# Chapter 3

# *S. mansoni* in experimentally infected mice

## 3.1 Introduction

### 3.1.1 Overview

Intra-mammalian stages of *S. mansoni* live among the components of host blood and encounter the vast network of the blood circulatory system where they eventually locate a suitable site for egg-laying. Interactions with the host environment are necessary for successful infection. However, much of the molecular biology underlying such interactions is still lacking. A better understanding of these processes will contribute to a holistic knowledge of *S. mansoni* biology, and could provide a basis for designing intervention strategies. In this chapter, I present a transcriptomic investigation of 6 intra-mammalian stages of *S. mansoni* infection, ranging from day 6 to day 35 post-infection. This dataset uses the patterns of gene expression over time to investigate processes at key stages of the infection particularly those related to host-parasite interactions. To this end, the data reflect processes known to be associated with key stages of the infection, represent a unique view into the resistance of the lung stage to attack by the host immune system, suggest potential novel players in host-parasite interactions, and propose mechanisms that govern mesenteric migration of adults.

### 3.1.2 Host-parasite interactions in intra-mammalian stages

Infection of the mammalian host with *S. mansoni* starts with cercariae penetrating through intact skin, entering the bloodstream, and migrating with blood through tissues around the body. After protracted stages in the lung and liver, the parasite develops into the adult and then migrates from the liver towards the mesenteric veins for mating and egg-laying (Wilson, 2009). The lodging of schistosomules in lung capillaries presents close physical contact between the schistosomule tegument and the endothelial wall of blood vessels inside the lung (Crabtree and Wilson, 1986a), interrupting blood flow and leading to potential accumulation of circulating immune

cells or coagulation; however, lung schistosomules are reported to be resistant to immune killing (Bentley *et al*., 1981; Crabtree and Wilson, 1986b; Mastin *et al*., 1985; McLaren and Terry, 1982). Interestingly, inflammation in the lung is not associated with the presence of parasites (Burke *et al*., 2011) raising the question on immune-evasion tactics that the parasites employ. Further, in mice vaccinated by irradiated cercariae, the lung is the main site of parasite attrition in challenge infection (Wilson et al., 1986); however, the attrition does not appear to be mediated directly by the immune-mediated cytotoxicity (Crabtree and Wilson, 1986b; Mastin *et al*., 1985). The lung also presents a unique environment with high oxygen pressure from respiration.

After the lung, development from schistosomules into adult stages is associated with localisation in the liver, and components of host immune responses are required for the development (Clegg, 1965; Shaker *et al*., 1998, 2011). It is thought that the arrival in the liver is a passive process and once the schistosomules are located in the liver, blood feeding and development commence (Clegg, 1965; Wilson, 2009). The signal for detecting liver environment upon arrival is not understood, and understanding the requirements and adaptations to the liver environment would have practical implications as the current line of treatment, praziquantel, has low efficacy against liver stages (Kusel *et al*., 2007).

The egg-laying sites are unique to particular species. For example, *S. mansoni* migrate towards the mesenteric vein close to the intestine, whereas *S. haematobium* migrates toward the venous plexus of the bladder. Even in species that reside in mesenteric veins for egg-laying, distinct locations within the mesenteric vein are used by different species. Furthermore wandering behaviour has been reported based on locations of eggs found in the mesenteric vein (Pellegrino and Coelho, 1978). The cue for this migration is currently unidentified. As well as development and migration to suitable sites, the parasites must adapt to their bloodstream environment, the different tissues through which the blood passes and, importantly, survive among the host immune responses and inflammatory processes that maybe caused by the damaged host tissue during the infection (Bloch, 1980; Crabtree and Wilson, 1986a).

### 3.1.3 Progress in transcriptomic and genomic approaches

Studying gene expression during infections provides a better understanding of the molecular processes underlying the interactions with host environment. Transcriptional changes of parasites in various environments and at different developmental stages could reflect molecular processes required for biologically relevant responses. Transcriptomic studies have been previously used to understand the biology of intra-mammalian stages as well as to predict drug and vaccine targets (Chai et al., 2006; Fitzpatrick et al., 2009a; Gobert et al., 2009b; Jolly et al., 2007; Parker-Manuel et al., 2011). However, many were done using microarray, or EST technologies which provide lower resolution than RNA-seq or cover only a subset of developmental stages. The developmental stages of parasite materials used for these studies also varied and usually combined materials obtained *in vivo* and *in vitro* (Chai et al., 2006; Fitzpatrick et al., 2009a). The lung stage is particularly challenging to study owing to the difficulty in obtaining sufficient biomaterials; past studies on lung stages of *S. mansoni* used biomaterial from *in vitro* cultivated schistosomules, mechanically transformed from cercariae, which may not fully represent the biology of the lung stage worms (Gobert *et al*., 2007). Transcriptional profiling of lung stage parasites *in vivo* has previously been performed (Chai *et al*., 2006) but using a microarray and in isolation, rather than alongside other time points from the lifecycle. The current status of the *S. mansoni* genome has enabled more information and confidence to be obtained from transcriptomic information. Further, the efficiency of library preparation means less materials are required to successfully produce transcriptomic data, therefore previously less accessible stages (such as the lung stage) can now be explored.

### 3.1.4 Aims and approaches

The aim of this chapter is to provide further insight into host-parasite interactions during the intra-mammalian stages of *S. mansoni*. Particular focus is on interactions in the lung stage's immune evasion strategies, migration and development in the liver, and migration of adults to an egg-laying site. To this end, transcriptomes were obtained from experimentally infected mice from which parasites were collected at days 6, 13, 17, 21, 28, and 35 post-infection, providing the first *ex vivo* lung transcriptome to be investigated as part of a comprehensive *ex vivo* intra-mammalian

timecourse. The selection of these time points was based on previous radio-isotope tracking studies that laid such time points on some key stages of the infection and migration path (Wilson, 2009). At day 6, parasites were mainly present in the lung, and after leaving the lung they circulate within the bloodstream with cardiac output and accumulate in liver until around day 21 when the highest number of parasites is reached (Clegg, 1965; Wilson, 2009). The parasites mature and start pairing by day 28 and egg-laying occur around day 35 (Clegg, 1965). With our initial interest on the intra-mammalian migration path of the parasite, the collection did not go beyond day 35 post-infection by which point, the parasites should already reside in the mesenteric vein (Bloch, 1980; Clegg, 1965; Georgi *et al*., 1982, 1986; Wilson *et al*., 1986).

### 3.1.5 Chapter outline

This chapter contains a brief description of the morphological heterogeneity of the parasites obtained at fixed time points post-infection. Transcriptomic data were investigated in detail, using pairwise comparisons, gene clustering, and functional enrichment, to select genes for more detailed bioinformatics investigation. The sections move from lung stage, to liver stages, and to adult stages. The lung stage transcriptome suggests potential mechanisms that allow the parasites to be resistant to immune killing. The liver stage transcriptome reflects growth and development as expected, and it suggest a possible adaptation to the liver environment. The adult stage transcriptome reflects reproduction as expected, and cues are proposed that enable the migration of the parasite to mesentery.

## 3.2 Methods

### 3.2.1 Mouse infection

#### *3.2.1.1 Infection procedures*

Schistosomes were collected from experimentally infected mice at 6 time points post-infection – days 6, 13, 17, 21, 28, and 35 – spanning lung stages through to the egg-laying adult stage. In order to obtain enough parasites at each time point for RNA sequencing, mice were infected with different numbers of cercariae shed from *B. glabrata* snails. Two thousand cercariae were used to infect mice used to provide day-6 and day-13 parasites, 500 were used to provide day-17 and day-21 parasites, 350 for day-28 and 300 for day 35. More mice were used for early time points due to the

greater uncertainty in acquiring the samples (Figure 3.1). Percutaenous infections were performed by Prof. Michael J. Doenhoff (University of Nottingham, Nottingham, UK) by applying suspensions of mixed-sex cercariae to the shaved abdomens of anesthetised mice and leaving for 30 minutes. Although much of the schistosome life cycle could be accessed locally, only sub-cutaneous infections was possible at the time of this experiment, and this type of infection route did not yield sufficient numbers of early-stage parasites. All mice used for this experiment were females of CD-1 outbred strain (Harlan Laboratories) aged between 8-12 weeks old. A pool of parasites from each mouse was treated as one biological replicate.

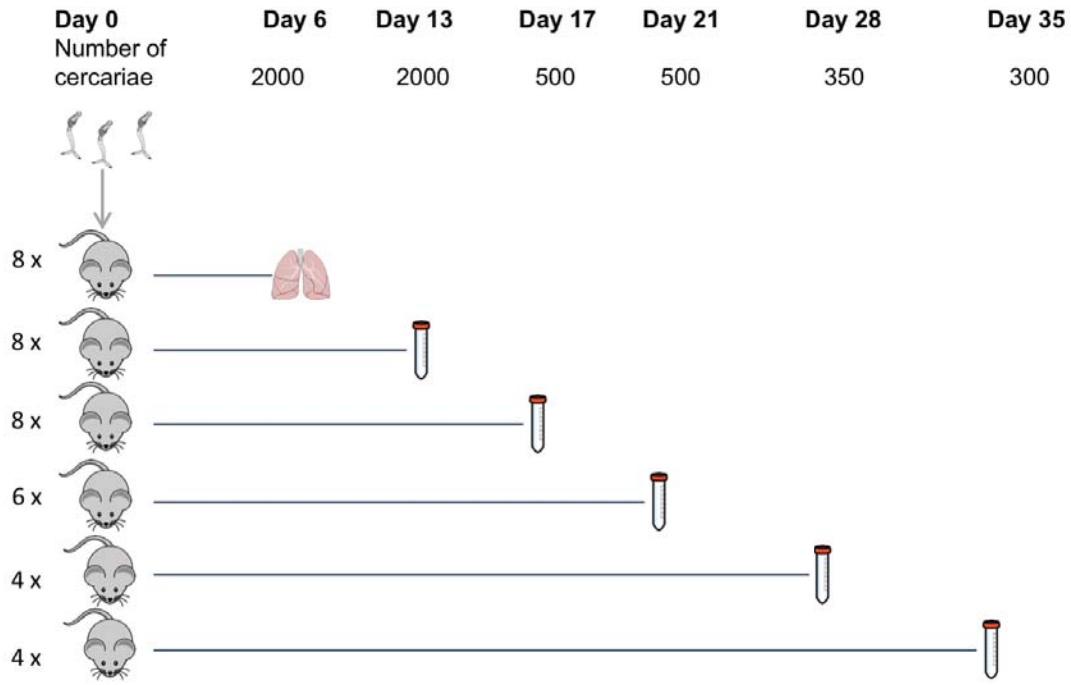


***Figure 3.1 Experiment layout***

Mice were infected with different numbers of cercariae for collection at six time points post-infection as shown at the top of the diagram. The different numbers of cercariae at each time point are necessary for sufficient amounts of parasites to be collected at the end of the infection and to maintain the health of the infected mice. A pool of parasites from one mouse is treated as one biological replicate and the number of mice used for each time point is shown on the left. Method for parasite collection at day 6 is mincing and incubation of the lung. Collection at other time points were done by portal perfusion through the heart and perfusate collected in Falcon tubes.

## 3.2.2 Parasite and tissue collection from infected mice

Lung stage parasites were collected from the incubated lung while other time points were collected by portal perfusion (Figure 3.1). The timing of infection was arranged such that the all stages could be collected on the same day. On the indicated day post-infection, infected mice were sacrificed using an overdose of pentobarbitone with addition of 10 U/ml heparin to prevent blood clotting. Then the parasites were collected, washed, imaged, collected into Eppendorf tubes, mixed with TRIzol, and stored in dry ice until later moved to -80 ºC storage. Details are explained below.

### *3.2.2.1 Lung stage collection and incubation*

Lung stage parasites were collected using a protocol adapted from the Biomedical Research Institute (Biomedical Research Institute, 2016); at day 6 post-infection, mice were dissected and lungs collected, minced into ~1 $mm^3$ pieces using scissors and added to a Falcon tube containing heparinised Basch media (10 U/ml heparin) and topped up to 50 ml with the heparinised Basch media (see Appendix D for Basch media conponents). The lungs from each mouse were incubated separately in individual Falcon tubes for ~1 hour at room temperature, followed by 37 ºC for 3 hours.

### *3.2.2.2 Lung stage after incubation*

After the 3-hour incubation, the media and lung tissue were resuspended, by turning the Falcon tube 2-3 times, and the media was passed through a 600 µm mesh into new 50 ml Falcon tubes. Larger pieces of lung remained on the mesh, while lung stage worms passed through the mesh together with small pieces of lung tissues. The protocol from BRI indicates that when incubated for a further 24 hours, more worms can be recovered. However, this was not feasible due to sample-collection logistics. The presence of worms after a 3-hour incubation was confirmed under a dissecting microscope and the remainder of the lung tissue was discarded. Next, the filtrate was centrifuged at 150 x g for 3 minutes at room temperature and approximately half of the supernatant gently decanted into a waste bucket. At this stage, the worms were still at the bottom of the Falcon tube and could be recovered using a Pasteur pipette. Removing all or most of the supernatant would run risk of losing the worm pellet from the bottom of the falcon tube because the worms were very small and easily

dislodged. Approximately 1-1.5 ml from the bottom of the tube was collected and added into a small Petri dish for imaging.

#### *3.2.2.3 Other time points - collection*

For other time points (days 13, 17, 21, 28, and 35), mice were euthanised and dissected to expose the portal vein and heart. The mice were clipped by their paws on to a plastic board so that the perfusate would run into a beaker placed underneath the animal. Then the portal vein was cut and ~ 30 ml perfusion media (Dulbecco's Modified Eagle's Medium (DMEM) high glucose, with 10 U/ml heparin) injected into the heart. Next, the surfaces of internal organs were briefly rinsed with more perfusion media to wash off any remaining parasites. The perfusate was collected into a Falcon tube. A pool of parasites from one mouse was treated as one biological replicate.

#### *3.2.2.4 Other time points - wash*

To wash the parasites, the perfusates were left at room temperature for 30 minutes for the worms to form pellets at the bottom of each tube. Supernatants were removed leaving ~5 ml, which was topped up to ~40 ml. Worms were left to sink for another 30 minutes. After which, half of the supernatant was removed and worms at the bottom were collected using a Pasteur pipette, and transferred to a small petri dish for imaging. With an exception for day-35 worms, which sank more quickly, worms from all other time points were left to sink for 30 minutes. Originally, centrifugation was used to collect the worms at each time point, however, blood cells from mice were also pelleted down and this prevented a clear view of the worms during imaging and would contaminate the parasite transcriptome data with mouse RNA.

#### *3.2.2.5 Imaging*

Worms were imaged on a dissecting microscope (Olympus SZ61), using Euromex Cmex10 camera and Image Focus 4.0 software. After the imaging step (~2 minutes per replicate), the content of the dish was transferred into a 2-ml Eppendorf tube, the dish was rinsed to recover residual worms and pooled into the Eppendorf tube. Then the tube was centrifuged at 150 x g for 3 minutes, the supernatant removed, 1 ml of TRIzol added, left at room temperature until transferred to dry ice, and later to storage at -80 ºC.

#### *3.2.2.6 Timing*

After collection, washing, and imaging TRIzol was added to each sample within 3-4 hours of perfusion. Prior to the addition of TRIzol, worms were at all times kept in DMEM high glucose media supplemented with heparin and were placed in a 37 ºC incubator when not being processed. After addition of TRIzol, the samples were kept at room temperature for 20–120 minutes before being stored in dry ice.

### 3.2.3 RNA extraction and library preparation

Parasite samples in TRIzol were next processed for RNA extraction. For samples of day 13 to day 35, three replicates were chosen at random for RNA extraction. For day-6 worms, contamination with mouse RNA was observed based on additional 28S rRNA peak on Agilent profile (electropherogram example shown in Figure 2.2D); therefore, all replicates were used to ensure the success of the process. There were 22 samples in total; therefore, RNA was extracted in batches of 6-8 samples. RNA extraction, library preparation, and sequencing were performed as described in chapter 2.

### 3.2.4 Morphology score

The images were used to determine morphology scores for all parasites captured on each image, based on the morphology categories from a published literature (Basch, 1981), diagram reproduced in Figure 3.2. In order to avoid bias from prior knowledge of the time points from which the images were obtained, all images were renamed to randomised numbers with metadata of the original names and the source of time points recorded.. An average of 31 worms were scored per replicate (range 18-75 worms). The number of parasites that fell into each morphology category were normalised by the total number of parasite in that replicate.

