Transcriptomic studies on host-parasite interactions in *Schistosoma mansoni* intramammalian stages

University of Cambridge
Pembroke College

A thesis submitted for the degree of

*Doctor of Philosophy*

Arporn Wangwiwatsin

The Wellcome Trust Sanger Institute
March 2017
Declaration

The work presented in this thesis was carried out at the Wellcome Trust Sanger Institute (Hinxton) between October 2012 and March 2017. This dissertation is the result of my own work – contributions from collaborations are clearly referenced. No part of this dissertation has been or is being submitted for any qualification in any other university. This thesis does not exceed the word limit established by the Biology Degree Committee.
Acknowledgements

First of all I would like to pass a massive gratitude to my supervisor, Matt, for giving me this opportunity, and for the understanding, patience, guidance, and support throughout. Thanks to everyone in the Parasite Genomics team including past and current members. Thanks for all the ideas, the fruitful discussions, and all the fun in team outing. Many people also helped read this thesis: Matt, Gabriel, Patrick, Adam, Kate, Anna, Avril, and Shona. Thank you to my thesis examiners for their contributions to the improvement of this thesis. Thank you Christina and Annabel from the Graduate Office for all their help with all questions I can ever ask. Thank you to Marc who has made every printing request possible.

Many people have been amazingly supportive and helpful throughout my work at the Sanger Institute, sharing their knowledge, helping make my experiment endeavour come true and many have become good friends. Thanks CGaP, Sally Forrest, Christ Kirton, Emma Goss, Theresa Feltwell, Liam Prestwood, Laura Wood, James Hewinson, Gavin Wright, Paul Kellam, Mercedes Pardo, Jyoti Choudhari, Lira Mamanova, Rachael Wash, Sarah Smith, Sally Linsdell, Derek Pickard, Dave Goulding Simon Clare, Cordelia Brandt, Colin Barker, Carol Smee, Dan Mead, Theodoros Roumeliotis (especially for his help with using the IPA), and the Sanger librarians.

Thanks for all support and joy beyond the work at Sanger Institute including Pembroke College, Public engagement teams at the Sanger and Cambridge, the CHaOS Society, Cambridge Thai Society, University Counselling Service, Polly Brown, family, and the husband.

It has been a wonderful experience.
Summary

The life cycle of the parasitic flatworm *Schistosoma mansoni* is split between snail and human hosts. In humans, it lives in the bloodstream and can survive for over 10 years, during which time, interactions with the host are essential for its survival. Starting the infection, cercariae penetrate human skin and become schistosomules which enter blood vessels. The schistosomules follow blood circulation to the lung and the liver where they develop into adults which migrate to the mesenteric venules for egg-laying. The association with certain host tissues and egg-laying site may involve interactions and adaptations. Moreover, the parasite needs to evade or modulate host immune responses and to optimise its acquisition of host metabolites. Molecular mechanisms involved in these interactions are not completely understood.

To gain deeper understanding into the interactions of *S. mansoni* and its mammalian host, I sought to understand what biological processes are required for successful intramammalian infection; specifically, what guides the tissue tropism of the parasite and what roles do the host tissues play in the infection? In the first part, parasite transcriptomes were produced from *S. mansoni* obtained from experimentally infected mice. The dataset includes novel transcriptomic profile of the lung stage, and covers developmental and egg-laying stages. In the next two parts, co-culture experiments were set up using mechanically transformed schistosomules and cells derived from human tissues. Transcriptomic profiles of the co-cultured schistosomules were investigated in the second part of this thesis; and transcriptomic profiles of the co-cultured human cells were investigated in the final part. The outputs of this thesis provide new insights into the infection biology, provide a large data resource for the research community, and propose avenues for further investigation and characterisation of interaction mechanisms.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>C1</td>
<td>Complement component 1</td>
</tr>
<tr>
<td>CFH</td>
<td>Complement factor H</td>
</tr>
<tr>
<td>CFI</td>
<td>Complement factor I</td>
</tr>
<tr>
<td>DAF</td>
<td>Decay-accelerating factor</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s Modified Eagle’s Medium</td>
</tr>
<tr>
<td>DMT</td>
<td>Divalent metal transporter</td>
</tr>
<tr>
<td>ES</td>
<td>Excretory/secretory products</td>
</tr>
<tr>
<td>ESTs</td>
<td>Express Sequence Tags</td>
</tr>
<tr>
<td>FPKM</td>
<td>Fragments Per Kilobase of transcript per Million mapped reads</td>
</tr>
<tr>
<td>GO</td>
<td>Gene Ontology</td>
</tr>
<tr>
<td>GPCR</td>
<td>G-protein coupled receptor</td>
</tr>
<tr>
<td>ICAM1</td>
<td>Intercellular adhesion molecule 1</td>
</tr>
<tr>
<td>IL-2</td>
<td>Interleukin-2</td>
</tr>
<tr>
<td>IPA</td>
<td>Ingenuity Pathways Analysis</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto Encyclopedia of Genes and Genomes</td>
</tr>
<tr>
<td>LDL</td>
<td>Low density lipoprotein</td>
</tr>
<tr>
<td>Log₂FC</td>
<td>Log₂ fold change</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>MEG</td>
<td>Micro-exon gene</td>
</tr>
<tr>
<td>NPC2</td>
<td>Niemann Pick type C2 protein</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>RNA-seq</td>
<td>RNA-sequencing</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen species</td>
</tr>
<tr>
<td>SELE</td>
<td>Selectin E</td>
</tr>
<tr>
<td>TAL</td>
<td>Tegumental-allergen-like</td>
</tr>
<tr>
<td>TGF-β</td>
<td>Transforming growth factor-beta</td>
</tr>
<tr>
<td>TNF-α</td>
<td>Tumour-necrosis factor-alpha</td>
</tr>
<tr>
<td>VCAM1</td>
<td>Vascular cell adhesion molecule</td>
</tr>
<tr>
<td>WTSI</td>
<td>Wellcome Trust Sanger Institute</td>
</tr>
</tbody>
</table>
# Table of Contents

## Chapter 1

**Introduction** ................................................................. 1

1.1 Overview............................................................................................ 1

1.2 Introduction to schistosomes ............................................................. 2

1.2.1 Schistosome life cycle ........................................................................ 2

1.2.2 Schistosomiasis ................................................................................. 5

1.3 *S. mansoni* life in bloodstream .......................................................... 7

1.3.1 Host-parasite interactions: the interfaces ........................................... 7

1.3.2 Motility, behaviour, and homeostasis: the nervous system ............... 11

1.3.3 Migration route and development ..................................................... 12

1.3.4 Metabolic requirements and acquisition ......................................... 17

1.3.5 Immune evasion ................................................................................. 22

1.3.6 Section summary ............................................................................... 30

1.4 *S. mansoni* genome and transcriptome ............................................ 30

1.4.1 Genome and gene annotation ........................................................ 30

1.4.2 Transcriptomes .................................................................................. 33

1.5 This thesis ......................................................................................... 34

## Chapter 2

**Materials and methods** .................................................. 35

2.1 Parasite materials ........................................................................... 35

2.1.1 Maintenance of the *S. mansoni* life cycle ........................................ 35

2.1.2 Snail husbandry ................................................................................ 36

2.1.3 Mouse maintenance ......................................................................... 37

2.2 Molecular methods ........................................................................... 37
### 2.2.1 RNA extraction from parasites

- RNA extraction from parasites

### 2.2.2 RNA extraction from human cells

- RNA extraction from human cells

### 2.2.3 Measurements of RNA concentration and purity

- Measurements of RNA concentration and purity

### 2.2.4 Library production and sequencing

- Library production and sequencing

### 2.2.5 Overall QC of sequencing outcome

- Overall QC of sequencing outcome

### 2.3 Data analysis

- Data analysis

#### 2.3.1 Mapping and quantifying read counts

- Mapping and quantifying read counts

#### 2.3.2 Read count and differential expression analysis

- Read count and differential expression analysis

#### 2.3.3 Genes clustering by timecourse expression profile

- Genes clustering by timecourse expression profile

#### 2.3.4 GO term enrichment

- GO term enrichment

#### 2.3.5 Pathway enrichment and pathway network

- Pathway enrichment and pathway network

#### 2.3.6 Pathway comparison between cell types

- Pathway comparison between cell types

#### 2.3.7 Protein structural prediction

- Protein structural prediction

#### 2.3.8 Protein domains and motif search

- Protein domains and motif search

#### 2.3.9 Gene phylogenetic tree

- Gene phylogenetic tree

#### 2.3.10 Artemis and BamView

- Artemis and BamView

#### 2.3.11 Seaview sequence alignment

- Seaview sequence alignment

### 2.4 Data presentation

- Data presentation

### Chapter 3

#### S. mansoni in experimentally infected mice

- S. mansoni in experimentally infected mice

#### 3.1 Introduction

- Introduction

##### 3.1.1 Overview

- Overview

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

- Host-parasite interactions in intra-mammalian stages

##### 3.1.3 Progress in transcriptomic and genomic approaches

- Progress in transcriptomic and genomic approaches

##### 3.1.4 Aims and approaches

- Aims and approaches

##### 3.1.5 Chapter outline

- Chapter outline

#### 3.2 Methods

- Methods

##### 3.2.1 Mouse infection

- Mouse infection

##### 3.2.2 Parasite and tissue collection from infected mice

- Parasite and tissue collection from infected mice
3.2.3 RNA extraction and library preparation ........................................... 54
3.2.4 Morphology score ............................................................................. 54

3.3 Results .................................................................................................. 55
3.3.1 Worm morphology ........................................................................... 55
3.3.2 RNA quantity and quality ................................................................. 58
3.3.3 Sequencing yields ............................................................................. 58
3.3.4 Sequencing data overall profiles ......................................................... 60
3.3.5 Lung stage ......................................................................................... 64
3.3.6 Liver stages ....................................................................................... 76
3.3.7 Adult stages ..................................................................................... 83

3.4 Discussion .......................................................................................... 94
3.4.1 Overview .......................................................................................... 94
3.4.2 Potential effect of collection procedures ............................................. 94
3.4.3 Lung stage signalling ......................................................................... 95
3.4.4 Lung stage immune evasion ............................................................... 96
3.4.5 Complement factor H ....................................................................... 96
3.4.6 Complement factor I ......................................................................... 97
3.4.7 Immunomodulation in liver stages ...................................................... 97
3.4.8 Implication for intervention ............................................................... 98
3.4.9 Expected changes in liver and adult stages ......................................... 98
3.4.10 Liver localisation ............................................................................. 98
3.4.11 Mesenteric migration ...................................................................... 99
3.4.12 Micro-exon genes .......................................................................... 101
3.4.13 Other ways to use this data ............................................................. 101
3.4.14 Summary ....................................................................................... 101

Chapter 4

S. mansoni in vitro culture with cell lines ......................... 103

4.1 Introduction .......................................................................................... 103
4.1.1 Overview .......................................................................................... 103
Chapter 5

Transcriptomes of cell lines exposed to schistosomules in vitro

5.1 Introduction

5.1.1 Overview

5.1.2 Host responses in S. mansoni infections

5.1.3 Host tissue responding to S. mansoni

5.1.4 Aims and approaches

5.1.5 Outline

5.2 Methods

5.3 Results

5.3.1 RNA quantity and quality

5.3.2 HUVEC and endothelial cell surface markers

5.3.3 HEPG2 and liver hepatocytes

5.3.4 Overall profiles of transcriptomes

5.3.5 Differential expression between co-cultured vs. worm-free cells

5.3.6 Pathways between cell types

5.4 Discussion

5.4.1 Overview

5.4.2 On in vitro-adapted cells

5.4.3 On aged media

5.4.4 Biological functions affected by parasite co-culture

5.4.5 On in vivo validation

5.4.6 Summary

Chapter 6

Discussion and conclusion

6.1 Findings and interpretations
6.1.1 Interactions with host defences .................................................................196
6.1.2 Host reparation ......................................................................................198
6.1.3 Responses to environments ...................................................................198
6.2 Limitations ...............................................................................................201
6.3 Concluding remarks ..............................................................................202

Chapter 7

References ...............................................................................................203

Chapter 8

Appendices .............................................................................................235
List of Figures

Figure 1.1 Life cycle of *Schistosoma* spp. ................................................................................. 3
Figure 1.2 The schistosome tegument ....................................................................................... 9
Figure 1.3 Coagulation and complement cascade........................................................................ 25
Figure 1.4 Coagulation pathway .............................................................................................. 28
Figure 2.1 Transfer of snail eggs using “The Octopus” ........................................................... 37
Figure 2.2 RNA qualities assessed from Agilent Bioanalyzer electropherograms .................. 39
Figure 3.1 Experiment layout .................................................................................................. 51
Figure 3.2 Diagram used for morphology scoring..................................................................... 55
Figure 3.3 Morphology of *in vivo* *S. mansoni* ................................................................... 57
Figure 3.4 PCA plots from total reads and down-sized reads.................................................. 59
Figure 3.5 Similarity between 96 clusters ................................................................................ 61
Figure 3.6 Clusters of genes based on timecourse expression pattern ..................................... 62
Figure 3.7 Clusters of genes based on timecourse expression pattern, with fixed y-axes ...... 63
Figure 3.8 Volcano plots of all consecutive pairwise comparison .......................................... 66
Figure 3.9 GO enrichment of genes up-regulated in day-6 compared to day-13 schistosomules .............................................................................................................................. 68
Figure 3.10 MEGs differentially expressed between day-6 and day-13 schistosomules ...... 69
Figure 3.11 Predicted structure of Smp_182770 aligned with structure of human CFH ...... 73
Figure 3.12 Homologous relationship of Smp_182770 and alignment of RNA-seq reads to genomic locations............................................................................................................. 75
Figure 3.13 Clusters of genes with high expression in liver stages and GO term enrichment 78
Figure 3.14 GO enrichment of genes up-regulated in day-13 compared to day-6 schistosomules .................................................................................................................................................. 81
Figure 3.15 GO enrichment of genes up-regulated in day-35 compared to day-28 worms.... 86
Figure 3.16 GO enrichment of genes up-regulated in day-28 compared to day-35 worms ..... 87
Figure 3.17 Expression profile of GPCRs up-regulated in day-28 compared to day-35 worms ................................................................................................................................................. 88
Figure 3.18 Log_{2}FC of genes in cluster 40 in response to pairing........................................... 93
Figure 4.1 Experiment set up ................................................................................................. 107
Figure 4.2 Gel image of PCR product from *Mycoplasma* tests ............................................ 114
Figure 4.3 Example images from each time point of each co-culture condition ................. 115
Figure 4.4 PCA of transcriptomes from *in vitro* and *in vivo* datasets .................................. 117
Figure 4.5 PCA of transcriptomes from all *in vitro* parasites and *in vivo* lung schistosomules ................................................................................................................................................. 118
Figure 4.6 PCA of transcriptomes from all in vitro schistosomules ........................................... 119
Figure 4.7 MA plots of pairwise comparisons between co-cultured and control schistosomules
.................................................................................................................................................... 120
Figure 4.8 Differentially expressed genes unique and common to each pairwise comparison
.................................................................................................................................................... 121
Figure 4.9 Differentially expressed genes in day-17 schistosomules comparing unique and
common changes between cell types ............................................................................................. 122
Figure 4.10 Enriched GO terms in genes up-regulated in day-17 co-cultured schistosomules
.................................................................................................................................................... 123
Figure 4.11 Enriched GO terms in genes down-regulated in day-17 co-cultured
schistosomules.................................................................................................................................. 124
Figure 4.12 Generic responses in schistosomules ........................................................................ 126
Figure 4.13 Homologous relationship of Smp_067800 ................................................................. 127
Figure 4.14 Genomic region of Smp_067800 and alignment of RNA-seq reads ....................... 129
Figure 4.15 Homologous relationship of Smp_052880 ............................................................... 131
Figure 4.16 Smp_052880 genomic region and amino acid sequence alignment with other
BLASTP hits .................................................................................................................................. 132
Figure 4.17 Volcano plot of pairwise comparison between HEPG2 vs. non-HEPG2
schistosomules............................................................................................................................... 134
Figure 4.18 Enriched GO terms in genes up-regulated in HEPG2 compared to non-HEPG2
schistosomules............................................................................................................................... 135
Figure 4.19 Enriched GO terms in genes down-regulated in HEPG2 compared to non-HEPG2
schistosomules............................................................................................................................... 135
Figure 4.20 Volcano plot of pairwise comparison between HUVEC vs. non-HUVEC
schistosomules............................................................................................................................... 141
Figure 4.21 Smp_123920 homologous relationship and structural prediction ......................... 143
Figure 4.22 Expression profiles of Smp_123920 in vivo and in vitro ......................................... 144
Figure 4.23 Smp_016490 domains and expression profiles ....................................................... 145
Figure 4.24 Smp_052880 in vivo expression profile .................................................................. 150
Figure 5.1 260/230 ratio of RNA extracted from human cells and its relationship with the
RNA concentration ......................................................................................................................... 159
Figure 5.2 Expression levels of endothelial cell surface markers in worm-free HUVEC .......... 160
Figure 5.3 PCA from transcriptomes of all cell types, and separate for each cell type ........... 162
Figure 5.4 Volcano plots for pairwise comparison in each cell type between co-cultured vs.
worm-free conditions .................................................................................................................... 164
Figure 5.5 Expression profiles of SELE, VCAM1, and ICAM1 .................................................. 165
Figure 5.6 Expression of endothelial adhesion molecules between co-cultured vs. worm-free HUVEC ......................................................................................................................... 166

Figure 5.7 GO enrichment of genes differentially expressed in co-cultured HUVEC compared to worm-free HUVEC ................................................................................................................. 168

Figure 5.8 Pathway enrichment of genes differentially expressed in co-cultured HUVEC compared to worm-free HUVEC ................................................................................................................. 169

Figure 5.9 Coagulation-related genes up-regulated in co-cultured HUVEC ......................................................... 170

Figure 5.10 GO enrichment of genes differentially expressed in co-cultured HEPG2 compared to worm-free HEPG2 ................................................................................................................. 174

Figure 5.11 Pathway enrichment of genes differentially expressed in co-cultured HEPG2 compared to worm-free HEPG2 ................................................................................................................. 175

Figure 5.12 GO enrichment of genes differentially expressed in co-cultured GripTite compared to worm-free GripTite ................................................................................................................. 179

Figure 5.13 Pathway enrichment of genes differentially expressed in co-cultured GripTite compared to worm-free GripTite ................................................................................................................. 180

Figure 5.14 Immune responses-related genes up-regulated in co-cultured GripTite ......................................................... 182

Figure 5.15 Log2FC of genes in extracellular matrix organisation pathways ......................................................... 184

Figure 5.16 Log2FC of genes in coagulation and complement cascade ......................................................... 186
List of Tables

Table 3.1 Average of total read counts ................................................................. 59
Table 3.2 Genes in cluster 8 (highly expressed in lung stage) ................................. 64
Table 3.3 Enriched GO terms of genes down-regulated after lung stage ............ 65
Table 3.4 Top 20 genes up-regulated in day-6 compared to day-13 schistosomules 67
Table 3.5 Genes in cluster 72 ............................................................................. 72
Table 3.6 Top 20 genes up-regulated in day-13 compared to day-6 schistosomules 80
Table 3.7 Enriched GO terms (biological process) of genes up-regulated in day-21 compared to day-28 worms ................................................................. 83
Table 3.8 Top 20 genes up-regulated in day-28 compared to day-21 worms .......... 84
Table 3.9 Enriched GO terms (biological process) of genes up-regulated in day-28 compared to day-21 worms ................................................................. 85
Table 3.10 Genes in cluster 40 .......................................................................... 90
Table 4.1 BLASTP output with Smp_067800 amino acid sequence as a query .... 127
Table 4.2 BLASTP output with Smp_052880 amino acid sequence as query .... 130
Table 4.3 Top 20 genes up-regulated in HEPG-2 compared to non-HEPG2 schistosomules 137
Table 4.4 Top 20 genes down-regulated in HEPG-2 compared to non-HEPG2 schistosomules ......................................................................................... 138
Table 4.5 Genes differentially expressed in HUVEC compared to non-HUVEC schistosomules ................................................................. 141
Table 5.1 Top differentially expressed genes in HEPG2 ...................................... 173
List of Appendices

Appendix A: Aquarium water, 10X ................................................................. 235
Appendix B: Reference genomes used by NPG QC ............................................. 236
Appendix C: Versions of R packages used in data analysis ............................... 238
Appendix D: Basch media components ............................................................. 239
Appendix E: S. mansoni genetrees downloaded from WormBase ParaSite release 9 240
Appendix F: Orthologues of Smp_041700 downloaded from WormBase ParaSite release 8 ......................................................... 251
Appendix G: List of endothelial cell surface marker from Durr et al., 2004 .......... 252
Appendix H: List of supplementary tables ......................................................... 255
Chapter 1

Introduction

1.1 Overview

The parasitic blood fluke *Schistosoma mansoni* has a complex and fascinating life cycle split between snail and human hosts. The parasites dwell in the bloodstream of the mammalian host, associated with particular niches during their early migration, development, and reproduction. Further, the parasites are in constant contact with host tissues at their site of residence, including the endothelial lining of blood vessels and circulating cells of the blood (Bloch, 1980; Crabtree and Wilson, 1986a). Within this environment, individual parasites can survive for more than a decade, demonstrating that they are well adapted to the mammalian host, having evolved strategies to evade host defence mechanisms (Kusel et al., 2007). Moreover, these are successful parasites infecting and reinfecting millions of people globally (Gryseels et al., 2006).

Previous work has shown that blood flukes employ multiple strategies to ensure their survival in the bloodstream (reviewed in Cai et al., 2016; Kusel et al., 2007; Wilson, 2012); however, further understanding of these mechanisms is an on-going process. This introductory chapter provides background, and sources of inspiration leading to the work in this thesis. It covers general background of blood flukes, in particular *S. mansoni*, including the disease (schistosomiasis) that it causes and the parasite life cycle. Aspects of the life of the parasites within the bloodstream are discussed, including interaction interfaces, migration paths during their development, acquisition and roles of molecules obtained from host environment, and evasion and modulation of defence mechanisms of their hosts. In addition, the use of genomics and transcriptomics in *S. mansoni* research is discussed in the context of host-parasite interactions. This thesis work aims to add insight into the molecular interactions between parasites and their hosts based on both *in vivo* and *in vitro* approaches using transcriptomes as the main tool of analysis.
1.2 Introduction to schistosomes

The three most common species that infect humans are *S. mansoni*, *Schistosoma japonicum*, and *Schistosoma haematobium*. The parasites belong to the class Trematoda and possess features associated with this group - complex life cycles requiring mollusc and mammalian hosts, flat bodies, and conspicuous suckers. However, they are unique from other members of the group in that members of this genus are dioecious with distinct sexual dimorphism genetically determined by Z and W chromosomes (females are heterogametic with a ZW karyotype), whereas other members of the group are hermaphroditic (Basch, 1991).

The parasites are the causative agent of a Neglected Tropical Disease (NTD) called schistosomiasis. Over 200 million people worldwide are infected with *Schistosoma* species (Gryseels et al., 2006). These are largely largely restricted to Africa, Middle East, South America, and Asia, due to the geographical distribution of the snail intermediate host, whose survival is largely determined by the climate (McCreesh et al., 2015). Recently however, cases of *S. haematobium* infections have been reported in Corsica (France) in individuals who had not been to an endemic area (Berry et al., 2014; Boissier et al., 2015). Parasitic infections are rarely fatal, instead the parasites produce chronic infections, surviving for a long time in humans and causing long-term morbidity (Gryseels et al., 2006).

1.2.1 Schistosome life cycle

The life cycle of *S. mansoni* is split between snail and mammalian host (Figure 1.1). Its natural mammalian hosts are humans and chimps, which is in contrast to *S. japonicum* where bovine and other mammalian hosts are also susceptible. Experimental models of *S. mansoni* are commonly laboratory mice, rats, and hamsters. Each schistosome species is carried by a specific snail species where the parasite asexually reproduces (Gryseels et al., 2006).
Figure 1.1 Life cycle of *Schistosoma* spp.

Life cycle of three main *Schistosoma* species that infect human. Diagram reproduced from Centers for Disease Control and Prevention (2012).

1.2.1.1 Intramolluscan stages

Snail hosts of *S. mansoni* belong to the group *Biomphalaria*, whereas snail hosts for *S. japonicum* and *S. haematobium* are *Oncomelania*, and *Bulinus*, respectively. This specificity to snail hosts limits the geographical distribution of each parasite species (Colley et al., 2014). Once eggs are deposited in water (from, for example, contamination of water with faecal waste from infected individuals), immediately miracidia hatch from eggs. The free-swimming miracidia are phototropic and propel themselves with cilia plates, following chemical cues, towards snails to penetrate (Haas et al., 1995). Once in the snail, a miracidium sheds its cilia plate and becomes a mother (primary) sporocyst; this starts to asexually reproduce in the snail. Daughter (secondary) sporocysts later bud off from the mother sporocyst and multiply into further daughter sporocysts. After four weeks, cercariae develop inside daughter sporocysts and leave the snail seeking a mammalian host (Basch, 1991). The shedding of cercariae from snails is stimulated by light and can follow circadian pattern of human behaviour (Lu et al., 2009).
1.2.1.2 Intramammalian stages

Upon leaving the snail hosts, cercariae swim upward following light and a thermal gradient, and are affected by chemical cues directing them to suitable mammalian hosts, all relying on cercariae energy reserves (Ghandour and Ibrahim, 1978). Using cues on mammalian skin, such as linoleic acid, cercariae are stimulated to release the content of acetabular glands which is at the head of the cercariae and contain proteases such as cercarial elastase, assisting the cercariae in penetration through the skin of their host (Haeberlein and Haas, 2008; Ingram et al., 2012; McKerrow and Salter, 2002). During penetration, cercariae lose their tails and become schistosomules. Skin excision (Miller and Wilson, 1978), radioisotope tracking (Georgi et al., 1982), and histological experiments (Wheater and Wilson, 1979) have shown that the parasites remain in the skin for around 3-4 days before finding and entering a blood vessel. Molecules from the host blood might serve as cues for finding blood vessels. An in vitro experiment where cercariae were stimulated to penetrate into agar gel (using linoleic acid) demonstrated that the resulting schistosomules navigated toward human serum, D-glucose, L-arginine, fibronectin, and bradykinin (Grabe and Haas, 2004; Haas et al., 2002).

Schistosomules enter blood vessels using secretions from head gland (Kusel et al., 2007). One of the main components secreted from the head gland is likely to be a protein product of micro-exon gene (MEG)-3 (DeMarco et al., 2010). Once in the bloodstream, schistosomules are swept by the circulation reaching multiple sites including the heart, the lung, and the liver where they develop into adults (Wilson, 2009). After the parasites develop into adult forms, in S. mansoni and S. japonicum, males and females pair and migrate toward mesenteric venules where egg-laying occurs (Wilson, 2009; Wilson et al., 2016). In contrast, pairs of adult S. haematobium migrate towards venous plexus for egg-laying. The pairing leads to changes in both the male and the female, but most spectacularly, the changes and development of the female reproductive organs and maturity only occur after pairing with a male (LoVerde and Chen, 1991). The size of adult worms varies depending on host type, but generally is around ~10mm in humans (Cheever, 1968) and ~7 mm in mice (Basch, 1981; Clegg, 1965a), with females being slightly (~1mm) longer than males (Basch, 1981; Cheever, 1968; Clegg, 1965). A pair of S. mansoni adult worms lays approximately 300 eggs per day and can live for over a decade (Chabasse et al., 1985;
Colley et al., 2014; LoVerde and Chen, 1991; Pellegrino and Coelho, 1978; Warren et al., 1974). Such a long life and continuous massive egg production would require ability to repair tissue damage, to compensate cellular aging, as well as to maintain proliferating germ cells (Collins et al., 2013; Wendt and Collins, 2016). The processes appear to involve a population of proliferative neoblast-like cells which differentiate into various somatic tissues such as tegemental cell body, intestinal lining cells, and muscle layers (Collins et al., 2013).

