Chapter 5

Exploration of the *S. mansoni* gonad transcriptome

5.1 Introduction

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

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

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

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

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

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

The signalling network that regulates female fertility (see Figure 5.1) has been studied intensively (Beckmann *et al.*, 2010; Knobloch *et al.*, 2007; LoVerde *et al.*, 2009; Andrade *et al.*, 2014) and is described in Chapter 1.5. It comprises over 40 known genes, many of which have been characterised using *in silico* approaches, RNAi, *in situ* hybridisation, yeast-two-hybrid screens or using inhibitors such as herbimycin (Beckmann *et al.*, 2010; Knobloch *et al.*, 2006).
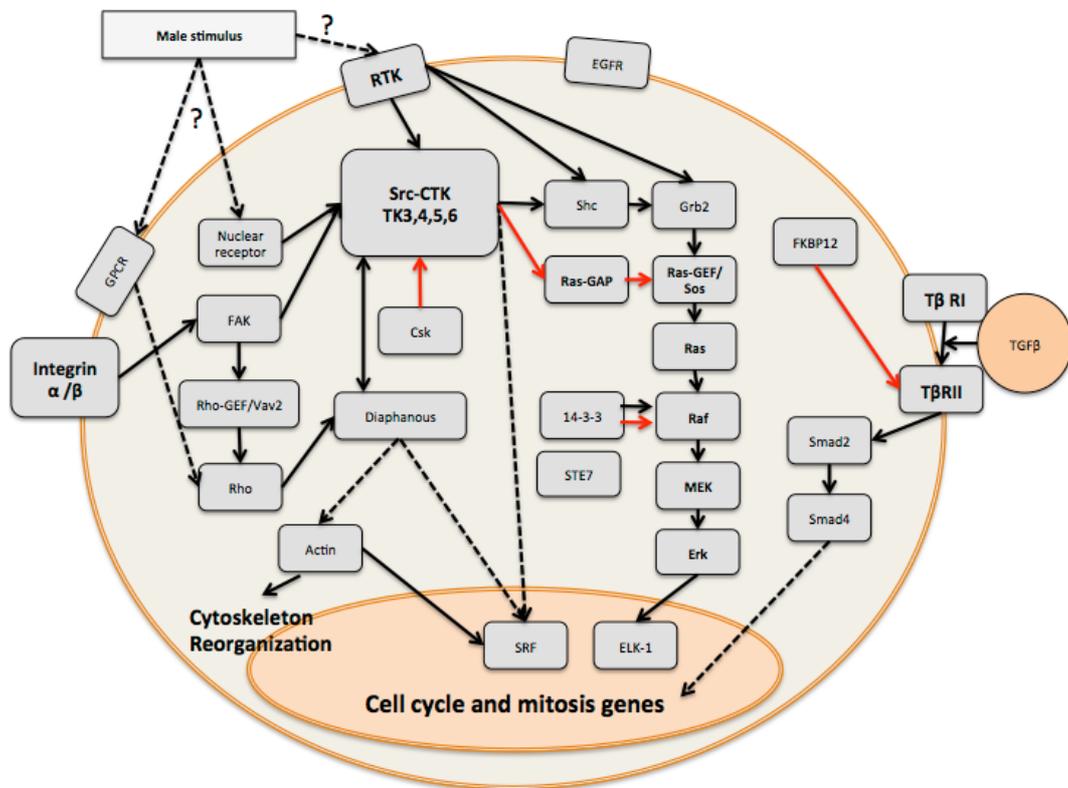
Figure 5.1: Representation of the MAPK, Rho and TGF-β signalling pathway inside a maturing oocyte from cell surface receptor to transcription regulators in the nucleus. Adapted from Knobloch *et al.* (2007). Dotted arrows indicate indirect interactions, solid arrows indicate direct interaction. Red arrows indicate inhibition and black arrows activation of the target.

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

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

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

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

5.2 Results

5.2.1 Sequencing and sample clustering

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

During the initial analysis, one sample (single sex ovary 3, sO3) was identified which had apparently been mishandled and appeared to be whole male RNA instead of an ovary sample. This became apparent on a PCA plot (Figure 5.2) and was confirmed by examining expression of known male specific transcripts (Fitzpatrick *et al.*, 2005) (see Figure 5.3) such as those coding for the tegument allergen like (TAL) proteins (Smp_045200, Smp_086480 and Smp_195090) and calpain (Smp_137410); the sample was excluded from subsequent analysis.
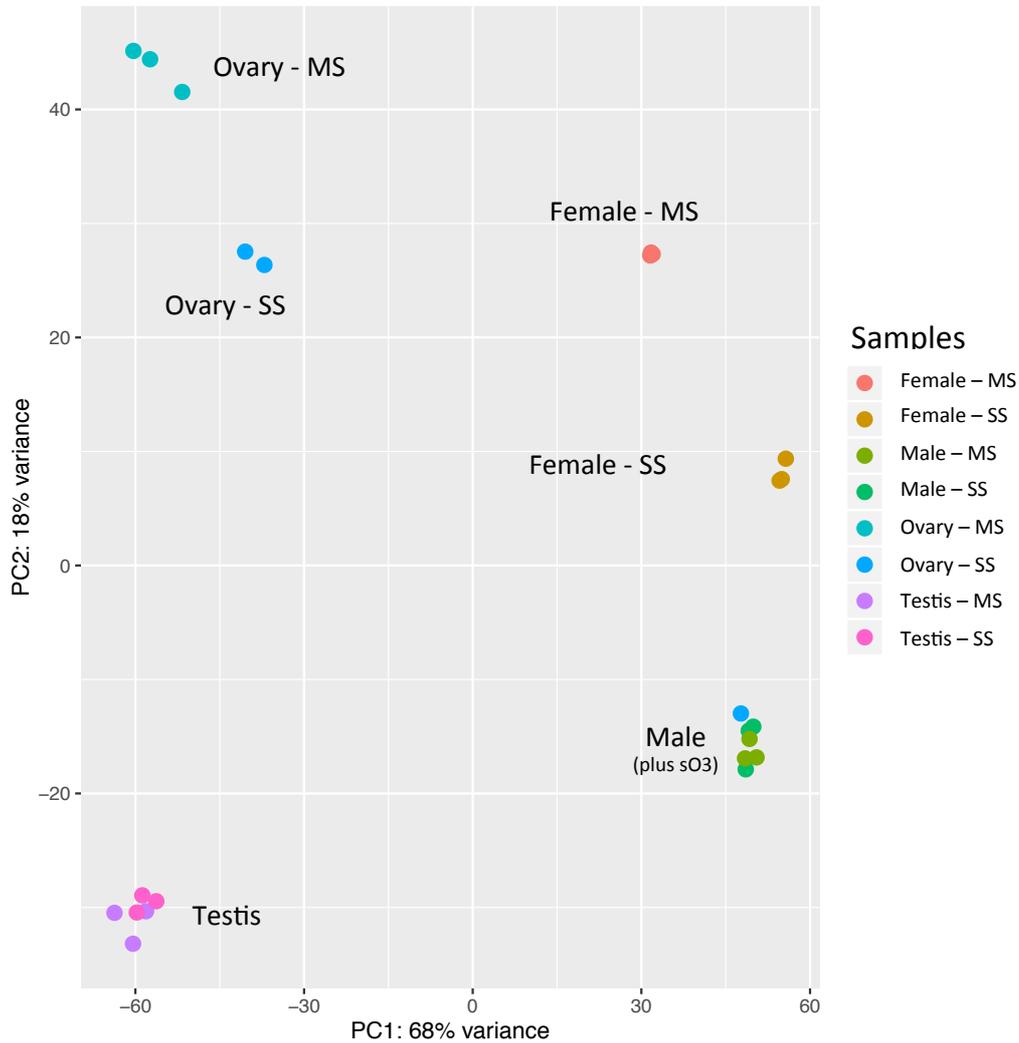
Figure 5.2: Samples cluster by type, male samples from mixed sex and single sex infection cluster together. The samples are either from single sex (SS) or mixed sex (MS) infections. In this principal component analysis of RNA-Seq samples, 86% of the variation among all samples is explained by the two most important principal components. The PCA plot is based on the 500 most variable genes in the data set, which contains three replicates per condition. Notably one of the SS ovary samples (sO3) clustered with the whole male samples. The three Female MS samples overlap tightly.
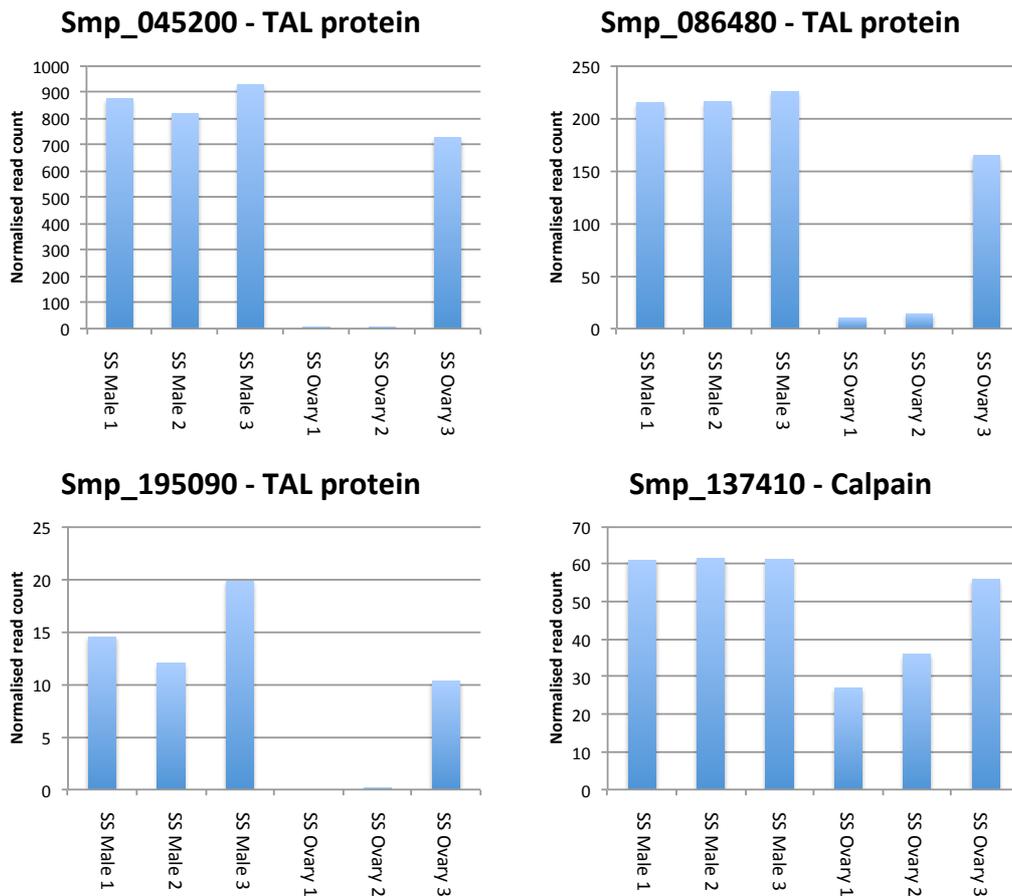
Figure 5.3: Single sex (SS) ovary replicate 3 was more similar to SS male replicates 1-3 than other SS ovary replicates (1-2). Here the sex-biased expression of genes is used to examine an outlier. Normalised read counts of SS male and SS ovary replicates were plotted for four transcripts with known male-biased expression patterns.

