CHAPTER 4

COMPARATIVE STUDY OF TRANSCRIPTOME PROFILES OF MECHANICAL- AND SKIN-TRANSFORMED SCHISTOSOMULA

4.1 Introduction

Before the development of a mechanical transformation method (Gazzinelli *et al.*, 1973; Howells *et al.*, 1974; Ramalho-Pinto *et al.*, 1974), schistosomula in large numbers could only be recovered by naturally allowing parasites to penetrate a layer of excised host skin. Such an approach was first reported in 1966 where a protocol using "scraped, dried, plucked skin" from the abdominal area of rats, mice and hamsters was described (Stirewalt *et al.*, 1966). Rat skin proved to be the most convenient given its larger size and availability. Despite providing large quantities of schistosomula (~45,000 per experiment) this protocol was extremely time- and labour-intensive taking 18 hours to dry the skin prior to the actual skin penetration experiment. However, this work allowed the authors to define schistosomula based on experimental observation of parasites obtained *in vitro* (with the described protocol) and *in vivo*. According to the authors, schistosomula are:

> *"saline- and serum-adapted, water-intolerant larval stage which develops from the cercaria after the tail has been cast, the pre- and post-acetabular glands have been evacuated, and there has been a change of surface resulting in altered permeability, and the loss of the precise cercarial shape and the capacity to form pericercarial sero-envelopes"* (Stirewalt *et al.*, 1966)

A later report from the same group (Stirewalt *et al.*, 1969) refined their own protocol (Stirewalt *et al.*, 1966) and tested the influence of different factors (such as temperature, light, number of cercariae applied, etc) on the yield of recovered schistosomula. The authors established that approximately 51% of the applied cercariae successfully transformed into schistosomula and that the final preparation lacked post-penetration larvae (cercariae that had penetrated through the layer of skin but did not comply with the definition of schistosomula). Later reports by Clegg and Smithers (1972) slightly modified the protocol from Stirewalt *et al.,* (1969) by using non-dried freshly excised skin from mouse abdomen. Schistosomula were phenotypically comparable to those obtained using rat skin (Stirewalt *et al.*, 1966) and also to those recovered from infected animals (*in vivo* collection) (Clegg *et al.*, 1972). However, the yield of schistosomula decreased using mouse skin (20-30%) compared to the optimised conditions listed above. In all the experiments and according to the authors, these approaches yielded tail-free and cercariae-free schistosomula preparations.

The *in vitro* mechanical transformation was introduced in 1974 (Ramalho-Pinto *et al.*, 1974). In this approach, the authors used centrifugation of freshly shed cold cercariae followed by incubation in culture media at 30°C for 40 minutes to produce viable schistosomula. Separation of the tails and isolation of cercariae bodies was achieved by sedimentation. Further incubation of the cercariae bodies in culture media and mild shaking induced changes in the parasites surface completing their transformation. This process is much more efficient than the skin penetration approaches,

transforming 50-70% of the initial cercariae subjected to the procedure. This incremental improvement in total number of recovered parasites and the ease of the experimental protocol represented the major attractions of the approach for subsequent investigators. Furthermore, these parasites fulfilled the criteria for schistosomula suggested by Stirewalt *et al.,* (1966). Brink *et al.,* (1977) improved the original mechanical transformation protocol by adding a Ficoll gradient step to obtain a schistosomula preparation virtually free from tails. To validate their results, the authors compared these "mechanically transformed schistosomula" (MT) with preparations obtained using the skin penetration approach (ST) (Clegg *et al.*, 1972) but observed that ST were different from MT in several aspects:

- Pre-acetabluar glands from ST were emptied after penetration, whilst the MT retained the contents in the glands.
- 3-hour old ST schistosomula showed evidence of presence of human A and B blood group-like antigens whilst MT schistosomula did not.
- In infectivity tests, a slightly higher percentage of ST parasites where recovered from mice compared to MT. However, the difference between these two groups was not significant.

On the other hand, several important factors did not differ between ST and MT, such as loss of the glycocalyx (as early as 2 hours) or the replacement of the trilaminate surface membrane by the heptalaminate membrane structure, which is a characteristic of transformation of cercariae into schistosomula. After 12 days, a higher percentage of ST (50-70%) had completed the process of closing the gut[1], while only 25-50% of the MT population had reached the same stage of development. This was not interpreted as a physiological difference but a delay in the development of the parasites. The authors explained this effect by arguing that during the ST there is a selection for the most "fit" cercariae as many of them die during the penetration process. In this case, the percentage of "gut-close" parasites in an entire population of ST is higher than that observed in MT where all cercariae (fit and no so fit) transform into schistosomula. This work showed that ST and MT are morphologically almost equivalent to each other (both change their surface membrane, both shed their glycocalyx) but that a slight "delay" can occur in their progression to the "gut-close" stage in the MT. Interestingly, MT schistosomula produced by the methods employed by Brink *et al.,* (1977) did not empty the contents of their pre-acetabular glands (Brink *et al.*, 1977), a requirement of the suggested definition of schistosomula (Stirewalt *et al.*, 1966). The authors argued that they do not consider the

[1] This stage is reached approximately 15 days after transformation and it describes parasites that have both the ends of the ceca joined. It is often called "gut-close" stage and it was defined in Clegg, (1965).

state of the pre-acetabular glands as a criterion to define a schistosomulum (Brink *et al.*, 1977). More recent studies have suggested that some of the contents of the acetabular gland complex are not emptied until the parasites are located deeper in the skin of the invaded host, probably because different secretion cocktails are needed to penetrate through different barriers (Crabtree *et al.*, 1985). This would support the claim of Brink *et al.,* (1977) that the status of the pre-acetabular glands after transformation is not a criterion for the definition of a schistosmulum.

The implementation of the MT together with the work of several groups that evaluated ST and MT parasites in terms of their phenotype and biochemistry provided enough evidence that these parasites were equivalent.

High throughput gene expression studies (i.e. microarrays) have relied on this equivalency, mainly because obtaining enough material from *in vivo* infections is not always feasible. Therefore, MT schistosomula have been used almost exclusively in high-throughput studies of gene expression (Verjovski-Almeida *et al.*, 2003; Dillon *et al.*, 2006; Fitzpatrick *et al.*, 2009; Gobert *et al.*, 2010; Farias *et al.*, 2011), identification of drug targets (Fitzpatrick *et al.*, 2009) and identification of effective drugs against schistosomes from a compound library (Abdulla *et al.*, 2009). In many cases, high-throughput gene expression studies are used as a "fishing expedition" where interesting genes, or at least those relevant for a particular process (i.e., invasion, host-parasite interaction, etc) can be pointed out based on their pattern of expression. Using inadequate samples could lead to miss interpretation of the results. For example, the mechanical transformation protocol might be subjecting the parasites to artificial sources of stress that could elicit a response from the parasites. In the case of high-throughput studies this could lead to the follow-up of genes expressed as a result of these artefacts instead of signals arising from the natural transformation of the cercariae into schistosomula. Another scenario would be to miss exploitable vulnerabilities that could arise from cues triggered by the skin barrier; such signals would be missed in MT parasites. As an example, cercariae of human-infectious *Schistosoma* spp. are known to react to fatty acids as chemoattractans (Haeberlein *et al.*, 2008) and these induce the release of contents of the acetabular glands (Stirewalt, 1978). It is expected that downstream signals deriving from these chemical cues be seen in the skin-transformed parasites but not in their mechanical counterparts. Therefore, validation at the transcriptional level of the equivalency of *in vitro* (mechanically transformed schistosomula) and *in vivo* obtained samples is important to justify these studies.

So far, there has been only one study reporting such a comparison. The work of Chai *et al.,* (2006) used a microarray platform to investigate the gene expression profile of

S. japonicum parasites from the lung stage (Chai *et al.*, 2006). They studied two lung-stage schistosomula parasite populations: one obtained *in vitro* (mechanical transformation followed by *in vitro* cultivation for three days, this preparation will be referred as MTS) and the other prepared *in vivo* (obtained from lungs from infected mice three days post infection, referred as IVS). A comparison with adult worms was also presented. Their results show that MTS and IVS transformed parasites were phenotypically identical with the exception that some MTS retained part of the cercarial glycocalyx. However, in terms of differential expression, the authors found that a striking 6,662 genes were differentially expressed, with a p-value ≤ 0.001 (but not corrected for multiple hypothesis/multiple testing), between MTS and IVS (3,207 and 3,455 genes with higher expression in the MTS and IVS respectively). As a comparison, the total number of differentially expressed genes observed between the MTS and the adult stage was found to be significantly less (just 3,777 genes) (Chai *et al.*, 2006).

One of the reasons that might have caused the large number of differentially expressed genes between MTS and IVS samples is the heterogeneity of the parasite population. As introduced in Chapter 1, the skin penetration and the passage through the endothelial wall are processes that vary greatly in time (skin penetration can occur between 0 and 40 hours and endothelial passage may take up to 8 hours) and therefore transformation of parasites within a population is not synchronised. This heterogeneity would most likely also be reflected in the repertoire of expressed genes.

On the other hand, mechanically transformed parasites are easier to synchronize and represent a more homogeneous population facilitating the analysis of differentially expressed genes. The large number of genes appearing as differentially expressed between these two populations might be an artefact of the heterogeneous population of *in vivo* recovered parasites used in this experiment. The authors conclude that MTS parasites do not represent their *in vivo* (IVS) age equivalents (Chai *et al.*, 2006).

The identification of transcripts appearing differentially expressed in either mechanically- or skin-transformed schistosomula provides the means to establish the differences between these two ways of preparing schistosomula from the gene expression perspective. Given that the early schistosomula stage represents the onset of infection where the schistosomes start their parasitic life in the mammalian host, it is important to understand and identify to what extent the mechanisms of transformation affect these parasites. In the following sections, 24-hour old skin-transformed and mechanically transformed schistosomula transcriptomes are compared and their differences are established.

.

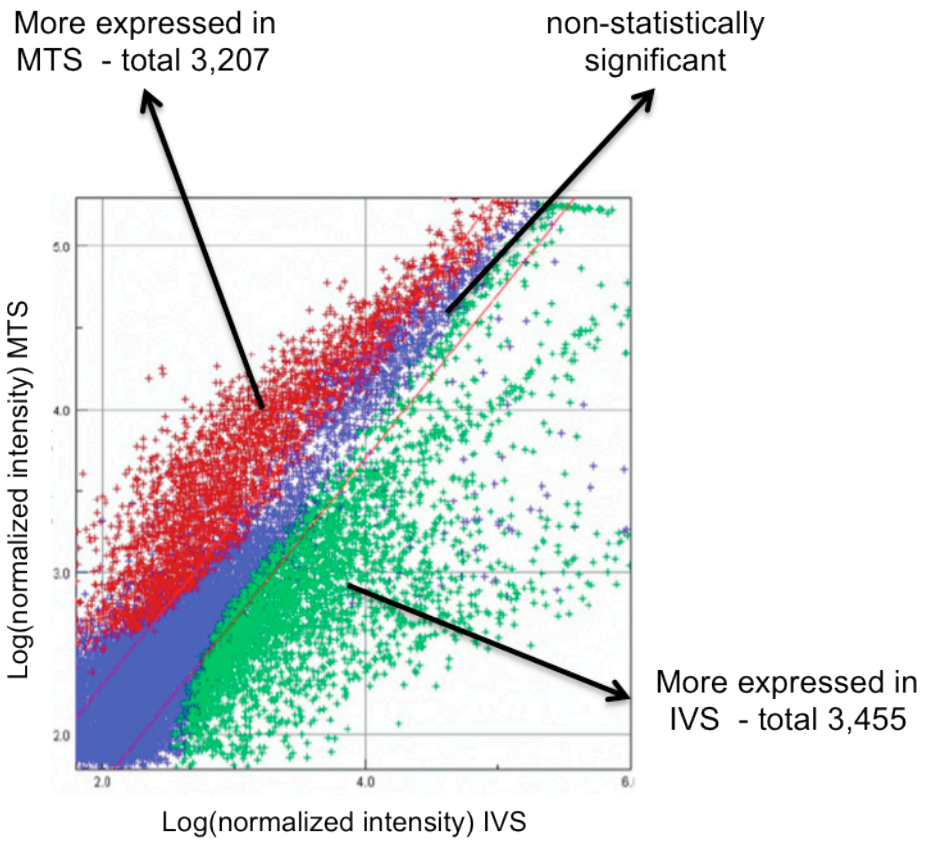


Figure 4.1 – Previous reports of comparison between *in vitro* and *in vivo* samples for *S. japonicum*. This scatter plot was taken from (Chai *et al.*, 2006) to illustrate the differences observed between lung-stage schistosomula obtained *in vivo* (x axis) and schistosomula obtained *in vitro* (y axis). The authors found 3,207 genes over expressed in the MTS (green crosses) and 3,455 genes over expressed in the IVS (red crosses) with p-value < 0.0016.