1.2.2 Schistosomiasis

1.2.2.1 Pathology

The infections consist of two phases: acute and chronic schistosomiasis. Acute schistosomiasis is known as Katayama syndrome and occurs before the egg-laying stage. This seems to be more common in individuals outside an endemic area (Colley et al., 2014). The symptoms can include fever, muscular pain, fatigue, diarrhoea and abdominal pain (Colley et al., 2014; Vale et al., 2017). The most frequent pathology of schistosomiasis results from chronic disease and is caused by eggs released from adult worms (Colley et al., 2014; Vale et al., 2017). For S. mansoni and S. japonicum, the eggs traverse the blood vessel wall, and enter into the lumen of the intestine to be excreted into the environment with faeces (Gryseels et al., 2006). However, about half of the eggs remain in the bloodstream and are carried with the bloodflow (Loeffler et al., 2002). Paired adult worms of S. mansoni and S. japonicum reside in the mesenteric veins, close to the intestine, and the bloodflow carries eggs through the portal vein to the liver, where the eggs can become trapped in liver tissue. This leads to liver inflammation, fibrosis and the formation of granulomas around the eggs. Over time, more eggs are deposited in the liver, disrupting blood flow across the liver, as well as normal liver function, eventually leading to other symptoms such as liver and spleen enlargement, or pulmonary hypertension (Colley et al., 2014; Vale et al., 2017). Adult S. haematobium dwell in the venous plexus draining the urogenital organs, and the eggs are normally released in urine. The traversing and lodging of eggs in the urogenital organs can lead to hematouria (bloody urine), cervical schistosomiasis, infertility, an increased risk of HIV transmission. Furthermore, the infections with S. haematobium are associated with squamous cell carcinoma of the bladder (Colley et al., 2014; Vale et al., 2017). Eggs can also lodge in other tissue,
such as in the lung (Cheever, 1968) and brain (Rose et al., 2014), disrupting the organ functions.

### 1.2.2.2 Immunology of schistosomiasis

During the early phase of infection, i.e. before eggs start to be laid, there is a low level inflammatory response with the production of cytokines that are characteristic of type-1 helper T cells (Th-1), such as tumour-necrosis factor alpha (TNF-α), interferon gamma, and interleukin-2 (IL-2) dominating (Hams et al., 2013; Pearce and MacDonald, 2002). However, a balanced Th-1/Th-2 response has also been suggested (Colley et al., 2014). Once eggs start to be laid, a type-2 T helper cell (Th-2) response dominates. The switch to Th-2, particularly the expression of IL-4, appears to be important for survival of the host by preventing inflammatory damage from eggs trapped in the internal organs (Colley et al., 2014; Hams et al., 2013; Pearce and MacDonald, 2002). The switch to a Th-2 response is induced by egg secretions and the main component of which is Omega-1 (Everts et al., 2009; Schramm et al., 2003). Th-2 cells can appear before egg-laying and this is thought to be a preventative mechanism to ensure a Th-2 polarisation at a later stage (de Oliveira Fraga et al., 2010a, 2010b). Alternatively, the early Th-2 responses may be beneficial for the parasite development, as a Th-2 cytokine IL-4 is involved in regulation of CD4+ immune cells that promote parasite development (Riner et al., 2013).

### 1.2.2.3 Diagnosis and treatment

The gold standard diagnostic analysis for schistosomiasis is still the identification of eggs in stool (for *S. mansoni* and *S. japonicum*) or urine (for *S. haematobium*) (Colley et al., 2014). Dominant secreted proteoglycan and other antigens from the parasites may also be used as diagnostic markers (Hamilton et al., 1998; Nash and Deelder, 1985). The severity of the pathology can be determined by imaging techniques such as ultrasound and CT-scans (Skelly, 2013). Currently, the main line of treatment is the use of a single drug, praziquantel, which is administered as a single dose, or in two doses separated by ~14 days to treat juvenile parasites which were not killed during the first dose (King et al., 2011). The drug is relatively inexpensive, with mild side effects, and generally efficacious (Cioli et al., 2014; Vale et al., 2017). However, cases of potential resistance or reduced responsiveness to drug treatment have also been reported (Crellen et al., 2016). Furthermore, the host immune system is slow to
develop resistance to reinfection which is common in endemic areas (Colley et al., 2014). Therefore, research is desperately needed to discover novel targets for drugs and vaccines.

1.2.2.4 Control and prevention
Current attempts to control schistosomiasis include mass drug administration programs (Secor, 2015). Behavioural changes, education, sanitation, infrastructure development, and control of the intermediate host are also important because the infection is caused by contact with contaminated water (Secor, 2014, 2015). Vaccine development is an on-going and challenging effort (Tebeje et al., 2016). The most successful ‘vaccine’ so far is composed of irradiated cercariae which leads to protection rates of over 80% in primate hosts (Bickle, 2009). However, such a method is not scalable because a large number of parasites and hosts are required to maintain the parasite production. Furthermore, introducing live cercariae to humans may not be socially acceptable. Proteins present on parasite tegument have been tested as potential vaccine targets. An example of these is tetraspanin-2 (TSP-2), which has entered phase I clinical trial (Cai et al., 2016; Tebeje et al., 2016).

1.3 S. mansoni life in bloodstream
This thesis concerns blood-dwelling intramammalian stages of S. mansoni, and particularly the interactions between the parasite and its host. In this section, the biology of S. mansoni intramammalian developmental stages is discussed in detail, aiming to describe key aspects of host-parasite interactions, including multiple mechanisms evolved by the parasite to interact (or interfere) with the host environment in order to thrive in the bloodstream.

1.3.1 Host-parasite interactions: the interfaces
While living in the bloodstream, the parasite interacts with the host environment to obtain nutrients, evades the host immune responses, migrates to seek a mate and find an egg-laying site. As S. mansoni is coated with a tegument and feeds on blood, the tegumental surface and lining of the gut form the interface with which the parasite interacts with its environment (Skelly and Wilson, 2006; Skelly et al., 2014). Further, excretory/secretory products (ES) are released by all developmental stages of the parasite and are relevant in parasite interactions with the host as well as with other
parasite (Cao et al., 2016; El Ridi and Tallima, 2009; Nowacki et al., 2015; Sotillo et al., 2016; Zhu et al., 2016).

1.3.1.1 Tegument

Early works focusing on the *S. mansoni* tegument (Hockley and Mclaren, 1973) described how the cercaria coating is replaced by an intramammalian tegument within 3 hours of skin penetration. A cercaria is covered with a thick glycocalyx layer that protects the parasite from osmotic pressure of fresh water (reviewed in Cai et al., 2016). Upon penetration, the cercaria loses its glycocalyx and the resulting schistosomule becomes covered with a double bilayer membrane, or with multiple membrane layers, with glycocalyx remains in some regions (Hockley and Mclaren, 1973). Proteins, carbohydrate, and glycoproteins, and host biomolecules, are found in the new tegument of the parasite, and are thought to be important for early immune evasion (reviewed in Kusel et al., 2007; Skelly and Wilson, 2006). The tegument was first known as “heptalaminate” for its appearance as seven layers on the electron microscopy (Hockley and Mclaren, 1973). Later, it was reported to be tightly opposed double bilayers lying over a syncitial cytoplasm (Mclaren and Hockley, 1977). There are cytoskeletal structures underneath the inner bilayer, and underneath this is a muscle layer (Skelly and Wilson, 2006). Cell bodies are embedded beneath the muscle tissue layer and connected to the syncitial cytoplasm by passing multilaminate vesicle into the tegumental layer and releasing their component to become part of the outer membrane, which is known as the membranocalyx (Skelly and Shoemaker, 2001; Wilson and Barnes, 1977) (Figure 1.2). Through this process, the tegument is constantly shed (tegument turnover) and the rate of turnover can be affected by binding of host molecules (reviewed in Van Hellemont et al., 2006). The schistosome tegument consists of pits, formed as invaginations of inner and outer membrane layers (Hockley and Mclaren, 1973). The parasite body is coated with spines and sensory papillae that contain nerve ending for sensing the external environment (Gustafsson, 1987). The pattern of coverage changes during development; for example, the mid-body spine and sensory papillae disappear in the lung stage (4-7 day post-infection) and appear again at day 10 and throughout adult stages (Crabtree and Wilson, 1980).
The tegument was originally thought to be an inert layer of membranocalyx offering protection from host antibody binding and hence antigen-mediated parasite killing, but later studies showed that there are proteins in the inner and outer layers, and in the space between them (Kusel et al., 2007; Wilson, 2012). With advances in proteomics, multiple tegument proteins have been identified using various methods (e.g. enzymatic shaving, biotin labelling, freeze-fracture, detergent) with some providing information on protein localisation on the tegument double bilayers and their abundance (Braschi and Wilson, 2006; Braschi et al., 2006a; Castro-Borges et al., 2011a, 2011b; Sotillo et al., 2015). These proteins include enzymes, transporters of glucose and amino acids, aquaporins, receptors, structural proteins, proteins of host origin, and proteins of unknown function (Braschi and Wilson, 2006; Braschi et al., 2006a; Castro-Borges et al., 2011a, 2011b; Sotillo et al., 2015). The tegument also consists of carbohydrates and lipids, particularly sphingomyelin and cholesterol (Kusel et al., 2007). Detailed information on the tegumental structure and components is covered by Skelly and Wilson (2006). Selected tegumental proteins with relevance to host-parasite interactions are explored in later sections.
1.3.1.2 Lining of the oesophagus and gut

The gut of the parasite is a blind-ended elongated tube lined with a layer of epithelium called gastrodermis (Basch, 1991). Surrounding the gut is a smooth muscle layer for peristalsis to move substances along the lumen. The gut lining is similar to the tegumental lining in that it is a syncitium layer of membrane. However, the gut lining has nuclei and “biosynthetic machinery” in the syncitial cytoplasm; whereas, the tegumental lining is linked to the cell body by cytoplasmic connections (reviewed in Skelly et al., 2014). Studies of the gut physiology, particularly the gastrodermis, have been challenging because of the inaccessibility of the organ (Gobert et al., 2009a). However, recent advances in laser microdissection have allowed the removal of the gastrodermis of *S. mansoni* and *S. japonicum* and led to finding of peptidase transcripts including transcripts encoding cathepsins, as well as transcripts encoding proteins related to lipid uptake, and antioxidant enzymes (Gobert et al., 2009a; Nawaratna et al., 2011). In addition to the gastrodermis, digestive enzymes are also secreted from a posterior oesophageal gland (Nawaratna et al., 2014) and the lysis of host erythrocytes is seen in the oesophagus (Li et al., 2013). Leukocytes are also retained in the posterior oesophagus possibly to prevent its oxidative damage to the gut (Li et al., 2013).

1.3.1.3 Excretory/secretory products

ES of parasites include substances from multiple sources. In *Schistosoma* species, ES could be a mixture of substrates secreted from oesophageal glands, gut regurgitation (or “vomitus”), tegumental secretion, or shed tegument. As schistosomes feed on blood, haem molecules are released from red blood cells. Free haems are converted to hemozoin and regurgitated into host blood as part of the vomitus. Vomitus from the parasite gut is a mixture of digestive waste and other proteins such as stress and tegumental proteins (Hall et al., 2011). The parasites may not be able to selectively retain the stress and tegumental proteins while regurgitating digestive waste, or their secretion maybe functional. In addition, *S. mansoni* ES contain exosome vesicles which carry miRNA and multiple proteins (Nowacki et al., 2015; Sotillo et al., 2016). Exosomes secreted from parasitic nematodes can modulate host immune responses (Buck et al., 2014), suggesting that exosomes secreted from *S. mansoni* may also have roles in host-parasite interactions (Nowacki et al., 2015; Sotillo et al., 2016).
1.3.2 Motility, behaviour, and homeostasis: the nervous system

The nervous system can be involved in host-parasite interactions, parasite-parasite interactions, as well as internal signalling within the parasite (Collins et al., 2010; Ribeiro and Patocka, 2013). Given the lack of an endocrine system in schistosomes, the nervous system and neuronal signalling are important for controlling motility, which is crucial for intramammalian migration, as well as other biological processes, such as sensing environment, behaviour, feeding, digestion, and excretion of waste (Ribeiro and Patocka, 2013). Moreover, sensing stimuli from the external environment could be relevant to activation of responses and guiding the behaviour of the parasite in processes such as lung, liver, and mesenteric migration, and mating between males and females (Kusel et al., 2007).

The layout of the central nervous system in *S. mansoni* is similar to that of other flatworms. Paired cephalic ganglia connect the nerve cord that runs through the body longitudinally, with transverse connections (“transverse commissures”) along the length forming the appearance of a “ladder” (Collins et al., 2011). Neuronal signalling involves neurotransmitters, transporters, and GPCR signalling (reviewed in Ribeiro and Patocka, 2013). Neurotransmitters used by *S. mansoni* are of various groups - acetylcholine, glutamate, biogenic amine (e.g. serotonin (5-HT), histamine, dopamine, and noradrenaline), and neuropeptides (reviewed in Ribeiro and Geary, 2010). The involvement of neurotransmitters and neuronal signalling in the motility of *S. mansoni* has been shown by RNAi of serotonin receptors affecting motility in 8-day schistosomules and adults (Patocka et al., 2014), and RNAi of acetylcholine GPCR leading to reduced motility in schistosomules (MacDonald et al., 2015). In addition, irradiated *S. mansoni* cannot migrate through the lung capillary network and a gene expression study shows that down-regulation of genes related to neuromuscular signalling may be involved (Dillon et al., 2008). Neuropeptides, also known as peptide hormones, are involved in developmental control and tissue differentiation in free-living flatworms such as planaria (Collins et al., 2010). *S. mansoni* may similarly rely on neuropeptide signalling (Collins et al., 2010). Attraction between male and female *S. mansoni* is also likely to involve sensory organs that connect to the nervous systems to alter behaviour and motility (reviewed in Kusel et al., 2007). Neuronal signalling is therefore likely to be a major component in many aspects of parasite biology.
1.3.3 Migration route and development

Upon entering the bloodstream, schistosomules are carried with blood circulation to multiple tissues from lung, heart, systemic organs, to liver, and migrate against the blood flow to the mesenteric venules for egg-laying (reviewed in Wilson, 2009). The encountering with various host tissues may act as signal for metabolic changes and development in the parasite (Kusel et al., 2007). Studies aiming to identify the migration route of the parasite have incorporated a range of techniques such as organ mincing to recover parasites from tissues, histology, and isotope tracking (e.g. Georgi et al., 1982; Miller and Wilson, 1978, 1980; Wheater and Wilson, 1979). Isotope tracking provides improved accuracy because it does not rely on parasites migrating out of tissues, or on investigators detecting parasites among a background of host tissues (Wilson, 2009). The technique involves incubating snail hosts with [$^{75}$Se]-methionine (a gamma-emitting isotope) to label cercariae and trace the parasites in pressed organs (Georgi et al., 1982). The path and timing of the migration has then been extensively described. However, only circumstantial evidence exists for the mechanisms that lead to the observed patterns of migration (Wilson, 2009).

1.3.3.1 Lung migration

Once schistosomules enter capillaries, it is thought that the parasites are carried by the blood circulation to enter the heart (right atrium) and then enter the lung capillary bed via the pulmonary arteries. Schistosomules can be found in the lung from day 4 and the number of parasites detected in the lung increases over time until it peaks around day 6 post-infection for S. mansoni (Clegg, 1965a; Georgi et al., 1986). During the lung migration, the schistosomules elongate to fit within narrow lung capillaries and are in very close contact with the lung capillary wall, prolonging the migration (Crabtree and Wilson, 1986a). The body length increases by 2-3 fold, relative to the skin stage (Clegg, 1965a; Wilson et al., 1978), but no cell division has been reported during this stage (Clegg, 1965a; Lawson and Wilson, 1980). The signals that trigger the morphological changes of schistosomule from the skin stage to the elongated lung stage are unclear, but it has been postulated that host proteins might be involved in this process (reviewed in Kusel et al., 2007). Alternatively, physical space in the lung and/or changes in the oxygen pressure might also have an effect.
In addition to elongation of the body, the parasites might also actively dilate blood vessels to allow the migration through the lung capillaries (Carvalho et al., 1998), and a use of lubricant has been proposed (Crabtree and Wilson, 1986a). Furthermore, the tegument in the lung stage lacks midbody spines, and the absence could provide room for contraction of the body helping with migration (Crabtree and Wilson, 1980). The spines are however present around the anterior and posterior ends of the parasites, and can cause damage to lung endothelial cell membranes (Crabtree and Wilson, 1986a) that may result in inflammation at a later stage (Crabtree and Wilson, 1986b). In contrast to the notion of prolonged lung migration, in vivo microscopy by Bloch (1980), observed schistosomule migration in anesthetised and surgically opened mice, and reported objects of schistosomule size passing through lung capillary without being detained. However, as the author commented, parasites in deeper layers of lung capillary may still have been detained and missed by this technique (Bloch, 1980).

1.3.3.2 Lung as a parasite attrition site

The migration in the lung capillaries is a challenge for the schistosomules (reviewed in Wilson et al., 2016). Radioisotope tracking showed that over 85% of the parasites can be counted in the lung, and only half of the parasites that reach the lung can be counted in the liver (Georgi et al., 1986). The number of parasites reaching the liver was similar to the number of adult parasites recovered by perfusion, suggesting that the attrition happened during or after the lung migration (Georgi et al., 1982). More evidence for parasite attrition in the lung came from histological examinations, reporting schistosomules accumulating in alveoli over time and none returning to blood vessels (Crabtree and Wilson, 1986a). The observation for infection in mice immunised with irradiated cercariae was similar, but with more host immune cells reaching the lung (Crabtree and Wilson, 1986b; Dean and Mangold, 1992). Similarly, in schistosome infections of rats, macrophages could be seen surrounding schistosomules in the lung vasculature in both primary and secondary infection, with fewer cases in primary infection (Bentley et al., 1981). Large inflammatory foci with macrophages and granulocytes were also found but this was not associated with the presence of schistosomules (Bentley et al., 1981). It is therefore clear that parasites are eliminated in the alveoli although the cause of death is not clear. Despite the inflammation around the lung tissue or around the parasites, most parasites in the alveoli were intact or appeared to die from autophagy (Crabtree and Wilson, 1986b;
Mastin et al., 1985). The accumulation of immune cells, instead of killing the parasites, is thought to either disrupt blood vessels, inducing them to become leaky and enabling the parasites to migrate into alveoli, or act as plugs that block migration (Dean and Mangold, 1992; Wilson, 2009). The departure of the parasites from the bloodstream may lead to eventual parasite death in alveoli. However, it is suggested that the parasites could be passed up the trachea and swallowed into the host gut where they were digested or passed out to the environment alive (Dean and Mangold, 1992).

It is intriguing that, although the lung schistosomules were surrounded by immune cells and inflammation in damaged lung tissues, the parasites seem to be resistant to the immune-mediate killing (Bentley et al., 1981; Crabtree and Wilson, 1986b; Mastin et al., 1985). This is consistent with observations that lung schistosomules are resistant to cytotoxic killing (e.g. Clegg and Smither, 1971; McLaren and Terry, 1982). Furthermore, accumulation of inflammatory immune cells was observed after the peak time of schistosomules migration through the lung, and the inflammatory foci were not always associated with the parasites. Therefore, the lung inflammatory responses could be towards damaged lung tissues instead of towards the parasites (Burke et al., 2011; Crabtree and Wilson, 1986b; Mastin et al., 1985). The mechanisms that protect lung schistosomules from inflammatory damage is unclear, although acquisition of host antigens for immunological camouflage may be involved (Clegg and Smither, 1971; McLaren and Terry, 1982). Furthermore, gene expression profiling using microarrays of in vitro-cultivated and in vivo lung schistosomules showed that genes responding to stress (hsp70) and other immunomodulation-related genes were upregulated in the in vivo lung schistosomules, suggesting that the host environment may also have a role in activation of immune evasion strategies in the parasite (Chai et al., 2006).

1.3.3.3 From the lung to the liver

From the lung, parasites are carried with systemic circulation and remain within blood vessels (Bloch, 1980; Wheater and Wilson, 1979) until they arrive at the liver (Bloch, 1980; Wilson, 2009). Schistosomules are found in liver from day 8 (Clegg, 1965a) and all schistosomules reach the hepatic portal system by day 21 (Georgi et al., 1986; Wilson et al., 1986). Appearance of schistosomules in other organs during this time also supports the hypothesis that migration between lung and liver is not an active,
direct passage (Wheater and Wilson, 1979; Wilson, 2009). A histological study of infected mice confirms systemic circulation of parasites, as schistosomules were found in the pulmonary veins, cerebral blood vessels, and in myocardium (Wheater and Wilson, 1979). Parasites are distributed to systemic organs according to the proportion of cardiac output that supplies the organs (Wilson et al., 1986). Sinusoidal networks may act as filters once the parasites grow large enough to be trapped within the portal system (reviewed in Wilson, 2009).

### 1.3.3.4 Liver schistosomules

Liver schistosomules are in ‘interlobular portal venules’ which is part of the intrahepatic radicle of the portal vein and are not found in the sinusoid. Furthermore, they are found to reside in blood vessel just large enough for their diameter (Bentley et al., 1981; Bloch, 1980). By residing in blood vessels, the parasites appear to reduce blood flow reaching the sinusoid adjacent to their portal triad, and hepatocytes become vacuolate, and develop midzonal necrosis (Bloch, 1980). Microscopic hemorrhage, damage, and fibrosis of liver are also observed before the egg laying stage (Bloch, 1980). This shows that the schistosomules also cause damage to the host liver, in addition to the damage of lung endothelial cells during the lung migration. Similarly to the lung stage, in schistosome infections of rats, eosinophils, mast cells, and mononuclear cells are seen infiltrating hepatic tissue surrounding the portal venule where schistosomules reside, but these cells were not observed bound on to schistosomules (Bentley et al., 1981).

Once in the liver, elongated lung schistosomules contract again into the size of skin schistosomules (Wilson, 2009). Although arrival in the liver may be a passive process, the liver environment could provide the parasites with signals to grow (reviewed in Kusel et al., 2007). Increased cell division was observed when schistosomules from 20 days post-infection were cultured with portal blood from humans or susceptible rodent hosts (Draz et al., 2008; Shaker et al., 1998). The portal-blood-derived molecules, have not been identified but are less than 50 kDa (Shaker et al., 1998, 2011). Mitosis and developmental changes start with the schistosomule body and gut beginning to extend (Clegg, 1965). First the gut appear as two fork ends, and by day 15 the posterior ends of the two forks join (Clegg, 1965a). Development is asynchronous both in vivo and in vitro (Basch, 1981; Clegg, 1965a). After 21 days
post-infection, reproductive organs start to form in both males and females. Mating can be observed from day 28 (Clegg, 1965a). Egg shell protein synthesis starts at day 30, vitelline tissues then develop in females, and oviposition is observed at day 35 (Clegg, 1965a).

1.3.3.5 Migration paths for egg laying

*S. mansoni* and *S. japonicum* dwell in mesenteric veins as a site for egg laying. In contrast, *S. haematobium* prefers the venus plexus near the bladder (Gryseels et al., 2006). This distinct migration path specific to different species suggests that some specific interactions might be happening. Although the chemical cues of the site determination is unknown, it is generally thought that male worms determine the site of migration for egg laying (Gryseels et al., 2006; Huyse et al., 2009; Webster et al., 2013) and that pairing occurs in portal veins before the migration towards mesenteric vein (Standen, 1953). However, it is worth noting that females in single-sex infections can also be found in mesenteric veins (Zanotti et al., 1982) and that unpaired females can be found in the mesenteric circulation (Bloch, 1980). Furthermore, the number of female worms found in the mesenteric veins increases over time (Zanotti et al., 1982) suggesting that the parasites migrate to the site after they mature. Transcriptomes of unpaired females are more similar to those of males than to those of mature females (Lu et al., 2016). This similarity in transcriptomes may also lead to similar behaviours of unpaired females migrating as males.

1.3.3.6 Imaging technology

Studies of *S. mansoni* migration pattern were advanced by the use of radioisotope tracking (Wilson, 2009). To study mechanisms that determine the migration site, similar experiments would need to be repeated with modified parasites (such as parasites with a candidate gene knocked down using RNAi). However, [*75-Se*]-methionine - a key substrate that was used to establish the current understanding on the migration - was withdrawn “around 1990” (Wilson, 2009). Therefore, a new imaging technique will be required to study effects of genes and signaling pathway on the migration inside the mammalian host. Detection of fluorescent trace emitted from parasites was able to show location of foci of adult parasites *in vivo* (reviewed in Skelly, 2013). This technique works by injecting fluorochrome ProSense 680 into infected mice. The fluorochrome is activated by cathepsin in the parasites, releasing
fluorescent signals which can be detected and imaged using Fluorescence Molecular Tomography (in vivo imaging). In addition, radioactive tagged glucose allows in vivo visualisation of adult stages (Salem et al., 2010). However, these techniques have not been demonstrated for early stage schistosomules.

1.3.4 Metabolic requirements and acquisition

The bloodstream environment is usually described as hostile for blood-dwelling parasites (e.g. Berriman et al., 2009; Cook et al., 2004; Wendt and Collins, 2016). However, schistosomes have evolved mechanisms to not only avoid host defence mechanisms but also to take advantage of bloodstream components as essential input for their development and reproduction (Kusel et al., 2007). *S. mansoni* requires multiple substrates from the host including glucose, amino acids, fatty acids, sterol, nucleotides, vitamins, iron, and other ions (Skelly et al., 2014). Some of these are acquired through tegument transporters or receptors, while some are obtained from blood feeding and likely to be absorbed through the gut surface (Kusel et al., 2007; Skelly et al., 2014).

1.3.4.1 Sugar and other substances absorbed through tegument

Schistosomes consume a large amount of glucose, approximately two fifth of their dry weight per hour in adults (Clegg, 1965a). *S. mansoni* adults mainly rely on anaerobic respiration to obtain energy from glucose (Da’dara et al., 2012; Tielens and van den Bergh, 1987). However, aerobic respiration through Krebs cycle can occur in the presence of oxygen (van Oordt et al., 1985). Ensuring constant supply of glucose, *S. mansoni* adults store glycogen in their body, and the degradation can be readily triggered by limited external glucose (Tielens and van den Bergh, 1987). It has been suggested that glucose requirement may explain the wandering behaviour of adult pairs in the mesenteric vein (Pellegrino and Coelho, 1978); parasites wander to new areas of glucose supply after their bodies have inhibited the blood flow and limited the glucose availability in their immediate surrounding (Tielens and van den Bergh, 1987).

*S. mansoni* obtains glucose (and other sugars) through transporters in their tegument (Skelly et al., 2014). Two of the sugar transporters in *S. mansoni* have been tested by RNAi and shown to transport glucose as well as other hexoses (Skelly et al., 1994).
Their silencing by RNAi led to a deleterious effect on the parasites especially when the parasites were subjected to in vivo environment or when they were maintained in vitro in a low glucose medium (Krautz-peterson et al., 2010). One of the transporters, SGTP4, is expressed on the outer layer of tegument, while the other one, SGTP1, is expressed in the inner layer, as well as in muscle (Jiang et al., 1996; Zhong et al., 1995). Such localisation would allow the sequential uptake of glucose from the bloodstream to pass through the outer layer of the tegument, into the interlayer space, and through the inner tegument layer into the parasite body (Da’dara et al., 2012; Skelly et al., 2014). The sugar transporters are expressed early after the skin penetration before the parasites reach the liver and start developing functional gut and oral sucker, suggesting the requirement of glucose even before the peak of development in the liver (Skelly and Shoemaker, 1996). Intriguingly, the *S. mansoni* genome also encodes two other sugar transporters, SGTP2 and SGTP3 (Berriman et al., 2009). However, biophysical modelling predictions suggest that the products encoded by these genes are not functional glucose transporters (Cabezas-Cruz et al., 2015), and previous characterisation has also confirmed the absence of glucose transport function in SGTP2 (Skelly et al., 1994).

Other substances acquired through the tegument include purine nucleotides and ions such as calcium, sodium, phosphorus, and copper (Skelly et al., 2014). Transporters of these ions and their associated ATPase have not been characterised and functionally tested but they have been described in tegumental proteomics (reviewed in Da’dara et al., 2012; Skelly and Wilson, 2006; Skelly et al., 2014). *S. mansoni* lacks pathways for de novo synthesis of purine nucleotides and needs to obtain purines from the host (Dovey et al., 1984). The process of acquisition may involve hydrolysis of phospho-nucleotide and diffusion through the tegumental surface (Levy and Read, 1975). It has subsequently been demonstrated that alkaline phosphatase and ATP diphosphohydrolase on the tegument can cleave nucleotide phosphate (e.g. ATP and ADP) into de-phosphorylated form (e.g. AMP), which can be transported across the membrane (Da’dara et al., 2014).

