Chapter 2

Materials & Methods

2.1 Parasite and snail maintenance

2.1.1 Parasites and snail origin

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

2.1.2 Miracidia infection of snails

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

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

Lepple water (1x)

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

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

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

<u>W2 primers:</u>

FWD: CTGTTTCGAATTTCACACTTCA (T_m = 61.7°C)

REV: CATTCACAGTTTGGCGAACA (T_m = 64.7°C)

Rho primers:

FWD: GACGGCCACACTAAAG (T_m = 54.7°C)

REV: AGTAAAATGGTCACTGCTAT (T_m = 52.8°C)

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

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

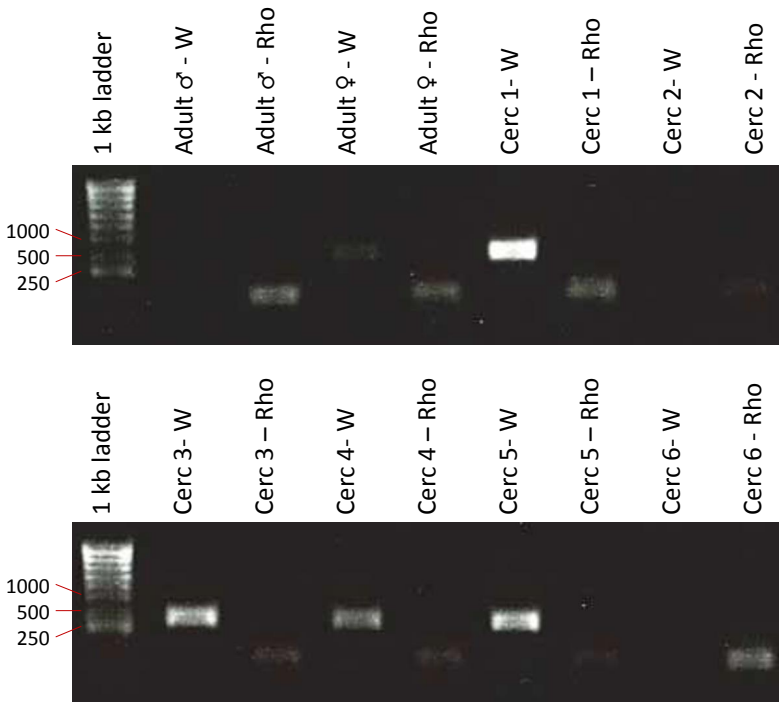

Figure 2.1: The sex of cercariae from single-miracidium infections was determined by PCR. Male *S. mansoni* samples do not have an amplicon in the W chromosome specific region, whereas female *S. mansoni* samples do have an amplicon for the W specific reaction. The Rho reactions serve as a positive control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

2.1.3 Detecting eggs in mouse livers

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

2.1.4 Collection of cercariae

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

2.1.5 Infections and perfusions

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

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

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

2.1.6 *In vitro* culture

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

Supplemented DMEM

0.02 M Hepes buffer (Sigma, 83264)

2 mM L-glutamine (Sigma, G7513)

10% Fetal calf serum* (Sigma, F3297)

1x Antibiotic-Anitimycotic (ThermoFisher, 15240062)

DMEM media** (ThermoFisher, 10270106)

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

**) High glucose DMEM media was used.

Basch media

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

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

2.1.7 Isolation of *S. mansoni* gonads

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

2.2 Molecular biology techniques

2.2.1 Cloning

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

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

FWD: ATGTGTACTGTCGTATTGAGATTAAC (T_M = 59.0°C)

REV: TTAAGGATATCCAGTGGTTAGTATC (T_M = 58.2°C)

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

FWD: ATGGCCTCTTTAAGCTGTGG (T_M = 63.1°C)

REV: TTAGTCAGAATCGCCAGATTTG (T_M = 63.0°C)

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

FWD: TTCTTCTTCGTTTCAGGGATGT (T_M = 63.7°C)

REV: TCACCGCGCTTTATAGCCAA (T_M = 67.5°C)

Boule (Smp_144860) primers:

FWD: GCCTTGGGGCATCTAAAGGT (T_M = 66.1°C)

REV: TGCTGCAGTGGATGCTGTTA (T_M = 65.7°C)

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

FWD: TCTTTCATGCAGGCCACAGT (T_M = 65.8°C)

REV: CCGGAACCAACCCGATTACA (T_M = 68.8°C)