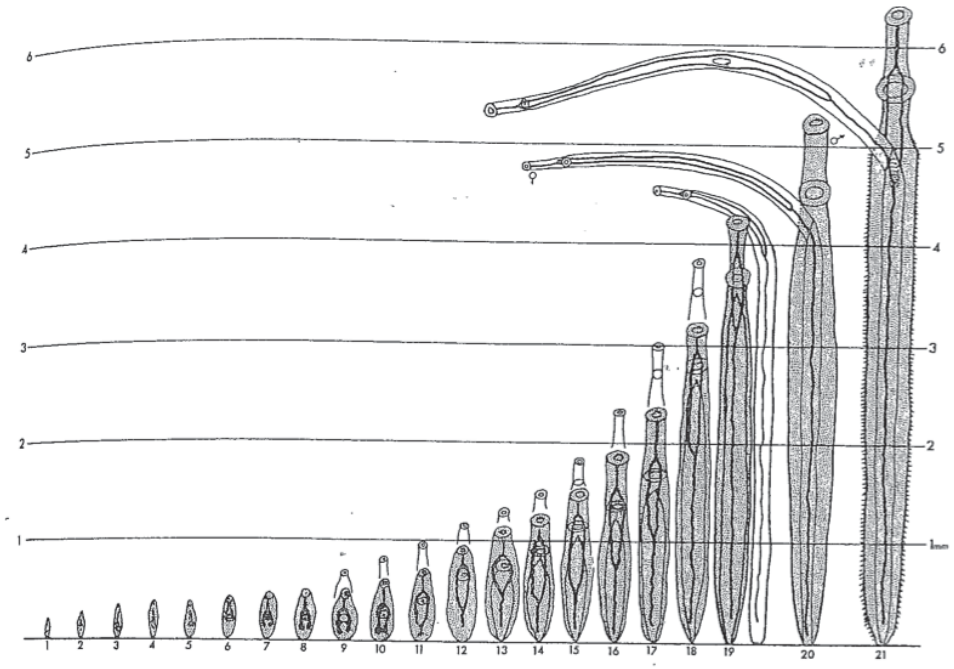


FIGURE 1. Standardized reference stages used to describe development of *S. mansoni*. Adapted from Faust et al. (1934).

***Figure 3.2 Diagram used for morphology scoring***

Morphology of *S. mansoni* from experimentally infected mice. The diagram is taken from Basch (1981) and is used for categorising *S. mansoni* in this thesis. Features of parasites used for categorising them into each morphology group are appearance of haemozoin-filled gut, the shape of the gut (fork end, closed end), the proportion of the extension of the gut end compared to the loop part, and whether the worms are paired or unpaired.

## 3.3 Results

### 3.3.1 Worm morphology

The majority of day-6 parasites were morphologically similar. It is worth bearing in mind that these parasites may not represent all morphological forms present at day 6 because these parasites were obtained from the lung only and not by perfusion; any parasite that had left the lung by day 6 were therefore not extracted. For this reason, day-6 parasite is referred to as ‘lung stage’. Developing parasites during day 13 to day 28 were morphologically heterogeneous, but with a clear trend of shifting toward more advanced developmental stages. At day 28 post infection, most parasites reached adult stages and some parasites formed pairs. Day-35 parasites have all reached fully mature adult stage and most were paired. Day-35 parasites were let to sink during the collection instead of centrifuged down; therefore, there might be some smaller immature parasites that were not collected, but according to the morphology

of the worms at day 28, which were collected by centrifugation, the number of smaller immature parasites should be minimal (Figure 3.3). The heterogeneity during development is consistent with observations made in previous studies (Basch, 1981; Clegg, 1965). This would add noise in the transcriptomes particularly in the liver stage and in adult stages where samples are mixed of male and female parasites, and was taken into account while analysing the data. Lung stage, however, should be the least affected by the heterogeneity.

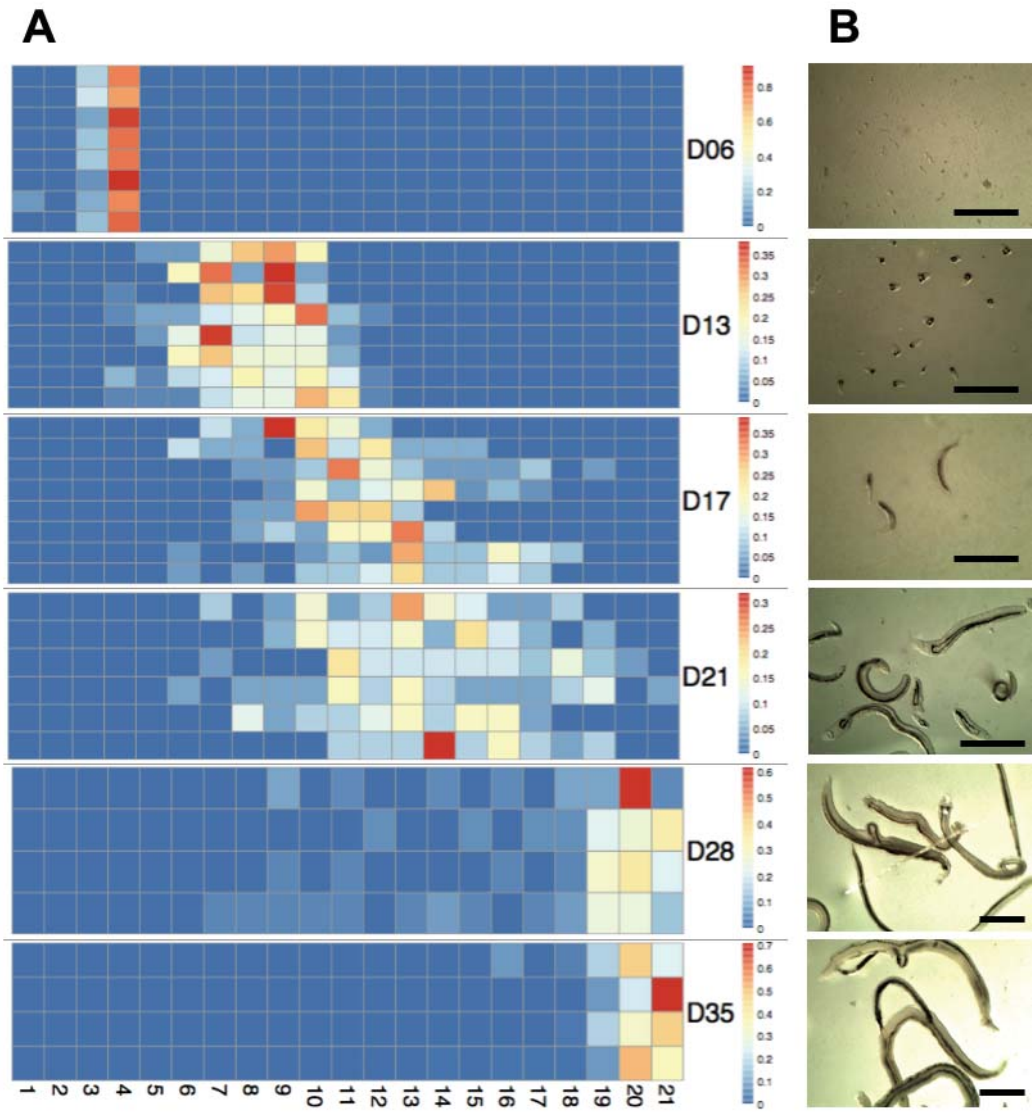


***Figure 3.3 Morphology of in vivo S. mansoni***

A) Morphology of *S. mansoni* at the six time points post-infection. Twenty-one morphology groups, based on morphology scores from Basch (1981), are shown at the bottom of the heatmap. Rows on the heatmap are replicates of the infections. An average of 31 worms per replicate (range 18-75 worms) were imaged for morphology scoring. Colours on the heatmap show the number of worms that fall into each morphology group, normalised by the number of worms categorised within a replicate. B) Representative images of worms from each time point. Scale bars: 1 mm.

## 3.3.2 RNA quantity and quality

All RNA from parasites of all time points showed clear rRNA peaks when run on Agilent Bioanalyzer chips (similar to electropherogram examples shown in Figure 2.2A and D), confirming high quality RNA with little or no degradation. The yield of extracted RNA ranged from 45 to 2844 ng/µl, in 30 µl nuclease-free water. NanoDrop measurements showed that 260/280 ratio of all samples were close to 2, but 260/280 ratio were ranged from 0.8 to ~2 (average 1.57) with no association with samples from any particular time points, suggesting that it could be TRIzol contamination from the extraction protocol.

## 3.3.3 Sequencing yields

### *3.3.3.1 Down sampling*

Fewer reads mapped to the *S. mansoni* genome from lung-stage parasites compared with other stages, and a large proportion of reads from day-6 worms mapped to the mouse genome (on average ~20 %). This was expected given that contamination of mouse RNA in the preparation of lung-stage worms was seen on RNA trace run on Agilent Bioanalyzer chips (similar to an example RNA trace shown in Figure 2.2D). Lung stage samples had an average of 520,214 read counts, while other samples (day-13 to day-35) had averages over 4,800,000 read counts (Table 3.1). To assess whether the number of reads were too low for the data analysis and whether re-sequencing of day-6 samples would be needed, the other samples were down-sampled to the read count level of day-6 samples (10%, 7.5%, 5%) and compared PCA plots resulting from the total reads, and the down-sized reads. All PCAs plots resemble each other in terms of variations between samples (Figure 3.4). However, down-sizing of other samples to the level of day-6 read counts reduced the number of differentially expressed genes detected (results not shown). The statistics tool used for analysing this dataset is designed to handle different numbers of reads between samples by estimated size factor normalisation. Therefore, downstream analysis were undertaken with the current dataset and further data for day-6 samples did not need to be generated.

*Table 3.1 Average of total read counts*

| Time point | Average counts between all replicates |
|---|---|
| **Day 6** | 520214 |
| **Day 13** | 7599869 |
| **Day 17** | 5938978 |
| **Day 21** | 4828234 |
| **Day 28** | 5938633 |
| **Day 35** | 5010390 |

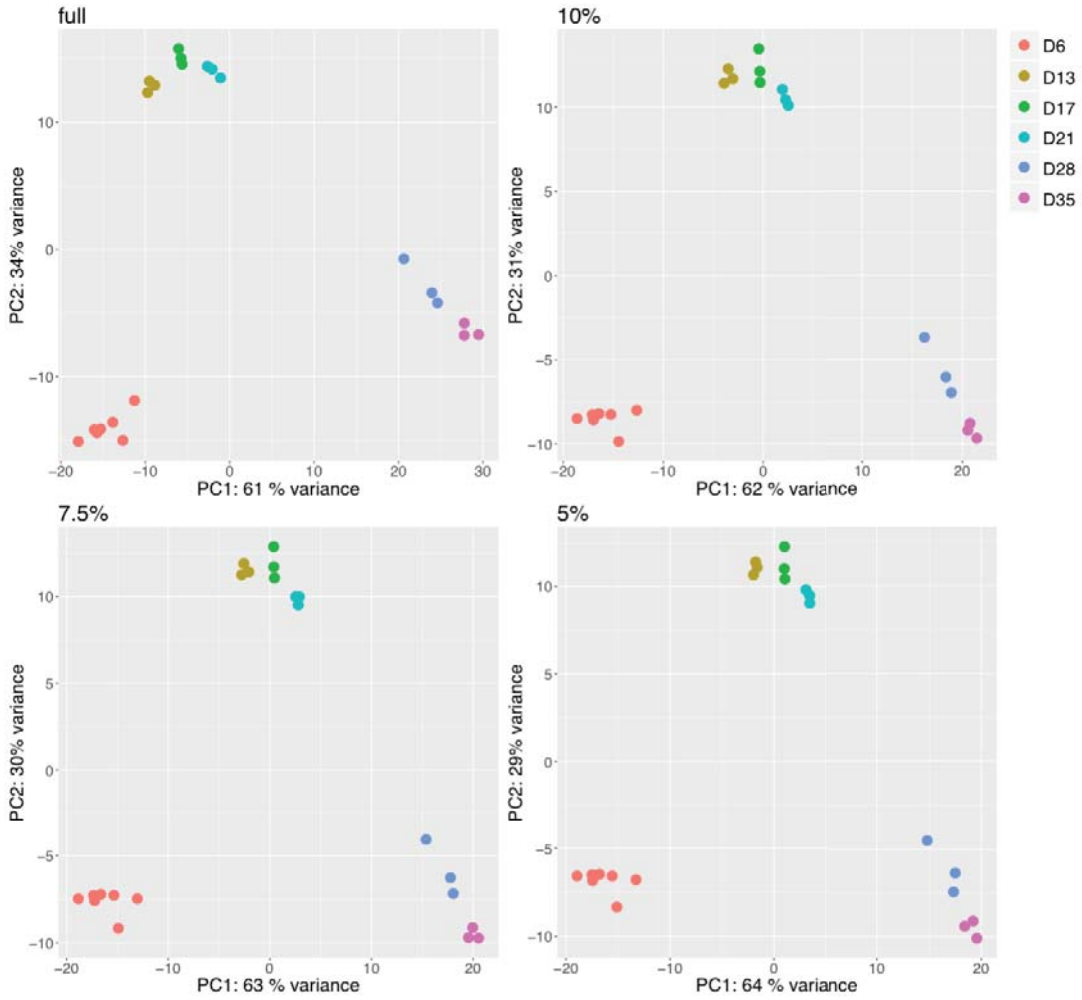


***Figure 3.4 PCA plots from total reads and down-sized reads***

PCA plot of samples that were proceeded to RNA-seq. Each dot is a pool of parasites from one mouse and represents one biological replicate. The percentages at the top left of each plot represent the proportion of down-sized reads, with 'full' being the whole dataset with all reads.

## 3.3.4 Sequencing data overall profiles

### *3.3.4.1 PCA of full* in vivo *dataset*

PCA was used to illustrate differences in transcriptomic profiles across all samples with larger distances between clusters representing larger variances in transcriptomic profiles. In this dataset, PCA plots show tight clustering of replicates for each time point and reassuringly indicate small variation between biological replicates compared with large variation between time points (Figure 3.4, full). All replicates from lung stage clustered separately from day-13 to day-21 groups. The closeness of day-13, 17, and 21 groups may be due in part to each sample representing developing stages that were dominated by similar gene expression. However, it is likely to also be due to the heterogeneity and morphological overlap between parasites from these three time points. Both day-28 and day-35 samples comprise adult parasites, many of which will have paired. The differences between them could be due to their stage of sexual maturity and egg laying which is onset at day 35.

Based on the distinctions between samples shown by the PCA, the following analyses grouped the samples into ‘lung stage’ (day-6), ‘liver stages’ (day-13, 17, and 21), and ‘adult stages’ (day-28 and 35).

### *3.3.4.2 Gene clustering by timecourse expression profile*

Using the whole dataset covering a timecourse of parasite development, I first sought a clustering method to group genes based on their changes in expression level over the timecourse. The clustering profiles were used for i) finding genes that had the biggest change over the timecourse, ii) inferring functions for genes of interest based on co-clustering genes with similar profiles, and iii) investigating genes that were expressed at high level in more than one time points (such as in liver stages). To this end, genes that were differentially expressed in at least one time point (adjusted p-value < 0.01, likelihood ratio test, 7987 genes) were grouped into 96 clusters and their similarity represented in Figure 3.5. Expression of genes in clusters are shown in Figure 3.6 (where y-axes were scaled based on the range of changes within a given cluster, which allows resolutions in clusters with smaller effect to show) and in Figure 3.7 (where all clusters are shown with the same y-axis, which allows clusters with the most extreme pattern to be seen). With this, two clusters showed the most striking changes and contained genes with expression specific to a particular time point.

Cluster 8 contains genes with high expression in the lung stage while expression levels in other time points are relatively low; cluster 96 contains genes with highest expression in adult stages (day 28 and day 35).

