Transcriptome of the parasitic flatworm *Schistosoma mansoni* during intra-mammalian development

Arporn Wangwiwatsin[1, 2], Anna V. Protasio[1, 3], Shona Wilson[3], Christian Owusu[1], Nancy E. Holroyd[1], Mandy J. Sanders[1], Jacqueline Keane[1], Mike J. Doenhoff[4], Gabriel Rinaldi[1], Matthew Berriman[1*]

[1] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK

[2] Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand

[3] Department of Pathology, Tennis Court Road, University of Cambridge, Cambridge, UK

[4] School of Life Sciences, University of Nottingham, University Park, Nottingham, UK

* Corresponding author

Email: mb4@sanger.ac.uk (MB)

Abstract

Schistosomes are parasitic blood flukes that survive for many years within the mammalian host vasculature. How the parasites establish a chronic infection in the hostile bloodstream environment, whilst evading the host immune response is poorly understood. The parasite develops morphologically and grows as it migrates to its preferred vascular niche, avoiding or repairing damage from the host immune system. In this study, we investigated temporal changes in gene expression during the intra-mammalian development of *Schistosoma mansoni*. RNA-seq data were analysed from parasites developing in the lung through to egg-laying mature adult worms, providing a comprehensive picture of *in vivo* intra-mammalian development. Remarkably, genes involved in signalling pathways, developmental control, and adaptation to oxidative stress were up-regulated in the lung stage. The data also suggested a potential role in immune evasion for a previously uncharacterised gene. This study not only provides a large and comprehensive data resource for the research community, but also reveals new directions for further characterising host–parasite interactions that could ultimately lead to new control strategies for this neglected tropical disease pathogen.

Author Summary

The life cycle of the parasitic flatworm *Schistosoma mansoni* is split between snail and mammalian (often human) hosts. An infection can last for more than 10 years, during which time the parasite physically interacts with its mammalian host as it moves through the bloodstream, travelling through the lungs and liver, to eventually establish a chronic infection in the blood vessels around the host gut. Throughout this complex journey, the parasite develops from a relatively simple larval form into a more complex, sexually reproducing adult. To understand the molecular basis of parasite interactions with the host during this complex journey we have produced genome-wide expression data from developing parasites. The parasites were collected from experimentally-infected mice over its developmental time-course

from the poorly studied lung stage, to the fully mature egg-laying adult worm. The data highlight many genes involved in processes known to be associated with key stages of the infection. In addition, the gene expression data provide a unique view of interactions between the parasite and the immune system in the lung, including novel players in host-parasite interactions. A detailed understanding of these processes may provide new opportunities to design intervention strategies, particularly those focussed on the early stages of the infection that are not targeted by current chemotherapy.

Introduction

The blood fluke *Schistosoma mansoni* is a major aetiological agent of hepatic and intestinal schistosomiasis, a Neglected Tropical Disease that affects over 200 million people around the world, largely in developing regions (1). Standard treatment of schistosomiasis relies on a single drug, praziquantel, and drug resistance is an ever-present threat with emerging reports of reduced efficacy to praziquantel in the field (2). A better understanding is needed of molecular and cellular mechanisms underlying the establishment of infections, so that vulnerabilities will be revealed with potential for targeting by new control tools.

Schistosomes are transmitted between mammalian hosts by susceptible aquatic snails. Each snail releases thousands of motile cercariae that seek out a host, penetrate through its skin and transform into larvae known as schistosomula. The schistosomula enter the bloodstream *via* the capillary beds and by day 6 post-infection are mainly present in the lung capillaries (3,4). After leaving the lung, the parasites circulate within the blood vessels throughout the body while developing organs, growing in size and eventually reaching the liver. Approximately 21 days post-infection, parasite accumulation in the liver peaks (3,5). In the case of *S. mansoni*, male and female parasites pair up over the following weeks, migrate from the liver to the mesenteric veins *via* the portal system and become sexually mature (3). By day-35 the parasites

reach full maturity in the mesenteric veins (3), where they release eggs that traverse the intestinal wall and reach the gut lumen. The eggs pass with faeces to the environment where they hatch free-living larvae that infect snails to continue the life cycle.

Comprehensive studies of intra-mammalian schistosomes have been compromised by difficulties in accessing all developmental stages of the parasite; consequently, studies have focused on juvenile forms just after skin penetration, or adult worms. Little is known about the intervening developmental stages; the lung stage, in particular, represents a key point of parasite attrition during primary and challenged infections of experimentally-infected animals (4,6,7).