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

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

5.2.2 Comparing RNA-Seq to published microarray & qPCR data

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

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

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

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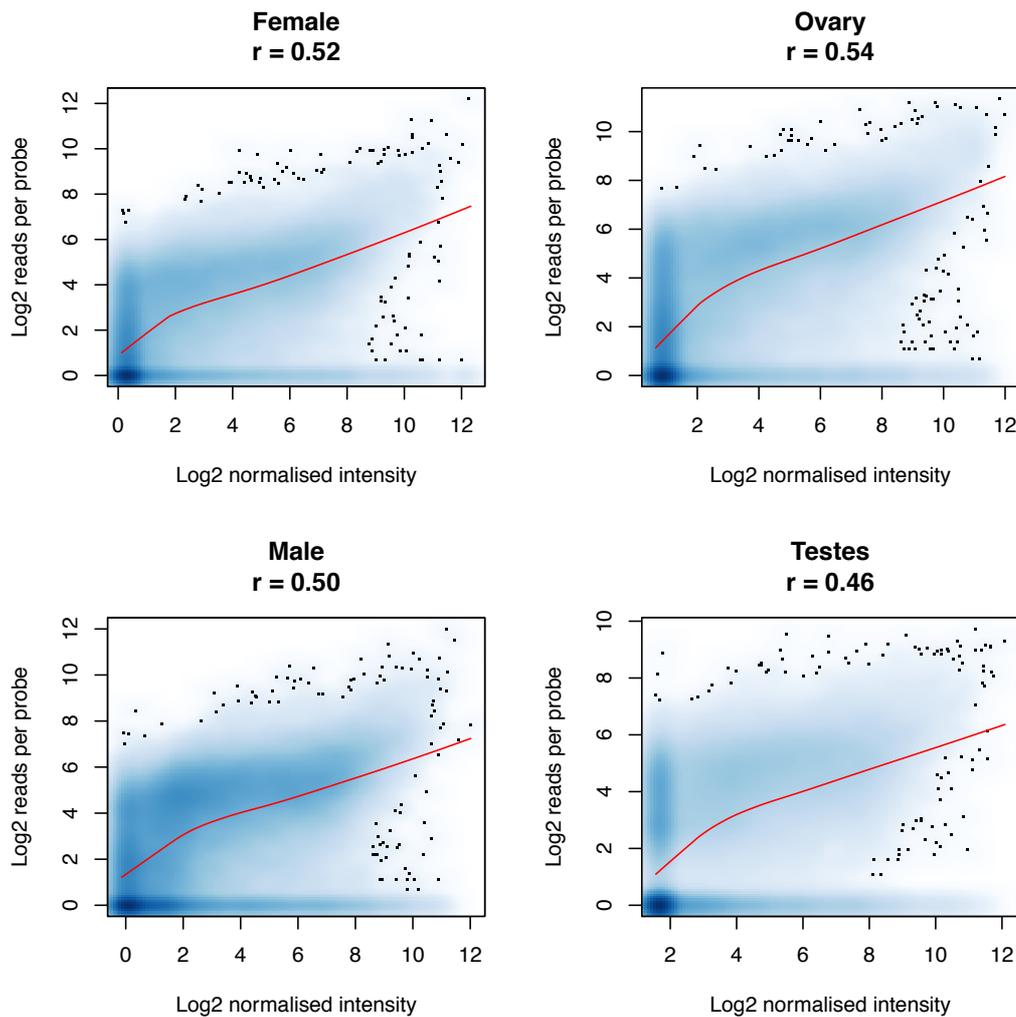
Figure 5.4: RNA-Seq and microarray data are moderately correlated. RNA-Seq counts were plotted against normalised intensity values of the microarray data sets (Nawaratna *et al.*, 2011); local regression line was plotted in red. A) Correlation of whole MS female samples. B) Correlation of ovaries from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

