Chapter 4

S. mansoni in vitro culture with cell lines

4.1 Introduction

4.1.1 Overview

During intramammalian stages, *S. mansoni* live in the host bloodstream and are in close contact with host tissues, such as the endothelial cells lining blood vessels and liver parenchyma close to portal venules. However, the effects of different host tissues on *S. mansoni* are not well understood. A better understanding of such effect may provide insights into the influence of host microenvironments on key changes and parasite survival during *S. mansoni* infection. In this chapter, I investigated the relationship by co-culturing *S. mansoni* larval stage (schistosomules) with three types of commercially-available mammalian cells, and studied transcriptional changes of the parasites. Both generic responses and cell-type specific responses were explored to provide further insights into the relationship between the parasites and the host micro-environments. *In vivo* expression profiles (from chapter 3) were included where relevant. The mammalian cells co-cultured with the schistosomules were studied in chapter 5.

4.1.2 *S. mansoni* and host tissues

While in the mammalian host, *S. mansoni* is always in close contact with host tissue. Living in the bloodstream, its immediate host tissue environment is endothelial cells lining blood vessels, and circulating cells in blood. Previous studies using *in vivo* microscopy (Bloch, 1980; Wilson, 2009) show that both schistosomules and adults sometimes reside in blood vessels that are only slightly larger than their size. Furthermore, certain tissues are associated with key stages of infection. In particular, the liver is widely accepted as the site where the parasites develop into the adult stage (Clegg, 1965; Wilson, 2009). During their development in the liver, the parasites remain in the blood circulation, but blood in the liver is separated from the liver parenchyma (hepatocytes) only by a porous layer of sinusoidal endothelial cells. The

porous nature of the liver sinusoids allows molecules to pass between blood and hepatocytes and might allow interactions between *S. mansoni* and host beyond blood vessels. Small schistosomules are thought to pass through liver multiple times until they become too large to enter sinusoidal network (Wilson, 2009). Thus, the schistosomules in early liver stages could be in close proximity with the hepatocytes. Developing schistosomules, however, are not found in the sinusoidal area, but in the venules connecting the portal vein and the sinusoidal area (Bloch, 1980). This proximity may allow movement of substrates between blood and hepatocytes. The host tissues with potential interaction with *S. mansoni* also have roles in host defence against infection (Pober and Sessa, 2007; Janeway 2001). For examples, endothelial cells interact with immune cells for their adhesion, and regulation of blood coagulation; liver hepatocytes are the main cellular factory for many components involved in immune response pathways and coagulation (Pober and Sessa, 2007; Janeway 2001).

A better understanding of how the host microenvironment affects parasites could provide insights into mechanisms that support *S. mansoni* survival, or how host tissue might influence molecular processes during the infections. Furthermore, liver is involved in other parasitic infections such as liver flukes, and *Plasmodium* spp. Understanding how *S. mansoni* respond to *in vitro* liver environment may further extend the concepts for investigation in other parasites.

4.1.3 Effects of host environments on *S. mansoni*

The effect of host environment on parasites have been investigated between different life stages of *S. mansoni* in different host types (Fitzpatrick *et al*., 2009; Gobert *et al*., 2009; Parker-Manuel *et al*., 2011). These studies have measured gene expression changes in responses to host environments and have shown some of the ways in which parasites adapt to ensure efficient infection (Fitzpatrick *et al*., 2009; Gobert *et al*., 2009; Parker-Manuel *et al*., 2011). However, the changes in gene expressions in these studies are combinations of responses to host environments, responses to host circulating molecules, and developmental regulation, which complicates the interpretation of such studies. In a previous study, the effect of host tissues on *S. japonicum* was investigated by co-culturing of schistosomules with mammalian cells and showed that co-cultured cell types affect schistosomule morphology and gene

expression (Ye *et al*., 2012). Information of gene expression was, however, limited because it was based on dot-blot hybridisation and comparison of bands present on electrophoresis gels (Ye *et al*., 2012). Changes at the molecular level can now be explored at more in-depth by analysing deep transcriptomic sequencing data.

4.1.4 Aims and approaches

The aim of this chapter is to investigate whether and how *S. mansoni* are affected by the host tissue environment. To this end, I set out to co-culture schistosomules with three types of mammalian cells over a timecourse, and investigate the schistosomule transcriptomes. The cells were originated from human tissues: endothelial cells (HUVEC; human umbilical vein endothelial cells), hepatocyte cancer cells (hepatocarcinoma; HEPG2), and adhesion-enhanced embryonic kidney cells (GripTite) which represented an irrelevant tissue. It should be noted that although these cells were derived from such tissues, they have been adapted to *in vitro* growth conditions and may have changed their physiology; therefore, it is likely that they do not function in exactly the same way as their *in vivo* counterpart. However, by using an *in vitro* approach, schistosomules were separated from other environmental factors, such as circulating immune cells and restricted space of blood vessels, enabling the effect of host cells on the schistosomules to be more directly measured.

4.1.5 Chapter outline

This chapter contains details of the experimental set up and investigation of schistosomule transcriptomes. First, I explore the overall differences in transcriptomic profiles and the effect of cell conditions and time points. Second I investigated changes that were common to all three cell conditions. And third, I examined responses specific to HUVEC and HEPG2 because these represent relevant tissues during an infection of *S. mansoni*. The results show that schistosomules displayed both generic and specific responses to each cell types. Furthermore, responses to HEPG2 may explain how liver environment affect development and other aspects of *S. mansoni* infections.

4.2 Methods

The methods described in this chapter encompass experimental preparation, set up and obtaining schistosomules to be processed. The methods for downstream processing including RNA extraction, sequencing library preparation, and bioinformatic analysis are as described in Chapter 2.

In addition to *in vitro* schistosomules from this chapter, samples in the downstream processing also included three biological replicates from day 17 *in vivo* worms. The *in vivo* samples were included so that technical variations between batches of RNA processing can be assessed and any comparisons between *in vivo* and *in vitro* samples can be done with an awareness on such variation. The day 17 *in vivo* worms were prepared and obtained in the same manner as that explained in chapter 3. The day-17 worms were kindly provided by Prof. Michael Doenhoff at the University of Nottingham.

4.2.1 Experimental design

In this experiment, mechanically transformed schistosomules were co-cultured with three types of human cells, or cell lines, in a modified Basch medium (Appendix D). The co-cultures were maintained over a timecourse, with erythrocytes being added to the media from day three, and schistosomules being transferred to new plates of cells at intervals. The timeline of the co-culture and the transfer is shown in Figure 4.1. HUVEC were purchased from ATCC; HEPG2 were originally purchased from ATCC and kindly provided by Drs James Hawison and David Adam (WTSI); GripTite cells were kindly provided by Dr Gavin Wright (WTSI). Each cell type formed an underlying layer on the plate with schistosomules staying on the top. Control schistosomules were cultivated without an underlying layer of cells. After 6, 10 and 17 days from the start of the co-culture set-up, both parasites and cells were collected for downstream processing for transcriptome analysis. At the point of sample collection (see Figure 4.1), the cells were four days old after being plated and were co-cultured with schistosomules for three days, regardless of the schistosomule time points. Therefore, controls for the human cells were four days old cultured without schistosomules for three days in modified Basch media with erythrocytes. The use of the co-cultured human cells and control human cells will be described in chapter 5.

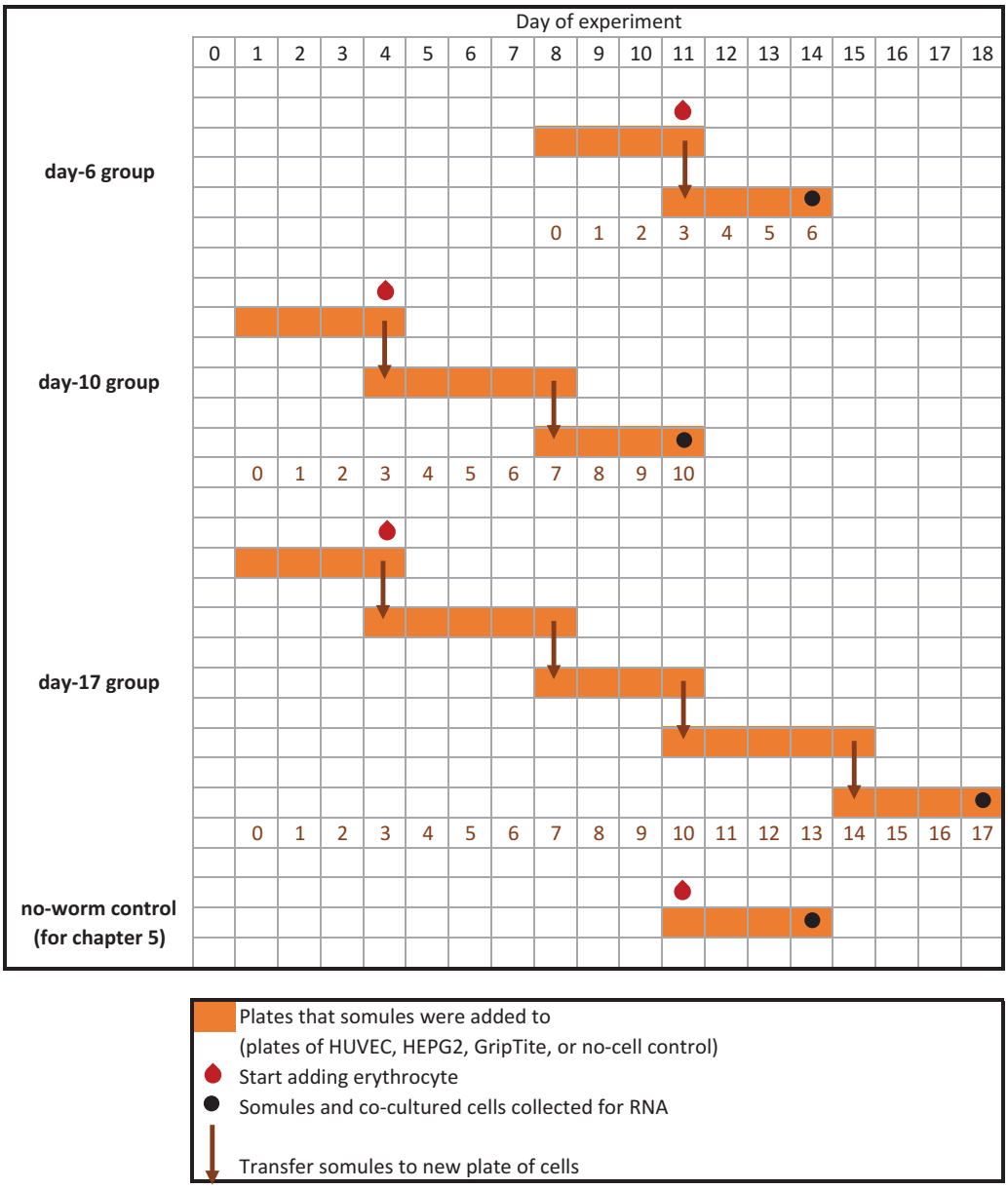

Figure 4.1 Experiment set up

Timeline for co-culture set up. Orange blocks represent plates containing HUVEC, or HEPG2, or GripTite, or were no-cell control. Arrow represent transfer of schistosomules to a new plate. Cells were added to plates the day before schistosomules were transferred to the plate (i.e. the day before the start of each block). Washed erythrocytes were added to the co-cultures on day 3 from the beginning of the set up as represented by the red drop. Both schistosomules and cells were collected at each time point specified. Numbers in dark orange at the bottom of each rows of blocks represent the age of schistosomules in the set up.

4.2.2 Preparing human cells: maintaining stock cells

Stock cells were maintained in either T25 or T75 tissue culture flasks. Fetal bovine serum was not heat-inactivated as this could negatively affect growth of the cells. Antibiotic was also not used during the culture of stock cells for the same reason (ATCC, 2014, 2015).

4.2.2.1 HUVEC (Human umbilical vein endothelial cells)

The protocol from ATCC for cultivation of HUVEC (CRL-1730) was followed. Briefly, once thawed, media change was required two days after thawing, and the cells were passaged twice a week. To passage the cells, old media was aspirated and the cells washed by adding 10-15 ml of sterile PBS and gently swirling. The PBS was aspirated and 1 ml of trypsin added to the cells. The flask was kept at room temperature for 2-3 minutes, checked under an inverted microscope for detachment of the cells, followed by addition of 3-4 ml of fresh media to stop the trypsinisation process, and resuspension into single cells by passing 4-5 times through pipette. The passaging density was at 1:2 - 1:3. When the cells did not reach 70% confluent growth, media change were performed instead of trypsinisation and passaging. To change the media, the old media was transferred into a new Falcon tube and centrifuged at 300 x g, 3 minutes to collect debris. The supernatant was then removed, replaced with pre-warmed fresh media, the debris pellet resuspended, and the fresh media with the resuspended debris added back to the flask. Presence of debris in the culture of HUVEC was normal and the debris should be returned to the cells when performing media change. The media contained F-12K (30-2004, ATCC) as a basal media, 0.1 mg/ml heparin, 0.05 mh/ml endothelial cells growth factors E-2759, Sigma, and 10% FBS). This cell line did not grow to more than 70-80% confluence.

4.2.2.2 HEPG2 (hepatocarcinoma cell lines)

The protocol from ATCC for cultivation of HEPG2 (HB-8065) was followed. To passage the cells, old media was aspirated and cells washed with sterile PBS as above. Trypsinisation required 37 ºC for 10 minutes with shaking and tapping every 3-4 minutes, followed by resuspending as individual cells by passing 8-10 times through a pipette. The passaging density was at 1:4. Media change was similar to HUVEC - the old media was centrifuged and the pellet resuspended in fresh media and transferred back to the flask. The recommended media contained EMEM as a basal media, and

10% FBS. However, it was found that the cells took over a week to reach the growth phase in this recommended media, and that addition of Basch media promoted growth. Thus the media used was 25% of the recommended media (EMEM + 10% FBS) and 75% of modified Basch media. The HEPG2 cells grew as clusters of cells rather than in a mono-layer.

4.2.2.3 GripTite (modified Human Embryonic Kidney (HEK-293) cells)

The GripTite™ 293 MSR cell is derived from HEK-293 cell line transfected with a macrophage scavenger receptor (hence MSR) which makes the cells adhere strongly to culture surface. The cells were obtained in growth phase. Detachment of the cells upon passaging requires a combination of trypsin and Versene. First, the old media was removed and cells were washed with sterile PBS. Then 1 ml of pre-warmed Versene was added and incubated at room temperature with the cells for 5 minutes. Next Versene was removed and 1 ml of trypsin was added and incubated at room temperature for 2-3 minutes. The passaging density was at 1:6. The media used contained DMEM as a basal media, with 10% FBS.

4.2.2.4 Mycoplasma test

All three cell lines were tested for contamination of *Mycoplasma* using PCR-based kit (LookOut Mycoplasma PCR Detection Kit (MP0035, Sigma-Aldrich)). Supernatant from two stock batches of each cell type were used as inputs for the test. The samples were added to reaction tubes containing dNTPs and primers, and DNA polymerase (JumpStart Taq DNA Polymerase (D9307, Sigma-Aldrich)), and the PCR cycles were performed as specified by the kit protocol. The primers provided with the kit amplifed internal control DNA (undisclosed sequence and source) as a positive control for each PCR reaction, and amplified conserved 16S rRNA coding region for the *Mycoplasma* test. The PCR products were run on 1% agarose gels and viewed under UV light on GelDoc-It Imaging System. A presence of the positive control band at 259 bp indicated success of the PCR reaction. A presence of the lower band at 481 bp indicated *Mycoplasma* contamination.

4.2.3 Preparing schistosomules: media for schistosomules and for the co-culture

Modified Basch media was used in the co-culture set up. The media composition followed to the original recipe (Basch, 1981) except that serotonin was not added (due to its short degradation time), fetal bovine serum was used instead of human serum, and DMEM was used as a basal media (Appendix D). The same batch of fetal bovine serum used for maintaining stock human cells was used for Basch media, and this was not heat-inactivated.