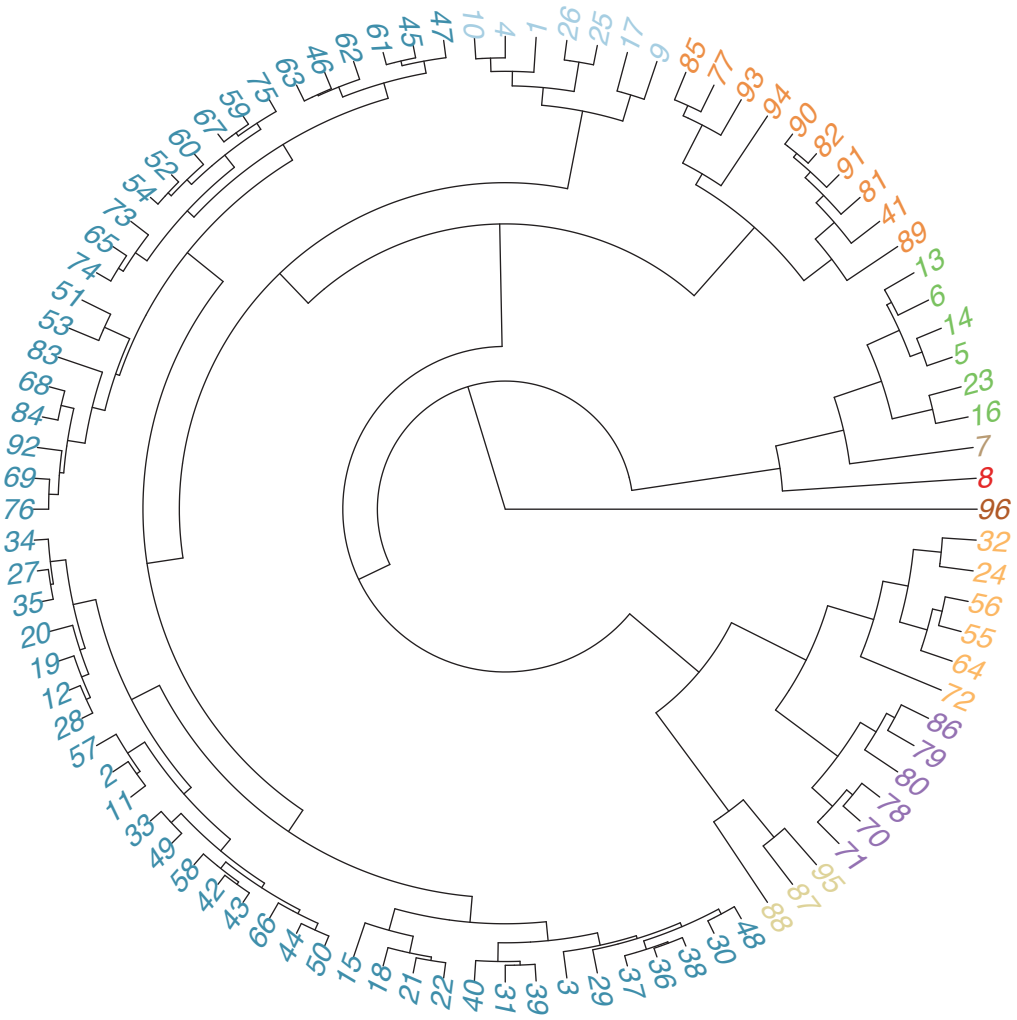


***Figure 3.5 Similarity between 96 clusters***

Clustering of genes based on their expression over the timecourse provides a representative profile for each of the 96 clusters. Hierarchical clustering was performed on the representative profiles and displayed on this dendrogram to illustrate relationship and similarity between clusters. The tip colours are for visualisation.

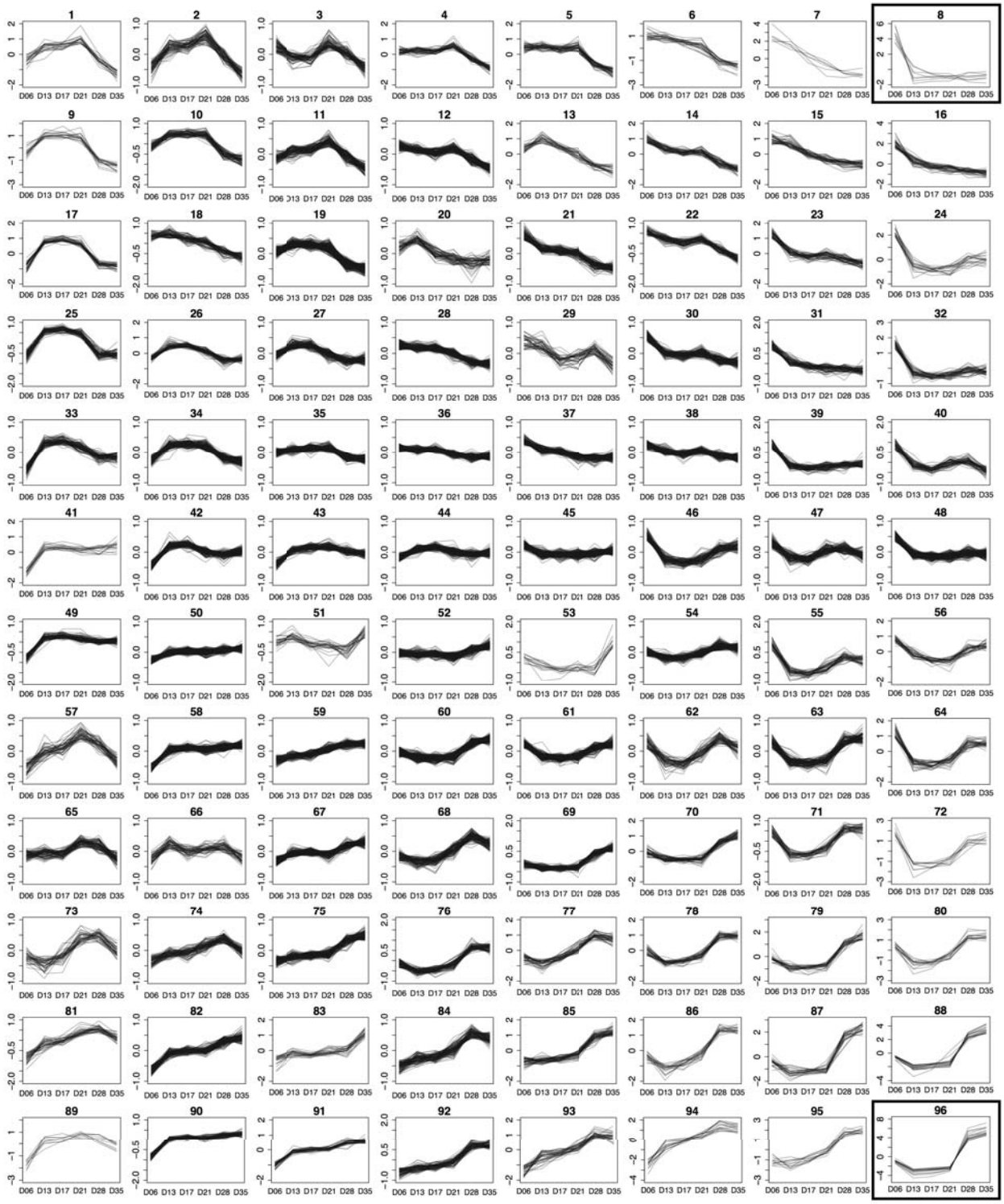


***Figure 3.6 Clusters of genes based on timecourse expression pattern***

Expression profile of genes differentially expressed in at least one time point clustered into 96 groups. The clustering was done on mean-normalised regularized log transformation (rlog transformed) of raw read counts. X-axes represent six time points from this dataset; y-axes represent the mean-normalised rlog transformed. Y-axes are not on the same scale so that differences between clusters can be displayed.

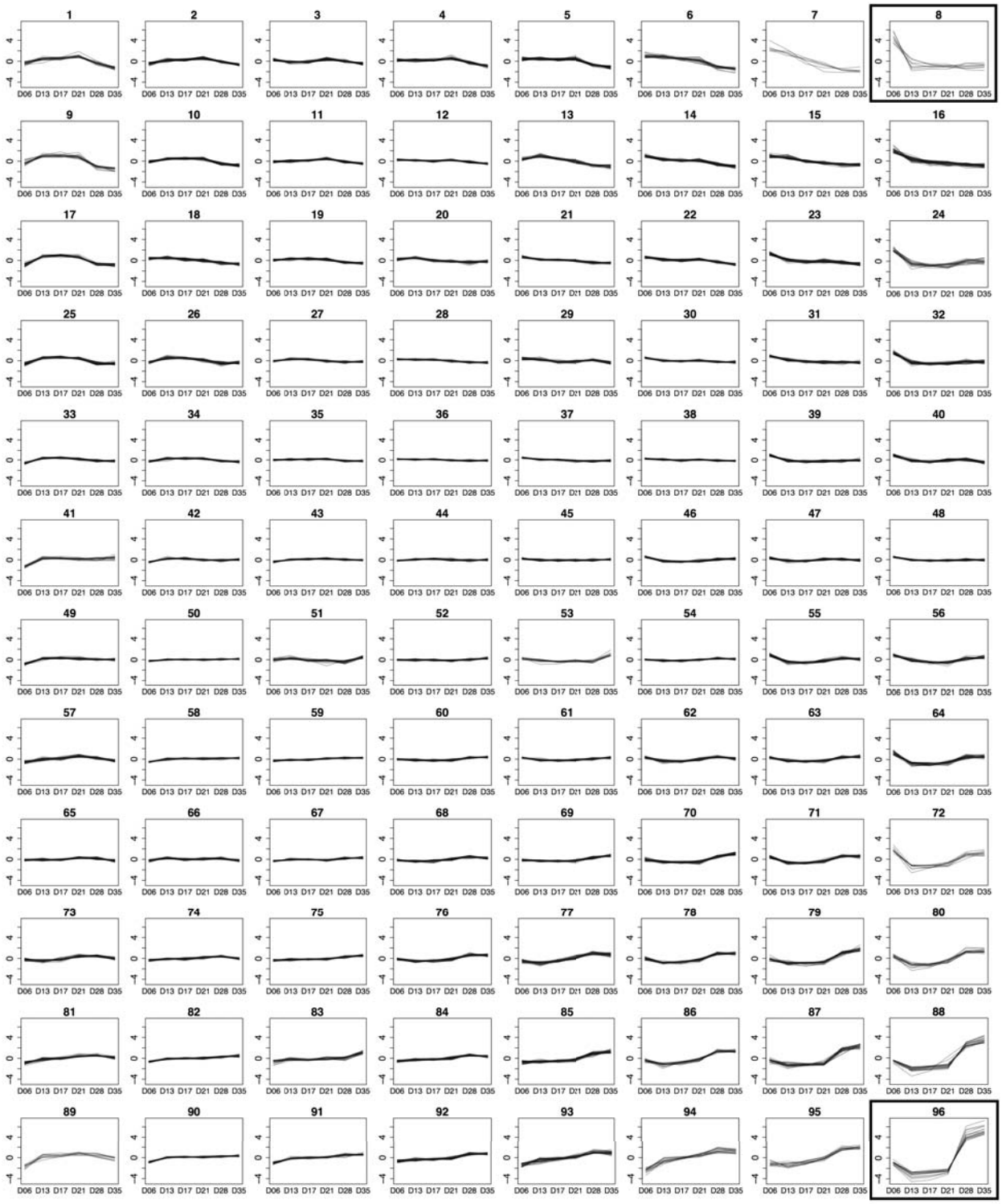


***Figure 3.7 Clusters of genes based on timecourse expression pattern, with fixed y-axes***

Timecourse expression profile of genes and clustering output as shown on Figure 3.6, but with the consistent range for y-axis across all clusters in order to visualise clusters with the largest changes over time.

## 3.3.5 Lung stage

### *3.3.5.1 Lung stage cluster*

The lung-specific clusters (cluster 8) contains 8 genes (Table 3.2). Four of these are *MEGs*, two from *MEG-2* family and two from *MEG-3* families. Signalling and the onset of development during the lung stage is indicated by the high expression of a calcium binding protein *calmodulin* (Smp_032970) and two adjacent genes annotated as *cat eye syndrome critical region proteins* (Smp_149860, Smp_149870) which in human has a role in development (Footz, 2001; Mitra et al., 2004). Genes responsible for developmental control are often tightly regulated and show temporal changes (Li et al., 2009). Expression of the genes with potential functions related to development could suggests that developmental processes were initiated at the lung stage even though mitosis has not occured (Clegg, 1965; Lawson and Wilson, 1980). Consistent with this, a group of clusters with high expression in lung stage and a steady decline toward adult (cluster 13,6,14,5,23,16,7) is enriched in genes with biological functions related to developmental control (Table 3.3). This includes several transcription factors; cell adhesion proteins involved in embryogenesis and neuronal development, such as *SOX* (Smp_148110, Smp_161600) and *procadherin family* (such as Smp_011430, Smp_141740, and Smp_155600) (Kamachi and Kondoh, 2013; Paulson et al., 2014); in addition *Wnt* and *frizzled receptors* that are important for cell-fate determination and control of development (Logan and Nusse, 2004; Park et al., 1994) were expressed in a similar pattern.

***Table 3.2 Genes in cluster 8 (highly expressed in lung stage)***

| Gene identifier | Product name |
|---|---|
| **Smp_032970** | calmodulin protein (calcium binding protein) |
| **Smp_138070** | MEG-3 (Grail) family |
| **Smp_138080** | MEG-3 (Grail) family |
| **Smp_149860** | cat eye syndrome critical region protein 5 |
| **Smp_149870** | cat eye syndrome critical region protein |
| **Smp_159800** | MEG-2 (ESP15) family |
| **Smp_159810** | MEG-2 (ESP15) family |
| **Smp_172460** | Krueppel factor 10 like |

*Table 3.3 Enriched GO terms of genes down-regulated after lung stage*

| GO identifier | GO description | p-value |
|---|---|---|
| **GO:0030238** | male sex determination | 0.0004 |
| **GO:0007223** | Wnt signaling pathway; calcium modulating pathway | 0.0004 |
| **GO:0030154** | cell differentiation | 0.0011 |
| **GO:0018958** | phenol-containing compound metabolic process | 0.0099 |
| **GO:0016055** | Wnt signaling pathway | 0.0145 |
| **GO:0006813** | potassium ion transport | 0.0174 |
| **GO:0007076** | mitotic chromosome condensation | 0.0178 |
| **GO:0007155** | cell adhesion | 0.0190 |
| **GO:0098609** | cell-cell adhesion | 0.0253 |
| **GO:0009072** | aromatic amino acid family metabolic process | 0.0332 |
| **GO:0006811** | ion transport | 0.0366 |

#### *3.3.5.2 Lung stage compared to early liver stage*

To further explore the lung stage, gene expression differences between the lung stage (day-6) with the next available stage after the parasite left the lung (day-13, early liver stage) was compared (Figure 3.8). Pairwise comparisons revealed 864 genes that were up-regulated in lung stage (Table 3.4, Table S3.1) compared to day-13 stage, and 686 genes were up-regulated in day-13 worms (Table 3.6, Table S3.2). Genes up-regulated in lung stages were further analysed by GO term enrichment (Figure 3.9, Table S3.3).

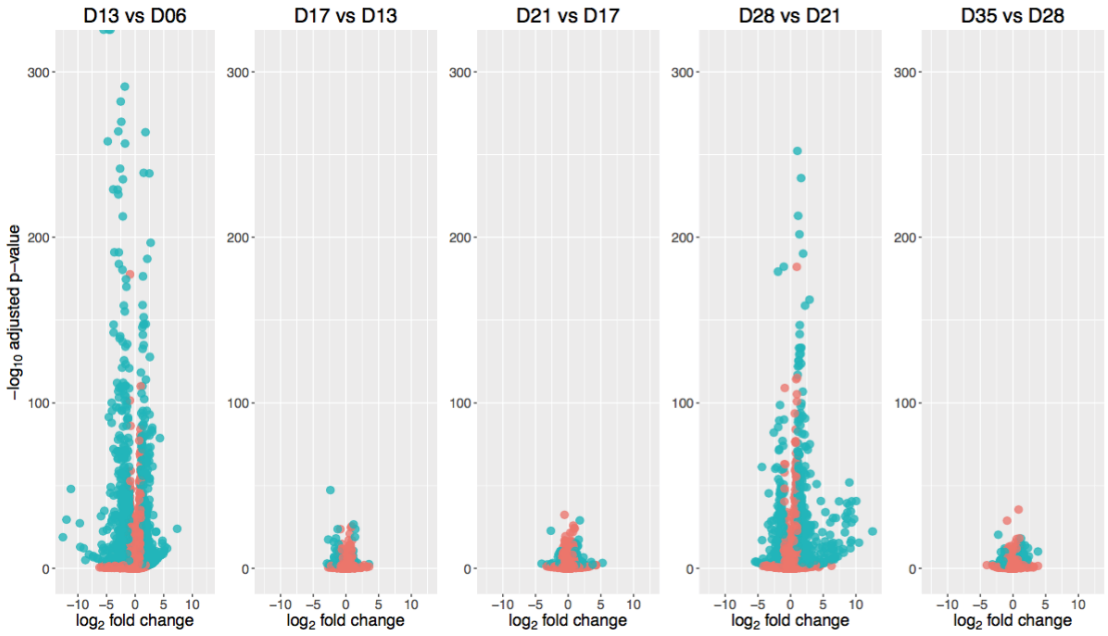


***Figure 3.8 Volcano plots of all consecutive pairwise comparison***

Pairwise comparisons between consecutive time points.. Each dot is a gene. Blue dots are differentially expressed genes (log2FC > 1 or < -1, and adjusted p-value < 0.01).