4.2 Results

4.2.1 Assessment of transformed parasites

The first step in the study of transcriptional differences between 24-hour old parasites recovered both *in vitro* or *in vivo* is the collection of the parasitic material. The protocols used for this were introduced in Chapter 2 sections 2.2. The following three sections in this chapter explain the steps taken to optimise the aforementioned protocols and the experimental approaches performed to evaluate the successful transformation of schistosomula.

4.2.1.1 Optimization of transformation protocols

Both mechanical and skin transformation protocols were subjected to optimization. In the case of the mechanical transformation two main aspects of the sample preparation were taken into close consideration for optimization: presence of contaminating tails plus the number and proportion of "healthy" parasites in the final schistosomula preparation. The concentration and temperature of Percoll were found to be important factors in obtaining tail-free schistosomula preparations. Percoll concentrations ranging from 60% to 80% were tested. The lowest percentage of Percoll tested (60%) resulted in an significant number of tails contaminating the final schistosomula preparation (**Figure 4.2A**) while the higher percentage of Percoll tested (80%) reduced the number of tails but also the number of schistosomula found after separation of tails and cercarial bodies (**Figure 4.2B**). A 70% Percoll solution was found to be a good compromise between the two extremes providing a virtually tail-free schistosomula preparation (~1% tails) with maximum yield of transformed parasites. Additionally, better separations were obtained when using an ice-cold solution of Percoll rather than one kept at room temperature.

The second aspect to be evaluated was the fraction of viable parasites obtained with the mechanical transformation protocol. The criterion used to catalogue viable individuals was introduced in Chapter 2 – section 2.2.4. Initially, 23 passages through a syringe needle were used to generate the separation of the heads and tails. This resulted in a major fraction of parasites showing characteristics of non-viable individuals (**Figure 4.2A,** ~15%): some would be dead, showing a uniform shape with granular appearance and lack of motility; while others would show evidence of morphological damage, also showing a granular appearance and adopting different shapes usually with some level of motility. By

introducing a step of vigorous shaking in a vortex the number of passages through the syringe could be reduced from 23 to 12 obtaining similar numbers of MT parasites with an increased proportion of viable individuals (from ~15% non-viable to just ~1%, **Figure 4.2B**). This improvement was probably due to the reduction in the number of passages through the syringe needle, which may represent a very harsh treatment to the parasites.

Regarding the skin transformation protocol, the main factor affecting the recovery of tail-free preparations of schistosomula was the number of cercariae deposited in the upper compartment of the transformation apparatus. Numbers ranging from 4,000 to 20,000 cercariae were tested and a maximum yield of transformed parasites, virtually free from tail contamination (~ 1 to 4%), could be obtained by placing approximately 14,000 cercariae in the upper compartment of the transformation apparatus. However, tail contamination varied greatly from one experiment to another even among transformation apparatuses within the same experiment. Therefore, schistosomula recovered from individual assemblies were observed under the light microscope to assess contamination and only those with < 4% tails were considered for further experimentation. The sample preparation procedure had to be undertaken approximately 20 times (with 12-13 transformation apparatuses each time) before sufficient number of parasites could be collected.

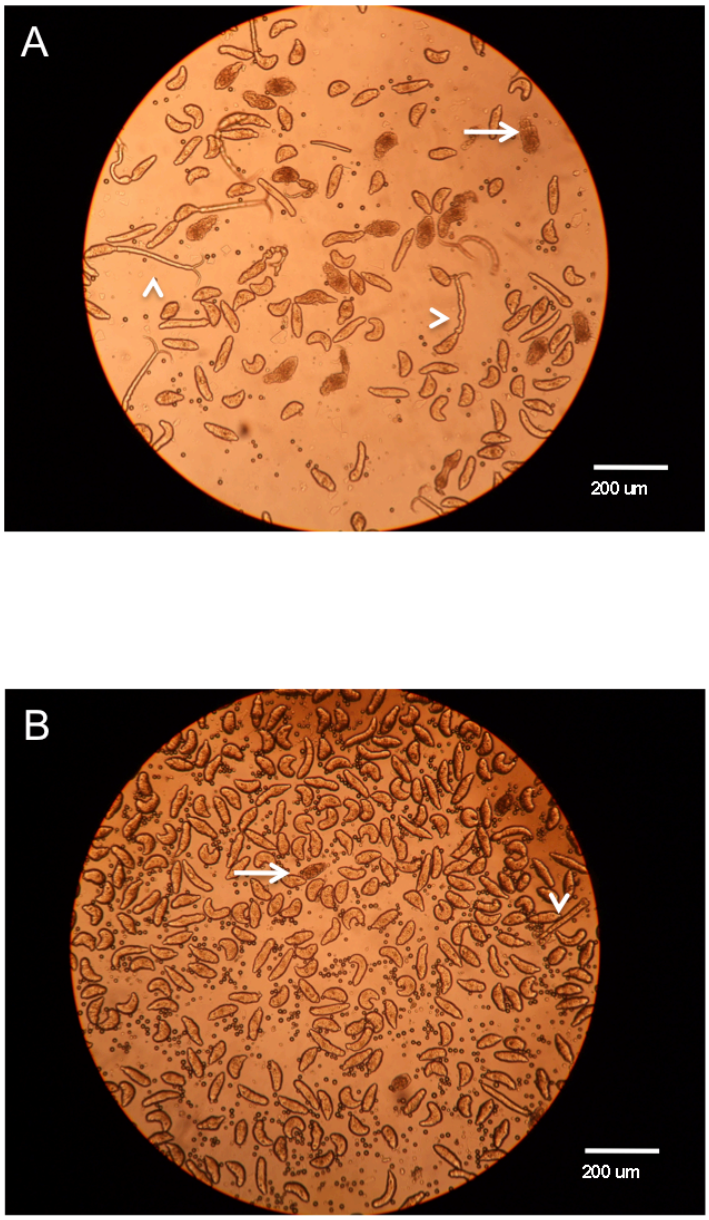


Figure 4.2 – Improving the mechanical transformation protocol. A – Parasites were transformed according to protocol described in Chapter 2 section 2.2.2 except that parasites were subjected to 23 passages through a syringe needle and separation of heads and tails was performed with 60% Percoll solution at room temperature. B - Transformation of parasite was performed using the optimized protocol described in Chapter 2 section 2.2.2 which involved shaking of the cercariae suspension in a vortex mixer followed by only 12 passages through a syringe needle. Separation of cercarial heads from tails was performed using 70% ice-cold Percoll solution. Arrows indicate damaged schistosomula, arrowheads point to contaminating cercaria or tails. Both panels (A and B) show light microscope images of 3-hour old mechanically transformed schistosomula.

4.2.1.2 Parasite viability

The quality of the transformed parasites, obtained either through MT or ST, was tested by phenotypic evaluation in an inverted microscope. Apart from testing for the characteristics of non-viable parasites (introduced in the Section 4.2.1.1), assessing the parasites' capability to remain alive when cultured *in vitro* for a given period of time can also be used as a parameter for evaluating fitness of parasites. Both MT and ST schistosomula were successful in progressing to the 2-weeks old stage. However, at this point both parasite populations were more phenotypically heterogeneous compared to those at 24 hours. These preparations contained parasites whose shape resembled the typical 2-weeks old worm grown in culture (much longer and thinner than the 24 hours schistosmulum) as well as parasites that showed a much more contracted shape resembling 3- or 24-hour old schistosomula. Microscopic evaluation of the phenotypic characteristics of both MT and ST parasites fulfilled the criteria used by others (Crabtree *et al.*, 1980) to define viable parasites at the lung stage indicating that both MT and ST protocols generated viable organisms.

4.2.1.3 Changes to the parasite surface

As introduced in Chapter 1 section 1.2.3, the parasites change the composition of their surface upon transformation. It has been shown that newly transformed schistosomula bind concanavilin A (Con A) to their surface (Samueleson *et al.*, 1989). This characteristic was used to evaluate whether the MT and ST transformation protocols triggered the remodelling of such structures. **Figure 4.3** shows Con A-FITC binding to both MT and ST schistosomula 5 hours after transformation. Consistent with previous reports (Samuelson *et al.*, 1982), Con A binding is positive for schistosomula but negative for the cercarial tails. Both MT and ST parasites bind Con A with the typical differential distribution of the Con A binding sites. Pronounced binding of Con A is observed at the apical end and the ventral sucker of the schistosomula, consistent with previous observations (Samuelson *et al.*, 1982).

A difference in the intensity of the fluorescence between samples (more in the MT parasites) can be seen in **Figure 4.3**. This might be due to differences in the sample itself or in the image acquisition procedure. Nevertheless, both samples positively bind Con A indicating that they have indeed transformed their surface membrane.