4.2.4 Preparing schistosomules: transforming cercariae into schistosomules

Co-cultured schistosomules were obtained by mechanically transforming cercariae through a fine needle. The protocol follows that of Protasio *et al*., (2013). Cercariae were obtained from shedding infected snails pooled in a glass beaker containing conditioned aquarium water. The water level was just enough to cover the snail and the snails were exposed to bright light for 2 hours in a 28 ºC temperature controlled room. Cercariae can be seen in water as cloudy moving particles. To count the cercariae obtained, the cercariae suspension was mixed homogenously by drawing pipette with 20P tip within the solution and immediately transferring 5x10 ul aliquots into a small petri dish followed by addition of 10 ul of Lugal® to each cercariae suspension droplet, which immediately killed the cercariae. The cercariae on the petridish were counted under microscope to calculate the total number of cercariae obtained from the shedding in order to determine the downstream volume and grouping needed for the transformation steps.

Cercariae suspension were transferred into Falcon tubes and left on ice for at least 30 minutes. This allowed cercariae to sink to the bottom by slowing down their activity. After 30 minutes, the supernatant was removed and 10 ml of transformation media added (DMEM (D6546, Sigma-Aldrich) with 2% Antibiotic-Antimycotic (15240062, Gibco). The cercariae pellet at the bottom was resuspended and transferred to a new 15 ml Falcon tube, capped tightly and vortexed at the maximum speed for 1 minute. The mix was then transferred back to 50 ml tube and passed through a 22-gauge needle attached to a 10 ml Luer Lock syringe, 7 times for every 100,000 cercariae to

remove the cercarial tails. The process should be done on ice as much as possible, as this is important for successful separation of cercariae and schistosomules. Cercaria tails and heads (schistosomule) were separated using ice-cold Percoll gradient. To make a Percoll gradient, 1.2 ml of 1 M NaCl was added to 6.5 ml of Percoll (P1644, Sigma), and topped up with the DMEM to 10 ml. For every 100,000 cercariae, a Percoll gradient was set up using 10 ml Percoll mix in a 15 ml Falcon tubes. The parasite suspension was loaded gently on top of the Percoll mix using disposable Pasteur pipettes and then centrifuged at 350 x g for 20 minutes, at 4ºC in a swing-bucket rotor. The tubes were carefully removed from the centrifuge. The tail layer (top) and supernatant containing undetached cercariae were removed and the pellet of schistosomules was washed three times in the 15 ml of fresh transformation media. Each time the schistosomules were centrifuged at 500 x g for 10 minutes at 4ºC. The wash is important because Percoll is toxic to schistosomules. After the final wash, the schistosomules were resuspended in the modified Basch media and transferred to a T25 with 2% antibiotic/antimicotic and kept at 37 ºC until being used in the co-culture experiment.

4.2.5 Co-culture

4.2.5.1 Cells plated for co-culture with schistosomules

To prepare the co-cultured cells for the experiment, human cells from the T25 or T75 flasks in their recommended media were trypsinised and plated onto 6-well plates, and to new flasks to maintain the stock cells. The plating was done one day before the experiment started or before the days of schistosomule transfer (Figure 4.1). The cells were kept in their recommended media until the the following day when schistosomules and Basch media were added to the cells.

Prior to setting up the experiment, the cells were passaged onto plates and grown in Basch media to assess their growth and to help decide the number of cells to be used in the co-culture experiment. The number of cells per well for each human cell type, therefore, varied depending on their growth rates in Basch media. For co-culture plates, HUVEC were plated at ~90,000 cells/well, HEPG2 at ~100,000 cells/well, and GripTite at ~50,000 cells/well. To start the co-culture, the growth media were removed from the cells. The cells were washed with PBS, replaced with fresh and pre-warmed Basch media, and 1,800 schistosomules were transferred to each well by

pipetting a fixed volume of schistosomule suspension. Since schistosomules can quickly sink to the bottom of its stock flask, the stock flask was resuspended every time before the schistosomules were withdrawn in order to ensure a homogeneous mix and to ensure consistent numbers of schistosomules were added to each well. The plates were maintained at 37 ºC, 5% CO_2, until the schistosomules were transferred to a new plate or collected.

4.2.5.2 Schistosomule transfer

Every three or four days, schistosomules were transferred to new plates of cells. For the transfer, half of the old media was removed, plates were swirled to resuspend the schistosomules, and the schistosomules were transferred to the new plates using a pipette (Figure 4.1). Cells that had been co-cultured with schistosomules at the point of transfer were discarded.

4.2.5.3 Use of blood

From day three of the co-culture, human red blood cells were added to the media at 1:1,000. Before adding to Basch medium, red blood cells were washed with 10 ml of wash media per 1 ml of packed red blood cells. The solution was gently mixed, centrifuged at 500 x g for 5 minutes and then the supernatant and top layer of the pellet removed. This washed blood could be used for up to one week. Wash medium was composed 1% v/v Hepes (sc-286961, Santa Cruz Biotechnology) and 1.5X Antibiotic/Antimycotic (15240-062, Invitrogen), diluted in DMEM, high glucose (D6546, Sigma). Anonymous human packed red blood cells, blood group O was obtained from Cambridge BioScience Ltd following authorised NRES Research Ethics Application according to The UK Human Tissue Act 2004.

4.2.6 Sample collection

On the day of sample collection, the plates were taken out of the incubator for imaging and proceeded straightaway to sample collection. To collect the schistosomules, half of the media was removed and the plates were swirled to resuspend the schistosomules. The media and schistosomules were transferred to Eppendorf tubes (1 well per tube, making one biological replicate). The tubes were centrifuged at 300 x g, 1 min, at room temperature to pellet the schistosomules. Supernatant was transferred to fresh Eppendorf tubes for storage and was not

currently used. To the schistosomule pellet, 1 ml of TRIzol was added followed by inverting the tube to mix and the tubes were placed at room temperature for 30-60 minutes and transferred to -80 ºC storage. For the human cells, once the schistosomule suspension and media had been removed, the plates were kept on dry-ice while schistosomules were being collected. After which, the plates were removed from the dry ice, and 1 ml of TRIzol added directly into each well. The mix was resuspended, transferred into fresh Eppendorf tubes, placed at room temperature for 30-60 minutes, and transferred to -80 ºC storage.

4.3 Results

4.3.1 *Mycoplasma* test

All samples were tested negative for *Mycoplasma* contamination. However, the test did not perform well with the supernatant from the HUVEC cultures - the internal positive control for the PCR reactions was not amplified, suggesting that HUVEC culture supernatant may contain a PCR inhibitor (Figure 4.2). Despite this, I proceeded to use the HUVEC in the experiment because the cells were purchased fresh from ATCC, and should therefore be *Mycoplasma*-free. Furthermore, the HUVEC were cultured in the same environment as other cell types (HEPG2, and GripTite) in which *Mycoplasma* was not detected. The absence of *Mycoplasma* in HUVEC was later confirmed in the RNA-seq results. Mapping to reference genomes was performed by the Sanger Institute NPG core quality control pipeline and showed only a minor percentage of total reads mapped to the *Mycoplasma hyopneumoniae* genome (less than 2% in all samples). This percentage was in the same range as found by the mapping pipeline using irrelevant genomes (mapping controls).

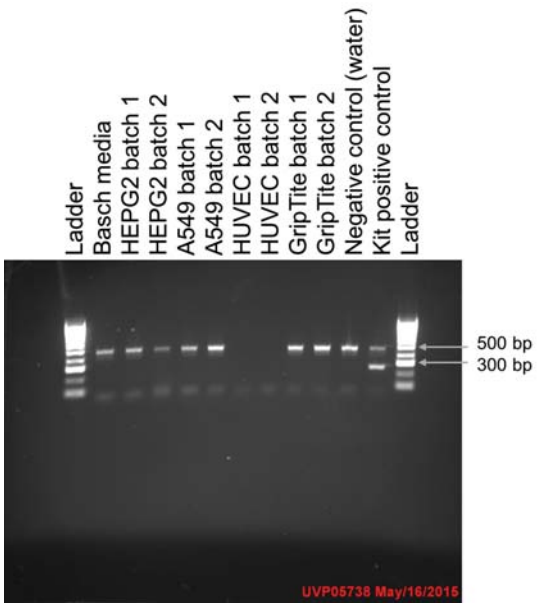

Figure 4.2 Gel image of PCR product from* Mycoplasma *tests

Products from *Mycoplasma* test (PCR-based) as run on 1.2% agarose gel stained with ethidium bromide (1:10,000) and viewed under UV light. Ladders (HyperLadder IV) run in bands of 100 bp increment. Bands in samples at 259 bp indicate presence of *Mycoplasma*. Bands in samples at 481 bp indicate success of PCR reactions.

4.3.2 Worm morphology

Systematic differences in the morphology of schistosomules from different culture conditions could not be discerned (see Figure 4.3 for representative examples). However, slight difference between time points could be seen in the quantities of haemozoin present (Figure 4.3). Despite the similar morphology of schistosomules from different conditions and time points, transcriptomes exhibited distinct profiles.

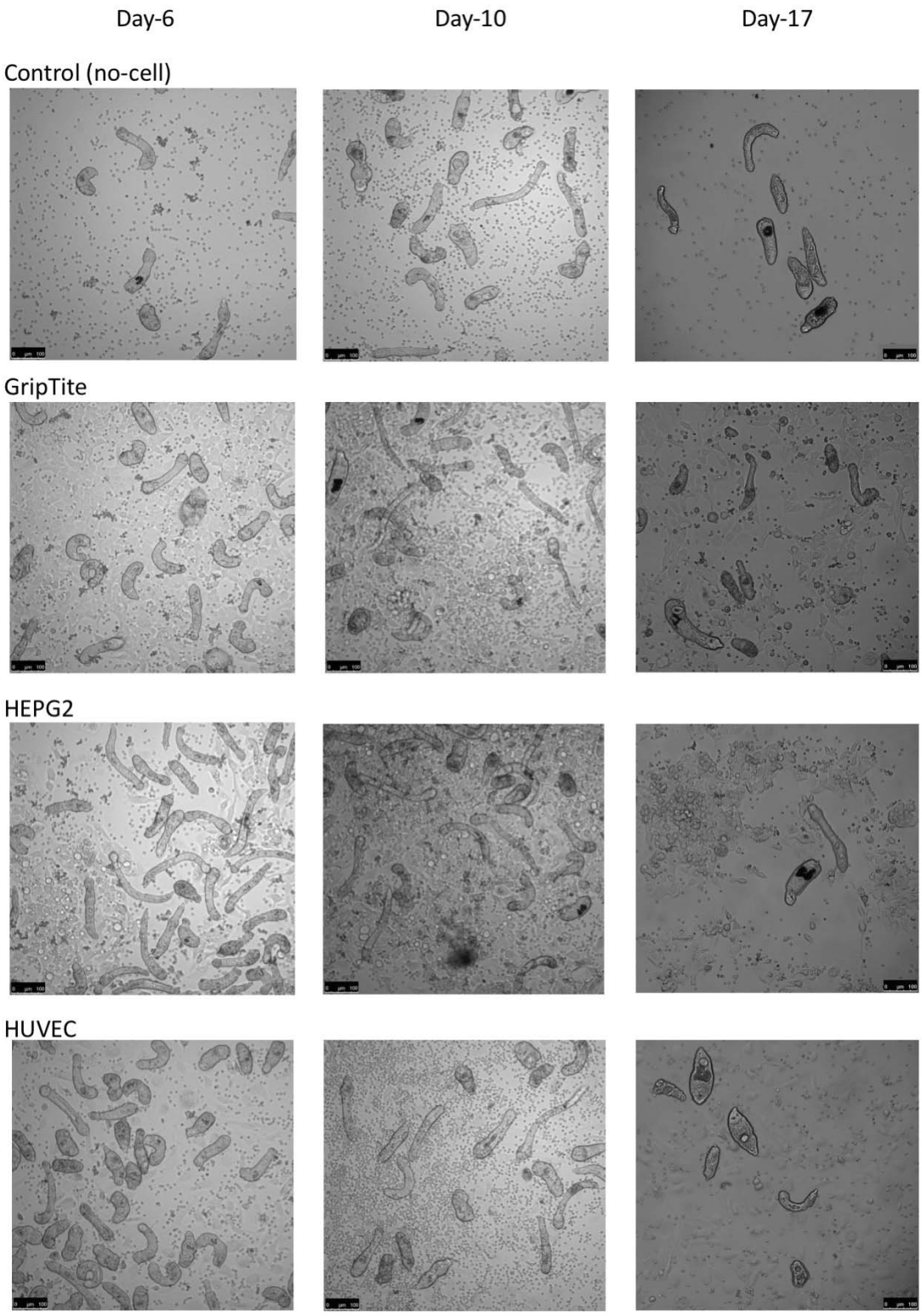

Figure 4.3 Example images from each time point of each co-culture condition

Images of schistosomules captured immediately before the schistosomules and cells were collected at each time point. Co-cultured human cells and erythrocytes can be seen in the background of each image. Scale bars are all 100 μm.

4.3.3 RNA quantity and quality

The yield of extracted RNA ranged from 11 to 54 ng/µl, in 30 µl nuclease-free water. Electropherograms from Agilent Bioanalyzer show a distinct peak of 18S rRNA size in every sample (similar to an example electropherogram in Figure 2.2A), indicating that the RNA samples were not degraded (data not shown). NanoDrop measurements were not taken for samples in this chapter.

4.3.4 Overall profiles of transcriptomes

4.3.4.1 PCA of in vitro *and* in vivo *parasites*

The transcriptomic profiles of *in vitro* schistosomules were compared to the *in vivo* dataset from chapter 3. All *in vitro* schistosomules were most similar to the earliest *in vivo* time point even after 17 days *in vitro*, post-transformation (Figure 4.4). This apparent failure to mature *in vitro* indicates that there are additional requirements of the parasite for development. Amongst the *in vitro* schistosomules, the schistosomules co-cultured with HEPG2 appeared to be most similar transcriptomically to the *in vivo* lung schistosomules (Figure 4.5).

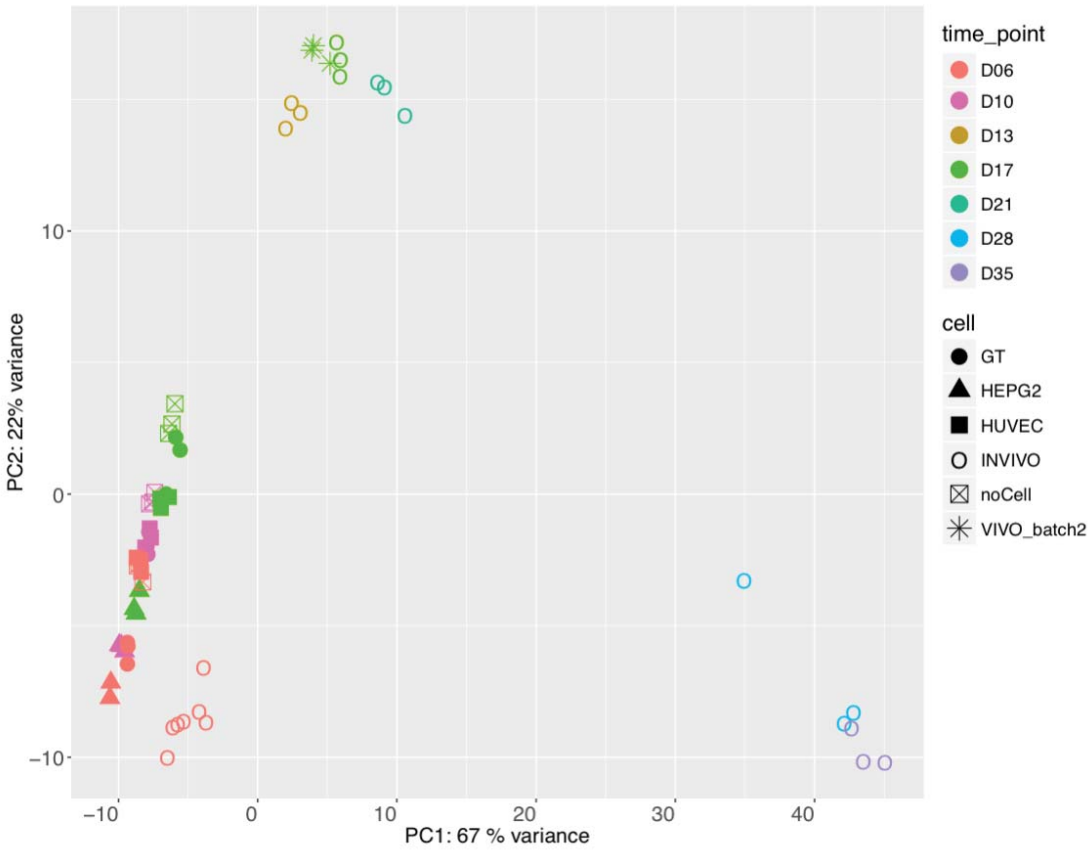

Figure 4.4 PCA of transcriptomes from* in vitro *and* in vivo *datasets

PCA plot of schistosomules from *in vitro* co-culture experiment (this chapter) and all *in vivo* samples (chapter 3). All the *in vivo* samples are shown as open circle. Other shapes represent conditions of the co-culture set up. Colours indicate time points. VIVO_batch2, shown by an asterisk symbol, are day-17 worms from *in vivo* that were processed together with schistosomules from the *in vitro* experiment.

