CHAPTER 5

TIME COURSE ANALYSIS OF DIFFERNTIAL EXPRESSION OF GENES: CERCARIAE, 3 HOUR OLD & 24 HOUR OLD SCHISTOSOMULA

5.1 Introduction

During the process of infection of the human host, *S. mansoni* parasites change from the cercariae free-living stage into a parasitic life form, the schistosomula. In natural conditions, this transition is triggered by the penetration of the skin barrier upon which the cercarial head invades the host while the cercarial tail is lost. This situation can be reproduced in the laboratory by the application of shear pressure as described in Chapter 2 section 2.2.2. Detailed descriptions of the changes that accompany these transformations were introduced in Chapter 1 sections 1.2.3.1 to 1.2.3.2. It is also known that this transformation is not dependent on the presence of the skin barrier. A brief recount of these changes is summarised here.

Once in the skin, the parasites loose their glycocalyx, the outer membrane changes from a single to a double bilayer, the contents of the acetabular glands are gradually emptied to assist the penetration process across the outermost layers (Holmfeldt *et al.*, 2007) and the deeper layers of skin (Wilson *et al.*, 1980).

Many groups have attempted to describe the gene expression changes during the transition from the cercariae to schistosomula. So far, approaches used to this end involved quantitative ESTs (Farias *et al.*, 2011) and high-throughput studies using microarrays (Chai *et al.*, 2006; Dillon *et al.*, 2006; Fitzpatrick *et al.*, 2009; Gobert *et al.*, 2010; Parker-Manuel *et al.*, 2011). The work of Fitzpatrick *et al.,* (2009) spanned 15 different life cycle time points including cercariae, 3-hour old and 24-hour old schistosomula; the same time points considered in this thesis. With a p-value cut off (non-adjusted) of < 0.05 the authors showed that 159 genes were differentially expressed at 3 hours after transformation (114 up regulated and 45 down regulated) and 321 genes at 24 hours after transformation (202 up regulated and 119 down regulated) compared to cercariae. However, after correcting the p-values for multiple testing [adjusted p-value (Benjamini *et al.*, 2001)] the number of differentially expressed genes was reduced to zero. These results are surprising considering cercariae and schistosomula are very structurally different stages of the parasite.

In a more recent microarray study performed by Gobert *et al.,* (2010), in which the cercariae *vs.* 3-hour old schistosomula comparison was also considered, the authors identified 2,791 differentially expressed genes (1,608 up regulated and 1,183 down regulated) between these two life cycle stages. The up-regulated dataset included genes related to the structure of the tegument (i.e. tetraspanins and calcium-binding proteins),

stress response proteins (i.e. HSP70), development (i.e. frizzle related protein) and enzymes involved in the blood meal digestion (i.e., range of cathepsins) among others (Gobert *et al.*, 2010). Unfortunately, no test statistic was applied to the analysis of fold changes in this report making the interpretation and comparison with other studies a difficult task.

The main motivation to re visit the question of which genes are differentially expressed during the cercariae to schistosomula transformation lies in the application of novel techniques (RNA-seq) and statistical approach [edgeR, (Robinson *et al.*, 2010)] that would allow a better resolution and improved understanding of the genes that have a role in shaping the adaptation of the parasite to the new environment. As exposed before, the microarrays studies that tried to tackle this question had some limitations that the RNA-seq approach can overcome. In the particular case of *Schistosoma* microarrays, these have relied on the existence of previously identified genes and genome sequences to generate the probes that form the array [i.e., (Fitzpatrick *et al.*, 2005)]. On the contrary, the generation of RNA-seq data is independent from previous sequence knowledge and offers the possibility of exploring coding sequences previously unknown. What is more, and as it was shown in Chapter 3 section 3.2.3, the dynamic range of RNA-seq surpasses that of microarrays providing a better tool for measuring very high and very low levels of expression with improved resolution.

This chapter present the transcriptome changes of the cercariae and the early stages of mechanically transformed schistosomula (3-hours and 24-hours old parasites). Briefly, transcriptome sampling of the cercariae, 3-hours old and 24-hours old schistosomula were taken and sequenced using the Illumina second-generation sequencing technology. The first part of this chapter offers an overview of the time course analysis including four time points in the parasites' life cycle: cercariae, 3-hours old and 24-hours old schistosomula and 7-week old mix sex adult worms. The latter sample was used as a reference to extrapolate the expression of genes found differentially expressed in the schistosomula stage. This is of particular importance because it helps to recognise genes that are expressed after transformation but are no longer needed in the parasites' adult life. The second part of this chapter studies the different biological processes found to be up regulated or down regulated in the schistosomula compared to the cercaria stage through the analysis of differentially expressed genes. The level of resolution achieved by these data has no precedent in any other parasitic worm.

5.2 Results – time course analysis of transcriptome changes.

The first step in the time course analysis was to identify transcripts that are differentially expressed between different life cycle time points. To this end, the edgeR package (Robinson *et al.*, 2010) was used (see Chapter 2 section 2.6.5). Briefly, edgeR takes the number of reads per transcript as input and uses the biological replicates to estimate the dispersion of the samples. The calculated dispersion serves to normalise the libraries and make them comparable to each other. A normalised list of reads per gene is generated and from it, differential expression for each gene/entry can be calculated together with a significance value (p-value).

In preparation for the edgeR analysis, all non-expressed genes (as defined in Chapter 3 section 3.2.5) were removed from the dataset leaving a total of 9,096 transcripts for which differential expression could be calculated. Four time points in the parasite's life cycle were analysed (**Table 5.1** and **Figure 5.1**).

- Cercariae vs. 3-hours old schistosomula
- 3-hours old schistosomula vs. 24-hours old schistosomula
- 24-hours old schistosomula vs. adult
- Cercariae vs. 24-hours old schistosomula

Table 5.1 - Number of differentially expressed genes (adjusted p-value <0.01).