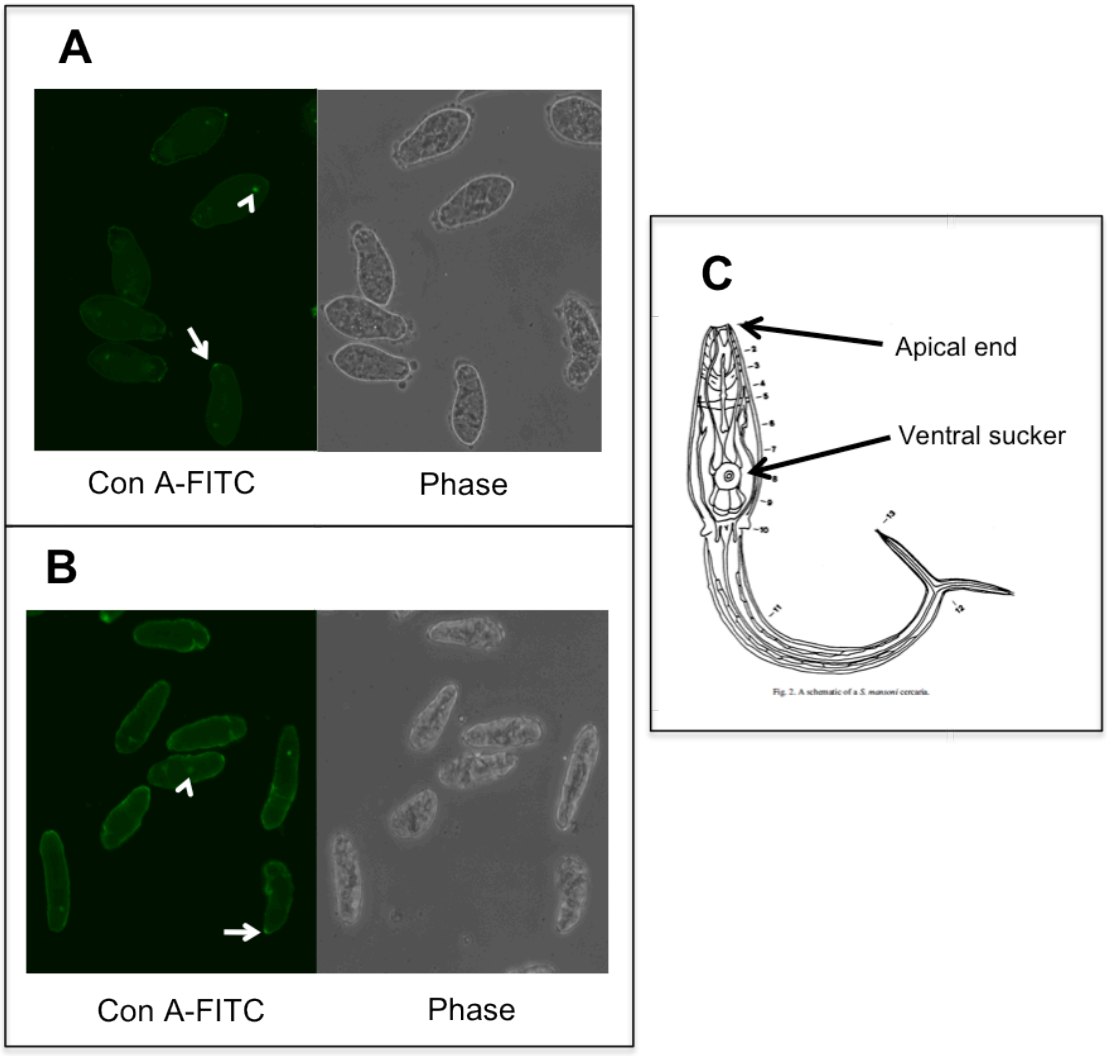


Figure 4.3 – Concanavilin A (Con A) binding to the surface of schistosomula but not to the cercariae (not shown) as a consequence of the uncovering of epitopes as evidence of transformation. Con A binds to both ST (A) and MT (B) schistosomula in similar fashion. (C) is included as a reference to the cercariae anatomy. Increased binding of Con A-FITC is observed in the apical end (arrows) and ventral sucker (arrowheads) of the parasites. Top and bottom panels are replicates of each sample. Note that the shape of the parasites in (A) and (B) are different and this is due to the sample preparation (see text).

A slight difference in the schistosomula size in the two samples can be observed in **Figure 4.3**. This was due to the differential conditions in which the samples were prepared for observation. MT schistosomula were left to move freely therefore appearing more elongated and slender, while the ST parasites were photographed using a cover slip limiting their movement and appearing rounder and shorter than the MT. It is not expected that these differences in specimen's preparation would affect the observations.

In summary, observations obtained from analysis and Con A binding as well as the parasite's capability of surviving for at east 2 weeks in culture led to the conclusion that the methods used to transform cercariae into schistosomula, both MT and ST, provide the necessary cues to trigger the transformation from the free-living cercariae to the parasitic schistosomula.

4.2.2 Differential expression between mechanical and skin transformed schistosomula.

After assessing that both parasite preparations fulfilled the criteria of schistosomula, samples were collected, RNA was extracted and RNA-seq libraries were prepared as described in Chapter 2.

In order to analyse the differences between MT and ST parasites at the transcriptional level, RNA-seq data for 24-hour old MT and ST schistosomula were generated. An overview of the results obtained from sequencing these libraries (number of reads, percentage mapped to genome, etc) were presented in Chapter 3 section 3.2.2. Reads per transcript and RPKM values for each gene were calculated as described in Chapter 2 section 2.6.1.1. Analysis (**Figure 4.4**) of the two 24-hour old schistosomula samples showed that they are highly correlated with Pearson's product and Spearman's rank coefficients of 0.98 and 0.99 respectively. Indeed, these correlation values resemble those obtained for biological replicates as previously shown in Chapter 3 section 3.2.2.1.3. Previous studies, for example the one from Chai *et al.,* (2006) presented in the introduction of this chapter, used deviations from the correlation to identify differentially expressed genes. Such approach cannot be used in this study because of the high correlation of the samples.

The software package edgeR (Robinson *et al.*, 2010) was used to achieve a more detailed analysis of transcripts differentially expressed between these two samples. First, all transcripts with levels of expression that could be attributed to noise were removed from the dataset by filtering out all transcripts with RPKM values lower than the

empirically determined background (see Chapter 2 section 2.6.2), corresponding to an RPKM of 2. This reduced the total number of transcripts from 10,852 to 8,715 (2,137 transcripts had negligible expression in both samples). After filtering the reads and using a statistical significance cut-off of adjusted p-value < 0.01, edgeR revealed 77 differentially expressed transcripts. By loosening the statistical criterion (adjusted p-value < 0.05) a total of 149 differentially expressed transcripts were found. Of these, 93 transcripts were more highly expressed in the ST parasites (**Table 4.1**) while 49 transcripts were more highly expressed in the MT (**Table 4.2**). A graphical representation of differentially expressed transcripts is shown in **Figure 4.5**).

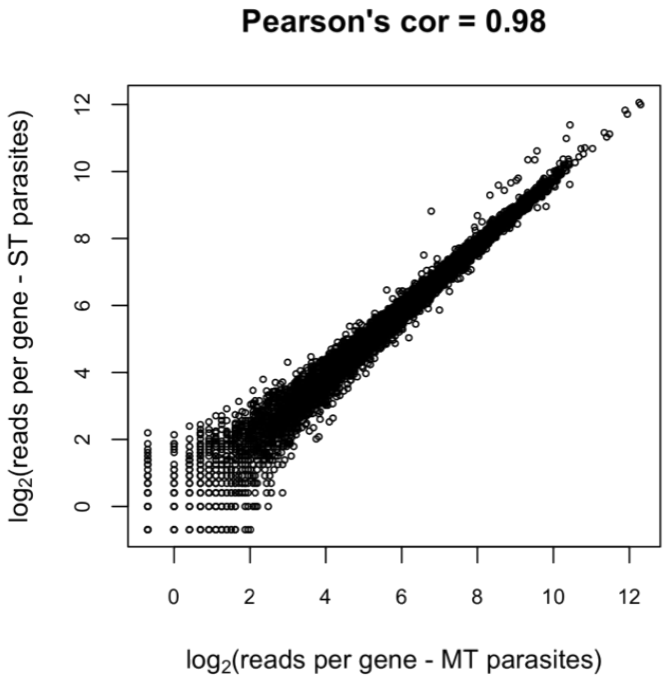


Figure 4.4 – Correlation of expression values of transcripts for the mechanically transformed (x-axis) and skin transformed (y-axis) schistosomula. Both Pearson and Spearman's correlations are high (0.98 and 0.99 respectively) indicating very low variability between these two samples.

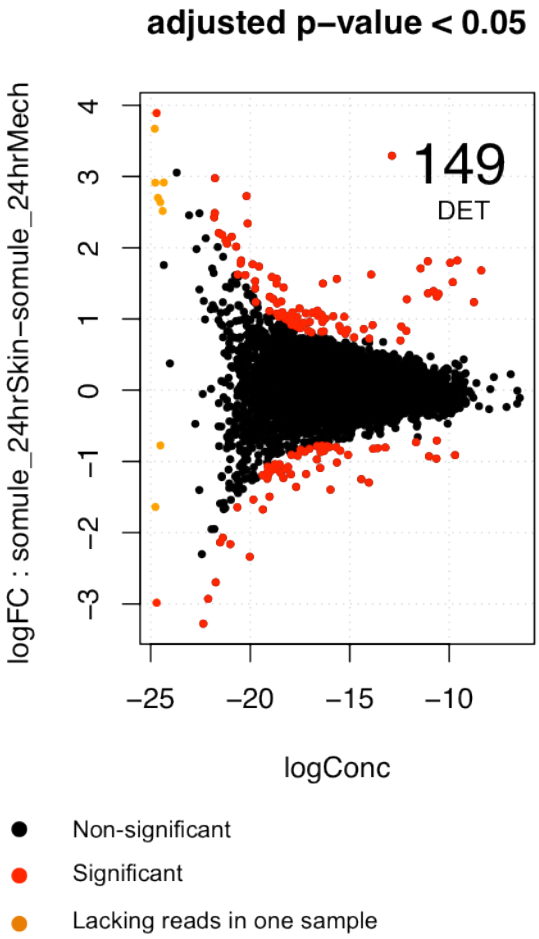


Figure 4.5 - Differential expression of transcripts. The comparison between MT and ST parasites at 24 hours after transformation (adjusted p-value < 0.05) is shown. Relative concentration (x axis) is plotted against fold change values (y axis) in the $\log_2$ scale. Positive $\log_2$ fold changes represent transcripts more highly expressed in ST schistosomula while negative $\log_2$ fold changes represent transcript more highly expressed in MT schistosomula. Red dots indicate differentially expressed transcripts below the statistic significance cut off value and therefore statistically differentially expressed; black dots represent transcripts above the statistic significance cut off value of differential expression; orange dots represent transcripts expressed in one sample but not in the other in which cases both the fold change value and the relative concentration are estimated. DET = differentially expressed transcripts.

Table 4.1 - Transcripts over expressed in ST *vs*. MT

91 genes; adjusted p-value cut-off is 0.05

GeneDB_ID	log_2FC	Product description
Smp_197430.1	28.57	Hypothetical protein
Smp_900100.1	3.29	NADH dehydrogenase subunit 3
Smp_199840.1	2.99	Nucleolar protein c7b, putative
Smp_204970.1	2.64	Na
Smp_144640.1	2.48	Hypothetical protein
Smp_172770.1	2.31	Hypothetical protein
Smp_200080.1	2.31	Na
Smp_119730.1	2.23	Hypothetical protein
Smp_205470.1	2.17	Na
Smp_202510.1	2.10	Na
Smp_139420.1	2.07	Hypothetical protein
Smp_177710.1	2.04	Hypothetical protein
Smp_205950.1	1.99	Na
Smp_900040.1	1.82	NADH dehydrogenase subunit 2
Smp_202050.1	1.82	Na
Smp_900020.1	1.81	NADH dehydrogenase subunit 6
Smp_900110.1	1.79	NADH dehydrogenase subunit 1
Smp_029780.1	1.76	Hypothetical protein
Smp_197440.1	1.75	Hypothetical protein
Smp_149340.1	1.73	Hypothetical protein
Smp_900060.1	1.71	Cytochrome c oxidase subunit 3
Smp_900050.1	1.68	NADH dehydrogenase subunit 5
Smp_028850.2	1.62	Hypothetical protein
Smp_151800.1	1.62	Hypothetical protein
Smp_159800.1	1.61	MEG-2 (ESP15) family
Smp_202920.1	1.60	Na
Smp_202120.1	1.56	Na
Smp_096750.1	1.54	Hypothetical protein
Smp_900090.1	1.52	NADH dehydrogenase subunit 4
Smp_169830.1	1.51	Hypothetical protein
Smp_159810.1	1.51	MEG-2 (ESP15) family
Smp_127860.1	1.46	Hypothetical protein
Smp_205870.1	1.46	Na
Smp_131830.1	1.40	Hypothetical protein
Smp_067800.1	1.39	Fibrillin 2, putative
Smp_900030.1	1.36	Atpase subunit 6
Smp_900070.1	1.35	Cytochrome B
Smp_900010.1	1.31	Cytochrome c oxidase subunit 2