Intra-mammalian development of *S. mansoni* has previously been investigated using transcriptomics (8–12) but an incomplete picture has been pieced together from material obtained *in vivo* and *in vitro.* Obtaining sufficient quantities of lung stage parasites in particular has been a major obstacle (8,10,11). Here, we describe the *S. mansoni* transcriptome during six stages of development within experimentally-infected mice, ranging from the early lung stage to sexually mature egg-laying worms. Gene expression changes correlated well with the known biology of egg production in the adult stages, and with growth and developmental control in the liver developmental stages of the parasite. Notably, the lung-stage transcriptome provided new insights into host immune-evasion by the parasite, oxidative stress regulation, as well as potential novel players in host-parasite interactions. This study not only provides novel resources for the community to understand schistosome biology, but also reveals molecular mediators at the host-parasite interface.

Materials and methods

Ethical statement

All procedures involving mice were performed by authorised personnel according to the UK Animals (Scientific Procedures) Act 1986 with project license held by MJD (number PPL 40/3595). The work was approved by the Ethical Review Committee of the University of Nottingham and was carried out in strict accordance with UK Home Office regulations for animal welfare and amelioration of suffering.

Mouse infection

Mice were infected with cercariae shed from *Biomphalaria glabrata* snails as described (13). In brief, percutaneous infections were performed by applying suspensions of mixed-sex cercariae to the shaved abdomens of anesthetised mice and leaving for 30 minutes. Mice were infected with the following numbers of cercariae: 2000 to provide day-6 and day-13 parasites, 500 for days 17 and 21, 350 for day 28, and 300 for day 35 (Fig 1). Four mice were used for adult stage parasites and up to 8 mice were used for juvenile stages (Fig 1). More mice were used for early time points due to the greater uncertainty in acquiring the samples. All mice were females of CD-1 outbred strain (Charles River, Harlow, UK) aged between 8-12 weeks by the time of infection. A pool of mixed-sex parasites from each mouse was considered a biological replicate.

Fig 1. The numbers of mice and cercariae used for infections. Mice were infected with indicated numbers of cercariae for parasite collection at six time points post-infection. The number of mice used for each time point is shown on the left. The method for parasite collection at day 6 involves mincing and incubation of the lung. Collections at other time points were done by portal perfusion.

Parasite material collection and imaging

On the indicated day post-infection, mice were culled using an overdose of pentobarbitone containing 10 U/ml heparin. Day-6 lung-stage parasites were collected as described (13). Briefly, lungs collected from infected mice were minced using a sterile scalpel cut into approximately 1 mm^3 pieces and incubated in 50 ml heparinised modified Basch media (10 U/ml heparin) (see supplementary information) for approximately 1 hour at room temperature, followed by 3 hours at 37 °C, 5% CO_2 to allow the parasites to exit the tissue. The tube contents were mixed by turning the tube 2-3 times before being passed through a 600 µm mesh into new 50 ml tubes to separate large pieces of tissue from worms. The filtrate was centrifuged at 150 x g for 3 minutes at room temperature and approximately half of the supernatant was discarded by gently decanting. Lung-stage worms were recovered using a Pasteur pipette to collect approximately 1-1.5 ml from the bottom of the tube. Given that parasites from day-6 post-infection were collected only from the lung and not by perfusion, circulating parasites that had left the lung were therefore excluded. For this reason, day-6 worms in the present study are hereafter referred to as 'lung stage'.

For all other time points, parasites were collected by portal perfusion with approximately 30 ml of perfusion media (Dulbecco's Modified Eagle's Medium (DMEM) high glucose, with 10 U/ml heparin). Parasites were left to settle for 30 minutes at room temperature and washed twice with DMEM before being recovered with a Pasteur pipette from the bottom of the tube.

Parasites were transferred to a Petri dish for imaging using an Olympus SZ61dissecting microscope with a Euromex Cmex10 camera and Image Focus 4.0 software. A subset of the parasites from each mouse was imaged for morphology scoring. After imaging, parasites were transferred into a 2 ml Eppendorf tube and centrifuged at 150 x g for 3 minutes before replacing the supernatant with 1 ml TRIzol (Thermo Fisher), left at room temperature for up to 1 hour, transferred to dry ice for transport, and later stored at -80 ºC until RNA extraction. An average

of 31 worms were imaged per mouse (range 18-75 worms); this did not represent a total number of parasites collected.

