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.

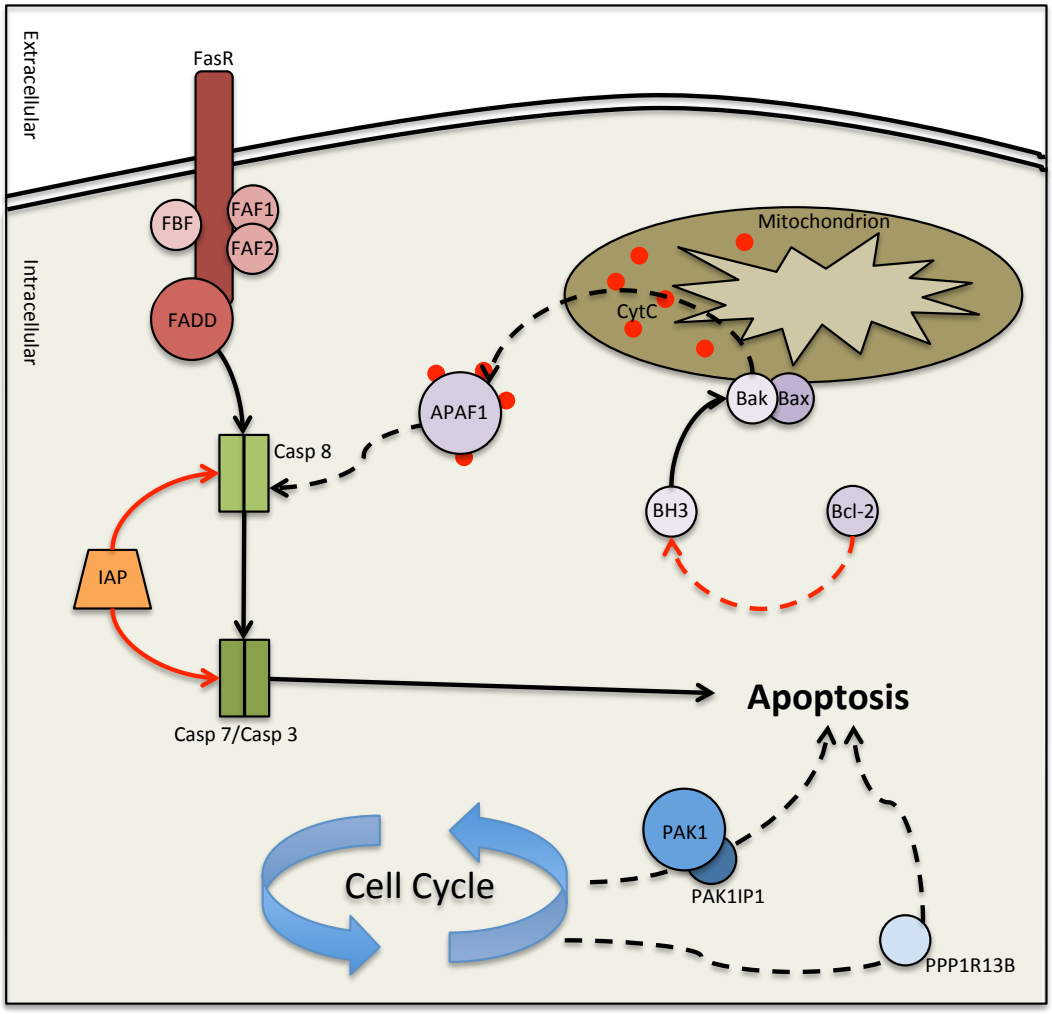

Figure 4.1: Apoptosis in *S. mansoni*. Black arrow – activation; red arrow – inhibition; dashed arrow – indirect interaction. The *S. mansoni* apoptosis pathway is similar to that of mammals, including a Fas receptor homologue and associated proteins, a caspase cascade, involvement of the mitochondria 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 ($p < 2.2E\text{-}16$).

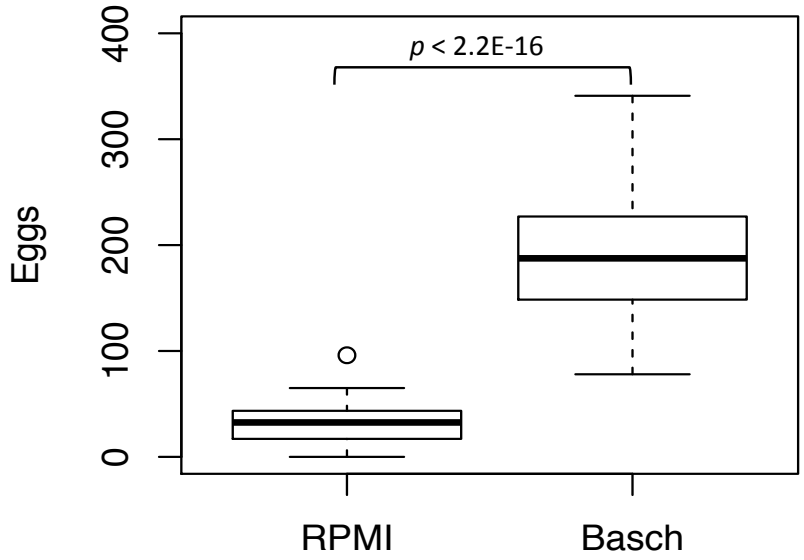

Figure 4.2: Significantly more eggs were laid in Basch media compared to RPMI. Box plot of eggs laid per female in RPMI based medium and Basch medium after 28 days in *in vitro* culture. Both groups contained 36 paired females. Using a Welch two sample *t*-test, a significant difference between the numbers 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% $p = 3.3E\text{-}04$.