### 1.3.4.2 Amino acid - through tegument and gut

During the development as well as maintenance of egg-laying and survival in adults, *S. mansoni* feeds on host blood and obtains essential nutrients and iron from ingested
erythrocytes (reviewed in Kusel et al., 2007; Skelly et al., 2014). Haemoglobin is a major source of protein and iron. In addition, plasma proteins, especially albumin, are also a source of amino acids (Delcroix et al., 2006; Hall et al., 2011). Proteomic studies of worm vomitus (Hall et al., 2011) and gene expression profiling of microdissected gastrodermis (Gobert et al., 2009a; Nawaratna et al., 2011) have revealed that schistosomes have an array of proteases in their guts to digest host erythrocytes and obtain nutrients (Gobert et al., 2009a; Hall et al., 2011; Nawaratna et al., 2011). The proteases involved in digestion of blood work as a cascade consisting of cathepsins (D, B1, L1, and C) and metallo-aminopeptidase (Delcroix et al., 2006). Suppression of these cathepsins (e.g. D and B1) affect parasite growth and digestion (Skelly et al., 2014). The expression of genes encoding proteins used for blood feeding – such as cathepsins and iron transport proteins – increases in vitro when schistosomules are fed erythrocytes, suggesting that these processes are stimulated by blood feeding (Gobert et al., 2010).

At least five amino acid transport systems have been suggested for S. mansoni adults (Asch and Read, 1975). So far, one system of amino acid transporter has been identified and it has been located on the tegument surface (Braschi et al., 2006b; Skelly et al., 1999) and in the gut lining (Skelly et al., 2014). The transporter is known as SPRM1 (schistosome permease 1) and it transports a range of amino acids (phenylalanine, arginine, lysine, alanine, glutamine, histidine, tryptophan, and leucine) (Krautz-Peterson et al., 2007). Other amino acid transporter systems have not yet been identified.

1.3.4.3 Iron - through gut and tegument
Blood feeding not only provides a source of amino acids for S. mansoni, but also a source of iron which the parasite requires for development and reproduction (Clemens and Basch, 1989; Glanfield et al., 2007; Jones et al., 2007). In addition, the competition between the host and the pathogen to sequester iron is one of the determinants of infection success (reviewed in Glanfield et al., 2007). In the liver, mammalian hepatocytes retain iron (by expression of hemopexin and hepcidin) competing with invading bacteria (reviewed in Zhou et al., 2016). Given that S. mansoni development is associated with liver and iron is required for parasite growth,
the parasites must have an efficient method for obtaining iron. However, the processes of iron uptake is not well defined.

It has been suggested that iron is taken up from blood feeding given that there is excess of iron in the haem which results from digested haemoglobin (Glanfield et al., 2007). However, obtaining iron from haem requires haem oxygenase which has not been identified in *S. mansoni* and not found in the genome (reviewed in Glanfield et al., 2007; Skelly et al., 2014). In addition, growth of early schistosomules is stimulated by presence of iron, suggesting that iron uptake happens before blood feeding starts (Clemens and Basch, 1989). This early sequestration of iron may also help the parasite evade the host immune responses by competing for iron with immune cells. For this purpose, the sequestration should take place as soon as the parasite enter the bloodstream. The system for uptake of iron through the tegument has been suggested to use a divalent metal transporter (DMT) (Clemens and Basch, 1989; Smyth et al., 2006). Such a transporter has been identified on the tegumental surface of *S. mansoni* and displays the ability to transport Fe$^{2+}$ (Smyth et al., 2006). In host blood, ferric ions (Fe$^{3+}$) are bound to transferrin. The Fe$^{3+}$ would be reduced to Fe$^{2+}$ which could be transported through DMT and stored in parasite’s ferritin (reviewed in Glanfield et al., 2007). Ferritin stores iron in a non-toxic form, which otherwise could lead to oxidative damage of surrounding tissues (Munro, 1990). Two types of ferritin are described in *S. mansoni*: ferritin-1 is expressed in reproductive organs, and ferritin-2 is expressed in somatic cells (Schüssler et al., 1996). Interestingly, ferritin genes in “higher animals” possess upstream sequences in their mRNA and translation into ferritin protein is regulated by the level of iron in the environment (Munro, 1990; Schüssler et al., 1996). In contrast, *S. mansoni* ferritins do not have such upstream sequences and it appears that *S. mansoni* ferritin levels are regulated using different, yet to be identified, mechanisms (Schüssler et al., 1996). Using a mechanism different from that used by the host may be beneficial in competition for iron.

### 1.3.4.4 Lipid through gut

*S. mansoni* is coated with a double lipid bilayer consisting of cholesterol as one of the major components (Skelly and Wilson, 2006). However, the parasite cannot produce fatty acid or sterol *de novo*, but can use complex lipid if provided with fatty acid...
(Meyer et al., 1970). Therefore, the starting molecules need to be obtained from their host (Meyer et al., 1970). So far there is no evidence of lipid uptake through the tegument. Low density lipoprotein (LDL) has been shown to bind to the tegument but the uptake of LDL through tegument has not been reported and a receptor for LDL has not been found (reviewed in Skelly et al., 2014). It is thought that lipid uptake happens in the gut using Niemann Pick type C2 protein (NPC2) and saposin (reviewed in Skelly et al., 2014). In eukaryotic cells, NPC2 is involved in traffic of cholesterol, sterol, and glycolipids and a homologue is reported in the S. mansoni gut vomitus (Hall et al., 2011). In addition, saposin that may work as a lipid-binding and interacting protein is found in the gut vomitus and is expressed in the gastrodermis (Don et al., 2008; Hall et al., 2011; Nawaratna et al., 2011). The role of saposins in S. mansoni has not been determined, but assuming similarity with other systems, they may be involved in lipid binding and uptake in the gut (reviewed in Skelly et al., 2014).

1.3.4.5 Hormone receptors and host molecules stimulating parasite development

In addition to nutrients, essential ions, and nucleotide that the parasite acquires from the host, other unknown host factors may display essential roles for the development and reproduction of schistosomes. As mentioned previously, a protein from portal blood can stimulate growth of schistosomules (Shaker et al., 1998, 2011). In addition, S. mansoni develops more slowly in vitro than in vivo and appears to be 20–40% smaller (Basch, 1981; Clegg, 1965a). Despite some in vitro culture conditions supporting the adult parasites, eggs laid by these in vitro-cultured worms are not fertile (Basch and Humbert, 1981). The presence of host molecules stimulates downstream effects that can lead to increase nutrient uptake, or trigger developmental progress (Hernandez et al., 2004; Saule et al., 2002, 2005).

The S. mansoni genome encodes genes that are homologous to receptors of host hormones such as host insulin receptor (SmIR1 and SmIR2) (Verjovski-Almeida et al., 2003) and transforming growth factor-beta (TGF-β) receptor (Beall and Pearce, 2001). SmIR1 and SmIR2 bind to insulin (Khayath et al., 2007) and affect regulation of glucose uptake in the parasite (Ahier et al., 2008). Extra supplies of insulin in infected mice during liver stage development (day 14 - day 21 post-infection) results in increased worm burden and size (Saule et al., 2005). Knocking down similar
receptors in \textit{S. japonicum} resulted in a negative impact on growth and reproduction of adult worms (You \textit{et al.}, 2015), together suggesting that the hormone may help with both growth and survival of the parasite. For TGF-\(\beta\), SmSK1 was identified as a receptor which activates SmSmad2 (Beall and Pearce, 2001). This in turn leads to gene expression of the gynaecophoral canal protein (GCP) (Osman \textit{et al.}, 2006).

In addition to host hormones and growth factors, \textit{S. mansoni} has integrated into the mammalian host environment to such an extent that components of the host immune system have become essential for development (Davies \textit{et al.}, 1980). In immunosuppressed mice, male and female development of \textit{S. mansoni} was delayed (Harrison and Doenhoff, 1983). Later, it was shown that, CD4+ lymphocytes, IL-7, and IL-2 are required for parasite development in the portal system and egg production (Blank \textit{et al.}, 2006; Davies \textit{et al.}, 2001; Hernandez \textit{et al.}, 2004). Such effects may also intertwine with the positive effect of IL-7 supplement, as well as thyroxine (T4), on the development of schistosomes (Saule \textit{et al.}, 2002). More recently, it has been shown that the effect of CD4+ cells involves stimulation from other innate immune cells such as macrophages and monocytes (Lamb \textit{et al.}, 2010). The effect on parasite development was shown to, at least in part, work through regulation of such immune cells by IL-4 (a Th-2 cytokine) which might explain appearance of Th-2 cytokine profiles before egg-laying (Riner \textit{et al.}, 2013).

Together, the roles of host molecules demonstrate intricate connections between the parasite and host environment, including the regulation of the host immune system for their development. Further it suggest how the parasites might benefit from the infiltration of some immune cells in the liver tissue outside the blood vessel where the worm resides (as described by Bentley \textit{et al.}, 1981), as well as a role of liver residential macrophages (Kupffer cells) in development of the parasites.

\subsection*{1.3.5 Immune evasion}

Adult pairs of \textit{S. mansoni} may live in the mammalian host over years with no sign of immune-mediated killing of the worms (Keating \textit{et al.}, 2006; Kusel \textit{et al.}, 2007; Pearce and MacDonald, 2002). Therefore, some mechanisms must have evolved to allow the evasion of the host immune response. During the migration, immune cells are in close proximity to the parasite but do not appear to induce any damage or
killing (Bentley et al., 1981; Crabtree and Wilson, 1986a). Early stages of *S. mansoni* are susceptible for immune killing by the complement cascade (Clegg and Smithers, 1972; Marikovsky et al., 1990). In particular, it has been shown that cercariae were killed by complement *in vitro* and the glycocalyx coating of the cercariae contributes to this killing (Marikovsky et al., 1990). Soon after the parasite penetrates the skin and lose their coating, schistosomules become resistant to complement killing (Clegg and Smithers, 1972; Kusel et al., 2007; Marikovsky et al., 1990; Wilson, 2012).

1.3.5.1 Camouflage

One of the very first proposed mechanisms of immune evasion involved the role of the tegument (Clegg and Smither, 1971). Host proteins are found on the parasite tegumental surface, leading to a proposed mechanism of immune evasion whereby the parasites camouflage underneath the host antigens (Clegg and Smither, 1971). This hypothesis suggested that *S. mansoni* is able to absorb host blood group antigens among other mammalian antigens on to their surface (Clegg and Smither, 1971; Goldring et al., 1976; Smithers et al., 1969). When worms were transferred from mice (hence coated with mouse antigen) to monkeys that had been immunised against mouse antigen, the worms were killed by the antigen-mediated immune responses (Clegg et al., 1970; Smithers et al., 1969). Similar findings were also described for schistosomules co-cultured with different human blood groups and tested in monkey immunised against the appropriate human blood group antigen (Goldring et al., 1976). Interestingly, when the lung stage parasites were transferred, the parasites were resistant to the antibody-mediated killing. However, the resistance was not observed when 15 days old parasites were used for the transfer (Clegg and Smither, 1971). *In vitro* observations are not fully consistent with these *in vivo* experiments. In a study where *S. mansoni* were recovered from mice and cultured with eosinophils and antibody against mouse protein, the lung stage was resistant to the antibody-mediated killing and, in contrast to *in vivo* observations, day-14 and day-21 worms were also resistant *in vitro* (McLaren and Terry, 1982). Only adults and larger liver developmental stages were killed by this process (McLaren and Terry, 1982). Despite the contradictions, both studies agree that the parasites are coated with host antigens, and further show that the lung stage might possess additional protective mechanisms against cytotoxic killing.
The camouflage strategy alone, however, could not suffice for immune evasion. The parasite tegument is constantly shed and renewed; therefore, phases where the parasites are not completely covered by host antigen will exist. In addition, the uptake of nutrients through the tegument may require some of the parasite transporters to be exposed to the host environment. Furthermore, activities of antibody-dependent cell-mediated cytotoxicity (ADCC) and complement are reduced in mice infected with schistosomes (Attallah et al., 1987), suggesting other immune-modulation strategies in schistosome infections.

1.3.5.2 Tegument complement inhibition

The complement system, or complement cascade, forms a part of immune defence in the mammalian host which S. mansoni needs to evade. The complement system consists of serum and surface proteins working together as protease activation cascades, and inhibitors exist at multiple levels (Markiewski et al., 2007). Molecules involved in the complement cascades are illustrated in Figure 1.3. The cascades ultimately result in the formation of membrane attack complexes (MAC) on the surface of invading pathogens, leading to the elimination of pathogens. In addition to MAC formation, an activated complement component (C3b) binds to pathogen surfaces, opsonizing them for attack by other immune cells (Abbas et al., 2014).
Figure 1.3 Coagulation and complement cascade.

Complement cascades consist of three pathways. The classical pathway is activated by an antigen-antibody complex interacting with complement component 1 (C1 subunits). The lectin pathway is activated by detection of carbohydrate on pathogen surfaces. The alternative pathway starts with the spontaneous activation of complement component 3 (C3), which forms C3b on pathogen surface. The three pathways converge at the formation of C3 and C5 convertase, leading to release of anaphylatoxin C3a and C5a that induce inflammatory responses. To the C5 convertase, other components (C6-C9) convene and form membrane attack complex which produces a pore on invading cells and kills them. Binding of C3b opsonized the surface for targeting by phagocytic cells. Represented in blue boxes are inhibitors of complement cascades. Diagram reproduced from Wagner and Frank (2010).

Schistosomes become resistant to complement killing soon after they transform into schistosomules (Marikovsky et al., 1990); however, this inhibition of complement seems to be reversed when parasites are treated with trypsin suggesting that some proteins on the tegument are involved in the resistance to complement killing (Marikovsky et al., 1990). Paramyosin (also known as schistosome complement inhibitory protein-1, or SCIP-1) has been identified as one of these tegumental proteins and it functions as an inhibitor of complement C9. The parasite paramyosin...
also cross reacts with antibody against human CD59 which is also a complement C9 inhibitor (Deng et al., 2003; Parizade et al., 1994). In addition to inhibiting complement cascade, paramyosin is also one of the main receptors for host immunoglobulin binding (Loukas et al., 2001). More recently, proteomic work showed that other host proteins also bind to parasite tegument but receptors have not been identified (Braschi et al., 2006b, 2006a; Wu et al., 2015).

The complement cascade is also inhibited at an early step. C3 and C4 were found on the outer layer of parasite tegument (Braschi et al., 2006b, 2006a; Castro-Borges et al., 2011a); however, the small amount of C3 suggests that complement activation, which results in amplification of C3, does not occur (Castro-Borges et al., 2011a). The inhibition of C3 and C4 activation has been partially illustrated (Da’dara et al., 2016; Fatima et al., 1991; Pearce et al., 1990); S. mansoni can recruit decay-accelerating factor (DAF) (70- kDa glycoprotein, inhibitor of C3b activation) from host erythrocytes onto their tegument and prevent activation of complement through C3b (Fatima et al., 1991; Pearce et al., 1990). Furthermore, when human serum was incubated with adult S. mansoni, multiple complement activating proteins, including C3 and C4, were degraded. The resulting fragments could derive from the involvement of factor I (complement inhibitor enzyme) and its cofactor, factor H. However, these proteins have not been identified in S. mansoni (Da’dara et al., 2016a). Binding of C3 to parasite tegument also stimulates membrane synthesis and shedding of the tegument outer layer, which might be another mechanism to prevent completion of complement cascade activation (Silva et al., 1993).

1.3.5.3 Modulation of immune responses through secretion

Another mechanism of protection from immune attack comes from the ES from the parasites. Complement receptor-related protein y (Cry, or CR1) inactivates C3-derived fragment and is found in the proteome of adult extracts (Braschi and Wilson, 2006; Braschi et al., 2006b). One of the major secreted antigens described in the vomitus (circulating anodic antigen, CAA) is able to bind complement C1q, suggesting that CAA prevents antibody-mediated complement attack in the parasite gut (van Dam et al., 1993). Apart from complement inhibition, superoxide dismutase, peroxidase, and thioredoxin are found in S. mansoni vomitome (Hall et al., 2011) and might be involved in protecting worms from oxidative stress (Gobert et al., 2009b). In
addition, in *S. japonicum*, thioredoxin peroxidase from 14 days old *ex vivo* schistosomules inhibits the expression of MHCII and CD80 genes in lipopolysaccharide-activated macrophage (LPS-activated macrophage), leading to reduced inflammation (Cao *et al*., 2015). During skin invasion, anti-inflammatory protein, Sm16, is secreted from schistosomules and induces production of host antinflammatory cytokines (Ramaswamy *et al*., 1995), and the protein is also expressed in other stages suggesting its function beyond skin invasion stage (Rao and Ramaswamy, 2000).

ES are also involved in the interaction with endothelial cells. However, the responses of endothelial cells to ES from lung schistosomules are different from the responses to adult parasites (Oliveira *et al*., 2011; Trottein *et al*., 1999a, 1999b). Lung schistosomule ES decrease the permeability of the endothelium lining (Trottein *et al*., 1999a), and reduce leukocyte adhesive protein VCAM and E-selectin, causing fewer leukocytes to bind to the treated endothelial cells (Trottein *et al*., 1999b). The key immunomodulatory molecule appears to be prostaglandin D2 (Angeli *et al*., 2001, 2001). In contrast, the permeability of the endothelium lining and the interactions with leukocytes are increased when ES collected from adult worms are employed (Oliveira *et al*., 2011), suggesting that the parasite interacts with host endothelial cells differently during its developmental stages.

### 1.3.5.4 Coagulation cascades

The coagulation process consists of platelet activation and formation of blood clots. Blood clots can form via extrinsic and intrinsic pathways. Extrinsic pathways are activated by tissue factors exposed when endothelial cells are damaged, and by tissue factors expressed on immune cells and on endothelial cell surfaces. The intrinsic pathway is activated when the factor 12 bind to collagen or a negatively charged surface. The accumulation of factor 12 can lead to auto-activation and activation of coagulation cascade (Figure 1.4).

Coagulation and innate immune responses are closely related. Multiple components in the coagulation and complement cascades activate components in the other pathways. For example, complement component C3 fragment (C3a) activates platelets and enhances platelet aggregation (reviewed in Esmon *et al*., 2011; and Markiewski *et al*., 2007); thrombin, which cleaves fibrinogen chains leading to formation of fibrin clot,
can also cleave C3 and C5 (reviewed in Esmon et al., 2011; and Markiewski et al., 2007). Given that schistosomes can inhibit complement cascades, it is likely that coagulation might be also affected. Coagulation is a well-known protective mechanism to prevent the spread of pathogens from the site of entry in the host (Markiewski et al., 2007). It is, however, less understood how coagulation affects macroparasites like parasitic worms. Presumably, blood coagulation hinders parasite migration during early stages, and blood clots might be deleterious for blood-feeding (reviewed in Mebius et al., 2013).

The three pathways that makeup the classical blood coagulation pathway

![Diagram of the coagulation pathway](https://via.placeholder.com/150)

**Figure 1.4 Coagulation pathway**

Both intrinsic and extrinsic pathways lead to cascades of protease activation which join at factor 10 (FX, shown as X). The coagulation pathway finishes with conversion of fibrinogen to fibrin which form blood clots. Platelet activation step is not included in the diagram. Diagram reproduced from Graham Beards, via Wikimedia Commons (2012).

### 1.3.5.5  *S. mansoni* interference with coagulation

*S. mansoni* resides in blood vessels just about the size of its body (Bentley et al., 1981; Bloch, 1980). This is expected to interrupt blood flow and cause damage to the endothelium, which should trigger blood coagulation cascades. However, blood
clotting is not observed around migrating parasites (reviewed in Da’dara and Skelly, 2011; and Mebius et al., 2013). Interactions of S. mansoni with coagulation cascades ranges from prevention of cascade activation to degradation of blood clots once the clots have been formed (Mebius et al., 2013). Vasodilators may be involved in prevention of coagulation by reducing the interruption of blood flow, and thereby preventing damage of host endothelium (Mebius et al., 2013). For example, S. mansoni adults express a tegumental enzyme with kallikrein-like activity (sK1) which can lead to production of bradykinin and cause vasodilation (Carvalho et al., 1998). Interestingly, bradykinin also stimulates inflammation and activation of neutrophils and mast cells (Hofman et al., 2016). Therefore, it seems that a fine balance between prevention of coagulation and regulation of inflammation is required for the S. mansoni survival in bloodstream.

In addition to vasodilators, a heparin-like protein is found in the tegumental proteome of S. mansoni although its function remains to be investigated (Castro-Borges et al., 2011b). The coagulation cascades can also be activated by ADP through platelet activation (Woulfe et al., 2001). The enzyme ATP-diphosphohydrolase on the parasite tegument, which assists hydrolysis of ATP and ADP to AMP for nucleotide uptake, may also have additional role in preventing activation of coagulation cascade by reducing availability of ADP (Da’dara et al., 2014; Mebius et al., 2013; Vasconcelos et al., 1993). Furthermore, a Kunitz type protease inhibitor, which is located in the tegument of adult worms and also secreted, inhibits a coagulation factor shared between both intrinsic and extrinsic pathways, factor 10a (Ranasinghe et al., 2015a). To prevent the formation of blood clots, secreted S. mansoni protein Sm22.6 inhibits protease activity of thrombin - an enzyme that cleaves fibrinogen chains for the formation of insoluble fibrin blood clot (Lin and He, 2006). Lastly, once the fibrin clots have been formed, increasing clot degradation (a process known as fibrinolysis) could be a counteraction strategy. Annexin in the tegument, which previously was thought to have structural function (reviewed in Wilson, 2012), was demonstrated in S. bovis to bind to plasminogen and enhance its conversion into plasmin (fibrinolysis enzyme). This could therefore lead to an increase in fibrinolysis rate and eventual reduction in coagulation (de la Torre-Escudero et al., 2012).
1.3.6 Section summary

*S. mansoni* faces multiple challenges during its infections in the mammalian host, and the interactions with the host environment is essential for parasite development and survival. The interactions could be manifested through many interfaces such as tegumental surface, gut lining surface, and multiple sources of ES. The parasites migrate through the blood circulation network to reach suitable sites for development and egg-laying. Throughout their life in the bloodstream, the parasites rely on their host for metabolic requirements that they obtain through the tegument and the gut. Although living in the bloodstream can provide ample metabolic resources, the parasites are also in close contact with circulating leukocytes and other protective systems of the host against pathogens. Multiple strategies have evolved in *S. mansoni* to exploit the molecular host-parasite interactions. Arguably, suppression of inflammation could be beneficial to both *S. mansoni* and its mammalian host because excessive responses to tissue damages and inflammation responses against pathogens may also harm the host. Many aspects of the mechanisms for successful infections have been explored, but they are not fully understood; for example, how the parasites navigate and anchor to suitable sites; how certain metabolites are obtained and become essential for different stages; and what other mechanisms are involved in the interfering with host defence systems. Recent development in genomic information of *S. mansoni* and other parasites now allow more information to be extracted.

1.4 *S. mansoni* genome and transcriptome

1.4.1 Genome and gene annotation

1.4.1.1 Genome and annotation

Studying a genome provides a broad picture of an organism and the mechanisms underlying its biology. Prior to the availability of *Schistosoma* genomes, gene discovery was done using Express Sequence Tags (ESTs) where mRNA fragments were cloned, sequenced and matched to databases of known genes (e.g. Verjovski-Almeida et al., 2003). As reviewed by Hoffmann & Dunne (2003), the approach led to improved details of molecular functions of the parasites and supports previous biochemical work. However, ESTs do not usually cover the whole length of genes, are expensive and labour-intensive, and many *S. mansoni* ESTs could not be matched
to known ESTs in databases, likely because of the specificity to parasite clades (Hoffmann and Dunne, 2003). The first *S. mansoni* draft genome was published in 2009 (Berriman *et al*., 2009). Later work on the *S. mansoni* genome dramatically improved its assembly and gene annotations giving an estimated genome size of 364.5 megabases (Protasio *et al*., 2012). This version is used in the analyses of this thesis. Presently, the genome of *S. mansoni* is undergoing further improvements which include incorporating new data from long-read technology to close gaps in the genome, and obtaining full-length transcripts from multiple stages of clonal infections (single-miracidium infection of snails) to aid gene finding. In the new assembly (unpublished; M Berriman, personal communication, 2017) the seven autosomes and single pair of sex chromosomes are close to completion (<300 gaps), with improved accuracy of repetitive regions, and clarification of many genes structures. In addition to the *S. mansoni* genome, genomes of other helminth species have become available as part of the 50 Helminth Genome Initiative (50 HGI, http://www.sanger.ac.uk/science/collaboration/50hgp). Even though many of these genomes have not been extensively curated, the current quality allows comparisons of genes across species, both parasitic and non parasitic, and provide additional information on gene homology which could help researchers infer functions of parasite genes (Howe *et al*., 2016a, 2016b).

Genomic information provides resource for investigating specific groups of genes. In the following sections, three major groups are covered that have potential roles in parasitism or host-parasite interactions. Moreover, these groups present a particular challenge to annotation efforts (due to their complex structures) and studies of them have substantially benefited from the availability of high quality genome data.

1.4.1.2 *S. mansoni* genome and GPCRs
Genomic sequences can be mined for members of an entire gene family. An example of a relevant protein family involved in host-parasite interaction is the G-protein coupled receptors (GPCRs). Prior to availability of the genome, the first GPCR characterised in schistosomes was identified from ESTs (Hoffmann *et al*., 2001). *S. mansoni* GPCRs are likely to be involved in multiple signalling pathways including chemosensory, regulation of physiological and developmental processes, and behavioural responses to stimuli (Chaisson and Hallem, 2012). In addition, GPCRs
are targets of multiple drugs (McVeigh et al., 2012a; Stevens et al., 2012). Multiple effort to mine GPCR-encoding genes from *S. mansoni* genomes have identified over 100 GPCRs (Berriman *et al*., 2009; Campos *et al*., 2014; Protasio *et al*., 2012; Zamanian *et al*., 2011). Some of these have been characterised for roles in parasite motility (e.g. MacDonald *et al*., 2015; Patocka *et al*., 2014). GPCRs for neurotransmitters and neuropeptides may regulate development and behaviours (Collins *et al*., 2010; Ribeiro and Patocka, 2013). Lastly, some GPCRs are localised to the tegument suggesting that their ligands are from external environment (e.g. Taman and Ribeiro, 2011).

### 1.4.1.3 *S. mansoni* genome and MEGs

Genomic data provide information on gene structures as well as their sequences. Micro-exon genes (MEGs) were identified in the genome of *S. mansoni* for their unique genic structure consisting of multiple small exons mainly comprising multiples (2-12) of three base pairs flanked at their 5’ and 3’ ends by longer exons of normal length (Berriman *et al*., 2009). Over 70 MEGs have been described and are currently grouped into families based on sequence similarity (Berriman *et al*., 2009; Chalmers *et al*., 2008; DeMarco *et al*., 2010).

MEGs have been postulated to have roles in host-parasites interactions owing to their presence being limited to parasitic flatworms (currently found in *Schistosoma* spp and tapeworm *Echinococcus*) (Tsai *et al*., 2013) and their expression patterns (different MEGs are expressed in specific stages of *S. mansoni* life cycle) (Parker-Manuel *et al*., 2011; Wilson, 2012). Moreover, with multiple exons and extensive alternative splicing, it has been proposed that proteins encoded by MEGs may contribute to antigenic polymorphism which could be a strategy for host immune evasion (Cai *et al*., 2016; DeMarco *et al*., 2010; Wilson, 2012). However, currently there is no report of host antibody recognition against MEGs; therefore, it is unlikely that antigenic polymorphism is a protective measure, for the protein products are not antigenic. However, some MEGs may, indeed, have roles in interactions with mammalian hosts (Philippsen *et al*., 2015; Wilson, 2012). Many MEGs are expressed in the oesophagus where many digestive enzymes are secreted, suggesting that MEGs may have roles in blood feeding or may be released into host environment (Li *et al*., 2014; Nawaratna *et al*., 2014). An example of this is a gene from the MEG-3 group.
secreted by skin stage and lung stage schistosomules that has been suggested to interact with endothelial cells (reviewed in Wilson, 2012). Additionally, MEG-14 is expressed in the oesophagus and its protein product binds to inflammatory protein preventing the protein from reaching the parasite gut (Orcia et al., 2016). Furthermore, a study of gene orthologues between Schistosoma species demonstrated that MEG-1, MEG-9, and MEG-15 may have been under selective pressure reflecting their possible role in host-parasite interactions (Philippson et al., 2015). Many MEGs have not been functionally characterised, and their amino acid sequences do not match any known protein domains (Berriman et al., 2009; DeMarco et al., 2010). However, some genes may have exonic structure resembling MEGs and also have exons encoding a conserved protein domain. An example of this is the venom allergen-like 6 gene (VAL6) which contains exons encoding a conserved CAP (cysteine-rich secretory protein/Antigen-5/Pathogenesis-related 1) domain adjacent to a long stretch of micro exons (Chalmers et al., 2008). Further investigation on this group of genes may provide insights into parasite-specific features and host-parasite interactions.

