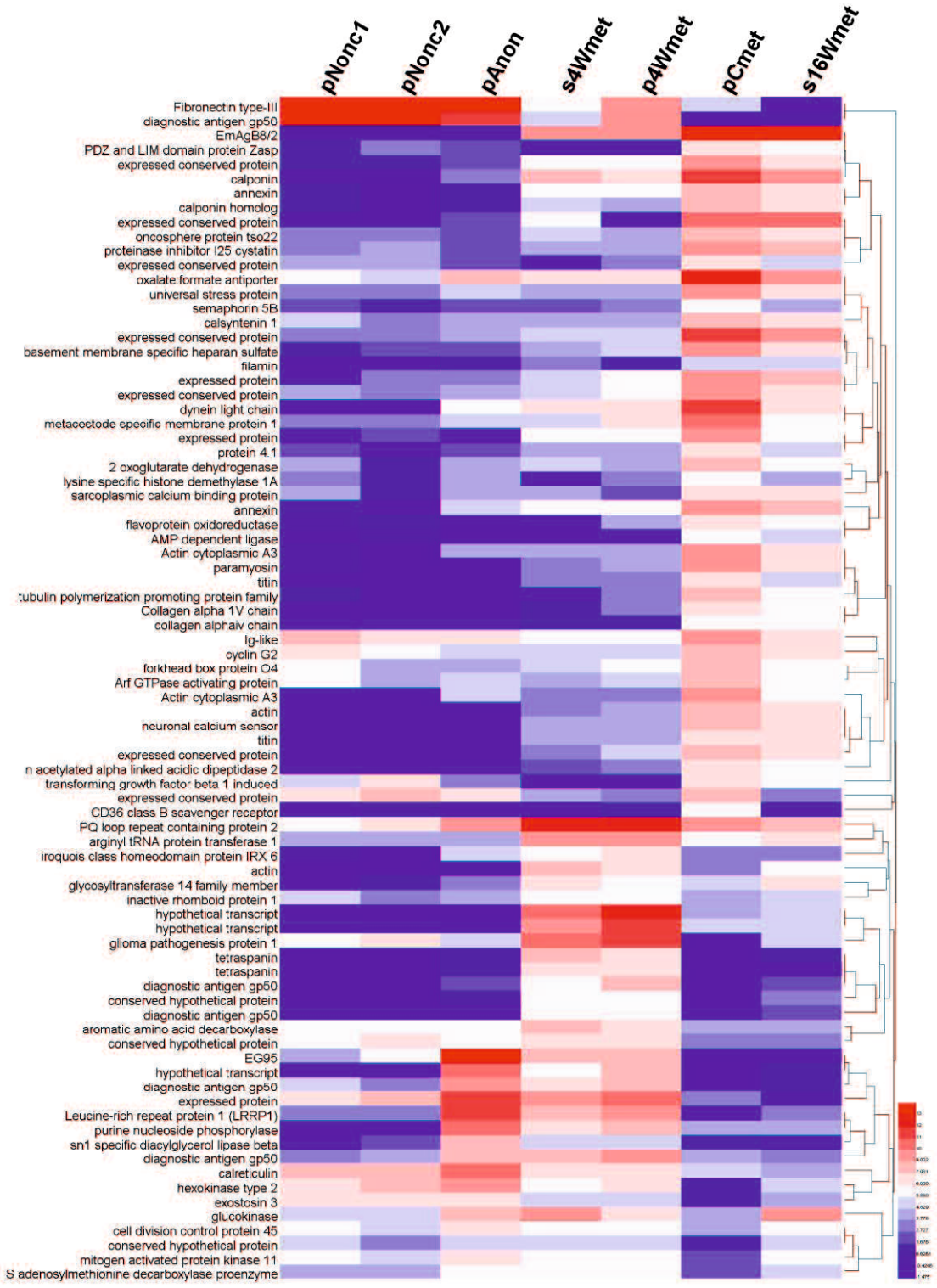


Figure 2-7. Heatmap of log-RPKM values for all DEGs in 4Wmet versus Cmet.

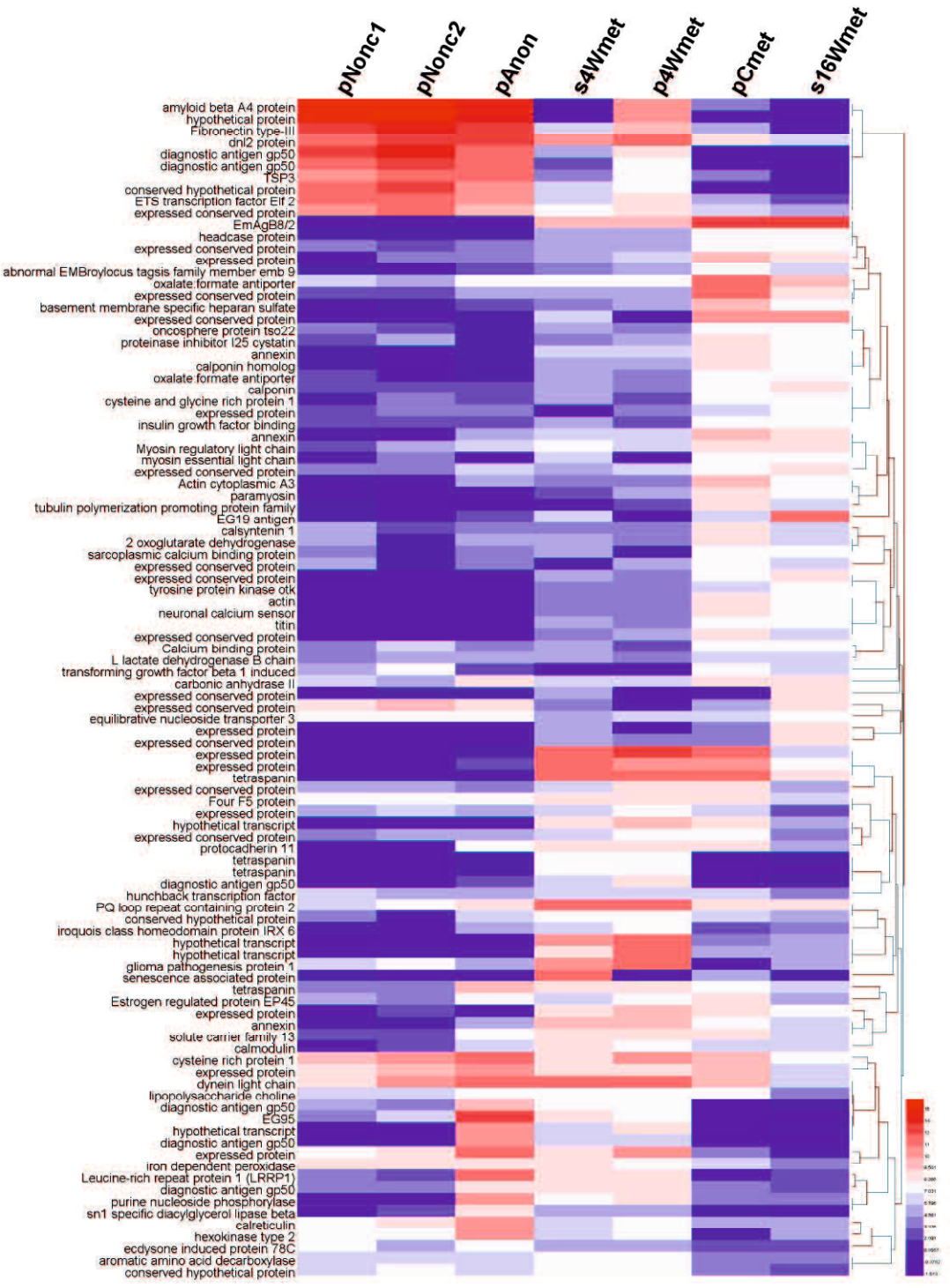


Figure 2-8. Heatmap of log-FPKM values for all DEGs in 4Wmet versus 16Wmet.

3.3 Gene Oncology (GO) term enrichment analysis

The GO term analysis show that the predominant terms in the reference transcriptome at level 2 for ‘molecular function’ were ‘binding’, for ‘biological process’ were ‘metabolic process’, for ‘cellular component’ were ‘cell’, respectively) (Figure2-9, 2-10, 2-11). Gene Ontology (GO) term enrichment analysis shown that a significant increase was observed in GO-terms associated with transmembrane transport (FDR=4.8E-3) of 84 up-regulated DEGs in Aonc when compared with Nonc (Figure 2-12) and the 752 up-regulated DEGs in metacestodes suggested the up-regulation of proteinaceous extracellular matrix (FDR=6.6E-4), plasma membrane (FDR=6.6E-4) and cell adhesion (FDR=4.2E-7) (Figure 2-13). In addition, the 135 up-regulated DEGs in 16Wmet suggested the up-regulation of proteinaceous extracellular matrix (FDR=2.6E-7) when compared with 4Wmet (Figure 2-14). However, there weren’t enrichment GO terms in 4Wmet compare to Cmet.

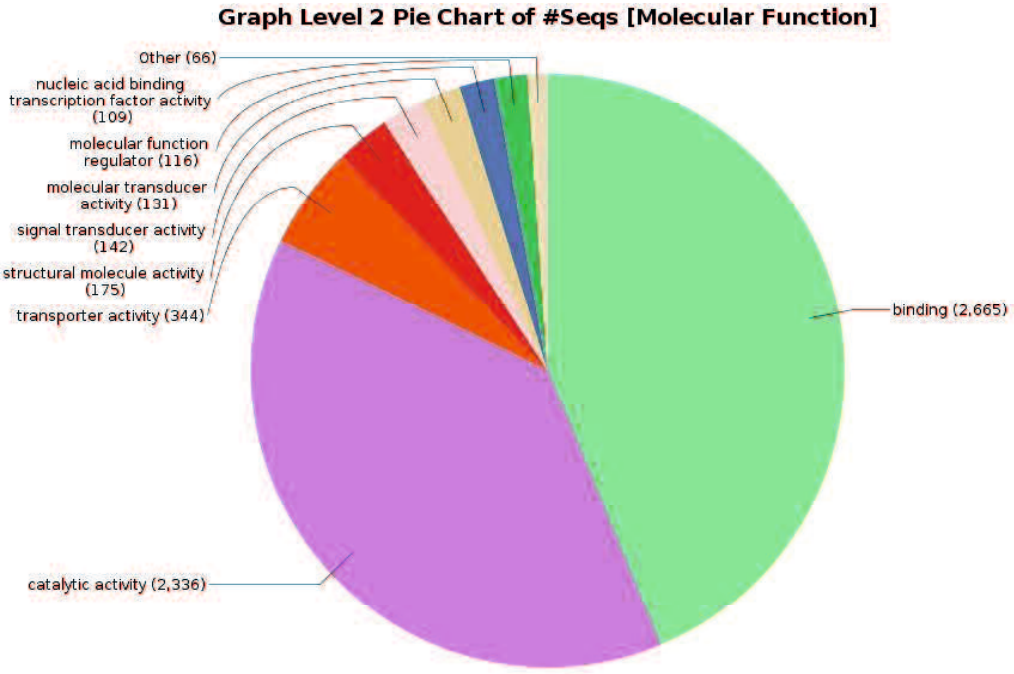


Figure 2-9. Pie charts level 2 GO distribution of annotated reference transcriptome in molecular function.

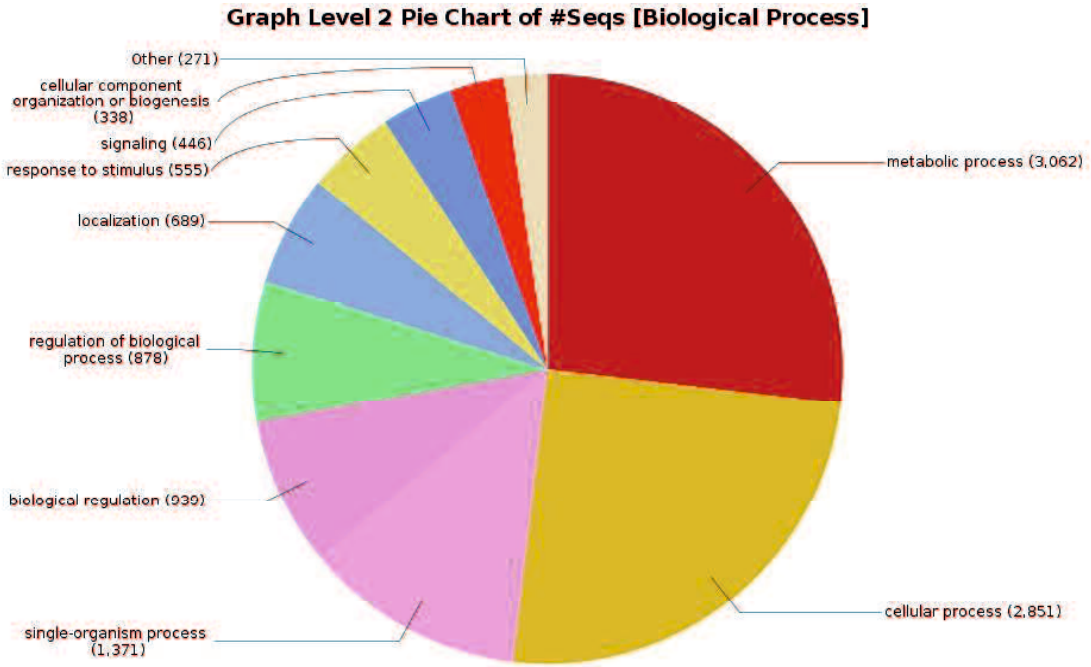


Figure 2-10. Pie charts for level 2 GO distribution of annotated reference transcriptome in biological process.

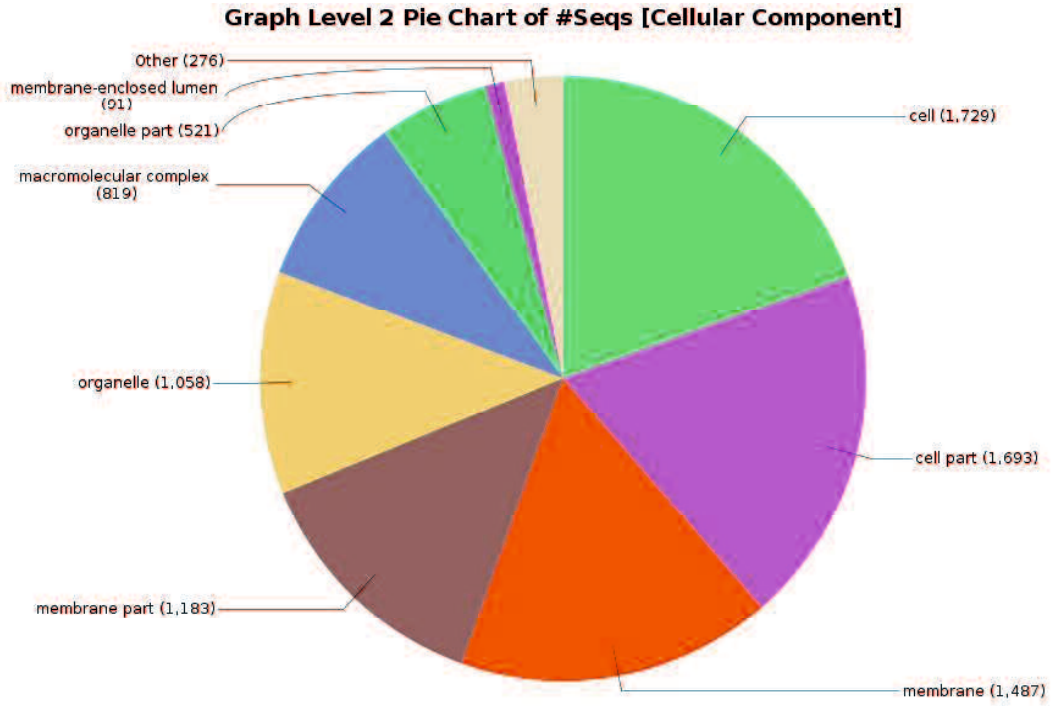


Figure 2-11. Pie charts for level 2 GO distribution of annotated reference transcriptome in cellular component.

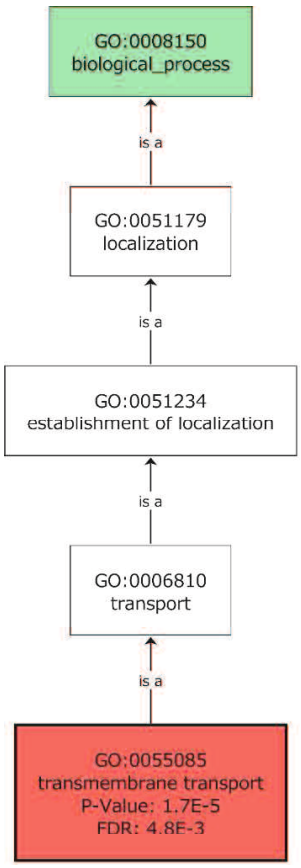


Figure 2-12. The GO enrichment of 84 significant highly expression genes (FDR<0.05) in Anoc when compared with Nonc.

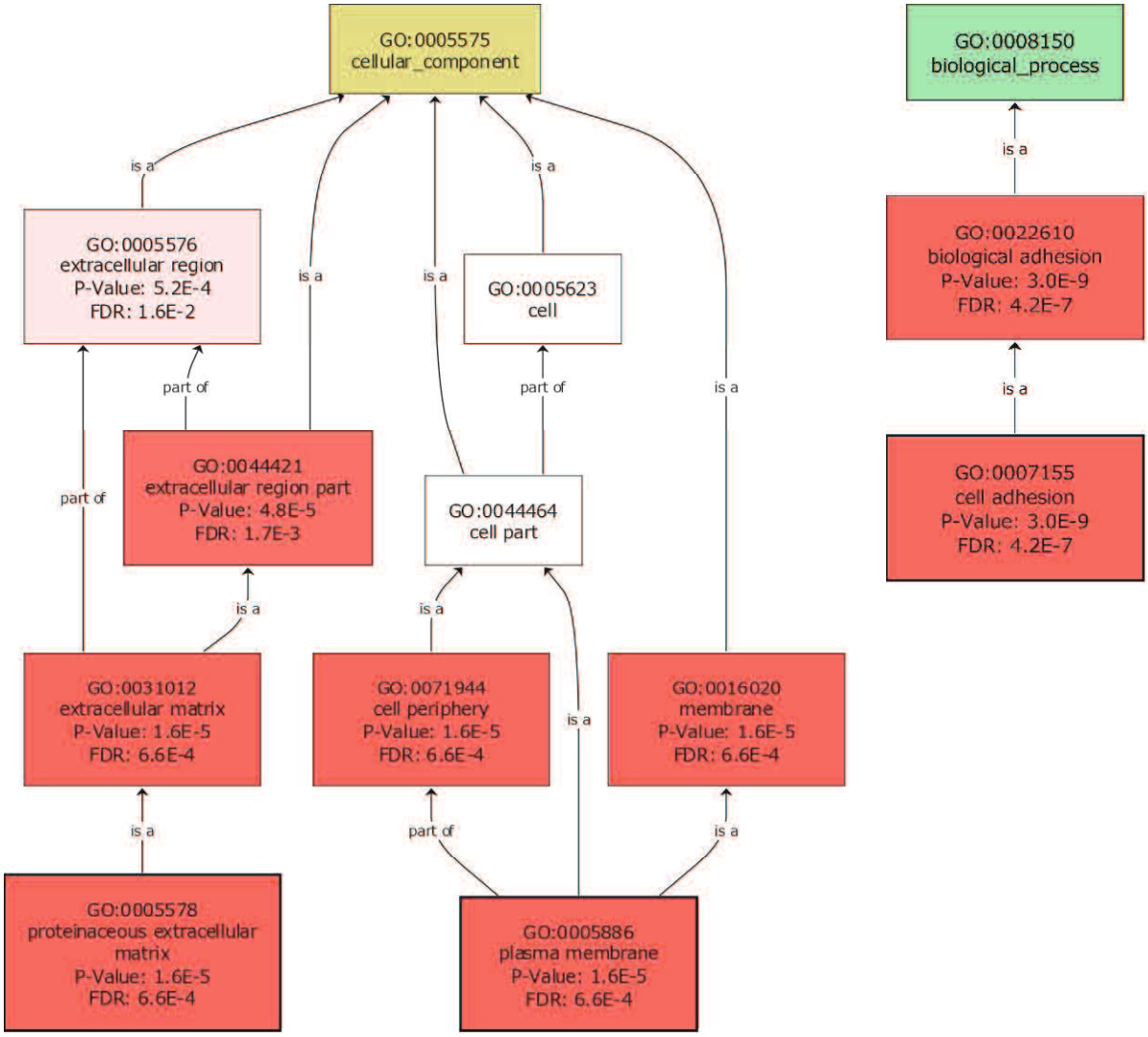


Figure 2-13. The GO enrichment of 752 significant highly expression genes (FDR<0.05) in metacestodes when compared with oncospheres.