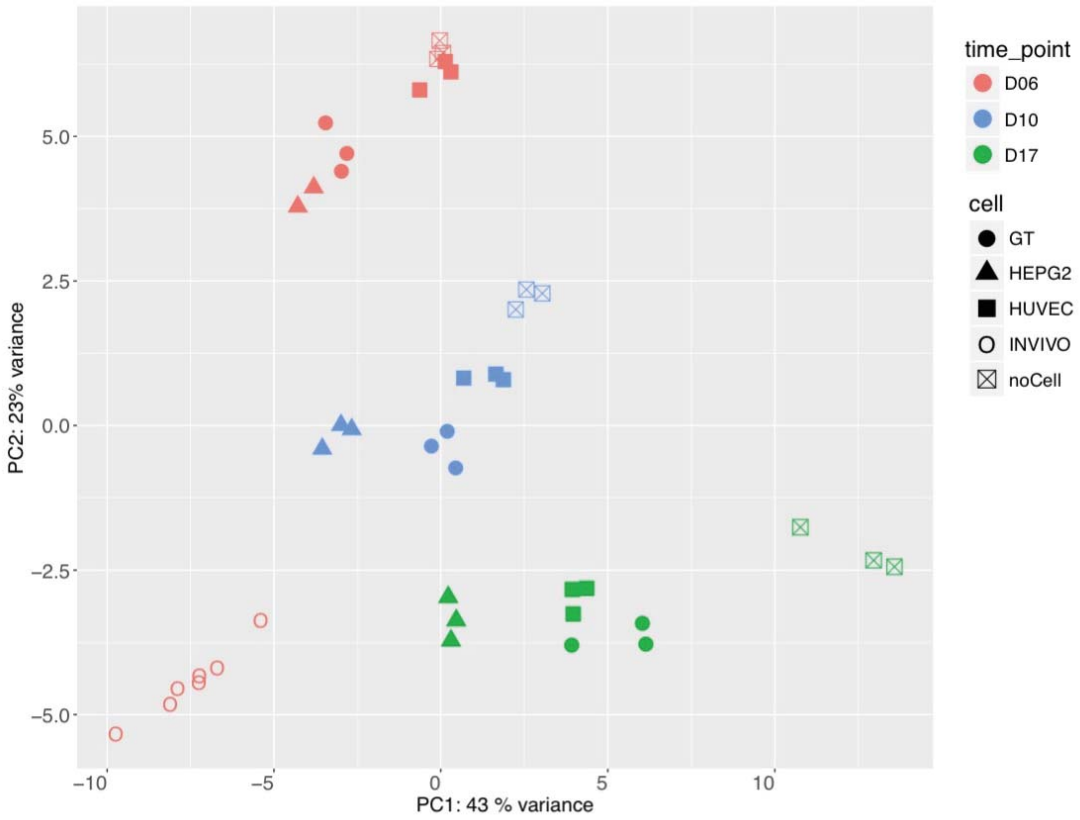

Figure 4.5 PCA of transcriptomes from all* in vitro *parasites and* in vivo *lung schistosomules

PCA plot of schistosomules from the *in vitro* co-culture experiment (this chapter) and day-6 schistosomules from *in vivo* from chapter 3. The *in vivo* day-6 samples are shown as open circle in D06 colour. Other shapes represent conditions of the co-culture set up. Colours indicate time points.

4.3.4.2 PCA of in vitro *parasites*

Considering only *in vitro* samples, the transcriptomes from different samples were separated by both cell-type and by time. The effect of time is largely seen in principal component 1 (PC1; x-axis) which explains 49% of the variance, whereas the effect of cell type is primarily seen in PC2 (y-axis), which explains a further 21% of the variance (Figure 4.6). The differences between co-cultured schistosomules and controls increased over time and became most remarkable at day 17. The differences between co-culture types appeared to slightly diminish over time. Consistently, HEPG2 co-cultured schistosomules were always the most different from the controls. And for day 6 and day 10, HUVEC co-cultured schistosomules were the most similar to the controls inferring few responses when exposed to the HUVEC.

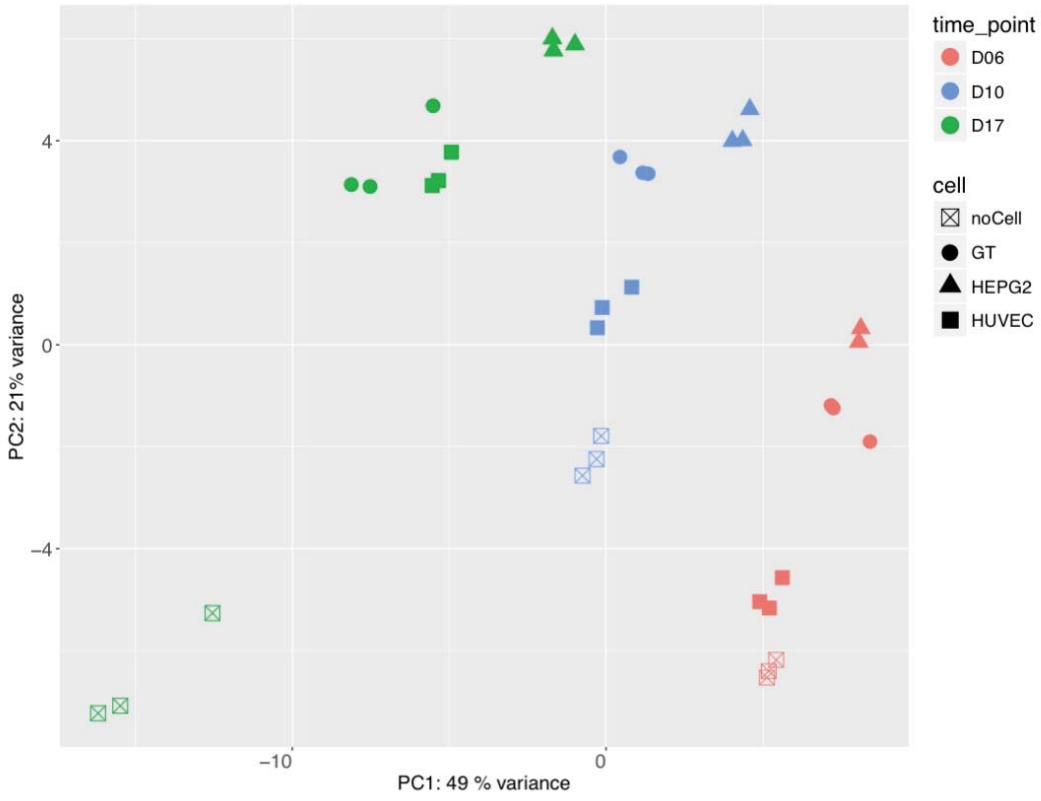

Figure 4.6 PCA of transcriptomes from all* in vitro *schistosomules

PCA plot of all *in vitro* schistosomules from the co-culture experiment. Shapes represent conditions of the co-culture set up. Colours indicate time points.

4.3.5 Overview of gene expression

To explore the effects of different co-cultured cell types on schistosomules, I performed pairwise comparison between co-cultured schistosomules and no-cell controls for each time point, keeping time as a constant variable. As already inferred from PCA plot (Figure 4.6). The number of differentially expressed genes varied with co-culture conditions. HEPG2 co-culture led to the highest number of differentially expressed genes, and HUVEC co-culture led to the smallest number (Figure 4.7, Figure 4.8). In all co-culture conditions at day 17, over 1,400 genes were differentially expressed between co-cultured and control schistosomules (Figure 4.8).

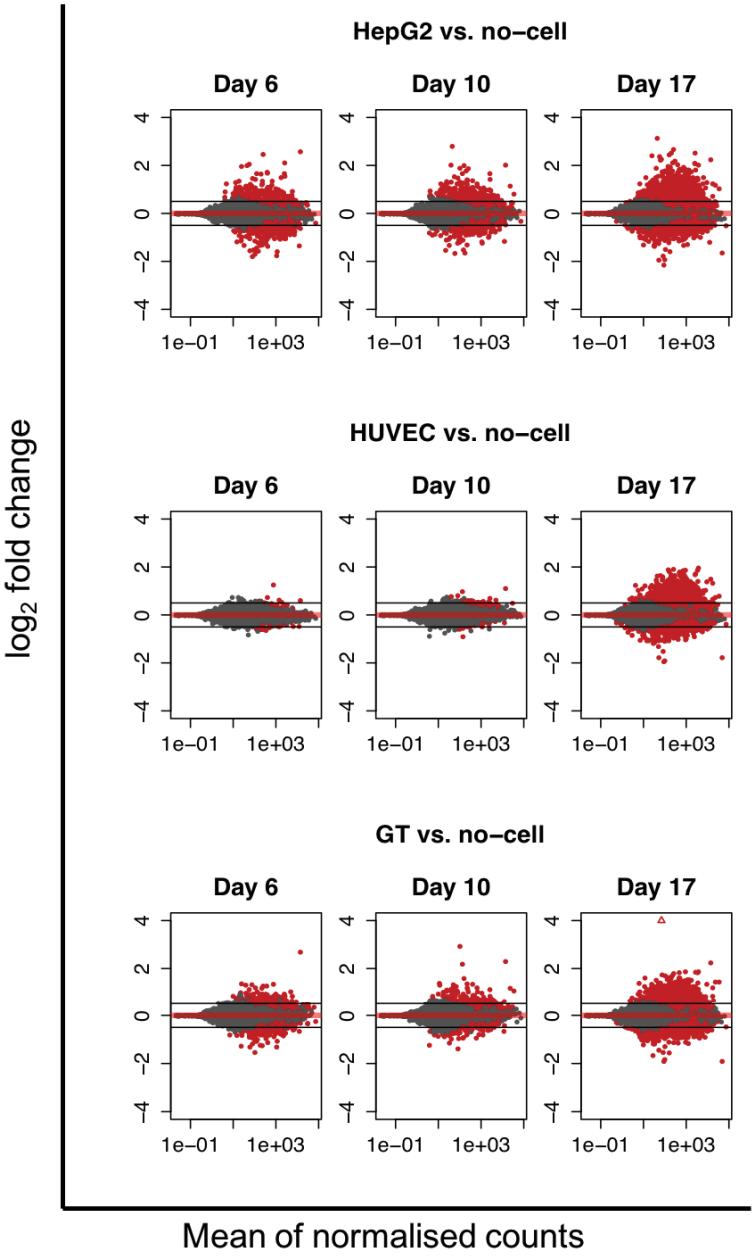

Figure 4.7 MA plots of pairwise comparisons between co-cultured and control schistosomules

MA plots of pairwise comparison between schistosomules co-cultured with each cell type compared to control schistosomules not in co-culture from the three time points. Each dot represents a gene. Dots in red are genes that pass $\log_2$FC cut-off at +/−0.5 (marked by the horizontal lines) and adjusted p-value cut-off at 0.01.

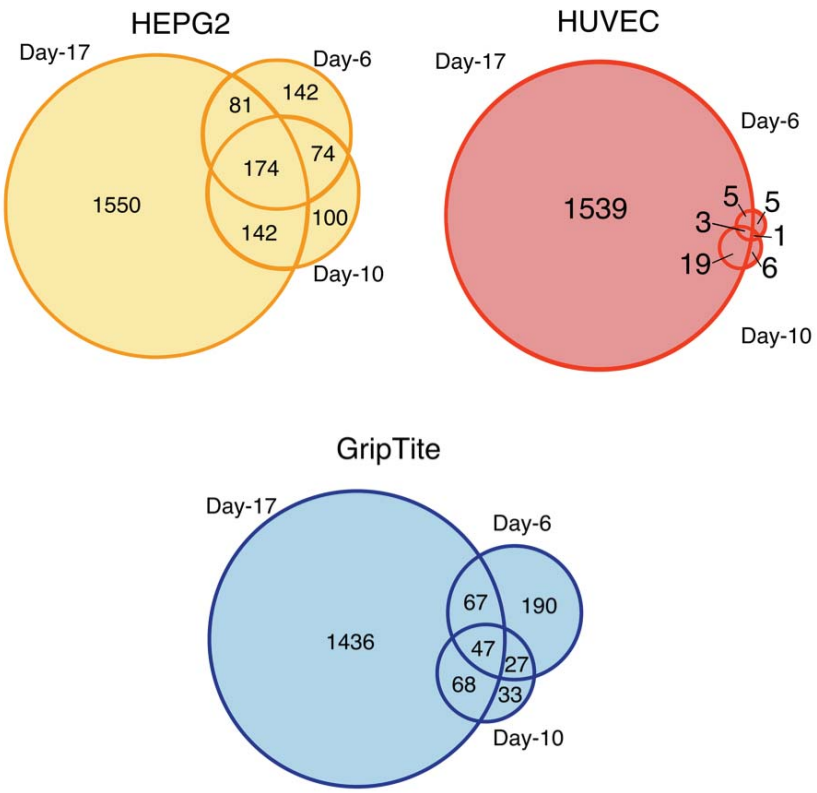

Figure 4.8 Differentially expressed genes unique and common to each pairwise comparison

The number of differentially expressed genes compared between schistosomules of each time point co-cultured with cell *vs.* schistosomules of the same time point not co-cultured with cells. Differentially expressed genes are genes that pass $\log_2$FC cut-off at +/−0.5 and adjusted p-value cut-off at 0.01.

4.3.5.1 Genes differentially expressed in all co-cultured schistosomules at day 17

Since all co-culture conditions led to a large number of differentially expressed genes at day 17 compared to controls, I asked whether the differentially expressed genes were common among all cell types and could therefore represent a generic innate response. 795 genes were differentially expressed only in the day-17 schistosomules but across all co-culture conditions; of these, 584 genes were up-regulated and 211 genes down-regulated (Figure 4.9).

