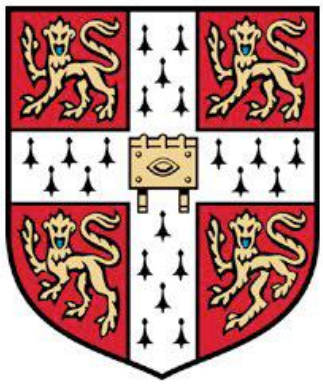
Transcriptome characterisation of the intra-mammalian stage of male and female *Schistosoma mansoni*



Andreas Florian Sessler

Darwin College

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

October 2017

Declaration

This dissertation is the result of my own work – all work done in collaboration being clearly references in the text. The work presented here was performed at the Wellcome Trust Sanger Institute (Hinxton). None of the work has been submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University. This thesis does not exceed the word limit established by the Biology Degree Committee.

Acknowledgements

During my time at the Sanger Institute I had the fortune of being helped and advised by many people. First, I'd like to thank my supervisor Matt (Dr. Matt Berriman) for his support and guidance over the last few years. This work would not have been possible without him. I also want to thank all of Team 133 for their immense help and friendship. In particular, Anna (Dr. Anna Protasio) for sharing her knowledge and experience in all things "Schisto" with me, Lia (Dr. Lia Chappell) and Hayley (Dr. Hayley Bennett) for all their help in the laboratory as well as Adam (Dr. Adam Reid) for sharing his bioinformatics skills with me.

Furthermore, I would like to thank several other people at the Sanger Institute, including the pathogen informatics team for their constant support. I also want to thank Dr. Simon Clare and Cordelia Brandt for their help with the infections and perfusions and Dave Goulding for his help with electron microscopy. Next, I would also like to thank Prof. Dr. Christoph Grevelding, Dr. Thomas Quack and Dr Zhigang Liu from the Institut für Parasitologie - Justus-Liebig-Universität Gießen for the productive collaboration, the results of which form part of this thesis. I also want to thank Prof. Andrew MacDonald and Prof. Mike Doenhoff for providing schistosome material at many points during my work as well as Dr Peter Olson for hosting me at the Natural History Museum to learn about *in situ* hybridisation.

Finally, I would like to thank Aki MacFarlane and my family for their continuous support throughout my work at the Sanger Institute.

Summary

Transcriptome characterisation of the intra-mammalian stage of male and female *Schistosoma mansoni*

Andreas Florian Sessler

Schistosoma mansoni is a member of a genus of platyhelminths whose members cause the disease *schistosomiasis*. Particularly prevalent in sub-Saharan Africa, it is thought to be directly responsible for approximately 5500 deaths per year, as well as contributing significantly to morbidity, being responsible for 3.3 million lost disability-adjusted life years. Schistosomes are dioecious and male and female worms find one another and pair in the blood vessels of the host's liver. This sets in motion a unique feature of schistosome biology, the pairing-dependent sexual maturation of the female worms. Over the course of the next three weeks, the females fully develop their reproductive organs, especially ovaries and vitellarian tissue, to allow for the production of large quantities of eggs, which not only play a crucial role in the transmission of the parasites but are also responsible for much of the pathology associated with schistosomiasis.

This thesis aims to explore the changes in gene expression which take place following pairing and result in the sexual maturation of females. To do so, RNA-Seq data was produced from male and female worms from mixed sex as well as single sex infections at 18, 21, 28, 35, 38 and 49 days *post* infection and analysed to understand when and how gene expression changes in paired worms. Then gene expression was examined in worms that had been removed from their partner to examine the process of regression, where female worms lose much of their reproductive tissue. The last experiments describe examine gene expression in the testes and ovaries of schistosomes, to reveal differences between the gonads of worms from mixed and single sex infections and understand in more detail how these worms may regulate the growth of their reproductive organs, contributing to our knowledge of schistosome biology.

Abbreviations

BME	Basal Medium Eagle
bp	Base pairs
Cdc	Cell division cycle
CDK	Cyclin Dependant Kinase
CO_2	Carbon dioxide
d.p.i.	Days *post* infection
DEG	Differentially Expressed Gene
DMEM	Dulbecco's Modified Eagle's Medium
dsRNA	Double stranded RNA
ECM	Extracellular Matrix
EGF	Epidermal Growth Factor
FGF	Fibroblast Growth Factor
GAP	GTPase-Activating Protein
GO	Gene Ontology
kb	Kilo bases
MAP	Mitogen-Activated Protein
MAPK	Mitogen-Activated Protein Kinases
Mb	Mega bases
MS	Mixed Sex
NaCl	Sodium Chloride
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
qRT-PCR	quantitative Reverse Transcription-Polymerase Chain Reaction
RNA-Seq	RNA Sequencing
RNAi	RNA interference
RPM	Revolutions Per Minute
RT	Room Temperature
SMAD	(Portmanteau of SMA and MAD)
Src	Sarcoma
SS	Single Sex
TBE	Tris Borate EDTA
TGF-β	Transforming Growth Factor-β
WNT	(Portmanteau of int and Wg)

Table of Contents

CHAPTER 1

INTRODUCTION 1
1.1 INTRODUCTION 1
1.2 BIOLOGY OF SCHISTOSOMES 2
1.2.1 Life cycle 3
1.3 IMPACT ON HUMAN WELFARE 10
1.3.1 Chemotherapy and control of schistosomiasis 11
1.3.2 Pathology of schistosomiasis 13
1.4 MALE-FEMALE INTERACTION 15
1.5 MOLECULAR BIOLOGY OF MALE-FEMALE INTERACTION 18
1.6 GENOMICS & TRANSCRIPTOMICS 24
1.6.1 Genome 24
1.6.2 Transcriptomics of sex-specific schistosome biology 25
1.6.3 Methods of transcriptomics 33
1.7 AIMS OF MY PROJECT 39

Chapter 2

MATERIALS & METHODS 42
2.1 PARASITE AND SNAIL MAINTENANCE 42
2.1.1 Parasite and snail origin 42
2.1.2 Miracidia infection of snails 42
2.1.3 Collection of cercariae 46
2.1.4 Infections and perfusions 46
2.1.5 *In vitro* culture 47
2.1.6 Isolation of *Schistosoma mansoni* gonads 48
2.2 MOLECULAR BIOLOGY TECHNIQUES 49
2.2.1 Cloning 49

2.2.2 RNA interference 51
2.2.3 RNA extractions 54
2.2.4 cDNA synthesis 55
2.2.5 qRT-PCR 56
2.2.6 Dot blot 58
2.2.7 Whole mount *in situ* hybridisation 58
2.2.8 Imaging whole mount *in situ* hybridisation specimen 62
2.3 RNA-SEQ LIBRARY PREPARATION & SEQUENCING 62
2.3.1 mRNA selection 62
2.3.2 mRNA fragmentation 63
2.3.3 Reverse transcription 63
2.3.4 Second strand DNA synthesis 64
2.3.5 End repair, dA tailing, adapter ligation, size selection 64
2.3.6 PCR amplification 65
2.3.7 Sequencing 65
2.4 BIOINFORMATICS 66
2.4.1 Aligning RNA-Seq reads to the genome 66
2.4.2 Sorting and merging of BAM files 67
2.4.3 Counting reads 67
2.4.4 Differential gene expression analysis 67
2.4.5 Correlation of RNA-Seq data with microarray data 70
2.4.6 Principal Component Analysis (PCA) 71
2.4.7 Heatmaps 71
2.4.8 Gene Ontology (GO) term enrichment 72
2.4.9 InterProScan and Pfam enrichment 72
2.4.10 KEGG pathway enrichment 73
2.4.11 Cluster analysis of RNA-Seq data 75
2.4.12 Gene models & annotation 76
2.5 SCANNING ELECTRON MICROSCOPY 77

CHAPTER 3

TRANSCRIPTOME ANALYSIS OF MALE AND FEMALE *SCHISTOSOMA MANSONI* DURING THE INTRA-MAMMALIAN DEVELOPMENT 78

3.1 INTRODUCTION 79

3.2 RESULTS 82

3.2.1 Egg laying 82

3.2.2 Sequencing & sample clustering 83

3.2.3 Gender-specific gene expression 85

3.2.4 Male development 95

3.2.5 Female development 99

3.2.6 Male time series analysis 104

3.2.7 Female time series analysis 107

3.2.8 Fertility-related genes 108

3.2.9 qRT-PCR analysis of CD63 antigen & CD63 receptor 117

3.2.10 Whole mount *in situ* hybridisation 119

3.2.11 RNA interference 122

3.3 DISCUSSION 133

CHAPTER 4

TRANSCRIPTOME ANALYSIS OF SEXUALLY REGRESSING *SCHISTOSOMA MANSONI* 141

4.1 INTRODUCTION 142

4.2 RESULTS 147

4.2.1 Optimising culture media 147

4.2.2 Apoptosis related genes 150

4.2.3 Sequencing & sample clustering 153

4.2.4 Comparison of paired and separated females at day 8 154

4.2.5 Comparison of paired and single males at day 8 159

4.2.6 Time series analysis 162

4.2.7 Comparison of females before and after *in vitro* culture 169

4.2.8 Comparison of males before and after *in vitro* culture 182

4.3 DISCUSSION 198

CHAPTER 5

EXPLORATION OF THE *S. MANSONI* GONAD TRANSCRIPTOME 205
5.1 INTRODUCTION 206
5.2 RESULTS 211
5.2.1 Sequencing & sample clustering 211
5.2.2 Comparing RNA-Seq to published DNA microarray & qRT-PCR data 214
5.2.3 Testes transcriptome 219
5.2.4 Ovary transcriptome 227
5.2.5 Comparing the testes and ovary transcriptome 232
5.2.6 Effect of pairing on testes transcriptome 235
5.2.7 Effect of pairing on ovary transcriptome 238
5.3 DISCUSSION 257

Chapter 6

CONCLUDING REMARKS 276

Chapter 7

APPENDICES 277
7.1 APPENDIX A 278
7.2 APPENDIX B 306
7.3 APPENDIX C 322

REFERENCES 348

Chapter 1

Introduction

1.1 Introduction

Schistosomiasis is caused by a genus of platyhelminths called *Schistosoma;* an estimated 200 million people require treatment for the disease annually, approximately 90% of whom live in sub-Saharan Africa (WHO, 2018). According to the Global Burden of Disease Study the death of 5500 people was directly caused by schistosomiasis in 2013 (Naghavi *et al.*, 2015). However the precise number is difficult to estimate because of the difficulties of measuring the contribution of chronic schistosomiasis to conditions such as kidney failure and bladder cancer and estimates are as high as 200,000 deaths per year globally WHO (2018).

1.2 Biology of schistosomes

Schistosoma mansoni as well as several other species of schistosomes, notably *S. japonicum* and *Schistosoma haematobium*, are responsible for schistosomiasis. Schistosomes belong to the class Trematoda, which usually have a flattened, oval body shape ranging in length from around 0.1 cm to 7 cm with a ventral as well as an oral sucker (Barnes and Robert, 1982). Schistosomes are around 0.6-1.1 cm long and 0.1 cm wide, with a long, slim body (Machado-Silva *et al.*, 1997). The morphology of male and female worms differs considerably (Loker & Brant, 2006). Male worms maintain the typical flattened body shape of trematodes but curled up to form the gynecophoral canal, whereas the female worms have a cylindrical profile (Figure 1.1).

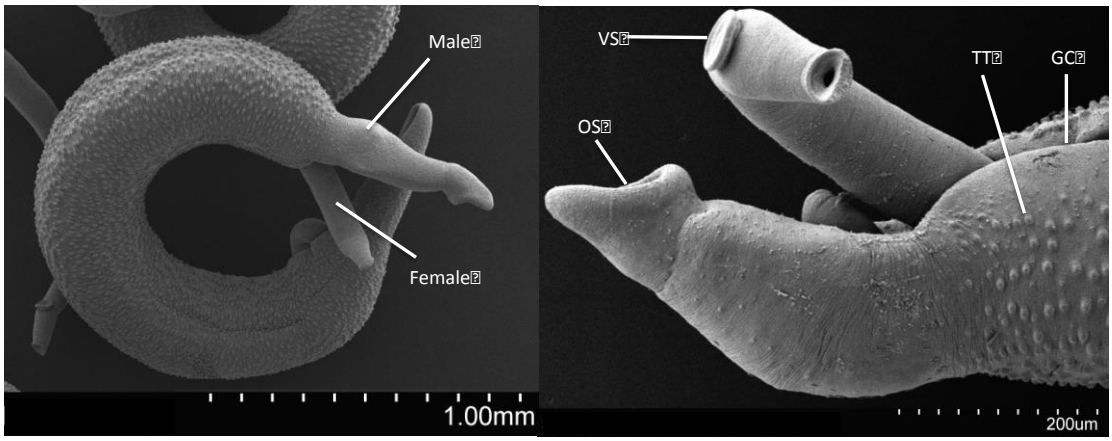


Figure 1.1: Scanning electron microscopy of adult *Schistosoma mansoni* couples. After pairing, the female worm resides inside the gynecophoral canal (GC) of the male. Both worms have an oral (OS) as well as a ventral sucker (VS). The tegument of male worms is covered with tegumental tubercles (TT). These images were taken in collaboration with Dave Goulding, WTSI, Hinxton.

1.2.1 Life cycle

The *S. mansoni* life cycle (see Figure 1.2) is similar to that of other trematodes and involves a mollusc and a mammalian host (Tucker *et al.*, 2013). However, unlike most trematodes, the members of the Schistosomatidae, the family *S. mansoni* belongs to, are dioecious (Loker & Brant, 2006). This means that there are distinct male and female worms instead of hermaphrodites. The life cycle of *S. mansoni* features two hosts, a molluscan intermediate host, in which the parasites undergo asexual reproduction, and the vertebrate definitive host, in which the adult worms reproduce sexually and lay eggs (Tucker *et al.*, 2013). When the eggs are placed in fresh water, ciliated larvae called miracidia (136 ± 3 µm in length and 54 ± 1 µm in width (Eklu-Natey *et al.*, 1985) are released. *S. mansoni* miracidia invade the intermediate host – fresh water snails of the genus *Biomphalaria* – by burrowing through the skin of its foot (Cosseau *et al.*, 2009). Inside the snail, the miracidia develop into primary (or mother) sporocysts,

which in turn produce secondary (or daughter) sporocysts by asexual reproduction (Cosseau *et al.*, 2009).

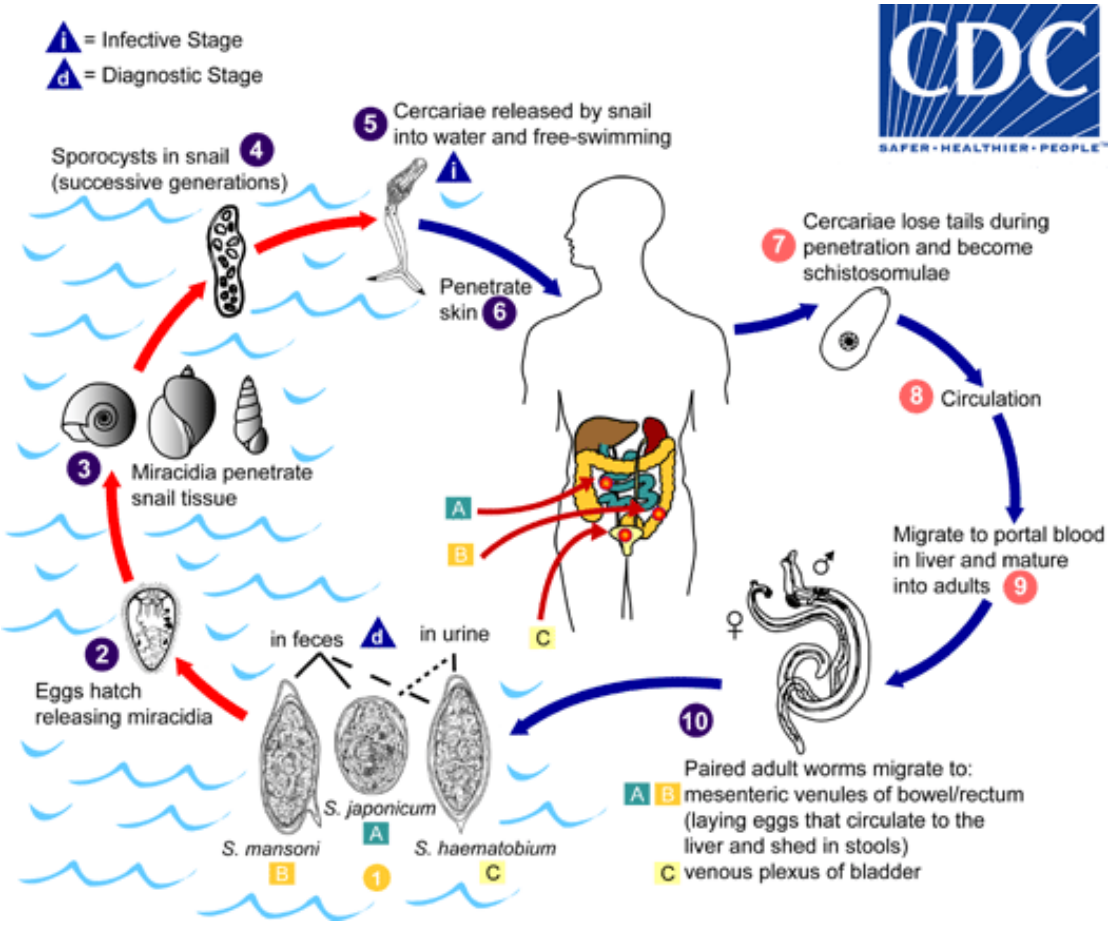


Figure 1.2: Life cycle of *Schistosoma* species. 1) Eggs are excreted via faeces or in the case of *S. haematobium* in the urine and reach fresh water. 2) Miracidia hatch from the eggs and seek a snail host. 3) The miracidia penetrate a snail host (in the case of *S. mansoni* a snail of the genus *Biomphalaria*) and 4) two generations of sporocysts develop from the miracidia inside the host. 5) Cercariae are produced and released into the water. 6) Cercariae penetrate the human skin, 7) shedding their tail and transforming into schistosomulae. 8) After moving through the circulation and lungs for several weeks, 9) male and female adult worms pair in the liver and 10) relocate to the mesenteric system of the bowel (or the venous plexus of the bladder in the case of *S. haematobium*). The couples then start egg-laying, thereby continuing the life cycle. **Reproduced from CDC (2012); https://www.cdc.gov/parasites/schistosomiasis/biology.html**

Each daughter sporocyst produces a clonal population of cercariae, the life cycle stage infecting humans (McKerrow & Salter, 2002). By shedding during daylight hours, *S. mansoni* maximise the chances of the cercariae being released when a suitable mammalian host might be nearby (Fingerut *et al.*, 2003). The cercariae use their forked tail for locomotion and use changes in light and the presence of medium-chain fatty acids in the skin to detect a host (McKerrow & Salter, 2002). When a cercaria comes into contact with a suitable host, it attaches to the skin and uses proteases to break into the skin, migrating into the subcutaneous tissue, losing its tail in the process (Salter *et al.*, 2000).

Following host invasion, the parasites transform into schistosomules. They spend around days traveling through the skin and circulation, migrate through the lungs (peaking here around 5-6 d.p.i.), then continuing over the course of the next two weeks until they enter the portal vessel of the liver (P. Miller & Wilson, 1980). There, they begin to feed on blood as the gut develops and they begin to grow. At this stage the male worms become visibly larger than the females (Basch, 1991) From around 21 day *post* infection (d.p.i.) onward the parasites mature and males and females commence pairing around 28 d.p.i. Basch, 1991). It is well established that female worms require pairing with a male partner to develop mature gonads and vitellaria to become sexually mature (Kunz, 2001). Without pairing, female worms remain stunted and infertile (Collins & Newmark, 2013). Around 35 d.p.i., *S. mansoni* pairs have usually migrated into the mesenteric blood vessels and maturing females produce the first eggs (Basch, 1991). However, there is considerable variation in the time it takes worms to migrate through different parts of their hosts' bodies, depending on the species

of worm in question as well as the host species – mice, rats, hamsters and humans – varying by as much as 20 days (Yolles *et al.,* 1949; Wilson, 1978; Crabtree & Wilson, 1986; Miller & Wilson, 1980).

After *S. mansoni* pairs have migrated to the mesentery and commenced egg production, they will often continue to produce eggs for years if left untreated (Harris *et al.*, 1984) Approximately half of the eggs get trapped in various organs (LoVerde *et al.*, 2005), where they cause considerable damage (see Chapter 1.3.2). The remaining eggs cross through the wall of the intestine into the lumen of the gut from where they are excreted with faeces (deWalick *et al.*, 2011) If eggs wash into fresh water, miracidia hatch, ready to invade the snail intermediate host, and the cycle repeats.

Aside from the migration through the host, another critical aspect of worm development is the morphological changes that take place during the first 6-7 weeks after infection. Cort (1921) describes the development of male and female schistosomes from mixed sex (MS) infections beginning at the lung stage (around 7 d.p.i) until the worms have matured fully and commence oviposition (around 35 d.p.i.). The author notes that at 7 d.p.i. the young worms remain at approximately the same size as schistosomules with the gut starting to develop. The first differences between male and female worms are described at 15 d.p.i., when the worms reach the portal vessel of the liver, commence feeding on blood and begin to grow. Males are larger and have more developed oral and ventral suckers at this time. At around 21 d.p.i. organogeny begins: male worms develop testes and in females a uterus becomes visible. After another week of maturation,

at around 28 d.p.i., gametogeny is thought to take place. Males have been shown to possess 8 testes at 28 d.p.i., some of which contain sperm, and females to have developed ovaries. At 28 d.p.i., Cort (1921) first observed pairing of worms, but also showed that maturation of the vitellarium in particular requires a further week. At 30 d.p.i., the author demonstrated the presence of vitelline globes, but found that oviposition generally begins to occur at 35 d.p.i. At this point, females have well-developed vitelline glands to provide a steady supply of egg shell proteins, whilst male worms produce large quantities of sperm for fertilisation. Very similar observations were made by Biolchini *et al.* (2006), but technological improvements over the 85 years that separated these two studies, most notably confocal microscopy, allowed them to study the development of the worm organs in more detail. Indeed, the latter authors noticed that at week three 47-59% of males had developed the gynecophoral canal, 5% of females had developed an ootype, and 33% of females had developed ovaries, whereas Cort (1921) did not observe ovaries at this time and made no mention of the gynecophoral canal development. At week four, Biolchini *et al.* (2006) found that 15-26% of male worms began to develop tegumental tubercles (Figure 1.1). Also, 69% of females were found to have ovaries and 12-75% an ootype, with a single female presenting with a fertilised egg. In addition, Biolchini *et al.* (2006) found the oral suckers of male worms to be significantly larger in size than that of females at 21-28 d.p.i., an important difference at 21 d.p.i. for distinguishing worms morphologically. Both papers noted the significance of pairing in the development of the females, which, unlike male worms, do not mature sexually until paired. In particular, development of the vitellarium and maturation of the ovaries was observed only following pairing (Biolchini *et al.*, 2006; Cort, 1921).

Notably, the development of *S. japonicum* appears to occur at a slightly faster pace, as most worms were found to be paired by 22 d.p.i. and the first sperm and mature oocytes are first observed at the same time, presumably in worms that paired particularly early, and sexually mature worms are found at 28 d.p.i. (Wang *et al.*, 2017).

The induced maturation of the reproductive organs of a paired female worm culminates in the production of eggs. The function of the ovary is to produce viable, fertilised oocytes, the vitellarium, on the other hand, produces eggshell precursor proteins and provides nutrient reserves for the developing embryo in the form of lipid droplets (LoVerde *et al.*, 2005). After oocytes have become fertilised with the male sperm inside the female oviduct, the fertilised cells move to the vitello-oviduct, where they are surrounded by approximately 38 vitellocytes (Loverde & Chen, 1991) (Figure 1.3). The maturation of vitellocytes is commonly divided into four stages (Erasmus, 1975): The first is a stem cell stage of undifferentiated, small cells near the periphery of the vitelline lobes (S1). Then the cell begins to differentiate and grow in size (S2). The cells then begin to synthesise large amounts of proteins and form vitelline droplets; while the rate of protein synthesis increases the vitelline droplets begin to form clusters. Subsequently, lipid droplets and a small number of β-glycogen particles accumulate to provide energy to the embryo (S3). The final stages of development, mature vitellocytes (S4), leave through the vitelline duct to reach the ootype (Figure 1.3) (Wang & Collins, 2016). Approximately 11,000 S4 vitellocytes are produced daily by mature females (LoVerde *et al*., 2005) and to

satisfy this demand the vitellarium takes up a large portion of the female worm's body (Galanti *et al.*, 2012).

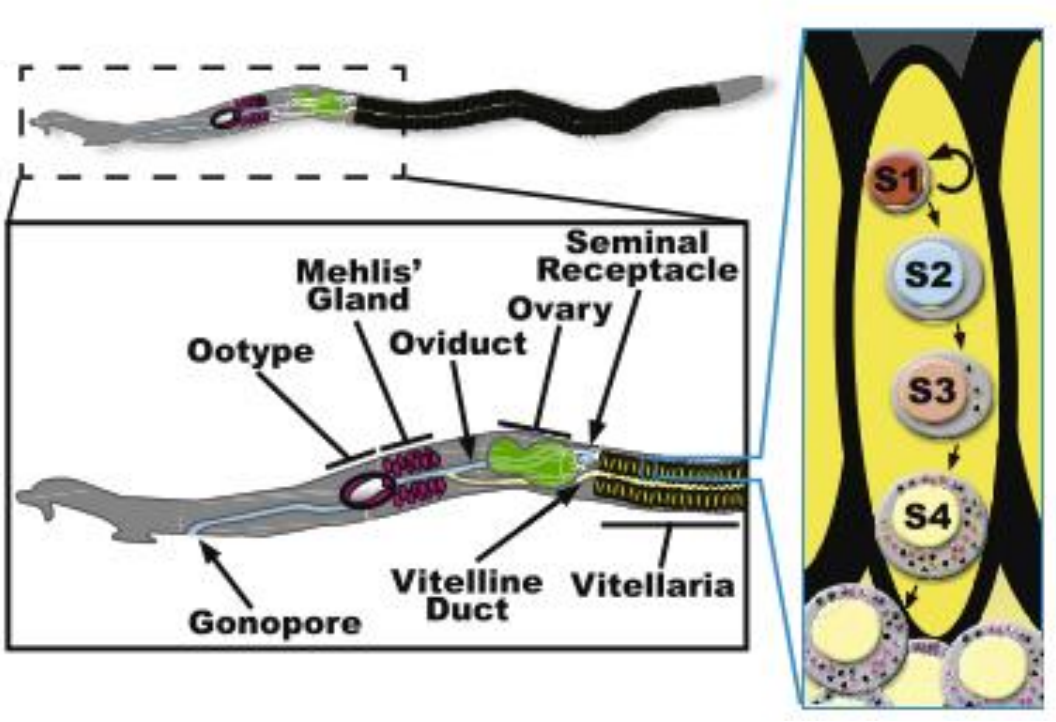


Figure 1.3: Diagram of the female reproductive organs, reproduced from Wang & Collins (2016). Left: The relative positions are shown for the gonopore, ootype, Mehlis' gland, oviduct, vitelline duct, ovary, vitellaria and seminal receptacle. Right: vitellocytes move towards the central vitelline duct as they mature through a sequence of stages (S1-S4); stem cell-like S1 cells, proliferating S2 cells, larger S3 vitellocytes accumulating granules, and finally mature S4 vitellocytes.

Each paired female has been estimated to lay around 300 eggs per day for several years (Loverde & Chen, 1991). Schistosome are estimated to live for 2.7-4.5 years (3.3 on average) (Goddard & Jordan, 1980) although there are reports of worms surviving and continuing to produce viable eggs for up to 47 years (Harris *et al.*, 1984; Markel *et al.*, 1978; Warren *et al.*, 1974). Adult *S. mansoni* cannot reproduce asexually and female reproductive maturity is critically dependent on pairing with a male worm. Not only does transmission rely on

efficient egg production but also the pathology caused by schistosomes is in large parts driven by the host immune reaction against the eggs of these parasites and the resulting inflammation and formation of granulomas in the liver and spleen (Gryseels *et al.*, 2006). For these reasons, the process of pairing-induced maturation is an attractive target for intervention and an intensely studied aspect of schistosome biology (Loverde & Chen, 1991).

1.3 Impact on human welfare

According to the WHO (2018) Approximately 200 million people required treatment for schistosomiasis in 2016, of which only about 90 million receive it. Furthermore, the WHO (2018) estimates that approximately 90% of those infected live in Africa, with limited access to medical care. Hence the number of schistosomiasis-related deaths is considerable, but much of it impact is related to chronic morbidity (WHO, 2016). To accurately reflect the impact of a disease not just in terms of lives lost but also taking into account the morbidity caused by diseases such as schistosomiasis a disability weight is used. This is a weight factor reflecting the magnitude of the health loss associated with the5 disease on a scale from 0-1, representing full health and death respectively (Salomon *et al.*, 2015). As part of the Global Burden of Disease Study, schistosomiasis was attributed a disability weight of 0.6% for the year 1990 (Murray & Lopez, 1997), although other studies put the disability weight as high as 2-15% (King *et al.*, 2005). In the most recent Global Burden of Disease Study, the estimate of total disability-adjusted life years (DALYs) lost due to schistosomiasis was raised from 2.125 million to 3.309 million DALYs (Murray *et al.*, 2012), reflecting the higher

estimates of disease prevalence as well as revised estimates of mortality and morbidity as measured by DALYs. Schistosomiasis transmission has been reported in 78 countries, including European countries such as Portugal and more recently France (Berry *et al.*, 2014). Schistosomiasis is considered to be endemic in 52 out of 78 countries (WHO, 2018). Although sub-Saharan Africa is most affected, schistosomiasis is also found in countries along the Atlantic coast of South America and the Caribbean as well as several Southeast Asian countries (Gryseels *et al.*, 2006).

1.3.1 Chemotherapy and control of schistosomiasis

Currently the only drug available to fight schistosomiasis effectively is praziquantel, which is used heavily in large parts of Africa with the declared aim of morbidity control (WHO, 2018). That means that rather than aiming to eradicate the disease in these areas, the aim is to reduce the intensity of the infection in the population and thereby lower the frequency of complications associated with schistosomiasis. The WHO considers schoolchildren to be target populations for these mass treatment programs, as well as at-risk adults such as fishermen or farmers who work in irrigation ditches and communities who heavily depend on water in endemic regions (WHO, 2018). Evidence has shown that pre-school children can also carry high intensity infections. For example, Dabo *et al.* (2011) found between 15.7% of preschool children in different villages in Mali to have high intensity infections, determined by an egg count of ≥50eggs/ml in their urine. However, there are concerns about the safety of praziquantel in very young children due to a lack of reliable safety data (Stothard

et al., 2013). This mass treatment strategy, although representing the most cost-effective method in high-prevalence settings (WHO, 2018), has major limitations: Firstly, it does not eradicate schistosomiasis in the regions covered but, as stated above, aims to control morbidity. Secondly, it leaves the vector population unaffected, including infected snails. For this reason, there is a risk of reinfection and so mass treatment needs to be repeated regularly to provide long-term benefit, which puts further strain on limited resources (WHO, 2018). Thirdly, due to mass drug administration, there is an increased pressure on parasite populations to become resistant to praziquantel - indeed reports of reduced sensitivity to praziquantel have been published in recent years (Crellen *et al.*, 2016; Wang *et al.*, 2014).

Unfortunately, despite decades of research into the development of a schistosomiasis vaccine, no successful formulation has been yet found (Hagan & Sharaf, 2003; McManus & Loukas, 2008). There is evidence that protective immunity against percutaneous infection with schistosomes can be raised with radiation-attenuated cercariae (Bickle, 2009; El Ridi & Tallima, 2015). Reductions in worm burdens up to 40-60% in treated mice were observed (Bickle *et al.*, 1979). However, no single- or multi-valent vaccine tested to date has been able to advance from clinical trials. As a result, the search for viable vaccine targets continues including "–omics" approaches such as transcriptomics, proteomics but also glycomics and immunomics for candidate generation (Loukas *et al.*, 2011) and several promising candidates are being investigated (Merrifield *et al.*, 2016). These include a fatty acid binding protein (Sm14), tetraspanin-2 (SmTSP-2), glutathione S-transferase (Sh28GST) and

calpain (Sm80) (Tebeje *et al.*, 2016). While vaccines using Sm14, Sh28GST and SmTSP-2 have already been tested in humans (Bourke *et al.*, 2014; Merrifield *et al.*, 2016; Santini-Oliveira *et al.*, 2016) the vaccine using calpain has been tested in mice and baboons (Ahmad *et al.*, 2010; Ahmad *et al.*, 2011) and is now undergoing preparations for human trials (Rojo *et al.*, 2017).

1.3.2 Pathology of schistosomiasis

Acute pathology

Individuals who are exposed to schistosomal cercariae for the first time, such as travelers or migrants are prone to developing acute symptoms in response to infection (Paris & Caumes, 2010; Ross *et al.*, 2007). It has been suggested that the low rate of acute schistosomiasis in exposed populations, especially children, is due to a combination of underdiagnoses and *in-utero* sensitization (Gryseels *et al.*, 2006). Within the first hours or days after penetration of the skin by cercariae, some may develop a temporary rash or lesions at the site of penetration (Gryseels *et al.*, 2006). Over the weeks or sometimes months following the infection, some individuals may experience an acute systemic hypersensitivity to the developing worms also known as Katayama fever (Ross *et al.*, 2007); symptoms include fever, fatigue, myalgia, malaise, cough, eosinophilia, and lung lesions (Gryseels *et al.*, 2006). Although most patients recover within 2-10 weeks, some go on to develop abdominal symptoms as well, including weight loss, diarrhoea, abdominal pain, hepatosplenomegaly, and rashes (Bottieau *et al.*, 2006).

Chronic pathology

As is the case for many infectious diseases, the severity of symptoms is determined by a combination of several factors: the intensity of infection, the host immune response and environmental factors. The symptoms of chronic schistosomiasis include abdominal pain, diarrhea and rectal bleeding (or haematuria in the case of an *S. haematobium* infection). Over the course of the infection the patients can develop anaemia, hepatosplenomegaly, an enlargement of liver and spleen as the result of fibrosis, and periportal fibrosis.

Most of the chronic pathology caused by the infection is not inflicted by the adult worms. Instead it results from the host immune response against the eggs during their perivesicular or periintestinal migration or once they have become trapped in organs such as liver, spleen and lung (Gryseels *et al.*, 2006). The human immune system responds to *S. mansoni* infections in a T helper 2-associated manner (Colley & Secor, 2014). This includes the use of parasite-specific antibodies, eosinophils and alternatively activated macrophages (Barron & Wynn, 2011; Colley & Secor, 2014), driven by the release of interleukins 4 and 13 (Pagán & Ramakrishnan, 2018). These interleukins help to modulate the immune repose to a state of chronic inflammation and the formation of granulomas around *S. mansoni* eggs (Colley & Secor, 2013). The eggs play an active role in this process through the secretion of proteolytic enzymes that provoke an eosinophilic pro-inflammatory and granulomatous response (Cass *et al.*, 2007; Jenkins *et al.*, 2007; Everts *et al.*, 2009). In the case of *S. haematobium*, the deposition of eggs around the bladder can also result in squamous cell carcinoma of the bladder (Nesi *et al.*, 2015). In the liver, this immune response can lead to

fibrosis and later cirrhosis (Wynn, 2015). This, in turn, can result in potentially lethal pulmonary hypertension and oesophageal varices (Colley & Secor, 2013). In other parts of the body this immune response can cause genital, urinary, intestinal and pulmonary schistosomiasis as well as neuroschistosomiasis (Gryseels *et al.*, 2006). Individuals with genital schistosomiasis may have symptoms including inflammatory lesions in reproductive organs that can lead to infertility as well as increasing the transmission of sexually transmitted infections, such as HIV (Gryseels, 2012).

Apart from the potentially irreversible damage to the internal organs, schistosomiasis can have a range of more subtle effects. These effects include, but are not limited to, fatigue, stunted physical and cognitive development of young children, as well as anaemia and exacerbation of malnutrition (Mcgarvey, 2000; Sousa-Figueiredo *et al.*, 2012). As these conditions affect large numbers of people in endemic areas, it becomes clear that successful intervention would be of great benefit to these communities (Gryseels *et al.*, 2006).

1.4 Male-female interaction

The dependence of female *S. mansoni* on pairing with a male for sexual maturation is long established (Kunz, 2001). Mating usually occurs in a parallel fashion (anterior to anterior), but about 12% of *S. mansoni* pairs are found to be paired in a reverse orientation (Faust *et al.*, 1934). Although it is usual for *S. mansoni* females to pair with one male only, it has been observed, both with freshly perfused worms as well as *in vitro*, that multiple males will clasp onto a

single female. (Basch, 1991). In other members of the family *Schistosomatidae*, the number of females paired with a single male can be greater. For example, Chu & Cutress (1954) found male *Austrobilharzia variglandis* to be paired with 3 to 8 females at once. Furthermore, Armstrong (1965) found a male *Heterobilharzia americana* to be paired simultaneously with 29 female *S. mansoni*, all of which appeared to be sexually mature. Male *S. mansoni* have a propensity to pair and in the absence of females will clasp onto objects of similar shape and size to female worms, *e.g.* made from cotton or agar (Basch & Nicolas, 1989). Even *in vitro,* worms pair or re-pair quickly: Michaels & Prata (1968) observed that when keeping male and a female worm together *in vitro*, the proportion of paired worms increased steadily. Out of a total of 140 potential pairs, after 1 day in culture 20.7% of worms were paired, 84.2% after 2 days, 94.5% after 3 days and 100% after 5 days.

When trying to determine the nature of the stimulus that leads to female sexual maturation, the first processes to be examined were insemination as well as the transfer of metabolites, such as carbohydrates, cholesterol, lipids, amino acids and polypeptides, to the female (Basch, 1991). Shaw (1977) demonstrated that pairing alone is enough to induce some development of vitelline tissue and occasionally even formation of an egg, without insemination. To do so, he allowed females from single sex (SS) infection and males from MS infections to pair, noting changes in female morphology within 5-6 days, whilst showing that ~85% of females remained unfertilised up to this point. Michaels (1969) also showed that males irradiated with X-rays readily paired and induced female maturation. Even without intact testes, this lead to successful female sexual

development and egg laying. The male stimulus was also shown not to be under the control of the cerebral ganglia of the nervous system, as decapitated males are still able to induce female maturation (Popiel & Basch, 1984); however, as the authors stated, this does not unequivocally rule out involvement of the central nervous system, as parts of the nervous system in the rest of the body were found to remain functional (Popiel & Basch, 1984). Furthermore, the stimulus is also not species-specific as females paired with males of other species (heterospecific pairing) will produce eggs, albeit at a reduced rate (Kunz, 2001).

Although the nature of the stimulation of female worms is unclear, it requires contact, occurs along the whole gynecophoral canal and is local, *i.e.* is not propagated longitudinally through the body of the female worm (Basch, 1991). The nature of the stimulus has been speculated to be physical, for example tactile, or involve transfer of hormones, nutrients and other signalling molecules (LoVerde *et al.*, 2005). Several studies have provided evidence for nutrients playing a key role in supporting female fertility (Pearce & Huang, 2015), especially the availability of red blood cells (Wang *et al.*, 2015) and fatty acids (Huang *et al.*, 2012). *S. mansoni* cannot synthesise fatty acids, relying instead on their host to provide them (Meyer *et al.*, 1970). Metabolites such as glucose (Cornford & Huot, 1981) and cholesterol (Popiel & Basch, 1986) can be transferred through the gynecophoral canal to the female which may help to fulfil the female nutritional demands.

1.5 Molecular biology of male-female interaction

Several signalling pathways thought to be involved in stimulating the growth and maturation of female *S. mansoni* have been characterised, including the transforming growth factor (TGF)-β/SMAD and the MAPK pathway (Beckmann *et al.*, 2010). In the context of female maturation they serve to transmit growth-inducing stimuli from the male worm to the nuclei of cells in the vitellarium and ovaries (Beckmann *et al.*, 2010).

The TGF-β pathway plays an important role in the development of multicellular organisms (Osman *et al.*, 2006). TGF-β is a polypeptide and a cytokine that helps to regulate cell proliferation, differentiation and apoptosis (Massagué, 2012). Signalling is initiated by transmembrane serine/threonine kinase receptors, the TGF-β receptors I & II (TβRI & TβRII) (Osman *et al.*, 2006). TGF-β signalling also relies on SMAD proteins, which were first detected in *Drosophila melanogaster* (Sekelsky *et al.*, 1995) and *Caenorhabditis elegans* (Savage *et al.*, 1996) where they were termed "mothers against decapentaplegic" (MAD) and "small body size" (SMA), respectively. Proteins found to be homologues of these two are now called SMAD by merging the two names. SMADs form trimers of two R-SMAD and one co-SMAD molecule. These act as transcription factors to activate expression of downstream genes (Shi & Massagué, 2003). Homologues of TβRs, SMADs, as well as of several other proteins that form part of the intracellular TGF-β signalling pathway, have been discovered in schistosomes (Loverde *et al.*, 2007). These include the TβRI-regulating protein FKBP12 as well as the proteins SMAD2, a receptor regulated SMAD (R-SMAD) (Wu *et al.*, 2001) and SMAD4, a

common-mediator-SMAD (co-SMAD) (Shi *et al.*, 1997). Many of these genes have been shown to be expressed in the ovaries and vitellarium of female and in the testis of male worms (Beckmann *et al.*, 2010).

To demonstrate the importance of TGF-β signalling for female reproductive activity, Knobloch *et al.* (2007) exposed *S. mansoni* pairs *in vitro* to the synthetic TβR-I inhibitor TRIKI and demonstrated a reduction in mitotic activity the vitelline cells of females by 46%, as well as a 28% reduction of laid eggs. However, as higher concentrations of TRIKI did not lead to a greater reduction in mitosis and egg laying, the authors concluded that alternative signalling pathways must be involved in the development of the vitellarium (Knobloch *et al.*, 2007).

Using another inhibitor, the Src-kinase specific herbimycin A, Knobloch *et al.* (2007, 2006) found egg laying *in vitro* to be reduced by 40-60% and mitotic activity in females (but not males) reduced by 25-75% in a dose-dependent manner. Src kinases are cytoplasmic tyrosine kinases (CTK) and hence belong to the superfamily of protein tyrosine kinases (Beckmann *et al.*, 2010). CTKs are divided into eight subcategories based on the presence or absence of domains as well as their structural arrangements and several members of the CTKs have been discovered and found to be expressed in schistosome reproductive tissues (Beckmann *et al.*, 2010).

The Src kinases SmTK3 (Kapp *et al.*, 2004) and SmTK6 (Beckmann *et al.*, 2011) have been localised to the gonads of *S. mansoni*, however, SmTK6 expression was

found to be absent from the vitellarian tissue (Beckmann *et al.*, 2011). The Fyn-type Src kinase SmTK5 is expressed in the gonads and vitellarium, and SmTK4, which belongs to the class of Syk kinases, is also expressed in the gonads but not the vitellarium (Knobloch *et al.*, 2002).. SmTK3, SmTK4 and SmTK6 were shown to interact directly with one another to regulate downstream signalling using yeast-2-hybrid screening (Beckmann *et al.*, 2010; Kapp *et al.*, 2004). Syk kinases regulate cell proliferation and progression through the cell cycle by activating expression of mitotic genes (Beckmann *et al.*, 2010) but also by regulating cytokinesis and the cytoskeleton rearrangements (Mócsai *et al.*, 2010). Syk kinases are activated by receptor tyrosine kinases (RTK), such as the epidermal growth factor receptor (EGFR) (Vicogne *et al.*, 2004) or venus kinase receptors (Vicogne *et al.*, 2003). Sky kinase activation leads to the activation of a cascade of mitogen activating protein kinases, known as the MAPK signalling pathway (Qi & Elion, 2005). At least 15 RTKs have been identified in the *S. mansoni* genome: This includes four EGFRs, two insulin receptors, two fibroblast growth factor receptors (FGFR), an Ephrin receptor, a RTK-like orphan receptor, a muscle-specific kinase receptor, a cholecystokinin tetrapeptide receptor families as well as an unknown receptor (Andrade *et al.*, 2011; Avelar *et al.*, 2011). A variety of extracellular stimuli can activate the MAPK pathway through RTKs in the cell membrane. Through a series of kinases activating one another, a MAPK such as the extracellular responsive kinase (ERK) is activated (Qi, 2005). ERK translocates into the nucleus where it phosphorylates the transcription factor ELK-1, thereby allowing it to form a heterodimer with the serum response factor and activate transcription by binding to the promoter of target genes such as c-Fos in humans (Cavigelli *et al.*, 1995).

Several components of the MAPK signalling cascade were discovered in the *S. mansoni* genome by *in silico* analysis (Andrade *et al.*, 2011) and their biological role has been further characterised since then: In planarians, ERK is now known to play a central role in the regeneration of tissue by stem cells (Tasaki *et al.*, 2011), a process thought to be similar to the growth of the female reproductive organs in *S. mansoni* in response to pairing (Collins & Newmark, 2013), whereas in *C. elegans* the MAPK pathway is known to be important in several developmental events as well as establishing meiosis in the germ line (Sundaram, 2006). Andrade *et al.* (2014) demonstrated that RNAi using long dsRNA molecules can be used *in vivo* in S*. mansoni* to investigate the role of the MAPK pathway. They successfully knocked down ERK and found no effect on worm survival but a 44% reduction in hepatic egg count relative to a negative control, suggesting that MAPK may play a role in female fertility.

Other proteins involved in regulation of the MAPK pathway for which homologues have been found in *S. mansoni* include the adapter proteins SHC and growth factor receptor-bound protein 2, which act by binding to the cytoplasmic portion of RTKs, and control recruitment of "Son of sevenless" (SOS), a guanine nucleotide exchange factor protein (Knobloch *et al.*, 2007). This in turn activates the membrane associated GTPase called Ras by catalysing the exchange of GDP with GTP, thereby activating Ras which goes on to activate the MAPK cascade (Beckmann *et al.*, 2010). Another protein, the GTPase-activating protein (GAP), promotes the exchange of GTP with GDP, thereby leading to deactivation of Ras, causing deactivation of the MAPK pathway (Beckmann *et al.*, 2010). However,

activated Src kinases can phosphorylate GAP to down-regulate its inhibitory effect on the MAPK pathway (Downward, 1997; Qi, 2005).

In addition to RTKs, the integrin receptors α and β have been found to activate focal adhesion kinase (FAK), another CTK (Guan, 1997; Mitra & Schlaepfer, 2006). Integrins are composed of two non-covalently associated transmembrane glycoprotein subunits, α and β, and primarily serve to attach the cytoskeleton to the extracellular matrix (Alberts *et al.*, 2007). They also play an important role in signalling by passing on extracellular stimuli to the FAK and often cooperate to promote cell growth, survival and proliferation (Alberts *et al.*, 2007). *S. mansoni* integrin receptors and their role in reproduction were described by Beckmann *et al.* (2012) who found that integrin-β-1 was expressed in the reproductive organs of *S. mansoni*, co-localising with SmTK3, SmTK4 and SmTK6. Using a yeast-2-hybrid screen the same authors showed that integrin-β-1 interacts with the SH2 domain of SmTK4 and with the SH3 domain of SmTK3 and SmTK6 (Beckmann *et al.*, 2012). Recently, the interaction between integrins and the Venus Kinase Receptor 1, a RTK, was shown to regulate proliferation of oocytes by apoptosis in *S. mansoni* (Gelmedin *et al.*, 2017).

Integrins often occur in complexes with tetraspanins (Tugues *et al.*, 2013). These are a group of proteins characterised by four transmembrane domains, which form so called tetraspanin enriched microdomains which allow other molecules to assemble into signalling complexes (Tugues *et al.*, 2013). Several tetraspanin homologues have been identified in *S. mansoni*, including members of the CD63 family of tetraspanins (Cogswell *et al.*, 2012; Tran *et al.*, 2010). One of the CD63

homologues is known to be expressed in a female-specific manner (Cogswell *et al.*, 2012). In humans, CD63 has been shown to down-regulate apoptosis by interacting with the protein "tissue inhibitor of metalloprotease-1" (Jung *et al.*, 2006).

Another group of kinases found in *S. mansoni* are polo like kinases (Long *et al.*, 2010), also known as Cdc5, which have been shown to regulate cell cycle progression in a wide range of organisms such as *Saccharomyces cerevisiae*, *D. melanogaster*, *C. elegans* and humans (Chase *et al.*, 2000; Llamazares *et al.*, 1991; Toczyski *et al.*, 1997). Polo-like kinases have a serine/threonine kinase domain as well as two signature motifs, known as polo boxes (Barr *et al.*, 2004). In model organisms they have been found to tightly control the G2/M transition in the cell cycle by phosphorylation of Cdc25 which in turn activates CDKs by dephosphorylation (Roshak *et al.*, 2000). This process drives the activation of mitotic events including centrosome maturation, bipolar spindle formation, activation of the anaphase promoting complex, chromosome segregation and actin ring formation in preparation for cytokinesis (Van De Weerdt & Medema, 2006). Polo-like kinases themselves are regulated by a group called polo-like kinase kinases, one member of which has been found in *S. mansoni* and is a homologue of human protein Ste20 (Yan *et al.*, 2007). Ste20-like kinases belong to the germinal center kinase subfamily and their function, like that of the integrin receptors, relates to the regulation of actin-mediated processes including cell shape and cytoskeletal rearrangements (Belkina *et al.*, 2009).

1.6 Genomics & Transcriptomics

1.6.1 Genome

S. mansoni is a diploid organism with seven pairs of autosomes, a pair of sex chromosomes and a mitochondrial genome (Swain *et al.*, 2011). Schistosomes have a ZW determination, meaning that the male is homogametic (ZZ) and the female heterogametic (ZW) (Short & Grossman, 1981).

Berriman *et al.* (2009) found the nuclear genome to be 30% larger than previously estimated, around 374.9 Mb, approximately one ninth of the human genome but considerably larger than the genome of *C. elegans* (~98 Mbp) (*C. elegans* Sequencing Consortium, 1998) or *D. melanogaster* (~120 Mbp) (Adams *et al.*, 2000). 40% of the *S. mansoni* genome is made up of repetitive sequences, including 72 families of long-terminal repeat (LTR) and non-LTR transposons each, comprising 15% and 5% respectively of the whole genome. In the most recently published version of the genome its size was revised down to 364.5 Mb and the number of genes now stands at 10,852, having been revised down from 11,807 in the original genome paper, after more than 45% of gene models were modified (Protasio *et al.*, 2012). These advances were due in part to RNA-Seq evidence that allowed for improved gene predictions and due to a less fragmented genome with fewer duplicated sequences that had resulted from polymorphism (Protasio *et al.*, 2012). The current genome assembly consists of 9203 contigs with an average length of 39.4 kb, a N50 length of 78.3 kb – a measure of assembly quality indicating that 50% of the genome sequence was

contained on contigs greater or equal to 78.3 kb – and 885 scaffolds with an average length of 411.9 kb and a N50 length of 32.1 Mb (Protasio *et al*., 2012).

Assembly of the Z and W chromosomes proved to be difficult as there are large regions of homology where the chromosomes are indistinguishable the Z and W chromosomes assembled together in 34 scaffolds with 59 Mb in total (Protasio *et al.*, 2012). The same authors used differences in coverage of male and female DNA to identify female-specific regions of the W chromosome. Overall, 30% of the Z/W chromosome sequence was unique to the Z chromosome, and 23 Z-specific genetic markers were identified in the assembled Z/W scaffolds (Protasio *et al.*, 2012). 114 unplaced scaffolds, totalling approximately 1.1 Mb, were found to be W-specific but contained approximately 90% repetitive sequence including known female-specific repeats (Portela *et al.*, 2010).

1.6.2 Transcriptomics of sex-specific schistosome biology

Due to its relevance for pathology and transmission, understanding the sexual dimorphism of schistosomes is critically important. This aspect of schistosome biology was and continues to be studied using different methods of transcriptome analysis (for example (Fitzpatrick *et al.*, 2005; Lu *et al.*, 2017)). In recent decades, several new methods of transcriptome analysis have become available, including methods based on quantitative real time PCR (qRT-PCR), expressed sequence tags (EST), serial analysis of gene expression (SAGE), cDNA microarrays and later RNA–Seq as well as variations of these methods (Lowe *et al*., 2017). In the following paragraphs, I will summarise schistosome

transcriptomics, the development of methods and their use to address increasingly detailed biological questions.

Early papers to examine the differential expression of genes in male and female *S. mansoni* examined individual genes such as actin (Davis *et al.*, 1985) or small groups of genes, due to a lack of high-throughput methods. In this case, Davis *et al.* (1985) used northern blot analysis to compare mRNA levels. The authors showed that two actin transcripts were more abundant in male worms than in females, cercariae and eggs. Using the same methodology the transcription of the gene coding for the egg shell protein p48 was studied (Chen *et al.*, 1992). The authors found it to be exclusively expressed in paired female worms. Furthermore Grevelding *et al.* (1997) identified three more genes, encoding egg shell protein p14, ferritin-1, and mucin-like protein A11, which they showed to only be expressed the vitelline tissue of paired female worms.

Thanks to the progress in human transcriptomics (Adams *et al.*, 1991, 1992, 1993; Hoog, 1991), increasing availability of genomic data (Franco *et al.*, 1995), programs and resources such as the WHO/UNDP/World Bank Schistosoma Genome Network (http://www.nhm.ac.uk/hosted_sites/schisto), the *Schistosoma mansoni* EST Genome Project (http://verjo18.iq.usp.br/schisto/) and the Institute for Genomic Research *Schistosoma mansoni* genome project (http://www.tigr.org/tdb/e2k1/sma1/index.shtml), it became feasible to design *S. mansoni* specific microarrays. This allowed scientists to examine hundreds of genes at the same time (Lowe *et al.*, 2017). Hoffmann *et al.* (2002) used a cDNA microarray with 576 probes and were able to double the number of genes

known to be differentially expressed between male and female schistosomes, identifying sex-specific expression for twelve female- and four new male-specific transcripts.

One year later, 16000 ESTs and 163 full-length cDNA had been sequenced, when Verjovski-Almeida *et al.* (2003) generated ten times more ESTs and provided the first sequencing-based comprehensive overview of a schistosome transcriptome. However most transcripts were still not sequenced along their full length, instead rather small, unique nucleotide sequences (tags) were sequenced, which allowed the corresponding genes to be identified (Lowe *et al.*, 2017)l.

Many of the subsequent publications that examined differences between male and female worms as well as paired and single worms used microarray or tag sequencing methods: Between 2004 and 2014 several such papers were published (Fitzpatrick *et al.*, 2004; Fitzpatrick *et al.*, 2005; Fitzpatrick & Hoffmann, 2006; Leutner *et al.*, 2013; Moertel *et al.*, 2006; Sun *et al.*, 2014) leading to a further expansion of genes known to be gender- or pairing-associated. When examining differences between male and female worms from the genus *Schistosoma*, the number of gender-associated genes discovered by such undertakings thus increased from 16 (four male- and twelve female-specific genes (Hoffmann *et al.*, 2002) to 30 (eight male- and 22 female- associated genes) in *S. japonicum* (Fitzpatrick *et al.*, 2004), then later 227 (86 male- and 141 female-specific genes) in *S. mansoni* (Fitzpatrick *et al.*, 2005) and finally 2179 (1163 male- and 1016 female-specific genes in *S. japonicum* (Moertel *et al.*, 2006) genes, an 136-fold increase of genes known to be differentially expressed

between male and female worms within four years.

Due to these investigations, we now know the transcripts most differentially expressed between sexually mature male and female worms. These differences closely reflect the biology of this parasite, with sexually mature females up-regulating transcription of genes associated with egg production, such as the major egg antigens other eggshell proteins, p48 (chorion), as well as tyrosinases (Fitzpatrick & Hoffmann, 2006). Male-specific genes on the other hand are dominated by tegument and muscle genes (e.g. tetraspanin genes, tegumental antigen as well as actin, dynein light chain 3, myosin and tropomyosin) (Fitzpatrick *et al.*, 2005, 2004; Moertel *et al.*, 2006). These may reflect the role of the male worm in anchoring the female inside blood vessels, moving her to suitable egg deposition sites (Fitzpatrick *et al.*, 2005).

The role of pairing on gene expression and the resulting sexual maturation has also generated considerable interest (Fitzpatrick & Hoffmann, 2006; Leutner *et al.*, 2013; Sun *et al.*, 2014). A number of different comparisons have been made: first Fitzpatrick & Hoffmann (2006) used a cDNA microarray to examine differential gene expression in seven week old adult males and sexually mature female from mixed sex (MS) and females from single sex (SS) infections. This allowed the author to identify 245 genes differentially expressed between females from MS infections (64 genes) and SS infections (181 genes), and 138 138 genes differentially expressed between males from equivalent infections (104 genes up-regulated in males from SS and 34 in males from MS infections). In 2013, Leutner *et al.* used a combination of oligonucleotide microarray and

SuperSAGE (a form of Serial Analysis of Gene Expression using high-throughput sequencing (Lowe *et al.*, 2017)) to examine males from MS and SS infections (using the terms pairing-experienced and inexperienced). Their microarray analysis allowed them to discover a total of 526 genes, which were differentially expressed in MS (229 genes) and SS males (297 genes) whereas their SuperSAGE analysis only detected 288 genes (253 genes in MS males and 35 in SS males). The discrepancy was explained partially by the fact that expression for over 4000 (out of a total 6326 examined) genes was only detected by one of the two methods as well as the use of stringent significance thresholds (Leutner *et al.*, 2013). The intersection of those two data sets contains a total of 29 genes, 21 of which had functional annotation at the time, from which 16 genes were up-regulated in males from SS, and 5 in males from MS infections (Leutner *et al.*, 2013). Sun *et al.* (2014) examined *S. japonicum* at 18 and 23 d.p.i. to get an insight into the early changes induced by pairing, as well as the changes taking place in females from SS and MS infections between 18 and 23 d.p.i. At 18 d.p.i. they found a total of 318 differentially expressed genes (241 in females from SS and 77 in females from MS infections). However, at 23 d.p.i. the number of differentially expressed genes had risen to 3446 genes (533 in females from MS and 2913 in females from SS infections) (Sun *et al.*, 2014). The numbers of genes differentially expressed between females from 18 and 23 d.p.i. are also large: for females from SS infections, 2545 genes were found to be differentially expressed (2193 up-regulated in females from 18 d.p.i. and 352 up-regulated in females from 23 d.p.i.) and for females from MS infections, 1998 genes were found to be differentially expressed (740 up-regulated in females at 18 d.p.i. and 1258 up-regulated at 23 d.p.i.) (Sun *et al.*, 2014). Due to an apparent lack of replicates in

this study, it is hard to estimate how much biological variability there is, but a general trend can be seen clearly: The transcriptome of paired female worms (MS infections) becomes increasingly divergent from females that remain unpaired (SS infections) (Sun *et al*., 2014). As paired females mature sexually the differences between them and females from single sex females become more noticeable both morphologically (Fitzpatrick & Hoffmann, 2006), but also at the transcriptome level and these differences increase as pairing continues until the paired female reaches complete sexual maturity around 5 weeks p.i. (Basch, 1991).

Recently, Wang *et al.* (2017) published a large RNA-Seq study in which they presented evidence that biogenic amines serve as the male stimulus for female sexual maturation. Having sequenced 48 *S. japonicum* transcriptomes from eight time points in the intra-mammalian stage (between 14-28 d.p.i.) and both sexes, with three biological replicates each, these authors found the expression of an aromatic-L-amino acid decarboxylase (AADC) to increase leading up to pairing, and showed the expression of AADC to be located in the gynecophoral canal by using whole mount *in situ* hybridisation (Wang *et al.*, 2017). Furthermore the authors found an allatostatin-A receptor-like gene (AlstR), coding for an insect-like hormone receptor, to be involved in the maturation of female reproductive organs (Wang *et al.*, 2017). Indeed, when AlstR was knocked down using RNAi, ovaries and vitelline glands failed to mature, in contrast to control females (Wang *et al*., 2017). Another study found evidence that histone modifications (H3K27me3) are different between male and female *S. mansoni* (Picard *et al.*, 2016). The authors found these epigenetic modifications to correlate with gene

expression showed there to be a sex-biased dynamic of histone methylation, in particular a depletion of H3K27me3 along the transcription unit in male worms (Picard *et al.*, 2016). However, as all RNA-Seq data in this study was derived from worms of SS infections, it remains to be seen if these results hold true for sexually mature females as well.

Publication	Method	Species	Number of probes/read	Comparison	Results
Hoffmann *et al.*, 2002	cDNA microarray	*S. mansoni*	576 ESTs	Adult male vs female	23 DEGs*
Verjovski-Almeida *et al.*, 2003	cDNA EST sequencing	*S. mansoni*	163000 ESTs sequenced	N/A	N/A
Fitzpatrick *et al.*, 2004	cDNA microarray	*S. japonicum*	457 probes	Adult male vs female	30 DEGs
Fitzpatrick *et al.*, 2005	Oligonucleotide microarray	*S. mansoni*	7638 probes	Adult male vs female	197 DEGs
Moertel *et al.*, 2006	Oligonucleotide microarray	*S. japonicum*	19221 probes	Adult male vs female	647/700 DEGs**
Fitzpatrick & Hoffmann, 2006	DNA microarray	*S. mansoni*	7638 probes	Single vs paired adult female	248 DEGs
				Single vs paired adult male	138 DEGs
Leutner *et al.*, 2013	Oligonucleotide microarray & SuperSAGE	*S. mansoni*	19197 probes; Number of sequenced tags not stated	Single vs paired adult male	29 DEGs
Sun *et al.*, 2014* **	Tag-Seq	*S. japonicum*	> 3 million tags sequenced	Paired vs single female, 21 d.p.i.	3446 DEGs
Picard *et al.*, 2016	RNA-Seq	*S. mansoni*	65.8 million reads per sample	Male vs female: Cercariae, Schistosomule, Adult worms	7168 DEGs
Wang *et al.*, 2017	RNA-Seq	*S. japonicum*	Number of sequenced reads not stated	Time course - Male	1934 DEGs
				Time course - Female	6535 DEGs

Table 1.1: Summary of the published literature on schistosome transcriptomic discussed in this chapter. The table provides the name of the first author and year of publication, the methods of transcriptomics used, the Schistosoma species examined in a given publication, the number of probes used (in the case of microarrays) or DNA reads sequenced (in the case of RNA-Seq), the samples which were compared and finally the number of differentially expressed genes (DEGs) observed. *) Including positive controls; **) Results for Philippine and Chinese Strains respectively; ***) Not all comparisons summarised here

1.6.3 Methods of transcriptomics

This thesis relies heavily on gene expression data, including RNA sequencing (RNA-Seq) first and foremost, but also used quantitative reverse transcription polymerase chain reaction (qRT-PCR) data, which is often considered the gold standard for comparing gene expression, for validation of RNA-Seq data. Microarray data is only used to briefly, by correlating previously published microarray data to RNA-Seq data. The following section introduces these methods for measuring changes in the transcriptome as well as their strengths and weaknesses.

Quantitative reverse transcription PCR

qRT-PCR is used frequently to compare expression of small numbers of genes between different samples. After reverse transcription of mRNA, the cDNA is used as template for a PCR reaction where, over the course of several rounds of amplification, the quantity of target cDNA is measured. This can for example be done using SYBR Green, a compound that fluoresces when bound to DNA. Over successive rounds of PCR this results in stronger fluorescence as more DNA has been produced. The quantities of DNA in the sample and control can then be

compared to determine the relative abundances of the target sequence in the cDNA. The method most commonly used for inferring differences in expression between samples is called ΔΔCt Calculation Method, and it makes use of one or more internal reference (IR) genes thought to be expressed at identical levels across all samples, to control for differences in the quantity of input cDNA (Kubista *et al.*, 2006). The IR is used for normalisation of the amount of input cDNA between different samples, so that any differences between target gene expression observed after normalisation represents the treatment effect (see Chapter 2.2.5). qRT-PCR has been widely used to measure gene expression in schistosomes. For example it has been used to identify or validate gender-specific gene expression by comparing male and female samples (Hoffmann *et al.*, 2002), pairing-induced gene expression by comparing cDNA samples from paired to those of virginal females (Fitzpatrick & Hoffmann, 2006). Another example is the use of qRT-PCR to compare gene expression in different organs (Hahnel *et al.*, 2014) and in RNAi knock-down experiments (Tran *et al.*, 2010).

Microarray

cDNA microarrays are a collection of DNA probes which are immobilised on a small chip to allow sample RNA (after reverse transcription to cDNA) to be bound by hybridisation with the complementary probe sequences. When applying sample cDNA to the microarray, the relative abundance of sample cDNA molecules is usually measured using fluorophore- or chemiluminescence-labelled probes (M. B. Miller & Tang, 2009). It is worth noting that reference sequences have to be known prior to the production of the microarray probes (Lowe *et al.*, 2017). This allows comparisons of transcript abundances of two or

more experimental conditions. As in the case of qRT-PCR, stringent normalisation is necessary to obtain accurate results. This includes measuring technical as well as biological variation across the transcriptome. Microarrays can be performed at very high throughput, due to the relatively low labour intensity (compared to RNA-Seq), but they require a relatively large amount of mRNA input (about 1 μg of mRNA) (Lowe *et al.*, 2017). This can limit the application of microarray, if the biological samples are difficult to obtain and very small in size.

RNA sequencing

RNA-Seq relies on high-throughput sequencing technology that can provide the sequences of vast numbers of short cDNA fragments to determine the abundance of cDNA molecules in each sample. Depending on the precise method used to make the cDNA library to be sequenced, the data may include information on the expression of not just messenger RNA (mRNA) but also non-coding RNAs such as ribosomal RNA (rRNA), transfer RNA (tRNA) and others. Careful normalisation is required for accurate results. The normalisation needs to correct for within-sample biases, caused by factors such as transcript length and GC-content, as well as between-sample biases, especially amount of sequenced cDNA. To correct within-sample biases RPKM (reads per kilobase of transcript per million mapped reads) or FPKM (fragments per kilobase of transcript per million mapped reads) are often chosen. This method scales the number of reads for a given gene to the length the given transcript as well as the overall number of reads across the whole transcriptome (Verk *et al.*, 2013). However, it has been demonstrated that this method skews calls for differential expression in favour of longer transcripts

as the number of reads mapping to a gene will be proportional to its length and higher read counts give more statistical power to determine significant differences (Oshlack & Wakefield, 2009). In cases where expression of a gene is compared across different samples (rather than different genes within one sample), only between-sample normalisation is necessary (Verk *et al.*, 2013). In such cases sequencing yield is the most important variable to take into account. The sequencing yield, i.e. the total number of transcript reads generated during sequencing, has to be considered during the normalisation as it often differs between replicates and treatment groups. Rather than dividing the read count per gene by the total number of reads for a given sample, more robust methods are used that can take extreme expression values into account (Love *et al.*, 2014). Such methods include the median-of-ratios (Love *et al.*, 2014) and the trimmed mean of M-values method (Robinson *et al.*, 2009). However, in this context it is also important to consider the effect of relative abundances of mRNAs in a transcriptome. If the real expression of a target gene is identical in two samples, but its relative abundance is higher in one sample (for example due to a large number of other genes becoming down-regulated) the target gene will appear to be differentially expressed. Technical variation in transcriptome analysis is the measured differences introduced by the method or transcriptome analysis and sample handling, rather than differences in the underlying biology. Technical variation is thought to be smaller in RNA-Seq than in microarray technology (Marioni *et al.*, 2008), biological variation has to be measured carefully to allow statistical inference of differential expression of a gene between two samples (Love *et al.*, 2014). Furthermore, during the final steps of RNA-Seq library production, the cDNA is usually amplified using PCR. During this process some

biases may be introduced as some GC-rich cDNA molecules don't become amplified at the same rate as other molecules (Aird *et al.*, 2011). However, this effect should be negligible when comparing expression of the same gene across different samples, as these samples should experience the same amplification biases at that locus.

RNA-Seq offers several advantages over microarrays. Firstly, microarrays rely on a library of probes that is pre-determined and based on existing genome annotations (Lowe *et al.*, 2017). Thus, genes that have not yet been discovered will be excluded from the analysis. Unlike microarrays, RNA-Seq data can be used to discover new transcripts and genes. This information can then be used to validate predicted gene models in the genome and contribute to improving the quality of genome annotation (Lowe *et al.*, 2017). Secondly, RNA-Seq allows to detect splice variants, which can remain undiscovered in microarrays as, in many cases, different variants will hybridise with the same probe. Another advantage of RNA-Seq over microarray data is a larger dynamic range, which means accurate quantification of genes with very low or high expression (Zhao *et al.*, 2014). In the absence of expression microarrays will often measure non-zero expression levels due to background noise such as non-specific hybridisation. Conversely at the high expression levels micro arrays can become saturated and unable resolve different expression levels anymore (Zhao *et al.*, 2014). This allows for a more quantitative measurement of low and very high abundance cDNA using RNA-Seq (Wang *et al.*, 2009).

Comparison of RNA-Seq and proteomics

A discrepancy between mRNA and protein levels is to be expected in any organism as there are many mechanisms regulating the abundance of proteins in a cell (Maier *et al.*, 2009). These include different efficiencies of mRNA cap binding proteins, the rate at which ribosomes are recruited to internal ribosome binding sites, codon usage, regulation by microRNAs, binding to RNA-binding proteins and mRNA stability which is mainly controlled by the mRNA poly-A tail (Vogel & Marcotte, 2012). Protein abundance is also regulated further down-stream of translation by degradation signals leading to ubiquitination and proteolysis (Vogel & Marcotte, 2012). All these factors combined mean that there is not a perfect correlation between mRNA and protein levels (Maier *et al.*, 2009). Schwanhaeusser *et al.* (2011) examined the correlation of steady-state mRNA and protein abundance in mouse fibroblasts, finding that, across 5028 genes, mRNA levels only explain between one and two thirds of the variation in protein levels observed (R^2=0.41). In human medulloblastoma cells, Vogel *et al.* (2010) reported that 66% of variation in protein levels could be explained by factors including mRNA concentration, UTR sequences and the coding sequence. However, these authors showed that mRNA concentration only explained 27% of the variance, 3' UTRs 8% and that 31% was explained by the coding sequence. Nonetheless, Vogel & Marcotte (2012) conclude that mRNA is a suitable proxy for the presence of a protein as mRNA levels detected by RNA-Seq directly correlated to the corresponding proteins in a proteomics experiment. However, work by Li *et al.* (2014) showed that due to problems with the normalisation of proteomics data in the aforementioned studies, mRNA concentrations predict around 84% of protein abundances (instead of 27%). Furthermore, the mRNA

and protein abundance were shown to correlate more strongly after when cells responded to a stimulus (as opposed to a being in a steady state) (Koussounadis *et al.* 2015). In such situations the mRNA abundance of differentially expressed genes is even better correlated to their protein products (Koussounadis *et al.* 2015).

1.7 Aims of my project

The aim of this thesis is to explore and characterise the differences between male and female worms during their intra-mammalian development at the transcriptome level. Using the vast quantities of expression data that can be generated using RNA-Seq and the unprecedented detail it affords, what factors involved in the process of female maturation can be identified from the transcriptome? How does this change our understanding of the role that the male-female interaction plays in the transmission and pathology of the disease? In particular, this thesis addresses the differential expression of genes between male and female schistosomes during their intra-mammalian development including the changes in gene expression induced by the male-female interaction. Three main topics are examined in this thesis.

The first is the development of male and female worms inside the mammalian host. Particularly, when and how do male and female worms start to diverge at the transcriptome level? And similarly, what are the early differences of female worms in MS infections compared to SS infections at the transcriptome level?

Furthermore, what roles do some of the genes identified play in the female fertility? These results are presented in Chapter 3.

Next, I wanted to discover the effect of cell proliferation and apoptosis on the maintenance of sexual maturity as well as regression of female reproductive tissues. Furthermore, how does pairing regulate these processes? Using an *in vitro* culture system of female worms in the presence or absence of male worms, the changes in expression of fertility-related genes were examined as well as genes that control apoptosis and cell proliferation. Results from this work comprise Chapter 4.

In Chapter 5, the transcriptomes of *S. mansoni* testes and ovaries were examined. This work was performed to learn how gene expression changed in the gonads in response to pairing. As pairing induces proliferation of the ovaries, it was of particular interest to us which transcriptome changes allowed for this proliferation to take place. These changes are obscured in the whole worms' transcriptome by the larger vitellaria but can be better resolved by examining the gonad transcriptomes.

Finally, Chapter 6 provides a summary of the findings in this thesis and places them in the context of the current understanding of the biology of schistosomes.

Chapter 2

Materials & Methods

2.1 Parasite and snail maintenance

2.1.1 Parasites and snail origin

The species of schistosome used throughout this thesis was *S. mansoni*. For Chapter 3 & 4 a parasite strain originally recovered from Puerto Rico were used. The parasites used in Chapter 5 originate from Liberia. *Biomphalaria glabrata* were used to maintain the *S. mansoni* life cycle throughout.

2.1.2 Miracidia infection of snails

Livers from *S. mansoni* infected (6-7 weeks p.i.) mice were soaked in sterile 1.2% NaCl solution with Penicillin/Streptomycin (100 U/ml; 0.05 mg/ml) at room temperature (RT) for 10 min. Livers were then transferred to a plastic beaker, and about 20 ml of sterile 1.2% NaCl solution added. A blender (Bosch MSM 6B150) was used to homogenise the livers for 3 min at the lowest speed and the resulting liquid then filtered through two sieves with a pore size of 180 µm, to remove larger liver particles, and then 45 µm to collect eggs. Eggs were collected in the lower sieve and further washed with 1.2% NaCl solution. The eggs were then transferred to a volumetric flask with a narrow neck and the flask filled with Lepple water (see below). After incubating for 1 hour at 28°C, all but the top 2 cm were covered in aluminium foil and the flask incubated for another 1 hour period. Next, miracidia were collected from the top layer of the solution. Five to six miracidia were used for mixed sex snail infections. For infections small snails (0.5 cm in diameter) were placed individually in the well of a 24-well plate together with 1 ml of Lepple water. The miracidia were then carefully added to each well and left to infect the snail for 2 hours. Next the snails were placed back

into an aquarium for 5 weeks when snails were ready to be used to obtain cercariae.

Lepple water (1x)

0.378 mM calcium chloride dihydrate
0.500 mM magnesium sulphate heptahydrate
0.025 mM potassium sulphate
0.500 mM sodium carbonate anhydrous
0.0002 mM ferric chloride

Lepple water was usually prepared as a 10x concentrated solution and diluted with water prior to use.

To obtain mice infected with male or female schistosomes only, 50 snails were infected with single miracidia. In the four weeks following infection, the snails were exposed to light for two hours to shed clonal, single sex, cercariae (Tucker et al., 2013). Approximately 500 cercariae from each snail, as well as single male and a single female adult worm (positive controls) were processed for DNA extraction with a DNA Mini kit (Qiagen, 51304) PCR was used to identify sex-specific W2 regions on the (female-specific) W-chromosome using the rhodopsin gene as positive control gene (Lepesant *et al.*, 2012). The following primers were used:

<u>W2 primers:</u>

FWD: CTGTTTCGAATTTCACACTTCA (T_m = 61.7°C)

REV: CATTCACAGTTTGGCGAACA (T_m = 64.7°C)

Rho primers:

FWD: GACGGCCACACTAAAG (T_m = 54.7°C)

REV: AGTAAAATGGTCACTGCTAT (T_m = 52.8°C)

PCR reactions were assembled using 10 µl of HiFi Ready Mix (Kapa, KK2602), 2.5 µl of genomic DNA solution, 1.0 µl of the primer mix (10 µM) and 6.5 µl of ddH_20 for a total of 20 µl. After denaturing the DNA for 5 min at 95°C the samples then underwent 35 cycles of amplification (95°C, 5 sec; 55°C, 5 sec; 68°C, 10 sec). Finally, the samples were kept ay 72°C for 1 min.

The PCR products were run on an 1% agarose gel in TBE buffer at 80 V and 100 mA for 50 min. Female samples showed amplification in both the Rhodopsin- as well as W2-specific PCR reaction, whereas male samples only had an amplicon in the Rhodopsin-specific reaction.

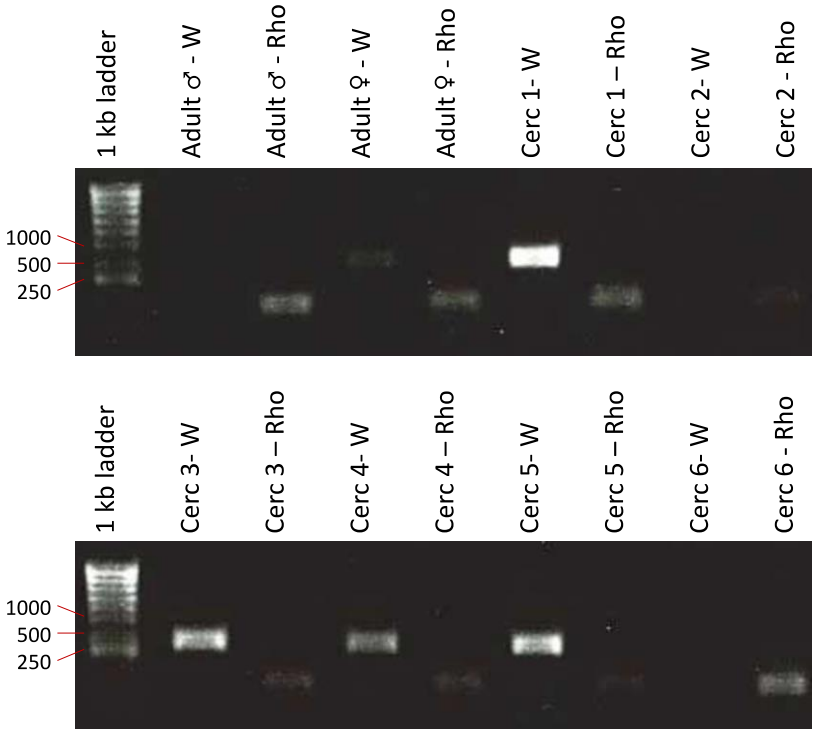


Figure 2.1: The sex of cercariae from single-miracidium infections was determined by PCR. Male *S. mansoni* samples do not have an amplicon in the W chromosome specific region, whereas female *S. mansoni* samples do have an amplicon for the W specific reaction. The Rho reactions serve as a positive control and should have an amplicon regardless of sex. The cercariae from four snails (Cerc 1, 3, 4 and 5) were found to be female, the other (Cerc 2 and 6) male.

2.1.3 Detecting eggs in mouse livers

To determine whether single sex infections with only male or only female worms had been achieved, the livers of sacrificed mice were blended as described above. Livers were transferred to a plastic beaker and homogenised using a blender (Bosch MSM 6B150) for 3 min at the lowest speed. The resulting liquid was then filtered through two sieves with a pore size of 180 µm, to remove larger liver particles, and then 45 µm to collect eggs. If any eggs were found, the corresponding mouse was classed as having had a mixed sex infection, otherwise (if any worms had been recovered from the mouse by perfusion) the mouse was classed as having had a single sex infection.

2.1.4 Collection of cercariae

Infected snails were kept in a dark cabinet until cercariae were required. Then they were moved into small glass beakers with enough Lepple water to cover all snails and left under a bright light for one hour allowing the cercariae to emerge from the snails. After one hour, the water was carefully poured into 50 ml falcon tubes and the snails transferred back into their tanks in the dark cabinet. Cercariae numbers were estimated by taking the average count from five 10 µl aliquots of the cercariae. The cercariae were killed with Lugol's iodine solution (Sigma, L6146) and then counted under a dissection microscope.

2.1.5 Infections and perfusions

All animal experiments were conducted under Home Office Project Licence No. 80/2596. All protocols were presented and approved by the Animal Welfare and Ethical Review Body (AWERB) of the Wellcome Trust Sanger Institute. The AWERB is constituted as required by the UK Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012.

BALB/c mice were infected with about 250 mixed sex or single sex cercariae via intraperitoneal injection. After 18 to 49 d.p.i., the mice were euthanised by injection with an overdose of sodium pentothal and perfused to recover adult worms residing in the blood system. The worms used in Chapter 5 were grown in Syrian hamsters (*Mesocricetus auratus*); all animal work associated with

Chapter 5 was performed by the group of Prof. Dr. Christoph Grevelding (Justus-Liebig-University, BFS, Institute for Parasitology, Gießen*, Germany)*.

2.1.6 *In vitro* culture

In Chapter 3 (RNAi experiments), worms were maintained *in vitro*. Two experiments were set up, one with a total duration of seven days, and one with a 48 h incubation. For this latter experiment parasites were kept in supplemented DMEM media (see below) in a dark incubator at 37°C with 5% CO_2. In the case of the seven day incubation, the media was replaced every two days (see RNAi methods (2.2.2) for details). Furthermore, in Chapter 4, the effect of culture in DMEM media and Basch media on worm fertility was tested. The following recipes were used to prepare these media.

Supplemented DMEM

0.02 M Hepes buffer (Sigma, 83264)

2 mM L-glutamine (Sigma, G7513)

10% Foetal calf serum* (Sigma, F3297)

1x Antibiotic-Anitimycotic (ThermoFisher, 15240062)

DMEM media** (ThermoFisher, 10270106)

*) Foetal calf serum was heat-deactivated before use.

**) High glucose DMEM media was used.

Basch media

In Chapter 4 worms were kept *in vitro* for seven days and Basch media (Basch, 1981). This media was used to was used to maintain female maturity as much as possible, as the female reproductive system usually regressed in culture. Worms were again maintained in an incubator at 37°C with 5% CO_2 and half the media was replaced every two days. Modified Basch media was prepared as described by Bhardwaj *et al.* (2011).

Directly prior to use, 1% (by volume) horse red blood cells (RBCs) were added to the media. For this whole blood (ThermoFisher, SR0050) was first centrifuged at 4°C and 500 g for 10 min, then serum and the upper layer of blood were removed. The remaining RBCs were then suspended in BME (ThermoFisher, 41010-109) to produce a 50% stock solution of RBCs.

2.1.7 Isolation of *S. mansoni* gonads

The isolation of gonads was performed by the Dr. Zhigang Lu from the group of Prof. Dr. Christoph Grevelding (Justus-Liebig-University, BFS, Institute for Parasitology, Gießen, Germany*)* on worms from both MS and SS infections. It was performed as previously described by Hahnel *et al.* (2013), first using detergents (Brij35, Nonidet P40-Substitute, Tween80 and TritonX-405) to dissolve the tegument and then Type IV elastase to digest muscle tissue.

2.2 Molecular biology techniques

2.2.1 Cloning

If a transcripts was 1300 bp or shorter, primers were designed to clone the whole transcripts. Otherwise, a fragment of the transcript around 1000 bp in length was chosen and primers designed manually to clone that fragment. The primer pairs were designed to have similar melting temperatures to facilitate PCR reactions. Total RNA was isolated from adult worms using the Trizol reagents (Invitrogen, 15596026) (see 2.2.3). cDNA was synthesised using the SuperScript® III First-Strand Synthesis System (ThermoFisher, 18080051) using the manufacturer's protocol, with slight alterations. Firstly, reagents were incubated for 2 min at 25°C before 1 µl of SuperScript® III was added. Secondly, samples were incubated at 25°C, 10min; 50°C, 60 min; 70°C, 15min. cDNAs were amplified using the ReadyMix Taq (Sigma, P4600) following manufacturer's instructions. The following primers were used:

<u>CD63 receptor (Smp_155310) primers:</u>

FWD: ATGTGTACTGTCGTATTGAGATTAAC (T_M = 59.0°C)

REV: TTAAGGATATCCAGTGGTTAGTATC (T_M = 58.2°C)

<u>CD63 antigen (Smp_173150) primers:</u>

FWD: ATGGCCTCTTTAAGCTGTGG (T_M = 63.1°C)

REV: TTAGTCAGAATCGCCAGATTTG (T_M = 63.0°C)

<u>TSP-2 (Smp_181530) primers:</u>

FWD: TTCTTCTTCGTTTCAGGGATGT (T_M = 63.7°C)

REV: TCACCGCGCTTTATAGCCAA (T_M = 67.5°C)

Boule (Smp_144860) primers:

FWD: GCCTTGGGGCATCTAAAGGT (T_M = 66.1°C)

REV: TGCTGCAGTGGATGCTGTTA (T_M = 65.7°C)

MXRA8b (*Danio rerio*; negative control) primers:

FWD: TCTTTCATGCAGGCCACAGT (T_M = 65.8°C)

REV: CCGGAACCAACCCGATTACA (T_M = 68.8°C)

Target genes were amplified from the cDNA using 35 cycles of PCR (95°C, 5 sec; variable temperature, 5 sec; 72°C, 30 sec) and finally 1 min at 72°C. The annealing temperature was generally set to be 5°C lower than the higher melting temperature of the two primers (see above).

PCR products were then run on a 1% agarose gel in TBE buffer for 40 min at 100 V and 120 mA to confirm product size. The PCR reaction was then cleaned up using the QIAquick PCR Purification Kit (Qiagen, 28104) following manufacturer's instructions.

The required volumes of insert and vector (3000 bp; 50 ng/µl) necessary for ligation at a 1:1 and 3:1 ratio were determined using Promega's Bio calculator. http://www.promega.com/a/apps/biomath/index.html?calc=ratio

The pGEM-T easy vector system (Promega, A1360) was used following manufacturer's instructions.

The reactions were then incubated for 1.5 h at RT (rather than 1 h according to the protocol). Luria-Bertani (LB) Ampicillin (0.1 mg/mL) agar plates were prepared by using 0.1M Isopropyl β-D-1-thiogalactopyranoside (IPTG) (ThermoFisher, R1171) and 20 mg/µl 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) (ThermoFisher, R0941) per plate which was spread with a sterile plate spreader. 50 µl of competent cells (Invitrogen, C4-04003) were used for cloning, following the manufacturer's instructions.

Next, 950 µl Super Optimal broth with Catabolite repression (SOC) medium (ThermoFisher, 15544034) were added to each tube of transformed bacteria; the samples were then placed in large 50 ml Falcon tubes and incubated at 37°C (shaking at 200 RPM) for 75 min. Transformed bacteria were removed from the incubator and centrifuged for 2 min at 1000 x g. 900 µl of clear supernatant was removed and the remainder used to resuspend the cells. Approximately 50 µl of suspended cells were plated onto the prepared LB agar plates and incubated at 37°C overnight.

2.2.2 RNA interference

This method relies on RNA molecules that are complementary to the target mRNA sequences. RNAi exploits a cellular mechanism to defend against RNA viruses to target specific mRNAs for degradation by the RNA-induced silencing complex, causing expression of the target gene to be reduced also known as "knocked down".

The genes of interest, CD63a (Smp_173150) and CD63R (Smp_155310), as well as a positive control, TSP-2 (Smp_181530), were cloned (see 2.2.1) into the pGEM-T easy vector (Promega, A1360). The zebrafish (*Danio rerio*) gene encoding the matrix-remodelling associated protein 8b, with no homology to *S. mansoni* proteins, was chosen as a negative control.

To create a linear template with T7 polymerase promoters on either end, primers were designed that had both a gene specific sequence (same as the original cloning primers) as well as the T7 promoter sequence.

CD63 receptor primers (T_m = 54°C):

FWD: TAATACGACTCACTATAGGGATGTGTACTGTCGTATTGAGATTAAC

REV: TAATACGACTCACTATAGGGGAACAACAAATTTGGCATC

CD63 antigen primers (T_m = 58°C):

FDW: TAATACGACTCACTATAGGGATGGCCTCTTTAAGCTGTGG

REV: TAATACGACTCACTATAGGGCAGGGATTGTTTGTCCACTC

TSP-2 primers (Tran *et al.,* 2010) (T_m = 59°C):

FWD: TAATACGACTCACTATAGGGTGATTGTGGTTGGTGCACTT

REV: TAATACGACTCACTATAGGGGACCAATGCGAACAGAAACA

Negative control (MXRA8b) primers (T_m = 60°C):

FWD: TAATACGACTCACTATAGGGTCTTTCATGCAGGCCACAGT

REV: TAATACGACTCACTATAGGGCCGGAACCAACCCGATTACA

All primers contain a T7 binding site (underlined). The PCR was performed following the instructions of the "MEGAscript T7 dsRNA" kit (ThermoFisher,

AM1626), first using the annealing temperatures (see above) of the gene-specific region of the primers for 5 cycles and then increasing the annealing temperature by 5°C for a further 30 cycles. The template was purified using a PCR clean up kit (Qiagen, 28104) following the manufacturer's instructions and was then used as input for the "MEGAscript T7 dsRNA" kit. Incubation of the *in vitro* transcription reaction was performed for 4 hours, and then the dsRNA was isolated from the reactions using the manufacturer's instructions. The kit used a column based method were the dsRNA is first precipitated with salts and ethanol, then bound to a column, washed several times and finally eluted. Length and concentration of the dsRNA was confirmed on the NanoDrop 1000 (ThermoFisher) and the Agilent tapestation.

dsRNA soaking

Soaking was performed as described by Tran *et al.* (2010) with slight modifications. Briefly, worms were perfused from mice at 49 d.p.i. and paired couples cultured in Basch media while being soaked for seven days in 10 μg/ml dsRNA. The following four dsRNA treatment groups were used: CD63R, CD63a, TSP-2 (positive control) and a negative control RNA (see above). Media was replaced every two days and new dsRNA added. After seven days worms were transferred into Trizol (Invitrogen, 15596026) for RNA extractions and the laid eggs counted in the plates.

dsRNA electroporation

For electroporation, worms were washed carefully in DMEM (ThermoFisher, 10270106) following perfusions, placed in a microcentrifuge tube in 100 μl

DMEM. Then, 20 µl of dsRNA solution (1 µg/ml) was added to the tube and incubated for 15 min at RT. The worms and the dsRNA solution were then transferred into 4 mm electroporation cuvettes (Bio-Rad).

20 milliseconds square wave pulse was delivered to the worms using the Bio-Rad Gene Pulser Xcell system at 125 V; worms were then transferred into pre-warmed Basch medium (37°C) and cultured *in vitro* for 48 h. After 48 h the worms were transferred into Trizol (Invitrogen, 15596026) for RNA extractions and their eggs counted on the plates.

2.2.3 RNA extractions

Samples were placed in 500 µl of Trizol (Invitrogen, 15596026) and stored at -80°C until further processing. Once ready, the samples were left to thaw on ice and then transferred to MagNA Lyser Green Beads tubes (Roche, 03358941001), containing small ceramic beads for mechanical homogenisation of tissue samples. The tubes were then placed in a FastPrep-24 instrument (MP Biomedicals) and homogenised three times for 20 sec at intensity 5, with 5 min rest on ice between runs. Following homogenisation, samples were incubated for 5 min at RT to allow dissociation of nucleoprotein complexes. Approximately 0.2 volumes of chloroform:isoamyl alcohol (24:1) (Sigma, C0549) was added and the samples shaken vigorously. After a further 3 min incubation at RT the samples were centrifuged at 14000 g for 15 min at 4°C.

Following centrifugation, the aqueous phase was transferred into a fresh RNase free micro centrifuge tube. Total RNA was isolated from the aqueous phase using

the RNA Clean & Concentrator kit (Cambridge Bioscience, R1019) following manufacturer's instructions. Samples for qPCR were treated with DNase in the isolation columns as described in the kit protocol. Total RNA was eluted in 22 µl of elution buffer and checked for degradation using a Nano Chip (Agilent, 5067-1512) on an Agilent 2100 Bioanalyzer. The RNA was deemed to be of good quality if one sharp ribosomal RNA peak could be seen. Samples were then stored at -80°C until further processing.

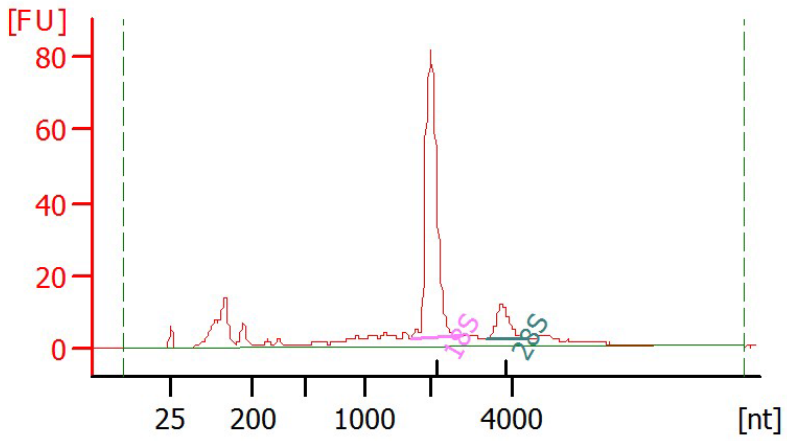


Figure 2.1: Example of good quality extracted total RNA. The sharp peak of RNA at 2000 nucleotides (nt) is the ribosomal RNA peak.

2.2.4 cDNA synthesis

cDNA was used for cloning as well as qPCR. 11 µl of DNase treated, full length total RNA was used as input. 1 µl dNTPs (10 mM for each nucleotide) (NEB, N0447S) was added as well as 1 µl oligo-dT_{18} primers (100 µM) (ThermoFisher, SO131). Samples were denatured for 5 min at 65°C and then cooled on ice. Then, 4 µl 5x First Strand buffer (ThermoFisher), 1 µl dithiothreitol (0.1 M; Sigma) and 1 µl SuperScript® III (ThermoFisher) were added. Samples were incubated at

25°C, 10 min; 50°C, 60 min; 70°C, 15 min and then stored at 4°C until further processing.

2.2.5 qPCR

Primer3 (http://bioinfo.ut.ee/primer3/) was used to design qPCR primers (see Table 2.1) for the target genes *cd63r*, *cd63a* and *tsp-2*. Settings were chosen to return primers providing approximately 100 bp long amplicons that did not overlap with the dsRNA probes used for RNAi (2.2.2). *psmd4* (Smp_090340) was used as reference gene in all qPCR reactions. For quantification of RNAi efficiency, two more genes were used as internal reference, *ndufv2* (Smp_069770) and *gapdh* (Smp_056970). Liu *et al.* (2012) found the house keeping genes *psdm4* and *ndufv2* to be particularly suitable as qPCR reference genes due to their stable expression.

Target	FWD	REV	Efficiency
cd63r **Smp_155310**	CCCCTGTTGTAAAATGGATGCC	CCTGGACATGTTGGATCGAGA	106%
cd63a **Smp_173150**	GTTGCCGTTTTAGTAGCGGC	CTCGGATAACACAGACAGCGA	95%
tsp-2 **Smp_181530**	CGAAATTGAACCCCCACTAC	CATGCTCCAAACATCCCTAAA	97%
psdm4 **Smp_000740**	CAAAGTTACAACGAGTTTCTGGCTCT	TGGCGTTTGCGATTTGCCAT	104%
ndufv2 **Smp_069770**	TGGAATGCCTTGGAGCTTGTG	TCCGCTTTGTGGACCAGGTT	98%
gapdh **Smp_056970**	TGTGAAAGAGATCCAGCAAAC	GATATTACCTGAGCTTTATCAATGG	101%

Table 2.1: **List of qPCR targets and their forward and reverse primers**. *tsp-2* primer sequences were first described by Tran *et al.* (2010). Primer efficiencies were calculated using a standard curve.

Primer efficiencies were determined using a dilution series of primer concentrations as described by the manufacturer (Applied Biosystems, 2010). All primer pairs were found to have an efficiency between 90-110%.

The qPCR was performed using KAPA SYBR FAST universal qPCR kit (KAPA Biosystems), following the manufacturer instructions. All cDNA samples were run in triplicates, as well as a triplicate negative control without cDNA template and a control with total RNA instead of cDNA to detect genomic DNA contamination. If genomic DNA has also been extracted during the RNA extraction, it could cause a false positive signal; the total RNA used in this control was not reverse transcribed and therefore should not allow for PCR amplification.

All reactions were 20 µl in volume and performed in sealed MicroAmp 48-Well skirted PCR plates (ThermoFisher, 4375816) and briefly centrifuged to remove bubbles. qPCR reactions were performed in an Applied Biosystems StepOnePlus machine using the thermocycling conditions recommended by the manufacturer (KAPA Biosystems, 2016) (denaturation: 95°C, 3 min; then 40 cycles of: 95°C, 1 sec; 60°C, 20 sec). The amplification data was analysed and plotted manually in Microsoft EXCEL (v14.2.3) using the ΔΔCt method (Livak & Schmittgen, 2001).

Initially the relative difference in expression between the internal reference and target gene was measured for both treated and control samples:

$$\Delta Ct_{treated} = Ct_{Internal\ reference,\ treated} - Ct_{target\ gene,\ treated}$$

$$\Delta Cq_{control} = Ct_{Internal\ reference,\ control} - Ct_{target\ gene,\ control}$$

For multiple reference genes the average Ct value was calculated:

$Ct_{reference\ gene} = (Ct_{GAPDH} + Ct_{PSMD4} + Ct_{NDUFV2}) / 3$

Finally, to find the relative difference in gene expression:

$\Delta\Delta Cq = mean(\Delta Ct_{treated}) - mean(\Delta Ct_{control})$

The fold-change was then calculated as follows:

Fold-change = 2 ^ ΔΔCt

2.2.6 Dot blot

To measure the efficiency of DIG labelling in the RNA probe, I used a dot blot assay, as described by Zimmerman *et al.* (2013). Briefly, serially diluted RNA probes were applied to a nylon membrane (Sigma, 11209272001) and cross-linked using UV light. The membrane was washed with maleic acid buffer, treated with blocking buffer and exposed to an anti-DIG antibody (Sigma, 11093274910) with a conjugated alkaline phosphatase enzyme (150 mU/ml) in blocking buffer. Samples were developed until the desired signal was reached.

2.2.7 Whole mount *in situ* hybridisation

Whole mount *in situ* hybridisations (WISH) were performed using a protocol provided by Dr. James Collins, (UTSouthwestern, Department of Pharmacology, Dallas, Texas). For the *in vitro* synthesis of DIG-labelled RNA probes, the insert of the pGEM-T easy plasmid (Promega, A1360) was amplified using M13 primers

(Sigma, P3098; Sigma, P2973). The amplicon included the insert flanked by a T7 and a SP6 promoter on either side to allow synthesis of a sense (negative control) and an anti-sense probe (95°C, 5 sec; 50°C 5 sec; 72°C, 30 sec for 35 cycles, then 72°C, 1 min).

The amplicon was then purified using a PCR purification Kit (Qiagen, 28104) and the DNA template concentration was measured using the Qubit 2.0 system (ThermoFisher).

A DIG RNA labelling KIT (Roche, 11175025910) was then used to synthesise DIG-labelled probes from the amplicons according to the manufacturer's protocol.

The solution was then cleaned up using the Cambridge Bioscience RNA clean and concentrator kit following manufacturer's instructions. Probe quality and concentration were determined using the Agilent Tapestation and the NanoDrop 1000 (ThermoFisher). DIG-labeling efficiency was tested using a dot blot test (see 2.2.9).