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

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

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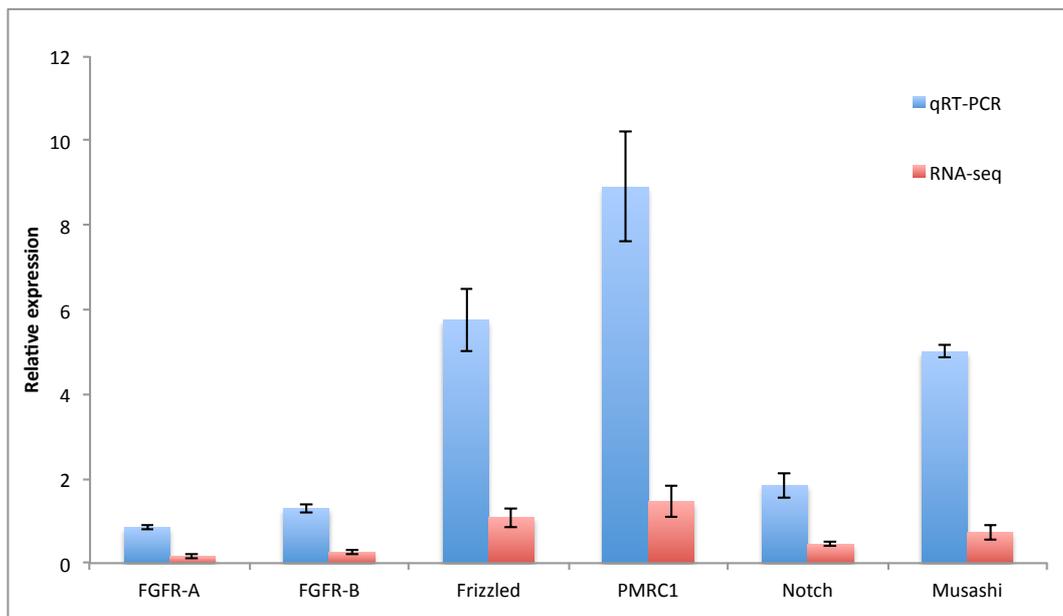
**Figure 5.5: Comparison of qRT-PCR and RNA-Seq reveals systematic differences.** Expression ratio in ovaries (expression in mixed sex ovaries divided by expression in single sex ovaries) as measured by qRT-PCR and RNA-seq. Generally RNA-seq ratios were all much lower suggesting less expression in paired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

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

Actin was not found to be expressed differentially to such large degrees in any other comparison of samples from SS versus MS infection but was still found to be significant (2.2-fold down-regulation in MS females; adjusted p value = 2.16E-13) in MS females (Appendix C.3).
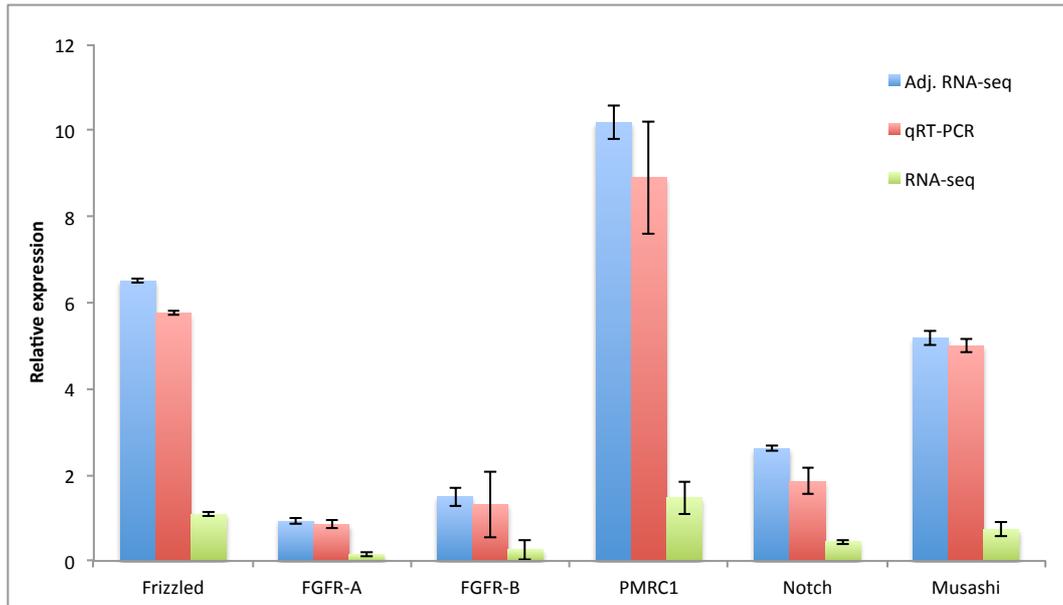
Figure 5.6: RNA-Seq data showed good agreement with qRT-PCR after adjustment for differential expression of qRT-PCR reference gene. Expression ratio in ovaries (expression in MS ovaries divided by expression in SS ovaries) as measured by qRT-PCR and RNA-Seq as well as an adjusted RNA-Seq ratio.

5.2.3 Testes Transcriptome

"Testes genes" were identified using a pair-wise comparison of gene expression in testes of MS males and whole MS males using DESeq2. To be considered as "testes gene" a gene had to be differentially expressed using an adjusted p-value cut-off smaller than 0.01 and be at least 2-fold more highly expressed in testes from MS infections than the whole male control. This threshold was used throughout this chapter to identify DEGs. 10270 out of 10828 transcripts were found to be expressed (*i.e.* non-zero read counts), of which 512 (5%) were filtered out due to low expression levels using the DESeq2 independent filtering algorithm. 1949 transcripts passed the criteria to be considered "testes genes" (Figure 5.6).
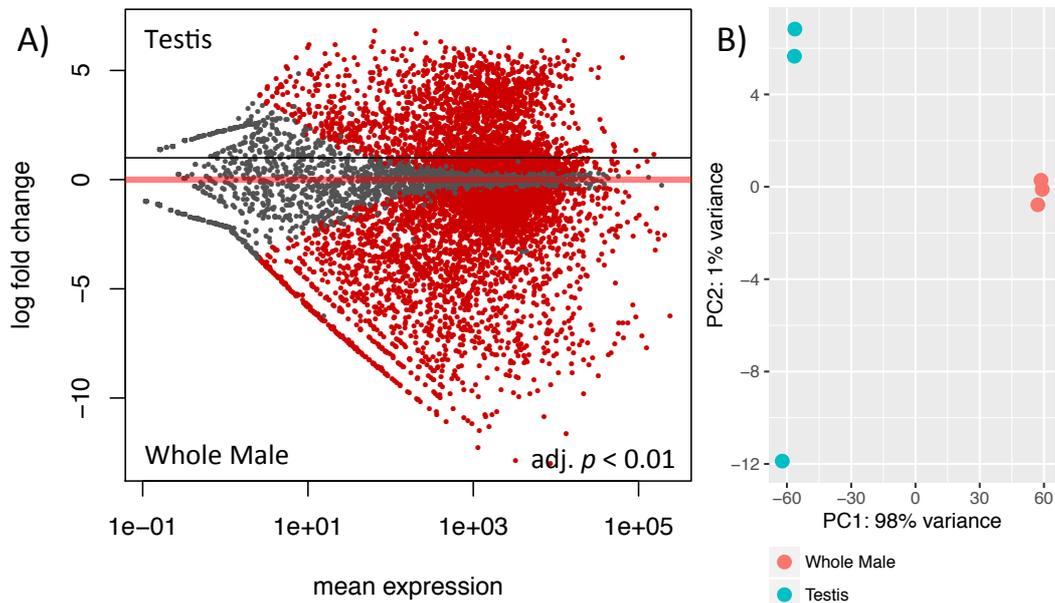
Figure 5.6: The testis transcriptome differs considerably from that of whole males. A) Log ratio against mean (MA) plot of the testis-whole male comparison. Genes with an adjusted p-value of < 0.01 are high-lighted in red; genes with a log-fold change of > 1 (fold-change of > 2) and adjusted p-value of < 0.01 were called as differentially expressed "testis genes". B) PCA plot of testes of paired males and whole male control samples.