### 1.4.1.4 S. mansoni genome and hypothetical proteins

Regarding another parasite-specific feature, approximately 30% of S. mansoni genes (~3000 genes) are annotated as hypothetical proteins due to absence of sequence similarity to known proteins and protein domains (Berriman et al., 2009; Protasio et al., 2012). Given their uniqueness to S. mansoni (and possibly to other parasitic species), they may explain parasite-specific biology or interactions. Finer details of their requirement and putative functions could be inferred from studying their expression at specific stages or in specific conditions.

### 1.4.2 Transcriptomes

Prior to the availability of RNA-sequencing (RNA-seq) technology for quantification of gene expression, DNA microarrays were used in many studies of S. mansoni and have provided insights into key changes in life stages as well as effects of the environment on parasites, and have been used for identifying drug targets (Chai et al., 2006; Fitzpatrick et al., 2009; Gobert et al., 2009b; Jolly et al., 2007; Parker-Manuel et al., 2011). As RNA-seq becomes more widely adopted, it has contributed to improvement of genome annotation and provided evidence for gene finding (e.g.
Almeida et al., 2012; Anderson et al., 2015; Protasio et al., 2012) as well as a deeper understanding of parasite biology (e.g. Lu et al., 2016; Picard et al., 2016; Protasio et al., 2013). RNA-seq has multiple benefits over microarrays: first, it allows the inclusion of more genes including genes that are not previously identified at the time of microarray probe design (i.e. RNA-seq is an unbiased approach, not relying on having previous information about the genes); second, it is less prone to resolution problems for genes with low expression levels (signal on microarray cannot be separated from background noises) and genes with high expression level (signal become saturated); and third, microarrays are more prone to technical variations whilst RNA-seq is highly reproducible (Wang et al., 2009). Current development of the RNA-seq technology, focussing on sample preparation procedures and sequencing chemistries, is driving down the requirements for input materials, allowing previously less accessible life-cycle stages to be investigated. Furthermore, multiple samples can be pooled into one sequencing run (multiplexing) reducing costs and the technical variation between samples.

1.5 This thesis

In this thesis, I have employed RNA-seq approaches (mRNA only), together with genomic information from available databases, to interrogate host-parasite interactions in S. mansoni during intramammalian stages. The next chapter covers generic methods used for this thesis such as parasite maintenance, preparation of samples for RNA-seq, and details of data analysis tools. In chapter 3, I investigate transcriptomic profiles of in vivo S. mansoni obtained from infected mice over a timecourse, covering the intramammalian life cycle stages (i.e. lung, liver, and adult stages) allowing exploration of changes associated with key stages of the infection. In chapter 4, I use co-cultured mechanically transformed S. mansoni schistosomules in vitro with three human-derived cell types to study how the parasite respond, transcriptomically, to different host environments. Then in chapter 5, I use the co-cultured cells to investigate how the parasite influence host environments.
Chapter 2

Materials and methods

This chapter contains generic methods including parasite maintenance, molecular procedures, and bioinformatic analyses of sequence data. Details of experimental designs and sample collection specific to each of the chapters are covered in each method section.

2.1 Parasite materials

Parasite materials for the in vivo timecourse dataset were provided by Prof. Michael Doenhoff from the University of Nottingham and the methods are provided in chapter 3. Parasite materials for the in vitro experiment (chapter 4 and 5) were obtained from the Wellcome Trust Sanger Institute (WTSI). All procedures involving mice were performed by authorised personnel according to the UK Animals (Scientific Procedures) Act 1986. The life cycle maintenance was a joint effort between multiple teams and is described below.

2.1.1 Maintenance of the S. mansoni life cycle

The life cycle at WTSI was initially set up by seeding Biomphalaria glabrata snails infected with S. mansoni (Puerto Rico stain) provided by Prof. Michael J. Doenhoff (University of Nottingham, UK) and naive B. glabrata snails from the Department of Pathology, Cambridge, UK. The parasites were propagated between snails and C57BL/6 or BALB/c mice. Uninfected snails were kept as a breeding colony in a light and temperature controlled room with a 12/12 hour light-dark cycle and an ambient temperature of 28 °C. Younger snails with sizes between 3-5 mm were collected for infection with miracidia. To perform the infection, one snail was placed into each well of a 24-well plate and up to 30 miracidia, counted and collected under a dissecting microscope, were added to each well. The plates were left at 28 °C for an hour before the snails were placed in a new tank and kept in the same light-dark cycle as the breeding colony. Shedding of cercariae started after 3-4 weeks post-infection. The snails were checked for patent infection and transferred to a dark cupboard reserved
for *S. mansoni* infected snails to maximise shedding capacity and for health and safety reasons.

To collect cercariae, snails were pooled into beakers containing 1X conditioned aquarium water (Appendix A) and left under bright light between 40 minutes to 2 hours. Exposure of intact skin to the cercariae could lead to an infection; therefore, appropriate personal protective equipment including long-cuff gloves, a face shield, and a plastic apron was used while handling the parasites. Following shedding, approximately 250 cercariae per mouse were used for fortnightly infections of mice by intraperitoneal injection, or by subcutaneous infection through tails. After 7-8 weeks, mice were sacrificed and adult worms collected by portal perfusion. Livers from infected mice were processed for egg collection by mincing the livers or by digesting the livers overnight in a collagenase enzyme, and filtering for eggs. Eggs were hatched in sterile water and miracidia collected for infecting snails.

### 2.1.2 Snail husbandry

The infected snails were kept in an aquarium tank lined with a disposable plastic insert, filled with 1X condition aquarium water (Appendix A), aerated, fed three times a week with fish pellets, and kept in a dark cabinet. Uninfected *B. glabrata* were maintained as a breeding colony in the same light- and temperature-controlled room. Small snails (3-5 mm), suitable for infection with miracidia, were maintained separately from the large breeder snails in order to minimise loss during routine tank cleaning and for ease of collecting snails for miracidial infection. To separate breeders from their offspring, egg clutches were regularly removed and transferred to new tanks. I observed that the breeder snails preferentially laid eggs on silicone tubes compared to the surface of the tank. To simplify the removal of egg clutches, I therefore developed “The Octopus” apparatus, comprising a polymer platform holding a number of long silicone tubes hanging on top of the tank into the water (Figure 2.1).
Figure 2.1 Transfer of snail eggs using “The Octopus”
A) “The Octopus” holds up to eight silicone tubes which are submerged in water and provide a substrate for snail egg-laying. B) The silicone tubes with snail eggs are transferred to a new tank for hatching and cultivating juvenile snails.

2.1.3 Mouse maintenance
The Research Service Facility staff of the WTSI maintained mouse colonies. Mice infected with *S. mansoni* were provided with food and water *ab libitum* and were checked daily for general health.

2.2 Molecular methods

2.2.1 RNA extraction from parasites
RNA was extracted from parasite materials with a modified phenol-chloroform method followed by column purification. Frozen samples in TRIzol® reagent (15596-026, Invitrogen) were thawed on ice. Next, the samples were resuspended by gently pipetting and transferred to 2 ml tubes containing ceramic beads (MagNA Lyser Green Beads, Roche). The parasite materials were homogenised in a MagNA Lyser Instrument (FastPrep-24) at maximum speed for 20 seconds twice, with 1 minute rest on ice in between. After this, 200 µl of chloroform-isoamyl alcohol 24:1 was added to each tube, followed by vigorous shaking by hand for 5 seconds. The tubes were
centrifuged at 13,000 x g for 15 minutes at 4°C to separate the aqueous and organic solvent layers. The top aqueous layer containing RNA was carefully transferred into a new RNase-free 1.5 ml tube. To these tubes, an equal volume of 100% ethanol was added and mixed by pipetting. The mixture was then transferred to Zymo RNA Clean & Concentrator-5 column (R1015, Zymo Research) and processed according to the manufacturer's protocol. Finally, 15 µl of RNase-free water was added to the column and centrifuged for 30 seconds to elute the RNA. The elution was repeated once using an additional 15 µl of the RNase–free water.

2.2.2 RNA extraction from human cells

In addition to cells which were part of the co-culturing experiments, spare cells of each cell type were cultured and collected to use as test samples for RNA extraction. This was not required for parasite materials as standard protocols were available in our laboratory. The yield of RNA from the test extractions ranged between 2.7 to 34 µg which exceeded the recommended amount for the Zymo Clean & Concentrator kit (5 µg). Therefore, QIAGEN RNAeasy Plus Universal mini kit (QIAGEN RNAeasy Plus Universal mini kit, QIAGEN), recommended for the extraction of up to 100 µg of RNA was used. The extraction followed the manufacturer’s protocol with minor modifications. Two hundred µl of chloroform-isoamyl alcohol 24:1 was used instead of 180 µl, and 100% ethanol was used instead of 70%. The elution was done twice, each time in 30 µl nuclease-free water. Batches of 11-12 samples were processed.

2.2.3 Measurements of RNA concentration and purity

Extracted RNA was measured for its concentration and integrity using a Agilent RNA 6000 Nano kit (5067-1511, Agilent Technologies), and assessed for its purity using a NanoDrop spectrophotometer. The integrity of the RNA was inferred from the presence of the distinctive 18S rRNA peak for parasite samples, and two rRNA peaks (18S and 28S) for human samples; broad peaks indicate that the RNA is degraded. For parasite samples, an extra peak of 28S rRNA (approximately 4700 nucleotides (nt)) infer contamination with mouse tissue (Figure 2.2). RNA contaminated with mouse tissue was commonly observed for *S. mansoni* schistosomule samples collected at day 6 post-infection. The source of the mouse tissue contamination was an artefact of the collection method. Minced lung tissues were incubated and filtered to obtain lung-stage schistosomules. Some of the minced lung tissues passed through
the filter and was collected with the schistosomules. Detailed methods for the
collection of lung schistosomules are presented in chapter 3. The purity of RNA was
determined from the 260/230 and 260/280 ratios measured by the Nanodrop
spectrophotometer. Ratios of ~2.0 indicated pure RNA and lower ratios suggested
contamination with chemical or biological molecules.

Figure 2.2 RNA qualities assessed from Agilent Bioanalyzer electropherograms
A) Electropherogram from Agilent Bioanalyzer output showing good quality RNA from S.
mansoni with little or no degradation inferred from a shape peak at approximately 4000 nt
representing schistosome 18S and nicked 28S ribosomal RNA (Tenniswood and Simpson,
1982). B) Good quality RNA from a human sample. C) An example of degraded RNA
inferred from the absence of a sharp rRNA peak and the shift of RNA abundance towards
shorter length. This example is from human RNA in a separate experiment. D) Good
quality RNA from S. mansoni but with contamination of mouse RNA.

2.2.4 Library production and sequencing

All RNA-seq libraries were produced by the library production team at WTSI.
Following quantity and quality checking, variable amounts of total RNA were sent for
library preparation. For chapter 3, up to 1 ug of RNA was used; for chapter 4, the entire RNA sample was used (range between 319 to 4,350 ng); and for chapter 5, 500 ng of RNA was required.

The libraries were produced using the Illumina ®TruSeq™ Stranded RNA Sample Preparation v2 Kits (RS-122-2101 and RS-122-2102, Illumina). Briefly, mRNA was pulled down from total RNA using oligo-dT beads. The mRNA was then fragmented into 200-300 base pairs fragments, and reverse-transcribed into cDNA using random hexamer primers and free nucleotides. The second strand synthesis was performed similarly but replacing dTTP with dUTP, which is essential to mark the second strand for quenching during the amplification. To improve the binding efficiency of an adapter with a T-overhang, an additional dATP was incorporated to the 3’ end of both strands. Different adapter index sequences were incorporated into samples to allow multiplexing of samples. After adapter ligation, a PCR reaction was performed with primers specific to the adapter sequences that also contain the Illumina adapter fork region. The PCR reaction was completed for 10-14 cycles followed by library clean up using Agencourt AMPure XP Beads (A63881, Beckman Coulter). The libraries were then quantified by qPCR and a suitable amount was loaded onto a sequencing lane.

The sequencing was performed at the WTSI by the sequencing facilities on an Illumina HiSeq 2500 platform using either rapid run (chapter 5) or normal run (chapter 3 and 4) modes. All sequencing data was produced as 75 bp paired-end reads.

2.2.5 Overall QC of sequencing outcome

Read outputs from the sequencing step were quality assessed using various parameters in an automatic standard pipeline managed by core sequencing informatics (NPG team), WTSI. Of relevance to this thesis is the mapping of reads to a set of reference genomes. A small randomly selected subset of reads were mapped to reference genomes from model organisms, pathogens, and other microbes. List of the genomes is available in Appendix B. These data were used to indicate possible contamination in the samples.
2.3 Data analysis

2.3.1 Mapping and quantifying read counts

2.3.1.1 Schistosome reads

Schistosome reads were mapped to the *S. mansoni* genome version 5 (GeneDB, Logan-Klumpler *et al.*, 2012) using TopHat (*Kim et al.*, 2013, version 2.0.8) with default parameters except the following: -g 1 (only report 1 alignment for each read); --library-type fr-firststrand (for dUTP Illumina library); -a 6 (minimum anchor length); -i 10 and --min-segment-intron 10 (minimum intron length); -I 40000 and --max-segment-intron 40000 (maximum intron length); --microexon-search (find alignment to micro-exons). The resulting BAM files of accepted hits were sorted using SAMtools (by read names; -n option), indexed, and used to obtain read counts per gene. A GFF file containing gene annotations and their genomic coordinates was downloaded from GeneDB.org and filtered to keep only the longest transcript for each gene. The number of read counts per gene was calculated using HT-Seq (*Anders et al.*, 2015, version 0.7.1) based on the GFF file. Next, the GFF file and sorted BAM files of mapped sequencing reads were used as inputs for HTSeq to obtain read counts per gene and used for analysis in R. HT-seq was run with default parameters except with strand option set to suit dUTP libraries (-s reverse), and alignment score cut-off increased (-a 30).

2.3.1.2 Human reads

In order to reduce computing power required for the mapping step, human reads were mapped to a reference transcriptome instead of to a reference genome (transcriptome version GRCh38, release 88 (Ensembl, Aken *et al.*, 2016)). The mapping and quantification of reads per transcripts was performed using Kallisto (*Bray et al.*, 2016, version 0.42.3). Kallisto uses pseudoalignment which match k-mers from each sequence read to reference transcripts, bypassing alignment by individual bases and by doing so, reducing computational power and time required. The output is a table of reads per transcript. Read counts per transcript were converted into read counts per gene by matching reference identifiers from Ensembl (Aken *et al.*, 2016). The read counts per gene were used as an input for gene expression analysis in R software environment (R Core Team, 2016).
2.3.2 Read count and differential expression analysis

Analyses were performed using RStudio version 0.99.489 (RStudio Team, 2016), with R version 3.3.1. Versions of R packages used for analyses and data visualisation are listed in Appendix C. DESeq2 (Love et al., 2014) was used to import read counts into the analysis environment (function: ‘DESeqDataSetFromHTSeqCount’), to investigate overall transcriptomic differences between samples using principal component analysis (PCA) (function: ‘plotPCA’), and to identify differentially expressed genes in the timecourse and in pair-wise comparison. PCA used regularized log–transformed read count data as input. Read counts were normalised based on negative binomial distribution and scale factors. Differential expression analyses were performed with either likelihood-ratio tests (when the whole timecourse was considered) or with the Wald test (when used with pairwise comparisons) and returned adjusted p-values according to the Benjamini–Hochberg procedure to control false discovery rate (Benjamini et al., 2001). Differentially expressed genes were defined as those with adjusted p-value < 0.01 and log₂ fold change (log₂FC) in expression > 1 or < -1 (i.e. 2-fold change) for chapter 3, and 5; or adjusted p-value < 0.01 and log₂FC in expression > 0.5 or < -0.5 (i.e. 1.4-fold change) for chapter 4.

2.3.3 Genes clustering by timecourse expression profile

Genes were clustered based on their expression profiles over the timecourse using self-organising maps constructed in the R package Kohonen (Wehrens and Buydens, 2007). The recommended inputs of mean-normalised, regularized log-transformed counts (rlog-transformed) were used for the self-organising maps. Rlog transformation is a robust transformation for stabilizing variances within genes especially when library sizes (size factor) vary between samples (Love et al., 2014). The library sizes between some samples varied up to 10-fold. To mean-normalise the regularized log-transformed counts, the mean value of a row was subtracted from each value within that row so that changes in expression fluctuated around zero and clustered based on expression pattern rather than absolute expression level. This normalised value was used to calculate means of replicates for each gene at each time point and used as input for clustering. Genes were grouped based on their expression pattern into 96 clusters (user defined). To reduce noise, only genes that were differentially expressed in at least one time point (likelihood ratio test, adjusted p-
value < 0.01) were used as inputs for clustering. Self-organising map outputs from the package Kohonen provide representative values (codebook vector) for each cluster which represent changes over time of genes in the cluster. This information was used to produce a dendrogram so that similarity between clusters could be used to find genes with similar expression patterns over the time points. Hierarchical clustering was used to group clusters based on their representative values.

2.3.4 GO term enrichment

Gene Ontology (GO) term enrichment (biological process terms) was performed using the TopGO R package (Alexa and Rahnenfuhrer, 2016). The test used a weight algorithm and Fisher’s exact test statistic to identify enrichment of GO terms among the input genes. To determine enrichment, all S. mansoni genes were used as a reference background. For the human dataset, expressed genes were used as a reference background. This was because each of the human cell types showed very distinct transcriptomic profiles. Separate lists of expressed genes were produced for each of the three cell types used. Expressed genes were genes that have FPKM > 0 in at least one replicate in a particular cell type. FPKM is Fragments Per Kilobase of transcript per Million mapped reads, and was calculated as

\[
\frac{\text{number of reads of a gene}}{\left(\frac{\text{transcript length}}{1000}\right) \times \left(\frac{\text{total number of reads in the library}}{1,000,000}\right)}
\]

GO annotations for S. mansoni were downloaded from GeneDB (January 2016). For human, GOslim annotations were downloaded from Ensembl GRCh38, release 88 (Ensembl, Aken et al., 2016) and supplemented with genes annotated for innate immune function from InnateDB (Breuer et al., 2013). To do this, GO:0002376 (immune system process) was added to the genes annotated with innate immune functions by InnateDB.

2.3.5 Pathway enrichment and pathway network

Pathway enrichment was used to provide another layer of specificity for deducing functional changes. Human transcriptome analysis benefits from richer pathway information and supporting evidence that is available due to the large size of the human-research community. Therefore, I used pathway databases for human
transcriptome analysis but not for schistosome transcriptome analysis because pathway descriptions in schistosomes primarily rely on homology based mapping to infer pathways from other species (Fabregat et al., 2016; Kanehisa et al., 2017a; Mi et al., 2017). It cannot necessarily be assumed that pathways in parasites follow those of better-characterised model species.

Pathway enrichment analysis were generated through the used of three databases - Reactome, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Ingenuity Pathways Analysis (IPA) (Fabregat et al., 2016; Ingenuity® Systems; Kanehisa et al., 2017b) Knowledge of human functional genomics, although growing rapidly, is still limited in scope. Hence combining information from various updated sources can provide further affirmation. The InnateDB online server was used for pathway analysis with the KEGG database. Reactome and IPA pathway analyses used their respective interfaces. Input to the analysis was a list of differentially expressed genes from the human dataset, given as Ensembl gene identifiers.

Many pathways from within the same database and between databases often shared similar genes, and pathways are sometimes named differently between databases. To incorporate this information, the enrichment outputs from the three databases were integrated by drawing a network of enriched pathways, linked by shared genes. To do this, each enriched pathway was compared to every other enriched pathway, and if there was at least one gene (differentially expressed input gene) in common between them, then the two pathways were linked on the network. The network was visualised on Cytoscape (Cline et al., 2007). Each network node represents a pathway, node colours represent sources of the enrichment analysis. Having edges between nodes indicated that the nodes shared at least one gene. A small distance between nodes indicated a large number of shared genes.

2.3.6 Pathway comparison between cell types

Changes of gene expression in two pathways were compared between cell types. The two pathways were chosen because of their occurrence in multiple pathway enrichment analyses, and their relevance to *S. mansoni* infection biology. These were the *extracellular matrix organisation* pathway and *coagulation and complement* pathway. The list of genes in the *extracellular matrix organisation* pathway was
obtained from the Reactome database (pathway identifier: R-HSA-1474244). Genes in coagulation and complement pathways were obtained from KEGG (pathway identifier: hsa04610). Gene names from the database were converted to ENSG identifiers. Changes in gene expression for each cell type were compared based on log$_2$FC of co-cultured vs. worm-free pairwise comparison. Only genes that were expressed (FPKM > 0) in at least one replicate for each cell type were included in the comparison. Log$_2$FC values were displayed only for genes with an adjusted p value of less than 0.01.

2.3.7 Protein structural prediction

I-TASSER online server (v5.0) (Roy et al., 2010; Yang et al., 2015; Zhang, 2008) was used to predict protein 3D structure from amino acid sequences, using default parameters. Amino acid sequences were obtained from GeneDB in November 2015. TM-scores indicate similarity between two structures. The value ranges between 0-1 with a higher value inferring better match.

2.3.8 Protein domains and motif search

InterProScan online server (v60 and v61) (Finn et al., 2017) was used to identify protein domains from amino acid sequences. Amino acid sequences were obtained from GeneDB in December 2015. SignalP 4.1 (Petersen et al., 2011) was used to resolve conflicts between transmembrane and signal peptide predictions from InterProScan (using Phobius predictor; Käll et al., 2004). MyHits (Pagni et al., 2004) and ScanProsite (de Castro et al., 2006) were used to identify functional motifs in amino acid sequences. CathDB (Sillitoe et al., 2015) was used to explore protein structural domains and to search by structural match.

2.3.9 Gene phylogenetic tree

Information on homologues, orthologues and paralogues of genes was obtained from WormBase ParaSite release 8 and 9 (Howe et al., 2016a), which uses the Compara pipeline to create gene trees (Vilella et al., 2009).

2.3.10 Artemis and BamView

Mapping of RNA-seq reads to genomic regions was visualised using BamView (Carver et al., 2010) within Artemis (Carver et al., 2008) to assess the accuracy of
gene models. Short read mappings came from *S. mansoni* data produced in this thesis. Long-read mappings (ISO-seq, or isoform sequencing) were obtained from a separate project of the Parasite Genomics team. In the project, *S. mansoni* RNA were processed into full-length RNA libraries and sequenced on a PacBio platform (Rhoads and Au, 2015) to obtain long reads covering full-length isoforms. Screenshots of long-read mapping were provided by Alan Tracey, Parasite Genomics team.

2.3.11 **Seaview sequence alignment**

Multiple sequence alignment was performed using SeaView version 4 (Gouy et al., 2010) with ClustalW alignment program (Sievers et al., 2014). Inputs were amino acid sequences from the reference *S. mansoni* genome (obtained December 2015) or from GenBank accession identifiers resulting from Basic Local Alignment Search Tool (BLAST).

2.4 **Data presentation**

R package ggplot2 (Wickham, 2009) was used to produce most plots and graphs in this chapter including PCA plots, volcano plots, gene expression dot plots, and graphical representation of GO term enrichment. PCA plots and gene expression dot plots used output from DESeq2 ‘plotPCA’ and normalised counts respectively. Volcano plots used pairwise differential expression as input. Gene clustering expression graphs used generic functions in R. Clustering dendrograms were created using R package APE (Paradis et al., 2004) using hierarchical clustering of cluster representative values (codebook vector). Heatmaps were generated with R package pheatmap (Kolde, 2015) using rlog transformed values of genes as input.
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). Percutaneous 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.

<table>
<thead>
<tr>
<th>Day 0 Number of cercariae</th>
<th>Day 6</th>
<th>Day 13</th>
<th>Day 17</th>
<th>Day 21</th>
<th>Day 28</th>
<th>Day 35</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>2000</td>
<td>500</td>
<td>500</td>
<td>350</td>
<td>300</td>
<td></td>
</tr>
</tbody>
</table>

*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 °C 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 mm³ 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 components). The lungs from each mouse were incubated separately in individual Falcon tubes for ~1 hour at room temperature, followed by 37 °C 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 °C 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 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

<table>
<thead>
<tr>
<th>Time point</th>
<th>Average counts between all replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 6</td>
<td>520214</td>
</tr>
<tr>
<td>Day 13</td>
<td>7599869</td>
</tr>
<tr>
<td>Day 17</td>
<td>5938978</td>
</tr>
<tr>
<td>Day 21</td>
<td>4828234</td>
</tr>
<tr>
<td>Day 28</td>
<td>5938633</td>
</tr>
<tr>
<td>Day 35</td>
<td>5010390</td>
</tr>
</tbody>
</table>

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) contain 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 suggest that developmental processes were initiated at the lung stage even though mitosis has not occurred (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)

<table>
<thead>
<tr>
<th>Gene identifier</th>
<th>Product name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smp_032970</td>
<td>calmodulin protein (calcium binding protein)</td>
</tr>
<tr>
<td>Smp_138070</td>
<td>MEG-3 (Grail) family</td>
</tr>
<tr>
<td>Smp_138080</td>
<td>MEG-3 (Grail) family</td>
</tr>
<tr>
<td>Smp_149860</td>
<td>cat eye syndrome critical region protein 5</td>
</tr>
<tr>
<td>Smp_149870</td>
<td>cat eye syndrome critical region protein</td>
</tr>
<tr>
<td>Smp_159800</td>
<td>MEG-2 (ESP15) family</td>
</tr>
<tr>
<td>Smp_159810</td>
<td>MEG-2 (ESP15) family</td>
</tr>
<tr>
<td>Smp_172460</td>
<td>Krueppel factor 10 like</td>
</tr>
</tbody>
</table>
Table 3.3 Enriched GO terms of genes down-regulated after lung stage

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

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

<table>
<thead>
<tr>
<th>Gene identifier</th>
<th>Log₂FC (day 13/day 6)</th>
<th>Adjusted p-value</th>
<th>Product name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smp_138080</td>
<td>-12.62</td>
<td>1.48E-19</td>
<td>MEG-3 (Grail) family</td>
</tr>
<tr>
<td>Smp_138070</td>
<td>-11.99</td>
<td>3.58E-30</td>
<td>MEG-3 (Grail) family</td>
</tr>
<tr>
<td>Smp_159810</td>
<td>-11.22</td>
<td>1.02E-48</td>
<td>MEG-2 (ESP15) family</td>
</tr>
<tr>
<td>Smp_159800</td>
<td>-9.66</td>
<td>5.10E-28</td>
<td>MEG-2 (ESP15) family</td>
</tr>
<tr>
<td>Smp_181510</td>
<td>-9.58</td>
<td>1.15E-13</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_032990</td>
<td>-8.98</td>
<td>7.14E-13</td>
<td>Calmodulin 4 (Calcium binding protein Dd112)</td>
</tr>
<tr>
<td>Smp_159830</td>
<td>-8.69</td>
<td>1.00E-05</td>
<td>MEG-2 (ESP15) family</td>
</tr>
<tr>
<td>Smp_138060</td>
<td>-8.06</td>
<td>3.76E-09</td>
<td>MEG-3 (Grail) family</td>
</tr>
<tr>
<td>Smp_203400</td>
<td>-7.34</td>
<td>1.88E-07</td>
<td>rhodopsin orphan GPCR</td>
</tr>
<tr>
<td>Smp_005470</td>
<td>-7.31</td>
<td>3.34E-08</td>
<td>dynein light chain</td>
</tr>
<tr>
<td>Smp_077610</td>
<td>-6.61</td>
<td>6.19E-07</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_166350</td>
<td>-6.59</td>
<td>1.31E-06</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_180330</td>
<td>-5.97</td>
<td>2.91E-32</td>
<td>MEG 2 (ESP15) family</td>
</tr>
<tr>
<td>Smp_205660</td>
<td>-5.84</td>
<td>4.52E-15</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_033250</td>
<td>-5.80</td>
<td>4.43E-05</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_132500</td>
<td>-5.73</td>
<td>4.50E-04</td>
<td>ras and EF hand domain containing protein</td>
</tr>
<tr>
<td>Smp_152730</td>
<td>-5.68</td>
<td>3.41E-11</td>
<td>histone lysine N methyltransferase MLL3</td>
</tr>
<tr>
<td>Smp_241430</td>
<td>-5.61</td>
<td>3.89E-23</td>
<td>Aquaporin 12A</td>
</tr>
<tr>
<td>Smp_125060</td>
<td>-5.55</td>
<td>6.70E-04</td>
<td>kinase suppressor of Ras (KSR)</td>
</tr>
<tr>
<td>Smp_198060</td>
<td>-5.51</td>
<td>4.73E-13</td>
<td>hypothetical protein</td>
</tr>
</tbody>
</table>
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, homophilic 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](image)

**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 (log₁₀ 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 O₂ 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 O₂ or into H₂O₂ 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 H₂O₂ (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.,
Other genes in cluster 72 are not as noticeable for their roles in host immune evasion.