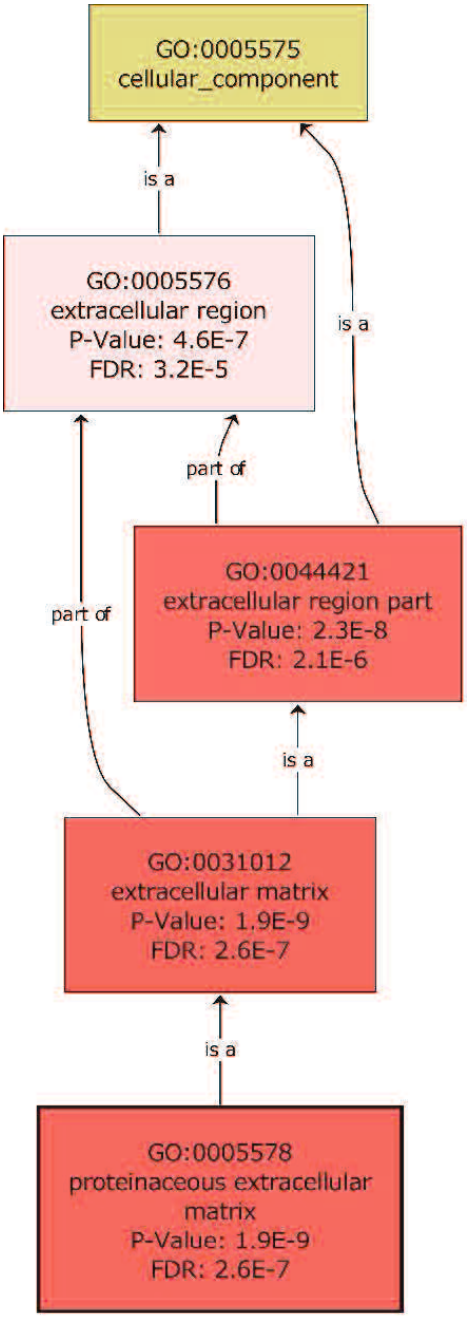


Figure 2-14. The GO enrichment of 135 significant highly expression genes (FDR<0.05) in 16Wmet when compared with 4Wmet.

3.4 Predicted *E. multilocularis* secretome and transmembranome size

There are 10,669 putative protein sequences of reference genome of *E. multilocularis*, a number of 853 sequences (8.0%) were predicted to contain a signal peptide cleavage

site by SignalP. Of the remaining sequences, SecretomeP classified 314 protein sequences (2.94%) as non-classical secreted proteins. The putative 1,167 secreted proteins were parsed to TargetP and TMHMM, Phobius, even though the result show that there are 15 protein sequences (7.8%) predicted to be of mitochondrial origin and 383 protein sequences (21.7%) predicted to contain transmembrane helices, there may contain mitochondrial and transmembrane proteins in the predicted secretome of *E. multilocularis* in the present study for the sensitive of the predict software. Potential mitochondrial and transmembrane proteins were excluded from the data set resulting in 769 *E. multilocularis* ES protein sequences, representing 7.21% of the putative protein dataset. TM protein resulted in 1,980 (18.56%) sequences out of the 10,669 *E. multilocularis* putative protein sequences and are similar to previous study (Wang et al., 2015a). The expression data of top 100 ES and TM proteins (Arranged by average log-RPKM value) are shown in Figure 2-15 and Figure 2-16. It is show that amyloid beta A4 protein (EmuJ_001136900.1), hypothetical protein (EmuJ_001142400.1) are highly expressed in oncospheres, however, MUC-1 (EmuJ_000742900.1), EmAgB8/1, 8/2, 8/3 (EmuJ_000381200.1, EmuJ_000381100, EmuJ_000381500.1) are highly expressed in metacestodes for the ES protein (Figure 2-15). As for predicted TM proteins, it is show that express protein (EmuJ_000312200.1) and tetraspanin (EmuJ_000355900.1) are highly expressed in metacestodes (Figure 2-16), however, peptidase inhibitor R3HDML (EmuJ_000651500.1) and conserved hypothetical protein (EmuJ_000072600.1) are highly expressed in oncospheres (Figure 2-16).

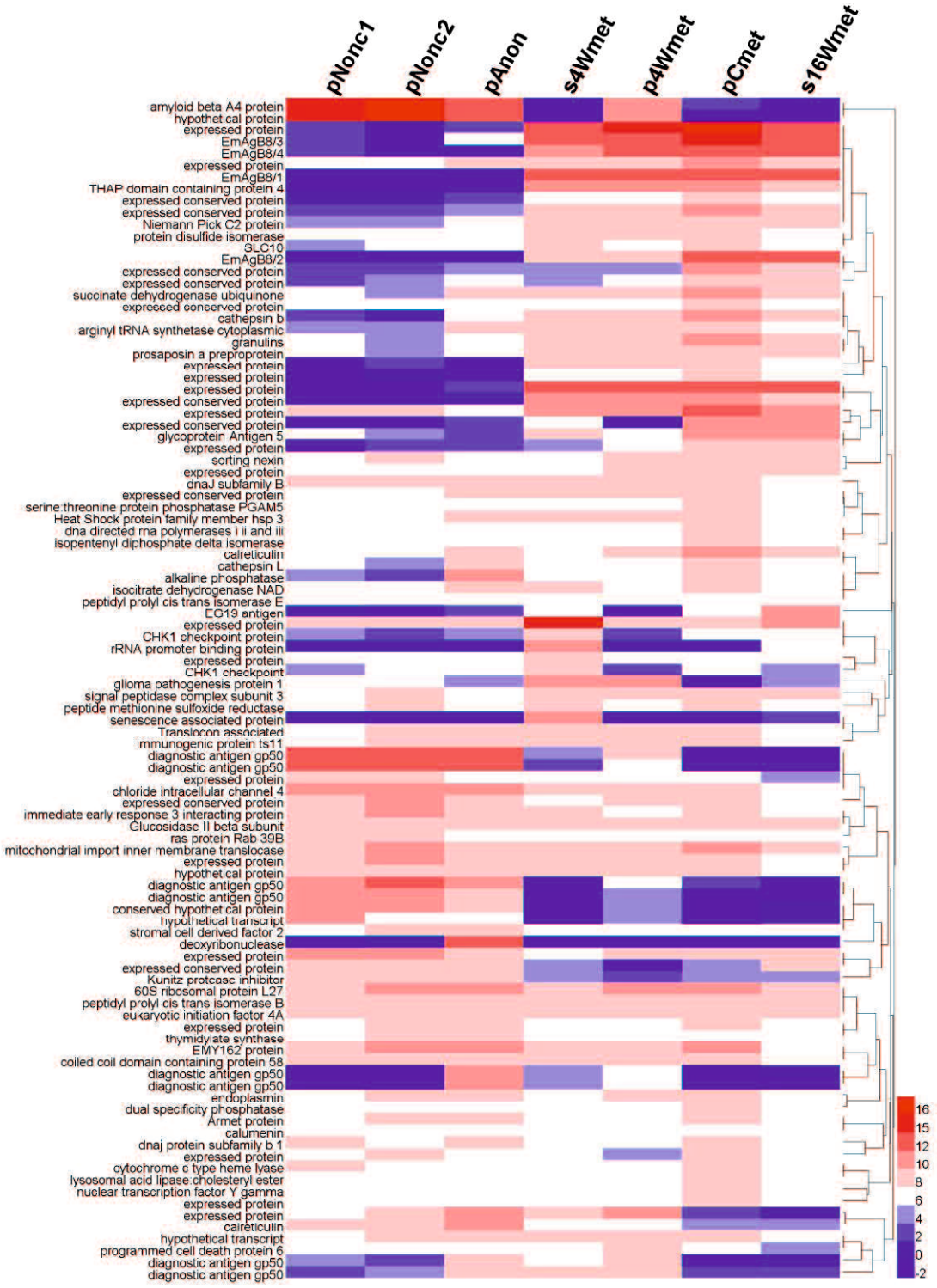


Figure 2-15. Heatmap of log-RPKM values for top 100 ES proteins. **Note:** ES proteins are arranged by the average log-RPKM value among all sequenced samples.

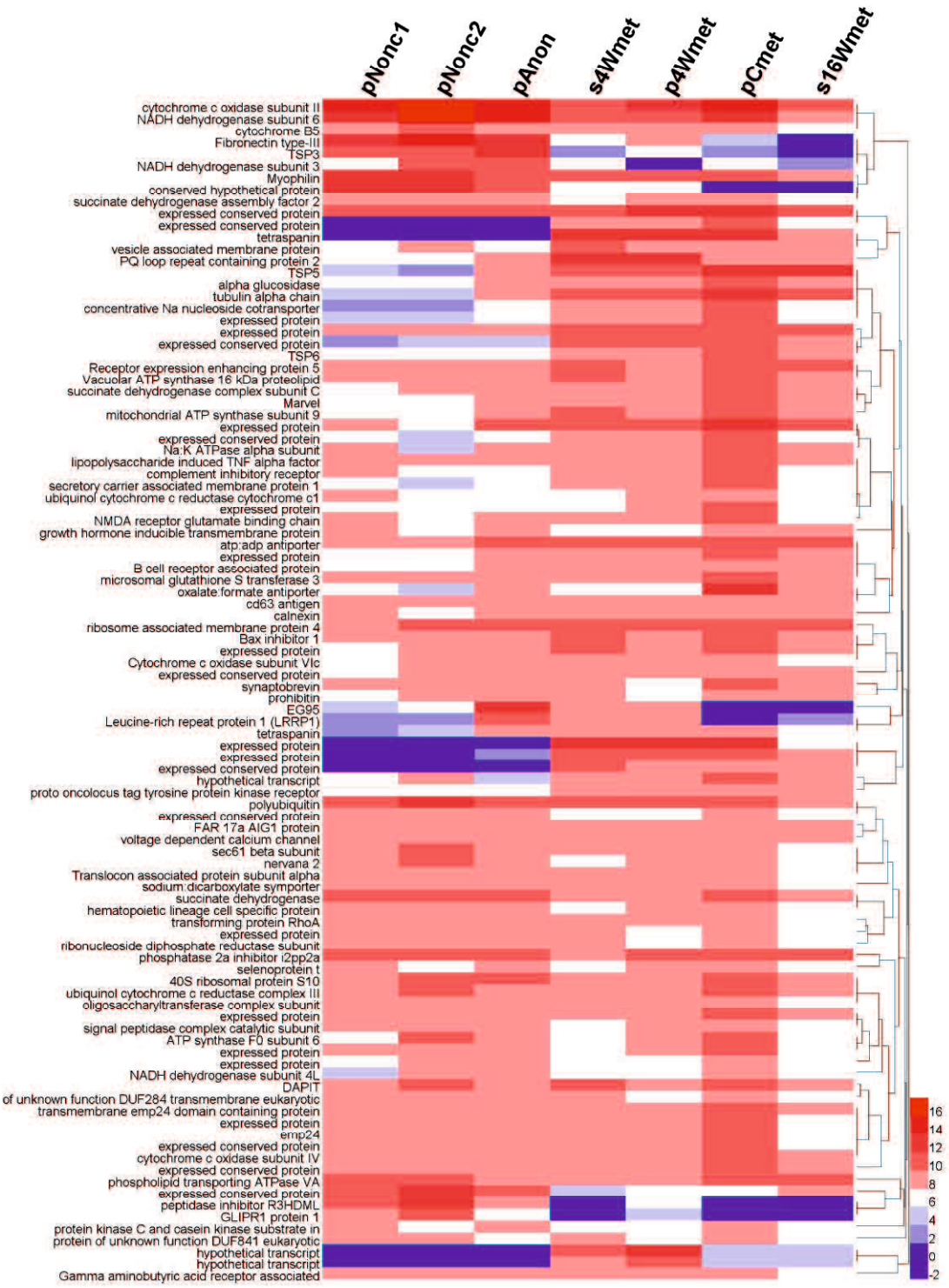


Figure 2-16. Heatmap of log-RPKM values for top 100 TM proteins. **Note:** TM proteins are arranged by the average log-RPKM value among all sequenced samples.

3.5 Functional annotation of *E. multilocularis* ES and TM proteins of the reference transcriptome

Functional annotations of *E. multilocularis* ES and TM proteins were based on GO terms, KEGG pathway and InterPro annotation. The enrichment analysis shown that a significant increase was observed in the GO-terms associated with ‘enzyme regulator activity’ (FDR=8.3E), ‘peptidase activity’ (FDR=3.3E-10) of ES proteins when compared with the reference transcriptome (Figure 2-17). And for TM protein, the GO-terms associated with ‘transmembrane transporter activity’ (FDR=8.36E-102), ‘transporter activity’ (FDR=1.43E-96) and ‘signal transducer activity’ (FDR=3.85E-32) were significant increased as well. (Figure 2-18).

InterPro annotation of predicted ES protein sequences resulted in 537 different assigned protein domains and families. The most represent domains and high ratio of ES proteins between secretome and transcriptome were ‘Cysteine peptidase’, ‘Proteinase’, ‘Pancreatic trypsin inhibitor Kunitz domain’ and ‘Taeniidae antigen’ (Table 2-3). As for TM proteins, ‘Major facilitator superfamily’, ‘G protein coupled receptor, rhodopsin-like’, ‘Ion transport domain’, ‘Tetraspanin/Peripherin’ and ‘Cadherin’ were most frequency occurring domain and shown high ratio between transmembranome and transcriptome at the same time (Table 2-4).

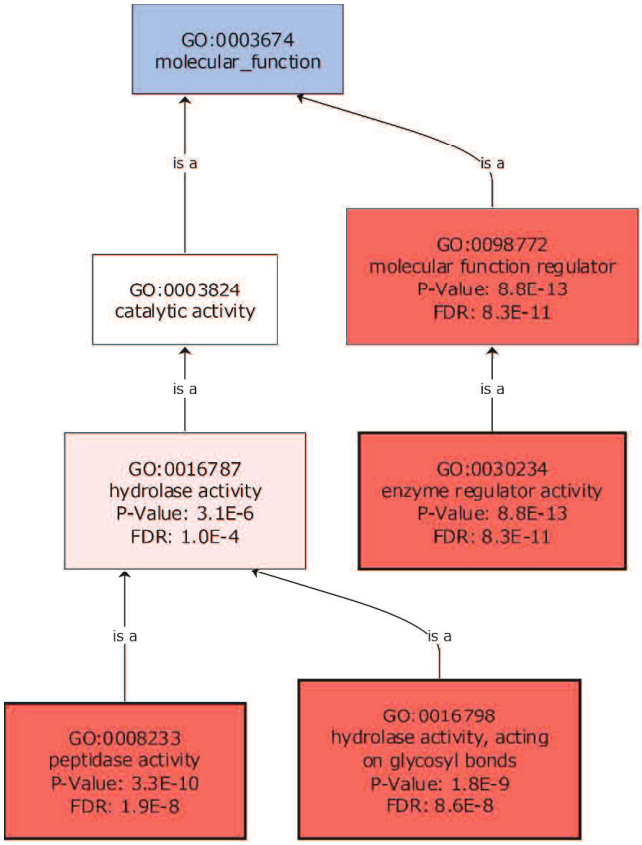


Figure 2-17. GO enrichment of predicted ES proteins.

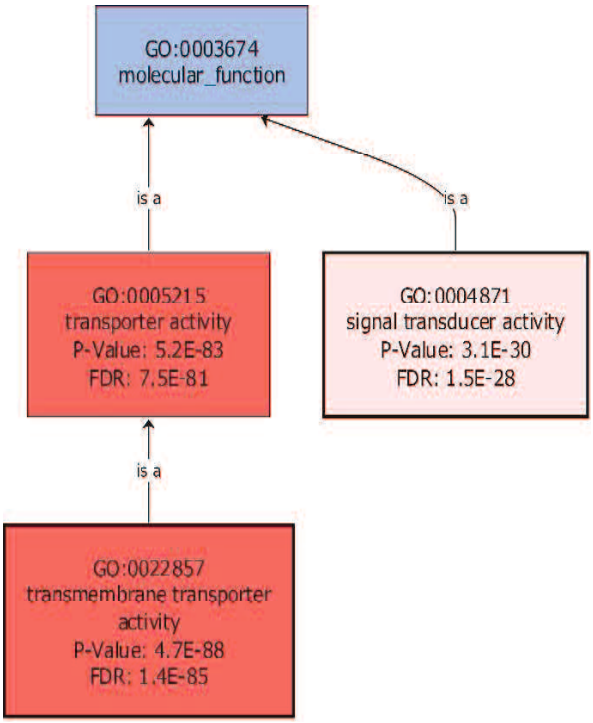


Figure 2-18. GO enrichment of TM proteins.

Table 2-3. Top 20 protein domains and families of predicted ES proteins from *E. multilocularis* reference transcriptome.

InterPro IDs	Description	No. of ES proteins
IPR013783	Immunoglobulin-like fold	23
IPR002223	Pancreatic trypsin inhibitor Kunitz domain	18
IPR007110	Immunoglobulin-like domain	16
IPR020901	Proteinase	16
IPR003599	Immunoglobulin subtype	13
IPR014044	CAP domain	12
IPR013032	EGF-like, conserved site	11
IPR003961	Fibronectin type III	11
IPR000742	EGF-like domain	10
IPR000169	Cysteine peptidase, cysteine active site	10
IPR000668	Peptidase C1A, papain C-terminal	10
IPR013128	Peptidase C1A	10
IPR025660	Cysteine peptidase, histidine active site	10
IPR017853	Glycoside hydrolase superfamily	9
IPR003598	Immunoglobulin subtype 2	9
IPR008860	Taeniidae antigen	9
IPR025661	Cysteine peptidase, asparagine active site	9
IPR013098	Immunoglobulin I-set	8
IPR009057	Homeodomain-like	8
IPR007087	Zinc finger, C2H2	7
IPR001283	Cysteine-rich secretory protein, allergen V5/Tpx-1-related	7
IPR013201	Cathepsin pro peptide inhibitor domain (I29)	7

Table 2-4. Top 20 protein domains and families of predicted TM proteins from *E. multilocularis* reference transcriptome.

InterPro IDs	Description	No. of TM proteins
IPR020846	Major facilitator superfamily	65
IPR013783	Immunoglobulin-like fold	60
IPR017452	G protein-coupled receptor, rhodopsin-like, 7TM	59
IPR000276	G protein-coupled receptor, rhodopsin-like	56
IPR005821	Ion transport domain	45
IPR007110	Immunoglobulin-like domain	44
IPR011701	Major facilitator superfamily	43
IPR018499	Tetraspanin/Peripherin	41
IPR002126	Cadherin	38
IPR015919	Cadherin-like	38
IPR020894	Cadherin conserved site	36
IPR008952	Tetraspanin, EC2 domain	34
IPR011009	Protein kinase-like domain	34
IPR013032	EGF-like, conserved site	32
IPR000719	Protein kinase domain	32
IPR000742	EGF-like domain	31
IPR003599	Immunoglobulin subtype	31
IPR027417	P-loop containing nucleoside triphosphate hydrolase	30
IPR003961	Fibronectin type III	28
IPR013083	Zinc finger, RING/FYVE/PHD-type	24

3.6 Predicted *E. multilocularis* protease analysis

There were 257 predicted proteases and 55 proteases inhibitor identified from the reference transcriptome of *E. multilocularis* (Appendix I). The 257 proteases constituted 2.41% of the 10,669 predicted protein-encoding transcripts of *E. multilocularis*. Proteases of five classes were characterized: 3.11%, 33.46%, 30.74%, 22.96% and 7.78% for aspartic, cysteine, metallo, serine, and threonine proteases, respectively (Figure

2-19). In addition, there were 28 transcripts coding proteases had no expression (RPKM<1) in oncosphere and metacestode (Appendix I) in the present study. And, we were able to assign KEGG (Kyoto Encyclopedia of Genes and Genomes) functional pathways to 212 *E. multilocularis* proteases using BlastKOALA analysis (Appendix I). 156 were predicted engage enzyme activity, and 101 were play roles in genetic information processes and 24 proteases likely perform functions in environmental information processes. (Figure 2-20). It was shown that genes that coding proteinases belong to the subfamily of C56, T01A, S26B were mostly conserve expressed in the sequenced sample. However, the antigen 5 and Mastin, which belong to the subfamily S01A, were mainly highly expressed in Met. And it is shown that *E. granulosus* antigen 5 is closely related to proteases of the trypsin family with only catalytic serine residue is replaced by threonine. As for the proteases inhibitor, it was shown that most genes that coding proteins belong to I02 family were highly expressed in oncospheres. And the genes coding 60S ribosomal protein L38 (I04) and Stefin B (I25A) are continuously expressed in all sequenced samples (Appendix I).