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

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

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

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

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

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

2.2.2 RNA interference

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

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

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

CD63 receptor primers (T_m = 54°C):

FWD: TAATACGACTCACTATAGGGATGTGTACTGTCGTATTGAGATTAAC

REV: TAATACGACTCACTATAGGGGAACAACAAATTTGGCATC

CD63 antigen primers (T_m = 58°C):

FDW: TAATACGACTCACTATAGGGATGGCCTCTTTAAGCTGTGG

REV: TAATACGACTCACTATAGGGCAGGGATTGTTTGTCCACTC

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

FWD: TAATACGACTCACTATAGGGTGATTGTGGTTGGTGCACTT

REV: TAATACGACTCACTATAGGGGACCAATGCGAACAGAAACA

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

FWD: TAATACGACTCACTATAGGGTCTTTCATGCAGGCCACAGT

REV: TAATACGACTCACTATAGGGCCGGAACCAACCCGATTACA

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

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

dsRNA soaking

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

dsRNA electroporation

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

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

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

2.2.3 RNA extractions

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

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

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

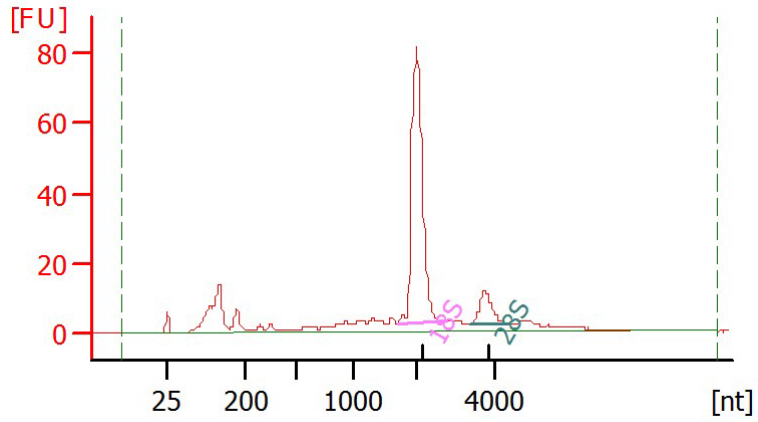

Figure 2.1: **Example of good quality extracted total RNA.** The sharp peak of RNA at 2000 nucleotides (nt) is the ribosomal RNA peak.

2.2.4 cDNA synthesis

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

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

2.2.5 qPCR

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

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

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

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

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

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

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

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

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

For multiple reference genes the average Ct value was calculated:

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

Finally to find the relative difference in gene expression:

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

The fold-change was then be calculated as follows:

Fold-change = 2 ^ ΔΔCt

2.2.6 Dot blot

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

2.2.7 Whole mount *in situ* hybridisation

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

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

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

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

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

1) Preserving specimens

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

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

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

2) Permeabilisation & hybridisation with probe

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

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

3) Removal of excess probe & incubation with antibody

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

4) Removal of excess antibody & developing AP staining

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

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

2.2.8 Imaging of WISH specimen – light microscopy

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

2.3 RNA-Seq library preparation & sequencing

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

2.3.1 mRNA selection

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

2.3.2 mRNA fragmentation

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

2.3.3 Reverse transcription

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

2.3.4 Second strand DNA synthesis

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

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

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

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

2.3.6 PCR amplification

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

2.3.7 Sequencing

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

2.4 Bioinformatics

2.4.1 Aligning RNA-Seq reads to the genome

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

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

2.4.2 Sorting and merging of BAM files

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

2.4.3 Counting reads

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

2.4.4 Differential gene expression analysis

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

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

Model and normalisation

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

Empirical Bayes shrinkage for dispersion estimation

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

Empirical Bayes shrinkage for fold-change estimation

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

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

Automatic independent filtering

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

Detection of count outliers

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

2.4.5 Correlation of RNA-Seq data with microarray data

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

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

2.4.6 Principal component analysis (PCA)

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

2.4.7 Heatmaps

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

2.4.8 Gene Ontology (GO) term enrichment

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

2.4.9 InterProScan & Pfam enrichment

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

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

2.4.10 KEGG pathway enrichment

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

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

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

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

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

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

2.4.11 Cluster analysis of RNA-Seq data

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

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

2.4.12 Gene models & annotation

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

2.5 Scanning Electron Microscopy

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