Stage comparison	Up regulated§	Down regulated§	Total
Cercariae - 3 hour schistosomula	964	522	1,486
3 hour schistosomula - 24 hour schistosomula	470	560	1,030
24 hour schistosomula - adult	1,120	960	2,080
Cercariae - 24 hour schistosomula	1,652	1,198	2,850

§ Fold change threshold is 2-fold

A total of 3,300 non-redundant transcripts (excluding alternative spliced forms) were found differentially expressed (adjusted (Benjamini *et al.*, 2001) p-value < 0.01) within the cercariae *vs.* 3-hours old schistosomula, 3-hours old schistosomula *vs.* 24-hours old schistosomula and 24-hours old schistosomula *vs.* adult comparisons. In this study, more differentially expressed transcripts are found than in the work of Fitzpatrick *et al.,* (2009) but less than in Gobert *et al.,* (2010).

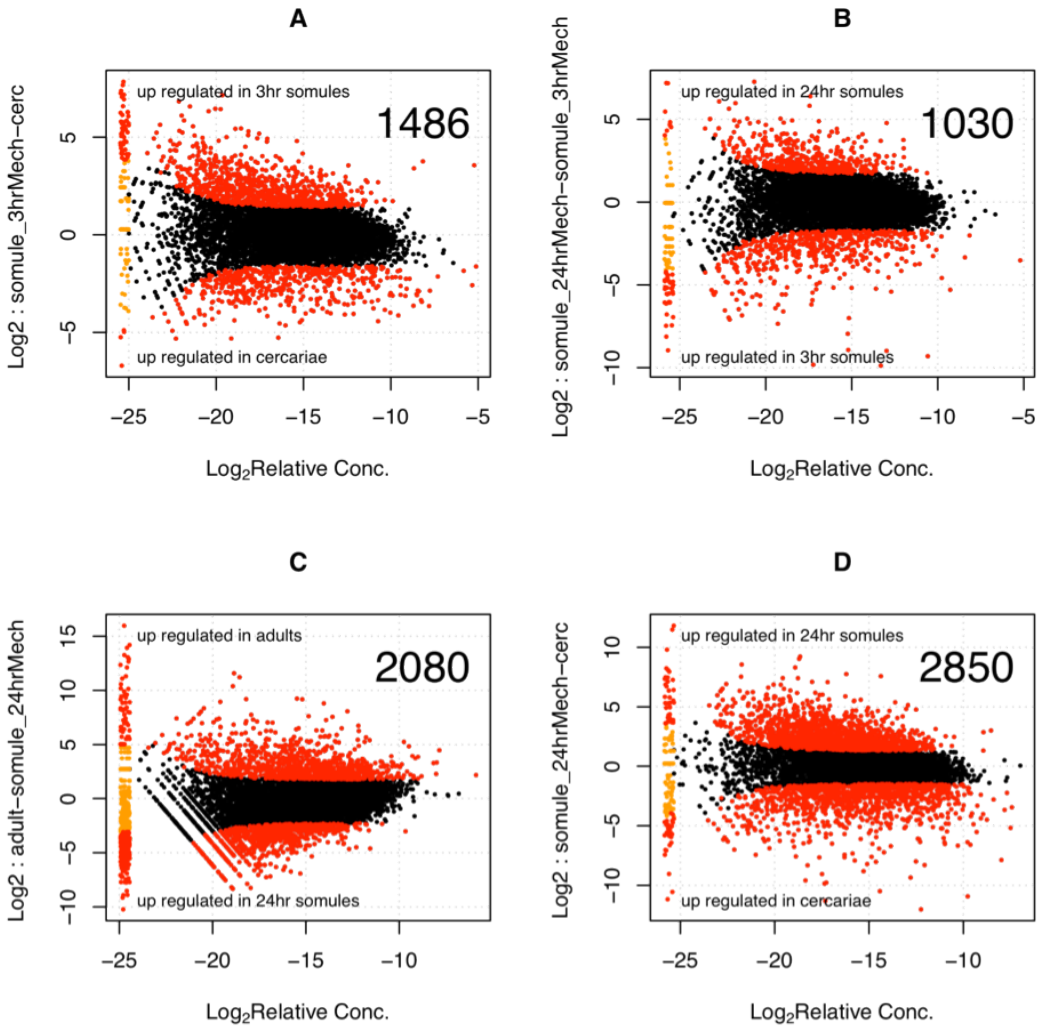

Figure 5.1 - Differential expression of transcripts across four time points in the life cycle of *S. mansoni*. Relative concentration (x axis) is plotted against fold change values (y axis) in the logarithmic scale in base 2. Legend: Red dots, significantly different (adjusted p-value <0.01); black dots, non-significant. Data points grouped in the left of each plot (red and orange) represent transcripts that have reads in one sample but not in the other and therefore the relative concentration and fold change cannot be calculated. Numbers in the top right corner of each plot represents the total number of significant differentially expressed transcripts. A – Differential expression between cercariae and 3-hours old schistosomula. B - Differential expression between 3-hours old and 24-hours old schistosomula. C - Differential expression between 24-hours old schistosomula and mix sex 7-week old adults. D - Differential expression between cercariae and 24-hours old schistosomula.

5.2.1 Genes with no change in expression

Fitzpatrick *et al.,* (2009) identified a set of 355 microarray probes with less than 1% variability in their expression values across 15 life cycle time points (Fitzpatrick *et al.*, 2009). The authors regarded genes represented by these probes as "constitutively expressed genes". It was found that 192 of these probes have an unambiguous match against the transcripts dataset (for Methods see Chapter 2 section 2.6.4) and 75 of them are differentially expressed at least in one of the comparisons. **Figure 5.2** shows an example of the distribution of the 192 transcripts in the cercariae to 3–hour old schistosomula comparison where only 25 transcripts were found differentially expressed. The variability between the results shown in this study and those shown in microarrays for the same genes might be attributed to the higher resolution that can be achieved through RNA-seq measurement. It is worth mentioning that most of the constitutively expressed genes/transcripts are found at relatively high concentration values meaning that they are relatively highly expressed compared to the rest of the transcripts. As shown before in Chapter 3 section 3.2.3, the capacity of microarrays in identifying differential expression in highly expressed genes is much reduced compared to RNA-seq. This limitation of the microarray platform could explain why these genes are found differentially expressed in the RNA-seq approach but not in the microarray.