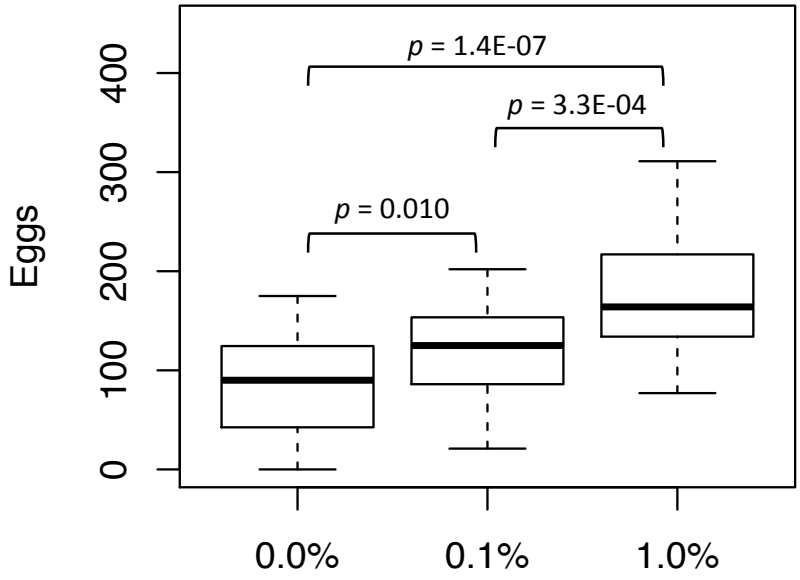

Figure 4.3: Significantly more eggs were laid in media enriched with 1.0% red blood cells compared to 0.0% and 0.1%. Box plot of eggs laid per female in Basch medium with varying amounts of red blood cells (by volume) after 28 days in *in vitro* culture. All groups contained 24 paired females. Using 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 apopotosis 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).

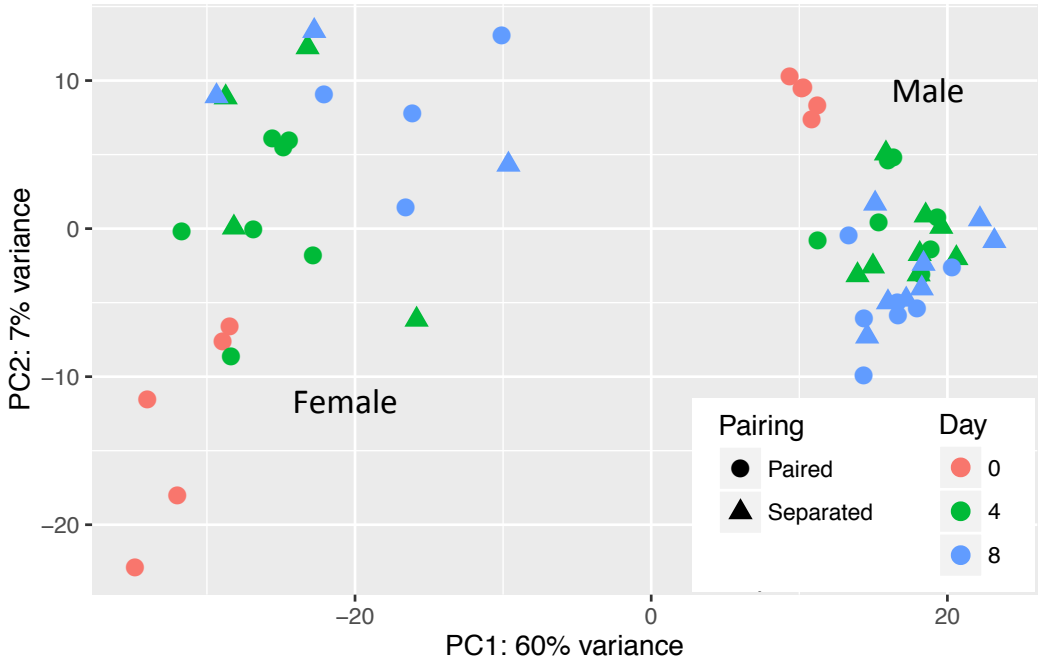

Figure 4.4: Samples clustered well by gender, but separated less well by time *in vitro* and pairing status. Principal component analysis (PCA) plot of RNA-Seq samples at 0, 4 and 8 days after perfusion. The RNA-seq samples separated into female (left) and male (right) clusters but samples separate poorly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