1) Preserving specimens

Parasites were collected by perfusion from mice seven weeks after infection with 250 mixed sex cercariae. The worms were then washed with DMEM to remove host material. Once clean, worms were killed using 0.6 M magnesium chloride for about 1 min and fixed for 4 h at room temperature in solution of 4% formaldehyde and 0.3% Triton-X 100 in phosphate buffered saline (PBSTx).

Next, the worms were rinsed with PBSTx, placed in 50% methanol in PBSTx for 5 min on a gentle shaker and then transferred to 50 ml tube with 100% methanol and stored at -20°C until needed.

The protocol for whole mount *in situ* hybridisation was run over 3 days; on day 1 the DIG-RNA probe is hybridised to the target RNA, on day 2 the DIG label is bound by the antibody and on day 3, the alkaline phosphatase on the antibody is used for detection.

2) Permeabilisation & hybridisation with probe

Samples were rehydrated in 50% methanol in PBSTx for 5min on a gentle shaker and then for 5 min in PBSTx at RT. Pigments were removed, especially from the guts and the vitellarian tissue of female worms by placing the worms in bleaching solution for 1 hour at RT under bright light. Next, samples were placed in small baskets, transferred to 24-well plates and rinsed in PBSTx. The baskets were used throughout the protocol to allow quicker and less damaging transfer of specimens between different buffers and solutions. The worms were permeabilised with proteinase K (5ug/ml) in PBSTx for 30 min at RT, without shaking. Following proteinase K treatment, the worms were rinsed in PBSTx and then fixed again in 4% formaldehyde in PBSTx for 10 min at RT. Then, the worms were transferred briefly to PBSTx and washed in a 50% PBSTx:50% PreHyb buffer solution for 5 min at RT while being gently shaken. The washing solution was then replaced with pre-warmed 100% PreHyb buffer and samples incubated at 52°C for 2 h while gently shaking. 1 h prior to adding the probe, 1000ng of DIG-RNA probe was mixed with 500 µl of hybridisation buffer and heated to

72°C for 5 min to denature the RNA probe using a heat block (Stuart scientific, SBH130D). The probe was then allowed to cool slowly to 52°C in the heat block and was held at that temperature until needed. The PreHyb buffer was then replaced with the hybridisation buffer with the probe. Next, the samples and probes were allowed to hybridise overnight at 52°C while being gently shaken.

3) Removal of excess probe & incubation with antibody

The hybridisation buffer was removed and samples washed at 52°C with the preheated solutions of wash hybridisation buffer (2 x 30 min), 2 x saline sodium citrate (SSC) buffer and 0.1% triton-X (2 x 30 min), and 0.2 x SSC buffer and 0.1% triton-X (2 x 30 min); finally, worms were washed twice in Tris-NaCl-Tween (TNT) buffer at RT while gently shaking. Samples were transferred to blocking solution for 2 h at RT, still shaking. Next, samples were transferred to blocking solution with anti-DIG-alkaline phosphatase antibody (1:2000 dilution) and incubated at 4°C overnight whilst gently shaking.

4) Removal of excess antibody & developing AP staining

Next, the samples were washed six times for 10 min at RT in TNT buffer.

Then, samples were developed in alkaline phosphatase (AP) buffer with nitro blue tetrazolium and 5-bromo-4-chloro-3-indolyl-phosphate (NBT/BCIP, Sigma, 11681451001) until the desired signal intensity had been reached. To stop the reaction, samples were transferred to PBSTx and placed in 100% ethanol for 10-20 min at RT. Samples were then removed from the ethanol, placed back in PBSTx for about 5 min and then stored in 80% glycerol in PBS.

2.2.8 Imaging of WISH specimen – light microscopy

A Leica M205 FA dissection microscope was used to examine the prepared specimen after staining, images were captured with a Leica DFC 340 FX digital camera.

2.3 RNA-Seq library preparation & sequencing

cDNA libraries were produced from pools of male or female worms, with the exception of the libraries in Chapter 4, where individual worms of both sexes were used, each representing a biological replicate. In the case of pooled worms, all worms originated from the same mouse, forming a biological replicate.

2.3.1 mRNA selection

100 ng of total RNA was used for each RNA-Seq libraries made from pooled worms (Chapter 3 & 5). For the cDNA libraries produced from single worms (Chapter 4) only 50 ng of total RNA was available per sample. Oligo dT beads (ThermoFisher, 61002) were used to increase the mRNA content of the samples, following the manufacturer's instruction. The method makes use of magnetic beads covered in oligo dT molecules, to capture the mRNA by binding to the polyA tails. The rRNA which is not bound is then be washed off.

2.3.2 mRNA fragmentation

mRNA was made up to 200 µl and sheared to around 200 bases using a AFA Covaris focused sonicator (Settings: Duty cycle – 10%; Intensity – 5; time – 60 sec; NB: Duty cycle refers to the proportion of treatment time during which the acoustic burst is active).

2.3.3 Reverse transcription

Similarly, to the cDNA synthesis step in the cloning section (2.2.1), RNA was used to create cDNA in this step. However, in this step the mRNA had been isolated (2.3.1) as well as fragmented (2.3.2) and rather than using oligo-dT primers, random hexamer primers (NEB, S1330S) were used to reverse transcribe all fragments. 1 µl random primers (3 µg/ml) and 1 µl dNTPs (10 mM for each dNTP) (NEB, N0446S) were added to11 µl of fragmented RNA. The mixture was denatured at 65°C for 5 min and chilled on ice. Next, 4 µl of 5x First Strand buffer (NEB, E6114), 1 µl of dithiothreitol (0.1 M; Sigma, D0632) and 1 µl of RNase inhibitor (40 U/µl; ThermoFischer, 10777019) were added. The sample was incubated at RT for 2 min following which 1 µl of SuperScript® III (200 U/µl; ThermoFischer, 18080093) enzyme was added, to a total volume of 20 µl. The sample was then incubated in a thermocycler (25°C, 10 min; 50°C, 60 min; 70°C, 15 min). This denatures the SuperScript® III enzyme but leaves the RNA/DNA duplex intact. After this step, the RNA/DNA duplex was cleaned using the RNAClean XP beads (Beckman Coulter, A63987) following manufacturer's instructions to remove any leftover dTTP nucleotides.

2.3.4 Second strand DNA synthesis

Second strand synthesis was performed with dUTP rather than dTTP allowing the two DNA strands to be differentiated into the sense strand (mRNA sequence – dUTP labelled) and antisense strand (complementary to mRNA – dTTP labelled). 22.6 µl of RNA/DNA duplex from the last step were mixed with 3 µl of buffer 2 (NEB, B7002S), 2 µl dNTP mix (dUTP, dATP; dGTP; dCTP at 10 mM each) (NEB, E6114), 0.4 µl RNase H (5000 U/ml) (NEB, M0297S) to nick (*i.e.* create a break in) the RNA strand, allowing the remaining RNA fragments to act as primers) and 2 µl of DNA pol I (5000 U/ml) (NEB, M0210S). The sample was incubated at 16°C for 2.5 h, then held at 4°C. After second strand synthesis, the sample was cleaned using AMPure XP beads (Agencourt, AQ 60050) (at a ratio of 1 volume of sample to 1.8 volumes of beads). The sample was resuspended in 40 µl of water at the end of the bead purification protocol.

2.3.5 End repair, dA tailing, adapter ligation, size selection

The Sanger Sequencing Kit I (NEB, E6000B-SS) was used to perform end repair, dA tailing, adapter ligation and size selection of the cDNA. All steps were performed following the manufacturer's instructions.

In the next step, the Uracil-specific excision reagent (USER) enzyme (NEB, M5505S) was used to digest the second strand of all DNA molecules in the sample. For this, 1 µl of USER enzyme was added to 10 µl of library. The sample was then incubated in a thermocycler for 15 min at 37°C, then for 10 min at 95°C, and finally held at 4°C.

2.3.6 PCR amplification

After removal of primer dimers and digestion with the USER enzyme, the sample was amplified to reach a sufficiently high DNA concentration for sequencing. For this, 25 µl of 2x KAPA HIFI HS Master mix was added to 22 µl of cDNA as well as 1 µl each of PE 1.0 Illumina primer (10mM), Illumina index primer (10 mM), and water. A unique barcode sequence was added to the cDNAs of each sample using the PCR primer. This allowed DNA sequences to be assigned to one sample *in silico* once sequencing was complete. This tagging allowed for mixing of libraries before sequencing, helping to control for biases, which could otherwise be introduced by differences across sequencing lane and runs. The sample was then amplified using a thermocycler (denaturation 95°C, 5min; 10 cycles of 95°C, 20 sec; 60°C, 15 sec; 72°C, 60 sec; and next 72°C, 5 min). Following the PCR, the samples were size selected, using 0.8x sample volume of Agencourt AMPure XP beads following manufacturer's instructions to remove adapter dimers, which were amplified by the PCR reaction.

2.3.7 Sequencing

All samples were sequenced on an Illumina HiSeq 2500 platform. Samples in Chapter 3 & 4 (60 and 70 samples respectively) were sequenced in three sequencing runs (six lanes in total). The biological replicates for each condition were distributed as evenly as possible across the runs. The 24 samples in Chapter 5 on the other hand were all sequenced together in one run (two lanes in total). Paired-end 100 bp reads were generated.

2.4 Bioinformatics

2.4.1 Aligning RNA-Seq reads to the genome

Tophat2 (v2.0.8b) was used to map RNA-Seq data to version 5.2 of the *S. mansoni* genome (Protasio *et al*., 2012). RNA-Seq reads span splice junctions where introns have been removed from the transcript. Many mapping tools do not take splicing into account when mapping reads to the genome as they are specialised for mapping genomic DNA reads. Tophat2 was specifically designed for RNA-Seq data to allow for optimal mapping of RNA-Seq data to the genome several parameters had to be set manually. By default, Tophat2 distributes a read that maps to multiple locations in the genome randomly to one of those locations. In this case this is undesirable as it could create the appearance of expression in loci that are not expressed, therefore the parameter was set to allow only for uniquely matching reads to be mapped (-g 1). Next, the appropriate library type was specified (--library-type fr-firststrand), based on the preparation method using dUTP. Also based on the method of library preparation, the expected (mean) inner distance between mate pairs was set to 200 (-r 200) and the mate standard deviation was set to 100 (--mate-std-dev 100). To accurately map spliced reads the minimum length of sequence on either side of the splice junction was set to be at least six bases long (-a 6). Finally, the minimum and maximum lengths of introns were also specified as 10 and 40,000 bases respectively (-i 10 -I 40000) and microexons, exons as short as 3 base pairs which are difficult to detect without specialised algorithms (Volfovsky *et al.*, 2003), were allowed for (--microexon-search). The output of this mapping of

RNA-Seq data to the *S. mansoni* genome was provided in Binary sequence Aligned Map (BAM) format.

2.4.2 Sorting and merging of BAM files

SAMtools (v0.1.19) was used for the processing of mapped RNA-Seq data in BAM format (Li *et al.*, 2009). All BAM files were sorted, and the two BAM files (one for each lane of Illumina HiSeq 2500) corresponding to each sample merged, generating a single BAM file per sample.

2.4.3 Counting reads

HTSeq (v0.5.4) was used to summarise the mapped data and produce a list of read counts per gene for each sample (Anders *et al.*, 2015). Using a file in Gene Transfer Format (GTF) as reference for gene boundaries, HTSeq takes into account the strandedness of the data, only counting reads in the orientation of the mRNA and not of antisense RNA. Reads were also only counted for the longest splice variant of each gene to avoid complications in regions where the splice variants overlap and reads could not be unambiguously assigned.

2.4.4 Differential gene expression analysis

RNA-Seq data was analysed in all three results Chapter (3, 4 & 5) to identify differentially expressed genes across groups of samples in R (v3.2.2) (R Core Team, 2015) using DESeq2 Love *et al.* (2014). DESeq2 was created specifically to address the challenges, *i.e.* small numbers of replicates, large dynamic range and

the presence of outliers within the replicates (Love *et al.*, 2014), that arise when analysing high throughput sequencing data, such as RNA-Seq data.

Model and normalisation

Using the output of HTSeq, the number of unambiguously mapped reads per gene, a count matrix is created containing the read count for each gene in each sample. DESeq2 fits a Generalised Linear Model (GLM) for each gene. It models the read counts to follow a negative binomial distribution and calculates a size factor using the median-of-ratios method described in DESeq (Love *et al.*, 2014). This corrects for the depth of sequencing across the different samples, allowing them to be compared against one another.

Empirical Bayes shrinkage for dispersion estimation

To model the variability between replicates, DESeq2 estimates a dispersion parameter, which is critical for proper statistical inference of differential expression. To do so, first the gene-wise dispersion is estimated using maximum-likelihood, then DESeq2 uses this information to estimate the expected dispersion value for genes of a given expression strength by fitting a smooth curve through the distribution of dispersion. DESeq2 then "shrinks the gene-wise dispersion estimates toward the values predicted" to obtain final dispersion values.

Empirical Bayes shrinkage for fold-change estimation

One of the biggest challenges when calculating fold changes using HTS data is the strong variance especially for genes with low read counts where the signal to

noise ratio is less favourable. DESeq2 shrinks log fold-change (LFC) estimates towards zero so that shrinkage is stronger for genes with low read counts, high dispersion, or fewer degrees of freedom. This is achieved by fitting each of the GLMs to obtain a maximum-likelihood estimation of the LFC and then fitting a zero-centred normalised distribution of all LFC. This distribution is then used in another round of GLM fitting, and the corrected estimates are kept as final LFC estimates.

Automatic independent filtering

As differential expression of over 10,000 *S. mansoni* genes were examined in each RNA-Seq experiment, multiple testing adjustments were performed to reduce false positive results using the Benjamini and Hochberg method (Benjamini & Hochberg, 1995). The automatic independent filtering method used by DESeq2 was designed to remove genes from the analysis that have little or no change of being differentially expressed. Importantly, these genes were excluded from further analysis using criteria independent of the statistics used to determine differential expression. The filtering relies on average expression strength for filtering, removing lowly expressed genes first. By default, DESeq2 removes as many genes from the analysis as necessary to maximise the number of genes found to be differentially expressed at the specified false discovery rate (FDR) value (default 10%). In this thesis the automatic independent filtering was most prominently used in Chapter 4 where high dispersion of gene expression presented a challenge for the identification of differentially expressed genes.

Detection of count outliers

To reduce the impact of outliers on the average distribution of dispersion and log fold change, DESeq2 detects and removes individual outliers that do not fit the assumptions of the model such as a replicate with a read count several orders of magnitude higher than all other samples. To achieve this, DESeq2 uses a standard outlier diagnostic called Cook's distance. It measures how much the GLM for a given gene would be affected, if a particular sample was removed. However, this could only be done were three or more replicates were available for a given condition, as outlier status cannot be determined otherwise.

2.4.5 Correlation of RNA-Seq data with microarray data

To give an estimate of the reproducibility of the RNA-Seq data from Chapter 5 – whole male and female worms as well as isolated *S. mansoni* gonads - sequencing data were compared to the microarray signal intensity of comparable samples published by Nawaratna *et al.* (2011). Microarray probe sequences were obtained from the Gene Expression Omnibus public database (accession numbers: GPL10875 & GSE23942). Probe sequences were mapped to the *S. mansoni* genome (version 5.2) using bowtie (v1.1.0), a fast and memory-efficient mapping tool for short reads (Langmead *et al.*, 2009), using default settings. A GTF file that contained the location of each mapped probe in the genome was created using the genome coordinates for each sequence from the BAM file and a custom python script. The number of RNA-Seq reads mapped to the microarray probe sequences was then counted as described in "Counting reads (HTSeq-count)" (2.4.3), using the newly created GTF file as reference. The number of RNA-Seq reads mapped to the probes was then compared to the normalised

signal intensity measured for the probe as reported by Nawaratna *et al.* (2011). The correlation coefficient was calculated for the RNA-Seq and microarray signal using the correlation function in R (R Core Team, 2015).

2.4.6 Principal component analysis (PCA)

PCA plots are used here to visualise data, especially the differences between experimental conditions or detect batch effects. The principal components are variables underlying the data set that explain the maximum amount of data variance with as few principal components as possible, with the aim to visualise multi-dimensional expression data in a two-dimensional plot. To create a PCA plot, the matrix of normalised counts created in DESeq2 was used. A regularised log transformation was performed on the data matrix. This has a variance stabilising effect, especially helping to minimize the differences between samples for rows with small counts. DESeq2 then provided a function to create the PCA plot based on the transformed data (Love *et al.*, 2014).

2.4.7 Heatmaps

The R package "pheatmaps" (V1.0.8) was used to draw heatmaps (Kolde, 2015). This software uses count data after regularised logarithmic transformation (performed by DESeq2) as input to draw heatmaps of a specified subset of genes. To produce a heatmap, the program then calculates a Z-score, *i.e.* the number of standard deviations a data point is from the mean, which allows for better scaling across all samples than for example plotting the log-fold change. K-means clustering is used by pheatmaps.

2.4.8 Gene Ontology (GO) term enrichment

To perform GO term enrichment analysis on the sets of differentially expressed genes, topGO (Alexa *et al.*, 2006) was used. GO terms provide a hierarchical structure to classify genes according to their molecular function, localisation and the biological process in which they are involved (GO Consortium, 2004). The analysis of GO terms in this thesis was restricted to those in the category "Biological Process". Genes significantly up or down-regulated (usually with an adjusted p-value of < 0.01) in a particular condition were used as input for topGO and GO terms were considered significantly enriched if their p-value was < 0.05.

2.4.9 InterProScan & Pfam enrichment

The program InterProScan 5.0.7 (Quevillon *et al.*, 2005; Zdobnov & Apweiler, 2001) was used to identify conserved protein domains. The sequences of all annotated *Schistosoma mansoni* proteins were used as input. This produces a list of all matches between the provided sequences and annotated protein domains, including Pfam matches. From the output, all Pfam (Finn *et al.*, 2014) domains were used that were identified with high confidence (p-value < 0.01). Duplicate domains were also removed, *i.e.* those that occurred more than once in a protein, to prevent the enrichment statistics to be skewed by such proteins. Furthermore, all domains, only found in a single *S. mansoni* protein were removed because no meaningful enrichment analysis could be performed on them. In total 1479 different domains in 6835 proteins were kept for further analysis.

Using a custom python script this information was combined with the results of differential expression analysis from DESeq2 to determine if the differentially expressed genes (DEGs) were enriched for proteins containing particular domains. The script created a table as output that was opened using Microsoft Excel to calculate if a domain was significantly more or less abundant than expected by change using a two-sided hypergeometric test. Domains were accepted to be significantly enriched if $p < 0.05$ and if the domain in question was found more frequently than the average domain. This resulted in excluding domains which were actually depleted in a sample, rather than enriched. For examples it might exclude the egg shell synthesis domains from being shown as significantly depleted in male worms; however, this domains would be shown as significantly enriched in female worms.

2.4.10 KEGG pathway enrichment

The Kyoto Encyclopedia of Genes and Genomes (KEGG) database online resource (Kanehisa & Goto, 2000) was used to provide pathway information, as they have reference pathways from hundreds of metabolic and signalling pathways. The html files of all KEGG entries of *S. mansoni* genes were downloaded and the corresponding GeneDB IDs as well as KEGG pathways associated with the gene in question were extracted from the files. Using this information, a table containing all *S. mansoni* genes and all pathways that the given gene belongs to was created. In total 2046 genes with KEGG pathway annotation were identified. In total genes belonging to 112 pathways were found.

After DESeq2 analysis of RNA-Seq data, a python script was used to count the number of DEGs found to belong to each pathway. The script would create a table that can be opened by Microsoft Excel to calculate if the number of genes belonging to a given pathway was significantly higher than expected by chance using a two-sided hypergeometric test. Pathways were accepted to be significantly enriched if $p < 0.05$ and if it was found more frequently that the average pathway. As in the case of domain enrichment, pathways which were actually depleted, rather than enriched were excluded from the analysis, as they would usually be found to be significantly enriched in the other condition of the DESeq2 analysis.

For example, in Chapter 5, in a comparison of female worms from SS and MS infections, 2726 DEGs were identified, of which 787 were up-regulated in MS females. Of the 787 DEGs in MS females, 365 were annotated as members of a KEGG pathway. Across all pathways, about 37% of annotated genes were up-regulated in MS females, including 75 of 110 (68%) genes associated with the Ribosome KEGG pathway (smm03010); a greater proportion of DEGs in this pathway are upregulated than expected by chance as was determined using a hypergeometric test, demonstrating that the genes up-regulated in MS females are significantly enriched with ribosome related genes ($p = 8.47E-12$).

On the other hand, in SS females, 1939 DEGs were identified, but none of the 110 ribosome-associated genes were found to be up-regulated. On average only about 14% of the genes associated with KEGG pathways were up-regulated in SS females. As a result, expression of ribosomal genes was also found to be

significantly different from the expected (14%) DEGs in SS females (p-value = 6.67E-08). However, in this case the proportion of DEGs in the Ribosome pathway is lower than expected by change (it is in fact 0). As a result, the hypergeometric test demonstrated that the genes up-regulated in SS females are significantly depleted of ribosome related genes.

Because the depletion of a pathway in one sample usually mirrored the enrichment of that same pathway in the other sample in a gene expression comparison, only significantly enriched pathways were considered in the results of the following chapters.

2.4.11 Cluster analysis of RNA-Seq data

Two R packages were used to cluster RNA-Seq data by the pattern of gene expression across different samples: MBCluster (Si *et al.*, 2013) and Kohonen (Wehrens, 2015). The MBCluster analysis was performed in R using counts normalised for library size as input; the size factors used to correct for library size were provided by DESeq2 (see 2.4.4). MBCluster models the data to follow a negative binomial distribution and uses an Expectation-Maximisation algorithm to estimate model parameters and divides genes into groups, or clusters, with similar expression patters across the different samples. Importantly, MBCluster was designed to cluster genes according to their expression profile (*i.e.* the relative changes in expression) but not absolute expression levels. Kohonen is also an R package used to group genes by their expression across different samples. Unlike MBCluster, Kohonen is based on a heuristic method designed to produce self-organising maps (SOM), a type of artificial neural network, from

data sets using unsupervised learning. It uses data that has undergone regularised logarithmic transformation as input (provided by DESeq2). The data was then processed to have a mean expression of zero for each gene. This allows Kohonen to cluster the genes by the magnitude of changes in their expression, not absolute expression. Kohonen initially uses a random distribution of data point across the map, but then trains the map until it reaches an optimal distribution of data point across the SOM.

2.4.12 Gene models & annotation

The annotation of DEGs was checked using a combination of BLAST, against gene sequences of model organisms, as well as Pfam to identify relevant domains in the predicted protein sequences. For Chapters 3 and 4 a list of putative apoptosis-related genes was compiled. These genes were identified from the literature (Lee *et al.*, 2011; Lee *et al.*, 2014; Peng *et al*., 2010), as well as by using BLAST to identify *S. mansoni* homologues of genes annotated in the KEGG database as apoptosis-related genes. Especially *Homo sapiens* and *Caenorhabditis elegans* gene sequences were used to BLAST against all *S. mansoni* genes on the GeneDB database. An e-value of 1.00E-05 was used as significance cut-off. Additionally, Pfam was used to confirm the presence of conserved domains, important to the function of protein product. Using the Artemis Comparison Tool, RNA-Seq evidence was used to improve predicted gene models of the apoptosis-related genes. Using the RNA-Seq data mapped by Tophat2, both exons that had not been previously annotated, as well as the exon-intron boundaries were checked and any changes submitted to GeneDB.

2.5 Scanning Electron Microscopy

Freshly perfused worms were washed in PBS and then fixed in 2.5% glutaraldehyde and 2% Paraformaldehyde in PBS for 1 h. Specimens were then prepared for scanning electron microscopy using the osmium tetroxide/thiocarbohydrazide method (Malick & Wilson, 1975). The following steps were performed by Dave Goulding, WTSI, Hinxton. The samples were rinsed in 0.1 M sodium cacodylate buffer and then alternated between incubation in 1% osmium tetroxide and incubation in 1% thiocarbohydrazide for a total of 5 incubations. Following these incubations, the specimens were dehydrated in an ethanol series (30%, 50%, 70%, 90% and 100%). Then a critical point drying was performed in a Bal-Tec CPD030 and specimens were mounted on aluminium stubs with silver dag. Finally, samples were coated with a 2 nm gold layer in a Bal-Tec SCD050 and examined in a Hitachi S-4800 scanning electron microscope.

Chapter 3

Transcriptome analysis of male and female *S. mansoni* during intra-mammalian development

3.1 Introduction

The aim of this Chapter is to explore the transcriptome changes that occur during the development of male and female schistosomes inside the mammalian host, with a particular emphasis on sexual maturation taking place after pairing. As discussed in detail in the introduction to this thesis, female maturation is dependent on stimulation by a male partner. Here, I examined the transcriptomes of female and male worms in the days leading up to pairing as well as the weeks following pairing. In particular, I wanted to explore how the male and female transcriptomes change over time and how the gene expression of worms from mixed sex (MS) infections differs from that of single sex (SS) infections.

In this Chapter, a variety of functional genomic tools and databases were used to investigate differentially expressed genes. Using a combination of different tools, such as GO terms, KEGG pathway and Pfam domain enrichment, I was able to get a more thorough understanding of the processes taking place in the developing worms.

While many previous studies on pairing-induced changes in gene expression used microarrays (Fitzpatrick & Hoffmann, 2006; Leutner *et al*., 2013), a recent study used RNA-Seq (Sun *et al.*, 2014). An analysis of female worms at 18 and 23 d.p.i. from both SS and MS infections was performed, to study female gene expression in the presence or absence of male worms, but also the changes that take place between the two time points. Although almost 3000 differentially expressed genes (DEGs) were identified in some of their comparisons, their

approach left room for improvement. Their method of RNA-Seq relied on sequencing tags, which are short, 17 base-pair sub-region of a whole cDNA sequence. As result the tags do not cover transcripts along their entire lengths which can make gene expression less quantifiable (Wang *et al.*, 2009). More importantly, the authors did not sequence any biological replicates for any of their samples for their conditions, although this has now become standard practice as it allows for a more robust statistical analysis by making it possible for the variation within treatment groups to be measured accurately (Schurch *et al.*, 2016). This is because variability between biological replicates is often higher than expected, so without biological replicates real differences in expression between treatment groups cannot be reliably distinguished from random fluctuations in expression (Schurch *et al.*, 2016). The method of determining differentially expressed genes is not described in detail in the methods (the authors refer to "a rigorous algorithm [that] was adopted to screen genes that were differentially expressed between samples"), making it difficult to replicate their analysis and their results (Sun *et al*., 2014). Finally, the authors performed a pathway analysis using the identified DEGs (Sun *et al.*, 2014). However, in their analysis Sun *et al.* (2014) did not evaluate the proportion of genes that were differentially expressed in each pathway and therefore did not examine the statistical significance of the up- or down-regulated pathways that they reported. For example, Sun *et al.* (2014) report that "pathways, such as cell adhesion molecules (CAMs), ribosome biogenesis in eukaryotes, the Notch signalling pathway" were up-regulated. However, it is not clear if what percentage of genes of a particular pathway were differentially expressed. A low proportion of DEGs

in a pathway could be the result of random chance, whereas a larger proportion could give greater confidence in the result.

Another study examining the development of *S. mansoni* from cercaria to adult worm (Picard *et al.*, 2016). However, the results were limited by the fact that female specimens did not experience pairing and thus did not commence sexual maturation over the course of the experiment (Picard *et al.*, 2016). Most recently, Wang *et al*. (2017) published a RNA-Seq time course of the intra-mammalian development of male and female *S. japonicum* from 14 to 28 d.p.i.. The results suggested that biogenic amines and insect-hormone-like receptors play crucial roles in male-female signalling and sexual maturation of ovaries and vitelline glands in female worms (see Chapter 1.6.2). Using RNAi targeted at the insect-hormone-like receptor, the authors provided evidence of a role of the insect-hormone-like receptor in the development of these reproductive tissues (Wang *et al.*, 2017). However, no RNAi results were reported for the AADC gene involved in the production of amines. The authors only analysed the transcriptomes of worms from MS infections. This makes it more difficult to distinguish the changes in expression linked to pairing from those that occur independently of pairing as the worms develop without the opposite sex. Notably, the development of *S. japonicum* as described by Wang *et al.* (2017) appears to be accelerated when compared to *S. mansoni* (see Chapter 1.2.1), as the worms were observed to pair and reach sexual maturity consistently earlier than in equivalent studies on *S. mansoni* (Biolchini *et al*, 2006). Differences in the speed of development, even between strains of the same species of schistosome, have been reported before (Hsü *et al.*, 1962).

In my experiment, male and female *S. mansoni* from MS and SS infections were collected and processed for RNA-Seq at six time points: days 18, 21, 28, 35, 38 and 49 d.p.i.. My analysis included not just the transcriptome changes taking place in female worms but also those of male worms. Robust statistical analyses were used to determine differentially expressed genes and pathway enrichment analysis. Additional bioinformatics tools were used, including Pfam domain analysis, GO terms and cluster analysis (see Chapter 2.4), to infer the potential roles of differentially expressed genes in the development and sexual maturation. Finally, I used whole mount *in situ* hybridisation and RNAi to characterised two genes found to be involved in female sexual maturity. This was done to explore the site of expression as well as the function of these genes in *S. mansoni* using respectively.

3.2 Results

3.2.1 Egg laying

The livers of infected mice were examined for the presence of *S. mansoni* eggs (Figure 3.1), as described in Chapter 2.1.3. Schistosome eggs were found exclusively in livers of mice infected with MS parasites at 35 and 38 d.p.i. As expected, eggs were not observed in the livers of mice with the SS infections. Although paired worms were recovered at 28 d.p.i. from mice with MS infections, no eggs could be recovered from the livers of these mice. Since pairing is thought to take place from around 28 d.p.i. (Basch, 1991) most female worms would not

started egg laying at this time point. However, a very low egg count could have gone undetected.

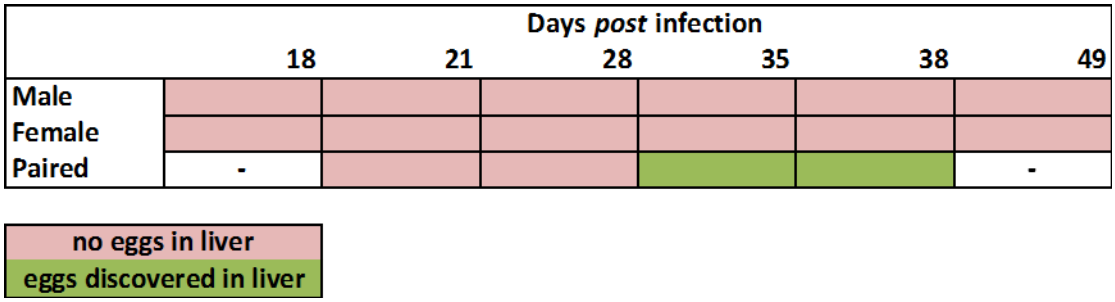


Figure 3.1: Eggs were only detected in the livers of mice with mixed sex infections. Single sex infections with male or female worms were found not to yield eggs at any time point, whereas mixed sex infections were found to yield eggs from 35 d.p.i., but not 21 and 28 d.p.i. At 18 and 49 d.p.i. no mixed sex infections were available ("-").

3.2.2 Sequencing and sample clustering

The RNA-Seq data was prepared as described in Chapter 2.4, yielding over 20 million reads for each of the 70 samples (see Appendix A.1 for details). On average, 79% of the reads could be mapped to the *S. mansoni* genome (v5.2) (Appendix A.1), with the remainder mostly comprising PCR artefacts, especially primer dimers, created during the amplification of the sequencing library. Samples were then analysed using DESeq2. First, a principal component analysis (PCA) plot (Figure 3.2) was created of all samples. This gives a simplified overview of how developing male and female worms differ at the transcriptome level across the time course from 18 to 49 d.p.i. in MS and SS infections. While male worms from SS and MS infections develop at a similar rate, females from SS infections are known to become arrested at an early developmental stage (Popiel, 1986). This was reflected at the transcriptome level by the PCA plot as well as the following differential expression analysis (see Chapters 3.2.4 and

3.2.5). From 21 d.p.i. the gene expression of SS females remained almost unchanged, developmentally arrested, and most similar to MS females at 21 d.p.i. Females from MS infections became very distinct from all other samples from 28 to 38 d.p.i. at the transcriptome level.

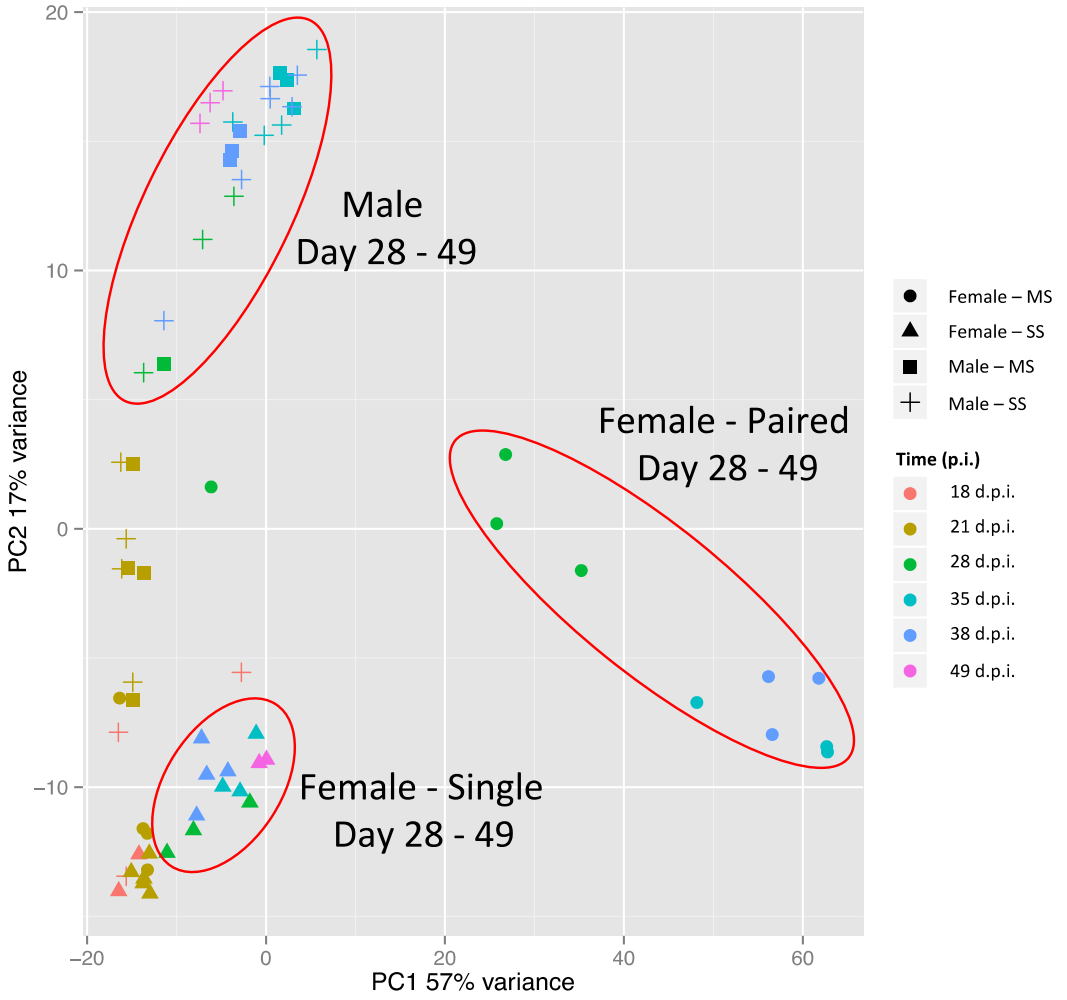


Figure 3.2: Samples group into male, as well as mixed and single sex female clusters. Principle Component Analysis (PCA) plot of gene expression in developing male and female *S. mansoni* from single sex (SS) and mixed sex (MS) infections at 18 – 49 days *post* infection (d.p.i.) PC1 explains 57% of the variance observed in the data set and broadly correlates with female sexual maturity over time, whereas PC2 explains only 17% of the variance and broadly correlates to male maturity. Together 74% of variance across the data set is captured by the first two principal components.

These findings reflected that female sexual development depends on pairing with a male whereas male maturation is largely independent of the female, as has been noted in the literature (Basch, 1991). In this context, it was also interesting to note that female worms from MS and SS infections were transcriptomically very similar at 21 d.p.i., but diverged significantly after 28 d.p.i., when pairing is thought to take place (Basch, 1991). It should also be noted that male and female samples from MS infections separate very well at all available time points (it should be noted that the males and females from SS infections at 18 d.p.i. did not separate). This shows that males and females were correctly identified even at 21 d.p.i. using morphological differences outlined by Basch (1991).

3.2.3 Gender-specific gene expression

During their intra-mammalian development, male and female worms grow from morphologically similar immature stages to distinct male and female worms that are usually reproductively active by 35 d.p.i. (Biolchini *et al.*, 2006). Pearce and Huang (2015) suggested that the sexual dimorphism of male and female worms can be explained by adaptations for different biological roles. This allows for a unique co-dependency that maximises egg production. The male, for instance, is responsible for moving the pair to the site of oviposition (Basch, 1991).

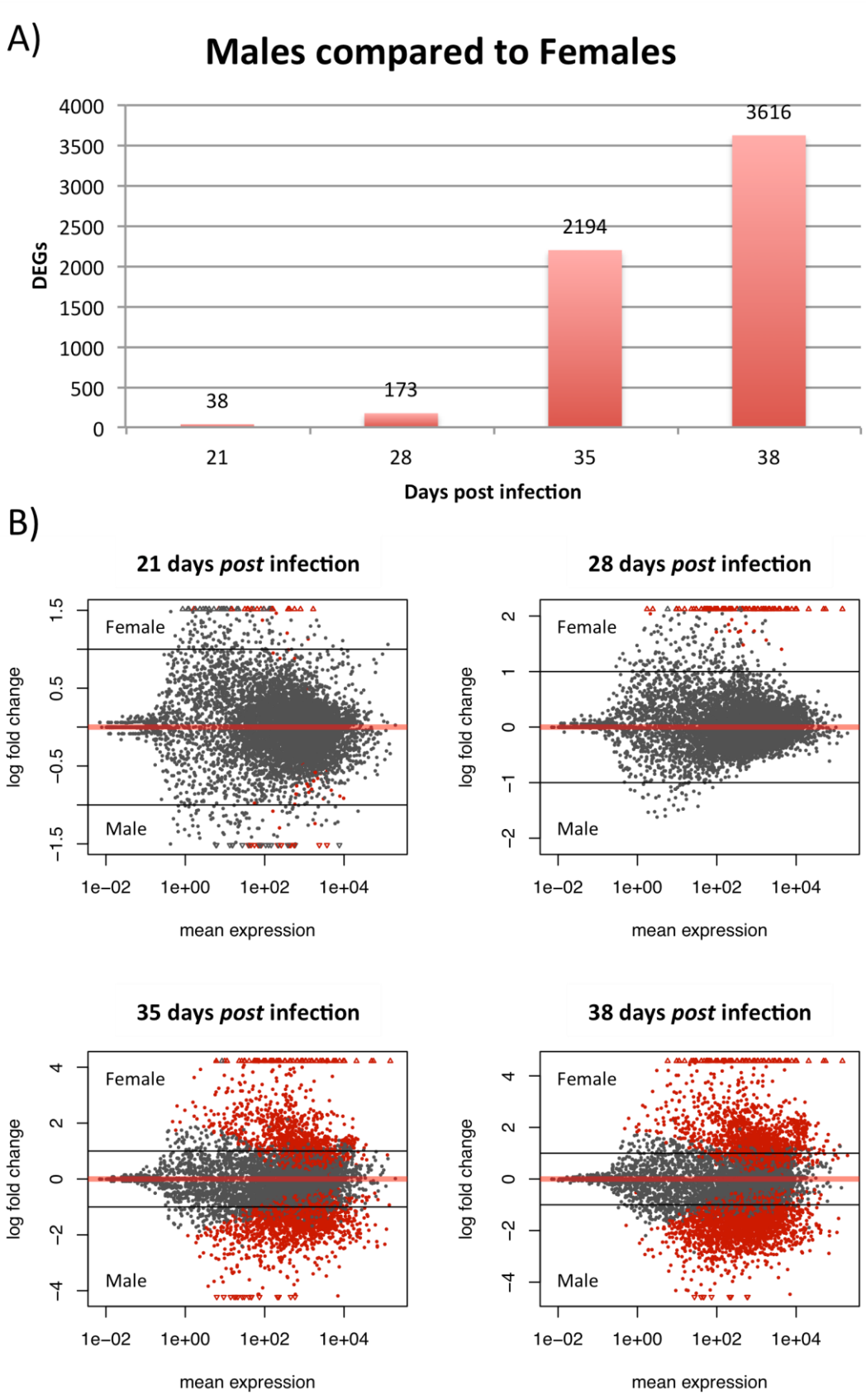

A)
Males compared to Females
DEGs
4000
3500
3000
2500
2000
1500
1000
500
0
38
173
2194
3616
21
28
35
38
Days post infection
B)
21 days post infection
28 days post infection
35 days post infection
38 days post infection
Female
Male
log fold change
mean expression
1e−02
1e+00
1e+02
1e+04

Figure 3.3: The male and female transcriptome becomes increasingly different, especially from 28 to 35 days *post* infection. A) Number of differentially expressed genes between male and female worms from mixed sex infections at 21 to 38 days *post* infection. Genes were considered to be differentially expressed if the adjusted p-value of < 0.01 and fold-change > 2. B) Log ratio vs mean (MA) plots of pair-wise comparisons of gene expression in male and female worms from MS infections from 21 to 38 d.p.i. Genes with an adjusted p-value of < 0.01 are highlighted in red, the others are grey. The black lines represent the fold change cut-off (log-fold change = ± 1). Triangles indicate genes outside the boundaries of the plots; like other genes, they are red if they are statistically significant, grey if they are not.

In order to identify genes differentially expressed between male and female worms from MS infections at 21, 28, 35 and 38 d.p.i. DESeq2 was used. The threshold for differential expression was an adjusted p-value of < 0.01 and a fold-change in expression > 2 throughout this Chapter. Pairwise comparisons of males and females at each time point showed that there was an increase from 38 DEGs at 21 d.p.i. to over 3600 DEGs at 38 d.p.i. (of 10828 genes in total). The biggest absolute increase occurred from 28 to 35 d.p.i., approximately 1270% (Figure 3.3), as males and females diverge transcriptomically. These changes are responsible for the development of sex-specific features and the increasing proportion of reproductive tissues such as ovaries and vitellaria in female worms.

At 21 d.p.i., male and female worms from MS infections were very similar at the transcriptome level. Of the 38 DEGs (see Figure 3.3a), 22 were up-regulated in females and 16 in males (see Appendix A.2 and A.3 for full lists of DEGs). In female worms several genes involved in cell division and cell fate determination

could be identified (Table 3.1). First, the gene coding for MutS, a protein implicated in mismatch DNA repair, was up-regulated (Table 3.1). Second, female worms express higher levels of the gene coding for Delta-like protein (Table 3.1), which is a ligand of Notch receptors involved in cell fate determination (Zhou *et al.*, 2010). Third, a gene coding for a *single minded*-like protein was up-regulated in female worms (Table 3.1). In *Drosophila*, *single minded* is required for the proper development of midline cells (Nambu *et al.*,1991). At 21 d.p.i. both male and female worms are in the process of developing their reproductive organs (Biolchini *et al.*, 2006) as well as their gut (Basch, 1991) and these gene may be involved in regulating these processes.

Gene ID	Product	Fold Change	Adjusted p-value
Smp_213480	MutS protein	6.54	9.36E-06
Smp_178780	Single minded	4.78	0.00160
Smp_141250	Jagged 1	2.59	0.00874

Table 3.1: **Cell division related genes up-regulated in females from mixed sex infections at 21 days *post* infection.** Fold changes are relative to expression in male worms from mixed sex infections at 21 d.p.i. (see Appendix A.3 for full list). The *p*-value has been adjusted for multiple hypothesis testing.

In males, two nervous system related genes were found to be up-regulated at 21 d.p.i. (Appendix A.3). One of the two genes encoded a sodium chloride dependent neurotransmitter (Smp_160360, 3.54-fold change, adjusted p-value = 3.36E-07). The other encoded a neuropeptide receptor (Smp_133550, 2.0-fold change, adjusted p-value = 0.0067). Collins *et al*. (2010) showed that neuropeptide Y is involved in the development and maintenance of the germline in planarians and identified peptide homologues in *S. mansoni*. This evidence suggests the differential use of neuropeptide receptors in male and female *S. mansoni* in the

mammalian stage of their life cycle. Smp_133550 remains significantly up-regulated in males at 38 d.p.i. (7.1-fold change, adjusted p-value = 2.79E-21) (Appendix A.4). Together, the differentially expressed genes suggest that the development of male and female worms including that of the nervous system has begun to diverge by 21 d.p.i., priming the otherwise transcriptomically very similar males and females for their sex-specific development.

At 38 d.p.i. approximately 33.4% (3616) of all genes were found to be differentially transcribed (Figure 3.3), revealing the extraordinary degree to which mature females from MS infections differ from their male counterparts. 1377 of the 3616 DEGs were up-regulated in females (Figure 3.3). Among the female DEGs were several encoding well known fertility-associated proteins including female specific protein 800 (fs800) and other trematode egg shell synthesis domain containing proteins, two tyrosinases and an extracellular superoxide dismutase all of which were at least 80-fold up-regulated compared to male worms (adjusted p-value < 4.00E-22) (Appendix A.5). Another gene up-regulated in females encodes a homologue of the human TNF receptor associated protein 1 (TRAP1; Smp_155740) (Appendix A.5).

At 38 d.p.i., MS male worms express 2239 genes at significantly higher levels than females (Figure 3.3). A selection of 29 genes pre-dominantly expressed in males was identified in the literature (Fitzpatrick *et al.*, 2005; Piao *et al.*, 2011) and tested for male-biased expression. All 29 genes were found to be differentially expressed (Table 3.2), including genes coding for collagen, calpain and musculature related proteins.

Gene ID	Product	Fold change	Adjusted p-value
Smp_006860	PDZ & LIM protein	4.66	9.42E-23
Smp_022340	PDZ & LIM protein	5.83	6.54E-22
Smp_166920	PDZ & LIM protein	7.50	8.16E-22
Smp_164590	Fibrillar collagen α	4.65	8.03E-12
Smp_123830	Collagen α (xi) chain	4.14	6.81E-12
Smp_196840	Collagen α 1(II) chain	2.59	7.50E-09
Smp_170340	Collagen α 1(IV) chain	5.68	2.01E-07
Smp_156770	Tegumental antigen Sm15	7.26	5.12E-14
Smp_179810	Tegumental antigen Sm8	5.06	8.86E-15
Smp_214180	Calpain	3.48	7.16E-18
Smp_214190	Calpain	3.03	2.08E-21
Smp_020550	Ldl receptor	3.85	1.15E-06
Smp_161930	Actin	4.56	1.49E-18
Smp_059170	Troponin	7.43	1.07E-22
Smp_174520	Dynein light chain	4.08	0.0031
Smp_176300	Dynein heavy chain	2.34	1.13E-06
Smp_212280	Dynein heavy chain	3.11	1.17E-13
Smp_140500	Paramyosin	2.13	0.0003
Smp_021920	Paramyosin	6.85	9.62E-08

Table 3.2: Differential expression analysis confirmed male-biased expression patterns of known male-specific genes from the literature (Fitzpatrick *et al.*, 2005; Piao *et al.*, 2011)**.** This was a comparison of male and female worms from mixed sex infections at 38 days *post* infection. Fold changes are provided relative to female expression; all genes in this table were up-regulated in male worms. The *p*-value has been adjusted for multiple hypothesis testing.

As in MS males at 21 d.p.i., the neuropeptide receptor (Smp_133550) was found to be up-regulated in MS males at 38 d.p.i. (fold change = 7.1, adjusted p-value = 2.79E-21). Additionally, another G-protein coupled receptor (Smp_170020, 9.1-fold change, adjusted p-value = 2.54E-30) with homology to neuropeptide receptors was found to be up-regulated in the male worms (Table 3.3). Overall, a high proportion (13 out of 16) of neurotransmitter-gated ion channels (identified by their PFAM domain annotation) were significantly up-regulated in male worms compared to female worms (Table 3.3). These include three

glutamate-gated chloride channel subunits (2, 3 and 4) (Smp_015630, Smp_104890 and Smp_099500), seven out of eight nicotinic acetylcholine receptors and receptor subunits (Smp_031680, Smp_012000, Smp_132070, Smp_180570, Smp_139330, Smp_157790 and Smp_176310), as well as an acetylcholine receptor subunit α (Smp_101990), a neural acetylcholine receptor subunit α 4 (Smp_037960) and a glycine receptor subunit β (Smp_096480) (Table 3.3). Of the remaining three neurotransmitter receptors identified in the *S. mansoni* genome, one is up-regulated in females (Smp_142690), while expression of the remaining two could not be detected at this stage of the life cycle (Smp_175110, Smp_197600) (Table 3.3). This striking enrichment in neurotransmitter receptor mRNA expression in male worms over female worms raises questions on the role of these proteins in the male reproductive process. When comparing expression of these 13 genes in males to female worms from SS infections, no such enrichment is found (Appendix A.6); only two of the genes, Smp_104890 and Smp_101990, are still expressed at significantly higher levels in male worms than in the immature females (Appendix A.6). This suggests that the relative abundance of different tissue types is responsible for the up-regulation of neurotransmitter receptor genes in male worms. The large proportion of ovarian and particularly vitellarian cells in mature females as well as the relatively lower proportion of muscle (Loker & Brant, 2006) may be responsible for the lower relative expression of neurotransmitter genes in whole females. In contrast, immature females are more male-like than mature females at the transcriptome level, due to the lack of developed reproductive organs.

Gene ID	Product	Average read count	Fold change	Adjusted p-value
Smp_015630	Glutamate-gated chloride channel subunit	353.74	4.61	1.23E-06
Smp_104890	Glutamate-gated chloride channel subunit	244.18	6.03	7.84E-24
Smp_099500	Glutamate-gated chloride channel subunit	38.57	2.75	0.0053
Smp_031680	Nicotinic acetylcholine receptor	176.02	5.13	4.92E-08
Smp_012000	Nicotinic acetylcholine receptor	92.58	3.52	8.81E-05
Smp_132070	Nicotinic acetylcholine receptor subunit	3562.48	2.67	5.57E-16
Smp_180570	Nicotinic acetylcholine receptor	132.26	4.04	9.15E-08
Smp_139330	Nicotinic acetylcholine receptor non α	811.23	4.32	1.98E-18
Smp_157790	Nicotinic acetylcholine receptor subunit	121.98	4.31	2.22E-06
Smp_176310	Nicotinic acetylcholine receptor subunit	699.76	3.58	6.37E-07
Smp_101990	Acetylcholine receptor subunit α	23.03	2.53	0.0053
Smp_037960	Neuronal acetylcholine receptor subunit α	1308.18	2.75	1.76E-05
Smp_096480	Glycine receptor subunit β	520.33	3.44	1.93E-11
Smp_142690	Neuronal acetylcholine receptor subunit α	1875.19	0.42	2.93E-21
Smp_175110	Neurotransmitter gated ion channel	0.00	N/A	N/A
Smp_197600	Nicotinic acetylcholine receptor subunit	0.68	1.5	0.3996

Table 3.3: Neurotransmitter-gated ion channels are largely up-regulated in male worms. The differential expression analysis compares male and female worms from mixed sex infections at 38 days *post* infection. Average read counts are a measure of absolute expression across all samples, and fold changes are provided relative to female expression (fold changes > 1 indicate up-regulation in males). If no expression was measured, fold change and p-value are marked as N/A. The *p*-value has been adjusted for multiple hypothesis testing.

To better understand what biological processes the DEGs play a role in in females and males I made use of the KEGG database and performed a pathway enrichment analysis to identify pathways with significantly more DEGs than expected by chance. At 38 d.p.i., the DEGs upregulated in MS females were found to be enriched for 13 pathways involved in DNA replication and protein synthesis (Table 3.4), reflecting the proliferation of ovarian and especially vitellarian cells to drive egg production. In contrast, several lipid-metabolic pathways, developmental signalling pathways and pathways involved in

endocytosis were significantly enriched in males (see Table 3.5). Amongst these, is the TGF-β pathway that has been implicated in the male-female interaction and the stimulation of female fertility (LoVerde *et al.*, 2009) (Table 3.5). Another up-regulated pathway is phosphatidylinositol signalling (smm04070, p-value = 0.020) (Table 3.5). It has been proposed that female fertility is supported by lipid transfer from the male to the female (Popiel & Basch, 1986). Several lipid metabolism-related pathways were indeed significantly enriched in the DEGs upregulated in males at 38 d.p.i., including glycerolipid and glycerophospholipid metabolism (smm00564, p-value = 0.00019; smm00561, p-value = 0.00059) as well as ether lipid metabolism (smm00565, p-value = 0.0082) (Table 3.5). However, this could also simply reflect a higher rate of tegument turnover in male worms, which have a thicker tegument as well as a larger surface area, due to their flattened morphology.

KEGG ID	Pathway	Total Genes	DEGs	Expected	p-value
smm03010	Ribosome	110	82	42.98	7.92E-15
smm03008	Ribosome biogenesis in eukaryotes	62	47	24.23	2.22E-09
smm03050	Proteasome	32	27	12.50	1.34E-07
smm03030	DNA replication	30	22	11.72	1.06E-04
smm03040	Spliceosome	105	59	41.03	1.11E-04
smm03060	Protein export	19	15	7.42	3.84E-04
smm00510	N-Glycan biosynthesis	30	21	11.72	4.13E-04
smm03440	Homologous recombination	18	14	7.03	7.83E-04
smm00970	Aminoacyl-tRNA biosynthesis	34	21	13.29	3.79E-03
smm03013	RNA transport	103	52	40.25	4.50E-03
smm04141	Protein processing in endoplasmic reticulum	89	45	34.78	7.02E-03
smm03410	Base excision repair	20	12	7.82	3.00E-02
smm00240	Pyrimidine metabolism	59	28	23.05	4.37E-02

Table 3.4: Female worms up-regulated KEGG pathways playing a role in protein synthesis and DNA replication. At 38 days *post* infection, genes significantly upregulated in females from mixes sex infections were enriched in 13 KEGG pathways. “Total genes” is the number of genes in the given pathway. “DEGs” provides the number of differentially expressed genes in that pathway. “Expected” provides the number of genes expected by chance to be differentially expressed in each pathway.

KEGG ID	Pathway	Total Genes	DEGs	Expected	p-value
smm04144	Endocytosis	64	26	10.82	3.15E-06
smm04512	ECM-receptor interaction	9	8	1.52	4.71E-06
smm04068	FoxO signalling pathway	38	18	6.43	8.63E-06
smm04310	Wnt signalling pathway	35	17	5.92	1.02E-05
smm04080	Neuroactive ligand-receptor interaction	15	10	2.54	2.10E-05
smm00564	Glycerophospholipid metabolism	35	15	5.92	1.88E-04
smm00561	Glycerolipid metabolism	17	9	2.87	5.93E-04
smm00500	Starch and sucrose metabolism	14	8	2.37	6.33E-04
smm00380	Tryptophan metabolism	10	6	1.69	2.29E-03
smm00340	Histidine metabolism	5	4	0.85	3.36E-03
smm04142	Lysosome	47	15	7.95	4.96E-03
smm00565	Ether lipid metabolism	9	5	1.52	8.19E-03
smm04340	Hedgehog signalling pathway	9	5	1.52	8.19E-03
smm00514	Other types of O-glycan biosynthesis	4	3	0.68	1.60E-02
smm00010	Glycolysis / Gluconeogenesis	27	9	4.57	1.85E-02
smm04070	Phosphatidylinositol signalling system	32	10	5.41	2.05E-02
smm04630	Jak-STAT signalling pathway	11	5	1.86	2.08E-02
smm04320	Dorso-ventral axis formation	8	4	1.35	2.71E-02
smm04350	TGF-β signalling pathway	16	6	2.71	2.91E-02

Table 3.5 Male worms upregulated KEGG pathways involved in cellular metabolism and various signalling pathways. At 38 days *post* infection, genes upregulated in males from mixed sex infections were enriched in 19 KEGG pathways. "Total genes" is the number of genes in the given pathway. "DEGs" provides the number of differentially expressed genes in that pathway. "Expected" provides the number of genes expected by chance to be differentially expressed in each pathway.

3.2.4 Male development

In male worms, the greatest difference in gene expression was found between 18 and 21 d.p.i. (1858 DEGs in total) (Figure 3.4), suggesting that this was when most of the male-specific developmental changes took place. The number of DEGs dropped sharply after that time, to 130 (21 vs 28), 59 (28 vs 35) and finally 17 DEGs (35 vs 38), (Figure 3.4). Below, some of the results of the differential expression analysis are discussed in detail.

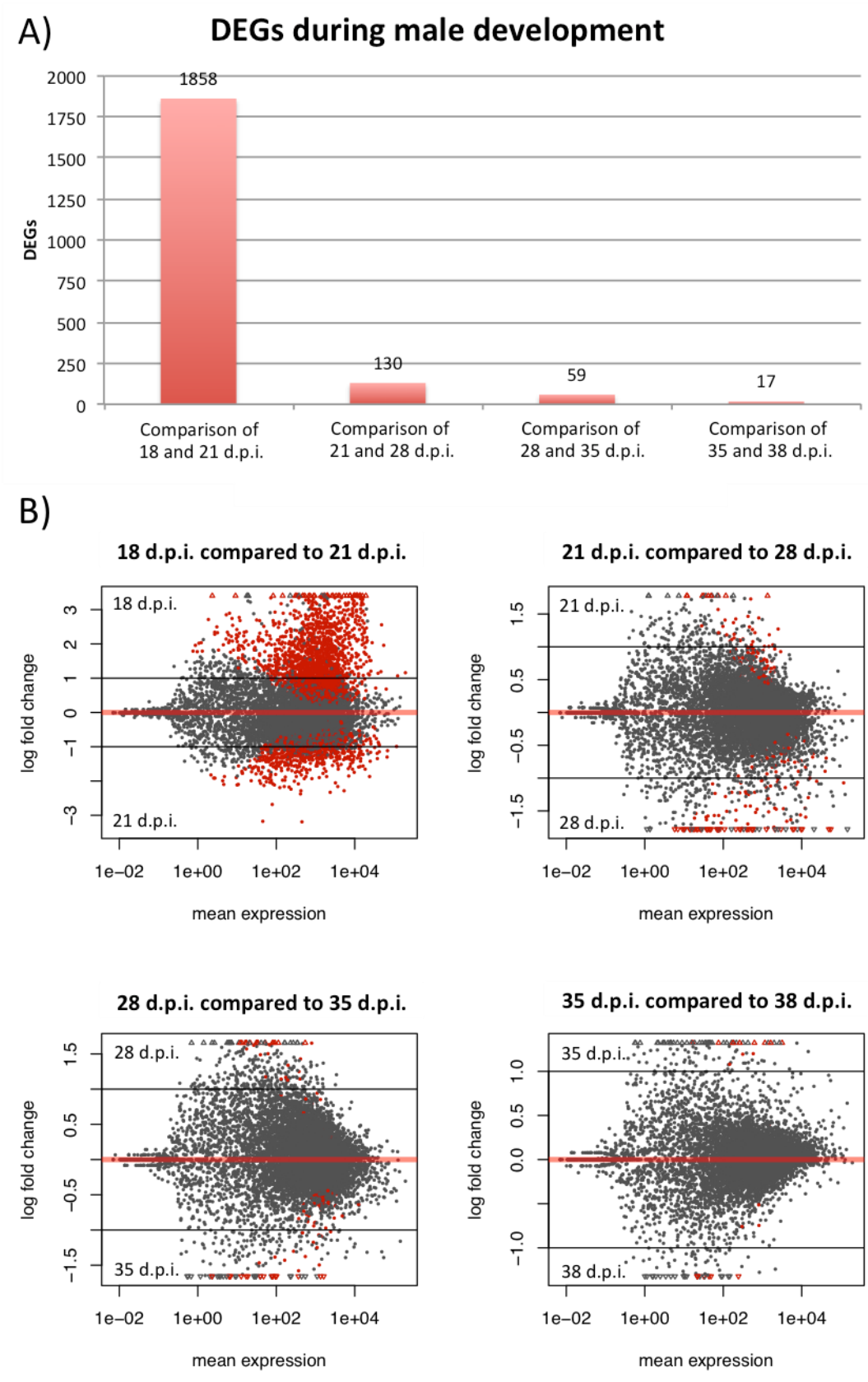

A)
DEGs during male development
DEGs
2000
1750
1500
1250
1000
750
500
250
0
1858
130
59
17
Comparison of 18 and 21 d.p.i.
Comparison of 21 and 28 d.p.i.
Comparison of 28 and 35 d.p.i.
Comparison of 35 and 38 d.p.i.
B)
18 d.p.i. compared to 21 d.p.i.
18 d.p.i.
21 d.p.i.
log fold change
mean expression
21 d.p.i. compared to 28 d.p.i.
21 d.p.i.
28 d.p.i.
log fold change
mean expression
28 d.p.i. compared to 35 d.p.i.
28 d.p.i.
35 d.p.i.
log fold change
mean expression
35 d.p.i. compared to 38 d.p.i.
35 d.p.i.
38 d.p.i.
log fold change
mean expression
1e−02
1e+00
1e+02
1e+04

Figure 3.4: The largest number of differentially expressed genes was found between 18 and 21 days *post* infection (d.p.i.) in developing male worms. A) This plot provides the number of differentially expressed genes found by pair-wise comparisons of males from mixed sex infections of different time points. B) Log ratio vs mean (MA) plots showing the mean expression and log-fold change of each pair-wise comparison. Genes marked as red dots have been assigned an adjusted p-value of < 0.01, the other are grey. Black lines represent the fold change cut-off (log-fold change = ±1). Triangles indicate genes outside the boundaries of the plots; like other genes, they are red if they are statistically significant, grey if they are not.

Cytochrome P450-dependent drug metabolism

Praziquantel is currently the main drug to treat schistosomiasis, especially as part of mass treatment programs in endemic areas (Crellen *et al.*, 2016). However, immature worms are known to be less sensitive to praziquantel, making it difficult to clear an infection completely and stop the parasite life cycle in endemic areas (Sabah *et al.*, 1986; Vimieiro *et al.*, 2013). At 18 d.p.i., three out of four genes belonging to the cytochrome P450 dependent drug metabolism pathway were found to be up-regulated in male worms compared to male worms at 21 d.p.i. (Appendix A.7). The relatively high expression of genes involved in cytochrome P450-dependent drug metabolism at 18 d.p.i. may be the reason why immature worms are less susceptible to praziquantel. A Glutathione S transferase (GST) was not only up-regulated in males at 18 d.p.i. compared to males at 21 d.p.i. (Smp_024010, 5.1-fold change, adjusted p-value = 4.50E-05) (Appendix A.8), but also compared to males at 38 d.p.i. (21.9-fold change, adjusted p-value = 1.34E-14) (Appendix A.8). The same gene was also up-regulated in female worms at 18 compared to 38 d.p.i. (18.2-fold change, adjusted p-value = 1.30E-11) (Appendix A.8).

Regulation of cell division in the testes

Compared to 21 d.p.i., the gene coding for the cell cycle regulator Cdc25 (Smp_152200, 6.0-fold change, adjusted p-value = 0.00061) was up-regulated at 28 d.p.i. in male worms, as was follistatin (Smp_123300, 2.3-fold change, adjusted p-value = 0.00027) (Appendix A.9). Follistatin is also associated with the proliferation of spermatocytes and was also found to be down-regulated from 28 to 35 d.p.i. (2.4-fold change, adjusted p-value = 0.00082) (Appendix A.10). Leutner *et al.* (2013) demonstrated that paired male worms express lower levels of follistatin than unpaired males. The down-regulation of the gene encoding follistatin in MS males at 35 d.p.i. compared to those at 28 d.p.i. is probably related to this observation (Appendix A.10). From 21 to 28 d.p.i., a gene encoding the homologue of basonuclin 2 (Smp_138350, 5.4-fold change, adjusted p-value = 1.86E-05) was up-regulated in male worms from mixed sex infections (Appendix A.16).

Aromatic-L-amino acid decarboxylase

Wang *et al.* (2017) hypothesised that the AADC gene, which they found to be expressed in *S. japonicum* males during mating (see Chapter 1.6.2), was responsible for inducing female sexual maturity. The AADC gene was also found to be up-regulated in male *S. mansoni*. The *S. mansoni* gene (Smp_135230) was significantly up-regulated in male worms at 38 d.p.i. compared to males at 18 d.p.i. (8.50-fold change, adjusted p-value = 2.26E-06) (Appendix A.11). Additionally, it was found to be significantly up-regulated in male worms from MS infections compared to males in SS infections at 35 d.p.i. (8.39-fold change, adjusted p-value = 1.23E-03) and 38 d.p.i. (5.33-fold change, adjusted p-value =

1.99E-03), falling just short of the significance threshold (adjusted p-value < 0.01) at 28 d.p.i. (2.32-fold change, adjusted p-value = 0.0359) (Appendix A.11). The lack of expression in males from SS infections as well as the confirmation of expression in mature MS males (here in *S. mansoni*) adds further weight to the hypothesis of AADC being involved the stimulation of female fertility in schistosomes.

3.2.5 Female development

Pairing is thought to commence between three and four weeks *post* infection, allowing female worms to reach sexual maturity around five weeks *post* infection (Biolchini *et al*. 2006). Accordingly, the greatest difference in gene expression in MS female worms was found between 28 and 35 d.p.i. (1545 DEGs) (Figure 3.5). By comparison, there were only 300 DEGs between 18 and 21 d.p.i., 385 between 21 and 28 d.p.i. and 29 between 35 and 38 d.p.i. (Figure 3.5).

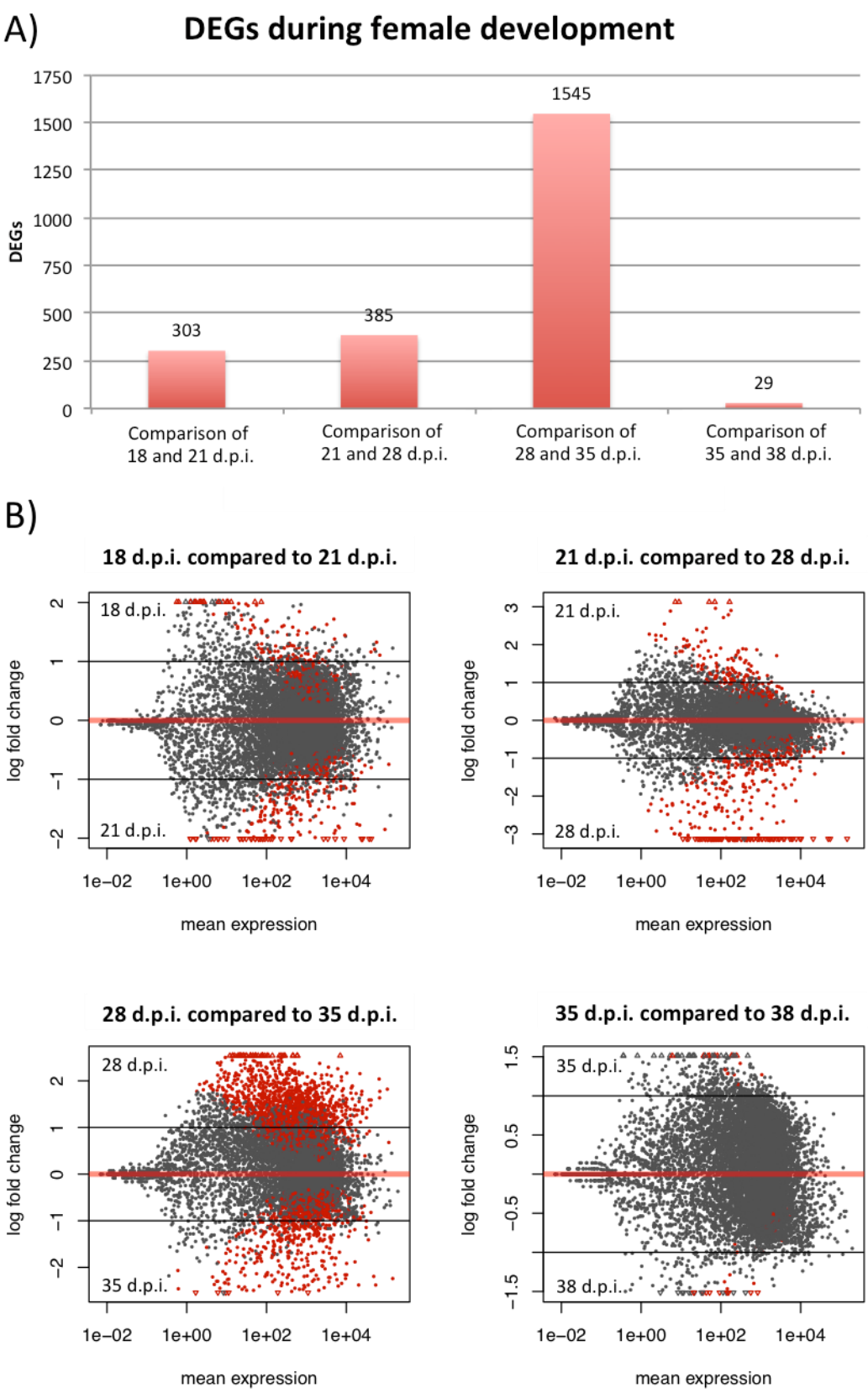

A)
DEGs during female development
DEGs
1750
1500
1250
1000
750
500
250
0
303
385
1545
29
Comparison of 18 and 21 d.p.i.
Comparison of 21 and 28 d.p.i.
Comparison of 28 and 35 d.p.i.
Comparison of 35 and 38 d.p.i.
B)
18 d.p.i. compared to 21 d.p.i.
21 d.p.i. compared to 28 d.p.i.
28 d.p.i. compared to 35 d.p.i.
35 d.p.i. compared to 38 d.p.i.
log fold change
mean expression
1e−02
1e+00
1e+02
1e+04
18 d.p.i.
21 d.p.i.
28 d.p.i.
35 d.p.i.
38 d.p.i.

Figure 3.5: The largest number of differentially expressed genes (DEGs) was found between 28 and 35 days *post* (d.p.i.) infection in female worms from mixed sex infections. A) DEGs found by pair-wise comparisons of females from mixed sex infections at different time points. B) Log ratio vs mean (MA) plots show the mean expression and log-fold change of each pair-wise comparison. Genes marked as red dots have been assigned an adjusted p-value of < 0.01, the others are grey. Black lines represent the fold change cut-off (log-fold change = ±1). Triangles indicate genes outside the boundaries of the plots; like other genes, they are red if they are statistically significant, grey if they are not.

Egg production related genes

During the maturation of paired female schistosomes, many fertility related genes are known to become up-regulated (Fitzpatrick & Hoffmann, 2006). This is largely due to the growth and maturation of the reproductive organs including the ovaries and the vitellarium (Fitzpatrick & Hoffmann, 2006). Many of the significantly up-regulated genes were also found in the comparison of male and female worms (Chapter 3.2.3). In this comparison of females from different time points, the most notable up-regulation of fertility-related genes occurred between days 21 and 28, where 266 DEGs were found to become up-regulated (Figure 3.5). The DEGs included all genes encoding trematode egg shell synthesis domains, a tyrosinase (Smp_050270), and the extracellular superoxide dismutase (Smp_095980) (Appendix A.12). Furthermore, two tetraspanins were up-regulated, CD63 antigen (Smp_173150) and a CD63 receptor (Smp_155310) (Appendix A.12). Although CD63 antigen is expressed in mature females, its function remains unclear (Cogswell *et al.*, 2012). At 38 d.p.i., females from MS infections expressed the CD63 receptor and antigen, at 191-fold (adjusted p-value = 5.98E-30) and 39-fold (adjusted p-value = 1.62E-42) higher levels

respectively, than females from SS infections (Appendix A.13). This made them the 21st and 51st of the most differentially expressed genes, the majority of which are thought to be expressed in the vitellarian tissue. Together the DEGs summarised in this paragraph confirmed that female schistosomes from MS infections have not only begun to pair by 28 d.p.i. but also that females have started to develop their reproductive organs at that time. As others recently observed (Biolchini *et al.*, 2006), it appears that pairing takes pace somewhat earlier than reported by Basch (1991). These genes became further up-regulated at 35 d.p.i. and 38 d.p.i. and were consistently expressed at significantly higher levels in females from MS infections rather than SS infections (see Chapter 3.2.8).

Ovary-specific gene expression

While the most differentially expressed genes up-regulated in mature females are thought to be largely expressed in the vitellaria [for example Fitzpatrick *et al.* (2007), Fitzpatrick & Hoffmann (2006), Köster *et al.* (1988)], at 21 d.p.i. (compared to 18 d.p.i.), a least two of the up-regulated genes probably play an important role in the ovaries of mature females. The first encoded a homologue of human sperm associated antigen (Smp_001160, 2.2-fold change, adjusted p-value = 2.08E-05) (Appendix A.14), which plays a role in fertility in mammals (Lin *et al.*, 2001). The other encoded *boule* (Smp_144860, 6.6-fold change, adjusted p-value = 6.34E-05) (Appendix A.14), a conserved transcription factor found in most animals. Amongst others, it has been shown to regulate male and female gametogenesis in the flatworm *Macrostomum ligano* (Kuales *et al.*, 2011). In *S. mansoni*, *boule* was also expressed specifically in the gonads (Chapter 5.2).

WNT signalling

At 18 d.p.i., females express two *wnt* family genes (Smp_196930, 2.5-fold change, adjusted p-value = 0.0027; Smp_167140, 3.1-fold change, adjusted p-value = 0.0067) that became down-regulated at 21 d.p.i. (Appendix A.15) Additionally, a gene (Smp_062560, 2.2-fold change, adjusted p-value = 0.0067) homologous to a *Secreted frizzled-related protein* 2, a regulator of the Wnt pathway, was up-regulated (Appendix A.15) Frizzled proteins are G-protein coupled receptors that regulate the Wnt signalling pathway (Nusse, 2009). Up-regulation of these three genes suggests, that the Wnt signalling pathway is more active during the earlier parts of their development, where it could regulate the formation of a body axis, cell fate and cell proliferation.

Differential expression of transcription factors

Across this RNA-Seq time course, many transcription factors were found to be differentially regulated. In particular a large number of genes coding for proteins regulating transcription became down-regulated at in females worms from MS infections at 28 d.p.i. when compared to 21 d.p.i. (Appendix A.17). Based on their GO term annotations, 37 transcription factor genes were identified: about a third of all the DEGs down-regulated at 28 d.p.i. (Appendix A.17). Such large-scale down-regulation of transcription factors in females from MS infections could suggest that some development processes reach their conclusion at this time. Of the 37 transcription factors, 22 were also expressed at significantly lower levels in females from SS infections at 28 d.p.i. and none were up-regulated (Appendix A.18).

Conversely, as the worms started to pair and females began to initiate the proliferation of their reproductive tissues, especially ovaries and vitellaria, another set of transcription factors was significantly up-regulated (Appendix A.19). Among the 100 most differentially expressed genes at 38 d.p.i. three genes with zinc finger domains were identified (Smp_166560, Smp_087320 and Smp_095350). All three zinc finger genes were found to be up-regulated more than 10-fold (adjusted p-value < 9.00E-22) (Appendix A.19) and it is tempting to speculate that they play a role in regulating transcription of genes involved in female maturation.

3.2.6 Male time series analysis

The previous analysis of pair-wise comparisons made it apparent that differential expression of genes can be found across the time course, with more differential expression found between MS and SS females than males. Next, I wanted to define sets of genes regulated in a pairing dependent manner throughout the course of infection. To do so, I performed a likelihood ratio test to identify genes that were up- or down-regulated over the course of infection in worms from SS or MS infections but not in both infections. Based on this test, 39 genes (Table 3.6) were identified as differentially expressed in males from MS and SS infections across the time course (adjusted p-value < 0.01) in the otherwise very similar male worms (Figure 3.6). These included several genes coding for proteins involved in signalling and signal transduction processes, such as the fibroblast growth factor (FGF) receptor activating protein (Smp_035730, adjusted p-value = 0.0053) and the epidermal growth factor (EGF) receptor (Smp_152680, adjusted p-value = 0.0069), which were found to be expressed at

higher levels in paired males (Table 3.6). These genes are shown to be expressed in the testes in Chapter 5 of this thesis. The FGF and the EGF signalling pathways regulate cell proliferation and differentiation (Eswarakumar *et al.*, 2005; Herbst, 2004). This suggests that paired males have a higher rate of cell proliferation in their testes, which could result in a greater number of spermatocytes being produced.

Gene ID	Product	MS / SS	Fold change 38 d.p.i.	Average read count	Adjusted p-value
Smp_005880	Phosphoenolpyruvate carboxykinase	MS	1.60	3571.83	2.86E-03
Smp_012000	Nicotinic acetylcholine receptor	SS	2.77	129.19	1.10E-05
Smp_012380	Voltage-gated potassium channel subunit β 2	SS	1.59	1174.94	6.87E-03
Smp_013950	Solute carrier family 43	MS	1.66	1119.06	2.03E-06
Smp_015020	Sodium potassium transporting ATPase α subunit	MS	1.71	2041.50	5.04E-03
Smp_019980	Vacuole membrane protein 1	MS	1.19	2745.53	9.17E-04
Smp_020840	Uncharacterised protein	MS	1.74	68.18	4.84E-03
Smp_021960	Uncharacterised protein	MS	1.02	318.04	1.10E-05
Smp_028800	Uncharacterised protein	SS	2.23	39.30	2.49E-04
Smp_035730	FGF receptor activating protein	MS	1.06	934.37	5.33E-03
Smp_044800	Ankyrin repeat and E domain containing protein	MS	1.15	327.46	8.81E-04
Smp_063560	Transmembrane protein 120b	MS	1.22	1080.30	5.33E-03
Smp_114430	Uncharacterised protein	MS	4.63	53.16	5.53E-03
Smp_125210	Aquaporin 9 (Small solute channel 1)	MS	1.49	119.40	6.47E-03
Smp_126290	Uncharacterised protein	SS	2.21	283.56	1.71E-03
Smp_128010	Uncharacterised protein	MS	4.10	146.98	1.10E-05
Smp_134990	Uncharacterised protein	MS	3.01	112.52	2.55E-08
Smp_135020	Putative oxalate:formate antiporter	MS	1.80	249.99	4.84E-03
Smp_135230	Tyrosine Decarboxylase	MS	3.15	440.13	1.27E-05
Smp_147070	Sodium-coupled neutral amino acid	MS	2.34	1640.36	4.29E-03
Smp_151490	Vertebrate nebulin	MS	1.87	3474.33	6.31E-03
Smp_152680	Epidermal growth factor receptor	MS	1.75	159.89	6.90E-03
Smp_152730	Histone-lysine N-methyltransferase	MS	1.83	39.41	8.32E-03
Smp_158480	AMP binding enzyme	MS	16.87	1332.34	2.12E-09
Smp_158720	SCY1-like protein kinase	SS	1.53	484.09	1.10E-05
Smp_160040	Uncharacterised protein	MS	3.44	15.31	3.10E-03
Smp_167160	Dynamin binding protein	MS	1.51	2472.35	6.90E-03
Smp_168670	cGMP dependent protein kinase 1	MS	1.16	381.69	5.53E-03
Smp_168940	Uncharacterised protein	MS	4.09	293.32	2.72E-03
Smp_169150	Short transient receptor potential channel 3	SS	2.00	144.97	9.86E-03

Smp_172960	Uncharacterised protein	MS	1.36	2005.73	1.24E-03
Smp_173290	Uncharacterised protein	MS	2.22	23.87	1.10E-05
Smp_173300	Uncharacterised protein	MS	2.35	16.12	1.14E-05
Smp_174870	Uncharacterised protein	MS	2.00	487.85	6.75E-08
Smp_176400	Uncharacterised protein	MS	6.56	29.18	7.16E-03
Smp_190210	Uncharacterised protein	SS	5.51	5.27	9.96E-04
Smp_194190	Endothelin-converting enzyme 2	MS	2.47	109.63	4.51E-03
Smp_212750	Uncharacterised protein	MS	2.43	75.40	2.64E-04
Smp_212760	Putative kinesin	SS	2.03	909.99	7.79E-04

Table 3.6: Significant genes found in the time series analysis of male worms. The "MS/SS" column reports whether expression was up-regulated in males from mixed sex or single sex infections at 38 days *post* infection (d.p.i.). The fold change is provided for 38 d.p.i. The average read count is a measure of absolute gene expression. The *p*-value has been adjusted for multiple hypothesis testing.

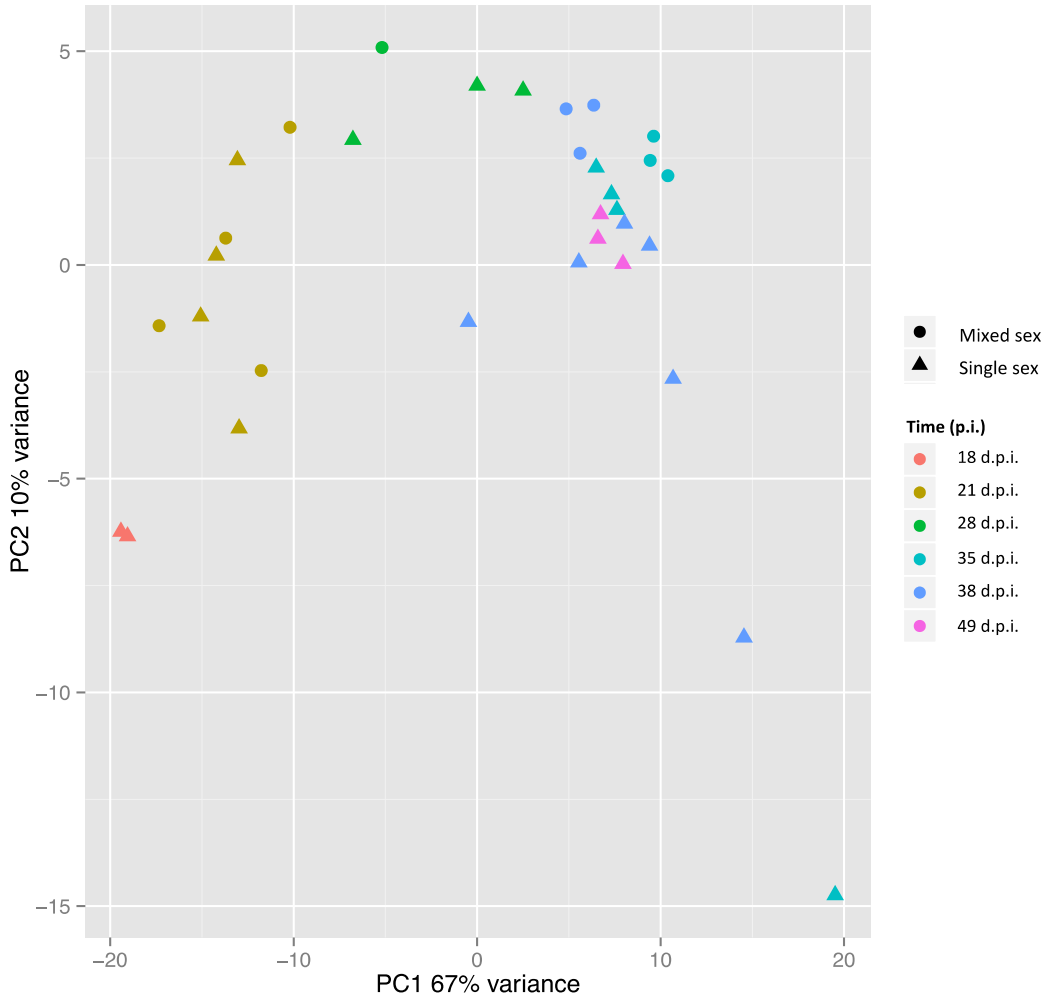


Figure 3.6: Male worms from mixed sex and single sex infections are very similar at the transcriptome levels throughout their development. Principal component analysis plot of male RNA-Seq samples. The first two principal components show 77% of the measured variance. The PC1 axis seems to broadly correlate with worm maturity. Time points indicate the number of days *post* infection (d.p.i.). Several outliers can be observed, especially at the late time points.

3.2.7 Female time series analysis

As for the male samples, a likelihood ratio test was used to identify pairing-controlled genes in females and 5173 such genes were discovered (adjusted p-value < 0.01). At 38 d.p.i., 2670 of the genes were up-regulated in MS females and 2503 up-regulated in SS females. The PCA plot (Figure 3.7) shows female worms from MS and SS infection forming separate clusters from 28 d.p.i. onwards. To analyse such a large number of genes, two different clustering methods were used to identify groups of genes with similar expression patterns.

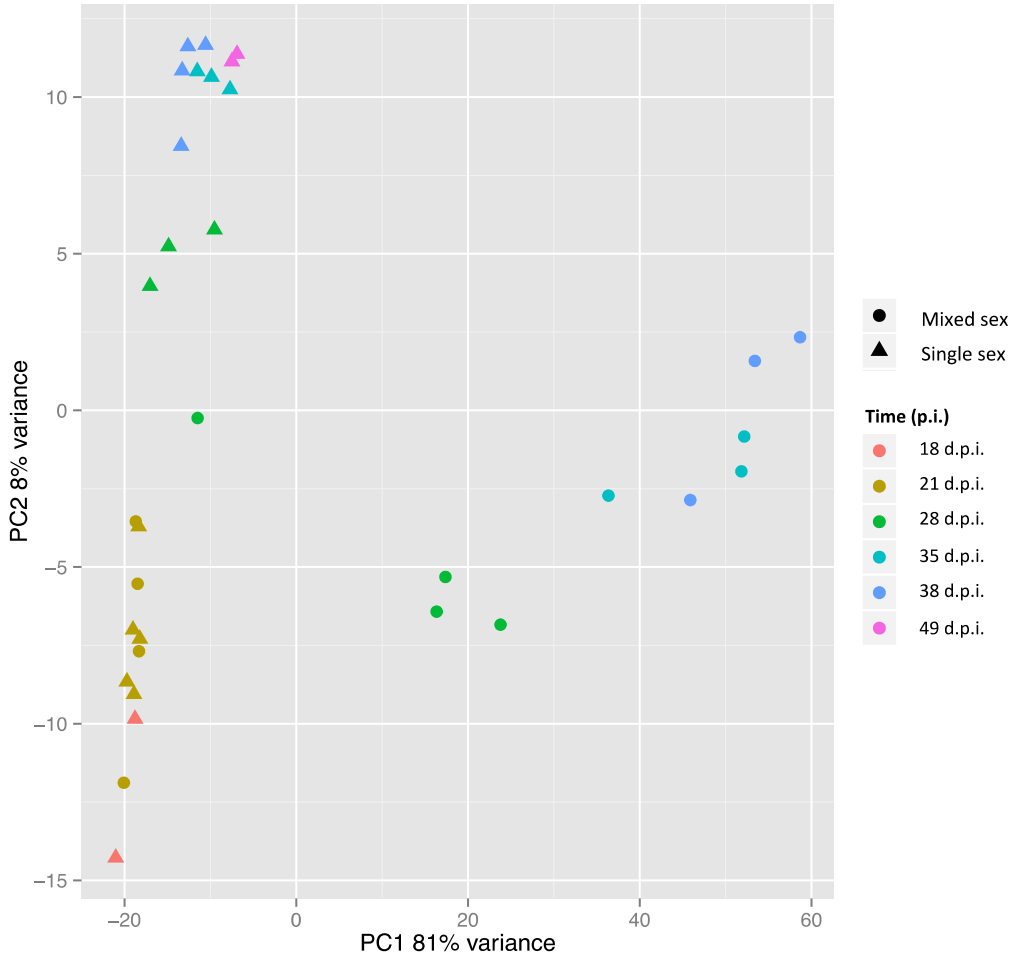


Figure 3.7: Female from mixed sex infections diverge from the single sex samples from 28 d.p.i. In this PCA plot of female RNA-Seq samples, the first two principal components capture 89% of the measured variance. Whereas paired females develop along PC1 over time, with sexual maturation contributing to 81% of the total variation, single females change primarily along PC2 (8%). Time points indicate the number of days *post* infection (d.p.i.).

3.2.8 Fertility-related genes

Having identified over 5000 pairing-regulated genes involved in females, I tried two different clustering methods, the Kohonen method (Wehrens & Buydens, 2007) and MBCluster (Si *et al.*, 2014) (Chapter 2.4.11), to split the genes into groups with similar patterns of expression. This was done to ensure the reproducibility of the clustering results and to define a set of putative fertility-related genes. Similar patterns of expression across several samples, especially

in time course experiments, can be evidence for co-expression and suggest roles in the same biological processes (Si *et al.*, 2014). Kohonen is based on a heuristic method designed to produce self-organising maps from data sets using unsupervised learning. Self-organising maps are a two-dimensional representation of multi-dimensional data that aim to place similar data points, e.g. genes close to one another thereby forming clusters of similar genes (Wehrens & Buydens, 2007). In contrast, MBCluster assumes the data to have a negative binomial distribution and uses an Expectation-Maximisation algorithm to estimate model parameters and cluster membership (Si *et al.*, 2014).

For both methods, genes were divided into 16 clusters (Figures 3.8 and 3.9) as higher numbers of clusters usually resulted only in an increasing number of clusters with very small number of genes, even as low as one or two genes (data not shown). In the MBCluster results, one cluster, containing 76 genes, was characterised by the largest overall fold changes and high expression levels in MS females at 28, 35 and 38 d.p.i. (see Figure 3.8). This group of genes included two of the clusters formed by the Kohonen method, which separated the genes further based on their relative expression (see Figure 3.9). Of the two Kohonen clusters, one (Cluster 13) contained genes with particularly high expression in mature females. The 25 genes in that cluster are among the 32 most up-regulated genes in paired females ranging from 540-fold to over 45350-fold up-regulation over the time course (Appendix A.20). Together, these genes account for over 14% of the sequencing reads in the transcriptome of mature females, compared to only 0.0065% of reads in immature females (Appendix A.21). Eleven of the genes were found by InterProScan to code for a trematode eggshell

synthesis domain (Appendix A.20). Also included were genes coding for two tyrosinases (Smp_013540 & Smp_050270) and an extracellular superoxide dismutase (Smp_095980) (Appendix A.20) all of which are known pairing-regulated genes (Fitzpatrick & Hoffmann, 2006).

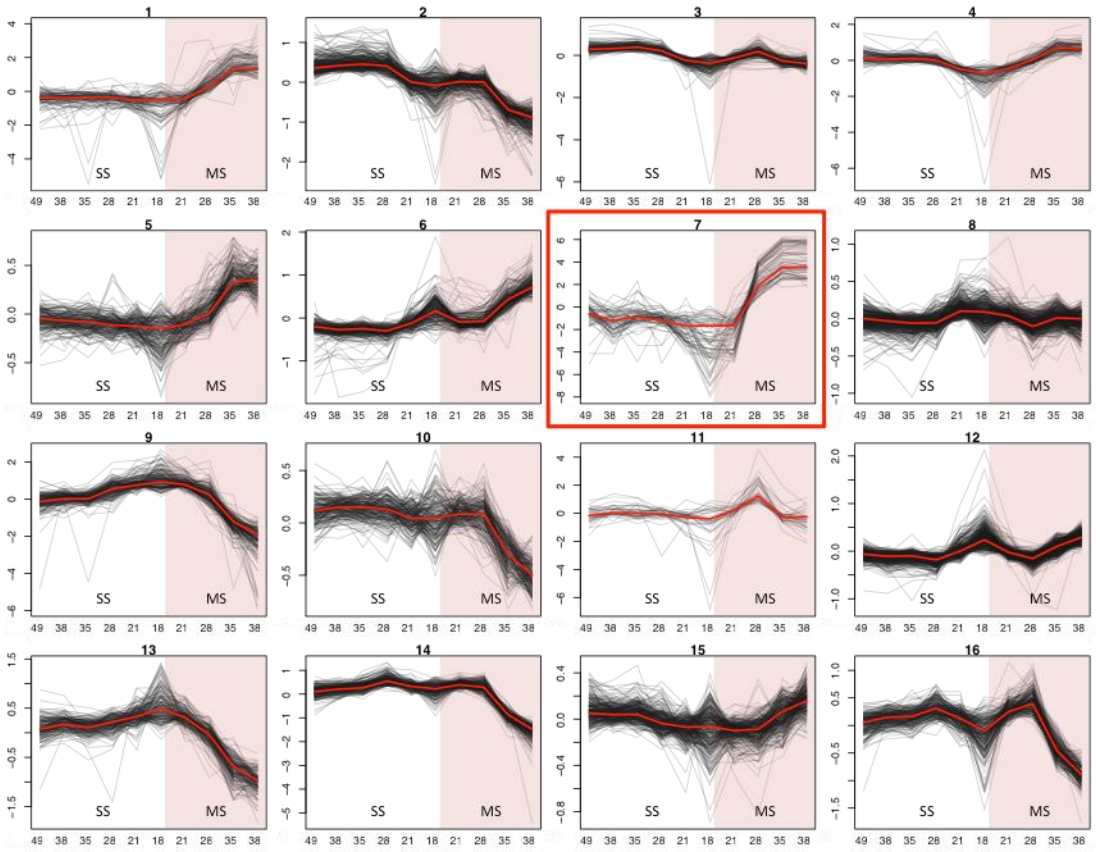


Figure 3.8: Genes were divided into 16 clusters based on expression profiles. The MBCluster plots show gene expression across all female RNA-Seq samples. For each gene in a cluster the log2 fold changes (Y-axis) are plotted against the days *post* infection (X-axis). Samples from single sex (SS) infections have a white background; samples from mixed sex (MS) infections have a red background. The red lines represent the cluster average for each time point. The cluster of fertility-related genes discussed above is contained within a red box.

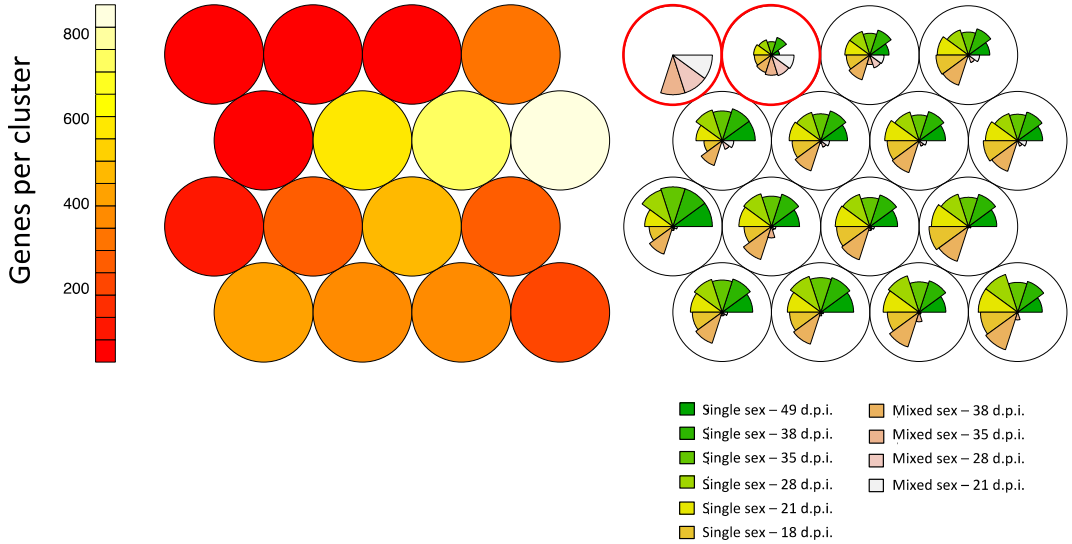


Figure 3.9: Genes were divided into 16 clusters based on their expression using the Kohonen method. Left – heatmap of number of genes per cluster. Right – Average $\log_2$-expression of genes inside each cluster for all time points and for mixed and single sex females. Cluster 13 is on the top left in both maps. The two clusters of fertility-related genes discussed in Chapter 3.2.8, Cluster 13 and 14, are contained in red circles.

To select gene clusters that exhibit significant changes in expression over the time course, genes were filtered based on their average fold changes relative to expression in female worms at 18 d.p.i. The expression levels of seven out of the 16 clusters (1, 4, 7, 9, 11, 13 and 14) identified by the MBCluster method at least doubled or halved over the time course (see Figure 3.10). It showed that the fold changes in gene expression were larger in females from MS than SS infections. A GO term analysis was performed on the sets of genes contained in each of these 7 clusters, identifying a total of 72 significantly enriched unique GO terms.

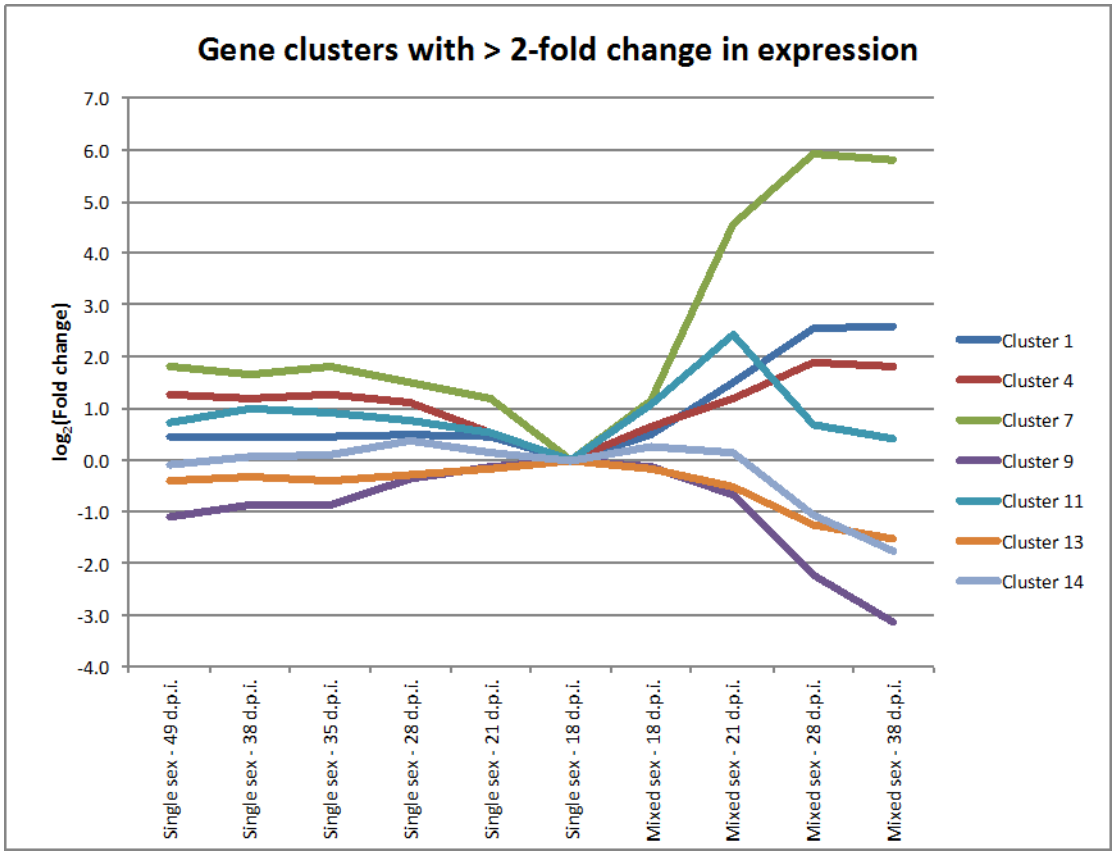


Figure 3.10: The magnitude of changes in gene expression is larger in females from mixed sex than females from single sex infections. Average expression of genes in the MBCluster clusters with greater than 2-fold changes in gene expression across the female samples of the time course. All expression fold changes are measured relative to females from single sex infections at 18 days *post* infection (d.p.i.).