5.2.2 Validation using known genes

In order to validate the RNA-seq approach in *S. mansoni*, genes with known expression profile were compared to the RNA-seq relative expression values. The control genes were chosen based on differences in their changes of expression shown by northern blotting. These genes are an 8 kDa calcium binding protein (Smp_033000.1), associated with tegument remodelling during cercariae transformation into schistosomula (Ram *et al.*, 1989; Ram *et al.*, 1994); a heat shock protein 70 (HSP70 - Smp_106930.1), active in very early (~3-hours old) schistosomula (Hedstrom *et al.*, 1987; Neumann *et al.*, 1992; Neumann *et al.*, 1993); and the tegument antigen Sm22.6 (Stein *et al.*, 1986) also known as SmTAL1 (for *S. mansoni* tegument allergen-like protein 1 - Smp_045200.1), associated with resistance to re-infection in adult patients of endemic areas (Dunne *et al.*, 1992). RNA-seq results broadly agreed (**Figure 5.3**) with relative gene expression measurements obtained through other approaches. The only case in which the data did not correlate was in the comparison of 24-hours old schistosomula *vs.* adult expression of HSP70 –

Smp_106930.1. The source of this discrepancy is unknown, although it might be related to the rapid control in the regulation of expression of this transcript.

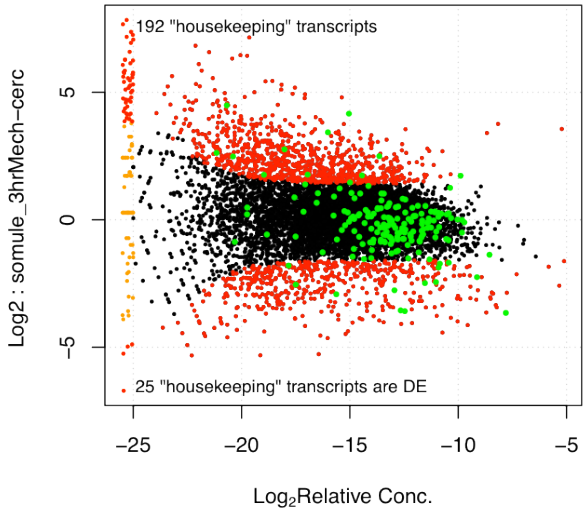

Figure 5.2 – Differential expression of transcripts in the cercariae *vs.* 3-hour old schistosomula. Green dots represent constitutively expressed as reported by Fitzpatrick *et al.,* (2009).

5.2.3 Analysis of differential expression in cercariae, 3 hour and 24 hour old schistosomula

As pointed out in section 5.2, there are ~3,300 non-redundant differentially expressed transcripts (adjusted p-value < 0.01) in the comparisons considered in this study. A breakdown of the biological processes over-represented among the groups of up- or down- regulated genes in the cercariae to 3-hours old and 24-hours old schistosomula transitions is presented.

5.2.3.1 Down regulated transcripts/processes in the schistosomula stage

In order to identify down regulated processes in the passage from the cercariae to the schistosomula stage, GO term enrichment analysis was performed using the dataset of down regulated genes from three pair-wise comparisons: the cercariae *vs.* 3-hour old schistosomula, 3-hour old schistosomula *vs.* 24-hour old schistosomula and cercariae *vs.*

24-hour old schistosomula. Results are shown in **Table 5.2**, **5.3** and **5.4** respectively. With few exceptions, down regulated processes do no differ much between the comparison cercariae *vs.* 3-hour old schistosomula and cercariae *vs.* 24-hour old schistosomula.

During the passage through the host skin, the cercariae lose their tails, and their heads transform into schistosomula. Tails are multi-cellular structures characterised by tissues resembling striated muscle fibres and packed with mitochondria. Tails are anatomically and physiologically different from the cercarial head (see Chapter 1 section 1.2.3.1) and therefore it is expected that they have a very different transcript repertoire. Because of this, transcripts that appear as down regulated in the schistosomula stage compared to the cercariae could arise from two scenarios: either transcripts could be truly down regulated in the schistosomula compared to the cercarial head or they could be exclusively or highly expressed in the cercarial tail and therefore appear as down regulated or are undetectable in the schistosomula stage.

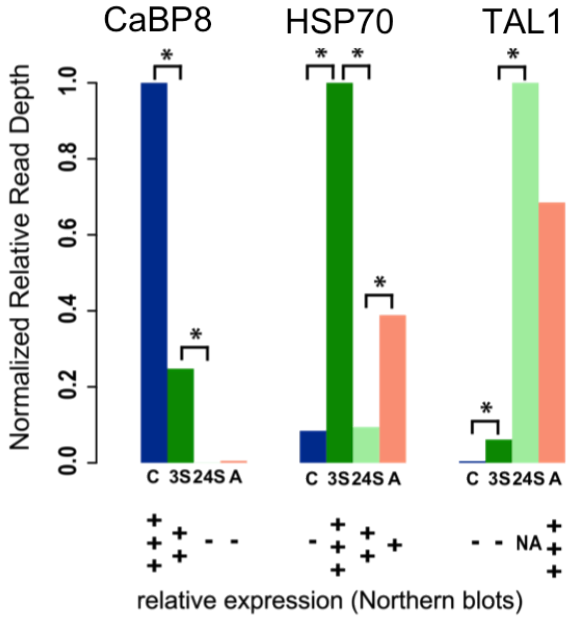

Figure 5.3 - Comparison of expression of genes previously identified to be developmentally regulated. Barplots represent relative normalized reads (from RNA-seq data) for 3 transcripts, asterisks represent comparisons where differential expression is significant (adjusted p-value <0.01). Relative expression reported in the literature (Stein *et al.*, 1986; Ram *et al.*, 1989; Neumann *et al.*, 1992) is shown at the bottom (+++, high expression, ++ medium expression, + some expression, - not expressed, NA, no information available). C = cercariae, 3S = 3-hour old schistosomula, 24S = 24-hour old schistosomula, A = adult.