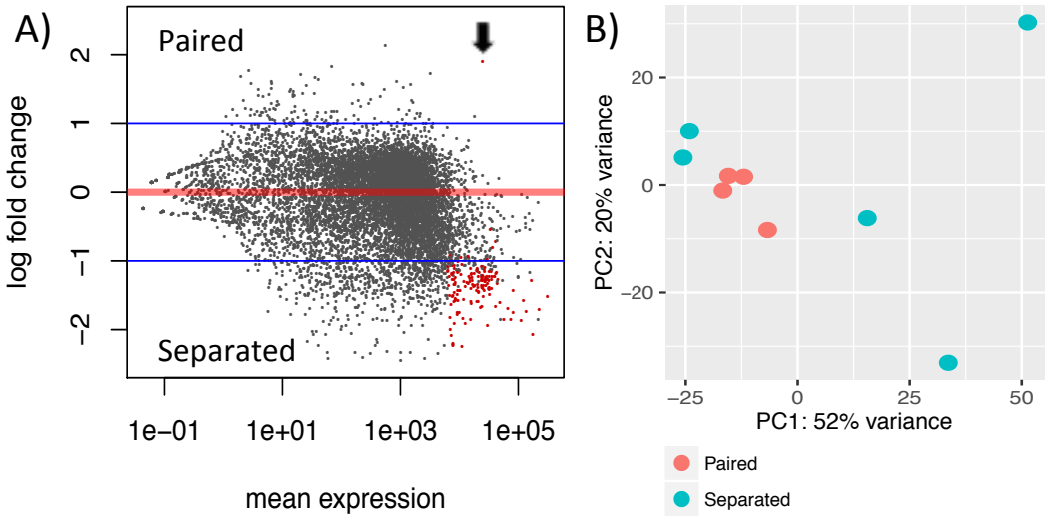

Figure 4.5: Most differentially expressed genes were up-regulated in separated females. Pair-wise comparison of paired and unpaired females at day 8 after perfusion. A) The Log ratio against mean (MA) plot showed only one differentially expressed gene (marked with arrow) was up-regulated in paired females, 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.

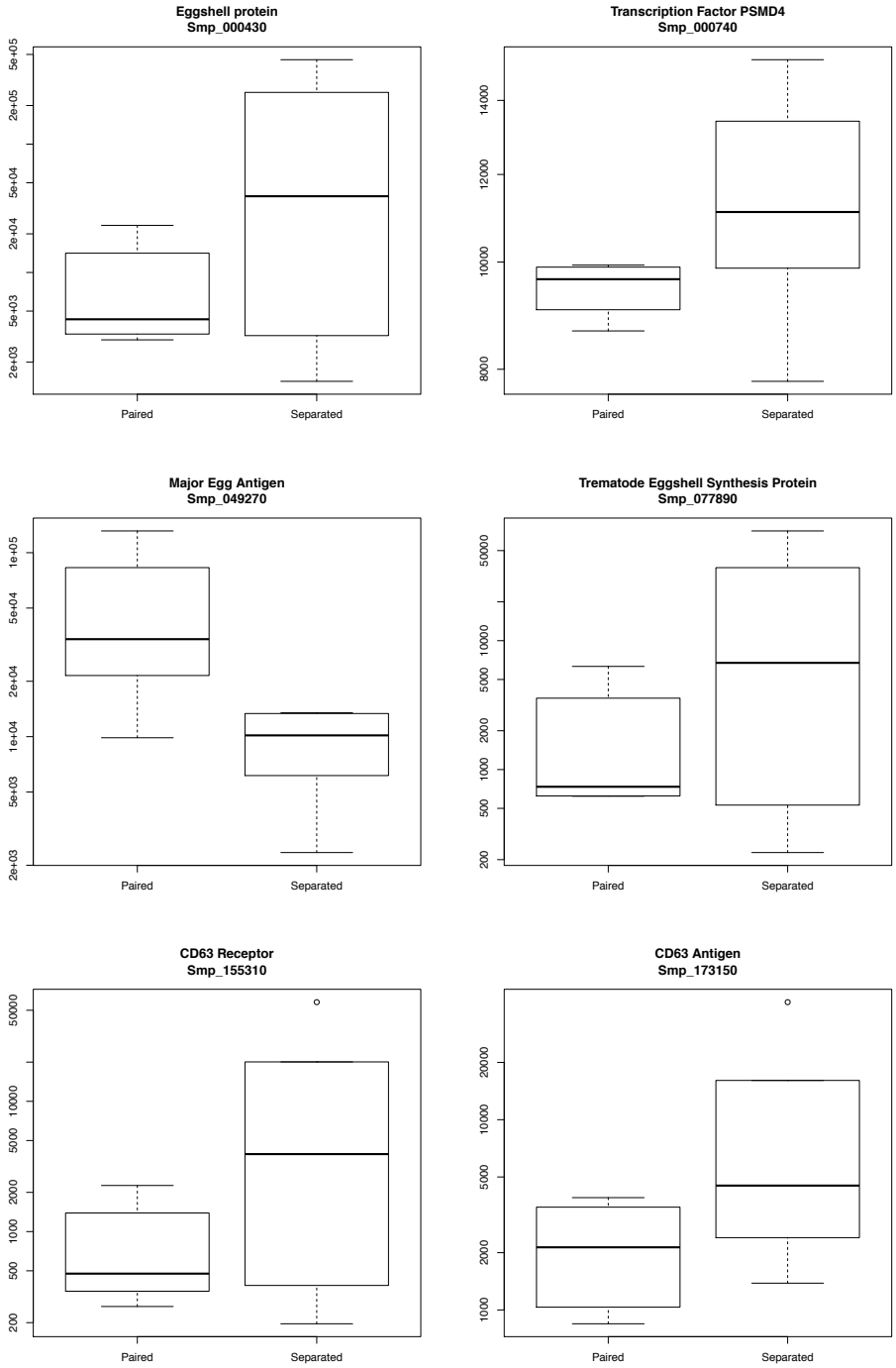

Figure 4.6: Biological replicates of separated females have great variability in gene expression. This is a box whisker plots of normalised read counts, a measure for absolute gene expression, for all RNA-seq samples of paired and separated females at day 8 after perfusion. In all genes, with the exception 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 tthe MA plot (Figure 4.7 A) shows, there were only small fold changes of gene expression between the two groups.