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

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

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

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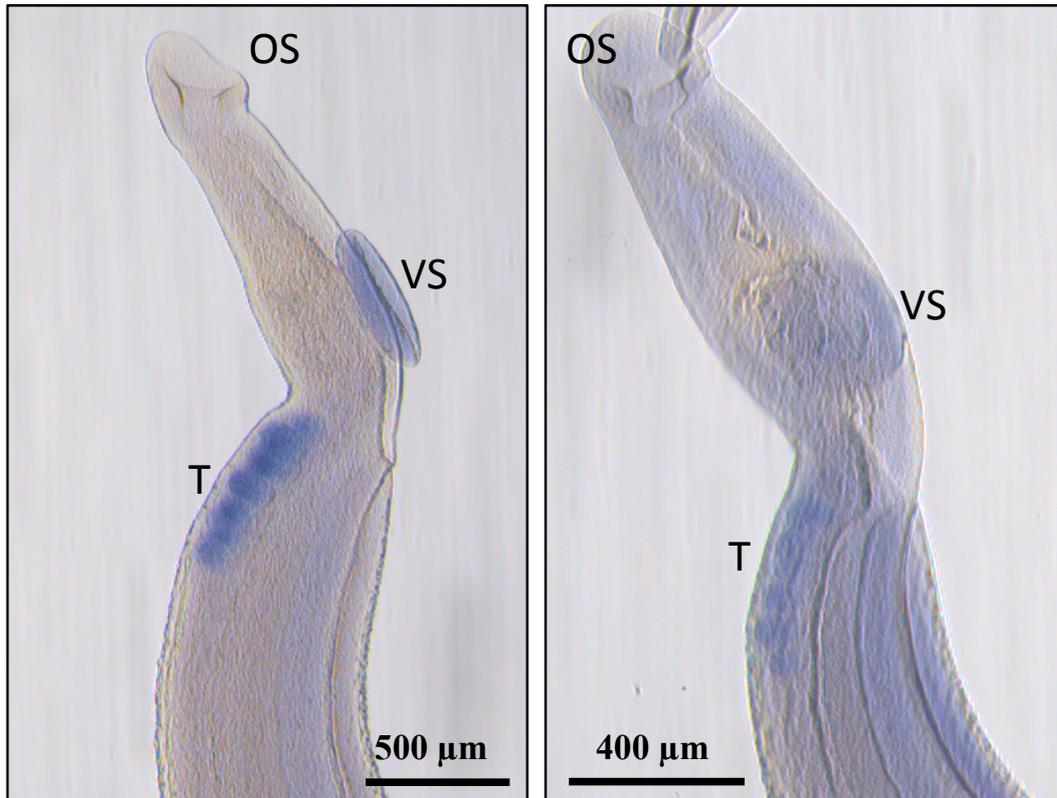
Figure 5.7: *In situ* **hybridisation targeting mRNA of transcription factor** *boule* **demonstrated expression in the testes of male worms.** Four out of twenty worms stained for boule also exhibited staining in the ventral sucker (left). Sense transcript of *boule* was used as negative control (right).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

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

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

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

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

The mitotic activity of testes is also evident in an analysis of over-represented protein domains in the testes transcriptome; for instance, out of 75 genes coding for helicase C-terminal domains, more than half are expressed at higher levels in testes than the whole male control (only three are expressed at higher levels in the whole male control) and 27 out of 55 genes coding for DEAD/DEAH box helicase domain containing proteins are up-regulated in testes (compared with 3 in whole males) (Table 5.4). Furthermore, all 10 genes coding for mini-chromosome maintenance domains are up-regulated in the testes, as are all genes coding for proteins with MutS domain II, III, IV and V (all involved in DNA mismatch repair) (Table 5.4). Many of proteins with these domains play an important role in the eukaryotic DNA replication machinery as can be seen in Figure 5.8.
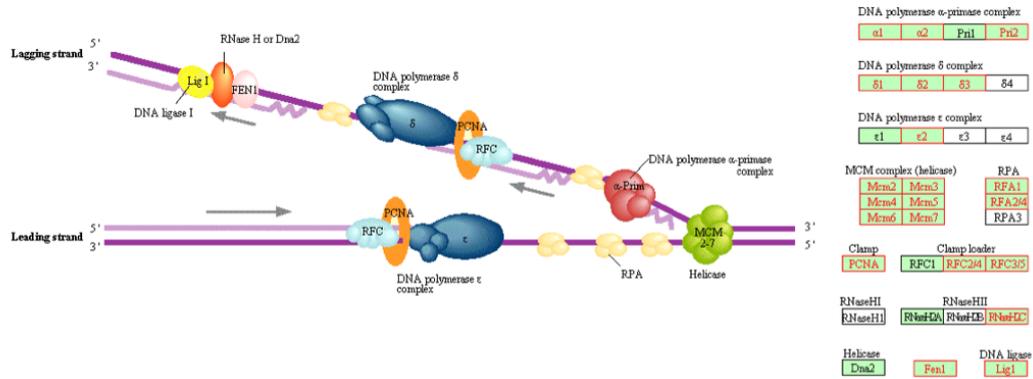
Figure 5.8: Most annotated *S. mansoni* homologues of DNA replication machinery genes are up-regulated in testes. Genes with a *Schistosoma mansoni* homologue are coloured in green, "testes genes" are highlighted in red. Most DNA replication genes are up-regulated in the testes transcriptome. This list highlights the important role played by many proteins with minichromosome maintenance (MCM) and helicase domains. Figure adapted from KEGG (https://www.genome.jp/dbget-bin/www_bget?hsa03030).

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

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

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

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

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

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

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

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

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

5.2.4 Ovary Transcriptome

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

adjusted p-value cut-off (< 0.01) and also have at least two-fold higher expression in ovaries than in the whole female control. Out of 10828 transcripts, 10213 were found to be expressed (*i.e.* non-zero read counts), of which 509 (5%) were filtered out due to low levels of expression using the DESeq2 independent filtering algorithm. 1272 transcripts passed the criteria to be considered "ovary genes" (Figure 5.9A).
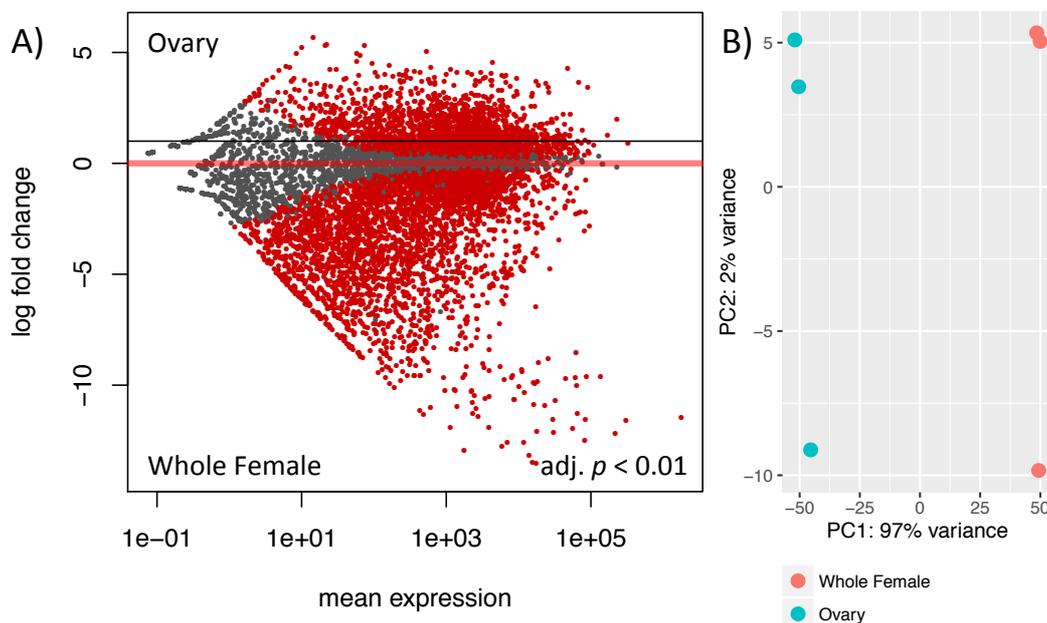
Figure 5.9: 1272 genes were identified as ovary-specific genes. A) Log ratio against mean (MA) plot of the ovary-whole female comparison. Genes with an adjusted p-value of < 0.01 are high-lighted in red; genes with a log-fold change of > 1 (fold-change of > 2) and adjusted p-value of < 0.01 were called as differentially expressed "ovary genes". B) PCA plot of ovaries of paired females and whole female control samples.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5.2.5 Comparing the Testes and Ovary Transcriptomes