*Table 3.4 Top 20 genes up-regulated in day-6 compared to day-13 schistosomules*

| Gene identifier | $Log_2FC$ (day 13/day 6) | Adjusted p-value | Product name |
|---|---|---|---|
| **Top 20 genes up-regulated in day 6** | | | |
| **Smp_138080** | -12.62 | 1.48E-19 | MEG-3 (Grail) family |
| **Smp_138070** | -11.99 | 3.58E-30 | MEG-3 (Grail) family |
| **Smp_159810** | -11.22 | 1.02E-48 | MEG-2 (ESP15) family |
| **Smp_159800** | -9.66 | 5.10E-28 | MEG-2 (ESP15) family |
| **Smp_181510** | -9.58 | 1.15E-13 | hypothetical protein |
| **Smp_032990** | -8.98 | 7.14E-13 | Calmodulin 4 (Calcium binding protein Dd112) |
| **Smp_159830** | -8.69 | 1.00E-05 | MEG-2 (ESP15) family |
| **Smp_138060** | -8.06 | 3.76E-09 | MEG-3 (Grail) family |
| **Smp_203400** | -7.34 | 1.88E-07 | rhodopsin orphan GPCR |
| **Smp_005470** | -7.31 | 3.34E-08 | dynein light chain |
| **Smp_077610** | -6.61 | 6.19E-07 | hypothetical protein |
| **Smp_166350** | -6.59 | 1.31E-06 | hypothetical protein |
| **Smp_180330** | -5.97 | 2.91E-32 | MEG 2 (ESP15) family |
| **Smp_205660** | -5.84 | 4.52E-15 | hypothetical protein |
| **Smp_033250** | -5.80 | 4.43E-05 | hypothetical protein |
| **Smp_132500** | -5.73 | 4.50E-04 | ras and EF hand domain containing protein |
| **Smp_152730** | -5.68 | 3.41E-11 | histone lysine N methyltransferase MLL3 |
| **Smp_241430** | -5.61 | 3.89E-23 | Aquaporin 12A |
| **Smp_125060** | -5.55 | 6.70E-04 | kinase suppressor of Ras (KSR) |
| **Smp_198060** | -5.51 | 4.73E-13 | hypothetical protein |

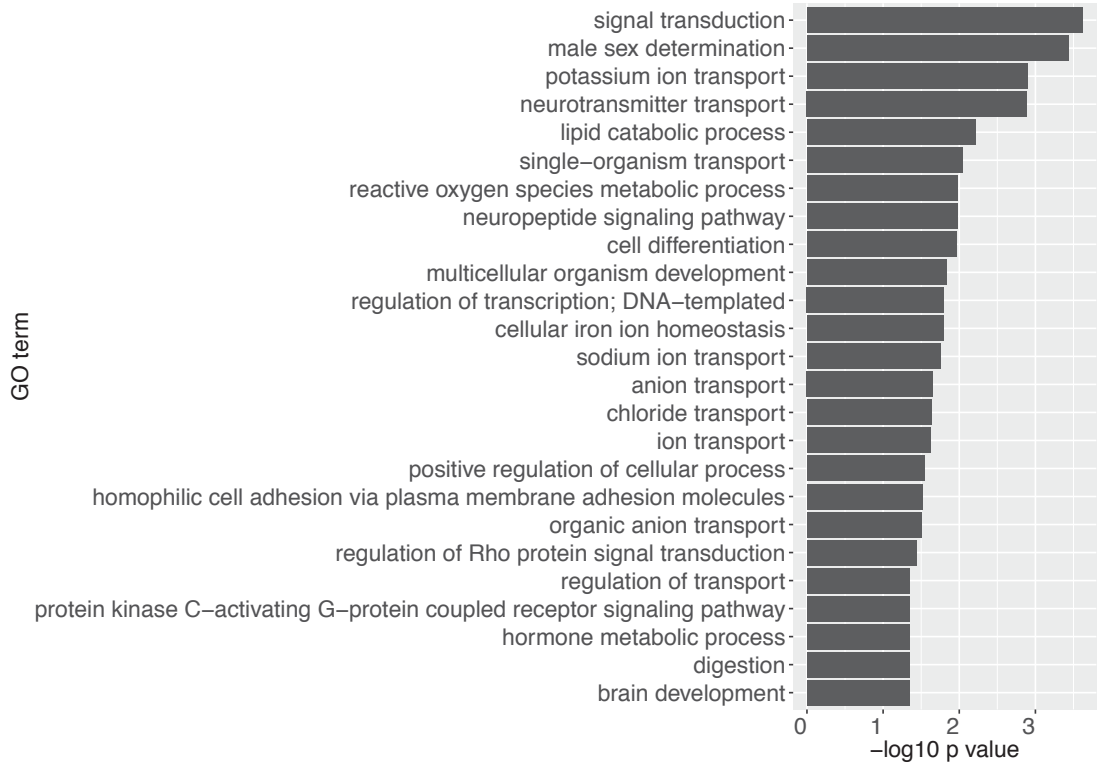


***Figure 3.9 GO enrichment of genes up-regulated in day-6 compared to day-13 schistosomules***
Bar chart shows enriched GO terms (biological processes terms) of genes that were up-regulated in day-6 compared to day-13 schistosomules, ranked by p-values obtained from the topGO package.

Signalling

Multiple signalling processes appeared to be up-regulated in the lung stage. The top four enriched GO terms were *signal transduction*, *male sex determination*, *potassium ion transport*, and *neurotransmitter transport*. Other enriched GO terms, though less significant, suggest that signalling processes may be up-regulated in lung stage compared to the early liver stage (day 13). Examples of these GO terms are *PKC-activating GPCR signalling pathway*, *regulation of Rho protein signal transduction*, *hormone metabolic process*. Such signalling may be related to responding to environment, or it could be part of developmental control as GO terms on development were also enriched; for example, *cell differentiation*, *homophillic cell adhesion*, *brain development*, *male sex determination*. Other types of signalling included potential neuromuscular signalling inferred from enriched GO term *neuropeptide signalling pathway*, *sodium ion transport*, *chloride transport*, and one of the top GO term mentioned previously, *neurotransmitter transport* (Figure 3.9). This may suggest an increased requirement for neuronal activities in lung stage compared to day-13 early migratory stage for locomotion that allow the worms to migrate out of

the lung and travel within the bloodstream. Consistent with the GO term enrichment, the top up-regulated genes in lung stage compared to the early liver stage included a *rhodopsin orphan GPCR* (Smp_203400), and two Ras related proteins (Smp_132500, Smp_125060) - all were up-regulated by more than 32 fold (Table 3.4).

*MEGs*

In total, there were 17 *MEGs* up-regulated in lung stage compared to day-13 schistosomules, with seven *MEGs* being among the top-20 lung-stage up-regulated genes. These are *MEG-2* (Smp_159810, Smp_159800, Smp_159830, Smp_180330) and *MEG-3* groups (Smp_138080, Smp_138070, Smp_138060). Other classes of *MEGs* were also up-regulated in lung stage compared to day-13 schistosomules; these were *MEG-10*, *MEG-14*, *MEG-15*, *MEG-17*, and other genes in *MEG-2* and *MEG-3* classes (Figure 3.10, Table S3.1). Some *MEGs*, such as *MEG-3* and *MEG-14*, were previously identified in the oesophagus of schistosomules and adults and were proposed to have a role in blood feeding and/or interactions with host through *MEGs* secreted into host environment (DeMarco et al., 2010; Orcia et al., 2016; Philippsen et al., 2015; Wilson, 2012).

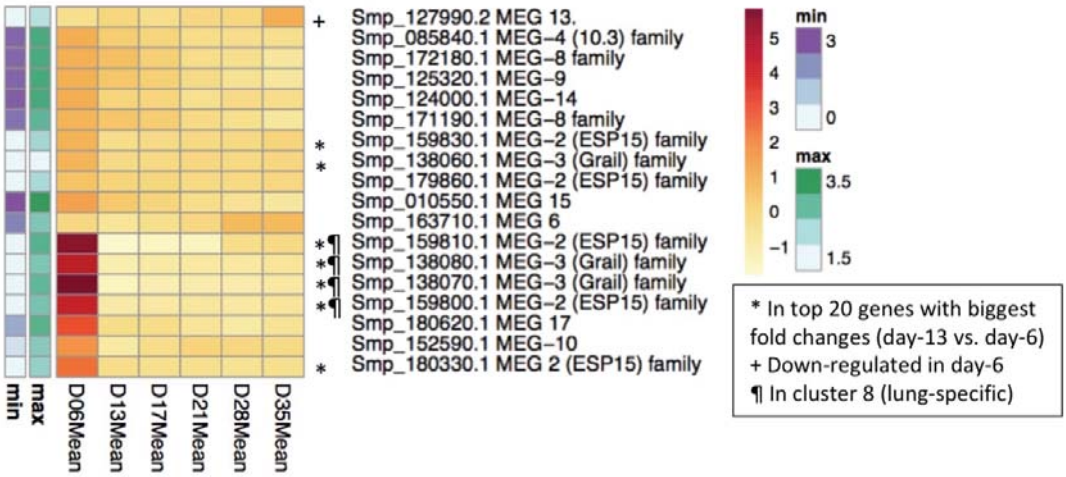


***Figure 3.10* MEGs *differentially expressed between day-6 and day-13 schistosomules***

*In vivo* timecourse expression of the *MEGs* that were differentially expressed between day-6 and day-13 schistosomules. The expression levels are shown as mean-normalised rlog-transformed read counts for each time point. Annotation columns adjacent to the heatmap (min, max) indicate expression level ($\log_{10}$ FPKM) - deeper colours indicate higher expression levels.

Iron homeostasis

Blood feeding is thought to start after the lung stage (Crabtree and Wilson, 1986a, 1980) and by day-13 schistosomules enter a growth spurt which requires iron. Intriguingly, however, GO term *cellular iron ion homeostasis* was enriched in genes up-regulated in lung stage compared to day-13 schistosomules (Figure 3.9). Genes contributing to this enriched GO term are *ferritin* genes (Smp_047650, Smp_047660, Smp_047680; fold change range between 2.1 - 7.2) (Table S3.1, Table S3.3). These three *ferritin* genes, although down-regulated in day-13 compared to day-6 schistosomules, were still expressed at low level during day 13 when iron is required for growth and development. Other genes related to iron-sequestration were expressed at a similar level in the lung stage and day-13 stage, such as *putative ferric reductase* (Smp_136400) and *DMT 1* (Smp_099850.3). The putative ferric reductase is hypothesised to cleave iron from host transferrin (glycoprotein iron carrier) before transporting into the parasite via a DMT (Glanfield *et al*., 2007; Smyth *et al*., 2006). This suggests that the lung stage may require additional iron storage for other purposes, possibly for sequestering iron away from immune cells.

Reactive oxygen species (ROS) and oxidative stress

Lung stage worms need to adapt to a unique environment where oxygen pressure is high and migration in tight capillaries damaging host cells can stimulate inflammation. Molecular $O_2$ can be reduced to ROS. In addition, ROS are involved in signalling of immune response and the neutralisation of ROS could be an immune evasion strategy of the parasites. Regulation of oxidative stress and ROS appeared to be up-regulated in lung stage. Two of *extracellular superoxide dismutase* were up-regulated in lung stages compared to day-13 stage (Smp_095980 and Smp_174810; 30 and 17 fold, respectively) (Table S3.1), which is consistent with an enriched GO term *ROS metabolic process*. Extracellular superoxide dismutase catalyse the change of ROS into molecular $O_2$ or into $H_2O_2$ which is less toxic than ROS (Afonso *et al*., 2007). Further, the product of *antioxidant thioredoxin peroxidase* (Smp_059480) has been previously characterised for its ability to neutralise $H_2O_2$ (Kwatia *et al*., 2000) and was more than 16-fold up-regulated in the lung stage compared to day-13 parasites (Table S3.1).

Kunitz protease inhibitor

Relating to immune evasion, a *single kunitz serine protease inhibitor* (Smp_147730) was up-regulated in lung stage by more than 25-fold (Table S3.1). Kunitz protease inhibitors have previously been shown to have anticoagulation and anti-inflammation properties by inhibiting FXa and kallikrein which are involved in coagulation cascade (Ranasinghe *et al.*, 2015). Its up-regulation in the lung stage could function to mediate inflammation and prevent coagulation which could result from schistosomules impeding blood flow and damaging the endothelial walls of blood vessels during migration.

#### *3.3.5.3 Expression of lung-stage anti-inflammatory genes*

Some genes that could be involved in interactions with host immune defence were up-regulated in lung stage compared to day 13. When considered over the timecourse, they were down-regulated during the liver stage (day 13 to day 21) and then up-regulated again at day 28. The gene *extracellular superoxide dismutase* (Cu Zn) (Smp_174810), *thioredoxin peroxidase* (Smp_059480), and *single kunitz protease inhibitor* (Smp_147730) fell into this expression pattern. Considering the nature of these genes and the expression pattern they shared, genes with similar expression pattern were explored further. Using results from the clustering done previously (section 3.3.4.2), both genes were grouped in cluster 72. With the potential roles in host immune evasion of the two genes, other genes in cluster 72 were investigated further.

Cluster with anti-inflammatory genes

Cluster 72 contains 10 genes (Table 3.5). In addition to three genes mentioned previously for host immune evasion, expression of *arginase* (Smp_059980) was hypothesised to be counteracting host immune response by depleting l-arginine from blood, thereby preventing l-arginine being used by macrophage in the production of nitric oxide (Fitzpatrick et al., 2009b; Hai et al., 2014; Ranasinghe et al., 2015a). A schistosome homologue of *early growth response protein* (EGR, Smp_134870) contains all of the expected domains – *Zinc finger, C2H2-like* (IPR015880), *Zinc finger C2H2-type/integrase DNA-binding domain* (IPR013087), and *Zinc finger, C2H2* (IPR007087) – and may therefore interfere with the endogenous function of host EGR-1 which is involved in expression of IL-2, IL-4, and TNF-α (Decker *et al.*,

2003; Lohoff *et al*., 2010). Other genes in cluster 72 are not as noticable for their roles in host immune evasion.

***Table 3.5 Genes in cluster 72***

| Gene identifier | Product name |
|---|---|
| **Smp_059480** | thioredoxin peroxidase |
| **Smp_059980** | arginase |
| **Smp_074570** | hypothetical protein |
| **Smp_114660** | hypothetical protein |
| **Smp_134870** | early growth response protein |
| **Smp_147730** | single kunitz protease inhibitor; serine type protease inhibitor |
| **Smp_156510** | PDZ and LIM domain protein 7 |
| **Smp_166920** | PDZ and LIM domain protein Zasp |
| **Smp_174810** | Extracellular superoxide dismutase (Cu Zn) |
| **Smp_182770** | hypothetical protein |

Hypothetical genes in the cluster

Given five out of 10 genes in this cluster have potential roles in interactions with host immune responses, the three that encode hypothetical proteins (Smp_074570, Smp_114660, Smp_182770) were investigated further. Peptides from Smp_074570 are abundant (top 10) in extracellular vesicles produced by *S. mansoni* schistosomules (Nowacki *et al*., 2015). No further published information was found for the other two hypothetical proteins and both contained no known signature domains. I-TASSER protein-structure prediction software was used to align putative structures against known structures in the Protein Data Bank (PDB) (Berman *et al*., 2000). No match was found for Smp_074570 and Smp_114660 but, surprisingly, Smp_182770 produced a single high confidence match (TM-score > 0.9) to human complement factor H (CFH) (PDB:3GAW; Okemefuna *et al*., 2009) (Figure 3.11).