Table 4.1 - Transcripts over expressed in ST *vs*. MT (cont)		
Smp_201190.1	1.31	Na
Smp_157330.1	1.30	Hypothetical protein
Smp_900080.1	1.28	NADH dehydrogenase subunit 4L
Smp_203200.1	1.27	Na
Smp_056260.1	1.26	Beta-1,4-galactosyltransferase, putative;with=uniprot:Q9GUM2
Smp_900000.1	1.24	Cytochrome c oxidase subunit 1
Smp_170650.1	1.23	Hypothetical protein
Smp_170630.1	1.15	Hypothetical protein
Smp_028840.1	1.13	Hypothetical protein
Smp_047400.1	1.11	Hypothetical protein
Smp_146940.1	1.09	Innexin, putative;with=Pfam:PF00876
Smp_146760.1	1.09	Hypothetical protein
Smp_131730.1	1.08	Hypothetical protein
Smp_107750.1	1.08	Hypothetical protein
Smp_133340.1	1.08	Hypothetical protein
Smp_007950.1	1.06	Beta-1,4-galactosyltransferase, putative;with=uniprot:Q80WN7
Smp_008400.1	1.06	Adenosine kinase, putative
Smp_195130.1	1.04	KRAB-A domain-containing protein
Smp_133660.1	1.04	Lin-9, putative
Smp_192220.1	1.04	Hypothetical protein
Smp_111640.1	1.02	Hypothetical protein
Smp_165250.1	1.02	Hypothetical protein
Smp_200450.1	1.02	Na
Smp_177250.1	1.02	Histone deacetylase, putative
Smp_204360.1	1.00	Na
Smp_183870.1	1.00	Hypothetical protein
Smp_142120.1	0.99	Achaete-scute transcription factor-related
Smp_200110.1	0.98	Na
Smp_168400.1	0.97	Hypothetical protein
Smp_169680.1	0.97	G-protein coupled receptor fragment, putative;with=uniprot:Q18179
Smp_194280.2	0.95	Hypothetical protein
Smp_193700.1	0.95	Hypothetical protein
Smp_200940.1	0.95	Na
Smp_051110.1	0.94	Hypothetical protein
Smp_144480.1	0.93	Hypothetical protein
Smp_117340.1	0.92	Hypothetical protein
Smp_022290.1	0.91	Hypothetical protein
Smp_131710.1	0.91	Hypothetical protein
Smp_151600.1	0.90	Neuronal calcium sensor, putative
Smp_057860.1	0.90	Hypothetical protein
Smp_126290.1	0.88	Hypothetical protein

Table 4.1 - Transcripts over expressed in ST *vs*. MT (cont)		
Smp_195030.1	0.86	ABC transporter subunit, putative
Smp_163800.1	0.85	Hypothetical protein
Smp_028860.1	0.83	Hypothetical protein
Smp_132670.1	0.83	Myosin regulatory light chain, putative
Smp_032970.1	0.83	Calmodulin, putative
Smp_164550.1	0.83	Hypothetical protein
Smp_039590.1	0.82	Hypothetical protein
Smp_166020.1	0.81	Hypothetical protein
Smp_155320.1	0.79	Hypothetical protein
Smp_144910.1	0.74	Collagen alpha-1(V) chain precursor, putative
Smp_163630.1	0.72	Expressed protein 10.3; MEG-4 (10.3) family
Smp_211020.1	0.70	Na

Na – Not assigned

Table 4.2 – Transcripts over expressed in MT vs. ST

58 genes; adjusted p-value cut-off is 0.05

GeneDB_ID	log_2FC	Product_description
Smp_180340.1	28.89	MEG-2 (ESP15) family
Smp_156200.1	3.30	Hypothetical protein
Smp_200150.1	2.78	Na
Smp_199820.1	2.66	Serine/threonine kinase;with=uniprot:P16912
Smp_204890.1	2.37	Na
Smp_116960.1	2.26	Hypothetical protein
Smp_172960.1	2.08	Kunitz-type protease inhibitor, putative;with=uniprot:P00978
Smp_199900.1	1.99	Phospholipid transport protein;with=uniprot:Q99J08
Smp_070940.1	1.62	Hypothetical protein
Smp_113760.1	1.61	Anti-inflammatory protein 16, putative
Smp_052760.1	1.56	TFIIH basal transcription factor complex TTD-A subunit (General transcription factor IIH polypeptide 5) (TFB5 ortholog), putative
Smp_203150.1	1.46	Na
Smp_047680.1	1.40	Ferritin, putative;with=uniprot:P25320
Smp_203400.1	1.34	Na
Smp_147730.1	1.29	Kunitz-type protease inhibitor, putative;with=uniprot:P00978
Smp_200290.1	1.28	Na
Smp_047660.1	1.25	Ferritin, putative;with=uniprot:P25320
Smp_013600.1	1.22	Nk homeobox protein, putative
Smp_201470.1	1.20	Na
Smp_018510.1	1.19	Hypothetical protein
Smp_159390.1	1.19	Hypothetical protein

Table 4.2 – Transcripts over expressed in MT vs. ST (cont)

Smp_195090.1	1.17	Hypothetical protein
Smp_195180.1	1.17	Surface membrane antigen;with=uniprot:Q04171
Smp_168610.1	1.17	Myelin transcription factor 1, myt1, putative
Smp_053950.1	1.16	Heterogeneous nuclear ribonucleoprotein, putative
Smp_205660.1	1.12	Na
Smp_146230.1	1.10	Hypothetical protein
Smp_125130.1	1.09	Hypothetical protein
Smp_170380.1	1.09	Hypothetical protein
Smp_100350.1	1.07	Hypothetical protein
Smp_198540.1	1.06	Hypothetical protein
Smp_120280.1	1.06	Hypothetical protein
Smp_079120.1	1.05	Hypothetical protein
Smp_204260.1	1.02	Na
Smp_147380.1	0.98	Hypothetical protein
Smp_113660.1	0.97	Cleavage and polyadenylation specificity factor, putative
Smp_002150.1	0.96	subfamily S1A unassigned peptidase (S01 family);with=UniProt:P20918
Smp_200850.1	0.93	Na
Smp_143190.1	0.92	hypothetical protein
Smp_089670.1	0.91	macroglobulin/complement, putative;with=UniProt:Q63041
Smp_134870.1	0.91	early growth response protein, putative
Smp_139500.1	0.90	hypothetical protein
Smp_089550.1	0.86	Eif5b-like protein, putative
Smp_158480.1	0.86	AMP dependent ligase, putative
Smp_179170.1	0.85	family C13 non-peptidase homologue (C13 family);with=UniProt:P09841
Smp_161650.1	0.84	hypothetical protein
Smp_125710.1	0.83	hypothetical protein
Smp_146360.1	0.82	hypothetical protein
Smp_063330.1	0.82	hypothetical protein
Smp_194050.1	0.81	Clumping factor A precursor (Fibrinogen-binding protein A) (Fibrinogen receptor A), putative
Smp_133830.1	0.80	hypothetical protein
Smp_064280.1	0.80	hypothetical protein
Smp_184280.1	0.79	hypothetical protein
Smp_158650.1	0.78	hypothetical protein
Smp_000790.1	0.78	actin binding LIM protein family member 2-related;with=UniProt:Q6H8Q1
Smp_151640.1	0.77	hypothetical protein
Smp_201910.1	0.73	Na
Smp_124000.1	0.71	MEG-14

Na – Not assigned

4.2.2.1 Genes more highly expressed in skin-transformed schistosomula

4.2.2.1.1 Mitochondrial transcripts

The results shown in Table 4.1 indicate that the skin-transformed parasites have higher expression of mitochondrial genes. These are also shown in **Figure 4.6A** where they cluster together at relative high concentrations and more highly expressed in the ST parasites. Their linear fold change values range from 2.3 to 9.8 and they are all statistically significant.

In order to investigate whether this difference in the rate of expression of the mitochondrial genes had any consequence on the metabolic activity of the parasites, the AlamarBlue® (AB) assay was used. As introduced before (Chapter 2 section 2.3.9), AB is a good indicator of mitochondrial activity (Springer *et al.*, 1998) through the measurement of redox species generated by the respiratory electron chain.

As a first step, a titration was performed to find the minimum number of parasites needed to identify a significant difference in absorbance between wells containing parasites and a blank. This is necessary because the number of parasites obtained from ST is very low and optimization of the protocol is a critical step in order not to waste valuable samples. For simplicity, these titration experiments were only performed using MT parasites. Although a statistical difference was found between wells with parasites and the blanks, no significant difference in absorbance could be detected among wells with different numbers of 3-hour old schistosomula (**Figure 4.6B**). However, experiments using 24-hour old parasites showed a significant difference between each population of parasites (250, 500 and 1000 parasites per well) and their blank. It was concluded that 250 parasites per well are sufficient to detect a difference in absorbance reflecting the metabolic activity in parasites compared to blanks at 24 hours after transformation.

Second, MT and ST parasites were tested (**Figure 4.6C**). A significant difference (t-test, p-value < 0.01) could be observed between 6-hour old parasites originating from MT and ST (red boxplots). Both 6-hour old MT and ST parasites were different from the blank and from each other, suggesting that parasites at this stage are already distinguishable in terms of metabolic activity. Interestingly, MT and ST parasites at 24 hours after transformation do not show significant difference between them (green boxplots) when incubated in AB for 3 hours. However, they are significantly different from their blank indicating that AB has reacted with species present in the sample. Increasing the

incubation time in AB is known to provide better resolution simply because there is a higher concentration of species that react with AB (manufacturer instructions). As expected, when parasites were left to grow for 24 hours in the presence of AB they showed a much more pronounced increment in the accumulation of metabolites (compared to blank wells) that reflected in a greater difference in the absorbance with respect to the blank. What is more, a significant difference between MT and ST could be established, suggesting that these two populations of parasites are indeed metabolically different with the ST parasites being more "metabolically active" than the MT after 24 hours in *in vitro* culture.

In summary, the differential expression analysis indicated an over expression of mitochondrial transcripts in the ST compared to the MT parasites. Increased numbers of mitochondrial transcripts could be responsible for the observed increment of mitochondrial activity, either by increased protein levels or because a higher number of active mitochondria are found in this parasites. This observation agrees with the explanation proposed by Brink *et al.,* (1977) – discussed earlier – that ST parasites are a selection of the most "fit" cercariae. If MT preparations contain a mixture of fit and less-fit parasites not all of them would be expected to develop at their maximum rate and, as seen here, their measurable metabolic activity would be reduced.

4.2.2.1.2 Other biologically relevant genes with higher expression in ST parasites

In order to identify other biological processes determined by the genes with higher expression in the ST parasites compared to the MT parasites, a GO term enrichment analysis of genes more highly expressed in the ST schistosomula was performed. These results are presented in **Table 4.3** and according to their description they can be further grouped as: mitochondrial function, G-protein couple receptors, microtubule movement, retrotransposon related functions, betaine synthesis and skeletal muscle development. An analysis of the higher rate of mitochondrial function inferred from the expression of mitochondrial genes was already presented in section 4.2.2.1.1.