Table 3.5 Genes in cluster 72

<table>
<thead>
<tr>
<th>Gene identifier</th>
<th>Product name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smp_059480</td>
<td>thioredoxin peroxidase</td>
</tr>
<tr>
<td>Smp_059980</td>
<td>arginase</td>
</tr>
<tr>
<td>Smp_074570</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_114660</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_134870</td>
<td>early growth response protein</td>
</tr>
<tr>
<td>Smp_147730</td>
<td>single kunitz protease inhibitor; serine type protease inhibitor</td>
</tr>
<tr>
<td>Smp_156510</td>
<td>PDZ and LIM domain protein 7</td>
</tr>
<tr>
<td>Smp_166920</td>
<td>PDZ and LIM domain protein Zasp</td>
</tr>
<tr>
<td>Smp_174810</td>
<td>Extracellular superoxide dismutase (Cu Zn)</td>
</tr>
<tr>
<td>Smp_182770</td>
<td>hypothetical protein</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Gene identifier</th>
<th>Log₂FC (day 13/day 6)</th>
<th>Adjusted p-value</th>
<th>Product name</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top 20 genes up-regulated in day 13</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smp_123200</td>
<td>7.35</td>
<td>1.17E-24</td>
<td>hypothetical protein (now as MEG-32.2)</td>
</tr>
<tr>
<td>Smp_154180</td>
<td>5.59</td>
<td>9.20E-13</td>
<td>25 kDa integral membrane protein</td>
</tr>
<tr>
<td>Smp_113760</td>
<td>5.21</td>
<td>7.69E-14</td>
<td>anti-inflammatory protein 16</td>
</tr>
<tr>
<td>Smp_087310</td>
<td>5.08</td>
<td>2.03E-11</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_060220</td>
<td>5.05</td>
<td>1.01E-10</td>
<td>lipopolysaccharide induced tumor necrosis</td>
</tr>
<tr>
<td>Smp_171130</td>
<td>4.9</td>
<td>2.19E-09</td>
<td>glial cells missing</td>
</tr>
<tr>
<td>Smp_158480</td>
<td>4.78</td>
<td>1.84E-17</td>
<td>AMP dependent ligase</td>
</tr>
<tr>
<td>Smp_045470</td>
<td>4.74</td>
<td>7.99E-16</td>
<td>prospero homeobox protein 2</td>
</tr>
<tr>
<td>Smp_127270</td>
<td>4.51</td>
<td>1.11E-08</td>
<td>Cytochrome b561:ferric reductase</td>
</tr>
<tr>
<td>Smp_161310</td>
<td>4.5</td>
<td>8.60E-11</td>
<td>IQ domain containing protein D</td>
</tr>
<tr>
<td>Smp_162370</td>
<td>4.48</td>
<td>1.42E-08</td>
<td>histone H1 gamma</td>
</tr>
<tr>
<td>Smp_086530</td>
<td>4.34</td>
<td>1.80E-09</td>
<td>tegument-allergen-like protein</td>
</tr>
<tr>
<td>Smp_123790</td>
<td>3.95</td>
<td>9.90E-07</td>
<td>glypican</td>
</tr>
<tr>
<td>Smp_011030</td>
<td>3.84</td>
<td>3.66E-08</td>
<td>heme binding protein 2</td>
</tr>
<tr>
<td>Smp_099670</td>
<td>3.82</td>
<td>1.02E-06</td>
<td>lipoprotein receptor</td>
</tr>
<tr>
<td>Smp_162860</td>
<td>3.69</td>
<td>5.95E-23</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_147890</td>
<td>3.61</td>
<td>3.57E-11</td>
<td>Rootletin</td>
</tr>
<tr>
<td>Smp_125250</td>
<td>3.59</td>
<td>9.97E-23</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_145760</td>
<td>3.57</td>
<td>2.89E-05</td>
<td>transmembrane protein 62</td>
</tr>
<tr>
<td>Smp_200900</td>
<td>3.48</td>
<td>2.15E-05</td>
<td>hypothetical protein</td>
</tr>
</tbody>
</table>
**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

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

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>
<thead>
<tr>
<th>Gene identifier</th>
<th>Log2FC (day 28/day 21)</th>
<th>Adjusted p-value</th>
<th>Product name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smp_050270</td>
<td>12.53</td>
<td>4.47E-23</td>
<td>Tyrosinase</td>
</tr>
<tr>
<td>Smp_000270</td>
<td>10.47</td>
<td>2.92E-16</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_138570</td>
<td>10.04</td>
<td>2.07E-41</td>
<td>spore germination protein</td>
</tr>
<tr>
<td>Smp_095980</td>
<td>9.72</td>
<td>1.46E-30</td>
<td>Extracellular superoxide dismutase (Cu Zn)</td>
</tr>
<tr>
<td>Smp_000430</td>
<td>9.36</td>
<td>2.86E-41</td>
<td>eggshell protein</td>
</tr>
<tr>
<td>Smp_191910</td>
<td>9.33</td>
<td>1.93E-28</td>
<td>Stress protein DDR48</td>
</tr>
<tr>
<td>Smp_077890</td>
<td>9.31</td>
<td>1.08E-17</td>
<td>Trematode Eggshell Synthesis domain containing protein</td>
</tr>
<tr>
<td>Smp_131110</td>
<td>9.22</td>
<td>3.63E-35</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_077920</td>
<td>9.20</td>
<td>2.14E-16</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_152150</td>
<td>9.16</td>
<td>8.02E-13</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_014610</td>
<td>8.99</td>
<td>1.18E-52</td>
<td>serine:threonine kinase 1</td>
</tr>
<tr>
<td>Smp_033250</td>
<td>8.99</td>
<td>7.77E-39</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_000420</td>
<td>8.71</td>
<td>1.74E-34</td>
<td>Pro His rich protein</td>
</tr>
<tr>
<td>Smp_165360</td>
<td>8.65</td>
<td>1.13E-30</td>
<td>histone acetyltransferase myst4</td>
</tr>
<tr>
<td>Smp_000280</td>
<td>8.64</td>
<td>4.27E-40</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_000390</td>
<td>8.47</td>
<td>2.26E-27</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_155310</td>
<td>8.39</td>
<td>3.90E-32</td>
<td>tetraspan CD63 receptor</td>
</tr>
<tr>
<td>Smp_144440</td>
<td>8.32</td>
<td>1.75E-22</td>
<td>replication A protein</td>
</tr>
<tr>
<td>Smp_130970</td>
<td>8.12</td>
<td>3.55E-10</td>
<td>G2:mitotic specific cyclin B3</td>
</tr>
<tr>
<td>Smp_000400</td>
<td>7.99</td>
<td>1.85E-12</td>
<td>hypothetical protein</td>
</tr>
</tbody>
</table>
Table 3.9 Enriched GO terms (biological process) of genes up-regulated in day-28 compared to day-21 worms

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

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.
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 3.16 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.
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 represents 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 log2FC 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

<table>
<thead>
<tr>
<th>Gene identifier</th>
<th>Product name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smp_000350</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_000755</td>
<td>family M13 non peptidase ue (M13 family)</td>
</tr>
<tr>
<td>Smp_001000</td>
<td>universal stress protein</td>
</tr>
<tr>
<td>Smp_015630</td>
<td>Glutamate-gated chloride channel subunit 2 isoform 1</td>
</tr>
<tr>
<td>Smp_016230</td>
<td>protocadherin 17</td>
</tr>
<tr>
<td>Smp_025030</td>
<td>dual specificity protein kinase</td>
</tr>
<tr>
<td>Smp_033000</td>
<td>calcium-binding protein</td>
</tr>
<tr>
<td>Smp_033010</td>
<td>16 kda calcium binding protein</td>
</tr>
<tr>
<td>Smp_034610</td>
<td>innexin</td>
</tr>
<tr>
<td>Smp_041700</td>
<td>G protein coupled receptor fragment</td>
</tr>
<tr>
<td>Smp_052660</td>
<td>metabotropic glutamate receptor 3</td>
</tr>
<tr>
<td>Smp_054170</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_057530</td>
<td>subfamily S9B unassigned peptidase (S09 family)</td>
</tr>
<tr>
<td>Smp_059570</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_068500</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_076030</td>
<td>tob protein; protein tob btg</td>
</tr>
<tr>
<td>Smp_078230</td>
<td>cGMP dependent protein kinase</td>
</tr>
<tr>
<td>Smp_119260</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_122870</td>
<td>regulator of G protein signaling</td>
</tr>
<tr>
<td>Smp_122880</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_124520</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_126500</td>
<td>tensin C1 domain containing phosphatase</td>
</tr>
<tr>
<td>Smp_126760</td>
<td>Proteasome activator complex subunit 4</td>
</tr>
<tr>
<td>Smp_127420</td>
<td>5' AMP activated protein kinase subunit gamma</td>
</tr>
<tr>
<td>Smp_136030</td>
<td>anion exchange protein</td>
</tr>
<tr>
<td>Smp_141070</td>
<td>glutamate receptor interacting protein 1</td>
</tr>
<tr>
<td>Smp_141980</td>
<td>cAMP specific 3'</td>
</tr>
</tbody>
</table>
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$_2$FC 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$_2$FC obtained from the published data. Column names denote pairwise comparison underlying each log$_2$FC. 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 tempting 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.
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 effects 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
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 °C 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 amplified 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 4°C 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 4°C. 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 °C 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% CO₂, 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 haemosporidin 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 log2FC 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.

![Bar chart of GO term enrichment](image)

**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 (log₂FC 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

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

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 preceding 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>
<thead>
<tr>
<th>Description</th>
<th>Query cover</th>
<th>E value</th>
<th>Ident</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>hypothetical protein Smp_052880 [Schistosoma mansoni]</td>
<td>99%</td>
<td>6.00E-75</td>
<td>100%</td>
<td>XP_018650687.1</td>
</tr>
<tr>
<td>hypothetical protein MS3_07516 [Schistosoma haematobium]</td>
<td>99%</td>
<td>5.00E-53</td>
<td>76%</td>
<td>XP_012798865.1</td>
</tr>
<tr>
<td>hypothetical protein [Schistosoma japonicum]</td>
<td>97%</td>
<td>8.00E-52</td>
<td>78%</td>
<td>CAX75147.1</td>
</tr>
<tr>
<td>hypothetical protein [Schistosoma japonicum]</td>
<td>97%</td>
<td>3.00E-51</td>
<td>77%</td>
<td>CAX75150.1</td>
</tr>
<tr>
<td>TPA: neuropeptide precursor-5 [Schmidtea mediterranea]</td>
<td>40%</td>
<td>0.005</td>
<td>39%</td>
<td>DAA33904.1</td>
</tr>
<tr>
<td>hypothetical protein CLF_105825 [Clonorchis sinensis]</td>
<td>24%</td>
<td>0.077</td>
<td>43%</td>
<td>GAA51282.1</td>
</tr>
<tr>
<td>hypothetical protein T265_07707 [Opisthorchis viverrini]</td>
<td>54%</td>
<td>0.21</td>
<td>31%</td>
<td>XP_009171559.1</td>
</tr>
<tr>
<td>neuropeptide-25 [Dugesia japonica]</td>
<td>30%</td>
<td>0.29</td>
<td>43%</td>
<td>BAV14854.1</td>
</tr>
<tr>
<td>Piso0_003363 [Millerozyma farinosa CBS 7064]</td>
<td>47%</td>
<td>4.2</td>
<td>37%</td>
<td>XP_004200765.1</td>
</tr>
</tbody>
</table>
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, log₂FC > 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<sub>2</sub>FC 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

<table>
<thead>
<tr>
<th>Gene identifier</th>
<th>Log2FC (HEPG2/non-HEPG2)</th>
<th>Adjusted p-value</th>
<th>Product name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smp_049850</td>
<td>1.01</td>
<td>5.22E-11</td>
<td>hypoxanthine guanine phosphoribosyltransferase</td>
</tr>
<tr>
<td>Smp_181510</td>
<td>0.96</td>
<td>1.00E-09</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_159810</td>
<td>0.94</td>
<td>2.75E-09</td>
<td>MEG-2 (ESP15) family</td>
</tr>
<tr>
<td>Smp_187410</td>
<td>0.90</td>
<td>1.06E-08</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_180620</td>
<td>0.90</td>
<td>1.84E-08</td>
<td>MEG 17</td>
</tr>
<tr>
<td>Smp_074560</td>
<td>0.89</td>
<td>3.05E-08</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_212180</td>
<td>0.88</td>
<td>1.43E-11</td>
<td>glucose dehydrogenase (acceptor)</td>
</tr>
<tr>
<td>Smp_180330</td>
<td>0.87</td>
<td>2.25E-10</td>
<td>MEG 2 (ESP15) family</td>
</tr>
<tr>
<td>Smp_126880</td>
<td>0.82</td>
<td>3.12E-07</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_136890</td>
<td>0.82</td>
<td>1.55E-07</td>
<td>universal stress protein</td>
</tr>
<tr>
<td>Smp_147740</td>
<td>0.82</td>
<td>1.33E-07</td>
<td>family M13 unassigned peptidase (M13 family)</td>
</tr>
<tr>
<td>Smp_138070</td>
<td>0.78</td>
<td>9.78E-07</td>
<td>MEG-3 (Grail) family</td>
</tr>
<tr>
<td>Smp_167120</td>
<td>0.76</td>
<td>3.94E-06</td>
<td>Peptidase M8</td>
</tr>
<tr>
<td>Smp_047680</td>
<td>0.74</td>
<td>1.32E-06</td>
<td>Ferritin-2 heavy chain</td>
</tr>
<tr>
<td>Smp_047660</td>
<td>0.72</td>
<td>9.87E-06</td>
<td>Ferritin-2 heavy chain</td>
</tr>
<tr>
<td>Smp_138080</td>
<td>0.70</td>
<td>2.17E-05</td>
<td>MEG-3 (Grail) family</td>
</tr>
<tr>
<td>Smp_193400</td>
<td>0.70</td>
<td>1.19E-05</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_062900</td>
<td>0.70</td>
<td>5.39E-07</td>
<td>tryparedoxin peroxidase</td>
</tr>
<tr>
<td>Smp_074570</td>
<td>0.69</td>
<td>1.43E-05</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_148820</td>
<td>0.67</td>
<td>2.51E-07</td>
<td>hypoxanthine guanine phosphoribosyltransferase</td>
</tr>
</tbody>
</table>
Table 4.4 Top 20 genes down-regulated in HEPG-2 compared to non-HEPG2 schistosomules

<table>
<thead>
<tr>
<th>Gene identifier</th>
<th>Log₂FC (HEPG2/non-HEPG2)</th>
<th>Adjusted p-value</th>
<th>Product name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smp_108550</td>
<td>-1.04</td>
<td>4.79E-16</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_045200</td>
<td>-0.87</td>
<td>2.61E-18</td>
<td>tegument-allergen-like protein</td>
</tr>
<tr>
<td>Smp_024180</td>
<td>-0.87</td>
<td>9.26E-11</td>
<td>placenta specific gene 8 protein</td>
</tr>
<tr>
<td>Smp_158480</td>
<td>-0.86</td>
<td>4.33E-10</td>
<td>AMP dependent ligase</td>
</tr>
<tr>
<td>Smp_166020</td>
<td>-0.84</td>
<td>4.56E-17</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_162500</td>
<td>-0.83</td>
<td>6.39E-10</td>
<td>drug efflux protein</td>
</tr>
<tr>
<td>Smp_010770</td>
<td>-0.81</td>
<td>6.39E-10</td>
<td>elongation of very long chain fatty acids</td>
</tr>
<tr>
<td>Smp_102190</td>
<td>-0.80</td>
<td>8.68E-10</td>
<td>steroid dehydrogenase</td>
</tr>
<tr>
<td>Smp_126290</td>
<td>-0.79</td>
<td>5.18E-07</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_175290</td>
<td>-0.79</td>
<td>6.72E-07</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_170280</td>
<td>-0.77</td>
<td>5.79E-23</td>
<td>integrin alpha ps</td>
</tr>
<tr>
<td>Smp_158510</td>
<td>-0.77</td>
<td>2.26E-14</td>
<td>diacylglycerol O-acyltransferase 1</td>
</tr>
<tr>
<td>Smp_175300</td>
<td>-0.75</td>
<td>3.86E-06</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_150640</td>
<td>-0.69</td>
<td>1.06E-08</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_161310</td>
<td>-0.68</td>
<td>2.85E-05</td>
<td>IQ domain containing protein D</td>
</tr>
<tr>
<td>Smp_197370</td>
<td>-0.68</td>
<td>1.99E-11</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_124600</td>
<td>-0.68</td>
<td>2.79E-10</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_046640</td>
<td>-0.66</td>
<td>4.30E-07</td>
<td>twik family of potassium channels</td>
</tr>
<tr>
<td>Smp_141690</td>
<td>-0.65</td>
<td>1.06E-04</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_160590</td>
<td>-0.65</td>
<td>9.35E-08</td>
<td>hypothetical protein</td>
</tr>
</tbody>
</table>

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, log2FC cut-offs were not used because all effects were small. The biggest change was of -1.27 fold (log2FC - 0.34, down-regulated in co-cultured schistosomules) (Table 4.5). All three genes were investigated as follows.
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 log2FC 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

<table>
<thead>
<tr>
<th>Gene identifier</th>
<th>Log2FC (HUVEC/non-HUVEC)</th>
<th>Adjusted p-value</th>
<th>Product name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smp_123920</td>
<td>-0.34</td>
<td>3.54E-03</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>Smp_016490</td>
<td>-0.28</td>
<td>3.82E-03</td>
<td>Saposin B domain containing protein</td>
</tr>
<tr>
<td>Smp_201210</td>
<td>0.32</td>
<td>8.09E-03</td>
<td>hypothetical protein</td>
</tr>
</tbody>
</table>

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 structure 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 interpret 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 may be 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 log2FC 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.
Chapter 5

Transcriptomes of cell lines exposed to schistosomules in vitro

5.1 Introduction

5.1.1 Overview

From chapter 4, I conducted a co-culture experiment between human-derived cells and *S. mansoni* schistosomules and studied how the parasites responded to the presence of the human cells. The three cell types used were HUVEC (human umbilical vein endothelial cells), HEPG2 (hepatocyte cancer cells), and GripTite (modified human embryonic kidney (HEK) cells). The overarching goal was to better understand the effect of the encounter in both the parasites and the host cells, and how this may contribute to the overall infection outcome. This chapter, therefore, covers the effect of the co-culture on the host cells. I investigated transcriptomic profiles of the co-cultured human cells, examined processes that were affected, and interpreted their relevance in the infections. This information, although derived from an *in vitro* system (with human cells forming an underlying layer and schistosomules placed on top), provides insights into host responses that the parasites encounter during an *in vivo* infection.

5.1.2 Host responses in *S. mansoni* infections

Systemic host responses have been explained largely in terms of changes in host immunological profiles during *S. mansoni* infections. The infections with *S. mansoni* appeared to suppress the inflammatory responses that would otherwise clear the infection (Pearce and MacDonald, 2002; Pearce *et al.*, 2004). It has been proposed that the parasites achieve this by camouflaging themselves with host molecules (Goldring *et al.*, 1976; Sell and Dean, 1972; Smithers *et al.*, 1969). Furthermore, the parasites secrete molecules that modulate host immune responses, reducing inflammation (Ranasinghe *et al.*, 2015a; Rao and Ramaswamy, 2000). Additionally, the parasites inhibit the attack by complement components by expressing genes that
prevent complement cascade activation, or degrade certain complement components (Da’dara et al., 2016a; Marikovsky et al., 1986; Schroeder et al., 2009; Skelly, 2004). The parasites clearly interact with their environment promoting the infection success. Moreover, the parasites are in close proximity to host tissues such as endothelial lining of blood vessels, and many cell types in the liver. These tissues are involved in pathogen infections; for examples, endothelial cells can express multiple cytokines and leukocyte adhesion molecules; and liver parenchymal tissue (hepatocytes) is the main factory of many molecules in complement and coagulation cascades. The proximity between *S. mansoni* and these tissues might allow interactions and modulation of such relevant processes.

### 5.1.3 Host tissue responding to *S. mansoni*

Endothelial cells are affected by *Schistosoma* infection. Treated with schistosomule ES, endothelial cells were polarised toward an anti-inflammatory profile that included changes to surface protein expression and second messenger signalling pathways (Angeli et al., 2001; Trottein et al., 1999a, 1999b). Furthermore, endothelial cells in the lung may be damaged during the lung phase migration due to the limited capillary spaces. However, lungs of infected rodents only show inflammation and wound healing processes after the parasites have left the tissue (Burke et al., 2011; Torre-escudero et al., 2014).

As for the liver tissues, information on the liver responses to schistosome infections have overwhelmingly focused on interactions with eggs and the formation of granuloma (e.g. Colley et al., 2014; Hams et al., 2013), with scarce information on how the liver is affected in early stages of an infection. Blood capillaries inside the liver are sinusoidal, allowing transport of large molecules between the blood vessels and the hepatocytes. The hepatocytes lying adjacent to blood vessels, therefore, might be affected by the migrating and developing parasites.

### 5.1.4 Aims and approaches

The aim of this chapter is to gain further insight into how cells of mammalian host respond to *S. mansoni*, in order to better understand the altered biological processes which may be important for the infections. Using co-culture set up between schistosomules and human-derived cells, I demonstrated in chapter 4 that
schistosomules changed their gene expression in response to different human cell types. In this chapter, I investigated transcriptional changes in the co-cultured human cells to find out how they were affected by the schistosomules.

### 5.1.5 Outline

In chapter 4, I described the co-cultured experiment where three types of *in vitro*-adapted human cells were co-cultured with schistosomules. Both the human cell and the schistosomule samples were processed for RNA-seq and in chapter 4 the transcriptomes of the schistosomules were investigated. In this chapter, the transcriptomes of the human cells were studied. Two types of the cells, although originating from tissues relevant to *S. mansoni* infections (endothelial cells and hepatocytes), have been adapted to an *in vitro* environment and may have changed in their physiology. Therefore, in the first part, I assessed HUVEC in comparison to *in vivo* endothelial cells, and HEPG2 in comparison to *in vivo* hepatocytes. In the second part, I explored variation in transcriptomic profiles of all the human cell samples and investigated each cell type individually using differential expression and functional enrichment analyses. Finally, responses of genes in two pathways, selected from pathway enrichment analyses and their relevance to *S. mansoni* infections, were compared between cell types. Overall, this chapter confirms some of the known key host responses but also augments our knowledge with additional details not previously reported.

### 5.2 Methods

The human cells used for producing transcriptome dataset for this chapter were obtained as described in chapter 4. The experimental design is summarised here.

Three types of cells derived from human tissue were co-cultured with mechanically transformed schistosomules over a period of time, in Basch media. The cell types were HUVEC (human umbilical vein endothelial cells), HEPG2 (hepatocyte cancer cells), and GripTite (modified human embryonic kidney cells, HEK293). Human erythrocytes were added to the media from day three onward. Every 3-4 days, the schistosomules were transferred to new plates of cells until day 6, 10, and 17 after the start of the experiment (i.e. age of schistosomules after the transformation) when the schistosomules and cells were collected for downstream processing for transcriptome
analysis (co-cultured groups). At these time points, human cells were four days old after being plated and were co-cultured with schistosomules for three days regardless of the schistosomule ages; therefore, control cells were four days old cultured without schistosomules for three days in Basch media supplemented with human erythrocytes (worm-free groups).

The methods for downstream processing including RNA extraction, sequencing library preparation, and bioinformatic analysis are as described in chapter 2.

5.3 Results

5.3.1 RNA quantity and quality

The yield of extracted RNA ranged from 17-750 ng/µl, with the majority having RIN number of 10 (the lowest was 9.10) indicating non-degraded RNA. The purity of RNA, measured by NanoDrop spectrometre 1.0, however, were compromised in various samples. The 260/230 ratio, but not 260/280 ratio, was low in RNA extracted from HUVEC and HEPG2 cells. (Figure 5.1). The low 260/230 ratio can be an indicator of contamination with carbohydrate, or residual phenol or guanidine from the RNA extraction (Thermo Scientific, 2011). Carbohydrate carryover is most likely. Hepatocytes, represented by HEPG2, store glycogen. And endothelial cells, represented by HUVEC, are coated with glycocalyx (Reitsma et al., 2007). The samples were proceeded into RNA library preparation without further clean up because mRNA would be pulled down with oligo-dT beads, removing the RNA from the contaminants. One sample failed at library preparation step but this was not related to the low 260/230 ratio.
Figure 5.1 260/230 ratio of RNA extracted from human cells and its relationship with the RNA concentration

Purity of RNA samples from three cell types as indicated by 260/280 ratio (A) and 260/230 ratio (B) measured from NanoDrop Spectrophotometer. C) Relationship between RNA concentration and the 260/230 ratio.

5.3.2 HUVEC and endothelial cell surface markers

Although HUVEC were derived from endothelial cells, they have been adapted to an in vitro culture environment and may only retain some aspects of true endothelial cells. The full extent of the differences could not be revealed without the original endothelial cells. Instead, HUVEC were assessed for their likeness to in vivo endothelial cells based on expression of known endothelial cell surface markers obtained from (Durr et al., 2004), containing 70 markers from humans, rats, and mice. Uniprot identifiers, if provided, were used to retrieve the identifiers of their human orthologues from the NCBI database. Where the identifiers were not provided, product names were used to search for human orthologues. The original list of genes from (Durr et al., 2004) and their matched human orthologues are provided in Appendix G. Defining expressed genes as having an FPKM greater than 0 in all replicates in worm-free HUVEC, 64 genes out of 70 genes (91.4%) were expressed in
worm-free HUVEC (Figure 5.2). Although, it is not possible to determine the full extent of differences between HUVEC and true endothelial cells nor the downstream processes once the surface interactions have been initiated, the presence of most of the cell surface markers confirms that the \textit{in vitro} cells are capable of many of the environmental interactions that their \textit{in vivo} counterparts perform.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure5_2.png}
\caption{Expression levels of endothelial cell surface markers in worm-free HUVEC}
\end{figure}

Endothelial surface marker genes from Durr \textit{et al}. (2004) and their expression levels in worm-free HUVEC. The expression levels shown on y-axis are the lowest FPKM amongst all three replicates of worm-free HUVEC. Each tick on x-axis represent one gene. Six genes at the right hand side were not expressed in any replicate of worm-free HUVEC. Two of the genes were expressed at a lower level that their levels are not visible on the plot. 91.4\% of the endothelial cell markers were expressed in worm-free HUVEC (at least one replicate with FPKM $> 0$).
5.3.3 HEPG2 and liver hepatocytes

Similarly, HEPG2 were derived from hepatocyte cancer cells and would also have adapted to in vitro environment. The comparison between HEPG2 and liver hepatocytes has been done previously using RNA-seq expression comparing HEPG2 with liver tissue (Tyakht et al., 2014). The study shows that 50 genes were up-regulated in HEPG2 cells while 608 genes were up-regulated in the liver tissue. The 50 genes were enriched in cell cycle and cell division, reflecting the carcinoma nature of the cell line (Tyakht et al., 2014) and positive selection for fast-growing cells in vitro. The 608 genes up-regulated in liver tissue, are enriched in processes of liver function such as metabolism and innate immune responses (Tyakht et al., 2014).