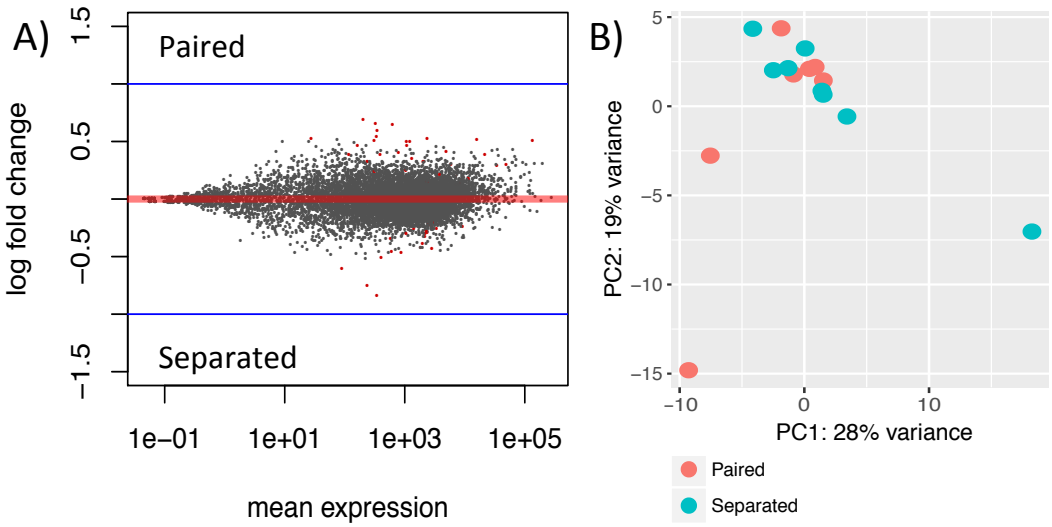

Figure 4.7: The low number of differentially expressed genes between separated and paired males has small fold changes. Pair-wise comparison of paired and separated males at day 8 after perfusion. A) The log ratio vs mean (MA) plot showed that none of the differentially expressed genes have a fold change 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 an 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 assembley 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 females 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.

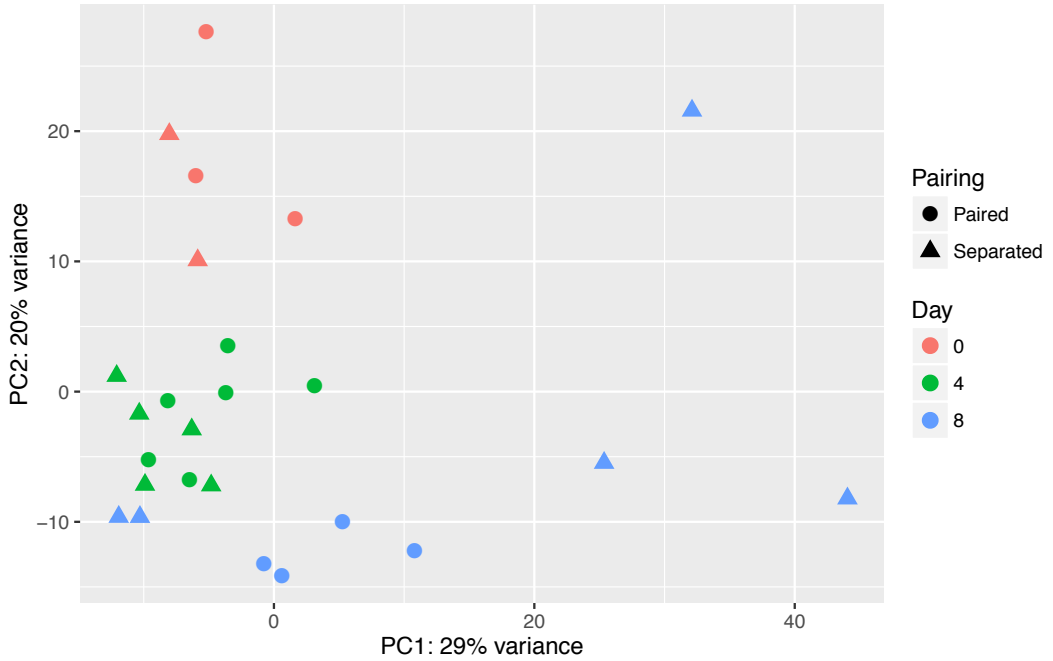

Figure 4.8: In this Principal component analysis plot female samples separate by time, but not pairing status. This is a PCA plot of a time series analysis of female samples at 0, 4 and 8 days after perfusion. The PCA shows high variance between the separated females at day 8 after perfusion, but paired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

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).