Because of the cercariae losing the tails, many of the transcripts down regulated in the schistosomula are related to biological processes known to occur mainly in the tail. This was shown by the GO enrichment analysis (see **Table 5.2** and **5.4)**. Higher expression of transcripts involved in aerobic respiration (ATP biosynthesis couple proton transport, TCA, glycolysis and electron transport chain) is detected in cercariae reflecting the metabolic pathway used to generate energy (Skelly *et al.*, 1993).

Consistently with loosing the mitochondria, the expression of solute and protein transporters related to the mitochondrial function are also down regulated. Transcripts included in this "transport" category (see "transport" in **Table 5.2**) encode a range of carrier proteins mostly with trans-membrane domains associated to the traffic of solutes across the mitochondrial membrane and therefore, expected to appear as down regulated once the cercarial tail is lost in the transformation.

As previously mentioned, the cercarial tail has a complex muscle structure formed by longitudinal and circular muscle layers. These types of structures are not found neither in the cercarial head nor the schistosomula. In accordance to this, results show down regulation of three calponin proteins and six transcripts containing collagen (grouped under phosphate transport, which could indicate wrong annotation). A calponin-like protein has been localised to the cercarial tail muscle in *S. japonicum* (Jones *et al.*, 2001) and proteins containing the calponin domain have been shown to bind actin and other components of the smooth muscle contraction apparatus in vertebrates (el-Mezgueldi, 1996). Collagen is a principal component of the connective tissue, also very abundant in the cercarial tail (Dorsey *et al.*, 2002). The differential expression pattern of genes related to the tail musculature is in agreement with loss of the tail.

Observations of the *in vivo* and *in vitro* schistosomula have suggested that there is no cell division until the fourth day after transformation (Clegg *et al.*, 1972). In this context, transcripts involved in DNA replication and cell division where investigated to asses whether the observations from Clegg *et al.,* (1972) were also valid at the transcriptional level. Histone protein expression is a good indicator of cell division. Histone mRNA accumulates during the DNA replication phase of the cell cycle and histone proteins are synthesised during the S phase [reviewed in (Osley, 1991)]. A decrease in the histone mRNA would be a good marker for absence of DNA replication. Consistent with the observations made by Clegg *et al.,* (1972) almost 40 years ago, RNA-seq data show that histone mRNAs (h1/h5, H2A, H2B, H3, H4 and linker histone H1) are down regulated (up to 40 times in linear scale) at 24 hours post transformation compared to the cercariae (**Table 5.5**). It is noteworthy that this effect is not detected when comparison cercariae

with 3-hours old schistosomula, indicating that this is not an effect of loosing the tail. It is possible that histone mRNA is still present in the cercariae and 3-hour old schistosomula as a reminiscent of the high rate of cell division observed in the germ ball (Parker-Manuel *et al.*, 2011); suggesting that the schistosomula shuts down it cell division machinery only after 3 hours post-transformation. Once the parasites has reached day 3 post-transformation the expression of histones mRNA is higher than that recorded for the cercariae stage (Parker-Manuel *et al.*, 2011) suggesting that cell cycle is resumed some time between 24-hours and 3-days old schistosomula.

Table 5.2 – Gene Ontology enrichment (Biological processes) among transcripts down regulated in the cercariae *vs.* 3-hours old schistosomula.

GO term ID	GO term description	p-value
GO:0015986	ATP synthesis coupled proton transport	0
GO:0000226	Microtubule cytoskeleton organization	0
GO:0006099	Tricarboxylic acid cycle	0
GO:0042773	ATP synthesis coupled electron transport	0.002
GO:0006810	Transport	0.003
GO:0006096	Glycolysis	0.003
GO:0022900	Electron transport chain	0.004
GO:0009098	Leucine biosynthetic process	0.004
GO:0007283	Spermatogenesis	0.007
GO:0006094	Gluconeogenesis	0.008
GO:0006108	Malate metabolic process	0.008

Table 5.3 – Gene Ontology enrichment (Biological processes) among transcripts down regulated in the 3-hours old schistosomula *vs.* 24-hours schistosomula.

GO term ID	GO term description	p-value
GO:0006412	Translation	0
GO:0006817	Phosphate transport	0
GO:0006334	Nucleosome assembly	0
GO:0006278	RNA-dependent DNA replication	0.001
GO:0008380	RNA splicing	0.001
GO:0008272	Sulfate transport	0.005
GO:0006542	Glutamine biosynthetic process	0.005
GO:0009097	Isoleucine biosynthetic process	0.005
GO:0051258	Protein polymerization	0.005

Table 5.4 – Gene Ontology enrichment (Biological processes) among transcripts down regulated in the cercariae *vs.* 24-hours old schistosomula.