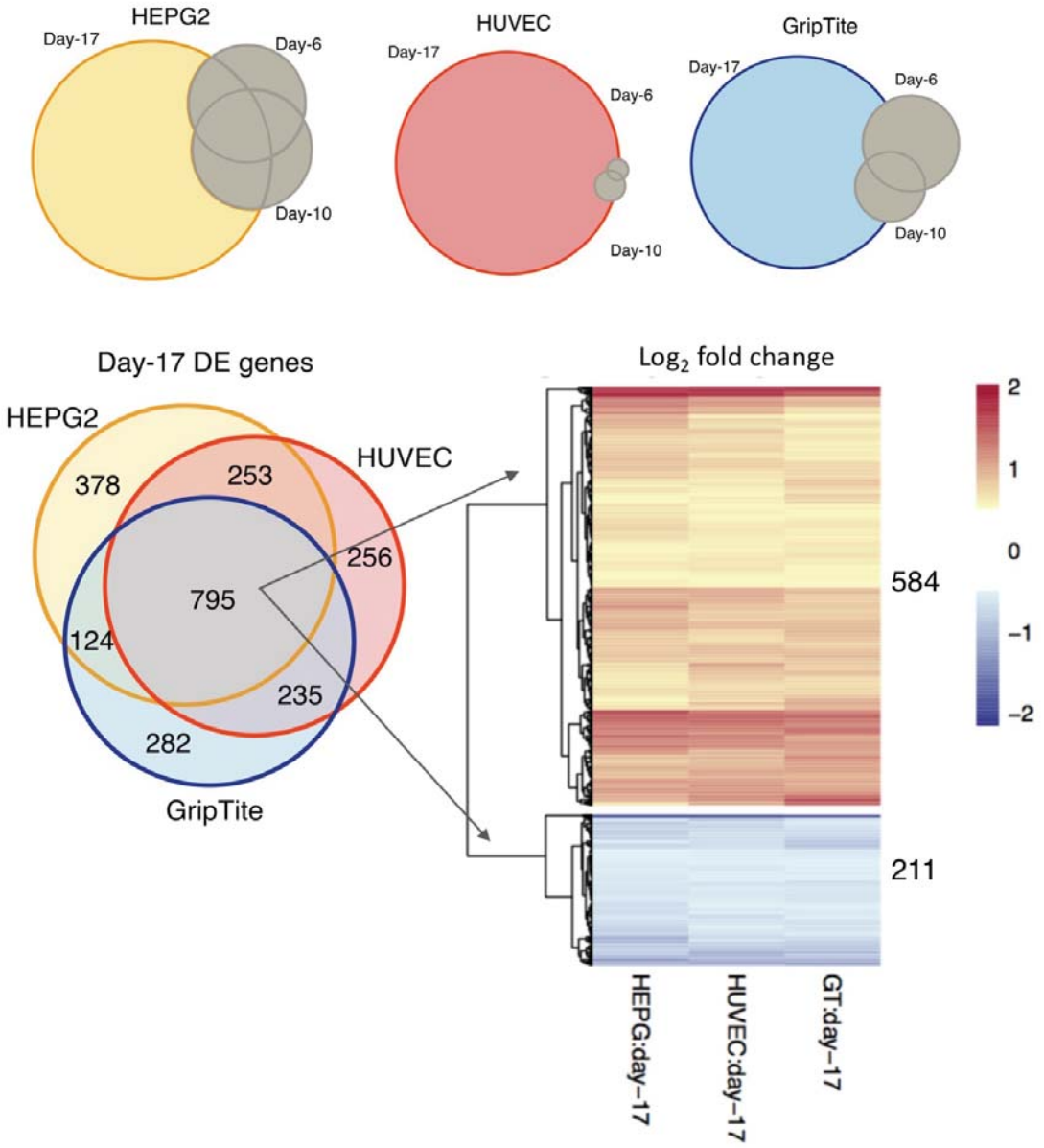

Figure 4.9 Differentially expressed genes in day-17 schistosomules comparing unique and common changes between cell types

The number of genes differentially expressed at day 17 comparing each group (cell types) of co-cultured schistosomules with control schistosomules. Heatmap shows 795 genes whose changes were observed at day 17 when schistosomules were co-cultured with any of the three types of the human cells.

Analysis of enriched GO term annotations amongst the genes up-regulated at day 17, revealed processes related to signalling and cell interactions, such as *protein phosphorylation, Rho protein signal transduction, cell-matrix adhesion, integrin-mediated signaling pathway, ion transport, regulation of small GTPase mediated*

signal transduction (Figure 4.10, Table S4.1). This information alone may suggest increase in signalling processes that might be a response to the cell-culture environment.

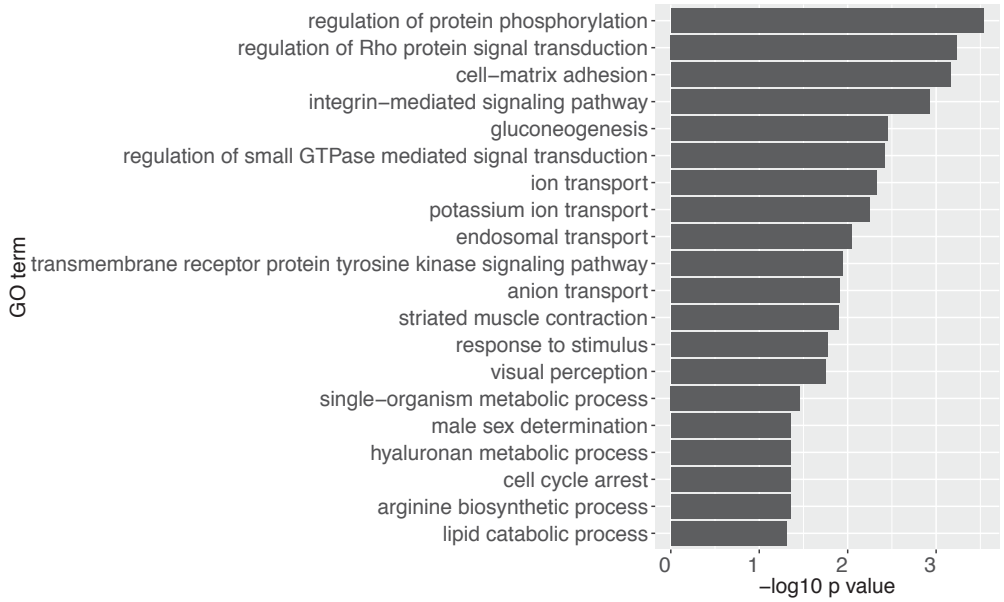

Figure 4.10 Enriched GO terms in genes up-regulated in day-17 co-cultured schistosomules

GO term enrichment of genes up-regulated in all co-cultured schistosomules at day 17 compared to control schistosomules at day 17. Genes input to GO term enrichment analysis are the 584 genes up-regulated in Figure 4.9. Bar chart shows enriched GO terms (biological process) ranked by p-values obtained from topGO package.

In contrast, when down-regulated genes were considered, biological processes related to core cell functions appeared to be inhibited. Examples of these included *processes of mitochondrion protein transport, intracellular protein transmembrane import, protein targeting to mitochondrion,* and *protein folding* (Figure 4.11, Table S4.1). Functions related to gene expression regulation, such as *rRNA processing, mRNA transport,* and *methylation of tRNA and rRNA,* were also affected. Furthermore, *mitotic cell cycle process* was strikingly enriched in down-regulated genes, consistent with the GO term *cell cycle arrest* being enriched in up-regulated genes. Together, this instead suggests that the core biological processes such as gene expression control, mitochondrion functions, and cell cycle might be impeded at day 17 in co-cultured schistosomules, and the increase in expression of components in signalling pathways might be a response to this change.

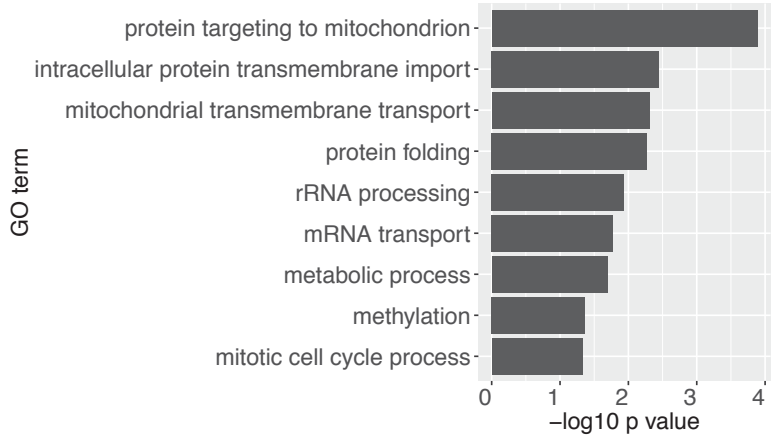

Figure 4.11 Enriched GO terms in genes down-regulated in day-17 co-cultured schistosomules

GO term enrichment of genes down-regulated in co-cultured schistosomules at day 17 compared to control schistosomules at day 17. Genes input to GO term enrichment analysis are from the 211 genes down-regulated in Figure 4.9. Bar chart shows enriched GO terms (biological process) ranked by p-values obtained from topGO package.

4.3.5.2 Changes at day 17 guided subsequent analyses

The dramatic changes in gene expression observed at day 17 are likely to be a consequence of the day-17 control. From the PCA plot (Figure 4.6), the day-17 control schistosomules appear to be an outlier group. In contrast, all day-17 co-cultured schistosomules were close to their counterparts from other time points. Thus, in the absence of co-cultured mammalian cells there appears to be a dramatic effect on the schistosomule leading to considerable transcriptional changes at day 17. Analysing the effects of co-cultured cells on day-17 schistosomules cannot therefore be interpreted just using the day-17 controls. By extension, I did not consider any time points independently, and, instead, focused on genes that showed consistent pattern of changes in all three time points.

In the following section, I investigated how schistosomules responded to the co-culture environment. First, I explored common effects on schistosomules in all co-culture conditions to understand generic responses to mammalian environment. Second I considered specific responses to HEPG2 and HUVEC co-culture conditions because of their relevant environment during their intramammalian stages. In a separate analysis, the generic responses and the cell-specific responses were

investigated when all day-17 samples were excluded from the dataset; similar results were observed (data not shown).

4.3.6 Schistosomules in co-cultured environment: generic responses

Generic responses were defined as genes that were differentially expressed in schistosomules from all three co-culture conditions compared to their respective no-cell control, and that the changes were observed at all time points (Figure 4.12). Only two genes were common across all co-cultured schistosomules and both were up-regulated (log_2FC range = 0.517-2.689). This was due to only three genes being differentially expressed in schistosomules co-cultured with HUVEC (*fibrillin 2* (Smp_067800), *hypothetical protein* (Smp_052880), *hypothetical protein* (Smp_141500)). The two generic responses genes were *fibrillin 2* (Smp_067800) and *hypothetical protein* (Smp_052880). The two genes were investigated in more details.

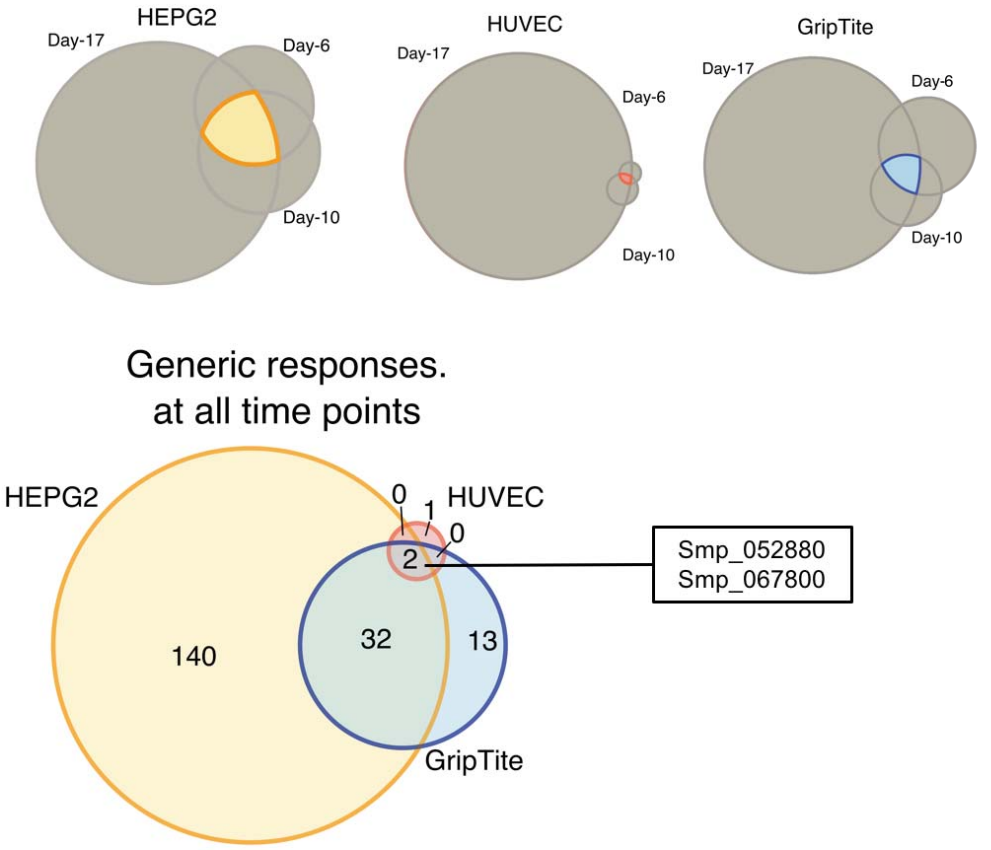

Figure 4.12 Generic responses in schistosomules

The number of genes that displayed shared and distinct changes in responses to each co-culture condition at all three time points. Generic responses were genes that were differentially expressed in all three time points and in all co-cultured conditions: two genes with this pattern are Smp_052880, and Smp_067800.

4.3.6.1 Smp_067800, potential non-coding gene

For the *fibrillin 2* (Smp_067800), previous transcriptomic works showed that the gene is highly expressed in miracidia (Taft *et al*., 2009), adults and 3hr schistosomules, and the expression reduces by almost five-fold in 24-hour schistosomules (Protasio *et al*., 2012). Its functions, however, have not been identified. BLASTP was used to search for sequence similarity in the NCBI non-redundant protein database. The returned hits were proteins in parasitic worms including three species of schistosomes (*S. mansoni*, *S. japonicum*, and *S. haematobium*), liver flukes *Opisthorchis viverrini*, and *Clonorchis sinensis*, as well as a member of roundworm group *Trichinella*

pseudospiralis, and tapeworm *Hymenolepis microstoma* (poor matching score) (Table 4.1). Gene trees in WormBase ParaSite revealed orthologues only in the *Schistosoma* genus and liver fluke *O.* and *C. sinensis* with consistent alignment across species. The gene Smp_067800 also has single copy in *S. mansoni* (no paralogues) (Figure 4.13). It appeared, therefore, that the gene is helminth-specific.

Table 4.1 BLASTP output with Smp_067800 amino acid sequence as a query

Description	Query cover	E value	Ident	Accession
putative fibrillin 2 [*Schistosoma mansoni*]	99%	2.00E-159	99%	XP_018651624.1
hypothetical protein MS3_00859 [*Schistosoma haematobium*]	98%	1.00E-129	91%	XP_012792489.1
fibrillin 2 [*Schistosoma mansoni*]	81%	4.00E-97	88%	AAA99800.1
fibrillin 2 [*Schistosoma japonicum*]	88%	2.00E-90	76%	CAX80657.1
fibrillin 2 [*Clonorchis sinensis*]	96%	4.00E-39	45%	GAA36848.1
hypothetical protein T265_15193 [*Opisthorchis viverrini*]	92%	1.00E-38	45%	XP_009175256.1
hypothetical protein T4E_7035 [*Trichinella pseudospiralis*]	55%	3.00E-17	43%	KRX82181.1
expressed protein [*Hymenolepis microstoma*]	80%	0.007	27%	CDS34135.1

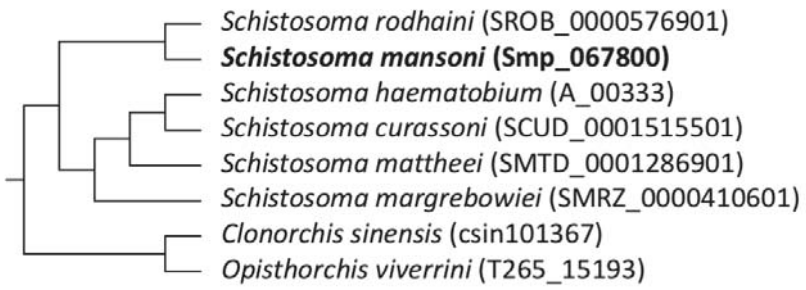

Figure 4.13 Homologous relationship of Smp_067800

Gene tree of Smp_067800 (annotated as *fibrillin 2*) and its homologues. The information of the tree 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.