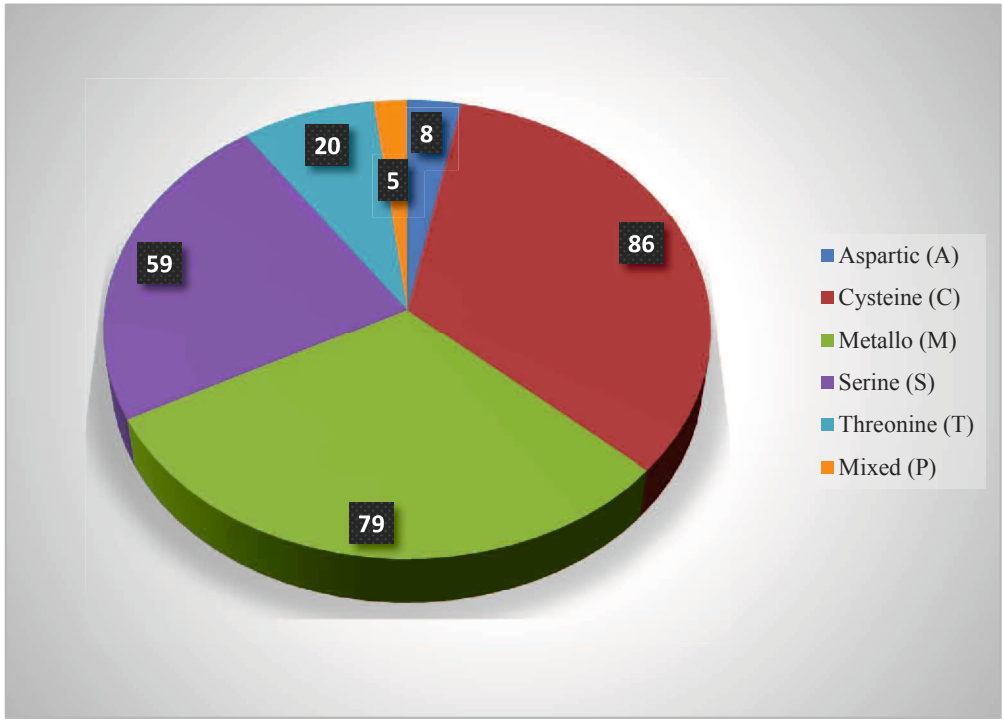


Figure 2-19. Proportions of protease families in the reference genomes of *E. multilocularis*.

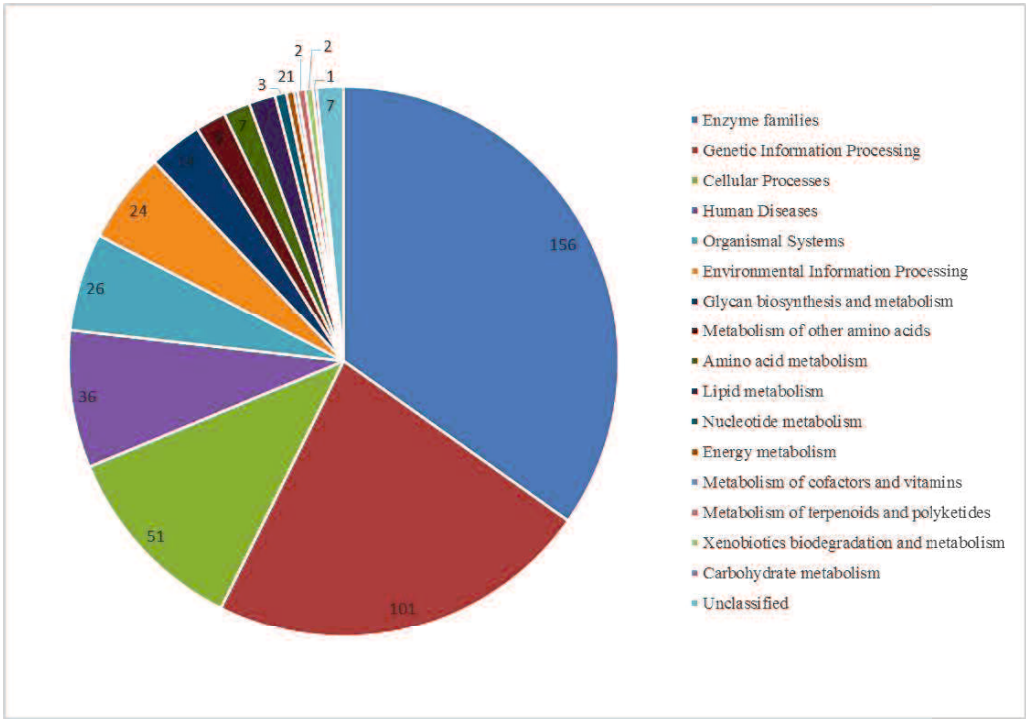


Figure 2-20. KEGG pathway interactions for predicted proteases from reference transcriptome of *E. multilocularis*. Graphic showing the number of proteases engaged in diverse signal processes and pathways.

3.7 Spliced-leader and trans-splicing genes analysis

Predicted were 968 SL-TS genes at the reference transcriptome level in the present study. 20% (2,177/10,669) genes in the reference transcriptome showed almost no expression (Average RPKM<1), 97% (938/968) of predicted trans-splicing genes were expressed in the stages of oncospheres and metacestodes. The *elp* gene (Antigen II/3), which identified as a trans-splicing gene in *E. multilocularis* was highly expressed among all sequenced samples (Figure 2-21) and solute carrier family 10 (SLC10) homology that identified as a trans-splicing gene in *T.solium* was expressed in all sequenced samples, as well (Figure 2-21).

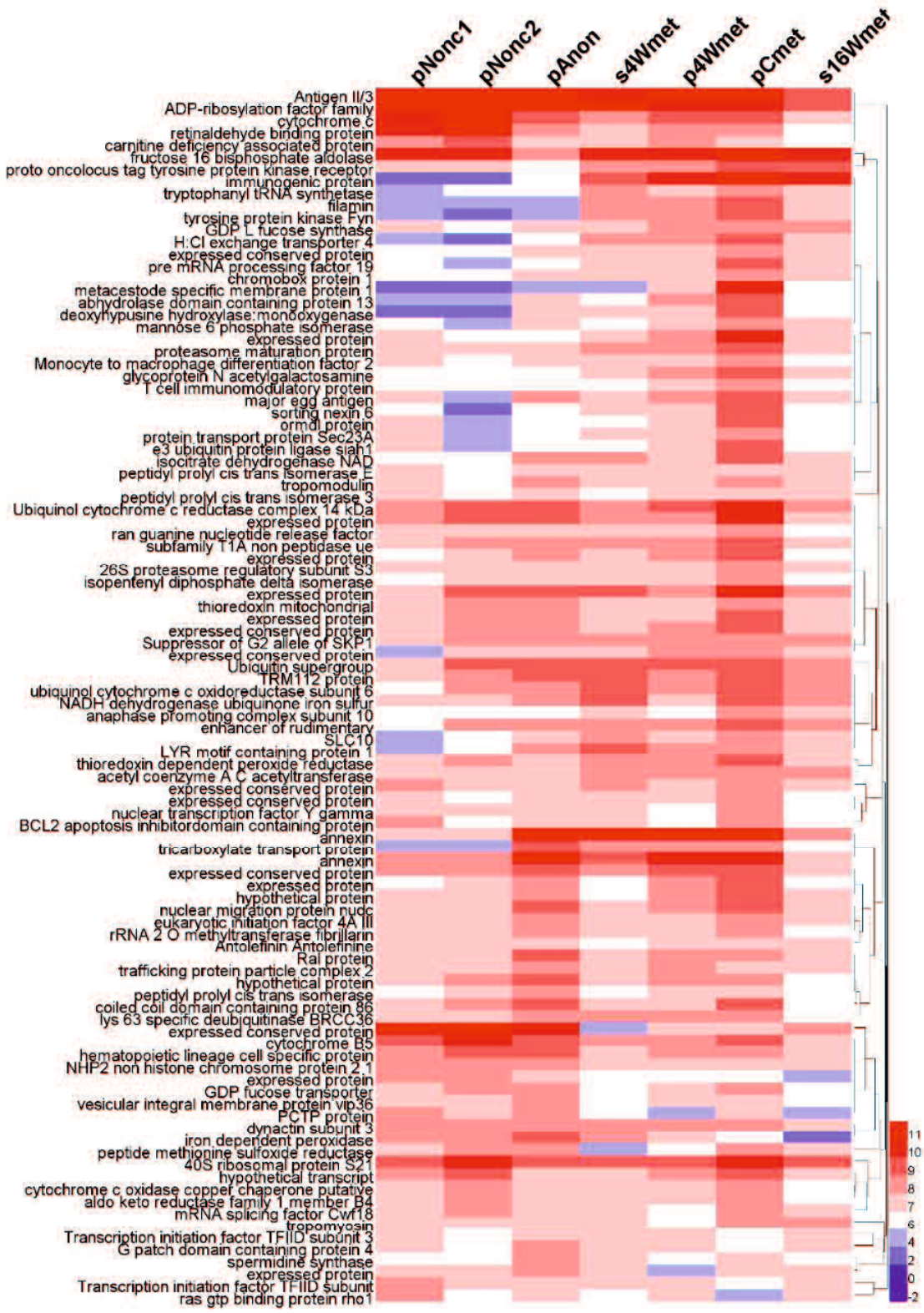


Figure 2-21. Heatmap of log-RPKM values for top 100 SL-TS genes. SL-TS genes are arranged by the average log-RPKM value among all sequenced samples

Discussion

The genus *Echinococcus* has been studied at the genomic level (Koziol and Brehm, 2015). Studying on genus *Echinococcus* have increased recently and several proteomes, transcriptomes, and even draft genomes have been published (Monteiro et al., 2010; Parkinson et al., 2012; Tsai et al., 2013; Zheng et al., 2013; Pan et al., 2014; Wang et al., 2015b; Huang et al., 2016). NGS has proved to be an appropriate approach to gain insights into the functional genomics of organisms such as *Echinococcus*, allowing the characterization of its transcriptome (Parkinson et al., 2012; Pan et al., 2014; Huang et al., 2016).

In present study, a global view of gene expression profiles and the stage-specific significant different express genes were demonstrated during the early invasion phases of the parasite. Putative ES and TM proteins were predicted using bioinformatics. Of these candidates “amyloid beta A4 protein” (EmuJ_001136900) and “WAP, Kazal, immunoglobulin, Kunitz and NTR” (EmuJ_001136500) which contain a pancreatic trypsin inhibitor Kunitz domain were highly expressed in Nonc and Aonc and poorly expressed in Cmet and 16Wmet. Real-time PCR (Huang et al., 2016) also validated amyloid beta A4 protein expression level in oncospheres. The previous study described two (EgKI-1, EgKI-2) secreted single domain Kunitz-type protease inhibitors from *E. granulosus* (Ranasinghe et al., 2015). It has been shown that the EgKI-1 which is highly expressed in oncospheres is a potent chymotrypsin and neutrophil elastase inhibitor that binds calcium and reduces neutrophil infiltration in a local inflammation model. It may also be involved in oncosphere host immune evasion by inhibiting neutrophil elastase and Cathepsin G once this stage is exposed to the mammalian blood system (Ranasinghe et al., 2015). The two candidate Kunitz-type protease inhibitors

analyzed in the present study were highly expressed in *E. multilocularis* oncospheres and both of them have a relatively close relationship to EgKI-1 indicting that they may both have similar functions to EgKI-1 in AE. This hypothesis requires further investigation to be confirmed.

Over the past decade, researches have been undertaken to develop vaccines and novel chemotherapeutic agents to prevent and control transmission of *Echinococcus* spp. *E. multilocularis* metacestode metabolites have been found to contain a cysteine protease that digests eotaxin, a C-C motif pro-inflammatory chemokine (Paredes et al., 2007; Mejri and Gottstein, 2009), and caspase 3 that can cause apoptosis was detected both in fertile and infertile *E. granulosus* cysts (Paredes et al., 2007). Furthermore, the serpin of *E. multilocularis* which highly expressed in oncospheres can readily inhibit trypsin and pancreatic elastase (Merckelbach and Ruppel, 2007; Huang et al., 2016). In present study, it was clear that Antigen 5 (S1-like), lysosomal protective protein (S10) and Cathepsin peptidase (C1) were highly expressed in Met. However, Kunitz inhibitor domain containing proteins (I2) were highly expressed in Nonc and Aonc, especially in Nonc.

Chapter 3. Transcriptome-wide based antigen candidate analysis for oncospheres and metacestodes of *E. multilocularis*

Abstracts

The transcriptome-wide based antigen candidate expression level of *E. multilocularis* oncospheres and metacestodes were tested using the Next-Generation Sequencing approach. The antigen profile analysis revealed that some diagnostic antigen GP50 isoforms, antigen EG95 family, major egg antigen (*HSP20*) and Tetraspanin 3 dominated in activated oncospheres, however, Antigen B subunits (EmAgB8/1, 2,3 and 4), Tetraspanin 5, 6, tegumental protein and Antigen 5 family dominated in metacestodes. Furthermore, heat shock proteins 70 family and antigen II/3(*elp*) were constantly expressed. Apomucins analysis showed that MUC-2 sub-family transcripts were present in all assayed samples and some of them were highly expressed in oncospheres, especially activated oncospheres. However, MUC-1 family transcripts presented only in *vitro*/*vivo* cultivated metacestodes. The identification of antigen expression profile during the parasite development stages, especially the stage of oncospheres, will give fundamental information for choosing candidate antigens used in vaccination and early diagnosis.

1. Introduction

It has been proven that infections can be blocked at the egg and early larval stages of *Echinococcus* and *Taenia* by antibodies and complement-dependent mechanisms (Rogan et al., 1992). Furthermore, *in vitro* hatching and activation of the oncospheres have been achieved, showing that oncospheres have an extended excretion apparatus and proteinases that may contribute to a considerable portion of the excreted proteins

during the penetration process (Lethbridge, 1980; Holcman et al., 1994; Hewitson et al., 2009; Santivanez et al., 2010). The fact that ES proteins produced in the early (oncosphere, immature metacestode) and chronic (mature metacestode) infectious stages by *E. multilocularis* can cause significant apoptosis of the dendritic cells (DC). The result suggests that the early infective stage of *E. multilocularis* is a strong inducer of tolerance in DC, which is most probably important for producing an immunosuppressive environment in the infection phase (Nono et al., 2012).

Immune response to larval *Echinococcus* spp. infections has been divided into "establishment" and "established metacestode" phases (Rogan et al., 1992; Siracusano et al., 2011). And it thought that the parasite is more susceptible to immune attack during early stages of infections ("establishment" phase) (Rogan et al., 1992; Siracusano et al., 2011). The immunogenic to the tested models of numbers of recombinant proteins are available. It was reported that vaccine Eg95, which is based on the recombinant protein cloned from mRNA from the oncosphere of *E. granulosus* and shown to be highly effective in vaccine trials of sheep and had induced a high level of protection (96–100%) for more than a year post-vaccination (Lightowlers and Heath, 2004). In addition, AgB (Ioppolo et al., 1996; Virginio et al., 2003), EmY162 (Katoh et al., 2008), P29 (Boubaker et al., 2014), EgEF (Margutti et al., 1999), Eg19 (Delunardo et al., 2010) and TSPs (Dang et al., 2012; Dang et al., 2009), derived from the *Echinococcus* spp., exhibit strong immunogenic properties in tested model, respectively. Furthermore, secondary AE, in which homogenates of the larval parasite are intraperitoneally, intravenously or intrahepatically injected into the host animals, is widely used; however, it does not reproduce the early stages of parasite development that occurs during natural infection via oral ingestion of the eggs (Matsumoto et al.,

2010). In addition, immunization with *E. multilocularis* 14-3-3 protein protected intermediate hosts from primary but not secondary challenge infection with AE (Siles-Lucas et al., 2003). Matsumoto et al. (2010) showed that the parasite lesions in the liver of primary AE at 4 weeks post-inoculation varied among the strains of mice suggesting that the resistance to the early stages of parasite infections, including parasite establishment in the liver, is genetically regulated.

Vaccination and early diagnosis are possible ways to prevent echinococcosis. Accurate immunodiagnosis of early stage of infection requires highly specific and sensitive antigens. At present, little gene expression data has been published for eggs and early larval stages. Thus, experiments on identifying antigens for use in immunodiagnostic assays are a crucial point in the improvement of the diagnostic tool and must be based on the developmental stage of the parasite.

As for mentioned above and gain understanding of the gene expression patterns for diagnostic assay and vaccine design, we analyzed the transcriptomes of Nonc, Aonc, 4Wmet, Cmet and 16Wmet to identify homologues of the various known antigens of tapeworms, especially *Echinococcus* spp.

2. Materials and Methods

2.1 Preparation of parasite samples

The parasite samples were prepared as written in chapter 1.

2.2 Antigen homologues in *E. multilocularis*

Putative antigen homologues of amino acid sequences in the *E. multilocularis* reference genome (German isolates) were identified using known antigen sequences

(accession numbers shown below). Briefly, BLASTP (Altschul et al., 1997) comparisons were carried out using the amino acid sequences of *E. multilocularis* genome version 4 as queries and the known antigens sequences as subjects. Sequences with an E-value < $1E^{-25}$, identity value & query coverage> 80% were considered to be homologues of matched antigens within *Echinococcus* spp. Furthermore, antigen EG95 and diagnostic antigen GP50 family homologues were queried using the same amino acid sequences as used previously inference genome of *E. multilocularis* (German isolates).

2.3 Accession numbers of published antigen candidates

Accession numbers for various known *E. multilocularis*, *E. granulosus* and *Taenia solium* antigens sequences used in this study were as follows: *E. multilocularis* (CAA59739, CAA10109, AAL51153, BAC11863, BAC66949, BAC77657, BAD89809, BAD89810, BAD89811, BAD89812, Q8WT41, BAF02516, BAF02517, BAF63674, BAF79609, ACJ02401, ACJ02402, ACJ02403, ACJ02404, ACJ02405, ACJ02406, ACJ02407, BAJ83490, BAJ83491, AER10547, AHA85399, Q07840, Q24895, Q24902, Q27652, Q8MM75, Q8WPI6, Q8WT42, Q9GP32, Q9NFZ5, Q9NFZ6, and Q9NFZ7); *E. granulosus* (AAF02297, AAL87239, CAF18421, AAX20156, AAX73175, ACA14465, ACA14466, ACA14467, ABI24154, AFI71096, AGE12481, AGE12482, O16127, O17486, O46119, P14088, P35417, P35432, Q02970, Q03341, Q03342, Q04820, Q07839, Q24789, Q24798, Q24799, Q24800, Q8MUA4, Q8T6C4, Q95PU1, Q9BMK3, Q9GP33, Q9GP38, Q9U408 and Q9U8G7); and *T. solium* (AAP49286, AAP49287, AAP49288, AAP49284 and AAP49285).

3. Results and Discussion

3.1 Apomucins

Since the laminated layer (LL) is synthesized by the tegumental syncytium located at the outer most part of germinal layer, genes coding for LL components should be expressed by the tegumental cells (Díaz et al., 2015) and essential in the interface with host. A survey of the *E. granulosus* transcriptome identified a family of apomucins among highly expressed Thr-rich proteins in the germinal layer (Parkinson et al., 2012) and only exists in *Echinococcus* spp. (Tsai et al., 2013; Zheng et al., 2013). There are two families of apomucin in *Echinococcus* spp. named as MUC-1 family and MUC-2 sub-family (Koziol et al., 2014;Díaz et al., 2015). And the MUC-1 family is highly expressed in metacestodes and only has a single gene in each of *E. granulosus* and *E. multilocularis* genomes and named EgrG_000742900.1 (EGR_08371) and EmuJ_000742900.1, respectively (Tsai et al., 2013; Zheng et al., 2013; Díaz et al., 2015). However, the MUC-2 sub-family with several similar genes and some of them appear to have an unpaired cysteine (Tsai et al., 2013). In this study, MUC-1 family was highly expressed in Cmet and 4Wmet, whereas one MUC-2 sub-family gene (EmuJ_000408200.1) was present in all assayed materials but highly expressed in oncospheres and 4Wmet (Appendix II). Most interesting, the rest MUC-2 genes were only detected in Aonc, suggesting these genes may have special function in Aonc. The LL is widely thought to be a key components in the host–parasite interaction in echinococcosis (Díaz et al., 2015). Its roles include shielding the parasite from direct attack by host immune cells, and probably down-regulating local inflammation (Díaz et al., 2015). MUC-1, as expected, is expressed in the tegument and thought to contribute to conventional glycocalyces (Koziol et al., 2014). The quotient of gene expression

level between the sum of MUC-2 members and MUC-1 was approximately 1/255 for *in vitro* metacestodes of *E. multilocularis* in a previous study (Tsai et al., 2013) which is in concordance with result for Cmet and 16Wmet. This suggests that it is highly likely that MUC-1 is a major LL constituent in *E. multilocularis* metacestodes. In addition, the formation of LL within 13 days of *in vitro* culture post-oncospheral has been observed (Gottstein et al., 1992). In the present study, MUC-2 sub-family members were highly expressed in oncospheres, especially in activated oncospheres, indicating that oncospheres are used to prepare materials, such as mRNA for synthesis of MUC-2 proteins, to construct LL during transformation to metacestodes of *E. multilocularis*. Furthermore, it thought that LL of post-oncospheres takes part in protecting the developing oncosphere from host immune reactions (Sakamoto and Sugimura, 1970), thus, MUC-2 may be a key factor in post-oncosphere host-parasite immune reactions.