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

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

Fewer genes had ovary-specific expression compared to the testes (Figure 5.10). In the differential expression analysis, ovary-expressed genes were found to have a lower median fold-change than testes genes (6.51 and 14.03-fold change respectively). As a result, more testes genes had a fold-change above the significance threshold (fold change > 2).
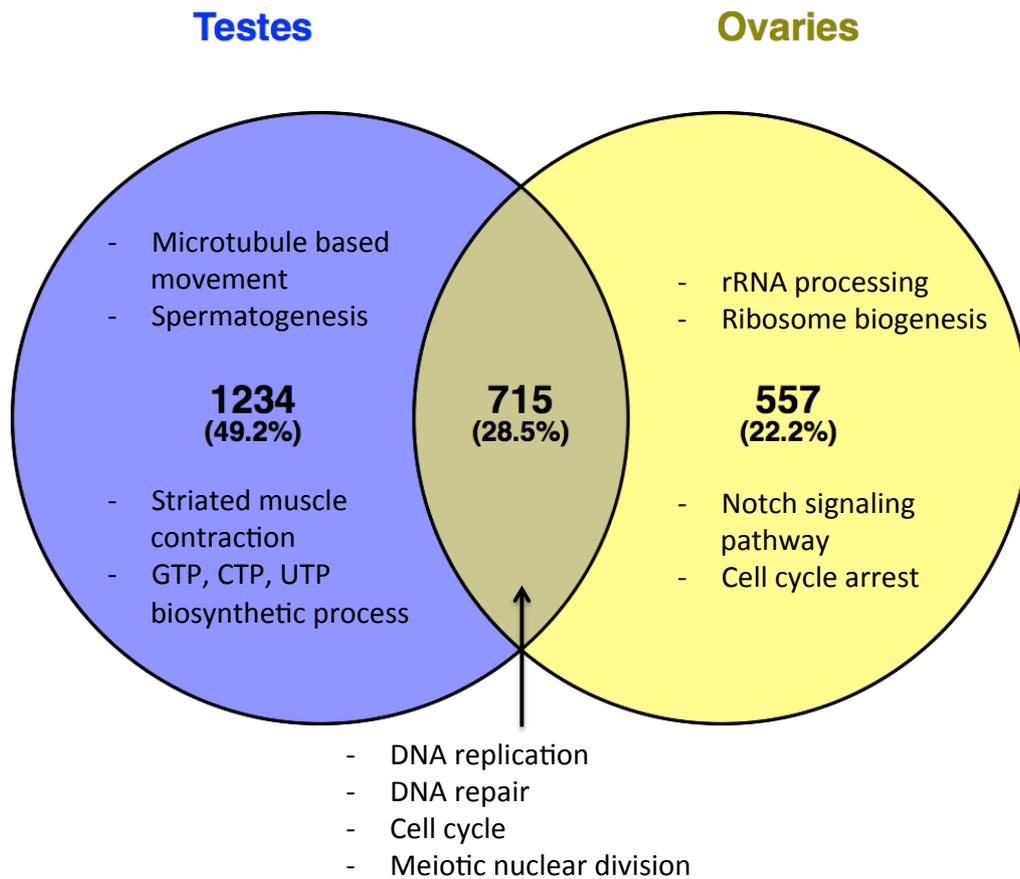
Figure 5.10: More testes-specific than ovary-specific genes were identified with distinct functions in the two groups of genes. The numbers indicate the testes and ovary "specific" genes as well as genes shared by both. A selection of significantly enriched GO terms is given for each category. The 715 shared genes were found to be strongly enriched for DNA replication related functions. This included repair and chromosome maintenance. The genes expressed in testes contained many spermatogenesis-specific genes whereas the genes expressed in ovaries included genes that regulate multicellular organismal development and Notch signalling pathway genes.

234

5.2.6 Effect of pairing on testes transcriptome

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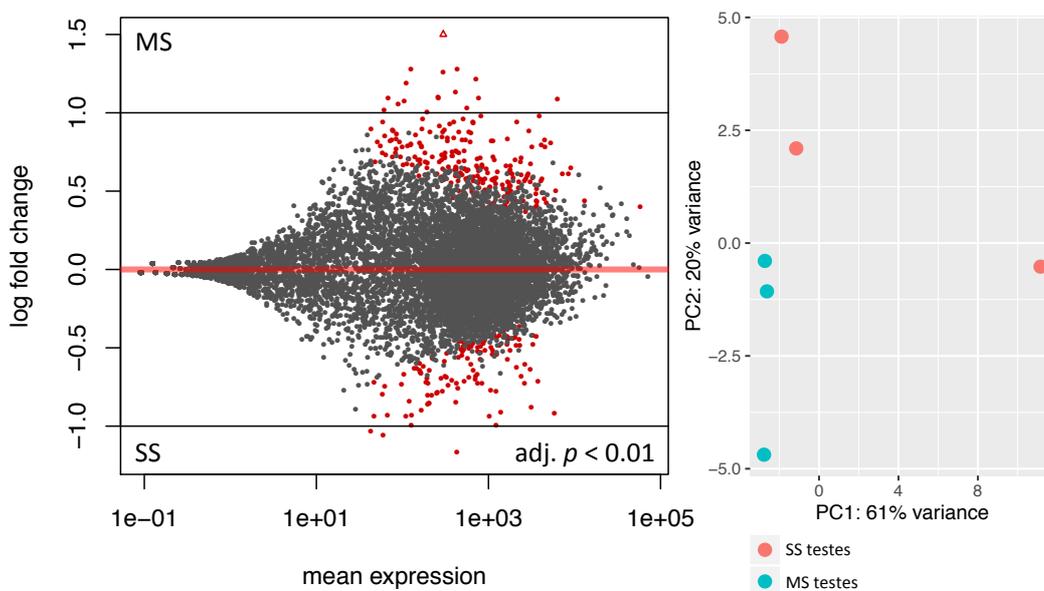

Figure 5.11 Few differentially expressed genes s were identified in comparison of testes from single sex (SS) and mixed sex (MS) infections. A) Log ratio vs mean (MA) plot of the SS testes – MS testes comparison. Genes with an adjusted p-value of < 0.01 are high-lighted in red; triangles indicate genes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