GO term ID	GO term description	p-value
GO:0006096	Glycolysis	0
GO:0006810	Transport	0.001
GO:0006817	Phosphate transport	0.002
GO:0007283	Spermatogenesis	0.003
GO:0042775	Mitochondrial ATP synthesis coupled electron transport	0.004
GO:0031032	Actomyosin structure organization	0.004
GO:0008380	RNA splicing	0.005
GO:0006334	Nucleosome assembly	0.006
GO:0015986	ATP synthesis coupled proton transport	0.006
GO:0000398	Nuclear mrna splicing, via spliceosome	0.006
GO:0022904	Respiratory electron transport chain	0.008

Other processes that seem to accompany this halt in the cell cycle are also identified as down regulated. A group of five transcripts encoding proteins known to participate in microtubule formation (grouped under the GO biological process of "microtubule cytoskeleton organization") are down regulated in the schistosomula. These proteins contain highly conserved tektin domains, which are involved in formation and stabilization of microtubules and centrioles (Amos, 2008). The down regulation of these transcripts in the schistosomula stage may be a consequence of the cell cycle arrest status. Transcripts related to the spliceosome, protein translation machinery and protein polymerisation are also down regulated (**Table 5.3** and **5.4**).

Because some enzymatic reactions appear in more than one pathway, it is common to find that the GO term analysis show certain processes as enriched while in reality they are just part of down stream steps of other processes. Leucine biosynthesis, malate metabolic process and gluconeogenesis (**Table 5.3**), and glutamine and isoleucine biosynthesis (**Table 5.4**) are shown as down regulated after transformation. In this case, these processes are shown in the GO analysis because the enzymes encoded by these transcripts are also part of the TCA cycle or the glycolysis pathway. Since the functional annotation of these genes is correct (data not shown) it would not be appropriate to consider them false positives.

However, true sources of false positives in GO enrichment analysis are transcripts associated with the wrong annotation. Spermatogenesis is a good example: there are five spermatogenesis-associated transcripts up regulated in this stage. Three of them have no

homology with any informative conserved domains (one of them has homology to one "domain of unknown function"). The other two transcripts encode a calponin domain and a DNA-binding zinc-finger domain respectively. An association between these transcripts and the process of spermatogenesis could not be found.

Table 5.5 – Nucleosome components are down regulated in the schistosomula stage

GeneDB	FC§ 3-hours old schistosomula	FC§ 24-hours old schistosomula	product description
Smp_002930.1	-11.48	-40.55	histone H2A, putative
Smp_003770.1	NS	-16.27	histone h1/h5, putative
Smp_036220.1	-3.71	-5.45	histone H2B, putative;with=UniProt:Q8I1N0
Smp_053390.1	-6.12	-9.55	histone H4, putative
Smp_054230.1	NS	-7.15	hypothetical protein
Smp_074610.1	-6.56	-11.57	histone H3, putative
Smp_082240.1	-9.24	-14.92	histone H3, putative
Smp_162370.1	-12.02	-34.11	Linker histone H1

§ Linear fold change of the gene expression in the schistosomula stage compared to cercariae. NS: Non-significantly.

5.2.3.2 Up regulated transcripts/processes in the schistosomula stage.

As previously pointed out, the establishment of the parasitic life inside the mammalian host starts with the penetration of the cercarial head in the host skin. In the previous section, transcriptional signals lost or down regulated in the schistosomula compared to the cercariae were discussed. In this section the focus is made in up regulated transcripts. As mentioned in the introduction of this chapter, there are only two previous high-throughput microarray studies that looked at the transcriptional changes in this very short time points during the transition. Among the 1,608 up regulated genes reported by Gobert *et al.,* (2010), the authors pointed out some examples of up regulated genes related to the tegument, gut function, stress and development (Gobert *et al.*, 2010). RNA-seq data could detect up regulated transcripts in all of these categories.

The GO term enrichment analysis applied in Gobert *et al.,* (2010) indicated genes enriched in rather general GO categories. The higher resolution of provided by the RNA-seq approach allowed to identify four main biological processes occurring in the schistosomula. These processes group genes involved in signal transduction (GPCRs, Wnt receptor signalling pathway, potassium/ion transport), carbohydrate transport (and monosaccharide transport), "regulation of transcription" and "tissue development"

(homophilic cell adhesion, cell differentiation, cell adhesion, cell-matrix adhesion, integrin-mediated signalling pathway among others). The full list of categories and how they are represented in the three comparisons analysed is presented in **Figure 5.4** and **Appendix C**.

Schistosoma G-protein couple receptors (GPCR) have been proposed as an important candidate group to search for new drug targets (Berriman *et al.*, 2009) mainly because they have been validated as such in a range of other organisms (Overington *et al.*, 2006). GPCRs are significantly enriched among up regulated transcripts in the 3-hour old schistosomula (32 transcripts) and 24-hours old schistosomula (61 transcripts) compared to cercariae. Closer inspection of these genes [performed in collaboration with Dr. Mostafa Zamanian, (Zamanian, 2011)] resulted in their classification into nine categories: Peptide, PROF1[1], Amine, Frizzled, Secretin, Glutamate, Other, "Unlikely GPCR" and "partial sequence". Among these, neuropeptide receptors are significantly over represented; they are discussed in following sections in the context of neuropeptide signalling.

Voltage-gated ion channels are the fourth most drugable targets in the pharmacological industry (Overington *et al.*, 2006). *S. mansoni* encodes a significant number of ion channels of which 81 are annotated as potassium channels (Berriman *et al.*, 2009). It is noteworthy that one third of them are up regulated in the schistosomula stage. Potassium ion channels are expressed in a wide range of tissues (i.e. nervous system, muscle) and their function is mainly to set the electrical resting potential of cells (Alberts *et al.*, 2002). Schistosomula express a number of genes involved in nervous system development. It is possible that the expression of potassium ion channels is related to the development of the parasite's nervous system; which is discussed later in this section

[1] PROF1 stands for "Platyhelminth Rhodopsin Orphan Family 1"; which constitute a divergent flatworm specific family of receptors with some resemblance to the rhodopsin receptor family.