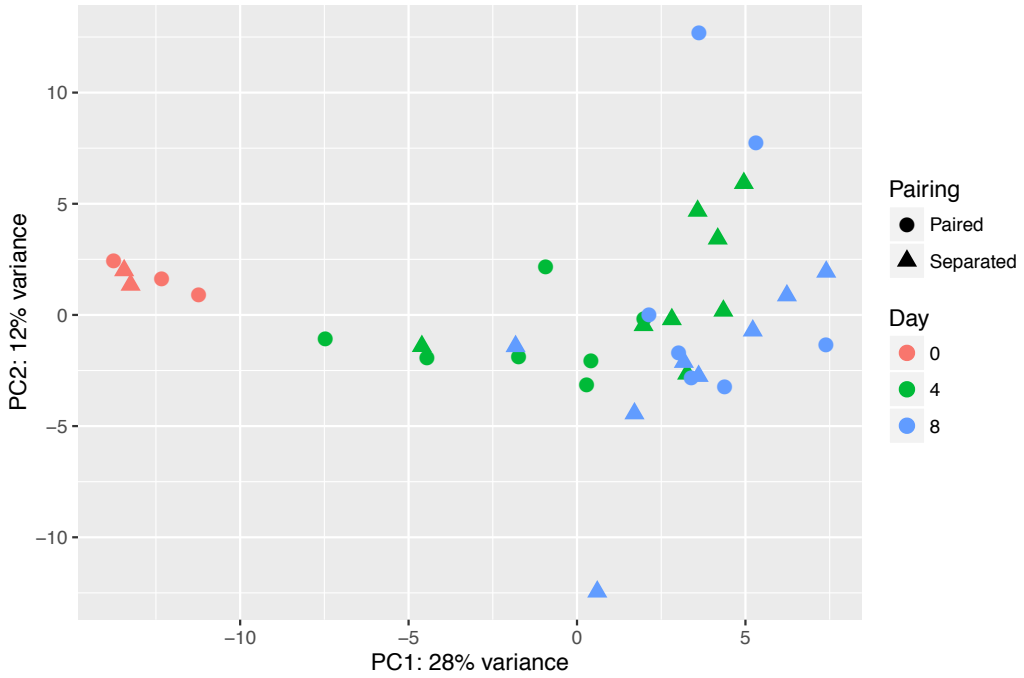

Figure 4.9: In this Principal component analysis male RNA-seq samples do not separate by pairing status. This is a PCA plot of the time series analysis of male samples at 0, 4 and 8 days after perfusion. The PCA shows more variance between samples at day 4 and especially day 8 when compared to the day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

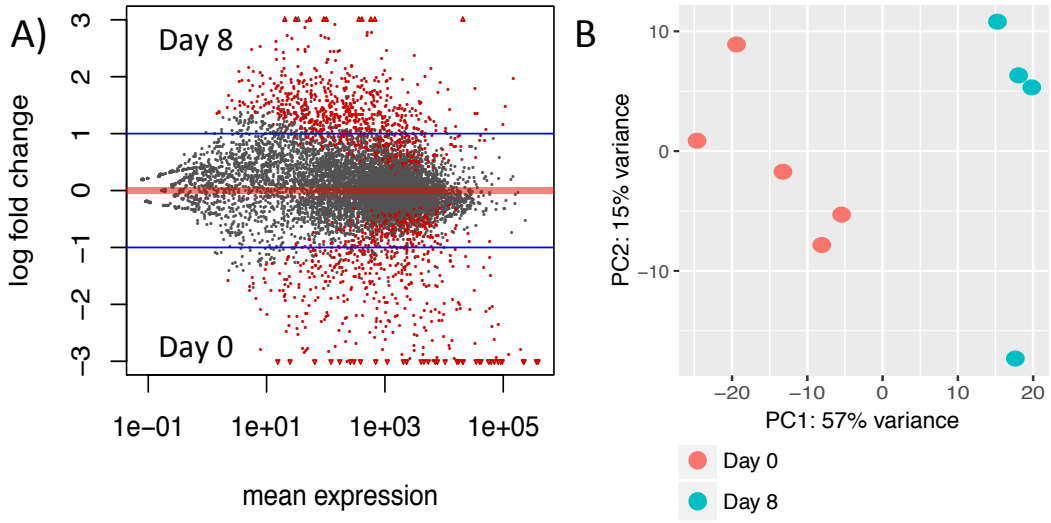

Figure 4.10: *In vitro* culture of *Schistosoma mansoni* induces a great magnitude of changes in gene expression in female worms. Pair-wise comparison of female worms at day 0 and day 8 after perfusion. A) A plot of the log ratio against the mean expression showed a large numbers of genes significantly 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 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).