Transcripts encoding putative G-protein couple receptors (GPCR - Smp_144640.1, Smp_117340.1 and Smp_169680.1) are more highly expressed in the skin-transformed parasites. Smp_144640.1 encodes a seven-trans-membrane domain and can be considered a putative GPCR receptor. The last two are better characterised: Smp_117340 is a PROF1 receptor and Smp_169680.1 is a neuropeptide receptor (Zamanian, 2011). PROF1 are a

group of GPCRs that are found exclusively in platyhelminths (Zamanian, 2011). Knowledge of the role of these GPCRs in helminths is still rudimentary; further experiments are needed to assess their roles in either cell-to-cell signalling or host-parasite interactions. Results presented here suggest that some of these GPCRs are more highly expressed when the parasites are transformed through a layer of skin.

Other GO enriched terms are microtubule-based movement. Two transcripts (Smp_131710.1 and Smp_170650.1) represent this term: the first of these transcripts is a small one-exon gene encoding a 891 amino acids polypeptide with no predicted conserved domains.

The second is a multi-exon gene encoding a conserved kinesin domain. Kinesin domains are involved in movement along microtubules and are often associated with proteins that have a role in transporting organelles along microtubules (Vale *et al.*, 2000). As previously described in the introduction of this chapter, chemoattractants found in the host skin induce the release of the contents of the acetabular glands (Stirewalt, 1978). Since these glands are unicellular, it would be expected that the vesicles containing secretory products (see Chapter 1 section 1.2.3.1) travel to the glands' openings transported by microtubules. Because the chemoattractant cues would be absent during the mechanical transformation the expression of these kinesin proteins may be induced at a lower level in MT parasites.

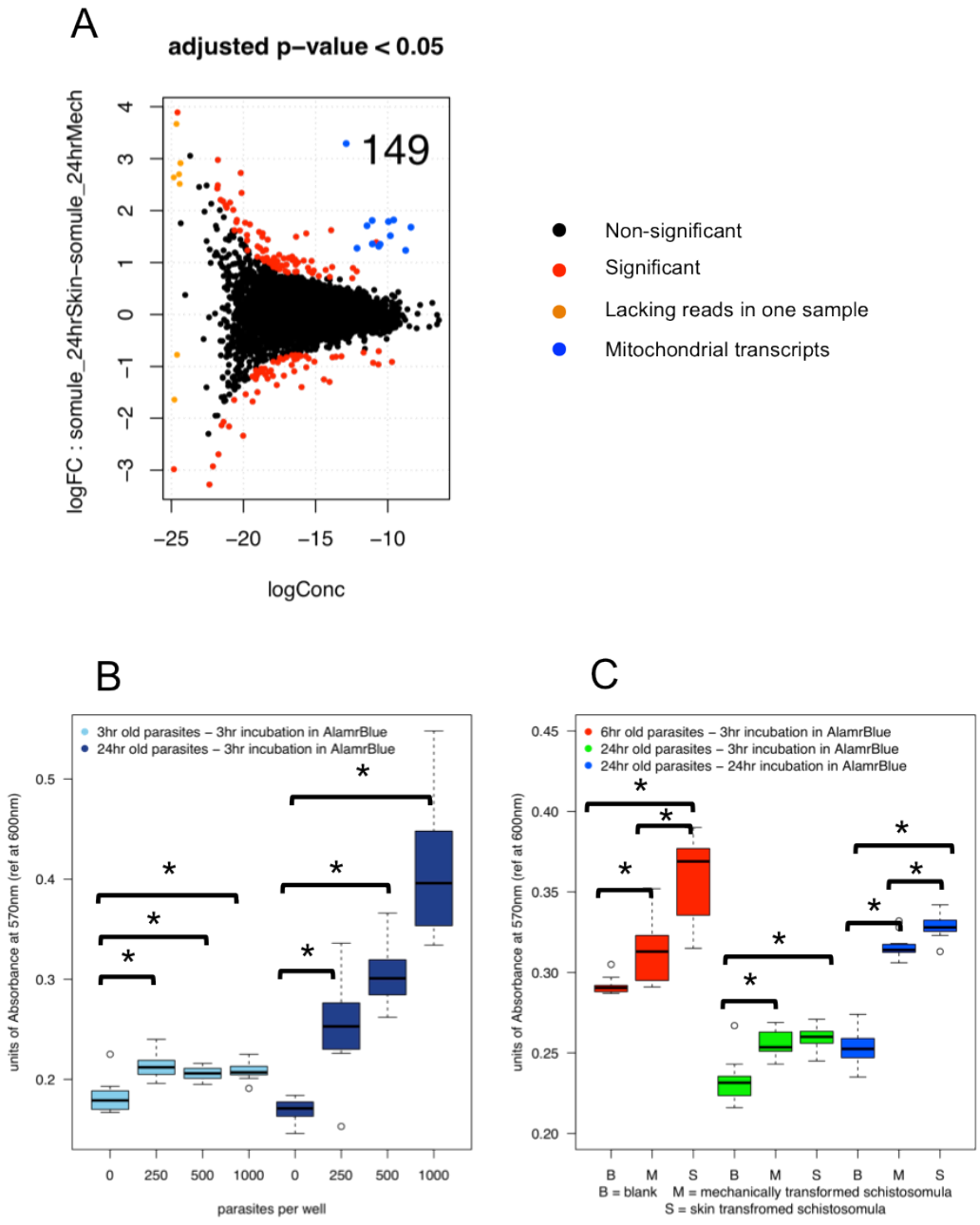


Figure 4.6 – Mitochondrial genes are differentially expressed in ST compared to MT. A – Mitochodrial transcripts (blue dots) are clearly separated from the rest of the differentially expressed transcripts. B – Titration of AlamarBlue® reactivity against different number of MT parasites at 3 hours after transformation (light blue) and 24 hours after transformation (dark blue). C – AlamarBlue® reactivity of 6-hour old and 24-hour old MT and ST schistosomula. Six-hours old with 3 hours incubation in AlamarBlue® (red boxplots) and 24-hours old schistosomula with 24 hours incubation in AlamarBlue® (blue boxplots) samples were statistically different from each other and compared to the blank. Twenty four-hour old schistosomula with 3 hours incubation in AlamarBlue® (green boxplots) were both statistically different from the blank but not between them (*t*-test, p.value = 0.25). Stars represent statistically different samples (*t*-test, p-value < 0.01).

Table 4.3 - Enriched Gene Ontology terms (Biological Processes).

A - 24 hours old SKIN transformed schistosomula

GO.ID	Term	p-value
GO:0022904	Respiratory electron transport chain	0
GO:0042773	ATP synthesis coupled electron transport	0
GO:0046797	Viral procapsid maturation [a]	0.008
GO:0015074	DNA integration [a]	0.011
GO:0032196	Transposition [a]	0.013
GO:0019285	Glycine betaine biosynthetic process from choline	0.014
GO:0006310	DNA recombination [a]	0.024
GO:0007018	Microtubule-based movement	0.032
GO:0019079	Viral genome replication [b]	0.037
GO:0006410	Transcription, RNA-dependent [b]	0.037
GO:0007186	G-protein coupled receptor protein signalling	0.040
GO:0042775	Mitochondrial ATP synthesis coupled electron transport	0.041
GO:0007519	Skeletal muscle tissue development	0.046

B - 24 hours old MECHANICALLY transformed schistosomula

GO.ID	Term	p-value
GO:0006879	Cellular iron ion homeostasis [c]	0
GO:0006826	Iron ion transport [c]	0.001
GO:0007548	Sex differentiation [d]	0.015
GO:0055114	Oxidation reduction [c]	0.018
GO:0006526	Arginine biosynthetic process	0.022
GO:0006269	DNA replication, synthesis of RNA primer	0.029
GO:0007530	Sex determination [d]	0.033
GO:0030154	Cell differentiation	0.037

Categories with the same super index letter are represented by the same transcripts.

Other differentially expressed transcripts not identified through the GO term analysis were also found biologically relevant in the process of transformation. EF-hand domains are calcium-binding domains and some can be associated to tegument antigens [such as Sm22.6/SmTAL1 (Dunne *et al.*, 1992; Fitzsimmons *et al.*, 2004)] in schistosomes. What is more, calcium-binding proteins have been associated in mechanisms of larval adaptation to the mammalian host (Kusel *et al.*, 2007). ST parasites show the over expression of four EF-hand encoding transcripts (see **Table 4.1** -Smp_151600.1 neuronal calcium sensor, putative; Smp_032970.1 calmodulin, putative; Smp_132670.1 myosin regulatory light chain, putative; Smp_137750.1 calbindin-32, putative). If these calcium-binding proteins have a role in the mechanisms of adaptation, it is possible that such mechanisms are triggered upon contact with the host skin. The lack/reduced expression of such transcripts in MT parasites might be associated with the lack of the appropriate signals.

4.2.2.2 Genes more highly expressed in mechanically-transformed schistosomula

In this section a description and analysis of the transcripts that are more highly expressed in the mechanically transformed schistosomula is presented. As done previously for the highly expressed transcripts in the ST parasites, GO term enrichment analysis was also used here as a guide for identifying biological processes among this group of genes. However, the study of the list of genes more highly expressed in the MT parasites alongside with their product description (**Table 4.2**) provided a good resource to associate these transcripts with processes that might be occurring in the parasites. Results emerging from these two sources are presented here.

4.2.2.2.1 Ferritins

The dataset of transcripts more highly expressed in the MT parasites present two transcripts associated with iron homeostasis and oxidation-reduction functions. These are ferritins and their main function is to bind iron and release it in a controlled way, avoiding the generation of reactive oxygen species (Chiancone *et al.*, 2004). In *S. mansoni* there are four ferritin genes (Smp_047640.1, Smp_047650.1, Smp_047660.1, Smp_047680.1) found in one cluster in Chromosome 2. The two ferritin genes found differentially expressed in the MT *vs.* ST (Smp_047660.1 and Smp_047680.1) share the higher percentage of similarity (98%) compared to the others in the cluster. Because 24-hours old schistosomula can ingest plasma molecules (Bennett *et al.*, 1991) it is possible that the role of ferritins is associated with uptake of iron from these sources or with later uptake of blood cells. This hypothesis is in agreement with the presence of ferritins in the gut vomit (Hall *et al.*, 2011), which suggests a role of these proteins in the process of feeding. However, it is not possible to infer why these transcripts are more highly expressed in the MT parasites.

4.2.2.2.2 Proteases

MT parasites show higher expression of transcripts encoding proteases. Proteases have a recognised important role in schistosomes. For example, adult worms use a set of aspartic proteases called cathepsins for the purpose of feeding (Brinkworth *et al.*, 2001) while cercariae use elastase and other proteases during the process of skin invasion (McKerrow *et al.*, 2002; McKerrow, 2003). Two proteases are shown more highly expressed in MT than in ST parasites.

The first one is encoded in the transcript Smp_002150.1, which is twice as much expressed in the MT parasites than in the ST parasites. The polypeptide product is a putative secreted serine protease (S1 family) from the trypsin family characterised by the conserved catalytic triad His/Asp/Ser (chymotrypsionogen numbering) towards their C-termini. Similarity searches using BLASTp (Altschul *et al.*, 1990) against the Uniprot database (Uniprot Consortium, 2009) showed similarity with other serine proteases in *S. japonicum, T. solium* and the *E. granulosus* antigen 5 precursor (Ag5). The latter is a well-characterised secreted antigen present in large quantities in the hydatid cyst fluid in the infected intermediate host. Interestingly, the serine in the catalytic triad of the *E. granulosus* Ag5 has been substituted by a threonine and probably compromises the serine protease activity of the enzyme (Lorenzo *et al.*, 2003). Contrary to Ag5, Smp_002150.1 has all residues from the catalytic site.