3.2 Em-alp

It is observed that alkaline phosphatase activity in metacestode very high (strong reaction in less than 5 minutes) and restricted to the distal syncytial tegument of the germinal layer (Koziol et al., 2014), but is not found in brood capsules. This may indicate that the gene coding alkaline phosphatase is a good histochemical marker for germinal layer. In *E. multilocularis* reference genome, there are four genes that code alkaline phosphatase named as *em-alp-1* (EmuJ_000393300.1), *em-alp-2* (EmuJ_000393400.1), *em-alp-3* (EmuJ_000752700.1) and *em-alp-4* (EmuJ_000752800.1). And the amino acid sequence analyses showed that em-alp-1 and em-alp-2 protein contain a signal peptide (Figure 3-1) at the N-terminal. Moreover, *em-alp-1* and *em-alp-2* are significant highly expressed when Nonc transforms to Aonc,

and has expressed in all sequenced samples of present study, whereas the *em-alp-3* and *em-alp-4* were almost no expressed in the sequenced samples of present study (Appendix II) and the previse study also show that *em-alp-1* and *em-alp-2* were found to be specifically expressed in the germinal layer (Koziol et al., 2014), while *em-alp-4* has substitutions of conserved catalytic amino acid residues, and cannot be detected by RT-PCR in the germinal layer or in protoscoleces (Koziol et al., 2014), suggesting that *em-alp-4* is a pseudogene, although expression was observed in RNA-Seq data of adult worms (Koziol et al., 2014). As for *em-alp-3*, it is show that it was only detected in protoscoleces, with a strong up-regulation after protoscolex activation that indicated the expression of *em-alp-3* need stimulate from their host, especially the definitive host. These data suggested that *em-alp-1* and *em-alp-2* were expressed in the oncospheres and tegumental cells of the germinal layer, and *em-alp-2* was another marker for the tegumental cells in the germinal layer, while *em-alp-3* was expressed in the protoscolex excretory system.

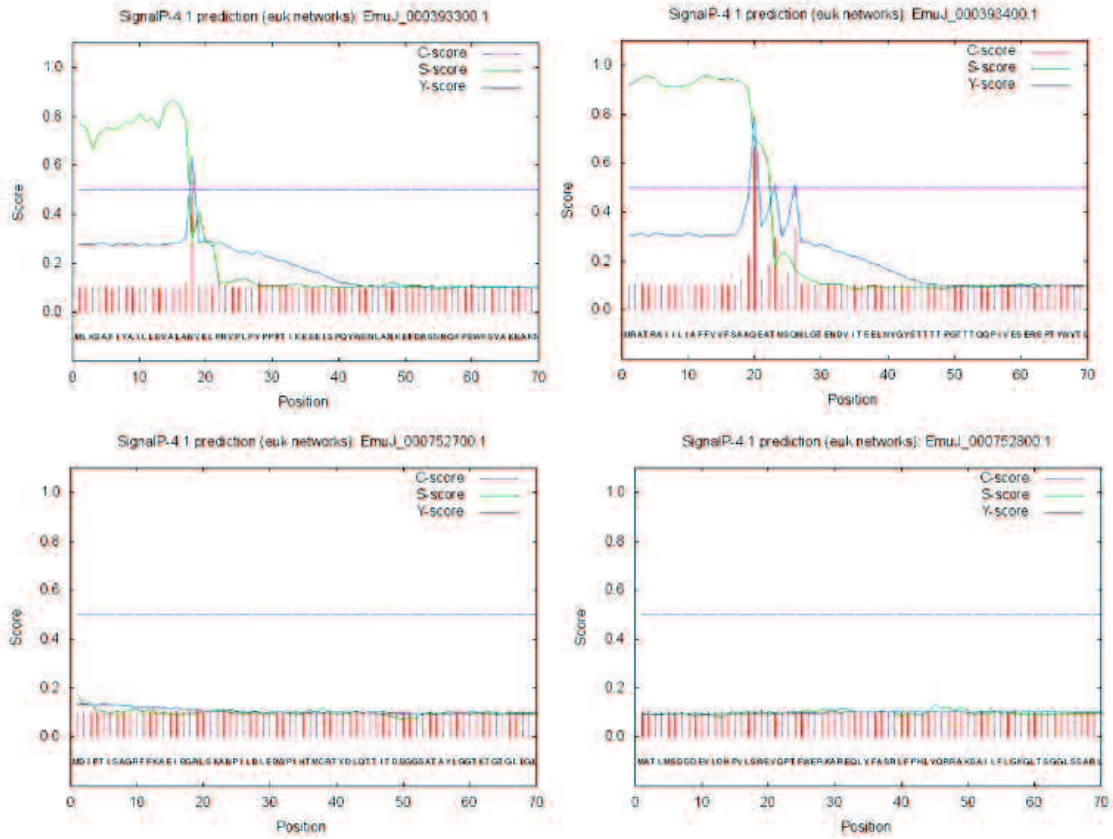


Figure 3-1. SignalP 4.1 prediction of cellular localization and signal sequence cleavage site of Em-alp.

3.3 Tubulin

Tubulin-α acetylation is a post-translational modification that occurs in highly stable microtubule. Because it is able in principle to label all nerve cells independently of their neurotransmitter, acetylated tubulin-α immunoreactivity has been used to describe in detail the complete nervous system of numerous invertebrates of *E. multilocularis* (Koziol et al., 2013). Benzimidazoles, particularly albendazole and mebendazole, are the most important forms of AE therapy (Brehm et al., 2000b). The main mode of benzimidazoles action on helminths involves direct binding of the drugs to β-tubulin, thus inhibiting the polymerization of microtubules (Lacey, 1990). In reference genome of *E. multilocularis*, there were 17 genes that coding proteins contain domains of alpha

tubulin (Appendix II) and two genes (EmuJ_000413200.1, EmuJ_000886400.1) were highly expressed among those alpha tubulin domain contain proteins (Appendix II). Furthermore, one gene (EmuJ_000886400.1) was significant low expressed at Nonc when compared to other sequenced samples (Appendix II). In addition, there were 10 genes that coding beta tubulin domain contain proteins. And one beta tubulin coding gene (EmuJ_000672200.1) was highly expressed in all sequenced samples, but beta tubulin coding genes (EmuJ_000202500.1, EmuJ_000202600.1) were significant low expressed in Nonc (Appendix II).

3.4 Actin

Actins constitute a highly conserved family of proteins found in all eukaryotes [l]. These proteins are found predominantly in the cytoplasm of cells where monomers polymerize to form microfilaments. Microfilaments of actin participate in various cell functions such as muscle contraction, cell cytoskeleton and motility (Oliveira and Kemp, 1995). Actin microfilaments also function in bundles where they form microvilli, stereocilia, and other cellular structures. The reference genome of *E. multilocularis* shows that there are 7 transcripts of actin, which show high homologies to *E. granulosus* actin. One actin (EmuJ_000036300.1) expressed highly in all sequenced samples in present study, but most of the other actin (EmuJ_000406900.1, EmuJ_000407200.1, EmuJ_000061200.1, EmuJ_000701700.1, EmuJ_000703300.1) expressed only in Cmet and 16Wmet (Appendix II). These results suggested that highly expression actin in Cmet and 16Wmet might locate in the muscle cell of the parasite.

3.5 Tropomyosin

It was shown that isoforms of two *tropomyosin* genes were strongly expressed in the suckers of *E. granulosus* protoscoleces (Alvite and Esteves, 2009). And a recent study has shown that tropomyosin isoforms can be found in the muscle fibers in the germinal layer, accumulating in the interior of brood capsules and in the muscle layers during protoscolex development in *E. multilocularis* (Koziol et al., 2014). There are two *tropomyosin* isoforms in *E. multilocularis* genome (Tsai et al., 2013), but previous study (Koziol et al., 2014) show that em-tpm-1 (EmuJ_000958100.1) was no expression using whole mount in situ hybridization (WMISH) method in the germinal layer. In present study, it was shown that *em-tpm-2* has a relative higher expression level in the cultivated metacestodes (no protoscolex) than *em-tpm-1*, but a lower expression level in the mature metacestodes (contain protoscolex), *in vivo* (Appendix II). In summary, em-tpm-1 protein could be used as a molecular marker for the development of muscle cells during brood capsule and protoscolex development, but not in the germinal layer.

3.6 Diagnostic antigen GP50

Taenia solium GP50 has been used for the diagnosis of cysticercosis (Levine et al., 2004). GP50 isoforms are species-specific antigens and may be stage-specific in *Cysticercus cellulosae* (Hancock et al., 2004) based on the lack of antibody reactivity with one serum sample from an individual confirmed to be taeniasis-positive but cysticercosis-negative (Hancock et al., 2004). A previous study showed that more than 90% of *E. multilocularis* GP50 isoforms were not expressed in metacestodes cultivated *in vitro* (Tsai et al., 2013), and our present work also corroborated to this finding, since few or no transcripts of GP50 were found in Cmet (Appendix II). Some GP50 isoforms

were expressed in 4Wmet from *in vivo* DBA/2 mice infections, suggested that these GP50 isoforms might be key factors in the host-parasite interface during the early stage of infection.

3.7 HSPs antigens

The putative HSP20 gene, which can express immunogenic products and stimulate the immune system, showed high expression in the oncosphere stage (Kouguchi et al., 2010; Merckelbach et al., 2003). The predicted HSP20 homologue (onco2) also showed the highest expression at Nonc (RPKM=9,125.50) and Met as well (Appendix II). Taken together with the findings from the published transcriptome of *E. multilocularis* (Tsai et al., 2013), it is clear that this molecule was expressed at almost all stages of *E. multilocularis*, including non-activated oncosphere, activated oncosphere, metacestode and adult worms.

The HSP70 family has been described as the major antigens in *Echinococcus* spp. (Mühlschlegel et al., 1995; Ortona et al., 2003) and the most striking gene family expansions with 22 full copies in *E. multilocularis* reference genomes. Furthermore, in various infectious disease models including echinococcosis, vaccination strategies using HSPs have been produced significant protection (Ortona et al., 2003; Zügeli and Kaufmann, 1999). The transcriptome datasets of the present study show that HSP70 homologues were constantly expressed in all stages (Appendix II). Continuous antigenic stimulation with parasite-derived HSP families would induce an apparent antibody response to these molecules in infected animals. These antibody responses create an opportunity to use HSPs in diagnostic assay and vaccine development for echinococcosis.

3.8 Antigen II/3 (*elp*)

Antigen II/3 share homology with the mammalian ezrin/radixin/moesin (ERM) protein family that is involved in several key processes related to cellular architecture, including cell-cell adhesion, membrane trafficking, microvillus formation and cell division (Louvet‐Vallée, 2000). Antigen II/3 is encoded by the *elp* gene and the antigens of Em10 and Em18 are thought to be homologues, which have also been used as important diagnostic antigens (Felleisen and Gottstein, 1993; Sako et al., 2002). In the present study, antigen II/3 was highly expressed in all sequenced samples. Previous studies proved that antigen II/3 can be expressed at the stages of protoscoleces, metacestodes and adult and localized within the germinal layer and parenchymal cell of protoscoleces and on the surface of calcareous corpuscles (Felleisen and Gottstein, 1993). It has been shown that antigen II/3 is also constantly expressed in the early stage metacestodes and adults (FPKM>200 (Tsai et al., 2013)).

The viability of protoscoleces was significantly reduced at day 10 after silencing the *elp* gene statistically (Mizukami et al., 2010). Together with the constantly high expression level of antigen II/3 at almost all life-cycle stages may hint that antigen II/3 has a fundamental role for supporting parasites, such that antigen II/3 can act not only as an important diagnostic antigen special for the oncosphere stage, but also as a vaccine candidate.

3.9 Antigen B subunits

Antigen B (AgB) was initially identified as major hydatid cyst fluid antigen of *E. granulosus* (Oriol et al., 1971). This antigen is a polymeric lipoprotein with a molecular weight of 120 KDa, and five 8 KDa subunits were identified as EmAgB 8/1,

EmAgB 8/2, EmAgB 8/3, EmAgB 8/4 and EmAgB 8/5 in *E. multilocularis* (Mamuti et al., 2007). And, there are seven isoforms that code antigen B subunits in *E. multilocularis* reference genome, of which EmAgB8/3 (EmuJ_000381500) had highest expression in metacestodes (Appendix II); even Aonc showed relative high expression (RPKM=186.79). But the other two isoforms (EmuJ_000381600.1 and EmuJ_000381700.1) that code EmAgB8/3 subunit were no expression expressed in all the sequenced stages (Appendix II). Unlike other AgB subunits, which were almost within the 2-fold expression level when comparing 4Wmet to Cmet or 16Wmet, EmAgB8/2 showed a more than 10-fold difference (Appendix II). Previous studies have shown that the sensitivity of EgAgB2 in *E. granulosus* was obviously different in different assays (Virginio et al., 2003; Jiang et al., 2012), and one reason may be that *E. granulosus* isolated from CE patients in different countries expresses differing levels of the AgB2 subunit (Jiang et al., 2012). The present data suggest this might be caused by differing expression of AgB2 within the early stage metacestodes. Furthermore, antibody responses to AgB in different larval stages (CE1-CE5) of different sensitivities (Zhang et al., 2012) also indicate that AgB subunits dynamically change in larval stages. In conclusion, from the perspective of expression level, we proposed that EmAgB8/3 may be expected to have essential metabolic functions throughout all life-cycle stages of the parasite, while EmAgB8/1, EmAgB8/2, and EmAgB8/4 may be essential factors for survival of larvae in intermediate hosts. EmAgB8/5, which was firstly detected to be highly expressed in the adult of *E. multilocularis* (Mamuti et al., 2007), but was not detected in this study.

3.10 EG95 (Fibronectin type III-like) antigen

Previous studies have described the effectiveness of Fibronectin type III domain-like protein vaccines against echinococcosis (Chow et al., 2001; Gauci et al., 2002; Katoh et al., 2008). These highly immunogenic proteins, which may be involved in host invasion, are encoded by a multigene family; EG95 vaccine is effective against *E. granulosus* (Chow et al., 2004) and EM95 is effective against *E. multilocularis* (Gauci et al., 2002). The antigen is a secreted protein with a GPI anchor that is upregulated during oncosphere activation (Chow et al., 2004; Zhang et al., 2003) and is probably involved in cell adhesion (Bonay et al., 2002). Three (EmuJ_000328500, EmuJ_000368620, EmuJ_000710400) out of five EG95 relatives followed the previous prediction (Tsai et al., 2013), and corresponded to the top 20 expressed proteins in Nonc and Aonc (Appendix II). Unlike EmuJ_000328500 (Identity=95.68%) and EmuJ_000368620 (Identity=99.36%), the highly expressed EmuJ_000710400 showed low identity (Identity=41.67%) with the published EM95 antigens, suggesting that it may be a new candidate antigen for vaccine development against alveolar echinococcosis. Most interestingly, EmuJ_000368620 that shows highest identity to EM95 was significantly expressed in Aonc (Appendix II). However, EmuJ_000328500 that shows highest identity to Onco1 (79.5% identity to EM95) was highest expression in Nonc (Appendix II). It is not surprised that EmuJ_000328500 has the highest expression level in Nonc in accordance with the data from previous study (Merckelbach et al., 2003).

EmY162, a potential vaccine candidate against *E. multilocularis*, showed 31.4% identity to the amino acid sequence of EM95, which is also a fibronectin type III-containing protein (Katoh et al., 2008). EmuJ_000564900 (85% identity to BAF79609) was expressed in most of the life-cycles stages, especially in Aonc

(Appendix II), and EmuJ_000515900 (98% identity to the BAF79609) primarily expressed in Cmet and 16Wmet, in the present study (Appendix II), which is consistent with findings in a previous study (Katoh et al., 2008; Tsai et al., 2013).

3.11 Serine protease inhibitors

Serpins (serine proteinase inhibitors) constitute a huge family of about 1,500 identified members. The function of serpins ranges from the regulation of proteinases from immune effector cells, blood coagulation and in the complement system in mammals (Law et al., 2006). The serpin of *E. multilocularis* (serpin^{Emu}) was the first member described from this class of cestodes (Merckelbach et al., 2003), and sequence analysis indicated that it was an intracellular serpin (Merckelbach and Ruppel, 2007; Merckelbach et al., 2003). However, the putative amino acid sequences of the parasite genome data and the *de novo* assembled data in the present study (Tsai et al., 2013) suggested that serpin^{Emu} with a signal peptide predicted by Phobius (Käll et al., 2007). In addition, *in vitro* assays have confirmed that serpin^{Emu} fails to inhibit Cathepsin G and chymotrypsin but could readily inhibit trypsin and pancreatic elastase (Merckelbach and Ruppel, 2007), both of which are digestive enzymes in the intestines of mammals. Therefore, an extracellular role of serpin^{Emu} might be possible. Previous descriptions of the ultrastructure of *E. granulosus* oncospheres have referred to the penetration gland cells (Holcman and Heath, 1997) and proteinases may make up a considerable portion of the excreted proteins during the penetration process that is hypothesized to involve the secretion that may help the parasite penetrate the intestinal wall of the intermediate host (Holcman et al., 1994; Holcman and Heath, 1997; Lethbridge, 1980; Reid, 1948). If serpin^{Emu} is excreted by penetration gland during the invasion of oncospheres, it

might be able to block the proteolytic attack of host digestive enzymes. If so, it may even be a target of the intestinal immune system and a vaccine candidate.

3.12 Tetraspanins

Tetraspanins (TSPs) are a superfamily of plasma membrane-associated proteins consisting of four conserved transmembrane (Seigneuret et al., 2013). They have been used as vaccine candidates against schistosomiasis, echinococcosis and as diagnostic antigens for cysticercosis (Zhu et al., 2004; Hancock et al., 2006; Dang et al., 2009; Dang et al., 2012). In addition, it was proven that Tetraspanins in the tegument of schistosomula and adult worms can act as receptors for host ligands, including MHC molecules, allowing parasites to mask their non-self-status and escape host immune responses (Tran et al., 2006). A total of 9 amino acid sequences (Appendix II) showed 91%-100% identity to the seven published Em-TSPs (Dang et al., 2009). In addition, there were two putative Em-TSP3 isoforms (Appendix II), and most mutation sites were located at the LEL variable region (Figure 3-2).

Previous transcriptome data (Tsai et al., 2013) and the present study showed that Em-TSP5 is expressed at almost all life-cycle stages and was significantly expressed at the stage of Aonc and 4Wmet compared with Nonc (Appendix II). Em-TSP5 was intensely stained in sections of the germinal layer of metacestode (Dang et al., 2009). Em-TSP5 is closely related to the T24 antigen of *T. solium*, a diagnostic antigen for cysticercosis (Hancock et al., 2006), which suggest that Em-TSP5 may be an important diagnostic candidate for detecting early stage infection.

Em-TSP1, one of the highly protective vaccine candidates (Dang et al., 2009), is located at the surface (germinal layer/tegument) of *E. multilocularis* larvae and the

tegument of the adult worms. Significantly high expression of Em-TSP1 in early stage metacestode compared with Nonc and Aonc was observed (Appendix II). A previous study showed that another protective effect vaccine candidate, Em-TSP3, is localized in the non-activated oncospheres and protoscoleces and the germinal layer of *E. multilocularis* cysts (Dang et al., 2012); the genome-mapped data in the present study showed relative higher expression in Aonc and 4Wmet than in Cmet (no protoscoleces), and the expression level of Em-TSP3 varied within Nonc (Appendix II).

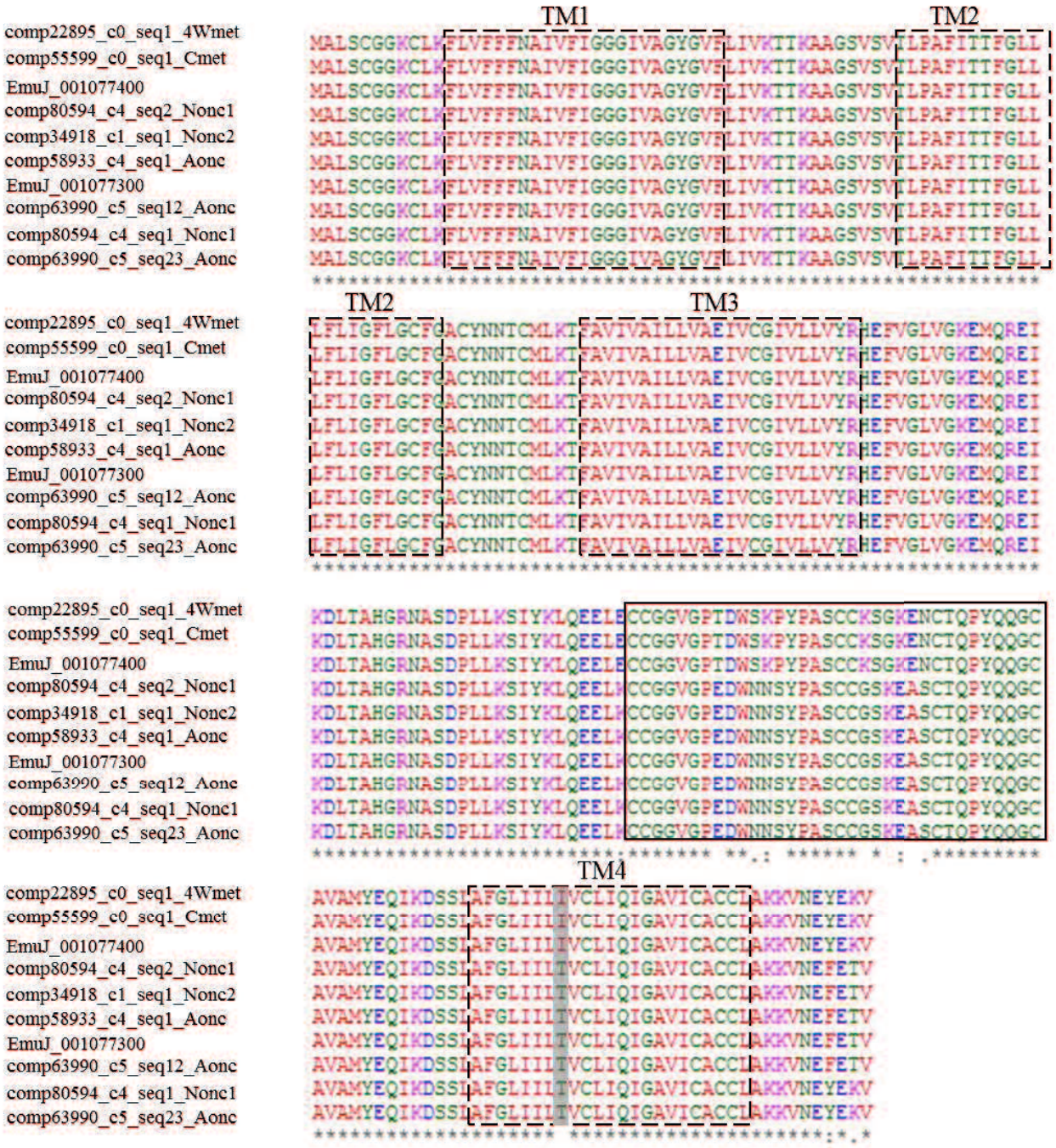


Figure 3-2. Protein alignment of putative Em-TSP3 isoforms with four transmembrane. Fully conserved residues are marked with (*), those replaced with amino acids of strongly similar properties with (:) and of weakly similar properties with (.) LEL variable region are in the solid line box and predicted transmembrane region are in the dashed line box.