Parasite morphology scoring

Parasite morphology was classified using numerical scores based on published categories (14). Parasite features used for categorisation were the presence of a haemozoin-filled gut, the shape and length of the gut (fork end, closed end, proportion of the posterior end of the gut to the anterior section), and whether the worms were paired or unpaired. The scoring was performed blindly, i.e. all images were renamed to randomised numbers. Different numbers of worms ranging from 18-75 were morphologically scored among replicates; hence, the number of parasites that fell into each morphology category was shown as a percentage of the total worms morphologically scored in that replicate.

RNA extraction

RNA was extracted from parasite material using a modified phenol-chloroform method and column purification. Briefly, frozen samples in TRIzol reagent were thawed on ice, resuspended by gently pipetting and transferred to 2 ml tubes containing ceramic beads (MagNA Lyser Green Beads, Roche). The parasites were homogenised in a MagNA Lyser Instrument (FastPrep-24) at maximum speed twice for 20 seconds, with a 1-minute rest on ice in between. Next, 200 µl of chloroform-isoamyl alcohol 24:1 was added to each tube, followed by vigorous shaking for 5 seconds. The tubes were centrifuged at 13,000 x g for 15 minutes at 4°C to separate the aqueous and organic solvent layers. The aqueous layer was transferred into a RNase-free 1.5 ml tube and one volume of 100% ethanol added and mixed by pipetting. The mixture was transferred to Zymo RNA Clean & Concentrator-5 column (Zymo Research) and processed according to the manufacturer's protocol. To elute the RNA, 15 µl of RNase-free water was added to the column and centrifuged for 30 seconds at 13,000 x g. The RNA concentration

and integrity were measured by Agilent RNA 6000 Nano kit (5067-1511, Agilent Technologies), and its purity assessed using a NanoDrop spectrophotometer.

Library preparation and sequencing

One to 2.8 µg of RNA was used to prepare each sequencing library. The libraries were produced using TruSeq Stranded RNA Sample Preparation v2 Kits (Illumina). Libraries were amplified using 10-14 cycles of PCR and cleaned using Agencourt AMPure XP Beads (Beckman Coulter). The libraries were quantified by qPCR before sequencing using the Illumina HiSeq 2500 platform. All sequencing data was produced as 75 bp paired-end reads and is available through ENA study accession number ERP113121.

Read mapping and quantifying read counts

Sequencing reads were mapped to the *S. mansoni* reference genome v5 (15) from WormBase ParaSite (16) using TopHat version 2.0.8 (17) with default parameters except the following: -g 1 (only report 1 alignment for each read); --library-type fr-firststrand (for dUTP Illumina library); -a 6 (minimum anchor length); -i 10 and --min-segment-intron 10 (minimum intron length); -I 40000 and --max-segment-intron 40000 (maximum intron length); --microexon-search (find alignment to micro-exons). The resulting BAM files of accepted hits were sorted by read name (-n option) and indexed using SAMtools (18). A GFF file of gene annotations from GeneDB.org was filtered to keep only the longest transcript for each gene. The GFF file and sorted BAM files were used as inputs for HTSeq-count version 0.7.1 (19) to obtain read counts per transcript and used for read count analysis. HTSeq-count was run with default parameters except with strand option set to suit dUTP libraries (-s reverse), and alignment score cut-off increased (-a 30).

Differential expression analysis

Analyses were performed using RStudio version 0.99.489 (20), with R version 3.3.1 (21). DESeq2 (version 1.12.3) (22) was used to import read counts, to investigate overall transcriptomic differences among samples using principal component analysis (PCA), to normalise read counts, and to identify genes differentially expressed in the time-course or in pairwise comparisons. PCA used regularized log–transformed (rlog-transformed) read count data as input. Differential expression analyses were performed with either likelihood-ratio tests (when the whole time-course was considered) or with the Wald test (when used with pairwise comparisons) and statistical significance was determined by adjusting p-values according to the Benjamini–Hochberg procedure to control false discovery rate (23). Differentially expressed genes were defined as those with adjusted p-value < 0.01 and $\log_2$ fold change in expression > 1 or < -1.

Gene clustering

Genes were clustered using self-organising maps constructed in the R package Kohonen (version 2.0.19) (24) based on their mean-normalised, rlog-transformed counts over the time-course. The mean-normalised counts were used to calculate means of replicates at each time point for each gene and used as input for clustering. Genes were grouped based on their expression pattern into 96 clusters. To reduce background signal, only genes that were differentially expressed in at least one time point (likelihood ratio test, adjusted p-value < 0.01, 7987 genes, S1 Table) were used as inputs for clustering.