The clusters 9, 13 and 14 were found to be enriched for genes annotated with developmental GO terms such as multicellular organismal development, Wnt signalling and sex determination (Tables 3.7, 3.8 and 3.9). From 18 d.p.i. onwards expression of these genes fell in females from MS as well as SS infections (Figure 3.10). Notably, the down-regulation of these genes was more pronounced in females from MS infections than those from SS infections (Figure 3.10).

GO term	Description	Annotated	Significant	Expected	p-value
GO:0007156	Homophilic cell adhesion	60	22	1.74	2.00E-19
GO:0007275	Multicellular organismal development	286	35	8.3	3.20E-13
GO:0006355	Regulation of transcription	800	57	23.22	9.80E-12
GO:0016055	Wnt signalling pathway	27	7	0.78	0.00026
GO:0007155	Cell adhesion	226	35	6.56	0.00027
GO:0030154	Cell differentiation	78	9	2.26	0.00038
GO:0030238	Male sex determination	6	3	0.17	0.00045

Table 3.7: Significant GO terms for the genes in cluster 9 of the MBCluster analysis. "Annotated" - total number of genes in the genome annotated with a given GO term. "Significant" - the number of genes annotated with that GO term in Cluster 9. "Expected" – the number of genes with that GO term expected by chance. The significance cut-off was p = 0.01. The GO terms are all in the "Biological Process" category.

GO term	Description	Annotated	Significant	Expected	p-value
GO:0006813	Potassium ion transport	77	8	1.99	0.00076
GO:0006941	Striated muscle contraction	145	11	3.74	0.00122
GO:0045596	Negative regulation of cell differentiation	3	2	0.08	0.00195
GO:0030514	Negative regulation of BMP signalling	3	2	0.08	0.00195
GO:0051216	Cartilage development	5	2	0.13	0.00629
GO:0007223	Wnt signalling pathway (Ca modulating)	6	2	0.15	0.00928

Table 3.8: Significant GO terms for the genes in cluster 13 of the MBCluster analysis. "Annotated" - total number of genes in the genome annotated with a given GO term. "Significant" - the number of genes annotated with that GO term in Cluster 13. "Expected" – the number of genes with that GO term expected by chance. The significance cut-off was p = 0.01. The GO terms are all in the "Biological Process" category.

GO term	Description	Annotated	Significant	Expected	p-value
GO:0007186	G-protein coupled receptor signalling	161	30	8.46	2.50E-09
GO:0007155	Cell adhesion	226	44	11.87	4.90E-08
GO:0007156	Homophilic cell adhesion	60	14	3.15	1.70E-06
GO:0006811	Ion transport	389	52	20.44	5.60E-06
GO:0006813	Potassium ion transport	77	14	4.05	3.70E-05
GO:0006468	Protein phosphorylation	337	34	17.71	0.00013
GO:0007160	Cell-matrix adhesion	8	4	0.42	0.00044
GO:0007229	Integrin-mediated signalling pathway	9	4	0.47	0.00076
GO:0006816	Calcium ion transport	49	9	2.57	0.00086
GO:0007165	Signal transduction	547	67	28.74	0.00143
GO:0030245	Cellulose catabolic process	6	3	0.32	0.00255
GO:0007169	Transmembrane receptor protein tyrosine kinase signalling	31	6	1.63	0.00483
GO:0006865	Amino acid transport	22	5	1.16	0.00486
GO:0006814	Sodium ion transport	64	9	3.36	0.0058

Table 3.9: Significant GO terms for the genes in cluster 14 of the MBCluster analysis. "Annotated" - total number of genes in the genome annotated with a given GO term. "Significant" - the number of genes annotated with that GO term in Cluster 14. "Expected" – the number of genes with that GO term expected by chance. The significance cut-off was p = 0.01. The GO terms are all in the "Biological Process" category.

Expression of Clusters 1, 4 and 7 increased over time in both SS and MS females. In all three the effect was greater in females from MS infections than females from SS infections (Figure 3.10). At 38 d.p.i. average expression was around 6- and 3.5-fold higher than at 18 d.p.i., compared with a 1.4- and 2.4-fold increase in SS females for clusters 1 and 4 respectively (Figure 3.10). Cluster 1 contained several genes with a female-biased expression, including gene coding for proteins such as ferritin (Smp_087760) (Appendix A.22), whereas cluster 4 contained several genes encoding saposins, cathepsins and a haemoglobinase, which may be related to digestive processes in mature females (Appendix A.23).

Cluster 7 expression displayed the greatest up-regulation of gene expression amongst all clusters (Figure 3.10). In MS females, expression was on average 60-fold up-regulated at 38 d.p.i. relative to expression at 18 d.p.i. (Figure 3.10). This cluster contained many genes associated with female fertility, particularly egg production, such as genes encoding eggshell synthesis domain containing proteins (Smp_077900, Smp_000290, Smp_191910, Smp_000430 and Smp_077890), tyrosinases (Smp_050270 and Smp_013540), and a superoxide dismutase (Smp_095980) (Appendix A.24). Cluster 7 was found to be significantly enriched for several GO terms. These included tyrosine metabolic process (GO:0006570, p = 0.00014), phenol-containing compound biosynthetic process (GO:0046189, p = 0.00014) and pigment biosynthetic process (GO:0046148, p = 0.00128) due to the presence of the two tyrosinases in this cluster (Appendix A.25). Both tyrosinases are known to be expressed predominantly in sexually mature females and are thought to catalyse the oxidation of phenolic compounds contributing to egg shell formation (Fitzpatrick *et al.*, 2007). This process can produce free radicals and other oxidising agents (Tada, Kohno, & Niwano, 2014). Superoxide metabolic process (GO:0006801, p = 0.020) (Appendix A.25) was also overrepresented. Superoxide dismutases that are known to be expressed in female worms in a pairing dependant manner (Sun *et al.*, 2014), are also found in this cluster and may serve to protect the female reproductive tissues from oxidative stress. Other up-regulated genes included genes coding for a protein kinase as well as the tetraspanin CD63 receptor (Smp_155310) and a tetraspanin CD63 antigen (Smp_173150) (Appendix A.24).

The role of eggshell synthesis domain containing proteins as well as the tyrosinases in egg production has been demonstrated previously (Fitzpatrick *et al*., 2007). On the other hand, the role of the membrane receptors CD63 antigen (CD63a) and CD63 receptor (CD63R) is not clear and will be the focus of the rest of the chapter. CD63a, but not CD63R, has previously been shown to have a pairing-dependent pattern of expression (Cogswell *et al.*, 2012). As revealed by the time course data, both genes are not only female-specific but also respond to pairing in their expression profiles (Cluster 7 in Figure 3.10).

According to GeneDB (Logan-Klumpler *et al.*, 2012), CD63a is a predicted 24.1 kDa protein with 225 amino acids and CD63R is a predicted 30.6 kDa protein of 276 amino acids. However, tetraspanins often undergo post-translational modifications (Termini & Gillette, 2017), so their actual mass may be substantially higher. Tetraspanins are a group of proteins defined by their four transmembrane regions, which anchor the proteins in the cell membrane (Termini & Gillette, 2017). They generally protrude around 3-5 nm from the cell membrane (Grove, 2014) with two extracellular regions and three short intracellular loops (Termini & Gillette, 2017). Their roles in cell biology involve association with other proteins, including other tetraspanins, and thereby enabling the formation of signalling complexes (Levy & Shoham, 2005).

Next I used qRT-PCR to measure the difference in gene expression between males and females at different d.p.i. for CD63a and CD63R. This was done to confirm the findings of the RNA-Seq analysis. Then, to further investigate the role of CD63a and CD63R in female fertility, *in situ* hybridisations and RNAi

knockdown experiments were performed to determine in which tissues these genes are expressed and whether an effect of these tetraspanins on fertility could be demonstrated.

3.2.9 qRT-PCR analysis of CD63 antigen & CD63 receptor

Pairing-dependent expression of CD63R and CD63a in female worms was confirmed by qRT-PCR (Figure 3.11). Expression in female worms was measured relative to expression in male worms at 21 d.p.i. using PSMD4 (a proteasome 26S subunit) as a reference gene (Liu *et al.*, 2012). CD63R expression in females from MS infections was found to increase dramatically from 21 to 28 d.p.i. and remained high thereafter with expression being over 800-fold higher in females than males at 40 d.p.i. (Figure 3.11).

CD63a expression also increased significantly in females from MS infections from 21 to 28 d.p.i compared to males and continued to be expressed highly until 40 d.p.i. (Figure 3.12). At 21 d.p.i., females expressed CD63a at 1.25-fold of the male expression, which increased to 56.6-fold at 40 d.p.i., a 45-fold increase in female expression (Figure 3.12).

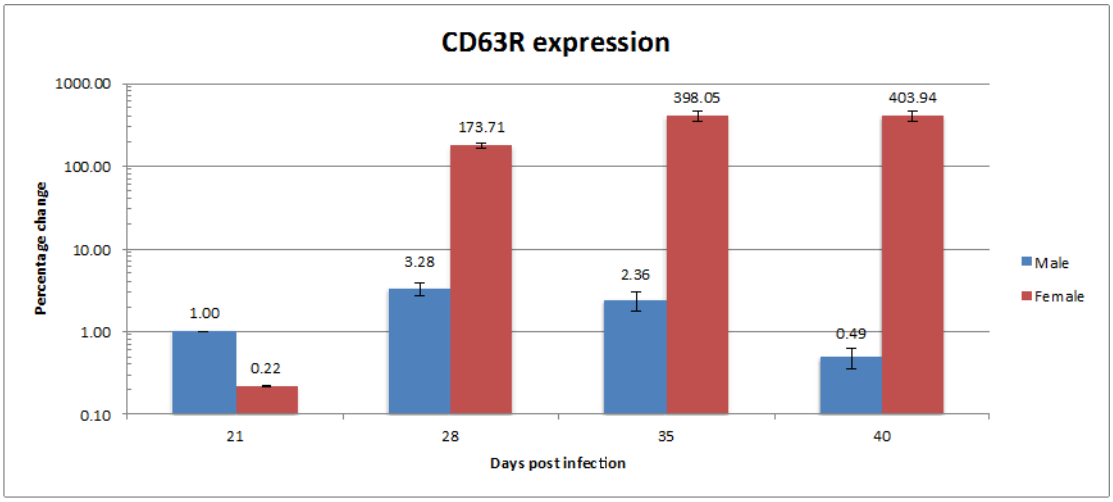


Figure 3.11: Expression of CD63R significantly higher in female worms from 28 d.p.i. onwards. Expression is give as a fold change relative to male expression at 21 d.p.i. PSMD4 was used as internal reference gene for this qRT-PCR analysis. Error bars are the standard error of the mean.

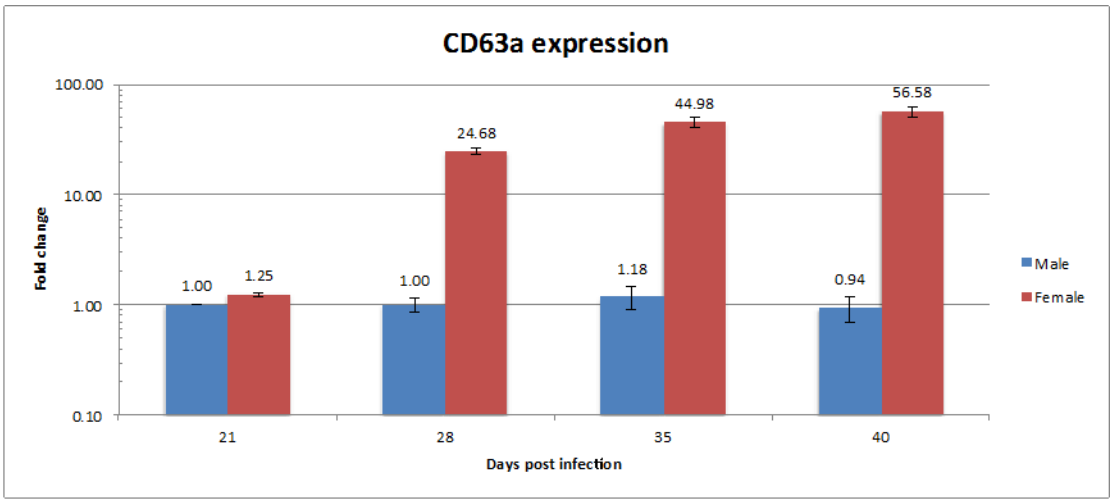


Figure 3.12: Expression of CD63a is higher in female than in male worms. Expression is given as a fold change relative to male expression at 21 d.p.i. PSMD4 was used as internal reference gene for this qRT-PCR analysis. Error bars are the standard error of the mean.

Both RNA-Seq and qPCR data show that there is a strong up-regulation of CD63R and CD63a expression in paired females, whereas expression in males (and, in the case of RNA-Seq data, females from SS infections) remains relatively constant at a low level (Figures 3.11 and 3.12; see Cluster 7 in Figure 3.10).

3.2.10 Whole mount *in situ* hybridisation

The cluster analysis of RNA-Seq data suggested that CD63a and CD63R could be expressed in the vitellaria of MS females and play a role in egg production. *In situ* hybridisation was used to identify the tissues in which expression of the two genes occurs. Tetraspanin-2 (TSP-2) was used as a positive control. TSP-2 is an abundant tegumental protein (Castro-Borges *et al.*, 2011) and its expression at the mRNA level can be demonstrated in the cell bodies of the tegument (see Figure 3.13). These cell bodies lie outside the light brown vitelline lobes but no staining was observed inside the vitelline tissue (see Figure 3.14 for a higher magnification).

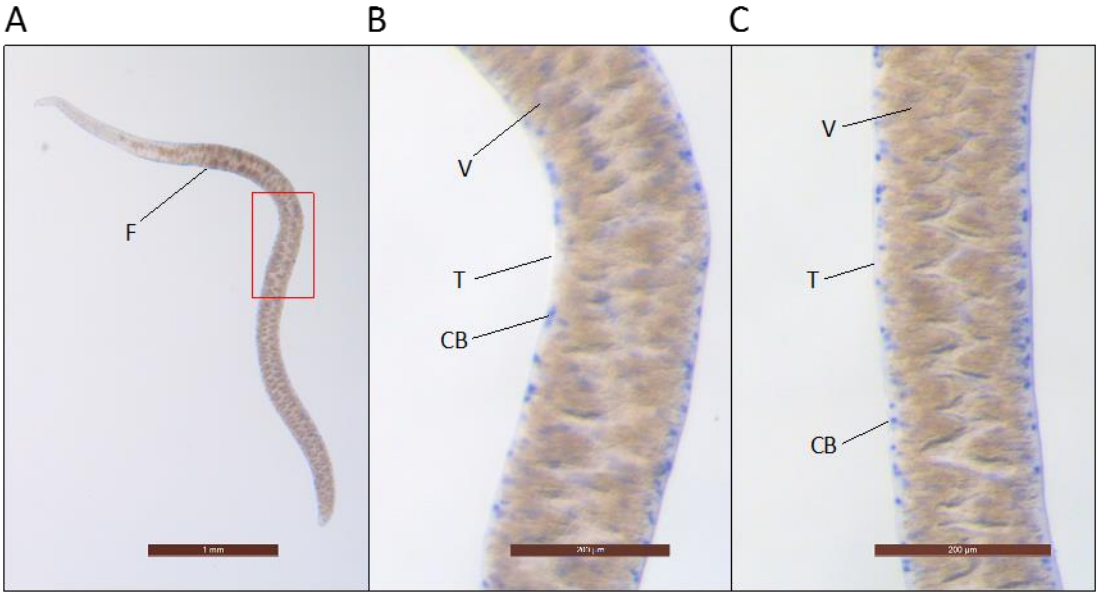


Figure 3.13: Tetraspanin-2 is expressed in the cell bodies of the worm tegument. Alkaline phosphatase staining (Chapter 2.27) of digoxigenin-labelled tetraspanin-2 mRNA in female worms. Expression was not observed in the vitelline tissue. F – Sexually mature female; V – Vitellarium; T – Tegument; CB – Cell body. A) Whole mature female with staining located in the cell bodies of the tegument. The red rectangle indicates where the close-up in B) is derived from. B) Close-up of the female specimen in A). Stained cell bodies can be seen in the tegument. C) Close-up of a different female specimen. Stained cell bodies can be seen in the tegument demonstrating tetraspanin-2 expression there.

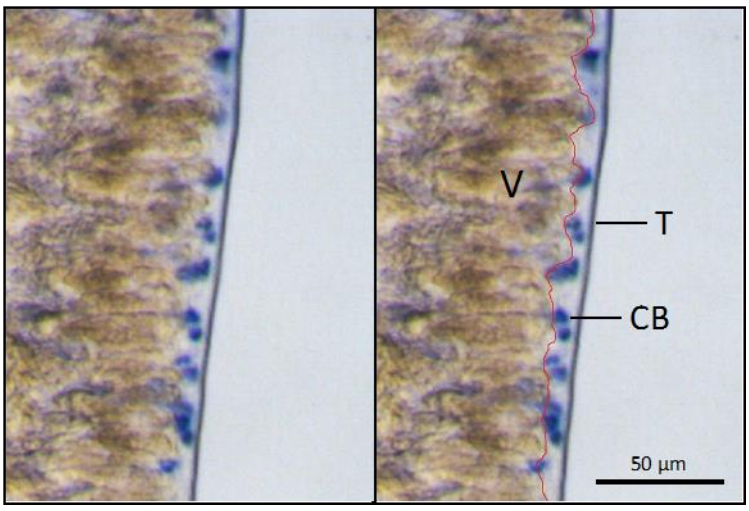


Figure 3.14: A close-up of the stained cell bodies indicates tetraspanin-2 is expressed in the cell bodies of the tegument. The original image is on the left; on the right the border of the vitellarium has been marked in red. V – vitellarium; T – tegument; CB – cell body.

In contrast to TSP-2, CD63a expression was detected inside the vitelline tissue of the 12 female specimens (see Figure 3.15). No expression was detected in the negative control (sense RNA) or in male worms (pictures not shown here).

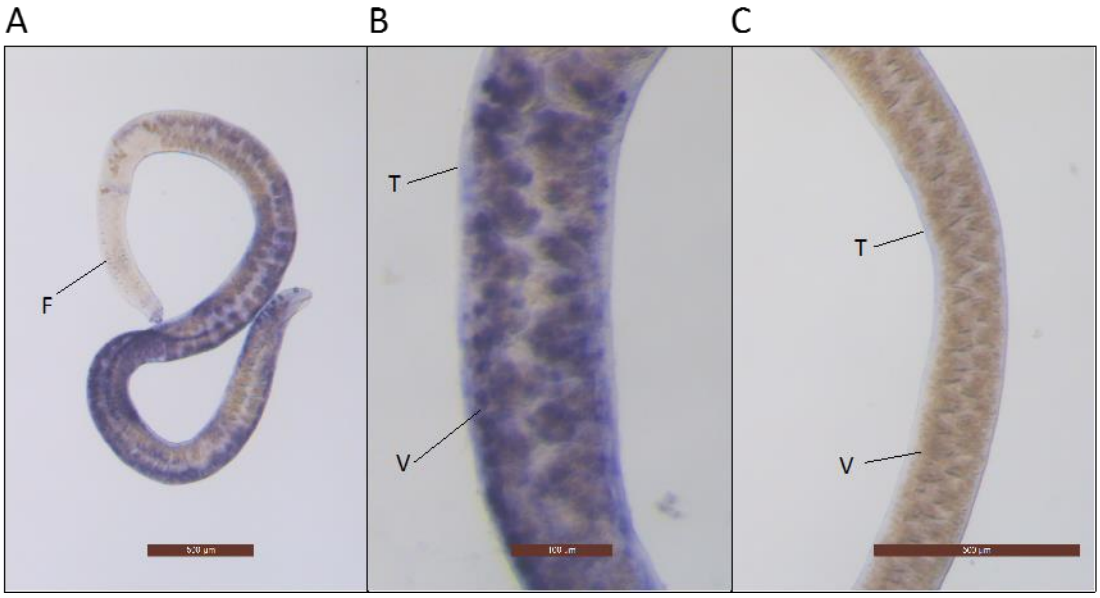


Figure 3.15: CD63 antigen is expressed in the vitelline tissue of female worms. Alkaline phosphatase staining (Chapter 2.27) of digoxigenin-labelled CD63 antigen mRNA in female worms. F – Sexually mature female; V – Vitellarium; T – Tegument. A) Whole female stained for CD63 antigen. B) Close-up female mid-section. C) No signal was found in the negative control.

The pattern of staining observed for the CD63R (see Figure 3.16) was similar to CD63a, showing expression in the vitellaria. Particularly, 14 female specimens displayed CD63R expression in the outer regions of the vitellaria, but little or no staining in the centre of the organ (Figure 3.15). No expression was observed in the female negative control or male worms (Figure 3.15). Erasmus (1975) described the organisation of vitellocytes. The author suggested that there is a gradient of maturity from the least mature (S1) cells at the periphery of the lobes, then S2 and S3 to the most mature (S4) cells in the centre of the female near the vitelline duct.

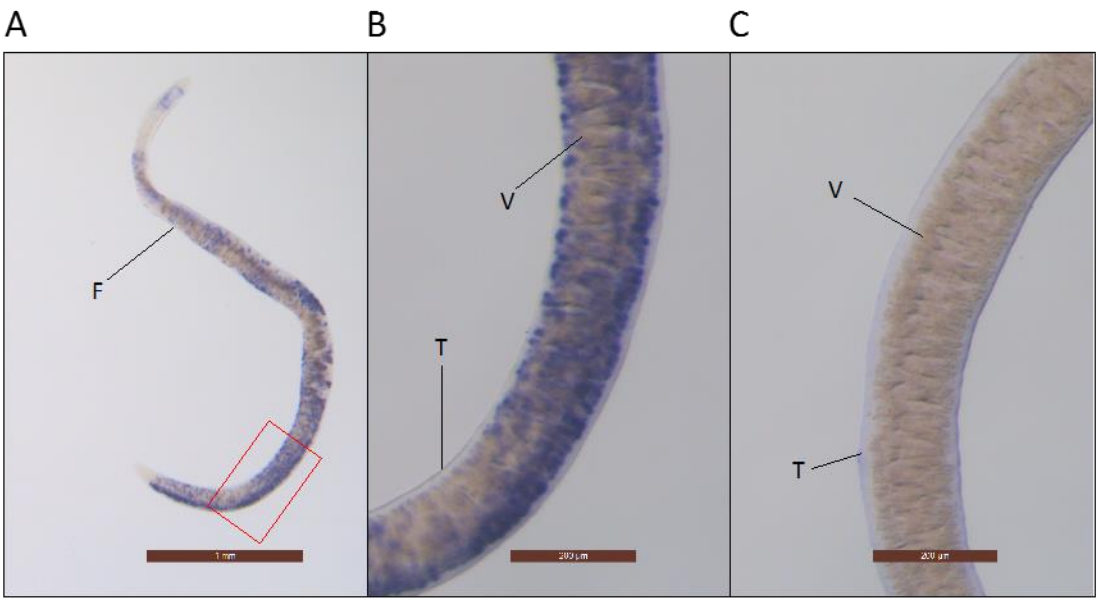


Figure 3.16: CD63 receptor is expressed in the vitelline tissue of female worms. Alkaline phosphatase staining (Chapter 2.27) of digoxigenin-labelled CD63 receptor mRNA in female worms. Staining was found to be most intense in the outer portions of the vitellaria. F – Sexually mature female; V – Vitellarium; T – Tegument. A) Whole female stained for CD63 receptor. A red rectangle indicates the section of the image that the close-up in B) was derived from. B) Close-up of female mid-section. C) No signal was found in the negative control.

The results of the whole mount *in situ* hybridisation confirm expression in the vitellaria of mature female worms, as suggested by the cluster analysis of RNA-Seq data of CD63a (Smp_173150) and CD63R (Smp_155310) (Appendix A.24). However, CD63R expression seemed to take place in the outer regions of the vitellaria (Figure 3.16), possibly due to expression in particular subpopulations of vitelline cells.

3.2.11 RNA interference

In order to determine if CD63a and CD63R play a significant role in the production of eggs, RNA interference (RNAi) was used (Chapter 2.2.2). Soaking in dsRNA was chosen, rather than electroporation, to maintain fertility as much

as possible and the number laid eggs was the measured phenotype. The positive control knockdown, TSP-2, was successful, with qRT-PCR showing a reduction of mRNA levels to 0.5% of those in the negative control (see Figure 3.17).

In contrast, CD63R expression was found to be higher in the treated sample compared to the negative control (see Figure 3.18).

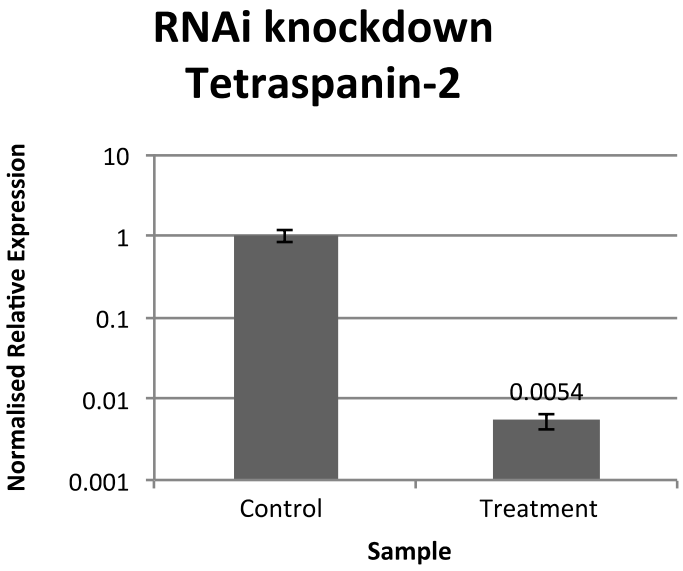


Figure 3.17: Tetraspanin-2 expression was reduced significantly to 0.5% of expression in negative control. Using the 'soaking method', an RNAi knockdown of tetraspanin-2 was performed in mature females. Knocked down tetraspanin-2 expression was measured relative to the negative control samples (soaked in non-specific dsRNA) using qRT-PCR. Two biological replicates were used for treatment and the control group. "Proteasome 26S Subunit, Non-ATPase 4" (PSMD4) was used as internal reference gene. The difference was found to be statistically significant ($p = 0.0083$). Error bars represent the standard error of the mean.

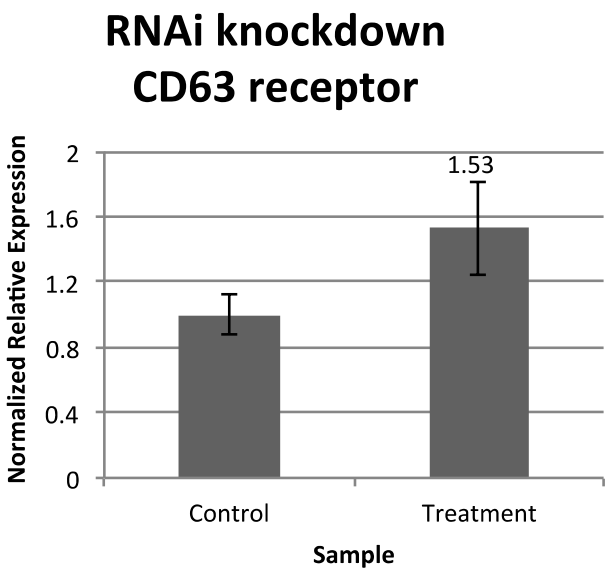


Figure 3.18: An increase in CD63R expression was observed relative to the negative control. RNAi knockdown in mature female worms, soaked in CD63a-specific dsRNA, was measured relative to negative control worms (soaked in non-specific dsRNA) using qRT-PCR. Two biological replicates were used for both treatment and control. "Proteasome 26S Subunit, Non-ATPase 4" (PSMD4) was used as internal reference gene. The difference was found to be statistically non-significant ($p = 0.086$). Error bars represent the standard error of the mean.

For CD63a, a moderate knockdown of around 35% was measured (see Figure 3.19). In negative controls, which did not undergo reverse transcription, amplification was only observed over six cycles after amplification in the samples suggesting that the DNase treatment successfully removed most potential gDNA contamination (Appendix A.25). In the no template control, amplification was only observed over 10 cycles after amplification in the sample showing that little non-specific amplification was taking place (Appendix A.25). TSP-2 is expressed in the tegumental cell bodies (Tran *et al.*, 2010), which may be more accessible to the dsRNA than the vitellarian tissue deeper inside the worm. This may explain why the TSP-2 knockdown was successful compared to the CD63a and CD63R knockdowns.

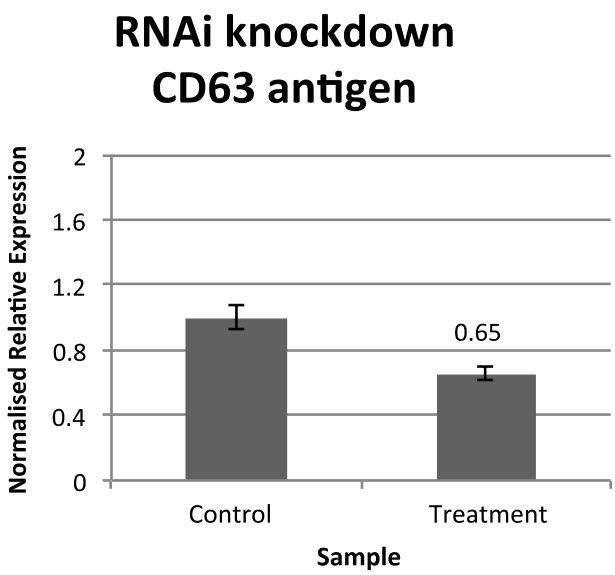


Figure 3.19: A moderate knock of CD63a expression of 35% was observed. RNAi knockdown of CD63R was measured in mature females relative to the negative control (worms soaked in non-specific dsRNA) using qRT-PCR. Two biological replicates were used for both treatment and control. "Proteasome 26S Subunit, Non-ATPase 4" (PSMD4) was used as internal reference gene. The difference was found to be just short of statistical significance ($p = 0.053$). Error bars represent the standard error of the mean.

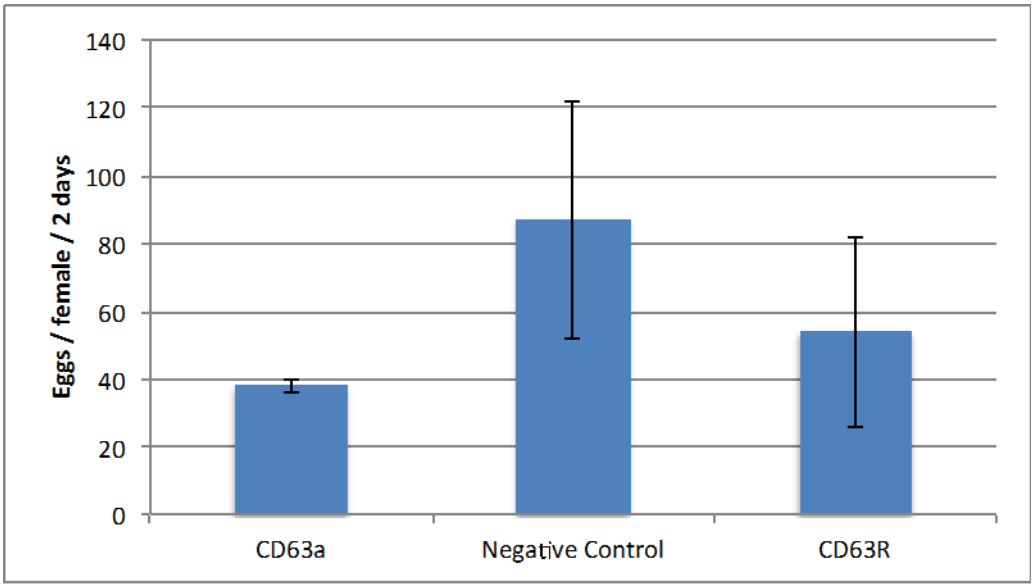


Figure 3.20: Reduction in laid eggs by the CD63a & CD63R knock-down worms was not statistically significant. An ANOVA showed that the treatment groups did not differ significantly from the negative control (p = 0.485). Egg production was measured as a phenotype of RNAi knockdown. Eggs were counted two days after RNAi knockdown began. The number of eggs is given per female. Error bars represent the standard error of the mean of the three biological replicates.

A moderate reduction in eggs was observed in both knockdown groups relative to the negative control group (see Figure 3.20), although no strong reduction in gene expression could be demonstrated and in the case of CD63R a 50% higher expression was measured compared to the negative control. However, because only two biological replicates could be set up for each condition, an analysis of variance (ANOVA) found there to be no statistically significant difference between the groups (p = 0.485).

In the RNA-Seq data, high levels of CD63R expression were observed in mature females. However, the qRT-PCR analysis indicated low expression of CD63R in the control and the treated samples. Therefore, I investigated whether there was a change in mRNA expression between freshly perfused worms and those

cultured *in vitro* for seven days, as had been done for the knockdown experiment. From the batch of worms used for the RNAi soaking experiment several worms had been placed in Trizol directly following the perfusion, which were used as *in vivo* control.

The qRT-PCR analysis showed that CD63R expression was 230-fold higher in freshly perfused worms compared to worms after seven days *in vitro*, whereas CD63a expression was only 14-fold higher in the freshly perfused worms compared to worms after *in vitro* culture (see Figure 3.21), probably due to the fact that CD63a is expressed at higher base levels in all worms compared to CD63R. Some down-regulation of these genes *in vitro* had been anticipated even in the negative control because of the regression of female fertility in culture (Galanti *et al.*, 2012). However, the extent to which expression of CD63a and CD63R declined was unexpected. The down-regulation in mRNA expression with such a large effect size both in the control and RNAi treated worms is likely to interfere with an accurate measurement of the effect of the RNAi treatment and obscure the effect of the RNAi knock down.

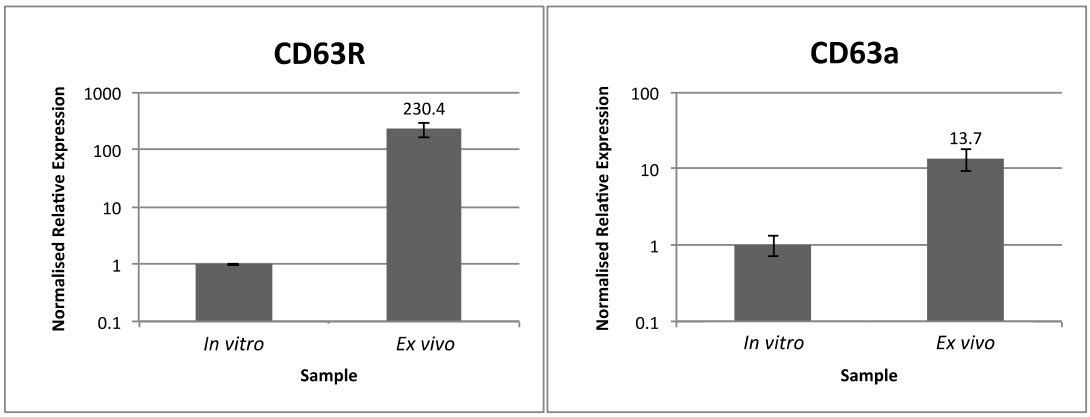


Figure 3.21: Expression of CD63 receptor (left) and CD63 antigen (right) was significantly higher prior to *in vitro* culture. Expression was measured in freshly perfused (*ex vivo*) worms and worms cultured for seven days (*in vitro*). For both qRT-PCR reactions PSMD4 was used as internal reference gene. The differences in expression were statistically significant for both genes (CD63R: $p = 0.016$; CD63a: $p = 0.036$). Error bars represent the standard error of the mean.

The experiment showed that TSP-2 mRNA levels could be depleted by approximately 99.5% by soaking worms in dsRNA (Figure 3.17). However, the soaking did not reduce CD63R mRNA and only resulted in a mild reduction of CD63a mRNA. To facilitate better access of dsRNA molecules to the deeper tissues such as the vitellarium, the treatment was repeated, using electroporation instead of simple soaking as well as a shortening of the experiment from one week to 48 hours. A potential disadvantage of electroporation is the greater level of stress caused to the female and the impact on their egg production this might have, which could obscure any subtle effect the knockdown might have on egg production. The length of the experiment was also shortened to lessen the effect of *in vitro* culturing and reduce the decline in CD63a and CD63R expression in the negative control that was demonstrated previously.

In this RNAi experiment, a very moderate reduction of about 16% in CD63R mRNA levels was observed (Figure 3.22). This reduction was well within the standard error of the normalised expression levels (see Figure 3.22). The results of this experiment did not allow the null hypothesis to be rejected. With more replicates it might be possible to determine, whether the small observed effect seen was reproducible or due to chance. Nonetheless, this result is an improvement on the previous soaking knockdown where an increase in CD63R mRNA levels was observed in the treated sample relative to the negative control.

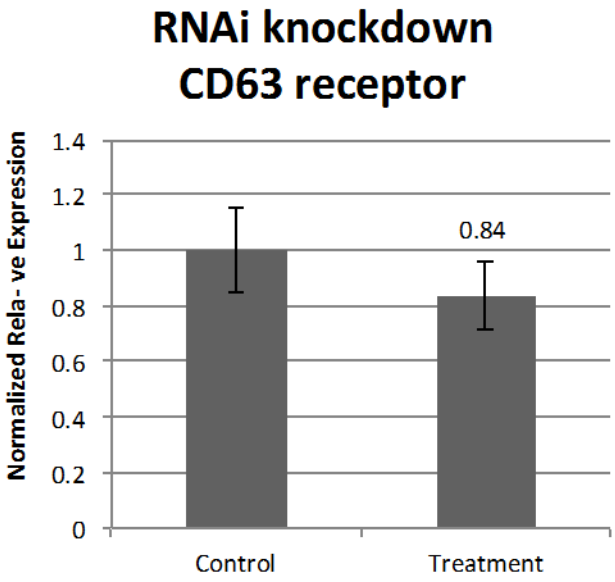


Figure 3.22: No significant knock-down of CD63 receptor expression was observed. RNAi knockdown of CD63 receptor using the 'electroporation method' was measured relative to the negative control sample (electroporated with non-specific dsRNA) using qRT-PCR. PSMD4 was used as internal reference gene. The difference in expression was not found to be statistical significant ($p = 0.822$). Error bars represent the standard error of the mean.

In the CD63a knockdown experiment, a statistically significant ($p = 0.044$) 47% reduction in CD63a mRNA levels was measured with much smaller standard errors measured compared to the CD63R knockdown (see Figure 3.23). In this experiment, the knockdown was found to be more successful than the soaking

experiment [reduction by ~35% (Figure 3.19)]. Taken together, these results suggest that CD63a is targetable by RNAi.

For measuring expression of both CD63R and CD63a, PSMD4 was used as a reference gene. The negative controls, that had not been reverse transcribed, only showed amplification about seven cycles after the samples suggesting that the DNase treatment successfully removed most gDNA contamination (Appendix A.26). The no-template controls showed no amplification until about ten cycles after the samples suggesting that there was little non-specific amplification (Appendix A.26).

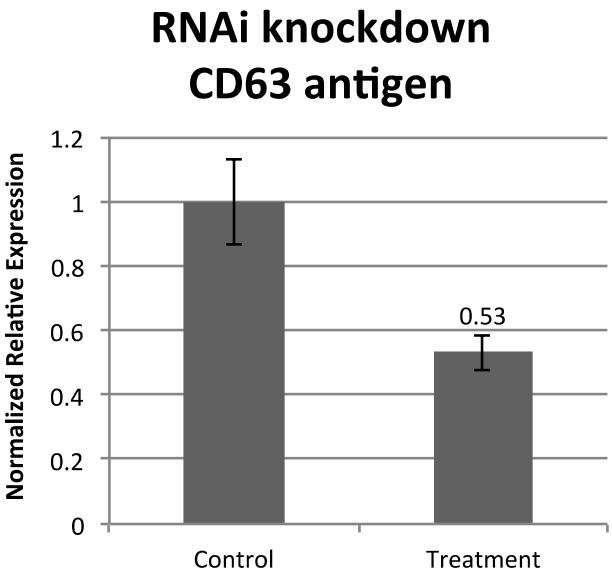


Figure 3.23: A significant knock-down in CD63 antigen expression was observed. RNAi knockdown of CD63 antigen using the 'electroporation methods' was measured relative to the negative control sample (electroporated with non-specific dsRNA) using qRT-PCR. PSMD4 was used as internal reference gene. The difference in gene expression was statistical significant ($p = 0.044$). Error bars represent the standard error of the mean.

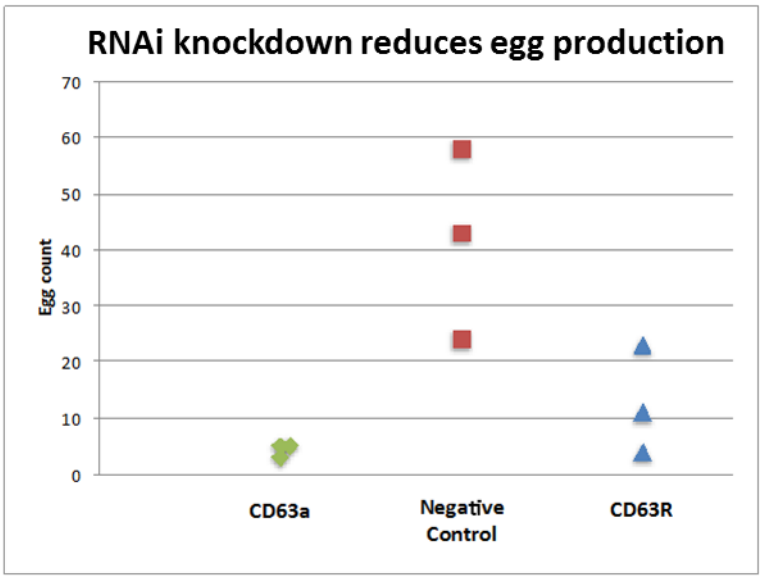


Figure 3.24: The number of laid eggs was reduced significantly in the CD63a and CD63R knockdown worms. The number of eggs per female is plotted for the knockdown of CD63 antigen, the negative control and the CD63 receptor; three replicates were available for each condition. The eggs were laid during the 48 hours following RNAi electroporation. An ANOVA showed that the treatment groups laid significantly fewer eggs compared to the negative control ($p = 0.0156$). The analysis also showed that both, the CD63a and CD63R knockdown resulted in a significant reduction of eggs compared to the negative control (CD63a: $p = 0.00679$; CD63R: $p = 0.0201$).

Again, a reduction in egg laying was observed in both knockdown experiments (see Figure 3.24). A 9.6-fold reduction of egg laying was observed in worms with knocked down CD63a expression and a 3.2-fold reduction in worms with knocked down CD63R expression, relative to the negative control (Figure 3.24). Here, three biological replicates were analysed and the observed reduction in eggs laid was greater than in the soaking experiment (Figure 3.20). As a result, an ANOVA showed that the reduction in eggs laid was found to be statistically significant ($p = 0.0156$) but also that CD63a and CD63R knockdowns significantly affected egg production compared to the negative control (CD63a: $p = 0.00679$; CD63R: $p = 0.0201$). Notably, the number of eggs was overall much lower (approximately 10%) compared to the egg counts in the soaking

knockdown experiment (compare Figures 3.20 and 3.24), which might be due to the effect of electroporation on the females. Although only a moderate reduction of mRNA was observed for CD63a, the significant reduction in egg production is strongly suggestive of a real effect and not chance. This is especially true, as CD63a expression was reduced to a similar extend in the soaking experiment where egg laying was also reduced (albeit not statistically significant) and therefore showing the same trend (compare Figures 3.20 and 3.24).

CD63R expression, however, was reduced to a much smaller extent (Figure 3.23), yet a phenotype was measured at the egg laying level (Figure 3.24). This was also observed in the soaking experiment, where a (non-significant) reduction in egg production was observed (Figure 3.20). At worst, this could suggest that the egg laying phenotype is not a suitable phenotype, as egg production can be sensitive to many other factors, such as stress and injury caused during the perfusion. If this is the case, then a lot of random fluctuations should be seen within replicates. To some extent this can be seen, certainly in the negative control, but also CD63R (Figure 3.24). However, the egg counts for the CD63a were remarkably similar (5, 5 and 3 eggs in each well respectively) and much smaller than even the lowest negative control sample (Figure 3.24). A similar trend was observed in the soaking experiment, giving me further confidence that the observed effect of the CD63a knockdown on egg production was real.

3.3 Discussion

The aim of this chapter was to perform an analysis of gene expression of the developing male and female worms inside the mammalian host. The differential expression analysis reproduced many findings in the published literature, validating the approach. The analysis of male worms showed clearly that the male transcriptome was largely unaffected by pairing, with only a handful of genes consistently differentially expressed between males from SS and MS infections (Leutner *et al.*, 2013). Females on the other hand are known to mature in a pairing dependent manner from the fourth week of infection (Biolchini *et al.*, 2006), with my RNA-Seq data showing the greatest increase in differentially expressed genes between females from SS and MS infections between 21 and 28 d.p.i. Also, in MS infections, egg laying was observed from 35 d.p.i. onwards in addition to egg production related gene expression reaching its peak at 35 and 38 d.p.i. This includes genes encoding eggshell synthesis domain containing proteins, tyrosinases, extracellular superoxide dismutases, as well as genes coding for the tetraspanin CD63 antigen and receptor, showing that paired female worms reached sexual maturity around this time. Another up-regulated gens was TRAP1, a heat shock protein mainly localised to the mitochondria, that protects cells from oxidative stress by inhibiting the respiratory complex II and protecting the cell from apoptosis (Guzzo *et al.*, 2014). Tyrosinases expressed in the vitellaria are essential for proper egg development in schistosomes (Fitzpatrick *et al.*, 2007). However, tyrosinase activity is generally thought to result in oxidative stress, as it has been shown in other systems such as mosquitoes (Christensen *et al.*, 2005). The extracellular superoxide dismutase

and TRAP1 may help cells in the vitellaria to cope with the oxidative stress resulting from the tyrosinase activity to prevent damage to the tissue and allow a high rate of continuous egg production. Together, the results confirmed many findings of the published literature (Fitzpatrick & Hoffmann, 2006; Leutner *et al.*, 2013) but provided a more comprehensive picture of when these sex specific genes become differentially expressed.

When comparing male and female worms at different time points, but particularly at 38 d.p.i., a significant up-regulation of nicotinic receptors was observed in male worms (Table 3.3). Nicotinic receptors represent a group of drug targets in many parasitic nematodes which are targeted by anthelmintics such as levamisole, pyrantel and oxantel (Williamson *et al.*, 2009). These drugs cause the ligand-gated ion channels to open, resulting in "prolonged muscle contraction and spastic paralysis of the worms" (Williamson *et al.*, 2009). Such receptors have recently also become of interest as a drug targets for *S. mansoni* as described by MacDonald *et al.* (2014). The authors observed that the *S. mansoni* nicotinic acetylcholine receptors do not exert an excitatory effect on the neuromuscular system as in mammals and nematodes, but instead function as inhibitory neuromuscular receptors (MacDonald *et al.*, 2014). MacDonald *et al.* (2014) localised two of these proteins to the worm surface as well as in nerves. Therefore, the up-regulation of neurotransmitter receptor genes in male worms could be the result of the greater surface area as well as greater proportion of muscle tissue in male worms in comparison to mature female worms.

During the analysis of the juvenile male worm, genes related to the cytochrome P450 related drug metabolism were found to be up-regulated. Cytochrome P450-dependent metabolism plays an important role in processing praziquantel in humans (Li *et al.*, 2003) but whether this is true for schistosomes is not clear. This pathway was found to contain several differentially expressed genes encoding glutathione S transferases, one of which (Smp_102070) has previously been shown to bind praziquantel and play an important role in the detoxification of *S. mansoni* (McTigue *et al.,* 1995; Cardoso *et al.*, 2003). Another GST encoding gene (Smp_024010) was also found to be down-regulated after 18 d.p.i.. From day 18 to 38, its expression decreased 21.9 and 18.2-fold in male and female worms respectively, potentially making the worm more susceptible to the effects of praziquantel. McTigue *et al.* (1995) showed that homodimers of GST bind praziquantel at the dimer interface, not in their substrate-binding site and concluded that praziquantel inhibits GST activity, which leads to a build-up of reactive oxygen species and toxic metabolites inside the parasite, in particular cytotoxic free radicals. Cardoso *et al.* (2003), on the other hand, concluded that the binding of praziquantel to GST had a protective effect on the worm by keeping praziquantel away from its real cellular target [for example voltage operated calcium ion channel (Nogi *et al.*, 2009)]. Given the rapid effect of praziquantel on schistosomes (Xiao *et al.*, 1985), a build-up of reactive oxygen species and toxic metabolites does not seems as plausible as the conclusion of Cardoso *et al.* (2003). However, in either case, a greater expression of GST encoding genes such as Smp_102070 and Smp_024010 would confer a protective effect on the schistosomes.

When examining the expression of testes related genes, a number of interesting similarities and differences were observed between males from SS and MS infections. The comparison between male worms from MS infections at 21 and 28 d.p.i. revealed that they appear to use a similar mechanism to prevent the premature proliferation of prospermatogonia as mammals. Basonuclin 2 (Smp_138350), a testes specific gene in mammals and *S. mansoni*, was found to be up-regulated in male worms from SS and MS infections prior to pairing at around 28 d.p.i. Vanhoutteghem *et al.* (2014) reported that "basonuclin 2 is required for proper mitotic arrest, prevention of premature meiotic initiation and meiotic progression" of germ cells. In knock out mice, prospermatogonia multiply excessively leading to accumulation of spermatocytes and ultimately apoptosis during meiotic prophase (Vanhoutteghem *et al.*, 2014). Adult male worms are known to produce sperm in the absence of a female partner (Armstrong, 1965); therefore, the role of this gene in schistosomes may be to allow male worms to mature sufficiently before investing energy into sperm production. However, differential regulation of genes potentially involved in sperm production were also observed. Neves *et al.* (2005) demonstrated morphological differences between the testes of males from SS and MS infections, despite both producing sperm. In the male time series analysis, genes encoding the FGF receptor activating protein (Smp_035730, adjusted p-value = 0.0053) and the epidermal growth factor receptor (Smp_152680, adjusted p-value = 0.0069) were found to be up-regulated in paired males. These genes were shown to be expressed in the testes at significantly higher levels than in other organs (see Chapter 5). Both FGF and the EGF signalling pathways regulate cell proliferation and differentiation (Eswarakumar *et al.*, 2005; Herbst, 2004).

This may assist to regulate sperm production in male worms and allow unpaired male worms to conserve energy.

In female worms from SS and MS infections, a much larger number of differentially expressed genes were observed than in male worms. When examining females between 21 and 28 d.p.i. a large proportion of the significantly regulated genes were found to encode transcription factors. This included approximately a third of the genes up-regulated at day 21 (37/119), suggesting that some developmental processes finished then, but also that the regulation of the transcriptome is changing significantly around the time pairing is thought to occur. Interestingly, a comparison of females from SS and MS infections at 28 d.p.i. showed that down-regulation of these genes encoding transcription factors occurs independently of pairing at this stage of the worm development. Female worms became developmentally arrested around this time in the absence of male worms (Figure 3.2), which stimulate further maturation. The down-regulation of these genes coding for transcription factors probably marks the beginning of this developmental arrest. However, at least six genes encoding transcription factors were found to be up-regulated at 28 d.p.i., four of which were found to be consistently associated with sexual maturity in female worms, with expression being pairing-dependent. Expression of these four genes was detected at very low levels in most samples, but a significant increase in expression in females from MS infections was measured from 28 d.p.i. onwards. Although their precise function remains unknown, their expression patterns over the RNA-Seq time course caused them to be clustered with genes expressed

predominantly in the vitellaria of mature female worms. This is further evidence of a changing regulation of transcription in paired females.

Using two clustering algorithms, MBCluster and Kohonen, a set of 76 female specific genes putatively involved in egg production was identified, a subset of which has previously been shown to play important roles in female reproduction (for example Fitzpatrick *et al.*, 2007; Fitzpatrick & Hoffmann, 2006). Due to their functions as cell surface receptors and their potential involvement in signalling, the CD63 antigen and tetraspanin CD63 receptor stood out. Further investigation with qRT-PCR confirmed the RNA-Seq expression patterns. Both genes were expressed in paired females, beginning around 28 d.p.i. To get more insight into the role of these genes and add evidence for their role in egg production whole mount *situ hybridisation* (WISH) and RNAi knockdown experiments were performed for both genes. The WISH experiments showed that both genes were expressed in the vitellarian tissue of the female worms, although the CD63 receptor was expressed in the outer portions of the vitellarium, whereas the CD63 antigen was found to be expressed throughout the vitellarian tissue.

The RNAi knockdown experiment proved challenging. In the first round of knockdowns, using a soaking technique, the positive control, TSP-2, worked well with a reduction of mRNA to 0.5% of the original expression. However, knockdown efficiency for CD63a and CD63R was far less satisfactory. CD63a expression had been reduced by 35% whereas expression of CD63R was measured to be 53% higher after RNAi treatment compared to the control samples. This unexpected result may have been caused by the effects of the *in*

vitro culture. Worms were found to have significantly lower expression of CD63a (~14x) and CD63R (~230x) after *in vitro* culture, compared to freshly perfused worms (Figure 3.21). It has previously been shown (Galanti *et al.*, 2012) that female worms lose maturity and regress outside their host even in the presence of male worms; however, the extent of this effect was greater than expected. In a repeat of this experiment, I opted to reduce the length of the *in vitro* culture after RNAi treatment to 48h, instead of 7 days. Furthermore, I chose to use electroporation rather than soaking in the second experiment to maximise probe penetration, at the expense of worm fitness. This experiment proved to be more successful with a measured reduction of CD63R mRNA of 16% (instead of a measured increase) and a reduction of CD63a mRNA of 47% (instead of 35%). Finally, the number of eggs laid by the RNAi treated worms was significantly reduced for both CD63R and CD63a by 3.2 and 9.6-fold respectively, suggesting that both genes play a role in the production of eggs. The CD63 antigen is known in humans to form complexes with integrins and act as part of signalling complexes that regulate cell proliferation and differentiation (Tugues *et al.*, 2013). For example CD63a is known to interact with syntenin-1 – an adaptor protein regulating vesicle formation and TGF-β mediated SMAD activation (Hwangbo *et al.*, 2016) – and other adaptor proteins of the clathrin-dependent endocytosis pathway (Charrin *et al.*, 2014) as well as the "tissue inhibitor of metalloproteinase-1" and integrins (Jung *et al.*, 2006). The interactions with syntenin-1 and the adaptor proteins occur at the same location on the CD63a protein and are thought to be mutually exclusive (Berditchevski & Odintsova, 2007). By interacting with syntenin-1, CD63a might play a role in the regulation of TGF-β signalling as well as the formation of vesicles, potentially in the

vitellocytes of the maturing female. Mature (S4) vitellocytes are known to contain a large number of vesicles containing eggshell protein precursors (Cogswell *et al.*, 2012). The CD63 receptor, on the other hand, may play a role in endocytosis via clathrin-dependant as well as independent pathways (Berditchevski & Odintsova, 2007) and may aid in the formation of the vitelline cells' secretory vesicles that will later go on to form the eggshells. Another mechanism by which CD63 antigen may regulate vitelline proliferation, namely apoptosis, will be examined in more detail in Chapter 5.

Chapter 4

Transcriptome analysis of sexually regressing *S. mansoni*

4.1 Introduction

The sexual maturity of female schistosomes is pairing-dependent, a trait of their biology that requires close contact to a male partner for the ovaries and vitelline tissue of the female to develop fully. Female reproductive tissues regress when the male partner is no longer present, and regenerate upon re-pairing.

Galanti *et al.* (2012) performed a series of experiments in order to examine the contribution of pairing on apoptosis and cell proliferation and the resulting effect on the growth and shrinkage of the female reproductive system. They involved culturing mature female worms *ex vivo* for up to 11 days either paired with a male or in the absence of a mate and recording their size and egg laying. They also measured the gene expression of an eggshell protein, p14, the rate of mitosis and apoptosis in female worms, which served as a marker of female sexual development. The results showed that both paired and single females regress sexually in *ex vivo* culture but, importantly, that paired females undergo this regression more slowly, retaining cellularity in their vitelline tissue and expressing p14 at higher levels, albeit lower than *in vivo*. Based on the relative amount of TUNEL (for apoptotic cells) and BrdU (for dividing cells) staining observed in paired and separated females, Galanti *et al.* (2012) also concluded that growth and shrinkage of the vitelline tissue is primarily regulated by the rate of apoptosis, not proliferation, and that unpaired females have more apoptotic cells. These results highlighted two things in particular: a pairing-specific effect on the females that can even be observed during *ex vivo* culture and secondly that in the culture conditions used by Galanti *et al.* (2012), females

do not retain their mature state independent of the pairing status. In their discussion, reference is made to a richer media (Basch, 1988) that better supports female fertility, although still not perfectly.

Work described in this chapter examined the changes in gene expression that take place in unpaired females compared to those that remain paired over time. By examining the mechanism by which female *S. mansoni* undergo regression of their reproductive tissues at the transcriptome level, new avenues for preventing or treating schistosomiasis may be uncovered.

Following a series of testes with different culture conditions Basch medium was chosen for this experiment to minimise the loss of fertility of females due to *ex vivo* culture. The length of the experiment was set to 8 days, based on qPCR results (Galanti *et al.*, 2012). These showed that single females expressed significantly lower levels of p14 than paired females from day 7 and that expression levels do not fall much further from then until day 11. Worms were perfused from mice and sorted by sex. Some males and females were immediately placed in separate microcentrifuge tubes containing Trizol (day 0) for RNA extractions. The remaining worms were split into three groups and maintained *ex vivo* using modified Basch medium (see methods). The three groups consisted of single male, single female and paired worms. After 4 and 8 days, half of each group was placed in Trizol (paired worms were separated beforehand), RNA was extracted from all samples and RNA-Seq libraries were prepared as before (see methods).

Based on the results described in Chapter 3, I expected to find a reduction of the expression of fertility related genes in single females when compared to paired females but also in paired females when compared to day 0 females. Examples of such fertility related genes include eggshell protein (Smp_000430), various eggshell synthesis domain containing proteins, major egg antigens and female specific genes (Smp_077900 & Smp_000290). Furthermore, based on the results by Galanti *et al.* (2012), I also expected to find evidence for increased levels of apoptosis in females cultured *in vitro* for 8 days - especially in single females – at the transcriptome level.

Previous publications have gathered a list of apoptosis related genes of *Schistosoma*. Peng *et al.* (2010) discovered and characterised an inhibitor of apoptosis (SjIAP) in *S. japonicum*, which they propose belongs to a conserved group of negative regulators of apoptosis with homologues in *H. sapiens*, *C. elegans* and *D. melanogaster*. Using both *S. japonicum* cell lysate and transformed human 293T cells, inhibition of apoptosis by SjIAP was demonstrated. Using the sequences of human genes regulating apoptosis, namely the Bcl-2 family, Lee *et al.* (2011) identified eight Bcl-2 related genes in *S. mansoni* with a variety of domain architectures. More recently Lee *et al.* (2014) published a summary of eleven apoptosis regulating genes, including five members of the Bcl-2 family: Bcl-2, Bak, Bcl-2 2, BH1 and BH3. They also identified an Apaf1 homologue, four caspase homologues and included the *S. mansoni* ortholog of the SjIAP gene characterised by Peng *et al.* (2010), which they noted shares no relation to other IAP proteins.

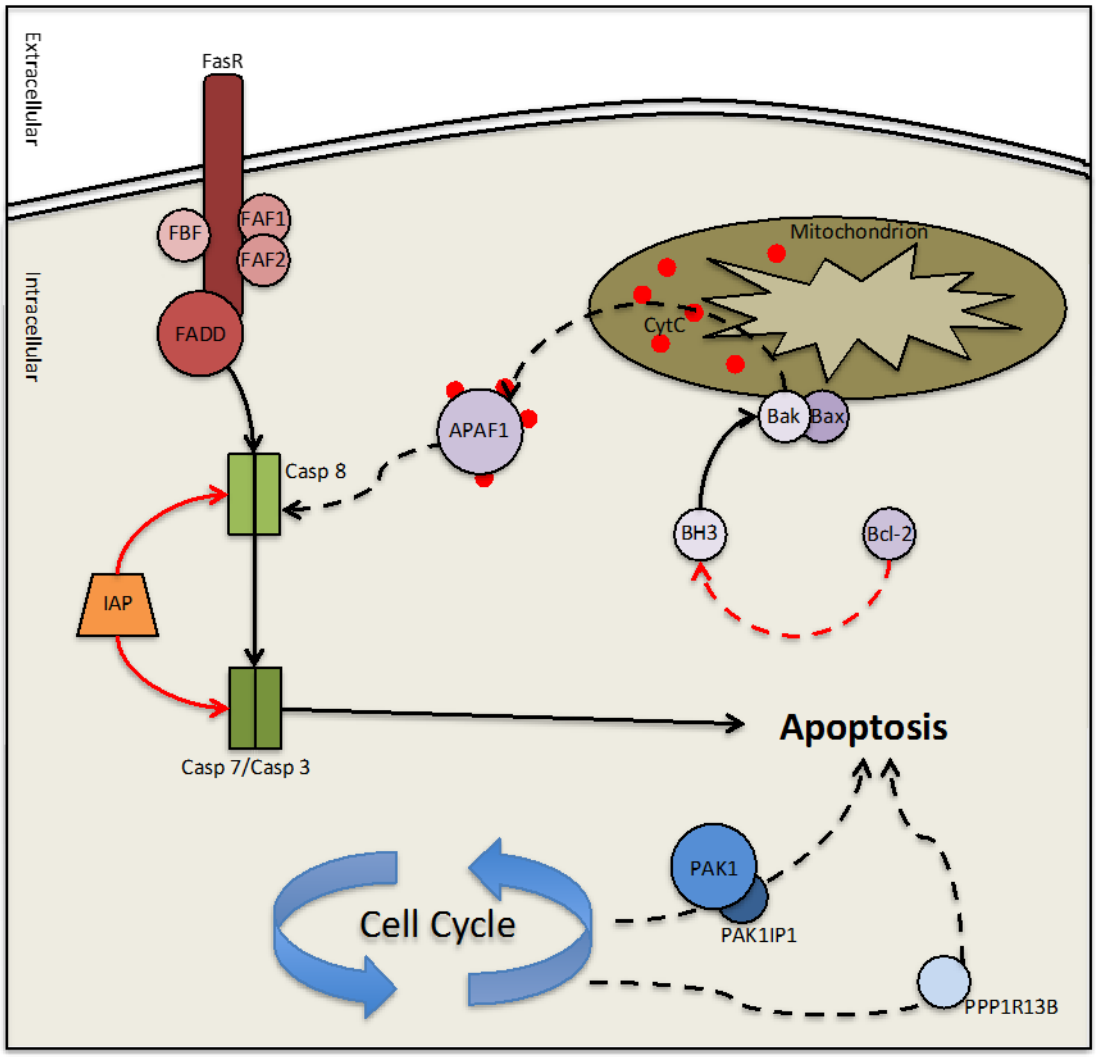


Figure 4.1: Apoptosis in *S. mansoni*. Black arrow – activation; red arrow – inhibition; dashed arrow – indirect interaction. The *S. mansoni* apoptosis pathway is similar to that of mammals, including a Fas receptor homologue and associated proteins, a caspase cascade, involvement of the mitochondria via Bak/Bax and Bcl-2 homologues as well as a cell cycle dependent apoptosis regulators such as PAK. Diagram modified from Lee *et al.* (2014) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) website (2012). Abbreviations: FasR, Fas Receptor; FAF, Fas Associated Factor; FBF, Fas Binding Factor; FADD, Fas Associated Death Domain; Casp, Caspase; IAP, Inhibitor of Apoptosis Protein; APAF1, Apoptotic Protease Activating Factor; CytC, Cytochrome C; Bak, Bcl-2 homologue antagonist/killer; Bcl, B Cell Lymphoma; BH3, Blc-2 homology 3; PAK1, p21 activated kinase 1; PAK1IP, PAK1 Interacting Protein; PPP1R13B, Apoptosis-Stimulating of p53 Protein1.

Despite these efforts, there are still considerable gaps in our understanding of the putative schistosome apoptosis pathway as summarised in Figure 4.1. The presence of Bak and Bax homologues in *S. mansoni* suggests involvement of the mitochondria in schistosome apoptosis as their function in other higher eukaryotes is to puncture the mitochondrial outer membrane and form pores in it (Westphal *et al.*, 2011). Furthermore, the Bcl-2 and BH3 homologues may be involved in the regulation of Bak/Bax activity, which leads to permeabilisation of the mitochondrial outer membrane and the release of cytochrome C and activation of the caspase cascade (Westphal *et al.*, 2011). In *S. mansoni* four caspase homologues have been identified which could be activated by cytochrome C via the apoptotic protease activating factor 1 (APAF1). Induction of apoptosis can occur either from within the cell, by proteins regulating the cell cycle and detecting critical DNA damage or by extracellular signals that stimulate the Fas receptor family. There are several apoptosis regulating genes that link the *S. mansoni* cell cycle to its apoptotic machinery including homologues of the p21 activating kinase (PAK) and an PAK interacting protein 1 (PAKIP1) as well as a p53 apoptosis stimulating factor (PPP1R13B). There is also a pro-apoptotic "death receptor", a homologue of FAS, as well as four homologues of FAS receptor interacting proteins including a FAS-Associated protein with Death Domain (FADD) homologue, a key mediator of pro-apoptotic signalling, which links the Fas receptor to caspase activation in mammals.

The ability of sexually mature female worms to regress to a more juvenile stage when unpaired is an intriguing aspect of schistosome biology. The work in this chapter aimed to examine how the expression of apoptosis-related genes

changes during the regression of female reproductive tissue. To do so, female worms were kept either by themselves or with a male worm *in vitro* for up to eight days. Several technical issues were encountered during the data analysis for this chapter, so that the focus of the chapter shifted to also examine the effect of *in vitro* culture on gene expression in male and female worms.

4.2 Results

4.2.1 Optimising culture media

In order to maintain worms *in vitro*, different culturing conditions were tested, particularly, different media and red blood cell concentrations. The number of eggs laid by female worms over the course of 28 days was used as a measure for the healthiness of cultured worms. The tests with different culturing conditions showed that Basch medium resulted in an approximately 6-fold more eggs being laid (see Figure 4.2) compared to a less complex, RPMI-based medium (see Chapter 2.1.5). While female worms in Basch medium were found to have laid approximately 190 eggs on average, females in the RPMI based media only laid around 30 eggs. Using a *t*-test the average number of eggs for both samples was significantly different ($p < 2.2E\text{-}16$).

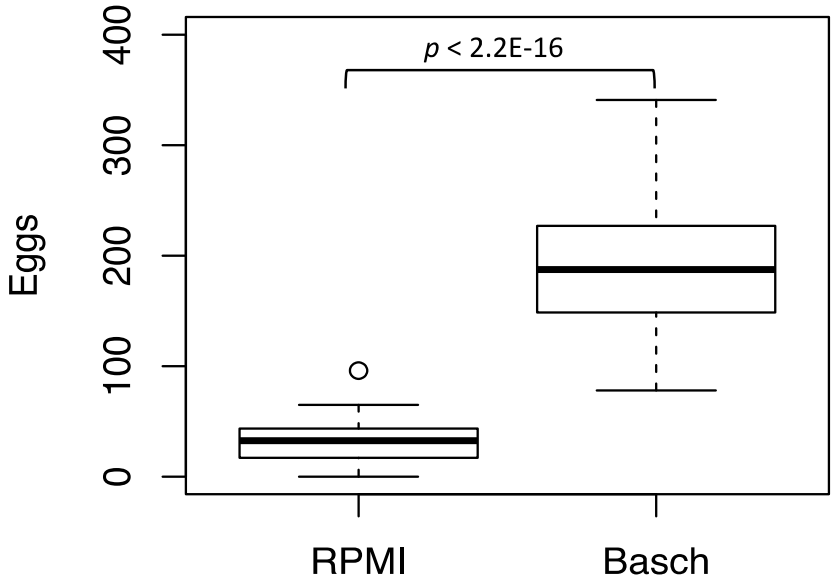


Figure 4.2: Significantly more eggs were laid in Basch media compared to RPMI. Box plot of eggs laid per female in RPMI based medium and Basch medium after 28 days in *in vitro* culture. Both groups contained 36 paired females. Using a Welch two sample *t*-test, a significant difference between the numbers of eggs laid by both groups was found; $p < 2.2E\text{-}16$.

Furthermore, the addition of red blood cells (RBC) was shown to increase the rate of egg laying. 1% RBC resulted in the highest number of eggs being laid (an average of 178 eggs), 0.1% resulted in fewer eggs (an average of 122 eggs) and no RBC resulted in the lowest number of eggs being laid (an average of 86 eggs) (see Figure 4.3). These differences were found to be statistically significant using a *t*-test; for 0.0% and 0.1% RBC $p = 0.010$; for 0.1% and 1.0% $p = 3.3E\text{-}04$.

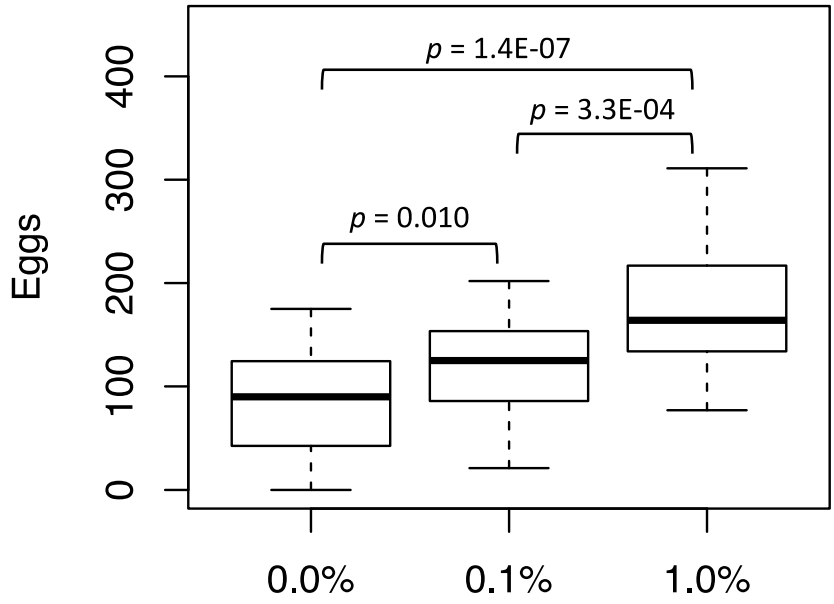


Figure 4.3: Significantly more eggs were laid in media enriched with 1.0% red blood cells compared to 0.0% and 0.1%. Box plot of eggs laid per female in Basch medium with varying amounts of red blood cells (by volume) after 28 days in *in vitro* culture. All groups contained 24 paired females. Using a Welch two sample *t*-test, significant differences between the numbers of eggs laid by the three groups were found.

For the main experiment, testing the effect of unpairing on the female transcriptome, the culture conditions with the best egg laying results were chosen: Basch medium with a 1% concentration of RBC. In total 100 worms, 50 female and 50 male, were used in this experiment. RNA-Seq libraries in this experiment were obtained from single worms. 10 biological replicates were set up for each time point and pairing condition, however due to the small quantities of RNA isolated from single worms only 60 out of the 100 libraries were completed successfully. While for some samples the RNA extraction failed entirely, other samples contained degraded RNA. Overall at least four biological replicates were successfully produced for each samples group. At the beginning of the experiment (day 0) the selected females were all sexually mature and paired. Both groups, paired and unpaired, remained visibly unchanged by day 8,

except that in both groups the stomachs of female worms appeared less darkly coloured. This suggests that despite the addition of RBC to the Basch medium the rate of feeding was lower during *in vitro* culture.

4.2.2 Apoptosis related genes

In preparation for the functional interpretation of the differential expression analysis, apoptosis related genes were identified. In order to determine whether any significant up-regulation of apoptosis occurred in the unpaired females when compared to the females that remained paired throughout the 8 days, I first compiled a list of apoptosis related genes both from the literature (Lee *et al.*, 2011, 2014; Peng *et al.*, 2010) as well as by using BLAST to identify other homologues of known apoptosis genes in the *S. mansoni* genome (see Chapter 2.4.12). In total, 29 genes were found but two were subsequently removed (see Appendix B.2-10 for details on the function and domains of the apoptosis related genes). These two genes, Smp_041630 and Smp_043360, had been initially identified from the literature, but had neither relevant Pfam domains nor significant BLAST hits to known apoptosis related genes (Table 4.1) and were excluded from the analysis. This was most likely the result of changes to the gene models.

Gene ID	Description	Species	Total score	Query cover	E value	Identity	Accession
Smp_002410	14-3-3 protein epsilon	*H. sapiens*	306	95%	2.00E-103	65%	NP_006752.1
Smp_009760	14-3-3 protein zeta/delta	*H. sapiens*	301	93%	3.00E-101	64%	NP_003397.1
Smp_034840	14-3-3 protein epsilon isoform transcript variant 1	*H. sapiens*	295	95%	2.00E-99	65%	AAX68683.1
Smp_032000	CED-3, isoform b	*C. elegans*	143	64%	5.00E-39	33%	NP_001255709.1
Smp_141270	CED-3, isoform b	*C. elegans*	115	32%	6.00E-28	28%	NP_001255709.1
Smp_028500	CED-3, isoform b	*C. elegans*	121	79%	9.00E-32	32%	NP_001255709.1
Smp_199580	Chain A, Crystal Structure Of Mutant Form Of Caspase-7	*H. sapiens*	207	77%	4.00E-63	40%	4HQ0_A
Smp_041630	Microtubule-associated protein 1A, isoform CRA_a	*H. sapiens*	34.3	28%	1.5	27%	EAW92606.1
Smp_095190	Bak-2 protein	*H. sapiens*	72	67%	2.00E-14	33%	AAA74467.1
Smp_168470	Apoptosis regulator Bcl-2 β isoform	*H. sapiens*	46.2	14%	6.00E-05	45%	NP_000648.2
Smp_213250	Chain A, Bak Domain Swapped Dimer Induced By Bidbh3 With Chaps	*H. sapiens*	90.5	24%	3.00E-19	32%	4U2U_A
Smp_213250	bcl-2 homologous antagonist/killer	*H. sapiens*	90.5	24%	6.00E-19	32%	NP_001179.1
Smp_043360	Chain A, Crystal Structure Of The Nudix Domain Of Nudt6	*H. sapiens*	29.3	70%	3.4	27%	3H95_A
Smp_044000	BAX inhibitor 1	*H. sapiens*	179	83%	2.00E-53	42%	AAU29521.1
Smp_072180	Apoptosis regulator BAX isoform delta	*H. sapiens*	48.9	41%	1.00E-06	32%	NP_620118.1
Smp_210790	Protein lifeguard 4 isoform b	*H. sapiens*	171	93%	5.00E-51	44%	NP_057140.2
Smp_084610	FAS-associated factor 2	*H. sapiens*	233	87%	4.00E-69	31%	NP_055428.1
Smp_213730	Chain A, Structure Of The Variant Histone H3.3-h4 Heterodimer In Complex With Its Chaperone Daxx	*H. sapiens*	146	28%	2.00E-38	39%	4HGA_A
Smp_213730	Death domain-associated protein 6 isoform c	*H. sapiens*	154	26%	3.00E-38	40%	NP_001241646.1
Smp_077540	FAS-associated factor 1	*H. sapiens*	182	45%	1.00E-29	28%	NP_008982.1
Smp_148130	FAS-binding factor 1	*H. sapiens*	83.2	53%	3.00E-15	26%	NP_001074011.1
Smp_207000	Anamorsin isoform 1	*H. sapiens*	122	93%	2.00E-31	31%	NP_064709.2
Smp_013040	Cathepsin D preproprotein	*H. sapiens*	419	85%	3.00E-142	53%	NP_001900.1
Smp_136730	Cathepsin D preproprotein	*H. sapiens*	315	88%	5.00E-102	41%	NP_001900.1

Smp_022110	PREDICTED: p21-activated protein kinase-interacting protein 1 isoform X1	*H. sapiens*	117	53%	2.00E-28	32%	XP_011513022.1
Smp_179800	Serine/threonine-protein kinase PAK 3 isoform c	*H. sapiens*	506	54%	6.00E-172	75%	NP_001121644.1
Smp_179800	Serine/threonine-protein kinase PAK 3 isoform a	*H. sapiens*	504	54%	7.00E-172	75%	NP_002569.1
Smp_179800	P21-activated kinase 3	*H. sapiens*	504	54%	1.00E-171	75%	AAF67008.1
Smp_129670	Apoptosis-stimulating of p53 protein 2 isoform X3	*H. sapiens*	196	11%	3.00E-50	46%	XP_011542572.1
Smp_137540	Chain A, Crystal Structure Of The Reduced Human Apoptosis Inducing Factor Complexed With Nad	*H. sapiens*	254	63%	3.00E-73	32%	4BUR_A
Smp_137540	Chain A, Crystal Structure Of Apoptosis Inducing Factor (Aif)	*H. sapiens*	253	63%	4.00E-73	32%	1M6I_A
Smp_197180	Apoptosis inhibitor 5 isoform b	*H. sapiens*	293	98%	1.00E-90	35%	NP_006586.1
Smp_168070	Fas Receptor	*H. sapiens*	60	21%	8.00E-09	33%	EAW50150.1
Smp_140260	Apoptotic peptidase activating factor, isoform CRA_h	*H. sapiens*	147	30%	1.00E-20	22%	EAW97610.1

Table 4.1: Summary of Schistosoma mansoni putative apoptosis-related genes. This table provides *S. mansoni* gene IDs, product names and their *Homo sapiens* or *Caenorhabditis elegans* homologues with BLAST scores, e-value and coverage. The total score is the number of sequence matches between the query and the subject sequences. Query coverage is the proportion of query sequence covered by the subject sequence. The Expected (E) value is a measure of how probable the given alignment is by chance. The identity is the extent to which query and subject sequences have the same amino acids in the same position, given as percentage.

4.2.3 Sequencing & sample clustering

Illumina libraries were prepared from 50ng of total RNA as outlined in Chapter 2. In total Illumina sequencing produced nearly 2.2 billion reads, with 10.7-75.8 million reads per sample (average 36.5 million) (see Appendix B.1). Of those reads, on average 90% were successfully mapped to the *S. mansoni* reference genome using TopHat2 (see Appendix B.1). The unmapped reads were mostly multimers of adapter and PCR primers, which cannot be completely removed during the size selection step of library preparation.

A principal component analysis was performed using DESeq2 It showed how well samples from each experimental group clustered together and how similar the groups were to one another (see Figure 4.4). Male and female samples clustered separately and the PCS suggests that the transcriptomes of paired and separated worms changed over the course of the 8-day experiment (Fig. 4.4). However, many samples did not cluster by treatment. While male worms from different time points separated relatively well, paired and separated males at each time point did not form discrete groups (Fig 4.4). Overall, the first two principal components explained 67% of the variability of samples, most of which was attributable to differences between male and female worms (Fig 4.4).

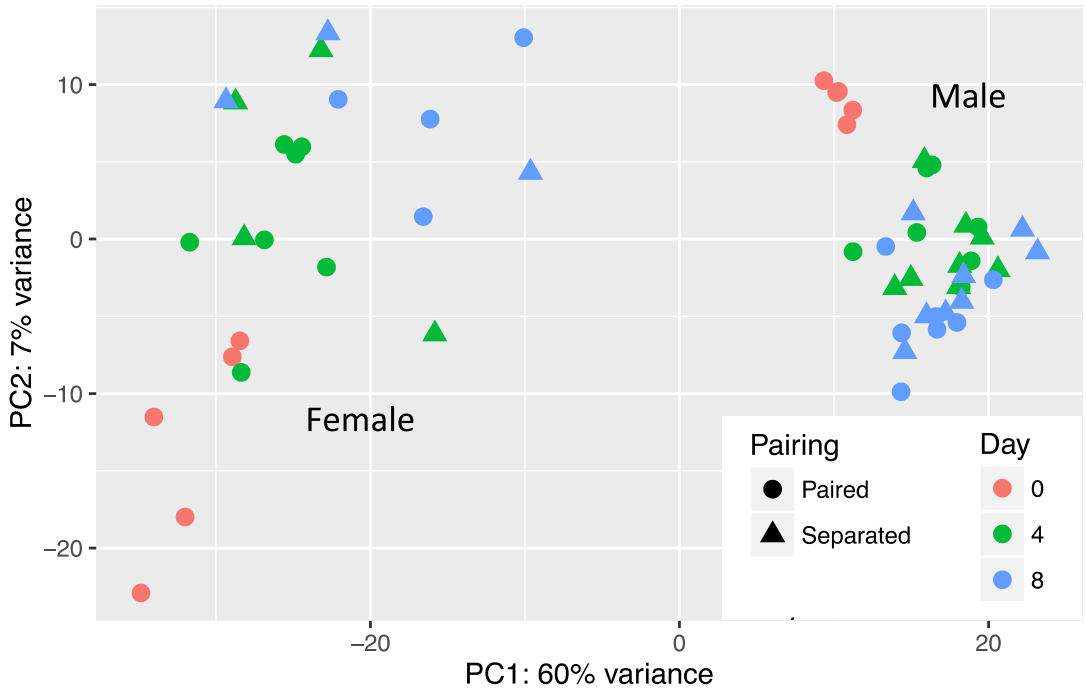


Figure 4.4: Samples clustered well by gender but separated less well by time *in vitro* and pairing status. Principal component analysis (PCA) plot of RNA-Seq samples at 0, 4 and 8 days after perfusion. The RNA-Seq samples separated into female (left) and male (right) clusters but samples separate poorly by treatment (paired and separated).

4.2.4 Comparison of pairing and separated females at day 8

The PCA plot (Figure 4.5 B) of paired and separated females at day 8 shows that the separated females (blue) showed greater transcriptome variability than paired females. However, the groups did not separate well and in fact overlapped. Compared to the paired females, the PCA plot showed that separated females responded to separation less uniformly and that the treatment introduced greater variability of gene expression at the transcriptome level.

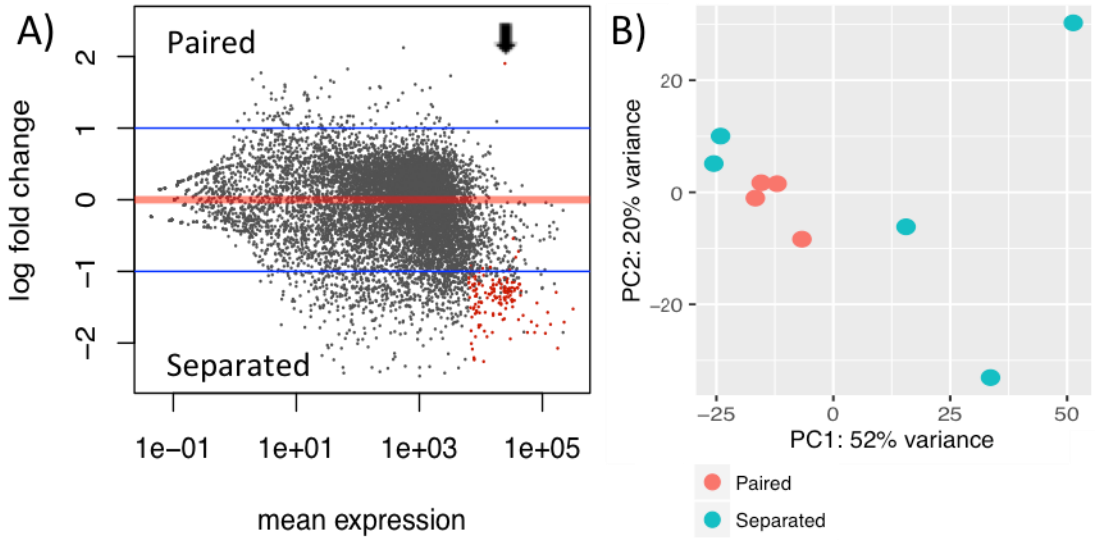


Figure 4.5: Most differentially expressed genes were up-regulated in separated females. Pair-wise comparison of paired and unpaired females at day 8 after perfusion. A) The Log ratio against mean (MA) plot showed only one differentially expressed gene (marked with arrow) was up-regulated in paired females, all others in separated females. Blue lines indicate 2-fold expression change. B) A principal component analysis of RNA-Seq data from paired and separated females. The gene expression data of separated females grouped less tightly but overlapped with the data of paired female samples.

Out of a total of 10,828 genes with non-zero read counts, 9464 (92%) were removed from the analysis using the DESeq2 independent filtering algorithm, and 297 were removed as outliers, based on their Cook cut-offs (Chapter 2.4.4), a single gene (Fig 4.6) was found to be up-regulated in paired females but 149 were up-regulated in single females (see Appendix B.11 for expression data of the DEGs of this chapter). Due to the high variability between biological replicates (Fig 4.5 B) the RNA-Seq analysis yielded a relatively low number of differentially expressed genes when compared to Chapters 3 and 5. Here, the default p-value threshold for the DESeq2 package (adjusted p-value < 0.1) was used to determine significance. In Chapters 3 and 5, the large number of identified DEGs allowed for a more stringent selection to remove a higher

proportion of false positive results. To increase the statistical power of this analysis, DESeq2 removed a large proportion of genes from this part of the analysis based on the independent filtering algorithm, as mentioned above. Genes with a smaller average expression than 6350 reads per sample were removed from this comparison. This is done to remove genes with low absolute expression, which have little chance of being found to be differentially expressed. As a result, the p-values of the remaining genes have to be adjusted less harshly for multiple hypothesis testing.

The gene found to be up-regulated in paired females codes for the major egg antigen (p40) (Smp_049270; 3.7x change, adjusted p-value = 0.046) (Table 4.2). Finding reproduction and fertility related genes up-regulated in the paired females was expected; however, many of the genes found to be up-regulated in the unpaired females were also associated with these aspects of female biology (Appendix B.11). These include genes coding for an eggshell protein (Smp_000430; 3.6x change, adjusted p-value = 0.089) and a trematode eggshell synthesis domain containing protein (Smp_077890; 3.0x change, adjusted p-value = 0.089). Furthermore, genes found in Chapter 3 to be associated with pairing *in vivo* were found to be expressed at significantly higher levels in the single females compared with the paired females, including the genes coding for the CD63 receptor (Smp_155310; 3.8x change, adjusted p-value = 0.078) and the CD63 antigen (Smp_173150; 3.1x change, adjusted p-value = 0.089) which were the 9th and 27th most up-regulated genes in single females respectively (Appendix B.11). A striking number of mitochondrial genes (8 out 150 DEGs), but even more ribosomal genes (48 out 150 DEGs) was found to be differentially

expressed, suggesting that unpaired females have a higher metabolic rate and greater rate of protein synthesis, both of which seems contradictory to the published literature given that pairing is associated with growth and reproduction (Fitzpatrick & Hoffmann, 2006; Sun *et al.*, 2015). None of the apoptosis related genes from my gene set were found to be differentially expressed (see Appendix B.11).

Gene ID	Average Read Count	Fold Change	Adjusted p-value	Product
Smp_074390	7340.6	4.242	0.033	Eukaryotic translation initiation factor binding protein
Smp_131110	177674.7	4.205	0.053	Uncharacterised protein
Smp_018930	7063.7	4.011	0.053	Uncharacterised protein
Smp_095980	30236.6	3.838	0.070	Extracellular superoxide dismutase (Cu Zn)
Smp_155310	8225.5	3.760	0.078	Tetraspanin CD63 receptor
Smp_023840	12609.0	3.758	0.053	Uncharacterised protein
Smp_900040	112222.7	3.631	0.064	Uncharacterised protein
Smp_000430	75350.6	3.607	0.089	Trematode eggshell synthesis protein
Smp_900100	57478.3	3.603	0.065	NADH dehydrogenase subunit 3
Smp_202690	15523.4	3.599	0.053	Putative universal stress protein
Smp_006040	11467.1	3.563	0.053	Uncharacterised protein
Smp_145370	6960.2	3.540	0.033	Uncharacterised protein
Smp_090520	9271.9	3.434	0.064	Purine nucleoside phosphorylase
Smp_241620	82078.6	3.373	0.089	Uncharacterised protein
Smp_054160	135853.5	3.328	0.053	Glutathione S-transferase
Smp_900050	227094.7	3.260	0.070	NADH dehydrogenase subunit 5
Smp_043120	41091.7	3.179	0.065	Universal stress protein
Smp_058690	20056.6	3.022	0.088	Glutathione peroxidase
Smp_158110	9316.0	2.257	0.053	Putative thioredoxin peroxidase
Smp_079770	41164.0	1.642	0.033	Putative disulfide-isomerase ER-60
Smp_049270	24812.2	0.267	0.045	Major egg antigen (p40)

Table 4.2: The 20 most differentially expressed genes identified in the comparison of paired and separated females at day 8 after perfusion. Average read counts provide a measure of absolute expression for a given gene. The fold changes reflect expression in separated females relative to paired females. Notably only a single gene (Smp_049270) was significantly up-regulated in paired females. The *p*-value has been adjusted for multiple hypothesis testing.

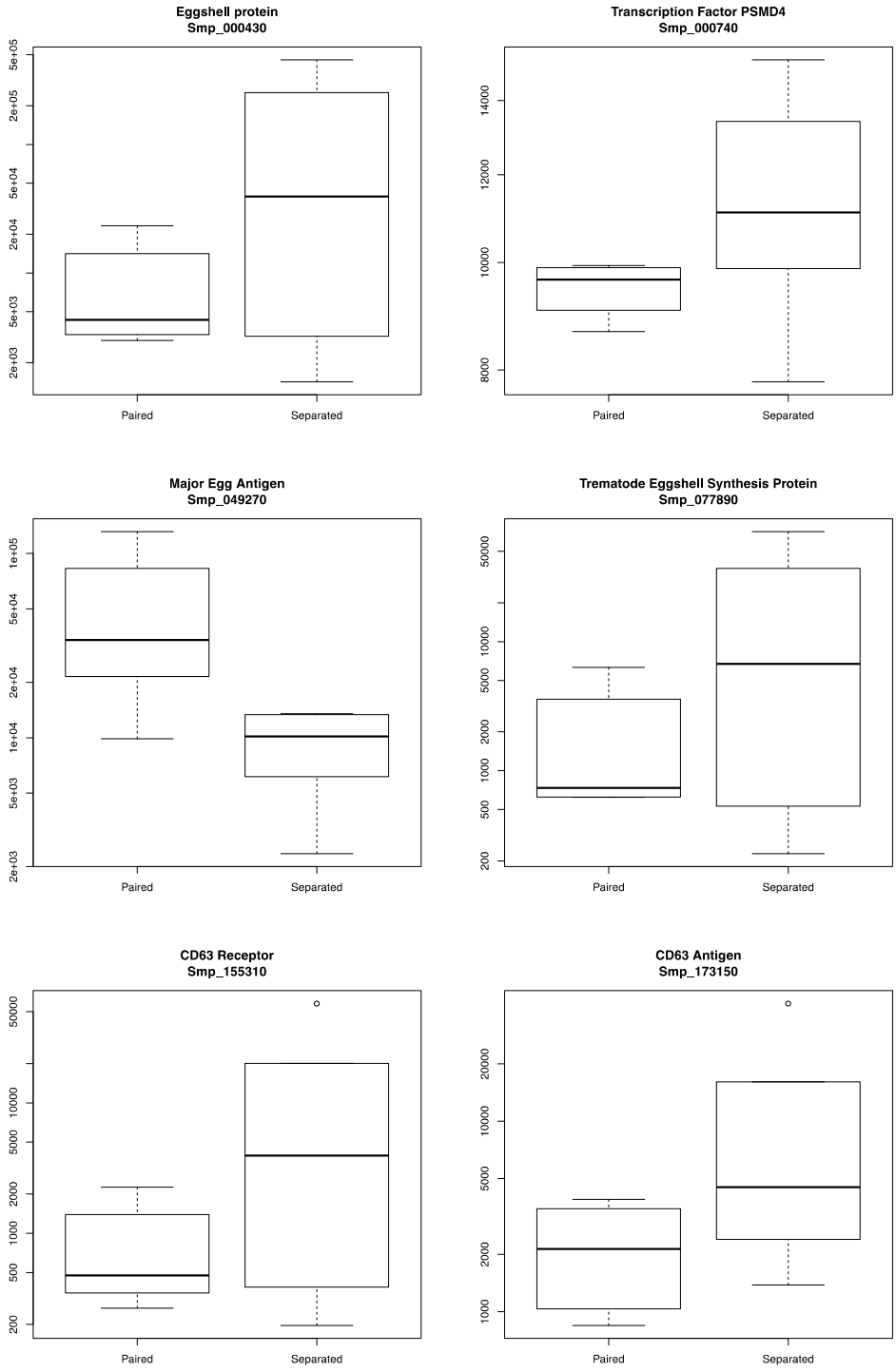


Figure 4.6: Biological replicates of separated females have great variability in gene expression. This is a box whisker plots of normalised read counts, a measure for absolute gene expression, for all RNA-Seq samples of paired and separated females at day 8 after perfusion. In all genes, with the exception of Smp_049270 (coding for the major egg antigen), the expression is less uniform and more variable in the separated females compared to paired females.

4.2.5 Comparison of pairing and separated males at day 8

The PCA plot (Figure 4.7 B) of paired and separated males at day 8 shows three relative outliers, one from the separated and two from the paired males. The remaining samples there do not separate into two groups but overlap. Given that pairing has been shown to have a very limited impact on the male transcriptome (Leutner *et al.*, 2013), it was not surprising that separation only had a small effect on the transcriptome. Furthermore, as the MA plot (Figure 4.7 A) shows, there were only small fold changes of gene expression between the two groups.

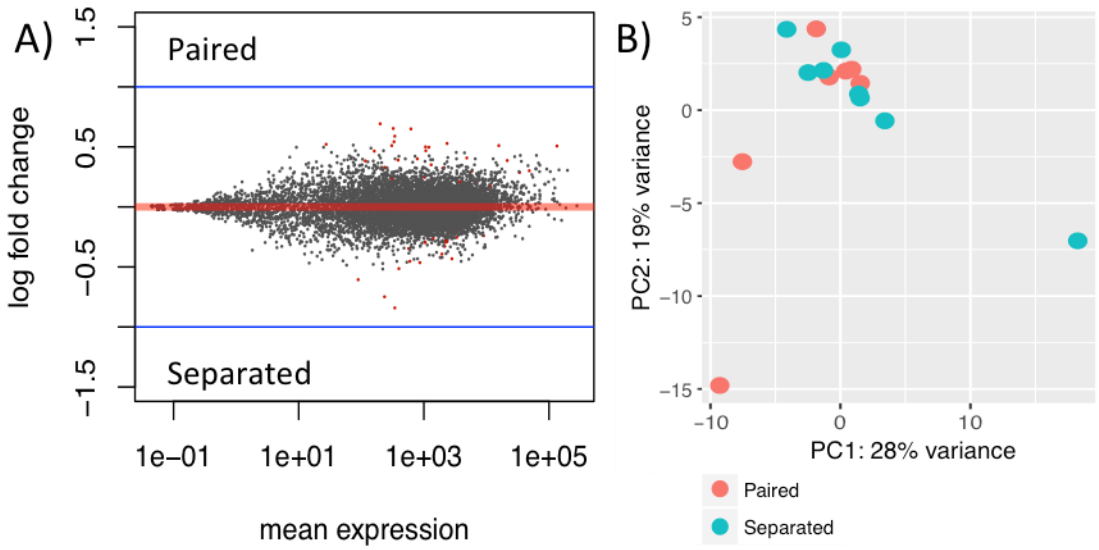


Figure 4.7: The low number of differentially expressed genes between separated and paired males has small fold changes. Pair-wise comparison of paired and separated males at day 8 after perfusion. A) The log ratio vs mean (MA) plot showed that none of the differentially expressed genes have a fold change greater than 2 (log2 fold change > 1), indicated by the blue lines. This small effect size stood in contrast to the female comparison where much bigger fold changes in expression were found. B) Principal component analysis of paired and separated males showed outliers in both groups but also that the two groups of males did not separate, suggesting a very limited effect of the treatment.

In the comparison of female samples, differential expression analysis identified only 150 DEGs despite there being over 900 genes with greater than two-fold change in expression (Figure 4.5), the result of high variance within biological replicates (Figure 4.6). In this comparison of male worms on the other hand, 42 DEGs were identified but not a single gene was found to have a greater than two-fold change in expression (Table 4.3).

In the comparison of paired and separated male worms at day 8, a total of 18 genes were found to be up-regulated in paired, and 24 genes up-regulated separated males (Fig. 4.7A and Table 4.3). This is out of a total of 10172 genes with non-zero read counts, of which 2029 (20%) were removed using the independent filtering algorithm of DESeq2. The cut-off used for filtering was a mean expression of fewer than 78 reads. The cut-off used to determine differential expression was again an adjusted p-value < 0.1, which is the DESeq2 default. Noticeably, unlike in the female comparison where the median fold change of DEGs was 2.5x change, the fold changes for the male DEGs were smaller, ranging from 1.8x change in expression to as small as 1.12x change, with a median fold change of only 1.3x for DEGs (Table 4.3). Several genes encoding proteins with structural functions were differentially expressed in separated males (Table 4.3). These included two actin homologues (Smp_183710 and Smp_046600), two tegument allergen-like proteins (Smp_086480 and Smp_077310) and two tubulin chains (Smp_103790 and Smp_192110) (Table 4.3). In contrast, the genes coding for an endothelin 2 homologue (Smp_122030) and an endothelin-converting protein (Smp_159370) were up-regulated.