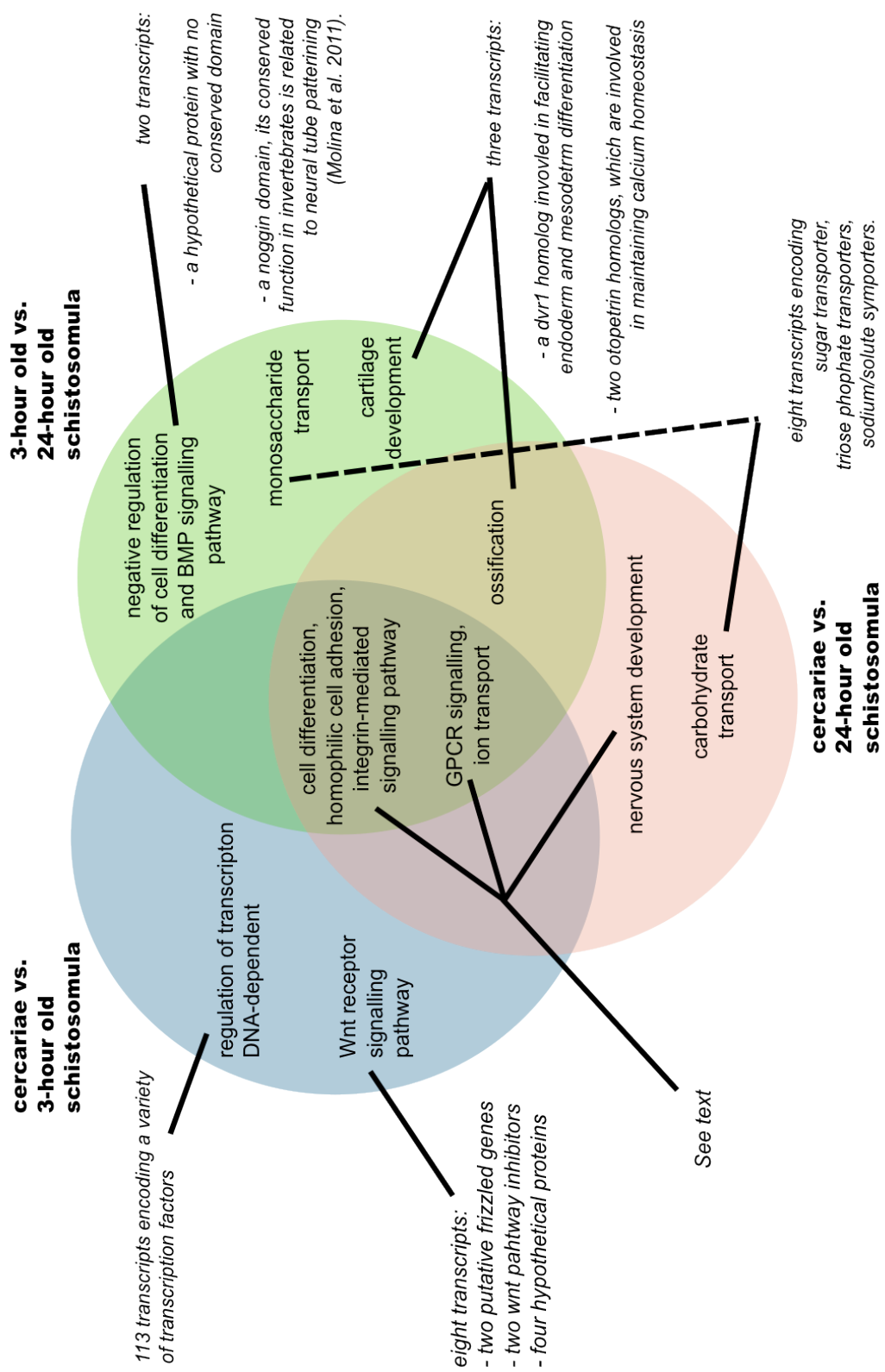

Figure 5.4 – Venn diagram representing biological processes enriched among up regulated genes in the cercariae to 3- and 24-hours old schistosomula.

Up regulated genes under the "transcription regulation DNA-dependent" category comprise a total of 113 transcripts with a range of fold change values between 2.7 and ~27 fold (linear scale) up regulated at 3 hours after transformation. Expression of these genes is either maintained or slightly up or down regulated at 24 hours, but none of them is down regulated with respect to the cercariae (**Figure 5.5**). These transcripts encode transcription factor (TF) domains that range from very conserved eukaryotic domains, such as the POU domain (see section 5.2.3.2.1), to less characterised ones such as DNA- or RNA-binding domains. Seventeen of the up regulated TF encode homeodomains. Genes encoding homeodomains are called homeobox or Hox genes; these are responsible for inducing cellular differentiation through the co-regulation of genes required for tissue and organ development and are key players in the establishment of body axis. Their study in parasitic worms is only but starting (Olson, 2008); therefore these findings are relevant to the advancement of the field.