Gene Ontology enrichment

Gene Ontology (GO) annotations for the *S. mansoni* genes were downloaded from GeneDB.org. Identification of enriched GO terms (biological process terms) was performed using topGO (version 2.24.0) (25) with a *weight* algorithm and Fisher's exact test. All *S. mansoni* genes were used as a reference background for the enrichment analysis.

Protein structure prediction

Amino acid sequences for proteins of interest were obtained from GeneDB.org. Protein 3D structures were predicted from amino acid sequences using I-TASSER online server (v5.0) (26–28) with default parameters. TM-scores indicate similarity between two structures. The value ranges between 0-1 with a higher value inferring better match. Images of the predicted structures and their alignment were from .pdb files obtained from I-TASSER predictions or from Protein Data Bank (PDB) (29) and reproduced using Chimera software (30).

Protein domain searching

Protein domains were identified from amino acid sequences using InterProScan online server (v60 and v61) (31). CathDB (32) was used to explore protein structural domains and to search by structural match.

Results

Parasite morphologies correlate with marked transcriptional signatures in developing parasites

Changes in the morphology and transcriptome during intra-mammalian development of *S. mansoni* were investigated in parasites collected from experimentally-infected mice at time points between 6-35 days post-infection (Fig 1).

Lung schistosomula were morphologically homogeneous, whereas circulating larvae that had left the lungs (days 13 to 28) were heterogeneous in size and developmental progression (Fig 2), consistent with previous reports (3,14). At 28 days post infection, most of the parasites had developed into adults and worm pairs started to become evident. All day-35 parasites were fully mature paired male and female adults (Fig 2).

Fig 2. Morphology of *in vivo S. mansoni*. A) Morphological scoring of *S. mansoni* parasites collected at the indicated time points post-infection (D06 to D35). Heatmap columns represent the twenty-one distinct morphological groups following published scores (14), and heatmap rows indicate biological replicates of the infections, i.e. parasites collected from individual mice. Colours on the heatmap show the number of worms that fell into each morphological group, normalised as a percentage of the total number of worms that were scored within the replicate. B) Representative images of worms from each time point. Scale bars: 1 mm.

Parasites from individual mice were pooled and each of these pools was considered a biological replicate. At least 3 biological replicates were obtained for each time point and changes in their transcriptomes were measured using RNA-seq. A principle components analysis showed tight clustering of biological replicates and a large variation among the time points (Fig 3). All replicates from lung schistosomula clustered separately from day-13 to day-21 groups, and the adult stages (days 28 and 35) clustered away from the other time points, indicating a good correlation between the morphological progress and marked transcriptional signatures of the developing parasite. The transcriptional differences observed between day-28 and day-35 parasites may be related to sexual maturation and egg laying, which is onset at day 35 (14). In contrast, the transcriptional similarity of parasites at days 13, 17, and 21 may reflect genuinely similar gene expression profiles or the heterogeneity in the overlapping morphologies of parasites at those three time points.

Fig 3. Principal component analysis of the transcriptomic data from all time points. PCA plot of all transcriptomic data based on rlog-transformed normalised read counts. Each dot represents the transcriptome from a pool of parasites collected from individual mice, i.e. one biological replicate.

Gene expression changes associated to developmental milestones of *S. mansoni*

Comparing the transcriptomes of parasites collected 6 and 13 days post-infection provided information on the transition between lung schistosomula and circulating juveniles, i.e. parasites that have already left the lungs and have entered the systemic circulation *via* the pulmonary veins. In the lung schistosomula, 864 genes were up-regulated compared with day 13 juveniles (S2 Table). The up-regulated genes related to multiple signalling processes including signal transduction and neuronal signalling pathways (S3 Table). This suggests that neuronal activities were increased in the lung stage, compared with day-13 stage, perhaps for sensing and locomotion to allow parasite migration through the lung capillary bed, to reach the pulmonary veins and continue through the rest of bloodstream circulation (33). In contrast, 686 genes were up-regulated in the day-13 schistosomula (S2 Table), including genes involved in mitosis and its associated processes, such as translation, post-translational modification, and transcriptional regulation (S3 Table). This is consistent with the growth phase described in day-13 worms and corroborates reports that mitosis was not detected in the lung stage (3,34).