Gene ID	Average Expression	Fold Change	Adjusted p-value	Product
Smp_169090	343.6	1.788	5.81E-04	Solute carrier family 43
Smp_128010	236.6	1.677	2.91E-03	Uncharacterised protein
Smp_074710	89.7	1.522	0.025	Putative pleckstrin homologue
Smp_159370	402.8	1.424	0.046	Endothelin-converting protein
Smp_151090	871.1	1.379	0.083	Large neutral amino acid transporter
Smp_005360	602.4	1.367	0.069	Uncharacterised protein
Smp_181140	2835.8	1.348	0.095	Uncharacterised protein
Smp_128680	1857.2	1.337	2.91E-03	SN1-specific diacylglycerol lipase β
Smp_158190	574.9	1.269	0.079	Transmembrane protein 184B
Smp_155750	2164.2	1.255	2.91E-03	Non-lysosomal glucosylceramidase
Smp_040720	2188.0	1.228	0.069	Uncharacterised protein
Smp_141380	1057.2	1.227	5.19E-03	Serine/threonine-protein kinase VRK1
Smp_162260	2299.6	1.215	0.027	Phosphatidylinositol-binding clathrin assembly protein unc-11
Smp_128300	1414.6	1.196	0.018	Putative ragulator complex protein LAMTOR2
Smp_008260	2864.2	1.190	0.089	Glycogen synthase kinase-3
Smp_034940	9109.6	1.179	0.068	Uncharacterised protein
Smp_122030	2532.3	1.162	0.095	Endophilin A2
Smp_164210	2893.6	1.151	5.19E-03	Synaptosomal-associated protein
Smp_085530	11478.5	0.886	0.080	Synapse-associated protein
Smp_052810	6968.3	0.871	0.052	Glucose-6-phosphate 1-dehydrogenase
Smp_046600	34020.0	0.819	0.018	Actin
Smp_073940	5863.5	0.813	0.027	Putative flare homologue
Smp_183710	48864.4	0.811	0.012	Actin
Smp_132550	236.0	0.795	0.088	Centrosomal protein POC5
Smp_128860	1315.1	0.784	0.025	Lysyl oxidase homologue
Smp_151800	117.0	0.765	0.038	Uncharacterised protein
Smp_141290	21544.6	0.764	0.079	Innexin unc-9
Smp_068180	388.4	0.763	0.015	Protein lysine oxidase
Smp_037900	788.4	0.756	0.033	β-lactamase-like protein
Smp_089320	4894.3	0.752	0.058	Uncharacterised protein
Smp_152990	162.3	0.724	0.089	Tetraspanin
Smp_103790	1076.8	0.722	5.81E-04	Tubulin α-1A chain
Smp_159090	1206.9	0.707	0.046	Uncharacterised protein
Smp_192110	1065.9	0.706	5.67E-03	Tubulin β
Smp_054160	134550.5	0.702	9.56E-03	Glutathione S-Transferase
Smp_212720	307.8	0.701	0.093	Methylmalonyl-CoA epimerase, mitochondrial precursor
Smp_086480	15779.4	0.701	0.069	Tegument allergen-like protein
Smp_150150	336.3	0.686	3.59E-03	Uncharacterised protein
Smp_015100	345.7	0.663	2.91E-03	Uncharacterised protein
Smp_077310	622.1	0.636	1.76E-03	Tegument allergen-like protein
Smp_157690	325.1	0.635	0.015	Putative zonadhesin
Smp_146610	201.4	0.619	2.91E-03	Uncharacterised protein

Table 4.3: A list of all genes differentially expressed between the RNA-Seq samples of paired and separated males at day 8 after perfusion. The average read counts provide a measure of absolute expression for a given gene. The fold changes reflect expression in separated males relative to paired males. That means fold changes greater than 0 are up-regulated separated males, and those smaller than 0 are down-regulated in separated males. The *p*-value has been adjusted for multiple hypothesis testing.

4.2.6 Time series analysis

Female Time Series Analysis

The pairwise comparison of female RNA-Seq samples at day 8 was noisy and highly enriched for ribosomal and mitochondrial genes. To overcome this problem, I analysed a time-series that combined the data of all female samples over the whole time course of day 0, 4 and 8. The aim was to generate a more accurate list of differentially expressed genes. This approach should allow for a better estimate of variation between biological replicates as well as between treatment groups and yield more statistical power to identify differentially expressed genes.

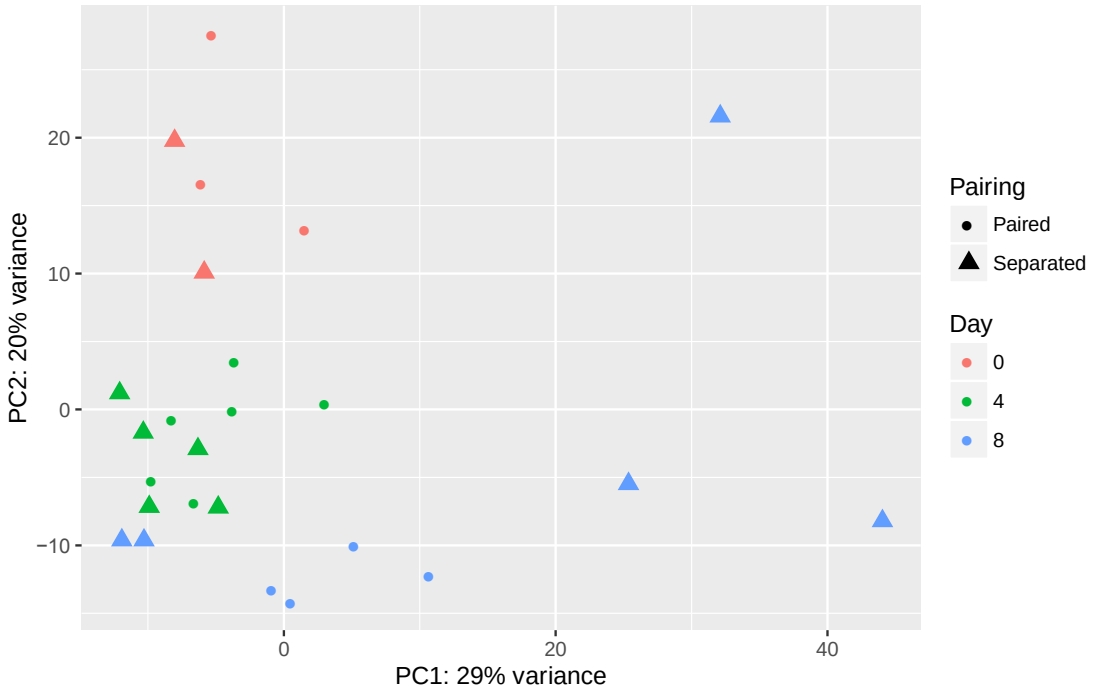


Figure 4.8: In this Principal component analysis plot female samples separate by time, but not pairing status. This is a PCA plot of a time series analysis of female samples at 0, 4 and 8 days after perfusion. The PCA shows high variance between the separated females at day 8 after perfusion but paired and separated samples overlap and do not separate into distinct groups.

By using a likelihood ratio test, genes were identified that were expressed in a pairing specific manner (Table 4.4). Paired and separated females at day 0, where no differences should be observed, were used to control for differences between the groups at the later time points. The day 0 samples are also compared against later time points to identify changes in gene expression over time. Out of 10439 genes with non-zero read counts, 9848 (94%; mean expression cut-off of fewer than 6545 reads) were filtered out by the DESeq2 independent filtering algorithm, 52 were removed as outliers and in total 103 DEGs were identified (adjusted p-value > 0.1), one of which was up-regulated in paired females. The remaining 102 DEGs were up-regulated in separated females (Table 4.4). As in the previously described pair-wise analysis, Smp_049270 (coding for the major egg antigen) was up-regulated in paired females (p-value =

0.017) but the number of genes up-regulated in separated females was 30% smaller compared to the pair-wise comparison (103 compared to 150 DEGs).

Gene ID	Average read count	Fold change	Adjusted p-value	Product
Smp_000280	17110.6	3.16	0.017	Trematode eggshell synthesis domain containing protein
Smp_000390	7664.0	3.36	0.017	Trematode eggshell synthesis domain containing protein
Smp_000420	17283.8	3.18	0.017	Uncharacterised protein
Smp_000430	45222.4	3.61	0.017	Trematode eggshell synthesis domain containing protein
Smp_006040	7682.4	3.56	0.019	Uncharacterised protein
Smp_009230	103686.8	3.60	0.081	Uncharacterised protein
Smp_014610	59748.7	3.37	0.017	P48 eggshell protein
Smp_023840	9114.0	3.76	0.017	Uncharacterised protein
Smp_049270	17664.8	0.267	0.017	Major egg antigen
Smp_054160	88898.5	3.33	0.025	Glutathione S-Transferase
Smp_058690	15222.5	3.02	0.025	Glutathione peroxidase
Smp_071000	6733.1	2.98	0.047	Proteasome subunit
Smp_077890	8954.8	2.97	0.025	Uncharacterised protein
Smp_087760	13520.1	3.06	0.071	Ferritin heavy chain
Smp_090520	7336.1	3.43	0.025	Purine nucleoside phosphorylase
Smp_095980	19016.7	3.84	0.017	Extracellular superoxide dismutase
Smp_131110	108668.3	4.21	0.017	Uncharacterised protein
Smp_138570	7169.2	3.36	0.017	Uncharacterised protein
Smp_173150	7836.9	3.11	0.024	Tetraspanin CD63 antigen
Smp_191910	53312.1	3.24	0.017	Uncharacterised protein
Smp_202690	10689.7	3.60	0.039	Uncharacterised protein
Smp_202770	9022.0	3.00	0.076	Uncharacterised protein
Smp_203090	23497.7	3.60	0.076	Uncharacterised protein
Smp_900040	67182.0	3.63	0.017	Uncharacterised protein
Smp_900100	34224.6	3.60	0.019	NADH dehydrogenase subunit 3

Table 4.4: Table of the 25 most differentially expressed genes identified in the female time series analysis. The average read counts provide a measure of absolute expression for a given gene. The fold changes provided are expression at day 8 in separated females relative to paired females. That means fold changes greater than 0 are up-regulated separated females, and those smaller than 0 are down-regulated in separated females. The *p*-value has been adjusted for multiple hypothesis testing.

The DEGs up-regulated in separated females still included a large number of fertility-related genes, such as genes coding for trematode eggshell synthesis domain containing proteins and an extracellular superoxide dismutase (Table 4.4). Furthermore, many ribosomal and mitochondrial genes were up-regulated in separated females compared to paired females. This resembles the results from earlier comparison of paired and separated females at day 8 (Chapter 4.2.4). The greatest factor influencing the transcriptome of these worms is certainly the time spent *in vitro*, as indicated by the results in Chapter 3.2.11 as well as the PCA plot (Figure 4.4). Together with the great variability in gene expression found in the separated females (see Figure 4.6), this may explain these unexpected results to some extent.

Male Time Series Analysis

The PCA plot (Figure 4.9) revealed that the variability of the biological replicates increased over time. Out of 10336 genes with non-zero read counts, none were filtered out by the DESeq2 independent filtering algorithm, compared to the 94% of genes filtered out in the female time series analysis (94%) as a result of lower overall variability across biological replicates. No genes were identified to be outliers based on the Cook cut-off. Using the same threshold for significance as for the female samples (adjusted p-value < 0.1) 23 genes were significantly up-regulated in separated males and 25 up-regulated in paired males.

Among the up-regulated DEGs in paired males was a genes coding for the serine:threonine kinase VRK1 (Smp_141380, 1.2x change, adjusted p-value = 0.084) (Table 4.5), which is known to be involved in regulating proliferation of

reproductive tissues in schistosomes (Vanderstraete *et al.*, 2014). Also, up-regulated were genes coding for a tyrosine phosphatase receptor (Smp_156870, 1.34x change, adjusted p-value = 0.004) as well as a DPY-30 homologue (Smp_026440, 1.8x change, adjusted p-value = 0.025) (Table 4.5).

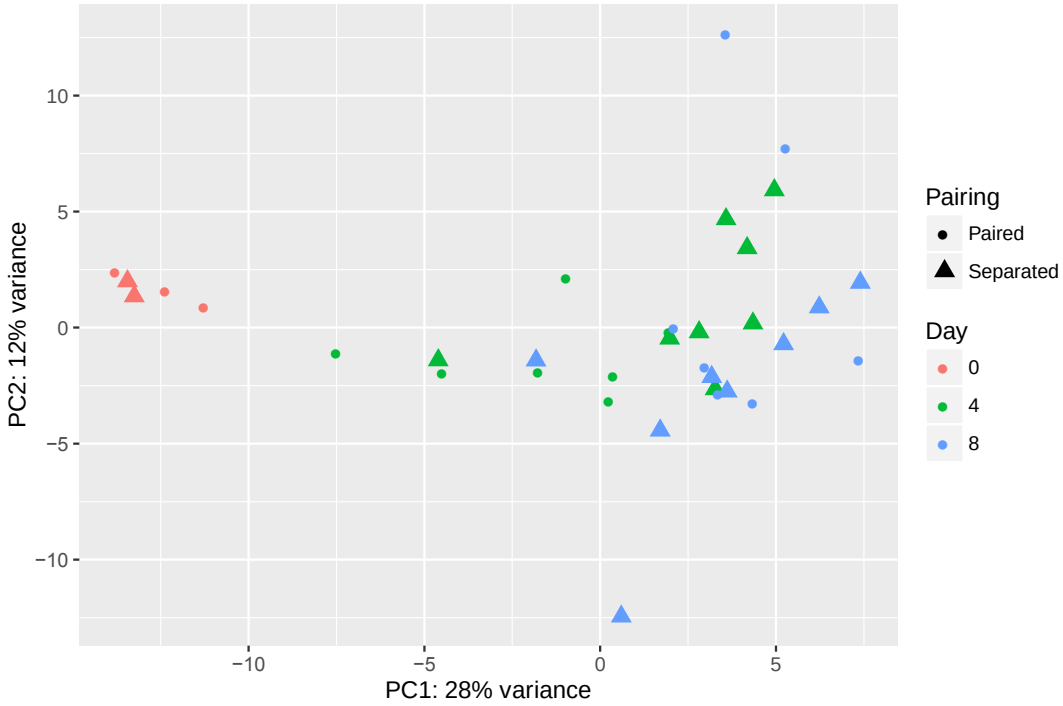


Figure 4.9: In this Principal component analysis male RNA-Seq samples do not separate by pairing status. This is a PCA plot of the time series analysis of male samples at 0, 4 and 8 days after perfusion. The PCA shows more variance between samples at day 4 and especially day 8 when compared to the day 0 samples.

Genes expressed significantly higher in separated males included genes coding for a tetraspanin (Smp_152990, 1.69x change, 0.024) and a spermatogenesis associated protein homologue (Smp_145360, 1.93x change, 0.030) (Table 4.5), a protein that in humans is needed for production of healthy spermatocytes (Bornstein *et al.*, 2011). Furthermore, two signalling related genes were found up-regulated (Table 4.5), coding for a putative glypican homologue

(Smp_245810, 1.6x change, adjusted p-value = 0.030), a group of heparin sulphate proteoglycans that regulate developmental pathways including Wnt (Filmus *et al.*, 2008), and a delta-like protein D homologue (Smp_135370, 3.3x change, adjusted p-value = 0.089), which is a member of the Notch signalling pathway (Gray *et al.*, 1999). The Notch and Wnt signalling pathways both play crucial roles in the regulation of cell fate determination in animal development (Collu *et al.*, 2014).

Gene ID	Average Read Count	Fold change	Adjusted p-value	Product
Smp_077610	70.4	1.33E-03	3.57E-12	Uncharacterised protein
Smp_128680	1850.3	0.84	6.63E-07	Putative SN1-specific diacylglycerol lipase β
Smp_159370	296.7	0.89	1.80E-04	Endothelin converting enzyme 1
Smp_130500	96.6	0.49	4.60E-04	Uncharacterised protein
Smp_155750	1925.7	0.79	3.40E-03	Putative non-lysosomal glucosylceramidase
Smp_156870	467.0	0.74	4.03E-03	Tyrosine phosphatase receptor
Smp_042210	528.5	2.60	1.21E-02	Origin recognition complex subunit
Smp_041700	174.8	0.50	1.21E-02	Uncharacterised protein
Smp_144380	98.9	1.69	2.24E-02	Homeobox protein
Smp_152990	123.0	2.01	2.24E-02	Tetraspanin
Smp_078640	1996.0	0.78	2.24E-02	Like-sM3, small nuclear ribonucleoprotein
Smp_076540	777.6	0.91	2.24E-02	Dynein regulatory complex subunit
Smp_026440	210.5	0.54	2.49E-02	DYP-30 protein
Smp_103790	883.2	1.38	2.56E-02	Tubulin-α chain
Smp_194720	1872.6	2.22	2.56E-02	Uncharacterised protein
Smp_245810	176.2	1.93	2.96E-02	Putative glypican homologue
Smp_145360	349.1	3.04	2.96E-02	Spermatogenesis associated protein
Smp_034940	7930.4	0.72	2.96E-02	Uncharacterised protein
Smp_187190	550.0	0.81	2.96E-02	Innexin unc 9
Smp_212110	99.1	1.14	2.98E-02	Uncharacterised protein
Smp_202120	40.9	0.96	3.78E-02	Hox class homeodomain protein
Smp_125070	1778.4	1.12	4.75E-02	Cystathionine-β synthase
Smp_037900	683.3	1.26	4.75E-02	β-lactamase domain containing protein
Smp_143300	323.3	0.91	5.42E-02	Neurogenic locus Notch homologue
Smp_063500	1304.9	1.11	7.71E-02	Uncharacterised protein

Table 4.5: The 25 most differentially expressed genes of the male time series analysis of paired and separated males at 0, 4 and 8 days after perfusion. The average read counts provide a measure of absolute expression for a given gene. The fold changes are provided for the expression in separated males relative to paired males. That means fold changes greater than 0 are up-regulated separated males, and those smaller than 0 are down-regulated in separated males. The *p*-value has been adjusted for multiple hypothesis testing.

I decided to also include in my thesis an analysis of the transcriptome changes that take place in male and female worms over the course of 8 days, grown *in vitro,* and how these changes relate to sexual regression, nutrient metabolism and other aspects of worm biology affected by the *in vitro* culture. For this

comparison samples of paired females at day 8 were used as they were shown to have relatively less variability across biological replicates compared to separated females (see Figure 4.5).

4.2.7 Comparison of females before and after *in vitro* culture

After 8 days, females in *in vitro* culture differentially expressed hundreds of genes compared to their day 0 counterparts, which were placed in Trizol immediately after perfusion and thus did not experience *in vitro* culture. Out of 10498 genes with non-zero read counts, DESeq2 filtered out 472 (4.5%) genes with a mean expression of fewer than 1.1 reads and 105 outliers were identified using the default Cook cut-off of DESeq2 and subsequently removed from the analysis. This reflects the relatively lower variability in these samples compared to the separated females (D8FS) where 94% of genes were filtered out. Of the remaining genes 696 (6.4%) were significantly down-regulated and 778 (7.2%) up-regulated, after 8 days *in vitro* (adjusted p-value < 0.01) (Fig 4.10).

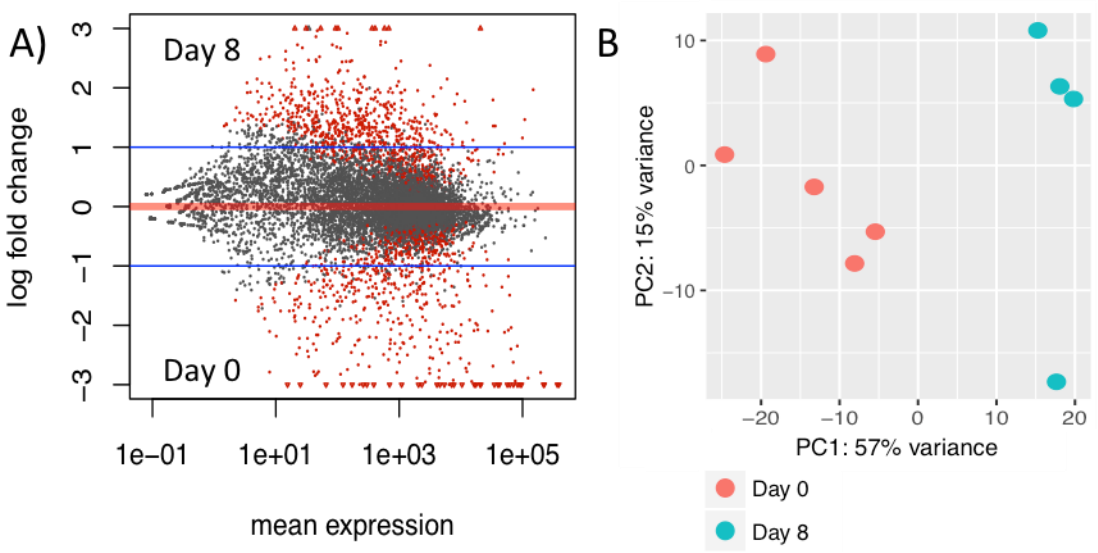


Figure 4.10: *In vitro* culture of *Schistosoma mansoni* induces a great magnitude of changes in gene expression in female worms. Pair-wise comparison of female worms at day 0 and day 8 after perfusion. A) A plot of the log ratio against the mean expression showed a large numbers of genes significantly up-regulated (red) in both groups with large fold changes. Triangles indicate genes outside the boundaries of the plots; like other genes, they are red if they are statistically significant, grey if they are not. B) The principal component analysis shows that females samples separate well into two groups based on principal component 1 (57% of observed variance). These results demonstrate the significant effect of *in vitro* culture on the female transcriptome.

Among the most down-regulated genes were several coding for proteins that are important for digestion, in particular lysosomes, such as saposin-related proteins, a lysosomal Pro-X carboxypeptidase and cathepsin B (Table 4.6). These proteins help to break down macromolecules such as proteins and lipids taken up in autophagic or phagocytic vacuoles to make these nutrients available to the cell metabolism (Luzio *et al.*, 2007). Other down-regulated genes included those coding for the transmembrane protein 60, 2 zinc finger proteins, a nogo B receptor homologue and a parathyroid hormone 2 receptor (Table 4.6).

Both CD63 genes that were found to be significantly up-regulated in mature, paired females in the previous chapter – Smp_173150 & Smp_155310 – were found to be significantly down-regulated in females after *in vitro* culture. Smp_173150 was 5.35x down-regulated (adjusted p-value = 9.21E-11) and Smp_155310 was 6.27x down-regulated (adjusted p-value = 0.00026) (Table 4.6). These results are similar to those from the RNAi experiment in Chapter 3 where *in vitro* culturing was found to lead to a significant down-regulation of the CD63 receptor and antigen transcripts. The lower p-value of Smp_155310 compared to Smp_173150 is the result of lower expression levels of Smp_155310 (~40% of the expression level of Smp_173150; see Table 4.6).

KEGG pathway enrichment analysis showed that the lysosome (smm04142) was the most significantly down-regulated of 14 pathways involving the metabolism of different amino acids, lipids and carbohydrates (Table 4.7). The second most significantly down-regulated pathway was sphingolipid metabolism (smm00600; p-value = 0.002): the genes coding for ceramide synthase (Smp_042440) and ceramide kinase (Smp_130110) were down-regulated 2.54x (adjusted p-value = 1.03E-06) and 2.06x (adjusted p-value = 2.91E-10) respectively (Table 4.6).

Pfam domain enrichment (Table 4.8) revealed a striking down-regulation of genes coding for proteins with trematode eggshell synthesis domains, with 10 out of 11 genes containing such a domain being down-regulated after eight days outside the host (p-value = 2.05E-11), which is consistent with a decline of egg being laid *in vitro* and a reduction in the expression of associated genes that has

been observed before (Galanti *et al.*, 2012). Smp_000430, the only gene of this group not to be down-regulated significantly, fell narrowly short of the significance threshold (adjusted p-value = 0.013) despite a 3.7x reduction of expression (Table 4.6). Across all of the differentially expressed genes found in this comparison of females before and after 8 days in *in vitro* culture, the down-regulation of genes coding for trematode egg shell synthesis domain-containing genes (Table 4.7) is the most compelling evidence that reproductive activity is being shut down in these females.

Gene ID	Average read count	Fold change	Adjusted p-value	Product
Smp_181070	25744.6	11.15	1.24E-35	Uncharacterised protein
Smp_137700	2087.4	10.19	3.24E-19	Zinc finger, C2H2 type domain containing protein
Smp_137720	8610.1	10.15	2.64E-23	Saposin B domain containing protein
Smp_016490	38084.7	10.10	5.52E-35	Saposin B domain containing protein
Smp_146180	1081.7	9.87	1.72E-16	Sterol esterase 2
Smp_139160	29285.6	9.74	3.13E-16	Cathepsin L
Smp_000500	40.4	9.48	2.47E-13	Uncharacterised protein
Smp_153070	2421.8	9.22	7.66E-29	Uncharacterised protein
Smp_136730	17432.3	9.16	1.61E-15	Streptomycin 3-adenylyltransferase 1
Smp_194910	168333.1	9.11	3.81E-23	Saposin B domain containing protein
Smp_130100	56762.1	9.08	1.67E-16	Saposin domain containing protein
Smp_105450	86662.2	8.85	6.73E-18	Saposin domain containing protein
Smp_202610	10354.1	8.34	2.44E-13	Uncharacterised protein
Smp_028870	28335.4	8.30	2.50E-17	Zinc finger, C2H2 type domain containing protein
Smp_139240	16693.6	8.17	1.03E-12	Cathepsin S
Smp_085180	1027.4	7.85	9.75E-13	Cathepsin B
Smp_002600	21358.8	7.67	7.59E-25	Lysosomal Pro X carboxypeptidase
Smp_158150	584.4	7.42	1.98E-07	Reticulocalbin 2
Smp_187140	17722.6	7.42	2.95E-11	Cathepsin L proteinase
Smp_131500	3612.4	6.14	7.47E-20	Transmembrane protein 60
Smp_126340	122.6	6.12	2.37E-07	Uncharacterised protein
Smp_170560	173.9	6.07	4.27E-16	Parathyroid hormone 2 receptor
Smp_155310	1513.5	5.30	2.58E-04	Tetraspanin CD63 receptor
Smp_173150	3910.5	4.84	9.21E-11	CD63 antigen
Smp_038330	1601.7	3.40	6.86E-17	Nogo B receptor
Smp_042440	1457.1	2.68	1.03E-06	Ceramide synthase
Smp_130110	1423.8	2.09	2.91E-10	Ceramide kinase

Table 4.6: A selection of significantly down-regulated genes from the comparison of female *Schistosoma mansoni* at day 8 compared to day 0 *in vitro*. This table contains the 20 most significantly down-regulated genes as well as those specifically mentioned in the results. The average read counts provide a measure of absolute expression for a given gene. The fold changes are the gene expression at day 8 relative to day 0. The *p*-value has been adjusted for multiple hypothesis testing.

KEGG ID	Pathway	Significant Genes	Total Genes	p-value
smm04142	Lysosome	15	47	1.66E-05
smm00600	Sphingolipid metabolism	6	15	1.72E-03
smm00330	Arginine and proline metabolism	6	18	4.69E-03
smm00531	Glycosaminoglycan degradation	2	2	9.64E-03
smm00630	Glyoxylate and dicarboxylate metabolism	4	10	1.04E-02
smm00561	Glycerolipid metabolism	5	17	1.62E-02
smm00920	Sulfur metabolism	3	7	2.19E-02
smm00604	Glycosphingolipid biosynthesis - ganglio series	2	3	2.61E-02
smm00830	Retinol metabolism	2	3	2.61E-02
smm00511	Other glycan degradation	3	8	3.16E-02
smm00480	Glutathione metabolism	4	14	3.30E-02
smm00350	Tyrosine metabolism	3	9	4.27E-02
smm00053	Ascorbate and aldarate metabolism	2	4	4.71E-02
smm00360	Phenylalanine metabolism	2	4	4.71E-02

Table 4.7: KEGG pathways enriched amongst the significantly down-regulated genes in females at day 8 after perfusion compared to females at day 0 females. The column "Significant Genes" provides the number of genes in a given pathway that were down-regulated in day 8 females; "Total Genes" is the total number of genes in the pathway.

Pfam ID	Domain	Significant Genes	Total Genes	p-value
PF08034	Trematode eggshell synthesis protein	10	11	2.05E-11
PF00112	Papain family cysteine protease	9	16	2.09E-07
PF00255	Glutathione peroxidase	3	3	3.15E-04
PF01008	Initiation factor 2 subunit family	3	3	3.15E-04
PF03982	Diacylglycerol acyltransferase	3	3	3.15E-04
PF04261	Dyp-type peroxidase family	3	3	3.15E-04
PF03062	MBOAT, membrane-bound O-acyltransferase family	4	8	1.13E-03
PF08127	Peptidase family C1 propeptide	3	4	1.17E-03
PF00106	Short chain dehydrogenase	7	27	1.45E-03
PF00789	UBX domain	3	5	2.74E-03

Table 4.8: A table of the 10 Pfam domains found to be most significantly enriched among the genes which were down-regulated in females on day 8 after perfusion compared to females at day 0. The column "Significant Genes" provides the number of genes encoding a given domain that were down-regulated in day 8 females; "Total Genes" is the total number of *S. mansoni* genes containing the domain.

Among the most up-regulated DEGs (Table 4.9) were several genes coding for proteins with proteolysis related functions such as E3 ubiquitin protein ligase and the peptidase inhibitor 16. Another gene, coding for an S9 family non-peptidase homologue, was also found up-regulated although, as the name suggests, it does not have peptidase activity. Furthermore, genes coding for a DC STAMP domain protein, otopetrin, and sortilin were up-regulated as well as many cell adhesion related proteins including innexin and protocadherin 1.

An analysis of KEGG pathways showed extracellular matrix (ECM) receptor interactions to be the most significantly up-regulated pathway after 8 days (p-value = 0.00013). The analysis also showed that many genes coding for protein domains associated with the ECM were expressed at higher frequency than expected by chance. This included genes coding for proteins with the following domains: cadherin (p-value = 1.38E-13), laminin N-terminal (p-value = 0.0053), laminin G (p-value = 0.013), annexin (p-value = 0.017), and fibronectin type III (p-value = 0.024). The above pathways are summarised in Table 4.10.

Pathway analysis further showed up-regulation of developmental signalling pathways, such as the Jak-STAT, Wnt, Notch, mTOR, FoxO and Hedgehog signalling pathway as well as neuroactive ligand-receptor interactions after 8 day *in vitro* (Table 4.10). Neural processes are involved in the maintenance of the female germline (Collins *et al.*, 2010) and their differential expression in this experiment may be further evidence for that function. Furthermore, the analysis of Pfam domains showed that genes coding for proteins with several closely related epidermal growth factor (EGF) domains were up-regulated at a higher

frequency than expected by chance. These included a total of 16 genes (Table 4.11). Up-regulation of growth factors was unexpected because these females were thought to be in the process of reducing their reproductive capacity, particularly egg production. Out of the 16 up-regulated genes with EGF domains, eight are involved in extracellular matrix receptor interactions, the pathway most significantly enriched with genes up-regulated in day 8 females (Table 4.10). These protein products of these genes play an important role in tissue remodelling and the regulation of cell survival (Gray *et al.*, 1999). Of the remaining eight genes, five code for neurogenic locus notch proteins, members of the significantly up-regulated notch signalling pathway (Table 4.10). Notch proteins are a group of proteins thought to be essential for maintenance of neuronal progenitor cells as well as neurogenesis (Zhou *et al.*, 2010). In total eight neurogenic locus notch proteins are annotated in the *Schistosoma* genome, five of them are differentially up-regulated after 8 days *in vitro* and the remaining 3 are not differentially expressed (Table 4.9). The Pfam domain enrichment analysis revealed another group of genes with neural related functions called semaphorins to be up-regulated significantly (Smp_159050, 1.89x change, adjusted p-value = 0.00053; Smp_158550, 1.4x change, adjusted p-value = 0.028; Smp_132520, 6.97x change, adjusted p-value = 1.22E-07) (Table 4.9). These genes encode proteins that are either membrane bound or secreted and act as inhibitors of axon growth, thus helping to regulate neurone development (Bashaw & Klein, 2010). In addition to the semaphorin genes a homologue of the SLIT gene was up-regulated in these female worms (Smp_147470, 2.30x change, adjusted p-value = 0.0013) (Table 4.9), which also regulates axon growth in *D. melanogaster* and *H. sapiens* (Bashaw & Klein, 2010).

Two genes coding for homologues of slit-interacting protein (Smp_132540, 3.57x change, adjusted p-value = 2.89E-08; Smp_028030, 1.88x change, adjusted p-value = 7.21E-05) were also up-regulated. Furthermore, the significantly up-regulated Wnt signalling pathway (Table 4.10), is known to help establish a growth axis within developing axons (Mulligan & Cheyette, 2012).

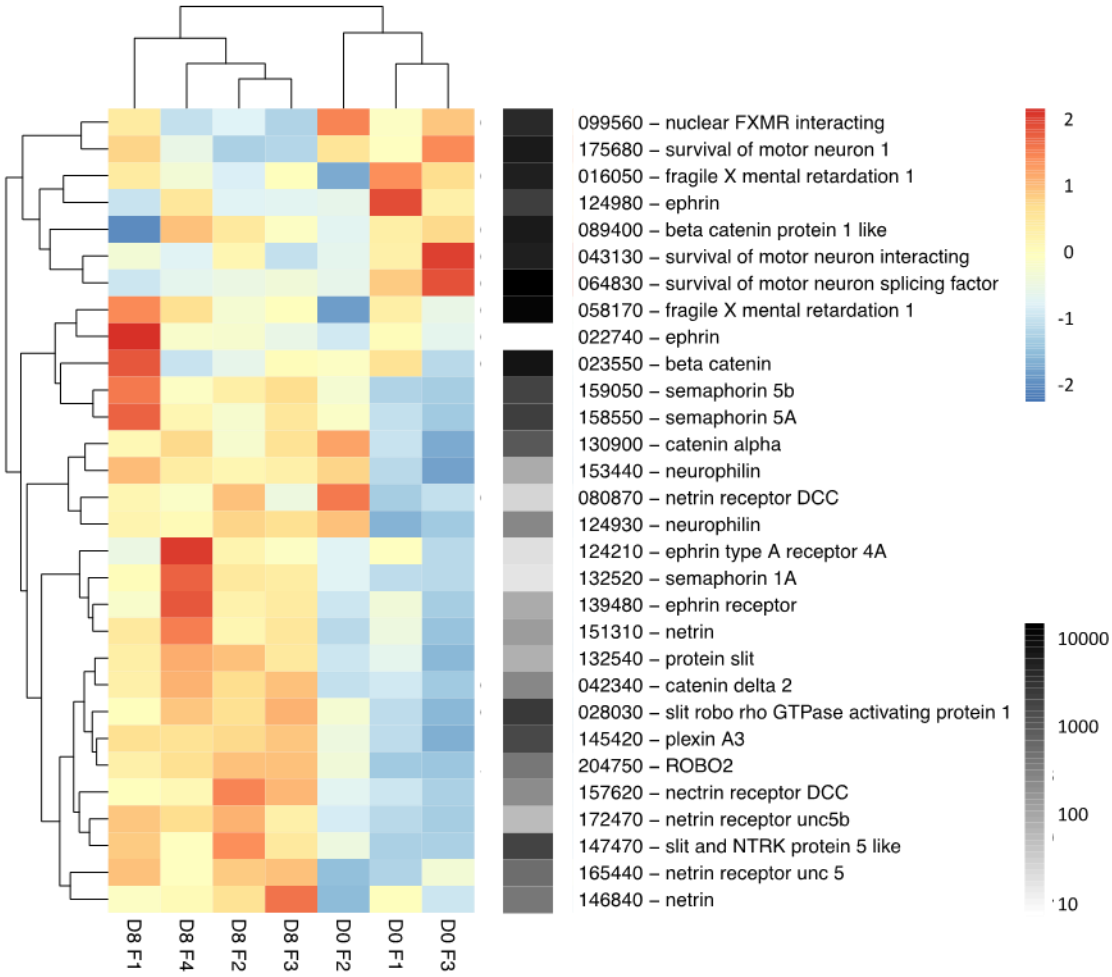


Figure 4.11: Many genes involved in axon development became up-regulated *in vitro*. Heatmap of neurogenesis related genes in female worms at day 0 (three replicates: D0 F1-3) and day 8 (four replicates: D8 F1-4). Colours represent a Z score of log transformed normalised read counts scaled by row, giving relative measures of gene expression in each sample. Average normalised read counts for each gene are in the row on the right on a logarithmic scale as a measure of absolute expression levels.

Several other groups of genes involved in the regulation of axon development were found to be expressed differently at day 8, including genes coding for netrins, netrin receptors, a plexin homologue, ephrins and their receptors, a slit robo homologue and a catenin delta homologue (Fig. 4.11), which play important roles in the guidance, pruning and deletion of axons, in some case regulated by the Wnt pathway (Bashaw & Klein, 2010; Vanderhaeghen & Cheng, 2010). In the heat map of neurogenesis related genes, samples clustered well by treatment but also revealed variability between biological samples (Fig. 4.11). Overall Fig. 4.11 shows a trend of genes coding for proteins involved in the processes of axon guidance and axon pruning to be expressed at higher levels at day 8 (for example ephrin and the ephrin receptors, netrins and their receptors, slit and ROBO as well as semaphorins, plexins and neurophilins (Bashaw & Klein, 2010)). In contrast, several genes known to be critical for neuron survival, such as survival of motor neuron 1 and fragile X mental retardation protein 1 as well as proteins interacting with them can be seen to be down-regulated at day 8 in Fig. 4.11.

Work by Galanti *et al.* (2012) showed that females begin to regress to a sexually immature state once removed from the mammalian host and that the rate of apoptosis increases, resulting in the reduction of reproductive tissue, especially in the vitellarium. A small number of the apoptosis related genes discussed at the beginning of this chapter were found to be differentially expressed in the female worms. Two genes coding for the Bcl-2 domain containing protein (Smp_043360) and the FAS associated factor 2 (Smp_084610) were found to be down-regulated after day 8 (Table 4.9). In contrast, two genes were found to be up-regulated in females after *in vitro* culture (Table 4.9): the genes coded for the

FAS associated factor 1 (Smp_077540), and the FAS binding factor 1 (Smp_148130). However, overall there was no clear pattern of up- or down-regulation of apoptosis related genes, one of the samples of female worms at day 0 did not even cluster with the other day 0 samples (Fig. 4.12).

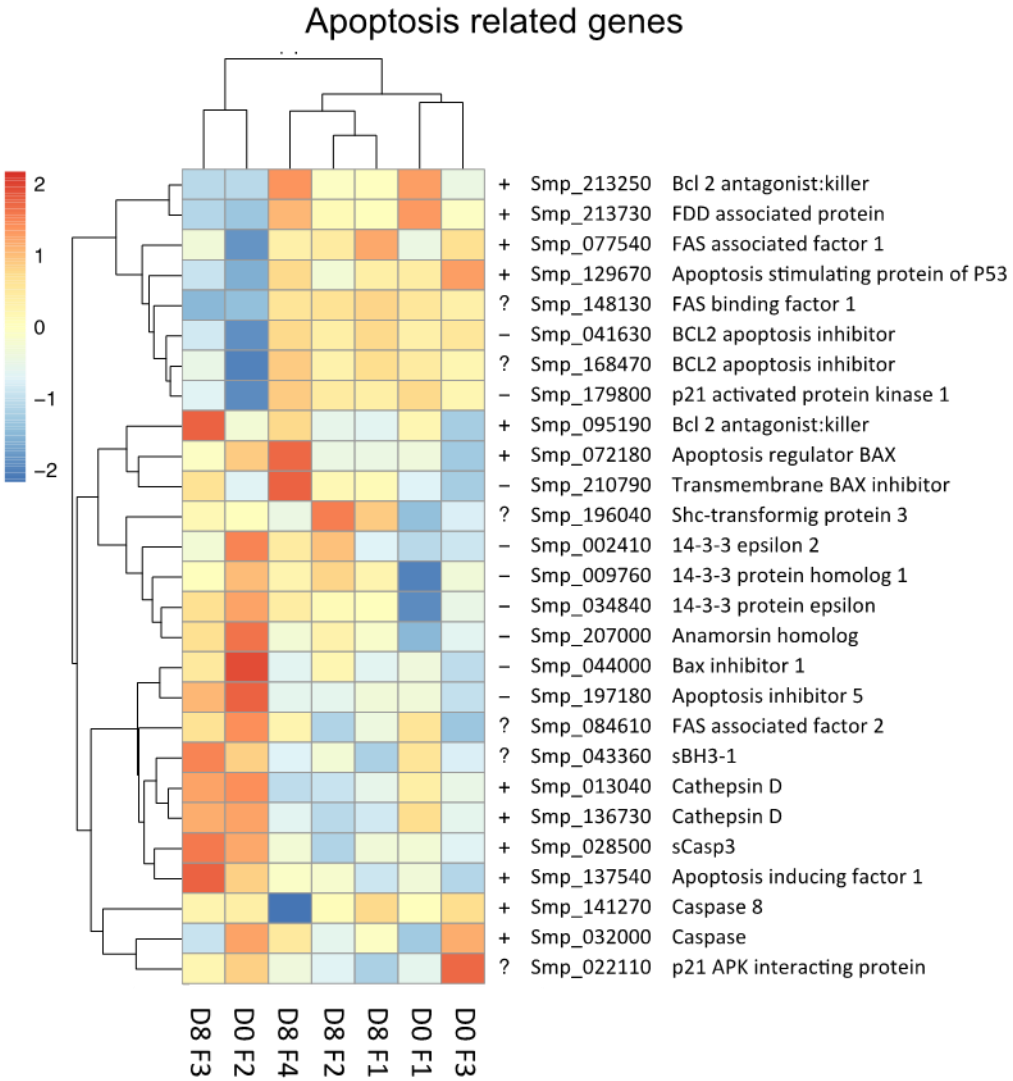


Figure 4.12: No clear trend was found in the expression of pro- and anti-apoptotic genes. Heatmap of apoptosis related genes in females *Schistosoma mansoni* at day 0 (three replicates: D0 F1-3) and day 8 (four replicates: D8 F1-4). Colours represent a Z score of log transformed normalised read counts scaled by row, giving relative measures of gene expression in each sample. The "+" and "-" to the right of the heat map indicate putative pro- and anti-apoptotic function; apoptosis related genes whose exact role could not be determined are marked "?".

Although a heat map of the apoptosis related genes and their expression in females at day 0 and 8 showed some consistent patterns, one of the day 0 females did not cluster with the others nor did a trend for pro- or anti-apoptotic genes to be expressed in either group become apparent (Fig. 4.12).

Gene ID	Average read count	Fold change	Adjusted p-value	Product
Smp_141520	593.9	4.32	1.61E-15	E3 ubiquitin protein ligase PDZRN3
Smp_134210	2229.9	4.21	3.66E-12	Peptidase inhibitor 16
Smp_141510	1025.4	4.98	6.58E-15	Family S9 non peptidase
Smp_152940	1372.9	5.99	6.66E-08	Otopetrin
Smp_162520	707.6	3.78	0.000185833	Protocadherin fat 4
Smp_141290	17226.1	2.72	3.87E-13	Innexin
Smp_179360	5635.2	3.09	5.44E-12	Sortilin
Smp_154790	634.8	12.87	2.47E-24	DC STAMP domain containing protein 2
Smp_105360	258.4	1.78	0.006098043	Neurogenic locus notch protein
Smp_140800	1165.0	2.37	6.34E-06	Neurogenic locus notch protein 2
Smp_128230	6954.8	1.96	0.000729261	Neurogenic locus notch protein
Smp_128220	1619.0	3.29	6.95E-05	Neurogenic locus notch protein
Smp_050520	3444.3	2.77	0.000747004	Neurogenic locus notch protein
Smp_159050	3041.2	1.89	0.000534464	Semaphorin 5B
Smp_158550	2399.4	1.42	0.028126836	Semaphorin 5A
Smp_132520	24.8	6.97	1.22E-07	Semaphorin 1A
Smp_147470	2426.8	2.30	0.001325385	Slit
Smp_132540	275.8	3.57	2.89E-08	Slit interacting protein
Smp_028030	3251.6	1.88	7.21E-05	Slit interacting protein

Table 4.9: Up-regulated genes of females at day 8 after perfusion compared to females at day 0. The average read counts provide a measure of absolute expression for a given gene. The fold changes represent changes in gene expression at day 8 relative to day 0. The *p*-value has been adjusted for multiple hypothesis testing.

KEGG ID	Pathway	Significant Genes	Total Genes	p-value
smm04512	ECM-receptor interaction	4	9	1.25E-04
smm04630	Jak-STAT signalling pathway	4	11	3.08E-04
smm04142	Lysosome	6	47	3.41E-03
smm04310	Wnt signalling pathway	5	35	4.61E-03
smm04330	Notch signalling pathway	3	13	7.47E-03
smm00500	Starch and sucrose metabolism	3	14	9.20E-03
smm04080	Neuroactive ligand-receptor interaction	3	15	1.11E-02
smm00563	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	3	17	1.56E-02
smm04150	mTOR signalling pathway	3	20	2.36E-02
smm02010	ABC transporters	2	8	2.55E-02
smm04068	FoxO signalling pathway	4	38	2.85E-02
smm04340	Hedgehog signalling pathway	2	9	3.17E-02
smm04144	Endocytosis	5	64	4.22E-02

Table 4.10: KEGG pathways significantly enriched amongst the genes up-regulated in females at day 8 after perfusion compared to females a day 0. The column "Significant Genes" provides the number of genes in a given pathway that were up-regulated in day 8 females; "Total Genes" is the total number of genes in the pathway.

Pfam ID	Domain	Significant Genes	Total Genes	p-value
PF00028	Cadherin domain	23	45	1.38E-13
PF00008	EGF-like domain	8	15	1.01E-05
PF07645	Calcium-binding EGF domain	5	7	8.24E-05
PF12661	Human growth factor-like EGF	5	7	8.24E-05
PF07679	Immunoglobulin I-set domain	12	45	2.36E-04
PF01414	Delta serrate ligand (Notch ligands)	4	6	6.87E-04
PF00053	Laminin EGF-like (Domains III and V)	4	7	1.46E-03
PF00503	G-protein α subunit	5	12	1.99E-03
PF00864	ATP P2X receptor	3	4	2.33E-03
PF01403	Sema domain	3	4	2.33E-03

Table 4.11: Pfam domains which were significantly enriched amongst the genes up-regulated in females at day 8 after perfusion compared to females at day 0. The column "Significant Genes" provides the number of genes encoding a given domain that were up-regulated in day 8 females; "Total Genes" is the total number of *Schistosoma mansoni* genes encoding the domain.

4.2.8 Comparison of males before and after *in vitro* culture

The comparison of day 0 and day 8 was performed for male worms as well, revealing transcriptome changes of a similar magnitude as in the female comparison. After removing 962 (9.2%) genes with low read counts and 105 (1%) outliers it was found that 493 (4.6%) genes were down-regulated and 717 (6.6%) genes up-regulated after 8 days in culture (adjusted p-value < 0.01) (Figure 4.13 A). As the MA plot revealed, the fold changes observed in males were relatively small compared to those in females (compare Figure 4.10 A and 4.13 A). The first two principal components, shown in the PCA plot, explain 72% of variability measured across the samples (Figure 4.13 B).

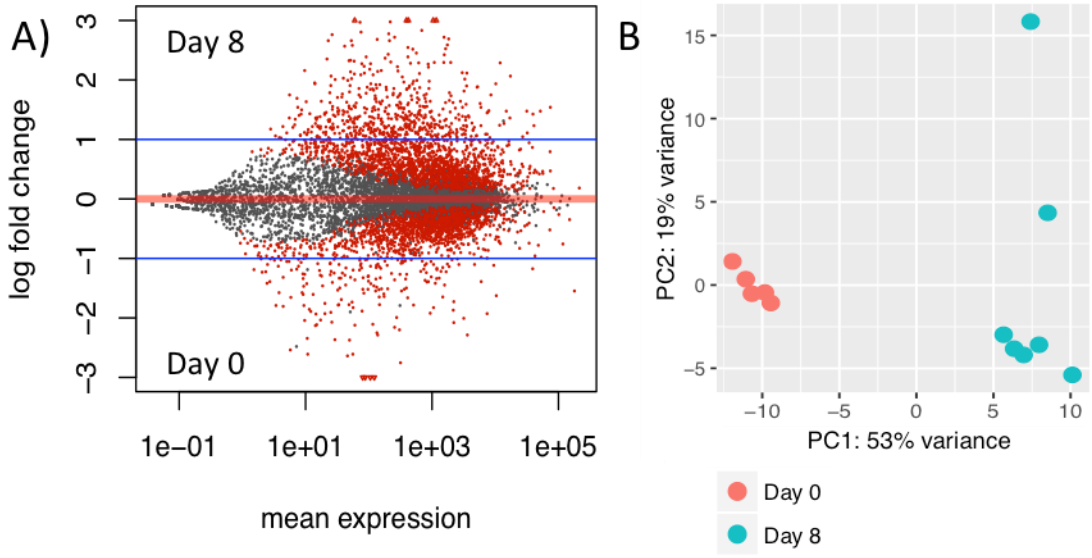


Figure 4.13: *In vitro* culture induced large changes in gene expression in male worms. Pair-wise comparison of male worms at day 0 and day 8 after perfusion. A) The log ratio against mean (MA) plot shows large numbers of genes significantly up-regulated (red) in both groups with large fold changes. Triangles indicate genes outside the boundaries of the plots; like other genes, they are red if they are statistically significant, grey if they are not. B) Principal component analysis shows that males samples also separate well into two groups based on principal component 1 (53% of observed variance). These results demonstrate a significant effect of *in vitro* culture on the male transcriptome.

Of the 717 genes up-regulated at day 8, the most significantly differentially expressed gene coded for a homologue of the interferon-related developmental regulator (Smp_004680; 2.13x change; adjusted p-value = 4.60E-17) (Table 4.12). Also differentially expressed were genes coding for a phenylalanyl-tRNA synthase β (Smp_094060; 1.41x change; adjusted p-value = 4.69E-17), a heat shock 70 protein (Smp_106930; 2.91x change; adjusted p-value = 4.21E-14)) and a γ-secretase subunit APH (Smp_050110;1.83x change; adjusted p-value = 4.31E-13) as well as many proteins with unknown function (Table 4.12).

After 8 days four pathways were significantly down-regulated in male worms (p-value < 0.05) compared to males at day 0 (Table 4.13). Down-regulation of RNA transport (smm03013; p = 0.0049), mRNA surveillance (smm03015; p = 0.028), and protein processing in endoplasmic reticulum (smm04141; p = 0.039) suggested a down-regulation of transcription. Also down-regulated was N-Glycan biosynthesis (smm00510; p = 0.034), which plays a role in the post-translation modification of glycoproteins (Bieberich, 2014). Furthermore, the Pfam domain enrichment analysis (Table 4.14) suggested that in addition to RNA processing, many genes coding for proteins with DNA replication and cell cycle control related function were down-regulated after 8 days outside the host. These relevant domains include MIF4G domain (PF02854; p = 0.0010), which forms part of translation initiation factors (Virgili *et al.*, 2013); Rad17 cell cycle checkpoint domain (PF03215; p = 0.0024), involved in cell cycle control (Post *et al.*, 2003); and BRCA1 C Terminus domain (PF00533; p = 0.0072), which occurs in proteins that help to regulate the cell cycle in response to DNA damage (Deng, 2006). Also, significantly down-regulated were genes coding for several heat

shock protein associated domains, including Hsp20 crystalline family (PF00011; p = 1.69E-07), DnaJ central domain (PF00684; p = 0.0024), DnaJ C terminal domain (PF01556; p = 0.013) and Hsp70 domain (PF00012; p = 0.039). Finally, GO term enrichment analysis showed, among others (Table 4.15), DNA replication (GO:0006260, p-value = 5.30E-05), cell division (GO:0051301, p-value = 0.0020), meiotic nuclear division (GO:0007126, p-value = 0.0056) and cell cycle (GO:0006269, p-value = 0.012) down-regulated, suggestion a reduction in sperm production as well as somatic cell division.

Gene ID	Average Read Count	Fold Change	p-value	Product
Smp_196900	92.7	0.087	6.63E-11	Uncharacterised protein
Smp_144740	173.9	0.112	1.42E-11	Uncharacterised protein
Smp_144730	166.2	0.115	6.77E-13	DnaJ protein
Smp_106930	162521.3	0.344	4.21E-14	Heat shock protein 70
Smp_124030	893.9	0.386	2.50E-11	Uncharacterised protein
Smp_004680	8356.1	0.469	4.60E-17	Interferon-related developmental regulator
Smp_168130	727.3	0.471	4.25E-11	Phosphatase and actin regulator 1
Smp_176490	693.6	0.491	4.60E-16	Uncharacterised protein
Smp_050110	3076.6	0.546	4.31E-13	γ-Secretase subunit APH
Smp_094060	3055.4	0.710	4.69E-17	Phenylalanyl-tRNA synthetase β

Table 4.12: The 10 genes most significantly down-regulated in males at day 8 after perfusion compared to males at day 0. Average read counts provide a measure of absolute expression for a given gene. Fold changes represent gene expression at day 8 after perfusion relative to day 0. The *p*-value has been adjusted for multiple hypothesis testing.

KEGG ID	Pathway	Significant Genes	Total Genes	p-value
smm03013	RNA transport	14	103	4.86E-03
smm03015	mRNA surveillance pathway	8	60	2.83E-02
smm00510	N-Glycan biosynthesis	5	30	3.43E-02
smm04141	Protein processing in endoplasmic reticulum	10	89	3.90E-02

Table 4.13: KEGG pathways significantly enriched amongst the genes down-regulated genes in males at day 8 after perfusion compared to males at day 0. The column "Significant Genes" provides the number of genes in a given pathway that were down-regulated in day 8 males; "Total Genes" is the total number of genes in the pathway.

Pfam ID	Domain	Significant Genes	Total Genes	p-value
PF00011	Hsp20 crystallin family	7	11	1.69E-07
PF00400	WD domain, G-β repeat	18	141	1.09E-04
PF02854	MIF4G domain	3	5	1.04E-03
PF00684	DnaJ central domain	2	2	2.37E-03
PF03215	Rad17 cell cycle checkpoint protein	2	2	2.37E-03
PF06584	DIRP	2	2	2.37E-03
PF06682	Protein of unknown function (DUF1183)	2	2	2.37E-03
PF09668	Aspartyl protease	2	2	2.37E-03
PF11461	Rab interacting lysosomal protein	2	2	2.37E-03
PF13481	AAA domain	2	2	2.37E-03

Table 4.14: Pfam domains enriched amongst the significantly down-regulated genes in males at day 8 after perfusion compared to males at day 0. The column "Significant Genes" provides the number of genes encoding a given domain that were down-regulated in day 8 males; "Total Genes" is the total number of *Schistosoma mansoni* genes encoding the domain.

Term	Description	Significant Genes	Expected Genes	p-value
GO:0006260	DNA replication	161	21	5.30E-05
GO:0009408	Response to heat	26	6	0.0014
GO:0051301	Cell division	123	15	0.002
GO:0007126	Meiotic nuclear division	8	3	0.0056
GO:0016070	RNA metabolic process	1303	76	0.0071
GO:0006457	Protein folding	85	10	0.0088
GO:0007049	Cell cycle	207	24	0.0098
GO:0000184	Nuclear-transcribed mRNA catabolic process	39	6	0.0115
GO:0007067	Mitotic nuclear division	88	9	0.0249
GO:0006269	DNA replication, synthesis of RNA primer	6	2	0.0321
GO:0031119	tRNA pseudouridine synthesis	6	2	0.0321
GO:0006506	GPI anchor biosynthetic process	25	4	0.0327
GO:0006879	Cellular iron ion homeostasis	7	2	0.0435
GO:0016485	Protein processing	17	3	0.0488
GO:0051168	Nuclear export	7	2	0.0494

Table 4.15: GO terms which were significantly enriched amongst the down-regulated genes in males at day 8 after perfusion compared to males at day 0. The column "Significant Genes" provides the number of genes associated with a given GO term that were down-regulated in males at day 8; "Expected Genes" provides the number of up-regulated genes annotated with the GO term as expected by chance. The GO terms are all in the "Biological Process" category.

The most up-regulated genes (Table 4.16) included several genes coding for proteins with neural function. This included a netrin homologue (Smp_146840, 5.62x change, adjusted p-value = 1.67E-20), and a catenin delta-2 homologue (Smp_042340, 3.04x change, adjusted p-value = 1.64E-13), both of which are thought to play important roles in neural development (Bashaw & Klein, 2010). Also up-regulated were genes coding for a stomatin homologue (Smp_003440, 5.72x change, adjusted p-value = 1.97E-16), a neurogenic locus Notch protein homologue (Smp_143300, 7.58x change, adjusted p-value 2.12E-16) and a calcium binding protein that confers elasticity to the ECM (Handford, 2000). Another up-regulated gene codes for a RNA polymerase associated factor 1

homologue (Smp_045940, 2.15x change, 2.52E-13), which is a component of the PAF1 complex that is thought to regulate development and maintenance of embryonic stem cell pluripotency in humans (Ding *et al.*, 2009). Also, among the ten most differentially expressed genes are three genes with unknown function (Table 4.16).

Gene ID	Mean Read Count	Fold change	Adjusted p-value	Product
Smp_175390	1144.8	9.19	2.07E-32	Uncharacterised protein
Smp_137300	51.7	8.54	1.96E-17	Amine GPCR
Smp_143300	220.8	7.58	2.12E-16	Fibrillin 1
Smp_084780	540.2	6.98	6.47E-15	Uncharacterised protein
Smp_003440	1455.0	5.73	1.97E-16	Stomatin
Smp_146840	584.5	5.62	1.67E-20	Netrin
Smp_162500	341.6	4.31	2.52E-13	Drug efflux protein
Smp_042340	783.0	3.05	1.64E-13	Catenin delta 2
Smp_045940	7329.6	2.16	2.52E-13	RNA polymerase II associated factor 1
Smp_060080	2182.4	1.81	3.56E-14	Uncharacterised protein
Smp_140800	1165.0	1.63	6.48E-03	Neurogenic locus notch protein
Smp_155050	3126.2	2.27	1.83E-05	Agrin
Smp_123860	798.8	2.75	7.61E-04	Neurexin-2 homologue
Smp_157620	553.2	3.26	3.24E-09	Netrin receptor DDC
Smp_172470	90.6	2.25	8.39E-05	Netrin receptor
Smp_165440	1102.1	2.12	4.57E-08	Netrin receptor unc 5
Smp_132540	275.8	1.97	1.43E-03	Slit
Smp_159050	3041.2	2.09	2.98E-06	Semaphorin

Table 4.16: Selection of the genes significantly up-regulated in males at day 8 after perfusion compared to males at day 0. Average read counts provide a measure of absolute expression for a given gene. Fold changes represent the gene expression in males at day 8 after perfusion relative to day 0. The *p*-value has been adjusted for multiple hypothesis testing.

After 8 days *in vitro* the up-regulated genes were enriched for pathways such as citrate cycle (smm00020; p = 0.00086), neuroactive ligand-receptor interaction (smm04080; p = 0.0044), wnt signalling pathway (smm04310; p = 0.0049), ECM-receptor interaction (smm04512; p = 0.0072), as well as two amino acid metabolism pathways (smm00330; p = 0.0084 & smm00250; p = 0.024) (Table 4.19). Three out of the six significantly up-regulated pathways in male worms were also up-regulated in female worms after 8 days: the neuroactive ligand-receptor interaction, the wnt signalling pathway and the ECM-receptor interaction (Tables 4.10 and 4.19).

After 8 days in culture, the up-regulated DEGs were enriched with genes coding for domains such as the neurotransmitter-gated ion-channel ligand binding domain (PF02931; p = 0.00021), the neurotransmitter-gated ion-channel transmembrane region (PF02932; p = 0.00068) and several EGF-related domains (as observed in the female worms; Table 4.11), including the calcium-binding EGF domain (PF07645; p = 0.0016), EGF-like domain (PF00008; p = 0.0064) and laminin EGF-like (Domains III and V) (PF00053; p = 0.017), in total ten genes with EGF domains (Table 4.17). Many of the genes up-regulated in males were involved in neural processes (Table 4.16). The annotation of the following genes was confirmed by checking for relevant PFAM domains as well as significant BLAST hits (data not shown here). They include a genes coding for a neurogenic locus notch protein (Smp_140800, 1.63x change, adjusted p-value = 0.0065), thought to be involved in neural cell progenitor maintenance (Louvi & Artavanis-Tsakonas, 2006), two homologues of netrin (Smp_146840, 5.62x change, adjusted p-value = 3.53E-24; Smp_151310, 3.37x change, adjusted p-

value = 5.01E-05), a gene involved axon guidance (Bashaw & Klein, 2010), and agrin (Smp_155050, 2.27x change, adjusted p-value = 1.83E-05), a large proteoglycan thought to play a role in the development of neuromuscular junctions (Bolliger *et al.*, 2010) and finally genes coding for a homolog of neurexin-2 (Smp_123860, 2.75x change, adjusted p-value = 0.00076) and a neuronal cell surface protein that mediates the assembly of presynaptic terminals (Dean *et al.*, 2003). In addition to the genes of both netrin homologues in the *S. mansoni* genome being up-regulated in male worms after 8 days, all three genes coding for homologues of the netrin receptor were also up-regulated in these worms (Smp_157620, 3.26x change, adjusted p-value = 3.24E-09; Smp_172470, 2.25x change, adjusted p-value = 8.39E-05; Smp_165440, 2.12x change, adjusted p-value = 4.57E-08). One other gene annotated as a netrin receptor homologue on GeneDB, Smp_080870, was found not to contain any conserved domains nor to have BLAST hits to relevant proteins and was excluded from this analysis. The gene coding for slit (Smp_132540, 1.97x change, p = 0.0014), another regulator of axon guidance (Bashaw & Klein, 2010), was also up-regulated in the male worms after 8 days, as was the gene of a semaphorin homologue (Smp_159050, 2.08x change, adjusted p-value = 2.98E-06) (Table 4.16).

GO term enrichment (Table 4.18) showed GO terms related to cell adhesion to be most significantly enriched (G0:0007155, p-value = 1.90E-17; GO:0007156, p-value = 2.00E-13), but also several development and regulatory GO terms to be up-regulated. These include G-protein coupled receptor signalling (GO:0007186, p-value = 5.20E-06), multicellular organismal development (GO:0007275, p-

value = 0.00031), negative regulation of cell differentiation (GO:0045596, p-value = 0.00043) and negative regulation of BMP signalling (GO:0030514, p-value = 0.00043) (Table 4.18). Furthermore, two pathways implicated in the development and growth of axons were found up-regulated. The first, Wnt signalling (GO:0016055, p-value = 0.0175), helps to establish the body's axes in developing organisms including *Schmidtea mediterranea* (Gurley *et al.*, 2010). Also up-regulated was gamma-aminobutyric acid (GABA) signalling (GO:0007214, p-value = 0.0162), which has been shown to play a role in axon pruning in mice (Wu *et al.*, 2012). The heat map of neurogenesis related genes (Figure 4.14) contains genes identified from the differential expression analysis as well as from the literature by including differentially expressed homologues of genes discussed by Bashaw & Klein (2010) and Vanderhaeghen & Cheng (2010). It showed that both male and female worms exhibited similar trends in gene expression (Figure 4.11). Genes coding for netrin and netrin receptor genes, as well as slit, ephrin receptor, semaphorin and plexin genes were up-regulated after 8 days *in vitro*. On the other hand, neuron survival genes such as genes coding for FXMR, SMN as well as most catenin genes were not differentially expressed, or even slightly down-regulated (Figure 4.14). Although this overall trend held true, and samples clustered well by treatment, variability within biological replicates could be clearly seen.

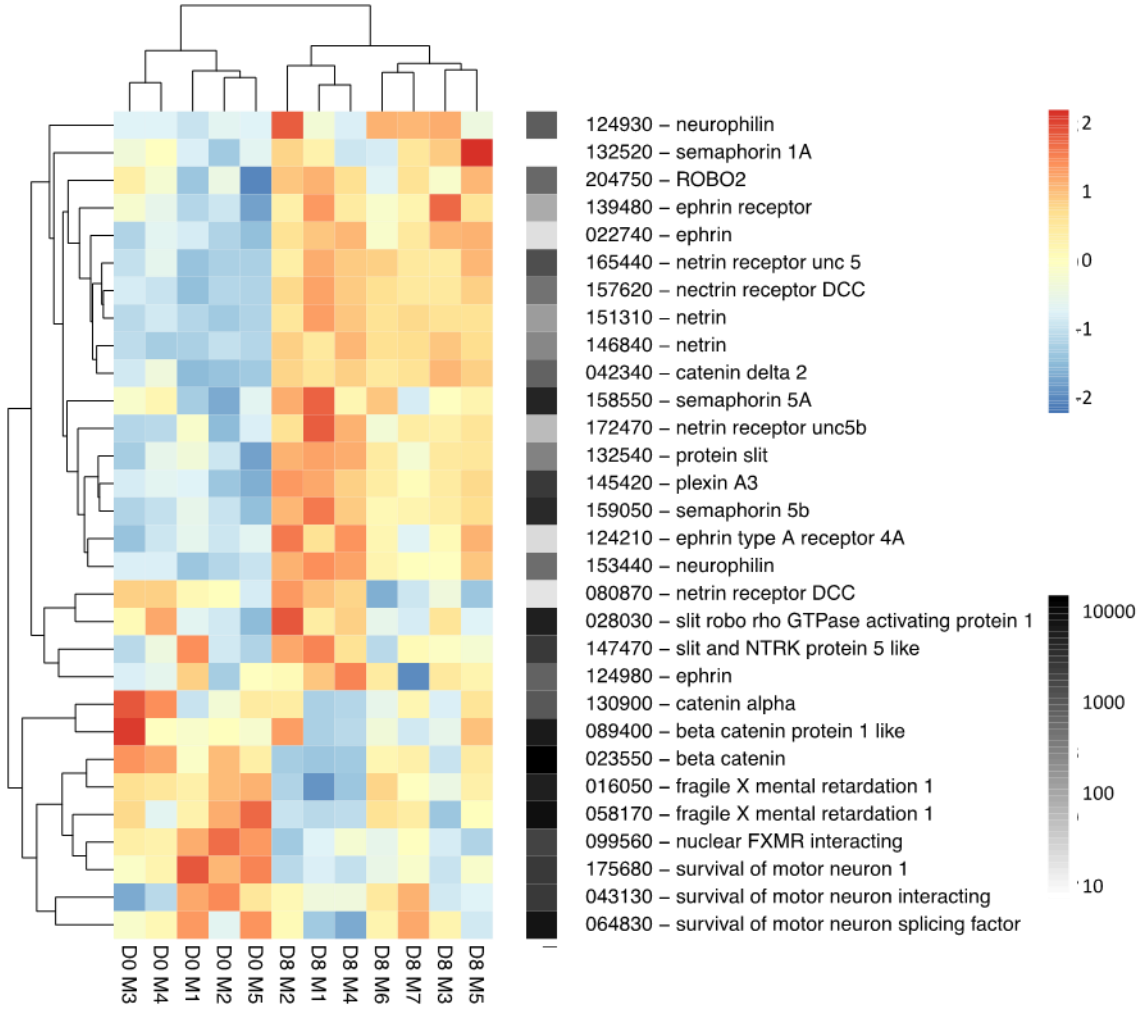


Figure 4.14: Axon proning-related genes became up-regulated *in vitro*. Heatmap of neurogenesis related genes in males worms at day 0 (five replicates: D0 M1-5) and day 8 (seven replicates: D8 M1-7). Colours represent a Z score of log transformed normalised read counts scaled by row, giving relative measures of gene expression in each sample. Average normalised read counts for each gene are in the row on the right on a logarithmic scale as a measure of absolute expression levels.

Unlike in female worms, no large changes in the set of apoptosis related genes were expected, as male worms are generally thought to remain sexually mature in the absence of a partner (Basch, 1991). The heat map of apoptosis related genes in male worms (see Figure 4.15) suggested that only small changes in the transcription of these genes took place between day 0 and 8 in culture, but no clear pattern of differential expression of pro- or anti-apoptotic genes emerged.

In total, nine out of 27 genes were differentially expressed (Table 4.20). This included four genes – two pro-apoptotic and two with unknown effect – down-regulated after 8 days and five genes up-regulated – three anti-apoptotic and two pro-apoptotic genes.

However, the fold changes were relatively small (Table 4.20). These data suggested that there is regulation of apoptosis at the transcriptome level. However, it did not conclusively indicate apoptosis being up-regulated during *in vitro* culture. Apoptosis after 8 days was expected to take place at a lower rate compared with females due to regression of the reproductive tissues taking place in females (Galanti *et al.*, 2012). However, due to high levels of variability, increased apoptotic activity could not be demonstrated for female worms.

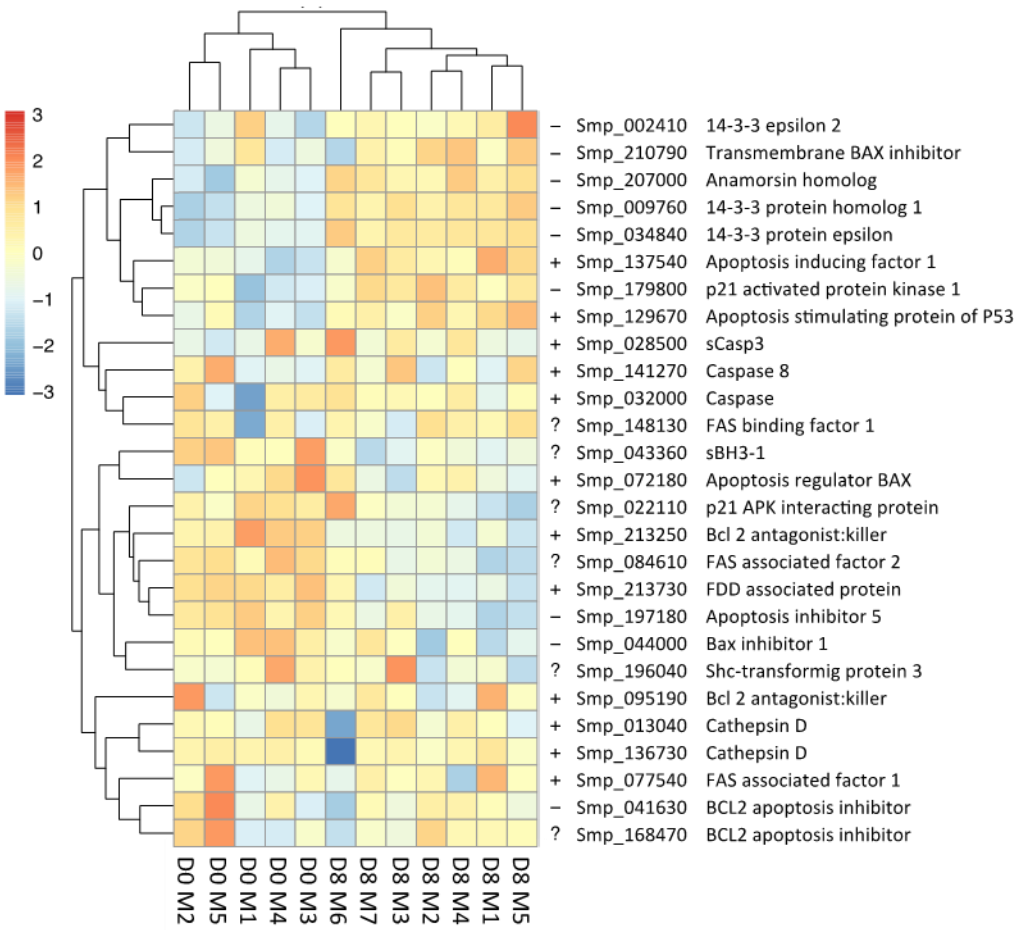


Figure 4.15: Expression of apoptosis-related genes does not show clear trends before and after *in vitro* cultrue. No Heatmap of apoptosis related genes in males worms at day 0 (five replicates: D0 M1-5) and day 8 (seven replicates: D8 M1-7). Colours represent a Z score of log transformed normalised read counts scaled by row, giving relative measures of gene expression in each sample. The "+" and "-" to the right of the heat map indicate putative pro- and anti-apoptotic function; apoptosis related genes whose exact role could not be determined are marked "?".

Pfam ID	Domain	Significant Genes	Total Genes	p-value
PF00028	Cadherin domain	21	45	2.49E-11
PF07679	Immunoglobulin I-set domain	18	45	1.33E-08
PF01410	Fibrillar collagen C-terminal domain	5	5	5.35E-06
PF00041	Fibronectin type III domain	11	29	1.61E-05
PF01391	Collagen triple helix repeat (20 copies)	7	12	2.06E-05
PF13895	Immunoglobulin domain	10	27	4.93E-05
PF00027	Cyclic nucleotide-binding domain	10	28	7.00E-05
PF00520	Ion transport protein	12	43	1.91E-04
PF02931	Neurotransmitter-gated ion-channel ligand binding domain	7	16	2.07E-04
PF00858	Amiloride-sensitive sodium channel	6	12	2.50E-04

Table 4.17: Pfam domains significantly enriched amongst the up-regulated genes in males at day 8 after perfusion compared to males at day 0. The column "Significant Genes" provides the number of genes coding for a given domain that were up-regulated in males at day 8; "Total Genes" is the total number of *Schistosoma mansoni* genes encoding the domain.

GO Term	Description	Significant Gene	Expected Genes	p-value
GO:0007155	Cell adhesion	226	73	1.90E-17
GO:0007156	Homophilic cell adhesion via plasma membrane	60	25	2.00E-13
GO:0006817	Phosphate ion transport	19	12	9.20E-10
GO:0007165	Signal transduction	547	91	2.60E-08
GO:0006811	Ion transport	389	75	9.20E-07
GO:0007186	G-protein coupled receptor signalling pathway	161	32	5.20E-06
GO:0006813	Potassium ion transport	77	17	4.30E-05
GO:0051216	Cartilage development	5	4	0.00015
GO:0006814	Sodium ion transport	64	14	0.00023
GO:0007275	Multicellular organismal development	286	41	0.00031
GO:0045596	Negative regulation of cell differentiation	3	3	0.00043
GO:0030514	Negative regulation of BMP signalling pathway	3	3	0.00043
GO:0050896	Response to stimulus	808	106	0.00112
GO:0010646	Regulation of cell communication	77	8	0.00568
GO:0023051	Regulation of signalling	77	8	0.00568
GO:0006163	Purine nucleotide metabolic process	101	10	0.0119
GO:0055114	Oxidation-reduction process	127	18	0.0159
GO:0006855	Drug transmembrane transport	3	2	0.01621
GO:0009098	Leucine biosynthetic process	3	2	0.01621
GO:0007214	γ-aminobutyric acid signalling pathway	3	2	0.01621
GO:0016055	Wnt signalling pathway	27	7	0.01749
GO:0006821	Chloride transport	8	3	0.01797
GO:0030154	Cell differentiation	78	14	0.02305
GO:0006108	Malate metabolic process	4	2	0.03081
GO:0005977	Glycogen metabolic process	24	4	0.0447

Table 4.18: GO terms which were significantly enriched amongst the genes up-regulated in males at day 8 after perfusion compared to males at day 0. The column "Significant Genes" provides the number of genes associated with a given GO term that were up-regulated in males at day 8; "Expected Genes" provides the number of up-regulated genes with the GO term as expected by chance. The GO terms are all in the "Biological Process" category.

KEGG ID	Pathway	Significant Genes	Total Genes	p-value
smm00020	Citrate cycle (TCA cycle)	6	25	8.62E-04
smm04080	Neuroactive ligand-receptor interaction	4	15	4.36E-03
smm04310	Wnt signalling pathway	6	35	4.89E-03
smm04512	ECM-receptor interaction	3	9	7.16E-03
smm00330	Arginine and proline metabolism	4	18	8.45E-03
smm00250	Alanine, aspartate and glutamate metabolism	3	14	2.43E-02

Table 4.19: KEGG pathways which were significantly enriched amongst the gene up-regulated in males at day 8 after perfusion compared to males at day 0. The column "Significant Genes" provides the number of genes in a given pathway that were up-regulated in males at day 8; "Total Genes" is the total number of genes in the pathway.

Gene ID	Average read count	Fold change	Adjusted p-value	Product
Smp_043360	97.2	1.64	0.022	sBH3-1
Smp_213730	1741.5	1.38	0.010	FDD associated protein
Smp_136730	17432.3	1.21	0.739	Cathepsin D
Smp_213250	3527.9	1.19	0.003	Bcl 2 antagonist:killer
Smp_084610	2237.6	1.18	0.002	FAS associated factor 2
Smp_197180	3314.5	1.14	0.053	Apoptosis inhibitor 5
Smp_041630	708.6	1.07	0.753	BCL2 apoptosis inhibitor
Smp_013040	45112.2	1.06	0.834	Cathepsin D
Smp_022110	2100.3	1.06	0.763	p21 APK interacting protein
Smp_044000	27649.2	1.05	0.713	Bax inhibitor 1
Smp_072180	2661.2	1.05	0.752	Apoptosis regulator BAX
Smp_168470	421.4	1.02	0.942	BCL2 apoptosis inhibitor
Smp_196040	4717.2	1.01	0.918	Shc-transformig protein 3
Smp_077540	2452.7	1.00	0.981	FAS associated factor 1
Smp_095190	512.4	0.99	0.961	Bcl 2 antagonist:killer
Smp_032000	555.9	0.95	0.713	Caspase
Smp_028500	395.6	0.94	0.666	sCasp3
Smp_141270	431.4	0.92	0.900	Caspase 8
Smp_148130	686.1	0.92	0.765	FAS binding factor 1
Smp_002410	12906.3	0.90	0.446	14-3-3 epsilon 2
Smp_210790	4704.8	0.88	0.367	Transmembrane BAX inhibitor
Smp_179800	824.4	0.86	0.465	p21 activated protein kinase 1
Smp_137540	2569.7	0.84	0.011	Apoptosis inducing factor 1
Smp_207000	1143.3	0.75	0.028	Anamorsin homolog
Smp_009760	50176.7	0.73	0.073	14-3-3 protein homolog 1
Smp_034840	44878.7	0.72	0.001	14-3-3 protein epsilon
Smp_129670	966.7	0.69	0.071	Apoptosis stimulating protein of P53

Table 4.20: Table of apoptosis-related gene expression in the comparison of male *Schistosoma mansoni* after 8 days *in vitro*. Average read counts provide a measure of absolute expression for a given gene. Fold changes express the gene expression at day 8 relative to day 0. The *p*-value has been adjusted for multiple hypothesis testing.

4.3 Discussion

Pairing-specific changes in gene expression are obscured by the effect of *in vitro* cultivation on the transcriptome

Maintenance of *S. mansoni in vitro* has been noted in the past to have adverse effects on female fertility (Basch, 1991) and regression of female reproductive tissue has been observed in the presence and absence of male worms (Galanti *et al.*, 2012). To optimally support female fertility in *in vitro* culture, Basch medium was tested here for the maintenance of *S. mansoni in vitro*. It was found to support female fertility to a significantly greater extent than a less rich medium. Overall females were found to lay six times as many eggs over four weeks in Basch medium. Supplementation of the growth media with red blood cells was also shown to have a significant effect on fertility, resulting in a 2-fold increase of egg production in media supplemented with 1% RBC compared to media without RBC.

The next challenge was the isolation of sufficient total RNA from single worms to prepare cDNA libraries. Using an extraction protocol that I had optimised (Chapter 2.2.3) at least 50 ng of total RNA could be isolated from 60 single worm samples. From the generated RNA-Seq data, 150 genes were identified to be differentially expressed between paired and separated females after 8 days in culture and 42 in between paired and separated males. Even more genes were identified to be differentially expressed in comparisons of worms before and after *in vitro* culture, 1474 DEGs in female worms and 1210 in males.

A considerable problem in this analysis was the large variability in gene expression between biological replicates, particularly of female worms. This made it very difficult to examine the effect of treatment, *i.e.*, the pairing or separation of pairs. The magnitude of variability increased over the course of the experiment. The PCA plot of all samples (Figure 4.4) as well as the PCA plots of the time series analysis (Figure 4.8 and Figure 4.9) suggested that the effect of *in vitro* culture on the transcriptome was greater than of the treatment itself, despite the optimisation efforts. The *in vitro* culture of worms is a long established technique but one that has been noted to have significant draw backs and weaknesses, most notably the decline of female fertility but also slow or absent maturation of immature worms (Galanti *et al.*, 2012). The variability within the biological replicates reduced the statistical power to detect differentially expressed genes and obscured the effect of unpairing on the transcriptome of worms, especially for genes with moderate to low expression. As a result, fewer than expected differentially expressed genes, 150 DEGs in total, could be identified between paired and separated females. Furthermore, these results included several unexpected genes, including those coding for putative fertility-related proteins (eggshell protein, Smp_000430; and eggshell synthesis domain containing protein, Smp_077890) being up-regulated in separated females compared to their paired counter parts at day 8. Had the effects of *in vitro* culture been smaller, more subtle effects could have been identified in the separated females. The process of sexual regression that female worms undergo when removed from their mate should ideally be studied *in vivo* to ensure proper host condition. This could be achieved by perfusing worms from mixed sex infections, isolating mature females and reimplanting them into a

new host before perfusing them again after a few days or weeks to extract the RNA of these separated worms from *in vivo* conditions. This method of reimplanting worms into a host surgically has been used in the past (*e.g.* Basch 1990). However, this would be an invasive procedure that could cause injury and stress to the female worms, and likely impact their transcriptomes.

In male worms, only 42 DEGs were identified, all of which had relatively small fold changes, ranging from 1.8x to 1.12x with a median fold change of 1.3x. These fold changes are consistent with small changes in expression in the male worms but could also have resulted from larger fold changes in particular tissues. Such changes would become masked by measuring the average expression across whole male worms. The fact that DEGs were identified by DESeq2 in the comparison of male worms, despite the overall small fold changes, could indicate that the relatively low number of DEGs in the female comparison could be due to high levels of variability across biological replicates. In male worms on the other hand, there seems to be genuinely low levels of differentially expressed genes.

Down-regulation of genes coding for two actin homologues, two tegument allergen-like proteins and two tubulin genes, would indicate a reduction of the size of the tegument and muscle tissue. The up-regulation of a gene coding for an endothelin 2 homologue as well as an endothelin-converting protein suggests an involvement of this peptide hormone in the response of male S. mansoni to being separated from their partner. In humans, endothelin primarily plays a role in the regulation of blood pressure and has also been shown stimulate the nervous system (Bruno *et al*., 2011).

Differential expression of apoptosis related genes does not conclusively show up or down-regulation

To study the role of apoptosis in regression of female reproductive tissues I carefully curated a list of 27 putative apoptosis-related genes, making sure that their gene models were supported by RNA-Seq evidence and that they had sufficient homology to known apoptosis effectors. The function as well as domain structure of these apoptosis-related genes is described in more detail in Appendix B. I was expecting to find pro-apoptotic genes to be up-regulated in separated females compared to paired females. However, the comparison between paired and separated females did not provide conclusive evidence for up-regulation of apoptotic genes nor did the comparison of worms from day 0 and day 8, although some differential regulation of apoptosis-related genes was observed. Again, this may to have been due to the high degree of variability between biological replicates introduced by *in vitro* culture. Notably, sphingolipid metabolism-related genes were found to be differentially expressed at a higher than expected rate. Sphingolipids are important components of membranes and help cells to withstand stress (Futerman & Hannun, 2004). They can also be involved in signalling as is the case with ceramide. Ceramide is a sphingolipid and an important positive regulator of apoptosis (Thomas D Mullen & Obeid, 2012; Tirodkar & Voelkel-Johnson, 2012). Two genes have been shown to be of particular importance for controlling ceramide levels in cells: Ceramide kinase that phosphorylates ceramide, greatly diminishing its pro-apoptotic effect, and ceramide synthase that helps to synthesis ceramide and has a pro-apoptotic effect (Jensen *et al.*, 2014). However, both genes coding for these protein were significantly down-regulated after *in vitro* culture. However, both

ceramide kinase and ceramide synthase were found to be down-regulated after in vitro culture. It was therefore not possible to reliably assess, whether the down-regulation of the sphingolipid metabolism pathway had a pro- or anti-apoptotic effect. Overall, differential expression of apoptosis-related genes in a manner consistent with my pro- and anti-apoptotic classification could not be demonstrated although several of these genes were found to be differentially expressed.

Neurogenesis related genes in male and female worms

One of the 42 DEGs in the comparison of paired and unpaired males was the gene coding for DYP-30 which was up-regulated in paired males. DPY proteins are involved in dosage compensation, but not sex determination, in *C. elegans* (Hsu & Meyer, 1994) and the protein DPY-30 has recently been shown to be critical for neural lineage differentiation of mouse embryonic stem cells by regulating histone methylation (Jiang *et al.*, 2011). Being located on the Z/W chromosome (Uniprot – G4VT64; https://www.uniprot.org/uniprot/G4VT64), it may play a similar role development of neural cells in *S. mansoni*.

However, the most striking finding was the differential expression of neurogenesis related genes in male and female worms, that suggested a regression of the nervous system during *in vitro* culture. As discussed in the introduction to this chapter, most of the genes involved in the process of neural development can function in several roles. In particular the ligand:receptor pairs known to control axon guidance, namely semaphorins and their plexin and neuropilin receptors, netrins and the DCC and UNC receptors, slit and its robo

receptors as well as ephrin and its receptors have also all been implicated in the pruning of axons and neural cell death (Vanderhaeghen & Cheng, 2010). Notably, *S. mansoni* homologues of all of these genes were found to be differentially expressed after *in vitro* culture. Given the context of this data set and the well documented regression observed in worms outside the mammalian host, it seems most likely that the sudden up-regulation of these neural genes is correlated to a regression of parts of the schistosome nervous system. First and foremost, several neuron survival genes are down-regulated at day 8 as can be seen from the heatmaps of both male and female worms (Figure 4.11 & 4.14). Secondly several further pieces of evidence point to this also, including the up-regulation of GABA signalling in male worms at D8, which has been implicated in axon pruning, as well as the GO terms such as negative regulation of cell differentiation as well as the up-regulation of genes coding for neurogenic locus notch proteins in both male and female worms, a group of proteins thought to be essential for maintenance of neuronal progenitor cells. Combined, this evidence suggests that parts of the worm nervous system regresses by axon pruning and neural cell death and that neural progenitor cells are maintained in an undifferentiated form.

This process may be similar to the regression of reproductive tissues observed in female worms after they become separated from the male (Galanti *et al.*, 2012). However, it was shown to be independent of worm gender as transcriptome changes consistent with neural regression were found in both males and females. As worms could not have evolved a response to *in vitro* culture, it seems likely that the regression was a response to the lack of a host factor. The host immune system is a plausible candidate, as schistosome development is known to be

modulated by B and T cells (Davies et al., 2001; Tang et al., 2013). Davies *et al.* (2001) found schistosomes derived from mice without B and T cells were smaller, uniform and undifferentiated at 26 days post infection compared to worms from wild type mice and remained significantly smaller and stunted at 42 and 56 days after infection. Similar observations were made by Tang *et al.* (2013). The authors suggest that this may be a mechanism to ensure host and thus worm survival. Davies *et al.* (2001) speculated that by arresting worm development and thus temporarily preventing reproduction, host survival might be ensured, until the host is healthy enough again to sustain egg production. An important part of the *S. mansoni* nervous system is made up of the neuromuscular system. It allows the worms to control their movement, but also plays a role in digestion and excretion as well as other body functions (Ribeiro & Patocka, 2013). During the process of regression any loss of muscle tissue would therefore likely be accompanied by a loss of neural cells involved on the neuromuscular processes. Like female worms regressing after becoming unpaired from the male worm, the neural regression observed here could be evidence that the immune system driven maturation is reversible in both genders.

Chapter 5

Exploration of the *S. mansoni* gonad transcriptome

5.1 Introduction

Schistosomes are a rare exception within the platyhelminths having evolved to be dioecious with distinct male and female individuals, however the sexual biology of schistosomes is not only of evolutionary interest but also a rich area of research. This is due to its potential as an intervention target in the *Schistosoma* life cycle as sexual reproduction is obligatory for schistosome reproduction but and because the eggs are essential for transmission and responsible most of the pathology caused by schistosomes (Gryseels, 2012). Few tissues play such a prominent role in reproduction as the testes and ovaries as they are responsible for generating the spermatocytes and oocytes required for fertilisation and the formation of embryos.

The gonads are highly specialised tissues in which cytokinesis takes place at a high rate and in which cells undergo meiosis (Beckmann *et al.*, 2010). But not only are the tissues themselves specialised, they also produce specialised cells, the spermatocytes and oocytes. So, it is to be expected that many genes will be uniquely expressed in the gonads or at least expressed differentially as compared to whole worms. In both genders the gonads make up a relatively small portion of the total organism, whereas the somatic worm tissues are comprised mostly of tegument, gut and muscle (especially in the male) and vitellarian tissue [in the female (Basch, 1990)].

The transcriptomes of both these organs have been examined in the past both in *S. mansoni* (Nawaratna *et al.*, 2011) and *S. japonicum* (Gobert *et al.*, 2009) using a

combination of laser micro dissection microscopy (LMM) and cDNA microarrays. These studies were great steps towards an improved understanding of the schistosome gonads but left room for improvement. The authors had to rely on less complete sequencing data as well as less up-to-date genome annotation for the construction of the microarrays. The study also lacked biological replicates and did not correct for multiple hypothesis testing. Nawaratna *et al.* (2011) found 1989 genes (4450 probes) to have at least 2-fold higher expression in ovaries of mature females compared to a whole females, whereas 1188 genes (2171 probes) were found to have at least 2-fold higher expression in the testes of males from MS infections compared to control samples from whole MS males (Nawaratna et al., 2011). To explore the functional relationship between these genes, GO term enrichment analysis was performed. This gave a broad overview of the different groups of genes involved in processes such as "binding", "catalytic activity", "cellular process" and "metabolic process". However, as the annotation of the *S. mansoni* genome has been steadily improved since 2011 (Protasio *et al*., 2012), including the GO term annotation of genes, it is now possible to perform more in-depth enrichment analysis to yield greater insight into the underlying biological processes of sets of up- or down-regulated genes than was achieved previously.

Another aspect of gonad biology that so far has remained relatively unexplored is the transcriptome changes induced by pairing. Hahnel *et al.* (2013) developed a new method for the isolation of whole organs from *S. mansoni* using an approach that used both detergent and proteases to dissolve the outer layers of tegument and muscle to gain access to the worm gonads. This approach allowed

isolation of those tissues at a greater scale than previously feasible by micro dissection and thus made it possible to gather enough material for RNA sequencing. The team could demonstrate that cells could be isolated intact and with high quality RNA (Hahnel *et al.*, 2013; Hahnel *et al.*, 2014) and measured differential genes expression between the ovaries and testes of worms from mixed sex (MS) and single sex (SS), worms by qRT-PCR. Six genes putatively influenced by pairing based on transcriptomic evidence published in previous studies (Collins *et al.*, 2013; Leutner *et al.*, 2013; Zamanian *et al.*, 2011) were chosen for closer examination: two genes coding for homologues of the fibroblast growth factor receptor (FGFR-A/B), and genes coding for one homologue each of frizzled (Fz1), membrane progestin receptor component 1 (PMRC1), notch and the RNA binding protein musashi. Most of the analysed genes were found to have up-regulated transcription upon pairing, supporting the hypothesis that these genes are regulated by pairing and play a role in the sexual maturation of females. While the authors demonstrated that high quality RNA can be extracted from the isolated gonads of *S. mansoni*, they did not show how the process of organ extraction effects the transcriptome (Hahnel *et al.*, 2014).

The signalling network that regulates female fertility (see Figure 5.1) has been studied intensively (Beckmann *et al.*, 2010; Knobloch *et al.*, 2007; LoVerde *et al.*, 2009; Andrade *et al.*, 2014) and is described in Chapter 1.5. It comprises over 40 known genes, many of which have been characterised using *in silico* approaches, RNAi, *in situ* hybridisation, yeast-two-hybrid screens or using inhibitors such as herbimycin (Beckmann *et al.*, 2010; Knobloch *et al.*, 2006).

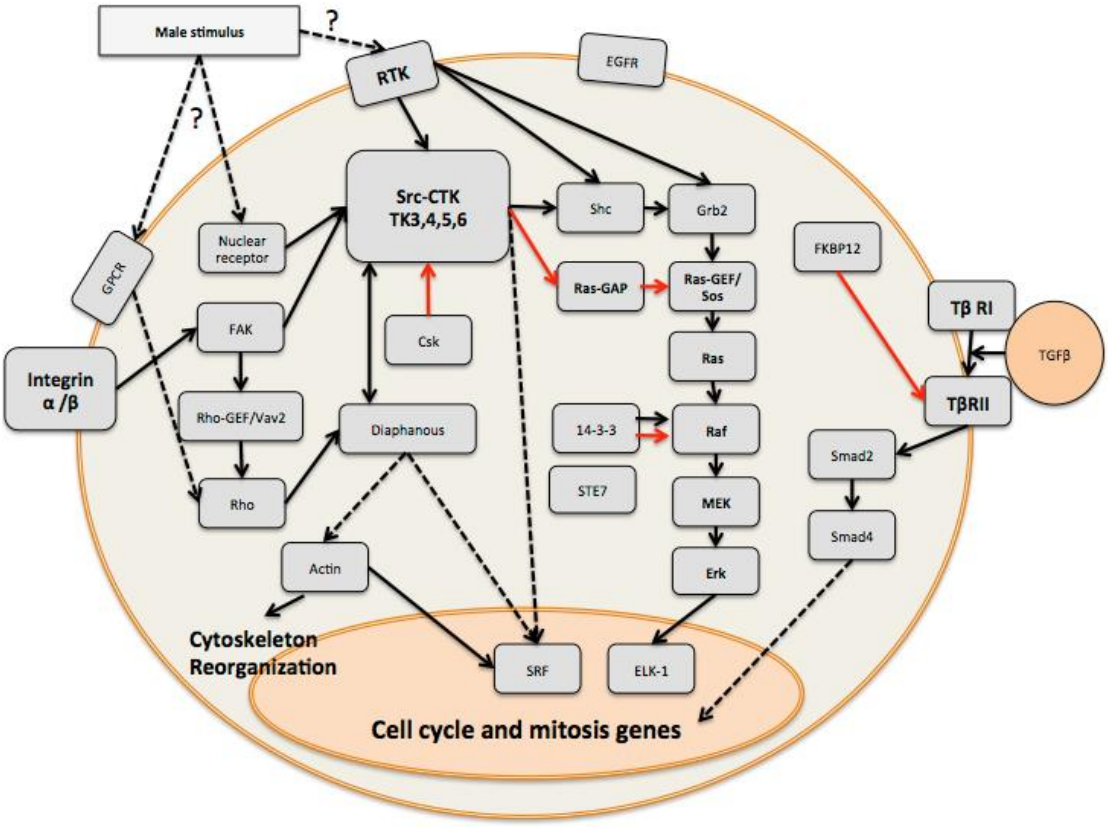


Figure 5.1: Representation of the MAPK, Rho and TGF-β signalling pathway inside a maturing oocyte from cell surface receptor to transcription regulators in the nucleus. Adapted from Knobloch *et al.* (2007). Dotted arrows indicate indirect interactions, solid arrows indicate direct interaction. Red arrows indicate inhibition and black arrows activation of the target.

Galanti *et al.* (2012) studied another aspect of schistosome reproductive biology: the female developmental regression by apoptosis, when separated from the male partner. The authors described the effect of pairing and un-pairing on gene expression in female worms *in vitro* and compared the rate of cell proliferation and apoptosis between females from single and mixed sex infections. Apoptosis was found to be responsible for the regression of the vitellarium as part of the developmental regression that the females undergoes in response to separation from their partner (Galanti *et al.*, 2012). Curiously, they also observed that the

rate of proliferation of the vitellarian tissue in females from SS and MS infections is very similar *in vivo,* but that the expansion of the vitellarium in SS females is prevented by a higher rate of apoptosis than in MS females (Galanti *et al.*, 2012). Like the vitellarium, the ovary also undergoes growth and development once induced by stable pairing of the female with a male. Therefore, it would be interesting to examine whether the transcriptome of ovaries from SS females reveals higher expression of pro-apoptotic genes compared to the ovaries from MS females. Apoptosis in *Schistosoma* has been previously studied by Lee *et al.* (2011, 2014) and Peng *et al*. (2010) who identified homologues of apoptosis related genes in the *S. mansoni* and *S. japonicum* genomes, respectively. Their results, as well as my efforts to identify other apoptosis related genes In the *S. mansoni* genome are discussed in detail in Chapter 4.2.

This work was a collaborative effort between our group and that of Prof. Dr. Christoph Grevelding. RNA was isolated from the gonads of worms from MS and SS infections as well as whole worm controls and provided to me. I used the RNA to synthesise stranded cDNA libraries for sequencing on Illumina HiSeq machines. Independent analyses of this data set (Lu *et al.*, 2016), as well as a gene expression atlas (Lu *et al.*, 2017) were recently published by both groups. The analysis placed greater emphasis on the evolutionary aspects of the gender interplay and examined the molecular mechanisms inducing sexual maturation. Since this experiment was conducted, our collaborators have also published a paper about the role of integrins and Venus Kinase Receptor 1 in the regulation of apoptosis in the ovaries (Gelmedin *et al*., 2017). The following analysis aims to define a set of ovary and testes expressed genes, as well as examining the

changes that take place in both gonads in response to pairing. To give proper context to these differentially expressed genes, I used several functional genomics approaches to show how groups of genes with similar biological functions are expressed in the gonads of *S. mansoni* and whole worms.

5.2 Results

5.2.1 Sequencing and sample clustering

In total, over 1.5 billion sequencing reads were produced, on average 63 million per sample (ranging from 42 million to 110 million), 84% of which could be mapped to the *S. mansoni* genome v5.2 (on average) (Appendix C.1).

During the initial analysis, one sample (single sex ovary 3, sO3) was identified which had apparently been mishandled and appeared to be whole male RNA instead of an ovary sample. This became apparent on a PCA plot (Figure 5.2) and was confirmed by examining expression of known male specific transcripts (Fitzpatrick *et al.*, 2005) (see Figure 5.3) such as those coding for the tegument allergen like (TAL) proteins (Smp_045200, Smp_086480 and Smp_195090) and calpain (Smp_137410); the sample was excluded from subsequent analysis.