Incorporating the data in this thesis, out of the 608 genes up-regulated in liver tissue compared to HEPG2 (Tyakht et al., 2014), 596 genes could be identified (by matching gene names) in the reference gene set used in this thesis. Out of the 596 genes, 319 (53.5%) were expressed in worm-free HEPG2 (FPKM > 0) (data not shown).

5.3.4 Overall profiles of transcriptomes

PCA showed distinct clusters of each cell type. From the Figure 5.3 (all cell types), each cell type forms a distinct cluster regardless of whether the cells were co-cultured with schistosomules or without. Most of the transcriptional differences are therefore determined by cell type. The differences between worm-free and co-cultured cells were only seen when each cell type was clustered separately (Figure 5.3, HEPG2, HUVEC, GripTite). With such marked differences between cell types, samples from each cell type were considered as separate experiments.
Figure 5.3 PCA from transcriptomes of all cell types, and separate for each cell type

Variation between samples shown as PCA plots when all samples (all three cell types) were considered together (All cell types). HEPG2, HUVEC, GripTite) variation within cell types as indicated by the plot titles. On each plot legend, ‘time’ refers to the age of schistosomules when the cells were collected. At the point of collection, all cells had been with schistosomules (regardless of schistosomule ages) for three days. NON: cells that were not co-cultured with schistosomules (worm-free cells).

When each cell type was treated individually, the samples that were co-cultured with worms formed distinct groups (filled shapes, Figure 5.3, HEPG2, HUVEC, GripTite), and all were separated from worm-free cells (crossed box, Figure 5.3 HEPG2, HUVEC, GripTite) along the first principal component axis (PC1) that explained 42-78% of transcriptional variation between samples. In contrast, the differences between cells that were exposed to the parasites of different time point were primarily separated along the PC2 axis which explained a much smaller percentage of variance (Figure 5.3). It could be seen here that the effect of time was relatively small compared to the effect of co-culture. Because of this distinct separation, I performed the analysis by placing cells into worm-free and co-cultured groups.
Next, I investigated the effect of the co-culture in each cell type separately; HUVEC, HEPG2, and lastly GripTite (modified HEK cells). Amongst the cell types used for this experiment, GripTite is the least similar to cell types that schistosomules would naturally encounter \textit{in vivo}.

\textbf{5.3.5 Differential expression between co-cultured vs. worm-free cells}

\textbf{5.3.5.1 HUVEC}

Comparing co-cultured vs. worm-free in HUVEC, the number of differentially expressed genes was the greatest between similar comparisons in the other two cell types; 453 genes were down-regulated and 624 genes were up-regulated (adjusted p-value cut-off = 0.01, log$_2$FC cut-off $+/-$ 1) (Figure 5.4, Table S5.1, Table S5.2).
Figure 5.4 Volcano plots for pairwise comparison in each cell type between co-cultured vs. worm-free conditions

Log2FC and adjusted p-value of differentially expressed genes in each cell type comparing between cells co-cultured with schistosomules vs. worm-free cells. Numbers at the top indicate the number of genes that pass cut-offs for calling genes differentially expressed with adjusted p-value < 0.01 and log2FC +/- 1 (shown as blue dots).

Endothelial cell adhesion molecules

Amongst the most down-regulated genes were those encoding adhesion proteins on endothelial cells. These are selectin E (SELE, 50.1 fold), vascular cell adhesion molecule 1 (VCAM1, 40.3 fold), and intercellular adhesion molecule 1 (ICAM1, 13.6 fold) (Figure 5.5, Figure 5.6, Table S5.2). In addition to the lower fold change of ICAM1, its expression in co-cultured HUVEC did not reduce to zero, unlike for VCAM1 and SELE (Figure 5.5).
Figure 5.5 Expression profiles of SELE, VCAM1, and ICAM1

Expression of three genes encoding endothelial cell adhesion molecules that were down-regulated in co-cultured HUVEC. Each dot is one biological replicate. Y-axis represent normalised read counts. The lowest read counts of ICAM1 in co-cultured HUVEC were approximately 200 where as, for the SELE and VCAM1, the lowest read counts were close to zero. SELE, selectin E; VCAM1, vascular cell adhesion molecule 1; ICAM1, intercellular adhesion molecule 1.

These genes are involved in adhesion of circulating leukocytes and have important roles for recruitment of circulating immune cells to sites of injury or tissue damage. All three genes have been established as being regulated at transcription level by cytokines, LPS, and “other mediators of inflammation”. In contrast, two other adhesion molecules ICAM2 and SELP are not found to be regulated by the inflammation-related signals (Carlos & Harlan 1994). ICAM2, SELP, and other endothelial cell adhesion molecules (SELL, PECAM1, MAdCAM1) (Carlos & Harlan 1994) were not down-regulated in co-cultured HUVEC cells (Figure 5.6), suggesting that this is a specific suppression on inflammation-related adhesion molecules.
Figure 5.6 Expression of endothelial adhesion molecules between co-cultured vs. worm-free HUVEC

Log2FC of eight genes encoding endothelial adhesion molecules, comparing between HUVEC co-cultured with schistosomules and worm-free HUVEC. SELE, selectin E; SELP, selectin P; SELL, selectin L; VCAM1, vascular cell adhesion molecule 1; ICAM1, intercellular adhesion molecule 1; ICAM2, intercellular adhesion molecule 2; MADCAM1, mucosal vascular addressin cell adhesion molecule 1; PECAM1, platelet/endothelial cell adhesion molecule 1.

Extracellular matrix

There is also transcriptional evidence that the extracellular matrix (ECM) is remodelled; for example, up-regulated genes were enriched for the GO term extracellular matrix organization and down-regulated genes were enriched for the GO term cell junction organisation (Figure 5.7, Table S5.3). Furthermore, all 16 enriched pathways from the down-regulated genes can be related to extracellular matrix organisation (Figure 5.8, Table S5.4). Genes encoding different types of collagen chains, integrin subunits, and metalloproteinases were among the down-regulated as well as up-regulated genes (Table S5.1, Table S5.2). Schistosomules and adult S. mansoni have previously been shown to degrade ECM by a process that involve metalloproteinases (Keene et al., 1983; McKerrow et al., 1983). In this thesis, I have shown that ECM may also be affected as a result of gene expression changes in host cells that were exposed to schistosomules. Thereby, S. mansoni might affect the
ECM environment in two ways: by altering existing ECM, and by influencing host gene expression in the presence of the parasite.

**Immune system**

The massive down-regulation of the three leukocyte adhesion molecules (Figure 5.5) suggests anti-inflammatory profile of the co-cultured HUVEC. In addition to this, down-regulated genes were enriched in *immune system processes* (Figure 5.7, Table S5.3). Some of the down-regulated genes annotated with this GO term are involved in differentiation and activation of immune cells. For example, expression of *colony stimulating factor 1* was reduced by 3.5 fold, and that of *IL-32* was reduced by 2.5 fold. However, down-regulated genes also include inhibitors of the complement cascade, and the lower level would promote complement activation (Table S5.2). For example, gene expression of gene encoding CFH which functions to inhibit complement component 3 (C3) was reduced by 2.8 fold. The GO term *immune system process* was also enriched in the up-regulated genes (Figure 5.7), reflecting the nature of extensive connections in the immune system.
Figure 5.7 GO enrichment of genes differentially expressed in co-cultured HUVEC compared to worm-free HUVEC

Bar chart of enriched GO term (biological process) of up-regulated and down-regulated genes in co-cultured HUVEC. Expressed genes were used as enrichment reference background. Expressed genes were genes that showed at least some level of expression in at least one replicate of either co-cultured or worm-free condition (FPKM > 0).
Figure 5.8 Pathway enrichment of genes differentially expressed in co-cultured HUVEC compared to worm-free HUVEC

Networks of enriched pathways from three pathway databases. Inputs to pathway enrichment analysis were differentially expressed genes between co-cultured and worm-free cells (HUVEC: adjusted p-value < 0.01, log₂FC +/-1). Each node represents a pathway and the links between nodes indicate that the nodes share at least one gene. Colours indicate sources of the analysis results: red, Reactome; green, KEGG (via InnateDB interface); blue, IPA. Numbers on each node match the numbers in Table S5.4 which provide more details on each enriched pathway.

Coagulation

Coagulation was an enriched pathway in the up-regulated genes (Figure 5.8). The up-regulated genes in the pathway included two plasminogen activator genes (PLAUR
and PLAT) (Table S5.4). Plasminogen activators promote the formation of plasmin which degrades fibrin clots, and thereby reverse the blood clotting process. In addition, thrombomodulin (THBD) and SERPIND1 were both up-regulated (Table S5.4) and their encoded proteins impede the coagulation cascade by reducing thrombin level (THBD), or by inhibiting thrombin functions (SERPIND1). Intriguingly, Factor 3 (F3) gene whose protein product initiates blood coagulation cascade was also up-regulated (Table S5.4). However, the four genes that work to inhibit (or reverse) blood clotting were up-regulated by a larger effect size (Figure 5.9). In fact, SERPIND1 was amongst the top 10 up-regulated genes in co-cultured HUVEC (up-regulated by 16.6 fold). Together this suggests inhibition of blood clotting process as well as degradation of fibrin clots once the clots have been formed.

![Figure 5.9](image)

**Figure 5.9 Coagulation-related genes up-regulated in co-cultured HUVEC**

Coagulation-related genes from the IPA database that were up-regulated in co-cultured HUVEC compared to worm-free HUVEC. F3, factor 3; PLAT, plasminogen activator (tissue type); PLAUR, plasminogen activator (urokinase receptor); SERPIND1, serpin peptidase inhibitor, clade D, member 1; THBD, thrombomodulin.

**Cell cycle**

Cell cycle was strikingly affected in co-cultured HUVEC. This was reflected in both enriched GO terms and enriched pathways, most recognisably for the up-regulated genes. (Figure 5.7, Figure 5.8). The top four enriched GO terms were chromosome
segregation, mitotic nuclear division, chromosome organization and cell division, and the enrichment of these GO terms were strongly significant (Figure 5.7). Many pathways related to cell cycle were enriched in the up-regulated genes (Figure 5.8). Up-regulated were genes encoding key regulatory functions for cell cycle such as cyclins and cyclin-dependent kinases (Table S5.4). However, in down-regulated genes, cell proliferation was also the most significantly enriched GO term (Figure 5.7). Cell cycle process is tightly controlled by many genes and check points; therefore, the presence of genes related to cell cycle in both up-regulated and down-regulated genes may not be surprising. Furthermore, up-regulated and down-regulated genes consist of both genes that promote cell cycle progression and genes that inhibit the progression (Table S5.4). As a result, interpreting phenotypic outcome for cell proliferation is challenging.

Notch signalling
In addition to the effect on cell cycle control and mitosis, up-regulated genes in co-cultured HUVEC were enriched in the Notch signalling pathway (Figure 5.8). Notch signalling has been described as a link between innate and adaptive immunity (Ito et al., 2012), as well as regulating cell-cell interaction between adjacent cells in a process involving differentiation, proliferation, and apoptosis (Chigurupati et al., 2007). Genes in the Notch signalling pathway that were up-regulated were Notch4 receptor, JAG1 (ligand of multiple Notch receptors), DDL4 (ligand of Notch1 and 4, and regulator of endothelial cell proliferation), and transcriptional repressors downstream of Notch signalling HEY2, HES7, HEY1 (Ranganathan et al., 2011) (Table S5.4). With these changes covering genes from Notch receptors to Notch-regulated transcriptional factors, Notch signalling and its role in controlling cell-cell interaction and cell proliferation appears to be affected.

HUVEC conclusion
Taken together, the HUVEC responses to the schistosomules primarily affected cell cycle and extracellular matrix organisation. Genes related to immune responses were affected but resulting phenotypes were unclear. However, changes in coagulation pathway genes suggest reduction of blood clotting, and down-regulation of leukocyte adhesion molecules suggests an anti-inflammatory profile.
5.3.5.2  HEpg2

Top differentially expressed genes

The effect of the co-culture on HEpg2 cells was the smallest among the three cell
types tested. Only five genes were up- and five genes were down-regulated (log2FC
cut-off +/-1) (Table 5.1). Strikingly, the five down-regulated genes included four
genes whose products can be associated with anti-microbial peptide or innate
responses to bacterial infection such as recognition of LPS, and are listed as follows.
Haptoglobin (2.8-fold down-regulated) is an antimicrobial peptide that also binds to
free heme reducing iron availability to pathogens (Yang et al., 1983). A form of
Kininogen 1 (2.4-fold down-regulated) is involved in coagulation and production of
bradykinin which is an antimicrobial peptide (Hofman et al., 2016). Hepcidin (2.2
fold down-regulated), in addition to being an antimicrobial peptide, functions in iron
homeostasis of liver and intestine (Ganz and Nemeth, 2012). Finally, LPS-binding
protein (2-fold down-regulated) is involved in acute responses to bacterial LPS
(Bochkov et al., 2002). Together, down-regulation of these genes suggests that part of
the innate immune response is suppressed in co-cultured HEpg2 (Table 5.1). Up-
regulated genes however did not show clear relevance to innate immune response but
instead include genes involved in modification of collagen, cell-cell interaction, cell
migration, proliferation and metabolism (Table 5.1).
To increase sensitivity for HEPG2, I reduced the log$_2$FC cut-off from $+/−1$ (2-fold change) to $+/−0.5$ (1.4-fold change). With this relaxed cut-off, 193 genes were detected as up-regulated (Table S5.5), and 137 genes were down-regulated (Table S5.6). The genes were used as inputs for GO term enrichment (Figure 5.10, Table S5.7) and pathway enrichment (Figure 5.11, Table S5.8). The results were consistent with the functions of top five up-regulated and top five down-regulated genes discussed previously but with additional biological processes identified; such as cell differentiation GO term and ECM-receptor interaction pathway for up-regulated genes, and with metabolic processes and complement and coagulation pathway in down-regulated genes (Figure 5.10, Figure 5.11).
Figure 5.10 GO enrichment of genes differentially expressed in co-cultured HEPG2 compared to worm-free HEPG2

Bar chart of enriched GO terms (biological process) of up-regulated and down-regulated genes in co-cultured HEPG2. Expressed genes were used as enrichment reference background. Expressed genes were genes that showed at least some level of expression in at least one replicate of either co-cultured or worm-free condition (FPKM >0).
Figure 5.11 Pathway enrichment of genes differentially expressed in co-cultured HEPG2 compared to worm-free HEPG2

Networks of enriched pathways from three pathway databases. Inputs to pathway enrichment analysis were differentially expressed genes between co-cultured and worm-free cells (HEPG2: adjusted p-value < 0.01, log₂FC +/- 0.5). Each node represents a pathway and the links between nodes indicate that the nodes share at least one gene. Colours indicate source of the analysis results: red, Reactome; green, KEGG (via InnateDB interface); blue, IPA. Numbers on each node match the numbers in Table S5.8 which provide more details on each enriched pathway.
Overview of up-regulated genes

The GO terms enriched amongst up-regulated genes included cell differentiation, cell death, development, and cell-cell signalling. These are also reflected in the results of pathway enrichment-analysis where ECM-receptor interactions and other cell-cell interaction pathways were highlighted (Figure 5.10, Figure 5.11). Many of the genes responsible for the enrichment encoded integrins (Table S5.7, Table S5.8). Integrins are localised at cell membrane and provide a link between the cytoskeleton of a cell and the ECM. This suggests increased interactions of HEPG2 cells with their extracellular matrix environment, which contrasts with the response of HUVEC. In co-cultured HUVEC, similar pathways and GO terms were enriched in down-regulated genes (Figure 5.7, Figure 5.8). Toward the end of this chapter, I explore changes of genes in ECM organisation pathway in more detail, comparing between three cell types.

Overview of down-regulated genes

The down-regulated genes fall into two main types according to functional enrichment: metabolic processes, and immune response and circulation processes (Figure 5.10, Figure 5.11). Metabolic processes include metabolism of amino acids (urea cycle, and phenylalanine metabolism pathways), lipid, and sulfur compounds. In particular, three genes out of six genes in the urea cycle were down-regulated and potentially could affect metabolism of amino acids (Table S5.8). Furthermore, some of the genes encoding enzymes in urea cycle have roles in immune responses - arginase 1 - and was down-regulated by 1.6-fold (Table S5.6). Immune response and circulation processes affected in co-cultured HEPG2 were genes in the complement and coagulation cascade (Figure 5.11). These are directly relevant to the infection with S. mansoni and are explored further as follows.

Coagulation

Regarding the coagulation, fibrinogen alpha, beta, and gamma chain were all down-regulated (fold changes between 1.6-1.7) (Table S5.6). Fibrinogens are required for formation of blood clot and are converted by thrombin into insoluble fibrin strand (Mebius et al., 2013). In addition to fibrinogen, genes encoding serine protease inhibitors, coagulation factor 7 and factor 12 that are situated at the beginning of coagulation cascade were down-regulated (Table S5.6). Factor 7 is responsible for the
initiation of extrinsic coagulation pathway (from tissue factor released from damaged tissue), and factor 12 is responsible for the initiation of intrinsic blood coagulation pathway (contact activation from damaged blood vessel surface revealing charged surface) (Renne et al., 2012). On the other hand, the down-regulated genes also included genes that regulate coagulation such as serine proteinase inhibitor SERPINC1 and SERPINF2 (Table S5.6); meanwhile, the up-regulated genes include genes that lead to degradation of blood clots, PLAU and PLAUR (Table S5.5). Together, changes of genes in this pathway suggest an anti-coagulation profile, which is in contrast to responses to microbes where coagulation is often activated (Esmon et al., 2011).

Complement cascades
For the complement cascade, genes that were down-regulated in co-cultured HEPG2 fell into the alternative pathway, classical pathway and the formation of membrane attack complex (genes encoding CFB, C1R, C1S, C4B, C5, C8B, C8G) (Table S5.8). Both pathways are capable of killing the parasites (Marikovsky et al., 1986; Ruppel et al., 1983; Santoro et al., 1979). However, and interestingly, C3 that is essential for complement cascade progression was not affected in co-cultured HEPG2. Nevertheless, C5 and C8 which are required for downstream steps of all complement pathways (after the activation of C3) were both down-regulated (Table S5.8).

HEPG2 conclusion
In summary, genes from HEPG2 that were most affected by the co-culture had roles in ECM organisation, cell-ECM interaction, metabolism, and immune responses. From the gene expression data and functional enrichment, it appeared that the coagulation pathway was impeded - genes that initiate the cascade were down-regulated, genes that lead to degradation of fibrin clots were up-regulated. For the complement cascade, initiators of the classical pathway were down-regulated as well as genes that are shared between all three complement pathways.

5.3.5.3 GripTite
GripTite, modified HEK293 cells, were used in this experiment to represent cells that the parasite would not normally encounter in natural infections. When co-cultured with schistosomules, the number of differentially expressed genes in GripTite cells fell in between the numbers seen for endothelial cells and HEPG2 cells. There was a
skew toward genes being up-regulated in co-cultured cells, with 292 genes up-regulated (Table S5.9) and only 58 genes down-regulated (Table S5.10) (log$_2$FC cut-off at +/-1, adjusted p-value cut-off at 0.01; Figure 5.4).

**Neuronal signalling**

Based on GO annotations, the differentially expressed genes in co-cultured GripTite cells are involved in signalling, cell differentiation, developmental control, metabolism, and carrier for ion transport (Figure 5.12, Table S5.11). The top 10 up-regulated genes are all involved in receptor signalling and transcriptional regulation (Table S5.9). Some of these top 10 up-regulated genes are related to nervous system functions; for example, *G protein-coupled receptor 3 (GPR3)* and *activity-regulated cytoskeleton-associated protein (ARC)* both function in neurons (Dynes and Steward, 2007; Kumar et al., 2015).

**JAK-STAT, MAPK, NF-κB signalling**

The nature of top 10 up-regulated genes were also supported in pathway enrichment based on KEGG, Reactome, and IPA databases where the majority of enriched pathways are related to receptor signalling including *JAK-STAT signalling pathway*, *MAPK pathway*, and *NF-κB signalling* (Figure 5.13, Table S5.12). All of these are signalling pathways that regulate transcription, cell survival, proliferation, differentiation, and immune responses (Dhillon et al., 2007; Hoesel and Schmid, 2013; Rawlings, 2004). In addition, the pathways are known to have roles in cancer progression and metastasis via regulation of cell proliferation, and survival (Dhillon et al., 2007; Hoesel and Schmid, 2013; Rawlings, 2004). GO term enrichment of down-regulated genes returned processes downstream of these pathways, such as *cell differentiation* and *anatomical structure development* (Figure 5.12), indicating the outcome that may be affected by the signalling.
Figure 5.12 GO enrichment of genes differentially expressed in co-cultured GripTite compared to worm-free GripTite

Bar chart of enriched GO terms (biological process) of up-regulated and down-regulated genes in co-cultured GripTite. Expressed genes were used as enrichment reference background. Expressed genes were genes that showed at least some level of expression in at least one replicate of either co-cultured or worm-free condition (FPKM > 0).
Figure 5.13 Pathway enrichment of genes differentially expressed in co-cultured GripTite compared to worm-free GripTite

Networks of enriched pathways from three pathway databases. Inputs to pathway enrichment analysis were differentially expressed genes between co-cultured and worm-free cells (GripTite: adjusted p-value < 0.01, log₂FC +/- 1). Each node represents a pathway and the links between nodes indicate that the nodes share at least one gene. Colours indicate sources of the analysis results: red, Reactome; green, KEGG (via InnateDB interface); blue, IPA. Numbers on each node match the numbers in Table S5.12 which provide more details on each enriched pathway.

Immune system process

Genes involved in immune responses were differentially expressed, particularly those with pro-inflammatory functions - contrary to the responses of HUVEC and HEPG2 cells where anti-inflammatory profiles are observed. In co-cultured GripTite, *immune system process* was the most significantly enriched GO term amongst the up-regulated genes (Figure 5.12, Table S5.11). Amongst immune-related genes, those encoding the following protein products were the most differentially expressed (Figure 5.14, Table S5.9, Table S5.11): TAC1 being a antimicrobial peptide (Sun and Bhatia, 2014);
RUNX1 regulating expression of IL2 and promoting T cell proliferation (Ono et al., 2007); IL6R involved in the acute phase response (Heinrich et al., 1990); FOS controlling cell proliferation (Shaulian and Karin, 2001); EPHA2 being a receptor involved in signalling; EGR3 involved in leukocyte development (Li et al., 2012); EGR1 regulating early immune response gene expression (Decker et al., 2003; McMahon and Monroe, 1996); and CD44 involved in lymphocyte activation and angiogenesis (van Royen et al., 2004; Stefanová et al., 1989). However, genes whose product might suppress immune responses or have both pro- and anti-inflammatory function were also up-regulated such as genes encoding ANXA3 (possible anticoagulant) (Tait et al., 1991), and ANXA1 (down-regulated immune responses) (Wallner et al., 1986).
Figure 5.14 Immune responses-related genes up-regulated in co-cultured GripTite

Log2.FC of genes that are in GO term immune system process and were up-regulated in co-cultured GripTite compared to worm-free GripTite.

Griptite summary

In summary, GripTite cells co-cultured with schistosomules seem to be altered in genes related to signalling processes and the immune responses. Multiple up-regulated genes can be related to signalling involved in cell proliferation. Amongst the down-regulated genes, metabolic pathway of acetyl-CoA and lipid might be
affected. However, the results were not consistent when different methods of enrichment analyses were considered.

5.3.6 Pathways between cell types

From the previous section, some common features emerged between co-cultured cell types. Specifically, some of the pathways were affected in all three cell types, although different genes in the pathways were affected. In this final section, two pathways were chosen to compare the effect of the co-culture between cell types: first, the *extracellular matrix organisation* pathway, affected in either up-regulated or down-regulated genes in all three cell types; and second, the *coagulation and complement cascades* affected in HUVEC and HEPG2 cells that is directly relevant to *S. mansoni* infections.

5.3.6.1 Extracellular matrix organisation

Among the differentially expressed genes, the largest changes for both up- and down-regulated genes were in HUVEC. Comparing between cell types and considering genes with large fold changes (denser colours on the heatmap in Figure 5.15), *ICAM1*, and *BMP4* were down-regulated in co-cultured HUVEC, but were up-regulated in HEPG2 (Figure 5.15).
Figure 5.15 Log$_2$FC of genes in extracellular matrix organisation pathways

Genes in Reactome pathway *extracellular matrix organisation* and their differential expression in HUVEC, HEPG2, and GripTite compared between co-cultured cells and worm-free cells within each cell type. Colour map on red-blue spectrum indicates log$_2$FC from pairwise comparison of co-cultured vs. worm-free conditions for each cell type. Colour map on yellow-black spectrum indicates expression level (log$_{10}$ of the highest FPKM). Only genes that were expressed in at least one replicate (FPKM > 0) in all cell types were included in the heatmap. The sections of the heatmap where no colour was displayed for the log$_2$FC are genes that were not differentially expressed.
ICAM1 and BMP4
ICAM1 is required for adhesion of leukocytes (Carlos and Harlan, 1994), while BMP4 induces activation of dendritic cells (Martínez et al., 2011). The down-regulation of these genes in co-cultured HUVEC suggests a shift toward an anti-inflammatory immune profile. However, for liver, ICAM1 has a role in liver regeneration (Selzner et al., 2003). Whereas BMP4 is essential for liver tissue repair following injury (Oumi et al., 2012), even though it has been also described as inhibiting liver regeneration (Do et al., 2012). Although the parasites were unlikely causing tissue damage in the in vitro set up, the up-regulation in liver might infer preparation for tissue repair.

5.3.6.2 Coagulation and complement cascades
Given that most of the genes in complement and coagulation pathway are expressed mainly in liver hepatocytes, it is reassuring that the coagulation and complement pathways were down-regulated in co-cultured HEPG2 but not in the other two cell types. However, some genes in the pathway (such as CLU, F2R, FXII, PLAUR, and PLAU) were also expressed in other tissues and may provide insight on the effect of co-culture between cell types (Figure 5.16).
Figure 5.16 Log₂FC of genes in coagulation and complement cascade

Genes in KEGG pathway *coagulation and complement cascades* and their differential expression in HUVEC, HEPG2, and GripTite compared between co-cultured cells and worm-free cells within each cell type. Colour map on red-blue spectrum indicates log₂FC from pairwise comparison of co-cultured vs. worm-free conditions for each cell type. Colour map on yellow-black spectrum indicates expression level (log₁₀ of the highest FPKM). Only genes that were expressed in at least one replicate (FPKM > 0) in all cell types were included in the heatmap. The sections of the heatmap where no colour was displayed for the log₂FC are genes that were not differentially expressed.
CLU (clusterin) and F2R (factor 2 receptor)

Considering genes with high expression levels in all three cell types, *clusterin (CLU)* and *coagulation factor 2 receptor (F2R)* were affected differently between the three cell types. Both genes were down-regulated in co-cultured HUVEC (Figure 5.16); whereas clusterin was also down-regulated in HEPG2 cells but up-regulated in GripTite (Figure 5.16). *F2R* was down-regulated in both co-cultured HEPG2 and GripTite cells (Figure 5.16). Both genes are involved in the coagulation and complement cascades but also regulate other processes such as cell proliferation and apoptosis (Reinhardt et al., 2012; Santilli et al., 2003). *Clusterin* also modulates the *NF-κB signalling pathway* (Santilli et al., 2003) and this may explain their up-regulation in co-cultured GripTite (where NF-κB is among the pathways up-regulated). *F2R* is a GPCR receptor that binds to thrombin and is thought to have a role in platelet activation, but also involved in vascular remodelling (Reinhardt et al., 2012). Down-regulation in HUVEC may prevent these processes, but the gene might have different functions in the other two cell types.

PLAU PLAUR (Plasminogen Activator, Urokinase and its receptor)

*Plasminogen Activator, Urokinase (PLAU)* and *Plasminogen Activator, Urokinase Receptor (PLAUR)* were up-regulated in co-cultured cells of all three cell types (Figure 5.16). Functions of these genes were discussed previously in HUVEC section regarding degradation of blood clot. The two genes together promote plasmin formation and later plasmin which is required to degrade blood clots. In addition, the genes are also involved in extracellular matrix remodelling (Gilkes et al., 2014; Martinez-Hernandez et al., 2011). The consistent pattern between cell types suggests that this could be a generic response to schistosomules. It further suggests that the regulation of coagulation in *S. mansoni* infections could provide a vulnerability with potential for therapeutic exploitation.