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

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

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

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

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

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

Follistatin - SmFst

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

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

5.2.7 Effect of pairing on ovary transcriptome

Unlike the small transcriptome changes found in the testes, in ovaries the changes induced by pairing are of a much greater magnitude. Out of 10828 examined genes, 10217 had non-zero read counts of which 1019 genes were filtered out due to low read counts using the DESeq2 independent filtering algorithm (Figure 5.12).
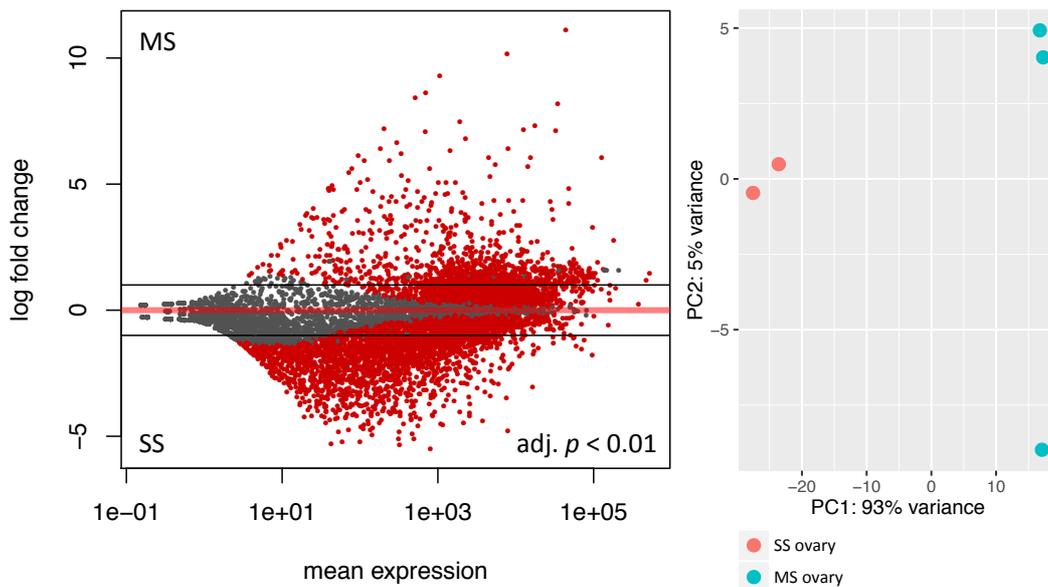
Figure 5.12 Over 2700 DEGs were identified in comparison of ovaries from single sex (SS) and mixed sex (MS) infections. A) Log ratio against mean (MA) plot of the SS ovaries – MS ovaries comparison. Genes with an adjusted p-value of < 0.01 are highlighted in red; triangles indicate genes outside the boundaries of the plots; like other genes, they are red if they are statistically significant, grey if they are not. Genes with a log-fold change of > 1 (fold-change of > 2) and adjusted p-value of < 0.01 were called as differentially expressed "ovary genes". B) PCA plot of ovaries of MS and SS females. 93% of variation across samples is explained by the experimental conditions in this PCA.

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

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

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

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

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

significantly enriched, including ATP synthase subunit C and mitochondrial carrier protein (Table 5.14). Highlighting the mitogenic activity within the mature ovaries, genes cofing for proteins with cyclin domains were differentially expressed at higher levels than expected by chance in the MS ovaries. Cyclins play a key role in regulating the cell cycle by controlling the progression from one stage to the next (Sclafani & Holzen, 2007). Comparing the expression levels of all genes annotated as cyclin, cyclin dependant kinases (CDK), or cell division cycle protein (CDC), there were a total of 38 genes that formed two distinct clusters based on their expression (Figure 5.13). This provided further evidence for differential regulation of the cell cycle in MS and SS ovaries.
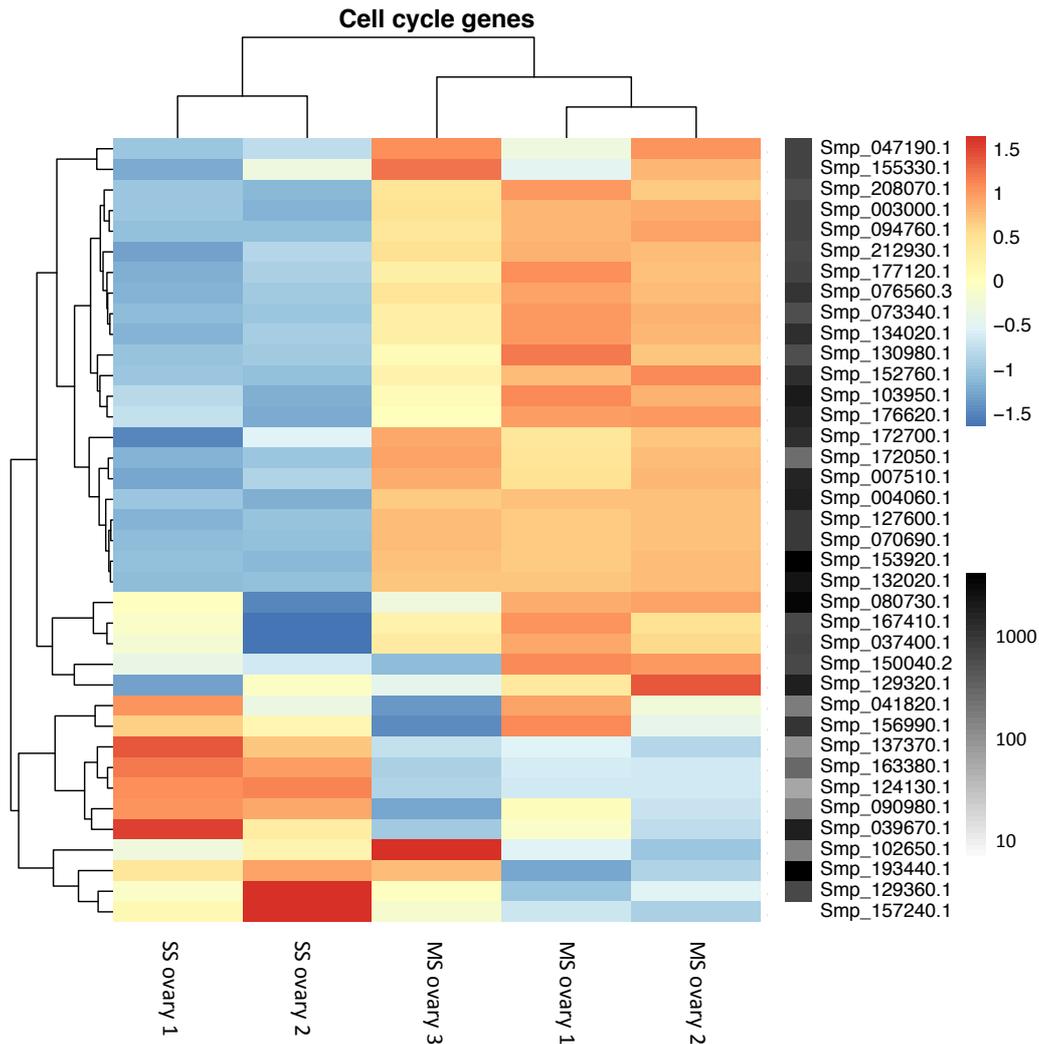
Figure 5.13: Many genes involved in cell cycle regulation, including cyclin, CDK and CDC genes are up-regulated in mix sex ovary samples. Three RNA-seq replicates of mixed sex (MS) ovary (MS ovary 1-3) and two replicates of single sex (SS) ovary (SS ovary 1-2) were used to generate this figure. Colours represent a Z-score of log transformed normalised read counts scaled by row, providing relative measures of gene expression in each sample (measured in standard deviations from the mean). Average normalised read counts for each gene are in the row on the right on a logarithmic scale as a measure of absolute expression levels. See Appendix C.20 for a description of these cell cycle genes.