Summary

Echinococcus multilocularis, a worldwide zoonotic parasite, causes alveolar echinococcosis that is of great public health concern. The parasite requires two mammalian hosts, a definitive host (carnivores) and an intermediate host (mostly wild rodents), to complete its life cycle. For decades, it has been developed as an experimental model to study host-parasite interplay. The excretory-secretory (ES) proteins of parasites have been found to be crucial for their survival inside and outside of the host organisms by acting as virulence factors or host immune responses modulators. In addition, transmembrane (TM) proteins, as a group of membrane proteins, are involved in many important biological processes. Thus, ES and TM proteins received great attention as antigen proteins for vaccine or drug development that aimed to cure or prevent *E. multilocularis* infection. However, only very few candidate proteins have been identified due to the lack of systematic understanding of *E. multilocularis* genome and gene expression details. Until recently, the *E. multilocularis* genome sequence has been resolved, which largely facilitated the identification of antigen families that possess high potentials for developing diagnostic assays, vaccines, and drugs. In order to diagnose and prevent the infection of *E. multilocularis* at a very early stage, discovery of/exploring potential antigen candidate proteins that are expressed during oncospheres and early metacestode larva stage would be important. Thus, it is invaluable to dissect the gene expression profile *E. multilocularis* at a stage-specific manner. In this thesis, we used next-generation sequencing approach to investigate gene expression dynamics at different stages of *E. multilocularis* (Nemuro strain).

In this work, seven *E. multilocularis* mRNA samples from non-activated oncospheres (Nonc), activated oncospheres (Aonc), 4-week immature metacestodes (4Wmet), 16-weeks mature metacestodes (16Wmet), and *in vitro* cultivated metacestodes small vesicles (Cmet) were collected to profiling the gene expression dynamics at different stages of the parasite development. The single-end (s4Wmet and s16Wmet) and pair-end (pNonc, pAonc, p4Wmet, pCmet) sequencing gave 700 million clean reads with > 90% of all bases having Phred scores above 30. Moreover, most of *de novo* assembled contigs could be matched to the reference genome of *E. multilocularis*, which indicated that all sequenced reads of the seven samples and the assembled contigs of pair-end sequencing samples were reliable.

The gene expression profile analysis revealed that amyloid beta A4 protein, some diagnostic antigen GP50 gene isoforms, antigen EG95 family, major egg antigen (HSP20) and *Tetraspanin 3* were dominantly expressed in activated oncospheres, while Antigen B subunits (EmAgB8/1, 2,3 and 4), Tetraspanin 5 and 6, tegumental protein and Antigen 5 family were highly expressed at metacestodes stage.

Furthermore, heat shock proteins 70 family and antigen II/3(*elp*) are constantly expressed in all stages. And apomucin analysis showed that MUC-2 sub-family transcripts were present in all sequenced stages and some of them were highly expressed in oncospheres, especially in Aonc. However, MUC-1 family transcripts only present in metacestodes. Functional annotation of *E. multilocularis* transcripts revealed that 769 predicted ES and 1980 predicted TM proteins were enriched with gene ontology term of ‘extracellular region’ and increased ‘transmembrane transporter activity’. And it was shown that the up-regulated genes in metacestodes were enrichment in ‘cell adhesion’ when compared with oncospheres which indicted that

many molecular that took part in the host-parasite interfaces were highly expression in the metacestodes to regulate the immune response for establishing the chronic infection. Strikingly, 97% (938/968) of the predicted trans-splicing genes were expressed at oncospheres and metacestodes, though 20% (2,177/10,669) genes in reference transcriptome were almost no expression. Furthermore, the protease analysis showed that there were 257 proteases and 55 proteases inhibitor in the reference transcriptome and most of these proteases had relatively higher expression levels in 16Wmet, which indicated they might play important roles in regulating host immune response during the chronic stage of larval echinococcosis. In contrast, proteases inhibitors, especially Kunitz-type protease inhibitors (I02), were highly expressed in oncospheres, suggesting they might play important roles to block the proteolytic attack in the host alimentary tract.

The results clearly showed that the expression dynamics of antigen candidates, ES proteins, TM proteins, and proteases in *E. multilocularis* at different developmental stages/growth phases were differentially regulated. These large sets of detailed and systematical results might provide novel insights for studying host-parasite interaction at different stages of the life cycle. In addition, it also serves as an invaluable resource for future experimental studies that aim to develop new intervention tools, including vaccines and drugs, against this parasite at its early infection phase.

Acknowledgements

I would first give my thanks to my supervisor Prof. Yuzaburo OKU, and other teachers of Prof. Hiroshi SATO, Prof. Toshihiro ITO, Prof. Takashi TAKEUCHI and Assoc.Prof. Kyeongsoon KIM for their supports in life and studies, and it is absolutely impossible for me to finish studies without their guidance and encouragement. I am indebted to Prof. Yutaka SUZIKI and Miss Terumi HORIUCHI for giving me chance to visit his lab and teach me how to analysis next-generation sequencing data, and to Kinpei YAGI, Hirokazu KOUGUCHI, Takao IRIE from Hokkaido Institute of Public Health, for providing of material, and to Bruno GOTTSTEIN from University of Bern, Switzerland, for generous providing of material. I would like to give my gratitude to everyone in our research group for their help in my experiments and life.

I extend my thanks to Zhisheng DANG from National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention and Prof. Wei LI who is my supervisor of master degree and it is impossible for me to go to Japan to obtain the PHD degree without Prof. LI' advice and persuade.

My thanks are especially given to my parents for their love and unwavering support. Their eager anticipation and smiles are main resources of impetus that drives me to surmount obstacles in life and study in Japan.

My study was financially supported by Japanese Government (Monbukagakusho) Scholarship and this work was funded by KAKENHI (No. 25450425) of Japan Society for the Promotion of Science (JSPS).

Reference

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, 25, 3389-3402.

Alvite, G., Esteves, A., 2009. *Echinococcus granulosus* tropomyosin isoforms: from gene structure to expression analysis. Gene, 433, 40-49.

Anders, S., Pyl, P.T., Huber, W., 2014. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics, 166-169.

Bendtsen, J.D., Jensen, L.J., Blom, N., Von Heijne, G., Brunak, S., 2004. Feature-based prediction of non-classical and leaderless protein secretion. Protein Engineering Design and Selection, 17, 349-356.

Blaxter, M., Liu, L., 1996. Nematode spliced leaders—ubiquity, evolution and utility. International Journal for Parasitology, 26, 1025-1033.

Blumenthal, T., 2004. Operons in eukaryotes. Briefings in Functional Genomics & Proteomics, 3, 199-211.

Bonay, P., González, L.M., Benítez, L., Foster, M., Harrison, L.J., Parkhouse, R.M.E., Gárate, T., 2002. Genomic and functional characterisation of a secreted antigen of *Taenia saginata* oncospheres. Molecular and Biochemical Parasitology, 121, 269-273.

Boubaker, G., Gottstein, B., Hemphill, A., Babba, H., Spiliotis, M., 2014. *Echinococcus* P29 antigen: molecular characterization and implication on post-surgery follow-up of CE patients infected with different species of the *Echinococcus granulosus* complex. PlOS ONE, 9, e98357.

Brehm, K., 2010. *Echinococcus multilocularis* as an experimental model in stem cell research and molecular host-parasite interaction. Parasitology, 137, 537-555.

Brehm, K., Jensen, K., Frosch, M., 2000a. mRNA Trans-splicing in the Human Parasitic Cestode *Echinococcus multilocularis*. Journal of Biological Chemistry, 275, 38311-38318.

Brehm, K., Kronthaler, K., Jura, H., Frosch, M., 2000b. Cloning and characterization of *β-tubulin* genes from *Echinococcus multilocularis*. Molecular and Biochemical Parasitology, 107, 297-302.

Brehm, K., Spiliotis, M., 2008. Recent advances in the *in vitro* cultivation and genetic manipulation of *Echinococcus multilocularis* metacestodes and germinal cells. Experimental Parasitology, 119, 506-515.

Brindley, P.J., Mitreva, M., Ghedin, E., Lustigman, S., 2009. Helminth genomics: The implications for human health. PLOS Neglected Tropical Diseases, 3, e538.

Brownlee, D., Fairweather, I., Johnston, C., Rogan, M., 1994. Immunocytochemical localization of serotonin (5-HT) in the nervous system of the hydatid organism, *Echinococcus granulosus*

(Cestoda, Cyclophyllidea). Parasitology, 109, 233-241.

Camicia, F., Herz, M., Prada, L., Kamenetzky, L., Simonetta, S., Cucher, M., Bianchi, J., Fernández, C., Brehm, K., Rosenzvit, M., 2013. The nervous and prenervous roles of serotonin in *Echinococcus* spp. International Journal for Parasitology, 43, 647-659.

Chow, C., Gauci, C.G., Cowman, A.F., Lightowlers, M.W., 2001. A gene family expressing a host-protective antigen of *Echinococcus granulosus*. Molecular and Biochemical Parasitology, 118, 83-88.

Chow, C., Gauci, C.G., Cowman, A.F., Lightowlers, M.W., 2004. *Echinococcus granulosus*: oncosphere-specific transcription of genes encoding a host-protective antigen. Experimental Parasitology, 106, 183-186.

Chu, Y., Corey, D.R., 2012. RNA sequencing: platform selection, experimental design, and data interpretation. Nucleic Acid Therapeutics, 22, 271-274.

Díaz, Á., Fernández, C., Pittini, Á., Seoane, P.I., Allen, J.E., Casaravilla, C., 2015. The laminated layer: Recent advances and insights into *Echinococcus* biology and evolution. Experimental Parasitology, 158, 23-30.

da Silva, C.M., Ferreira, H.B., Picón, M., Gorfinkiel, N., Ehrlich, R., Zaha, A., 1993. Molecular cloning and characterization of actin genes from *Echinococcus granulosus*. Molecular and Biochemical Parasitology, 60, 209-219.

Dang, Z., Yagi, K., Oku, Y., Kouguchi, H., Kajino, K., Matsumoto, J., Nakao, R., Wakaguri, H., Toyoda, A., Yin, H., 2012. A pilot study on developing mucosal vaccine against alveolar echinococcosis (AE) using recombinant tetraspanin 3: vaccine efficacy and immunology. PLOS Neglected Tropical Diseases, 6, e1570.

Dang, Z., Yagi, K., Oku, Y., Kouguchi, H., Kajino, K., Watanabe, J., Matsumoto, J., Nakao, R., Wakaguri, H., Toyoda, A., 2009. Evaluation of *Echinococcus multilocularis* tetraspanins as vaccine candidates against primary alveolar echinococcosis. Vaccine, 27, 7339-7345.

Davidson, R.K., Romig, T., Jenkins, E., Tryland, M., Robertson, L.J., 2012. The impact of globalisation on the distribution of *Echinococcus multilocularis*. Trends in Parasitology, 28, 239-247.

Delunardo, F., Ortona, E., Margutti, P., Perdicchio, M., Vacirca, D., Teggi, A., Sorice, M., Siracusano, A., 2010. Identification of a novel 19kDa *Echinococcus granulosus* antigen. Acta Tropica, 113, 42-47.

Deplazes, P., Eckert, J., 2001. Veterinary aspects of alveolar echinococcosis—a zoonosis of public health significance. Veterinary Parasitology, 98, 65-87.

Eckert, J., Gemmell, M.A., Meslin, F.-X., Pawłowski, Z.S., 2001. WHO/OIE Manual on Echinococcosis in Humans and Animals: A Public Health Problem of Global Concern. Pairs: OIE-WHO.

Eckert, J., Deplazes, P., 2004. Biological, epidemiological, and clinical aspects of echinococcosis, a zoonosis of increasing concern. Clinical Microbiology Reviews, 17, 107-135.

Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols, 2, 953-971.

Fairweather, I., McMullan, M., Johnston, C., Rogan, M., Hanna, R., 1994. Serotoninergic and peptidergic nerve elements in the protoscolex of *Echinococcus granulosus* (Cestoda, Cyclophyllidea). Parasitology Research, 80, 649-656.

Felleisen, R., Gottstein, B., 1993. *Echinococcus multilocularis*: molecular and immunochemical characterization of diagnostic antigen II/3-10. Parasitology, 107, 335-342.

Flisser, A., Gauci, C.G., Zoli, A., Martinez-Ocana, J., Garza-Rodriguez, A., Dominguez-Alpizar, J.L., Maravilla, P., Rodriguez-Canul, R., Avila, G., Aguilar-Vega, L., 2004. Induction of protection against porcine cysticercosis by vaccination with recombinant oncosphere antigens. Infection and Immunity, 72, 5292-5297.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research, 36, 3420-3435.

Garg, G., Ranganathan, S., 2011. In silico secretome analysis approach for next generation sequencing transcriptomic data. BMC Genomics, 12, 1.

Gauci, C., Merli, M., Muller, V., Chow, C., Yagi, K., Mackenstedt, U., Lightowlers, M.W., 2002. Molecular cloning of a vaccine antigen against infection with the larval stage of *Echinococcus multilocularis*. Infection and Immunity, 70, 3969-3972.

Gauci, C., Vural, G., Öncel, T., Varcasia, A., Damian, V., Kyngdon, C.T., Craig, P.S., Anderson, G.A., Lightowlers, M.W., 2008. Vaccination with recombinant oncosphere antigens reduces the susceptibility of sheep to infection with *Taenia multiceps*. International Journal for Parasitology, 38, 1041-1050.

Gonzalez, A.E., Gauci, C.G., Barber, D., Gilman, R.H., Tsang, V.C., Garcia, H.H., VERASTEGUI, M., Lightowlers, M.W., 2005. Vaccination of pigs to control human neurocysticercosis. The American Journal of Tropical Medicine and Hygiene, 72, 837-839.

Gottstein, B., Deplazes, P., Aubert, M., 1992. *Echinococcus multilocularis*: immunological study on the “Em2-positive” laminated layer during *in vitro* and *in vivo* post-oncospheral and larval development. Parasitology Research, 78, 291-297.

Gottstein, B., Hemphill, A., 2008. *Echinococcus multilocularis*: the parasite–host interplay. Experimental Parasitology, 119, 447-452.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29, 644-652.

Hancock, K., Pattabhi, S., Greene, R.M., Yushak, M.L., Williams, F., Khan, A., Priest, J.W., Levine, M.Z., Tsang, V.C., 2004. Characterization and cloning of GP50, a Taenia solium antigen diagnostic for cysticercosis. Molecular and Biochemical Parasitology, 133, 115-124.

Hancock, K., Pattabhi, S., Whitfield, F.W., Yushak, M.L., Lane, W.S., Garcia, H.H., Gonzalez, A.E., Gilman, R.H., Tsang, V.C., 2006. Characterization and cloning of T24, a *Taenia solium* antigen diagnostic for cysticercosis. Molecular and Biochemical Parasitology, 147, 109-117.

Harris, A., Heath, D., Lawrence, S., Shaw, R., 1989. *Echinococcus granulosus*: ultrastructure of epithelial changes during the first 8 days of metacestode development in vitro. International Journal for Parasitology, 19, 621-629.

Harrison, G., Heath, D., Dempster, R., Gauci, C., Newton, S., Cameron, W., Robinson, C., Lawrence, S., Lightowlers, M., Rickard, M., 1996. Identification and cDNA cloning of two novel low molecular weight host-protective antigens from *Taenia ovis* oncospheres. International Journal for Parasitology, 26, 195-204.

Heath, D., Lawrence, S., 1976. *Echinococcus granulosus*: development in vitro from oncosphere to immature hydatid cyst. Parasitology, 73, 417-423.

Heath, D., Smyth, J., 1970. *In vitro* cultivation of *Echinococcus granulosus*, *Taenia hydatigena*, *T. ovis*, *T. pisiformis* and *T. serialis* from oncosphere to cystic larva. Parasitology, 61, 329-343.

Hewitson, J.P., Grainger, J.R., Maizels, R.M., 2009. Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. Molecular and Biochemical Parasitology, 167, 1-11.

Holcman, B., Heath, D., Shaw, R., 1994. Ultrastructure of oncosphere and early stages of metacestode development of *Echinococcus granulosus*. International Journal for Parasitology, 24, 623-635.

Holcman, B., Heath, D.D., 1997. The early stages of *Echinococcus granulosus* development. Acta Tropica, 64, 5-17.

Huang, F., Dang, Z., Suzuki, Y., Horiuchi, T., Yagi, K., Kouguchi, H., Irie, T., Kim, K., Oku, Y., 2016. Analysis on Gene Expression Profile in Oncospheres and Early Stage Metacestodes from *Echinococcus multilocularis*. PLOS Neglected Tropical Diseases, 10, e0004634.

Ioppolo, S., Notargiacomo, S., Profumo, E., Franchi, C., Ortona, E., Rigano, R., Siracusano, A., 1996. Immunological responses to antigen B from *Echinococcus granulosus* cyst fluid in hydatid patients. Parasite Immunology, 18, 571-578.

Jabbar, A., Crawford, S., Młocicki, D., Świderski, Z.P., Conn, D.B., Jones, M.K., Beveridge, I., Lightowlers, M.W., 2010. Ultrastructural reconstruction of *Taenia ovis* oncospheres from serial sections. International Journal for Parasitology, 40, 1419-1431.

Jiang, L., Zhang, Y.-g., Liu, M.-x., Feng, Z., 2012. Analysis on the reactivity of five subunits of antigen B family in serodiagnosis of echinococcosis. Experimental Parasitology, 131, 85-91.

Johnson, K., Harrison, G., Lightowlers, M., O'hoy, K., Cougle, W., Dempster, R., Lawrence, S., Vinton, J., Heath, D., Rickard, M., 1989. Vaccination against ovine cysticercosis using a defined recombinant antigen. Nature, 338, 585 - 587

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics, 30, 1236-1240.

Käll, L., Krogh, A., Sonnhammer, E.L., 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Research, 35, W429-W432.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2015. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Aesearch, 44, D457-D462.

Katoh, Y., Kouguchi, H., Matsumoto, J., Goto, A., Suzuki, T., Oku, Y., Yagi, K., 2008. Characterization of emY162 encoding an immunogenic protein cloned from an adult worm-specific cDNA library of *Echinococcus multilocularis*. Biochimica et Biophysica Acta (BBA)-General Subjects, 1780, 1-6.

Kouguchi, H., Irie, T., Matsumoto, J., Nakao, R., Sugano, Y., Oku, Y., Yagi, K., 2016. The timing of worm exclusion in dogs repeatedly infected with the cestode *Echinococcus multilocularis.* Journal of Helminthology, 1-7.

Kouguchi, H., Matsumoto, J., Katoh, Y., Suzuki, T., Oku, Y., Yagi, K., 2010. *Echinococcus multilocularis*: two-dimensional Western blotting method for the identification and expression analysis of immunogenic proteins in infected dogs. Experimental Parasitology 124, 238-243.

Koziol, U., Brehm, K., 2015. Recent advances in *Echinococcus* genomics and stem cell research. Veterinary Parasitology, 213, 92-102.

Koziol, U., Krohne, G., Brehm, K., 2013. Anatomy and development of the larval nervous system in *Echinococcus multilocularis*. Frontiers in Zoology, 10, 1.

Koziol, U., Rauschendorfer, T., Rodríguez, L.Z., Krohne, G., Brehm, K., 2014. The unique stem cell system of the immortal larva of the human parasite *Echinococcus multilocularis*. EvoDevo, 5, 10.

Lacey, E., 1990. Mode of action of benzimidazoles. Parasitology Today, 6, 112-115.

Lascano, E., Coltorti, E., Varela-Diaz, V., 1975. Fine structure of the germinal membrane of *Echinococcus granulosus* cysts. The Journal of Parasitology, 853-860.