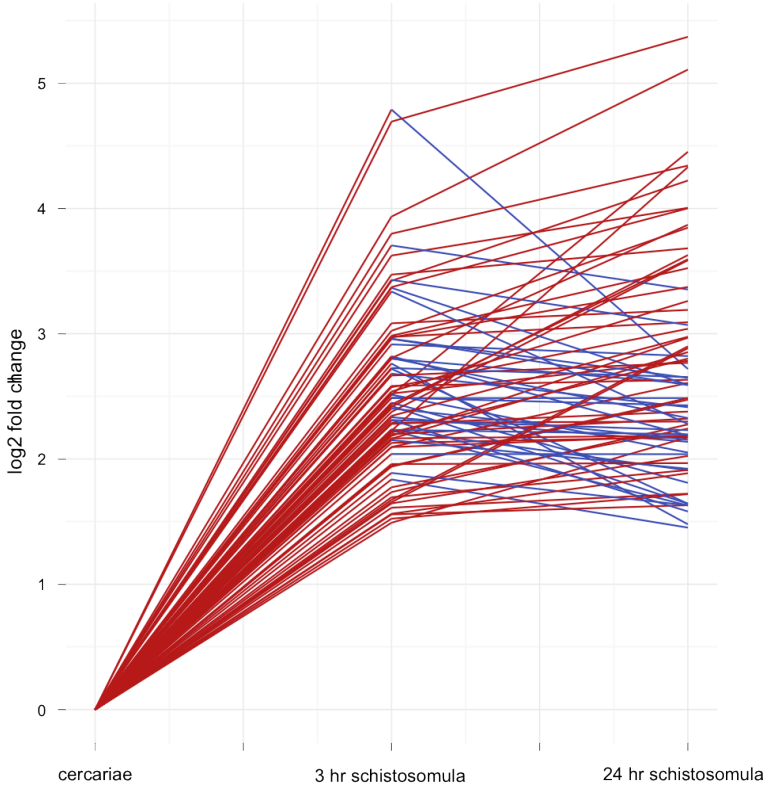

Figure 5.5 – Fold change values of up regulated genes (red) in the category "regulation of transcription DNA-dependent". These genes encode a range of transcription factors and some are slightly down regulated at 24-hours compared to 3-hours old schistosomula (bule).

Consistently with the up regulation of transcriptome factors involved in development, genes involved in many aspects of cell differentiation and tissue development (multicellular organismal development, homophilic cell adhesion, cell differentiation, cell adhesion, cell-matrix adhesion, integrin-mediated signalling pathway among others) are also up regulated indicating that schistosomula have started the programmatic chain of events that could lead to the development of tissues and organs as early as 3 hours after transformation. As mentioned earlier, mitosis does not occur in schistosomula until these are at least 3-4 days old (Clegg *et al.*, 1972); which is reflected in the inhibition of DNA replication. Hence, it is not possible that cell proliferation is actively occurring at this stage. However, as indicated by the GO term enrichment analysis, there is a profound up regulation (15 out of the 61 annotated in the genome) of transcripts encoding putative cadherins, integrins and laminins. Moreover, all these transcripts have their peak of expression at the schistosomula stage (either 3 hours or 24 hours old) suggesting that mRNAs necessary for the formation of such structures are ready and probably put "on hold" during the cell arrest period. Once this is gone, the whole machinery could be quickly activated to generate tissue development.

It was previously noticed that the spliceosome machinery is down regulated suggesting that pre-mRNA processing might be compromised or occur at a slower rate than in the schistosomula stage. Hence, it is possible to hypothesise that the up regulated TF might be inhibiting expression rather then inducing it. However, it was not possible to identify transcription repressors among these genes. It is worth mentioning that the down regulation of the spliceosome machinery seems to occur after the schistosomula have initiated transformation and therefore cannot be attributed to the loss of the tail ("spliceosome" category does not show in the Table 5.2 where the cercariae and 3-hours old schistosomula comparison is presented). The relationship between the down regulation of spliceosome machinery and the up regulation of TF remains to be explored.

5.2.3.2.1 Nervous system development is triggered in the early schistosomula

Consistent with the tissue developmental programme, transcripts involved nervous system development are up regulated in schistosomula. These are grouped in three GO categories: "nervous system development", "negative regulation of cell differentiation" and "negative regulation of BMP (bone morphogenic pathway) signalling pathway". The last two terms are closely related as they hold the same transcripts; one of them encodes a

hypothetical protein while the other encodes a *noggin* domain. The latter has a role in regulating neural development in invertebrates (Molina *et al.*, 2011) and therefore was considered within the nervous system development.

Evidence of the up regulation of genes involved in the nervous system development led to investigate whether any of the up regulated transcription factors were related to neural tissue development. Of the 20 genes involved in "early neural patterning" present in the of *S. mansoni* genome (Berriman *et al.*, 2009), RNA-seq data showed that 15 of them are expressed above background in at least one of the life cycle time points studied. What is more, 12 of them have their peak of expression in either 3-hours old or 24-hours old schistosomula suggesting a role in the development of the parasite at this stage. As an example, a well-conserved eukaryotic transcription factor (POU domain) responsible for specifying the identity of individual neural cells [reviewed in (Hobert, 2010)] is found among the up regulated TF, suggesting that genes expressed at this stage play a key role in determining the fate of neural cells in *S. mansoni*.

This wave of activation of transcription of genes involved in nervous system development led to investigate whether other aspects of neuron proliferation and function were also up regulated. One important aspect to the physiology of the nervous system is communication between neuronal cells, which is mediated in part by neuropeptides and neuropeptide receptors.

S. mansoni has approximately 30 GPCR belonging to the neuropeptide receptors sub-classification [(Berriman *et al.*, 2009) and (Zamanian, 2011)], 18 of which are up regulated in 24-hours old parasites with fold change values ranging from 2.6 to 55 times (linear scale). What is more, all except one are down regulated in adult worms, suggesting an important role during the schistosomula stage. The up regulation of these neuropeptide receptors occurs at the same time with the expression of nervous system development transcripts providing possible grounds for transcriptional co-regulations and further evidence of the active development of this tissue.

The distinct profile of expression of the neuropeptide receptors led to investigate their potential ligands. McVeigh *et al.,* (2009) reported a very comprehensive analysis of the neuropeptide precursor (npp) sequences found in platyhelminths including *S. mansoni* (McVeigh *et al.*, 2009). The authors found 14 npp that they grouped into 11 families. Five of the 14 sequences had already been associated to a gene model in GeneDB (GeneDB, 2011); here the other 9 matches are reported (for full description of Methods see Chapter 2 section 2.6.9). The rest of the npp reported for other platyhelminth species by McVeigh *et al.,* (2009) could not be found in the genome.