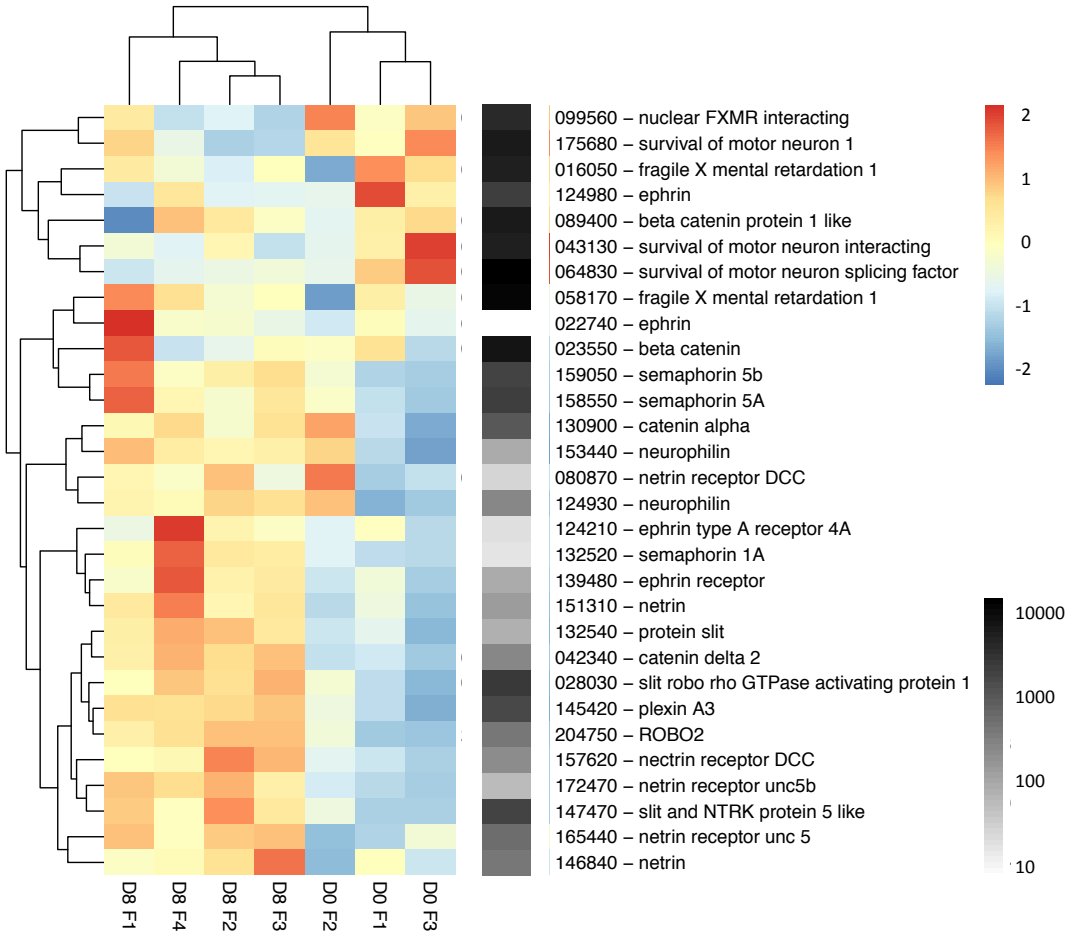

Figure 4.11: Many genes involved in axon development became up-regulated *in vitro*. Heatmap of neurogenesis related genes in female worms at day 0 (three replicates: D0 F1-3) and day 8 (four replicates: D8 F1-4). Colours represent a Z score of log transformed normalised read counts scaled by row, giving 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).