Lasda, E.L., Blumenthal, T., 2011. Trans‐splicing. Wiley Interdisciplinary Reviews: RNA 2, 417-434.

Law, C.W., Alhamdoosh, M., Su, S., Smyth, G.K., Ritchie, M.E., 2016. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Research, 5, 1408.

Law, R.H., Zhang, Q., McGowan, S., Buckle, A.M., Silverman, G.A., Wong, W., Rosado, C.J., Langendorf, C.G., Pike, R.N., Bird, P.I., 2006. An overview of the serpin superfamily. Genome Biology, 7, 216.

Lethbridge, R., 1980. The biology of the oncosphere of cyclophyllidean cestodes. In: Helminthological Abstracts, Series A, 49, 59-72.

Levine, M.Z., Calderón S, J., Wilkins, P.P., Lane, W.S., Asara, J.M., Hancock, K., Gonzalez, A.E., Garcia, H.H., Gilman, R.H., Tsang, V.C., 2004. Characterization, cloning, and expression of two diagnostic antigens for Taenia solium tapeworm infection. Journal of Parasitology, 90, 631-638.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics, 25, 2078-2079.

Lightowlers, M., Heath, D., 2004. Immunity and vaccine control of *Echinococcus granulosus* infection in animal intermediate hosts. Parassitologia, 46, 27-31.

Lightowlers, M., Lawrence, S., Gauci, C., Young, J., Ralston, M., Maas, D., Heath, D., 1996a. Vaccination against hydatidosis using a defined recombinant antigen. Parasite Immunology, 18, 457-462.

Lightowlers, M.W., Rolfe, R., Gauci, C.G., 1996b. *Taenia saginata*: vaccination against cysticercosis in cattle with recombinant oncosphere antigens. Experimental Parasitology, 84, 330-338.

Liu, W., Zhao, R., McFarland, C., Kieft, J., Niedzwiecka, A., Jankowska-Anyszka, M., Stepinski, J., Darzynkiewicz, E., Jones, D.N., Davis, R.E., 2009. Structural insights into parasite eIF4E binding specificity for m7G and m2, 2, 7G mRNA caps. Journal of Biological Chemistry, 284, 31336-31349.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods, 25, 402-408.

Louvet‐Vallée, S., 2000. ERM proteins: from cellular architecture to cell signaling. Biology of the Cell, 92, 305-316.

Mühlschlegel, F., Frosch, P., Castro, A., Apfel, H., Müller, A., Frosch, M., 1995. Molecular cloning and characterization of an *Echinococcus multilocularis* and *Echinococcus granulosus* stress protein homologous to the mammalian 78 kDa glucose regulated protein. Molecular and Biochemical Parasitology, 74, 245-250.

Mamuti, W., Sako, Y., Bart, J.-M., Nakao, M., Ma, X., Wen, H., Ito, A., 2007. Molecular characterization of a novel gene encoding an 8-kDa-subunit of antigen B from *Echinococcus granulosus* genotypes 1 and 6. Parasitology International, 56, 313-316.

Margutti, P., Ortona, E., Vaccari, S., Barca, S., Rigano, R., Teggi, A., Muhschlegel, F., Frosch, M., Siracusano, A., 1999. Cloning and expression of a cDNA encoding an elongation factor 1β/δ

protein from *Echinococcus granulosus* with immunogenic activity. Parasite Immunology, 21, 485-492.

Matsumoto, J., Kouguchi, H., Oku, Y., Yagi, K., 2010. Primary alveolar echinococcosis: course of larval development and antibody responses in intermediate host rodents with different genetic backgrounds after oral infection with eggs of *Echinococcus multilocularis*. Parasitology International, 59, 435-444.

Maule, A.G., Marks, N.J., 2006. Parasitic flatworms: molecular biology, biochemistry, immunology and physiology. Cambridge:CABI.

McManus, D., 2009. Reflections on the biochemistry of *Echinococcus*: past, present and future. Parasitology, 136, 1643-1652.

Mejri, N., Gottstein, B., 2009. *Echinococcus multilocularis* metacestode metabolites contain a cysteine protease that digests eotaxin, a CC pro-inflammatory chemokine. Parasitology Research, 105, 1253-1260.

Merckelbach, A., Ruppel, A., 2007. Biochemical properties of an intracellular serpin from *Echinococcus multilocularis*. Molecular and Biochemical Parasitology, 156, 84-88.

Merckelbach, A., Wager, M., Lucius, R., 2003. Analysis of cDNAs coding for immunologically dominant antigens from an oncosphere-specific cDNA library of *Echinococcus multilocularis*. Parasitology Research, 90, 493-501.

Monteiro, K.M., de Carvalho, M.O., Zaha, A., Ferreira, H.B., 2010. Proteomic analysis of the *Echinococcus granulosus* metacestode during infection of its intermediate host. Proteomics, 10, 1985-1999.

Morin, R.D., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T.J., McDonald, H., Varhol, R., Jones, S.J., Marra, M.A., 2008. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques, 45, 81.

Nono, J.K., Pletinckx, K., Lutz, M.B., Brehm, K., 2012. Excretory/secretory-products of *Echinococcus multilocularis* larvae induce apoptosis and tolerogenic properties in dendritic cells in vitro. PLOS Neglected Tropical Diseases, 6, e1516.

Oliveira, G.C., Kemp, W.M., 1995. Cloning of two actin genes from *Schistosoma mansoni.* Molecular and Biochemical Parasitology, 75, 119-122.

Oriol, R., Williams, J., Perez Esandi, M., Oriol, C., 1971. Purification of lipoprotein antigens of *Echinococcus granulosus* from sheep hydatid fluid. American Journal of Tropical Medicine and Hygiene, 20, 569-574.

Ortona, E., Margutti, P., Delunardo, F., Vaccari, S., Rigano, R., Profumo, E., Buttari, B., Teggi, A., Siracusano, A., 2003. Molecular and immunological characterization of the C‐terminal region of a new *Echinococcus granulosus* heat shock protein 70. Parasite Immunology, 25, 119-126.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature Genetics, 40, 1413-1415.

Pan, W., Shen, Y., Han, X., Wang, Y., Liu, H., Jiang, Y., Zhang, Y., Wang, Y., Xu, Y., Cao, J., 2014. Transcriptome profiles of the protoscoleces of *Echinococcus granulosus* reveal that excretory-secretory products are essential to metabolic adaptation. PLOS Neglected Tropical Diseases 8, e3392.

Paredes, R., Jimenez, V., Cabrera, G., Iragüen, D., Galanti, N., 2007. Apoptosis as a possible mechanism of infertility in *Echinococcus granulosus* hydatid cysts. Journal of Cellular Biochemistry, 100, 1200-1209.

Parkinson, J., Wasmuth, J.D., Salinas, G., Bizarro, C.V., Sanford, C., Berriman, M., Ferreira, H.B., Zaha, A., Blaxter, M.L., Maizels, R.M., 2012. A transcriptomic analysis of Echinococcus granulosus larval stages: implications for parasite biology and host adaptation. PLOS PLOS Neglected Tropical Diseases, 6, e1897.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods, 8, 785-786.

Pruitt, K.D., Tatusova, T., Maglott, D.R., 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research, 35, D61-D65.

Ranasinghe, S.L., Fischer, K., Zhang, W., Gobert, G.N., McManus, D.P., 2015. Cloning and characterization of two potent Kunitz type protease inhibitors from *Echinococcus granulosus*. PLOS Neglected Tropical Diseases, 9, e0004268.

Rawlings, N.D., Barrett, A.J., Finn, R., 2015. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Research, 44, D343-d350.

Rawlings, N.D., Morton, F.R., 2008. The MEROPS batch BLAST: a tool to detect peptidases and their non-peptidase homologues in a genome. Biochimie, 90, 243-259.

Reid, W.M., 1948. Penetration glands in cyclophyllidean onchospheres. Transactions of the American Microscopical Society 67, 177-182.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140.

Rogan, M., Craig, P., Zehyle, E., Masinde, G., Wen, H., Zhou, P., 1992. In vitro killing of taeniid oncospheres, mediated by human sera from hydatid endemic areas. Acta Tropica, 51, 291-296.

Sakamoto, T., Sugimura, M., 1970. Studies on echinococcosis XXIII: electron microscopical observations on histogenesis of larval *Echonococcus multilocularis*. Japanese Journal of Veterinary Research, 18, 131-144.

Sako, Y., Nakao, M., Nakaya, K., Yamasaki, H., Gottstein, B., Lightowers, M.W., Schantz, P.M., Ito, A., 2002. Alveolar echinococcosis: characterization of diagnostic antigen Em18 and serological evaluation of recombinant Em18. Journal of Clinical Microbiology, 40, 2760-2765.

Santivanez, S., Hernandez-Gonzalez, A., Chile, N., Oleaga, A., Arana, Y., Palma, S., Verastegui, M., Gonzalez, A., Gilman, R., Garcia, H., 2010. Proteomic study of activated *Taenia solium* oncospheres. Molecular and Biochemical Parasitology, 171, 32-39.

Seigneuret, M., Conjeaud, H., Zhang, H.-T., Kong, X.-P., 2013. Structural bases for tetraspanin functions. Berditchevski F., Rubinstein E. Tetraspanins. Heidelberg: Springer Netherlands. 1-29.

Siles-Lucas, M., Merli, M., Mackenstedt, U., Gottstein, B., 2003. The *Echinococcus multilocularis* 14-3-3 protein protects mice against primary but not secondary alveolar echinococcosis. Vaccine, 21, 431-439.

Siracusano, A., Delunardo, F., Teggi, A., Ortona, E., 2011. Host-parasite relationship in cystic echinococcosis: an evolving story. Clinical and Developmental Immunology 2012, 639362.

Smith, S., Richards, K., 1993. Ultrastructure and microanalyses of the calcareous corpuscles of the protoscoleces of *Echinococcus granulosus*. Parasitology Research, 79, 245-250.

Sonnhammer, E.L., Von Heijne, G., Krogh, A., 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Ismb, 6, 175-182.

Spiliotis, M., Mizukami, C., Oku, Y., Kiss, F., Brehm, K., Gottstein, B., 2010. *Echinococcus multilocularis* primary cells: improved isolation, small-scale cultivation and RNA interference. Molecular and Biochemical Parasitology, 174, 83-87.

Spiliotis, M., Tappe, D., Sesterhenn, L., Brehm, K., 2004. Long-term in vitro cultivation of *Echinococcus multilocularis* metacestodes under axenic conditions. Parasitology Research, 92, 430-432.

Stojkovic, M., Junghanss, T., 2012. Cystic and alveolar echinococcosis. Handbook of Clinical Neurology, 114, 327-334.

Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science, 321, 956-960.

Swiderski, Z., 1983. *Echinococcus granulosus*: hook-muscle systems and cellular organisation of infective oncospheres. International Journal for Parasitology, 13, 289-299.

Swiderski, Z., Miquel, J., Azzouz-Maache, S., Petavy, A.F., 2016. *Echinococcus multilocularis* (Cestoda, Cyclophyllidea, Taeniidae): oncospheral hook morphogenesis. Parasitology Research, 115, 3715-3721.

Tappe, D., Kern, P., Frosch, M., Kern, P., 2010. A hundred years of controversy about the taxonomic

status of *Echinococcus* species. Acta Tropica, 115, 167-174.

Torgerson, P.R., Keller, K., Magnotta, M., Ragland, N., 2010. The global burden of alveolar echinococcosis. PLOS Neglected Tropical Diseases, 4, e722.

Tran, M.H., Pearson, M.S., Bethony, J.M., Smyth, D.J., Jones, M.K., Duke, M., Don, T.A., McManus, D.P., Correa-Oliveira, R., Loukas, A., 2006. Tetraspanins on the surface of *Schistosoma mansoni* are protective antigens against schistosomiasis. Nature Medicine, 12, 835-840.

Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics, 25, 1105-1111.

Tsai, I.J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K.L., Tracey, A., Bobes, R.J., Fragoso, G., Sciutto, E., 2013. The genomes of four tapeworm species reveal adaptations to parasitism. Nature, 496, 57-63.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Hieter, P., Vogelstein, B., Kinzler, K.W., 1997. Characterization of the yeast transcriptome. Cell, 88, 243-251.

Virginio, V., Hernandez, A., Rott, M., Monteiro, K., Zandonai, A., Nieto, A., Zaha, A., Ferreira, H., 2003. A set of recombinant antigens from *Echinococcus granulosus* with potential for use in the immunodiagnosis of human cystic hydatid disease. Clinical & Experimental Immunology, 132, 309-315.

Wang, S., Wei, W., Cai, X., 2015a. Genome-wide analysis of excretory/secretory proteins in Echinococcus multilocularis: insights into functional characteristics of the tapeworm secretome. Parasites & Vectors 8, 1.

Wang, Y., Xiao, D., Shen, Y., Han, X., Zhao, F., Li, X., Wu, W., Zhou, H., Zhang, J., Cao, J., 2015b. Proteomic analysis of the excretory/secretory products and antigenic proteins of *Echinococcus granulosus* adult worms from infected dogs. BMC Veterinary Research, 11, 119.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10, 57-63.

Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Research, 34, D187-D191.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C.-Y., Wei, L., 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Research, 39, W316-W322.

Yassour, M., Kaplan, T., Fraser, H.B., Levin, J.Z., Pfiffner, J., Adiconis, X., Schroth, G., Luo, S., Khrebtukova, I., Gnirke, A., 2009. Ab initio construction of a eukaryotic transcriptome by

massively parallel mRNA sequencing. Proceedings of the National Academy of Sciences, 106, 3264-3269.

Zügeli, U., Kaufmann, S.H., 1999. Immune response against heat shock proteins in infectious diseases. Immunobiology, 201, 22-35.

Zhang, W., Li, J., You, H., Zhang, Z., Turson, G., Loukas, A., Mcmanus, D.P., 2003. Short report: Echinococcus granulosus from Xinjiang, PR China: cDNAs encoding the EG95 vaccine antigen are expressed in different life cycle stages and are conserved in the oncosphere. The American Journal of Tropical Medicine and Hygiene, 68, 40-43.

Zhang, W., Ross, A.G., McManus, D.P., 2008. Mechanisms of immunity in hydatid disease: implications for vaccine development. The Journal of Immunology, 181, 6679-6685.

Zhang, W., Li J., Lin, R., Wen, H, McManus, D. P., 2012. Recent Advances in the Immunology and Serological Diagnosis of Echinococcosis, Serological Diagnosis of Certain Human, Animal and Plant Diseases. 2012; Dr. Moslih Al-Moslih (Ed.), ISBN: 978-953-51-0370-7, InTech, Available from: http://www.intechopen.com/books/serological-diagnosis-of-certain-human-animal-and-plant-diseases/recentadvances-in-the-serological-diagnosis-of-echinococcosis.

Zheng, H., Zhang, W., Zhang, L., Zhang, Z., Li, J., Lu, G., Zhu, Y., Wang, Y., Huang, Y., Liu, J., 2013. The genome of the hydatid tapeworm *Echinococcus granulosus*. Nature Genetics, 45, 1168-1175.

Zheng, Y., 2012. The definition of *Echinococcus multilocularis* differentially expressed molecules using deep sequencing. University of Nottingham, Doctor Thesis.

Zhu, Y., Ren, J., Da'dara, A., Harn, D., Xu, M., Si, J., Yu, C., Liang, Y., Ye, P., Yin, X., 2004. The protective effect of a *Schistosoma japonicum* Chinese strain 23kDa plasmid DNA vaccine in pigs is enhanced with IL-12. Vaccine, 23, 78-83.

Appendix I

Predicted protease of *E. multilocularis*

Transcript ID	Description	Protease ID	KO ID	pNonc1	pNonc2	pAonc	s4Wmet	p4Wmet	pCmet	s16Wmet
EmuJ_000970500.1	cathepsin d lysosomal aspartyl protease	A01A	K01379	64	97	96	187	88	285	196
EmuJ_000780700.1	presenilin	A22A	K04505	54	13	44	17	25	56	44
EmuJ_000084000.1	Minor histocompatibility antigen H13	A22B	K09595	337	515	217	197	271	314	182
EmuJ_000955800.1	signal peptide peptidase 2A	A22B	K09596	35	19	51	27	37	48	30
EmuJ_000898400.1	protein ddi1 2	A28A	K11885	67	62	93	49	46	133	49
EmuJ_000201800.1	RNA directed DNA polymerase reverse transcriptase	A28B	—	2	0	0	10	0	0	1
EmuJ_000696400.1	conserved hypothetical protein	A28B	—	2	0	2	0	0	0	0
EmuJ_000676400.1	RNA directed DNA polymerase reverse transcriptase	A28B/A02D	—	0	0	3	0	0	0	0
EmuJ_000790200.1	cathepsin b	C01A	K01363	8	3	139	417	299	1085	336
EmuJ_000477200.1	cathepsin L	C01A	K01365	82	63	401	125	157	973	211
EmuJ_000790300.1	cathepsin b	C01A	K01363	8	7	61	32	24	36	24
EmuJ_000989200.1	cathepsin l cysteine peptidase	C01A/I29	K01365	73	73	101	32	77	133	66

EmuJ_000967900.1	cathepsin L	C01A/I29	—	1	0	0	0	0	4	10
EmuJ_000654500.1	cysteine protease	C01A/I29	K01365	0	0	1	0	0	0	0
EmuJ_000654100.1	cathepsin l cysteine peptidase	C01A/I29	K01365	0	0	0	0	0	0	0
EmuJ_000654200.1	cathepsin L cysteine protease	C01A/I29	K01365	0	0	0	0	0	0	0
EmuJ_000654600.1	cysteine protease	C01A/I29	K01365	0	0	0	0	0	0	0
EmuJ_000654800.1	cathepsin L cysteine protease	C01A/I29	K01365	0	0	0	0	0	0	0
EmuJ_000719700.1	calpain	C02A	K08585	128	74	193	252	314	1242	400
EmuJ_000911200.1	calpain A	C02A	K08585	206	67	76	284	339	1134	304
EmuJ_000319100.1	family C2 unassigned peptidase (C02 family)	C02A	K08585	21	11	43	71	75	296	72
EmuJ_001000650.1	calpain 7 C02 family	C02A	K08576	73	17	66	51	52	76	36
EmuJ_000205500.1	calpain 5	C02A	K08574	53	22	36	20	42	55	37
EmuJ_002206900.1	family C2 unassigned peptidase (C02 family)	C02A	—	0	0	3	0	50	0	6
EmuJ_000253150.1	hypothetical protein	C02A	—	1	1	4	10	9	5	23
EmuJ_000765200.1	Transglutaminase	C110	—	7	0	25	58	78	346	49
EmuJ_000836300.1	ubiquitin protein ligase BRE1	C110	—	19	15	30	66	48	169	43

EmuJ_000656610.1	putative aldehyde dehydrogenase	C110	—	13	0	6	9	15	9	9
EmuJ_000765300.1	Transglutaminase	C110	—	1	0	3	7	2	13	5
EmuJ_001029600.1	ubiquitin carboxyl terminal hydrolase isozyme	C12	K05609	199	364	204	257	271	419	147
EmuJ_000141500.1	ubiquitin carboxyl terminal hydrolase isozyme	C12	K05610	149	163	143	114	172	528	86
EmuJ_000151200.1	ubiquitin carboxyl terminal hydrolase BAP1	C12	K08588	15	5	9	14	20	31	16
EmuJ_000869600.1	GPI anchor transamidase	C13	K05290	112	196	112	81	90	126	122
EmuJ_000462900.1	caspase 3 apoptosis cysteine peptidase	C14A	K02187	92	62	100	24	98	191	117
EmuJ_000417900.1	caspase 8	C14A	—	1	3	19	29	34	90	40
EmuJ_000113100.1	caspase 2	C14A	K02186	33	11	28	7	14	22	27
EmuJ_001067200.1	caspase	C14A	—	4	2	7	0	8	30	15
EmuJ_001067500.1	caspase 3	C14A	—	0	0	0	0	0	0	1
EmuJ_000217600.1	Peptidase C15 pyroglutamyl peptidase I	C15	K01304	253	305	133	39	46	26	23
EmuJ_000337400.1	Peptidase C19 ubiquitin carboxyl terminal hydrolase 2	C19	K11833	495	507	375	64	181	196	94
EmuJ_000837800.1	ubiquitin carboxyl terminal hydrolase 4	C19	K11835	89	133	141	122	101	123	111
EmuJ_000314400.1	expressed protein	C19	—	177	106	123	55	77	98	38