Using InterProScan to identify signature domains, no domains were found in accordance with the annotated name *fibrillin 2*. Signature domains for *fibrillin 2* are

transforming growth factor beta binding protein (TB) domain and two *hydrid* domains (Robertson *et al*., 2011). Previous work showed the absence of fibrillin in *Drosophila, C. elegans*, and other nematode species (Piha-Gossack *et al*., 2012). The current gene model only contains exon 2, with exon 1 previously removed (Figure 4.14). The gene model is consistent with the RNA-seq evidence - both from deep-coverage short-read mapping, and from sparser coverage long-read data (Figure 4.14A and B). Small numbers of reads mapped to genomic region preceeding 5' end of the gene but the number of reads mapped was very low compared to reads mapped to the Smp_067800 exon region, and no mate-pair connection was observed between the two regions. Intriguingly, the earliest start codon in this exon 2 region would yield polypeptide of (82 aa long) with 5'UTR being twice as long as the coding region. The mapping of RNA-seq data, however, did not resemble mapping normally observed at a 5'UTR. Mapping at 5'UTR tends to show a slope in the depth of coverage, but in this case, the mapping showed a sharp boundary which resembled the coverage often observed at the start of an exon (Figure 4.14A). It may be that this gene was once a protein coding gene but became a non-coding RNA. Search on Rfam for known RNA families (Nawrocki *et al*., 2015) yield no match for this gene. Currently, the knowledge on non-coding RNA, even in model organisms, is limited and it is not possible to predict the function of this potentially non-coding RNA. Another explanation for the missing first exon could be an error in assembly. The gene is located in genomic regions with a gap near its 5' end where coding exon seemed to be missing. Future versions of the *S. mansoni* genome may help solve the ambiguity.

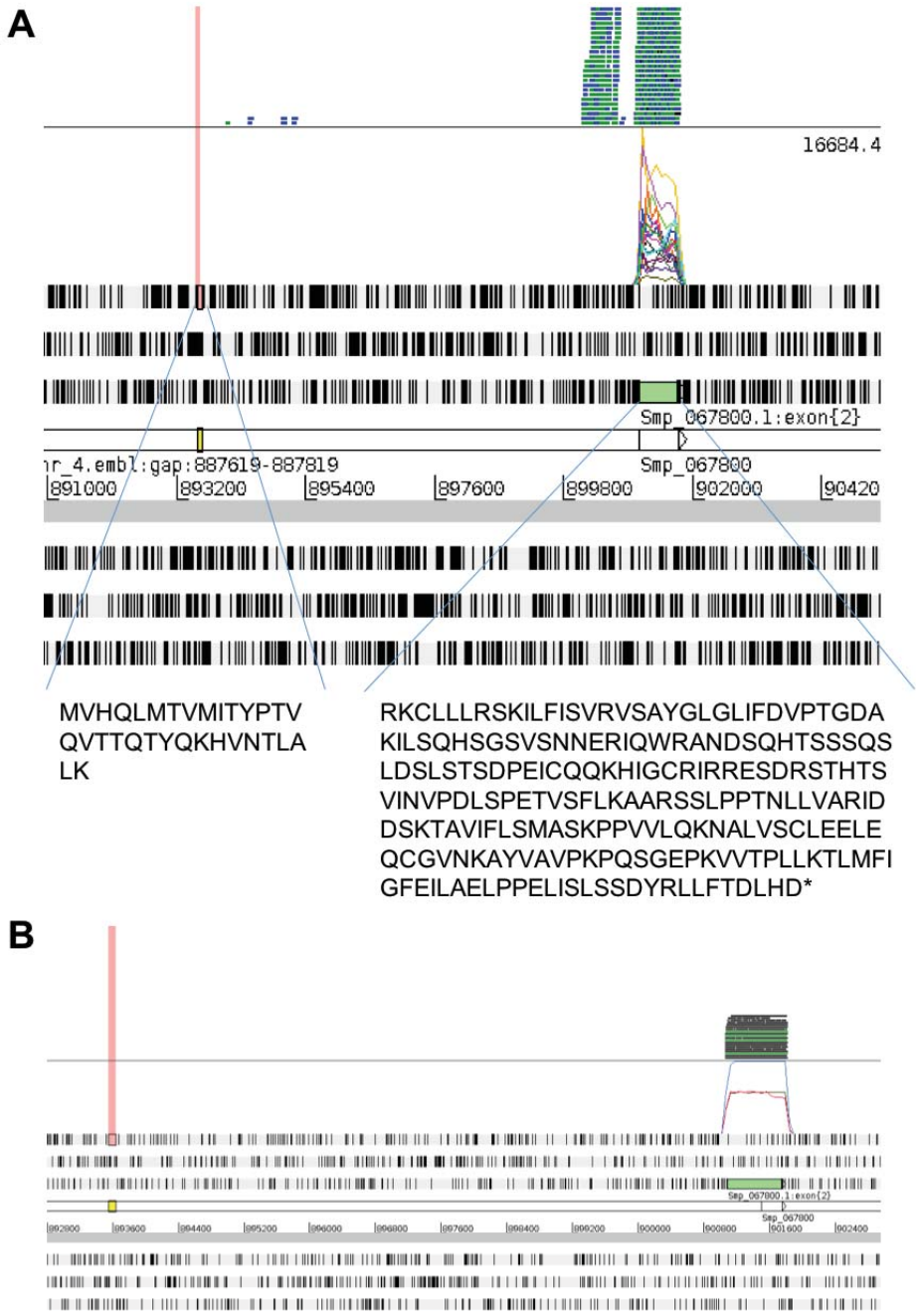

Figure 4.14 Genomic region of Smp_067800 and alignment of RNA-seq reads

Screenshot of genomic region and RNA-seq read mapping to gene Smp_067800. A) Short read mapping. B) Iso-seq read mapping. Amino acid sequence for the missing first exon (with no RNA-seq read mapped to the region) come from old GenBank accession identifier (CCD79016.1) that is the top BLASTP hit to this gene. Amino acid sequence for the exon 2 (with RNA-seq reads mapped to it) come from the version of *S. mansoni* annotation used in this analysis.

4.3.6.2 Smp_052880, potential neuropeptide

For the *hypothetical protein* (Smp_052880), RNA-seq short read mapping supported that this is a complete gene model with coding potential (Figure 4.16A). A BLASTP search (amino acid sequence) on NCBI non-redundant protein database found matches in three species of schistosomes, two other parasitic flatworms *O. viverrini*, and *C. sinensis*, and two free-living flatworms *Dugesia japonica*, and *Schmidtea mediterranea*. All of the hits to the parasitic flatworms were hypothetical proteins, whereas both hits to free living flatworms were neuropeptides. The last hit was a gene from the fungi phylum, and was not included in further investigations due to its high E value (Table 4.2). Genetree from WormBase ParaSite revealed orthologues limited to *Schistosoma* species and avian schistosome (Figure 4.15).

Table 4.2 BLASTP output with Smp_052880 amino acid sequence as query

Description	Query cover	E value	Ident	Accession
hypothetical protein Smp_052880 [*Schistosoma mansoni*]	99%	6.00E-75	100%	XP_018650687.1
hypothetical protein MS3_07516 [*Schistosoma haematobium*]	99%	5.00E-53	76%	XP_012798865.1
hypothetical protein [*Schistosoma japonicum*]	97%	8.00E-52	78%	CAX75147.1
hypothetical protein [*Schistosoma japonicum*]	97%	3.00E-51	77%	CAX75150.1
TPA: neuropeptide precursor-5 [*Schmidtea mediterranea*]	40%	0.005	39%	DAA33904.1
hypothetical protein CLF_105825 [*Clonorchis sinensis*]	24%	0.077	43%	GAA51282.1
hypothetical protein T265_07707 [*Opisthorchis viverrini*]	54%	0.21	31%	XP_009171559.1
neuropeptide-25 [*Dugesia japonica*]	30%	0.29	43%	BAV14854.1
Piso0_003363 [*Millerozyma farinosa* CBS 7064]	47%	4.2	37%	XP_004200765.1

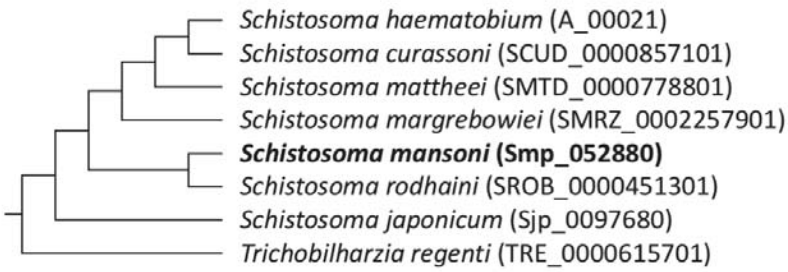

Figure 4.15 Homologous relationship of Smp_052880

Smp_052880 (*hypothetical protein*) and its homologues. The information of the tree 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.

The alignment of the query sequence Smp_052880 to all the NCBI BLASTP hits, including the neuropeptide in free-living flatworms, revealed the amino acid motif D+PWGKR was common to most of the sequences. Using motif search and ProSite tool, queried with amino acid sequence of the Smp_052880 genes, the gene product contain an amidation site, recognised as x-G-[RK]-[RK] motif. All orthologues from WormBase ParaSite genetree also contain this amidation site, except avian schistosome *Trichobilharzia regenti*. For *T. regenti*, the last amino acid of a potential amidation site was replaced by a stop codon, possibly because its genome is in a draft form. The presence of such motif in all *Schistosoma* species is interesting because amidation at C terminal of a peptide is a post-translational modification essential for turning a precursor neuropeptide into an active neuropeptide (The UniProt Consortium, 2017) and neuropeptide signalling is involved in multiple processes in *S. mansoni* (Collins *et al*., 2010; Ribeiro and Patocka, 2013).

This amidation site is conserved in all of the genes from BLASTP results, and the homologues of Smp_052880 are limited to schistosomes suggesting that all the *hypothetical proteins* on this list (Table 4.2, Figure 4.15) could be a precursor of neuropeptides. However, it is worth noting that neuropeptides in free-living flatworm such as one from *S. mediterranea* tend to have more than one amidation site, leading to many neuropeptides upon activation (Collins *et al*., 2010). With all the matches from the BLASTP search, all the parasite genes have only one amidation site, and one incomplete amidation site. An exception was *O. viverrini* where two sites were found (Figure 4.16). Given that the gene in *S. mansoni* was up-regulated when exposed to

the human cell co-culture, its orthologues might be involved in responding to host environment in other species of parasites.

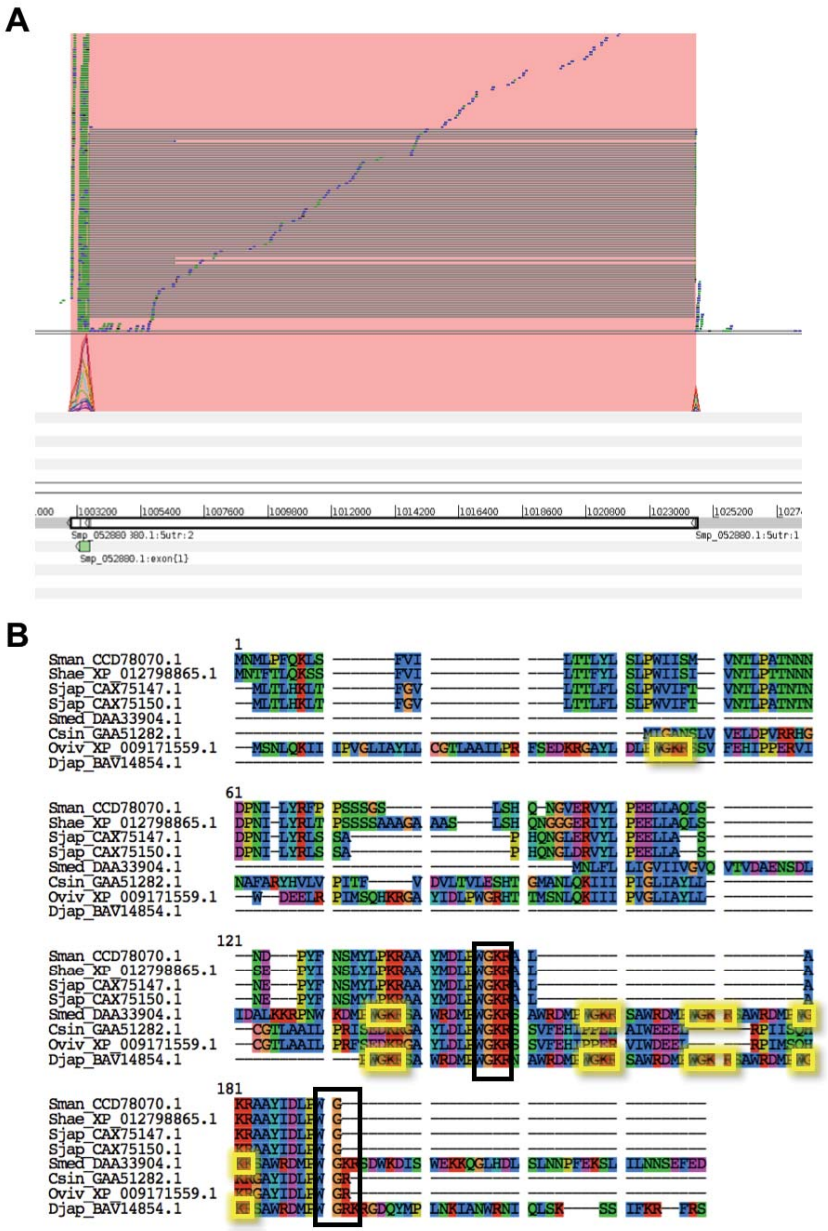

Figure 4.16 Smp_052880 genomic region and amino acid sequence alignment with other BLASTP hits

A) Screenshot of genomic region and RNA-seq short read mapping of gene Smp_052880. B) Alignment of amino acid sequences of genes from BLASTP hits of Smp_052880 (Table 4.2), excluding the fungus gene which has poor matching score. Black and yellow boxes marks amidation sites. Yellow boxes emphasise multiple amidation sites in free-living species and an additional site in a parasitic species *O. viverrini*.

4.3.7 Schistosomule adaptation to HEPG2 environment

4.3.7.1 Determining cell-specific responses

To identify specific changes in response to HEPG2 environment, schistosomule transcriptomes were grouped as being from HEPG2 or from non-HEPG2 (schistosomules co-cultured with HUVEC, with GripTite, and control schistosomules). Pairwise comparison between the HEPG2 and non-HEPG2 groups was used to determine HEPG2-specific responses. This approach treated transcriptomes from different time points as replicates, thereby increasing the number of replicates per group and increasing the power for differential expression analysis. In addition, genes with drastic shift in expression level between time point would result in large variation and likely to not pass cut-off of the adjusted p-values.

4.3.7.2 Differentially expressed genes in HEPG2 co-culture condition

A total of 129 genes are differentially expressed between HEPG2 vs. non-HEPG2 groups (adjusted p-value < 0.01, $\log_2$FC > 0.5 or < -0.5), with 67 genes being up-regulated and 62 genes being down-regulated (Figure 4.17, Table 4.3, Table 4.4, Table S4.2, Table S4.3). Up-regulated genes were enriched in GO terms *iron transport*, *GPCR signalling*, and *oxidation-reduction process*. In addition, *purine ribonucleoside salvage*, and multiple metabolic processes were among the enriched GO terms. In contrast, down-regulated genes were enriched in GO terms with biological functions related to *regulation of signal transduction*, *development*, *leukotriene biosynthesis*, *cell-matrix adhesion*, and *fatty acid biosynthetic process* (Figure 4.18, Figure 4.19, Table S4.4). There were two enriched GO terms from up-regulated genes that should not be included because the genes involved were annotated as *hypothetical proteins* and InterProScan found no signature domain associated with such GO terms. These are *ROS metabolic process*, and *sterol biosynthetic process* (Table S4.4). The GO terms were assigned by the original genome project based on domain-sensitivity thresholds determined by InterPro. Over time, these thresholds are occasionally revised causing functions assigned to genes to become out of date. This observation stresses the importance of not drawing conclusions from an enriched GO term itself, rather the genes contributing to that term must be directly assessed.