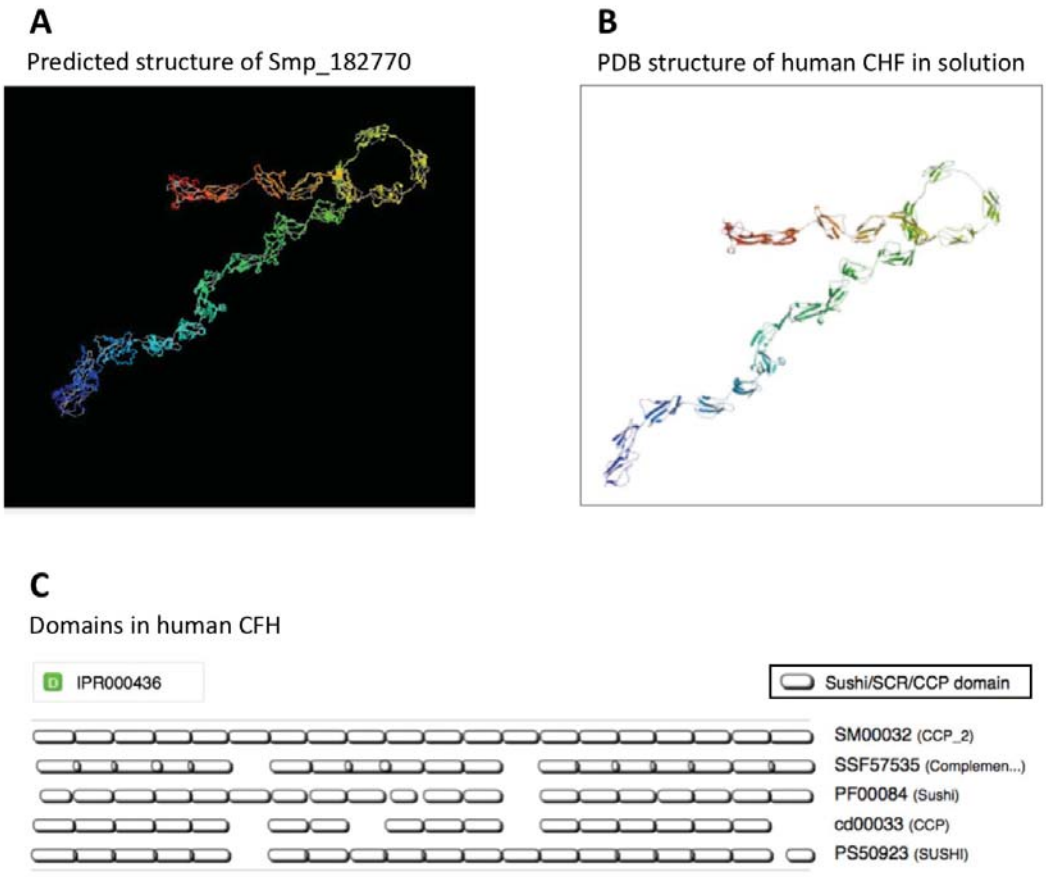


***Figure 3.11 Predicted structure of Smp_182770 aligned with structure of human CFH***

A) Predicted 3D structure of Smp_182770 by I-TASSER server from the input amino acid sequence. B) 3D structure of human CFH obtained from PDB (PDB identifier: 3GAW). C) Domain component of human CHF identified by multiple databases through InterProScan. SCR, short consensus repeat; CCP, complement control protein. Both SCR and CCP are alternative name of the sushi domain.

Putative CFH

CFH is a regulator of the complement cascade and is normally found on the surface of human cells. CFH is a cofactor of complement factor I (CFI) and together, they prevent complement attack on self-cells. While human CFH (amino acid sequence from 3GAW entry, PDB) contains *sushi/ccp* domain repeats that are involved in regulating complement cascade (ccp = component control protein), Smp_182770 does not (Figure 3.11). The gene does however encode a pattern of tandem repeats somewhat similar to those expected from mammalian CFH genes. Given the nature of other genes in this cluster and the structural match to human CFH, the *S. mansoni* hypothetical protein might mimic the function of host CFH and inhibit activation of

complement cascade, preventing membrane attack complex from forming and damaging the parasite surface. In addition, amongst other helminths, homologues of the Smp_182770 are present in other *Schistosoma* species and liver flukes (*Fasciola*, *Echinostoma*, *Opisthorchis*, and *Clonorchis*) (Figure 3.12A). Smp_182770 has a paralogue (Smp_038730) that is immediately adjacent to it on the genome and RNA-seq mapping suggest that the two genes might in fact be part of the same gene (Figure 3.12B). The paralogue Smp_038730 is in cluster 24, which is part of neighbour branches of cluster 72 and contains genes with high expression in the lung stage (Figure 3.5). Two of the ferritin, previously discussed for possible roles in scavenging iron in lung to prevent it from being used by host immune system and prevent production of ROS, are also in cluster 24. Potentially, more genes in the neighbouring cluster could be explored individually for possible roles in parasite survival and evasion of host immune responses. Furthermore, the nature of genes with this expression profile suggests that defence against the immune system is very important in the lung stage.

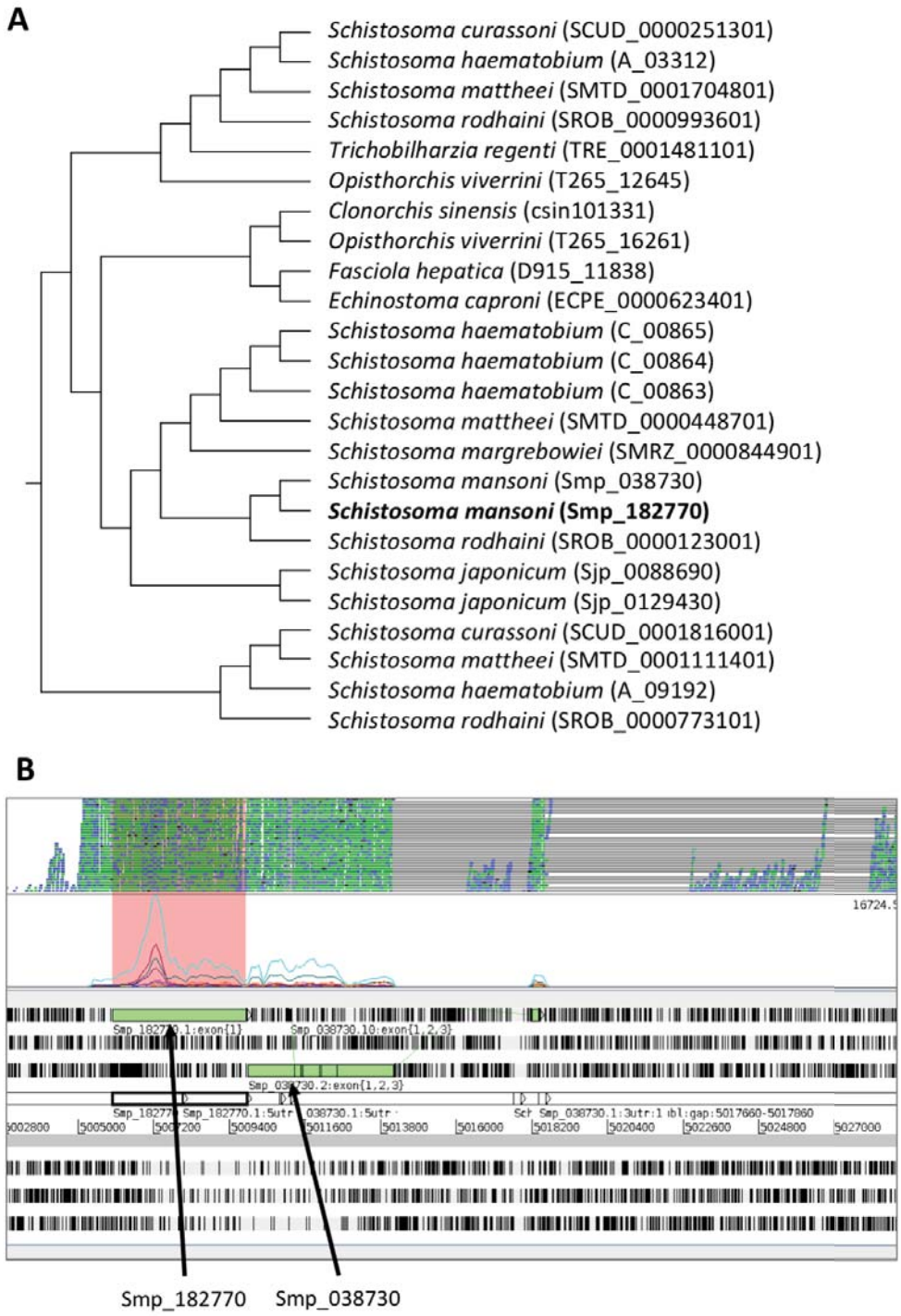


***Figure 3.12 Homologous relationship of Smp_182770 and alignment of RNA-seq reads to genomic locations***

A) Homologues of Smp_182770. The information was obtained from WormBase ParaSite release 8 (Howe *et al.*, 2016) and the tree was regenerated for clarity. Gene identifiers from the WormBase ParaSite are shown in the parentheses after the species names. B) Genomic region where Smp_182770 is located and mapping of RNA-seq reads from multiple RNA-seq libraries (two top rows), showing the location adjacent to its paralogue Smp_038730.

#### *3.3.5.4 Lung stage conclusion*

In summary, lung-stage gene expression is dominated by an up-regulation of multiple signalling processes – including those involved in developmental control, cell differentiation and neuropeptide signalling – and likely reflects the onset of development and motility control in schistosomules in the lung capillary network. Multiple *MEGs* were up-regulated specifically in the lung stage. However, MEG function is unclear despite host-parasite interaction roles having been postulated (DeMarco et al., 2010; Wilson, 2012). Immune modulation and evasion might be increased for the lung stage as reflected in up-regulation of iron homeostasis, oxidative stress-related genes, and anti-inflammatory genes. Further investigation into immune-related genes led to identification of putative regulator of complement factor, CFH. A set of genes with immune-related function appeared to be down-regulated in liver stages compared to the lung stage. Although they were up-regulated again in adult stages, the expression in lung stage was highly up-regulated, reflecting resistance of the lung stage to immune-mediated cytotoxic killing (Bentley *et al*., 1981; Crabtree and Wilson, 1986b; Mastin *et al*., 1985; McLaren and Terry, 1982). The lower expression level compared to the lung stage could imply that additional mechanisms may be in place and organs or cell types that only exist in adult stages may be involved.

### 3.3.6 Liver stages

#### *3.3.6.1 Rationale*

After the worms leave the lung, they enter the circulation through the cardiac output and can be found in various tissues (Bloch, 1980; Crabtree and Wilson, 1980; Wheater and Wilson, 1979). Schistosomules can be observed in the liver from day 14 post-infection and the number of schistosomules found in the liver increases until day 21, during which time the parasites develop into adult form (Clegg, 1965). Clusters of genes with prominent expression during the liver stages were investigated to seek insight into signals of liver localisation/tropism and interaction with the liver environment. Two approaches were used for the investigation: i) genes that were expressed mainly during liver stages (clusters with high expression between day 13 to day 21) and ii) changes in gene expression as the parasites enter and leave the liver (pairwise comparison between day 13 and lung stage, and pairwise comparison

between day 21 and day 28). Due to the developmental heterogeneity between days 13 and 21, pairwise comparisons would be prone to noises and were not examined.

#### *3.3.6.2 Liver stage cluster*

Selecting clusters with expression mainly in liver stages, cluster 17 was chosen (Figure 3.6) with its six neighbouring clusters (Figure 3.13A), covering 356 genes (Table S3.4). Analysis of enriched GO terms amongst these genes suggested that many had a role in growth and cell division, with the top five GO terms related to cell replication and regulation of related processes (Figure 3.13B). The term *neuropeptide signalling pathway* was also enriched and contained genes *lipoprotein receptor* (Smp_099670) and *neuropeptide Y prohormone 1* (Smp_159950) (Table S3.5). In the free-living flatworm *S. mediteranea*, neuropeptide Y is involved in developmental control (Collins *et al*., 2010), thus the schistosome homologue may be involved in control of development during liver stages.

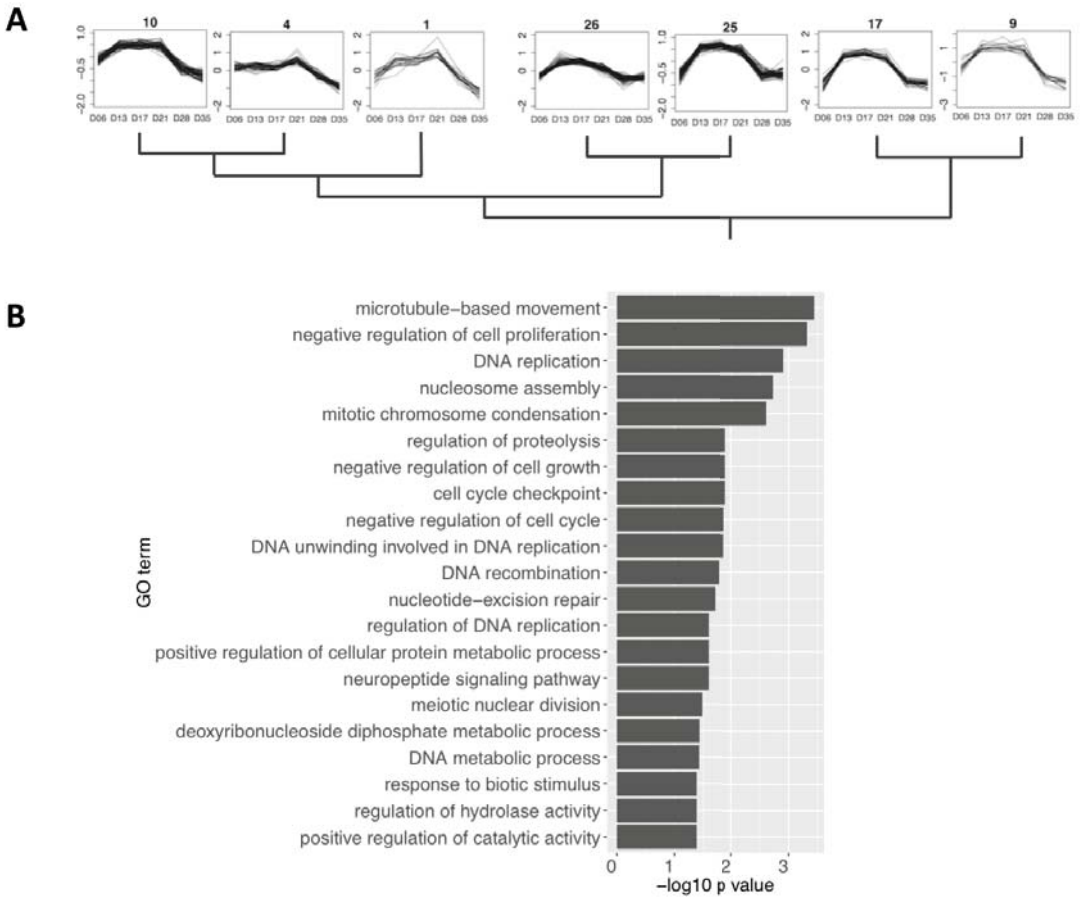


***Figure 3.13 Clusters of genes with high expression in liver stages and GO term enrichment***

**A**) Genes with high expression in the liver stages were selected for functional analysis. Cluster 17 from Figure 3.6 was selected as the best cluster fitting this pattern. Neighbouring clusters to the cluster 17 according to the dendrogram (Figure 3.5) are cluster 1,4,9,10,25, and 26. Genes in the clusters were pooled for functional analysis using GO term enrichment. **B**) Bar chart shows enriched GO terms (biological processes) of the genes with high expression in the liver stages, ranked by p-values obtained from topGO package.

### *3.3.6.3 Entering the liver*

Growth, developmental control, cellular respiration

Similarly, among 686 genes up-regulated in early liver stage compared to lung stage (Figure 3.8, Table S3.2), the most striking effect were up-regulation of mitotic cell division and its associated processes such as translation, post-translational modification, and transcriptional regulation (Figure 3.14). A top up-regulated gene *Rootletin* (Smp_147890), with possible role in mitosis (Bahe *et al*., 2005), was up-

regulated over 12-fold (Table 3.6). This is expected for the developmental period in the liver and consistent with the report that mitosis start in the liver stage once the gut has formed (Clegg, 1965; Lawson and Wilson, 1980). Homologues of genes with potential involvement in developmental control were also up-regulated; for example, *glial cells missing* gene (Smp_171130), *prospero homeobox protein 2* (Smp_045470), *and IQ domain containing protein D* (Smp_161310). The major tegument components *tegumental-allergen-like protein 3* (*TAL-3*; Smp_086530) and *25 kDa integral membrane protein* (Smp_154180) were up-regulated by more than 16 fold (Table 3.6). Similarly, synthesis of cholesterol – another major part of the tegument – was up-regulated, including two key enzymes in supplying the chemical backbone for sterol synthesis: *farnesyl pyrophosphate synthase* (Smp_070710), and *HMG-CoA synthase (*Smp_198690) (Table S3.2, Table S3.3). Consistent with rapid growth, genes related to cellular respiration were also up-regulated, such as genes encoding components of the Krebs cycle (or Citric acid cycle): *malate dehydrogenase* (Smp_129610) and *isocitrate dehydrogenase* (Smp_163050); and genes encoding components of cellular respiration: *cytochrome b c1 complex subunit 6* (Smp_059930), *ubiquinol cytochrome c reductase* (Smp_061870) (Table S3.2, Table S3.3).