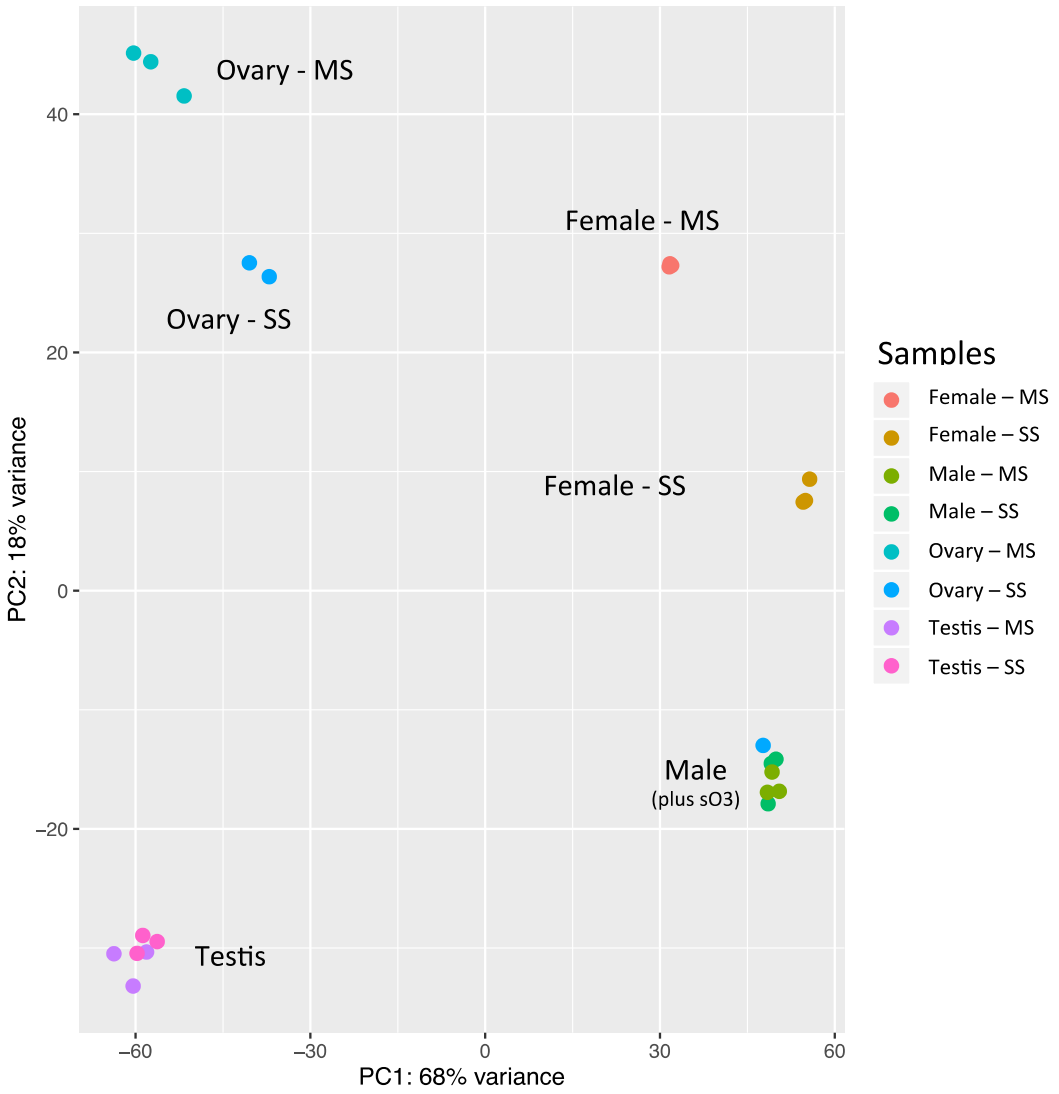


Figure 5.2: Samples cluster by type, male samples from mixed sex and single sex infection cluster together. The samples are either from single sex (SS) or mixed sex (MS) infections. In this principal component analysis of RNA-Seq samples, 86% of the variation among all samples is explained by the two most important principal components. The PCA plot is based on the 500 most variable genes in the data set, which contains three replicates per condition. Notably one of the SS ovary samples (sO3) clustered with the whole male samples. The three Female MS samples overlap tightly.

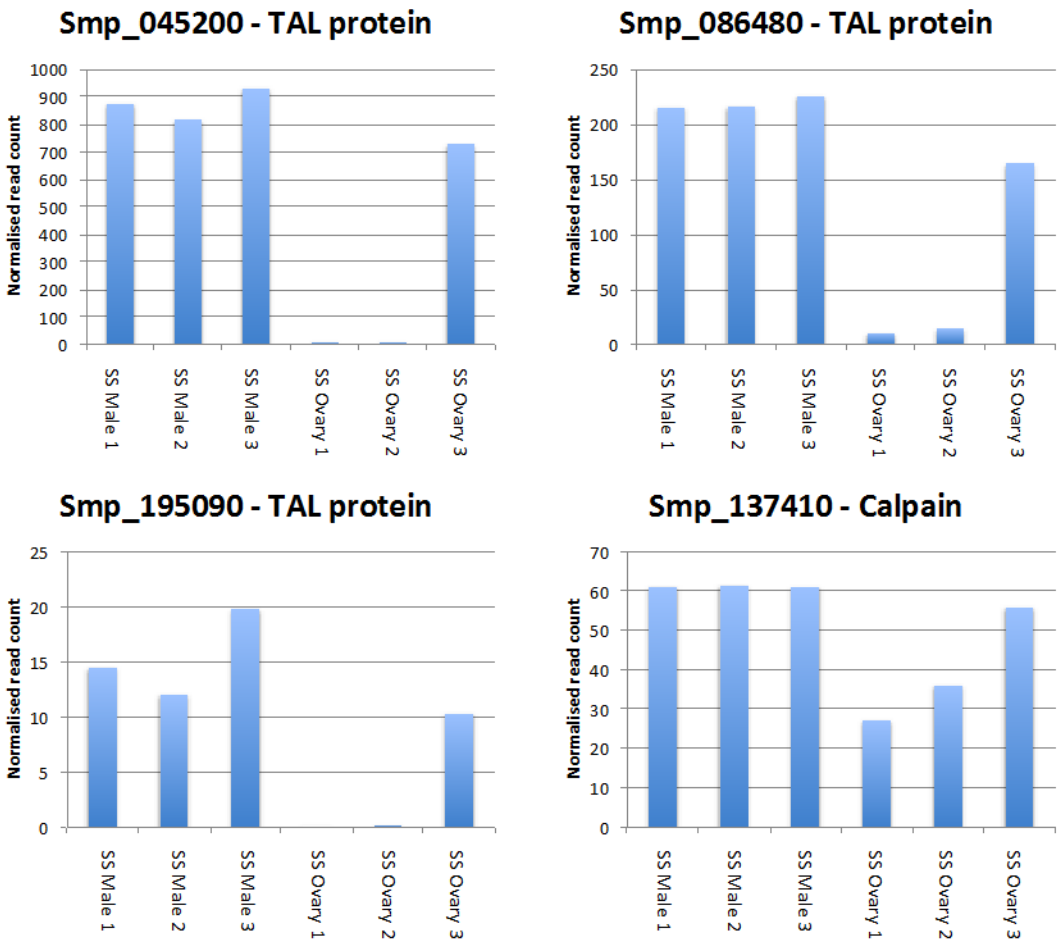


Figure 5.3: Single sex (SS) ovary replicate 3 was more similar to SS male replicates 1-3 than other SS ovary replicates (1-2). Here the sex-biased expression of genes is used to examine an outlier. Normalised read counts of SS male and SS ovary replicates were plotted for four transcripts with known male-biased expression patterns.

The principal component analysis based on the 500 genes with the greatest variance showed that samples from tight clusters. As expected, male and female samples formed discrete clusters and, as observed in Chapter 3.2, females from MS and SS infections form separate clusters reflecting the transcriptomic (and physiological differences) between sexually mature and immature females. Again, unsurprisingly, both males from MS as well as SS infections clustered very closely, despite a small number of consistently differentially expressed genes.

Furthermore, the clusters of gonad samples resembled those of the whole worms with ovaries of MS and SS females clustering away from one another, whereas both groups of testes showed very similar transcriptomes (Figure 5.2). Generally, the organ samples clustered less tightly and showed more variation within groups than those of whole worms (Figure 5.2). This could be due to differences in RNA quality caused by the procedure used for organ extraction, although apart from lower RNA concentrations no signs of low quality were observed (Appendix C.2).

5.2.2 Comparing RNA-Seq to published microarray & qPCR data

To examine how the gonad RNA-Seq data compared to similar data sets in the literature, I compared the present results to the most recent large scale effort to define the *S. mansoni* testes and ovary transcriptome (Nawaratna *et al.*, 2011). In their paper, Nawaratna *et al.* (2011) use a powerful combination of laser microdissection and a small oligonucleotide microarray to isolate tissue from the testes and ovary of sexually mature worms from MS infections. They then compared the expression levels of "39342 probes representing 19907 putative genes" (Nawaratna *et al.*, 2011) in the gonads with whole worm controls, defining genes expressed 2-fold higher in the gonads as compared with the whole worm control as "testes" and/or "ovary" genes.

To compare this data set with my RNA-Seq data the probe sequences were mapped to the *S. mansoni* reference genome v5.2 to find regions that uniquely corresponded to each probe. Splicing had to be taken into account, therefore Tophat2 was used for this process. Having determined the positions of the

probes in the genome allowed me to count RNA-Seq reads overlapping the position of these mapped probe sequences as described in the methods (Chapter 2.4.5). RNA-Seq data and microarray data are expected not to correlate linearly as RNA-Seq is expected to be more accurate at the lower and higher ends of expression (Zhao *et al.*, 2014) as discussed in the introduction (Chapter 1.6.3). Therefore, I chose Spearman's rank correlation coefficient to measure the correlation between RNA-Seq and microarray expression and found there was a weak positive correlation across all samples (r=0.46 to r=0.54), both whole worms and gonads (Figure 5.4).

In a previous analysis of the correlation between *S. mansoni* RNA-Seq to microarray results Protasio (2011) found that the results produced by both methods correlated well, with the correlation coefficient r ranging from 0.66-0.69. Therefore, a slightly higher correlation was anticipated in the present study, but there are several possible contributing factors to explain the lower correlation that I observed (see Figure 5.4). First, it is possible that the method of obtaining the organs, laser microdissection and protease/detergent treatment, had an impact on the transcriptome. However, this is unlikely as the correlation between whole worm samples was as low as the correlation of the gonad samples. Some of the variation between samples can be explained by the use of different parasite strains; Nawaratna *et al.* (2012) used the Puerto Rican strain of *S. mansoni*, whereas the strain used in our experiment on the gonad transcriptome originated in Liberia. Another reason could be differences in the quality of the RNA used in the two experiments, but there is no way to directly compare RNA quality across the two studies.

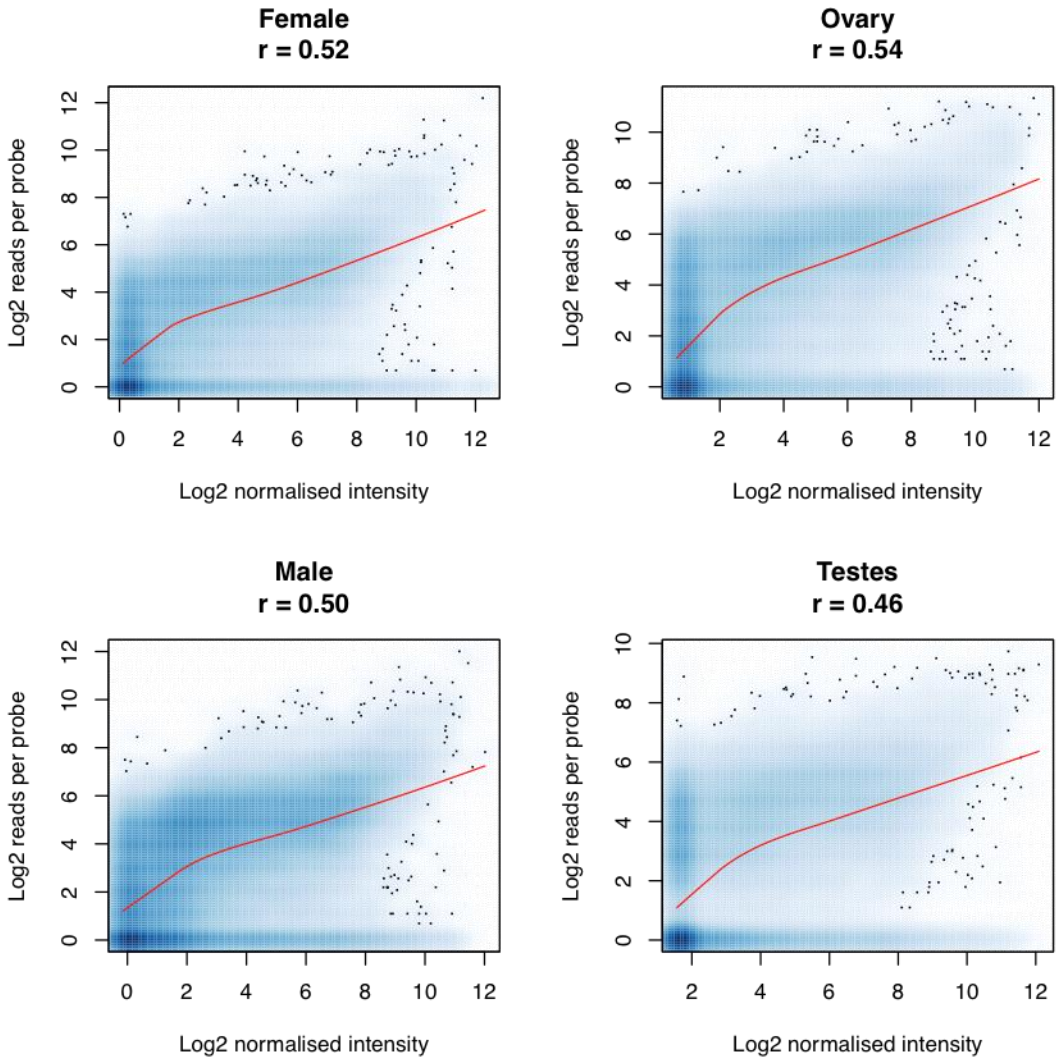


Figure 5.4: RNA-Seq and microarray data are moderately correlated. RNA-Seq counts were plotted against normalised intensity values of the microarray data sets (Nawaratna *et al.*, 2011); local regression line was plotted in red. A) Correlation of whole MS female samples. B) Correlation of ovaries from MS females. C) Correlation of whole MS males. D) Correlation of testes from MS males.

RNA-Seq counts were plotted against signal intensity of the microarray probes (see Figure 5.4). A local regression line (red) was fit onto the plots showing a weak correlation and the correlation coefficient was close to r=0.5 for all four samples Figure 5.4). In all cases, no signal was detected for a large number of probes using one or both methods of expression analysis (Figure 5.4). Another

publication that examined gene expression in the testes and ovaries of *S. mansoni* by Hahnel *et al.* (2014) used qRT-PCR to measure the expression of six genes coding for the following proteins: Frizzled (Smp_173940), FGFR-A (Smp_175590), FGFR-B (Smp_157300), PMRC1 (Smp_093700), Notch (Smp_050520), Musashi (Smp_157750) using actin (Smp_161930) as a reference gene. Next, the published qRT-PCR fold changes were comparable to the fold changes measured by RNA-Seq (Appendix C.3).

The initial comparison revealed striking differences in the ratio of expression (mature ovary gene expression divided by immature ovary gene expression) (Figure 5.5).

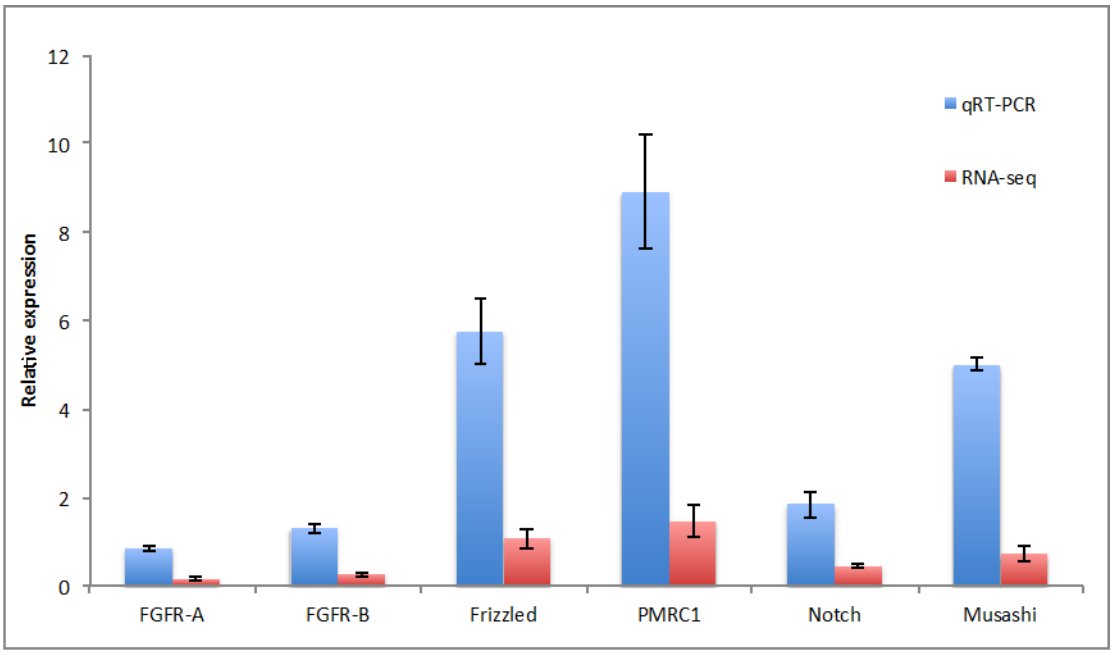


Figure 5.5: Comparison of qRT-PCR and RNA-Seq reveals systematic differences. Expression ratio in ovaries (expression in mixed sex ovaries divided by expression in single sex ovaries) as measured by qRT-PCR and RNA-seq. Generally, RNA-Seq ratios were all much lower suggesting less expression in paired ovaries to be lower than measured by qRT-PCR, although the trend is the same, indicating a normalisation issue.

When examining expression of the reference gene (actin) however, the RNA-Seq data suggested significant differential expression of actin between ovaries from MS and SS infections (8.1-fold higher expression in SS ovaries than in MS ovaries; adjusted p-value 6.33E-14). Eleven other published qRT-PCR reference genes (Liu *et al.*, 2012) were examined to see whether they displayed a similar behaviour in the RNA-Seq data which would indicate of a problem with the normalisation of expression across samples. However, the other reference genes were not found to be differentially expressed and had smaller fold changes (1.3-fold change on average) (Appendix C3). When taking into account the 8.1-fold expression change of the qRT-PCR reference gene in the comparison of gene expression between the two methods, the "adjusted" expression ratios were much closer to those reported by Hahnel *et al.* (2014) (Figure 5.6) and therefore provide excellent validation of the results discussed in the following sections.

Actin was not found to be expressed differentially to such large degrees in any other comparison of samples from SS versus MS infection but was still found to be significant (2.2-fold down-regulation in MS females; adjusted p value = 2.16E-13) in MS females (Appendix C.3).

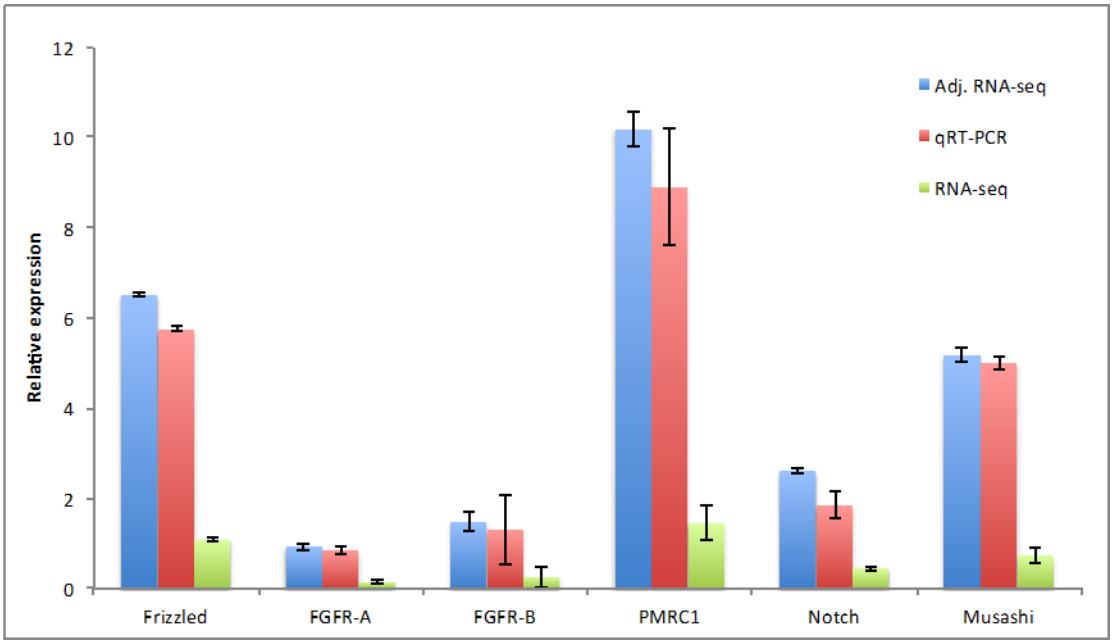


Figure 5.6: RNA-Seq data showed good agreement with qRT-PCR after adjustment for differential expression of qRT-PCR reference gene. Expression ratio in ovaries (expression in MS ovaries divided by expression in SS ovaries) as measured by qRT-PCR and RNA-Seq as well as an adjusted RNA-Seq ratio.

5.2.3 Testes Transcriptome

"Testes genes" were identified using a pair-wise comparison of gene expression in testes of MS males and whole MS males using DESeq2. To be considered as "testes gene" a gene had to be differentially expressed using an adjusted p-value cut-off smaller than 0.01 and be at least 2-fold more highly expressed in testes from MS infections than the whole male control. This threshold was used throughout this chapter to identify DEGs. 10270 out of 10828 transcripts were found to be expressed (*i.e.* non-zero read counts), of which 512 (5%) were filtered out due to low expression levels using the DESeq2 independent filtering algorithm. 1949 transcripts passed the criteria to be considered "testes genes" (Figure 5.6).

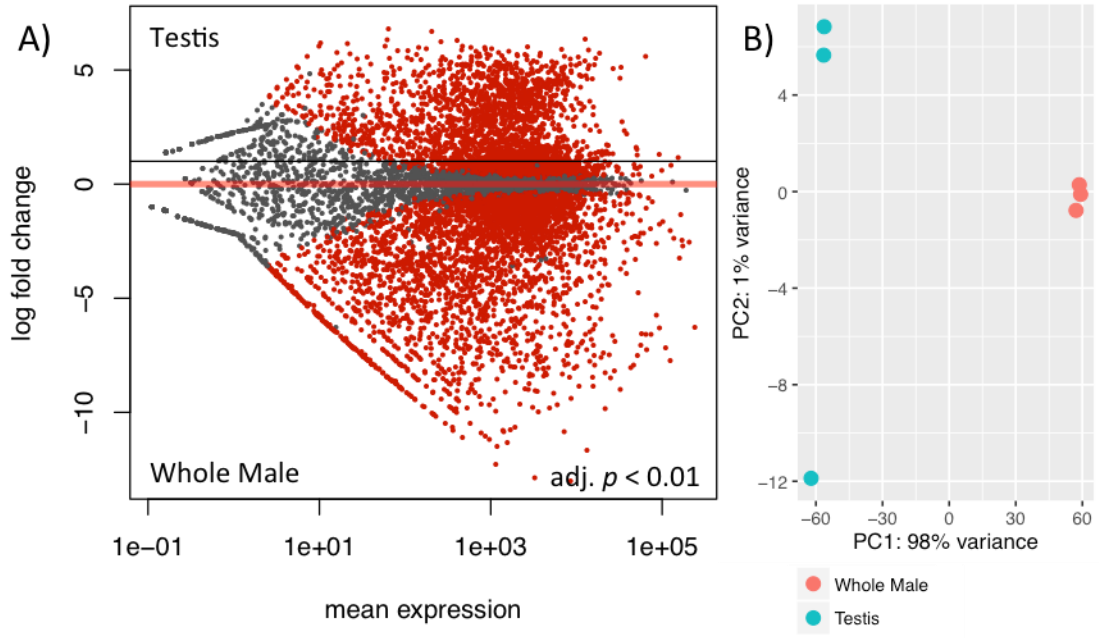


Figure 5.6: The testis transcriptome differs considerably from that of whole males. A) Log ratio against mean (MA) plot of the testis-whole male comparison. Genes with an adjusted p-value of < 0.01 are high-lighted in red; genes with a log-fold change of > 1 (fold-change of > 2) and adjusted p-value of < 0.01 were called as differentially expressed "testis genes". B) PCA plot of testes of paired males and whole male control samples.

When plotting the mean expression (A) of genes against their log fold change (M) (Figure 5.6 A), a large number of genes have large fold changes, especially those up-regulated in whole males. This is because genes expressed exclusively outside the testes should be completely absent from the testes samples, unlike the testes-specific genes, which also appear as expressed in the whole males. 99% of variation of the RNA-Seq count data between whole male and testes samples is explained by the first two principal components, most notably, the first principal component that separates the male from the testes samples explains 98% of the total variation (Figure 5.6 B). This suggests that there are large differences in

gene expression between whole male worms and testes samples and that the expression data has relatively little noise.

Amongst the most differentially expressed testes genes (adjusted p-value < 1.30E-186; Table 5.1) were DEGs coding for a cyclin dependent kinase (Smp_080730), a RNA helicase (Smp_172150), HORMA domain containing protein (Smp_169930), which is involved in DNA replication and recognition of damaged DNA (Muniyappa *et al.*, 2014), a DNA primase subunit (Smp_079050) and a hormone receptor 4 (Smp_041540). As expected, this list suggests that the transcriptome of the testes is dominated by factors involved in mitosis. Also, a "testis expressed protein" (Smp_131630) was found to be expressed more strongly in testes than whole male. Another interesting gene found to be differentially expressed codes for the transcription factor *boule* (Smp_144860), which is expressed in the gonads of other Platyhelminthes, such as free-living *Macrostomum lignano* (Kuales *et al.*, 2011). The transcription factor *boule* is also widely conserved across other animals including mammals due to its importance in the regulation of meiosis (Shah *et al.*, 2010). I performed a whole mount *in situ* hybridisation to examine the specificity of *boule* expression in 20 male worms and detected a robust signal in the testes of all worms but no other tissues of the worm, except in the ventral sucker of four worms which are common sites of non-specific staining (see Figure 5.7).

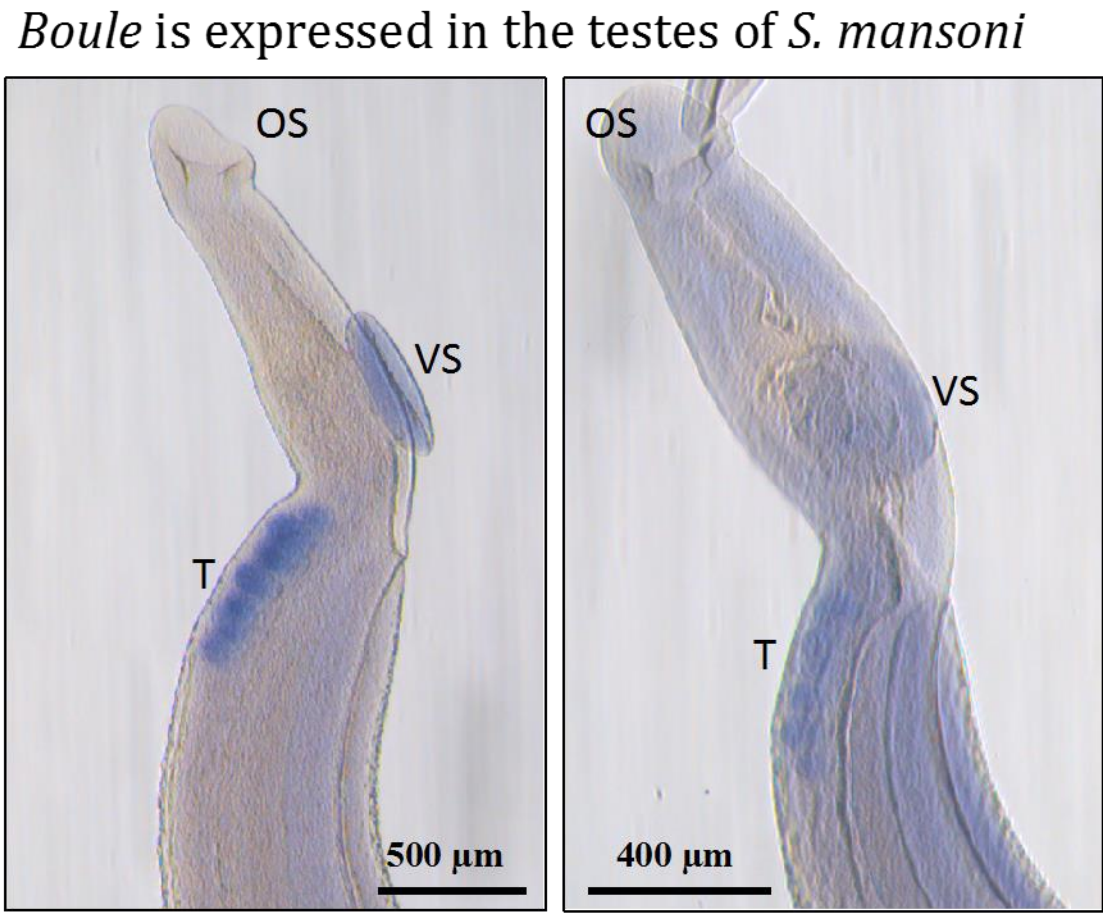


Figure 5.7: *In situ* hybridisation targeting mRNA of transcription factor *boule* demonstrated expression in the testes of male worms. Four out of twenty worms stained for boule also exhibited staining in the ventral sucker (left). Sense transcript of *boule* was used as negative control (right). Even after prolonged development no distinct expression signal was noticeable in the testes. OS – Oral Sucker; VS – Ventral Sucker; T – Testes.

A GO term enrichment analysis was performed using topGO (Alexa *et al.*, 2006) in order to get a more comprehensive picture of the function of testes genes. By testing for statistically significant (p-value < 0.01) enrichment of GO terms I aimed to get a more comprehensive picture of the biological processes taking place in the testes. The testes genes were significantly enriched with DEGs from a total of 32 GO terms (Table 5.2), whereas genes up-regulated in whole males were enriched with genes from 37 GO terms (Appendix C.4). Notably, the list of

testes GO terms was dominated by mitosis related processes such as “DNA replication initiation”, “spermatogenesis”, “DNA repair”, “cell cycle”, “mitotic nuclear division” and “miss match repair”, all reflecting the primary function of testes, the production of sperm. Furthermore, “microtubule-based movement” and “striated muscle contraction” were found to be up-regulated significantly in testes (Table 5.2). Spermatocytes are known to rely on microtubules for locomotion; in human sperm cells eleven microtubules are arranged into an axoneme which forms the central part of the sperm flagellum (Kierszenbaum, 2002). Furthermore, microtubules play an important role in the separation of chromosomes during meiosis and mitosis, which are expected to occur at a high rate in the testes. Striated muscle, on the other hand, has been observed to surround the testes of *D. melanogaster* and mammals (Kuckwa *et al.*, 2016). “Protein glycosylation” (GO:0006486) and “protein refolding” (GO:0042026) were also significant GO terms (Table 5.2). The latter includes several genes coding for chaperone proteins such as heat shock protein 70, which in mice plays an essential role in spermatogenesis by allowing Cdc2 and cyclin B1 to form a complex without which the developing spermatocytes become arrested at the G2-M-phase transition and undergo apoptosis (Eddy, 1999)

A KEGG pathway was performed, aimed at identifying pathways significantly enriched in testes-specific genes, including metabolic and signalling pathways among others (Chapter 2.4.10). It showed that DNA replication, mismatch repair and homologous recombination but also the Fanconi anaemia (FA) pathway play an important role in the testicular tissue (Table 5.3). The FA pathway is necessary to allow the repair of damaged DNA, especially interstrand cross-links,

which are recognised by the FA core complex, which in turn activates various DNA repair proteins (Rodríguez & D'Andrea, 2017). Additionally, ribosomes, ribosome biogenesis and tRNA biosynthesis were up-regulated in testes compared to whole worms (Table 5.3). Together with mitosis-related GO terms and KEGG pathways, these biological processes such as ribosome and protein production are also indicative of rapid cell proliferation in the testes.

The mitotic activity of testes is also evident in an analysis of over-represented protein domains in the testes transcriptome; for instance, out of 75 genes coding for helicase C-terminal domains, more than half are expressed at higher levels in testes than the whole male control (only three are expressed at higher levels in the whole male control) and 27 out of 55 genes coding for DEAD/DEAH box helicase domain containing proteins are up-regulated in testes (compared with 3 in whole males) (Table 5.4). Furthermore, all 10 genes coding for mini-chromosome maintenance domains are up-regulated in the testes, as are all genes coding for proteins with MutS domain II, III, IV and V (all involved in DNA mismatch repair) (Table 5.4). Many of proteins with these domains play an important role in the eukaryotic DNA replication machinery as can be seen in Figure 5.8.

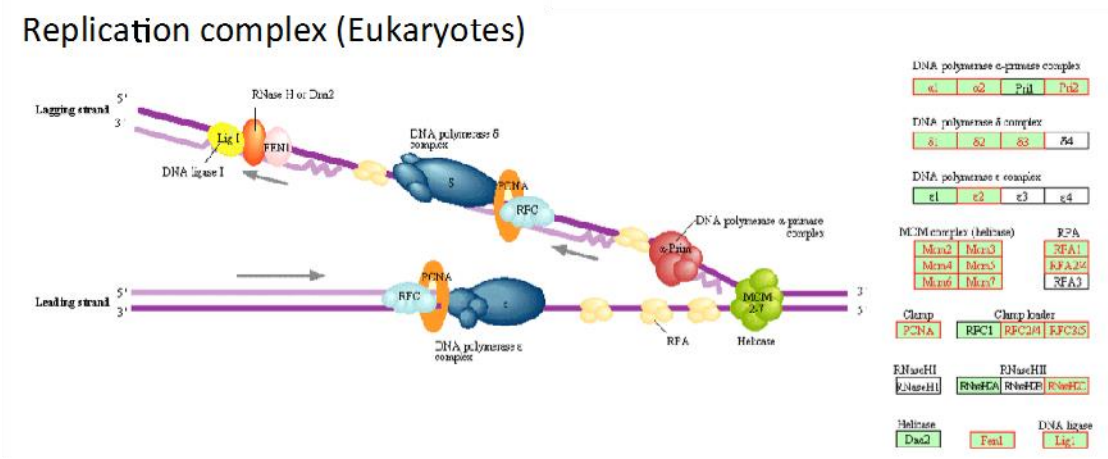


Figure 5.8: Most annotated *S. mansoni* homologues of DNA replication machinery genes are up-regulated in testes. Genes with a *Schistosoma mansoni* homologue are coloured in green, "testes genes" are highlighted in red. Most DNA replication genes are up-regulated in the testes transcriptome. This list highlights the important role played by many proteins with minichromosome maintenance (MCM) and helicase domains. Figure adapted from KEGG (https://www.genome.jp/dbget-bin/www_bget?hsa03030).

Notably absent from the testes genes but instead over represented in the whole male transcriptome were genes mapped to the following KEGG pathways: lysosomes, phagosome and endocytosis as well as various metabolic pathways including starch, sucrose, galactose, glucose, glycerophospholipid and amino acid metabolism (Appendix C.5), all of which are responsible for core physiology of the parasite. The whole male transcriptome was also enriched for various signalling pathways including Wnt, FoxO and hedgehog signalling (Appendix C.5) which play important roles during development of many organisms (Eijkelenboom & Burgering, 2013; Ingham *et al.*, 2011; Nusse & Varmus, 1992; Nusse & Varmus, 2012). This is probably caused by the whole worms containing a wide variety of tissues and being more complex, including at the transcriptome level, than the testes in isolation.

Gene ID	Description	Average Read Count	Fold Change	adjusted p-value
Smp_169930	HORMA domain containing protein	5313.73	34.78	4.14E-215
Smp_131630	Testis expressed protein	1452.24	34.07	1.27E-146
Smp_172150	Putative ATP-dependent RNA helicase	4062.9	32.22	2.67E-223
Smp_041540	Hormone receptor 4 (dHR4)	7395.01	31.12	2.39E-205
Smp_144860	Boule	4169.81	27.19	2.61E-166
Smp_079050	DNA primase large subunit	3908.09	24.93	1.27E-210
Smp_080730	Cyclin dependent kinase	4902.06	22.78	1.64E-233

Table 5.1: List of differentially expressed genes with testes-specific expression (see above). Average read counts are a measure of absolute gene expression; they are the average of normalised read counts across all samples. The fold changes reflect expression in the testes relative to whole male worms. The *p*-value has been adjusted for multiple hypothesis testing. See Appendix C.6 for an expanded list.

GO term	Description	Total Genes	DEGs	Expected	p-value
GO:0007018	Microtubule-based movement	60	36	9.32	3.40E-15
GO:0006941	Striated muscle contraction	145	59	22.52	1.20E-13
GO:0006270	DNA replication initiation	15	13	2.33	2.20E-09
GO:0007283	Spermatogenesis	10	9	1.55	4.40E-07
GO:0006281	DNA repair	122	46	18.95	7.00E-07
GO:0007049	Cell cycle	207	78	32.15	2.40E-06
GO:0007067	Mitotic nuclear division	88	37	13.67	3.50E-06
GO:0006260	DNA replication	161	64	25.01	7.50E-05
GO:0007126	Meiotic nuclear division	8	6	1.24	0.00029
GO:0006298	Mismatch repair	11	7	1.71	0.0004
GO:0006486	Protein glycosylation	58	15	9.01	0.00897
GO:0042026	Protein refolding	8	4	1.24	0.024

Table 5.2: List of GO terms enriched in testes genes (see above). The column "Total Genes" provides the number of genes associated with a particular GO term. "DEGs" provides the number of genes associated with a given GO term that had significantly higher expression in testes than whole males; "Expected Genes" provides the number of differentially expressed genes associated with a given GO term that are expected by chance. The GO terms are all in the "Biological Process" category. See Appendix C.7 for the complete list.

Pathway	Description	Total Genes	DEGs	Expected	p-value
smm03030	DNA replication	30	24	7	5.80E-11
smm03430	Mismatch repair	18	16	4.2	5.93E-09
smm03440	Homologous recombination	18	15	4.2	1.07E-07
smm03010	Ribosome	110	47	25.68	1.88E-06
smm03460	Fanconi anaemia pathway	27	17	6.3	9.39E-06
smm03008	Ribosome biogenesis in eukaryotes	62	22	14.47	9.86E-03
smm00970	Aminoacyl-tRNA biosynthesis	34	13	7.94	2.09E-02

Table 5.3: List of pathways enriched amongst the testes expressed genes (see above). The column "Total Genes" provides the number of genes in the KEGG pathway; "DEGs" provides the number of differentially expressed genes in that pathway; "Expected Genes" provides the number of differentially expressed genes in a given pathway that is expected by chance. See Appendix C.8 for the complete list.

Domain	Description	Total Genes	DEGs	Expected	p-value
PF00271	Helicase conserved C-terminal domain	75	38	17.68	2.03E-07
PF00270	DEAD/DEAH box helicase	55	27	12.97	2.19E-05
PF00493	MCM2/3/5 family	10	10	2.36	5.20E-07
PF00488	MutS domain V	4	4	0.94	3.08E-03
PF05188	MutS domain II	3	3	0.71	1.31E-02
PF05190	MutS family domain IV	3	3	0.71	1.31E-02
PF05192	MutS domain III	3	3	0.71	1.31E-02

Table 5.4: List of domains enriched amongst testes expressed genes discussed above. The column "Total Genes" provides the number of genes with a given Pfam domain; "DEGs" provides the number of differentially expressed genes with that domain; "Expected" provides the number of differentially expressed genes encoding a given domain that is expected by chance. See Appendix C.9 for an expanded list.

5.2.4 Ovary Transcriptome

"Ovary genes" were identified using the same method as testes genes, by doing a pair-wise comparison between ovaries of MS females and whole paired females using DESeq2. "Ovary genes" had to be differentially expressed using the same

adjusted p-value cut-off (< 0.01) and also have at least two-fold higher expression in ovaries than in the whole female control. Out of 10828 transcripts, 10213 were found to be expressed (*i.e.* non-zero read counts), of which 509 (5%) were filtered out due to low levels of expression using the DESeq2 independent filtering algorithm. 1272 transcripts passed the criteria to be considered "ovary genes" (Figure 5.9A).

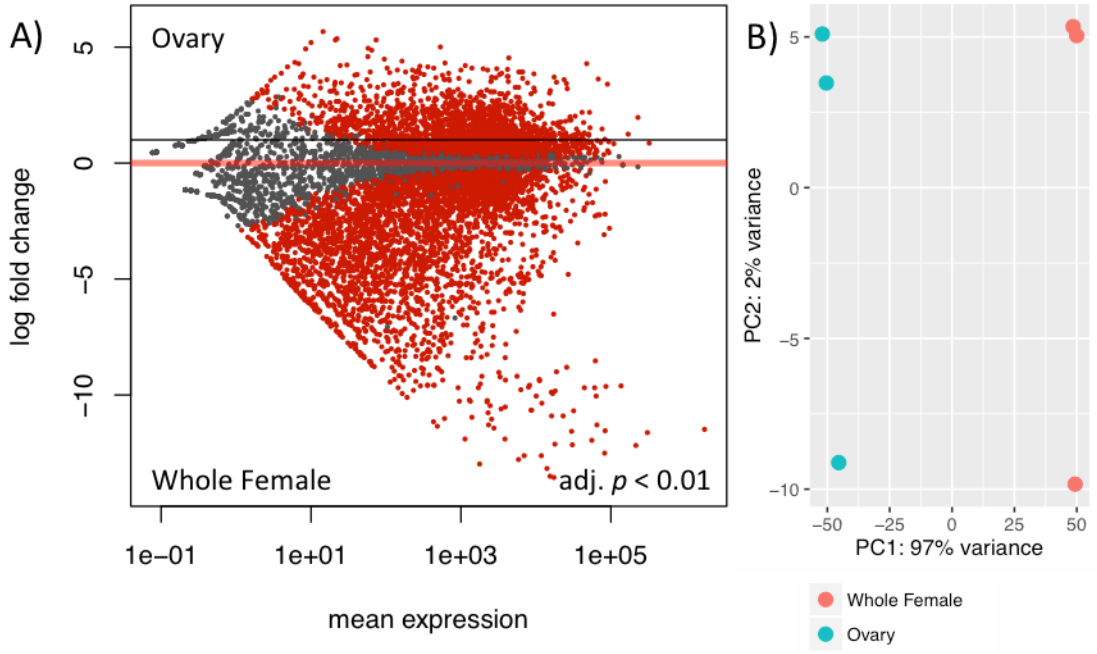


Figure 5.9: 1272 genes were identified as ovary-specific genes. A) Log ratio against mean (MA) plot of the ovary-whole female comparison. Genes with an adjusted p-value of < 0.01 are high-lighted in red; genes with a log-fold change of > 1 (fold-change of > 2) and adjusted p-value of < 0.01 were called as differentially expressed "ovary genes". B) PCA plot of ovaries of paired females and whole female control samples.

Similar to the testes vs whole male comparison, the comparison of ovaries vs whole female shows a pattern, where genes up-regulated in whole females have greater fold changes than those up-regulated in ovaries (Figure 5.9B). Similarly to the testes genes, the most differentially expressed ovary genes contain genes

coding for proteins related to DNA replication and repair such as cytospin-A, as well as cell cycle controlling transcripts such as cyclins (Table 5.5) which promotes the anaphase stage of the cell cycle (Yuan & O'Farrell, 2015). The gene *boule* was also found to be differentially expressed in ovaries as compared to whole worms, indicating that - like its homologue in hermaphroditic flatworms – it is expressed in both testes and ovaries (Kuales et al., 2011). A "testis expressed protein" gene (Smp_131630) was found to be up-regulated in the ovaries as well (Table 5.5). In mice it contributes to the formation of intercellular bridges, making it essential for spermatogenesis. It plays the same role in the development of oocytes, although knock-outs of testis expressed gene 14 does not lead to sterility in females (Greenbaum *et al.*, 2009).

The results of the GO term analysis of ovary genes (Table 5.6) were very similar to that of the testes genes (Table 5.2). The ovary genes are enriched for 36 GO terms, 20 of which are shared with the GO terms enriched in the testes genes and include: "DNA replication initiation" and "DNA repair", "mitotic nuclear division", "miss-match repair" and "protein glycosylation". Other GO terms enriched in the testes and ovary genes include cell cycle control and regulation of cell proliferation: "cell proliferation", "negative regulation of cell proliferation", "cell cycle" and "meiotic nuclear division". Unlike testes genes, which are not over-represented in GO terms for signalling, the ovary genes are over-represented in "regulation of Rho protein signal transduction". Almost four times more ovary genes associated with that GP-term are up-regulated significantly when compared to the expected number of such genes (23 up-regulated, 6 expected). The results above are summarised in Table 5.6.

The KEGG pathway analysis also highlighted the importance of DNA replication, the FA pathway and mismatch repair pathway, which were significantly up-regulated in the ovary transcriptome (Table 5.7). Additionally, the analysis indicated a high rate of protein synthesis in this tissue, with ovary genes being over represented in Ribosome, Ribosome Biogenesis and other related pathways (Table 5.7).

Even more so than in the testes genes, the ovary genes were significantly enriched with genes coding for helicase (both C-terminal helicase domains and DEAD/DEAH box helicase domains) and MCM domains (both the N-terminal domain and the MCM2/3/5 family domain). These were the three most significantly enriched domains, with between 77-100% (compared to 30% expected by chance) of genes coding for such domains being expressed at significantly higher levels in ovaries compared to whole females (Table 5.8). Another significant domain was the BRCA1 C Terminus domain (BRCT), belonging to a group of genes involved in DNA repair and is most known for its role in breast cancer (Friedenson, 2007). All nine *S. mansoni* genes coding for this domain had at least 2-fold higher expression in the ovaries than whole females (Table 5.8).

Genes with noticeably low expression in the ovaries included those involved in digestion and metabolism of nutrients (in particular KEGG pathways such as lysosome, phagosome and endocytosis as well as amino acid and sugar metabolism) (Appendix C.10. Other pathways which were also expressed at low

levels in ovaries were signalling pathways such as FoxO (smm04068), hedgehog (smm04340) and neuroactive ligand-receptor interactions (smm04080). These results were found to be similar to those of the testes transcriptome, reflecting the function of germ cell production common to both gonad tissues.

Gene ID	Product	Average Read Count	Fold Change	adjusted p-value
Smp_144860	Boule	1998.21	16.06	1.34E-147
Smp_131630	Testis expressed protein	1524.23	15.04	7.49E-162
Smp_018010	Cytospin-A	9979.58	8.34	2.59E-211
Smp_047620	Cyclin	6150.25	8.05	5.09E-162

Table 5.5: Differentially expressed ovary-genes (see above). Average read counts provide a measure of absolute expression for a given gene. The fold changes reflect expression in ovaries relative to whole females. The *p*-value has been adjusted for multiple hypothesis testing. See Appendix C.11 for an expanded list.

GO term	Description	Total Genes	DEGs	Expected	p-value
GO:0006281	DNA repair	122	45	14.67	2.500E-10
GO:0006270	DNA replication initiation	15	12	1.8	2.700E-09
GO:0007049	Cell cycle	207	71	24.89	8.700E-08
GO:0006260	DNA replication	161	57	19.36	1.500E-07
GO:0007067	Mitotic nuclear division	88	33	10.58	1.300E-06
GO:0007126	Meiotic nuclear division	8	6	0.96	6.700E-05
GO:0006298	Mismatch repair	11	6	1.32	8.000E-04
GO:0008285	Negative regulation of cell proliferation	5	3	0.6	1.436E-02
GO:0006486	Protein glycosylation	58	11	6.98	4.348E-02
GO:0035023	Regulation of Rho protein signal transduction	23	6	2.77	4.972E-02

Table 5.6: GO terms enriched amongst genes up regulated in the ovaries of paired females (see above). The column "Total Genes" provides the number of genes associated with a particular GO term; "DEGs" provides the number of genes associated with a given GO term that had significantly higher expression in ovaries; "Expected Genes" provides the number of differentially expressed genes associated with a given GO term that are expected by chance. The GO terms are all in the "Biological Process" category. See Appendix C.12 for the complete list.

Pathways	Description	Total Genes	DEGs	Expected	p-value
smm03008	Ribosome biogenesis in eukaryotes	62	48	20.60	5.076E-13
smm03030	DNA replication	30	25	9.97	1.650E-08
smm03010	Ribosome	110	62	36.55	1.712E-07
smm03460	Fanconi anaemia pathway	27	21	8.97	2.040E-06
smm03430	Mismatch repair	18	15	5.98	1.495E-05

Table 5.7: KEGG pathways enriched amongst the genes up-regulated in the ovaries of paired females (see above). The column "Total Genes" provides the number of genes in the KEGG pathway; "DEGs" provides the number differentially expressed genes in that pathway; "Expected Genes" provides the number of differentially expressed genes in a given pathway that is expected by chance. See Appendix C.13 for the complete list.

Domain	Description	Total Genes	DEGs	Expected	p-value
PF00271	Helicase conserved C-terminal domain	75	58	22.29	1.360E-17
PF00270	DEAD/DEAH box helicase	55	43	16.35	1.126E-13
PF00493	MCM2/3/5 family	10	10	2.97	5.309E-06
PF00533	BRCA1 C Terminus (BRCT) domain	9	9	2.68	1.791E-05
PF14551	MCM N-terminal domain	6	6	1.78	6.865E-04

Table 5.8: Pfam domains enriched amongst the genes up-regulated in the ovaries of paired females (see above). The column "Total Genes" provides the number of genes encoding the Pfam domain; "DEGs" provides the number of differentially expressed genes coding for the domain; "Expected" provides the number of differentially expressed genes coding for the domain that is expected by chance. See Appendix C.14 for an expanded list.

5.2.5 Comparing the Testes and Ovary Transcriptomes

To identify genes that serve common function in the testes and ovary of *S. mansoni*, I examined the intersection between the sets of "testes genes" and "ovary genes". Out of 1949 testes genes and 1272 ovary genes, 715 were shared (Figure 5.10). The 715 DEGs were over-represented in many mentioned GO terms discussed in Chapters 5.2.3 and 5.2.4. This includes DNA repair and

replication, both mitotic and meiotic nuclear division, striated muscle contraction and microtubule based movement (Figure 5.10). However, there are interesting differences when looking at the genes expressed in ovaries and testes. While "testes genes" showed an expected enrichment for spermatogenesis, the "ovary genes" included an enrichment for genes of the Notch signalling pathway (Figure 5.10) which is known to be interconnected with Rho signalling and is a highly conserved pathway found in all metazoans (Boureux *et al.*, 2007; Schwartz, 2004). The Notch signalling pathway (GO:0007219) was represented by 22 genes (4 expected, p = 0.033) (Appendices C.15-16).

Fewer genes had ovary-specific expression compared to the testes (Figure 5.10). In the differential expression analysis, ovary-expressed genes were found to have a lower median fold-change than testes genes (6.51 and 14.03-fold change respectively). As a result, more testes genes had a fold-change above the significance threshold (fold change > 2).

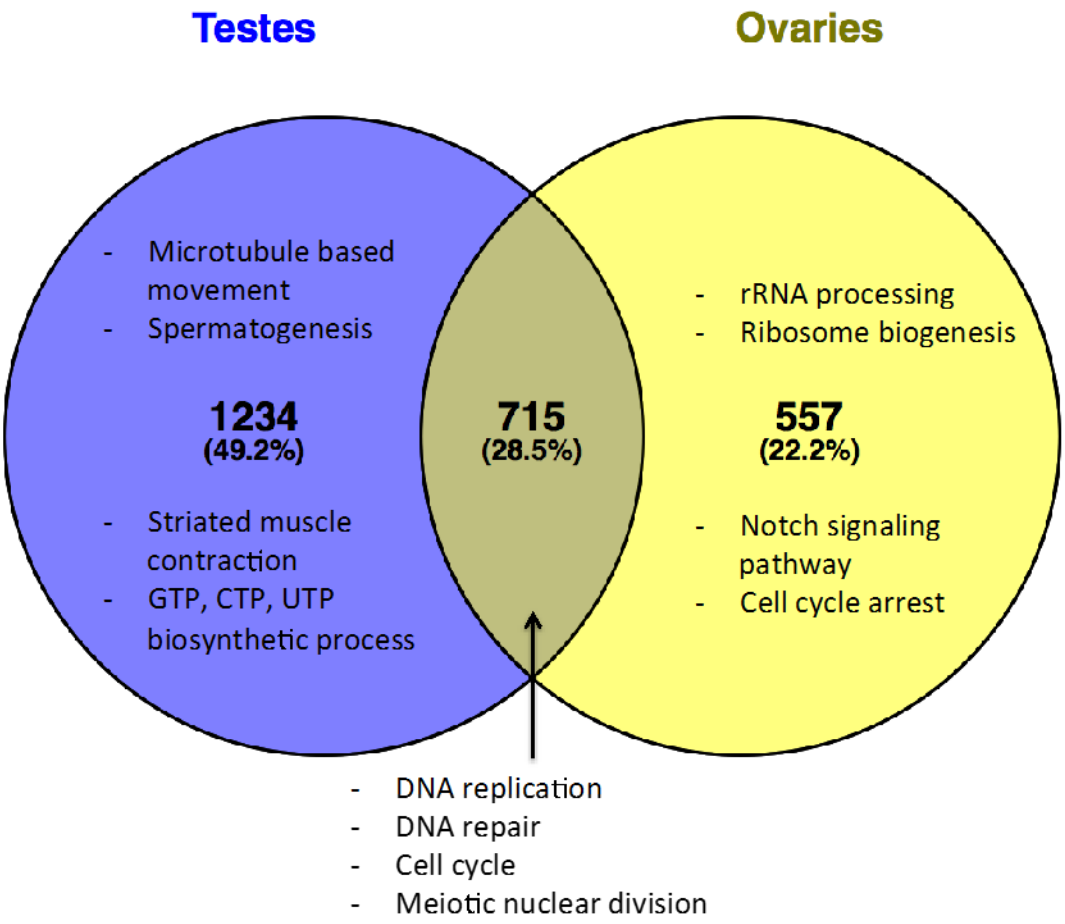


Figure 5.10: More testes-specific than ovary-specific genes were identified with distinct functions in the two groups of genes. The numbers indicate the testes and ovary "specific" genes as well as genes shared by both. A selection of significantly enriched GO terms is given for each category. The 715 shared genes were found to be strongly enriched for DNA replication related functions. This included repair and chromosome maintenance. The genes expressed in testes contained many spermatogenesis-specific genes whereas the genes expressed in ovaries included genes that regulate multicellular organismal development and Notch signalling pathway genes.

5.2.6 Effect of pairing on testes transcriptome

The effect of pairing on the testes transcriptome is, like the effect of pairing on the male overall, relatively small. The aim was to identify pairing-related transcriptome changes in the testes. By excluding all non-testes specific transcripts, small transcriptome changes in the testes should be more easily detected, allowing more differentially expressed testes-specific genes to be identified. Genes differentially expressed between testes of males from MS and SS infections were identified by doing a pair-wise comparison using DESeq2.

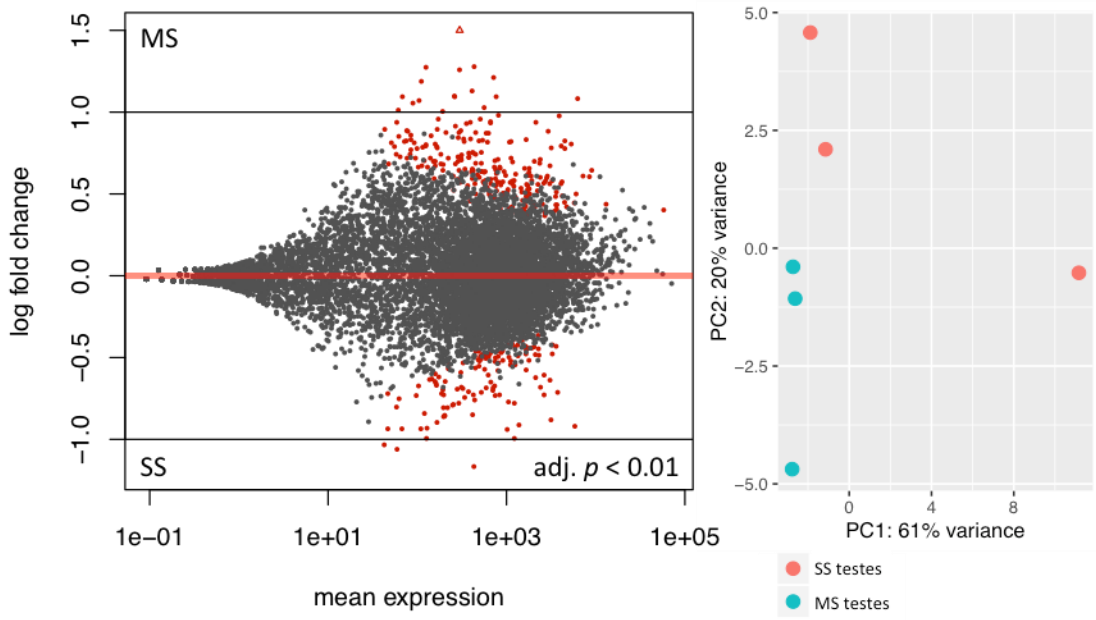


Figure 5.11 Few differentially expressed genes s were identified in comparison of testes from single sex (SS) and mixed sex (MS) infections. A) Log ratio vs mean (MA) plot of the SS testes – MS testes comparison. Genes with an adjusted p-value of < 0.01 are high-lighted in red; triangles indicate genes outside the boundaries of the plots; like other genes, they are red if they are statistically significant, grey if they are not. Genes with a log-fold change of > 1 (fold-change of > 2) and adjusted p-value of < 0.01 were called as differentially expressed "testes genes". B) PCA plot of testes of MS males and testes of SS males. 81% of variation across samples is explained by the two components.

10348 out of 10828 transcripts were found to be expressed (*i.e.* non-zero read counts), of which 2555 (25%) were filtered out due to low levels of expression using the DESeq2 independent filtering algorithm (Figure 5.11). 17 transcripts were found to be significantly up-regulated in testes of MS worms (Table 5.9), whereas only 3 were found to be down-regulated (Table 5.10). None of these genes had been found to be differentially expressed in the comparison of whole males from MS and SS infections.

Gene ID	Description	Average read count	Fold change	Adjusted p-value
Smp_179370	Putative low-density lipoprotein receptor	298.44	3.94	9.67E-13
Smp_105420	Uncharacterised protein	298.24	2.39	7.07E-06
Smp_075800	Legumain preprotein	433.98	2.43	1.20E-05
Smp_103610	Cathepsin B-like cysteine proteinase	718.52	2.33	2.06E-05
Smp_179170	Legumain preprotein	125.36	2.43	7.15E-05
Smp_191250	Uncharacterised protein	412.72	2.19	9.85E-05
Smp_198240	Anosmin-1	193.43	2.01	1.09E-04
Smp_067060	Cathepsin B-like cysteine proteinase	770.84	2.14	3.75E-04
Smp_179970	Uncharacterised protein	111.34	2.28	4.49E-04
Smp_162070	Uncharacterised protein	562.93	2.04	4.84E-04
Smp_084450	Asteroid-like protein	260.89	2.14	4.84E-04
Smp_169920	Translation machinery associated protein	6312.85	2.13	5.20E-04
Smp_199790	α-(1,3)-fucosyltransferase	262.87	2.14	7.52E-04
Smp_004530	Uncharacterised protein	68.27	2.14	2.09E-03
Smp_014570	Saposin-like protein	104.84	2.10	2.09E-03
Smp_139530	Tumour protein p63	89.43	2.08	2.40E-03
Smp_053520	Homeobox protein aristaless-like	61.39	2.03	5.27E-03

Table 5.9: Differentially expressed genes up-regulated in testes of males from mixed sex infections. Average read counts provide a measure of absolute expression for a given gene. The fold changes reflect expression in testes of males from mixed sex infections compared to testes of males from single sex infections. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Description	Average read count	Fold change	Adjusted p-value
Smp_166530	Phospholipase A	431.09	2.23	4.45E-05
Smp_130280	Uncharacterised protein	59.57	2.08	2.72E-03
Smp_018170	Krupple-like zinc finger protein	42.67	2.04	4.12E-03

Table 5.10: Genes up-regulated in testes of males from single sex infections. Average read counts provide a measure of absolute expression for a given gene. The fold changes reflect expression in testes of males from single sex infections compared to testes of males from mixed sex infections. The *p*-value has been adjusted for multiple hypothesis testing.

For this comparison the number of differentially expressed genes was so low that no reliable results could be obtained from the enrichment analysis. All three types of analysis, GO term, KEGG pathway and Pfam domain enrichment, returned statistically significant results for both MS and SS testes, but often these were based on as few as a single gene so that the results were not convincing from a biological point of view.

Follistatin - SmFst

Leutner *et al.* (2013) suggested in their paper on pairing experienced, *i.e.* MS, and inexperienced, *i.e.* SS, males that follistatin or SmFst (Smp_123300) might play a regulatory role in the testes of adult *S. mansoni*. This was based on the observation that SmFst was consistently up-regulated in SS males when compared to MS males (an effect that was replicated in this data set also (2.3x expression in SS males compared to MS males; adjusted p-vale = 1.50E-08), as well as an *in situ* hybridisation experiment suggesting a testes-specific pattern of SmFst expression (Leutner *et al.*, 2013). However, the testes-specific nature of SmFst expression could not be detected using RNA-Seq, instead our data suggested an almost 10-fold higher expression of SmFst outside the testes

(adjusted p-value = 4.29E-11) (Appendix C.17). Furthermore, our data revealed that SmFst expression is not limited to male worms but is similarly down-regulated in MS females (12-fold decrease in MS females; adjusted p-value = 9.46E-31) and ovaries of MS and SS females (9-fold decrease in ovaries from MS worms; adjusted p-value = 8.58E-12) (Appendix C.18).

5.2.7 Effect of pairing on ovary transcriptome

Unlike the small transcriptome changes found in the testes, in ovaries the changes induced by pairing are of a much greater magnitude. Out of 10828 examined genes, 10217 had non-zero read counts of which 1019 genes were filtered out due to low read counts using the DESeq2 independent filtering algorithm (Figure 5.12).

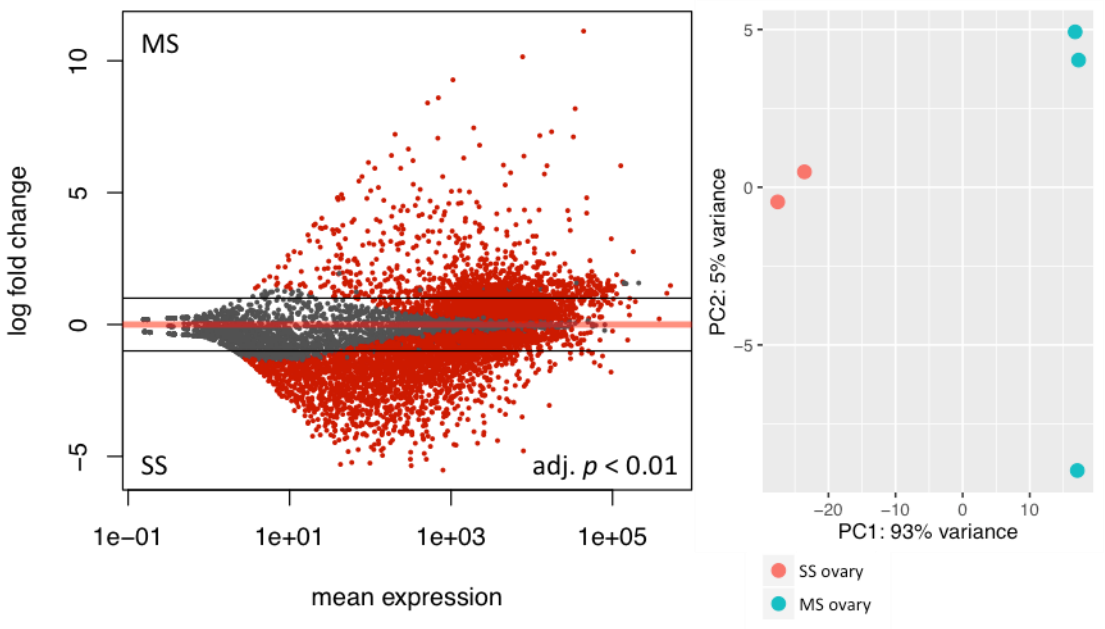


Figure 5.12 Over 2700 DEGs were identified in comparison of ovaries from single sex (SS) and mixed sex (MS) infections. A) Log ratio against mean (MA) plot of the SS ovaries – MS ovaries comparison. Genes with an adjusted p-value of < 0.01 are highlighted in red; triangles indicate genes outside the boundaries of the plots; like other genes, they are red if they are statistically significant, grey if they are not. Genes with a log-fold change of > 1 (fold-change of > 2) and adjusted p-value of < 0.01 were called as differentially expressed "ovary genes". B) PCA plot of ovaries of MS and SS females. 93% of variation across samples is explained by the experimental conditions in this PCA.

Of the remaining genes, 2769 were found to be differentially expressed in response to pairing (966 up-regulated in MS ovaries and 1803 up in ovaries of SS females) (Figure 5.12A). Many of the genes differentially expressed between the ovaries of MS and SS females had not been identified as "ovary genes" in Chapter 5.2.4. Instead they were expressed in the rest of the female worm as well. Out of the 2769 DEGs in this comparison, 574 were "ovary genes". It shows that pairing dependant regulation of gene expression is not limited to "ovary genes" but also affects many genes expressed throughout the body of female worm. Particular

the vitellarian tissue undergoes significant growth and cell proliferation though at an even larger scale than the ovaries.

Initially, I examined the genes up-regulated in MS ovaries (summaries in Table 5.11). The list of GO term these genes were enriched with was dominated by all aspects of translation, but also included "cell-matrix adhesion", "integrin-mediated signalling pathway" and "neurotransmitter transport" (Table 5.12 and Appendix C.19). Female maturation is linked to a neural signal, neuropeptide-Y 8 (Collins & Newmark, 2013) and cell proliferation of oogonia is thought to be stimulated by integrin signalling among others (Beckmann *et al.*, 2012).

Genes up-regulated in MS ovaries coded for proteins that for part of the KEGG pathways responsible for gene expression and protein production: "RNA transport", "RNA polymerase" and "pyrimidine metabolism" indicated a high rate of transcription whereas "ribosome", "ribosome biogenesis" and functionally related pathways such as "aminoacyl-tRNA biosynthesis" represented the translational activity of the mature ovaries, all of which is fuelled by a more highly transcribed "oxidative phosphorylation" pathway (Table 5.13).

When examining enriched protein domains, these same processes were found to be enriched (Table 5.14). Amongst genes up-regulated in MS ovaries were many coding for proteins with domains such as helicase domains and RNA polymerase Rpb2 domains as well as RNA recognition motifs, 50S ribosome-binding GTPase domains, tRNA synthetase class I domains and different ribosomal protein domains (Table 5.14). Additionally, mitochondrial specific domains were

significantly enriched, including ATP synthase subunit C and mitochondrial carrier protein (Table 5.14). Highlighting the mitogenic activity within the mature ovaries, genes coding for proteins with cyclin domains were differentially expressed at higher levels than expected by chance in the MS ovaries. Cyclins play a key role in regulating the cell cycle by controlling the progression from one stage to the next (Sclafani & Holzen, 2007). Comparing the expression levels of all genes annotated as cyclin, cyclin dependant kinases (CDK), or cell division cycle protein (CDC), there were a total of 38 genes that formed two distinct clusters based on their expression (Figure 5.13). This provided further evidence for differential regulation of the cell cycle in MS and SS ovaries.

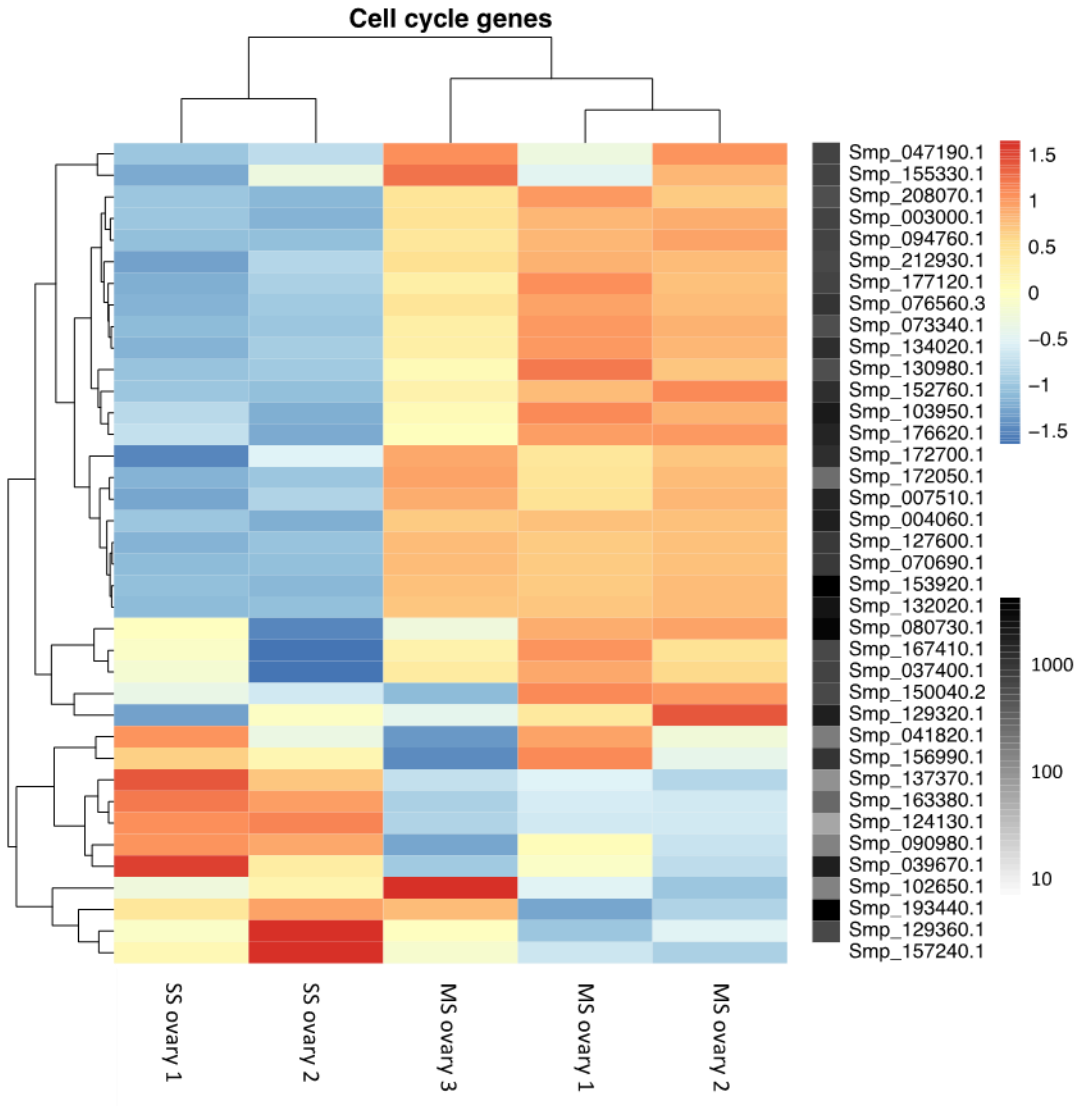


Figure 5.13: Many genes involved in cell cycle regulation, including cyclin, CDK and CDC genes are up-regulated in mix sex ovary samples. Three RNA-seq replicates of mixed sex (MS) ovary (MS ovary 1-3) and two replicates of single sex (SS) ovary (SS ovary 1-2) were used to generate this figure. Colours represent a Z-score of log transformed normalised read counts scaled by row, providing relative measures of gene expression in each sample (measured in standard deviations from the mean). Average normalised read counts for each gene are in the row on the right on a logarithmic scale as a measure of absolute expression levels. See Appendix C.20 for a description of these cell cycle genes.

The MAPK, Rho and TGF-β signalling pathways have been examined closely and are known to play an important role in the development of the female sexual maturity (Andrade *et al.*, 2014; Osman *et al.*, 2006; Santos *et al.*, 2002). To examine the processes that lead to the initiation of the MAPK pathway in the cells of the ovaries and induce maturation of the tissue I compiled a list of the well-established *S. mansoni* genes involved in the signalling network that transmits the male stimulus to the nuclei of ovary cells (Andrade *et al.*, 2014; Beckmann *et al.*, 2010; Knobloch *et al.*, 2007; Oliveira *et al.*, 2012; Osman *et al.*, 2006) and created a heat map of normalised read counts for the ovary samples (Figure 5.14). This set of genes was also clustered into two expression groups, one up-regulated in the ovaries of MS females and one down-regulated when compared to ovaries of female worms from SS infections (Figure 5.14). The signalling network features over 40 members with different regulatory function that, in response to pairing, drive the cell cycle towards DNA replication and the reorganisation of the cytoskeleton to allow for cell division (Beckmann *et al.*, 2010).

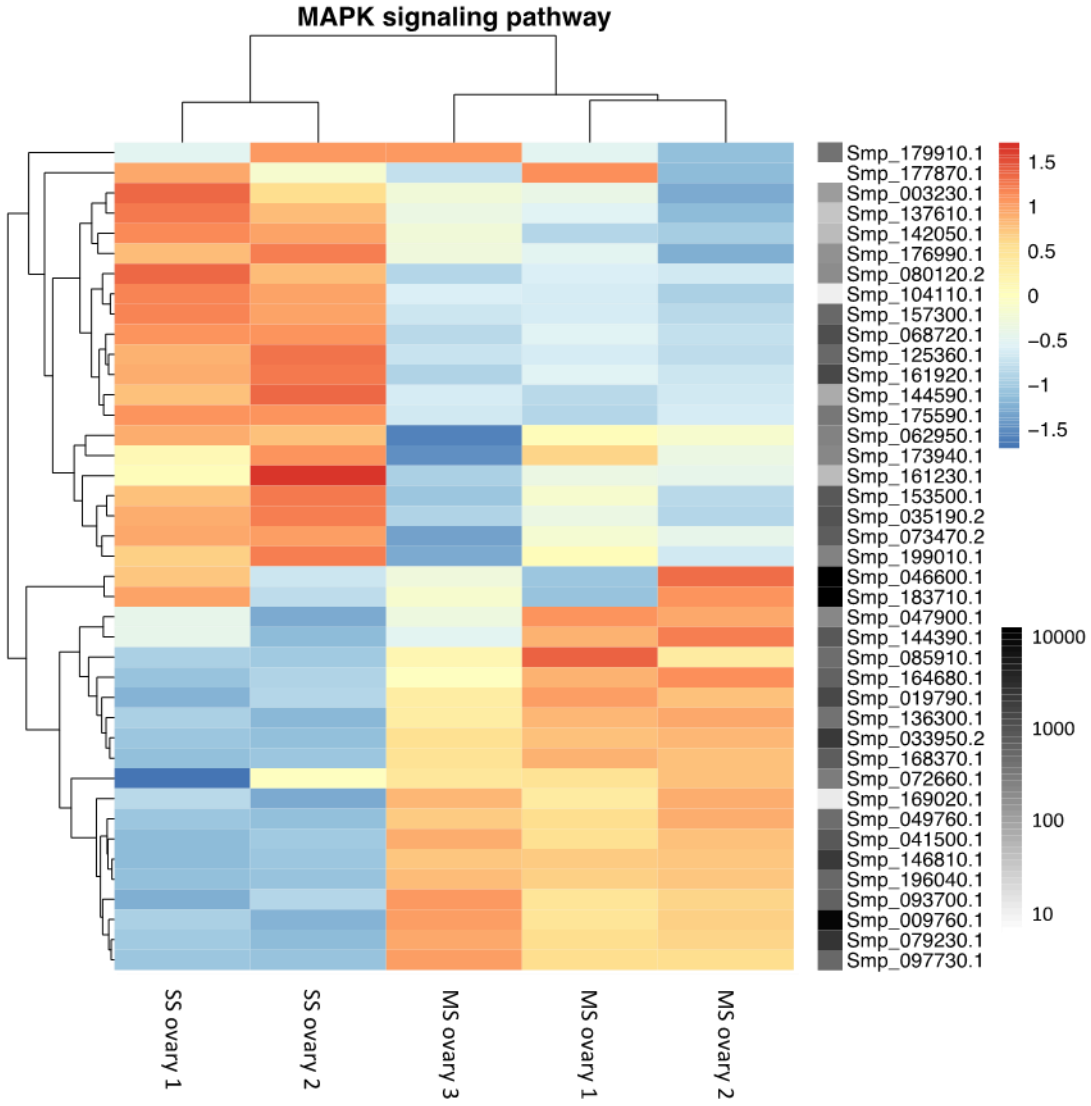


Figure 5.14: Many genes of the MAPK, Rho and TGF-β signalling pathways are differentially expressed in the mixed sex (MS) and single sex (SS) ovary samples. Three RNA-seq replicates of MS ovary (MS ovary 1-3) and two replicates of SS ovary (SS ovary 1-2) were used to generate this figure. Colours represent a Z-score of log transformed normalised read counts scaled by row, giving relative measures of gene expression in each sample. Average normalised read counts for each gene are in the row on the right on a logarithmic scale as a measure of absolute expression levels. See Appendix C.21 for a description of these MAPK signalling genes.

Unlike in Figure 5.13, where the expression of genes coding for cyclins, CDCs and CDKs was compared, the picture emerging from a comparison of genes coding for proteins involved in MAPK, TGFβ and integrin signalling was less clear. Although a large number of these genes were up-regulated in the MS ovaries, a similar number of genes had greater expression in the ovaries of SS females (Figure 5.14). Up-regulation of genes in the SS ovaries was by no means limited to genes coding for proteins with inhibitory functions but included genes coding for some of the key regulators of the MAPK signalling pathway such as homologues of Rho, Ras-guanine nucleotide exchange factor, Raf and Erk (Figure 5.15). Others, such as genes coding for growth factor receptor-bound protein 2 and Ras, showed no differentially expression (Figure 5.15). In contrast, the genes coding for SMAD2 and 4 as well as Transforming Growth Factor-β Receptor II, Ras-GAP and Scr-cytoplasmic tyrosine kinase homologues were found to be up-regulated in MS ovaries (Figure 5.15). Ultimately however, the induction of cell proliferation is dependent on receiving specific extracellular stimuli provided by pairing as well as host factors to initiate signalling and mitogenic activity and cell differentiation (Beckmann *et al.*, 2010).

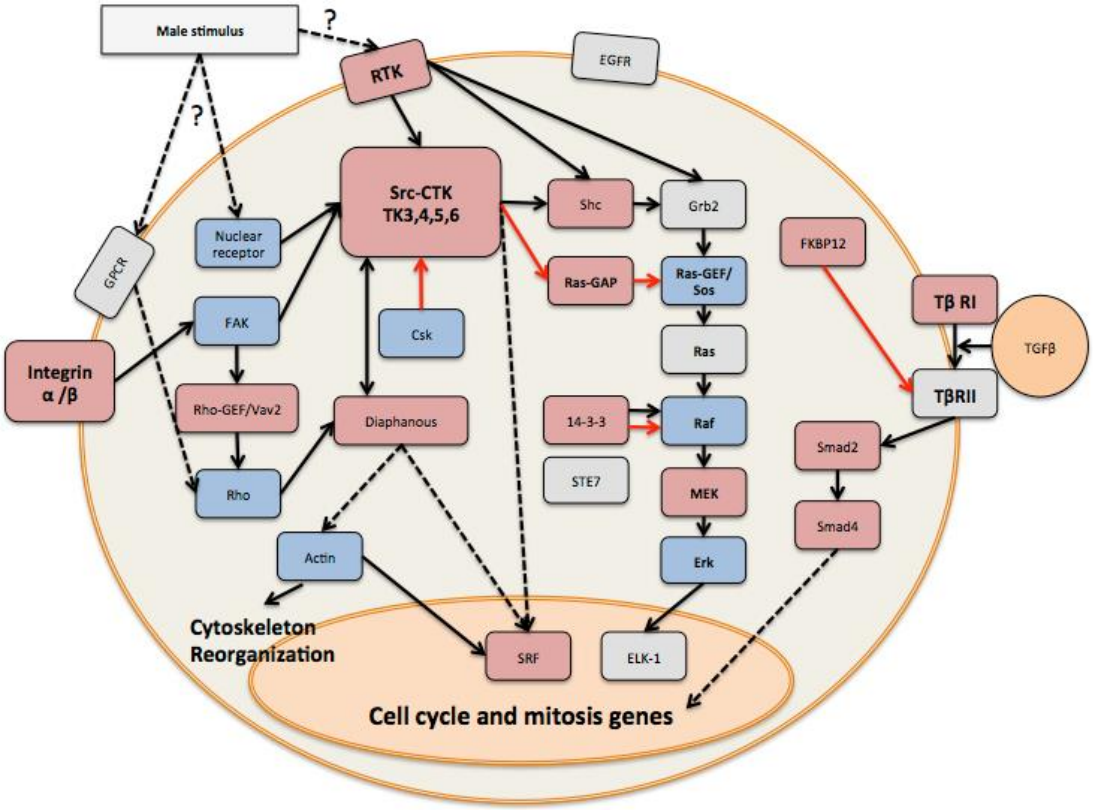


Figure 5.15: Many members of the MAPK, Rho and TGF-β signalling pathway are differentially expressed in ovaries. Adapted from Knobloch *et al.* (2007). Dotted arrows indicate indirect interactions, solid arrows indicate direct interaction. Red arrows indicate inhibition and black arrows activation of the target. Red coloured genes are up-regulated in the ovaries of MS females, blue coloured genes are up-regulated in the ovaries of SS females, grey coloured genes are not differentially expressed.

Gene ID	Description	Average Read Count	Fold change	Adjusted p-value
Smp_095350	Zinc finger protein	16607.90	985.26	0
Smp_070360	Cytoplasmic polyadenylation element-binding protein	3765.62	576.03	4.63E-153
Smp_145490	Putative poly-rC-binding protein	7793.72	557.04	2.12E-298
Smp_148390	Uncharacterised protein	13636.61	220.84	0
Smp_164140	Uncharacterised protein	428.29	191.09	2.49E-43
Smp_129920	Sodium- and chloride-dependent transporter	624.76	163.98	1.97E-59
Smp_141570	Potassium channel of the TWiK family	278.68	154.52	5.77E-34
Smp_013540	Tyrosinase	1012.14	138.29	2.77E-96
Smp_172980	Uncharacterised protein	6231.87	112.03	0
Smp_176940	Permease solute carrier	1183.57	98.93	1.48E-125
Smp_134490	Thyroid receptor α	395.32	96.86	1.47E-49
Smp_129960	Uncharacterised protein	14093.01	87.57	0
Smp_080360	Uncharacterised protein	3841.06	82.10	0
Smp_154390	Uncharacterised protein	2326.47	75.57	2.65E-221
Smp_145400	Putative Rcd1-domain protein	730.32	72.26	6.68E-88

Table 5.11: Most differentially expressed genes up-regulated in the ovaries of females from mixed sex infections compared to ovaries from single sex infections. Average read counts provide a measure of absolute expression for a given gene. The fold changes reflect expression in ovaries of females from mixed sex infections compared to ovaries of females from single sex infections. The *p*-value has been adjusted for multiple hypothesis testing. Some p-values were calculated to be equal to 0, highlighting how statistically significant the differential expression of these genes was.

GO term	Description	Total Genes	DEGs	Expected	p-value
GO:0006412	Translation	393	83	41.42	2.000E-09
GO:0042254	Ribosome biogenesis	137	31	14.44	5.900E-06
GO:0006414	Translational elongation	27	10	2.85	4.600E-04
GO:0007160	Cell-matrix adhesion	8	4	0.84	6.040E-03
GO:0006446	Regulation of translational initiation	5	3	0.53	9.890E-03
GO:0007229	Integrin-mediated signalling pathway	9	4	0.95	9.980E-03
GO:0006836	Neurotransmitter transport	30	8	3.16	1.038E-02
GO:0006413	Translational initiation	37	11	3.9	1.510E-02
GO:0006364	rRNA processing	108	19	11.38	2.259E-02
GO:0008033	tRNA processing	61	10	6.43	2.631E-02
GO:0006633	Fatty acid biosynthetic process	51	10	5.37	3.746E-02

Table 5.12: GO terms enriched amongst the genes up-regulated in ovaries of females from mixed sex (MS) infections compared to ovaries from single sex infections (see above). The column "Total Genes" provides the number of genes associated with a particular GO term. "DEGs" provides the number of genes associated with a given GO term that had significantly higher expression in ovaries of MS females; "Expected Genes" provides the number of differentially expressed genes associated with a given GO term that are expected by chance. The GO terms are all in the "Biological Process" category. See Appendix C.19 for the complete list.

Pathway	Description	Total Genes	DEGs	Expected	p-value
smm03010	Ribosome	110	91	55.90	7.516E-13
smm03008	Ribosome biogenesis in eukaryotes	62	57	31.51	1.798E-12
smm03013	RNA transport	103	73	52.34	1.118E-05
smm00190	Oxidative phosphorylation	69	52	35.06	1.377E-05
smm03020	RNA polymerase	24	20	12.20	7.794E-04
smm00970	Aminoacyl-tRNA biosynthesis	34	23	17.28	1.972E-02
smm00240	Pyrimidine metabolism	59	35	29.98	4.411E-02

Table 5.13: KEGG pathways enriched amongst the genes up-regulated in the ovaries of females from mixed sex infections compared to ovaries from single sex infections (see above). The column "Total Genes" provides the number of genes in the KEGG pathway. "DEGs" provides the number differentially expressed genes (DEGs) in that pathway; "Expected Genes" provides the number of DEGs in a given pathway that is expected by chance. See Appendix C.22 for the complete list.

Domain	Description	Total Genes	DEGs	Expected	p-value
PF00270	DEAD/DEAH box helicase	55	33	18.16	2.375E-05
PF14259	RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	29	19	9.57	2.547E-04
PF00153	Mitochondrial carrier protein	26	16	8.58	1.896E-03
PF00271	Helicase conserved C-terminal domain	75	35	24.76	4.477E-03
PF13893	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	13	9	4.29	6.669E-03
PF01926	50S ribosome-binding GTPase	11	8	3.63	6.965E-03
PF00137	ATP synthase subunit C	4	4	1.32	1.186E-02
PF00579	tRNA synthetases class I (W and Y)	4	4	1.32	1.186E-02
PF00134	Cyclin, N-terminal domain	13	8	4.29	2.442E-02
PF00004	ATPase family associated with various cellular activities (AAA)	34	16	11.22	3.222E-02
PF00164	Ribosomal protein S12/S23	3	3	0.99	3.595E-02
PF00380	Ribosomal protein S9/S16	3	3	0.99	3.595E-02
PF00562	RNA polymerase Rpb2, domain 6	3	3	0.99	3.595E-02

Table 5.14 Pfam domains enriched amongst the genes up-regulated in ovaries the of females from mixed sex infections compared to ovaries from single sex infections (see above). The column "Total Genes" provides the number of genes with a given Pfam domain. "DEGs" provides the number of differentially expressed genes with that domain; "Expected" provides the number of DEGs with a given domain that is expected by chance. See Appendix C.23 for an expanded list.

Next, I examined the genes upregulated in SS ovaries, the most differentially expressed of which are summarised in Table 5.15. Genes up-regulated in the ovaries of SS females were enriched for a total of 38 GO terms (Table 5.16). Interestingly these GO terms included "spermatogenesis" and "male sex determination" (Table 5.16), two biological processes that intuitively are associated with testes rather than ovaries. In the case of the "male sex determination" the genes in question included negative regulators of transcription (Logan-Klumpler *et al.*, 2012) via the MAPK signalling pathway such as a capicua (Smp_142600) and a SOX family transcription factor homologue (Smp_148110) (Hargrave *et al.*, 2000; Jimenez *et al.*, 2012) as well as

two members of the Wnt signalling pathway (Logan-Klumpler *et al.*, 2012) responsible for regulating embryogenesis in *D. melanogaster*, a pangolin (Smp_131260) and transcription factor 7-like-2 (Smp_148190) homologue (Jin & Liu, 2008; Ravindranath & Cadigan, 2014) which also play a role in the ovaries. The "spermatogenesis" genes include three homologues of the gene coding for the "spermatogenesis associated proteins" (Logan-Klumpler *et al.*, 2012) but their function is not necessarily involved exclusively in spermatogenesis; they fulfil varied roles, such as apoptosis regulation and degradation of mitochondrial proteins (Jiang *et al.*, Wang, 2013; Puri *et al.*, 2011). Additionally, two nanos homologues (Smp_051920, Smp_055740) are included (Logan-Klumpler *et al.*, 2012), which play an important role in spermatogenesis but were also shown to be involved in oocyte development by maintaining the adult germlines in *D. melanogaster* (Bhat, 1999). Furthermore the "spermatogenesis" GO term included two lin54 protein homologues (Smp_022870, Smp_178740) (Logan-Klumpler *et al.*, 2012), which are both involved in the regulation of the cell cycle. They are components of the DREAM complex, which regulates gene expression at various points throughout the cell cycle (Schmit *et al.*, 2009). Other interesting GO terms included the "Wnt signalling pathway", "cell cycle arrest", and "cell differentiation" (Table 5.16). These GO terms were further evidence for the strict control of cell proliferation in the ovaries of SS females. Finally, a significant proportion of genes annotated with the GO terms "Fatty acid biosynthetic process" and "Sphingolipid metabolism" were up-regulated (Table 5.16).

Next, a pathway analysis was performed using the KEGG reference pathways for *S. mansoni.* The aim was to identify any pathways containing more up-regulated

genes than expected by chance. The pathway analysis of the SS ovary genes revealed 14 pathways to be significantly up-regulated (Table 5.17). Metabolic pathways, including amino acid metabolism, carbohydrate metabolism and lipid metabolism notably including "sphingolipid metabolism" were found to be enriched with DEGs (Table 5.17). Several developmental signalling pathways were also found to be over-represented among the DEGs, such as Wnt signalling and dorsal-ventral axis formation, similar to the GO terms found up-regulated in the SS ovaries (Table 5.17). The SS ovary genes were also examined for Pfam domain enrichment with 77 different domains found to be enriched (Table 5.18). Among them were genes coding for domains that control apoptosis, such as the Bcl-2 domain and growth factor domains, such as Calcium-binding EGF domains, complement clr-like EGF-like domains and growth factor receptor domain IV (Table 5.18).

Since Galanti *et al.* (2012) observed apoptosis to be occurring at a higher rate in the vitellarian tissue of SS females compared to MS females, I examined whether similar activity could be detected in the ovary transcriptome. I first noted that all four genes known to code for Bcl-2 domains were up-regulated in ovaries of SS females (Table 5.18). Furthermore, ceramide, one of the central metabolites of the sphingolipid metabolism pathway that was significantly up-regulated in the ovaries of SS females (Table 5.17), is also an important regulator of apoptosis (Bose *et al.*, 1995; Jensen *et al.*, 2014; Separovic *et al.*, 2012). Bose *et al.* (1995) showed that ceramide synthase mediates an alternative pathway of apoptosis activation. Sparovic *et al.* (2012) showed that human cells became more resistant to apoptosis when ceramide synthase was knocked down using RNAi.

Finally, Jensen *et al.* (2014) discovered that a member of the Bcl-2 family, a group of apoptosis regulating protein, can inhibit ceramide synthase. The mitochondria associated protein Bcl-2-like 13 blocks apoptosis by inhibiting ceramide synthase, thereby linking sphingolipid metabolism and the Bcl-2 regulated apoptosis machinery (Jensen *et al.*, 2014). In the ovaries of SS females, a homologue of ceramide synthase (Smp_042440) was 2.5-fold up-regulated (adjusted p value = 1.74E-07) (Table 5.15) and as can be seen in Figure 5.16, several other genes involved in the biosynthesis of ceramide were up-regulated as well.

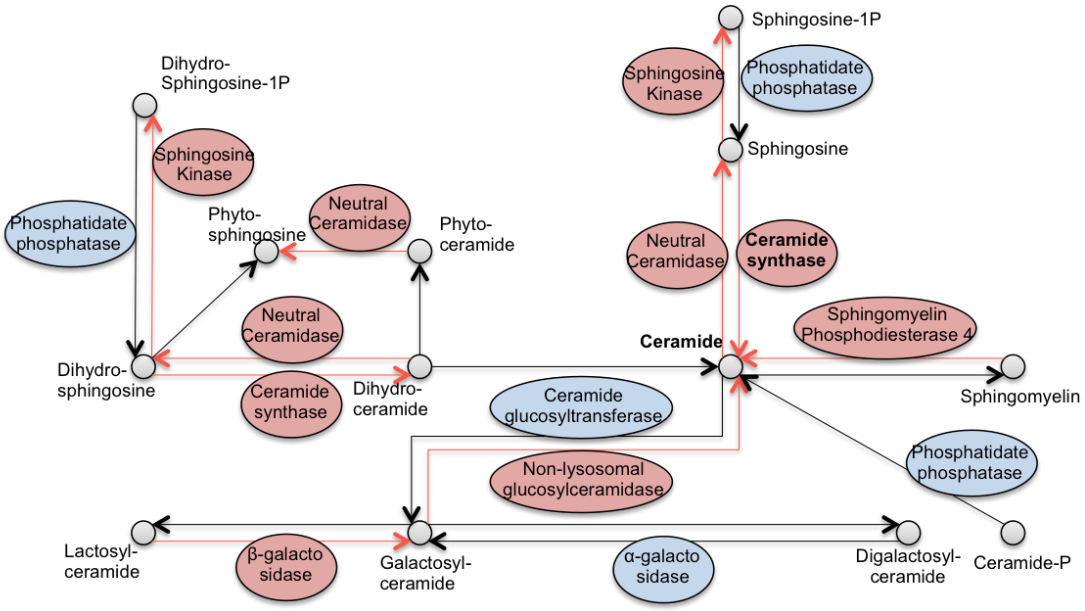


Figure 5.16: Many genes involved in sphingolipid metabolism are differentially expressed in ovaries from single sex (SS) and mixed sex (MS) infections. The genes encoding red coloured enzymes are up-regulated in ovaries of SS females, genes coding for blue coloured enzymes are down-regulated in the ovaries of SS females. Arrows indicate the catalysed reaction. Adapted from KEGG reference pathway. URL: http://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=smm00600&keyword=sphingolipid.

I used the list of 28 putative apoptosis-related genes from Chapter 4.2.2 to examine the apoptosis in the ovaries. These genes had been identified using a combination of previously published literature (Lee *et al.*, 2011, Lee *et al.*, 2014; Peng *et al.*, 2010) as well as Interpro Scan, Pfam and BLAST comparison (see Chapter 2.4.12 for method; see Appendix B2-10 for more details). Several groups of genes, such as cytochrome C genes (21 genes in total), known to be involved in apoptosis were excluded from this heat map due to the size of these gene families and due to their role in many other cellular processes. In total, six out of 13 putative pro-apoptotic factors were significantly up-regulated in the ovaries of SS females, whereas only one was significantly up-regulated in the ovaries of MS females (six not significantly differentially expressed) (Appendix C.24). In contrast, six out of nine putative anti-apoptotic genes were significantly up-regulated in the ovaries of MS females and only one in the ovaries of SS females (two were not significantly differentially expressed) (Appendix C.24). Of the six genes for which it was unknown whether they are pro- or anti-apoptotic two – Smp_168470 & Smp_148130 – were significantly up-regulated in the ovaries of SS females and four – Smp_043360, Smp_022110, Smp_084610 & Smp_196040 – were up-regulated in the ovaries of MS females. If the assumption that pro-apoptotic genes are up-regulated in SS ovaries and anti-apoptotic genes are down-regulated is correct, Smp_043360 (sBH3), Smp_022110 (p21 activated protein kinase interacting protein), Smp_084610 (FAS associated factor 2) and Smp_196040 (SHC transforming protein 3) may have anti-apoptotic activity, whereas Smp_168470 (Bcl-2 family protein) and Smp_148130 (FAS binding factor) could have pro-apoptotic functions (see Figure 5.17). However, to

confirm or dismiss these predictions careful characterisation of these genes using RNAi or other methods should to be performed.

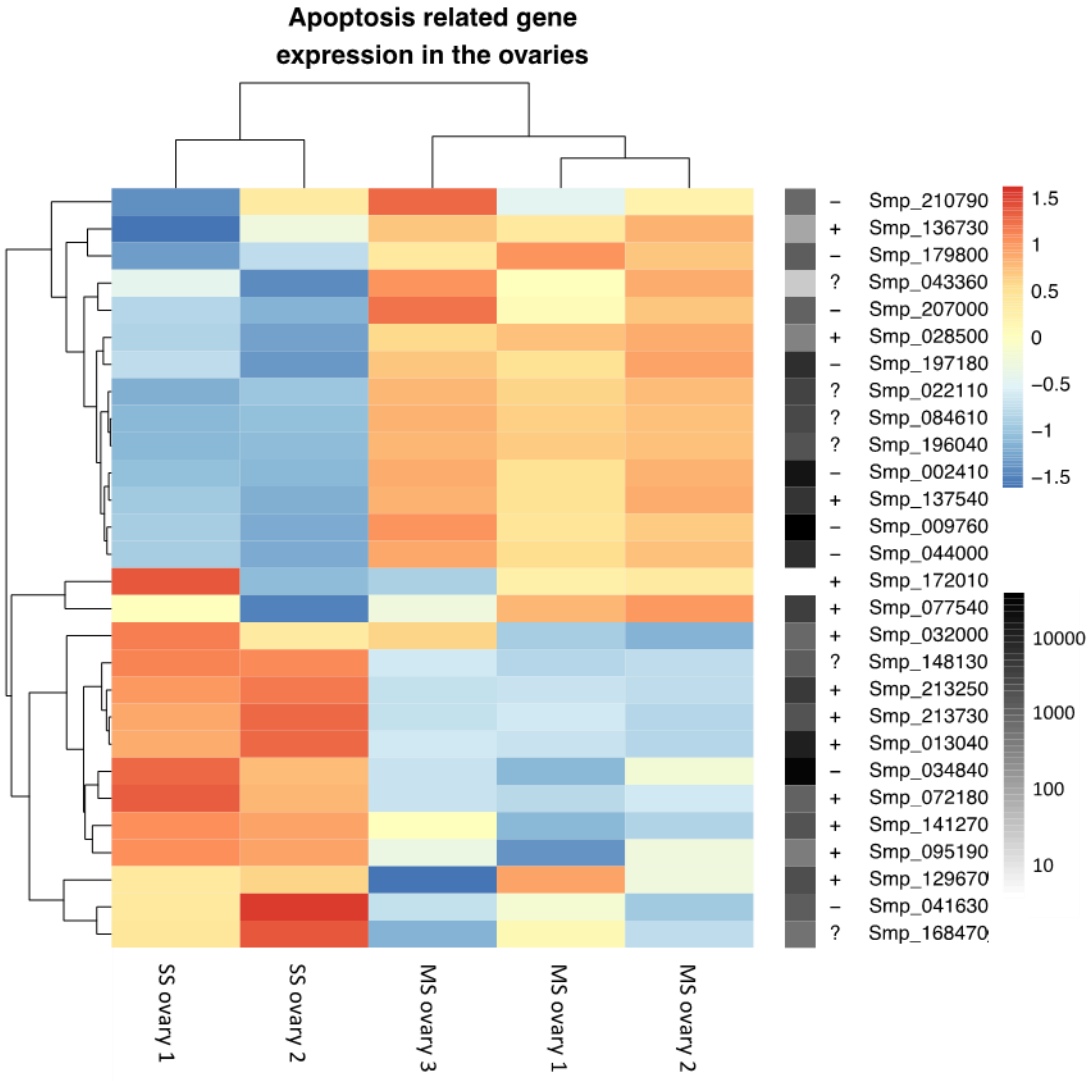


Figure 5.17 Expression of apoptosis-related genes points to higher levels of apoptosis in the ovaries of females from single sex (SS) infections than those of females form mixed sex (MS) infections. Three RNA-seq samples of MS ovary (MS ovary 1-3) and two SS ovary (SS ovary 1-2) were used to generate this figure. Colours represent a Z-score of log transformed normalised read counts scaled by row, giving relative measures of gene expression in each sample. Average normalised read counts for each gene are in the row on the right on a logarithmic scale as a measure of absolute expression levels. Samples were clustered by similarity of expression patterns, forming the expected groups of single sex and mixed sex ovaries. Genes were also clustered by expression. Genes were either categorised as pro-apoptotic (+), anti-apoptotic (-) or with unknown regulatory function (?).See Appendix C.24 for a description of these putative apoptosis-related genes and their expression in ovaries.