The second protease is Smp_179170.1, encodes a legumain C13 asparginyl endopeptidase domain. It is highly similar to another *S. mansoni* asparginyl endopeptidase Sm32 (Smp_075800 – former hemoglobinase), which is known to participate in activating gut-related proteins [cathepsin B and F (Dalton *et al.*, 1995)]. These two genes are located adjacent to each other approximately 8 kB distance in Chromosome 3. SignalP results (Emanuelsson *et al.*, 2007) suggest that Sm32 is secreted (0.96 probability) while the same prediction for the product of Smp_179170.1 resulted in inconclusive (0.5 probability). Closer analysis of their protein sequence revealed that they are highly similar in both protein (98%) and mRNA (93%) sequences. Their intronic sequences are different suggesting this is not an assembly error but a possible gene duplication. As these transcripts share only the last half (corresponding to the last 177 aa) of Sm32, Smp_179170.1 seems to be a truncated version of Sm32. However, it encodes the protein domain characteristic of these endopeptidases. Because of their high similarity, it is not possible to say which of the transcripts is actually more highly expressed in the MT schistosomula. However, it is possible to say that mRNA encoding this particular endopeptidase domain is differentially expressed (whether it comes from the expression of Smp_179170 or Sm32) indicating that, provided the protein in synthesised, this asparaginyl endopeptidase function is present in the parasites. Sm32 has been localized to the acetabular glands of the cercariae, the gut lumen and gut epithelium of the adult worm using northern and western blotting, suggesting a role of both nutrition and defense/invasion for this endopeptidase (el Meanawy *et al.*, 1990). Given the high sequence similarity of Smp_179170.1, it is speculated it could carry out similar functions.

Because proteases are important in the process of invasion, one of the hypotheses is that these proteases are remnants of those used by the cercariae during the skin penetration process. The pattern of expression for both transcripts indicates that they are developmentally up regulated after the transformation in schistosomula (**Figure 4.7**) discarding the former hypothesis. Smp_002150.1, the putative trypsin protease, shows an impressive 30 times fold increase (linear scale) between 3-hours and 24-hours old parasites; while the asparaginyl endopeptidase Smp_179170.1 is up regulated almost 4 times between 3-hours old schistosomula and cercariae.

As for the ferritins, no hypothesis can be generated as this point that would explain why these proteases are more highly expressed in the MT parasites.

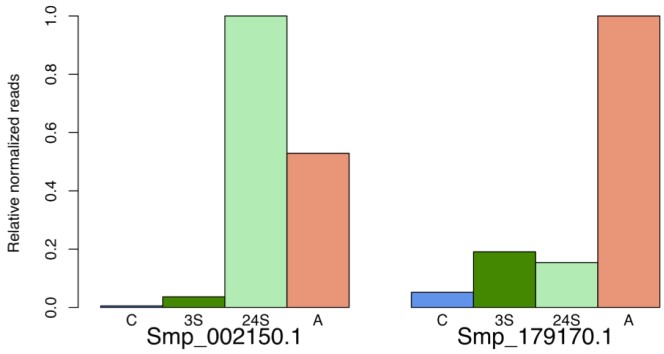


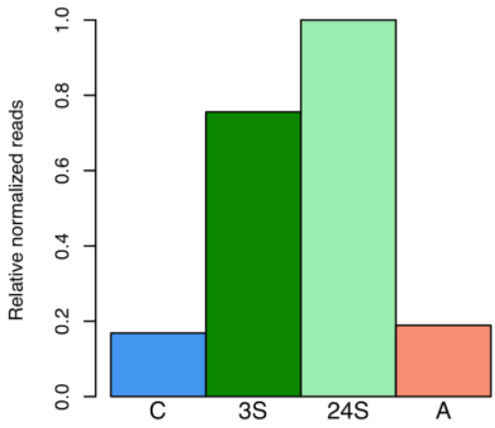
Figure 4.7 – Differential expression of two proteases with higher expression in ST parasites.

4.2.2.2.3 Protease inhibitors

It is interesting to find that protease inhibitors are also differentially expressed alongside proteases. Protease inhibitors can neutralise the action of host- and/or parasite-derived proteases. There are two functional classes of protease inhibitors: the "active-site" inhibitors which bind to the active site of the protease with such affinity that it inactivates their activity, and the alpha-macroglobulins, which modulate the protease activity through an entrapment mechanism [reviewed in (Armstrong, 2006)]. Macroglobulins are also capable of inhibiting coagulation by the inhibition of thrombin (de Boer *et al.*, 1993). ST schistosomula show higher expression of two protease inhibitors, one from each type.

Smp_089670.1 encodes a 1,800 amino acids polypeptide with high similarity to an alpha-macroglobulin protein. Closer examination of the protein sequence showed that it has the basic structure of an alpha-macroglobulin protein: it encodes a signal peptide and

5 functional domains: 2 macroglobulin domains, a thiol-ester bond-forming region with the sequence GCGEQNM strictly conserved among alpha-macroglobulins (Armstrong, 2006), an A-macroglobulin complement component and a macroglobulin receptor binding site. An interesting difference between the *S. mansoni* alpha-macroglobulin and its homolog proteins in other organisms is the gene organisation: in *S. mansoni* this gene is encoded in only two exons while in rat and in *C. elegans* this gene is encoded in 36 and 17 exons respectively.



comp	log2FC	adj–p.val
3S/C	2.1648	1.085E–04
24S/3S	0.4037	7.260E–01
A/24S	–2.4025	1.187E–02
24S/C	2.5686	1.342E–06

Figure 4.8 – The alpha-macroglobulin transcript Smp_089670.1 is highly expressed in the schistosomula stage (mechanically transformed parasires only). The barplot (top) shows the relative expression in each life cycle stage. C, cercariae; 3S, 3 hours old schistosomula; 24S, 24 hours old schistosomula; A, mix sex adult. The table (bottom) shows Log_2 fold changes for each of the comparisons and the corresponding adjusted p-values.

The mechanism of action of macroglobulins is complex but its explanation is necessary to understand why this protease inhibitor may have a *regulatory* rather than inhibitory effect. Macroglobulins function as a cage for proteases: proteases cleave the exposed "bait site" and this produces a rapid conformational change. This has two consequences; first the conformational change is such that the protease is now trapped inside the cage formed by the macroglobulin. The entrapment of the protease regulates its activity in the sense that the protease is still active but only on substrates that are small enough to enter the cage and interact with the protease. Second, the conformational change causes the macroglobulin receptor-binding site to be now exposed; this guarantees that the "used" macroglobulins are taken out from circulation [reviewed in (Armstrong, 2006)].

It is worth mentioning that Parker-Manuel *et al.,* (2011) found this transcript up regulated in the 3-days old schistosomula compared to cercariae and Hall *et al.,* (2011) found this protein in proteome analysis of the gut contents of adult worms. RNA-seq data presented in this thesis indicate that Smp_089670.1 has its peak of expression in the 24-hour old MT schistosomula where it is up regulated compared to the cercariae stage (**Figure 4.8**).

Another inhibitor of protease activity is Smp_147730.1. This is a kunitz-type inhibitor and it is just one of a larger family of kunitz domain containing proteins. These are known for their capability to inhibit serine proteases and are thought to have a role in the host parasite-interaction in the cestode *E. granulosus* (Gonzalez *et al.*, 2009). Smp_147730.1 fulfils the criteria that define a kunitz inhibitor protein: it is small (~150 amino acids) and contains both a signal peptide and a kunitz inhibitor domain with six conserved cysteines that form disulfide bonds. Kunitz-inhibitors act by competing for the active site of serine proteases. Although it has expression below background (RPKM < 2) in cercariae, it shows rapid up regulation as early as 3 hours after transformation. At 24 hours after transformation, its expression reaches 55-fold (linear scale) compared to cercariae. The early up regulation at the onset of infection suggests an important role of the kunitz inhibitor in the development of the schistosomula. Similar increments are observed for the ST sample compared to cercariae but with slightly less magnitude: the fold change between MT over ST being 2.3 fold. Interestingly, serine protease inhibitors (also called serpins), as well as the previously described alpha-macroglobulin, have been identified as up regulated in 3 days old parasites (Parker-Manuel *et al.*, 2011) and the corresponding protein has been found present in the contents of the gut (Hall *et al.*, 2011). However, both

these records correspond to another serine protease (Smp_090080) and not a kuntiz-type inhibitor.

Results presented here suggest that some of the genes more highly expressed in the MT parasites may have a role in feeding or tissue invasion process. Ferritins, proteases and protease inhibitors are thought to have a role in uptake of nutrients or perhaps a role in host tissue invasion. These results complement those reported by Parker-Manuel *et al.,* (2011) - where the authors described up regulation of the same transcripts in 3-days old schistosomula – and agrees with functional studies where schistosomula were seen to actively ingest plasma proteins only 24-hours after transformation (Bennett *et al.*, 1991).

The reason why these transcripts are differentially more highly expressed in the MT compared to the ST parasites is not clear. One of the possibilities could be that the MT population has more parasites in synch with each other than the ST population. Therefore, all the transcripts that show rapid up regulation in the early schistosomula stage would appear as "not so expressed" in the ST parasites. Generating a time course experiment that would reflect several time points in the development o these parasites my help answer this question. Independently of whether digestive-related transcripts are differentially expressed between "types" of parasites, it is important to note that they are expressed in very early stages of the schistosomula transformation; and that their expression is independent from the presence of a skin barrier. Further research into the localization of this proteins in the schistosomula, for example by using WISH (whole *in situ* hybridization), would shed more light on the functions of these genes.

4.2.2.3 Microexon genes are expressed at 24 hours post-transformation

Study of the differentially expressed transcripts presented in **Table 4.1** showed six microexon genes appearing as differentially expressed in both the MT and ST parasites.

Microexons are small exons of < 36 bp that are found in genes present in many eukaryotes (Volfovsky *et al.*, 2003). Typically, microexons contribute a small proportion of the coding sequence of genes but in *S. mansoni* microexons are found as mayor components (~75%) of the coding sequence of some genes; these genes are therefore called microexon genes (MEGs). They were first reported in Berriman *et al.*, (2009) with a more detailed description in DeMarco *et al.*, (2010). The most prominent characteristics of these genes are:

- Approximately 75% of the coding region is encoded in microexons.
- Microexons are typically < 36 bp, with bases in multiples of 3.
- MEGs have conventional (> 36 bp) exons and /or UTRs in their 3' and 5' ends.

- They share little or no sequence similarity among them.

Despite the last point, microexon genes are grouped in families based on the criterion that members of a family should have a BLASTp e-value < $1e^{-4}$ (Berriman *et al.*, 2009). No MEG homologs could be found in other species except for *S. japonicum* and *S. haematobium*.

MEGs show expression of several transcript variants for each locus at a given time. This is achieved by microexon skipping that, because each microexon is a multiple of three, does not change the ORF of the rest of the sequence. This has led to the hypothesis that splice variants change along the life cycle of the parasites providing them with the potential to generate antigenic variation of MEG product (DeMarco *et al.*, 2010).

Six MEGs from three different families appeared as differentially expressed in the comparison of MT *vs.* ST schistosomula at 24 hours after transformation. A summary is presented in **Table 4.4** and **Figure 4.9A**.

Table 4.4 – Summary of MEGs found as differentially expressed in ST and MT.