5.4 Discussion

5.4.1 Overview

In this chapter, I studied transcriptomic changes in human-derived cells co-cultured with schistosomules. The aim was to gain further insights into host responses to the parasites and how the responses might affect outcome of the infection. I have shown
that the gene expression of the co-cultured HUVEC and HEPG2 exhibited anti-inflammatory and anti-coagulation profiles, whereas changes in co-cultured GripTite may suggest activation of immune responses. In addition, changes were observed in genes involved in extracellular matrix organisation and signalling related to cell proliferation. Changes within the same pathway, however, can be different between cell types and mechanisms underlying these differences are not known.

5.4.2 On in vitro-adapted cells

Two of the relevant cell types in this experiment were HUVEC and HEPG2, which are derived endothelial cells and hepatocytes, and have been adapted into in vitro environment. The extent to which HUVEC and HEPG2 cells represent true (in vivo) endothelial cells and hepatocytes was evaluated. HUVEC (in Basch media) expressed most of the genes of endothelial cell surface marker (over 90%). The high percentage may not be surprising because HUVEC is not a transformed cell line; therefore, the cells might not have adapted to a prolonged in vitro environment. Arguably, the differences between HUVEC and true endothelial cells may be in the gene expression levels.

For HEPG2, the cell line has been used by others for studying immune responses and responses to liver pathogen (Conceição et al., 2008; Israelow et al., 2014; Shavva et al., 2013) and most of the complement component were expressed in HEPG2 (C1q and C9 missing) (Morris et al., 1982). Some biological processes are clearly less complete in HEPG2 cells than in native liver cells (Tyakht et al., 2014) and it might therefore not be possible to draw a full picture of hepatocyte responses to schistosomules (or any treatment in vitro) when using HEPG2 as a model. However, some parts of important processes were transcriptionally active and presumably functioning and therefore can be useful to provide insights into real hepatocyte responses.

GripTite was included in the experiment to represent irrelevant cells where S. mansoni are unlikely to encounter in the infections. Although it is known as a modification of human embryonic kidney cells, it should not be considered as an in vitro counterpart of kidney epithelial or fibroblast. Instead it appears to have neuronal origin and possibly came from adrenal gland which locates close to kidney
(Stepanenko and Dmitrenko, 2015). Despite being irrelevant to *S. mansoni* infection, the cells were affected by the presence of the parasites and appeared to induce activation of immune responses and multiple signalling processes. This is likely to be generic innate responses to the parasites, but other irrelevant cell types may respond differently.

### 5.4.3 On aged media

In addition to the *in vitro* nature of the cells, their functions may be affected by media used in the experiment. The procedure transferred half of the old media to a new plate, meaning that overall media became older over time (as mentioned previously in chapter 4). In contrast, the worm-free cells were grown in fresh media for four days. Ideally, there should be a conditioned media control where worm-free cells received media that have been transferred over the same timecourse but without the presence of schistosomules. Alternatively, the media could be completely changed during the schistosomule transfer, but this would pose a risk of parasite loss or damage. The aged media may affect the cells, hence, it is important to bear in mind that the effect observed between the co-cultured and worm-free cells could be caused by the parasites, as well as by aged media. However, Basch media used in the co-culture was very rich in nutrient and still supported growth of co-cultured cells in the later time points. PCA also shows that time had little effect of transcriptomic profiles compared to the co-culture/worm-free factor. It is therefore not possible to entirely rule out the effect of aged media in this experiment, but it is unlikely to have been a major confounder.

### 5.4.4 Biological functions affected by parasite co-culture

#### 5.4.4.1 Endothelial-leukocyte adhesion molecules

HUVEC co-cultured with schistosomules exhibited down-regulation in genes encoding endothelial-leukocyte adhesion molecules. Previous studies investigated three adhesion molecules (VCAM1, ICAM1, and SELE) and reported two of the adhesion molecules VCAM1 and SELE, being down-regulated in human lung endothelial cells treated with lipophilic portion ES from lung stage schistosomules (Trottein *et al.*, 1999b). While our data is consistent regarding the down-regulation of VCAM1 and SELE, I also found ICAM1 being down-regulated. Multiple factors could contribute to the additional down-regulated gene, including the use of whole
schistosomules instead of ES portion, and different sources of endothelial cells. Expression of SELE, ICAM1, and VCAM1 on endothelial cells are involved in recruitment of leukocytes to the site of infection. SELE and ICAM1 in particular are relevant for recruitment of neutrophil (Carlos and Harlan, 1994; Lomakina and Waugh, 2009) which is capable of killing schistosomules (Incani and McLaren, 1981). For eosinophil, another effector cell that could mediate killing of schistosomules (Capron et al., 1979), the tethering is more dependent on selectin-P (SELP) than SELE (Kitayama et al., 1997) but SELP was not affected by the co-culture. The down-regulation of these adhesion molecules suggest an anti-inflammatory phenotype of co-cultured HUVEC.

5.4.4.2 Antimicrobial peptide

Similar to the observations in HUVEC, co-cultured HEPG2 cells also showed an anti-inflammatory responses. In particular, four genes with massive down-regulation encodes proteins which can function as antimicrobial peptides. The effects of antimicrobial peptides on schistosomes are not well known but they may have inhibitory roles by inducing inflammation around the parasites or by directly weakening the parasites. For example, an invertebrate antimicrobial peptide (dermaseptin) is lethal to S. mansoni adults in vitro and it also leads to reduced egg production and tegumental damage (de Moraes et al., 2011). Using antimicrobial peptide as an alternative treatment for schistosome infections has previously been proposed (Oyinloye et al., 2014). The four peptides down-regulated in the present study would clearly be good starting points for such a study.

5.4.4.3 Complement

In addition to cell-mediated immune responses, S. mansoni also needs to evade damage from complement cascade activation. Multiple strategies are known to be in place such as tegument receptor for complement components, acquiring inhibitor from host cells, and degradation of complement components (Da’dara et al., 2016a; Schroeder et al., 2009; Skelly, 2004). Our data show that S. mansoni may also modulate host complement production in HEPG2. Complement component genes down-regulated were CFB, C1R, C1S, C4B, C5, C8B, and C8G encompassing classical and alternative pathway, and formation of membrane attack complex. Interestingly, C3 was not affected at transcript level despite its importance at the start
of the alternative pathway and as a hub where all complement pathways converge. In comparison, C5, which also connects the three complement pathways, were down-regulated. Indeed, *S. mansoni* seem to have other strategies for regulating C3, such as receptor binding, C3 degradation, and tegument turn-over (Da’dana *et al*., 2016a; Schroeder *et al*., 2009). Additionally, C3 may be required in the hosts for other processes including developmental control, and liver regeneration (Hess and Kemper, 2016; Inal, 2004), implying that any response that *S. mansoni* provokes, needs to balance parasite and host survival.

### 5.4.4.4 Coagulation

Residing in blood vessels, *S. mansoni* would be expected to induce coagulation through interrupted blood flow and through damage induced from anchoring and migration. However, modulating coagulation may be important for *S. mansoni* infection due to the cross-talk between coagulation and innate immune response pathways; components in each of the two pathways can stimulate activation of the other pathway (Esmon *et al*., 2011; Markiewski *et al*., 2007), and enabling easy migration in the bloodstream. Multiple parasite-derived molecules can modulate the process (Mebius *et al*., 2013; Ranasinghe *et al*., 2015b). Here, our data show that *S. mansoni* might also regulate expression of components in coagulation pathways at their site of production (hepatocytes). The component affected in co-cultured HEPG2 ranged from those that initiate the coagulation cascade, formation of blood clots, and degradation of formed blood clots. Amongst these, down-regulation of three different fibrinogen chains – key components of blood clots – strongly suggests an anti-coagulation profile. Inhibitors of plasmin, an enzyme that degrades blood clots, were down-regulated in all three co-cultured cell types. This may explain a previous study in which blood from infected mice is more readily degraded (Da’dara *et al*., 2016b), but the effect of older parasites (liver stages and adults) on hepatocyte gene expression would need further investigation.

### 5.4.4.5 Tissue repair, signalling, cell cycle, and extracellular matrix reorganisation

Some genes responded differently between co-cultured cell types and the responses reflect physiological roles of the genes in that tissues. Examples of this are the down-regulation of *ICAM1* and *BMP4* in co-cultured HUVEC, but their up-regulation in co-
cultured HEPG2, with both genes having roles in liver regeneration (Oumi et al., 2012; Selzner et al., 2003). In addition, complement component C3, as mentioned previously, was not down-regulated in co-cultured HEPG2 and the expression level may be maintained for liver regeneration. Mechanisms that govern these differences are unknown but this differences between cell types may infer specific responses between parasites and different host tissues.

During infections with schistosomes, host tissues such as lung endothelial cells and liver hepatocytes are damaged (Bloch, 1980; Crabtree and Wilson, 1986; Torre-escudero et al., 2014). In this experiment, however, it is unlikely that the co-cultured cells were damaged by the parasites (and require tissue regeneration), because the parasites remained small and their oral and ventral suckers were not well developed. Although microscopic damages in vitro cannot be ruled out, the responses of liver-regenerating genes in HEPG2 may suggest that the presence of parasites could prompt the host cells to prepare for potential damage. Alternatively, the expression of genes related to liver regeneration may reflect HEPG2 origin of carcinoma cells. Another example of a preparation for tissue repair may come from the changes in co-cultured HUVEC. Notch signalling was one of the pathways up-regulated in co-cultured HUVEC. For endothelial cells, Notch signalling is important for homeostasis and development of blood vessel (Hofmann and Iruela-Arispe, 2007), regulating cell cycle (Noseda et al., 2004), and wound healing (Chigurupati et al., 2007). Furthermore, cell cycle genes were clearly affected in co-cultured HUVEC, and ECM organisation was affected by the production of collagen and proteases that reorganise the ECM.

The changes in these processes, however, could have other functions apart from preparation for tissue repair. For example, interference with host ECM is thought to be important for stages that invade host tissues, such as passing of eggs, miracidia, and sporocysts (Yoshino et al., 2014). In addition, schistosomules also degrade ECM (McKerrow et al., 1983). Such process may help with migration through capillary, or help with adhesion to host tissues. Indeed, schistosomules show preference for adhesion to endothelial cells compared to irrelevant cells (Trottein et al., 1999a). Moreover, in bacteria and protozoa, the pathogens modified host ECM to enhance their adhesion to the host tissues (Singh et al., 2012).
5.4.5 On in vivo validation

Given the in vitro nature of the experiment, cautions must be taken in interpreting its relevance in vivo. First, the host cells will have adapted to in vitro growth, with selection for rapidly growing cells; and second, additional factors in vivo could influence gene expression. Validation from in vivo system should be used to confirm the findings. However, many in vivo observations use rodents as an experimental model; whereas the in vitro experiment in this thesis used cells derived from humans. This could lead to inconsistent findings due to differences between organisms. Yu et al. (2010) investigated differences in livers of humans, mice, and rats and showed that there are differences in basic cellular processes (such as tricarboxylic acid cycle), cytokine-and cytokine receptor interaction, and complement and coagulation cascade (Yu et al., 2010). Many of these could be altered in S. mansoni infection; therefore, where possible, validation of effects represent in this chapter should derive from infection in human.

5.4.6 Summary

In this chapter, I investigated transcriptomic profiles of in vitro human cells co-cultured with schistosomules and showed that multiple biological processes related to infection of S. mansoni were affected, including immune responses, coagulation, alteration of host environment and signalling pathways. This information provides further understanding on how host biological processes might be modulated by the parasites. It is important to bear in mind the artificial aspects of the in vitro system and that follow up work should be performed, preferably from human infection, to fully understand the effect of the host-parasite interactions.
The discovery of *Schistosoma* spp. was made over a century ago. Since then multiple efforts have been made to decipher its life cycle and mechanisms required for successful infections. Although mammalian host bloodstream is often referred to as a “hostile environment”, interactions and co-evolution between the parasite and the host have reached the point that the parasite is not only able to evade attack of the host immune responses, but also dependent on components of the immune responses and the bloodstream environment for their normal development. Furthermore, although the parasite lives at the expense of its host, it is important for *S. mansoni*, as a long-living parasite, that the host alive over a long period of infection. As a result, the balance between exploitation and reparation needs to be maintained.

The recent sequencing of the genome has moved *S. mansoni* research to a new level. In addition, since its development in 2008, RNA-seq technology has revolutionised large-scale studies of gene structures and expression. Its standardised methodologies yield vast amounts of data with minimal technical variabilities. Arguably, transcriptomic approach does have limitations. The approach cannot explain additional processes that occur after transcription of genes; it does not explain the extent to which a gene is translated. Furthermore, stability of protein products may also affect biological outcomes. However, transcriptomic approach has a huge benefit of being able to cover a broader scope compared to other high-throughput methods. It also requires minimal starting materials, meaning that samples which are challenging to obtain in sufficient quantities can be much more easily studied.

In this work, I incorporated high-throughput RNA-seq and available genomic information to deepen the understanding of *S. mansoni* interactions with its mammalian hosts. In chapter 3, I investigated changes in gene expression profiles of the parasite during its development in infected mice. One of the features that sets the dataset from the *in vivo* parasite apart from previous attempts is the inclusion of the lung stage as part of the infection timecourse, providing a novel resource for.
investigating this early stage of infection. In chapter 4 and 5, *in vitro* experiments were set up where schistosomules were co-cultured with three types of human cells representing tissues that the parasite may encounter during its intra-mammalian infections. Transcriptomic profiles of the co-cultured schistosomules and co-cultured human cells were then explored.

6.1 Findings and interpretations

The findings in this thesis reflected key processes that could be important for *S. mansoni* infections and demonstrated that interactions with its host occurred in many biological aspects and through several changes in both the host and the parasite. One of the most fascinating aspects in the host-parasite interactions is how the parasite evade host defence mechanisms. This work agrees that multiple strategies for immune evasion are in place and it also predict a potential new player used by the parasite. The parasite also appeared to adapt to host environment in a way that it may benefit from metabolic resources available from the host. Finally, the interactions may not only support parasite survival, but also reduce pathology in the host.

6.1.1 Interactions with host defences

The interactions with host defences happen at multiple levels. Acquisition of host antigens alone cannot provide full protection from the host defences given that the parasite tegument is constantly shedded and that multiple parasite proteins and transporters are also present on tegument. Immunomodulatory strategies, therefore, could have essential roles. Majorities of the modulatory mechanisms discovered so far include parasite secreted and tegumental molecules that interact with immunological proteins produced by the host. In this thesis, additional potential player of the immunomodulation was revealed in chapter 3, where a predicted protein with no apparent sequence similarity but expressed at high level in lung stage, was found to have a clear structural similarity to the human CFH, an inhibitor of the complement cascade. Inhibition of complement cascade is likely to be important for *S. mansoni* survival, as demonstrated by multiple strategies of the parasite to inhibit host complement activation.

Interestingly, while the *in vivo* lung stage schistosomules and the adult stages up-regulated expression of a putative *S. mansoni* CFH-like gene (chapter 3), the exposure
of HUVEC to schistosomules *in vitro* led to down-regulation of host CFH-encoding genes (chapter 5). The down-regulation in HUVEC could be a secondary effect following the down-regulation of other inflammatory genes (leading to down-regulation of a regulator). Expression of *CFH* in endothelial cells during *S. mansoni in vivo* infections has not been investigated. However, if the down-regulation also occurred *in vivo*, the up-regulation of the putative parasite *CFH-like* gene may not only protect the parasite but also supply CFH to the host, reducing host pathology. Further experiments on CFH level in endothelial cells of infected mice and on molecular activities of putative parasite CFH would be needed to support the proposition.

In addition, this thesis demonstrated that the modulation of the host immune responses may also happen via changes in host gene expression. In chapter 5, transcriptomes of the co-cultured HEPG2 cells, originated from hepatocytes, the main source of complement proteins, showed down-regulation of multiple components in complement and coagulation cascades. A previous study on livers from infected mice did not report reduction in gene expression of either complement or coagulation components (Wijayawardena *et al.*, 2016). However, unlike the work in this thesis, the livers were obtained from mice infected with egg-laying adult worms which reside further away from the liver. The stages of the parasite, its distance from the liver, as well as the effects of eggs lodged in the liver might contribute to differences between studies. It would be interesting to investigate gene expression in the liver tissue during early liver stage. Alternatively, the gene expression of HEPG2 cells could be investigated when the cells were co-cultured with liver stage schistosomules obtained from infected mice.

Understanding how *S. mansoni* interacts with the host immune responses could lead to insights into the regulations of the immune system processes and how it can be modulated. The processes by which *S. mansoni*, and other parasites, modulate the host immune responses can be exploited for other practical purposes. For example, anti-coagulants from the parasite could lead to development of new treatment in blood-clotting disorders (Mebius *et al.*, 2013); and immunomodulatory proteins from parasites could be used to treat immune-related diseases (Artis and Pearce, 2013; Croese *et al.*, 2015).
6.1.2 Host reparation

Interactions with the host environment are required for parasite survival, and they may also help ensure survival of the host. For example, the immune responses against pathogens can damage host tissues; therefore, interference with the host immune responses may be beneficial to both the parasite and its host. HUVEC, originated from endothelial cells, displayed changes in gene expression related to the cell cycle and extracellular matrix remodelling, both of which are important for wound repair. Interestingly, these in vitro schistosomules were transcriptomically similar to in vivo lung schistosomules, which is the stage where schistosomules have been observed causing damages to endothelial cells in the lung (Crabtree and Wilson, 1986a). Changes in co-cultured HEPG2 also demonstrated that genes involved in liver tissue regeneration were up-regulated in the presence of schistosomules. Although damage is unlikely to have occurred within the in vitro co-culture, the presence of the parasites may trigger the host cells to prepare for potential damage. A better understanding on this interaction and how the parasite might trigger the responses could also be applicable for other research on tissue repair.

6.1.3 Responses to environments

The work in this thesis showed that different mammalian tissues can induce distinct responses in the parasite in vitro; however, other host factors are required for development of the parasite as shown previously by many investigators (e.g. Clegg, 1965b; Shaker et al., 1998, 2011). Growth of schistosomules in vitro was limited to the early in vivo lung stage as demonstrated by the morphologies and transcriptomic profiles. Some of the changes in the in vitro schistosomules, however, suggested that the HEPG2 environment may have induced responses in schistosomules most similarly to the in vivo environment. Not only were the transcriptomes of the HEPG2 schistosomules most similar to the lung stage schistosomules in vivo, but all of the MEGs that were up-regulated in the lung stage schistosomules were also up-regulated in the HEPG2 schistosomules.

This up-regulation of specific MEGs in HEPG2 schistosomules and in vivo lung schistosomules provides striking evidence that MEGs are involved in host-parasite interactions. It supports previous work that characterised the roles of MEGs in host tissue invasion (DeMarco et al., 2010), immune evasion (Lopes et al., 2013), and
rapid evolutionary changes of certain MEGs that suggest diversifying selection from host interactions (Philippsen et al., 2015). In addition, this thesis work showed that the HEPG2 environment may provide certain signals that stimulate responses similar to the lung environment.

I have also demonstrated that *S. mansoni* may adapt to oxidative stress in their environment, consistent with existing knowledge on the importance of oxidation-reduction processes in schistosomes (Li et al., 2015). In chapter 3 and 4, genes involved in neutralisation of ROS and iron storage were up-regulated in the stages where oxidative stress could be a major challenge (in the *in vivo* lung, and in the HEPG2 co-culture). The lung may present an oxidative challenge as a result of high oxygen pressure, a by-product of the cellular respiration and of the immune responses against the parasite during its delayed migration in the lung. Liver hepatocytes, represented in this thesis with HEPG2, are the main site for detoxification and this leads to an environment with high oxidative stress. In addition, hepatocytes can respond to pathogens by producing ROS. During the *in vitro* co-culture, it is not clear whether the production of oxidative species were induced in HEPG2 cells. However, one of the GO terms enriched in the up-regulated genes was *responses to stress*. In addition, a gene encoding aldehyde dehydrogenase was up-regulated in *in vivo* schistosomules from the liver stages. As described in chapter 3, this gene may be involved in optimising aldehyde metabolism in host liver to obtain substrates for parasite energy requirement. However, aldehyde can also cause oxidative stress and is toxic; therefore, the roles of the parasite enzyme may include regulations of oxidative stress. Together, this highlights the importance of regulating oxidative stress in *S. mansoni* and demonstrates that the regulations may be responsive to their environment.

On the other hand, specific responses of co-cultured cell types show that host tissues may sense the presence of the schistosomules, although the mechanisms of detection are unclear. Some of the responses between the schistosomules and the host cells may be related. Co-cultured HUVEC (chapter 5) were up-regulated in multiple biological processes including *Notch signalling pathway*. The genes up-regulated included gene encoding Notch receptor (Notch4) and the effector transcription factors. In schistosomules co-cultured with HUVEC cells (chapter 4), down-regulated were
genes structurally similar to the immunoglobulin part of Notch ligands. Immunoglobulin domains could be involved in many other cell-cell interactions. The immunoglobulin part in the parasite protein might be able to bind to Notch receptors and its down-regulation in HUVEC schistosomules may prevent interference with host Notch signalling. However, the gene was up-regulated in the in vivo lung stage (chapter 3), when the parasite is frequently found in tight capillary spaces, and the expression was declined as the parasites developed into adult forms. Therefore, this suggested that the gene may require additional stimuli to promote its expression. Together with the down-regulation in HUVEC environment, it is possible that the gene could be involved in responding to and interacting with host endothelial cells and could be explored further.

In addition to Notch signalling, one of the key changes in co-cultured HUVEC (chapter 5) was genes involved in ECM organisation. Interactions with host ECM are not well described for *S. mansoni*, but they are thought to be important during invading stages that include penetration through host tissues (Yoshino *et al.*, 2014). However, the roles of host ECM interactions during intramammalian stages are not established. The interactions appeared to occur at multiple levels. Schistosomules and adults of *S. mansoni* can degrade ECM as a result of their shed tegumental membrane (Keene *et al.*, 1983). In this thesis, expression of genes related to ECM organisation (including collagens and protease enzymes involving in the remodelling) was affected in co-cultured HUVEC as well as HEPG2. The effects on host gene expression suggested that the remodelling of and interactions with host extracellular matrix could be important for *S. mansoni* infections. Some of the possible explanations are that ECM interactions may help the parasite adhere to host blood vessels, or prepare host tissues for reparation of damage as explained previously in chapter 5.

In contrast to co-cultured HUVEC, co-cultured HEPG2 (chapter 5) exhibited a considerably smaller number of genes affected by the co-cultured with schistosomules. A possible explanation is that the HEPG2 were present with schistosomules transcriptomically similar to the lung stage (chapter 4) instead of liver stages. Alternatively, HEPG2 do not grow in a monolayer, but instead grow in clusters and often form lumps. Cells at the centre of the lumps might not be affected.
by the co-cultured schistosomules. As a result, the effects of schistosomules on cells on the periphery may be diluted.

### 6.2 Limitations

The possible dilution effect of HEPG2 can occur with any transcriptomic studies that include pooling of individual cells, or organisms. For example, *S. mansoni* samples from the *in vivo* timecourse were used as whole worms which contained mixtures of tissue types. If the changes over the timecourse only happened in certain tissues, they may have been masked out by changes in tissues which were more abundant or fluctuated over time. Such an effect has been described as Simpson’s paradox, which explains a common problem in identifying individual pattern when the data is a pool of multiple individual set of data (Simpson, 1951; Trapnell, 2015). The trend of changes might present when each dataset is considered separately, but disappear when all datasets are pooled. For example, given a tissue of mixed cell types, if a gene was up-regulated in one cell type but down-regulated in another cell type, the changes may not be detectable when RNA came from the whole tissue. Likewise, if one cell type is more abundant than others, its effects may represent the whole tissue even though other cell types might respond differently.

Current transcriptomic approaches cannot avoid the issues of pooled cell types, but modifications and technological development could improve future attempts. For example, in the case of HEPG2, a different type of hepatocyte-derived cells could be used that grow as a monolayer; alternatively, a single-cell transcriptomic approach could separate affected and un-affected cells. For *S. mansoni*, techniques have been developed for isolation of certain tissues such as reproductive organs (Lu *et al*., 2016) and gastrodermis (Gobert *et al*., 2009; Nawaratna *et al*., 2014). However, isolation of other tissues, such as cells of the nervous system, is still limited. Potential next steps to solve these issues and to obtain increasingly informative data could involve obtaining transcriptomes from single worms, or single cells from dissociated worms. Current constraints for this endeavour are laboratory procedures for dissociating each stage of the parasite into single cells effectively and optimally, and the handling and analysing massive amount of information that this approach will lead to.
Validating host responses *in vivo*, if using a transcriptomic approach, is also limited by the issue of pooling cell types. Extracting RNA from a mass of tissue types would not be ideal for studying the effects on the host cells that are in contact with the parasite. A proteomic study has shown that surface proteins could be isolated from other cellular proteins using biotin tagging followed by isolation of the proteins from dissociated cells (Torre-escudero *et al*., 2014). It could be appealing to experiment whether surface cells could be isolated from dissociated organs using a similar approach.

### 6.3 Concluding remarks

Transcriptomic data from this thesis provide further understanding of biological processes involved in intramammalian stages of *S. mansoni* infections, particularly on the aspects of immune evasion, and the responses between the parasite and its host. The datasets produced, and the genes and biological processes identified will provide a rich new resource for data mining and functional characterisation by the research community. Understanding the functions of parasite genes could lead to identifying targets for intervention and provide better overall insights into how the parasite functions within its host.

Finally, this work reflects the importance of incorporating multiple approaches into a research problem. The value of using multiple bioinformatic tools for functional predictions are demonstrated, but the idea can be extended to the incorporation between the high-throughput approaches and the functional characterisation of an individual gene or pathway. To study host-parasite molecular interactions, for *S. mansoni* and mammalian hosts, each organism possesses over 10,000 genes responsible for multiple interlinking biological processes. High-throughput approaches have provided a broad perspective, identifying putative processes that may be interconnected, and determining genes or processes that may have key roles. The putative new roles and functions from the current study will need to be validated, but the dataset has provided new starting points in the exploration of *S. mansoni* biology and in the search for vulnerabilities that may ultimately be exploited to combat and control schistosomiasis.
Chapter 7

References


splicing and distinct sub-groupings in members of the *Schistosoma mansoni* venom allergen-like (SmVAL) gene family. BMC Genomics 9, 89.


Lin, Y.-L., and He, S. (2006). Sm22.6 antigen is an inhibitor to human thrombin. Molecular and Biochemical Parasitology 147, 95–100.


## Chapter 8

### Appendices

**Appendix A: Aquarium water, 10X**

(Prepared by the Wellcome Trust Sanger Institute (WTSI) media team)

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Amount for 5 litre</th>
<th>Final conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CaCl2</td>
<td>2.78 g</td>
<td>5.00 mM</td>
</tr>
<tr>
<td>MgSO4.7H2O</td>
<td>6.14 g</td>
<td>4.98 mM</td>
</tr>
<tr>
<td>K2SO4</td>
<td>0.215 g</td>
<td>0.25 mM</td>
</tr>
<tr>
<td>NaHCO3</td>
<td>2.1 g</td>
<td>5.00 mM</td>
</tr>
<tr>
<td>*FeCl3.6H2O</td>
<td>240 ul</td>
<td>0.009 mM</td>
</tr>
</tbody>
</table>

*The FeCl3.6H2O is prepared by dissolving 5 g in 100 ml water.
The final pH of the aquarium water is pH 7 +/- 0.5.
Appendix B: Reference genomes used by NPG QC

The NPG QC team at the WTSI constantly adds more species to the pool of reference genomes. The species listed below are those included during the work of this thesis.