As the parasites develop from circulating juveniles to adult forms, up-regulated genes identified in the liver stage (day 21) compared to pre-egg-laying adult stage (day 28) were involved in cell division, differentiation, and developmental regulation (S4 and S5 Tables). In contrast, gene expression in the day-28 worms showed a massive up-regulation of genes involved in egg production such as *tyrosinase* (Smp_050270, Smp_013540), *eggshell protein* (Smp_000430), and *Trematode Eggshell Synthesis domain containing protein* (Smp_077890) (S4 Table). The Gene Ontology (GO) terms *metabolic processes* and *biosynthesis processes* were enriched in up-regulated genes at day 28 (S5 Table), possibly also reflecting synthesis of compounds necessary for egg production (*tyrosinase* genes are annotated with the terms *organic hydroxy compound biosynthetic process*). However, the increased biosynthesis also likely reflected increased nutrient and energy requirements, and scavenging of host-derived

280 substrates by the parasites because GO terms for lipid metabolic process, glycerol metabolic
281 process, purine ribonucleoside salvage, and carbohydrate transport were also enriched (S5
282 Table). The genes contributing to the enriched GO term carbohydrate transport encode two
283 confirmed glucose transporters (Smp_012440, Smp_105410) (35) and a third, non-confirmed,
284 putative glucose transporter (Smp_139150).

285 Between days 28 and 35, the parasites become fully established in the portal system
286 within the mesentery veins and lay large numbers of eggs. Expectedly, amongst the 72 genes
287 that were up-regulated during the progression from day 28 to day 35, many were related to egg
288 production (S6 Table). Proteases involved in blood feeding (36), *cathepsin B, D,* and *L* were also
289 up-regulated (approximately 2.5-fold; S6 Table), consistent with the high nutrient requirement
290 of egg-producing females. In addition, down regulation of genes involved in signalling and
291 developmental control was evident (S6 and S7 Tables).

292 To further explore transcriptomic changes across all developmental stages analysed,
293 genes were clustered into 96 groups based on their expression profile over the whole time-
294 course (Fig 4 and S1 Fig). Specific expression patterns became evident for multiple clusters; for
295 instance, clusters 1, 2, 9, 10, 17, 25, 26, 27, 33, and 34 showed increased expression in the
296 developing parasites, from day 13 to day 28 post-infection (Fig 4 orange boxes and S1 Fig).
297 These clusters comprised a total of 737 genes with the top five enriched GO terms related to cell
298 replication and regulation (S8 and S9 Tables). Striking up-regulation in adult stages (day 28 and
299 day 35) was seen in six clusters, particularly cluster 96 and to a lesser extent, clusters 79, 80, 87,
300 88 and 95 (Fig 4 pink boxes, S1 Fig and S8 Table). These clusters comprised genes involved in
301 egg production, such as two tyrosinase genes (Smp_013540 and Smp_050270) involved in
302 protein cross-linking during egg-shell synthesis (37). Two of the egg-shell precursors
303 (Smp_131110 and Smp_014610) (38) were also present in cluster 96 (S8 Table). In addition,
304 multiple genes of unknown function shared expression patterns with these genes related to egg-

305 production, suggesting that they may have a related function or share the same pathway(s) (S8
306 Table).

307 **Fig 4. Clusters of genes based on time-course expression pattern.** Expression profiles of
308 genes showing differential expression in at least one time point, clustered into 96 groups. The
309 clustering was based on mean-normalised regularized log (rlog-transformed) of raw read
310 counts over six time points. The y-axes are scaled independently to emphasise the differences
311 between clusters. Plots with a single y-axis scale are shown in S1 Fig. The coloured boxes mark
312 clusters that were part of the GO term enrichment analyses or were discussed in detail; orange,
313 clusters of genes with increased expression during the liver stage; blue, high expression in the
314 lung stage and steadily declined toward adult stages; yellow, high expression in the lung stage;
315 green, high expression in the lung and adult stages with low expression during liver stages;
316 pink, high expression in adult stages.

317 Genes involved in signalling pathways, iron homeostasis and micro-
318 exon genes (MEGs) up-regulated in lung schistosomula

319 Given the scarcity of information about lung schistosomula, we further investigated the
320 expression data using pairwise comparisons between the lung stage and day-13 parasites, and
321 clustering across multiple time points. Lung stage-specific expression was seen mainly in three
322 clusters (8, 24 and 32), where high expression on day 6 precipitously dropped to a low baseline
323 for the rest of the time-course (Fig 4 yellow boxes, S1 Fig and S8 Table). The three clusters
324 contained a total of 72 genes, and a GO term enrichment analysis suggested that many of these
325 were involved in signalling, metabolism, transport and iron homeostasis (S10 Table).