GeneDB_id	Gene name	Log2-FC	RPKM MT	RPKM ST
Smp_159800.1	MEG-2.2	1.50	4.7	14.2
Smp_159810.1	MEG-2.4	1.33	20.8	59.2
Smp_180340.1	MEG-2.?	-28.93	3.27	< 2
Smp_124000.1	MEG-14	-0.72	2275.5	1396.1
Smp_163630.1	MEG-4	0.93	120.5	197.8
Smp_138080.1	MEG-3	0.96	245.8	161.8

Smp_124000 is a member of the MEG-14 family (DeMarco *et al.*, 2010) where Smp_124000.1 is the only annotated splice variant in GeneDB. Although TopHat reads alignment for this model broadly agrees with the current annotation (**Figure 4.9B**), the profile of microexons expressed in this variant is different from the one present in the database, suggesting this is a different variant compared to that previously identified. This new splice variant shows three extra microexons (black arrows) and lacks the expression of others (not shown). Interestingly, the transcript expressed in the ST sample differs from that expressed in the MT sample. There are three exons clearly expressed in the ST but

absent in the MT schistosomula sample (**Figure 4.10**). Analysis of the transcript expression level showed that this is high in the schistosomula stage and relatively low in cercariae, the same pattern as previously reported (DeMarco *et al.*, 2010). The sampling of close time points in this experiment suggests that this MEG-14 variant is up-regulated as early as 3 hours after transformation (**Figure 4.9A**). Its expression peaks, at least within the samples here studied, at 24 hours post transformation and it is more highly expressed in the MT compared to the ST. Adult worms present low levels of expression of this gene. From the group of MEGs found to be differentially expressed in MT and ST, this is the only one with higher expression in MT.

Smp_163630.1 and Smp_138080.1 are members of the MEG-4 and MEG-3 families respectively. They both have higher expression in the ST samples with similar fold changes (nearly twice as much expression in the ST compared to MT). As in the previous example, close inspection of the reads alignment suggests that although the reads broadly support the current gene models, some differences are noted and these suggest the presence of a different variant from the one annotated in GeneDB (GeneDB, 2011).

The current annotation for Smp_163630.1 shows 16 exons, of which 14 (all except the 3'- and 5' most exons) are microexons. RNA-seq data showed expression of eight exons of the current annotated model for Smp_163630 plus two new exons. This new variant has its peak of expression at 3 hours post transformation, representing another example of an early up regulated transcript. Its expression is down regulated at 24 hours post transformation and goes up again in adult worms to a level comparable to that seen at 3 hours (**Figure 4.9A**).

Smp_138080.1 has 20 exons of which only 15 are expressed in this life cycle stage; RNA-seq data suggest that the 5 exons at the 5'end of the models are not expressed. This member of the MEG-3 family is up regulated in 3-hours old schistosomula [(DeMarco *et al.*, 2010) and **Figure 4.9A**] and has its maximum expression (within the samples here analysed) at 24 hours after transformation.

The other three differentially expressed transcripts are members of the MEG-2 family. Close inspection of the reads alignment showed that the coverage for these genes neither support the current gene models nor identify a new variant. The coverage plot and read alignment for Smp_159800.1 is illustrated in **Figure 4.10** as an example; Smp_159810.1 and Smp_180340.1 show the same pattern of read coverage (not shown). In order to investigate whether expression of the members of the MEG-2 family could be representing non-coding RNA species, the intronic regions with RNA-seq reads coverage were used to

query the Rfam database[1] (Gardner *et al.*, 2011) but no homology with known non-coding RNAs could be found.

According to the data presented here, only three (Smp_124000.1, Smp_163630.1 and Smp_138080.1) out of six MEGs identified as differentially expressed between MT and ST parasites show read alignments that support the annotated gene model. Data suggest that in two out of these three cases, the variants expressed are novel splice forms. What is more, in the case of Smp_124000, the variant expressed in the ST sample has three extra exons compared to the one expressed in the MT sample suggesting that the differential treatment received by these two populations of parasites may have an effect in the regulation of the expression of different isoforms.

4.2.2.4 Differentially expressed transcripts likely to be false positives

The functions represented by these transcripts vary greatly. In the case of genes more highly expressed in the ST parasites, transcripts encoding proteins related to retrotransposon functions, betaine synthesis and muscle development were identified. Other potentially interesting transcripts are those encoding an innexin, a betagalactosyltranserase and an ornithine decarboxylase.

In the genes found more highly expressed in the MT parasites, transcripts encoding proteins with DM domains (domain involved in sexual differentiation in *D. melanogaster*), a myelin transcription factor and a protein involved in arginine biosynthesis.

The gene structures, RNA-seq read coverage and similarity searches were performed for all the transcripts mentioned above. With exception of those encoding retrotransposon elements, all the rest of the transcripts showed either a model that could not be confirmed with RNA-seq data (usually because the reads coverage was very low) or they had no similarity with any other protein that could provide any information about their function. A summary of these transcripts and the reasons why they are not regarded as biological relevant is presented in **Table 4.5**.

[1] The Rfam database Gardner, P. P., J. Daub, J. Tate, B. L. Moore, I. H. Osuch, S. Griffiths-Jones, . . . A. Bateman (2011). "Rfam: Wikipedia, clans and the "decimal" release." Nucleic Acids Res **39**(Database issue): D141-145. holds a catalogue of known non-coding RNA motifs.

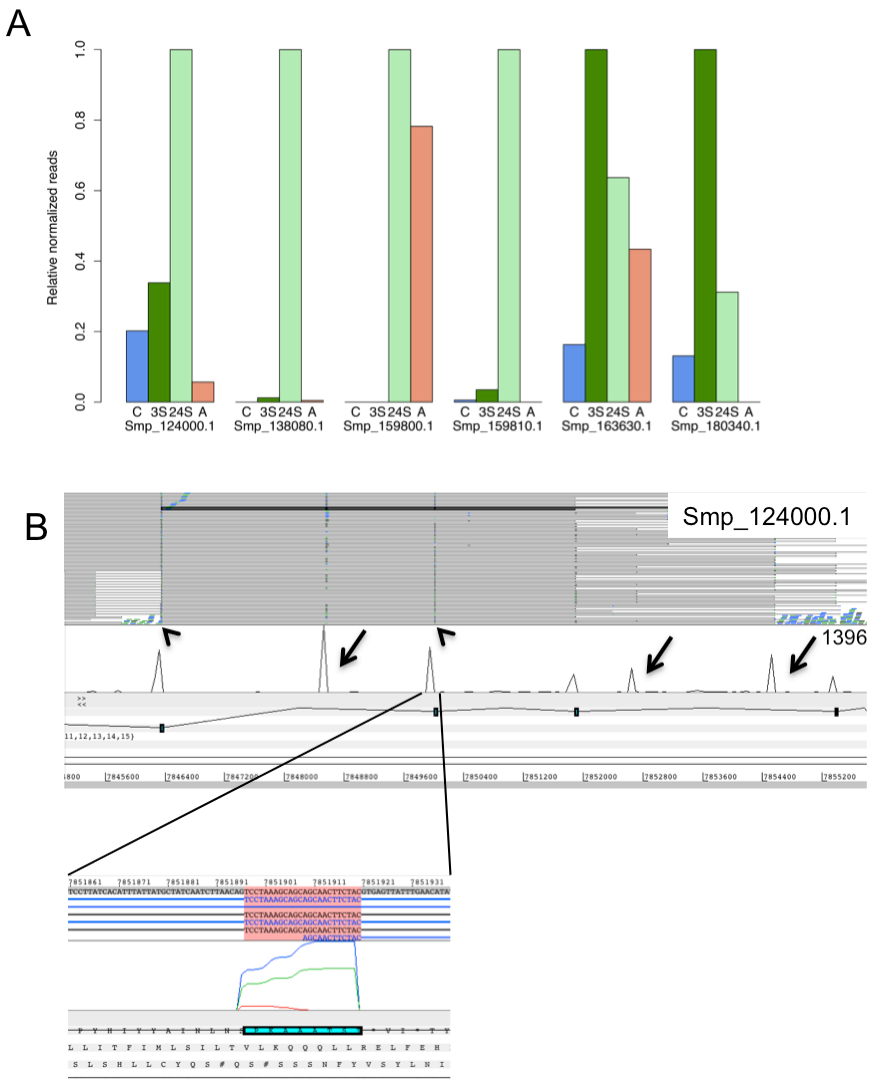


Figure 4.9 – Some microexon genes are differentially expressed in ST and MT parasites. A - Relative expression of microexon genes found differentially expressed in the MT *vs.* ST comparison; the bar plots show their relative expression across the life cycle time points surveyed in this thesis. C: cercariae, 3S: 3-hour old schistosomula, 24S: 24-hour old schistosomula (MT), A: 7-week old mixed sex adult worms. B - Artemis view of a region of the microexon gene Smp_124000. Top panel shows blue and green lines representing reads and grey lines representing split/paired reads (alignment of reads performed using TopHat); middle panel shows a plot representing the coverage of reads (maximum value of 1,512 reads). Arrows indicate the presence of novel exons expressed in this variant. Note how split reads are found mapped to micro exons (arrowheads). In some cases the coverage plot appears to be offset with the annotated exon; closer inspection of the alignment (zoom-in inset) shows that the offset is just an artefact of the visualization tool.

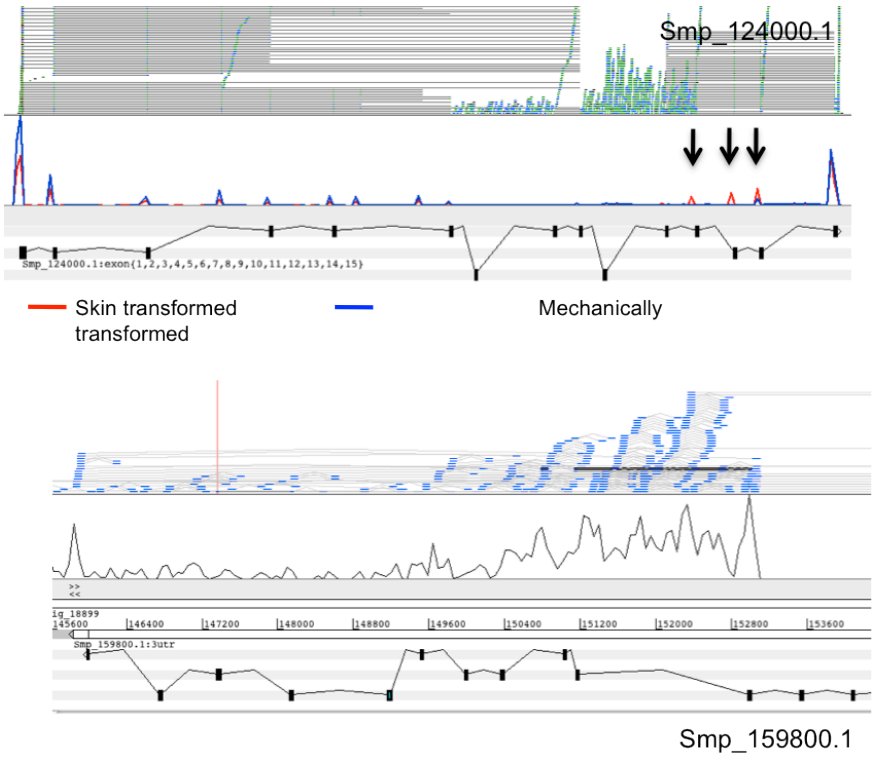


Figure 4.10 - Some microexon genes are differentially expressed in ST and MT parasites. Top panel: Artemis view of a region of the microexon gene Smp_124000 where the ST sample shows expression of three exons that are not present in the MT sample (arrows). Bottom panel: Artemis view of a region of the microexon gene Smp_159800. Top panel shows blue lines representing reads and grey lines representing split or paired reads. The middle panel shows a graphical representation of the coverage of reads. Note how there is no clear peaks of increased expression for the exons of this microexon gene compared to introns.