EmuJ_000580300.1	ubiquitin carboxyl terminal hydrolase 14	C19	K11843	65	67	141	21	53	124	51
EmuJ_000875300.1	ubiquitin carboxyl terminal hydrolase 7	C19	K11838	63	10	50	59	114	142	70
EmuJ_000167900.1	ubiquitin carboxyl terminal hydrolase 36	C19	K11855	100	25	71	39	62	128	52
EmuJ_000915600.1	ubiquitin carboxyl terminal hydrolase 12	C19	K11842	95	39	48	14	52	145	46
EmuJ_000627400.1	u4:u6.u5 tri snrnp associated protein 2	C19	K12847	98	25	63	29	35	116	71
EmuJ_000175500.1	ubiquitin specific peptidase c family	C19	K11841	38	10	79	36	52	125	51
EmuJ_000102100.1	Zinc finger MYND type	C19	—	66	61	68	39	52	55	46
EmuJ_000095700.1	ubiquitin carboxyl terminal hydrolase 5	C19	K11836	32	11	40	28	31	191	41
EmuJ_000957900.1	ubiquitin carboxyl terminal hydrolase 8	C19	K11839	80	51	68	41	40	30	51
EmuJ_000065800.1	ubiquitin carboxyl terminal hydrolase 4	C19	—	46	9	21	56	43	80	26
EmuJ_000199000.1	SET and MYND domain containing protein 3	C19	K11426	26	26	30	43	46	66	31
EmuJ_000711400.1	Ubiquitin carboxyl terminal hydrolase 48	C19	K11858	46	8	28	55	32	25	62
EmuJ_000696800.1	ubiquitin carboxyl terminal hydrolase	C19	K11840	39	9	35	42	45	39	19
EmuJ_000929000.1	ubiquitin carboxyl terminal hydrolase 16	C19	K11844	33	8	33	33	44	29	42
EmuJ_001005800.1	ubiquitin carboxyl terminal hydrolase 7	C19	K11838	26	3	20	31	41	45	36

EmuJ_000729100.1	ubiquitin specific protease 41	C19	K11833	85	27	22	11	13	27	12
EmuJ_000570500.1	ubiquitin carboxyl terminal hydrolase 20	C19	K11848	28	6	20	6	23	31	23
EmuJ_000042300.1	ubiquitin carboxyl terminal hydrolase 22	C19	K11366	28	8	8	23	12	12	13
EmuJ_000915300.1	ubiquitin carboxyl terminal hydrolase 4	C19	K11835	0	0	0	7	7	6	20
EmuJ_000695600.1	ubiquitin carboxyl terminal hydrolase 24	C19	—	9	3	3	3	4	3	6
EmuJ_001182600.1	ubiquitin carboxyl terminal hydrolase 47	C19	K11869	4	1	4	0	14	0	1
EmuJ_001164900.1	ubiquitin specific peptidase 42 C19 family	C19	—	0	0	0	0	0	0	0
EmuJ_001198000.1	GMP synthase glutamine hydrolyzing	C26	K01951	84	73	133	120	149	249	168
EmuJ_000849200.1	CTP synthase 1	C26	K01937	46	16	88	61	61	135	50
EmuJ_000097800.1	Glutamine:fructose 6 phosphate aminotransferase	C44	K00820	51	54	212	557	572	485	133
EmuJ_000644000.1	glutamate synthase	C44	K00264	6	1	55	282	188	557	68
EmuJ_000668900.1	asparagine synthetase domain containing protein	C44	—	57	28	128	64	57	94	128
EmuJ_001200300.1	hedgehog	C46	—	0	1	0	15	8	10	52
EmuJ_000304000.1	expressed protein	C48	K14764	97	65	104	90	114	57	93
EmuJ_000060300.1	sentrin specific protease 7	C48	K08596	53	29	94	35	41	53	58

EmuJ_000338300.1	sentrin specific protease 1	C48	K08592	34	11	14	27	19	54	28
EmuJ_001069800.1	sentrin specific protease 8	C48	K08597	6	13	13	24	0	24	17
EmuJ_002197000.1	sentrin specific protease	C48	K03345	0	0	0	0	0	0	0
EmuJ_000702900.1	separin	C50	K02365	18	7	19	27	19	47	35
EmuJ_000458100.1	cysteine protease atg4b	C54	K08342	70	24	65	95	59	48	58
EmuJ_000098800.1	Peptidase C54	C54	K08342	38	16	27	26	3	39	27
EmuJ_000135900.1	protein DJ 1	C56	K05687	481	845	468	896	747	1561	735
EmuJ_000999800.1	ES1 protein mitochondrial	C56	—	97	180	110	297	211	427	278
EmuJ_001071800.1	Zinc finger RanBP2 type	C64	K11862	56	10	14	29	41	85	21
EmuJ_000540000.1	Ubiquitin thioesterase otubain protein	C65	K09602	170	156	216	136	230	376	103
EmuJ_000586500.1	Ufm1 specific protease 2	C78	K01376	20	15	53	36	36	57	31
EmuJ_000799700.1	ufm1 specific protease 1	C78	K01376	17	13	35	0	32	25	14
EmuJ_000255400.1	phytochelatin synthase	C83	K05941	46	9	35	29	14	54	32
EmuJ_000728500.1	OTU domain containing protein 6B	C85A	K18342	158	73	174	69	97	101	59
EmuJ_001090200.1	OTU domain containing protein 5 A	C85A	K12655	47	22	99	50	39	134	70

EmuJ_000725600.1	OTU domain containing protein 3	C85A	K13717	8	0	34	11	35	29	26
EmuJ_000825650.1	zinc finger C2H2 type	C85B	K13719	14	16	43	58	67	62	34
EmuJ_001174800.1	Atxn3 protein	C86	K11863	99	92	230	114	129	157	158
EmuJ_000701300.1	josephin 2	C86	K15235	207	176	83	37	81	144	45
EmuJ_000220500.1	LAMA protein 2	C95	—	163	94	38	41	34	206	51
EmuJ_000438500.1	PPPDE peptidase domain containing protein 2	C97	—	45	44	88	99	76	173	60
EmuJ_000378500.1	PPPDE peptidase domain containing protein 1	C97	—	29	9	44	51	0	57	41
EmuJ_000061400.1	agrin	I01	K06254	0	0	1	16	15	24	27
EmuJ_000278400.1	expressed protein follistatin	I01	K04661	2	5	2	0	0	21	44
EmuJ_001136900.1	amyloid beta A4 protein	I02	—	60128	123301	16480	0	1429	15	0
EmuJ_000419100.1	Kunitz protease inhibitor	I02	—	284	666	483	22	11	22	47
EmuJ_001181950.1	Papilin	I02	—	41	60	215	102	123	352	104
EmuJ_000077800.1	Papilin	I02	—	282	290	59	20	30	60	25
EmuJ_001137100.1	expressed protein	I02	—	0	0	0	32	0	0	443
EmuJ_000929500.1	SPONdin extracellular matrix glycoprotein	I02	—	2	0	1	64	4	115	163

EmuJ_000534800.1	expressed protein	I02	—	53	30	80	0	0	18	137
EmuJ_000077700.1	Kunitz:Bovine pancreatic trypsin inhibitor	I02	—	122	109	29	0	0	0	1
EmuJ_000302900.1	expressed protein	I02	—	0	1	3	56	0	56	86
EmuJ_001136500.1	WAP kazal immunoglobulin kunitz and NTR	I02	—	35	102	6	0	0	0	0
EmuJ_000225800.1	Papilin	I02	—	0	0	5	6	2	18	10
EmuJ_001137200.1	collagen alpha 3VI chain	I02	—	0	0	17	0	0	0	3
EmuJ_001137000.1	Kunitz protease inhibitor	I02	—	0	14	0	0	0	0	0
EmuJ_000548800.1	collagen alpha 1XXVIII chain	I02	—	0	0	0	0	0	0	8
EmuJ_000549400.1	collagen alpha 1XXVIII chain	I02	—	0	0	0	0	0	0	3
EmuJ_001136700.1	proteinase inhibitor I2 Kunitz metazoa	I02	—	0	0	0	0	0	0	1
EmuJ_001136800.1	WAP kazal immunoglobulin kunitz and NTR	I02	—	0	0	0	0	0	0	1
EmuJ_000419200.1	trypsin inhibitor	I02	—	0	0	0	0	0	0	1
EmuJ_001136600.1	Kunitz protease inhibitor	I02	—	0	0	0	0	0	0	0
EmuJ_001137300.1	collagen alpha 3VI chain	I02	—	0	0	0	0	0	0	0
EmuJ_001137400.1	collagen alpha 3VI chain	I02	—	0	0	0	0	0	0	0

EmuJ_000255800.1	egf domain protein	I02/I08	—	3	0	12	26	25	135	47
EmuJ_000488100.1	60S ribosomal protein L38	I04	K02923	456	1554	735	491	570	799	388
EmuJ_000824100.1	Estrogen regulated protein EP45	I04	K13963	38	11	251	90	279	324	25
EmuJ_001193100.1	serine protease inhibitor	I04	K13963	318	564	45	0	12	27	8
EmuJ_001193200.1	serine protease inhibitor	I04	K13963	218	296	38	0	12	12	3
EmuJ_000824000.1	Estrogen regulated protein EP45	I04	—	9	19	50	27	7	3	30
EmuJ_000068100.1	laminin	I15	—	0	0	5	17	16	46	25
EmuJ_001132400.1	laminin	I15	K05637	17	6	7	11	13	23	22
EmuJ_000119200.1	neuroendocrine protein 7b2	I21	—	45	43	57	8	4	16	14
EmuJ_000159200.1	Cystatin B Stefin B	I25A	K13907	195	320	290	357	249	512	477
EmuJ_000698600.1	inhibitor of apoptosis protein	I32	—	412	227	179	1023	849	2664	990
EmuJ_001004100.1	baculoviral IAP repeat containing protein	I32	K08731	40	14	24	0	16	16	30
EmuJ_000641100.1	alpha 2 macroglobulin	I39	—	45	16	14	46	62	311	63
EmuJ_000610100.1	cd109 antigen	I39	—	1	0	2	0	1	9	3
EmuJ_000060000.1	mitochondrial ribosomal protein L38	I51	K17419	112	27	105	48	121	157	76

EmuJ_000677600.1	phospholipase A2 receptor 1	I63	K06560	67	16	60	150	140	200	79
EmuJ_000392700.1	protein jagged 2	I84	—	134	33	141	24	24	15	40
EmuJ_000951400.1	stomatin protein 2	I87	—	86	63	122	110	135	123	80
EmuJ_000764700.1	mechanosensory protein 2	I87	K17286	1	0	1	30	30	49	52
EmuJ_000194700.1	stomatin	I87	K17286	7	2	1	14	29	0	7
EmuJ_000382000.1	mechanosensory protein 2	I87	K17286	5	2	1	16	4	8	12
EmuJ_000450500.1	MEChanosensory abnormality family member	I87	K17286	13	8	11	6	0	0	6
EmuJ_000469900.1	mechanosensory protein 2	I87	K17286	0	0	0	0	0	12	29
EmuJ_000205700.1	mechanosensory protein 2	I87	K17286	1	1	3	0	0	0	22
EmuJ_000523300.1	mechanosensory protein 2	I87	K17286	0	0	0	0	14	0	8
EmuJ_000636500.1	frizzled	I93	K02842	5	2	4	25	4	60	50
EmuJ_000682100.1	frizzled	I93	K02432	8	1	5	23	8	16	46
EmuJ_000085700.1	frizzled 10	I93	K02842	0	0	1	17	0	0	54
EmuJ_001023000.1	secreted frizzled protein 5	I93	—	1	3	3	0	0	7	26
EmuJ_000838700.1	secreted frizzled protein	I93	K02176	1	0	3	6	0	3	16

EmuJ_000996400.1	frizzled 5	I93	K02375	3	2	5	4	0	2	11
EmuJ_000438200.1	frizzled 4	I93	—	0	0	2	0	0	2	1
EmuJ_000972400.1	leukotriene A 4 hydrolase	M01	K01254	103	60	78	130	79	262	99
EmuJ_000101000.1	transcription initiation factor TFIID subunit 2	M01	K03128	26	7	32	19	28	44	52
EmuJ_000350500.1	puromycin sensitive aminopeptidase	M01	K08776	2	0	8	42	30	92	16
EmuJ_000356700.1	puromycin sensitive aminopeptidase	M01	K08776	5	1	1	21	20	20	24
EmuJ_001105200.1	puromycin sensitive aminopeptidase	M01	K08776	2	0	3	3	12	50	12
EmuJ_001063900.1	aminopeptidase N	M01	K11140	13	3	12	11	10	7	16
EmuJ_000350600.1	puromycin sensitive aminopeptidase	M01	K08776	0	0	0	3	3	8	3
EmuJ_000350300.1	puromycin sensitive aminopeptidase	M01	K08776	0	0	0	0	0	0	1
EmuJ_001105000.1	puromycin sensitive aminopeptidase	M01	K08776	0	0	0	0	0	0	0
EmuJ_000466300.1	oligopeptidase	M03A	K01414	59	15	33	53	41	99	56
EmuJ_000491800.1	mitochondrial intermediate peptidase	M03A	K01410	15	3	16	35	50	45	27
EmuJ_000839500.1	invadolysin M08 family	M08	K13539	15	7	30	12	20	6	17
EmuJ_000939100.1	matrix metallopeptidase 7 M10 family	M10A	—	0	0	4	29	29	39	26

EmuJ_000640700.1	Tolloid protein 1	M12A	K09608	0	0	3	42	35	14	29
EmuJ_001195900.1	astacin protein	M12A	—	1	0	2	5	0	0	15
EmuJ_000832900.1	subfamily M12B unassigned peptidase	M12B	—	99	27	41	52	59	139	65
EmuJ_001069300.1	disintegrin and metalloproteinase	M12B	K06704	21	4	12	45	31	68	28
EmuJ_000347500.1	adam	M12B	—	2	1	7	20	18	54	52
EmuJ_000892800.1	adam 17 protease	M12B	K06059	7	2	25	14	12	11	19
EmuJ_001046600.1	a disintegrin and metalloproteinase with	M12B	—	7	5	27	6	4	5	5
EmuJ_000892700.1	Blood coagulation inhibitor Disintegrin	M12B	K06059	0	0	0	0	0	4	1
EmuJ_001177500.1	endothelin converting enzyme 1	M13	K01415	19	11	18	50	58	201	48
EmuJ_000814400.1	subfamily M14A unassigned peptidase	M14A	—	10	11	6	0	33	42	35
EmuJ_000421900.1	Zinc carboxypeptidase family protein	M14B	K01294	5	7	4	96	32	80	68
EmuJ_000352300.1	cytosolic carboxypeptidase 1	M14B	—	6	2	9	10	6	11	23
EmuJ_001068000.1	cytosolic carboxypeptidase protein 5	M14B	—	0	0	1	0	0	1	1
EmuJ_000765600.1	insulin degrading enzyme	M16A	K01408	36	22	100	40	27	148	52
EmuJ_002210500.1	nardilysin	M16A	—	0	0	0	0	0	0	0

EmuJ_000113700.1	mitochondrial processing peptidase beta subunit	M16B	K17732	271	191	439	391	470	899	355
EmuJ_000091700.1	Cytochrome b c1 complex subunit 2	M16B	K00415	232	168	316	281	207	615	220
EmuJ_000058800.1	mitochondrial processing peptidase	M16B	K01412	41	24	70	26	61	99	65
EmuJ_000921800.1	nardilysin	M16B	K01411	43	9	28	22	22	42	40
EmuJ_000935700.1	nardilysin	M16B/M16A	K01411	43	12	57	69	47	119	72
EmuJ_000503400.1	presequence protease mitochondrial	M16C	K06972	141	50	108	89	125	237	102
EmuJ_001133600.1	leucine aminopeptidase protein	M17	K11142	72	63	157	121	108	315	108
EmuJ_001031700.1	leucyl aminopeptidase	M17	K01255	12	12	86	79	59	69	27
EmuJ_000855000.1	leucyl aminopeptidase	M17	K01255	4	1	33	15	22	37	21
EmuJ_000374300.1	leucyl aminopeptidase	M17	K01255	1	0	1	0	0	0	0
EmuJ_001083400.1	aspartyl aminopeptidase	M18	K01267	88	33	65	84	99	216	74
EmuJ_000675100.2	0	M20A	K14677	5	7	15	25	12	87	42
EmuJ_000675100.1	aminoacylase 1	M20A	K14677	0	0	0	0	0	0	0
EmuJ_000992300.1	cytosolic non-specific dipeptidase	M20F	K08660	17	11	55	11	32	16	14
EmuJ_000820900.1	methionyl aminopeptidase 2	M24A	K01265	99	70	173	102	106	100	103

EmuJ_000923200.1	methionyl aminopeptidase 1 M24 family	M24A	K01265	27	15	70	13	29	132	34
EmuJ_000651700.1	methionine aminopeptidase type I	M24A	K01265	25	23	12	0	8	21	12
EmuJ_000025900.1	methionine aminopeptidase 1	M24A	—	3	8	0	0	0	0	0
EmuJ_000025800.1	methionyl aminopeptidase 1 M24 family	M24A	K01265	0	0	0	0	0	0	0
EmuJ_000864100.1	xaa pro aminopeptidase	M24B	K01262	95	54	122	164	70	166	200
EmuJ_000621300.1	xaa Pro dipeptidase	M24B	K14213	39	23	26	33	22	77	91
EmuJ_001034400.1	Xaa Pro aminopeptidase 3	M24B	K01262	53	22	37	46	23	32	30
EmuJ_000921400.1	xaa Pro dipeptidase	M24B	K14213	5	0	2	11	0	36	12
EmuJ_000781200.1	Proliferation associated protein 2G4	M24X	—	131	322	313	49	187	123	108
EmuJ_001138800.1	FACT complex subunit SPT16	M24X	—	43	19	48	45	51	99	57
EmuJ_002197100.1	proliferation-associated protein 2g4	M24X	—	0	0	0	0	0	6	1
EmuJ_000908900.1	n acetylated alpha linked acidic dipeptidase 2	M28B	K01301	0	0	0	4	7	128	119
EmuJ_000253000.1	glutaminyl peptide cyclotransferase	M28X	K00683	142	40	79	88	62	75	76
EmuJ_000890300.1	endoplasmic reticulum metallopeptidase 1	M28X	—	78	19	35	46	99	123	37
EmuJ_000825400.1	nicalin	M28X	—	51	23	53	34	33	100	54

EmuJ_000953400.1	dihydropyrimidinase	M38	K01464	101	42	98	146	225	1309	222
EmuJ_000953200.1	dihydropyrimidinase 2	M38	—	0	0	4	24	14	104	92
EmuJ_001045500.1	imidazolonepropionase	M38	K01468	3	4	0	12	25	59	46
EmuJ_001020900.1	dihydropyrimidinase protein 4	M38	—	14	7	19	9	2	9	20
EmuJ_000616300.1	afg3 protein 2	M41	K08956	90	35	77	128	84	224	106
EmuJ_000847400.1	ATP dependent zinc metalloprotease YME1	M41	K08955	93	35	95	63	63	124	108
EmuJ_000927200.1	paraplegin	M41	K09552	70	14	37	43	94	85	70
EmuJ_001023600.1	caax prenyl protease 1	M48A	K06013	47	62	103	61	133	126	85
EmuJ_001028100.1	dipeptidyl peptidase 3	M49	K01277	239	344	234	120	179	320	151
EmuJ_000468800.1	membrane bound transcription factor site 2	M50A	K07765	90	13	47	101	59	135	64
EmuJ_000796000.1	lys 63 specific deubiquitinase BRCC36	M67A	K11864	186	231	344	275	265	283	102
EmuJ_001166800.1	26S proteasome regulatory subunit N11	M67A	K03030	110	182	213	199	257	371	163
EmuJ_000331200.1	COP9 signalosome complex subunit 6	M67A	K12179	217	47	87	88	137	240	80
EmuJ_000687300.1	26S proteasome non ATPase regulatory subunit 7	M67A	K03038	83	119	157	108	143	163	122
EmuJ_000831400.1	COP9 signalosome complex subunit 5	M67A	K09613	43	52	80	54	58	167	108

EmuJ_000795700.1	26S proteasome regulatory subunit N11	M67A	—	45	44	105	64	78	84	71
EmuJ_000723000.1	STAM binding protein	M67C	K11866	376	290	69	190	188	401	110
EmuJ_000151700.1	pre mRNA processing splicing factor 8	M67C	K12856	58	10	65	75	79	294	67
EmuJ_000068700.1	eukaryotic translation initiation factor 3	M67X	K03249	230	332	186	159	220	460	126
EmuJ_000117100.1	eukaryotic translation initiation factor 3	M67X	K03247	163	232	185	122	197	238	84
EmuJ_000535900.1	CAAX prenyl protease 2	M79	K08658	57	18	77	37	33	37	55
EmuJ_000665500.1	cadherin EGF LAG seven pass G type receptor 1	P02A	—	1	0	2	2	1	1	1
EmuJ_000168400.1	polycystin 1	P02A	—	0	0	0	0	0	0	0
EmuJ_000396200.1	hypothetical protein PKD	P02A	—	0	0	0	0	0	0	0
EmuJ_000723600.1	polycystin 1	P02A	—	0	0	0	0	0	0	0
EmuJ_000622600.1	receptor for egg jelly 6	P02B	—	0	0	0	0	0	0	0
EmuJ_000184900.1	glycoprotein Antigen 5	S01A	—	94	33	14	308	212	2038	1049
EmuJ_000085400.1	Mastin	S01A	—	7	1	78	136	143	424	220
EmuJ_000165600.1	enteropeptidase	S01A	—	0	3	9	7	0	11	8
EmuJ_001046200.1	subfamily S1A unassigned peptidase S01 family	S01A	—	3	2	8	0	0	4	9