Gene ID	Description	Average Read Count	Fold change	Adjusted p-value
Smp_079970	Tubulin-β chain	31.51	44.38	5.27E-16
Smp_079960	Tubulin-β chain	41.05	37.93	2.23E-17
Smp_135500	Phosphodiesterase	387.74	36.16	9.33E-101
Smp_127170	Putative potassium channel	147.33	35.77	6.53E-51
Smp_038730	Uncharacterised protein	4801.90	34.18	3.69E-30
Smp_184180	Uncharacterised protein	213.38	32.46	2.69E-33
Smp_022480	Putative transient receptor cation channel	32.34	32.29	8.95E-14
Smp_133900	Indoleamine 2,3-dioxygenase	91.77	30.77	6.21E-31
Smp_168610	Putative suppression of tumorigenicity protein	73.08	30.49	9.33E-26
Smp_182770	Putative neuroblast differentiation associated protein	5961.91	28.93	1.00E-44
Smp_173070	Invasolysin-like protease	247.08	28.56	5.82E-62
Smp_162190	Cystein and glycine-rich protein	61.27	28.34	3.58E-19
Smp_180890	Rho/GAP domain-containing protein	237.53	28.33	2.78E-41
Smp_128250	Uncharacterised protein	158.60	27.06	4.93E-45
Smp_174340	Frizzled protein	82.08	26.62	9.44E-22
Smp_042440	Ceramide synthase	151.44	2.56	1.74E-07

Table 5.15: Most differentially expressed genes up-regulated in the ovaries of females from single sex (SS) infections compared to ovaries from mixed sex (MS) infections. Average read counts provide a measure of absolute expression for a given gene. The fold changes reflect expression in ovaries of SS females compared to ovaries of MS females. The *p*-value has been adjusted for multiple hypothesis testing.

GO ID	Description	Total Genes	DEGs	Expected	p-value
GO:0030154	Cell differentiation	78	29	11.89	2.000E-06
GO:0007283	Spermatogenesis	10	5	1.52	1.053E-02
GO:0016055	Wnt signalling pathway	27	9	4.12	3.073E-02
GO:0007050	Cell cycle arrest	6	3	0.91	4.934E-02
GO:0030238	Male sex determination	6	3	0.91	4.934E-02

Table 5.16 GO terms enriched in genes up-regulated in ovaries of ovaries of females from single sex (SS) infections compared to ovaries from mixed sex (MS) infections (see above). The column "Total Genes" provides the number of genes associated with a particular GO term. "DEGs" provides the number of genes associated with a given GO term that had significantly higher expression in ovaries of SS females; "Expected Genes" provides the number of differentially expressed genes associated with a given GO term that are expected by chance. The GO terms are all in the "Biological Process" category. See Appendix C.25 for an expanded list.

Pathway	Description	Total Genes	DEGs	Expected	p-value
smm00330	Arginine and proline metabolism	18	9	2.84	5.943E-04
smm00052	Galactose metabolism	14	6	2.21	1.152E-02
smm00600	Sphingolipid metabolism	15	6	2.37	1.621E-02
smm04320	Dorso-ventral axis formation	8	4	1.26	2.164E-02
smm04310	Wnt signalling pathway	35	10	5.52	2.345E-02
smm00561	Glycerolipid metabolism	17	6	2.68	2.854E-02
smm00500	Starch and sucrose metabolism	14	5	2.21	4.143E-02
smm00062	Fatty acid elongation	6	3	0.95	4.675E-02
smm00564	Glycerophospholipid metabolism	35	9	5.52	4.878E-02

Table 5.17: KEGG pathways enriched for in genes up-regulated in ovaries of females from single sex infections compared to ovaries from mixed sex infections (see above). The column "Total Genes" provides the number of genes in the KEGG pathway. "DEGs" provides the number differentially expressed genes in that pathway; "Expected Genes" provides the number of differentially expressed genes in a given pathway that is expected by chance. See Appendix C.26 for the complete list.

Domain	Description	Total Genes	DEGs	Expected	p-value
PF07645	Calcium-binding EGF domain	7	6	1.94	2.310E-03
PF13405	EF-hand domain	12	8	3.33	4.741E-03
PF00452	Apoptosis regulator proteins, Bcl-2 family	4	4	1.11	5.938E-03
PF14843	Growth factor receptor domain IV	3	3	0.83	2.140E-02
PF12662	Complement Clr-like EGF-like	3	3	0.83	2.14E-02

Table 5.18 Pfam domains enriched amongst the in genes up-regulated in ovaries of females from single sex infections compared to ovaries from mixed sex infections (see above). The column "Total Genes" provides the number of genes encoding a given Pfam domain. "DEGs" provides the number of differentially expressed genes with that domain; "Expected" provides the number of differentially expressed genes encoding a given domain that is expected by chance. See Appendix C.27 for an expanded list.

5.3 Discussion

When examining the published results (Hahnel *et al.*, 2014) on differential expression in mature (MS) and immature (SS) gonads, differences emerged both in terms of direction and magnitude of the observed changes. These discrepancies could be resolved by taking into account changes in the expression of the reference gene, actin, which was shown to be distinctly differentially expressed across mature and immature ovaries. After applying this calibration, both sets of data agreed well, not only in terms of the direction of differential expression but also the magnitude of the changes measured, strengthening the confidence that can be placed in this data set as well as highlighting the importance of choosing appropriate internal reference genes for qPCR analysis.

There are several papers discussing the role of pairing-induced signalling through various pathways on actin and its role in the remodelling of the cell cytoskeleton (Beckmann *et al.*, 2010; Faix & Grosse, 2006; Knobloch *et al.*, 2007; Olson, 2003; Wallar & Alberts, 2003). As a result, differential expression of actin between ovaries of SS and those of MS females is not entirely unexpected, although the magnitude of the change was surprising. This analysis suggests that actin should be used very carefully as a reference gene for qRT-PCR. Instead a more rational choice of internal reference gene, such PSMD4, another gene recommended by Liu *et al.* (2012), may more suitable, especially when comparing samples from different life cycle stages, different genders or different pairing statuses.

Comparing the number genes identified to be gonad specific by Nawaratna *et al.* (2011) to this analysis several differences become apparent. Nawaratna *et al.* (2011) identified 2141 probes, corresponding to 1188 genes, to be testes-specific and 4352 probes, 1989 genes, to be ovary-specific. In the present study, 1949 testes-specific and 1272 ovary-specific genes were found, despite using the same threshold for gonad specificity (a two-fold higher expression in the gonads compared to the whole worm control). In this context it is interesting to note that Nawaratna *et al.* (2011) had two biological replicates for testes, but only one sample for ovaries (and only single sample for both male and female whole worm controls). This may explain some of the discrepancy between the microarray and RNA-Seq results.

The transcriptome analysis of ovaries and testes reflected their known biological functions closely with both highly enriched for genes involved in DNA replication and cell proliferation/division genes as suggested by GO term, KEGG pathway and Pfam domain enrichment analysis. The domain enrichment analysis highlighted a significant enrichment of genes involved in faithful DNA replication compared to the whole worm transcriptome, showing all genes with MutS and BRCA domains being expressed at higher levels in the gonads than the whole worm controls. These findings do not come as a great surprise, given that germ cell maintenance as well as spermatocyte and oocyte production are the main function of the testes and ovaries respectively. Similar to Nawaratna *et al.* (2011), this analysis showed cyclins, CDCs and CDKs to be expressed at higher levels in the gonads than in the somatic tissues of the adult worms.

Few changes in gene expression were found in the testes in response to pairing. This is in good agreement with the literature that suggests little change except a small shrinkage of testes, which was observed in MS male worms. Follistatin, which was hypothesised by Leutner *et al.* (2013) to be testes specific, was not found to be expressed at higher levels in the testes than whole males nor was it found to be differentially expressed in the testes of MS and SS males. Nonetheless a down-regulation of follistatin in MS males was observed in this experiment, as had been previously reported by Leutner *et al.* (2013). This suggests that SmFst plays a role in the response of male worms to pairing that is unrelated to the testes. Rather its function appears to take place in the somatic tissues of male worms, which was unexpected.

The small transcriptome changes in testes stood in stark contrast to those in the ovaries. In the ovaries of MS females, clear up-regulation of cell cycle regulating genes was seen, including cyclins, CDCs and CDKs, which play a pivotal in the tight control of DNA replication and repair in eukaryotes (Sclafani & Holzen, 2007) to allow for successful mitosis and meiosis. The higher expression of those genes in MS females emphasised the higher rate of cell division and proliferation in females worms from MS infections compared to female worms from SS infections, as expected from our current knowledge of the effect of pairing on the ovaries.

In addition to the cell cycle regulating proteins, many genes involved in Rho and MAPK signalling pathway - known to regulate proliferation and maturation of

ovaries in response to pairing (Beckmann *et al.*, 2010) - were differentially regulated between SS and MS ovaries. Although no clear pattern of up- or down-regulation of this group of genes emerged, the large, statistically significant proportion of these genes expressed differentially between immature and mature ovaries, reinforces the key role that the MAPK pathway plays in regulating ovary growth. As discussed in the introduction, Rho is thought to be an important member of the group of genes regulating cell proliferation in the ovaries in response to pairing (Beckmann *et al.*, 2010). Rho signalling allows cells to respond to a variety of extracellular signals, including growth factors, cytokines and hormones (Schwartz, 2004). This could be part of the mechanism that regulates the pairing-induced fertility in female worms.

The changes that took place in the ovary transcriptome in response to pairing were also highly suggestive of apoptosis playing an important role in controlling ovary maturation before pairing. Two important aspects of this process are the sphingolipid metabolism (Tirodkar & Voelkel-Johnson, 2012) and secondly a set of proteins including Bcl-2 domain containing proteins, caspases, BAX domain containing proteins and apoptosis inhibitors (Lee *et al.*, 2011). The sphingolipid metabolism pathway was significantly up-regulated in SS females, with ceramide synthase, a known positive regulator of apoptosis in human cells (Separovic *et al.*, 2012), being 2.5-fold more expressed in SS ovaries than MS ones. On the other hand, neutral ceramidase, known to have a protective effect on cells in rat cells (Zhu et al., 2008), was also up-regulated, suggesting that proliferation and apoptosis were carefully balanced in SS ovaries.

Further evidence of the important role apoptosis plays in the regulation of ovary proliferation before pairing was based on classification of apoptosis genes in schistosomes by Lee *et al.* (2014) as well as BLAST and Pfam evidence. Many putative pro-apoptotic genes were found to be expressed in the ovaries of SS females in contrast to the high level of expression of putative anti-apoptotic genes in the ovaries of MS females. These findings were consistent with the high levels of apoptosis reported in vitellarian tissue of SS females by Galanti *et al.* (2012) and point to a common regulatory mechanism for the proliferation of both these tissues, which are intimately involved in egg production.

Chapter 6

Concluding remarks

Sexual maturation of female schistosomes is critically dependant on stimulation by a male partner, as is well established (Basch, 1991). In the absence of a male stimulus, the female worms remain developmentally stunted without mature ovaries and vitellaria on which egg production depends. While the differences between fully mature females and immature females had already been explored extensively [for example Fitzpatrick & Hoffmann (2006) and Waisberg *et al.* (2007)] when I started working on this thesis, a time course of gene expression, measured using RNA-Seq, over the course of female development was not available. This time course revealed the changes in gene expression as they occurred over time after pairing commenced, rather than just the gene expression of fully mature females. Many studies have examined the signalling cascade that leads to the proliferation of cells in the female reproductive organs (see Chapter 1.5) and there was already evidence for the involvement of neuropeptides (Collins *et al.*, 2010) as well as apoptosis (Galanti *et al.*, 2012) playing an important role in female development, and particularly regulating cell proliferation in female reproductive organs.

Recently, several papers examined the changes in genes expression over the course of schistosome maturation (Cai *et al.*, 2016; Wang *et al.*, 2017). One of their aims was to discover the precise stimulus used by male worms to induce maturation in females. The nature of this stimulus has been the subject to research and speculation of years but has remained elusive (Chapter 1.4). Wang *et al.* (2017) used their expression data to identify genes whose expression was correlated with paring or vitelline development. Wang *et al.* (2017) hypothesised that an aromatic-L-amino acid decarboxylase (AADC) gene, which they showed

to be expressed in the gynecophoral canal of male worms using WISH, was responsible for the stimulation of female worms. The authors did not present RNAi evidence showing that knock down of AADC prevented female maturation as they did for another gene in the same paper. Thus, the evidence for AADC inducing female fertility seems somewhat circumstantial at this point, as it appears to be based on a correlation of AADC expression with female fertility, as well as the expression in the gynecophoral canal. Furthermore, data from Chapter 3 showed that the *S. mansoni* homologue of AADC is expressed in a paring-dependant manner. However, that would raise the question of how male worms are stimulated to express AADC. More concrete evidence, such as RNAi experiments, similar to that provided for the allatostatin receptor in the same publication (Wang *et al*., 2017), would greatly strengthen this hypothesis.

This thesis examined the events leading to maturation of female worms starting with pairing and the male-female interaction. This is an important aspect of schistosome biology for many reasons. In particular, if untreated, adult schistosomes survive for years inside the mammalian host, and their egg production is responsible for much of the pathology associated with schistosomiasis as well as its transmission. Hopefully the information gathered here can help to identify new avenues in the search for novel drug targets and therapies. Praziquantel, the current drug of choice for the treatment of schistosomiasis, is not effective against developing parasites and also faces the challenge of emerging resistance (Crellen *et al.*, 2016).

Divergent development of males and females

In Chapter 3, the development of male and female worms was examined at the transcriptome level, to determine when changes in gene expression take place and how they differ in worms from mixed sex and single sex infections. Male and female *S. mansoni* are still highly similar at the transcriptome level at 18 d.p.i. with only 38 DEGs identified between them. As pairing commences, their differences increase to 173 DEGs at 28 d.p.i. and continues to 2194 DEGs at 35 d.p.i. when sexual maturation of the female worms has significantly progressed. Both male and female worms expressed genes involved in the cytochrome P450-dependent drug metabolism pathway at higher levels at 18 d.p.i. than at later time points. Of these genes, glutathione S transferases (including Smp_024010 and Smp_102070) in particular are thought to play a key role in the mediation of xenobiotic metabolism in schistosomes, and have been shown to bind praziquantel at a non-substrate site (Cardoso *et al.*, 2003). Their high expression levels at the earliest point during the time course suggests that these glutathione S transferases may be responsible for the resistance of immature worms to praziquantel. Expression of genes involved in spermatocyte production such as the testes expressed transcription factors *boule* (Smp_144860) and basonuclin 2 (Smp_138350) was found to change from 18 d.p.i. onwards in male worms. The expression of boule, which in *Schmidtea mediterranea* is responsible for meiotic progression of male germ cells, rose significantly over the following time points. In contrast, the expression of basonuclin 2, which prevents premature maturation of spermatocytes in mammals, was found to decline significantly after 18 d.p.i.

Relatively few differences were found between males of mixed sex and single sex infections during their development. A notable exception to this was an aromatic-L-amino acid decarboxylase (AADC) described by Wang *et al.* (2017) (see above). Female worms are well known to up-regulate fertility related genes such as eggshell proteins and tyrosinases once paired, and this was observed consistently from 28 to 38 d.p.i. in mixed sex infections. A large number of transcription factors became down-regulated around the time pairing commences in females from both mixed sex as well as single sex infections. This is interesting because it showed that the down-regulation occurs in a pairing-independent fashion (Chapter 3.2.5). After an approximately three week period of maturation the female worms become developmentally arrested, which coincides with this down-regulation of transcription factors. Only after pairing, three zinc-finger transcription factors (Smp_166560, Smp_087320 and Smp_095350) became up-regulated. Their homologues in *D. melanogaster* regulate expression of neuropeptides (Vogler & Urban, 2008) and the response to stimuli such as growth factors (Baou *et al.*, 2009). If this function were conserved in schistosomes it would potentially allow them to regulate the development of female schistosomes to reach sexual maturity.

In vitro culture has significant effects on neural genes

Having examined the process of maturation in Chapter 3, Chapter 4 aimed to examine the opposite: regression of female worms in the absence of a male partner. However, when comparing RNA-Seq data of paired and single worm, it became apparent that *in vitro* culture has an extensive effect on gene expression. Changes due to *in vitro* culture, especially of fertility related genes, had been

expected in female worms as reported in the literature (Galanti *et al.*, 2012). However, the changes in gene expression measured here were neither limited to fertility related genes nor to female worms. Using RNA-Seq the whole transcriptome could be examined rather than being limited to small numbers of genes as in previous studies. As a result, a large group of neural genes involved in regulating the development of the nervous system was found to be up-regulated *in vitro*. These genes, such as semaphorins and their plexin and neuropilin receptors, netrins and the DCC and UNC5 receptors, slit and its robo receptors as well as ephrin and its receptors, have all been implicated in the pruning of axons and controlled neural cell death (Vanderhaeghen & Cheng, 2010). On the other hand, neuron survival genes, such as those coding for motor neuron 1 and fragile X mental retardation protein 1, were found to be down-regulated. Together these expression results suggested that parts of the worm nervous system regressed together with the reproductive tissues. This may be of particular interest as the nervous system is thought to be involved in the regulation of female fertility (Collins *et al.*, 2010; Wang *et al.*, 2017).

The role of CD63 receptor and antigen in female fertility

The analysis of RNA-Seq data revealed the up-regulation of two tetraspanins in females from mixed sex infections from 28 d.p.i.: CD63 receptor, previously reported by Fitzpatrick *et al.* (2009), and CD63 antigen. Using RNA-Seq and qRT-PCR data, mature paired females were shown to express the CD63 receptor and antigen at 800- and 50-fold higher levels than male worms as well as at over 190- and 39-fold higher levels than females from single sex infections. Also, both genes were found to be depleted in the ovaries of paired females, compared to

whole worms and the *in situ* experiments showed expression predominantly in the vitellarian tissue. Furthermore, in Chapter 3, both genes clustered with several well-known fertility related genes strongly suggesting a role in the development or maintenance of female fertility. RNAi provided some evidence of a RNA knock down affecting egg production. However, the RNAi experiment was performed *in vitro* and was therefore confounded by the associated problems of reduced fertility in the negative control. Female fertility is particular difficult to study *in vitro*, as regression of female reproductive tissues occurs quickly outside the host although Galanti *et al.* (2012) showed that this process is slowed in the presence of male worms. Although only a moderate reduction of expression was achieved for the CD63 antigen and no significant reduction for the CD63 receptor, both knockdowns were found to result in a significantly reduced production of eggs (CD63 antigen: 9.6-fold; CR63 receptor: 3.2-fold), compared to the negative control.

Given this evidence, a role of both CD63 receptor and antigen in egg production seems near certain. However, their exact function in this process is unclear. In humans the CD63 receptor is involved in forming signalling complexes of integrins and growth factors and thereby regulating downstream gene expression and cell proliferation. If the schistosome homologues performed similar functions, such a role could be crucial to regulate the development of vitellarian tissue in paired female worms. Furthermore, both the CD63 receptor and antigen have been implicated in the regulation of cell survival by down-regulating integrin-β signalling and consequently down-regulating apoptosis in human cell lines. Up-regulation of these two genes in paired females could allow

for proliferation of vitellocytes by reducing the rate of apoptosis in this tissue. The CD63 receptor has also been associated with intracellular vesicles and regulation of protein transport in and out of such vesicles. During their maturation vitellocytes become filled with intracellular vesicles in which precursors of the egg shell proteins are stored. This could mean that CD63 receptors play a role in the formation of such vesicles.

To further explore the role of CD63 receptor and antigen it would be useful to study their interactions with other proteins or maybe even with one another. This might be possible by using a yeast two-hybrid assay which has been used identify other protein interactions in *S. mansoni* (Vanderstraete et al., 2014). Another possible method might be the use of fluorescence resonance energy transfer to observe co-localisation of CD63R and CD63a, however this would require the availability of antibodies with binding specificity to the schistosome genes. A similar approach has been used successfully before to study the interaction of tetraspanins with other proteins (Termini and Gillette, 2017).

Apoptosis in *S. mansoni*

Finally, in Chapter 5, gene expression in the gonads of worms from mixed and single sex infections was examined. This enabled a set of genes to be produced with ovary- and testes-biased expression and the effect of pairing in the worm gonads to be analysed, which had previously been obscured by the somatic tissue. Several apoptosis genes have been identified in schistosomes previously and apoptosis has been shown to play an important role in regulating the growth of vitellarian tissue. In Chapter 2, many genes were examined for their sequence

homology and the presence of appropriate domains in particular, to expand the list of schistosome apoptosis genes. Among the 15 identified genes were homologues of the FAS associated factors 1 and 2, a homologue of the tumour necrosis factor receptor as well as two Bax inhibitors, adding considerable detail to how apoptosis potentially functions in *S. mansoni*. Having divided these genes in to pro- and anti-apoptotic genes based on the reported functions of orthologs, the expression of such genes was observed in several experiments. No consistent pattern of apoptosis related genes was found *in vitro* (see Chapter 4). However, a significant down-regulation of putative pro-apoptotic genes and, vice versa, an up-regulation of putative anti-apoptotic genes was observed in the ovaries of females from mixed sex infections, compared to ovaries from single sex infections (see Figure 5.17). This further strengthened my confidence that these genes are involved in the regulation of apoptosis in *S. mansoni*.

The limitations of bulk RNA-Seq

RNA-Seq is a powerful method to study gene expression, making it possible to measure changes in mRNA levels for all genes at once. And as discussed in the introduction (see Chapter 1.6.3), RNA-Seq offers some significant advantages over other methods used to measure gene expression. However, there are also some limitations to RNA-Seq as used in this thesis. For one, studies by Schwanhaeusser *et al.* (2011) and Vogel *et al.* (2010) found the steady-state mRNA concentrations to only explain between one and two thirds of the variation in observed protein levels. However, work by Li *et al.* (2014) showed that due to problems with the normalisation of proteomics data in the aforementioned studies, mRNA may explain around 84% of protein abundances.

Furthermore, the correlation of mRNA and protein abundance were shown to correlate strongly when specifically looking at changes in gene expression, meaning that the mRNA abundance of differentially expressed genes is in fact very well correlated to their protein products (Koussounadis *et al.* 2015).

Another limitation is the use of RNA-Seq on mRNA extracted from whole worms, rather than individual organs or even single cells, although this was done to some extent in Chapter 5 of this thesis. When examining genes that are only or predominantly expressed in a subset of organs, changes in gene expression may be more difficult to detect. This is because in a sample derived from a whole worm, the mRNA of such organ-specific genes will make up a lower proportion of total reads. Another variation of this problem occurs when comparing samples from different developmental stages. As organs such as the vitellarium develop, they start to make up an increasingly large share of the sequenced mRNA. Due to this, it may appear that certain genes in these female worms have become down-regulated, when in reality the absolute expression has not changed but the gene's mRNA now makes up a relatively smaller proportion of the total mRNA pool. For instance, in Chapter 3.2.8, the reads from just 27 genes, most of which are known to be expressed in vitellarian cells, made up >14% of the transcriptome of mature females. As a result, many other genes appeared to be expressed at lower levels in mature females compared to females from single sex infections. Lastly, when performing RNA-Seq on whole worms, in essence the average gene expression across all tissues is measured, despite there potentially being considerable variation in expression in organs or even adjacent cells. In such samples the average of gene expression in all tissues may show a different

qualitative trend than when all tissues had been measured individually, an effect known as Simpson's paradox (Trapnell *et al.*, 2014). Although RNA-Seq data from different tissues or even single cells would provide an even better and set of data with greater resolution of schistosome transcriptome, RNA-Seq data derived from whole worms still holds a wealth of information. Arguably, for genes expression in a single tissue, such as many genes expressed in the reproductive organs, there must be a very strong change in expression for it to be detected as significantly expressed.

As the above summary of results shows, the work in this thesis succeeded in generating high quality RNA-Seq data that illuminate the changes in gene expression that result in female maturation, following their development from approximately one week before pairing occurs to sexual maturity. The data generated here also covered transcriptome changes *in vitro* as well as gene expression in the gonads of male and female worms. The analysis presented in this thesis also showed numerous examples of differences between male and female schistosomes during their development as well as the changes induced by the male-female interaction.

In the future, the use of single cell RNA sequencing (scRNA-Seq) could address some of the shortcomings of bulk RNA-seq. Most importantly, it would provide a finer resolution of gene expression, not providing transcriptome data of whole worms or individual organs but of single cells. With this technology, it could be possible to detect and quantify the abundance of mRNAs coding for the CD63 receptor and antigen in different subpopulations of vitellocytes. The vitelline

tissue could be dissociated using a collagenases or dispases (Freshney, 1987) and fluorescence activated cell sorting (FACS) could be used to sort cells by their granularity (Basu *et al.*, 2010). This may allow vitelline S1, S2, S3 and S4 cells to be separated as they contain increasing numbers of vesicles (Erasmus, 1975), with S4 cells being most densely packed with eggshell protein containing vesicle (Köster *et al.*, 1988). With this method the localisation of CD63R to particular portion of the vitellaria could potentially be confirmed. Also, very small subpopulations of cells could potentially be examined, such as stem cells, which have previously been observed (Collins *et al.*, 2013), and could play a role in the ability of *S. mansoni* to regenerate reproductive tissues after losing it. Similarly, scRNA-Seq could provide unprecedented resolution of the development of cells in the ovary and testes. This may make it possible to better understand the role of apoptosis in the development of ovarian tissue.

However, not all of the goals of this thesis could be met. In particular, examining the regression of sexually mature females as a result of separation from male worms in Chapter 4 proved challenging. Despite several biological replicates, the number of differentially expressed genes found between treatment groups was low. This most likely resulted from the large magnitude of transcriptome changes brought on by *in vitro* culture, which was shown to have extensive effects on the transcriptome of male and female worms. Due to this, the focus of the chapter changed to examining the effects of *in vitro* culture on gene expression, especially neural and apoptosis-related genes. The transcriptome changes observed in Chapter 4, most likely brought on by a lack of host factors, highlights the need for a better *in vitro* culturing system. Alternatively, worms

could have been re-implanted into mice (Basch & Humbert, 1981) after separation from their partner, to provide a host environment for the rest of the experiment. However, this method would have required more mice to be subjected to invasive procedures, moreover it would expose the worms themselves to considerable stress, which might have interfered with gene expression. Another improvement could have been to perform the RNAi experiment described in Chapter 3 *in vivo*, as described by (Wang *et al.*, 2017). As described above, the *in vitro* conditions caused changes in gene expression that made it difficult to detect the changes in gene expression brought on by the dsRNA treatment compared to the control. Performing RNAi *in vivo* circumvents that problem by making *in vitro* culture during or after exposure to dsRNA unnecessary.

The results in this thesis, including a better understanding of the male-female interaction as well as how apoptosis functions in schistosomes, will hopefully prove valuable for the development of therapeutics in the future. For example, in addition to GST28 (Smp_054160), a long-standing vaccine target, other glutathione S transferases, such as Smp_024010, may also be suitable therapeutic targets. Also, an improved understanding of how schistosomes regulate apoptosis could make it possible to exploit differences between schistosome and mammalian apoptosis to kill worms. The data generated for Chapter 5, in collaboration with Professor Grevelding's group, has also been made available to the research community as a resource for gene expression data in the gonads of adult male and female worms (Lu *et al.*, 2017).

To further investigate apoptosis in the ovaries, but also the vitelline tissue, a number of approached could be taken in the future.

To determine how much apoptosis is taking place in the ovaries of females from MS and SS infections, terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) staining could be used, as Galanti *et al.* (2012) did to detect apoptosis in the vitelline tissue. TUNEL staining is a method that detects the DNA fragmentation that occurs during apoptosis (Kyrylkova *et al.*, 2012). This is done by using the terminal deoxynucleotidyl transferase to conjugate labelled nucleotides, for example with a bromodeoxyuridine (BrdU) label. In the next step, these nucleotides can then be detected with a method appropriate for the chosen label (Kyrylkova *et al.*, 2012). In the case of BrdU, this can be done with antibodies with conjugated alkaline phosphatase, allowing BrdU detection using a method like that used in WISH (Chapter 2.2.7). Another method which could be used to study apoptosis is an assay for caspase 3 activity (Galanti *et al.*, 2012).

In Chapter 5, differences in the expression of gene coding for proteins of the sphingolipid metabolite pathway were observed in the ovaries of females of SS and MS infections. Two important enzymes regulating the abundance of ceramide in cells are ceramide synthase and ceramidase (Levy & Futerman, 2010; Mullen *et al.*, 2012). Inhibitors of these enzymes could be used to investigate if these enzymes also play a role in regulating apoptosis in *S. mansoni*, as they do in mammalian cells (Separovic *et al.*, 2012; Zhu *et al.*, 2008). Fumonisin B, a fungal toxin, has been shown to inhibit ceramide synthase and reduce apoptosis in human cells (Boppana *et al.*, 2014), whereas carmofur is a

potent ceramidase inhibitor (Realini *et al.*, 2013). Using TUNEL staining, it may be possible to examine, if an effect on apoptosis can be observed in females from SS and MS infections when exposed to these inhibitors. Treatment of females from SS infections may prevent apoptosis of vitellarian cells and allow maturation of the organ. On the other hand, treatment of females from MS infections with carmofur, may result in higher levels of apoptosis and lead to a shrinkage of the vitellarian tissue.

There are also fluorescence-based assays available to measure the activity of ceramide synthase (Kim *et al.*, 2012) and ceramidase (Bedia *et al.*, 2010), which could be used to confirm the specificity of the inhibitors to the schistosome homologues of these genes.

By addressing the limitations discussed above, new insights can be gained. Single cell RNA-Seq especially will open up new avenues for characterising gene expression at much greater resolution and improve our understanding of schistosome biology. Secondly improving the available *in vitro* culture methods will allow for better characterisation of genes, which may become targets for future interventions.

Chapter 7

Appendices

Appendix A

Sample name	D18 F1 SS	D18 F2 SS	D18 M1 SS	D18 M2 SS	D18 M3 SS	D21 F1 SS	D21_F2 SS	D21 F3 SS	D21 F4 SS	D21 F5 SS	D21 M1 SS	D21 M2 SS	Average
Total reads	2080114	2062052	2040192	2309320	2148352	2119452	2009938	2120300	2152774	2123504	2048618	2096230	**2073810.5**
Mapped	1540672	1630884	1587154	1682538	1729541	1714133	1491835	1526086	1704036	1691104	1602862	1677186	**1641899.8**
Prop. Paired	1449174	1544283	1504942	1595907	1637981	1625847	1409668	1441913	1618474	1597515	1512372	1586614	**1549813.5**
Unmapped	539442	431168	453038	626782	418811	405319	518103	594214	448738	432400	445756	419044	**431910.79**
% Unmapped	25.9	20.9	22.2	27.1	19.5	19.1	25.8	28	20.8	20.4	21.8	20	**20.8**
% Mapped	74.1	79.1	77.8	72.9	80.5	80.9	74.2	72	79.2	79.6	78.2	80	**79.2**
% Paired	69.7	74.9	73.8	69.1	76.2	76.7	70.1	68	75.2	75.2	73.8	75.7	**74.7**

Sample name	D21 M3 SS	D21 M4 SS	D21 F1 MS	D21 F2 MS	D21 F3 MS	D21 F4 MS	D21 M1 MS	D21 M2 MS	D21 M3 MS	D21 M4 MS	D28 F1 SS	D28 F2 SS	Average
Total reads	2069646	2089104	2049258	2297626	2017852	2086646	2075498	2067554	2075286	2005388	2060388	2113702	**2073810.5**
Mapped	1675116	1662779	1621071	1842584	1589133	1634689	1667968	1666472	1651356	1616338	1653169	1677707	**1641899.8**
Prop. Paired	1581325	1572093	1532912	1736812	1498166	1537935	1579703	1578906	1563079	1517476	1563201	1584245	**1549813.5**
Unmapped	394530	426325	428187	455042	428719	451957	407530	401082	423930	389050	407219	435995	**431910.79**
% Unmapped	19.1	20.4	20.9	19.8	21.2	21.7	19.6	19.4	20.4	19.4	19.8	20.6	**20.8**
% Mapped	80.9	79.6	79.1	80.2	78.8	78.3	80.4	80.6	79.6	80.6	80.2	79.4	**79.2**
% Paired	76.4	75.3	74.8	75.6	74.2	73.7	76.1	76.4	75.3	75.7	75.9	75	**74.7**

Sample name	D28 F3 SS	D28 M1 SS	D28 M2 S	D28 M3 SS	D28 F1 MS	D28 F2 MS	D28 F3 MS	D28 F4 MS	D28 M1 MS	D28 M2 MS	D28 M3 MS	D28 M4 MS	Average
Total reads	2053026	2058386	2052242	2003982	2048124	2043236	2096656	2069750	2135820	2070970	2043092	2064518	**2073810.5**
Mapped	1522288	1639142	1621913	1524462	1680771	1617647	1677258	1653050	1664760	1686818	1648057	1635606	**1641899.8**
Prop. Paired	1428607	1548632	1529460	1433195	1599259	1525497	1585061	1554242	1565877	1595903	1557610	1541694	**1549813.5**
Unmapped	530738	419244	430329	479520	367353	425589	419398	416700	471060	384152	395035	428912	**431910.79**
% Unmapped	25.9	20.4	21	23.9	17.9	20.8	20	20.1	22.1	18.5	19.3	20.8	**20.8**
% Mapped	74.1	79.6	79	76.1	82.1	79.2	80	79.9	77.9	81.5	80.7	79.2	**79.2**
% Paired	69.6	75.2	74.5	71.5	78.1	74.7	75.6	75.1	73.3	77.1	76.2	74.7	**74.7**

Sample name	D35 F1 SS	D35 F2 SS	D35 F3 SS	D35 M1 SS	D35 M2 SS	D35 M3 SS	D35 M4 SS	D35 F1 MS	D35 F2 MS	D35 M3 MS	D35 M1 MS	D35 M2 MS	Average
Total reads	2060052	2096878	2039540	2008772	2017950	2050816	2018606	2088038	2080764	2057080	2066508	2010962	**2073810.5**
Mapped	1545483	1704745	1629324	1600736	1635793	1632836	1610231	1710015	1668174	1669591	1641929	1622566	**1641899.8**
Prop. Paired	1441668	1613167	1539189	1506134	1556801	1538332	1513943	1620797	1574130	1581931	1544947	1532003	**1549813.5**
Unmapped	514569	392133	410216	408036	382157	417980	408375	378023	412590	387489	424579	388396	**431910.79**
% Unmapped	25	18.7	20.1	20.3	18.9	20.4	20.2	18.1	19.8	18.8	20.5	19.3	**20.8**
% Mapped	75	81.3	79.9	79.7	81.1	79.6	79.8	81.9	80.2	81.2	79.5	80.7	**79.2**
% Paired	70	76.9	75.5	75	77.1	75	75	77.6	75.7	76.9	74.8	76.2	**74.7**

Sample name	D35 M3 MS	D38 F1 SS	D38 F2 SS	D38 F3 SS	D38 F4 SS	D38 M1 SS	D38 M2 SS	D38 M3 SS	D38 M4 SS	D38 M5 SS	D38 M6 SS	D38 F1 MS	Average
Total reads	2002350	2084778	2089938	2086154	2099738	2094136	2085988	2070088	2066106	2094480	2093118	2058490	**2073810.5**
Mapped	1591148	1656569	1663267	1627644	1686112	1638494	1647282	1590347	1638961	1655896	1684572	1660113	**1641899.8**
Prop. Paired	1495252	1563930	1570778	1528519	1593739	1543889	1549069	1496552	1544909	1560467	1589377	1574793	**1549813.5**
Unmapped	411202	428209	426671	458510	413626	455642	438706	479741	427145	438584	408546	398377	**431910.79**
% Unmapped	20.5	20.5	20.4	22	19.7	21.8	21	23.2	20.7	20.9	19.5	19.4	**20.8**
% Mapped	79.5	79.5	79.6	78	80.3	78.2	79	76.8	79.3	79.1	80.5	80.6	**79.2**
% Paired	74.7	75	75.2	73.3	75.9	73.7	74.3	72.3	74.8	74.5	75.9	76.5	**74.7**

Sample name	D38 F2 MS	D38 F3 MS	D38 M1 MS	D38 M2 MS	D38 M3 MS	D49 F1 SS	D49 F2 SS	D49 M1 SS	D49 M2 SS	D49 M3 SS	Average
Total reads	2051724	2127108	2064036	2035976	2046108	2132220	2023748	2012990	2030664	2062238	**2073810.5**
Mapped	1620590	1725363	1631237	1632715	1633668	1681676	1589172	1606279	1632246	1660024	**1641899.8**
Prop. Paired	1525409	1634703	1531509	1543353	1539579	1584697	1493600	1514002	1541721	1568474	**1549813.5**
Unmapped	431134	401745	432799	403261	412440	450544	434576	406711	398418	402214	**431910.79**
% Unmapped	21	18.9	21	19.8	20.2	21.1	21.5	20.2	19.6	19.5	**20.8**
% Mapped	79	81.1	79	80.2	79.8	78.9	78.5	79.8	80.4	80.5	**79.2**
% Paired	74.3	76.9	74.2	75.8	75.2	74.3	73.8	75.2	75.9	76.1	**74.7**

Appendix A.1: Summary of sequencing statistics for sequnced RNA-Seq samples. Samples names provide the day post infections (e.g. D18 stands for 18 d.p.i.), the gender and replicate number (e.g. F1 is female replicate 1) and the pairing status (MS for mixed sex infections, SS for single sex infection). Total reads is the number of all reads sequenced. Mapped is the total number of reads mapped to the *S. mansoni* reference genome v5.2. 'Prop. Paired' refers to the total number of reads mapped together with a mate-paired read in close proximity and in the correct orientation. Unmapped is the total number of reads that could not be mapped to the reference by Tophat2.

Gene ID	Product	Fold Change	Adjusted p-value
Smp_070540	Uncharacterised protein	5.58	1.62E-06
Smp_213480	Muts protein	6.54	9.36E-06
Smp_124750	Uncharacterised protein	8.53	3.93E-05
Smp_135290	Macrophage expressed gene 1 protein	3.30	3.93E-05
Smp_191690	Tumour susceptibility gene 101 protein	8.96	3.93E-05
Smp_082420	Uridine phosphorylase	2.98	3.63E-04
Smp_147320	Camp dependent protein kinase regulatory chain	5.11	5.22E-04
Smp_174530	Aminopeptidase PILS (M01 family)	5.74	5.22E-04
Smp_007760	Alanine aminotransferase 2	2.20	6.61E-04
Smp_172340	Uncharacterised protein	3.48	1.09E-03
Smp_151840	Uncharacterised protein	5.33	1.13E-03
Smp_032980	Calmodulin protein	6.17	1.49E-03
Smp_015100	Uncharacterised protein	2.32	1.60E-03
Smp_178780	Single minded	4.78	1.60E-03
Smp_042150	Tegument-allergen-like protein	3.52	1.79E-03
Smp_162840	Uncharacterised protein	3.37	1.79E-03
Smp_163170	Sodium:potassium dependent ATPase beta subunit	3.61	2.31E-03
Smp_153070	Uncharacterised protein	3.43	2.32E-03
Smp_118960	Jun protein	6.60	2.60E-03
Smp_158150	Reticulocalbin 2	4.95	5.31E-03
Smp_141810	Uncharacterised protein	2.76	7.79E-03
Smp_141250	Protein jagged 1	2.59	8.74E-03

Appendix A.2: List of the most differentially expressed genes upregulated in females compared to male worms from mixed sex infections at 21 days *post* infection. Fold changes measure expression in female relative to male worms. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Fold Change	Adjusted p-value
Smp_211240	Beta 1,4 N acetylgalactosaminyltransferase	13.37	1.64E-07
Smp_160360	Sodium:chloride dependent neurotransmitter	3.54	3.36E-07
Smp_074560	Uncharacterised protein	3.59	9.85E-07
Smp_129960	Uncharacterised protein	5.07	3.52E-05
Smp_195090	Tegument-allergen-like protein	6.09	7.41E-05
Smp_084270	Rhodopsin orphan GPCR	2.92	2.28E-04
Smp_153930	Invadolysin (M08 family)	4.84	3.35E-04
Smp_168730	Carbonic anhydrase	2.35	3.63E-04
Smp_105780	Uro adherence factor A	6.74	8.16E-04
Smp_120020	Uncharacterised protein	6.44	1.13E-03
Smp_212180	Glucose dehydrogenase (acceptor)	5.03	1.19E-03
Smp_056280	Zinc finger CCCH domain containing protein 31	4.96	1.39E-03
Smp_091210	Uncharacterised protein	6.12	1.49E-03
Smp_151660	Uncharacterised protein	3.32	2.60E-03
Smp_161190	Uncharacterised protein	2.44	4.58E-03
Smp_071810	Uncharacterised protein	2.11	7.11E-03

Appendix A.3: List of the most differentially expressed genes upregulated in males compared to female worms from mixed sex infections at 21 days *post* infection. Fold changes measure expression in male relative to female worms. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Fold Change	Adjusted p-value
Smp_147670	Receptor kinase I-interacting protein SIP	6.49	6.85E-36
Smp_159510	Uncharacterised protein	6.98	1.14E-35
Smp_017610	amiloride sensitive amine oxidase	66.88	5.87E-33
Smp_075280	Uncharacterised protein	3.78	1.77E-31
Smp_175090	long-chain-fatty-acid--CoA ligase	5.63	3.41E-31
Smp_170020	tachykinin receptor protein like	9.10	2.54E-30
Smp_125940	Uncharacterised protein	10.53	9.28E-28
Smp_168730	carbonic anhydrase	9.29	4.26E-27
Smp_169570	glycerol 3 phosphate dehydrogenase	14.15	1.24E-26
Smp_161030	Uncharacterised protein	14.02	2.02E-26
Smp_093210	FMRFamide activated amiloride sensitive sodium	4.92	2.45E-26
Smp_059570	Uncharacterised protein	7.04	3.15E-26
Smp_139380	Spindle assembly checkpoint component MAD1	55.18	5.20E-26
Smp_034550	Alpha actinin	6.54	1.93E-25
Smp_167160	Dynamin binding protein	2.91	4.29E-25
Smp_089670	Alpha 2 macroglobulin	4.54	6.92E-25
Smp_197860	Gelsolin	11.28	4.05E-24
Smp_134710	Uncharacterized MFS type transporter C19orf28	16.55	4.12E-24

Smp_104890	Glutamate-gated chloride channel subunit 3	6.03	7.84E-24
Smp_198930	Uncharacterised protein	6.54	1.13E-23
Smp_028670	Carbonic anhydrase II	4.23	1.14E-23
Smp_146200	Zyxin:trip6	2.88	1.61E-23
Smp_153780	Glutamate receptor, ionotropic kainate 3	6.68	1.89E-23
Smp_168980	Adenylate cyclase	4.06	2.09E-23
Smp_134130	Ring canal kelch	5.60	3.63E-23
Smp_068280	Prolyl 4 hydroxylase subunit alpha 2	3.99	4.79E-23
Smp_194050	Clumping factor A	5.09	6.12E-23
Smp_160360	Sodium:chloride dependent neurotransmitter	9.63	9.19E-23
Smp_006860	PDZ and LIM domain protein Zasp	4.66	9.42E-23
Smp_059170	Troponin i	7.43	1.07E-22
Smp_155410	Coiled coil domain containing protein 102A	4.80	1.22E-22
Smp_074560	Uncharacterised protein	10.69	1.26E-22
Smp_136930	Uncharacterised protein	3.58	1.42E-22
Smp_053560	MAP kinase activated protein kinase 2	3.15	2.44E-22
Smp_016230	Protocadherin 17	12.59	4.94E-22
Smp_164240	Junctophilin 2	5.43	6.27E-22
Smp_000170	KV channel interacting protein 2 (KChIP2)	15.56	6.44E-22
Smp_022340	PDZ and LIM domain protein Zasp	5.83	6.54E-22
Smp_126360	Uncharacterised protein	10.50	8.02E-22
Smp_166920	PDZ and LIM domain protein Zasp	7.50	8.16E-22
Smp_214190	Calpain	3.03	2.08E-21
Smp_121190	Voltage gated potassium channel	7.48	2.44E-21
Smp_133550	Neuropeptide receptor type; neuropeptide ff receptor	7.09	2.79E-21
Smp_072450	Rhodopsin orphan GPCR	10.11	2.95E-21
Smp_126320	D glucuronyl C5 epimerase	2.41	3.34E-21
Smp_174870	Uncharacterised protein	5.16	4.25E-21
Smp_046640	Twik family of potassium channels	6.67	4.77E-21
Smp_164960	Phosphatidylinositol 4,5 bisphosphate 3 kinase	1.84	5.65E-21
Smp_151100	cGMP dependent protein kinase	5.43	7.80E-21
Smp_163090	Voltage gated potassium channel	5.45	9.67E-21

Appendix A.4: List of the most differentially expressed genes upregulated in male worms compared to female worms from mixed sex infections at 38 days *post* infection. Fold changes measure expression in male relative to female worms. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Fold Change	Adjusted p-value
Smp_212820	Zinc finger protein	20.55	7.07E-67
Smp_038030	Adenylosuccinate lyase	6.35	3.02E-63
Smp_179710	Maternal Effect Lethal family member	32.86	4.82E-62
Smp_159920	Protocadherin 11	41.35	4.03E-49
Smp_090520	Purine nucleoside phosphorylase	18.90	2.20E-47
Smp_126210	Uncharacterised protein	20.18	7.25E-41
Smp_160160	Solute carrier family 17	13.56	7.68E-41
Smp_212090	Calcium dependent activator protein	13.74	3.75E-40
Smp_042700	Protein FAM81B	13.50	1.96E-39
Smp_002080	S adenosylmethionine synthetase	3.03	1.28E-38
Smp_210370	Ormdl protein	2.95	1.67E-38
Smp_148770	Ribose 5 phosphate isomerase	6.94	6.99E-36
Smp_173150	CD63 antigen	34.92	1.03E-35
Smp_041540	Hormone receptor 4	13.98	1.18E-35
Smp_080730	Cyclin dependent kinase 1	6.84	1.18E-35
Smp_056700	Membrane associated RING finger protein	36.62	1.14E-34
Smp_048430	Thioredoxin glutathione reductase	3.05	1.59E-33
Smp_092390	N acetylglucosamine kinase	5.33	5.87E-33
Smp_026400	Thyrotroph embryonic factor	23.27	7.70E-33
Smp_077900	Female specific protein 800	82.82	1.00E-32
Smp_155740	TNF receptor associated protein 1	2.22	1.79E-32
Smp_075870	Translocation associated membrane protein 1	3.65	2.47E-32
Smp_000290	Female-specific protein 800	84.02	6.32E-32
Smp_021070	Guanine nucleotide binding protein	3.79	7.89E-32
Smp_035470	Dolichyl diphosphooligosaccharide protein	3.36	7.89E-32
Smp_055800	Pescadillo	3.57	7.89E-32
Smp_146480	Uncharacterised protein	60.43	7.89E-32
Smp_198170	Protein FAM81B	13.66	1.05E-31
Smp_105680	Dolichyl diphosphooligosaccharide protein	2.97	3.41E-31
Smp_171150	Protein arginine N methyltransferase 5	3.75	5.04E-31
Smp_209060	Alpha 1,3 fucosyltransferase B	10.74	2.70E-30
Smp_148390	Uncharacterised protein	193.90	3.22E-30
Smp_142760	Centrosomal protein of 95 kDa	5.04	5.55E-30
Smp_011870	Uncharacterised protein	139.19	5.80E-30
Smp_075330	Uncharacterised protein	13.02	6.68E-30
Smp_135840	Uncharacterised protein	5.07	7.55E-30
Smp_082490	G2:mitotic specific cyclin B2	7.21	1.25E-29
Smp_062070	Sulfide quinone reductase	3.58	1.56E-29
Smp_175560	Beta site APP cleaving enzyme 1	60.07	2.12E-29
Smp_169970	Kinetochore protein ndc80	5.18	4.08E-29
Smp_076320	Uncharacterised protein	70.01	4.83E-29
Smp_083770	Anoctamin 7	14.31	6.92E-29
Smp_054010	Y+L amino acid transporter 2	35.13	2.94E-28
Smp_199540	Inositol type trisphosphate receptor	40.29	3.73E-28

Smp_058240	Reticulon 4 (Neurite outgrowth inhibitor)	3.74	4.33E-28
Smp_141160	E3 ubiquitin protein ligase RNF168	11.23	5.97E-28
Smp_048660	Nucleolar protein 56	3.27	1.02E-27
Smp_013970	Uncharacterised protein	6.43	1.06E-27
Smp_199830	Leucine rich repeat protein	7.36	1.25E-27
Smp_031730	Signal peptidase 18 kDa subunit	3.43	1.34E-27

Appendix A.5: List of the most differentially expressed genes upregulated in females compared to male worms from mixed sex infections at 38 days *post* infection. Fold changes measure expression in female relative to male worms. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Average read count	Fold change	Adjusted p-value
Smp_015630	Glutamate-gated chloride channel subunit	353.74	1.07	0.929
Smp_104890	Glutamate-gated chloride channel subunit	244.18	0.41	5.58E-08
Smp_099500	Glutamate-gated chloride channel subunit	38.57	2.17	3.45E-02
Smp_031680	Nicotinic acetylcholine receptor	176.02	0.95	0.948
Smp_012000	Nicotinic acetylcholine receptor	92.58	1.18	0.779
Smp_132070	Nicotinic acetylcholine receptor subunit	3562.48	1.02	0.937
Smp_180570	Nicotinic acetylcholine receptor	132.26	1.08	0.900
Smp_139330	Nicotinic acetylcholine receptor non α	811.23	0.72	0.106
Smp_157790	Nicotinic acetylcholine receptor subunit	121.98	2.32	4.21E-03
Smp_176310	Nicotinic acetylcholine receptor subunit	699.76	1.23	0.637
Smp_101990	Acetylcholine receptor subunit α	23.03	0.27	9.56E-05
Smp_037960	Neuronal acetylcholine receptor subunit α	1308.18	1.33	0.387
Smp_096480	Glycine receptor subunit β	520.33	1.03	0.951
Smp_142690	Neuronal acetylcholine receptor subunit α	1875.19	1.33	5.15E-03
Smp_175110	Neurotransmitter gated ion channel	0	1.00	N/A
Smp_197600	Nicotinic acetylcholine receptor subunit	0.68	0.78	0.819

Appendix A.6: Expression of genes encoding neurotransmitters in females compared to males from single sex infections at 38 days *post* infection. Fold changes measure expression in male relative to female worms; that mean genes with a fold change of > 1 are up regulated in males. The *p*-value has been adjusted for multiple hypothesis testing.

Pathway	Description	Total Genes	DEGs	Expected	p-value
smm03010	Ribosome	110	88	22.49	1.66E-43
smm00190	Oxidative phosphorylation	69	38	14.11	9.48E-11
smm03030	DNA replication	30	14	6.13	7.75E-04
smm03020	RNA polymerase	24	12	4.91	8.73E-04
smm00240	Pyrimidine metabolism	59	21	12.06	2.64E-03
smm03040	Spliceosome	105	32	21.47	3.93E-03
smm03050	Proteasome	32	12	6.54	1.20E-02
smm03060	Protein export	19	8	3.88	1.83E-02
smm00982	Drug metabolism - cytochrome P450	4	3	0.82	2.71E-02
smm03008	Ribosome biogenesis in eukaryotes	62	18	12.68	3.00E-02
smm04122	Sulfur relay system	7	4	1.43	3.06E-02

Appendix A.7: KEGG pathways which are significantly enriched amongst the genes up regulated in males from single sex infections at 18 days *post* infection compared to males at 21 days *post* infection. "Total genes" is the number of genes in the given pathway. "DEGs" provides the number of differentially expressed genes in that pathway. "Expected" provides the number of genes expected by chance to be differentially expressed in each pathway.

Comparison	Fold change	Adjusted p-value
Expression in male worms at 18 d.p.i. compared to males at 21 d.p.i.	5.10	6.45E-06
Expression in male worms at 18 d.p.i. compared to males at 38 d.p.i.	21.95	1.34E-14
Expression in female worms at 18 d.p.i. compared to females at 38 d.p.i.	18.16	1.30E-11

Appendix A.8: Expression of the gene encoding the microsomal glutathion S transferase (Smp_024010) in different gene expression comparisons of male and females worms at different time points. The fold change provides the gene expression in the first condition relative to the second condition in each comparison. Smp_024010 was found up-regulated in male or female worms at 18 days *post* infection (d.p.i.) compared to later time points in all the comparisons. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Fold Change	Adjusted p-value
Smp_139380	Spindle assembly checkpoint component MAD1	14.15	2.35E-14
Smp_136100	Kelch protein 10	8.63	3.13E-09
Smp_155570	Endoglycoceramidase	2.78	3.18E-08
Smp_074570	Uncharacterised protein	4.85	3.43E-07
Smp_042160	Fructose-bisphosphate aldolase	2.10	1.55E-05
Smp_105410	Glucose transport protein	3.14	1.55E-05
Smp_169090	Solute carrier family 43	3.59	1.55E-05
Smp_201700	Uncharacterised protein	2.68	3.77E-05
Smp_017610	Amiloride sensitive amine oxidase	4.95	5.81E-05
Smp_133970	Uncharacterised protein	5.26	7.10E-05
Smp_142970	Palmitoyl protein thioesterase 1	2.01	7.10E-05
Smp_134710	Uncharacterized MFS type transporter C19orf28	3.20	7.58E-05
Smp_056460	Uncharacterised protein	6.61	1.04E-04
Smp_154800	DC STAMP domain containing protein 2	6.54	1.23E-04
Smp_128510	Uncharacterised protein	7.51	1.56E-04
Smp_144440	Replication A protein	6.92	1.99E-04
Smp_123300	Follistatin	2.35	2.67E-04
Smp_046790	Solute carrier family 2 facilitated glucose	3.34	2.72E-04
Smp_127880	Uncharacterised protein	3.43	2.72E-04
Smp_143590	Uncharacterised protein	2.30	2.79E-04
Smp_005470	Dynein light chain	4.89	2.80E-04
Smp_084270	Rhodopsin orphan GPCR	2.80	2.80E-04
Smp_169570	Glycerol 3 phosphate dehydrogenase	2.86	2.80E-04
Smp_104450	Uncharacterised protein	4.83	3.09E-04
Smp_084440	Uncharacterised protein	2.32	3.23E-04
Smp_045490	Annexin	2.96	3.31E-04
Smp_017620	Uncharacterised protein	5.13	4.21E-04
Smp_013040	Cathepsin D	2.06	4.36E-04
Smp_194950	ELAV protein 2	3.57	5.00E-04
Smp_212180	Glucose dehydrogenase	5.23	5.43E-04
Smp_091460	Glutamine synthetase	2.45	5.61E-04
Smp_128470	Cation transporting ATPase worm	2.37	5.61E-04
Smp_034550	Alpha actinin	2.09	6.04E-04
Smp_152200	Cdc25	6.03	6.11E-04
Smp_161240	Uncharacterised protein	3.94	6.25E-04
Smp_042910	Uncharacterised protein	2.08	6.41E-04
Smp_091210	Uncharacterised protein	5.55	6.78E-04
Smp_153930	Invadolysin (M08 family)	4.32	7.48E-04
Smp_163360	Dystrophin	2.35	7.55E-04
Smp_048030	Uncharacterised protein	3.05	8.06E-04
Smp_007760	Alanine aminotransferase 2	2.19	8.84E-04
Smp_148030	DC STAMP domain containing protein 1	4.24	8.84E-04

Smp_161220	Adipocyte plasma membrane associated protein	2.42	9.68E-04
Smp_172960	Serine type protease inhibitor	3.84	9.88E-04
Smp_129960	Uncharacterised protein	3.67	1.00E-03
Smp_145900	Sphingolipid delta(4) desaturase:C4 hydroxylase	2.96	1.03E-03
Smp_071610	Lysosomal Pro X carboxypeptidase	2.13	1.48E-03
Smp_151920	Paladin	3.80	1.51E-03
Smp_191910	Stress protein DDR48	5.82	1.52E-03
Smp_033250	Uncharacterised protein	5.98	1.57E-03

Appendix A.9: Genes up-regulated in males from from mixed sex infections at 28 days *post* infection (d.p.i.) compared to males at 21 d.p.i. Fold changes provide gene expression in males at 28 d.p.i. compared to males at 21 d.p.i. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Fold Change	Adjusted p-value
Smp_185130	Uncharacterised protein	5.46	2.19E-03
Smp_199670	Cation transporting ATPase	5.09	9.85E-04
Smp_141740	Protocadherin 11	4.49	2.51E-04
Smp_165240	Uncharacterised protein	4.43	8.32E-03
Smp_022740	Tumour necrosis factor alpha induced protein	4.41	5.41E-03
Smp_164080	Bromodomain testis specific protein	4.09	4.78E-04
Smp_135630	Protocadherin gamma a10	4.07	1.97E-03
Smp_137820	Uncharacterised protein	3.80	2.89E-03
Smp_012000	Nicotinic acetylcholine receptor	3.55	6.18E-04
Smp_198650	Ankyrin :unc	3.46	6.05E-03
Smp_187740	Forkhead box protein j1 b	3.38	2.06E-03
Smp_170270	Uncharacterised protein	3.28	9.72E-03
Smp_012950	Cytoplasmic polyadenylation element binding	3.23	2.85E-03
Smp_160590	Uncharacterised protein	3.20	3.74E-07
Smp_163790	Nuclear receptor 2DBD gamma	3.20	2.12E-03
Smp_124600	Uncharacterised protein	3.15	2.10E-03
Smp_193630	Endoglycoceramidase	3.10	3.73E-03
Smp_177020	Phospholipase C Like family member (pll 1)	3.09	8.88E-03
Smp_183130	Liprin alpha 2	2.97	2.10E-03
Smp_161640	Transient receptor potential cation channel	2.84	1.43E-03
Smp_203940	Short chain dehydrogenase	2.81	2.06E-03
Smp_126290	Uncharacterised protein	2.70	5.06E-03
Smp_142940	Paired box protein pax	2.54	7.61E-03
Smp_123300	Follistatin	2.39	8.16E-04
Smp_168650	Microtubule actin cross linking factor 1	2.26	2.71E-03

Smp_170620	Rhodopsin orphan GPCR	2.26	4.44E-03
Smp_174340	Frizzled 9	2.22	5.41E-03
Smp_158720	Protein kinase	2.21	4.77E-03
Smp_173450	Chromosome associated kinesin KIF4A	2.19	8.60E-03

Appendix A.10: Genes up-regulated in males from from mixed sex infections at 28 days *post* infection (d.p.i.) compared to males at 35 d.p.i. Fold changes provide gene expression in males at 28 d.p.i. compared to males at 35 d.p.i. The *p*-value has been adjusted for multiple hypothesis testing.

Comparison	Fold change	Adjusted p-value
Males from single sex infections at 38 d.p.i. compared to Males from single sex infections at 18 d.p.i.	7.54	2.26E-06
Males from mixed sex infections at 35 d.p.i. compared to males from single sex infections at 35 d.p.i.	7.45	1.23E-03
Males from mixed sex infections at 38 d.p.i. compared to males from single sex infections at 38 d.p.i.	5.33	1.99E-03

Appendix A.11: Expression of the gene encoding the aromatic-L-amino acid decarboxylase (Smp_135230) in different gene expression comparisons of male worms at different time points. The fold change provides the gene expression in the first condition relative to the second condition in each comparison. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Fold Change	Adjusted p-value
Smp_191910	Stress protein DDR48	165.06	1.85E-33
Smp_144440	Replication A protein	118.44	2.30E-33
Smp_033250	Uncharacterised protein	135.80	3.73E-30
Smp_014610	Serine:threonine kinase 1	141.61	8.74E-30
Smp_129960	Uncharacterised protein	27.58	2.90E-27
Smp_000420	Pro His rich protein	116.10	3.56E-27
Smp_131110	Uncharacterised protein	116.50	5.99E-27
Smp_077900	Female specific protein 800	41.86	2.52E-26
Smp_165360	Histone acetyltransferase myst4	118.07	2.52E-26
Smp_151150	Arabinogalactan protein	57.98	1.06E-25
Smp_169090	Solute carrier family 43	10.13	6.27E-25
Smp_076320	Uncharacterised protein	39.53	1.16E-24
Smp_000290	Female-specific protein 800	36.61	1.32E-23
Smp_130970	G2:mitotic specific cyclin B3	40.30	2.40E-23
Smp_000410	Trematode Eggshell Synthesis domain containing protein	64.17	2.91E-23
Smp_050750	Uncharacterised protein	52.20	4.24E-23

Smp_138570	Spore germination protein	84.73	2.53E-22
Smp_077890	Trematode Eggshell Synthesis domain containing protein	71.24	8.35E-22
Smp_151160	Stress protein DDR48	54.68	9.80E-22
Smp_192220	Uncharacterised protein	59.23	1.23E-21
Smp_000430	Eggshell protein	75.11	5.84E-21
Smp_050270	Tyrosinase	75.26	7.31E-21
Smp_175560	Beta site APP cleaving enzyme 1	21.44	2.76E-19
Smp_137460	Cytoplasmic polyadenylation element binding	46.34	7.38E-19
Smp_202830	Uncharacterised protein	52.80	1.01E-18
Smp_095980	Extracellular superoxide dismutase (Cu Zn)	59.48	1.72E-18
Smp_000400	Uncharacterised protein	58.60	1.83E-18
Smp_143970	Kelch protein 10	38.47	1.87E-18
Smp_000270	Uncharacterised protein	56.03	2.51E-18
Smp_013540	Tyrosinase	53.82	3.17E-18
Smp_155310	Tetraspanin CD63 receptor	56.65	1.23E-17
Smp_000280	Uncharacterised protein	54.16	1.55E-17
Smp_056700	Membrane associated RING finger protein	9.84	2.75E-17
Smp_095350	Acetyl coenzyme A carboxylase; methylcrotonyl CoA	44.73	1.00E-16
Smp_125740	Uncharacterised protein	3.02	1.61E-16
Smp_195090	Tegument-allergen-like protein	19.98	1.61E-16
Smp_074710	Def8 protein	11.41	6.55E-16
Smp_007670	Uncharacterised protein	5.83	1.36E-15
Smp_058160	Uncharacterised protein	24.34	1.43E-15
Smp_112450	Trematode Eggshell Synthesis domain containing protein	41.41	2.01E-15
Smp_145490	Poly(rC) binding protein 2:3:4	39.76	3.52E-15
Smp_077920	Uncharacterised protein	40.27	9.28E-15
Smp_159920	Protocadherin 11	6.03	1.57E-14
Smp_211240	Beta 1,4 N acetylgalactosaminyltransferase	23.57	3.77E-14
Smp_196950	One cut domain family member	14.57	4.00E-14
Smp_136100	Kelch protein 10	17.81	6.25E-14
Smp_130780	Monocarboxylate transporter	30.67	9.01E-14
Smp_065530	Acrosin	29.09	9.98E-14
Smp_132800	STARP antigen	18.57	1.89E-13
Smp_199540	Type inositol trisphosphate receptor; inositol trisphosphate receptor type	10.61	2.28E-13
Smp_173150	CD63 antigen	6.33	5.24E-12

Appendix A.12: The most differentially expressed genes in female worms from mixed sex infections at 28 days *post* infection (d.p.i.) compared to females at 21 d.p.i. The fold change provides a measure of expression in females at 28 d.p.i. relative to females at 21 d.p.i. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Fold change	Adjusted p-value
Smp_131110	Uncharacterised protein	616.92	5.50E-48
Smp_077900	Female specific protein 800	504.95	4.69E-65
Smp_000290	Female-specific protein 800	463.72	5.79E-61
Smp_191910	Stress protein DDR48	431.31	3.31E-46
Smp_014610	Serine:threonine kinase 1	427.38	1.30E-43
Smp_000430	Eggshell protein	374.21	5.91E-39
Smp_151150	Arabinogalactan protein	362.41	8.12E-50
Smp_013540	Tyrosinase	353.16	1.35E-38
Smp_077890	Trematode Eggshell Synthesis domain containing protein	351.06	4.31E-40
Smp_192220	Uncharacterised protein	350.92	1.63E-40
Smp_095980	Extracellular superoxide dismutase (Cu Zn)	323.47	9.38E-37
Smp_202830	Uncharacterised protein	293.50	2.46E-37
Smp_050750	Uncharacterised protein	265.17	1.30E-43
Smp_000280	Uncharacterised protein	262.95	5.31E-34
Smp_138570	Spore germination protein	261.01	1.15E-34
Smp_000400	Uncharacterised protein	252.51	7.23E-34
Smp_000270	Uncharacterised protein	243.96	8.24E-34
Smp_050270	Tyrosinase	243.54	1.61E-33
Smp_151160	Stress protein DDR48	230.02	1.82E-38
Smp_000420	Pro His rich protein	214.78	3.81E-34
Smp_155310	Tetraspanin CD63 receptor	190.75	5.98E-30
Smp_165360	Histone acetyltransferase myst4	175.53	1.24E-30
Smp_144440	Replication A protein	169.70	2.82E-37
Smp_000370	Uncharacterised protein	152.97	5.61E-30
Smp_000410	Trematode Eggshell Synthesis domain containing protein	146.36	2.85E-32
Smp_175560	Beta site APP cleaving enzyme 1	138.39	4.93E-45
Smp_058160	Uncharacterised protein	112.56	8.89E-33
Smp_143970	Kelch protein 10	109.35	2.49E-30
Smp_129960	Uncharacterised protein	101.63	2.15E-48
Smp_199540	Inositol type trisphosphate receptor	50.27	1.72E-34
Smp_074710	Def8 protein	38.78	5.84E-33
Smp_173150	CD63 antigen	38.76	1.62E-42
Smp_054010	Y+L amino acid transporter 2	37.52	1.66E-32
Smp_159920	Protocadherin 11	30.49	5.42E-47
Smp_169090	Solute carrier family 43	30.16	5.37E-49
Smp_212820	Zinc finger protein	19.90	1.63E-74
Smp_075330	Uncharacterised protein	19.25	4.86E-44
Smp_179710	Maternal Effect Lethal family member (mel 32)	14.46	5.02E-42
Smp_129610	Malate dehydrogenase	10.57	8.24E-34
Smp_209060	Alpha 1,3 fucosyltransferase B	9.25	1.90E-30
Smp_199830	Leucine rich repeat protein	7.03	5.61E-30
Smp_148770	Ribose 5 phosphate isomerase	6.93	4.35E-41
Smp_038030	Adenylosuccinate lyase	5.45	5.79E-61

Smp_135840	Uncharacterised protein	5.22	3.63E-35
Smp_062070	Sulfide quinone reductase	4.44	9.54E-46
Smp_058240	Reticulon 4 (Neurite outgrowth inhibitor)	4.00	3.24E-35
Smp_011960	Inositol pentakisphosphate 2 kinase	3.78	1.46E-33
Smp_075870	Translocation associated membrane protein 1	3.49	1.52E-34
Smp_031730	Signal peptidase 18 kDa subunit	3.36	2.02E-30
Smp_035470	Dolichyl diphosphooligosaccharide protein	3.33	7.22E-36
Smp_056760	Protein disulfide isomerase	3.26	1.46E-33
Smp_075470	Cysteine desulfurase, mitochondrial	3.21	2.74E-41
Smp_048660	Nucleolar protein 56	3.17	5.78E-30
Smp_048430	Thioredoxin glutathione reductase	2.70	1.08E-30
Smp_210370	Ormdl protein	2.63	1.48E-35

Appendix A.13: The most differentially expressed genes in female worms from mixed sex infections at 38 days *post* infection (d.p.i.) compared to females from single sex infections. The fold change provides a measure of expression in females from mixed sex infections relative to females from single sex infections. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Fold Change	Adjusted p-value
Smp_185340	Uncharacterised protein	8.98	7.08E-06
Smp_121070	Uncharacterised protein	8.82	2.08E-05
Smp_172340	Uncharacterised protein	6.87	7.56E-07
Smp_141840	Uncharacterised protein	6.74	8.37E-07
Smp_144860	Protein boule	6.61	6.34E-05
Smp_012660	Sushi:SCR:CCP,domain containing protein	6.15	6.52E-06
Smp_191250	Reverse transcriptase	6.15	6.09E-05
Smp_124870	Uncharacterised protein	5.99	5.42E-05
Smp_145580	Progesterone induced blocking factor	5.55	3.33E-06
Smp_164590	Fibrillar collagen chain FAp1 alpha	5.54	3.96E-11
Smp_196840	Collagen alpha 1(II) chain	5.21	1.06E-19
Smp_123830	Collagen alpha (xi) chain	4.90	3.96E-11
Smp_197370	Uncharacterised protein	4.66	3.96E-11
Smp_135560	Collagen alpha 2(I) chain	4.60	1.49E-14
Smp_110280	Nucleolar phosphoprotein p130	4.42	4.00E-09
Smp_043390	Beta D xylosidase 2	4.29	6.94E-06
Smp_050520	Neurogenic locus notch protein	3.97	1.14E-10
Smp_166540	Serine:threonine protein kinase Nek11	3.85	7.08E-06
Smp_124040	FMRFamide activated amiloride sensitive sodium	3.72	2.10E-06
Smp_062720	e1b 55 kDA associated protein	3.42	1.10E-06
Smp_128470	Cation transporting ATPase worm	3.39	2.08E-05
Smp_165040	Tbc1 domain family member	3.36	2.08E-05
Smp_133340	Uncharacterised protein	3.35	1.54E-05

Smp_137550	Dynein heavy chain	3.27	1.40E-07
Smp_168590	Pema SRCR protein	3.23	5.15E-07
Smp_188670	Uncharacterised protein	3.23	1.14E-10
Smp_062120	Serine protease inhibitors serpins	3.08	7.32E-10
Smp_143910	Coiled coil and C2 domain containing protein 2A	2.98	3.33E-06
Smp_164000	Uncharacterised protein	2.94	3.50E-05
Smp_123010	High affinity cationic amino acid transporter 1	2.79	9.13E-06
Smp_068720	Formin Homology 2 Domain containing protein	2.76	4.83E-08
Smp_213300	Putative transcription elongation regulator 1	2.66	7.56E-07
Smp_091820	Bestrophin 3	2.50	4.15E-10
Smp_136850	Centrin 1	2.50	5.09E-05
Smp_120320	Splicing factor 3A subunit 1	2.43	2.34E-09
Smp_189850	Uncharacterised protein	2.39	2.33E-06
Smp_025130	RNA binding motif protein 22	2.37	4.31E-06
Smp_151730	Rhabdoid tumour deletion region protein 1	2.35	9.33E-09
Smp_013860	Glutamate cysteine ligase catalytic subunit	2.34	5.25E-08
Smp_105950	Armadillo repeat containing protein 3	2.33	5.36E-07
Smp_021540	Zinc finger protein	2.28	3.34E-05
Smp_134660	Radial spoke head protein 4 A	2.25	3.62E-06
Smp_131890	Sodium:chloride dependent transporter	2.25	2.02E-06
Smp_156080	Dynein heavy chain	2.24	1.48E-05
Smp_175090	Long-chain-fatty-acid--CoA ligase	2.20	2.19E-05
Smp_180810	Papilin	2.18	1.22E-05
Smp_001160	Sperm associated antigen 16 protein	2.18	2.08E-05
Smp_005860	Heterogeneous nuclear ribonucleoprotein k	2.16	2.19E-05
Smp_165060	Uncharacterised protein	2.16	1.04E-05
Smp_136550	Tyrosine protein kinase transmembrane receptor	2.08	4.83E-08
Smp_067420	Splicing factor 3b subunit 4	2.07	2.14E-05
Smp_007640	Uncharacterised protein	2.03	3.50E-05

Appendix A.14: The most differentially expressed genes in female worms from mixed sex infections at 21 days *post* infection (d.p.i.) compared to females at 18 d.p.i. The fold change provides a measure of expression in females at 21 d.p.i. relative to females at 18 d.p.i. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Fold Change	Adjusted p-value
Smp_150630	Dissatisfaction (Dsf)	8.56	1.43E-04
Smp_180330	MEG 2 (ESP15) family	5.90	1.74E-03
Smp_134100	G protein coupled receptor fragment	4.54	4.01E-03
Smp_138350	Zinc finger protein basonuclin 2	4.32	1.61E-04
Smp_083880	G protein coupled receptor No9	3.89	5.00E-08
Smp_128710	G protein coupled receptor	3.86	3.85E-04
Smp_167140	WNT	3.07	6.70E-03
Smp_177670	Placental protein 11	2.79	4.84E-04
Smp_062080	Serine protease inhibitors serpins	2.75	5.57E-08
Smp_196930	WNT	2.49	2.72E-03
Smp_167400	Transcription factor sum 1	2.36	1.52E-03
Smp_062560	Secreted frizzled protein	2.21	6.67E-03
Smp_041700	G protein coupled receptor fragment	2.08	6.48E-03
Smp_157300	Basic fibroblast growth factor receptor 1 A	2.03	3.80E-05

Appendix A.15: The most differentially expressed genes in female worms at 18 days *post* infection (d.p.i.) compared to females from mixed sex infections at 21 d.p.i. The fold change provides a measure of expression in females at 18 d.p.i. relative to females at 21 d.p.i. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Fold change	Adjusted p-value	Product
Smp_062620	12.61	1.73E-07	Uncharacterised protein
Smp_138350	5.43	1.86E-05	Zinc finger protein basonuclin 2
Smp_000620	5.13	2.15E-04	Class E basic helix loop helix protein 23
Smp_136830	4.90	3.09E-04	Subfamily A1A unassigned peptidase (A01 family)
Smp_203270	4.61	3.23E-04	Twist protein
Smp_158480	3.43	4.75E-04	AMP dependent ligase
Smp_119140	3.32	1.72E-06	Uncharacterised protein
Smp_203660	2.76	3.69E-06	Uncharacterised protein
Smp_174320	2.70	3.10E-05	Protein lozenge
Smp_201920	2.66	1.04E-04	Uncharacterised protein
Smp_124300	2.42	1.23E-04	Lamin B receptor (ERG24)
Smp_026670	2.40	1.54E-05	Transmembrane protein 26
Smp_157370	2.32	4.82E-05	Ribosomal protein s kinase alpha
Smp_202620	2.32	1.56E-04	Uncharacterised protein
Smp_202400	2.29	9.58E-06	Uncharacterised protein
Smp_130870	2.20	2.56E-04	ETS domain containing protein Elk 4
Smp_130050	2.18	5.61E-04	PDZ domain containing RING finger protein
Smp_135730	2.15	2.67E-04	Protocadherin alpha 7
Smp_086220	2.12	3.60E-04	Uncharacterised protein
Smp_137110	2.02	4.50E-05	Uncharacterised protein

Appendix A.16: The most differentially expressed genes in male worms at 28 days *post* infection (d.p.i.) compared to males from mixed sex infections at 21 d.p.i. The fold change provides a measure of expression in males at 28 d.p.i. relative to males at 21 d.p.i. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Fold change	Adjusted p-value
Smp_000620	Class E basic helix loop helix protein 23	3.89	0.000546398
Smp_001980	Barh 2 homeobox protein	3.98	0.000461429
Smp_009020	Zinc finger protein	2.07	0.006630352
Smp_009040	POU domain, class 4, transcription factor 2	3.16	7.04E-05
Smp_027990	Homeobox protein NK 2	2.70	0.008565627
Smp_035650	Dorsal root ganglia homeobox protein	2.73	0.001246198
Smp_045470	Prospero homeobox protein 2	9.87	1.12E-06
Smp_049160	Zinc finger protein	2.17	0.000610056
Smp_051130	ERG transcription factor	2.52	2.27E-06
Smp_060220	Lipopolysaccharide induced tumour necrosis	3.73	0.002346011
Smp_072470	Neurogenic differentiation factor	2.40	0.001806143
Smp_075230	B cell lymphoma/leukaemia 11B	5.04	3.13E-05
Smp_076600	Transcription factor SOX 14	2.33	0.000577322
Smp_124090	Uncharacterised protein	2.53	0.002649938
Smp_124600	Uncharacterised protein	3.67	1.25E-06
Smp_125400	Neurogenic differentiation 6	5.89	0.001685335
Smp_126560	OTP paired class homeobox protein	9.09	3.52E-07
Smp_130870	ETS domain containing protein Elk 4	2.12	2.78E-05
Smp_134690	EMX homeobox protein	3.45	0.00905731
Smp_136120	N twist	4.37	0.009792167
Smp_136900	Homeobox protein distal less dlx	7.42	8.67E-06
Smp_138230	Doublesex and mab-3 related transcription factor 1	3.23	2.04E-05
Smp_138350	Zinc finger protein basonuclin 2	7.81	7.45E-11
Smp_140980	MDS1 and evi1 complex locus protein evi1	2.35	0.004593848
Smp_142120	Achaete scute transcription factor	9.13	4.22E-06
Smp_148110	Transcription factor SOX 14	3.50	0.008318475
Smp_153850	Zinc finger ch type domain containing protein	2.17	5.83E-05
Smp_155690	Sox family of transcription factor	3.71	0.003879486
Smp_162760	Homeobox protein prospero	3.13	4.79E-06
Smp_165410	LIM class homeodomain transcription factor	2.78	0.001109032
Smp_168600	Aryl hydrocarbon receptor	2.42	0.002250422
Smp_170750	Paired box protein pax 6	5.27	4.30E-07
Smp_171130	Glial cells missing	10.95	8.12E-07
Smp_174320	Protein lozenge	3.02	1.02E-08
Smp_175410	Zinc finger protein 398	4.26	6.23E-05
Smp_189380	Homeobox protein distal less	4.44	0.006044648
Smp_203270	Twist protein	4.33	4.67E-05

Appendix A.17: List of genes involved in the regulation of transcription (annotated with GO term G0:0006355) which were up-regulated in females from mixed sex infections at 21 days *post* infection (d.p.i.) compared to females at 28 d.p.i. The fold change provides a measure of expression in females at 21 d.p.i. relative to females at 28 d.p.i. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Fold change	Adjusted p-value
Smp_000620	Class E basic helix loop helix protein 23	1.90	0.114
Smp_001980	Barh 2 homeobox protein	2.72	0.020
Smp_009020	Zinc finger protein	1.23	0.604
Smp_009040	POU domain, class 4, transcription factor 2	1.87	0.066
Smp_027990	Homeobox protein NK 2	2.13	0.059
Smp_035650	Dorsal root ganglia homeobox protein	2.13	0.025
Smp_045470	Prospero homeobox protein 2	4.50	3.67E-03
Smp_049160	Zinc finger protein	1.74	0.026
Smp_051130	ERG transcription factor	1.90	3.68E-03
Smp_060220	Lipopolysaccharide induced tumour necrosis	4.00	1.40E-03
Smp_072470	Neurogenic differentiation factor	3.18	5.47E-05
Smp_075230	B cell lymphoma/leukaemia 11B	3.70	1.87E-03
Smp_076600	Transcription factor SOX 14	2.75	5.01E-05
Smp_124090	Uncharacterised protein	1.63	0.187
Smp_124600	Uncharacterised protein	2.47	2.60E-03
Smp_125400	Neurogenic differentiation 6	7.92	1.11E-04
Smp_126560	OTP paired class homeobox protein	1.62	0.416
Smp_130870	ETS domain containing protein Elk 4	1.24	0.419
Smp_134690	EMX homeobox protein	4.15	2.58E-03
Smp_136120	N twist	0.66	0.645
Smp_136900	Homeobox protein distal less dlx	3.96	3.50E-03
Smp_138230	Doublesex and mab-3 related transcription factor 1	2.37	4.14E-03
Smp_138350	Zinc finger protein basonuclin 2	7.94	3.98E-10
Smp_140980	MDS1 and evi1 complex locus protein evi1	1.58	0.209
Smp_142120	Achaete scute transcription factor	5.21	1.37E-03
Smp_148110	Transcription factor SOX 14	2.55	0.064
Smp_153850	Zinc finger ch type domain containing protein	1.87	2.91E-03
Smp_155690	Sox family of transcription factor	1.80	0.285
Smp_162760	Homeobox protein prospero	2.56	5.44E-04
Smp_165410	LIM class homeodomain transcription factor	2.38	8.69E-03
Smp_168600	Aryl hydrocarbon receptor	1.56	0.203
Smp_170750	Paired box protein pax 6	1.76	0.148
Smp_171130	Glial cells missing	3.05	0.051
Smp_174320	Protein lozenge	3.75	8.37E-11
Smp_175410	Zinc finger protein 398	2.84	8.82E-03
Smp_189380	Homeobox protein distal less	1.36	0.694
Smp_203270	Twist protein	2.30	0.041

Appendix A.18: List of genes from Appendix A.17 expressed in female worms from single sex infections at 21 days *post* infection (d.p.i.) compared to females at 28 d.p.i. These genes are involved in the regulation of transcription (annotated with GO term G0:0006355). The fold change provides a measure of gene expression in females from single sex infections at 21 d.p.i. compared to females at 28 d.p.i. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Product	Fold change	Adjusted p-value
Smp_191910	Stress protein DDR48	455.30	1.26E-41
Smp_131110	Uncharacterised protein	387.32	3.95E-38
Smp_202830	Uncharacterised protein	331.05	4.34E-31
Smp_050750	Uncharacterised protein	315.19	1.52E-39
Smp_144440	Replication A protein	281.18	2.53E-38
Smp_151150	Arabinogalactan protein	266.26	2.97E-39
Smp_014610	Uncharacterised protein	246.64	5.15E-33
Smp_033250	Uncharacterised protein	243.38	1.78E-33
Smp_077890	Trematode Eggshell Synthesis domain containing protein	224.72	8.07E-32
Smp_095980	Extracellular superoxide dismutase	213.90	1.57E-30
Smp_000410	Trematode Eggshell Synthesis domain containing protein	203.71	2.72E-32
Smp_151160	Stress protein DDR48	203.06	1.98E-32
Smp_000420	Pro His rich protein	199.29	1.57E-30
Smp_000430	Eggshell protein	197.15	1.05E-29
Smp_192220	Uncharacterised protein	178.59	1.66E-30
Smp_076320	Uncharacterised protein	173.56	6.66E-37
Smp_050270	Tyrosinase	172.20	3.14E-28
Smp_165360	Histone acetyltransferase myst4	170.11	2.15E-28
Smp_077900	Female specific protein 800 (fs800)	168.73	4.01E-38
Smp_138570	Spore germination protein	167.61	5.00E-28
Smp_013540	Tyrosinase	165.63	4.89E-28
Smp_000280	Uncharacterised protein	155.70	4.07E-27
Smp_000270	Uncharacterised protein	147.29	9.32E-27
Smp_000400	Uncharacterised protein	143.08	2.21E-26
Smp_155310	Tetraspanin CD63 receptor	140.39	4.62E-26
Smp_148390	Uncharacterised protein	140.06	1.87E-26
Smp_000290	Female-specific protein 800 (fs800)	139.81	1.37E-34
Smp_095350	Tis11 zinc finger protein	46.36	2.34E-16
Smp_087320	TPA induced protein 11B mouse	37.10	8.14E-18
Smp_166560	Zinc finger homeobox protein 1	12.20	1.13E-10

Appendix A.19: The most differentially expressed genes in females from mixed sex infections at 38 days *post* infection (d.p.i.) compared to females at 18 d.p.i. Three putative transcription factor genes (Smp_095350, Smp_087320 and Smp_166560) were added to this list despite ranking lower on the list of most differentially expressed genes but are in the 100 most differentially expressed genes. Fold changes provide a measure of gene expression in females at 38 d.p.i. compared to females at 18 d.p.i. The *p*-value has been adjusted for multiple hypothesis testing.

Gene ID	Fold change	Adjusted p-value	Product	Rank	Eggshell synthesis domain
Smp_000270	243.96	8.24E-34	Uncharacterised protein	17	Yes
Smp_000280	262.95	5.31E-34	Uncharacterised protein	14	Yes
Smp_000290	463.72	5.79E-61	Female-specific protein 800	3	Yes
Smp_000390	142.88	3.93E-26	Uncharacterised protein	28	Yes
Smp_000400	252.51	7.23E-34	Uncharacterised protein	16	
Smp_000410	146.36	2.85E-32	Trematode Eggshell Synthesis domain containing protein	27	Yes
Smp_000420	214.78	3.81E-34	Pro His rich protein	20	
Smp_000430	374.21	5.91E-39	Eggshell protein	6	Yes
Smp_013540	353.16	1.35E-38	Tyrosinase	8	
Smp_014610	427.38	1.30E-43	Serine:threonine kinase 1	5	
Smp_050270	243.54	1.61E-33	Tyrosinase	18	
Smp_050750	265.17	1.30E-43	Uncharacterised protein	13	
Smp_077890	351.06	4.31E-40	Trematode Eggshell Synthesis domain containing protein	9	Yes
Smp_077900	504.95	4.69E-65	Female specific protein 800	2	Yes
Smp_077920	183.03	2.02E-29	Uncharacterised protein	22	Yes
Smp_095980	323.47	9.38E-37	Extracellular superoxide dismutase (Cu Zn)	11	
Smp_112450	179.17	9.08E-30	Trematode Eggshell Synthesis domain containing protein	23	Yes
Smp_131110	616.92	5.50E-48	Uncharacterised protein	1	
Smp_138570	261.01	1.15E-34	Spore germination protein	15	
Smp_144440	169.70	2.82E-37	Replication A protein	25	
Smp_151150	362.41	8.12E-50	Arabinogalactan protein	7	
Smp_151160	230.02	1.82E-38	Stress protein DDR48	19	
Smp_155310	190.75	5.98E-30	Tetraspanin CD63 receptor	21	
Smp_165360	175.53	1.24E-30	Histone acetyltransferase myst4	24	
Smp_191910	431.31	3.31E-46	Stress protein DDR48	4	Yes
Smp_033250	96.72	4.80E-26	Uncharacterised protein	36	
Smp_152150	93.32	2.47E-22	Uncharacterised protein	38	

Appendix A.20: Expression of genes of Cluster 13 from the Kohonen analysis in the comparison of female worms from mixed sex infections to males worms at 38 days *post* infection. Fold changes provide a measure of gene expression in females worms relative to male worms. The *p*-value has been adjusted for multiple hypothesis testing. The Rank provides the position of the gene within the list of differentially expressed gene when ordered by fold change. All genes in Cluster 13 where amongts the 40 most differentially expressed genes. The final column indicates the presence or absence of a trematode eggshell synthesis domain in the encoded protein.

Gene ID	Reads in females from mixed sex infections at 38 d.p.i.	Reads in females from single sex infections at 38 d.p.i.
Smp_000270	24679	14
Smp_000280	143373	49
Smp_000290	13081	17
Smp_000390	67711	14
Smp_000400	26606	14
Smp_000410	128053	501
Smp_000420	208495	299
Smp_000430	483455	97
Smp_013540	31888	8
Smp_014610	1045457	415
Smp_033250	66492	416
Smp_050270	92273	61
Smp_050750	39007	75
Smp_077890	44949	23
Smp_077900	16593	20
Smp_077920	18893	4
Smp_095980	204041	87
Smp_112450	16580	12
Smp_131110	3988553	1125
Smp_138570	229180	122
Smp_144440	129993	498
Smp_151150	5309	6
Smp_151160	44898	92
Smp_152150	10866	25
Smp_155310	84156	60
Smp_165360	50134	77
Smp_191910	1321831	1072
Sum of fertility gene reads	8536546	5203
Total of all gene reads	59621509	84076361
Percentage of reads	14.31%	0.0062%

Appendix A.21: Numbers of reads in females from mixed and single sex infections at 38 days *post* infection for the 27 genes in Cluster 13 of the Kohonen analysis. This is the raw read count, not normalised by gene length.

Smp_000950	Smp_075330	Smp_136870	Smp_152310	Smp_185440
Smp_005710	Smp_078200	Smp_137700	Smp_152910	Smp_186940
Smp_007670	Smp_083770	Smp_139700	Smp_154030	Smp_187360
Smp_011180	Smp_084890	Smp_139720	Smp_154830	Smp_194840
Smp_011960	Smp_085180	Smp_140570	Smp_154900	Smp_198170
Smp_012930	Smp_087760	Smp_141340	Smp_156250	Smp_198260
Smp_020840	Smp_089840	Smp_141900	Smp_158120	Smp_198900
Smp_025830	Smp_095360	Smp_142020	Smp_158360	Smp_199830
Smp_026400	Smp_101100	Smp_142620	Smp_158510	Smp_200150
Smp_029930	Smp_121070	Smp_142680	Smp_158520	Smp_200310
Smp_034680	Smp_125060	Smp_144270	Smp_159470	Smp_202450
Smp_036660	Smp_125740	Smp_145240	Smp_160160	Smp_202690
Smp_038030	Smp_126210	Smp_145290	Smp_163060	Smp_203200
Smp_038970	Smp_127130	Smp_145760	Smp_164140	Smp_204130
Smp_041540	Smp_128190	Smp_146180	Smp_170320	Smp_204250
Smp_042700	Smp_128480	Smp_146810	Smp_172830	Smp_205410
Smp_056290	Smp_129920	Smp_148770	Smp_172980	Smp_206250
Smp_058690	Smp_132660	Smp_149610	Smp_173520	Smp_206290
Smp_058700	Smp_135020	Smp_149780	Smp_175310	Smp_209060
Smp_061100	Smp_135840	Smp_151660	Smp_179710	Smp_212090
Smp_062070	Smp_136730	Smp_151860	Smp_181390	Smp_212820
Smp_063520				

Appendix A.22: All genes belonging to Cluster 1 of the MBCluster analysis.

Smp_000500	Smp_056680	Smp_123880	Smp_141460	Smp_165500
Smp_000850	Smp_058380	Smp_124070	Smp_142490	Smp_167000
Smp_001410	Smp_059980	Smp_125070	Smp_143540	Smp_167960
Smp_004420	Smp_063530	Smp_125190	Smp_144260	Smp_169020
Smp_011270	Smp_063630	Smp_125250	Smp_144770	Smp_170560
Smp_012560	Smp_073990	Smp_126230	Smp_147450	Smp_171270
Smp_016490	Smp_074000	Smp_126820	Smp_148680	Smp_171290
Smp_018870	Smp_074570	Smp_128110	Smp_151120	Smp_172740
Smp_019030	Smp_075800	Smp_128300	Smp_153070	Smp_174500
Smp_026810	Smp_077610	Smp_128350	Smp_155510	Smp_177540
Smp_027340	Smp_089240	Smp_129270	Smp_155570	Smp_179170
Smp_030000	Smp_089290	Smp_130100	Smp_155900	Smp_181060
Smp_033930	Smp_090010	Smp_131410	Smp_157080	Smp_181070
Smp_037720	Smp_090110	Smp_131470	Smp_158620	Smp_181470
Smp_042140	Smp_091750	Smp_132800	Smp_161240	Smp_187140
Smp_042910	Smp_094500	Smp_133170	Smp_162320	Smp_191970
Smp_043120	Smp_096290	Smp_134590	Smp_162390	Smp_195070

Smp_043290	Smp_096840	Smp_137720	Smp_163310	Smp_197730
Smp_049150	Smp_102430	Smp_139160	Smp_163380	Smp_199290
Smp_052470	Smp_105420	Smp_139240	Smp_163700	Smp_201330
Smp_054620	Smp_105450	Smp_140460	Smp_163720	Smp_203910
Smp_055740	Smp_121930	Smp_140840	Smp_164320	Smp_900040
Smp_055990	Smp_123720			

Appendix A.23: All genes belonging to Cluster 4 of the MBCluster analysis.

Gene ID	Average read count	Fold change	Gene ID	Average read count	Fold change
Smp_000270	1098.89	243.96	**Smp_129960**	414.79	101.63
Smp_000280	5844.21	262.95	**Smp_130370**	143.49	24.89
Smp_000290	520.85	463.72	**Smp_130780**	22.76	94.63
Smp_000370	40.72	152.97	**Smp_131110**	146222.11	616.92
Smp_000390	2812.64	142.88	**Smp_137460**	167.86	35.98
Smp_000400	1056.45	252.51	**Smp_138570**	10245.16	261.01
Smp_000410	5371.41	146.36	**Smp_140720**	336.99	22.60
Smp_000420	9619.11	214.78	**Smp_141450**	104.15	98.45
Smp_000430	20613.88	374.21	**Smp_144440**	6322.04	169.70
Smp_011870	222.98	47.15	**Smp_145020**	135.79	23.62
Smp_013540	1311.58	353.16	**Smp_145490**	63.19	113.36
Smp_014610	49190.98	427.38	**Smp_146480**	99.57	17.68
Smp_029620	4.49	10.18	**Smp_148390**	317.14	71.70
Smp_033250	2959.04	96.72	**Smp_151150**	237.48	362.41
Smp_041880	56.69	30.47	**Smp_151160**	1990.40	230.02
Smp_050270	3936.38	243.54	**Smp_151850**	85.29	28.57
Smp_050750	1575.47	265.17	**Smp_152150**	443.42	93.32
Smp_054010	876.51	37.52	**Smp_154390**	30.35	20.51
Smp_055260	226.77	22.05	**Smp_155310**	3646.08	190.75
Smp_056280	82.41	5.48	**Smp_157730**	9.54	24.66
Smp_056700	1482.19	17.68	**Smp_159480**	57.64	28.80
Smp_058160	199.83	112.56	**Smp_159920**	1085.76	30.49
Smp_065530	107.12	69.46	**Smp_164400**	424.85	40.10
Smp_070360	30.37	14.18	**Smp_165360**	2524.98	175.53
Smp_074710	246.16	38.78	**Smp_166560**	217.40	29.12
Smp_076320	648.59	25.03	**Smp_167830**	28.59	17.07
Smp_077890	2068.34	351.06	**Smp_169090**	551.51	30.16
Smp_077900	675.45	504.95	**Smp_170770**	11.19	21.11
Smp_077920	916.91	183.03	**Smp_170840**	47.48	19.16
Smp_078720	58.79	22.57	**Smp_173150**	3694.12	38.76
Smp_080360	54.01	41.78	**Smp_175560**	457.47	138.39
Smp_087320	75.30	15.68	**Smp_191910**	54945.48	431.31
Smp_095350	151.74	13.84	**Smp_194060**	6.05	4.10

Smp_095980	7743.79	323.47	**Smp_194750**	137.24	86.72
Smp_105780	68.84	1.99	**Smp_195090**	551.03	3.99
Smp_112450	827.17	179.17	**Smp_196950**	203.53	7.69
Smp_122280	66.19	44.71	**Smp_199540**	163.74	50.27
Smp_124510	92.96	43.69	**Smp_202840**	40.55	113.42

Appendix A.24: Summary of the 76 genes in Cluster 7 generated by MBCluster. The column "Average read count" provides a measure of absolute expression, the average number of reads mapped to the gene across all RNA-Seq samples used in Chapter 3. The fold changes provided reflect expression in females from mixed sex infections compared to females from single sex infections at 38 days *post* infection.

GO term	Description	Annotated	Significant	Expected	p-value
GO:0006570	Tyrosine metabolic process	4	2	0.02	1.40E-04
GO:0046189	Phenol-containing compound biosynthetic process	4	2	0.02	1.40E-04
GO:0044550	Secondary metabolite biosynthetic process	7	2	0.03	5.00E-04
GO:0006875	Cellular metal ion homeostasis	10	2	0.05	1.05E-03
GO:0046148	Pigment biosynthetic process	11	2	0.05	1.28E-03

Appendix A.25: Significantly enriched GO terms amongst the genes in Cluster 7 of the MBCluster analysis. "Annotated" provides the total number of genes in the genome annotated with a given GO term. "Significant" provides the number of genes annotated with that GO term in Cluster 9. "Expected" provides the number of genes with that GO term expected by chance. The significance cut-off was p = 0.01.

Target Gene	CD63R Replicate 1	CD63R Replicate 2	CD63a Replicate 1	CD63a Replicate 2	Control Replicate 1	Control Replicate 2	No template (negative control)
CD63R	20.86	19.52	22.05	23.04	21.51	21.27	37.15
	20.82	19.37	21.8	22.84	21.37	21.3	> 40.00
	20.8	19.41	21.9	22.83	21.49	21.44	35.97
CD63a	22.95	22.49	23.41	22.96	22.42	21.62	36
	22.91	22.54	23.34	23.17	22.69	21.62	35.92
	22.93	22.43	23.41	23.18	22.5	21.74	34.24
PDMS4 (House keeping gene)	17.52	16.65	17.2	18.21	16.68	16.56	> 40.00
	17.65	16.57	17.03	18.26	16.52	16.66	34.4
	17.65	16.57	17.23	18.18	16.55	16.52	35.8
No reverse transcription (negative control)	34.4	28.61	28.4	> 40.00	27.84	29.02	> 40.00
	31	28.56	35.5	> 40.00	28.25	33.65	37.14
	30.15	28.99	29.55	> 40.00	29.03	> 40.00	> 40.00

Appendix A.26: This table provides the C_T for all of the qRT-PCR reactions used to measure gene expression the RNAi knockdown expreiment using the soaking method. Three PCR replicates are given for each combination of target gene and treatment replicate (e.g. CD63R 1 and CD63 R2). C_T were not determined beyond 40 cycles of replication; samples that did not show amplification were assigend " > 40 " as C_T value. For each sample replciate a negative control was run which had not been reverse transcribed to detect gDNA contamination in the total RNA extracted from the worms. Another negative control with out any added template was run together with the samples.

Target Gene	CD63R Replicate 1	CD63R Replicate 2	CD63R 3 Replicate 3	CD63a Replicate 1	CD63a Replicate 2	CD63a Replicate 3	Control Replicate 1	Control Replicate 2	Control Replicate 3	No template (negative control)
CD63a	24.67	23.1	23.91	23.05	22.93	22.66	23.49	22.66	23.87	31.92
	24.64	22.84	23.64	23.12	22.95	22.66	23.53	22.59	23.84	34.47
	24.41	22.78	23.63	23	22.76	22.6	23.34	22.66	23.79	33.22
CD63R	19.88	18.97	19.86	18.91	19.02	19.71	19.46	19.51	19.75	33.97
	19.9	18.97	19.54	18.9	19.07	19.45	19.4	19.51	19.77	22.93
	19.76	18.92	19.23	18.76	19.11	19.53	19.35	19.43	19.78	34.18
PDMS4 (House keeping gene)	19.27	18.77	18.74	18.45	18.29	18.42	16.24	15.82	16.48	> 40.00
	19.32	18.86	19	18.5	18.33	18.54	16.41	15.82	16.57	> 40.00
	19.38	18.97	18.9	18.57	18.3	18.53	16.43	15.91	16.59	30.63
No reverse transcription (negative control)	33.46	32.81	32.84	33.61	30.82	33.41	30.84	29.18	29.89	> 40.00
	31.79	32.79	34.81	32.9	30.52	33.94	31.28	28.99	28.44	37.1
	33.59	34.17	32.55	33.64	30.43	32.87	30.3	30.26	29.12	> 40.00

Appendix A.27: This table provides the C_T for all of the qRT-PCR reactions used to measure gene expression the RNAi knockdown expreiment using the electroporation method. Three PCR replicates are given for each combination of target gene and treatment replicate (e.g. CD63R 1 and CD63 R2). C_T were not determined beyond 40 cycles of replication; samples that did not show amplification were assigend " > 40 " as C_T value. For each sample replciate a negative control was run which had not been reverse transcribed to detect gDNA contamination in the total RNA extracted from the worms. Another negative control with out any added template was run together with the samples.