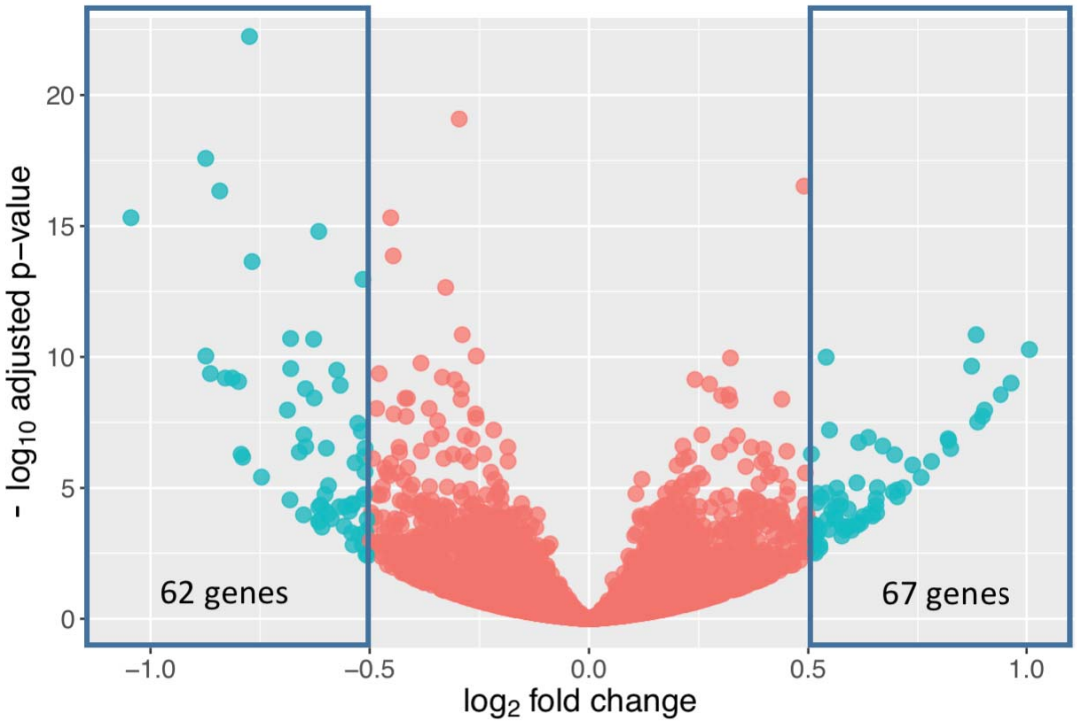

Figure 4.17 Volcano plot of pairwise comparison between HEPG2* vs. *non-HEPG2 schistosomules

Volcano plots of genes from pairwise comparison of HEPG2 and non-HEPG2 schistosomules. Each dot is a gene. In blue are genes that pass log_2FC cut-off at +/−0.5 and adjusted p-value cut-off at 0.01 and therefore were counted as differentially expressed genes.

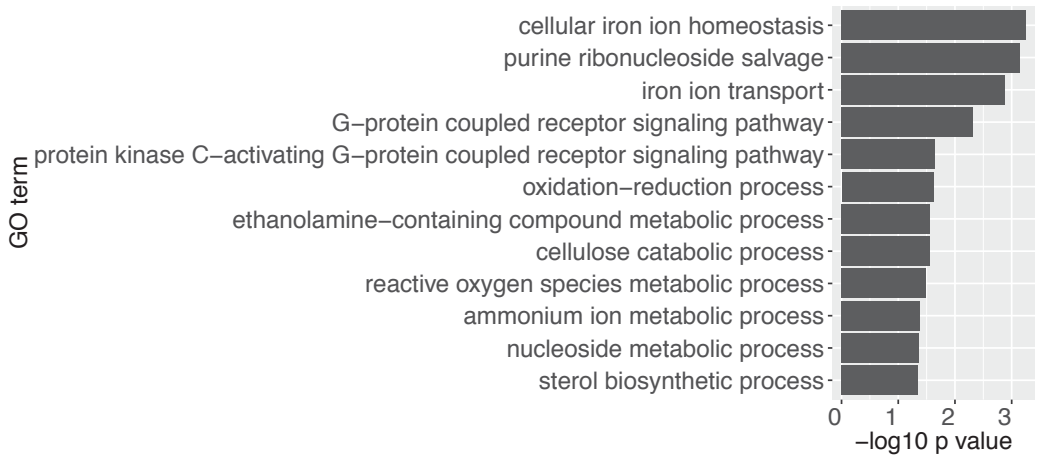

Figure 4.18 Enriched GO terms in genes up-regulated in HEPG2 compared to non-HEPG2 schistosomules

Bar chart shows enriched GO terms (biological processes) of genes that were up-regulated in HEPG2 schistosomules compared to non-HEPG2 schistosomules, ranked by p-values obtained from topGO package.

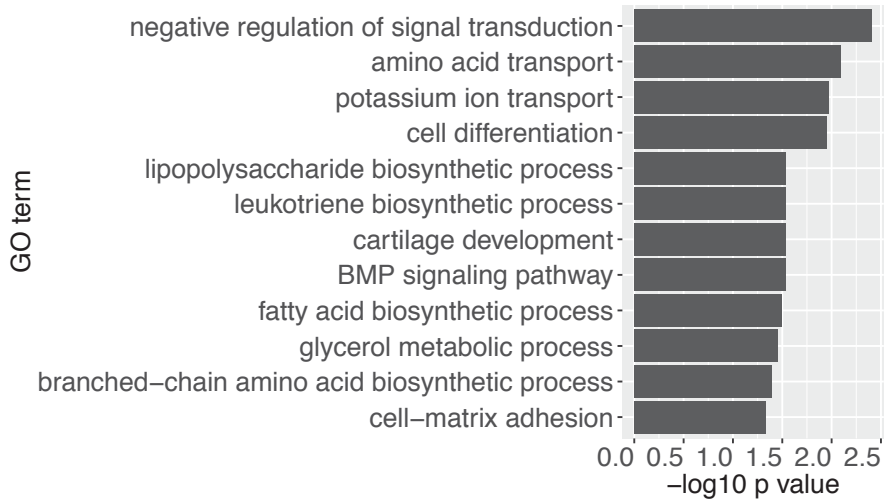

Figure 4.19 Enriched GO terms in genes down-regulated in HEPG2 compared to non-HEPG2 schistosomules

Bar chart shows enriched GO terms (biological processes) of genes that were down-regulated in HEPG2 schistosomules compared to non-HEPG2 schistosomules, ranked by p-values obtained from topGO package.

Oxidation-reduction and stress responses

Up-regulated genes related to oxidation-reduction processes were the two *Ferritin-2 heavy chain* (Smp_047660, Smp_047680), and *tryparedoxin peroxidase* (*Peroxiredoxin*, Smp_062900) (Table 4.3, Table S4.2). *Tryparedoxin peroxidase* was also previously found in proteomic analysis of *S. mansoni* egg (Abdulla *et al*., 2011). Although its functions have not been empirically characterised, an analysis of its amino acid sequence on the InterProScan showed that the protein contains signature domains for peroxiredoxin. An paralogue of this gene (Smp_158110, Appendix E) is a potential antischistosomiasis drug target (Li *et al*., 2015). In addition, up-regulated *glucose dehydrogenase* (Smp_212180) contains conserved protein domains that suggested its role in oxidation-reduction processes. Among the top 10 up-regulated genes was *universal stress protein* (Smp_136890) (Table 4.3). This gene was not previously found expressed during intramammalian stages except for low expression in schistosomules at 3- and 24- hour (Isokpehi *et al*., 2011; Protasio *et al*., 2012). Consistently, the gene was barely expressed in the transcriptomes from *in vivo S. mansoni* (Chapter 3, data not shown). Its up-regulation *in vitro* when co-cultured with HEPG2 could therefore be an *in vitro* artefact.

Table 4.3 Top 20 genes up-regulated in HEPG-2 compared to non-HEPG2 schistosomules

Gene identifier	Log_2FC (HEPG2/non-HEPG2)	Adjusted p-value	Product name
Top 20 genes up-regulated in HEPG2 schistosomules			
Smp_049850	1.01	5.22E-11	hypoxanthine guanine phosphoribosyltransferase
Smp_181510	0.96	1.00E-09	hypothetical protein
Smp_159810	0.94	2.75E-09	MEG-2 (ESP15) family
Smp_187410	0.90	1.06E-08	hypothetical protein
Smp_180620	0.90	1.84E-08	MEG 17
Smp_074560	0.89	3.05E-08	hypothetical protein
Smp_212180	0.88	1.43E-11	glucose dehydrogenase (acceptor)
Smp_180330	0.87	2.25E-10	MEG 2 (ESP15) family
Smp_126880	0.82	3.12E-07	hypothetical protein
Smp_136890	0.82	1.55E-07	universal stress protein
Smp_147740	0.82	1.33E-07	family M13 unassigned peptidase (M13 family)
Smp_138070	0.78	9.78E-07	MEG-3 (Grail) family
Smp_167120	0.76	3.94E-06	Peptidase M8
Smp_047680	0.74	1.32E-06	Ferritin-2 heavy chain
Smp_047660	0.72	9.87E-06	Ferritin-2 heavy chain
Smp_138080	0.70	2.17E-05	MEG-3 (Grail) family
Smp_193400	0.70	1.19E-05	hypothetical protein
Smp_062900	0.70	5.39E-07	tryparedoxin peroxidase
Smp_074570	0.69	1.43E-05	hypothetical protein
Smp_148820	0.67	2.51E-07	hypoxanthine guanine phosphoribosyltransferase

Table 4.4 Top 20 genes down-regulated in HEPG-2 compared to non-HEPG2 schistosomules

Gene identifier	Log_2FC (HEPG2/non-HEPG2)	Adjusted p-value	Product name
Top 20 genes down-regulated in HEPG2 schistosomules			
Smp_108550	-1.04	4.79E-16	hypothetical protein
Smp_045200	-0.87	2.61E-18	tegument-allergen-like protein
Smp_024180	-0.87	9.26E-11	placenta specific gene 8 protein
Smp_158480	-0.86	4.33E-10	AMP dependent ligase
Smp_166020	-0.84	4.56E-17	hypothetical protein
Smp_162500	-0.83	6.39E-10	drug efflux protein
Smp_010770	-0.81	6.39E-10	elongation of very long chain fatty acids
Smp_102190	-0.80	8.68E-10	steroid dehydrogenase
Smp_126290	-0.79	5.18E-07	hypothetical protein
Smp_175290	-0.79	6.72E-07	hypothetical protein
Smp_170280	-0.77	5.79E-23	integrin alpha ps
Smp_158510	-0.77	2.26E-14	diacylglycerol O-acyltransferase 1
Smp_175300	-0.75	3.86E-06	hypothetical protein
Smp_150640	-0.69	1.06E-08	hypothetical protein
Smp_161310	-0.68	2.85E-05	IQ domain containing protein D
Smp_197370	-0.68	1.99E-11	hypothetical protein
Smp_124600	-0.68	2.79E-10	hypothetical protein
Smp_046640	-0.66	4.30E-07	twik family of potassium channels
Smp_141690	-0.65	1.06E-04	hypothetical protein
Smp_160590	-0.65	9.35E-08	hypothetical protein

Signalling

Signalling-related genes were differentially expressed in HEPG2 schistosomules and can be found in both up-regulated and down-regulated groups. GPCRs were up-regulated but were not among the top differentially expressed genes. This is understandable because GPCRs are at the start of signal transduction pathways and such pathways involve amplification of the signal as it cascades through thus GPCR will only be expected to make a minor contribution to the overall changes in transcriptome. The three GPCRs up-regulated in HEPG2 schistosomules are all in Rhodopsin class and have not been characterised in *S. mansoni* (Smp_012920, Smp_041700, and Smp_170610) (Table S4.2). All of the three GPCRs contain homologues only in invertebrates and all are limited to helminth group (Appendix E). Interestingly, in WormBase ParaSite release 8, Smp_041700 orthologues in molluscs

(*C. gigas*) include genes annotated as a *growth hormone receptor*, and a *chemokine receptor* - both could be related to the schistosomule development and host immune modulation (Appendix F). *Calcium binding proteins* and *calmodulin* were also up-regulated in HEPG2 (Smp_033000, Smp_033010, Smp_134500) (Table S4.2).

Amongst down-regulated genes, *potassium ion transport* was an enriched GO term, with twik family of potassium channels being the genes down-regulated (Smp_046640, Smp_046650) (Figure 4.19, Table 4.4, Table S4.3). This family of potassium ion transport functions in neuronal signal conduction and in setting voltage differences across cell membrane. Another enriched GO term was *negative regulation of signal transduction*, encompassing genes *noggin* (Smp_099440) and *suppressor of cytokine signalling 6* (Smp_194390), both of which could be involved in developmental control as well as inflammatory responses. In addition, *leukotriene A 4 hydrolase* (Smp_007550) was down-regulated. The enzyme leukotriene A 4 hydrolase is required for the biosynthesis of proinflammatory molecule leukotriene B4 (Sharma and Mohammed, 2006) which is involved in cytotoxicity of *S. mansoni* (Rogerio and Anibal, 2012).

MEGs and TALs

Differentially expressed genes in HEPG2 schistosomules also included genes with potential roles in host-parasite interactions although their functions are largely uncharacterised. Among the top 10 up-regulated genes, ranked by fold changes, were *MEGs* (Table 4.3). The schistosomules co-cultured with HEPG2 may have detected certain host factors and expressed *MEGs* in response to the environmental cues. In contrast, schistosomules co-cultured with HUVEC (section 4.3.8) did not up-regulate expression of any *MEGs*. This specificity is consistent with the notion that *MEGs* have roles in host-parasite interactions. A total of 10 *MEGs* from multiple groups were up-regulated in HEPG2 schistosomules. These were *MEG-2* (Smp_159800, Smp_159810, Smp_159830, Smp_179860, Smp_180330), *MEG-3* (Smp_138070, Smp_138080, Smp_138060), *MEG-17* (Smp_180620), and *MEG-10* (Smp_152590). Out of the 10 genes, six genes contain signal peptide as predicted by SignalP v4.1 (Petersen *et al*., 2011). These were *MEG-17* (Smp_180620), some of the *MEG-2* (Smp_159800, Smp_159830), and all of the *MEG-3* (Smp_138070, Smp_138080, Smp_138060). In addition, *TAL-1* (Smp_045200) was the second most down-

regulated gene (1.74 fold). *TAL* genes are developmentally regulated (Fitzsimmons *et al*., 2012) and the presence of HEPG2 appeared to stimulate the change in gene expression.

4.3.7.3 Summary

In summary, the schistosomules alter their gene expression in response to HEPG2 environment. Up-regulated genes were enriched in biological functions related to oxidation-reduction processes, stress responses, GPCR signalling pathway, and acquisition of purine nucleotides. Down-regulated genes included those with potential roles in developmental control, inflammation, and transport of sodium, potassium, and amino acids (which may be involved in nerve signal transduction). Genes with unidentified functions such as *MEGs* and a *TAL* were among the top differentially expressed genes.

Next, I explored the specific responses of schistosomules to HUVEC co-culture condition. The schistosomule transcriptomes were placed into HUVEC and non-HUVEC groups.

4.3.8 Schistosomule adaptation to HUVEC environment

4.3.8.1 Differentially expressed genes in HUVEC co-culture condition

Comparing HUVEC and non-HUVEC schistosomules, three genes were found to be differentially expressed with adjusted p-value < 0.01 (two down-regulated, and one up-regulated) (Figure 4.20, Table 4.5). In this comparison, $\log_2$FC cut-offs were not used because all effects were small. The biggest change was of -1.27 fold ($\log_2$FC -0.34, down-regulated in co-cultured schistosomules) (Table 4.5). All three genes were investigated as follows.