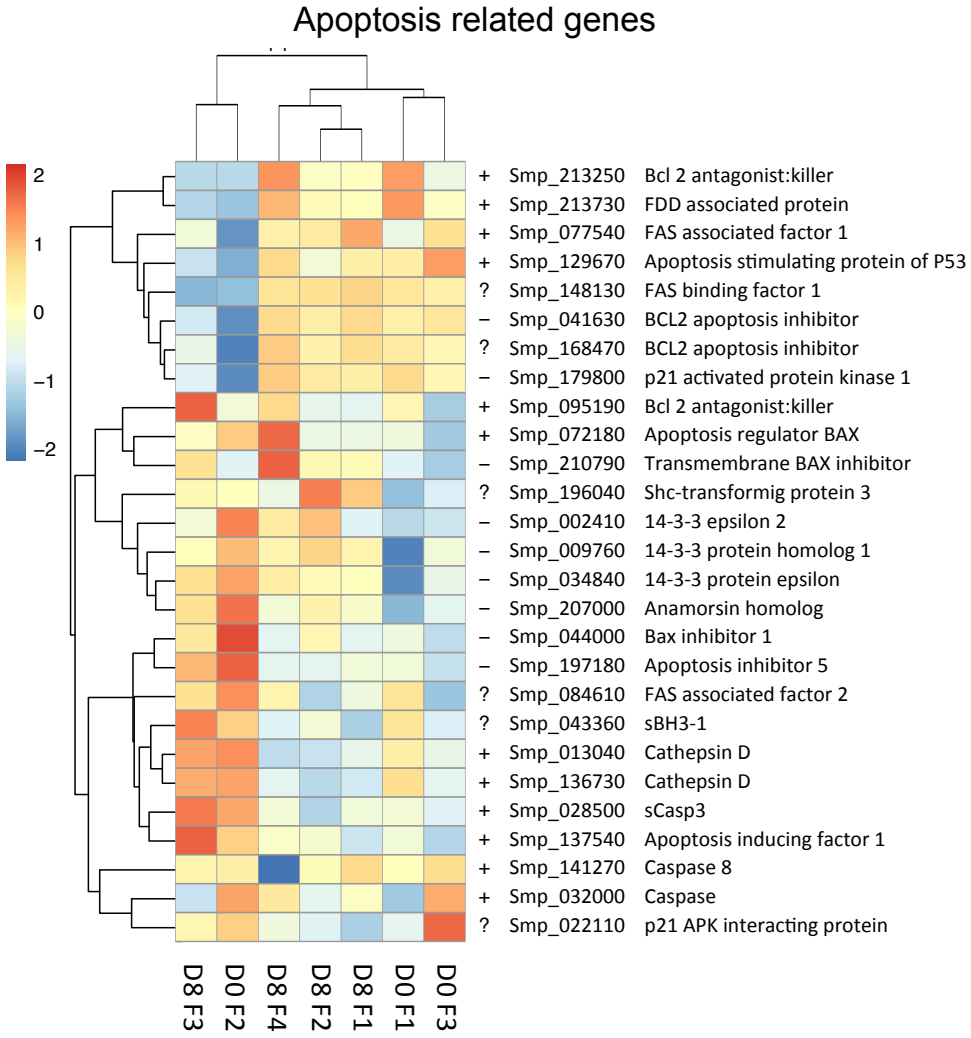

Figure 4.12: No clear trend was found in the expression of pro- and anti-apoptotic genes. Heatmap of apoptosis related genes in females *Schistosoma mansoni* at day 0 (three replicates: D0 F1-3) and day 8 (four replicates: D8 F1-4). Colours represent a Z score of log transformed normalised read counts 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 signaling pathway	4	11	3.08E-04
smm04142	Lysosome	6	47	3.41E-03
smm04310	Wnt signaling pathway	5	35	4.61E-03
smm04330	Notch signaling 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 signaling pathway	3	20	2.36E-02
smm02010	ABC transporters	2	8	2.55E-02
smm04068	FoxO signaling pathway	4	38	2.85E-02
smm04340	Hedgehog signaling 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).

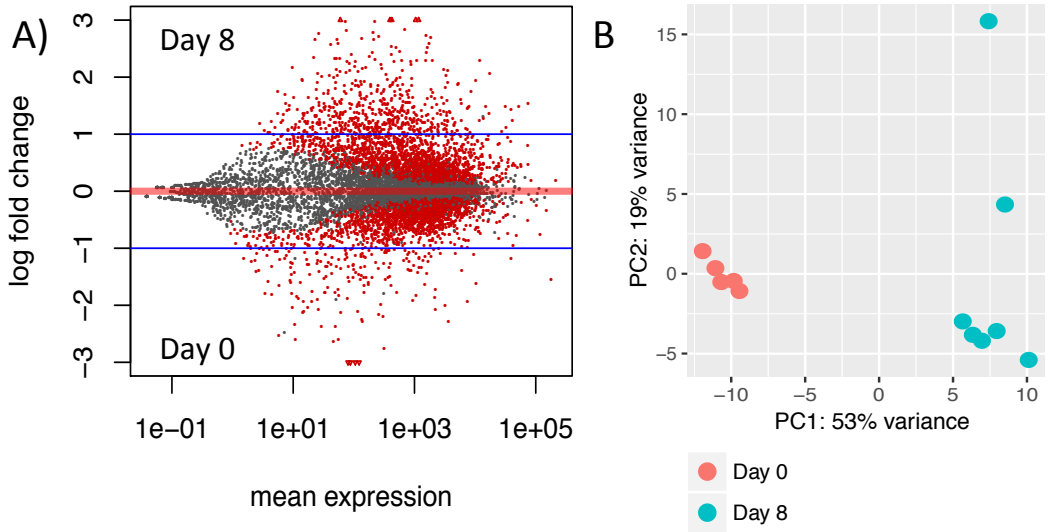

Figure 4.13: *In vitro* culture induced large changes in gene expression in male worms. Pair-wise comparison of male worms at day 0 and day 8 after perfusion. A) The log ratio against mean (MA) plot shows large numbers of genes significantly up-regulated (red) in both groups with large fold changes. 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.