The MAPK, Rho and TGF-β signalling pathways have been examined closely and are known to play an important role in the development of the female sexual maturity (Andrade *et al.*, 2014; Osman *et al.*, 2006; Santos *et al.*, 2002). To examine the processes that lead to the initiation of the MAPK pathway in the cells of the ovaries and induce maturation of the tissue I compiled a list of the well established *S. mansoni* genes involved in the signalling network that transmits the male stimulus to the nuclei of ovary cells (Andrade *et al.*, 2014; Beckmann *et al.*, 2010; Knobloch *et al.*, 2007; Oliveira *et al.*, 2012; Osman *et al.*, 2006) and created a heat map of normalised read counts for the ovary samples (Figure 5.14). This set of genes was also clustered into two expression groups, one up-regulated in the ovaries of MS females and one down-regulated when compared to ovaries of female worms from SS infections (Figure 5.14). The signalling network features over 40 members with different regulatory function that, in response to pairing, drive the cell cycle towards DNA replication and the reorganisation of the cytoskeleton to allow for cell division (Beckmann *et al.*, 2010).
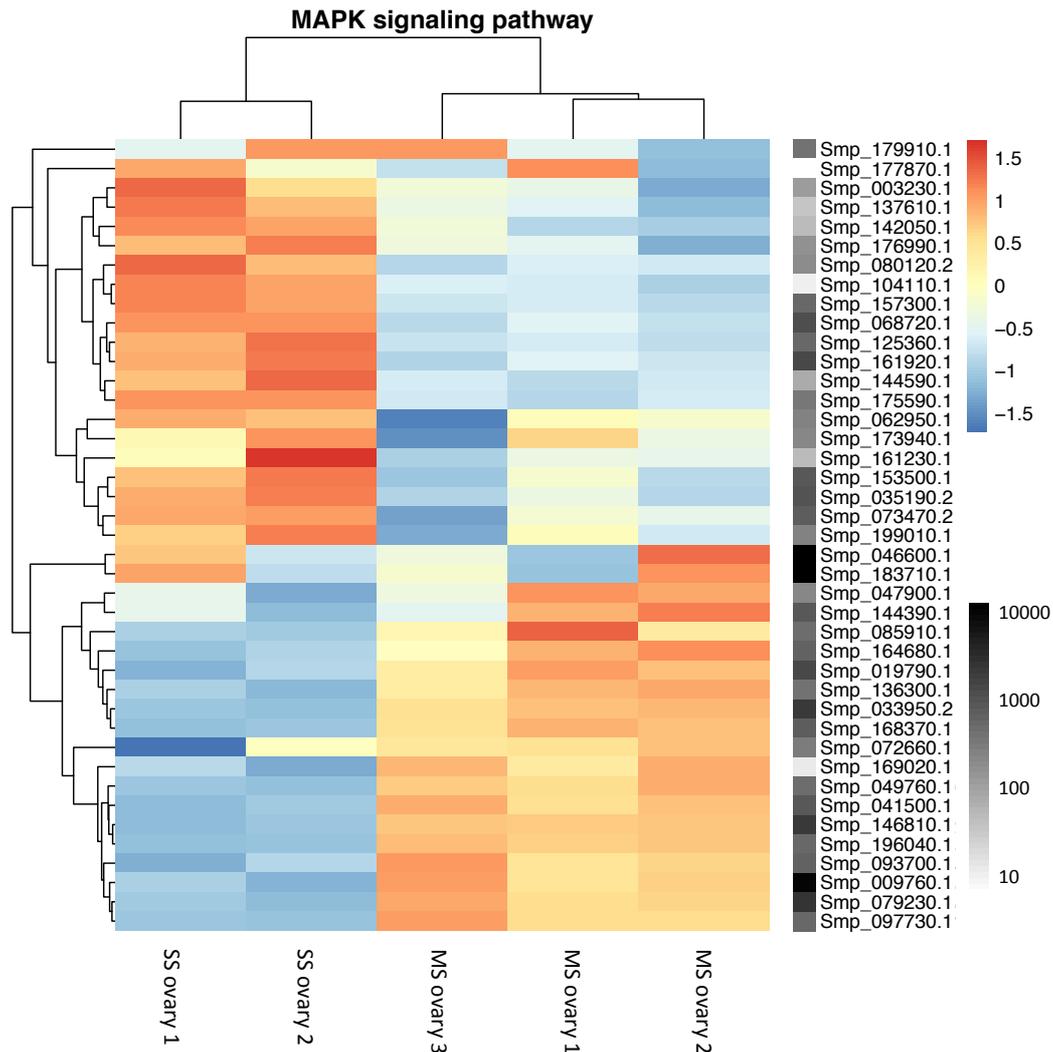
Figure 5.14: Many genes of the MAPK, Rho and TGF-β signalling pathways are differentially expressed in the mixed sex (MS) and sngle sex (SS) ovary samples. Three RNA-seq replicates of MS ovary (MS ovary 1-3) and two replicates of SS ovary (SS ovary 1-2) were used to generate this figure. Colours represent a Z-score of log transformed normalised read counts scaled by row, giving relative measures of gene expression in each sample. Average normalised read counts for each gene are in the row on the right on a logarithmic scale as a measure of absolute expression levels. See Appendix C.21 for a description of these MAPK signalling genes.

Unlike in Figure 5.13, where the expression of genes coding for cyclins, CDCs and CDKs was compared, the picture emerging from a comparison of genes coding for proteins involved in MAPK, TGFb and integrin signalling was less clear. Although a large number of these genes were up-regulated in the MS ovaries, a similar number of genes had greater expression in the ovaries of SS females (Figure 5.14). Up-regulation of genes in the SS ovaries was by no means limited to genes coding for proteins with inhibitory functions but included genes coding for some of the key regulators of the MAPK signalling pathway such as homologues of Rho, Ras-guanine nucleotide exchange factor, Raf and Erk (Figure 5.15). Others, such as genes coding for growth factor receptor-bound protein 2 and Ras, showed no differentially expression (Figure 5.15). In contrast, the genes coding for SMAD2 and 4 as well as Transforming Growth Factor-β Receptor II, Ras-GAP and Scr-cytoplasmic tyrosine kinase homologues were found to be up-regulated in MS ovaries (Figure 5.15). Ultimately however, the induction of cell proliferation is dependent on receiving specific extracellular stimuli provided by pairing as well as host factors to initiate signalling and mitogenic activity and cell differentiation (Beckmann *et al.*, 2010).
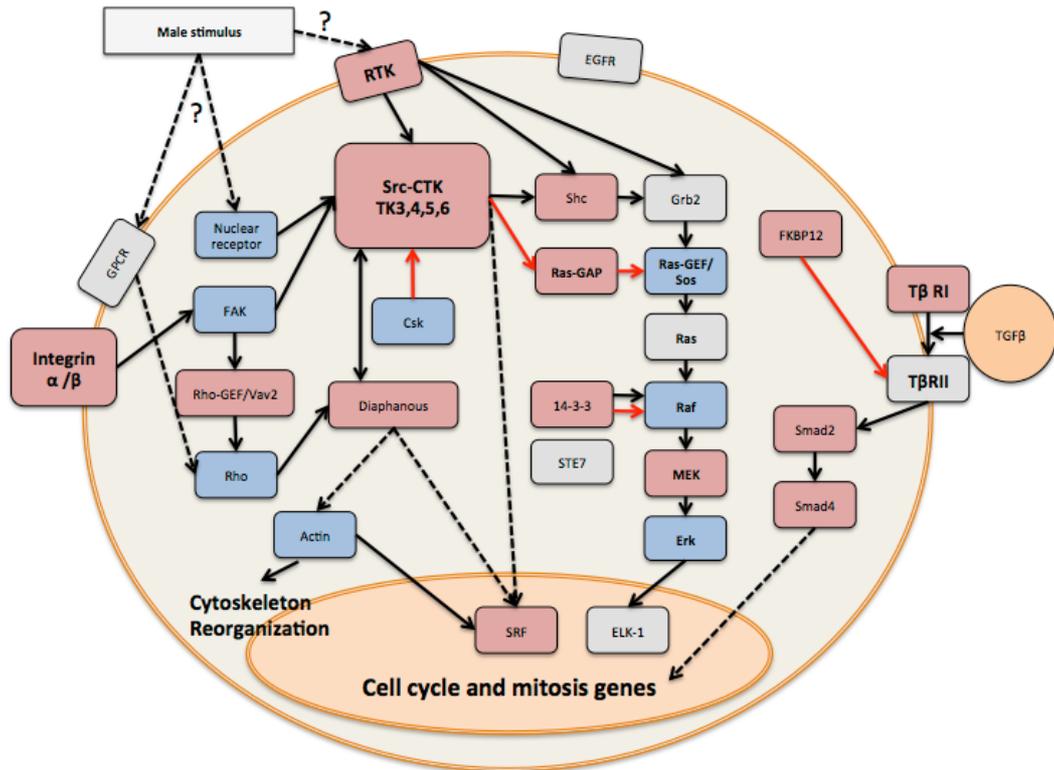
Figure 5.15: Many members of the MAPK, Rho and TGF-β signalling pathway are differentially expressed in ovaries. Adapted from Knobloch *et al.* (2007). Dotted arrows indicate indirect interactions, solid arrows indicate direct interaction. Red arrows indicate inhibition and black arrows activation of the target. Red coloured genes are up-regulated in the ovaries of MS females, blue coloured genes are up-regulated in the ovaries of SS females, grey coloured genes are not differentially expressed.