Appendix B

Samples	D0 F10	D0 F4	D0 F5	D0 F8	D0 F9	D0 M10	D0 M1	D0 M2	D0 M5	D0 M9	D4 FP10	D4 FP2	Average
Total Reads	42,839,264	40,885,356	40,355,812	18,025,886	49,625,790	14,063,226	42,329,640	24,937,274	27,513,990	67,601,548	23,971,200	23,260,574	36,514,504
Mapped	38916416	36646230	36493578	16072299	43413686	12496871	39181978	22370714	24462457	61949189	21860348	21076223	32715432
Prop. Paired	81.43%	78.82%	79.91%	78.76%	76.58%	79.18%	82.01%	79.81%	79.05%	82.28%	79.84%	79.27%	79.63%
Unmap.	3746955	4111222	3698354	1887292	6056171	1514468	2973642	2474627	2964568	5375856	2022005	2110713	3661759
% Mapped	90.84%	89.63%	90.43%	89.16%	87.48%	88.86%	92.56%	89.71%	88.91%	91.64%	91.19%	90.61%	89.60%
% Unmapped	8.75%	10.06%	9.16%	10.47%	12.20%	10.77%	7.02%	9.92%	10.77%	7.95%	8.44%	9.07%	10.03%
Failed Quality Control	0.41%	0.31%	0.41%	0.37%	0.31%	0.37%	0.41%	0.37%	0.32%	0.41%	0.37%	0.32%	0.38%

Samples	D4 FP3	D4 FP4	D4 FP5	D4 FP9	D4 FS10	D4 FS3	D4 FS6	D4 FS8	D4 FS9	D4 MP10	D4 MP1	D4 MP2	Average
Total Reads	27,937,410	38,813,550	35,647,926	48,950,808	30,331,014	20,601,208	21,341,254	38,262,878	61,440,742	39,086,800	19,033,002	41,838,666	36,514,504
Mapped	24681965	35272275	31809561	43934269	27024218	18506319	18967782	34423953	55229205	36277200	17194451	37390396	32715432
Prop. Paired	78.71%	80.24%	78.89%	78.38%	78.61%	78.84%	78.54%	78.97%	80.46%	82.61%	80.96%	80.17%	79.63%
Unmap.	3139249	3397391	3725939	4816482	3211224	2019624	2307076	3682758	5984085	2664807	1767972	4317276	3661759
% Mapped	88.35%	90.88%	89.23%	89.75%	89.10%	89.83%	88.88%	89.97%	89.89%	92.81%	90.34%	89.37%	89.60%
% Unmapped	11.24%	8.75%	10.45%	9.84%	10.59%	9.80%	10.81%	9.62%	9.74%	6.82%	9.29%	10.32%	10.03%
Failed Quality Control	0.42%	0.37%	0.32%	0.41%	0.32%	0.37%	0.31%	0.41%	0.37%	0.37%	0.37%	0.31%	0.38%

Samples	D4 MP3	D4 MP4	D4 MP5	D4 MP8	D4 MS10	D4 MS1	D4 MS2	D4 MS3	D4 MS4	D4 MS5	D4 MS8	D4 MS9	Average
Total Reads	31,991,790	22,369,714	24,146,306	37,815,740	35,386,012	37,843,540	51,749,442	12,819,534	26,090,006	50,489,724	20,859,514	38,584,968	36,514,504
Mapped	29114450	20082407	21549878	34139827	32264435	33920979	47483220	11495623	23269207	46193807	18537694	33844860	32715432
Prop. Paired	81.69%	80.09%	79.23%	80.15%	81.66%	79.87%	82.06%	80.61%	78.99%	81.28%	78.98%	77.17%	79.63%
Unmap.	2747238	2204828	2520353	3521572	2978354	3803876	4034666	1276989	2737848	4089750	2245034	4617616	3661759
% Mapped	91.01%	89.77%	89.25%	90.28%	91.18%	89.63%	91.76%	89.67%	89.19%	91.49%	88.87%	87.72%	89.60%
% Unmapped	8.59%	9.86%	10.44%	9.31%	8.42%	10.05%	7.80%	9.96%	10.49%	8.10%	10.76%	11.97%	10.03%
Failed Quality Control	0.41%	0.37%	0.32%	0.41%	0.40%	0.31%	0.45%	0.37%	0.32%	0.41%	0.37%	0.32%	0.38%

Samples	D8 FP10	D8 FP2	D8 FP4	D8 FP5	D8 FS1	D8 FS2	D8 FS3	D8 FS5	D8 FS8	D8 MP1	D8 MP2	D8 MP4	Average
Total Reads	47,008,208	24,550,604	36,397,450	44,329,112	10,712,766	31,616,670	32,081,458	68,971,632	50,614,744	25,197,978	36,872,448	35,550,458	36,514,504
Mapped	42150511	21965618	32047016	40412159	9292832	28375108	25469702	61525934	40167006	23003747	33246349	31627528	32715432
Prop. Paired	78.69%	79.58%	77.84%	81.71%	73.13%	79.92%	63.39%	79.96%	63.83%	81.14%	80.89%	79.98%	79.63%
Unmap.	4685762	2494367	4234879	3735666	1380139	3141867	6320535	7192351	10275620	2090259	3489994	3812064	3661759
% Mapped	89.67%	89.47%	88.05%	91.16%	86.75%	89.75%	79.39%	89.20%	79.36%	91.29%	90.17%	88.97%	89.60%
% Unmapped	9.97%	10.16%	11.64%	8.43%	12.88%	9.94%	19.70%	10.43%	20.30%	8.30%	9.47%	10.72%	10.03%
Failed Quality Control	0.37%	0.37%	0.32%	0.41%	0.37%	0.32%	0.91%	0.37%	0.34%	0.41%	0.37%	0.31%	0.38%

Samples	D8 MP5	D8 MP7	D8 MP8	D8 MP9	D8 MS10	D8 MS1	D8 MS2	D8 MS3	D8 MS5	D8 MS7	D8 MS8	D8 MS9	Average
Total Reads	42,434,848	17,753,578	25,831,508	43,743,710	38,120,912	41,733,998	45,553,098	75,767,817	42,100,608	41,931,428	55,188,534	47,965,560	36,514,504
Mapped	38840596	16174219	23345922	38874466	35194743	36987220	42074247	66679467	37881514	38121241	49698796	42757518	32715432
Prop. Paired	82.32%	81.83%	80.87%	80.44%	82.71%	78.82%	82.08%	90.16%	80.78%	81.77%	81.42%	80.16%	79.63%
Unmap.	3420972	1514069	2404286	4666249	2770194	4615435	3293712	8795063	4086470	3608151	5285191	5056152	3661759
% Mapped	91.53%	91.10%	90.38%	88.87%	92.32%	88.63%	92.36%	88.00%	89.98%	90.91%	90.05%	89.14%	89.60%
% Unmapped	8.06%	8.53%	9.31%	10.67%	7.27%	11.06%	7.23%	11.61%	9.71%	8.60%	9.58%	10.54%	10.03%
Failed Quality Control	0.41%	0.37%	0.31%	0.46%	0.41%	0.31%	0.41%	0.39%	0.32%	0.48%	0.37%	0.32%	0.38%

Table B.1: Summary of sequencing statistics for sequnced RNA-Seq samples. Total reads is the number of all reads sequenced. Mapped is the total number of reads mapped to the *S. mansoni* reference genome v5.2. 'Prop. Paired' refers to the total number of reads mapped together with a mate-paired read in close proximity and in the correct orientation. Unmapped is the total number of reads that could not be mapped to the reference by Tophat2.

Apoptosis-related proteins

14-3-3 proteins play an important role in relaying apoptosis signals within cells and exert pro-apoptotic effects at different points of the apoptosis pathway (Rosenquist, 2003). Three 14-3-3 proteins have been annotated in *S. mansoni* all of which were found to contain a 14-3-3 Pfam domain as well as significant BLAST hits (Table 4.1) against 14-3-3 proteins of model organisms.

The next group of genes I examined are caspases which are key mediators of apoptosis as they form the apoptotic protease cascade that lead to controlled cell death (McIlwain *et al.*, 2013). Four caspase homologues have been identified in *S. mansoni*, which all contain a caspase domain. Blast results suggest caspase-3, -7 and -8 homologues are represented in the *S. mansoni* genome. Three of these caspases (Smp_028500, Smp_141270 & Smp_032000) were previously described by Lee *et al.* (2014).

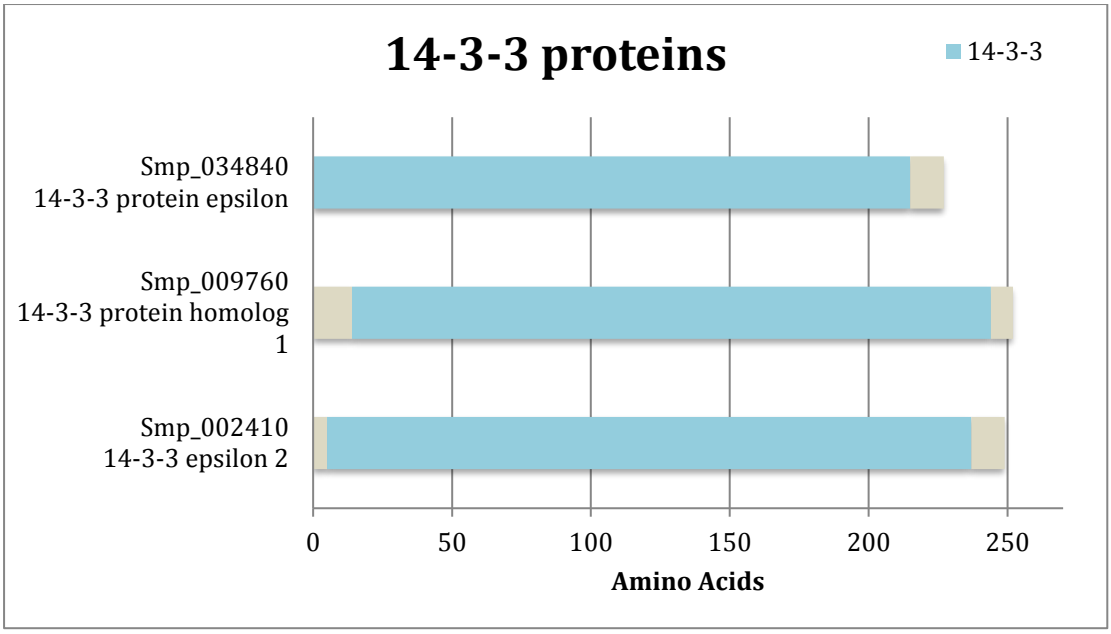


Figure B.2: Diagram of 14-3-3 family members showing length and presence of PFAM domains.

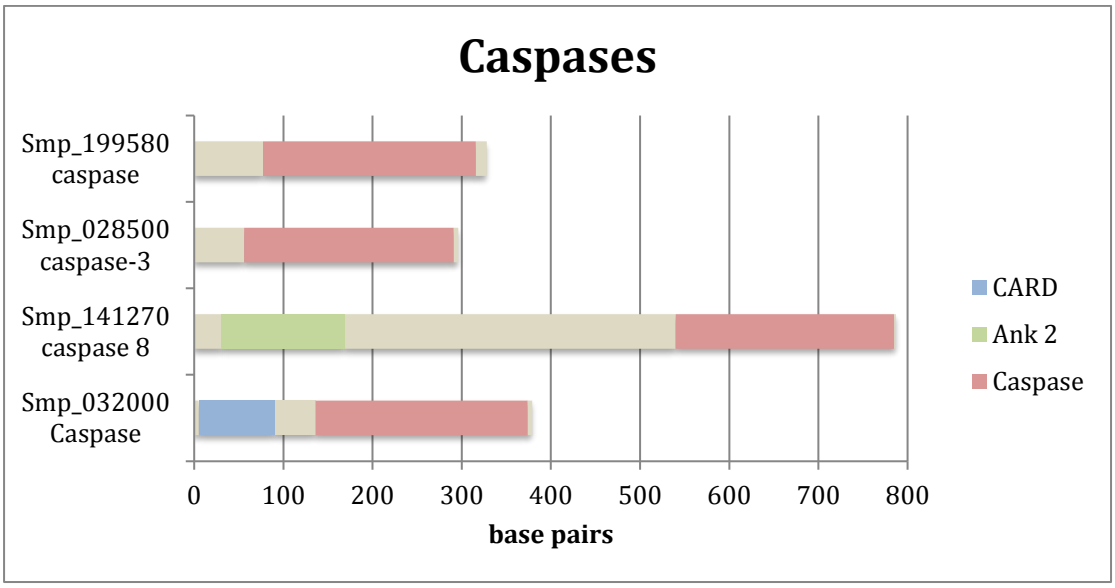


Figure B.3: Diagram of caspase family members showing length and presence of PFAM domains.

Another group of proteins that function as regulators of apoptosis, in both pro- and anti-apoptotic roles are Bcl-2 domain containing proteins. Four Bcl-2 domain containing proteins were identified by InterProScan, a further two were found to be annotated as Bcl-2 proteins, one in the genome (Smp_041630)

(Protasio *et al.*, 2012) and the other by Lee *et al.* (2011) (Smp_043360). However, InterProScan did not identify Bcl-2 domains in the latter two, nor did BLAST find significant homology to such proteins (Table B.5), which is why I decided to exclude them from the down-stream work flow of my analysis. One of the genes identified by InterProScan showed significant homology to apoptosis regulator Bax, also known as Bcl-2-like protein 4, a known pro-apoptotic member of the bcl-2 like family (Pawlowski & Kraft, 2000). Five of the Bcl-2 family proteins (Smp_043360, Smp_072180, Smp_095190, Smp_213250 & Smp_168470) were described previously (Lee *et al.*, 2014).

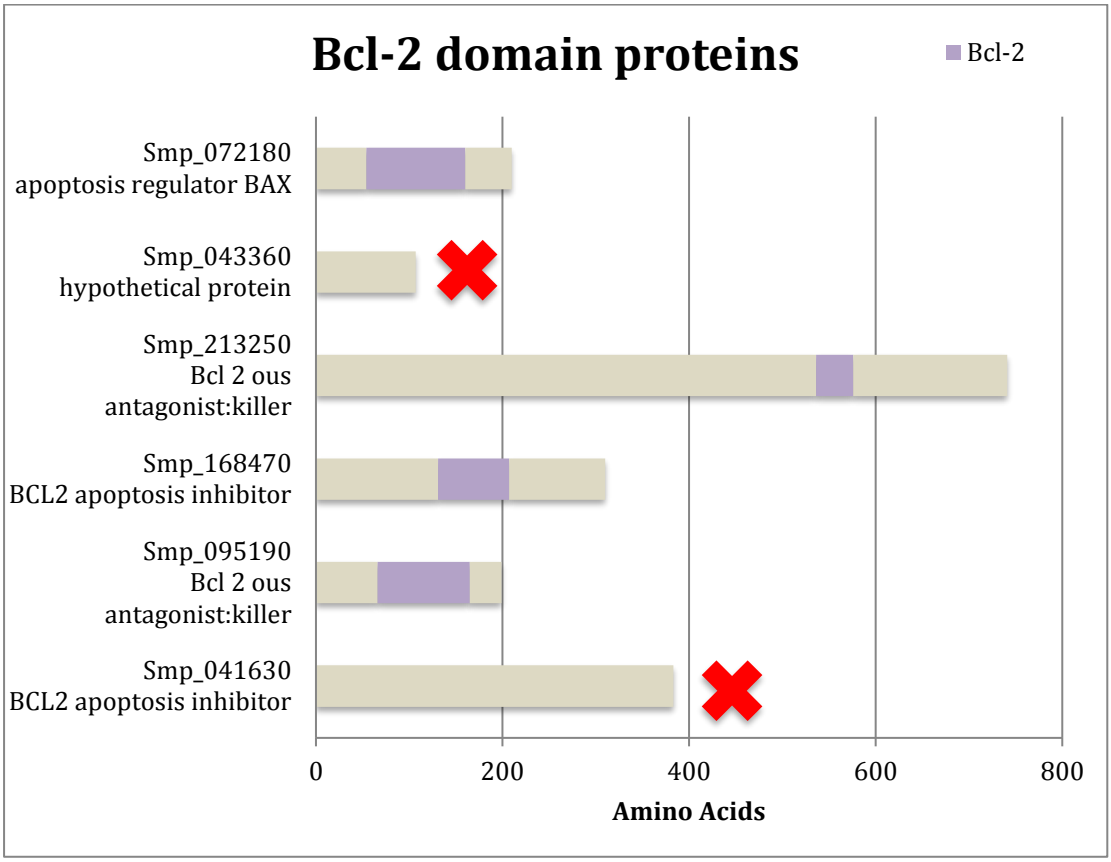


Figure B.4: Diagram of Bcl-2 domain containing protein showing length and presence of PFAM domains. Two genes, marked with 'X', failed the quality control and were excluded from the following analysis.

Two genes were found to contain Bax-Inhibition domains. Bax is known to have an important pro-apoptotic domain and its activity is tightly regulated by proteins such as Bax inhibitor 1. This group of proteins was first identified as an inhibitor of Bax but was later recognised to inhibit apoptosis through interactions with several other pro-apoptotic processes as well (Robinson *et al.*, 2011).

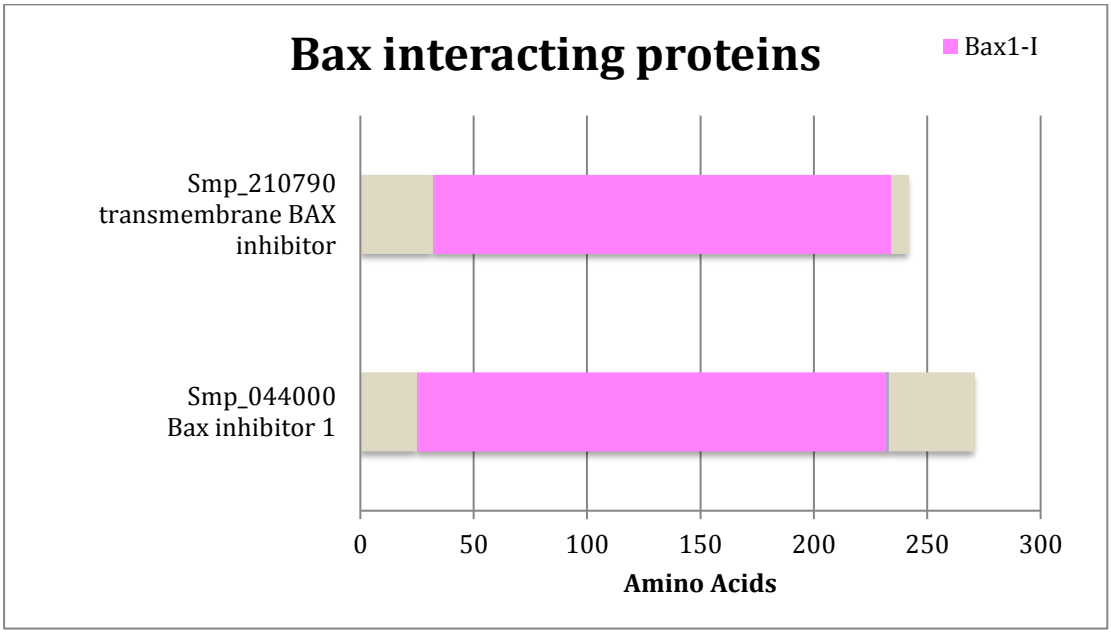


Figure B.5: Diagram of BAX interacting proteins showing length and presence of PFAM domains.

The Fas ligand:receptor interaction was one of the first aspects of the pro-apoptotic signalling cascade to be studied (Nagata, 1997, 1999). A homologue of the Fas receptor as well as several homologues of Fas associated proteins with putative roles in the regulation of apoptosis have been identified. Determining whether these genes should be assigned putative pro- or anti-apoptotic function based on domains was difficult as two of the genes did not contain known domains (see Figure B.7) so that their function could only be inferred by

homology to genes of model organisms. However, the function of the human fas binding factor 1 (FAB1) and fas associated factor 2 (FAF2) are only poorly understood. FAF2 is thought to confer resistance to apoptosis to T cells in patients with atopic dermatitis (Koelsch *et al.*, 2013). FAB1 on the other hand was first described in 2000 but no definitive function has yet been assigned (Schmidt *et al.*, 2000).

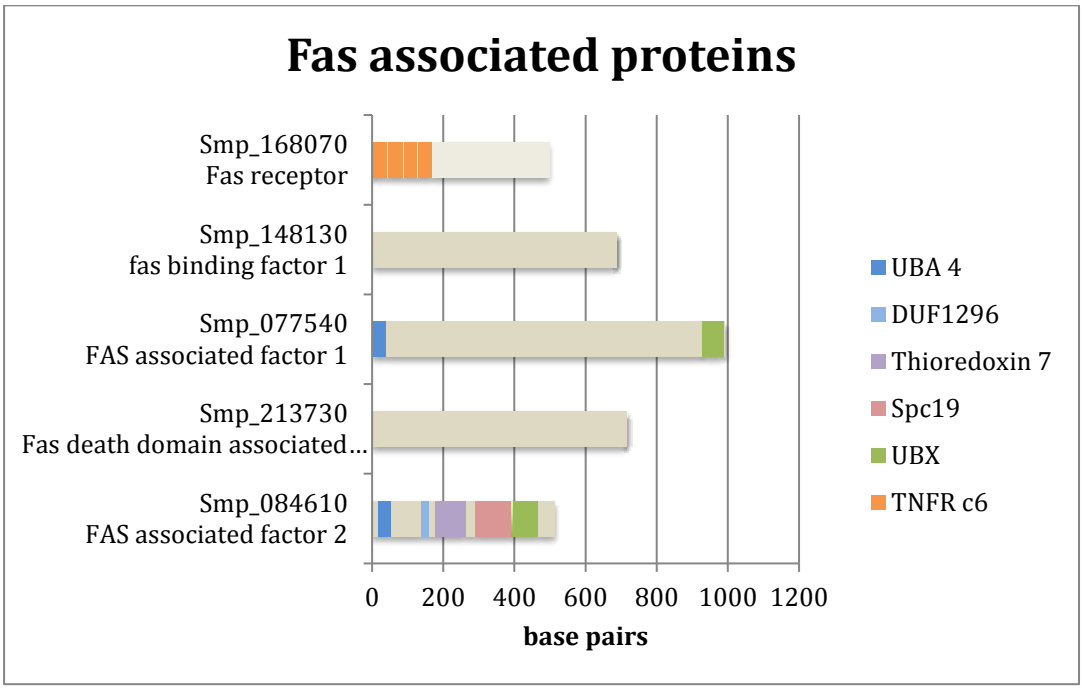


Figure B.6: Diagram of Fas receptor and associated proteins, showing length and presence of PFAM domains.

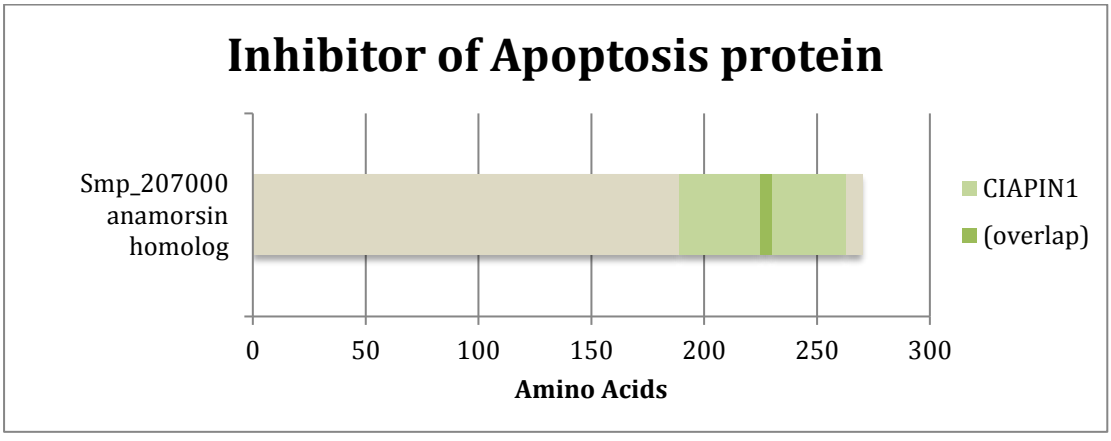


Figure B.7: Diagram of inhibitor of apoptosis protein, showing length and presence of PFAM domains.

The schistosome homologue of the Inhibitor of Apoptosis protein (IAP) was identified by Peng *et al.* (2010) and contains two overlapping "cytokine induced apoptosis inhibitor 1" domains, but lacks the Baculo virus IAP repeat (BIR) domain which is a hallmark of the family of IAPs. Rather than a member of the IAP family, Smp_207000 shares greater sequence similarity with the human anamorsin gene, a cytokine induced inhibitor of apoptosis. Given that the schistosome homologue was shown to inhibit caspases and thus apoptosis both in human cells as well as schistosome lysate (Peng *et al.*, 2010) it can be classified confidently as an anti-apoptotic gene.

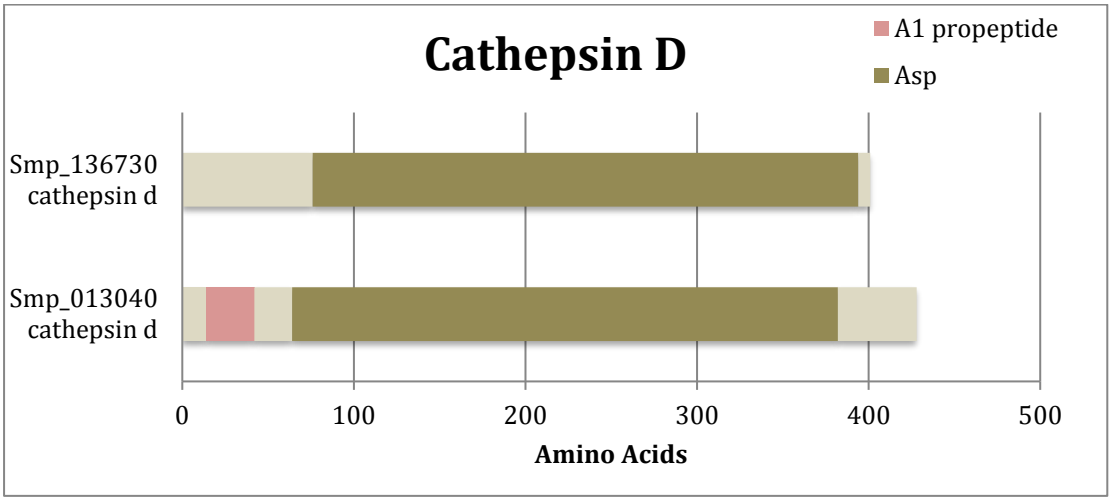


Figure B.8: Diagram of cathepsin D homologues, showing length and presence of PFAM domains.

Cathepsin D is a lysosomal protease that besides its role in digestion also plays a regulatory function in apoptosis. When activated, for example by ceramide (Heinrich *et al.*, 1999), cathepsin D mediates oxidative stress induced apoptosis by translocation to the cytoplasm (Watchorn *et al.*, 2001) as well as mediation of cytochrome C release and caspase activation (Conus *et al.*, 2008; Johansson *et al.*,

2003). Both cathepsin D homologues in *S. mansoni* contain aspartate protease domains as expected from cathepsins.

p21 is a cyclin dependent kinase (CDK) inhibitor and an important regulator of cell fate, controlling both division and apoptosis depending on the cell state and extracellular signals. By impeding CDKs it can arrest the cell cycle and is considered an important proliferation inhibitor (Gartel & Tyner, 2002). Despite this, p21 is a negative regulator of p53-dependent apoptosis. p21-activated kinase 1 (PAK1) are serine/threonine protein kinases serving as mediators of Rac and Cdc42 GTPase function as well as Ras-driven tumorigenesis by protecting cells from apoptosis (Ong *et al.*, 2011). As the schistosome homologue of PAK1 has a protein kinase domain as well as significant sequence similarity to the human PAK1, I assigned it a putative anti-apoptotic function. The PAK1 interacting protein (PAK1IP) homologue on the other hand contains a conserved WD40 repeat, as does the human PAK1IP. PAK1IP is a negative regulator of PAK1, binding to the N-terminal region of PAK1 to inactivate it, thus acting in a pro-apoptotic function (Xia *et al.*, 2001).

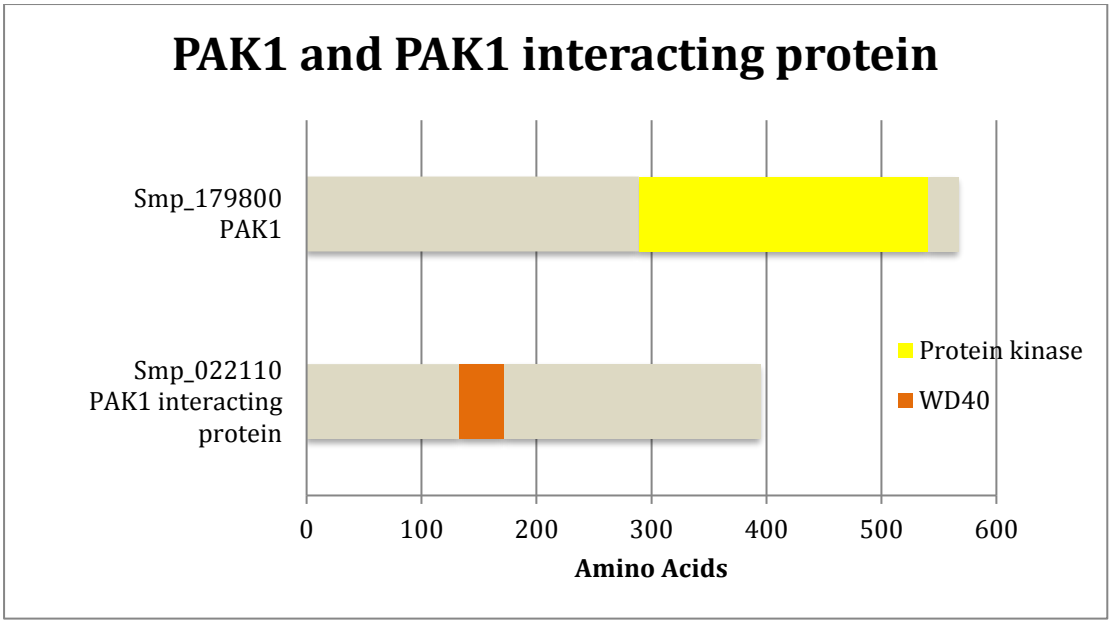


Figure B.9: Diagram of PAK1 and PAK1IP, showing length and presence of PFAM domains.

Five other *S. mansoni* genes were found to be homologous to apoptosis regulating genes of humans and *C. elegans*. These did not form a related group of genes and thus contain a range of different, unrelated protein domains, such as WD40, SH2 and Sh3 domains. The apoptotic protease activating factor (APAF1) like gene (Smp_140260) contains several WD domains that allow its human and *C. elegans* homologues to bind to caspase proteins and activate them, thereby promoting apoptosis (Qin *et al.*, 1999; Yakovlev *et al.*, 2001). The schistosome homologue of APAF1 (Smp_140260) was first described by Lee *et al*. (2014).

Almost the entire length of the schistosome homologue of apoptosis inhibitor 5 (AIP5) is made up of a conserved apoptosis inhibitory protein 5 domain, which, together with the sequence homology to human AIP5 suggests that it has an anti-apoptotic effect, acting as a suppressor of E2-promoter binding factor dependent apoptosis (Morris *et al.*, 2006).

In humans, the apoptosis inducing factor 1 is associated with the mitochondria and involved in a caspase independent activation of apoptosis (Joza *et al.*, 2001; Susin *et al.*, 1999). The schistosome homologue (Smp_137540) shares a "C-terminal mitochondrion-associated apoptosis-inducing factor"-domain as well as a pyridine nucleotide-disulphide oxidoreductase domain.

Like the human apoptosis-stimulating of p53 protein (PPP1R13B), the schistosome homologue has two domains near its C-terminus: an ankyrin repeat and a SH3 domain. In humans, these domains allow PPP1R13B to interact with p53/TP53 and regulate it by enhancing the DNA binding and transactivation function of TP53 on the promoters of pro-apoptotic genes *in vivo*, thereby specifically enhancing apoptosis but not cell cycle arrest (Samuels-Lev *et al.*, 2001).

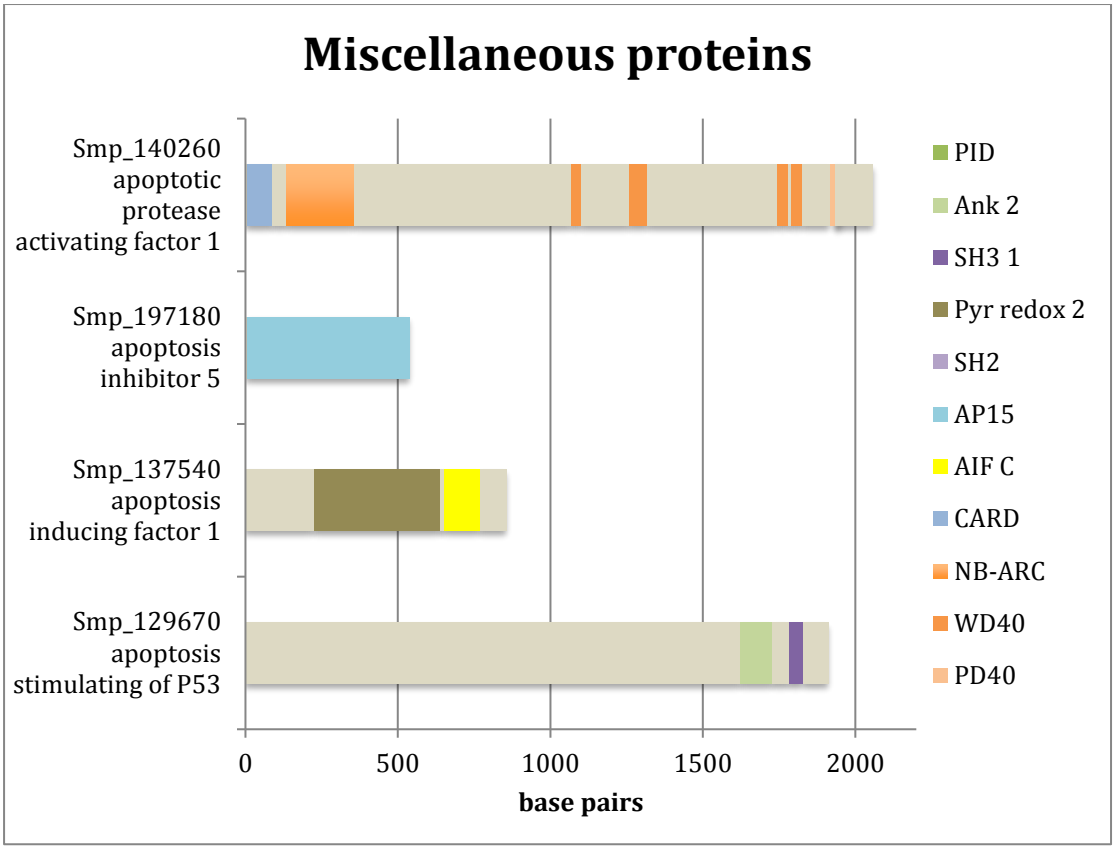


Figure B.10: Diagram of miscellaneous apoptosis related proteins, showing length and presence of PFAM domains.

Gene ID	Average Read Count	Fold change	Adjusted p-value	Product
Smp_107750	11120.5	4.764	0.033	Uncharacterised protein
Smp_108010	7839.4	4.696	0.033	Uncharacterised protein
Smp_116960	7850.2	4.665	0.033	Uncharacterised protein
Smp_100450	7759.5	4.605	0.033	Uncharacterised protein
Smp_074390	7340.6	4.242	0.033	Eukaryotic translation initiation factor
Smp_145370	6960.2	3.540	0.033	Uncharacterised protein
Smp_079770	41164.0	1.642	0.033	Probable protein disulfide-isomerase ER-60
Smp_049270	24812.2	0.267	0.045	Major egg antigen (p40)
Smp_131110	177674.7	4.205	0.053	Uncharacterised protein
Smp_018930	7063.7	4.011	0.053	Uncharacterised protein
Smp_023840	12609.0	3.758	0.053	Uncharacterised protein
Smp_202690	15523.4	3.599	0.053	Universal stress protein
Smp_006040	11467.1	3.563	0.053	Uncharacterised protein
Smp_054160	135853.5	3.328	0.053	Glutathione S-transferase class-mu 28 kDa isozyme
Smp_158110	9316.0	2.257	0.053	Thioredoxin peroxidase
Smp_900040	112222.7	3.631	0.064	NADH dehydrogenase subunit 2
Smp_090520	9271.9	3.434	0.064	Purine nucleoside phosphorylase
Smp_900100	57478.3	3.603	0.065	NADH dehydrogenase subunit 3
Smp_043120	41091.7	3.179	0.065	Universal stress protein
Smp_900050	227094.7	3.260	0.070	NADH dehydrogenase subunit 5
Smp_095980	30236.6	3.838	0.070	Extracellular superoxide dismutase (Cu Zn)
Smp_155310	8225.5	3.760	0.078	Tetraspanin CD63 receptor
Smp_058690	20056.6	3.022	0.088	Glutathione peroxidase
Smp_000430	75350.6	3.607	0.089	Eggshell protein
Smp_014610	82078.6	3.373	0.089	Serine:threonine kinase 1
Smp_138570	9976.6	3.364	0.089	Spore germination protein
Smp_000390	9269.4	3.361	0.089	Uncharacterised protein
Smp_191910	66486.1	3.244	0.089	Stress protein DDR48
Smp_000420	22883.9	3.184	0.089	Pro His rich protein
Smp_000280	24250.8	3.163	0.089	Uncharacterised protein
Smp_173150	7196.8	3.111	0.089	CD63 antigen
Smp_087760	15931.7	3.059	0.089	Ferritin 1 heavy chain
Smp_080920	48931.4	3.050	0.089	Uncharacterised protein
Smp_008360	21077.1	3.020	0.089	Adenosine kinase
Smp_033610	8093.4	3.007	0.089	NADH dehydrogenase (ubiquinone) 1 beta
Smp_202770	15844.8	3.002	0.089	Uncharacterised protein
Smp_071000	9542.5	2.977	0.089	26 proteasome complex subunit DSS1
Smp_016490	8570.6	2.970	0.089	Saposin B domain containing protein
Smp_900020	123222.9	2.967	0.089	NADH dehydrogenase subunit 6
Smp_077890	11865.3	2.966	0.089	Trematode Eggshell Synthesis domain containing protein
Smp_036940	8638.4	2.944	0.089	Protein transport protein Sec61 subunit
Smp_135740	8509.8	2.918	0.089	Splicing factor 3B subunit 5
Smp_900000	316106.4	2.876	0.089	Cytochrome c oxidase subunit I
Smp_066590	7573.6	2.875	0.089	LYR motif containing protein 4

Smp_105450	25570.2	2.857	0.089	Saposin containing protein
Smp_158910	6850.8	2.835	0.089	Cyclin dependent protein kinase CDC28 regulatory
Smp_028190	9228.7	2.816	0.089	Methylthioadenosine phosphorylase
Smp_053390	32861.4	2.813	0.089	Histone H4
Smp_033400	21383.4	2.755	0.089	Uncharacterised protein
Smp_079270	8197.8	2.741	0.089	LIM, zinc binding domain containing protein
Smp_079230	21402.5	2.717	0.089	Immunophilin FK506 binding protein FKBP12, putative
Smp_004940	26544.0	2.708	0.089	60S ribosomal protein L35
Smp_036220	19780.6	2.707	0.089	Histone H2B
Smp_900110	82393.2	2.706	0.089	NADH dehydrogenase subunit 1
Smp_019060	10298.8	2.699	0.089	Sec61 beta subunit
Smp_179460	9344.1	2.683	0.089	Uncharacterised protein
Smp_007900	32378.3	2.676	0.089	Large subunit ribosomal protein 23
Smp_029820	35411.7	2.645	0.089	Ribosomal protein, large P2
Smp_205570	15821.5	2.640	0.089	Small nuclear ribonucleoprotein e
Smp_050280	25043.2	2.637	0.089	Ribosomal protein L31
Smp_066990	31353.8	2.631	0.089	40S rRNA protein homolog
Smp_146190	20760.2	2.627	0.089	40S ribosomal protein S21
Smp_035790	32243.6	2.623	0.089	60S ribosomal protein L44
Smp_074780	24348.9	2.614	0.089	Ribosomal protein S18
Smp_089570	17532.3	2.586	0.089	Stress associated endoplasmic reticulum protein
Smp_075560	12705.4	2.583	0.089	Small nuclear ribonucleoprotein Sm D2
Smp_059930	6598.7	2.568	0.089	Cytochrome b c1 complex subunit 6
Smp_035800	23949.4	2.566	0.089	60S ribosomal protein L37
Smp_041650	18343.7	2.563	0.089	Ribosomal protein S27
Smp_091640	28194.6	2.553	0.089	Ribosomal protein S24
Smp_031310	25869.9	2.553	0.089	Ribosomal protein, Small subunit
Smp_014650	20471.8	2.544	0.089	Small subunit ribosomal protein s30e
Smp_210630	8450.9	2.535	0.089	histone H2A
Smp_032060	8096.4	2.532	0.089	14 kDA subunit splicing factor 3b
Smp_039400	44213.9	2.518	0.089	Eukaryotic translation initiation factor 5A
Smp_089430	30223.8	2.518	0.089	Ubiquitin (ribosomal protein L40)
Smp_113620	14145.9	2.500	0.089	Splicing factor, arginine:serine rich 2
Smp_176200	13488.6	2.484	0.089	Superoxide dismutase [Cu-Zn]
Smp_058700	8579.6	2.480	0.089	Glutathione peroxidase
Smp_001830	27123.9	2.456	0.089	Ribosomal protein L24
Smp_087550	11928.4	2.452	0.089	Uncharacterised protein
Smp_022560	15102.5	2.451	0.089	60S ribosomal protein L21
Smp_900090	167743.1	2.444	0.089	NADH dehydrogenase subunit 4
Smp_050940	24957.3	2.442	0.089	60S ribosomal protein L11
Smp_063350	23711.0	2.428	0.089	Ribosomal protein L27
Smp_054780	45354.7	2.410	0.089	40S ribosomal protein S8
Smp_060910	7628.8	2.406	0.089	Small nuclear ribonucleoprotein f
Smp_074470	26940.0	2.406	0.089	Ribosomal protein, Small subunit
Smp_019440	8101.0	2.405	0.089	Protein translation factor sui1
Smp_097380	15244.8	2.378	0.089	Heat shock 10 kDa protein 1

Smp_151810	6755.1	2.340	0.089	Potassium voltage gated channel subfamily H
Smp_068420	7552.8	2.287	0.089	Small nuclear ribonucleoprotein sm d3
Smp_210310	18054.2	2.265	0.089	Ribosomal protein l7a
Smp_055210	9659.2	2.221	0.089	Microsomal glutathione S transferase 3
Smp_018020	6429.4	2.211	0.089	NADH dehydrogenase (ubiquinone) iron sulfur
Smp_020770	7586.2	2.095	0.089	Dolichyl diphosphooligosaccharide protein
Smp_074000	36900.7	1.742	0.089	Uncharacterised protein
Smp_049550	33633.6	1.454	0.089	78 kDa glucose regulated protein
Smp_054560	8523.3	2.363	0.089	Huntingtin interacting protein K
Smp_175740	35702.1	2.351	0.091	Ribosomal protein L14
Smp_061200	27528.7	2.469	0.092	Ribosomal protein L35A
Smp_130100	7798.6	2.830	0.092	Saposin containing protein
Smp_105320	16857.7	2.532	0.092	60S ribosomal protein L37a
Smp_213130	31795.4	2.436	0.092	Ribosomal protein S25
Smp_044580	24381.9	2.423	0.092	Ribosomal protein L30
Smp_098960	35400.1	2.408	0.092	60S ribosomal protein L26
Smp_032760	16458.3	2.369	0.092	Ribosomal protein S11
Smp_012750	39845.5	2.348	0.092	60S ribosomal protein L12
Smp_053830	29481.0	2.347	0.092	40S ribosomal protein S14
Smp_098330	6356.7	2.229	0.092	H:ACA ribonucleoprotein complex subunit 2
Smp_054240	35052.9	2.217	0.092	Translationally-controlled tumour protein homolog
Smp_082240	19731.3	2.210	0.092	Histone H3
Smp_002180	11152.9	2.068	0.092	Basic transcription factor 3 4
Smp_042400	6909.0	1.929	0.092	Arginine rich, mutated in early stage tumours
Smp_030690	16512.0	2.426	0.093	Eukaryotic translation elongation factor 1 beta
Smp_174950	28022.6	2.367	0.093	Ribosomal protein S19e
Smp_027610	32304.2	2.335	0.093	Ribosomal protein S3
Smp_022640	39721.0	2.334	0.093	60S ribosomal protein L13
Smp_036400	9180.3	2.199	0.093	NADH dehydrogenase (ubiquinone) 1 beta
Smp_027880	8201.2	2.153	0.093	Prefoldin subunit 6
Smp_138210	10112.2	2.691	0.094	Uncharacterised protein
Smp_055050	24033.0	2.372	0.094	40S ribosomal protein S17
Smp_211250	25308.8	2.343	0.094	Ribosomal protein L34a
Smp_085300	16087.7	2.260	0.094	Ribosomal protein L22
Smp_098890	6376.0	2.120	0.094	Arginine:serine rich splicing factor
Smp_079430	29778.5	1.970	0.094	GTP binding nuclear protein Ran
Smp_900030	29615.6	2.412	0.095	ATP synthase F0 subunit 6
Smp_076740	21127.7	2.344	0.095	Ribosomal protein saa; ribosomal protein rpsa
Smp_096750	14596.0	2.338	0.095	40S ribosomal protein S13
Smp_018990	24725.4	2.313	0.095	Ribosomal protein L9
Smp_194840	41848.5	2.263	0.095	Niemann Pick C2 protein
Smp_119920	21711.9	2.259	0.095	Ribosomal protein S16
Smp_037700	8014.7	2.254	0.095	Nuclear transport factor 2
Smp_066890	23966.5	2.120	0.095	40S ribosomal protein S10
Smp_055900	8220.4	2.101	0.095	Leptin receptor overlapping transcript
Smp_179650	9467.6	2.008	0.095	Chromobox protein

Smp_066940	10628.5	2.290	0.096	60S ribosomal protein L29
Smp_011570	27137.1	2.138	0.096	Ribosomal protein S4
Smp_092380	6836.2	2.042	0.096	Protein bud31
Smp_046020	10706.6	1.941	0.097	Protein phosphatase 1 regulatory subunit 7
Smp_179300	14257.1	1.915	0.097	Cellular nuclear acid binding protein
Smp_058240	6942.3	1.902	0.098	Reticulon 4 (Neurite outgrowth inhibitor)
Smp_082030	15560.6	2.169	0.098	Uncharacterised protein
Smp_041500	7421.3	2.265	0.099	Mitogen activated protein kinase
Smp_130850	43065.9	2.257	0.099	40S ribosomal protein S15
Smp_008070	27008.4	2.221	0.099	Thioredoxin
Smp_089290	15876.3	2.286	0.100	Alpha galactosidase:alpha n
Smp_132300	32382.5	2.225	0.100	Kinesin protein KIF3A like
Smp_024850	26420.3	2.176	0.100	60S ribosomal protein L17
Smp_038560	20226.0	2.159	0.100	Prothymosin alpha B

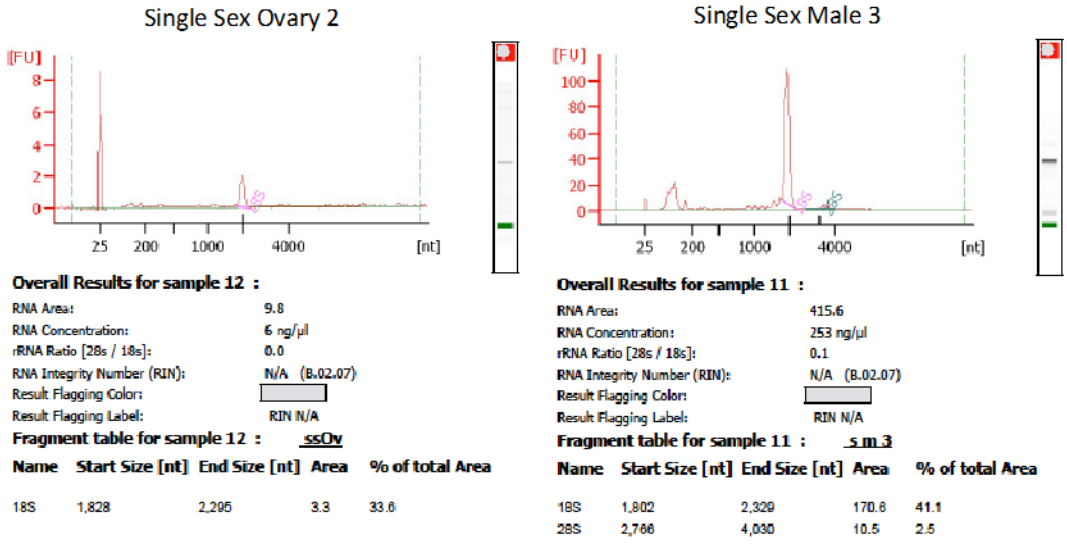
Appendix B11: All 150 genes found to be differentially expressed between paired and separated females 8 days after perfusion. Genes are ranked by adjusted p-value, from smallest to largest. Average read count provides a measure of absolute expression. The fold change is expression in separated females, relative to paired females. The *p*-value has been adjusted for multiple hypothesis testing.

Appendix C

Sample name	M1 MS	M2 MS	M3 MS	M1 SS	M2 SS	M3 SS	T1 MS	T2 MS	T3 MS	T1 SS	T2 SS	T3 SS	Average
Total reads	82,157,621	61,413,861	58,786,566	63,319,818	63,888,404	54,258,554	43,373,743	43,840,647	65,099,133	44,595,566	56,767,832	68,875,916	63,005,559
Mapped	71,801,485	51,649,806	51,179,327	53,852,252	53,738,964	47,564,797	38,119,993	32,048,217	48,212,772	38,876,508	49,722,004	56,457,897	53,160,332
Prop. Paired	57,615,006	42,911,100	42,452,604	45,111,588	44,763,160	38,200,040	30,926,682	26,063,698	24,158,538	29,982,900	40,360,354	46,779,174	41,746,213
Singletons	5,011,317	2,509,444	2,580,515	2,701,856	2,691,940	2,435,665	2,511,925	2,199,055	2,332,190	2,376,020	3,168,100	2,740,915	3,165,597
Unmapped	10,356,136	9,764,055	7,607,239	9,467,566	10,149,440	6,693,757	5,253,750	11,792,430	16,886,361	5,719,058	7,045,828	12,418,019	9,845,227
% Unmapped	12.60%	15.90%	12.90%	15.00%	15.90%	12.30%	12.10%	26.90%	25.90%	12.80%	12.40%	18.00%	15.70%
Mapped to feature	53,899,496	39,306,965	38,169,249	39,427,831	39,494,229	34,877,418	31,264,617	28,735,955	28,719,005	27,747,160	42,043,347	42,473,097	40,888,434
% Mapped to feature	65.60%	64.00%	64.90%	62.30%	61.80%	64.30%	72.10%	65.50%	44.10%	62.20%	74.10%	61.70%	64%

Sample name	F1 MS	F2 MS	F3 MS	F1 SS	F2 SS	F3 SS	O1 MS	O2 MS	O3 MS	O1 SS	O2 SS	O3 SS	Average
Total reads	42,049,886	56,720,483	48,461,371	62,724,794	111,348,638	80,824,593	60,987,961	83,483,600	73,599,576	62,449,538	68,107,686	54,997,636	63,005,559
Mapped	36,204,716	45,482,737	39,592,784	54,360,068	94,437,658	61,321,111	54,988,691	74,676,180	66,211,423	53,494,629	58,533,720	43,320,228	53,160,332
Prop. Paired	19,131,978	33,138,056	28,397,394	45,920,862	78,762,590	52,127,848	29,669,282	63,194,578	55,763,652	43,577,282	47,044,524	35,856,214	41,746,213
Singletons	2,683,370	4,739,835	4,281,282	2,718,634	4,987,700	2,868,193	2,960,611	4,419,900	3,971,051	3,214,019	3,714,416	2,156,378	3,165,597
Unmapped	5,845,170	11,237,746	8,868,587	8,364,726	16,910,980	19,503,482	5,999,270	8,807,420	7,388,153	8,954,909	9,573,966	11,677,408	9,845,227
% Unmapped	13.90%	19.80%	18.30%	13.30%	15.20%	24.10%	9.80%	10.50%	10.00%	14.30%	14.10%	21.20%	15.70%
Mapped to feature	23,079,604	33,100,676	28,860,010	41,880,335	74,704,667	50,759,202	33,648,187	63,218,280	56,989,452	43,844,868	49,958,368	35,120,404	40,888,434
% Mapped to feature	54.90%	58.40%	59.60%	66.80%	67.10%	62.80%	55.20%	75.70%	77.40%	70.20%	73.40%	63.90%	64%

Appendix C.1: Summary of sequencing statistics for sequnced RNA-Seq samples. Samples are named as follows: Sample type (Males – M; Female – F; Testes – T; Ovaries – O), replicate number (1 to 3) and infection status (Mixed sex infections – MS; Single sex infection – SS). Total reads is the number of all reads sequenced. Mapped is the total number of reads mapped to the *Schistosoma mansoni* reference genome v5.2. 'Prop. Paired' refers to the total number of reads mapped together with a mate-paired read in close proximity and in the correct orientation. 'Singletons' refers to the total number of reads with a correctly mapped mate but themselves unmapped. Unmapped is the total number of reads that could not be mapped to the reference by Tophat2. Mapped to feature is the total number of reads mapped to exons of annotated genes in the reference, excluding introns as well as 5'- and 3'-UTRs.



Appendix C.2: Bioanalyzer data shows that ovary and whole male worm samples form single sex infections had good quality total RNA. RNA was extracted from whole worms and isolated organs and the quality examined as described in Chapter 2.2.3. All samples showed a single, sharp ribosomal RNA peak and no signs of degradation.

Gene ID	Description	Average Gene Count	Fold Change	Adjusted p-value
Smp_173940	Frizzled	636.07	1.08	4.38E-01
Smp_175590	FGFR-A	977.81	0.16	9.34E-70
Smp_157300	FGFR-B	1537.22	0.27	2.23E-47
Smp_093700	PMRC1	1974.06	1.46	4.15E-09
Smp_050520	Notch	286.33	0.45	1.29E-06
Smp_157750	Musashi	799.75	0.73	2.66E-04
Smp_161930	Actin	2205.63	0.12	6.33E-14
Smp_156670	EIF4E	8964.66	0.92	3.27E-01
Smp_069770	NDUFV2	3069.44	1.13	1.33E-01
Smp_000740	PSMD4	7314.60	1.02	7.34E-01
Smp_047200	RPL3	46765.34	2.19	3.09E-32
Smp_091740	TPC2L	1194.72	1.30	8.45E-03
Smp_090120	TUBA	40164.97	0.85	5.61E-03
Smp_023160	ATPF	1261.13	1.08	3.68E-01
Smp_056970	GAPDH	28090.64	1.96	8.25E-45
Smp_003770	HID	21443.85	0.42	1.42E-44
Smp_007630	PABP2	3694.28	2.14	2.76E-36
Smp_002180	BTF3L4	9920.43	2.15	1.14E-54

Appendix C.3: List of genes examined in Chapter 5.2.2. Expression data is provided for the comparison of ovaries from females from single sex (SS) infections compared to the ovaries from mixed sex (MS) infections. Average read counts provide a measure of absolute expression for a given gene. The fold changes reflect expression in ovaries from SS females relative to ovaries of MS females. The *p*-value has been adjusted for multiple hypothesis testing.

GO term	Description	Total Genes	DEGs	Expected	topGO
GO:0007155	Cell adhesion	226	128	55.92	2.50E-14
GO:0007156	Homophilic cell adhesion	60	40	14.85	6.90E-12
GO:0007186	G-protein coupled receptor signalling pathway	161	73	39.84	2.50E-09
GO:0006811	Ion transport	389	167	96.25	2.00E-05
GO:0007165	Signal transduction	581	242	143.76	3.90E-05
GO:0006810	Transport	1184	391	292.96	0.00011
GO:0007264	Small GTPase mediated signal transduction	168	65	41.57	0.00012
GO:0006813	Potassium ion transport	77	34	19.05	0.00014
GO:0006817	Phosphate ion transport	19	12	4.7	0.00043
GO:0006897	Endocytosis	52	24	12.87	0.0006
GO:0006865	Amino acid transport	22	13	5.44	0.00061
GO:0006120	Mitochondrial electron transport	9	7	2.23	0.00124
GO:0007229	Integrin-mediated signalling pathway	9	7	2.23	0.00124
GO:0007017	Microtubule-based process	101	24	24.99	0.00196
GO:0008203	Cholesterol metabolic process	18	9	4.45	0.00257

GO:0006814	Sodium ion transport	64	26	15.84	0.00355
GO:0007160	Cell-matrix adhesion	8	6	1.98	0.00397
GO:0050790	Regulation of catalytic activity	60	18	14.85	0.0044
GO:0006629	Lipid metabolic process	241	83	59.63	0.00476
GO:0006508	Proteolysis	318	85	78.68	0.0055
GO:0006493	Protein O-linked glycosylation	12	7	2.97	0.00941
GO:0051216	Cartilage development	5	4	1.24	0.01499
GO:0046165	Alcohol biosynthetic process	13	6	3.22	0.01511
GO:0031032	Actomyosin structure organization	3	3	0.74	0.01512
GO:0045596	Negative regulation of cell differentiation	3	3	0.74	0.01512
GO:0030514	Negative regulation of BMP signalling pathway	3	3	0.74	0.01512
GO:0030301	Cholesterol transport	3	3	0.74	0.01512
GO:0006855	Drug transmembrane transport	3	3	0.74	0.01512
GO:0015904	tetracycline transport	8	5	1.98	0.02604
GO:0007169	Transmembrane receptor protein tyrosine signalling	31	13	7.67	0.02627
GO:0022904	Respiratory electron transport chain	19	12	4.7	0.03574
GO:0006665	Sphingolipid metabolic process	9	6	2.23	0.03606
GO:0007223	Wnt signalling pathway, calcium modulating	6	4	1.48	0.03618
GO:0035023	Regulation of Rho protein signal transduction	23	10	5.69	0.03787
GO:0042632	Cholesterol homeostasis	4	3	0.99	0.04929
GO:0001503	Ossification	4	3	0.99	0.04929
GO:0006518	Peptide metabolic process	13	5	3.22	0.04931

Appendix C.4: Expanded list of GO terms enriched amongst genes expressed more strongly in whole male, compared to testes. The column "Total Genes" provides the number of genes associated with a particular GO term. "DEGs" provides the number of genes associated with a given GO term that had significantly higher expression in testes; "Expected Genes" provides the number of differentially expressed genes associated with a given GO term that are expected by chance.

Pathway	Description	Total Genes	DEGs	Expected	p-value
smm04142	Lysosome	47	37	14.25	5.41E-12
smm04144	Endocytosis	64	38	19.41	7.10E-07
smm00564	Glycerophospholipid metabolism	35	21	10.61	1.74E-04
smm00500	Starch and sucrose metabolism	14	11	4.25	2.35E-04
smm00010	Glycolysis / Gluconeogenesis	27	17	8.19	3.28E-04
smm00565	Ether lipid metabolism	9	8	2.73	4.37E-04
smm04145	Phagosome	38	21	11.52	7.45E-04
smm00561	Glycerolipid metabolism	17	11	5.16	2.74E-03
smm00520	Amino sugar and nucleotide sugar metabolism	24	14	7.28	2.83E-03
smm04070	Phosphatidylinositol signalling system	32	17	9.70	3.71E-03
smm04512	ECM-receptor interaction	9	7	2.73	4.06E-03
smm04310	Wnt signalling pathway	35	18	10.61	4.37E-03
smm00562	Inositol phosphate metabolism	26	14	7.88	6.81E-03
smm00052	Galactose metabolism	14	9	4.25	7.00E-03
smm04141	Protein processing in endoplasmic reticulum	89	36	26.99	1.04E-02
smm00511	Other glycan degradation	8	6	2.43	1.05E-02
smm04068	FoxO signalling pathway	38	18	11.52	1.11E-02
smm04080	Neuroactive ligand-receptor interaction	15	9	4.55	1.22E-02
smm00330	Arginine and proline metabolism	18	10	5.46	1.57E-02
smm00620	Pyruvate metabolism	21	11	6.37	1.87E-02
smm04340	Hedgehog signalling pathway	9	6	2.73	2.19E-02
smm00020	Citrate cycle (TCA cycle)	25	12	7.58	2.83E-02
smm00340	Histidine metabolism	5	4	1.52	2.93E-02
smm00190	Oxidative phosphorylation	69	26	20.93	4.18E-02
smm04320	Dorso-ventral axis formation	8	5	2.43	4.84E-02

Appendix C.5: Complete list of KEGG pathways enriched amongst the genes up regulated the wholes males compared to testes. The column "Total Genes" provides the number of genes in the KEGG pathway. "DEGs" provides the number differentially expressed genes (DEGs) in that pathway; "Expected Genes" provides the number of DEGs in a given pathway that is expected by chance.

Gene ID	Description	Average read count	Fold Change	Adjusted p-value
Smp_202140	Uncharacterised protein	14.69	149.01	2.68E-06
Smp_184400	Uncharacterised protein	34.97	126.87	8.23E-12
Smp_130450	RNA-binding protein bicaudal C	500.41	111.34	1.26E-105
Smp_015190	Uncharacterised protein	10050.53	108.74	1.43E-105
Smp_205500	Uncharacterised protein	10.22	106.35	2.19E-05
Smp_193990	Uncharacterised protein	18.90	101.49	1.49E-07
Smp_138560	Uncharacterised protein	178.30	89.95	3.31E-42
Smp_194950	Putative ELAV-like RNA-binding protein	4441.96	87.64	1.93E-109
Smp_145730	Putative EGF-domain containing protein	575.65	79.13	5.22E-68

Smp_135030	Uncharacterised protein	97.10	78.96	3.48E-24
Smp_154260	Putative glioma pathogenesis-related protein	35.79	78.74	1.72E-13
Smp_125190	Uncharacterised protein	1628.78	77.56	7.55E-110
Smp_166570	Zinc finger protein	229.83	73.38	5.82E-50
Smp_152310	Uncharacterised protein	527.06	68.84	1.19E-92
Smp_202660	Uncharacterised protein	6.07	67.51	2.51E-04
Smp_134110	Voltage-gated hydrogen channel	12.18	67.26	3.98E-06
Smp_185360	Uncharacterised protein	14.76	65.70	8.63E-06
Smp_161840	Uncharacterised protein	5577.64	64.96	7.27E-106
Smp_027290	Uncharacterised protein	1467.26	64.13	2.82E-58
Smp_142240	Uncharacterised protein	266.80	63.42	2.37E-64
Smp_062490	Putative heart- and neural crest derivatives-expressed protein	421.89	60.19	1.14E-95
Smp_182990	Uncharacterised protein	5.84	59.98	4.56E-04
Smp_123970	Uncharacterised protein	103.54	59.58	1.21E-26
Smp_185830	Uncharacterised protein	18.29	59.28	1.07E-06
Smp_205410	Uncharacterised protein	7080.46	59.11	1.59E-94
Smp_128140	Putative ATP-dependant DNA helicase	94.14	57.69	9.08E-31
Smp_203270	Twist-related protein	11.16	57.14	2.55E-05
Smp_168440	Uncharacterised protein	16.07	56.10	4.87E-07
Smp_061400	Uncharacterised protein	20.61	55.82	1.46E-08
Smp_161060	Uncharacterised protein	1305.71	55.33	3.75E-83
Smp_155350	Putative coiled-coil domain containing protein	482.40	54.92	9.27E-69
Smp_131240	Uncharacterised protein	222.27	54.09	2.28E-32
Smp_205470	Uncharacterised protein	28.75	52.86	9.76E-08
Smp_054680	Archipelago-like protein	28.81	52.24	1.55E-11
Smp_203840	Uncharacterised protein	19.99	52.04	1.96E-07
Smp_143940	Cytoplasmic dynein-1 light intermediate chain	66.98	50.90	1.30E-17
Smp_142220	APOBEC1 complementation factor-like RNA-binding protein	1999.21	50.66	1.15E-96
Smp_028130	Uncharacterised protein	425.61	50.49	4.11E-54
Smp_123300	Follistatin	99.62	109.80	4.29E-11

Appendix C.6: Expanded list of DEGs with testes-biased expression compared to whole males. Average read counts are a measure of absolute gene expression; they are the average of normalised read counts across all RNA-Seq samples used in Chapter 5. The fold changes reflect expression in the testes relative to whole male worms. The *p*-value has been adjusted for multiple hypothesis testing.

GO term	Description	Total Genes	DEGs	Expected	topGO
GO:0007018	Microtubule-based movement	60	36	9.32	3.40E-15
GO:0006941	Striated muscle contraction	145	59	22.52	1.20E-13
GO:0006270	DNA replication initiation	15	13	2.33	2.20E-09
GO:0007283	Spermatogenesis	10	9	1.55	4.40E-07
GO:0006281	DNA repair	122	46	18.95	7.00E-07
GO:0007049	Cell cycle	207	78	32.15	2.40E-06
GO:0007067	Mitotic nuclear division	88	37	13.67	3.50E-06
GO:0006260	DNA replication	161	64	25.01	7.50E-05
GO:0007126	Meiotic nuclear division	8	6	1.24	2.90E-04
GO:0006298	Mismatch repair	11	7	1.71	4.00E-04
GO:0030030	Cell projection organization	9	5	1.4	5.80E-04
GO:0000226	Microtubule cytoskeleton organization	17	7	2.64	1.65E-03
GO:0008285	Negative regulation of cell proliferation	5	4	0.78	2.54E-03
GO:0007064	Mitotic sister chromatid cohesion	3	3	0.47	3.74E-03
GO:0007076	Mitotic chromosome condensation	9	5	1.4	6.52E-03
GO:0006278	RNA-dependent DNA replication	32	11	4.97	6.67E-03
GO:0006183	GTP biosynthetic process	6	4	0.93	6.67E-03
GO:0006269	DNA replication, synthesis of RNA primer	6	4	0.93	6.67E-03
GO:0006228	UTP biosynthetic process	6	4	0.93	6.67E-03
GO:0006241	CTP biosynthetic process	6	4	0.93	6.67E-03
GO:0006486	Protein glycosylation	58	15	9.01	8.97E-03
GO:0007601	Visual perception	10	5	1.55	0.011
GO:0006310	DNA recombination	39	12	6.06	0.012
GO:0009809	Lignin biosynthetic process	4	3	0.62	0.013
GO:0006268	DNA unwinding involved in DNA replication	11	5	1.71	0.018
GO:0042026	Protein refolding	8	4	1.24	0.024
GO:0042592	Homeostatic process	58	9	9.01	0.024
GO:0016568	Chromatin modification	111	24	17.24	0.028
GO:0007059	Chromosome segregation	18	12	2.8	0.029
GO:0006955	Immune response	11	5	1.71	0.029
GO:0018298	Protein-chromophore linkage	9	4	1.4	0.038
GO:0006261	DNA-dependent DNA replication	42	24	6.52	0.050

Appendix C.7: Expanded list of GO terms enriched amongst the testes genes. The column "Total Genes" provides the number of genes associated with a particular GO term. "DEGs" provides the number of genes associated with a given GO term that had significantly higher expression in testes; "Expected Genes" provides the number of differentially expressed genes associated with a given GO term that are expected by chance.

Pathway	Description	Total Genes	DEGs	Expected	p-value
smm03030	DNA replication	30	24	7.00	5.80E-11
smm03430	Mismatch repair	18	16	4.20	5.93E-09
smm03440	Homologous recombination	18	15	4.20	1.07E-07
smm03010	Ribosome	110	47	25.68	1.88E-06
smm03420	Nucleotide excision repair	32	19	7.47	9.28E-06
smm03460	Fanconi anaemia pathway	27	17	6.30	9.39E-06
smm03410	Base excision repair	20	12	4.67	3.71E-04
smm00230	Purine metabolism	82	30	19.14	2.15E-03
smm00240	Pyrimidine metabolism	59	23	13.77	2.70E-03
smm03008	Ribosome biogenesis in eukaryotes	62	22	14.47	9.86E-03
smm03040	Spliceosome	105	33	24.51	0.013
smm00970	Aminoacyl-tRNA biosynthesis	34	13	7.94	0.021

Appendix C.8: Complete list of KEGG pathways enriched amongst the testes expressed genes. The column "Total Genes" provides the number of genes in the KEGG pathway. "DEGs" provides the number differentially expressed genes in that pathway; "Expected Genes" provides the number of differentially expressed genes in a given pathway that is expected by chance.

Domain	Description	Total Genes	DEGs	Expected	p-value
PF00271	Helicase conserved C-terminal domain	75	38	17.68	2.03E-07
PF00493	MCM2/3/5 family	10	10	2.36	5.20E-07
PF00270	DEAD/DEAH box helicase	55	27	12.97	2.19E-05
PF00225	Kinesin motor domain	19	13	4.48	3.67E-05
PF08393	Dynein heavy chain, N-terminal region 2	13	10	3.06	6.67E-05
PF05186	Dpy-30 motif	6	6	1.41	1.71E-04
PF14551	MCM N-terminal domain	6	6	1.41	1.71E-04
PF13414	TPR repeat	40	20	9.43	1.73E-04
PF00852	Glycosyltransferase family 10 (fucosyltransferase)	12	9	2.83	2.18E-04
PF03028	Dynein heavy chain and region D6 of dynein motor	12	9	2.83	2.18E-04
PF12774	Hydrolytic ATP binding site of dynein motor region D1	12	9	2.83	2.18E-04
PF12775	P-loop containing dynein motor region D3	12	9	2.83	2.18E-04
PF12777	Microtubule-binding stalk of dynein motor	12	9	2.83	2.18E-04
PF12780	P-loop containing dynein motor region D4	12	9	2.83	2.18E-04
PF12781	ATP-binding dynein motor region D5	12	9	2.83	2.18E-04
PF00533	BRCA1 C Terminus (BRCT) domain	9	7	2.12	8.45E-04
PF00929	Exonuclease	7	6	1.65	9.14E-04
PF00077	Retroviral aspartyl protease	14	9	3.30	1.16E-03
PF12799	Leucine Rich repeats (2 copies)	20	11	4.72	1.85E-03
PF00400	WD domain, G-beta repeat	141	47	33.24	2.15E-03
PF07728	AAA domain (dynein-related subfamily)	10	7	2.36	2.15E-03
PF02493	MORN repeat	8	6	1.89	2.79E-03
PF00488	MutS domain V	4	4	0.94	3.08E-03
PF06220	U1 zinc finger	4	4	0.94	3.08E-03
PF08542	Replication factor C C-terminal domain	4	4	0.94	3.08E-03
PF00612	IQ calmodulin-binding motif	9	6	2.12	6.41E-03
PF03133	Tubulin-tyrosine ligase family	9	6	2.12	6.41E-03
PF03148	Tektin family	7	5	1.65	8.91E-03
PF00078	Reverse transcriptase (RNA-dependent DNA polymerase)	24	11	5.66	9.40E-03
PF00334	Nucleoside diphosphate kinase	5	4	1.18	0.012
PF01798	Putative snoRNA binding domain	5	4	1.18	0.012
PF02463	RecF/RecN/SMC N terminal domain	5	4	1.18	0.012
PF13181	Tetratricopeptide repeat	5	4	1.18	0.012
PF00385	Chromo (Chromatin Organisation Modifier) domain	10	6	2.36	0.012

Appendix C.9: Expanded list of Pfam domains enriched amongst the testes expressed genes. The column "Total Genes" provides the number of genes with a given Pfam domain. "DEGs" provides the number of differentially expressed genes with that domain; "Expected" provides the number of differentially expressed genes encoding a given domain that is expected by chance.

Pathway	Description	Total Genes	DEGs	Expected	p-value
smm04142	Lysosome	47	39	13.59	9.79E-15
smm04141	Protein processing in endoplasmic reticulum	89	48	25.74	2.65E-07
smm04145	Phagosome	38	26	10.99	3.48E-07
smm04144	Endocytosis	64	35	18.51	7.26E-06
smm00520	Amino sugar and nucleotide sugar metabolism	24	17	6.94	1.98E-05
smm00510	N-Glycan biosynthesis	30	19	8.68	6.65E-05
smm00052	Galactose metabolism	14	11	4.05	1.48E-04
smm00500	Starch and sucrose metabolism	14	11	4.05	1.48E-04
smm00010	Glycolysis / Gluconeogenesis	27	16	7.81	6.85E-04
smm00511	Other glycan degradation	8	7	2.31	9.43E-04
smm00600	Sphingolipid metabolism	15	10	4.34	2.17E-03
smm04512	ECM-receptor interaction	9	7	2.60	3.03E-03
smm00051	Fructose and mannose metabolism	14	9	4.05	5.03E-03
smm04130	SNARE interactions in vesicular transport	14	9	4.05	5.03E-03
smm04320	Dorso-ventral axis formation	8	6	2.31	8.18E-03
smm00562	Inositol phosphate metabolism	26	13	7.52	1.19E-02
smm00564	Glycerophospholipid metabolism	35	16	10.12	1.44E-02
smm03060	Protein export	19	10	5.50	1.72E-02
smm04340	Hedgehog signalling pathway	9	6	2.60	1.75E-02
smm00250	Alanine, aspartate and glutamate metabolism	14	8	4.05	1.87E-02
smm04140	Regulation of autophagy	14	8	4.05	1.87E-02
smm00604	Glycosphingolipid biosynthesis - ganglio series	3	3	0.87	2.41E-02
smm00740	Riboflavin metabolism	5	4	1.45	2.47E-02
smm04070	Phosphatidylinositol signalling system	32	14	9.25	2.85E-02
smm04080	Neuroactive ligand-receptor interaction	15	8	4.34	2.86E-02
smm04068	FoxO signalling pathway	38	16	10.99	2.87E-02
smm00020	Citrate cycle (TCA cycle)	25	11	7.23	4.40E-02

Appendix C.10: Complete list of KEGG pathways enriched amongst the genes up-regulated in whole females compared to ovaries from mixed sex infections. The column "Total Genes" provides the number of genes in the KEGG pathway. "DEGs" provides the number differentially expressed genes in that pathway; "Expected Genes" provides the number of differentially expressed in a given pathway that is expected by chance.

Gene ID	Description	Average Gene Count	Fold Change	Adjusted p-value
Smp_054430	Scratch-family zinc finger protein	37.84	46.88	1.66E-10
Smp_136900	Homeobox protein DLX	9.77	36.34	5.04E-05
Smp_198590	Kinesin-like protein	54.56	30.95	4.24E-14
Smp_032970	Calmodium-like protein	15.05	29.95	3.45E-06
Smp_126070	Uncharacterised protein	45.96	28.29	1.16E-12
Smp_204040	Uncharacterised protein	14.31	28.25	4.59E-06
Smp_082410	Uncharacterised protein	64.87	27.36	1.36E-14
Smp_134490	Thyroid hormone receptor	265.72	26.97	4.95E-48
Smp_143820	Uncharacterised protein	6.90	26.82	3.24E-04
Smp_154860	Zinc finger CW-type protein	358.40	26.30	3.17E-42
Smp_196000	Aquaporin	20.93	26.14	2.11E-07
Smp_147880	Rootletin	41.89	24.34	2.60E-11
Smp_078590	Uncharacterised protein	5.91	22.28	9.25E-04
Smp_032990	Calmodulin-like protein	24.19	21.25	6.62E-07
Smp_145230	T-box transcription factor	108.28	19.67	1.41E-14
Smp_145570	Uncharacterised protein	418.42	19.32	9.13E-60
Smp_036660	Uncharacterised protein	2385.46	18.72	1.29E-76
Smp_193990	Uncharacterised protein	5.18	18.41	2.47E-03
Smp_043430	Uncharacterised protein	788.64	17.21	6.06E-89
Smp_160590	Uncharacterised protein	2797.25	16.88	1.09E-100
Smp_203840	Uncharacterised protein	5.71	16.71	4.25E-03
Smp_161050	Uncharacterised protein	187.69	16.44	2.26E-41
Smp_152490	E3 ubiquitin-protein ligase	301.44	16.37	3.48E-58
Smp_165230	Location of vulva defective-like protein	65.10	16.10	1.78E-18
Smp_144860	Boule-like protein	1998.21	16.06	1.34E-147
Smp_145280	NACHT and WD repeat domain-containing protein	223.85	15.80	1.71E-42
Smp_172150	Putative ATP-dependant RNA helicase	975.82	15.78	1.29E-136
Smp_170960	Uncharacterised protein	709.57	15.77	2.35E-117
Smp_141570	TWiK family potassium channel	192.15	15.34	6.45E-44
Smp_145580	Uncharacterised protein	1351.25	15.31	6.99E-58
Smp_131630	Testes expressed protein-like serine/threonine-protein kinase	1524.23	15.04	7.49E-162
Smp_132660	Rho guanine nucleotide exchange factor (RhoGEF)	2296.39	14.95	1.42E-79
Smp_165430	Sperm-tail PG-rich repeat-containing protein	81.11	14.82	2.12E-21
Smp_185360	Uncharacterised protein	3.88	14.69	6.27E-03
Smp_162740	Meiosis-specific with OB domain-containing protein	190.11	14.44	2.21E-24
Smp_129750	α-(1,3)-fucosyltransferase	3.70	14.35	6.92E-03

Appendix C.11: Expanded list of differentially expressed ovary-genes compared to whole females from mixed sex infections. Average read counts provide a measure of absolute expression across all RNA-Seq samples used in Chapter 5 for a given gene. The fold changes reflect expression in ovaries relative to whole females. The *p*-value has been adjusted for multiple hypothesis testing.

GO term	Description	Total Genes	DEGs	Expected	p-value
GO:0006412	Translation	393	83	41.42	2.00E-09
GO:0042254	Ribosome biogenesis	137	31	14.44	5.90E-06
GO:0006414	Translational elongation	27	10	2.85	4.60E-04
GO:0006839	Mitochondrial transport	16	7	1.69	1.11E-03
GO:0016246	RNA interference	3	3	0.32	1.17E-03
GO:0006465	Signal peptide processing	4	3	0.42	4.29E-03
GO:0007160	Cell-matrix adhesion	8	4	0.84	6.04E-03
GO:0006446	Regulation of translational initiation	5	3	0.53	9.89E-03
GO:0007229	Integrin-mediated signalling pathway	9	4	0.95	9.98E-03
GO:0006836	Neurotransmitter transport	30	8	3.16	0.010
GO:0005975	carbohydrate metabolic process	246	30	25.92	0.011
GO:0006875	Cellular metal ion homeostasis	10	3	1.05	0.011
GO:0006413	Translational initiation	37	11	3.9	0.015
GO:0006401	RNA catabolic process	59	9	6.22	0.018
GO:0019321	Pentose metabolic process	6	3	0.63	0.018
GO:0006364	rRNA processing	108	19	11.38	0.023
GO:0008033	tRNA processing	61	10	6.43	0.026
GO:0006098	Pentose-phosphate shunt	7	3	0.74	0.029
GO:0006637	Acyl-CoA metabolic process	3	2	0.32	0.031
GO:0006855	Drug transmembrane transport	3	2	0.32	0.031
GO:0008612	Peptidyl-lysine modification to peptidyl-hypusine	3	2	0.32	0.031
GO:0006633	Fatty acid biosynthetic process	51	10	5.37	0.037
GO:0042026	Protein refolding	8	3	0.84	0.044
GO:0006821	Chloride transport	8	3	0.84	0.044
GO:0006534	Cysteine metabolic process	8	3	0.84	0.044

Appendix C.12: Expanded list of GO terms enriched amongst the genes up-regulated in ovaries of females from mixed sex infections compared to whole females. The column "Total Genes" provides the number of genes associated with a particular GO term. "DEGs" provides the number of genes associated with a given GO term that had significantly higher expression in ovaries of females from mixed sex infections; "Expected Genes" provides the number of differentially expressed genes associated with a given GO term that are expected by chance.

Pathway	Description	Total Genes	DEGs	Expected	p-value
smm03008	Ribosome biogenesis in eukaryotes	62	48	20.60	5.08E-13
smm03030	DNA replication	30	25	9.97	1.65E-08
smm03013	RNA transport	103	60	34.22	5.56E-08
smm03010	Ribosome	110	62	36.55	1.71E-07
smm03460	Fanconi anaemia pathway	27	21	8.97	2.04E-06
smm03430	Mismatch repair	18	15	5.98	1.50E-05
smm00970	Aminoacyl-tRNA biosynthesis	34	23	11.30	2.92E-05
smm03420	Nucleotide excision repair	32	22	10.63	2.99E-05
smm03040	Spliceosome	105	53	34.89	7.50E-05
smm00240	Pyrimidine metabolism	59	33	19.60	1.49E-04
smm03440	Homologous recombination	18	13	5.98	6.55E-04
smm03020	RNA polymerase	24	15	7.97	2.20E-03
smm03018	RNA degradation	38	21	12.63	2.52E-03
smm03410	Base excision repair	20	12	6.65	8.81E-03
smm00230	Purine metabolism	82	34	27.25	0.026
smm03450	Non-homologous end-joining	5	4	1.66	0.041

Appendix C.13: Complete list of KEGG pathways enriched amongst the genes upregulated in the ovaries of females from mixed sex infections compared to whole females. The column "Total Genes" provides the number of genes in the KEGG pathway. "DEGs" provides the number differentially expressed genes in that pathway; "Expected Genes" provides the number of differentially expressed genes in a given pathway that is expected by chance.

Domain	Description	Total Genes	DEGs	Expected	p-value
PF00271	Helicase conserved C-terminal domain	75	58	22.29	1.36E-17
PF00270	DEAD/DEAH box helicase	55	43	16.35	1.13E-13
PF00493	MCM2/3/5 family	10	10	2.97	5.31E-06
PF00533	BRCA1 C Terminus (BRCT) domain	9	9	2.68	1.79E-05
PF00176	SNF2 family N-terminal domain	19	14	5.65	8.21E-05
PF00400	WD domain, G-beta repeat	141	62	41.91	1.02E-04
PF00856	SET domain	19	13	5.65	4.55E-04
PF00481	Protein phosphatase 2C	6	6	1.78	6.87E-04
PF14551	MCM N-terminal domain	6	6	1.78	6.87E-04
PF00852	Glycosyltransferase family 10 (fucosyltransferase)	12	9	3.57	1.37E-03
PF00679	Elongation factor G C-terminus	5	5	1.49	2.31E-03
PF00752	XPG N-terminal domain	5	5	1.49	2.31E-03
PF02463	RecF/RecN/SMC N terminal domain	5	5	1.49	2.31E-03
PF08389	Exportin 1-like protein	5	5	1.49	2.31E-03
PF13589	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase	5	5	1.49	2.31E-03
PF00929	Exonuclease	7	6	2.08	3.38E-03
PF01926	50S ribosome-binding GTPase	11	8	3.27	3.47E-03

PF03144	Elongation factor Tu domain 2	11	8	3.27	3.47E-03
PF00076	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	84	35	24.97	5.80E-03
PF14259	RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	29	15	8.62	6.87E-03
PF00136	DNA polymerase family B	4	4	1.19	7.79E-03
PF00383	Cytidine and deoxycytidylate deaminase zinc-binding region	4	4	1.19	7.79E-03
PF00488	MutS domain V	4	4	1.19	7.79E-03
PF00867	XPG I-region	4	4	1.19	7.79E-03
PF02984	Cyclin, C-terminal domain	4	4	1.19	7.79E-03
PF03104	DNA polymerase family B, exonuclease domain	4	4	1.19	7.79E-03
PF06470	SMC proteins Flexible Hinge Domain	4	4	1.19	7.79E-03
PF08542	Replication factor C C-terminal domain	4	4	1.19	7.79E-03
PF13959	Domain of unknown function (DUF4217)	4	4	1.19	7.79E-03
PF01336	OB-fold nucleic acid binding domain	8	6	2.38	9.51E-03
PF01398	JAB1/Mov34/MPN/PAD-1 ubiquitin protease	8	6	2.38	9.51E-03
PF00133	tRNA synthetases class I (I, L, M and V)	6	5	1.78	9.76E-03
PF05383	La domain	6	5	1.78	9.76E-03
PF08264	Anticodon-binding domain of tRNA	6	5	1.78	9.76E-03

Appendix C.14: Expanded list of Pfam domains enriched amongst the genes up-regulated in the ovaries of females from mixed sex infections compared to whole females. The column "Total Genes" provides the number of genes encoding a given Pfam domain. "DEGs" provides the number of differentially expressed genes with that domain; "Expected" provides the number of differentially expressed genes with a given domain that is expected by chance.

GO ID	Term	Annotated	Significant	Expected	p-value
GO:0007018	Microtubule-based movement	60	24	5.52	1.20E-10
GO:0006941	Striated muscle contraction	145	32	13.33	1.80E-06
GO:0007283	Spermatogenesis	10	7	0.92	5.00E-06
GO:0030030	Cell projection organization	9	5	0.83	7.00E-05
GO:0006183	GTP biosynthetic process	6	4	0.55	9.10E-04
GO:0006228	UTP biosynthetic process	6	4	0.55	9.10E-04
GO:0006241	CTP biosynthetic process	6	4	0.55	9.10E-04
GO:0000226	Microtubule cytoskeleton organization	17	6	1.56	1.09E-03
GO:0009809	Lignin biosynthetic process	4	3	0.37	2.88E-03
GO:0006955	Immune response	11	4	1.01	6.69E-03
GO:0007275	Multicellular organismal development	286	37	26.29	8.21E-03
GO:0007602	Phototransduction	6	3	0.55	1.25E-02
GO:0006278	RNA-dependent DNA replication	32	7	2.94	2.35E-02
GO:0042026	Protein refolding	8	3	0.74	3.04E-02
GO:0006334	Nucleosome assembly	34	7	3.13	3.21E-02
GO:0015074	DNA integration	28	6	2.57	3.85E-02
GO:0018298	Protein-chromophore linkage	9	3	0.83	4.26E-02
GO:0006108	Malate metabolic process	4	2	0.37	4.46E-02
GO:0009073	Aromatic amino acid family biosynthetic ...	4	2	0.37	4.46E-02
GO:0006563	L-serine metabolic process	4	2	0.37	4.46E-02

Appendix C.15: Expanded list of GO terms enriched amongst the genes up-regulated in testes compared to ovaries. The column "Total Genes" provides the number of genes associated with a particular GO term. "DEGs" provides the number of genes associated with a given GO term that had significantly higher expression in testes compared to ovaries; "Expected Genes" provides the number of differentially expressed genes associated with a given GO term that are expected by chance.

GO ID	Term	Annotated	Significant	Expected	p-value
GO:0006364	rRNA processing	108	18	6.14	3.10E-05
GO:0042254	ribosome biogenesis	137	26	7.79	1.10E-04
GO:0009086	methionine biosynthetic process	5	3	0.28	1.67E-03
GO:0006468	protein phosphorylation	337	30	19.16	8.79E-03
GO:0001682	tRNA 5'-leader removal	3	2	0.17	9.30E-03
GO:0006996	organelle organization	328	32	18.65	1.07E-02
GO:0007420	brain development	4	2	0.23	1.79E-02
GO:0006338	chromatin remodelling	4	2	0.23	1.79E-02
GO:0006333	chromatin assembly or disassembly	63	7	3.58	2.21E-02
GO:0007275	multicellular organismal development	286	26	16.26	3.23E-02
GO:0007219	Notch signalling pathway	22	4	1.25	3.33E-02
GO:0051028	mRNA transport	84	10	4.78	3.37E-02
GO:0007050	cell cycle arrest	6	2	0.34	4.15E-02
GO:0016485	protein processing	15	3	0.85	4.98E-02

Appendix C.16: Expanded list of GO terms enriched in genes up-regulated in ovaries compared to testes. The column "Total Genes" provides the number of genes associated with a particular GO term. "DEGs" provides the number of genes associated with a given GO term that had significantly higher expression in ovaries compared to the testes; "Expected Genes" provides the number of differentially expressed genes associated with a given GO term that are expected by chance.

Gene ID	Description	Average Gene Count	Fold Change	Adjusted p-value
Smp_075370	Uncharacterised protein	2546.97	281.45	3.20E-161
Smp_123830	Collagen alpha (xi) chain	12391.26	247.02	1.23E-204
Smp_214190	Calpain	7468.55	223.17	1.28E-194
Smp_161790	Uncharacterised protein	2177.89	170.65	8.69E-168
Smp_130280	Neurogenic locus notch protein	2542.52	165.81	2.54E-157
Smp_084090	Uncharacterised protein	2213.81	158.02	1.07E-155
Smp_018250	Troponin i	12126.92	129.72	2.78E-215
Smp_045200	Tegument-allergen-like protein	7170.98	107.84	6.62E-286
Smp_144910	Collagen type I:II:III:V:XI alpha	9835.41	105.69	2.09E-187
Smp_006860	PDZ and LIM domain protein Zasp	3301.43	72.02	1.59E-193
Smp_074560	Uncharacterised protein	2562.52	70.69	9.58E-160
Smp_144280	Tensin	3376.96	67.30	4.93E-190
Smp_156960	Nardilysin (M16 family)	4146.65	66.18	1.13E-187
Smp_020070	Uncharacterised protein	6841.37	65.49	4.86E-219
Smp_045550	Annexin	3514.23	49.80	5.55E-182
Smp_196250	Supervillin	4252.86	47.39	2.39E-194
Smp_179810	Oncosphere protein tso22e	17872.05	44.39	0
Smp_085540	Myosin heavy chain, putative	77964.61	43.45	1.65E-185
Smp_086330	Myophilin	5101.28	37.76	2.37E-184
Smp_171780	SPARC protein	2760.33	31.76	7.19E-189
Smp_049270	Major egg antigen (p40)	6759.56	31.12	2.09E-227
Smp_037230	Fimbrin	7826.62	24.88	2.47E-215
Smp_087250	Muscle LIM protein	2504.41	24.05	9.54E-158
Smp_044010	Tropomyosin	17835.33	17.14	2.26E-159
Smp_123300	Follistatin	99.62	9.67	4.29E-11

Appendix C.17: Genes up-regulated in whole males from mixed sex infections compared to testes. Average read counts are a measure of absolute expression across all samples, and fold changes provided a measure pf gene expression in whole males relative to testes. The *p*-value has been adjusted for multiple hypothesis testing.

Comparison	Fold Change	Adjusted p-value
Expression in ovaries from single sex infections compared to ovaries from mixed sex infections	9.07	8.58E-12
Expression in females from single sex infections compared to females from mixed sex infections	12.34	9.46E-31

Appendix C.18: Expression of the gene encoding follistatin (Smp_122300) in female worms and their ovaries. The first column specifies the samples in which expression of Smp_122300 was compared. The fold change provides a measure of gene expression in the first sample compared to the second. Smp_122300 expression was up regulated in the single sex sample in both cases. The *p*-value has been adjusted for multiple hypothesis testing.

GO term	Description	Total Genes	DEGs	Expected	p-value
GO:0006281	DNA repair	122	45	14.67	2.50E-10
GO:0006270	DNA replication initiation	15	12	1.8	2.70E-09
GO:0007049	Cell cycle	207	71	24.89	8.70E-08
GO:0006260	DNA replication	161	57	19.36	1.50E-07
GO:0007067	Mitotic nuclear division	88	33	10.58	1.30E-06
GO:0007126	Meiotic nuclear division	8	6	0.96	6.70E-05
GO:0006941	Striated muscle contraction	145	34	17.44	7.50E-05
GO:0006364	rRNA processing	108	26	12.99	3.30E-04
GO:0006298	Mismatch repair	11	6	1.32	8.00E-04
GO:0007059	Chromosome segregation	18	12	2.16	9.00E-04
GO:0009086	Methionine biosynthetic process	5	4	0.6	9.40E-04
GO:0042254	Ribosome biogenesis	137	36	16.48	1.19E-03
GO:0051276	Chromosome organization	168	39	20.2	1.61E-03
GO:0007064	Mitotic sister chromatid cohesion	3	3	0.36	1.73E-03
GO:0007018	Microtubule-based movement	60	15	7.22	4.00E-03
GO:0006268	DNA unwinding involved in DNA replication	11	5	1.32	6.15E-03
GO:0008285	Negative regulation of cell proliferation	5	3	0.6	0.014
GO:0006547	Histidine metabolic process	6	3	0.72	0.014
GO:0000184	Nuclear-transcribed mRNA catabolic process	39	10	4.69	0.015
GO:0006310	DNA recombination	39	10	4.69	0.015
GO:0042592	Homeostatic process	58	5	6.98	0.015
GO:0016568	Chromatin modification	111	23	13.35	0.015
GO:0007076	Mitotic chromosome condensation	9	4	1.08	0.016
GO:0006333	Chromatin assembly or disassembly	63	12	7.58	0.018
GO:0006261	DNA-dependent DNA replication	42	22	5.05	0.018
GO:0006269	DNA replication, synthesis of RNA primer	6	3	0.72	0.026
GO:0030212	Hyaluronan metabolic process	6	3	0.72	0.026
GO:0006435	Threonyl-tRNA aminoacylation	6	3	0.72	0.026
GO:0008283	Cell proliferation	11	6	1.32	0.040
GO:0031570	DNA integrity checkpoint	3	2	0.36	0.040
GO:0001682	tRNA 5'-leader removal	3	2	0.36	0.040
GO:0006529	Asparagine biosynthetic process	3	2	0.36	0.040
GO:0006486	Protein glycosylation	58	11	6.98	0.043

Appendix C.19: Expanded list of GO terms enriched amongst the gene up regulated in the ovaries of females from mixed sex infections compared to those from single sex infections. The column "Total Genes" provides the number of genes associated with a particular GO term. "DEGs" provides the number of genes associated with a given GO term that had significantly higher expression in ovaries; "Expected Genes" provides the number of differentially expressed genes associated with a given GO term that are expected by chance.

Gene ID	Description
Smp_130980	Cyclin k
Smp_163380	Cyclin dependent kinase 2
Smp_076560	Cyclin t1
Smp_124130	Cyclin dependent kinase 1
Smp_073340	Cyclin dependent kinase 5
Smp_152760	Cyclin protein FAM58A
Smp_137370	Cyclin dependent kinase 16
Smp_090980	Cyclin dependent kinase 20
Smp_172050	Cyclin y
Smp_103950	Cyclin dependent kinase 10
Smp_172700	Cyclin dependent kinase 6
Smp_127600	Cyclin y
Smp_176620	Cyclin dependent kinase 11
Smp_129360	Cyclin G2
Smp_150040	Cyclin dependent kinase 7
Smp_153920	Cyclin G2
Smp_003000	Cyclin dependent kinase 9
Smp_041820	Cyclin dependent kinase 16
Smp_102650	Cyclin dependent kinase
Smp_157240	Cyclin dependent kinase 1
Smp_156990	Cyclin dependent kinase 19
Smp_080730	Cyclin dependent kinase 1
Smp_094760	Cyclin dependent kinase 9
Smp_047190	Cyclin g associated kinase
Smp_039670	Cyclin L1
Smp_212930	Cyclin h
Smp_155330	Cyclin dependent kinase 14
Smp_167410	Cell division cycle protein 123
Smp_004060	Cell division cycle 20
Smp_134020	Cell division cycle protein 27
Smp_070690	Cell division cycle 16
Smp_208070	CDC16 cell division cycle 16
Smp_007510	Cell division cycle 2 protein kinase 5
Smp_177120	Cell division cycle 7 protein
Smp_129320	Cell division cycle 5 protein
Smp_193440	Cell division cycle and apoptosis regulator
Smp_132020	Cell division cycle 20 protein 1 cofactor
Smp_037400	Cell division cycle protein 23

Appendix C.20: Description of cyclins other cell cycle-related genes.

Gene ID	Description
Smp_161920	Actin
Smp_175590	FGFR-A
Smp_157300	FGFR-B
Smp_142050	ERK1
Smp_144590	Ras-GAP
Smp_104110	Rho
Smp_137610	FAK
Smp_161230	Ras-GEF/Sos
Smp_035190	Ras-GEF/Sos
Smp_125360	Csk
Smp_003230	Grb2
Smp_068720	Diaphanous
Smp_176990	Raf
Smp_199010	Ras-GEF/Sos
Smp_046600	Actin
Smp_177870	Ras-GEF/Sos
Smp_183710	Actin
Smp_080120	TβRII
Smp_073470	Nuclear receptor
Smp_153500	Tyrosine Kinase
Smp_062950	Growth factor receptor-bound protein
Smp_179910	Ras
Smp_173940	Frizzled
Smp_072660	Ras-GAP
Smp_144390	Type IIb activin receptor (SmRK2)
Smp_164680	Ras-GAP
Smp_047900	ERK2
Smp_009760	14-3-3
Smp_093700	PMRC1
Smp_136300	TK5
Smp_085910	Smad 2
Smp_097730	SRF
Smp_019790	SmRTK1
Smp_049760	TβRI
Smp_079230	SmFKBP12
Smp_168370	Vav2
Smp_033950	Smad 4
Smp_041500	MEK
Smp_196040	Shc
Smp_169020	Ras-GAP
Smp_146810	Diaphanous 2

Appendix C.21: Descriptions of genes of the MAPK signalling pathway.