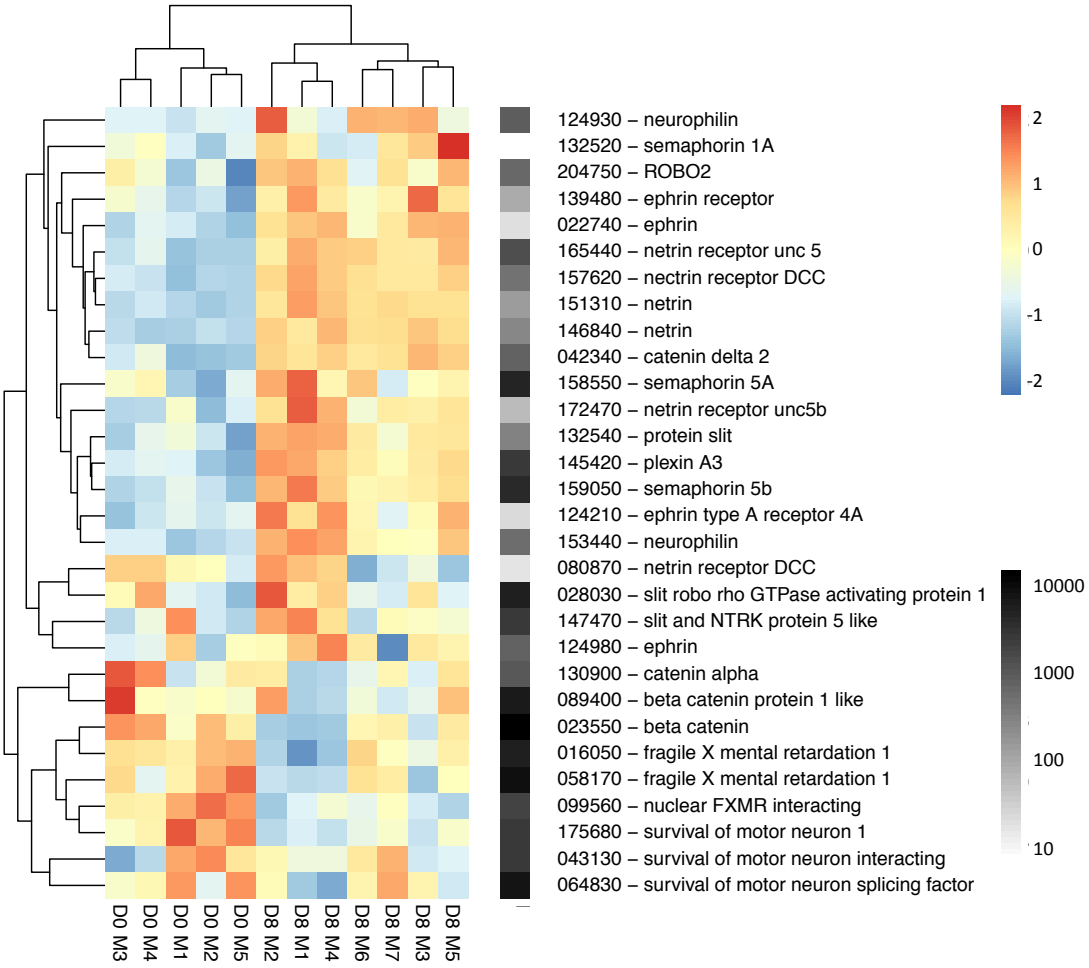

Figure 4.14: Axon proning-related genes became up-regulated *in vitro*. Heatmap of neurogenesis related genes in males worms at day 0 (five replicates: D0 M1-5) and day 8 (seven replicates: D8 M1-7). Colours represent a Z score of log transformed normalised read counts scaled by row, giving relative 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 apoptosic activity could not be demonstrated for female worms.

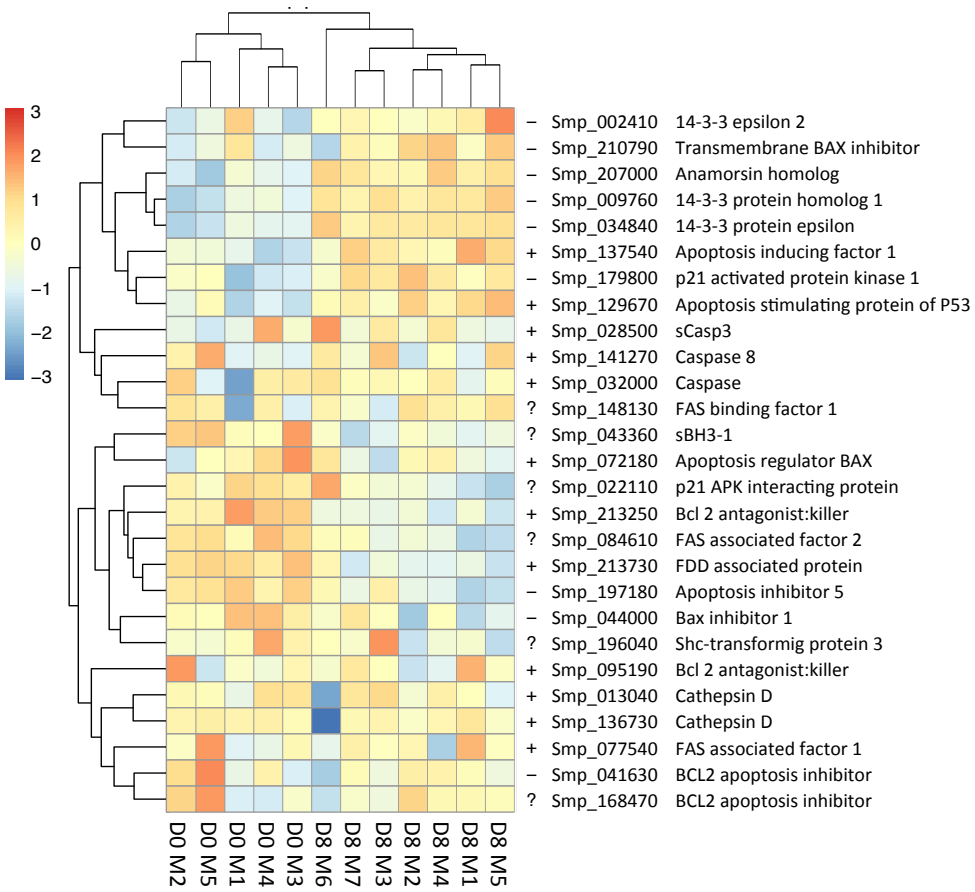

Figure 4.15: Expression of apoptosis-related genes does not show clear trends before and after *in vitro* cultrue. No Heatmap of apoptosis related genes in males worms at day 0 (five replicates: D0 M1-5) and day 8 (seven replicates: D8 M1-7). Colours represent a Z score of log transformed normalised 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 signaling 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 tge 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, sphingo lipid 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 paried 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 a 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.