Table 4.5 –Some transcripts found differentially expressed between ST and MT samples have inaccurate functional annotation and are therefore not considered for the analysis. A summary of these is presented in this table.

Transcripts with higher expression in ST schistosomula		
GeneBD_ID	Function	Description
Smp_142120.1	Muscle development	DNA binding domain but no significant matches in protein database – function cannot be inferred.
Smp_212180.1	Glycine betaine biosynthesis	Encodes a choline dehydrogenase without any simililarities in protein databases except for a hit in *Clonorchis sinensis*.
Smp_007950.1	Glycosylation	Read coverage does not support the gene model – new gene model cannot be inferred
Smp_146940.1	Innexin	Gene model is truncated - encodes a trans-membrane domain of the four that are characteristic of innexin proteins
Smp_067800.1	Ornithine decarboxylase	No significant similarities; possible wrong functional annotation

Transcripts with higher expression in MT schistosomula		
GeneBD_ID	Function	Description
Smp_143190.1	Sex determination	Encodes DM domain (sex differentiation domain in *D. melanogaster*); read coverage does not support the gene model, new gene model cannot be inferred
Smp_168610.1	Myelin transcription factor	Read coverage does not support the gene model, new gene model cannot be inferred

4.2.2.5 Transcripts of unknown function

From the total of 142 differentially expressed transcripts (adjusted p-value < 0.05) between ST and MT parasites, many of them have no associated function and therefore are described as "hypothetical proteins". The percentage of "hypothetical proteins" found within the differentially expressed transcripts (44%) is similar to that found for the rest of the transcriptome.

As presented in Chapter 3, the newer version of the genome annotation contains 504 entirely new genes derived from RNA-seq evidence. These gene models represent 5% of the total gene complement and most of them have no function associated with them. Within the significant differentially expressed genes found in the comparison of ST and MT it was found that 17% of them (26 in total) are new RNA-seq derived predictions. Of these, only two of them have a significant match in the InterPro database; both these transcripts have their peak of expression at 24 hours post transformation and they are approximately 2 fold over expressed in the ST compared to the MT parasites. Smp_202120.1 found in Chromosome 1 has a conserved homebox domain with a putative transcription factor function. Smp_200450.1 is found in Chromosome 3 and has a conserved galactosyltransferase domain suggesting it may have a role in the biosynthesis of polysaccharides. Another related enzyme, a beta-1,4-galactosyltransferase, is also more highly expressed in the ST schistosomula. Because reorganization of the outer membrane is very important in the early stages of the schistosomula (see Chapter 1 sections 1.2.3.1.3), it is possible that the expression of these genes is related to a higher activity in the generation of polysaccharides for the schistosomula coating.

4.3 Discussion

The process of cercarial invasion and schistosomula migration, especially during the early stages of infection, are relevant in the study of intervention strategies against *Schistosoma* parasites. As introduced in Chapter 1, the skin schistosomula stage is regarded as a vulnerable stage for parasite killing (Wilson *et al.*, 2009). Understanding the biology of the parasites at this stage would improve the chances of finding the appropriate targets for intervention.

Many high-throughput gene expression studies have been performed in which different schistosomula stages were analysed (Chai *et al.*, 2006; Dillon *et al.*, 2006; Fitzpatrick *et al.*, 2009; Gobert *et al.*, 2010; Parker-Manuel *et al.*, 2011). With one

exception (Chai *et al.*, 2006), all studies used mechanically transformed parasites. Different schistosomula stages (i.e., skin, lung, etc) were obtained by varying the time parasites were incubated after transformation or after recovery from the host. Chai *et* al., (2006) showed that indeed lung stage (3-days post infection) obtained from infected animals and mechanically transformed parasites have different profiles of gene expression. These findings prompted the question of whether skin transformed and mechanically transformed schistosomula in the skin stage (24-hours post invasion) would also present differential expression of genes.

In order to address this question, transcriptome samples from both mechanically transformed and skin transformed schistosomula were sequenced and transcriptomes quantitatively compared.

The experimental approach presented in this thesis differs from that of Chai *et al.,* (2006) in many aspects. First, the work of Chai *et al.,* (2006) was performed in another species. The authors used *S. japonicum*, where significant differences in the speed of migration through the skin have been reported between this species and *S. mansoni* (Wang *et al.*, 2005). Second, the authors focus on 3-days old schistosomula residing in the lung tissue. Many publications have feature the differences between the skin- and lung-stages of the parasite (Crabtree *et al.*, 1980; Wilson *et al.*, 1980; Crabtree *et al.*, 1985; Crabtree *et al.*, 1986) and therefore these cannot be easily compared. Thirdly, the authors used a microarray platform where oligonucleotides sequences were generated from are *S. japonicum* and *S. mansoni* EST databases.

I the work here presented, the first challenge was to obtain the biological material. As shown before, (see section 4.2.1.1), mechanical transformed preparations of schistosomula can present a significant number of damaged parasites or contaminating tails. Therefore, optimization of the protocol was required to obtain healthy non-contaminated populations of transformed schistosomula. This was achieved by replacing part of the harsh needle-passage step by vigorous shaking of parasites in a vortex mixer. At the same time, optimization of the percentage of Percoll® used in the separation of cercarial heads and tails yielded a preparation virtually free from tails and cercariae. These modifications to the protocol resulted in fewer number of damaged parasites and less contamination of tails in the schistosomula preparation. The skin transformation protocol was also subjected to optimization; yet the schistosomula obtained from these experiments were usually contaminated with tails or cercariae and assessment of individual preparations was performed in order to obtain tail-free and cercariae-free schistosomula preparations.

The skin transformation protocol presents some disadvantages: 1) it is laborious and time consuming, taking approximately 9 hours to take the experiment to completion compared to the mechanical transformation (~ 3 hours); 2) it is possible that skin-transformed parasites are slightly off synchronization because transformation can occur at any time within a 3-hour window. On the other hand, skin transformation provides a more natural way of transforming the cercariae into schistosomula. Compared to the more widely used mechanical transformation, the skin transformation protocol allows the parasite to penetrate a layer of skin; which mimics the natural barrier present in the human host. The advantages of the mechanical transformation are evident: it is a quick protocol that allows researchers to obtain typically hundred of thousands of schistosomula per experiment. Once the protocol is optimized, parasites are generally healthy and phenotypically closely resemble the skin-transformed parasites even though they have been subjected to artificially imposed environmental variables such as low temperatures, centrifugation forces, shaking and squeezing through a syringe needle. In spite of being very different, both transformation protocols yielded healthy schistosomula that were capable of surviving *in vitro* for several days.

Although many studies have focused on the differences and similarities between these two types of schistosomula, none before had focused on the comparison of their transcriptome profiles. As mentioned earlier, the work of Chai *et al.*, (2006) is the closest to the experiment presented in this chapter. In their work, the authors found that IVS[1] parasites express transcripts encoding protaglandins, glutathione-S-transferase Sm28GST, paramyosin, stress related proteins and transcripts related to markers of anti-inflammatory and immunomodulatory processes. In the case of MTS, the authors show that these schistosomula show higher expression of transcripts involved in glucose transport, and fatty acids transport and haemoglobin digestion (Chai *et al.*, 2006).

The RNA-seq differential expression analysis of MT *vs.* ST parasites performed in this thesis showed that 149 transcripts are differentially expressed (adjusted p-value < 0.05). Transcripts encoded in the mitochondrial genome (mitochondrial genes) were found among the most differentially expressed transcripts being all of them more highly

[1] IVS refers to schistosomula obtained from lungs from infected mice three days post infection; MTS refers to mechanically transformed schistosomula followed by *in vitro* cultivation for three days Chai, M., D. P. McManus, R. McInnes, L. Moertel, M. Tran, A. Loukas, . . . G. N. Gobert (2006). "Transcriptome profiling of lung schistosomula,in vitro cultured schistosomula and adult Schistosoma japonicum." Cell Mol Life Sci **63**(7-8): 919-929..

expressed in the ST parasites. This over expression of mitochondrial transcripts was correlated with higher metabolic rate in the ST parasites compared to the MT measured using the AlamarBlue® essay (section 4.2.2.1.1) suggesting that skin transformed parasites could be regarded as more metabolically active than their mechanical counterparts. The reasons behind the increased abundance of mitochondrial transcripts could be many. One possibility could be that the mitochondrial genes are subjected to higher transcription rate in the ST parasites. Another possibility could be that the ST parasites simply have more functional mitochondria than the MT parasites. Counting the mitochondria in both schistosomula preparations using a mitochondrial marker could test this hypothesis. Contraction and extension movements generated by the musculature structure demands important quantities of ATP and the increased metabolic activity could be a reflection of such need. Unfortunately, it was not possible to quantitatively evaluate the rate of movement of MT and ST parasites, data from such experiments would provide more evidence to back up this hypothesis.

In order to study which other processes were represented among transcripts more highly expressed in ST parasites, Gene Ontology enrichment was performed. This analysis showed a potpourri of transcripts associated to different functions such as GPCR signalling, microtubule-related movement, retrotransposon related functions, betaine synthesis and muscle development. In the first two functional categories, the transcripts products description correlated well with the associated GO terms, providing a back up of the functional annotation of such genes. Additionally, their up-regulation in the ST schistosomula could be associated with processes likely to be taking place in these parasites. The rest of the categories were discarded based on the lack of reliability of either the gene models or their annotation. In the case of retrotransposon related function it was not possible to associate these transcripts to any known biological process in the schistosomula. Their function in this or any other stages of development remains unknown.

Mechanically transformed parasites showed higher expression of ferritins, proteases and protease inhibitors. Ferritins (Hall *et al.*, 2011) and certain proteases [reviewed in (Caffrey *et al.*, 2004)] are thought to be involved in uptake of nutrients from the host. At the same time, proteases have been linked to host tissue invasion (Salter *et al.*, 2000; Hansell *et al.*, 2008). Chai *et al.*, (2006) reported the differential expression of transcripts involved in nutrients uptake. However, these transcripts are not the same as those reported in this chapter. Nevertheless, it is interesting to notice that results presented here agree with those from Chai *et al.*, (2006) in that proteases are more highly expressed

in mechanically transformed parasites and the authors attribute this to culturing conditions. This explanation cannot be extrapolated to the results presented in this thesis regarding the differential expression of ferritins and proteases at 24 hours post transformation because both types of parasites were exposed to the same culturing conditions.

Independently from their higher expression in MT or ST parasites, it is noteworthy that all these transcripts are being up-regulated in the early schistosomula and their higher expression in MT parasites may reflect the higher synchronization in their development compared to the ST parasites. The expression of these transcripts in such an early stage of development is very interesting because of their possible implications in the survival of the parasite.

Some members of the microexon family 2 (MEG-2) were found over expressed in the ST parasites. What is more, the variants found here are different from the ones previously reported confirming that different splice variants from the same loci are expressed at different time points. Interestingly, one of the variants expressed in the ST parasites has a different exon profile than in the MT transcript. This is an important finding because it suggests that different cues from the environment might be triggering the differential expression of certain exons in the microexon gene repertoire. The function of microexon genes remains to be unknown.

Both populations of parasites showed expression of "false positives" that could be discarded based on the crosscheck of the putative functions associated to the products. Future upgrades in the annotation of the genome promise the opportunity to further improve the interpretation of these results.