Pathway	Description	Total Genes	DEGs	Expected	p-value
smm03010	Ribosome	110	91	55.90	7.52E-13
smm03008	Ribosome biogenesis in eukaryotes	62	57	31.51	1.80E-12
smm03050	Proteasome	32	29	16.26	1.52E-06
smm03013	RNA transport	103	73	52.34	1.12E-05
smm00190	Oxidative phosphorylation	69	52	35.06	1.38E-05
smm03020	RNA polymerase	24	20	12.20	7.79E-04
smm03060	Protein export	19	15	9.66	8.63E-03
smm00970	Aminoacyl-tRNA biosynthesis	34	23	17.28	0.020
smm04141	Protein processing in endoplasmic reticulum	89	51	45.23	0.040
smm00240	Pyrimidine metabolism	59	35	29.98	0.044

Appendix C.22: Complete list of KEGG pathways enriched amongst the genes up-regulated in ovaries of females mixed sex infections compared to those of single sex infections. The column "Total Genes" provides the number of genes in the KEGG pathway. "DEGs" provides the number differentially expressed genes in that pathway; "Expected Genes" provides the number of differentially expressed genes in a given pathway that is expected by chance

Domain	Description	Total Genes	DEGs	Expected	p-value
PF00227	Proteasome subunit	14	14	4.62	1.781E-07
PF00400	WD domain, G-beta repeat	141	72	46.55	3.273E-06
PF00270	DEAD/DEAH box helicase	55	33	18.16	2.375E-05
PF03357	Snf7	9	9	2.97	4.612E-05
PF00179	Ubiquitin-conjugating enzyme	26	18	8.58	1.341E-04
PF01398	JAB1/Mov34/MPN/PAD-1 ubiquitin protease	8	8	2.64	1.400E-04
PF14259	RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)	29	19	9.57	2.547E-04
PF10584	Proteasome subunit A N-terminal signature	7	7	2.31	4.249E-04
PF01399	PCI domain	14	11	4.62	5.494E-04
PF03144	Elongation factor Tu domain 2	11	9	3.63	1.141E-03
PF00009	Elongation factor Tu GTP binding domain	17	12	5.61	1.384E-03
PF00153	Mitochondrial carrier protein	26	16	8.58	1.896E-03
PF00085	Thioredoxin	16	11	5.28	2.966E-03
PF00467	KOW motif	5	5	1.65	3.910E-03
PF00679	Elongation factor G C-terminus	5	5	1.65	3.910E-03
PF05193	Peptidase M16 inactive domain	5	5	1.65	3.910E-03
PF08389	Exportin 1-like protein	5	5	1.65	3.910E-03
PF00271	Helicase conserved C-terminal domain	75	35	24.76	4.477E-03
PF00753	Metallo-beta-lactamase superfamily	7	6	2.31	6.050E-03
PF13893	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	13	9	4.29	6.669E-03
PF00118	TCP-1/cpn60 chaperonin family	11	8	3.63	6.965E-03
PF01926	50S ribosome-binding GTPase	11	8	3.63	6.965E-03
PF00076	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	84	37	27.73	9.403E-03
PF00137	ATP synthase subunit C	4	4	1.32	0.012
PF00383	Cytidine and deoxycytidylate deaminase zinc-binding region	4	4	1.32	0.012
PF00579	tRNA synthetases class I (W and Y)	4	4	1.32	0.012
PF00613	Phosphoinositide 3-kinase family, accessory domain (PIK domain)	4	4	1.32	0.012
PF01237	Oxysterol-binding protein	4	4	1.32	0.012
PF03129	Anticodon binding domain	4	4	1.32	0.012
PF01423	LSM domain	14	9	4.62	0.013
PF00443	Ubiquitin carboxyl-terminal hydrolase	21	12	6.93	0.013
PF00133	tRNA synthetases class I (I, L, M and V)	6	5	1.98	0.016
PF08264	Anticodon-binding domain of tRNA	6	5	1.98	0.016
PF00675	Insulinase (Peptidase family M16)	8	6	2.64	0.016

Appendix C.23: Expanded list of Pfam domains enriched amongst the genes up-regulated in ovaries of females from mixed sex infections compared to those of single sex infections. The column "Total Genes" provides the number of genes encoding a given Pfam domain. "DEGs" provides the number of differentially expressed genes with that domain; "Expected" provides the number of differentially expressed genes with a given domain that is expected by chance.

Gene ID	Description	Fold Change	Adjusted p-value
Smp_210790	Transmembrane BAX inhibitor motif containing	1.18	1.05E-01
Smp_136730	Cathepsin D	1.66	2.73E-02
Smp_179800	p21 activated protein kinase 1 Dpak1	2.01	1.09E-13
Smp_043360	sBH3-1	4.49	3.77E-04
Smp_207000	Anamorsin homologue	1.84	4.99E-08
Smp_028500	Caspase 3	1.35	2.69E-02
Smp_197180	Apoptosis inhibitor 5	1.51	2.51E-09
Smp_022110	p21 activated protein kinase interacting protein	1.95	1.50E-28
Smp_084610	FAS associated factor 2	2.01	5.71E-29
Smp_196040	Uncharacterised protein	2.67	3.71E-48
Smp_002410	14-3-3 epsilon 2	1.85	6.72E-42
Smp_137540	Apoptosis inducing factor 1, mitochondrial	2.29	5.03E-29
Smp_009760	14-3-3 protein homolog 1	1.45	1.12E-11
Smp_044000	Bax inhibitor 1	1.34	5.66E-08
Smp_172010	Caspase 7	0.95	9.41E-01
Smp_077540	FAS associated factor 1	1.20	2.46E-02
Smp_032000	Caspase	0.97	8.35E-01
Smp_148130	FAS binding factor	0.32	1.23E-39
Smp_213250	Bcl 2 ous antagonist:killer	0.85	6.12E-03
Smp_213730	FAS death associated protein	0.83	5.69E-03
Smp_013040	Cathepsin D	0.77	9.99E-08
Smp_034840	14-3-3 epsilon	0.81	1.08E-04
Smp_072180	Apoptosis regulator BAX	0.70	1.87E-05
Smp_141270	Caspase 8	0.82	6.50E-03
Smp_095190	Bcl 2 ous antagonist:killer	0.53	2.40E-05
Smp_129670	Apoptosis stimulating of P53	0.97	7.51E-01
Smp_041630	BCL2 apoptosis inhibitor domain containing protein	0.86	8.99E-02
Smp_168470	BCL2 apoptosis inhibitor domain containing protein	0.76	9.36E-03

Appendix C.24: Expression of putative apoptosis related genes in the ovaries of females from single sex infections, compared to ovaries from mixed sex infections. The fold change provides a measure of expression in the ovaries of ovaries from single sex infections compared to those from mixed sex infections (That means fold changes greater than 1 indicate an up regulation in the ovaries from single sex infections). The *p*-value has been adjusted for multiple hypothesis testing.

GO term	Description	Total Genes	DEGs	Expected	p-value
GO:0007155	Cell adhesion	226	91	34.46	3.80E-17
GO:0006278	RNA-dependent DNA replication	32	19	4.88	1.20E-08
GO:0030154	Cell differentiation	78	29	11.89	2.00E-06
GO:0006941	Striated muscle contraction	145	44	22.11	2.40E-06
GO:0007156	Homophilic cell adhesion	60	22	9.15	3.60E-05
GO:0007165	Signal transduction	581	128	88.6	4.00E-05
GO:0007169	Transmembrane receptor protein tyrosine kinase signalling pathway	31	14	4.73	6.90E-05
GO:0006813	Potassium ion transport	77	24	11.74	3.00E-04
GO:0006811	Ion transport	389	103	59.32	3.30E-04
GO:0015074	DNA integration	28	12	4.27	4.20E-04
GO:0006814	Sodium ion transport	64	20	9.76	9.20E-04
GO:0006816	Calcium ion transport	49	16	7.47	1.75E-03
GO:0007018	Microtubule-based movement	60	18	9.15	2.71E-03
GO:0006537	Glutamate biosynthetic process	3	3	0.46	3.54E-03
GO:0006508	Proteolysis	318	58	48.49	4.19E-03
GO:0007017	Microtubule-based process	101	32	15.4	5.89E-03
GO:0007283	Spermatogenesis	10	5	1.52	0.011
GO:0007601	visual perception	10	5	1.52	0.011
GO:0009073	Aromatic amino acid family biosynthetic process	4	3	0.61	0.013
GO:0009809	Lignin biosynthetic process	4	3	0.61	0.013
GO:0006879	Cellular iron ion homeostasis	7	4	1.07	0.013
GO:0006826	Iron ion transport	11	5	1.68	0.017
GO:0042278	Purine nucleoside metabolic process	128	21	19.52	0.023
GO:0007517	Muscle organ development	12	3	1.83	0.023
GO:0009060	Aerobic respiration	20	3	3.05	0.023
GO:0009154	Purine ribonucleotide catabolic process	44	5	6.71	0.023
GO:0007275	Multicellular organismal development	286	62	43.61	0.025
GO:0009165	Nucleotide biosynthetic process	115	18	17.54	0.026
GO:0006013	Mannose metabolic process	5	3	0.76	0.028
GO:0016055	Wnt signalling pathway	27	9	4.12	0.031
GO:0016042	Lipid catabolic process	26	8	3.96	0.031
GO:0006754	ATP biosynthetic process	50	8	7.62	0.032
GO:0006120	Mitochondrial electron transport, NADH to ubiquinone	9	4	1.37	0.036

Appendix C.25: Expanded list of GO terms enriched amongst the genes up-regulated in ovaries of female single sex infections compared to those from mixed sex infections. The column "Total Genes" provides the number of genes associated with a particular GO term. "DEGs" provides the number of genes associated with a given GO term that had significantly higher expression in ovaries of SS females; "Expected Genes" provides the number of differentially expressed genes associated with a given GO term that are expected by chance.

Pathway	Description	Total Genes	DEGs	Expected	p-value
smm04142	Lysosome	47	20	7.41	6.70E-06
smm00330	Arginine and proline metabolism	18	9	2.84	5.94E-04
smm00760	Nicotinate and nicotinamide metabolism	10	6	1.58	1.59E-03
smm00511	Other glycan degradation	8	5	1.26	3.21E-03
smm04512	ECM-receptor interaction	9	5	1.42	6.10E-03
smm00052	Galactose metabolism	14	6	2.21	0.012
smm00600	Sphingolipid metabolism	15	6	2.37	0.016
smm02010	ABC transporters	8	4	1.26	0.022
smm04320	Dorso-ventral axis formation	8	4	1.26	0.022
smm04310	Wnt signalling pathway	35	10	5.52	0.023
smm00561	Glycerolipid metabolism	17	6	2.68	0.029
smm00500	Starch and sucrose metabolism	14	5	2.21	0.041
smm00062	Fatty acid elongation	6	3	0.95	0.047
smm00564	Glycerophospholipid metabolism	35	9	5.52	0.049

Appendix C.26: Complete list of KEGG pathways enriched amongst the genes up-regulated in ovaries of females from single sex infections compared to those of mixed sex infections. The column "Total Genes" provides the number of genes in the KEGG pathway. "DEGs" provides the number differentially expressed genes in that pathway; "Expected Genes" provides the number of differentially expressed genes in a given pathway that is expected by chance.

Domain	Description	Total Genes	DEGs	Expected	p-value
PF00078	Reverse transcriptase (RNA-dependent DNA polymerase)	24	21	6.67	1.480E-09
PF00665	Integrase core domain	13	13	3.61	5.693E-08
PF00041	Fibronectin type III domain	29	22	8.05	8.660E-08
PF00077	Retroviral aspartyl protease	14	13	3.89	5.767E-07
PF13895	Immunoglobulin domain	27	19	7.50	4.238E-06
PF07679	Immunoglobulin I-set domain	45	25	12.50	5.559E-05
PF00011	Hsp20/alpha crystallin family	11	9	3.06	2.792E-04
PF09380	FERM C-terminal PH-like domain	13	10	3.61	2.908E-04
PF13499	EF-hand domain pair	66	31	18.33	3.916E-04
PF07686	Immunoglobulin V-set domain	8	7	2.22	7.318E-04
PF07645	Calcium-binding EGF domain	7	6	1.94	2.310E-03
PF13405	EF-hand domain	12	8	3.33	4.741E-03
PF00452	Apoptosis regulator proteins, Bcl-2 family	4	4	1.11	5.938E-03
PF01413	C-terminal tandem repeated domain in type 4 procollagen	4	4	1.11	5.938E-03
PF11901	Protein of unknown function (DUF3421)	8	6	2.22	6.679E-03
PF00008	EGF-like domain	15	9	4.17	6.944E-03

PF00595	PDZ domain (Also known as DHR or GLGF)	54	23	15.00	7.106E-03
PF01048	Phosphorylase superfamily	6	5	1.67	7.141E-03
PF01392	Fz domain	6	5	1.67	7.141E-03
PF13927	Immunoglobulin domain	6	5	1.67	7.141E-03
PF01221	Dynein light chain type 1	31	15	8.61	7.360E-03
PF00520	Ion transport protein	43	19	11.94	8.619E-03
PF09379	FERM N-terminal domain	16	9	4.44	0.011
PF00373	FERM central domain	19	10	5.28	0.013
PF00617	RasGEF domain	9	6	2.50	0.014
PF03133	Tubulin-tyrosine ligase family	9	6	2.50	0.014
PF07714	Protein tyrosine kinase	46	19	12.78	0.017
PF07690	Major Facilitator Superfamily	43	18	11.94	0.017
PF00754	F5/8 type C domain	7	5	1.94	0.018
PF00501	AMP-binding enzyme	12	7	3.33	0.020
PF01391	Collagen triple helix repeat (20 copies)	12	7	3.33	0.020
PF00036	EF hand	3	3	0.83	0.021
PF00992	Troponin	3	3	0.83	0.021
PF01596	O-methyltransferase	3	3	0.83	0.021
PF03188	Eukaryotic cytochrome b561	3	3	0.83	0.021
PF04434	SWIM zinc finger	3	3	0.83	0.021
PF04712	Radial spokehead-like protein	3	3	0.83	0.021

Appendix C.27: Expanded list of Pfam domains enriched amongst the genes up-regulated in ovaries of females from single sex infections compared to those of mixed sex infections. The column "Total Genes" provides the number of genes encoding a given Pfam domain. "DEGs" provides the number of differentially expressed genes with that domain; "Expected" provides the number of differentially expressed genes with a given domain that is expected by chance.

References

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., *et al.* (2000). The Genome Sequence of Drosophila melanogaster. *Science*, *287*, 2185–2196.

Adams, M. D., Dubnick, M., Kerlavage, A. R., Moreno, R., Kelley, J. M., Utterback, T. R., *et al.* (1992). Sequence identification of 2375 human brain genes. *Nature*, *355*, 632–634.

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., *et al.* (1991). Complementary Sequencing : Expressed Sequence Tags and Human Genome Project, *49*(1990), 1651–1656.

Adams, M. D., Kerlavage, A. R., Fields, C., & Venter, J. C. (1993). 3400 new expressed sequence tags identify diversity of transcripts in human brain. *Natur Genetics*, *4*, 256–267.

Ahmad, G., Zhang, W., Torben, W., Ahrorov, A., Damian, R. T., Wolf, R. F., *et al*. (2011). Preclinical prophylactic efficacy testing of Sm-p80-based vaccine in a nonhuman primate model of schistosoma mansoni infection and immunoglobulin G and e responses to Sm-p80 in human serum samples from an area where schistosomiasis is endemic. *Journal of Infectious Diseases*, *204*(9), 1437–1449. http://doi.org/10.1093/infdis/jir545

Ahmad, G., Zhang, W., Torben, W., Noor, Z., & Siddiqui, A. (2010). Protective effects of Sm-p80 in the presence of resiquimod as an adjuvant against challenge infection with Schistosoma mansoni in mice. *International Journal of Infectious Diseases*, *14*(9), e781–e787. http://doi.org/10.1016/j.ijid.2010.02.2266

Aird, D., Ross, M. G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., *et al.* (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, *12*(2), R18. http://doi.org/10.1186/gb-2011-12-2-r18

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2007). *Molecular Biology of the Cell*. Garland Science.

Alexa, A., Rahnenführer, J., & Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, *22*(13), 1600–1607. http://doi.org/10.1093/bioinformatics/btl140

Anders, S., Pyl, P. T., & Huber, W. (2015). Genome analysis HTSeq — a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. http://doi.org/10.1093/bioinformatics/btu638

Andrade, L. F. De, Mourão, M. D. M., Geraldo, J. A., Coelho, F. S., Silva, L. L., Neves, R. H., *et al.* (2014). Regulation of Schistosoma mansoni Development and Reproduction by the Mitogen-Activated Protein Kinase Signaling Pathway. *PLoS Neglected Tropical Diseases*, *8*(6), e2949. http://doi.org/10.1371/journal.pntd.0002949

Andrade, L. F., Nahum, L. a, Avelar, L. G. a, Silva, L. L., Zerlotini, A., Ruiz, J. C., & Oliveira, G. (2011). Eukaryotic protein kinases (ePKs) of the helminth parasite Schistosoma mansoni. *BMC Genomics*, *12*(1), 215. http://doi.org/10.1186/1471-2164-12-215

Armstrong, J. C. (1965). Mating Behavior and Development of Schistosomes in the Mouse. *The Journal of Parasitology*, *51*(4), 605–616.

Avelar, L. G. a, Nahum, L. a, Andrade, L. F., & Oliveira, G. (2011). Functional Diversity of the Schistosoma mansoni Tyrosine Kinases. *Journal of Signal Transduction*, *2011*, 603290. http://doi.org/10.1155/2011/603290

Baou, M., Jewell, A., & Murphy, J. J. (2009). TIS11 family proteins and their roles in posttranscriptional gene regulation. *Journal of Biomedicine and Biotechnology*, *2009*. http://doi.org/10.1155/2009/634520

Barr, F. a, Silljé, H. H. W., & Nigg, E. a. (2004). Polo-like kinases and the orchestration of cell division. *Nature Reviews. Molecular Cell Biology*, *5*(6), 429–440. http://doi.org/10.1038/nrm1401

Barron, L., & Wynn, T. A. (2011). Macrophage activation governs schistosomiasis-induced inflammation and fibrosis. *European Journal of Immunology*, *41*(9), 2509–2514. http://doi.org/10.1002/eji.201141869

Basch, P. (1991). *Schistosomes*. New York, Oxford: Oxford University Press.

Basch, P. (1988). Schistosoma mansoni: nucleic acid synthesis in immature females from single-sex infections, paired in vitro with intact males and male segments. *Comparative Biochemistry and Physiology. B, Comparative Biochemistry*, *90*(2), 389–392.

Basch, P., & Humbert, R. (1981). Cultivation of Schistosoma mansoni In vitro. III. Implantation of Cultured Worms into Mouse Mesenteric Veins. *The Journal of Parasitology*, *67*(2), 191. http://doi.org/10.2307/3280634

Bashaw, G. J., & Klein, R. (2010). Signaling from axon guidance receptors. *Cold Spring Harbor Perspectives in Biology*, *2*(5), a001941. http://doi.org/10.1101/cshperspect.a001941

Basu, S., Campbell, H. M., Dittel, B. N., & Ray, A. (2010). Purification of Specific Cell Population by Fluorescence Activated Cell Sorting (FACS). *Journal of Visualized Experiments*, (41), 3–6. http://doi.org/10.3791/1546

Beckmann, S., Buro, C., Dissous, C., Hirzmann, J., & Grevelding, C. G. (2010). The Syk Kinase SmTK4 of Schistosoma mansoni Is Involved in the Regulation of Spermatogenesis and Oogenesis. *PLoS Pathogens*, *6*(2), e1000769. http://doi.org/10.1371/journal.ppat.1000769

Beckmann, S., Hahnel, S., Cailliau, K., Vanderstraete, M., Browaeys, E., Dissous, C., & Grevelding, C. G. (2011). Characterization of the Src/Abl hybrid kinase SmTK6 of Schistosoma mansoni. *Journal of Biological Chemistry*, *286*(49), 42325–42336. http://doi.org/10.1074/jbc.M110.210336

Beckmann, S., Quack, T., Burmeister, C., Buro, C., Long, T., Dissous, C., & Grevelding, C. G. (2010). Schistosoma mansoni: signal transduction processes during the development of the reproductive organs. *Parasitology*, *137*(3), 497–520. http://doi.org/10.1017/S0031182010000053

Beckmann, S., Quack, T., Dissous, C., Cailliau, K., Lang, G., & Grevelding, C. G. (2012). Discovery of platyhelminth-specific α/β-integrin families and evidence for their role in reproduction in Schistosoma mansoni. *PloS One*, *7*(12), e52519. http://doi.org/10.1371/journal.pone.0052519

Bedia, C., Camacho, L., Abad, J. L., Fabriàs, G., & Levade, T. (2010). A simple fluorogenic method for determination of acid ceramidase activity and diagnosis of Farber disease. *Journal of Lipid Research*, *51*(12), 3542–3547. http://doi.org/10.1194/jlr.D010033

Belkina, N. V, Liu, Y., Hao, J.-J., Karasuyama, H., & Shaw, S. (2009). LOK is a major ERM kinase in resting lymphocytes and regulates cytoskeletal rearrangement through ERM phosphorylation.

Proceedings of the National Academy of Sciences of the United States of America, *106*(12), 4707–4712. http://doi.org/10.1073/pnas.0805963106

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistics Society*, *57*(1), 289–300.

Berditchevski, F., & Odintsova, E. (2007). Tetraspanins as regulators of protein trafficking. *Traffic*, *8*(2), 89–96. http://doi.org/10.1111/j.1600-0854.2006.00515.x

Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A., Dillon, G. P., Cerqueira, G. C., *et al.* (2009). The genome of the blood fluke Schistosoma mansoni. *Nature*, *460*, 352–8. http://doi.org/10.1038/nature08160

Berry, A., Mone, H., Iriart, X., Mouahid, G., Aboo, O., Boissier, J., *et al.* (2014). Schistosomiasis Haematobium , Corsica , France. *Emerging Infectious Diseases*, *20*(9), 1–3.

Bhardwaj, R., Krautz-Peterson, G., & Skelly, P. J. (2011). Therapeutic Oligonucleotides - Methods and Protocols, 223–239. http://doi.org/10.1007/978-1-61779-188-8

Bhat, K. M. (1999). The posterior determinant gene nanos is required for the maintenance of the adult germline stem cells during drosophila oogenesis. *Genetics*, *151*(4), 1479–1492.

Bickle, Q. D. (2009). Radiation-attenuated schistosome vaccination--a brief historical perspective. *Parasitology*, *136*(12), 1621–1632. http://doi.org/10.1017/S0031182009005848

Bickle, Q. D., Taylor, M. G., Doenhoff, M. J., & Nelson, G. S. (1979). Immunization of mice with gamma-irradiated intramuscularly injected schistosomula of Schistosoma mansoni. *Parasitology*, *79*(2), 209–222. http://doi.org/10.1017/S0031182000053294

Bieberich, E. (2014). Synthesis, Processing, and Function of N-glycans in N-glycoproteins. In R. K. Yu & C.-L. Schengrund (Eds.), (Vol. 9, pp. 47–70). New York, NY: Springer New York. http://doi.org/10.1007/978-1-4939-1154-7_3

Biolchini, C. D. L., Neves, R. H., Hulstijn, M., Gomes, D. C., & Machado-Silva, J. R. (2006). Development of Schistosoma mansoni worms in mice analyzed by bright field and confocal microscopy. *Memorias Do Instituto Oswaldo Cruz*, *101*(SUPPL. 1), 261–265. http://doi.org/10.1590/S0074-02762006000900040

Applied Biosystems (2010). Applied Biosystems StepOne™ and StepOnePlus™. ThermoFisher. Retrieved from https://tools.thermofisher.com/content/sfs/manuals/cms_046736.pdf

KAPA Biosystems (2016). KAPA SYBR® FAST qPCR Kits. Retrieved from http://www.kapabiosystems.co.uk/product-applications/products/qpcr-2/sybr-fast-one-step/

Bolliger, M. F., Zurlinden, a., Luscher, D., Butikofer, L., Shakhova, O., Francolini, M., *et al.* (2010). Specific proteolytic cleavage of agrin regulates maturation of the neuromuscular junction. *Journal of Cell Science*, *123*(22), 3944–3955. http://doi.org/10.1242/jcs.072090

Boppana, N. B., Kodiha, M., Stochaj, U., Lin, H., Haimovitz-Friedman, A., Bielawska, A., *et al.* (2014). Ceramide synthase inhibitor fumonisin B1 inhibits apoptotic cell death in SCC17B human head and neck squamous carcinoma cells after Pc4 photosensitization. *Photochemistry & Photobiological Sciences*, *13*(11), 1621–1627. http://doi.org/10.1039/C4PP00292J

Bornstein, C., Brosh, R., Molchadsky, a., Madar, S., Kogan-Sakin, I., Goldstein, I., *et al.* (2011). SPATA18, a Spermatogenesis-Associated Gene, Is a Novel Transcriptional Target of p53 and p63. *Molecular and Cellular Biology*, *31*(8), 1679–1689. http://doi.org/10.1128/MCB.01072-10

Bose, R., Verheij, M., Haimovitz-Friedman, a, Scotto, K., Fuks, Z., & Kolesnick, R. (1995). Ceramide synthase mediates daunorubicin-induced apoptosis: an alternative mechanism for generating death signals. *Cell*, *82*(3), 405–414. http://doi.org/10.1016/0092-8674(95)90429-8

Bottieau, E., Clerinx, J., Rosario, M., Vega, D., Enden, E. Van Den, Colebunders, R., & Esbroeck, M. Van. (2006). Imported Katayama fever : Clinical and biological features at presentation and during treatment. *Journal of Infection*, *52*, 339–345. http://doi.org/10.1016/j.jinf.2005.07.022

Boureux, A., Vignal, E., Faure, S., & Fort, P. (2007). Evolution of the Rho family of Ras-like GTPases in eukaryotes. *Molecular Biology and Evolution*, *24*(1), 203–216. http://doi.org/10.1093/molbev/msl145

Bourke, C. D., Nausch, N., Rujeni, N., Appleby, L. J., Trottein, F., Midzi, N., *et al.* (2014). Cytokine Responses to the Anti-schistosome Vaccine Candidate Antigen Glutathione-S-transferase Vary with Host Age and Are Boosted by Praziquantel Treatment. *PLoS Neglected Tropical Diseases*, *8*(5). http://doi.org/10.1371/journal.pntd.0002846

Bruno, R. M., Sudano, I., Ghiadoni, L., Masi, L., & Taddei, S. (2011). Interactions between sympathetic nervous system and endogenous endothelin in patients with essential hypertension. *Hypertension*, *57*(1), 79–84. http://doi.org/10.1161/HYPERTENSIONAHA.110.163584

C. elegans_Sequencing Consortium (1998). Genome Sequence of the Nematode *C. elegans* : A Platform for Investigating Biology. *Science*, *282*, 2012–2018.

Cai, P., Liu, S., Piao, X., Hou, N., Gobert, G. N., McManus, D. P., & Chen, Q. (2016). Comprehensive Transcriptome Analysis of Sex-Biased Expressed Genes Reveals Discrete Biological and Physiological Features of Male and Female Schistosoma japonicum. *PLoS Neglected Tropical Diseases*, *10*(4), 1–24. http://doi.org/10.1371/journal.pntd.0004684

Cardoso, R. M. F., Daniels, D. S., Bruns, C. M., & Tainer, J. a. (2003). Characterization of the electrophile binding site and substrate binding mode of the 26-kDa glutathione S-transferase from Schistosoma japonicum. *Proteins*, *51*(1), 137–46. http://doi.org/10.1002/prot.10345

Cass, C. L., Johnson, J. R., Califf, L. L., Xu, T., Hernandez, H. J., Stadecker, M. J., *et al.* (2007). Proteomic analysis of Schistosoma mansoni egg secretions. *Molecular and Biochemical Parasitology*, *155*(2), 84–93. http://doi.org/10.1016/j.molbiopara.2007.06.002

Castro-Borges, W., Simpson, D. M., Dowle, A., Curwen, R. S., Thomas-Oates, J., Beynon, R. J., & Wilson, R. A. (2011). Abundance of tegument surface proteins in the human blood fluke Schistosoma mansoni determined by QconCAT proteomics. *Journal of Proteomics*, *74*(9), 1519–33. http://doi.org/10.1016/j.jprot.2011.06.011

Cavigelli, M., Dolfi, F., Claret, F. X., & Karin, M. (1995). Induction of c-fos expression through JNK-mediated TCF/Elk-1 phosphorylation. *The EMBO Journal*, *14*(23), 5957–5964. http://doi.org/8846788

Center for Disease Control and Prevention. (2012). Schistosomiasis - Biology. Retrieved from: https://www.cdc.gov/parasites/schistosomiasis/biology.html

Charrin, S., Jouannet, S., Boucheix, C., & Rubinstein, E. (2014). Tetraspanins at a glance. *Journal of Cell Science*, *127*(17), 3641–8. http://doi.org/10.1242/jcs.154906

Chase, D., Serafinas, C., Ashcroft, N., Kosinski, M., Longo, D., Ferris, D. K., & Golden, A. (2000). The Polo-like kinase PLK-1 is required for nuclear envelope breakdown and the completion of meiosis in Caenorhabditis elegans. *Genesis*, *26*(1), 26–41. http://doi.org/10.1002/(SICI)1526-968X(200001)26:1<26::AID-GENE6>3.0.CO;2-O

Chen, L. L., Rekosh, D. M., & LoVerde, P. T. (1992). Schistosoma mansoni p48 eggshell protein gene: characterization, developmentally regulated expression and comparison to the p14 eggshell protein gene. *Molecular and Biochemical Parasitology*, *52*(1), 39–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1625706

Christensen, B. M., Li, J., Chen, C., & Nappi, A. J. (2005). Melanization immune responses in mosquito vectors. *Trends in Parasitology*, *21*(4), 192–199. http://doi.org/10.1016/j.pt.2005.02.007

Chu, G. W. T. C., & Cutress, C. E. (1954). Austrobilharzia variglandis (Miller and Northup, 1926) Penner, 1953, (Trematoda: Schistosomatidae) in Hawaii with Notes on Its Biology. *Journal of Parasitology*, *40*(5), 515–524.

Cogswell, A. a, Kommer, V. P., & Williams, D. L. (2012). Transcriptional analysis of a unique set of genes involved in Schistosoma mansoni female reproductive biology. *PLoS Neglected Tropical Diseases*, *6*(11), e1907. http://doi.org/10.1371/journal.pntd.0001907

Colley, D. G., & Secor, W. E. (2014). Immunology of human schistosomiasis. *Parasite Immunology*, *36*(8), 347–357. http://doi.org/10.1111/pim.12087

Collins, J. J., Hou, X., Romanova, E. V., Lambrus, B. G., Miller, C. M., Saberi, A., *et al.* (2010). Genome-Wide Analyses Reveal a Role for Peptide Hormones in Planarian Germline Development. *PLoS Biology*, *8*(10), e1000509. http://doi.org/10.1371/journal.pbio.1000509

Collins, J. J., & Newmark, P. a. (2013). It's no fluke: the planarian as a model for understanding schistosomes. *PLoS Pathogens*, *9*(7), e1003396. http://doi.org/10.1371/journal.ppat.1003396

Collins, J. J., Wang, B., Lambrus, B. G., Tharp, M. E., Iyer, H., & Newmark, P. a. (2013). Adult somatic stem cells in the human parasite Schistosoma mansoni. *Nature*, *494*(7438), 476–9. http://doi.org/10.1038/nature11924

Collu, G. M., Hidalgo-Sastre, A., & Brennan, K. (2014). Wnt–Notch signalling crosstalk in development and disease. *Cellular and Molecular Life Sciences*, *71*(18), 3553–3567. http://doi.org/10.1007/s00018-014-1644-x

Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acid Research*, *32*, 258–261. http://doi.org/10.1093/nar/gkh036

Conus, S., Perozzo, R., Reinheckel, T., Peters, C., Scapozza, L., Yousefi, S., & Simon, H.-U. (2008). Caspase-8 is activated by cathepsin D initiating neutrophil apoptosis during the resolution of inflammation. *The Journal of Experimental Medicine*, *205*(3), 685–698. http://doi.org/10.1084/jem.20072152

Cornford, E. M., & Huot, M. E. (1981). Glucose Transfer from Male to Female Schistosomes Glucose Transfer from Male to Female Schistosomes. *Science*, *213*(4513), 1269–1271.

Cort, W. (1921). the development of the japanese blook-fluke, Schistosoma japonicum katsurada, in its final host. *American Journal of Hygiene*, *1*, 1–38.

Cosseau, C., Azzi, A., Smith, K., Freitag, M., Mitta, G., & Grunau, C. (2009). Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of Schistosoma mansoni: Critical experimental parameters. *Molecular and Biochemical Parasitology*, *166*(1), 70–6. http://doi.org/10.1016/j.molbiopara.2009.02.015

Crellen, T., Walker, M., Lamberton, P. H. L., Kabatereine, N. B., Tukahebwa, E. M., Cotton, J. a., & Webster, J. P. (2016). Reduced Efficacy of Praziquantel Against Schistosoma mansoni Is

Associated with Multiple Rounds of Mass Drug Administration. *Clinical Infectious Diseases*, *63*(9), 1151–1159. http://doi.org/10.1093/cid/ciw506

Dabo, A., Badawi, H. M., Bary, B., & Doumbo, O. K. (2011). Urinary schistosomiasis among preschool-aged children in Sahelian rural communities in Mali. *Parasites & Vectors*, *4*(1), 21. http://doi.org/10.1186/1756-3305-4-21

Davies, S. J., Grogan, J. L., Blank, R. B., Lim, K. C., Locksley, R. M., & McKerrow, J. H. (2001). Modulation of blood fluke development in the liver by hepatic CD4+ lymphocytes. *Science (New York, N.Y.)*, *294*(5545), 1358–1361. http://doi.org/10.1126/science.1064462

Davis, A. H., Ronald, B., & Klich, P. (1985). Stage and sex specific differences in actin gene expression in schistosoma mansoni. *Molecular and Biochemical Parasitology*, *17*, 289–298.

Dean, C., Scholl, F. G., Choih, J., DeMaria, S., Berger, J., Isacoff, E., & Scheiffele, P. (2003). Neurexin mediates the assembly of presynaptic terminals. *Nature Neuroscience*, *6*(7), 708–716. http://doi.org/10.1038/nn1074

Deng, C. X. (2006). BRCA1: Cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. *Nucleic Acids Research*, *34*(5), 1416–1426. http://doi.org/10.1093/nar/gkl010

Dewalick, S., Bexkens, M. L., van Balkom, B. W. M., Wu, Y.-P., Smit, C. H., Hokke, C. H., *et al.* (2011). The proteome of the insoluble Schistosoma mansoni eggshell skeleton. *International Journal for Parasitology*, *41*(5), 523–32. http://doi.org/10.1016/j.ijpara.2010.12.005

Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., Slabicki, M. M., Heninger, A. K., Vries, I. De, *et al.* (2009). A Genome-Scale RNAi Screen for Oct4 Modulators Defines a Role of the Paf1 Complex for Embryonic Stem Cell Identity. *Cell Stem Cell*, *4*(5), 403–415. http://doi.org/10.1016/j.stem.2009.03.009

Downward, J. (1997). Cell cycle: routine role for Ras. *Current Biology*, *7*(4), R258–R260. http://doi.org/10.1016/S0960-9822(06)00116-3

Eddy, E. (1999). Role of heat shock protein HSP70-2 in spermatogenesis. *Reviews of Reproduction*, *4*(1), 23–30. http://doi.org/10.1530/ror.0.0040023

Eijkelenboom, A., & Burgering, B. M. T. (2013). FOXOs: signalling integrators for homeostasis maintenance. *Nature Reviews. Molecular Cell Biology*, *14*(2), 83–97. http://doi.org/10.1038/nrm3507

Eklu-Natey, D. T., Wuest, J., Swiderski, Z., Striebel, H. P., & Huggel, H. (1985). Comparative scanning electron microscope (SEM) study of miracidia of four human schistosomes species. *International Journal for Parasitology*, *15*(1), 33–42.

El Ridi, R., & Tallima, H. (2015). Why the radiation-attenuated cercarial immunization studies failed to guide the road for an effective schistosomiasis vaccine: A review. *Journal of Advanced Research*, *6*(3), 255–267. http://doi.org/10.1016/j.jare.2014.10.002

Erasmus, D. a. (1975). *Schistosoma mansoni*: Development of the vitelline cell, its role in drug sequestration, and changes induced by Astiban. *Experimental Parasitology*, *38*(2), 240–256. http://doi.org/10.1016/0014-4894(75)90027-2

Eswarakumar, V. P., Lax, I., & Schlessinger, J. (2005). Cellular signaling by fibroblast growth factor receptors. *Cytokine & Growth Factor Reviews*, *16*(2), 139–149. http://doi.org/http://dx.doi.org/10.1016/j.cytogfr.2005.01.001

Everts, B., Perona-Wright, G., Smits, H. H., Hokke, C. H., Ham, A. J. Van Der, Fitzsimmons, C. M., *et al.* (2009). Omega-1 , a glycoprotein secreted by *Schistosoma mansoni* eggs , drives Th2 responses. *Journal of Experimental Medicine*, *206*(8), 1673–1680. http://doi.org/10.1084/jem.20082460

Faix, J., & Grosse, R. (2006). Staying in Shape with Formins. *Developmental Cell*, *10*(6), 693–706. http://doi.org/10.1016/j.devcel.2006.05.001

Filmus, J., Capurro, M., & Rast, J. (2008). Glypicans. *Genome Biology*, *9*(5), 224. http://doi.org/10.1186/gb-2008-9-5-224

Fingerut, J. T., Zimmer, C. A. N. N., & Zimmer, R. K. (2003). Patterns and Processes of Larval Emergence in an Estuarine Parasite System. *Biological Bulletin*, *205*, 110–120.

Fitzpatrick, J. M., Hirai, Y., Hirai, H., & Hoffmann, K. F. (2007). Schistosome egg production is dependent upon the activities of two developmentally regulated tyrosinases. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *21*(3), 823–35. http://doi.org/10.1096/fj.06-7314com

Fitzpatrick, J. M., & Hoffmann, K. F. (2006). Dioecious Schistosoma mansoni express divergent gene repertoires regulated by pairing. *International Journal for Parasitology*, *36*(10-11), 1081–9. http://doi.org/10.1016/j.ijpara.2006.06.007

Fitzpatrick, J. M., Johansen, M. V., Johnston, D. a., Dunne, D. W., & Hoffmann, K. F. (2004). Gender-associated gene expression in two related strains of *Schistosoma japonicum*. *Molecular and Biochemical Parasitology*, *136*(2), 191–209. http://doi.org/10.1016/j.molbiopara.2004.03.014

Fitzpatrick, J. M., Johnston, D. a, Williams, G. W., Williams, D. J., Freeman, T. C., Dunne, D. W., & Hoffmann, K. F. (2005). An oligonucleotide microarray for transcriptome analysis of Schistosoma mansoni and its application/use to investigate gender-associated gene expression. *Molecular and Biochemical Parasitology*, *141*(1), 1–13. http://doi.org/10.1016/j.molbiopara.2005.01.007

Fitzpatrick, J. M., Peak, E., Perally, S., Chalmers, I. W., Barrett, J., Yoshino, T. P., *et al.* (2009). Anti-schistosomal intervention targets identified by lifecycle transcriptomic analyses. *PLoS Neglected Tropical Diseases*, *3*(11). http://doi.org/10.1371/journal.pntd.0000543

Franco, G. R., Adams, M. D., Soares, M. B., Simpson, A. J. G., Venter, J. C., & Pena, S. D. J. (1995). Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. *International Journal for Parasitology*, *152*, 141–147.

Freshney, R. (1987). *Culture of Animal Cells: A Manual of Basic Technique*. (A. R. Liss, Ed.). New York.

Friedenson, B. (2007). The BRCA1/2 pathway prevents hematologic cancers in addition to breast and ovarian cancers. *BMC Cancer*, *7*, 152. http://doi.org/10.1186/1471-2407-7-152

Futerman, A. H., & Hannun, Y. a. (2004). The complex life of simple sphingolipids. *EMBO Reports*, *5*(8), 777–782. http://doi.org/10.1038/sj.embor.7400208

Galanti, S. E., Huang, S. C.-C., & Pearce, E. J. (2012). Cell death and reproductive regression in female Schistosoma mansoni. *PLoS Neglected Tropical Diseases*, *6*(2), e1509. http://doi.org/10.1371/journal.pntd.0001509

Gartel, A. L., & Tyner, A. L. (2002). The role of the cyclin-dependent kinase inhibitor p21 in apoptosis. *Molecular Cancer Therapeutics*, *1*(8), 639–649.

Gobert, G. N., McManus, D. P., Nawaratna, S., Moertel, L., Mulvenna, J., & Jones, M. K. (2009). Tissue specific profiling of females of Schistosoma japonicum by integrated laser microdissection microscopy and microarray analysis. *PLoS Neglected Tropical Diseases*, *3*(6), e469. http://doi.org/10.1371/journal.pntd.0000469

Goddard, M. J., & Jordan, P. (1980). On the longevity of Schistosoma in man on St Lucia, West Indies. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *79*, 185–191.

Gray, G. E., Mann, R. S., Mitsiadis, E., Henrique, D., Carcangiu, M. L., Banks, a, *et al.* (1999). Human ligands of the Notch receptor. *The American Journal of Pathology*, *154*(3), 785–94. http://doi.org/10.1016/S0002-9440(10)65325-4

Greenbaum, M. P., Iwamori, N., Agno, J. E., & Matzuk, M. M. (2009). Mouse TEX14 is required for embryonic germ cell intercellular bridges but not female fertility. *Biology of Reproduction*, *80*(3), 449–457. http://doi.org/10.1095/biolreprod.108.070649

Grevelding, C. G., Sommer, G., & Kunz, W. (1997). Female-specific gene expression in Schistosoma mansoni is regulated by pairing. *Parasitology*, *115*, 635–640.

Grove, J. (2014). Super-resolution microscopy: A virus' eye view of the cell. *Viruses*, *6*(3), 1365–1378. http://doi.org/10.3390/v6031365

Gryseels, B. (2012). Schistosomiasis. *Infectious Disease Clinics of North America*, *26*(2), 383–397. http://doi.org/10.1016/j.idc.2012.03.004

Gryseels, B., Polman, K., Clerinx, J., & Kestens, L. (2006). Human schistosomiasis. *Lancet*, *368*(9541), 1106–18. http://doi.org/10.1016/S0140-6736(06)69440-3

Guan, J. L. (1997). Role of focal adhesion kinase in integrin signaling. *The International Journal of Biochemistry & Cell Biology*, *29*(8-9), 1085–1096. http://doi.org/10.1016/S1357-2725(97)00051-4

Gurley, K. a., Elliott, S. a., Simakov, O., Schmidt, H. a., Holstein, T. W., & Alvarado, A. S. (2010). Expression of secreted Wnt pathway components reveals unexpected complexity of the planarian amputation response. *Developmental Biology*, *347*(1), 24–39. http://doi.org/10.1016/j.ydbio.2010.08.007

Guzzo, G., Sciacovelli, M., Bernardi, P., & Rasola, A. (2014). Inhibition of succinate dehydrogenase by the mitochondrial chaperone TRAP1 has anti-oxidant and anti-apoptotic effects on tumor cells. *Oncotarget*, *5*(23), 11897–908. http://doi.org/10.18632/oncotarget.2472

Hagan, P., & Sharaf, O. (2003). Schistosomiasis vaccines. *Expert Opinion on Biological Therapy*, *3*(8), 1271–1278. http://doi.org/10.4161/hv.7.11.17017

Hahnel, S., Lu, Z., Wilson, R. A., Grevelding, C. G., & Quack, T. (2013). Whole-organ isolation approach as a basis for tissue-specific analyses in Schistosoma mansoni. *PLoS Neglected Tropical Diseases*, *7*(7), e2336. http://doi.org/10.1371/journal.pntd.0002336

Hahnel, S., Quack, T., Parker-Manuel, S. J., Lu, Z., Vanderstraete, M., Morel, M., *et al.* (2014). Gonad RNA-specific qRT-PCR analyses identify genes with potential functions in schistosome reproduction such as SmFz1 and SmFGFRs. *Frontiers in Genetics*, *5*, 170. http://doi.org/10.3389/fgene.2014.00170

Handford, P. a. (2000). Fibrillin-1, a calcium binding protein of extracellular matrix. *Biochim Biophys Acta*, *1498*(2-3), 84–90. http://doi.org/S0167-4889(00)00085-9 [pii]

Hargrave, M., Karunaratne, a, Cox, L., Wood, S., Koopman, P., & Yamada, T. (2000). The HMG box transcription factor gene Sox14 marks a novel subset of ventral interneurons and is regulated by sonic hedgehog. *Developmental Biology*, *219*(1), 142–153. http://doi.org/10.1006/dbio.1999.9581

Harris, A. R. C., Russell, R. J., & Charters, A. D. (1984). A review of schistosomiasis in immigrants in Western Australia, demonstrating the unusual longevity of *Schistosoma mansoni*. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *78*(3), 385–388. http://doi.org/10.1016/0035-9203(84)90129-9

Heinrich, M., Wickel, M., Schneider-Brachert, W., Sandberg, C., Gahr, J., Schwandner, R., *et al.* (1999). Cathepsin D targeted by acid sphingomyelinase- derived ceramide. *The EMBO Journal*, *18*(19), 5252–5263.

Herbst, R. S. (2004). Review of epidermal growth factor receptor biology. *Radiation Oncology*, *59*(2), S21–S26. http://doi.org/http://dx.doi.org/10.1016/j.ijrobp.2003.11.041

Hewitson, J. P., Hamblin, P. a., & Mountford, a. P. (2005). Immunity induced by the radiation-attenuated schistosome vaccine. *Parasite Immunology*, *27*(7-8), 271–280. http://doi.org/10.1111/j.1365-3024.2005.00764.x

Hoffmann, K., Johnston, D. A., & Dunne, D. W. (2002). Identification of *Schistosoma mansoni* gender-associated gene transcripts by cDNA microarray profiling. *Genome Biology*, 1–12.

Hoog, C. (1991). Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy, *19*(22), 6123–6127.

Hsu, D. R., & Meyer, B. J. (1994). The Dpy-30 Gene Encodes an Essential Component of the *Caenorhabditis-elegans* Dosage Compensation Machinery. *Genetics*, *137*(4), 999–1018. Retrieved from <Go to ISI>://WOS:A1994NX91200011

Hsü, S. Y. L., Hsü, H. F., & Chu, K. Y. (1962). Interbreeding of geographic strains of *Schistosoma japonicum*. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *56*(5), 383–385.

Huang, S. C.-C., Freitas, T. C., Amiel, E., Everts, B., Pearce, E. L., Lok, J. B., & Pearce, E. J. (2012). Fatty acid oxidation is essential for egg production by the parasitic flatworm *Schistosoma mansoni*. *PLoS Pathogens*, *8*(10), e1002996. http://doi.org/10.1371/journal.ppat.1002996

Hwangbo, C., Tae, N., Lee, S., Kim, O., Park, O. K., Kim, J., *et al.* (2016). Syntenin regulates TGF-β1-induced Smad activation and the epithelial-to-mesenchymal transition by inhibiting caveolin-mediated TGF-β type I receptor internalization. *Oncogene*, *35*(3), 389–401. http://doi.org/10.1038/onc.2015.100

Ingham, P. W., Nakano, Y., & Seger, C. (2011). Mechanisms and functions of Hedgehog signalling across the metazoa. *Nature Reviews. Genetics*, *12*(6), 393–406. http://doi.org/10.1038/nrg2984

Jenkins, S. J., Hewitson, J. P., Jenkins, G. R., & Mountford, A. P. (2007). Modulation of the host's immune response by schistosome larvae. *Parasite Immunology*, *27*(10-11), 385–393. http://doi.org/10.1111/j.1365-3024.2005.00789.x.Modulation

Jensen, S. a, Calvert, A. E., Volpert, G., Kouri, F. M., Hurley, L. a, Luciano, J. P., *et al.* (2014). Bcl2L13 is a ceramide synthase inhibitor in glioblastoma. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(15), 5682–7. http://doi.org/10.1073/pnas.1316700111

Jiang, H., Shukla, A., Wang, X., Chen, W., Bernstein, B. E., & Roeder, R. G. (2011). Role for Dpy-30 in ES Cell-Fate Specification by Regulation of H3K4 Methylation within Bivalent Domains. *Cell*, *144*(4), 513–525. http://doi.org/10.1016/j.cell.2011.01.020

Jiang, J., Zhang, N., Shiba, H., Li, L., & Wang, Z. (2013). Spermatogenesis Associated 4 Promotes Sertoli Cell Proliferation Modulated Negatively by Regulatory Factor X1. *PLoS ONE*, *8*(10), 1–10. http://doi.org/10.1371/journal.pone.0075933

Jimenez, G., Shvartsman, S. Y., & Paroush, Z. (2012). The Capicua repressor - a general sensor of RTK signaling in development and disease. *Journal of Cell Science*, *125*(6), 1383–1391. http://doi.org/10.1242/jcs.092965

Jin, T., & Liu, L. (2008). The Wnt signaling pathway effector TCF7L2 and type 2 diabetes mellitus. *Molecular Endocrinology (Baltimore, Md.)*, *22*(11), 2383–2392. http://doi.org/10.1210/me.2008-0135

Johansson, A.-C., Steen, H., Ollinger, K., & Roberg, K. (2003). Cathepsin D mediates cytochrome c release and caspase activation in human fibroblast apoptosis induced by staurosporine. *Cell Death and Differentiation*, *10*(11), 1253–1259. http://doi.org/10.1038/sj.cdd.4401290

Joza, N., Susin, S. a, Daugas, E., Stanford, W. L., Cho, S. K., Li, C. Y., *et al.* (2001). Essential role of the mitochondrial apoptosis-inducing factor in programmed cell death. *Nature*, *410*(6828), 549–554. http://doi.org/10.1038/35069004

Jung, K.-K., Liu, X.-W., Chirco, R., Fridman, R., & Kim, H.-R. C. (2006). Identification of CD63 as a tissue inhibitor of metalloproteinase-1 interacting cell surface protein. *The EMBO Journal*, *25*(17), 3934–3942. http://doi.org/10.1038/sj.emboj.7601281

Kanehisa Laboratories. (2012). KEGG: Apoptosis - Reference pathway. Retrieved from: https://www.genome.jp/kegg/pathway/hsa/hsa04210.html

Kapp, K., Knobloch, J., Schüssler, P., Sroka, S., Lammers, R., Kunz, W., & Grevelding, C. G. (2004). The *Schistosoma mansoni* Src kinase TK3 is expressed in the gonads and likely involved in cytoskeletal organization. *Molecular and Biochemical Parasitology*, *138*(2), 171–82. http://doi.org/10.1016/j.molbiopara.2004.07.010

Kierszenbaum, A. L. (2002). Sperm axoneme: A tale of tubulin posttranslation diversity. *Molecular Reproduction and Development*, *62*(1), 1–3. http://doi.org/10.1002/mrd.10139

Kim, H. J., Qiao, Q., Toop, H. D., Morris, J. C., & Don, A. S. (2012). A fluorescent assay for ceramide synthase activity. *Journal of Lipid Research*, *53*(8), 1701–1707. http://doi.org/10.1194/jlr.D025627

King, C. H., Dickman, K., & Tisch, D. J. (2005). Reassessment of the cost of chronic helmintic infection: A meta-analysis of disability-related outcomes in endemic schistosomiasis. *Lancet*, *365*(9470), 1561–1569. http://doi.org/10.1016/S0140-6736(05)66457-4

Knobloch, J., Beckmann, S., Burmeister, C., Quack, T., & Grevelding, C. G. (2007). Tyrosine kinase and cooperative TGFbeta signaling in the reproductive organs of *Schistosoma mansoni*. *Experimental Parasitology*, *117*(3), 318–36. http://doi.org/10.1016/j.exppara.2007.04.006

Knobloch, J., Kunz, W., & Grevelding, C. G. (2006). Herbimycin A suppresses mitotic activity and egg production of female *Schistosoma mansoni*. *International Journal for Parasitology*, *36*(12), 1261–72. http://doi.org/10.1016/j.ijpara.2006.06.004

Knobloch, J., Winnen, R., Quack, M., Kunz, W., & Grevelding, C. G. (2002). A novel Syk-family tyrosine kinase from *Schistosoma mansoni* which is preferentially transcribed in reproductive organs. *Gene*, *294*(1-2), 87–97. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12234670

Koelsch, K. a., Webb, R., Jeffries, M., Dozmorov, M. G., Frank, M. B., Guthridge, J. M., *et al.* (2013). Functional characterization of the MECP2/IRAK1 lupus risk haplotype in human T cells and a human MECP2 transgenic mouse. *Journal of Autoimmunity*, *41*, 168–174. http://doi.org/10.1016/j.jaut.2012.12.012

Köster, B., Dargatz, H., Schröder, J., Hirzmann, J., Haarmann, C., Symmons, P., & Kunz, W. (1988). Identification and localisation of the products of a putative eggshell precursor gene in the vitellarium of *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, *31*(2), 183–98. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2847044

Koussounadis, A., Langdon, S. P., Um, I. H., Harrison, D. J., & Smith, V. A. (2015). Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Scientific Reports*, *5*(May), 10775. http://doi.org/10.1038/srep10775

Kuales, G., De Mulder, K., Glashauser, J., Salvenmoser, W., Takashima, S., Hartenstein, V., *et al.* (2011). Boule-like genes regulate male and female gametogenesis in the flatworm *Macrostomum lignano*. *Developmental Biology*, *357*(1), 117–132. http://doi.org/10.1016/j.ydbio.2011.06.030

Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., *et al.* (2006). The real-time polymerase chain reaction. *Molecular Aspects of Medicine*, *27*(2-3), 95–125. http://doi.org/10.1016/j.mam.2005.12.007

Kuckwa, J., Fritzen, K., Buttgereit, D., Rothenbusch-Fender, S., & Renkawitz-Pohl, R. (2016). A new level of plasticity: *Drosophila* smooth-like testes muscles compensate failure of myoblast fusion. *Development*, *143*(2), 329–338.

Kunz, W. (2001). Schistosome male-female interaction: induction of germ-cell differentiation. *Trends in Parasitology*, *17*(5), 227–31. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11323306

Kyrylkova, K., Kyryachenko, S., Leid, M., & Kioussi, C. (2012). Detection of Apoptosis by TUNEL Assay. *Methods in Molecular Biology*, *887*, 41–47. http://doi.org/10.1007/978-1-61779-860-3_5

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Software. 10*(3). http://doi.org/10.1186/gb-2009-10-3-r25

Lee, E. F., Clarke, O. B., Evangelista, M., Feng, Z., Speed, T. P., Tchoubrieva, E. B., *et al.* (2011). Discovery and molecular characterization of a Bcl-2-regulated cell death pathway in schistosomes. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(17), 6999–7003. http://doi.org/10.1073/pnas.1100652108

Lee, E. F., Young, N. D., Lim, N. T. Y., Gasser, R. B., & Fairlie, W. D. (2014). Apoptosis in schistosomes: toward novel targets for the treatment of schistosomiasis. *Trends in Parasitology*, *30*(2), 75–84. http://doi.org/10.1016/j.pt.2013.12.005

Lepesant, J. M. J., Cosseau, C., Boissier, J., Freitag, M., Portela, J., Climent, D., *et al.* (2012). Chromatin structural changes around satellite repeats on the female sex chromosome in *Schistosoma mansoni* and their possible role in sex chromosome emergence. *Genome Biology*, *13*(2), R14. http://doi.org/10.1186/gb-2012-13-2-r14

Leutner, S., Oliveira, K. C., Rotter, B., Beckmann, S., Buro, C., Hahnel, S., *et al.* (2013). Combinatory microarray and SuperSAGE analyses identify pairing-dependently transcribed genes in

Schistosoma mansoni males, including follistatin. *PLoS Neglected Tropical Diseases*, *7*(11), e2532. http://doi.org/10.1371/journal.pntd.0002532

Levy, M., & Futerman, A. H. (2010). Mammalian Ceramide Sythases. *IUBMB Life*, *62*(5), 347–356. http://doi.org/10.1002/iub.319.Mammalian

Levy, S., & Shoham, T. (2005). The tetraspanin web modulates immune-signalling complexes. *Nature Reviews Immunology*, *5*(2), 136–148. http://doi.org/10.1038/nri1548

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–9. http://doi.org/10.1093/bioinformatics/btp352

Li, J. J., Bickel, P. J., & Biggin, M. D. (2014). System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ*, *2*, e270. http://doi.org/10.7717/peerj.270

Li, X. Q., Björkman, A., Andersson, T. B., Gustafsson, L. L., & Masimirembwa, C. M. (2003). Identification of human cytochrome P450s that metabolise anti-parasitic drugs and predictions of in vivo drug hepatic clearance from in vitro data. *European Journal of Clinical Pharmacology*, *59*(5-6), 429–442. http://doi.org/10.1007/s00228-003-0636-9

Lin, W., Zhou, X., Zhang, M., Li, Y., Miao, S., Wang, L., *et al.* (2001). Expression and function of the HSD-3.8 gene encoding a testis-specific protein. *Molecular Human Reproduction*, *7*(9), 811–818.

Liu, S., Cai, P., Hou, N., Piao, X., Wang, H., Hung, T., & Chen, Q. (2012). Genome-wide identification and characterization of a panel of house-keeping genes in *Schistosoma japonicum*. *Molecular and Biochemical Parasitology*, *182*(1-2), 75–82. http://doi.org/10.1016/j.molbiopara.2011.12.007

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and. *Methods*, *25*, 402–408. http://doi.org/10.1006/meth.2001.1262

Llamazares, S., Moreira, a., Tavares, a., Girdham, C., Spruce, B. a., Gonzalez, C., *et al.* (1991). Polo encodes a protein kinase homolog required for mitosis in *Drosophila*. *Genes and Development*, *5*(12 A), 2153–2165. http://doi.org/10.1101/gad.5.12a.2153

Logan-Klumpler, F. J., De Silva, N., Boehme, U., Rogers, M. B., Velarde, G., McQuillan, J. a., *et al.* (2012). GeneDB-an annotation database for pathogens. *Nucleic Acids Research*, *40*(D1), 98–108. http://doi.org/10.1093/nar/gkr1032

Loker, E. S., & Brant, S. V. (2006). Diversification, dioecy and dimorphism in schistosomes. *Trends in Parasitology*, *22*(11), 521–8. http://doi.org/10.1016/j.pt.2006.09.001

Long, T., Cailliau, K., Beckmann, S., Browaeys, E., Trolet, J., Grevelding, C. G., & Dissous, C. (2010). Schistosoma mansoni Polo-like kinase 1: A mitotic kinase with key functions in parasite reproduction. *International Journal for Parasitology*, *40*(9), 1075–86. http://doi.org/10.1016/j.ijpara.2010.03.002

Loukas, A., Gaze, S., Mulvenna, J. P., Gasser, R. B., Brindley, P. J., Doolan, D. L., *et al.* (2011). Vaccinomics for the major blood feeding helminths of humans. *Omics : A Journal of Integrative Biology*, *15*(9), 567–77. http://doi.org/10.1089/omi.2010.0150

Louvi, A., & Artavanis-Tsakonas, S. (2006). Notch signalling in vertebrate neural development. *Nature Reviews Neuroscience*, *7*(2), 93–102. http://doi.org/10.1038/nrn1847

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. http://doi.org/10.1186/s13059-014-0550-8

LoVerde, P. T., Andrade, L. F., & Oliveira, G. (2009). Signal transduction regulates schistosome reproductive biology. *Current Opinion in Microbiology*, *12*(4), 422–8. http://doi.org/10.1016/j.mib.2009.06.005

Loverde, P. T., & Chen, L. (1991). Schistosome Female Reproductive Development. *Parasitology today. 7*(11), 303–8.

LoVerde, P. T., Niles, E. G., Osman, A., & Wu, W. (2005). *World Class Parasites: Volume 10 - Schistosomiasis*. (S. J. Black, J. R. Seed, W. E. Secor, & D. G. Colley, Eds.). Springer.

Loverde, P. T., Osman, A., & Hinck, A. (2007, November). *Schistosoma mansoni*: TGF-beta signaling pathways. *Experimental Parasitology*. http://doi.org/10.1016/j.exppara.2007.06.002

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLoS Computational Biology*, *13*(5), 1–23. http://doi.org/10.1371/journal.pcbi.1005457

Lu, Z., Sessler, F., Holroyd, N., Hahnel, S., Quack, T., Berriman, M., & Grevelding, C. G. (2016). Schistosome sex matters: a deep view into gonad-specific and pairing-dependent transcriptomes reveals a complex gender interplay. *Scientific Reports*, *6*(1), 31150. http://doi.org/10.1038/srep31150

Lu, Z., Sessler, F., Holroyd, N., Hahnel, S., Quack, T., Berriman, M., & Grevelding, C. G. (2017). A gene expression atlas of adult *Schistosoma mansoni* and their gonads. *Scientific Data*, *4*, 170118. http://doi.org/10.1038/sdata.2017.118

Luzio, J. P., Pryor, P. R., & Bright, N. a. (2007). Lysosomes: fusion and function. *Nature Reviews Molecular Cell Biology*, *8*(8), 622–632. http://doi.org/10.1038/nrm2217

MacDonald, K., Buxton, S., Kimber, M. J., Day, T. a, Robertson, A. P., & Ribeiro, P. (2014). Functional characterization of a novel family of acetylcholine-gated chloride channels in *Schistosoma mansoni*. *PLoS Pathogens*, *10*(6), e1004181. http://doi.org/10.1371/journal.ppat.1004181

Machado-Silva, J. R., Lanfredi, R. M., & Gomes, D. C. (1997). Morphological study of adult male worms of *Schistosoma mansoni* Sambon, 1907 by scanning electron microscopy. *Memórias Do Instituto Oswaldo Cruz*, *92*(5), 647–53. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9566233

Maier, T., Güell, M., & Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Letters*, *583*(24), 3966–3973. http://doi.org/10.1016/j.febslet.2009.10.036

Malick, L., & Wilson, R. (1975). Modified thiocarbohydrazide procedure for scanning electron microscopy: routine use for normal, pathological, or experimental tissues. *Stain Technology*, *50*(4), 265–9.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., & Gilad, Y. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, *18*(9), 1509–1517. http://doi.org/10.1101/gr.079558.108

Markel, S. F., LoVerde, P. T., & Britt, E. M. (1978). Prolonged latent schistosomiasis. *Journal of the American Medical Association*, *30*, 45–54.

Massagué, J. (2012). TGFβ signalling in context. *Nature Reviews Molecular Cell Biology*, *13*(10), 616–630. http://doi.org/10.1038/nrm3434

Mcgarvey, S. T. (2000). Schistosomiasis : Impact on Childhood and Adolescent Growth , Malnutrition , and Morbidity. *Seminars in Pediatric Infectious DIseases*, *11*(4), 269–274. http://doi.org/10.1053/spid.2000.9642

McIlwain, D. R., Berger, T., & Mak, T. W. (2013). Caspase functions in cell death and disease. *Cold Spring Harbor Perspectives in Biology*, *5*(4), 1–28. http://doi.org/10.1101/cshperspect.a008656

McKerrow, J., & Salter, J. (2002). Invasion of skin by Schistosoma cercariae. *Trends in Parasitology*, *18*(5), 193–195. http://doi.org/10.1016/S1471-4922(02)02309-7

McManus, D. P., & Loukas, A. (2008). Current status of vaccines for schistosomiasis. *Clinical Microbiology Reviews*, *21*(1), 225–242. http://doi.org/10.1128/CMR.00046-07

McTigue, M. a, Williams, D. R., & Tainer, J. a. (1995). Crystal structures of a schistosomal drug and vaccine target: glutathione S-transferase from *Schistosoma japonica* and its complex with the leading antischistosomal drug praziquantel. *Journal of Molecular Biology*, *246*(1), 21–7. http://doi.org/10.1006/jmbi.1994.0061

Merrifield, M., Hotez, P. J., Beaumier, C. M., Gillespie, P., Strych, U., Hayward, T., & Bottazzi, M. E. (2016). Advancing a vaccine to prevent human schistosomiasis. *Vaccine*, *34*(26), 2988–2991. http://doi.org/10.1016/j.vaccine.2016.03.079

Meyer, F., Meyer, H., & Bueding, E. (1970). Lipid metabolism in the parasitic and free-living flatworms, *Schistosoma mansoni* and *Dugesia dorotocephala*. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, *210*(2), 257–266. http://doi.org/10.1016/0005-2760(70)90170-0

Miller, M. B., & Tang, Y. W. (2009). Basic concepts of microarrays and potential applications in clinical microbiology. *Clinical Microbiology Reviews*, *22*(4), 611–633. http://doi.org/10.1128/CMR.00019-09

Miller, P., & Wilson, R. (1980). Migration of the schistosomula of Schistosoma mansoni from the lungs to the hepatic portal system. *Parasitology*, *80*, 267–288.

Mitra, S. K., & Schlaepfer, D. D. (2006). Integrin-regulated FAK-Src signaling in normal and cancer cells. *Current Opinion in Cell Biology*, *18*(5), 516–523. http://doi.org/10.1016/j.ceb.2006.08.011

Mócsai, A., Ruland, J., & Tybulewicz, V. L. J. (2010). The SYK tyrosine kinase: a crucial player in diverse biological functions. *Nature Publishing Group*, *10*(6), 387–402. http://doi.org/10.1038/nri2765

Moertel, L., McManus, D. P., Piva, T. J., Young, L., McInnes, R. L., & Gobert, G. N. (2006). Oligonucleotide microarray analysis of strain- and gender-associated gene expression in the human blood fluke, *Schistosoma japonicum*. *Molecular and Cellular Probes*, *20*(5), 280–9. http://doi.org/10.1016/j.mcp.2006.02.002

Morris, E. J., Michaud, W. a., Ji, J. Y., Moon, N. S., Rocco, J. W., & Dyson, N. J. (2006). Functional identification of Api5 as a suppressor of E2F-dependent apoptosis in vivo. *PLoS Genetics*, *2*(11), 1834–1848. http://doi.org/10.1371/journal.pgen.0020196

Mullen, T. D., Hannun, Y. A., & Obeid, L. M. (2012). Ceramide synthases at the centre of sphingolipid metabolism and biology. *Biochemistry Journal*, *441*(3), 789–802. http://doi.org/10.1042/BJ20111626.Ceramide

Mullen, T. D., & Obeid, L. M. (2012). Ceramide and Apoptosis: Exploring the Enigmatic Connections between Sphingolipid Metabolism and Programmed Cell Death. *Anti-Cancer Agents in Medicinal Chemistry*, *12*(4), 340–363.

Mulligan, K. a., & Cheyette, B. N. R. (2012). Wnt Signaling in Vertebrate Neural Development and Function. *Journal of Neuroimmune Pharmacology*, *7*(4), 774–787. http://doi.org/10.1007/s11481-012-9404-x

Muniyappa, K., Kshirsagar, R., & Ghodke, I. (2014). The HORMA domain: An evolutionarily conserved domain discovered in chromatin-associated proteins, has unanticipated diverse functions. *Gene*, *545*(2), 194–197. http://doi.org/10.1016/j.gene.2014.05.020

Murray, C. J. L., & Lopez, a. D. (1997). Global mortality, disability and the contributions of risk factors: Global Burden of Disease Study. *The Lancet*, *349,* 1436–1442.

Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., *et al.* (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, *380*(9859), 2197–2223. http://doi.org/10.1016/S0140-6736(12)61689-4

Nagata, S. (1997). Apoptosis by death factor. *Cell*, *88*(3), 355–365. http://doi.org/10.1016/S0092-8674(00)81874-7

Nagata, S. (1999). Fas Ligand-Induced Apoptosis. *Annual Reviews Genetics*, *33*, 29–55.

Naghavi, M., Wang, H., Lozano, R., Davis, A., Liang, X., Zhou, M., *et al.* (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, *385*(9963), 117–171. http://doi.org/10.1016/S0140-6736(14)61682-2

Nambu, J. R., Lewis, J. O., Wharton, K. A., & Crews, S. T. (1991). The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell*, *67*(6), 1157–1167.

Nawaratna, S. S. K., McManus, D. P., Moertel, L., Gobert, G. N., & Jones, M. K. (2011). Gene Atlasing of digestive and reproductive tissues in *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases*, *5*(4), e1043. http://doi.org/10.1371/journal.pntd.0001043

Nesi, G., Nobili, S., Cai, T., Caini, S., & Santi, R. (2015). Chronic inflammation in urothelial bladder cancer. *Virchows Archiv*, *467*(6), 623–633. http://doi.org/10.1007/s00428-015-1820-x

Neves, R. H., de Lamare Biolchini, C., Machado-Silva, J. R., Carvalho, J. J., Branquinho, T. B., Lenzi, H. L., *et al.* (2005). A new description of the reproductive system of *Schistosoma mansoni* (Trematoda: Schistosomatidae) analyzed by confocal laser scanning microscopy. *Parasitology Research*, *95*(1), 43–9. http://doi.org/10.1007/s00436-004-1241-2

Nogi, T., Zhang, D., Chan, J. D., & Marchant, J. S. (2009). A Novel Biological Activity of Praziquantel Requiring Voltage-Operated Ca2+ Channel β Subunits: Subversion of Flatworm Regenerative Polarity. *PLoS Neglected Tropical Diseases*, *3*(6), e464. http://doi.org/10.1371/journal.pntd.0000464

Nusse, R. (2012). Wnt Signaling. *Cold Spring Harbor Perspectives in Biology*, *4*(5), a011163. http://doi.org/10.1101/cshperspect.a011163

Nusse, R., & Varmus, H. (2012). Three decades of Wnts: a personal perspective on how a scientific field developed. *The EMBO Journal*, *31*(12), 2670–2684. http://doi.org/10.1038/emboj.2012.146

Nusse, R., & Varmus, H. E. (1992). Wnt genes. *Cell*, *69*(7), 1073–1087. http://doi.org/10.1016/0092-8674(92)90630-U

Oliveira, K. C., Carvalho, M. L. P., Verjovski-Almeida, S., & LoVerde, P. T. (2012). Effect of human TGF-β on the gene expression profile of *Schistosoma mansoni* adult worms. *Molecular and Biochemical Parasitology*. http://doi.org/10.1016/j.molbiopara.2012.02.008

Olson, M. F. (2003). GTPase signalling: New functions for Diaphanous-related formins. *Current Biology*, *13*(9), 360–362. http://doi.org/10.1016/S0960-9822(03)00277-X

Ong, C. C., Jubb, A. M., Haverty, P. M., Zhou, W., Tran, V., Truong, T., *et al.* (2011). Targeting p21-activated kinase 1 (PAK1) to induce apoptosis of tumor cells. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(17), 7177–7182. http://doi.org/10.1073/pnas.1103350108

Oshlack, A., & Wakefield, M. J. (2009). Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct*, *4*(1), 14. http://doi.org/10.1186/1745-6150-4-14

Osman, A., Niles, E. G., Verjovski-Almeida, S., & LoVerde, P. T. (2006). *Schistosoma mansoni* TGF-beta receptor II: role in host ligand-induced regulation of a schistosome target gene. *PLoS Pathogens*, *2*(6), e54. http://doi.org/10.1371/journal.ppat.0020054

Pagán, A. J., & Ramakrishnan, L. (2018). The Formation and Function of Granulomas. *Annual Review of Immunology*, *36*(1), annurev–immunol–032712–100022. http://doi.org/10.1146/annurev-immunol-032712-100022

Paris, L., & Caumes, E. (2010). Acute schistosomiasis , a diagnostic and therapeutic challenge. *European Society of Clinical Microbiology and Infectious Disease*, *16*, 225–231. http://doi.org/10.1111/j.1469-0691.2009.03131.x

Pawlowski, J., & Kraft, a S. (2000). Bax-induced apoptotic cell death. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(2), 529–531. http://doi.org/10.1073/pnas.97.2.529

Pearce, E. J., & Huang, S. C. C. (2015). The metabolic control of schistosome egg production. *Cellular Microbiology*, *17*(6), 796–801. http://doi.org/10.1111/cmi.12444

Peng, J., Yang, Y., Feng, X., Cheng, G., & Lin, J. (2010). Molecular characterizations of an inhibitor of apoptosis from *Schistosoma japonicum*. *Parasitology Research*, *106*(4), 967–976. http://doi.org/10.1007/s00436-010-1752-y

Piao, X., Cai, P., Liu, S., Hou, N., Hao, L., Yang, F., *et al.* (2011). Global Expression Analysis Revealed Novel Gender-Specific Gene Expression Features in the Blood Fluke Parasite *Schistosoma japonicum*. *PLoS ONE*, *6*(4), e18267. http://doi.org/10.1371/journal.pone.0018267

Picard, M. a L., Boissier, J., Roquis, D., Grunau, C., Allienne, J.-F., Duval, D., *et al.*(2016). Sex-Biased Transcriptome of *Schistosoma mansoni*: Host-Parasite Interaction, Genetic Determinants and Epigenetic Regulators Are Associated with Sexual Differentiation. *PLoS Neglected Tropical Diseases*, *10*(9), e0004930. http://doi.org/10.1371/journal.pntd.0004930

Poole, H., Terlouw, D. J., Naunje, A., Mzembe, K., Stanton, M., Betson, M., *et al.* (2014). Schistosomiasis in pre-school-age children and their mothers in Chikhwawa district, Malawi with notes on characterization of schistosomes and snails. *Parasites & Vectors*, *7*(1), 153. http://doi.org/10.1186/1756-3305-7-153

Popiel, I. (1986). Male-stimulated female maturation in *Schistosoma*: a Review. *Journal of Chemical Ecology*. *12*(8).

Popiel, I., & Basch, P. F. (1984). Reproductive development of female *Schistosoma mansoni* (Digenea: Schistosomatidae) following bisexual pairing of worms and worm segments. *The Journal of Experimental Zoology*, *232*(1), 141–50. http://doi.org/10.1002/jez.1402320117

Popiel, I., & Basch, P. F. (1986). *Schistosoma mansoni*: cholesterol uptake by paired and unpaired worms. *Experimental Parasitology*, *61*(3), 343–347.

Portela, J., Grunau, C., Cosseau, C., Beltran, S., Dantec, C., Parrinello, H., & Boissier, J. (2010). Whole-genome in-silico subtractive hybridization (WISH) - using massive sequencing for the identification of unique and repetitive sex-specific sequences: the example of *Schistosoma mansoni*. *BMC Genomics*, *11*(1), 387. http://doi.org/10.1186/1471-2164-11-387

Post, S. M., Tomkinson, A. E., & Lee, E. Y. H. P. (2003). The human checkpoint Rad protein Rad17 is chromatin-associated throughout the cell cycle, localizes to DNA replication sites, and interacts with DNA polymerase E. *Nucleic Acids Research*, *31*(19), 5568–5575. http://doi.org/10.1093/nar/gkg765

Protasio, A. (2011). Transcriptome characterisation of cercariae and skin-stage schistosomula in the parasitic helminth Schistosoma mansoni. University of Cambridge. Retrieved from: ftp://ftp.sanger.ac.uk/pub/resources/theses/ap6/thesis.pdf

Protasio, A. V, Tsai, I. J., Babbage, A., Nichol, S., Hunt, M., Aslett, M. a, *et al.* (2012). A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases*, *6*(1), e1455. http://doi.org/10.1371/journal.pntd.0001455

Puri, P., Acker-Palmer, A., Stahler, R., Chen, Y., Kline, D., & Vijayaraghavan, S. (2011). Identification of testis 14-3-3 binding proteins by tandem affinity purification. *Spermatogenesis*, *1*(4), 354–365. http://doi.org/10.4161/spmg.1.4.18902

Qi, M. (2005). MAP kinase pathways. *Journal of Cell Science*, *118*(16), 3569–3572. http://doi.org/10.1242/jcs.02470

Qin, H., Srinivasula, S. M., Wu, G., Fernandes-Alnemri, T., Alnemri, E. S., & Shi, Y. (1999). Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. *Nature*, *399*(6736), 549–557. http://doi.org/10.1038/21124

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Ravindranath, A., & Cadigan, K. M. (2014). Structure-Function Analysis of the C-clamp of TCF/Pangolin in Wnt/ß-catenin Signaling. *PLoS ONE*, *9*(1), e86180. http://doi.org/10.1371/journal.pone.0086180

Realini, N., Solorzano, C., Pagliuca, C., Pizzirani, D., Armirotti, A., Luciani, R., *et al.* (2013). Discovery of highly potent acid ceramidase inhibitors with in vitro tumor chemosensitizing activity. *Scientific Reports*, *3*(Table 1). http://doi.org/10.1038/srep01035

Ribeiro, P., & Patocka, N. (2013). Neurotransmitter transporters in schistosomes: Structure, function and prospects for drug discovery. *Parasitology International*, *62*(6), 629–638. http://doi.org/10.1016/j.parint.2013.06.003

Robinson, K. S., Clements, a, Williams, a C., Berger, C. N., & Frankel, G. (2011). Bax inhibitor 1 in apoptosis and disease. *Oncogene*, *30*(21), 2391–2400. http://doi.org/10.1038/onc.2010.636

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. http://doi.org/10.1093/bioinformatics/btp616

Rodríguez, A., & D'Andrea, A. (2017). Fanconi anemia pathway. *Current Biology*, *27*(18), R986–R988. http://doi.org/10.1016/j.cub.2017.07.043

Rojo, J. U., Melkus, M. W., Kottapalli, K. R., Okiya, O. E., Sudduth, J., Zhang, W., *et al.* (2017). Sm-p80-based Schistosomiasis vaccine mediated epistatic interactions identified potential immune signatures for vaccine efficacy in mice and baboons. *PLoS ONE*, *12*(2), 1–28. http://doi.org/10.1371/journal.pone.0171677

Rosenquist, M. (2003). 14-3-3 Proteins in Apoptosis. *Brazilian Journal of Medical and Biological Research*, *36*(4), 403–408. http://doi.org/10.1590/S0100-879X2003000400001

Roshak, A. K., Capper, E. a., Imburgia, C., Fornwald, J., Scott, G., & Marshall, L. a. (2000). The human polo-like kinase, PLK, regulates cdc2/cyclin B through phosphorylation and activation of the cdc25C phosphatase. *Cellular Signalling*, *12*(6), 405–411. http://doi.org/10.1016/S0898-6568(00)00080-2

Ross, A. G., Vickers, D., Olds, G. R., Shah, S. M., & Mcmanus, D. P. (2007). Katayama syndrome. *Lancet Infectious Diseases 7*, 218–24.

Sabah, A. A., Fletcher, C., Webbe, G., & Doenhoff, M. J. (1986). *Schistosoma mansoni*: Chemotherapy of infections of different ages. *Experimental Parasitology*, *61*(3), 294–303.

Salomon, J. a., Haagsma, J. a., Davis, A., de Noordhout, C. M., Polinder, S., Havelaar, A. H., *et al.* (2015). Disability weights for the Global Burden of Disease 2013 study. *The Lancet Global Health*, *3*(11), e712–e723. http://doi.org/10.1016/S2214-109X(15)00069-8

Salter, J. P., Lim, K. C., Hansell, E., Hsieh, I., & McKerrow, J. H. (2000). Schistosome invasion of human skin and degradation of dermal elastin are mediated by a single serine protease. *Journal of Biological Chemistry*, *275*(49), 38667–38673. http://doi.org/10.1074/jbc.M006997200

Samuels-Lev, Y., O'Connor, D. J., Bergamaschi, D., Trigiante, G., Hsieh, J. K., Zhong, S., *et al.* (2001). ASPP proteins specifically stimulate the apoptotic function of p53. *Molecular Cell*, *8*(4), 781–794. http://doi.org/10.1016/S1097-2765(01)00367-7

Santini-Oliveira, M., Coler, R. N., Parra, J., Veloso, V., Jayashankar, L., Pinto, P. M., *et al.* (2016). Schistosomiasis vaccine candidate Sm14/GLA-SE: Phase 1 safety and immunogenicity clinical trial in healthy, male adults. *Vaccine*, *34*(4), 586–594. http://doi.org/10.1016/j.vaccine.2015.10.027

Santos, T. M., Machado, C. R., Franco, G. R., & Pena, S. D. J. (2002). Characterization and comparative functional analysis in yeast of a *Schistosoma mansoni* Rho1 GTPase gene. *Molecular and Biochemical Parasitology*, *125*(1-2), 103–112. http://doi.org/10.1016/S0166-6851(02)00218-9

Savage, C., Das, P., Finelli, a L., Townsend, S. R., Sun, C. Y., Baird, S. E., & Padgett, R. W. (1996). *Caenorhabditis elegans* genes sma-2, sma-3, and sma-4 define a conserved family of transforming growth factor beta pathway components. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(2), 790–794. http://doi.org/10.1073/pnas.93.2.790

Schmidt, T., Karsunky, H., Fraß, B., Baum, W., Denzel, A., & Möröy, T. (2000). A novel protein (Fbf-1) that binds to CD95/APO-1/FAS and shows sequence similarity to trichohyalin and plectin. *Biochimica et Biophysica Acta - Gene Structure and Expression*, *1493*(1-2), 249–254. http://doi.org/10.1016/S0167-4781(00)00163-9

Schmit, F., Cremer, S., & Gaubatz, S. (2009). LIN54 is an essential core subunit of the DREAM/LINC complex that binds to the cdc2 promoter in a sequence-specific manner. *FEBS Journal*, *276*(19), 5703–5716. http://doi.org/10.1111/j.1742-4658.2009.07261.x

Schröder, H. M., Hoffmann, S. C., Hecker, M., Korff, T., & Ludwig, T. (2013). The tetraspanin network modulates MT1-MMP cell surface trafficking. *International Journal of Biochemistry and Cell Biology*, *45*(6), 1133–1144. http://doi.org/10.1016/j.biocel.2013.02.020

Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., *et al.* (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *Rna*, *22*(6), 839–851. http://doi.org/10.1261/rna.053959.115

Schwartz, M. (2004). Rho signalling at a glance. *Journal of Cell Science*, *117*(Pt 23), 5457–8. http://doi.org/10.1242/jcs.01582

Sclafani, R. A., & Holzen, T. M. (2007). Cell Cycle Regulation of DNA Replication. *Annual Reviews Genetics*, *41*, 237–280.

Sekelsky, J. J., Newfeld, S. J., Raftery, L. a., Chartoff, E. H., & Gelbart, W. M. (1995). Genetic characterization and cloning of mothers against dpp, a gene required for decapentaplegic function in *Drosophila melanogaster*. *Genetics*, *139*(3), 1347–1358.

Separovic, D., Breen, P., Joseph, N., Bielawski, J., Pierce, J. S., Van Buren, E., & Gudz, T. I. (2012). siRNA-mediated down-regulation of ceramide synthase 1 leads to apoptotic resistance in human head and neck squamous carcinoma cells after photodynamic therapy. *Anticancer Research*, *32*(7), 2479–2485.

Shah, C., VanGompel, M. J. W., Naeem, V., Chen, Y., Lee, T., Angeloni, N., *et al.* (2010). Widespread presence of human BOULE homologs among animals and conservation of their ancient reproductive function. *PLoS Genetics*, *6*(7), 1–16. http://doi.org/10.1371/journal.pgen.1001022

Shi, Y., Hata, a, Lo, R. S., Massagué, J., & Pavletich, N. P. (1997). A structural basis for mutational inactivation of the tumour suppressor Smad4. *Nature*, *388*(6637), 87–93. http://doi.org/10.1038/40431

Shi, Y., & Massagué, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell*, *113*(6), 685–700. http://doi.org/10.1016/S0092-8674(03)00432-X

Short, R., & Grossman, A. (1981). Conventional Giemsa and C-Banded Karyotypes of *Schistosoma mansoni* and *S. rodhaini*. *The Journal of Parasitology*, *67*(5), 661–671.

Si, Y., Liu, P., Li, P., & Brutnell, T. P. (2014). Model-based clustering for RNA-seq data. *Bioinformatics*, *30*(2), 197–205. http://doi.org/10.1093/bioinformatics/btt632

Smithers, S. R. (1962). Immunizing effect of irradiated cercariae of *Schistosoma mansoni* in rhesus monkeys. *Nature*, *194*, 1146–1147.

Sousa-Figueiredo, J. C., Gamboa, D., Pedro, J. M., Fançony, C., Langa, A. J., Soares Magalhães, R. J., *et al.* (2012). Epidemiology of Malaria, Schistosomiasis, Geohelminths, Anemia and Malnutrition in the Context of a Demographic Surveillance System in Northern Angola. *PLoS ONE*, *7*(4), e33189. http://doi.org/10.1371/journal.pone.0033189

Stothard, J. R., Sousa-Figueiredo, J. C., Betson, M., Bustinduy, A., & Reinhard-Rupp, J. (2013). Schistosomiasis in African infants and preschool children: Let them now be treated! *Trends in Parasitology*, *29*(4), 197–205. http://doi.org/10.1016/j.pt.2013.02.001

Sun, J., Li, C., & Wang, S. (2015). The Up-Regulation of Ribosomal Proteins Further Regulates Protein Expression Profile in Female *Schistosoma japonicum* after Pairing. *Plos One*, *10*(6), e0129626. http://doi.org/10.1371/journal.pone.0129626

Sun, J., Wang, S.-W., Li, C., Hu, W., Ren, Y.-J., & Wang, J.-Q. (2014). Transcriptome profilings of female *Schistosoma japonicum* reveal significant differential expression of genes after pairing. *Parasitology Research*, *113*(3), 881–92. http://doi.org/10.1007/s00436-013-3719-2

Sundaram, M. (2006). RTK/Ras/MAPK signaling. *WormBook*, 1–19. http://doi.org/10.1895/wormbook.1.80.1

Susin, S. a, Lorenzo, H. K., Zamzami, N., Marzo, I., Snow, B. E., Brothers, G. M., *et al.* (1999). Molecular characterization of mitochondrial apoptosis-inducing factor. *Nature*, *397*(6718), 441–446. http://doi.org/10.1038/17135

Swain, M. T., Larkin, D. M., Caffrey, C. R., Davies, S. J., Loukas, A., Skelly, P. J., & Hoffmann, K. F. (2011). *Schistosoma* comparative genomics: integrating genome structure, parasite biology and anthelmintic discovery. *Trends in Parasitology*, *27*(12), 555–64. http://doi.org/10.1016/j.pt.2011.09.003

Tada, M., Kohno, M., & Niwano, Y. (2014). Alleviation effect of arbutin on oxidative stress generated through tyrosinase reaction with l -tyrosine and l -DOPA. *BMC Biochemistry*, *15*(1), 1–7. http://doi.org/10.1186/1471-2091-15-23

Takino, T., Miyamori, H., Kawaguchi, N., Uekita, T., Seiki, M., & Sato, H. (2003). Tetraspanin CD63 promotes targeting and lysosomal proteolysis of membrane-type 1 matrix metalloproteinase. *Biochemical and Biophysical Research Communications*, *304*(1), 160–166. http://doi.org/10.1016/S0006-291X(03)00544-8

Tang, H., Ming, Z., Liu, R., Xiong, T., Grevelding, C. G., & Dong, H. (2013). Development of Adult Worms and Granulomatous Pathology Are Collectively Regulated by T- and B-Cells in Mice Infected with Schistosoma japonicum, *8*(1), 1–8. http://doi.org/10.1371/journal.pone.0054432

Tasaki, J., Shibata, N., Nishimura, O., Itomi, K., Tabata, Y., Son, F., & Umesono, Y. (2011). ERK signaling controls blastema cell differentiation during planarian regeneration. *Development (Cambridge, England)*, *138*(12), 2417–2427. http://doi.org/10.1242/dev.060764

Tebeje, B. M., Harvie, M., You, H., Loukas, A., & McManus, D. P. (2016). Schistosomiasis vaccines: where do we stand? *Parasites and Vectors*, *9*(1), 1–15. http://doi.org/10.1186/s13071-016-1799-4

Termini, C. M., & Gillette, J. M. (2017). Tetraspanins Function as Regulators of Cellular Signaling. *Frontiers in Cell and Developmental Biology*, *5*(April), 1–14. http://doi.org/10.3389/fcell.2017.00034

Tirodkar, T. S., & Voelkel-Johnson, C. (2012). Sphingolipids in apoptosis. *Experimental Oncology*, *34*(3), 231–242.

Toczyski, D. P., Galgoczy, D. J., & Hartwell, L. H. (1997). Control adaptation to the yeast DNA damage checkpoint. *Cell*, *90*(1), 1097–1106.

Tran, M. H., Freitas, T. C., Cooper, L., Gaze, S., Gatton, M. L., Jones, M. K., *et al.* (2010). Suppression of mRNAs encoding tegument tetraspanins from *Schistosoma mansoni* results in impaired tegument turnover. *PLoS Pathogens*, *6*(4), e1000840. http://doi.org/10.1371/journal.ppat.1000840

Tucker, M. S., Karunaratne, L. B., Lewis, F. a., Freitas, T. C., & Liang, Y. S. (2013). *Schistosomiasis*. *Current Protocols in Immunology*. http://doi.org/10.1002/0471142735.im1901s103

Tugues, S., Honjo, S., König, C., Padhan, N., Kroon, J., Gualandi, L., *et al.* (2013). Tetraspanin CD63 promotes vascular endothelial growth factor receptor 2-β1 integrin complex formation, thereby regulating activation and downstream signaling in endothelial cells in Vitro and in Vivo. *Journal of Biological Chemistry*, *288*(26), 19060–19071. http://doi.org/10.1074/jbc.M113.468199

Van De Weerdt, B. C. M., & Medema, R. H. (2006). Polo-like kinases: A team in control of the division. *Cell Cycle*, *5*(8), 853–864. http://doi.org/10.4161/cc.5.8.2692

Vanderhaeghen, P., & Cheng, H.-J. (2010). Guidance Molecules in Axon Pruning and Cell Death. *Cold Spring Harbour Perspectives in Biology*, *2*, a001859.

Vanderstraete, M., Gouignard, N., Cailliau, K., Morel, M., Hahnel, S., Leutner, S., *et al.* (2014). Venus kinase receptors control reproduction in the platyhelminth parasite *Schistosoma mansoni*. *PLoS Pathogens*, *10*(5), e1004138. http://doi.org/10.1371/journal.ppat.1004138

Vanhoutteghem, A., Messiaen, S., Hervé, F., Delhomme, B., Moison, D., Petit, J.-M., *et al.* (2014). The zinc-finger protein basonuclin 2 is required for proper mitotic arrest, prevention of premature meiotic initiation and meiotic progression in mouse male germ cells. *Development (Cambridge, England)*, (2014), 4298–4310. http://doi.org/10.1242/dev.112888

Verjovski-Almeida, S., DeMarco, R., Martins, E. a L., Guimarães, P. E. M., Ojopi, E. P. B., Paquola, A. C. M., *et al.* (2003). Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nature Genetics*, *35*(2), 148–57. http://doi.org/10.1038/ng1237

Verk, M. C. Van, Hickman, R., Pieterse, M. J., & Van Wees, S. C. M. (2013). RNA-Seq : revelation of the messengers. *Trends in Plant Sciences*, *18*(4), 175–179. http://doi.org/10.1016/j.tplants.2013.02.001

Vicogne, J., Cailliau, K., Tulasne, D., Browaeys, E., Yan, Y. T., Fafeur, V., *et al.* (2004). Conservation of epidermal growth factor receptor function in the human parasitic helminth *Schistosoma mansoni*. *The Journal of Biological Chemistry*, *279*(36), 37407–37414. http://doi.org/10.1074/jbc.M313738200

Vicogne, J., Pin, J. P., Lardans, V., Capron, M., Noël, C., & Dissous, C. (2003). An unusual receptor tyrosine kinase of *Schistosoma mansoni* contains a Venus Flytrap module. *Molecular and Biochemical Parasitology*, *126*(1), 51–62. http://doi.org/10.1016/S0166-6851(02)00249-9

Vimieiro, A. C. S., Araújo, N., Katz, N., Kusel, J. R., & Coelho, P. M. Z. (2013). Schistogram changes after administration of antischistosomal drugs in mice at the early phase of *Schistosoma mansoni* infection. *Memorias Do Instituto Oswaldo Cruz*, *108*(7), 881–886. http://doi.org/10.1590/0074-0276130135

Virgili, G., Frank, F., Feoktistova, K., Sawicki, M., Sonenberg, N., Fraser, C. S., & Nagar, B. (2013). Structural analysis of the DAP5 MIF4G domain and its interaction with eIF4A. *Structure*, *21*(4), 517–527. http://doi.org/10.1016/j.str.2013.01.015

Vogler, G., & Urban, J. (2008). The transcription factor Zfh1 is involved in the regulation of neuropeptide expression and growth of larval neuromuscular junctions in *Drosophila melanogaster*. *Developmental Biology*, *319*(1), 78–85. http://doi.org/10.1016/j.ydbio.2008.04.008

Volfovsky, N., Haas, B. J., & Salzberg, S. L. (2003). Computational discovery of internal micro-exons. *Genome Research*, *13*(6 A), 1216–1221. http://doi.org/10.1101/gr.677503

Waisberg, M., Lobo, F. P., Cerqueira, G. C., Passos, L. K. J., Carvalho, O. S., Franco, G. R., & El-Sayed, N. M. (2007). Microarray analysis of gene expression induced by sexual contact in *Schistosoma mansoni*. *BMC Genomics*, *8*, 181. http://doi.org/10.1186/1471-2164-8-181

Wallar, B. J., & Alberts, A. S. (2003). The formins: Active scaffolds that remodel the cytoskeleton. *Trends in Cell Biology*, *13*(8), 435–446. http://doi.org/10.1016/S0962-8924(03)00153-3

Wang, J., & Collins, J. J. (2016). Identification of new markers for the *Schistosoma mansoni* vitelline lineage. *International Journal for Parasitology*, *46*(7), 405–410. http://doi.org/10.1016/j.ijpara.2016.03.004

Wang, J., Wang, S., Liu, X., Xu, B., Chai, R., Zhou, P., *et al.* (2015). Intake of Erythrocytes Required for Reproductive Development of Female Schistosoma japonicum. *PLOS ONE*, *10*(5), e0126822. http://doi.org/10.1371/journal.pone.0126822

Wang, J., Yu, Y., Shen, H., Qing, T., Zheng, Y., Li, Q., *et al.* (2017). Dynamic transcriptomes identify biogenic amines and insect-like hormonal regulation for mediating reproduction in *Schistosoma japonicum*. *Nature Communications*, *8*(May 2016), 14693. http://doi.org/10.1038/ncomms14693

Wang, W., Li, H.-J., Qu, G.-L., Xing, Y.-T., Yang, Z.-K., Dai, J.-R., & Liang, Y.-S. (2014). Is there a reduced sensitivity of dihydroartemisinin against praziquantel-resistant *Schistosoma japonicum*? *Parasitology Research*, *113*(1), 223–228. http://doi.org/10.1007/s00436-013-3647-1

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, *10*(1), 57–63. http://doi.org/10.1038/nrg2484

Warren, K. S., Mahmoud, A. A. F., Cummings, P., Murphy, D. J., & Houser, H. B. (1974). *Schistosomiasis mansoni* in Yemeni in California: duration of infection, presence of disease, therapeutic management. *American Journal of Tropical Medicine and Hygiene*, *23*, 902–909.

Watchorn, T. M., Waddell, I., Dowidar, N., & Ross, J. a. (2001). Proteolysis-inducing factor regulates hepatic gene expression via the transcription factors NF-(kappa)B and STAT3. *The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *15*(3), 562–564. http://doi.org/10.1096/fj.00

Wehrens, R., & Buydens, L. M. C. (2007). Self- and super-organizing maps in R: The kohonen package. *Journal of Statistical Software*, *21*(5), 1–19. http://doi.org/10.18637/jss.v021.i05

Westphal, D., Dewson, G., Czabotar, P. E., & Kluck, R. M. (2011). Molecular biology of Bax and Bak activation and action. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1813*(4), 521–531. http://doi.org/10.1016/j.bbamcr.2010.12.019

Williamson, S. M., Robertson, A. P., Brown, L., Williams, T., Woods, D. J., Martin, R. J., *et al.* (2009). The Nicotinic Acetylcholine Receptors of the Parasitic Nematode *Ascaris suum*: Formation of Two Distinct Drug Targets by Varying the Relative Expression Levels of Two Subunits. *PLoS Pathogens*, *5*(7), e1000517. http://doi.org/10.1371/journal.ppat.1000517

World Health Organisation. (2015). Schistosomiasis - Strategy. Retrieved from http://www.who.int/schistosomiasis/strategy/en/

World Health Organisation. (2016). Global Health Estimates 2015: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2015. Retrieved from: http://www.who.int/healthinfo/global_burden_disease/GlobalDALY_method_2000_2016.pdf?ua=1

World Health Organisation. (2018). Schistosomiasis fact sheet. Retrieved May 2, 2017, from http://www.who.int/mediacentre/factsheets/fs115/en/

Wu, J. W., Hu, M., Chai, J., Seoane, J., Huse, M., Li, C., *et al.* (2001). Crystal structure of a phosphorylated Smad2: Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-β signaling. *Molecular Cell*, *8*(6), 1277–1289. http://doi.org/10.1016/S1097-2765(01)00421-X

Wu, X., Fu, Y., Knott, G., Lu, J., Di Cristo, G., & Huang, Z. J. (2012). GABA Signaling Promotes Synapse Elimination and Axon Pruning in Developing Cortical Inhibitory Interneurons. *Journal of Neuroscience*, *32*(1), 331–343. http://doi.org/10.1523/JNEUROSCI.3189-11.2012

Wynn, T. A. (2015). Type 2 cytokines: mechanisms and therapeutic strategies. *Nature Reviews Immunology*, *15*(5), 271–282. http://doi.org/10.1038/nri3831

Xia, C., Ma, W., Stafford, L. J., Marcus, S., Xiong, W. C., & Liu, M. (2001). Regulation of the p21-activated kinase (PAK) by a human Gbeta -like WD-repeat protein, hPIP1. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(11), 6174–6179. http://doi.org/10.1073/pnas.101137298

Xiao, S. H., Catto, B. a, & Webster, L. T. (1985). Effects of praziquantel on different developmental stages of *Schistosoma mansoni in vitro* and *in vivo*. *The Journal of Infectious Diseases*, *151*(6), 1130–7. Retrieved from http://jid.oxfordjournals.org/content/151/6/1130.short\nhttp://www.ncbi.nlm.nih.gov/pubmed/3998507

Yakovlev, a G., Ota, K., Wang, G., Movsesyan, V., Bao, W. L., Yoshihara, K., & Faden, a I. (2001). Differential expression of apoptotic protease-activating factor-1 and caspase-3 genes and susceptibility to apoptosis during brain development and after traumatic brain injury. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *21*(19), 7439–7446.

Yan, Y., Tulasne, D., Browaeys, E., Cailliau, K., Khayath, N., Pierce, R. J., *et al.* (2007). Molecular cloning and characterisation of SmSLK, a novel Ste20-like kinase in Schistosoma mansoni. *International Journal for Parasitology*, *37*(14), 1539–1550. http://doi.org/10.1016/j.ijpara.2007.06.001

Yuan, K., & O'Farrell, P. H. (2015). Cyclin B3 Is a Mitotic Cyclin that Promotes the Metaphase-Anaphase Transition. *Current Biology*, *25*(6), 811–816. http://doi.org/10.1016/j.cub.2015.01.053

Zamanian, M., Kimber, M. J., McVeigh, P., Carlson, S. a, Maule, A. G., & Day, T. a. (2011). The repertoire of G protein-coupled receptors in the human parasite *Schistosoma mansoni* and the model organism *Schmidtea mediterranea*. *BMC Genomics*, *12*(1), 596. http://doi.org/10.1186/1471-2164-12-596

Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE*, *9*(1), e78644. http://doi.org/10.1371/journal.pone.0078644

Zhou, Z. D., Kumari, U., Xiao, Z. C., & Tan, E. K. (2010). Notch as a molecular switch in neural stem cells. *IUBMB Life*, *62*(8), 618–623. http://doi.org/10.1002/iub.362

Zhu, Q., Jin, J.-F., Shan, X.-H., Liu, C.-P., Mao, X.-D., Xu, K.-F., & Liu, C. (2008). Chronic activation of neutral ceramidase protects β -cells against cytokine-induced apoptosis. *Acta Pharmacologica Sinica*, *29*(5), 593–599. http://doi.org/10.1111/j.1745-7254.2008.00781.x

Zimmerman, S. G., Peters, N. C., Altaras, A. E., & Berg, C. A. (2013). Dot blot to estimate RNA-probe concentration and DIG-labeling efficiency. *Nature Protocols*, *8*, 2158–2179.