Acinetobacter baumannii
Actinobacillus pleuropneumoniae
Aeromonas hydrophila
Anopheles gambiae
Aspergillus fumigatus
Bacillus thuringiensis
Betacoronavirus
Bordetella bronchiseptica
Bordetella pertussis
Bos taurus
Brucella abortus
Brucella canis
Brugia malayi
Brugia pahangi
Burkholderia cenocepacia
Burkholderia gladioli
Burkholderia pseudomallei
Caenorhabditis elegans
Callithrix jacchus
Campylobacter fetus
Campylobacter jejuni
Canis familiaris
Cavia porcellus
Chlamydia trachomatis
Chlamydophila abortus
Citrobacter rodentium
Clostridium difficile
Cryptococcus neoformans
Cryptosporidium parvum
Danio rerio
Dracunculus medinensis
Drosophila melanogaster
Echinococcus granulosus
Echinococcus multilocularis
Enterococcus casseliflavus
Enterococcus faecalis
Enterococcus faecium
Enterococcus hirae
Equus caballus
Escherichia coli
Felis catus
Giardia intestinalis
Gorilla beringei
Gorilla gorilla
Haemonchus contortus
Haemophilus influenzae
Haemophilus parasuis
Heligmosomoides polygyrus
Hepatitis C
Heterocephalus glaber
HIV 1
Homo sapiens
Human herpesvirus 1
Human herpesvirus 2
Human herpesvirus 3
Human herpesvirus 4
Human herpesvirus 5
Human herpesvirus 6
Human herpesvirus 7
Human herpesvirus 8
Human papillomavirus
Hymenolepis microstoma
Influenza A
Klebsiella pneumoniae
Lactobacillus casei
Lambda
Legionella pneumophila
Leishmania braziliensis
Leishmania donovani
Leishmania infantum
Leishmania major
Leptospira interrogans
Macaca fascicularis
Melissococcus plutonius
MERS coronavirus
Monodelphis domestica
Mus musculus
Mustela putorius
Mycobacterium abscessus
Mycobacterium africanum
Mycobacterium avium
Mycobacterium bovis
Mycobacterium tuberculosis
Mycobacterium ulcerans
Mycoplasma agalactiae
Mycoplasma bovis
Mycoplasma genitalium
Mycoplasma hyopneumoniae
Myotis lucifugus
Neisseria gonorrhoeae
Neisseria meningitidis
Norwalk virus
Onchocerca volvulus
Oryctolagus cuniculus
Oryzias latipes
Ovis aries
Paenibacillus larvae
Pan troglodytes
Plasmodium berghei
Plasmodium chabaudi
Plasmodium falciparum
Plasmodium knowlesi
Plasmodium vivax
Propionibacterium acnes
Proteus mirabilis
Pseudomonas aeruginosa
Pseudomonas fluorescens
Rattus norvegicus
Rhabditophanes sp
Rhesus macaque
Saccharomyces cerevisiae
Saliva composite
Salmonella bongori
Salmonella enterica
Salmonella pullorum
Sarcophilus harrisii
Schistoscephalus solidus
Schistosoma mansoni
Schizosaccharomyces pombe
Serratia proteamaculans
Shigella boydii
Shigella flexneri
Shigella sonnei
Staphylococcus aureus
Staphylococcus haemolyticus
Staphylococcus saprophyticus
Streptococcus agalactiae
Streptococcus dysgalactiae
Streptococcus equi
Streptococcus pneumoniae
Streptococcus pyogenes
Streptococcus suis
Streptococcus uberis
Streptomyces coelicolor
Streptomyces venezuelae
Strongyloides ratti
Sus scrofa
Teladorsagia circumcincta
Trichobilharzia regenti
Trichobilharzia szidati
Trichuris muris
Trypanosoma brucei
Tupaia belangeri
Tursiops truncatus
Vibrio cholerae
Wolbachia endosymbiont of
Drosophila melanogaster
Xenopus tropicalis
Yersinia enterocolitica
Yersinia pseudotuberculosis
Appendix C: Versions of R packages used in data analysis

R version 3.3.1 (2016-06-21)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: OS X 10.11.1 (El Capitan)

locale:

attached base packages:
grid parallel stats4 stats graphics grDevices utils datasets methods base

other attached packages:
ape_3.5 VennDiagram_1.6.17 futile.logger_1.4.3
RCOLORBrewer_1.1-2 gridExtra_2.2.1 dplyr_0.5.0
vioplot_0.2 sm_2.2-5.4 kohonen_2.0.19
MASS_7.3-45 class_7.3-14 heatmap_1.0.8
topGO_2.24.0 SparseM_1.7 GO.db_3.3.0
AnnotationDbi_1.34.4 graph_1.50.0 ggplot2_2.2.1
DESeq2_1.12.3 SummarizedExperiment_1.2.3 Biobase_2.32.0
GenomicRanges_1.24.2 GenomeInfoDb_1.8.1 IRanges_2.6.1
S4Vectors_0.10.2 BiocGenerics_0.18.0

loaded via a namespace (and not attached):
genefilter_1.54.2 locfit_1.5-9.1 splines_3.3.1
survival_2.39-5 XML_3.98-1.4 foreign_0.8-66
lambda.r_1.1.9 matrixStats_0.50.2 plyr_1.8.4
gtable_0.2.0 latticeExtra_0.6-28 geneplotter_1.50.0
xtable_1.8-2 scales_0.4.1 Hmisc_3.17-4
tools_3.3.1 magrittr_1.5 lazyeval_0.2.0
futile.options_1.0.0 Formula_1.2-1 cluster_2.0.4
assertthat_0.1 R6_2.1.2 rpart_4.1-10
colorspace_1.2-6 munsell_0.4.3 XVector_0.12.0
BiocParallel_1.6.2 acepack_1.3-3.3 RSQLite_1.0.0
lattice_0.20-33 Rcpp_0.12.9.1 Matrix_1.2-6
DBI_0.4-1 annotate_1.50.0 nlme_3.1-128
zlibbioc_1.18.0 tibble_1.2 data.table_1.10.0
nnet_7.3-12
Appendix D: Basch media components

0.5 gr  Lactalbumin hydrolysate powder (L9010, Sigma)

250ul  Hypoxanthine (1 mM) (H9377, Sigma)

500ul  Insulin (8mg/ml) (I0516, Sigma)

500ul  Hydrocortisone (1 mM) (H0888, Sigma)

500ul  Triiodothyronine (0.2 mM) (T5516, Sigma)

2.5 ml  MEM Vitamins (100X) (M6895, Sigma)

25 ml  Schneiders Drosophila Medium (21720024, Invitrogen)

5 ml  Hepes Buffer (sc-286961, Santa Cruz Biotechnology, Inc.)

50 ml  Fetal bovine serum (F0926, Sigma)

10 ml  Antibiotic-Antimycotic (15240-062, Invitrogen)

to 500 ml  DMEM (D6546, Sigma)

Filter through 0.22 μm membrane and store at 4 °C.
Appendix E: *S. mansoni* genetrees downloaded from WormBase ParaSite release 9

The genetrees are grouped by chapters and ordered by *S. mansoni* gene identifier.

Legend for all genetrees

<table>
<thead>
<tr>
<th>Branch Length</th>
<th>Genes</th>
<th>Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1 branch length</td>
<td>Gene ID gene of interest</td>
<td>gene node</td>
</tr>
<tr>
<td>x10 branch length</td>
<td>Gene ID within-sp. paralog</td>
<td>speciation node</td>
</tr>
<tr>
<td>x100 branch length</td>
<td></td>
<td>duplication node</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collapsed Nodes</th>
<th>Collapsed Alignments</th>
<th>Expanded Alignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>collapsed sub-tree</td>
<td>0 - 33% Aligned AA</td>
<td>Gap</td>
</tr>
<tr>
<td>collapsed (paralog)</td>
<td>33 - 66% Aligned AA</td>
<td></td>
</tr>
<tr>
<td>(gene of interest)</td>
<td>66 - 100% Aligned AA</td>
<td></td>
</tr>
</tbody>
</table>

**Smp_008850 (chapter 3)**

A fully expanded tree is available at

http://parasite.wormbase.org/Multi/GeneTree/Image?gt=WBGT00840000210809
A fully expanded tree is available at http://parasite.wormbase.org/Multi/GeneTree/Image?gt=WBGT00800000173209
Smp_123350 (chapter 3)

A fully expanded tree is available at http://parasite.wormbase.org/Multi/GeneTree/Image?gt=WBGT0000000014882
Smp_140250 (chapter 3)

A fully expanded tree is available at
http://parasite.wormbase.org/Multi/GeneTree/Image?gt=WBGT00000000006409
Smp_150180 (chapter 3)

A fully expanded tree is available at
http://parasite.wormbase.org/Multi/GeneTree/Image?gt=WBG0080000173094
Smp_173010 (chapter 3)

A fully expanded tree is available at http://parasite.wormbase.org/Multi/GeneTree/Image?gt=WBGT0080000173235
Smp_062900 (chapter 4)

A fully expanded tree is available at
http://parasite.wormbase.org/Multi/GeneTree/Image?gt=WBGT00800000173326
Smp_012920 (chapter 4)

A fully expanded tree is available at
http://parasite.wormbase.org/Multi/GeneTree/Image?gt=WBGT00000000012253
Smp_041700 (chapter 4)

A fully expanded tree is available at
http://parasite.wormbase.org/Multi/GeneTree/Image?gt=WBGT00880000282628
Smp_170610 (chapter 4)

A fully expanded tree is available at
http://parasite.wormbase.org/Multi/GeneTree/Image?gt=WBGT00000000007301
Smp_201210 (chapter 4)

The tree is available at
http://parasite.wormbase.org/Multi/GeneTree/Image?gt=WBGT00860000263648

- SROB_0002124301, Schistosoma rodhaini (PRJEB526)
- Smp_201210, Schistosoma mansoni (PRJEA36577)
### Appendix F: Orthologues of Smp_041700 downloaded from WormBase ParaSite release 8

<table>
<thead>
<tr>
<th>Gene stable identifier of orthologues</th>
<th>Gene description</th>
<th>Genome project</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_02595</td>
<td>Rhodopsin-like orphan GPCR, putative (inferred by orthology to a <em>S. mansoni</em> protein)</td>
<td>schistosoma_haematobium_prjna78265</td>
</tr>
<tr>
<td>CapteG188111</td>
<td></td>
<td>capitella_teleta</td>
</tr>
<tr>
<td>CapteG190027</td>
<td></td>
<td>capitella_teleta</td>
</tr>
<tr>
<td>CapteG192465</td>
<td></td>
<td>capitella_teleta</td>
</tr>
<tr>
<td>CapteG197443</td>
<td></td>
<td>capitella_teleta</td>
</tr>
<tr>
<td>CapteG200779</td>
<td></td>
<td>capitella_teleta</td>
</tr>
<tr>
<td>CapteG206916</td>
<td></td>
<td>capitella_teleta</td>
</tr>
<tr>
<td>CapteG207119</td>
<td></td>
<td>capitella_teleta</td>
</tr>
<tr>
<td>CapteG208374</td>
<td></td>
<td>capitella_teleta</td>
</tr>
<tr>
<td>CapteG211298</td>
<td></td>
<td>capitella_teleta</td>
</tr>
<tr>
<td>CGI_10000319</td>
<td>Growth hormone secretagogue receptor type 1</td>
<td>crassostrea_gigas</td>
</tr>
<tr>
<td>CGI_10009331</td>
<td>Growth hormone secretagogue receptor type 1</td>
<td>crassostrea_gigas</td>
</tr>
<tr>
<td>CGI_10012154</td>
<td>C-C chemokine receptor type 5</td>
<td>crassostrea_gigas</td>
</tr>
<tr>
<td>CGI_10022932</td>
<td>Neuromedin-U receptor 1</td>
<td>crassostrea_gigas</td>
</tr>
<tr>
<td>CGI_10025276</td>
<td>Growth hormone secretagogue receptor type 1</td>
<td>crassostrea_gigas</td>
</tr>
<tr>
<td>CGI_10026655</td>
<td></td>
<td>crassostrea_gigas</td>
</tr>
<tr>
<td>D915_05109</td>
<td>Rhodopsin-like orphan GPCR, putative (inferred by orthology to a <em>S. mansoni</em> protein)</td>
<td>fasciola_hepatica_prjna179522</td>
</tr>
<tr>
<td>ECPE_0000219101</td>
<td>Rhodopsin-like orphan GPCR, putative (inferred by orthology to a <em>S. mansoni</em> protein)</td>
<td>echinostoma_caproni_prjeb1207</td>
</tr>
<tr>
<td>PXEA_0002723401</td>
<td>Rhodopsin-like orphan GPCR, putative (inferred by orthology to a <em>S. mansoni</em> protein)</td>
<td>protopolystoma_xenopodis_prjeb1201</td>
</tr>
<tr>
<td>SCUD_0000629701</td>
<td>Rhodopsin-like orphan GPCR, putative (inferred by orthology to a <em>S. mansoni</em> protein)</td>
<td>schistosoma_curassoni_prjeb519</td>
</tr>
<tr>
<td>Sip_0057450</td>
<td>Rhodopsin-like orphan GPCR, putative (inferred by orthology to a <em>S. mansoni</em> protein)</td>
<td>schistosoma_japonicum_prjea34885</td>
</tr>
<tr>
<td>SMRZ_0001551401</td>
<td>Rhodopsin-like orphan GPCR, putative (inferred by orthology to a <em>S. mansoni</em> protein)</td>
<td>schistosoma_margrebowiei_prjeb522</td>
</tr>
<tr>
<td>SMTD_0000981201</td>
<td>Rhodopsin-like orphan GPCR, putative (inferred by orthology to a <em>S. mansoni</em> protein)</td>
<td>schistosoma_mattheei_prjeb523</td>
</tr>
<tr>
<td>SROB_0001283401</td>
<td>Rhodopsin-like orphan GPCR, putative (inferred by orthology to a <em>S. mansoni</em> protein)</td>
<td>schistosoma_rodhaini_prjeb526</td>
</tr>
<tr>
<td>T265_15440</td>
<td>Rhodopsin-like orphan GPCR, putative (inferred by orthology to a <em>S. mansoni</em> protein)</td>
<td>opisthorchis_viverrini_prjna222628</td>
</tr>
<tr>
<td>TRE_0000682001</td>
<td>Rhodopsin-like orphan GPCR, putative (inferred by orthology to a <em>S. mansoni</em> protein)</td>
<td>trichobilharzia_regenti_prjeb4662</td>
</tr>
</tbody>
</table>
Appendix G: List of endothelial cell surface marker from Durr et al., 2004

<table>
<thead>
<tr>
<th>Product description</th>
<th>Produce name</th>
<th>Species</th>
<th>Uniprot identifier</th>
<th>Ensembl identifier of Human orthologue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endothelial plasminogen activator inhibitor</td>
<td>SERPINE1</td>
<td>R</td>
<td>P20961</td>
<td>ENSG00000106366</td>
</tr>
<tr>
<td>Endothelial actin-binding protein</td>
<td>FLNA</td>
<td>H</td>
<td>P21333</td>
<td>ENSG00000196924</td>
</tr>
<tr>
<td>MUC18</td>
<td>MCAM</td>
<td>R</td>
<td>x</td>
<td>ENSG00000076706</td>
</tr>
<tr>
<td>Integri alpha-5 (CD49e)</td>
<td>ITGA5</td>
<td>M</td>
<td>P11688</td>
<td>ENSG00000161638</td>
</tr>
<tr>
<td>Platelet endothelial tetraspan antigen-3</td>
<td>CD151</td>
<td>R</td>
<td>Q9QZA6</td>
<td>ENSG00000177697</td>
</tr>
<tr>
<td>Vascular endothelial-cadherin 1</td>
<td>CDH5</td>
<td>M</td>
<td>P55284</td>
<td>ENSG00000179776</td>
</tr>
<tr>
<td>EDG-1</td>
<td>S1PR1</td>
<td>R</td>
<td>P48303</td>
<td>ENSG00000170989</td>
</tr>
<tr>
<td>MAC-inhibitor (CD59)</td>
<td>CD59</td>
<td>R</td>
<td>P27274</td>
<td>ENSG00000085063</td>
</tr>
<tr>
<td>MRP-1 (CD9)</td>
<td>CD9</td>
<td>R</td>
<td>P40241</td>
<td>ENSG00000010278</td>
</tr>
<tr>
<td>Platelet-endothelial cell adhesion molecule-1 (CD31)</td>
<td>PECAM1</td>
<td>R</td>
<td>x</td>
<td>ENSG00000261371</td>
</tr>
<tr>
<td>Aminopeptidase N (CD13)</td>
<td>ANPEP</td>
<td>R</td>
<td>P97449</td>
<td>ENSG00000166825</td>
</tr>
<tr>
<td>Von Willebrand factor</td>
<td>VWF</td>
<td>R</td>
<td>P04275</td>
<td>ENSG00000110799</td>
</tr>
<tr>
<td>Caveolin-1</td>
<td>CAV1</td>
<td>R</td>
<td>Q8VIK9</td>
<td>ENSG00000105974</td>
</tr>
<tr>
<td>Endothelial cell-selective adhesion molecule</td>
<td>ESAM</td>
<td>M</td>
<td>Q925F2</td>
<td>ENSG00000149564</td>
</tr>
<tr>
<td>Integri beta-1 (CD29)</td>
<td>ITGB1</td>
<td>R</td>
<td>P49134</td>
<td>ENSG00000150093</td>
</tr>
<tr>
<td>Annexin V</td>
<td>ANXA5</td>
<td>R</td>
<td>P14668</td>
<td>ENSG00000164111</td>
</tr>
<tr>
<td>Podocalyxin</td>
<td>PODXL</td>
<td>R</td>
<td>x</td>
<td>ENSG00000128567</td>
</tr>
<tr>
<td>Intercellular adhesion molecule-1 (I-CAM1; CD54)</td>
<td>ICAM1</td>
<td>R</td>
<td>Q00238</td>
<td>ENSG00000090339</td>
</tr>
<tr>
<td>Endothelial differentiation-related factor 1</td>
<td>EDF1</td>
<td>H</td>
<td>O60869</td>
<td>ENSG00000107223</td>
</tr>
<tr>
<td>Intercellular adhesion molecule-2 (I-CAM2; CD102)</td>
<td>ICAM2</td>
<td>M</td>
<td>P35330</td>
<td>ENSG00000108622</td>
</tr>
<tr>
<td>Na+K+ transporting ATPase alpha 1</td>
<td>ATP1A1</td>
<td>M</td>
<td>P06685</td>
<td>ENSG00000163399</td>
</tr>
<tr>
<td>Vascular endothelial cell specific protein 11</td>
<td>SEPT2</td>
<td>R</td>
<td>x</td>
<td>ENSG00000168385</td>
</tr>
<tr>
<td>Protein Name</td>
<td>Gene Name</td>
<td>Gene ID</td>
<td>Protein ID</td>
<td>Gene ID</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>-----------</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>Integrin alpha V</td>
<td>ITGAV</td>
<td>M</td>
<td>P43406</td>
<td>ENSG00000138448</td>
</tr>
<tr>
<td>Endothelin converting enzyme</td>
<td>ECE1</td>
<td>R</td>
<td>P42893</td>
<td>ENSG00000117298</td>
</tr>
<tr>
<td>Thrombomodulin</td>
<td>THBD</td>
<td></td>
<td>x</td>
<td>ENSG00000178726</td>
</tr>
<tr>
<td>Scavenger receptor class F</td>
<td>SCARF1</td>
<td>H</td>
<td>x</td>
<td>ENSG00000074660</td>
</tr>
<tr>
<td>Microvascular endothelial differentiation gene 1</td>
<td>DNAJB9</td>
<td>R</td>
<td>P97554</td>
<td>ENSG00000128590</td>
</tr>
<tr>
<td>Integrin alpha-3 (CD49c)</td>
<td>ITGA3</td>
<td>M</td>
<td>Q62470</td>
<td>ENSG00000005884</td>
</tr>
<tr>
<td>5'-nucleotidase (CD73)</td>
<td>NT5E</td>
<td>R</td>
<td>P21588</td>
<td>ENSG00000135318</td>
</tr>
<tr>
<td>EDG-2</td>
<td>BUD31</td>
<td>H</td>
<td>P41223</td>
<td>ENSG00000106245</td>
</tr>
<tr>
<td>H-CAM (CD44)</td>
<td>CD44</td>
<td>R</td>
<td>O08779</td>
<td>ENSG00000026508</td>
</tr>
<tr>
<td>Tyrosine-protein kinase receptor TIE-2</td>
<td>TEK</td>
<td>R</td>
<td>Q9QW24</td>
<td>ENSG00000120156</td>
</tr>
<tr>
<td>Transferrin receptor (CD71)</td>
<td>TFRC</td>
<td>H</td>
<td>P02786</td>
<td>ENSG00000072274</td>
</tr>
<tr>
<td>Angiotensin-converting enzyme (CD134)</td>
<td>ACE</td>
<td>R</td>
<td>P12821</td>
<td>ENSG00000159640</td>
</tr>
<tr>
<td>Tight junction protein 2 - ZO2</td>
<td>TJP2</td>
<td>R</td>
<td>Q9UDY2</td>
<td>ENSG00000119139</td>
</tr>
<tr>
<td>Endothelial collagen</td>
<td>COL8A1</td>
<td>H</td>
<td>P27658</td>
<td>ENSG00000144810</td>
</tr>
<tr>
<td>Sialomucin (CD34)</td>
<td>CD34</td>
<td>M</td>
<td>Q64314</td>
<td>ENSG000000174059</td>
</tr>
<tr>
<td>Tumor endothelial marker 4 (TEM4)</td>
<td>ARHGEF17</td>
<td>H</td>
<td>Q96PFE</td>
<td>ENSG000000110237</td>
</tr>
<tr>
<td>APC protein</td>
<td>APC</td>
<td>H</td>
<td>P25054</td>
<td>ENSG00000134982</td>
</tr>
<tr>
<td>PAR-1B alpha</td>
<td>MARK2</td>
<td>H</td>
<td>x</td>
<td>ENSG00000072518</td>
</tr>
<tr>
<td>Endomucin</td>
<td>EMCN</td>
<td>R</td>
<td>x</td>
<td>ENSG000000164035</td>
</tr>
<tr>
<td>Annexin IV</td>
<td>ANXA4</td>
<td>R</td>
<td>P55260</td>
<td>ENSG00000196975</td>
</tr>
<tr>
<td>Vascular cell adhesion protein 1</td>
<td>VCAM1</td>
<td>R</td>
<td>P29534</td>
<td>ENSG000000162692</td>
</tr>
<tr>
<td>Chemokine receptor CCX CKR</td>
<td>ACKR4</td>
<td>M</td>
<td>-</td>
<td>ENSG000000129048</td>
</tr>
<tr>
<td>Vascular endothelial-cadherin 2</td>
<td>PCDH12</td>
<td>H</td>
<td>Q9NPG4</td>
<td>ENSG00000113555</td>
</tr>
<tr>
<td>Tight junction protein 1 - ZO1</td>
<td>TJP1</td>
<td>M</td>
<td>P39447</td>
<td>ENSG000000104067</td>
</tr>
<tr>
<td>Purinergic receptor 5</td>
<td>LPAR6</td>
<td>H</td>
<td>P43657</td>
<td>ENSG000000139679</td>
</tr>
<tr>
<td>Ecto-apyrase (CD39)</td>
<td>ENTPD1</td>
<td>M</td>
<td>P55772</td>
<td>ENSG000000138185</td>
</tr>
<tr>
<td>MECA32, PV-1</td>
<td>PLVAP</td>
<td>M, R</td>
<td>Q9WV78</td>
<td>ENSG000000130300</td>
</tr>
<tr>
<td>Scavenger receptor class B type I</td>
<td>SCARB1</td>
<td>R</td>
<td>x</td>
<td>ENSG000000073060</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>protein name</td>
<td>Protein ID</td>
<td>ENSG ID</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>RAGE</td>
<td>AGER</td>
<td>R</td>
<td>Q63495</td>
<td></td>
</tr>
<tr>
<td>Tumor endothelial marker 6 (TEM6, Tensin 3)</td>
<td>TNS3</td>
<td>H</td>
<td>Q96PE0</td>
<td></td>
</tr>
<tr>
<td>MDR 1A</td>
<td>ABCB1</td>
<td>R</td>
<td>P21447</td>
<td></td>
</tr>
<tr>
<td>Integrin alpha-1 (CD49a)</td>
<td>ITGA1</td>
<td>R</td>
<td>P18614</td>
<td></td>
</tr>
<tr>
<td>Alpha-2 macroglobulin</td>
<td>A2M</td>
<td>M</td>
<td>P28666</td>
<td></td>
</tr>
<tr>
<td>Vascular endothelial junction-associated molecule</td>
<td>JAM2</td>
<td>H</td>
<td>P57087</td>
<td></td>
</tr>
<tr>
<td>Aquaporin-CHIP</td>
<td>AQP1</td>
<td>R</td>
<td>P29975</td>
<td></td>
</tr>
<tr>
<td>Dipeptidyl peptidase IV (CD26)</td>
<td>DPP4</td>
<td>R</td>
<td>P14740</td>
<td></td>
</tr>
<tr>
<td>Nicotinic acetylcholine receptor alpha 3</td>
<td>CHRNA3</td>
<td>M</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Platelet-derived growth factor receptor</td>
<td>PDGFRB</td>
<td>M</td>
<td>P05622</td>
<td></td>
</tr>
<tr>
<td>P2Y purinoceptor 6</td>
<td>P2RY6</td>
<td>R</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Nitric-oxide synthase</td>
<td>NOS1</td>
<td>R</td>
<td>P29476</td>
<td></td>
</tr>
<tr>
<td>Scavenger receptor (CD36)</td>
<td>CD36</td>
<td>R</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>EGFR</td>
<td>EGF</td>
<td>H</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Vascular adhesion protein-1</td>
<td>AOC3</td>
<td>M</td>
<td>O70423</td>
<td></td>
</tr>
<tr>
<td>Muscarinic acetylcholine receptor M3</td>
<td>CHRM3</td>
<td>R</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Angiotensin II receptor (1 or 2)</td>
<td>AGTR1</td>
<td>R</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Nicotinic acetylcholine receptor beta 3</td>
<td>CHRN3</td>
<td>R</td>
<td>P12391</td>
<td></td>
</tr>
<tr>
<td>Carbonic anhydrase IV</td>
<td>CA4</td>
<td>R</td>
<td>P48284</td>
<td></td>
</tr>
<tr>
<td>Bradykinin receptor B2</td>
<td>BDKRB2</td>
<td>R</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Adrenergic receptor alpha 2B</td>
<td>N/A</td>
<td>H</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>
Appendix H: List of supplementary tables

Supplementary files listed below and this list are available at
https://doi.org/10.17863/CAM.10322 in .csv and .pdf format.

Chapter 3

Table S3.1  D13 vs D6 down-regulated genes.csv
Table S3.2  D13 vs D6 up-regulated genes.csv
Table S3.3  D13 vs D6 GO term enrichment.csv
Table S3.4  Genes with high expression during liver stages D13 to D21.csv
Table S3.5  GO term enrichment of genes with high expression during liver stages.csv
Table S3.6  D28 vs D21 down-regulated genes.csv
Table S3.7  D28 vs D21 up-regulated genes.csv
Table S3.8  D28 vs D21 GO term enrichment.csv
Table S3.9  D35 vs D28 down-regulated genes.csv
Table S3.10 D35 vs D28 up-regulated genes.csv
Table S3.11 D35 vs D28 GO term enrichment.csv

Chapter 4

Table S4.1  GO term enrichment of genes differentially expressed at day 17.csv
Table S4.2  HEPG2 vs non-HEPG2 up-regulated genes.csv
Table S4.3  HEPG2 vs non-HEPG2 down-regulated genes.csv
Table S4.4  HEPG2 vs non-HEPG2 GO term enrichment.csv

Chapter 5

Table S5.1  HUVEC co-culture vs worm-free up-regulated genes.csv
Table S5.2  HUVEC co-culture vs worm-free down-regulated genes.csv
Table S5.3  HUVEC co-culture vs worm-free GO enrichment.csv
Table S5.4  HUVEC co-culture vs worm-free pathway enrichment.csv
Table S5.5  HEPG2 co-culture vs worm-free up-regulated genes.csv
Table S5.6  HEPG2 co-culture vs worm-free down-regulated genes.csv
Table S5.7  HEPG2 co-culture vs worm-free GO enrichment.csv
<table>
<thead>
<tr>
<th>Table S5.8</th>
<th>HEPG2 co-culture vs worm-free pathway enrichment.csv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table S5.9</td>
<td>GripTite co-culture vs worm-free up-regulated genes.csv</td>
</tr>
<tr>
<td>Table S5.10</td>
<td>GripTite co-culture vs worm-free down-regulated genes.csv</td>
</tr>
<tr>
<td>Table S5.11</td>
<td>GripTite co-culture vs worm-free GO enrichment.csv</td>
</tr>
<tr>
<td>Table S5.12</td>
<td>GripTite co-culture vs worm-free pathway enrichment.csv</td>
</tr>
</tbody>
</table>