EmuJ_000436600.1	Transcription factor FAR1 lipoprotein receptor ldl	S01A	—	0	0	1	3	0	6	8
EmuJ_000924600.1	transmembrane protease serine 3	S01A	K09634	1	0	3	7	0	0	3
EmuJ_000820800.1	Peptidase S1 S6 chymotrypsin Hap	S01A	—	0	0	0	0	0	0	2
EmuJ_000166800.1	peptidase s8 s53 subtilisin kexin sedolisin	S08A	K01280	23	5	30	30	29	75	38
EmuJ_001148400.1	membrane bound transcription factor site 1	S08A	K08653	15	7	19	17	4	44	23
EmuJ_000998900.1	proprotein convertase subtilisin:kexin type 5	S08B	K08654	62	13	43	81	51	53	39
EmuJ_000412100.1	neuroendocrine convertase 2	S08B	K01360	2	6	16	68	47	26	27
EmuJ_000896900.1	furin 1 S08 family	S08B	K01349	6	5	8	13	10	5	12
EmuJ_000086100.1	Furin 1	S08B	—	1	1	2	0	0	2	9
EmuJ_000896900.2	Uncharacterized protein	S08B	K01349	0	0	0	0	0	0	0
EmuJ_000668800.1	prolyl endopeptidase	S09A	K01322	120	78	186	141	171	337	168
EmuJ_000297600.1	dipeptidyl aminopeptidaseprotein	S09B	—	32	36	31	6	12	39	27
EmuJ_000362100.1	Dipeptidyl peptidase 9	S09B	K08656	26	6	18	11	4	64	18
EmuJ_001178900.1	abhydrolase domain containing protein 13	S09C	K06889	37	42	137	99	265	550	88
EmuJ_000958300.1	Phospholipase carboxylesterase	S09C	—	31	26	44	47	50	24	55

EmuJ_000217800.1	monoacylglycerol lipase abhd12	S09C/S33	K13704	15	33	182	33	54	8	11
EmuJ_000843300.1	acylamino acid releasing enzyme	S09C/S33	K01303	40	18	40	46	35	85	61
EmuJ_000465300.1	acyl protein thioesterase 1	S09X	—	156	210	205	167	276	384	199
EmuJ_000053900.1	acyl protein thioesterase 12	S09X	K06130	81	47	108	70	111	157	100
EmuJ_000445400.1	abhydrolase domain containing protein 16A	S09X	—	55	19	111	51	79	82	81
EmuJ_000438400.1	S formylglutathione hydrolase	S09X	K01070	32	24	64	67	63	101	98
EmuJ_000106200.1	para nitrobenzyl esterase	S09X	—	78	52	129	18	42	15	41
EmuJ_001075400.1	acetylcholinesterase	S09X	K01049	20	16	19	4	2	0	3
EmuJ_000297300.1	BC026374 protein (S09 family)	S09X	K01050	0	0	0	0	4	7	6
EmuJ_000845300.1	neuroligin	S09X	K07378	0	0	0	3	0	0	2
EmuJ_000962700.1	family S9 non peptidase ue S09 family	S09X	—	0	0	1	0	0	0	0
EmuJ_000783300.1	carboxylesterase 5A	S09X	—	0	0	0	0	0	0	0
EmuJ_000966500.1	hormone sensitive lipase	S09X/S09C	K07188	24	3	10	30	15	68	32
EmuJ_000732400.1	acetylcholinesterase	S09X/S09C	K01049	1	3	3	0	4	2	18
EmuJ_000170200.1	lysosomal protective protein	S10	K13289	63	42	86	53	93	344	192

EmuJ_000939400.1	beta LACTamase domain containing family member	S12	K17382	6	1	9	61	3	47	22
EmuJ_000242400.1	ATP dependent Clp protease proteolytic subunit	S14	K01358	90	87	74	88	100	167	107
EmuJ_000324100.1	Lon protease homolog	S16	K08675	50	22	125	65	102	208	76
EmuJ_001050600.1	mitochondrial inner membrane protease subunit	S26A	K09647	116	95	107	32	56	81	97
EmuJ_000341800.1	signal peptidase complex catalytic subunit	S26B	K13280	257	495	482	236	312	818	190
EmuJ_001028800.1	Lysosomal Pro X carboxypeptidase	S28	K01285	149	91	193	143	172	108	204
EmuJ_000456150.1	Lysosomal Pro X carboxypeptidase	S28	K01276	12	10	28	40	110	61	40
EmuJ_000190200.1	protein NDRG3	S33	—	395	466	426	632	573	698	218
EmuJ_001065500.1	Ndr	S33	—	363	197	89	342	299	1272	383
EmuJ_000917500.1	lysosomal acid lipase:cholesteryl ester	S33	K01052	169	171	92	149	164	379	75
EmuJ_000295800.1	Alpha	S33	K13696	384	271	53	57	64	96	64
EmuJ_000708700.1	abhydrolase domain containing protein 11	S33	K13703	72	40	100	48	120	110	136
EmuJ_000955700.1	abhydrolase domain containing protein 8	S33	K13701	190	133	227	7	13	0	2
EmuJ_000295900.1	Alpha	S33	K13696	32	35	83	32	3	93	56
EmuJ_000947700.1	Ndr	S33	—	11	6	6	15	0	8	30

EmuJ_000295600.1	Alpha	S33	K13696	2	0	16	7	18	11	18
EmuJ_000684300.1	abhydrolase domain containing protein	S33/S09C	K13699	23	29	222	86	68	87	64
EmuJ_000879100.1	protein phosphatase methylesterase 1	S33/S09C	K13617	114	22	35	29	47	157	35
EmuJ_000612200.1	Der1 domain family	S54	K11519	119	65	103	54	38	120	74
EmuJ_001179900.1	inactive rhomboid protein 1	S54	—	38	9	25	70	83	22	40
EmuJ_000102400.1	stem cell tumor	S54	K02857	68	9	43	40	13	64	27
EmuJ_000435100.1	presenilin associated rhomboid	S54	K09650	38	31	85	15	20	27	22
EmuJ_001126200.1	stem cell tumor	S54	K02857	6	1	8	45	8	37	20
EmuJ_001034300.1	nuclear pore complex protein Nup98 Nup96	S59	K14297	42	7	19	45	32	51	34
EmuJ_000569300.1	family S60 non peptidase ue S60 family	S60	K06569	0	1	2	0	2	9	12
EmuJ_000481200.1	proteasome prosome macropain subunit beta	T01A	K02734	227	612	380	471	363	657	266
EmuJ_001182700.1	subfamily T1A non peptidase ue	T01A	K02729	225	350	475	473	434	813	187
EmuJ_000230600.1	proteasome subunit beta type 6	T01A	K02738	286	609	449	224	385	622	210
EmuJ_001120300.1	proteasome subunit alpha type 2	T01A	K02726	227	254	388	368	302	487	220
EmuJ_000877400.1	proteasome subunit alpha type 6	T01A	K02730	292	418	299	307	315	408	206

EmuJ_000062300.1	proteasome subunit beta type 7	T01A	K02739	187	269	563	166	388	476	166
EmuJ_000864950.1	proteasome prosome macropain subunit alpha	T01A	K02728	200	363	359	164	212	460	235
EmuJ_000590200.1	proteasome prosome macropain	T01A	K02737	143	282	310	121	269	629	134
EmuJ_000252500.1	proteasome prosome macropain subunit beta	T01A	K02732	85	149	208	415	234	292	380
EmuJ_001064900.1	proteasome prosome macropain subunit beta	T01A	K02735	151	218	286	237	326	326	151
EmuJ_000196100.1	Proteasome subunit alpha beta	T01A	K02727	70	135	194	233	181	362	188
EmuJ_000682700.1	Proteasome subunit alpha type 7	T01A	K02731	22	17	38	30	51	51	10
EmuJ_000682300.1	Proteasome subunit alpha type 7	T01A	K02731	0	0	0	0	0	0	0
EmuJ_000682550.1	Proteasome subunit alpha type 7	T01A	K02731	0	0	0	0	0	0	0
EmuJ_001150600.1	proteasome prosome macropain subunit beta type	T01X	K02736	252	315	280	366	279	408	273
EmuJ_000750500.1	20S proteasome subunit alpha 6	T01X	K02725	100	143	252	207	199	512	157
EmuJ_000325200.1	N4 Beta N acetylglucosaminyl L asparaginase	T02	K01444	21	11	22	38	19	34	73
EmuJ_000829500.1	threonine aspartase 1	T02	K08657	7	9	15	0	11	11	18
EmuJ_000761500.1	gamma glutamyltransferase 1	T03	—	22	16	31	25	26	53	22
EmuJ_000806300.1	gamma-glutamyltranspeptidase	T03	K18592	0	0	1	0	0	2	1

Appendix II

The antigen homologues matched to *E. multilocularis* reference transcriptome

Transcripts ID	Description	pNonc1	pNonc2	pAonc	s4Wmet	p4Wmet	pCmet	s16Wmet
EmuJ_000364000.1	14-3-3 protein homolog 2	2745	4812	1919	1458	2003	8364	1422
EmuJ_001192500.1	14-3-3 protein zeta	1733	4117	2462	3000	2727	3986	2245
EmuJ_000036300.1	ACTI	34252	20266	15162	6288	7838	18887	3311
EmuJ_000406900.1	ACTI	0	0	35	7	11	907	68
EmuJ_000407200.1	ACTI	1	0	25	14	22	925	163
EmuJ_000061200.1	ACTII	0	0	0	123	7	505	354
EmuJ_000407300.1	ACTII	0	3	1	0	0	0	1
EmuJ_000701700.1	ACTII	0	0	0	12	18	364	212
EmuJ_000703300.1	ACTIII	0	0	1	29	22	192	107
EmuJ_000413200.1	Alpha tubulin	356	110	898	1220	1672	7856	1665
EmuJ_000886400.1	Alpha tubulin	42	43	435	2191	1739	5280	1129
EmuJ_000476400.1	Alpha tubulin	4	5	11	0	3	18	8
EmuJ_000040900.1	Alpha tubulin	7	3	15	12	0	6	4
EmuJ_000339900.1	Alpha tubulin	6	1	5	6	1	7	4
EmuJ_000042200.1	Alpha tubulin	1	3	15	0	0	3	2
EmuJ_002136500.1	Alpha tubulin	2	3	2	5	0	2	1
EmuJ_000352800.1	Alpha tubulin	1	1	5	0	0	2	1
EmuJ_000042600.1	Alpha tubulin	1	0	1	6	0	0	0
EmuJ_000042500.1	Alpha tubulin	2	0	2	0	0	0	1
EmuJ_000359200.1	Alpha tubulin	0	0	0	4	0	0	0
EmuJ_001070900.1	Alpha tubulin	0	0	2	0	0	0	0
EmuJ_000340050.1	Alpha tubulin	0	0	0	0	0	0	1
EmuJ_000588000.1	Alpha tubulin	0	0	1	0	0	0	0
EmuJ_000042700.1	Alpha tubulin	0	0	0	0	0	0	0
EmuJ_000735000.1	Alpha tubulin	0	0	0	0	0	0	0
EmuJ_001124100.1	Alpha tubulin	0	0	0	0	0	0	0
EmuJ_000184900.1	Antigen 5	94	33	14	308	212	2038	1049
EmuJ_000485800.1	Antigen II/3（elp）	4994	6465	5166	1694	2662	2058	840
EmuJ_000672200.1	Beta tubulin	784	861	1180	1761	1783	6685	1514
EmuJ_000202500.1	Beta tubulin	17	18	327	936	814	3012	251
EmuJ_000202600.1	Beta tubulin	1	1	11	401	382	1251	389

EmuJ_001126150.1	Beta tubulin	11	5	18	26	13	4	21
EmuJ_000569000.1	Beta tubulin	1	0	1	0	6	12	3
EmuJ_000041100.1	Beta tubulin	4	0	8	0	3	0	3
EmuJ_001081200.1	Beta tubulin	4	0	8	0	0	0	1
EmuJ_000955100.1	Beta tubulin	0	0	0	0	0	3	10
EmuJ_000069900.1	Beta tubulin	0	1	2	0	6	0	1
EmuJ_000617000.1	Beta tubulin	0	0	0	0	0	3	0
EmuJ_000601200.1	Calcineurin A	78	18	52	58	54	85	59
EmuJ_000447500.1	Calcineurin B	148	144	150	260	175	189	168
EmuJ_000454300.1	Calcineurin B	180	136	127	32	137	129	101
EmuJ_000920600.1	Cyclophilin	2732	5395	4481	4143	4436	10470	4275
EmuJ_000009600.1	cytoplasmic antigen 1	0	0	0	0	0	0	1
EmuJ_000517100.1	EF-1	1432	2701	1272	639	916	1276	575
EmuJ_000982200.1	EF1a	5598	7644	5178	5789	5220	11147	4723
EmuJ_000911600.1	EF-2	0	0	1	0	0	0	0
EmuJ_000342900.1	EG19	3	0	9	128	0	72	2440
EmuJ_000328500.1	EG95 (Onco1)	9212	19004	9414	121	732	35	0
EmuJ_000368620.1	EG95	19	78	8422	367	286	0	0
EmuJ_000710400.1	EG95	4855	15169	6447	0	104	0	1
EmuJ_000381200.1	EmAgB8/1	0	0	0	19211	5046	22858	26592
EmuJ_000381100.1	EmAgB8/2	0	0	0	836	752	9748	12075
EmuJ_000381500.1	EmAgB8/3	10	0	187	8091	13182	51220	9222
EmuJ_000381600.1	EmAgB8/3	0	0	0	0	0	0	0
EmuJ_000381700.1	EmAgB8/3	0	0	0	0	0	0	0
EmuJ_000381400.1	EmAgB8/4	5	0	0	3351	4411	25827	8418
EmuJ_000381800.1	EmAgB8/5	0	0	0	0	0	0	0
EmuJ_000393300.1	Em-alp-1	1	0	259	56	44	179	54
EmuJ_000393400.1	Em-alp-2	18	14	1022	90	118	544	124
EmuJ_000752700.1	Em-alp-3	0	0	0	0	0	0	0
EmuJ_000752800.1	Em-alp-4	0	0	0	0	0	0	0
EmuJ_000362600.1	Em-bruno1	3	2	3	11	5	8	29
EmuJ_000943000.1	Em-bruno2	2	0	21	25	13	32	118
EmuJ_000942800.1	Em-bruno3	0	0	0	0	0	0	0
EmuJ_000790200.1	EmCBP1	8	3	139	417	299	1085	336
EmuJ_000790300.1	EmCBP2	8	7	61	32	24	36	24

EmuJ_000654100.1	EmCLP1	0	0	0	0	0	0	0
EmuJ_000654200.1	EmCLP1	0	0	0	0	0	0	0
EmuJ_000654500.1	EmCLP1	0	0	1	0	0	0	0
EmuJ_000654600.1	EmCLP1	0	0	0	0	0	0	0
EmuJ_000654800.1	EmCLP1	0	0	0	0	0	0	0
EmuJ_000989200.1	EmCLP2	73	73	101	32	77	133	66
EmuJ_000590100.1	EmDLC	29	100	60	89	223	402	104
EmuJ_000940800.1	EmDLC	0	0	27	0	44	265	67
EmuJ_000940900.1	EmDLC	32	31	300	9523	3108	8115	1268
EmuJ_000941100.1	EmDLC	0	0	76	205	147	2311	179
EmuJ_000946700.1	EmDLC	475	1153	952	60	76	61	8
EmuJ_001060400.1	EmDLC	140	320	291	483	639	2038	1279
EmuJ_000538300.1	EmGST1	1716	2815	2416	4681	4157	9607	2258
EmuJ_001102300.1	Em-hdac1	85	39	89	83	118	175	114
EmuJ_000606200.1	Em-nanos	0	0	0	0	6	57	25
EmuJ_000791700.1	EmTRX	1166	2479	2680	5757	3436	5885	3135
EmuJ_000355800.1	Em-TSP1	0	0	15	391	347	456	167
EmuJ_001070300.1	Em-TSP2	33	47	100	197	357	456	560
EmuJ_001077400.1	Em-TSP3	1399	3778	4156	11	173	12	1
EmuJ_001077300.1	Em-TSP3	26	20	18	0	0	0	0
EmuJ_001021500.1	Em-TSP4	33	60	132	617	646	418	119
EmuJ_001077100.1	Em-TSP5	16	13	903	3631	1514	7559	4800
EmuJ_001021300.1	Em-TSP6	116	195	123	759	857	1492	994
EmuJ_000834300.1	Em-TSP7	92	108	169	177	161	601	391
EmuJ_000328400.1	Em-TSP8	52	104	426	36	90	0	1
EmuJ_000515900.1	EmY162	0	0	0	0	0	62	34
EmuJ_000564900.1	EmY162	346	1009	2690	453	825	2931	139
EmuJ_000550000.1	FABP1	0	0	0	0	398	72	59
EmuJ_000549800.1	FABP2	0	0	0	162	204	439	80
EmuJ_002165500.1	FABP2	0	0	0	0	0	0	1
EmuJ_000905600.1	FBPA	1127	1077	490	1023	1437	8319	1899
EmuJ_000382200.1	Ferritin	5740	11631	5426	6622	7002	34365	6409
EmuJ_000254600.1	GAPDH	7746	1411	2707	2805	3751	22243	3630
EmuJ_000032300.1	GP50	11485	19983	4034	37	671	0	1
EmuJ_000295100.1	GP50	7356	11796	7346	9	228	0	1

EmuJ_000289400.1	GP50	3493	6128	1667	0	84	5	0
EmuJ_000293700.1	GP50	1073	1889	389	0	28	0	0
EmuJ_000261100.1	GP50	0	0	3100	20	77	0	1
EmuJ_000050100.1	GP50	628	894	643	0	12	0	0
EmuJ_000681200.1	GP50	13	16	396	458	682	14	7
EmuJ_000049700.1	GP50	44	14	799	217	270	0	2
EmuJ_000512300.1	GP50	0	0	1140	59	100	0	1
EmuJ_001120900.1	GP50	0	0	4	118	322	0	4
EmuJ_001120700.1	GP50	0	0	0	83	109	0	5
EmuJ_000401200.1	GP50	31	10	31	0	0	0	1
EmuJ_001201600.1	GP50	0	0	0	9	0	0	15
EmuJ_000515550.1	GP50	6	4	11	0	0	0	2
EmuJ_000289600.1	GP50	0	0	7	0	0	0	1
EmuJ_000480200.1	GP50	0	2	2	0	0	0	2
EmuJ_000566700.1	GP50	0	0	0	0	0	0	3
EmuJ_000324200.1	GP50	0	0	2	0	0	0	1
EmuJ_000047000.1	GP50	3	0	0	0	0	0	0
EmuJ_000304800.1	GP50	0	0	2	0	0	0	1
EmuJ_000520550.1	GP50	0	0	0	0	0	0	1
EmuJ_000564000.1	GP50	0	0	0	0	0	0	0
EmuJ_000743500.1	GP50	0	0	0	0	0	0	0
EmuJ_000743600.1	GP50	0	0	0	0	0	0	0
EmuJ_000249600.1	GRP78(HSP70)	195	250	424	297	305	499	234
EmuJ_000472800.1	Histone H2B	245	605	291	552	512	1224	467
EmuJ_000566500.1	Histone H2B	3	4	7	39	20	40	9
EmuJ_000212700.1	HSP20(Onco2)	7699	10552	1705	52	587	252	440
EmuJ_000723700.1	HSP70	4	2	10	11	0	16	7
EmuJ_001085100.1	HSP70	248	48	981	313	475	836	295
EmuJ_001085400.1	HSP70	3141	1825	2949	975	1256	3429	879
EmuJ_001136500.1	Kunitz	35	102	6	0	0	0	0
EmuJ_001136800.1	Kunitz	0	0	0	0	0	0	1
EmuJ_001084300.1	M123	0	0	0	0	0	0	0
EmuJ_001084400.1	M9	0	5	0	0	0	0	3
EmuJ_000417100.1	MDH	238	289	610	987	1139	6351	1762
EmuJ_001185000.1	Mdhm	73	95	465	497	315	614	225

EmuJ_001185100.1	Mdhm	24	54	236	260	193	277	105
EmuJ_000742900.1	MUC-1	6	0	13	14540	35610	66669	4624
EmuJ_000315800.1	MUC-2	0	0	254	0	0	0	0
EmuJ_000315900.1	MUC-2	0	0	17	0	0	0	0
EmuJ_000316000.1	MUC-2	0	0	164	0	0	0	0
EmuJ_000408150.1	MUC-2	0	0	138	0	12	0	1
EmuJ_000408200.1	MUC-2	211	350	3733	534	1287	12	2
EmuJ_000653900.1	Myophilin	11	40	8	254	33	818	446
EmuJ_000861500.1	nanos-like protein	4	3	15	0	0	54	21
EmuJ_000550800.1	P29	528	716	619	269	628	1845	413
EmuJ_000763300.1	Paramyosin	1	0	0	9	24	561	195
EmuJ_000517700.1	PMI	82	64	200	106	252	255	154
EmuJ_000738700.1	prohibitin protein WPH	735	1261	995	508	832	2925	464
EmuJ_000450400.1	Pumilio 2	145	32	48	95	53	108	84
EmuJ_000878100.1	Pumilio 2	50	11	17	11	38	29	22
EmuJ_001006600.1	Rab-4A	116	85	99	146	141	165	79
EmuJ_001193100.1	SerpinEmu	318	564	45	0	12	27	8
EmuJ_001193200.1	SerpinEmu	218	296	38	0	12	12	3
EmuJ_000882500.1	Severin	922	1358	947	1070	1449	3846	616
EmuJ_000958100.1	Tropomyosin	210	134	189	248	106	274	413