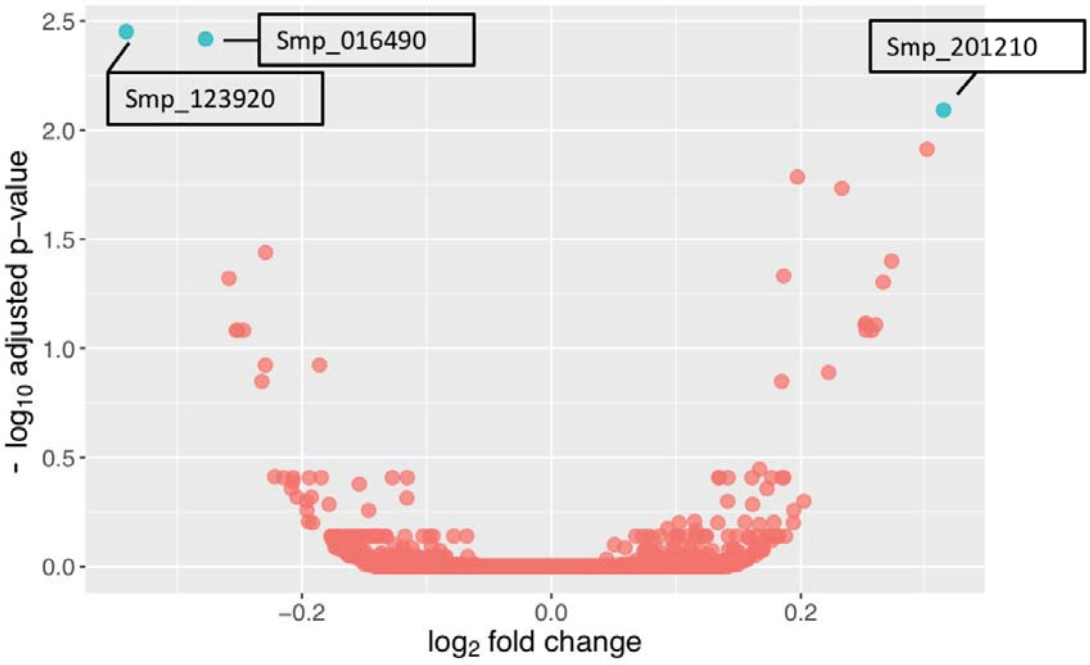

Figure 4.20 Volcano plot of pairwise comparison between HUVEC* vs. *non-HUVEC schistosomules

Volcano plots of genes from pairwise comparison of HUVEC and non-HUVEC schistosomules. Each dot is a gene. In blue are genes that pass adjusted p-value cut-off at 0.01. No log_2FC cut-off was used for HUVEC analysis due to small effect size.

Table 4.5 Genes differentially expressed in HUVEC compared to non-HUVEC schistosomules

Gene identifier	Log_2FC (HUVEC/non-HUVEC)	Adjusted p-value	Product name
Smp_123920	-0.34	3.54E-03	hypothetical protein
Smp_016490	-0.28	3.82E-03	Saposin B domain containing protein
Smp_201210	0.32	8.09E-03	hypothetical protein

4.3.8.2 Smp_123920, hypothetical protein

The gene with the biggest fold change was a *hypothetical protein* (Smp_123920) which was down-regulated by 1.27 fold. InterProScan identified no signature domain, but it reported presence of a signal peptide (based on Phobius tool (Käll *et al*., 2004)) and a transmembrane helix (based on TMHMM (Krogh *et al*., 2001) (Figure 4.21A). However, a separate signal peptide search using SignalP v4.1, which was designed to distinguish signal peptides from transmembrane domains, predicted no signal peptide

(Petersen *et al*., 2011). Based on sequence homology, the gene is limited to *Schistosoma* species suggesting its potential roles in parasitism (Figure 4.21B). Gene model of the Smp_123920 is supported by RNA-seq mapping evidence, which inferred reliability of the encoded amino acid sequences. To investigate this gene further, I used the I-TASSER server which predicts protein functions by measuring structural similarity (Yang *et al*., 2015). The tool computes protein structures based on their amino acid sequences and aligns the predicted structure to all entries in the PDB (Berman *et al*., 2000).

The top three structural matches were related to Notch complex-signalling (TM-score between 0.791 - 0.701; a TM score range between 0-1 where 1 indicate perfect match). However, the predicted structure of the query protein (Smp_123920) matched only partially to the top three hits (Figure 4.21C). Therefore, it is not certain whether the query *S. mansoni* hypothetical protein would function as the Notch complex or Notch ligands. One of the top structural match (4cc0) has been curated on CathDB (Sillitoe *et al*., 2015) allowing further exploration. The region that matched to Smp_123920 predicted structure contained immunoglobulin-like domain. To further confirm this, the Smp_123920 structure predicted from the I-TASSER was used to perform a structural search against the CathDB database (v4.1) (Sillitoe *et al*., 2015). The top three hits were to immunoglobulin-like domains: *1fnfA03* domain (SSAP score 78.8), *1wfjA01* domain (SSAP score 78.2), and *2nsmA02* domain (SSAP score 78); a SSAP score range between 0-100 with 80-100 being a highly similar match. Immunoglobulin folds (IPR007110, IPR013783) are used across a broad range of proteins, with immunological roles as well as cell-cell recognition, cell surface receptor signalling, and cell adhesion (Artero *et al*., 2001; Finn *et al*., 2017).

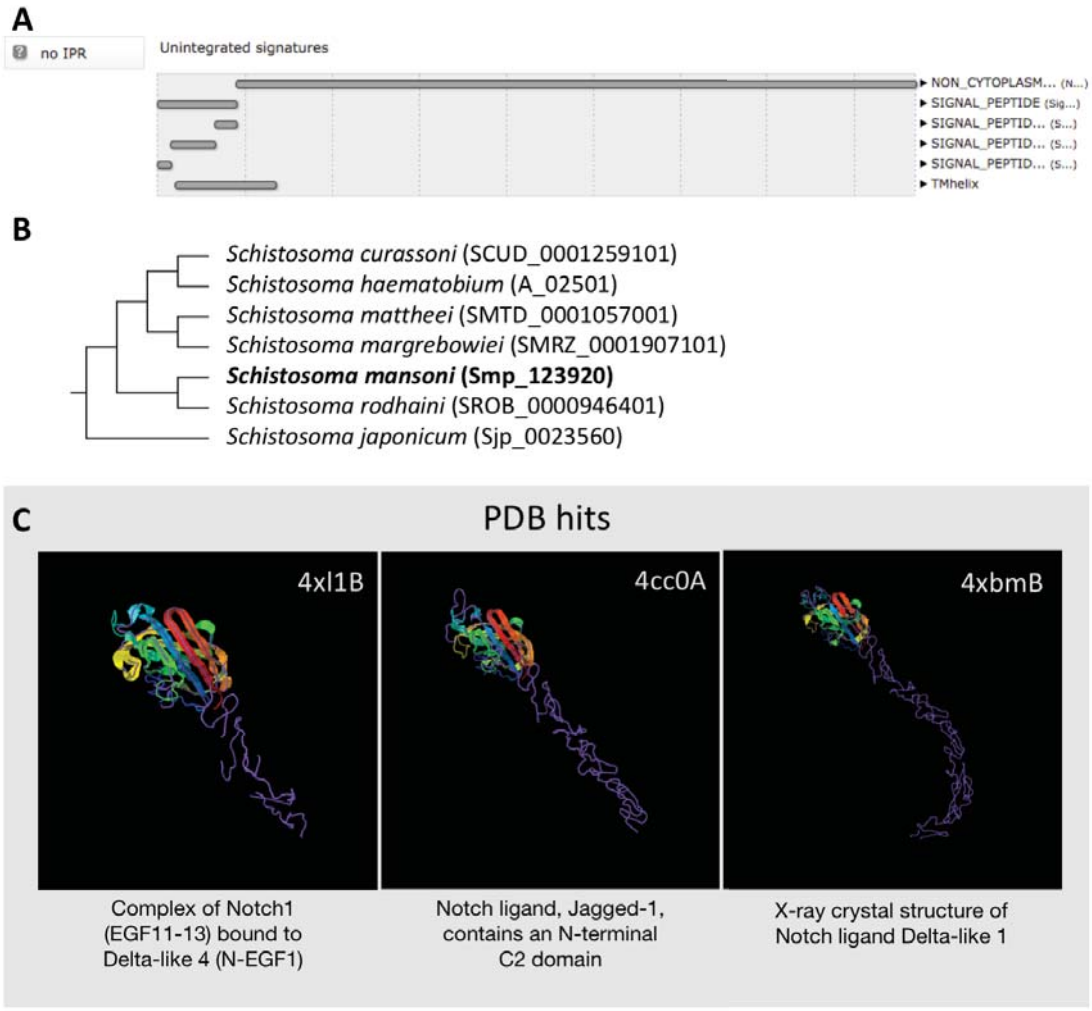

Figure 4.21 Smp_123920 homologous relationship and structural prediction

A) Domains identified by InterProScan based on amino acid sequence of the gene Smp_123920 (*hypothetical protein*). B) Gene tree of Smp_123920 and its homologues. The information of the tree was obtained from WormBase ParaSite release 8 (Howe et al., 2016b) and the tree was regenerated for clarity. Gene identifiers from the WormBase ParaSite are shown in the parentheses after the species names. C) I-TASSER structural alignment of predicted 3D strucuture of the gene Smp_123920 and the top three matches from PDB. In colour are predicted structure of the query sequence; in purple are traces of structures of the matched proteins. Identifiers in each box and description underneath are from PDB.

From the *in vivo* timecourse data from chapter 3, Smp_123920 was highly expressed in the lung stage but expression reduced to zero in adult stages (Figure 4.22). This is consistent with previous gene expression data (Protasio *et al.*, 2012) where high expression was seen in 24hr schistosomules. Expression *in vitro* also declined over time (Figure 4.22). It is possible that this gene is involved in interactions with the host

immune system, or in developmental processes where cell-cell interactions have major roles but given the broad prevalence of immunoglobulin domains, further experimentation would be required to narrow down the range of functional possibilities.

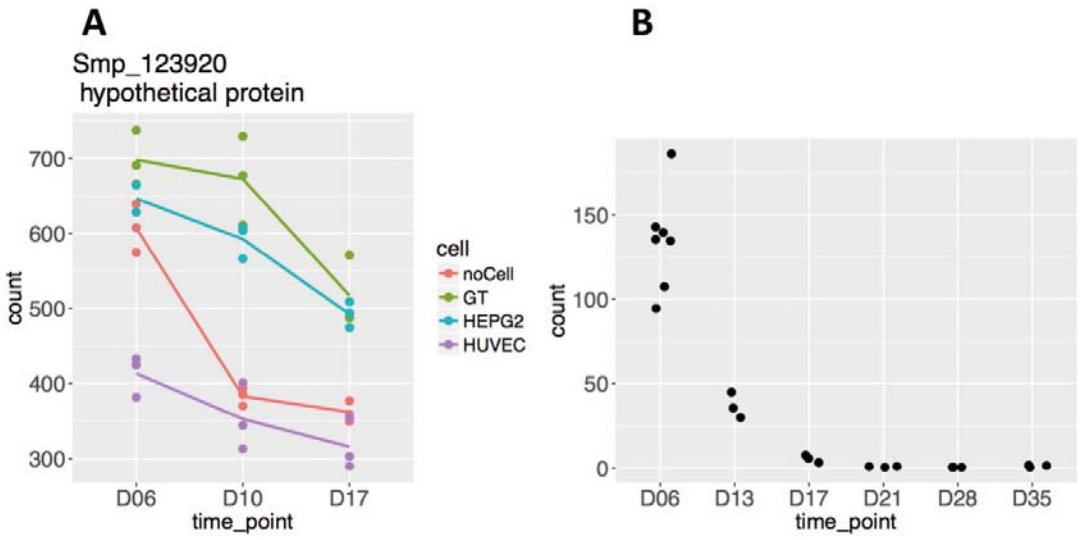

***Figure 4.22 Expression profiles of Smp_123920* in vivo *and* in vitro**

A) *In vitro* expression profiles. Colour represent co-culture conditions. Each dot is a biological replicate. B) *In vivo* expression profiles from the dataset in chapter 3. Each dot represent one biological replicate (a pool of *S. mansoni* from one infected mouse). For both A and B, y-axes indicate normalised read counts.

4.3.8.3 Smp_016490, saposin B domain-containing protein

A gene encoding saposin-like protein (Smp_016490) was down-regulated by 1.21 fold in schistosomules co-cultured with HUVEC (Table 4.5, Figure 4.23B). The encoded protein contained *saposin B type* (IPR008139) and *saposin-like* (IPR011001) domains (Figure 4.23A). This gene has not been functionally and molecularly characterised. However, saposins are lipid-interacting proteins (Bruhn, 2005) and the saposin-like gene in *S. mansoni* may have related functions. Four other saposins were previously identified in the gut vomitus from *S. mansoni* (Hall *et al*., 2011), but the down-regulated Smp_016490 was not amongst the list and its expression site is currently unknown. However, using the *in vivo* transcriptome dataset from the chapter 3, its expression pattern over the timecourse was similar to three out of the four gut vomitus saposins (Smp_105450, Smp_014570, Smp_130100); the highest expression

level was at day 35 (Figure 4.23C). With the similarity in expression pattern and the presence of relevant protein domains, the HUVEC-responsive saposin-like gene (Smp_016490) may encode genes with lipid-interacting roles similar to other saposins although it may not be involved in gut functions.

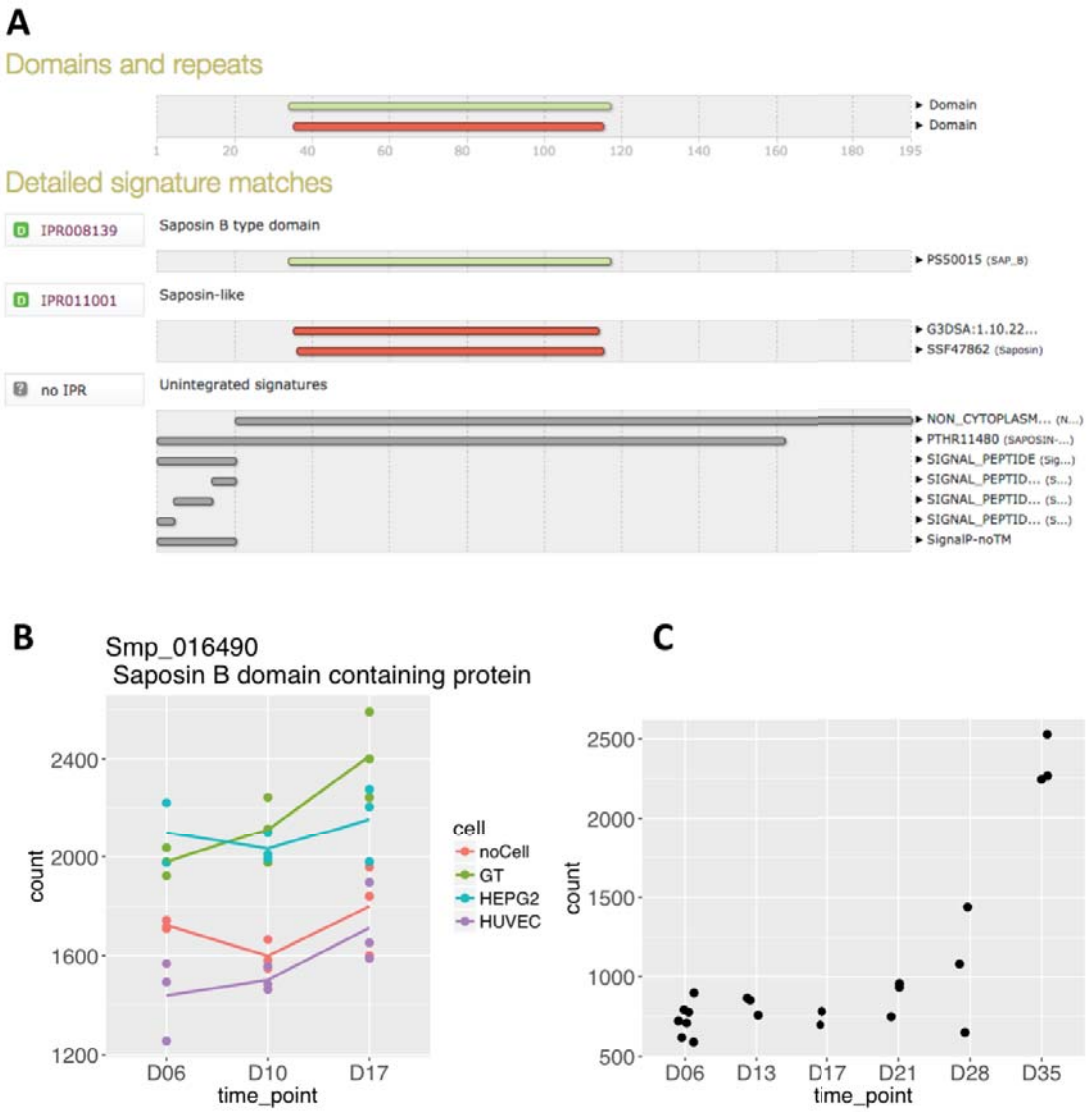

Figure 4.23 Smp_016490 domains and expression profiles

A) Domains identified by InterProScan based on amino acid sequence of Smp_016490. B) *In vitro* expression profiles. Colour represent co-culture conditions. Each dot is a biological replicate. C) *In vivo* expression profiles from the dataset in chapter 3. Each dot represent one biological replicate (a pool of *S. mansoni* from one infected mouse). For both B and C, y-axes represent normalised read counts.