Of all the neuropeptide precurosrs, Sm-npp-20b is the only significantly up regulated - approximately 3 times (adjusted p-value < 0.01) compared to cercariae - in 3-hours old schistosomula and rapidly down regulated at 24 hours after transformation remaining low in the adult. This pattern of expression suggests that Sm-npp-20b has a role in the developing schistosomula; which remains so far unknown. Sm-npp-20b belongs to the neuropeptide F (NPF) subfamily of neuropeptidedes; whose individual peptide products are similar to those found in vertebrates [reviewed in (McVeigh *et al.*, 2009)] – these are called NPYs. The function of vertebrate NPYs involves inhibiting the accumulation of cyclic adenosine monophosphate (cAMP) in a concentration-dependent manner. The structural similarity between NPFs and NPYs suggest a similar role in *S. mansoni* (Humphries *et al.*, 2004; McVeigh *et al.*, 2009). However, whether Sm-npp-20b acts as a ligand of any of the neuropeptide receptors up regulated at 3 or 24 hours after transformation will require further investigations. Given the large number of neuropeptide receptors expressed in the schistosomula, it is possible that *S. mansoni* has a larger battery of npp than the one identified so far and perhaps the npp acting upon the neuropeptide receptor identified in the skin schistosomula stage are among those missing. These neuropeptide receptors could also have a role in sensing host-derived neuropeptides.

5.3 Discussion

Previous high-throughput studies aimed at describing transcriptional changes occurring during the transformation of the cercariae into the skin stage schistosomula (Fitzpatrick *et al.*, 2009; Gobert *et al.*, 2010). Although some insightful results could be drawn from these works, they were limited by the inherent disadvantages of microarrays; some of which can be overcome by the use of sequencing - instead of hybridisation - in the analysis of gene expression. The improvement of the *S. mansoni* genome (Protasio *et al.*, 2012) together with the advancement of RNA-seq as the cutting edge technique for analysing messenger RNA samples made it possible to address the question of which genes are developmentally regulated in the early skin schistosomula stage without the restriction of surveying only known features. Moreover, the greater dynamic range and digital nature of the RNA-seq approach allow more accurate measurements of gene expression at both ends of the spectrum.

In this chapter, focus was made on identifying signals that are important for the developmental stage of the parasite. To this end, gene expression from schistosomula

stage was compared to that of cercariae and adult worms. This allowed the identification of genes that are triggered by the transformation process and that might be relevant for establishment of infection.

As a first step, data from RNA-seq data were compared to genes regarded as constitutively expressed (Fitzpatrick *et al.*, 2009). In this case, data broadly agrees with the results from Fitzpatrick *et al.,* (2009); discrepancies may be due to the known limitation of microarrays in discriminating signals from highly expressed genes (as demonstrated in Chapter 3 section 3.2.3). Then, genes known as differentially expressed [8 kDa CaBP (Ram *et al.*, 1989), HSP70 (Hedstrom *et al.*, 1987) and Sm22.6/SmTAL1 (Stein *et al.*, 1986)] were also investigated. Results showed that data obtained from the measurement of expression of individual mRNAs from independent experiments correlate with those of RNA-seq (section 5.2.2). Apart from correlation of individual genes, some of the processes found to change in the cercariae to schistosomula transformation, such as the down regulation of transcripts involved in aerobic metabolism (Skelly *et al.*, 1993; Parker-Manuel *et al.*, 2011) gut function and gut tegument (Gobert *et al.*, 2010) were also found.

Most importantly, with the RNA-seq differential expression approach it was possible to identify groups of genes involved in other processes that were either unknown to happen in the schistosomula or not so well explored. Some examples were presented in this chapter and are discussed here.

Firstly, many signals indicating a state of cell arrest were identified in this study. The most prominent of these signals is the down regulation of all the constituents of the nucleosome. Results shown in this chapter suggest that schistosomula down regulate histone mRNA synthesis probably as a consequence of the state of cell cycle arrest (Clegg *et al.*, 1972). Other signals related to cell division and cell function are also down regulated.

A battery of transcription factors found up regulated in 3-hour old schistosomula, whose expression is maintained in the 24 hours-old parasites and decays in adult worms. Among these, transcription factors involved in the nervous system development were found, indicating that among tissue development signals, neural development is important in this early stage of the schistosomula. If tissue development is important at this stage so cell communication should be. This led to find that many G-protein couple receptors were up regulated in the 3- and 24-hours old parasites. It is known that these receptors have a central role in the cell-to-cell communication in all metazoan organisms (Alberts *et al.*, 2002) and therefore their expression is expected in all stages of the parasite development.

However, the preferential up regulation of the expression of neuropeptide receptors in 24-hours old schistosomula has not been reported previously and open new questions regarding the role of these receptors in the development of the skin stage schistosomula. On the other hand, GPCR proteins and more specifically neuropeptide receptors have been tagged as potential drug targets in other systems (Overington *et al.*, 2006) and also in helminths (Greenwood *et al.*, 2005). Knowing when and which of these receptors are expressed at any time represents an opportunity to understand their function as well as it narrows the list of potential drug targets that would need to be empirically tested.

Regarding the potential ligands for these receptors, only one of the known neuropeptide precursors was found coregulated with the neuropeptide receptors. The participation of other neuropeptides so far unknown as well as the interaction with neuropeptides derived from host proteins cannot be discarded.

On the one hand, parasites seem to be in a state of cell cycle arrest; which has been previously described and can be detected here at the molecular level. Together with cell cycle arrest, other functions related to cell division, such as generation of microtubules and protein synthesis in general seem to be also at halt. On the other hand, the massive up regulation of transcriptome factors, many of them related to tissue development, suggests that there is an underlying programmed fate for these parasite to undergo development. Reduced protein synthesis suggest that this process is regulated post transcriptionally but before translation, probably with the objective of saving resources in case they are needed. Having said that, it is hypothesised that during the parasites first 24 hours in the mammalian host, these organisms suppress parasites' growth during the early skin stage, and this is reflected in reduction of elements needed for cell division. Nevertheless, parasites are getting ready for swift transition into development by preparing part of the machinery (i.e., transcription factors, GPCRs) needed to undergone development.