*Table 3.6 Top 20 genes up-regulated in day-13 compared to day-6 schistosomules*

| Gene identifier | $Log_2FC$ (day 13/day 6) | Adjusted p-value | Product name |
|---|---|---|---|
| **Top 20 genes up-regulated in day 13** | | | |
| **Smp_123200** | 7.35 | 1.17E-24 | hypothetical protein (now as MEG-32.2) |
| **Smp_154180** | 5.59 | 9.20E-13 | 25 kDa integral membrane protein |
| **Smp_113760** | 5.21 | 7.69E-14 | anti-inflammatory protein 16 |
| **Smp_087310** | 5.08 | 2.03E-11 | hypothetical protein |
| **Smp_060220** | 5.05 | 1.01E-10 | lipopolysaccharide induced tumor necrosis |
| **Smp_171130** | 4.9 | 2.19E-09 | glial cells missing |
| **Smp_158480** | 4.78 | 1.84E-17 | AMP dependent ligase |
| **Smp_045470** | 4.74 | 7.99E-16 | prospero homeobox protein 2 |
| **Smp_127270** | 4.51 | 1.11E-08 | Cytochrome b561:ferric reductase |
| **Smp_161310** | 4.5 | 8.60E-11 | IQ domain containing protein D |
| **Smp_162370** | 4.48 | 1.42E-08 | histone H1 gamma |
| **Smp_086530** | 4.34 | 1.80E-79 | tegument-allergen-like protein |
| **Smp_123790** | 3.95 | 9.90E-07 | glypican |
| **Smp_011030** | 3.84 | 3.66E-08 | heme binding protein 2 |
| **Smp_099670** | 3.82 | 1.02E-06 | lipoprotein receptor |
| **Smp_162860** | 3.69 | 5.95E-23 | hypothetical protein |
| **Smp_147890** | 3.61 | 3.57E-11 | Rootletin |
| **Smp_125250** | 3.59 | 9.97E-23 | hypothetical protein |
| **Smp_145760** | 3.57 | 2.89E-05 | transmembrane protein 62 |
| **Smp_200900** | 3.48 | 2.15E-05 | hypothetical protein |

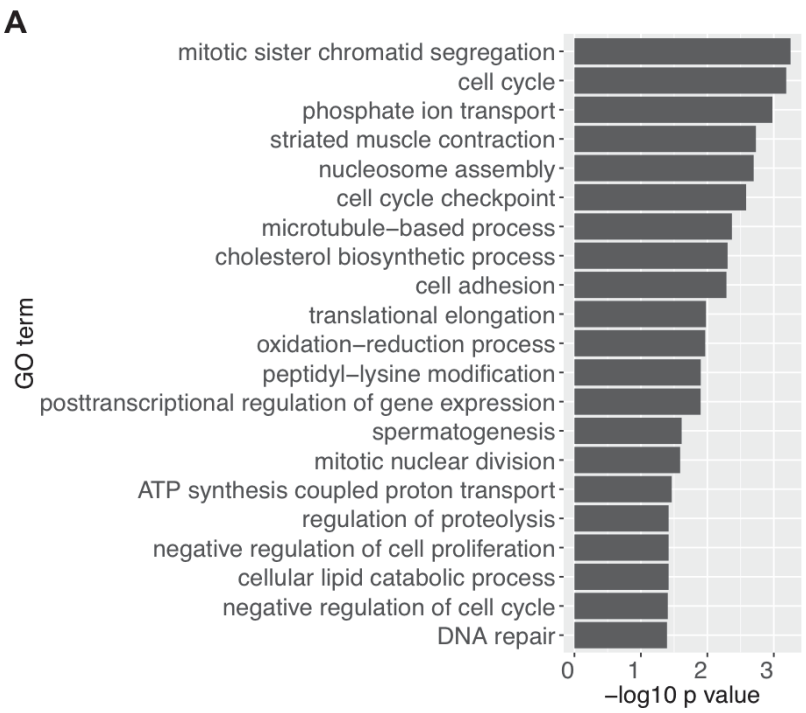


***Figure 3.14 GO enrichment of genes up-regulated in day-13 compared to day-6 schistosomules***

Bar chart shows enriched GO terms (biological processes) of genes up-regulated in day-13 compared to day-6 schistosomules, ranked by p-values obtained from topGO package.

Aldehyde metabolism

Two *aldehyde dehydrogenase* genes were up-regulated at day 13 compared to lung stage (Smp_022960 (2 fold), Smp_050390 (4 fold) (Table S3.2). The liver is a major site of aldehyde detoxification by oxidation to supply acetate for energy metabolism (Cederbaum, 2012). Aldehyde may therefore be taken up from the host, oxidised to acetate and acetyl-coA by the parasite's aldehyde dehydrogenase, supplying parasite acetyl-coA for use in the Krebs cycle, which is also upregulated (Table S3.2, Table S3.3).

Host-parasite interactions

Regarding host-parasite interactions, three genes were among the top up-regulated genes at day 13. *MEG-32.2* (Smp_123200) was up-regulated 157-fold; anti-inflammatory protein Sm16 (Smp_113760) (Rao and Ramaswamy, 2000) and *LPS-induced tumor necrosis factor alpha* homologue (Smp_060220) were up-regulated by over 32 fold (

Table 3.6). All three genes were also expressed at similar levels in day 17 and day 21. *MEG-32.2* was previously identified to be more highly expressed in the head part of adult *S. mansoni* but no function has yet been identified (Wilson *et al*., 2015).

It is intriguing that Sm16 was up-regulated in day-13 worms and in the liver stages but barely expressed in lung stage, because this suggests that lung-stage and liver-stage parasites employ different strategies for immune evasion. Up-regulation of the *LPS-induced tumor necrosis factor alpha* homologue is interesting because this is a proinflammatory cytokine in human (Ploder *et al*., 2006), while the parasites would be expected to inhibit inflammation. The *S. mansoni* gene contains relevant domain (*LPS-induced tumour necrosis factor alpha factor* (IPR006629)) which is normally associated with responses to LPS from bacteria. Its up-regulation in the liver stage might have roles in interacting with CD4+ cells which promote growth of *S. mansoni* (Riner et al., 2013).

#### *3.3.6.4 Leaving the liver*

Genes up-regulated in late liver stage (day-21) parasites compared to pre-pairing adult stage (day-28) are involved in cell division, differentiation, developmental regulation (Table 3.7, Table S3.6). This is expected because, by day 28, the parasites have already morphed into adult stages, and their major tissues established. Though there is further development upon pairing, this is clearly minimal compared to the development during liver period.

***Table 3.7 Enriched GO terms (biological process) of genes up-regulated in day-21 compared to day-28 worms***

| GO identifier | GO description | p-value |
|---|---|---|
| **GO:0007067** | mitotic nuclear division | 0.0002 |
| **GO:0008285** | negative regulation of cell proliferation | 0.0006 |
| **GO:0030154** | cell differentiation | 0.0011 |
| **GO:0007223** | Wnt signaling pathway; calcium modulating pathway | 0.0011 |
| **GO:0030245** | cellulose catabolic process | 0.0011 |
| **GO:0006281** | DNA repair | 0.0020 |
| **GO:0007076** | mitotic chromosome condensation | 0.0030 |
| **GO:0051276** | chromosome organization | 0.0043 |
| **GO:0007530** | sex determination | 0.0046 |
| **GO:0006259** | DNA metabolic process | 0.0046 |
| **GO:0006260** | DNA replication | 0.0097 |
| **GO:0007018** | microtubule-based movement | 0.0102 |
| **GO:0006941** | striated muscle contraction | 0.0111 |
| **GO:0000075** | cell cycle checkpoint | 0.0146 |
| **GO:0006268** | DNA unwinding involved in DNA replication | 0.0164 |
| **GO:0040007** | growth | 0.0211 |
| **GO:0006695** | cholesterol biosynthetic process | 0.0214 |
| **GO:0007126** | meiotic nuclear division | 0.0378 |
| **GO:0016055** | Wnt signaling pathway | 0.0420 |

#### *3.3.6.5 Liver stage conclusion*

Overall, the liver stage transcriptome represented a period of growth coupled with control of development and increase in energy requirement. Further, biosynthesis of cholesterol and expression of genes encoding tegumental proteins reflect the increasing tegument surface. Although host-parasite interactions did not seem to be a major feature of the transcriptome, three potent genes involved in host-parasite interactions were up-regulated in liver stages compared to lung stage, with one potentially stimulating inflammation.

### 3.3.7 Adult stages

#### *3.3.7.1 Becoming adults*

Egg shell proteins, metabolic requirements, and biosynthesis processes

As the parasites developed into adults (pairwise comparison between days 28 and 21), an expected massive up-regulation was seen in the expression of genes involved in

egg production such as *tyrosinase* (Smp_050270, Smp_013540), *eggshell protein* (Smp_000430), and *Trematode Eggshell Synthesis domain containing protein* (Smp_077890) (Table 3.8, Table S3.7). In addition, metabolic processes and biosynthesis processes were up-regulated at day 28 (Table S3.8). Some of these may reflect production of compounds for egg production (such as organic hydroxy compound biosynthetic process, mainly *tyrosinase* genes), while some may reflect increased nutrient and energy requirements, and scavenging of host-derived substrate by the parasites. Enriched GO terms related to this include *carbohydrate transport* (including three genes encoding glucose transporters, Smp_012440, Smp_105410, Smp_139150, and the function of the first two has been confirmed (Krautz-peterson et al., 2010), *biosynthesis processes*, *lipid metabolic process*, *glycerol metabolic process*, and *purine ribonucleoside salvage* (Table 3.9, Table S3.8).

***Table 3.8 Top 20 genes up-regulated in day-28 compared to day-21 worms***

| Gene identifier | $Log_2FC$ (day 28/day 21) | Adjusted p-value | Product name |
|---|---|---|---|
| **Top 20 genes up-regulated in day 28** | | | |
| **Smp_050270** | 12.53 | 4.47E-23 | Tyrosinase |
| **Smp_000270** | 10.47 | 2.92E-16 | hypothetical protein |
| **Smp_138570** | 10.04 | 2.07E-41 | spore germination protein |
| **Smp_095980** | 9.72 | 1.46E-30 | Extracellular superoxide dismutase (Cu Zn) |
| **Smp_000430** | 9.36 | 2.86E-41 | eggshell protein |
| **Smp_191910** | 9.33 | 1.93E-28 | Stress protein DDR48 |
| **Smp_077890** | 9.31 | 1.08E-17 | Trematode Eggshell Synthesis domain containing protein |
| **Smp_131110** | 9.22 | 3.63E-35 | hypothetical protein |
| **Smp_077920** | 9.20 | 2.14E-16 | hypothetical protein |
| **Smp_152150** | 9.16 | 8.02E-13 | hypothetical protein |
| **Smp_014610** | 8.99 | 1.18E-52 | serine:threonine kinase 1 |
| **Smp_033250** | 8.99 | 7.77E-39 | hypothetical protein |
| **Smp_000420** | 8.71 | 1.74E-34 | Pro His rich protein |
| **Smp_165360** | 8.65 | 1.13E-30 | histone acetyltransferase myst4 |
| **Smp_000280** | 8.64 | 4.27E-40 | hypothetical protein |
| **Smp_000390** | 8.47 | 2.26E-27 | hypothetical protein |
| **Smp_155310** | 8.39 | 3.90E-32 | tetraspanin CD63 receptor |
| **Smp_144440** | 8.32 | 1.75E-22 | replication A protein |
| **Smp_130970** | 8.12 | 3.55E-10 | G2:mitotic specific cyclin B3 |
| **Smp_000400** | 7.99 | 1.85E-12 | hypothetical protein |

***Table 3.9 Enriched GO terms (biological process) of genes up-regulated in day-28 compared to day-21 worms***

| GO identifier | GO description | p-value |
|---|---|---|
| **GO:0007155** | cell adhesion | 0.0006 |
| **GO:0018958** | phenol-containing compound metabolic process | 0.0011 |
| **GO:0006875** | cellular metal ion homeostasis | 0.0015 |
| **GO:0008643** | carbohydrate transport | 0.0030 |
| **GO:1901617** | organic hydroxy compound biosynthetic process | 0.0078 |
| **GO:0019438** | aromatic compound biosynthetic process | 0.0119 |
| **GO:1901362** | organic cyclic compound biosynthetic process | 0.0120 |
| **GO:0009058** | biosynthetic process | 0.0127 |
| **GO:0009698** | phenylpropanoid metabolic process | 0.0143 |
| **GO:0007586** | digestion | 0.0143 |
| **GO:0055114** | oxidation-reduction process | 0.0171 |
| **GO:0006071** | glycerol metabolic process | 0.0209 |
| **GO:0006629** | lipid metabolic process | 0.0210 |
| **GO:0072593** | reactive oxygen species metabolic process | 0.0284 |
| **GO:0006836** | neurotransmitter transport | 0.0320 |
| **GO:0015711** | organic anion transport | 0.0346 |
| **GO:0006879** | cellular iron ion homeostasis | 0.0370 |
| **GO:0006166** | purine ribonucleoside salvage | 0.0463 |

#### *3.3.7.2 Becoming reproductively mature and migration toward mesenteric veins*

As a reproductively mature stage, day-35 *S. mansoni* are known to be egg-laying and should reside in the mesenteric vein (Clegg, 1965). Liver pathology observed while collecting worms from infected mice in this thesis is consistent with this notion. In contrast, some day-28 parasites were a mix of unpaired worms and parasites from this time point would not have started laying eggs (Clegg, 1965). A study by Zanotti *et al* (1982), in which *S. mansoni* were recovered from experimentally infected mice, showed an increasing number of paired worms in mesenteric vein from days 28 to 35 (Zanotti *et al*., 1982).

Pairwise comparisons yielded 72 genes up-regulated at day 35 and 113 genes up-regulated at day 28 (Figure 3.8, Table S3.9, Table S3.10). The differences between the two time points appeared to be an increase in egg production-related genes at day 35, and a decrease of signalling and developmental control (Figure 3.15, Table S3.11). In addition, up-regulation of *cathepsin B, D,* and *L* (approximately 2.5-fold) at

day 35, reflects increase in blood feeding and the requirement for iron and amino acids in egg production. The genes down-regulated at day 35 can be related to GPCR signalling, transcriptional regulation, and developmental control (Figure 3.16, Table S3.11). By day 35, *S. mansoni* should be fully mature. However, between day 28 and 35 the parasite is still maturing, particularly with the female developing large vitellaria, and this may explain the changes in the development-related genes. The genes involved in GPCR signalling, with higher expression at day 28 than 35, may be relevant for host-parasite interactions because the time between days 28 and 35 coincides with the parasites' migration from the portal to the mesenteric vein (Clegg, 1965; Zanotti et al., 1982). With this, the GPCR up-regulated in day-28 compared to day-35 worms were investigated further.

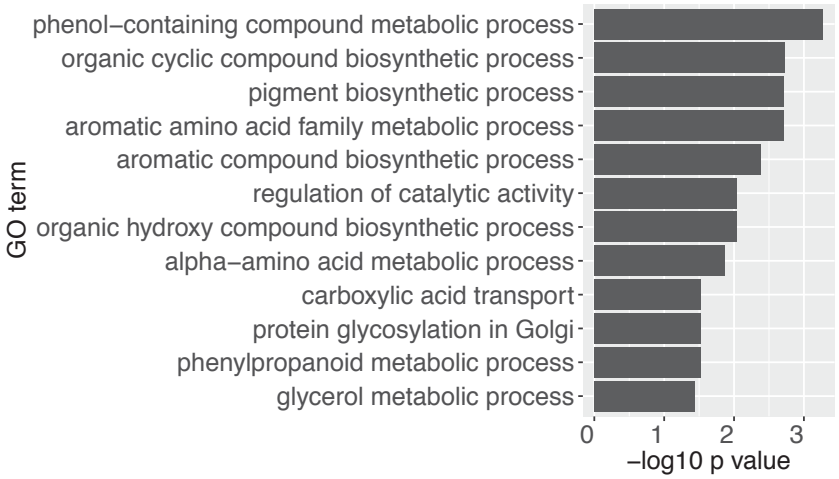


***Figure 3.15 GO enrichment of genes up-regulated in day-35 compared to day-28 worms***
Bar chart shows enriched GO terms (biological processes) of genes up-regulated in day-35 compared to day-28 worms, ranked by p-values from topGO package.

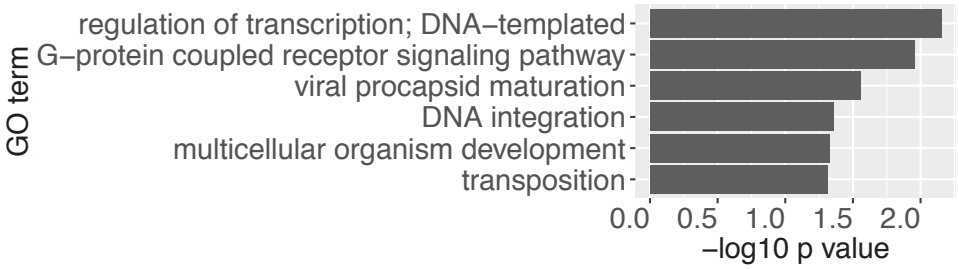


***Figure 3.16 GO enrichment of genes up-regulated in day-28 compared to day-35 worms***
Bar chart shows enriched GO terms (biological processes) of genes up-regulated in day-28 compared to day-35 worms, ranked by p-values from topGO package.