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

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

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

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

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

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

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

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

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

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

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

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

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

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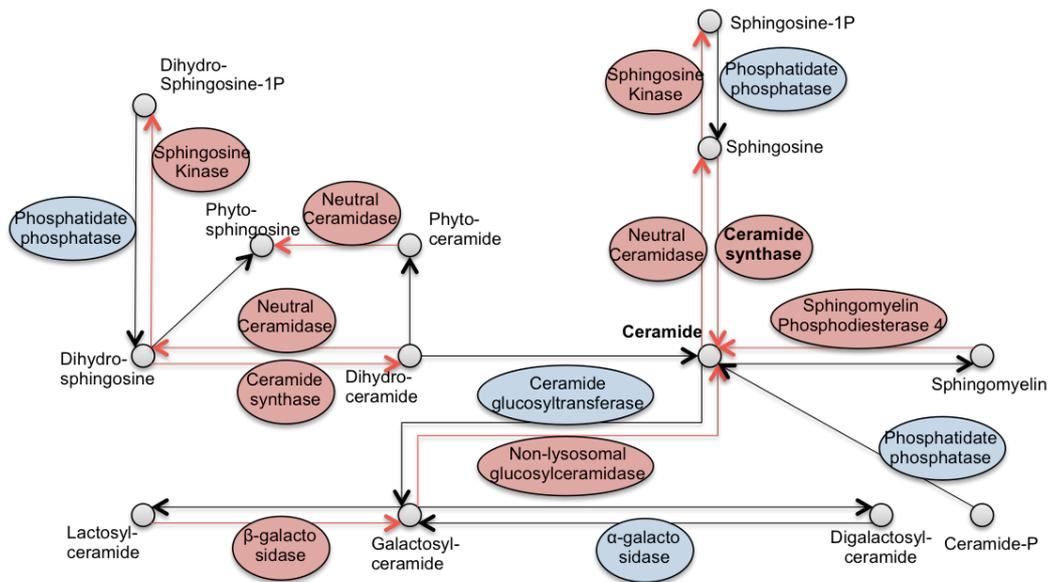

Figure 5.16: Many genes involved in sphingolipid metabolism are differentially expressed in ovaries from single sex (SS) and mixed sex (MS) infections. The genes encoding red coloured enzymes are up-regulated in ovaries of SS females, genes coding for blue coloured enzymes are down-regulated in the ovaries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

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

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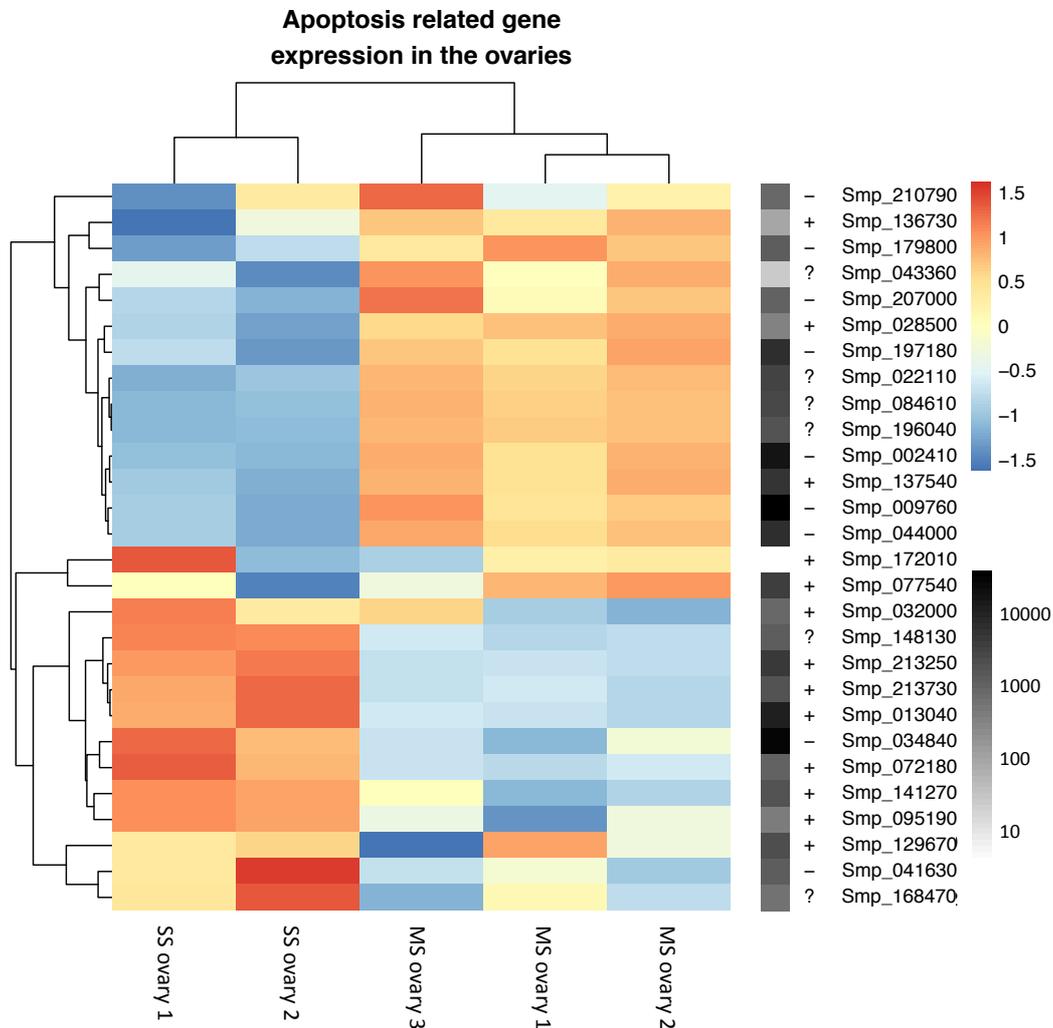
Figure 5.17 Expression of apoptosis-related genes points to higher levels of apoptosis in the ovaries of females from single sex (SS) infections than those of females form mixed sex (MS) infections. Three RNA-seq samples of MS ovary (MS ovary 1-3) and two SS ovary (SS ovary 1-2) were used to generate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

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

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

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

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

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

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

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

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

5.3 Discussion

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

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

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

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

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

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

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

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

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

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