4.3.8.4 Smp_201210, hypothetical protein and the only up-regulated genes in HUVEC schistosomules

The only up-regulated genes in HUVEC vs. non-HUVEC comparison was Smp_201210 (1.24 fold) (Table 4.5). The gene has been annotated as *hypothetical protein*, and InterProScan as well as SignalP v4.1 predicted signal peptides but no other signature domains to classify the gene into a protein family. However, the locus for this gene is flanked by gaps and the gene model could be changed in a future genome version (data not shown). Based on current information, protein structural prediction and alignment to 3D structure database (using I-TASSER server) did not yield any convincing matches. The top structural match was to pollen allergen but the score was low (TM-score 0.675, RMSD score 2.58) and other match with similar scores were not related to allergen protein. Interestingly, the gene has only one homologue in *S. rodhaini* and not in any other species (Appendix E), but this could be a consequence of gaps flanking the gene model, affecting its sequence.

4.3.8.5 Summary

In summary, schistosomules co-cultured with HUVEC were hardly different transcriptomically from control schistosomules. Only three genes were differentially expressed. Biological functions of the genes require further investigation.

4.4 Discussion

4.4.1 Overview

S. mansoni lives in close contact with host cells during infections and this presents possibilities of interactions. To gain a better understanding on the effects of host environments on the parasite, I conducted a co-culture experiment with schistosomules co-cultured with one of the three cell types (HEPG2, HUVEC, and GripTite), or no underlying cells as a control. The co-culture spanned timecourses of 6, 10, or 17 days. Comparing transcriptomic profiles of the parasites, it appeared that both cell conditions and time points affected the gene expression profiles. Control schistosomules were affected considerably by time, particularly at day 17. The realisation of this effect guided the rest of the analyses in this chapter. Considering generic responses, few genes were affected by three co-cultured cells, owing to small effect in schistosomules co-cultured with HUVEC. In contrast, HEPG2 condition

induced the highest number of differentially expressed genes. Genes affected by the co-culture may have roles in interacting or responding to host environment, and would benefit from further investigation and refined experiments.

4.4.2 *In vitro* schistosomules and effect of culture methods

From the morphologies and the transcriptomic profiles, it is clear that the schistosomules in this *in vitro* experiment did not develop as they would in an *in vivo* infection. This observation has been reported previously elsewhere (e.g. Basch, 1981; Clegg, 1965) and it signifies an important limitation of deriving biological conclusions from an *in vitro* experiments. In addition, as shown in this chapter, genes that were up-regulated in HUVEC schistosomules did not all have the same expression profile *in vivo*. This suggests that there are other environmental cues, and possibly interplay with the presence of host tissues, that can affect their gene expression. *In vivo* approaches, or an *in vitro* environment that can support better parasite growth, would help solidify functional interpretation of the differentially expressed genes affected by the co-culture.

Regarding effects of the time points, control schistosomules at day 17 became hugely different from the controls for day 6 and day 10. In contrast, the transcriptomes of the co-cultured schistosomules, although shifted with time, did not display such drastic shift as the day-17 control. As a result, a large number of genes were differentially expressed when each of the three co-cultured schistosomules were compared to control schistosomules at day 17 time point. While common up-regulated genes at day 17 seemed to suggest increase in signalling and responding to environment, the common down-regulated genes inferred that many core cellular processes were impeded.

At least three scenerios might explain the deviation of day-17 controls from other samples. Possibly, day-17 control schistosomules had adapted to *in vitro* environment in absence of mammalian cells whereas control schistosomules in other time point had not. Alternatively, a critical molecule for schistosomules may have become depleted or accumulated over time, and host cells might provide this factor or metabolise the accumulated molecules for co-cultured schistosomules. On the other hand, the use of non-heat-inactivated serum might have negative impact on

schistosomules. The complement components in the serum may still be active and could bind and accumulate on schistosomules over time. Such binding may affect biological processes leading to observed changes in gene expression. In addition, complement components in the serum, coming from a bovine origin, might bind to human erythrocytes in the culture media, causing agglutination of erythrocytes. The agglutination of erythrocytes would prevent ingestion by the schistosomules. This might also explain limited increase in size of schistosomules and limited accumulation of hemozoin over the timecourse. Others investigators have shown more advanced development in schistosomules at day 14 when similar media, but with heat-inactivated serum, were used (Mann *et al.*, 2010). If the complement components were the main cause, the day-17 control schistosomules would be affected more substantially than the co-cultured counterparts. This is because, in the co-culture set up, the complement components would also bind to the underlying mammalian cells, diluting available complement components that could bind to erythrocytes and schistosomules. An additional factor regarding media changing method may also affect changes in gene expression over time. This, however, is unlikely to be the cause of the shift in gene expression in day-17 control. When the schistosomules were transferred to a new plate during the experiment, half of the old media was transferred with the schistosomules and an equal volume of fresh media added to the new plates. This means that the media as a whole get older over time and may accumulate waste products, or lessen required nutrients. With these concerns, interpreting changes between time points should be avoided, and future experiment of similar timecourse could be refined by using heat-inactivated serum and by completely replace the old media.

With the possible confounding effect of time points, analyses that followed took the issue into account. Instead of using changes between time points to intepret biological meaning, the consistent changes unaffected by time became indicators for real effect of co-cultured human cells. Genes identified as generic responses included genes whose differences between co-cultured and control schistosomules were not affected by time (i.e. they were differentially expressed in all three time points). And to investigate cell-specific effect, samples were grouped by culture conditions regardless of their time points. This way, genes with large variation between time points would likely fail adjusted p-value cut-off for calling differentially expressed genes.

4.4.3 Generic responses

Generic responses to mammalian cell co-culture were up-regulation of two *hypothetical proteins* (Smp_052880 and Smp_067800). Smp_067800 could encode a non-coding RNA but the gene model is potentially inaccurate; its close proximity to a gap at its 5' end suggest a first exon could be missing. The other gene (Smp_052880) is a protein coding gene and contains a potential amidation site. Amidation is essential for activation of many neuropeptide by cleaving the peptide from its chain of precursors. Such amidation site was conserved in other *Schistosoma* and liver fluke species but at lower frequency than in free-living species. The encoded product of this gene, although annotated as hypothetical protein, has been identified as neuropeptide precursor-5 (npp-5) using EST (McVeigh *et al*., 2009) and further validated with genomic sequences (Collins *et al*., 2010). This is consistent with the BLASTP results of this gene with the match with *S. meditereania* being the npp-5 gene and the presence of an amidation site. The npp-5 gene in *S. mediterranea* is expressed in "cells surrounding the ventral midline" (Collins *et al*., 2010), though this may be different in *S. mansoni*. Functions of npp-5 protein in flatworm have not been identified, but its up-regulation in response to mammalian cells may suggest its role in adapting to host environment. *In vivo* expression were high in lung stage and adult stages, and one of the common features of the two stages is locomotion (Figure 4.24). Given that multiple processes in *S. mansoni* are regulated by neuropeptide signalling, such as behaviour, locomotion, reproduction, host invasion, and development (Ribeiro and Patocka, 2013), this potential neuropeptide could be interesting for follow up.

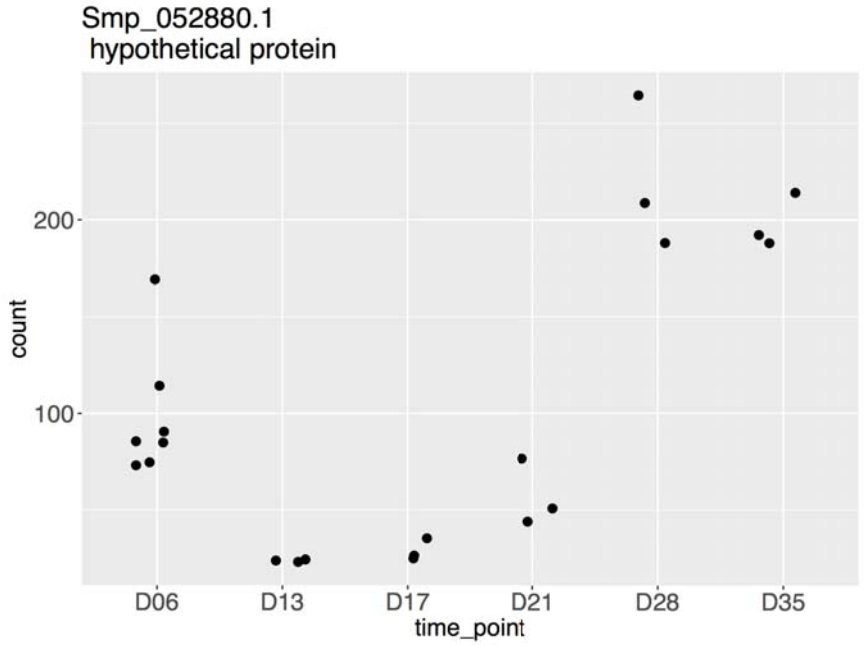

Figure 4.24 Smp_052880* in vivo *expression profile

In vivo timecourse expression profile of Smp_052880. The information came from transcriptome dataset covered in chapter 3. Y-axis indicate normalised read counts.

4.4.4 HEPG2-specific responses

Specific responses to HEPG2 co-culture condition included genes that may have roles in developmental control and in responding and interacting with environment. It is generally accepted that *S. mansoni* remains in the liver during its development into adult stages. Changes in expression of potential developmental regulators agree with the involvement of liver environment in developmental processes. The up-regulation of GPCR signalling pathway genes may suggest liver environment triggering downstream effect in molecular processes of the parasite. Further, *hypoxanthine guanine phosphoribosyltransferase* is involved in *purine salvage pathway*. *S. mansoni* is not able to synthesise its own purine nucleotide which it requires for development and gene expression (Dovey *et al*., 1984). Given that liver is one of the major sites for *de novo* purine biosynthesis (Angstadt, 1997), the up-regulation of this gene in HEPG2 schistosomules reflects environmental responses and a possible role of liver environment in supporting *S. mansoni* in their growth phase. The schistosomules in this experiment, however, were transcriptomically similar to *in vivo* lung stage rather than developing liver stage. To further investigate possible roles of liver environment

in development, similar co-culture set up could be done with schistosomules from *ex vivo* liver stage (e.g. day-13 schistosomules).

Genes involved in oxidation-reduction processes were up-regulated in HEPG2 schistosomules. Up-regulation of *ferritins* and *peroxiredoxin* suggest strategies for balancing oxidative stress in liver. Alternatively, *ferritin* up-regulation could be an adaptation to HEPG2 environment, ensuring that the parasite retains enough iron for their development and cellular process. HEPG2, having derived from liver hepatocytes, can store iron, and this would reduce available iron in the culture media, which the parasite may counteract by increasing ferritin expression. Level of ferritin in model organisms is regulated in response to iron levels in the environment (Schüssler *et al*., 1996). The mechanism for such regulation, however, is absent in *S. mansoni* (Schüssler *et al*., 1996) and a regulator of *ferritin* expression in *S. mansoni* has not yet been found. As a result, it cannot be certain whether the up-regulation of *ferritins* were in response to the reduction in available iron, or in response to the presence of HEPG2 cells. The up-regulation of genes encoding ferritins may also be a preparation to continue development because iron is a critical factor for *S. mansoni* development (Clemens and Basch, 1989; Glanfield et al., 2007). Furthermore, this emphasises the importance of oxidation-reduction control in *S. mansoni* (Simeonov *et al*., 2008).

4.4.5 HUVEC-specific responses

In contrast to HEPG2, HUVEC co-culture condition induced changes in only a small number of genes. This may be because schistosomules during intramammalian infection are in host blood vessels; therefore, endothelial cells are their generic environment. Additional signals from other tissues maybe required to generate site-specific responses. However, liver endothelial cells, as well as other cell types, do induce changes in schistosomules in term of size and gene expression (Ye *et al*., 2012), suggesting that endothelial cells from different source may lead to different effects. Alternatively, a small number of differentially expressed genes may be caused by missed contact between the schistosomules and the cells. Compared to other cell types used in this experiment, HUVEC only grew to 70-80% confluence; therefore, they did not cover the whole surface where schistosomules could contact. In contrast, both HEPG2 and GripTite grew to 100% confluence at the point of sample collection.

One of the *hypothetical proteins* up-regulated in HUVEC schistosomules did not match convincingly to any known protein structure. This is understandable as the protein product was small and the predicted structure consisted of few short helix loops - a feature that could match to many proteins. Another *hypothetical protein* was down-regulated, and had structural domain similarity to an immunoglobulin domain. Proteins containing immunoglobulin domains function in a range of biological processes with the majority being immunoglobulin molecules and others involved in interactions such as in receptor binding (Artero *et al*., 2001). Lastly, a gene encoding saposin-like protein was down-regulated in HUVEC schistosomules. Although *S. mansoni* saposins are thought to have roles in lipid binding for uptake in the gut (Hall *et al*., 2011), the functions and localisations of this gene is currently unknown. Nevertheless, information from other saposins in *S. mansoni* suggests that saposins may function at host-parasite interfaces. Other saposins were found in the proteome of gut vomitus and in secreted extracellular vesicles of adults *S. mansoni* (Figueiredo *et al*., 2015; Hall *et al*., 2011; Sotillo *et al*., 2016). One of the saposin found in gut vomitome (Sm-SLP-1, Smp_105450) is immunogenic in mice (Don *et al*., 2008). None of these however, are paralogues to the down-regulated saposin-like protein (Smp_016490.2). In fact, the gene has no detectable paralogues, nor do the two other saposins previously identified in vomitome (Hall *et al*., 2011). Some of the saposins appear to have rapidly evolved and a role in lysing ingested red blood cells has been proposed (Philippsen *et al*., 2015). Given the role of other saposin related genes in *S. mansoni*, this saposin-like protein may serve a similar function. However, it is not clear how its down-regulation in response to HUVEC environment could provide fitness advantages, similarly for the down-regulation of the *hypothetical protein* with immunoglobulin domain (Smp_123920). Comparing the *in vitro* expression in other cell conditions, it appeared that both genes also increased in their expression in HEPG2 and GripTite schistosomules (but did not pass the $\log_2$FC or adjusted p-value cut-offs). This suggested that their expression may require additional environmental cues which was not present in the HUVEC environment.

4.4.6 Summary

In this chapter, I have shown that *in vitro* schistosomule transcriptomes were affected by cell types in co-culture, and by time in culture. The change with time may be confounded by multiple factors, but this has been minimised by the chosen

approaches for the analysis. Using this dataset, I have shown that multiple processes of relevance to *S. mansoni* infection were affected by co-cultured cells, and that expression changes in some genes may suggest their roles in host interactions, but further investigation particularly for their roles *in vivo* would be required.

In the next chapter, the mammalian cells from the co-cultures in this chapter were subjected to RNA-seq and transcriptomic analysis. The transcriptomes were analysed to find out how the presence of schistosomules may influence gene expression of the host cells *in vitro*. Together, this could provide better understanding of host-parasite interactions, and may lead to topics for further investigation in the *in vivo* system.