#### *3.3.7.3 GPCR up-regulated in immature adults*

Six GPCRs were up-regulated in day-28 worms and each were up-regulated by ~2 fold (Table S3.9). GPCR signalling, however, could be involved in many aspects, particularly neuropeptide signalling which could regulate locomotion, reproduction, to germline development (Collins *et al*., 2010; Liang *et al*., 2016; McVeigh *et al*., 2005). Apart from the postulated migration, key differences between day-28 and day-35 parasites are egg-laying and male–female interaction which could also be reflected in signalling. However, the expression of these six GPCRs over the full timecourse shows that they each were highly expressed in the lung stage (Figure 3.17) and therefore are unlikely to function solely in signalling during reproduction. Another piece of evidence that supports this notion is demonstrated in section 3.3.7.4 that will follow.

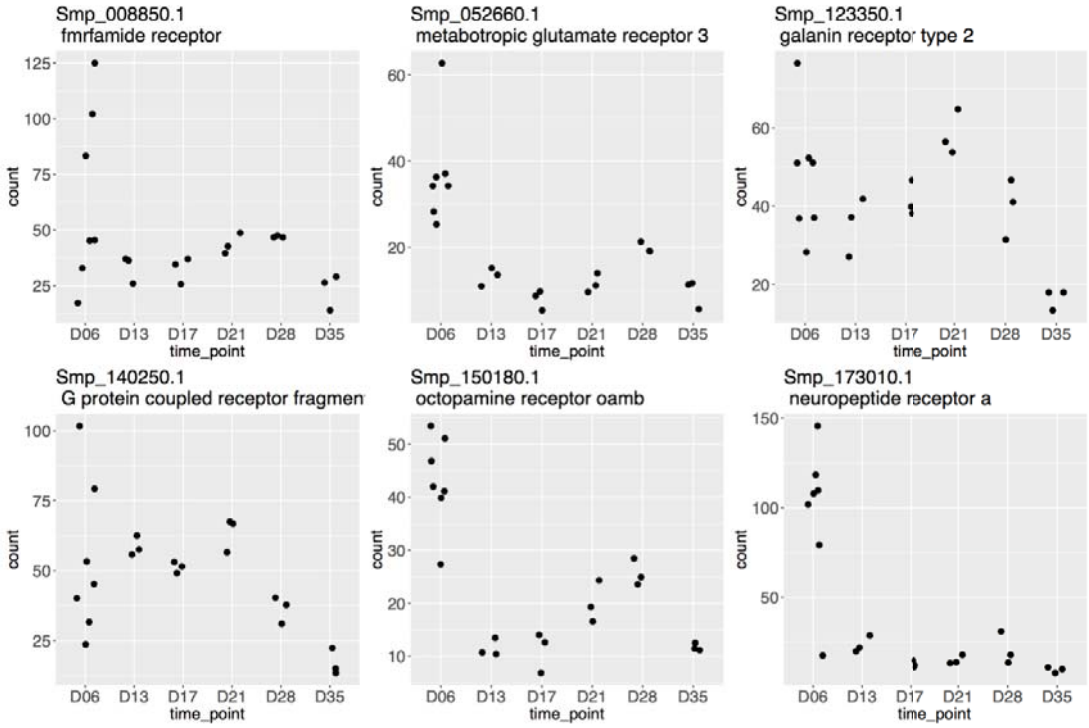


***Figure 3.17 Expression profile of GPCRs up-regulated in day-28 compared to day-35 worms***

Six genes annotated as encoding GPCRs were up-regulated in day-28 compared to day-35 worms. Each plot shows their expression over the whole *in vivo* timecourse. X-axes denote time points; y-axes indicate normalised read counts. Each dot represent the expression level from one biological replicate.

Metabotropic glutamate receptor 3

Of these six GPCRs, only *metabotropic glutamate receptor 3* (Smp_052660) has been characterised in *S. mansoni* (Taman and Ribeiro, 2011). The protein is located in the tegument membrane and in the dorsal tubercles of adult males, and receptor binding to glutamate has been confirmed (Taman and Ribeiro, 2011). The automatically generated phylogenetic tree for Smp_052660 that is available from WormBase ParaSite (Howe *et al*., 2016a, 2016b) shows that homologues of this gene are hugely expanded across a wide range of metazoan species from mammals to placozoa - the simplest form of invertebrate (Appendix E). Homologues in mammals and zebrafish include calcium sensing, olfactory, gustatory and pheromone receptors (Appendix E). A paralogue in *S. mansoni*, Smp_150370, followed a similar expression pattern to Smp_052660 but fell below the $\log_2$FC threshold that I have used to define differentially expressed genes.

Putative octopamine receptor

*Octopamine receptor* (Smp_150180), similarly, has paralogously expanded in a wide ranges of species (Appendix E). However, unlike the metabotropic glutamate receptor 3 (Smp_052660), homologues are not identified in earlier branching metazoa. This gene has not been characterised in *S. mansoni* nor tested for its putative octopamine ligand. However, its homologues include dopamine receptors, and serotonin (5-HT) receptors; similar group of molecules to octopamine.

Other GPCRs

The other four GPCRs are not as extensively expanded across species (Appendix E) and have not been previously characterised in *S. mansoni*. The first is a putative *Neuropeptide receptor A* (Smp_173010) with homologues in *Drosophila*, molluscs (*Crassostrea gigas*), and all main groups of worms (roundworm, flukes, tapeworm, annelids). The second, *Fmrfamide receptor* (Smp_008850), has homologues in both free-living and parasitic nematodes and platyhelminths, annelids and molluscs. Fmrfamide has been shown to stimulate muscle contraction in *S. mansoni* (Day *et al*., 1994, 1997). However, the exact ligand for this receptor needs to be determined. The third, a gene *G protein coupled receptor fragment* (Smp_140250), is specific to platyhelminths and spans flukes, tapeworms and free living *S. mediterranea*. Lastly, *galanin receptor type 2* (Smp_123350) is also largely platyhelminths-specific, but with 2 homologues in molluscs (*C. gigas*). The gene has homologues in trematodes and monogeneans, but not cestodes. In the free living fluke *S. mediterranea,* seven paralogues were found compared to only one in parasitic species.

Postulation

Given their homology to olfactory and gustatory receptors, I hypothesised that the GPCR *metabotropic glutamate receptor 3* (Smp_052660) may be involved in sensing cues or chemical gradient from the mesenteric vein and used for guiding the migration. Further, *octopamine receptor* (Smp_150180) is likely to have biogenic amines as ligand and such ligand is involved in neuromuscular function in *S. mansoni* (Ribeiro *et al*., 2012). These two GPCRs were therefore investigated in more details.

### *3.3.7.4 Cluster of postulated GPCRs*

Signalling-related genes

Both GPCRs were in cluster 40, which consist of 59 genes (Table 3.10; Figure 3.6). Using the gene clustering data, I investigated other genes with similar expression profiles over the whole timecourse to the two GPCRs. Additional GPCRs were in the clusters but were not identified previously amongst genes up-regulated at day 28 due to the $log_2$FC cut-off. Many other genes in the cluster may be related to neuropeptide receptor and processing of signalling – such as those annotated as GPCR regulatory protein, cAMP, kinase, and ion channel – and provide further evidence of signalling processes that may be related to host-pathogen interactions.

***Table 3.10 Genes in cluster 40***

| Gene identifier | Product name |
|---|---|
| **Smp_000350** | hypothetical protein |
| **Smp_000755** | family M13 non peptidase ue (M13 family) |
| **Smp_001000** | universal stress protein |
| **Smp_015630** | Glutamate-gated chloride channel subunit 2 isoform 1 |
| **Smp_016230** | protocadherin 17 |
| **Smp_025030** | dual specificity protein kinase |
| **Smp_033000** | calcium-binding protein |
| **Smp_033010** | 16 kda calcium binding protein |
| **Smp_034610** | innexin |
| **Smp_041700** | G protein coupled receptor fragment |
| **Smp_052660** | metabotropic glutamate receptor 3 |
| **Smp_054170** | hypothetical protein |
| **Smp_057530** | subfamily S9B unassigned peptidase (S09 family) |
| **Smp_059570** | hypothetical protein |
| **Smp_068500** | hypothetical protein |
| **Smp_076030** | tob protein; protein tob btg |
| **Smp_078230** | cGMP dependent protein kinase |
| **Smp_119260** | hypothetical protein |
| **Smp_122870** | regulator of G protein signaling |
| **Smp_122880** | hypothetical protein |
| **Smp_124520** | hypothetical protein |
| **Smp_126500** | tensin C1 domain containing phosphatase |
| **Smp_126760** | Proteasome activator complex subunit 4 |
| **Smp_127420** | 5' AMP activated protein kinase subunit gamma |
| **Smp_136030** | anion exchange protein |
| **Smp_141070** | glutamate receptor interacting protein 1 |
| **Smp_141980** | cAMP specific 3' |

| | |
|---|---|
| **Smp_142280** | hyperpolarization activated cyclic |
| **Smp_142290** | hyperpolarization activated cyclic |
| **Smp_142820** | hypothetical protein |
| **Smp_143710** | gamma aminobutyric acid receptor subunit |
| **Smp_144290** | hypothetical protein |
| **Smp_144310** | potassium voltage gated channel subfamily KQT |
| **Smp_144650** | hypothetical protein |
| **Smp_146750** | sodium:hydrogen exchanger 2 (nhe2) |
| **Smp_147800** | Lmp3 protein |
| **Smp_148490** | kazrin |
| **Smp_148500** | hypothetical protein |
| **Smp_148870** | hypothetical protein |
| **Smp_149170** | Probable G protein coupled receptor B0563 |
| **Smp_149280** | kinase D interacting substrate of 220 kDa |
| **Smp_150180** | octopamine receptor oamb |
| **Smp_150220** | protein tyrosine phosphatase |
| **Smp_150300** | MAP kinase activating death domain |
| **Smp_151600** | neuronal calcium sensor 2 |
| **Smp_155240** | tubulin tyrosine ligase family |
| **Smp_158990** | neuronal calcium sensor |
| **Smp_159790** | kinesin family A |
| **Smp_161450** | small conductance calcium activated potassium |
| **Smp_161510** | hypothetical protein |
| **Smp_163750** | otoferlin |
| **Smp_164010** | serine:threonine protein phosphatase 6 |
| **Smp_167700** | hypothetical protein |
| **Smp_170780** | hypothetical protein |
| **Smp_177020** | PhosphoLipase C Like family member (pll 1) |
| **Smp_187090** | hypothetical protein |
| **Smp_198960** | hypothetical protein |
| **Smp_204230** | hypothetical protein |
| **Smp_211310** | soluble guanylate cyclase gcy |

Potential functions of the signalling on reproduction

The evidence of differentially expressed signalling genes in cluster 40, between days 28 and 35, could alternatively be the result of changes upon pairing and egg-laying processes. Furthermore, although the genes were up-regulated in the lung stage, they may employ different functions as genes in signalling pathways often cross-talk. Consequently, involvement of these genes in reproductive signalling could not be ruled out simply because the genes were up-regulated in the lung stage. To determine

whether expression differences were indeed due to reproductive signalling, published data was used that compared the transcriptomes of worms and gonads from mixed and single-sex infections (Lu *et al*., 2016).

Existing data on effect of pairing

In the previously published study (Lu *et al*., 2016), 44 genes, out of the 59 in cluster 40, were affected by pairing. Mostly, they were up-regulated in single females compared to paired females (36 genes). However, half were not changed in ovary as a result of pairing while the other half were affected and can therefore be ruled out as likely reproductive signalling proteins (Figure 3.18). Of the GPCRs in the cluster, none were differentially expressed in ovary as a result of pairing but some are differentially expressed at whole-worm level: the *metabotropic glutamate receptor 3* (Smp_052660), *octopamine receptor oamb* (Smp_150180), and *gamma aminobutyric acid receptor subunit* (Smp_143710). The changes upon pairing in whole females but not in ovaries, and their high expression in lung stage, suggests that these three GPCRs may be involved in processes other than pairing and reproduction. In contrast to many genes being affected in females, only five genes are up-regulated in paired male worms.

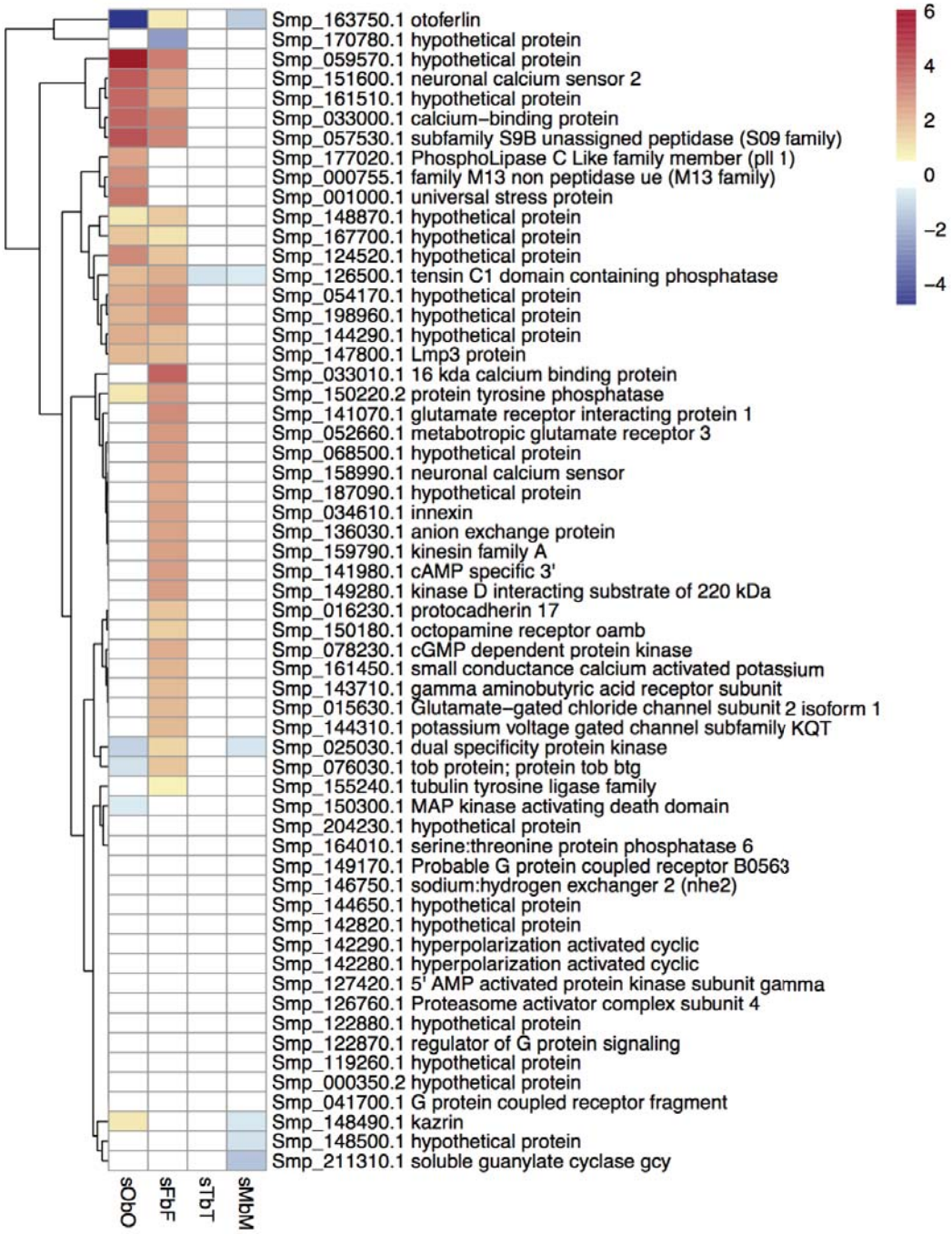


***Figure 3.18 $Log_2FC$ of genes in cluster 40 in response to pairing***

Expression of genes in cluster 40 in response to male-female pairing according to published RNA-seq data (Lu et al., 2016). Colours on the heatmap represented $log_2FC$ obtained from the published data. Column names denote pairwise comparison underlying each $log_2FC$. sObO: ovaries, single-sex infection *vs.* mixed-sex infection. sFbF: females, single-sex infection *vs.* mixed-sex infection. sTbT: testes, single-sex infection *vs.* mixed-sex infection. sMbM: males, single-sex infection *vs.* mixed-sex infection.

#### *3.3.7.5 Adult stage conclusion*

In summary, gene expression in adult stages shows, as expected, up-regulation in egg-production and acquisition of macromolecules and ions essential for the reproduction process, and down-regulation of developmental regulations as the parasites progressed into fully mature adults. In addition, GPCR signalling was down-regulated in paired and mature (day 35) compared to immature (day 28) worms, with one GPCR being a homologue of olfactory and taste receptors, and one potentially involved in neuromuscular signalling. Considering the functions of genes with similar expression pattern over the timecourse, the effect of pairing (from published data) on these genes, and their overall expression over the timecourse, it does appear that there is some involvement of the GPCRs in detecting host-derived signals.

## 3.4 Discussion

### 3.4.1 Overview

In this chapter, I sought insight into host-parasite interactions of *S. mansoni* during key stages of intra-mammalian infection. How the parasite evades host immune attack and localises to the liver and mesenteric veins are of particular interests. I analysed transcriptomic profiles from a timecourse of parasites developed *in vivo* and covering the lung stage through to the egg-laying adult stage. Information on protein domains, predicted protein structure, and phylogenetic relationships were incorporated in order to decipher gene function. The transcriptome dataset produced in this chapter covers key stages in intra-mammalian infection and includes the first RNA-seq transcriptome of *in vivo* lung stage as part of a developmental timecourse. Together, the data reflected the expected nature of key stages but also provide new details on underlying processes. In addition, the data highlight striking signalling events in the lung stage alongside possible strategies for evading immune attack. Lastly, adaptive changes in the liver stages and signalling during the migration towards the mesentery are proposed.

### 3.4.2 Potential effect of collection procedures

It is possible that the collection procedures themselves might have impacted the transcriptomes of the parasite. As detailed in the Methods (section 3.2.2.6), after the parasites were perfused from the mice, they were imaged and washed, meaning that

they were outside their mammalian host for a couple of hours before being collected into TRIzol. Future attempts could avoid this extended time, perhaps by collecting from fewer mice each day, or by randomly halving each replicate so that the half for RNA-seq could be processed straightaway into TRIzol. Additionally, the protocol of collecting lung stage worms involved incubation of minced lungs in media for three hours, which may also affect the transcriptomes. An alternative approach was considered whereby the whole lung tissue would be extracted for RNA which would contain RNA from murine host lungs and lung stage worms. This approach, however, would likely yield data that are overwhelmed by reads from the host materials. To assess potential changes in the transcriptomes caused by the incubation, a portion of lung tissues could be collected and extracted for RNA. Gene expression could then be compared between the incubated lung stage worms and the worms in the fresh lung tissue. However, despite the potential effects of the collection procedures, the transcriptomes still reflected the expected biology of their respective stages (such as development in the growing liver phase, and egg laying process in adult stages), which suggests that the parasite transcriptomes were not drastically affected.

### 3.4.3 Lung stage signalling

Lung schistosomules, overall, showed increased expression of genes in signalling processes and ion transport. Signalling via neuropeptides and neurotransmitters may be involved in motility and migration through the lung capillaries. Neuropeptide signalling is one of the biological processes disturbed in irradiated *S. mansoni* (Dillon *et al.*, 2008) and is reasoned for the loss of irradiated parasites during the lung migration, causing them to exit capillaries into the alveoli and eventually die (Crabtree and Wilson, 1986a; Dean and Mangold, 1992; Mastin *et al.*, 1985). In addition, signalling processes control development and differentiation. Upon entering the lung, schistosomules are required to change their shape into an elongated form and obtain the machinery for migrating inside lung capillary (Crabtree and Wilson, 1986a). Although the cues that stimulate the changes to their shape are unknown (Kusel *et al.*, 2007), it is understood that the schistosomules remodel themselves by differentiating existing cells and do not perform mitosis at this stage (Clegg, 1965; Lawson and Wilson, 1980). Signalling processes may have an important role in these changes. Furthermore, signalling in the lung might trigger the onset of development.

### 3.4.4 Lung stage immune evasion

In addition to the migration challenge, lung schistosomules are at risk of oxidative damage in the lung (Dirks and Faiman, 1982; Kobayashi-Miura *et al*., 2007), and of blood coagulation from inhibited blood flow (Mebius *et al*., 2013), their transcriptomes were expected to reflect adaptation to the environment. Up-regulated in the lung stage were a group of genes functioning in neutralising oxidative stress, including *superoxide dismutase* and *thioredoxin peroxidase* (Glanfield *et al*., 2007). Such up-regulation is possibly in response to changes in environment rather than a hard-wired gene expression programme. In a mammalian system, expression of thioredoxin is induced by oxidative stress (Kobayashi-Miura *et al*., 2007). Furthermore, lung schistosomules, when obtained from *in vitro* cultivation, express genes related to stress and immune evasion at a lower level (Chai *et al*., 2006). In addition to *superoxide dismutase* and *thioredoxin*, *ferritins* were up-regulated in the lung stage, possibly for scavenging iron from immune cells and from inducing formation of ROS (Glanfield *et al*., 2007).

### 3.4.5 Complement factor H

The prominence of genes with immune-related roles in the timecourse led me to question whether there were uncharacterised genes (annotated as encoding "hypothetical protein") with similar expression patterns that could have similar roles. The particular profile of interest was composed of genes with high expression in the lung stage, down-regulated during liver stages, and up-regulated once the parasites passed the liver stages. Of the three hypothetical proteins in this cluster, one was previously found in secreted extracellular vesicles from *in vitro* schistosomules (Nowacki *et al*., 2015) further suggesting its role in environmental interaction. Using protein structural prediction, another hypothetical protein (Smp_182770) was found to be structurally similar to human CFH, a well-characterised regulator of the complement cascade (Medjeral-Thomas and Pickering, 2016; Morgan *et al*., 2011). Despite minimal sequence similarity and an absence of signature domains for complement control proteins, the parasite gene displayed a pattern of tandem repeats as expected from mammalian CFH. Presence of its homologues in other *Schistosoma* spp. and liver flukes suggest a common strategy for living under similar pressure of host immune attack (Schroeder *et al*., 2009; Sirisinha *et al*., 1986).

### 3.4.6 Complement factor I

In the mammalian system, CFH is described as a co-factor of CFI (Medjeral-Thomas and Pickering, 2016) and works downstream of DAF (Morgan *et al*., 2011). So far, there is no description of *S. mansoni* genes with sequence or structure similar to CFI, but there is a serine protease, m28, that cleaves C3bi (another function of mammalian CFI (Fishelson, 1995; Morgan *et al*., 2011). However, molecular weight of the m28 is 28kDa, not matching with factor I which has molecular weight of 38 kDa for its serine protease light chain (Alba-Domínguez *et al*., 2012). For DAF, previous work has shown that schistosomes can acquire DAF from host blood cells and such acquisition promotes parasite survival *in vitro* (Horta *et al*., 1991; Ramalho-Pinto *et al*., 1992). Validation of the putative CFH involvement in host immunomodulation requires further laboratory investigation - such as by incubating a recombinant protein produced from the putative CFH gene with CFI and its C3b substrate, or by functional knock-down of the gene. In addition, further structural prediction could be across all the *S. mansoni* genes to search for a potential CFI-encoding gene. However, such computationally intensive searches at such a large scale are currently prohibitively slow.

### 3.4.7 Immunomodulation in liver stages

In contrast to the profile with up-regulation in lung and adult stages, some genes with possible immunomodulation effect were up-regulated in liver and down-regulated in lung and adult stages. The gene encoding anti-inflammatory protein Sm16 is a key example of this. The gene product has been shown previously to be a major part of cercaria ES and interferes with signalling in monocytes and macrophages, proposing to delay antigen recognition upon *S. mansoni* infection to prevent inflammation during skin invasion (Sanin and Mountford, 2015). The gene was previously shown to be developmentally regulated with high expression in cercariae (Protasio et al., 2013; Sanin and Mountford, 2015). In this thesis, the expression of this gene was also increased between day 13 and day 21 and was lower at day 6 and in adults. Many genes with this expression profile are genes involved in growth and developmental control. However, annotated product names of a few genes suggested that some genes might be involved in immunomodulation. Examples of these are *microsomal glutathione S transferase 1* (Smp_024010), *Lamin B receptor* (ERG24;

Smp_124300), *Uveal autoantigen with coiled coil domains* (Smp_129710), and *LPS-induced tumor necrosis factor* (Smp_060220). Furthermore, *aldehyde dehydrogenase* gene, which was up-regulated in liver stages, may also be involved in regulating oxidative stress (Singh *et al*., 2013), in addition to its proposed role of providing acetate for energy generation via the Kreb's cycle.

### 3.4.8 Implication for intervention

These different profiles of genes with potential functions in immune evasion suggest existence of multiple methods that the parasites may use at different stages even though all are residing in the mammalian bloodstream. A better understanding of these different strategies will be important for designing an effective vaccine against certain stages of the parasite. However, it is worth noting that *S. mansoni* in this timecourse was obtained from mice infected with different numbers of cercariae that may induce different immune response profiles in the host and therefore affect the parasite's immune evasion strategy.

### 3.4.9 Expected changes in liver and adult stages

Despite the parasite samples being from mixed-sex infection and each pool not being morphologically homogeneous, the transcriptomic data reflected major biological features expected for key stages of development. In liver schistosomules, highly expressed genes were enriched in biological processes important for growth (mitosis) and control of development, whereas expression of egg production-related genes predominated in adult stages. Neuropeptide signalling was enriched in liver stages and may be part of developmental control, orientation, motility, feeding (Collins *et al*., 2010; McVeigh *et al*., 2012). In the planarian *S. mediterranea*, neuropeptide Y may be involved in regeneration and it is possible that this signalling is also of developmental control in *S. mansoni* (Sandmann *et al*., 2011). In adult stages, multiple hypothetical genes were strongly up-regulated amongst other egg-production-related genes which suggest their potential functions.

### 3.4.10 Liver localisation

This dataset has not, however, provided a tentative explanation on signals involved in liver localisation. This lack of signal from the transcriptomic data may support the idea that schistosomules arrive at the liver passively through blood circulation from

portal vein (Bloch, 1980; Kusel *et al*., 2007; Wilson, 2009). When they reach a sufficiently large size, they become trapped in the portal triad, with the fine sinusoid network acting as a filter preventing large schistosomules from exiting with blood flow (Wilson, 2009), as opposed to detecting liver environment and actively anchoring to the blood vessel near liver. Alternatively, even if a sign for liver localisation is present, it may be drowned out in the transcriptomic signal related to growth and development, particularly, if the sensing only involve small parts of the worms. A future approach for solving this issue may require studying of isolated parasite tissues or single cells.

### 3.4.11 Mesenteric migration

Migration of adult worm pairs toward the mesenteric vein, however, is thought to involve active migration following certain cues - firstly because the migration is against the blood flow, and secondly because another species of schistosome (*S. haematobium*) migrates to a different site. Nevertheless, such cues have not been identified. From this dataset, multiple GPCRs were differentially expressed between days 28 and 35 which would cover this migration. *Metabotropic glutamate receptor* (Smp_052660) binds to glutamate and is localised on tubercles of adult male *S. mansoni*, absent from neuromuscular structure, and expressed at low levels in paired female (Taman and Ribeiro, 2011). The authors also suggested its role in chemical signalling from “endogenous signals or host-derived glutamate”. Interestingly, the gene is down-regulated once females pair with males (Lu *et al*., 2016). The absence in the female parasite in the study of Taman and Ribeiro (2011) could be down to the source of female parasites coming from mixed sex infection. The expression of this gene in females raises questions about its role in migration as it is generally thought that the parasites migrate as a pair, assisted by the male. However, lone females can migrate toward mesenteric vein, though at much smaller number than paired worms and eventually returned to the portal vein (Zanotti *et al*., 1982).

Together with the pattern of gene expression in our dataset, its homologues with olfactory receptors, and the fact that glutamate is an excitatory neurotransmitter, it is temping to speculate that the worms might use a glutamate gradient as a cue to migrate toward mesenteric veins. Intriguingly an orthologue of this receptor is present in *S. haematobium* genome, but *S. haematobium* migrates to a different egg-laying

site to *S. mansoni* (Gryseels et al., 2006). A possible explanation is that the same receptor can trigger a different downstream signalling pathway (Lodish et al., 2000). In *S. haematobium*, it may stimulate the movement away from the mesenteric vein instead of towards it. Furthermore, the expression site of the orthologous receptor in *S. haematobium* has not been established and the receptor might not be situated at a site suitable for detecting external ligands and may have other roles in *S. haematobium*. Attraction towards, or repulsion away from, glutamate could be tested using adult schistosomes and a simple decision chamber that was employed for male-female attraction studies (Eveland et al., 1982; Imperia et al., 1980). Afterward, the specificity for the receptors for attraction behaviour could be characterised using functional genomic tools available for schistosomes (Alrefaei et al., 2011; MacDonald et al., 2015).

#### *3.4.11.1 Glutamate*

Glutamate is an amino acid abundant in diet and passed from intestine into blood. Despite only 2-20% of intake glutamate is passed onto blood, glutamate gradient might be possible between mesenteric vein and portal vein (Burrin and Stoll, 2009; Janeczko *et al*., 2007; Windemueller and Spaeth, 1980; Windmueller and Spaeth, 1975). Roles of glutamate and glutamate receptors in control of directional locomotion have been shown in other invertebrates (Croset et al., 2016; Hills et al., 2004; Vezina and Kim, 1999). Two types of glutamate receptors are involved in such process - metabotropic glutamate receptors which are GPCRs, and ionotropic glutamate receptors which are glutamate-gated ion channels. Metabotropic glutamate receptors are involved in locomotor behaviour (Vezina and Kim, 1999) and glutamate as a neurotransmitter controls directional foraging behaviour in *C. elegans* (Hills *et al*., 2004). *Drosophila* larvae show attraction toward glutamate (Croset *et al*., 2016). However, in both *C. elegans* and *Drosophila*, the receptors involved are ionotropic glutamate receptors.

#### *3.4.11.2 Octopamine receptor and GABA*

Another GPCR with similar expression to the metabotropic glutamate receptor is *octopamine receptor* (Smp_150180) and *gamma aminobutyric acid (GABA) receptor subunit* (Smp_143710). Though the exact ligand has not been identified for the octopamine receptor, biogenic amine signalling can modulate behaviour and

locomotion in *C. elegans* (Chase, 2007). GABA is found at oral and ventral suckers and is thought to be involved in attaching to host blood vessel (Mendonça-Silva *et al*., 2004). However, this GABA receptor has not been confirmed for GABA binding but is, sequence-wise, similar to an ion channel; therefore, it could be involved in neuromuscular signalling.

### 3.4.12 Micro-exon genes

MEGs have been implicated in host-parasite interactions. In this transcriptomic dataset, some *MEGs* were strongly up-regulated at specific time points. Up-regulation of *MEG-2* and *MEG-3* groups in the lung stage extended from what has been shown previously that these MEGs are highly expressed in early stage schistosomules compared to pre-intramammalian stages (DeMarco *et al*., 2010). *MEG-14*, one of the *MEGs* up-regulated in lung stage, can bind to host inflammatory protein in the worm oesophagus to prevent internal damage to the parasites (Orcia *et al*., 2016) and appears to become important for the parasites particularly in lung stage. Further, *MEG-32.2*, previously localised to the head part of adults (Wilson *et al*., 2015) is highly up-regulated in day-13 (early liver stage) compared to lung stage.

### 3.4.13 Other ways to use this data

In addition to investigating host-parasite interactions, this transcriptomic dataset provides a resource on other aspects of *S. mansoni* biology and is an important resource for the research community. This does however mean that in teasing apart signals related to host-parasite interactions is somewhat confounded by signals from other processes such as development, and parasite-parasite interactions. For functional genomics and transgenesis in *S. mansoni*, this dataset will be useful for identifying promoters of genes specifically expressed at different stages or promoters of constitutively expressed genes. Lastly, the analyses approach in this chapter also demonstrated the benefits of incorporating additional information from expression patterns over a timecourse, in addition to using pairwise comparisons, to interpret potential roles of genes and identify relevant molecular processes.

### 3.4.14 Summary

In summary, I have produced and analysed the transcriptomic profile of *S. mansoni* during its intra-mammalian stage aiming for a better understanding of its host-parasite

interactions. I interpret that the lung stage schistosomules employed up-regulation of genes to counteract immune attack and present potential new player in immune evasion. Cues that stimulate schistosomule migration along the lung capillary are still not clear, but many signalling processes may be involved. Liver stage schistosomule migration appeared to follow passive migration as many have previously thought. Receptors and chemical cues involved in mesenteric migration is proposed.

In the next chapter, I moved to an *in vitro* approach of co-culture between schistosomules and human cells/cell lines and use transcriptomic information to find out how schistosomules might respond to host environment that they may encounter during their intra-mammalian stages.