326 Genes related to developmental control were also over-represented in 11 clusters that
327 showed high expression in lung stage followed by a steady decline towards adult (cluster 5, 6, 7,
328 13, 14, 15, 16, 21, 22, 23, 31; Fig 4 blue boxes, S1 Fig and S11 Table), including several
329 transcription factors and cell adhesion proteins involved in embryogenesis and neuronal

development, such as *SOX* (Smp_148110, Smp_161600) and *procadherin family* (such as Smp_011430, Smp_141740, and Smp_155600) (39,40). In addition, *Wnt, nanos,* and *frizzled receptors,* important for cell-fate determination and control of development (41–43), showed a similar expression pattern.

Compared to day-13 stage, multiple signalling processes appeared to be up-regulated in the lung stage; amongst the 864 genes up-regulated genes, the top-four enriched GO terms were *signal transduction*, *male sex determination*, *potassium ion transport*, and *neurotransmitter transport* (S2 and S3 Tables). The enrichment of several other GO terms, albeit with weaker statistical support (p-values between 0.036-0.045), provided further evidence for the prominent role of signalling processes in the lung stage (S2 and S3 Tables). Several developmental terms were also enriched (e.g. *cell differentiation*, *homophilic cell adhesion*, *brain development*, *male sex determination*). The pronounced expression of genes involved in neuronal signalling – inferred from the GO terms *neuropeptide signalling pathway*, *sodium ion transport*, *chloride transport* and *neurotransmitter transport* (S2 and S3 Tables) – may reflect the parasite's responsiveness to environmental cues. Consistent with the broad picture provided by GO term enrichment, the top up-regulated genes in the lung stage compared to the day-13 stage included a *rhodopsin orphan GPCR* (Smp_203400), and two Ras related proteins (Smp_132500, Smp_125060), all of which were up-regulated more than 32-fold (Table 1).

Table 1. Top 20 genes up-regulated in day-6 compared to day-13 schistosomula

Gene identifier	Log_2FC (D6/D13)	Adjusted p-value	Product name
Smp_138080	12.62	1.48E-19	MEG-3 (Grail) family
Smp_138070	11.99	3.58E-30	MEG-3 (Grail) family
Smp_159810	11.22	1.02E-48	MEG-2 (ESP15) family
Smp_159800	9.66	5.10E-28	MEG-2 (ESP15) family
Smp_181510	9.58	1.15E-13	hypothetical protein
Smp_032990	8.98	7.14E-13	Calmodulin 4 (Calcium binding protein Dd112)

Smp_159830	8.69	1.00E-05	MEG-2 (ESP15) family
Smp_138060	8.06	3.76E-09	MEG-3 (Grail) family
Smp_203400	7.34	1.88E-07	rhodopsin orphan GPCR
Smp_005470	7.31	3.34E-08	dynein light chain
Smp_077610	6.61	6.19E-07	hypothetical protein
Smp_166350	6.59	1.31E-06	hypothetical protein
Smp_180330	5.97	2.91E-32	MEG 2 (ESP15) family
Smp_205660	5.84	4.52E-15	hypothetical protein
Smp_033250	5.80	4.43E-05	hypothetical protein
Smp_132500	5.73	4.50E-04	ras and EF hand domain containing protein
Smp_152730	5.68	3.41E-11	histone lysine N methyltransferase MLL3
Smp_241430	5.61	3.89E-23	Aquaporin 12A
Smp_125060	5.55	6.70E-04	kinase suppressor of Ras (KSR)
Smp_198060	5.51	4.73E-13	hypothetical protein

Log_2FC (D6/D13), log_2 fold change of expression level between day-6 and day-13 schistosomula

The GO term *cellular iron ion homeostasis* was enriched amongst genes up-regulated in lung stage compared to day-13 schistosomula (S3 Table). For example, the *ferritin*-related genes (Smp_047650, Smp_047660, Smp_047680) displayed fold changes in the range 2.1–7.2 (S2 and S3 Tables). These three *ferritin* genes, although down-regulated in day-13 compared to lung stage schistosomula, were still expressed at a low level during day 13 when iron is required for growth and development (S2 Fig). In contrast, other genes related to iron-sequestration were expressed at a similar level in the lung stage and day-13 stage, such as *putative ferric reductase* (Smp_136400) and *Divalent Metal Transporter* (*DMT 1)* (Smp_099850) (S3 Fig). Both genes are hypothesised to be involved in the same pathway, with the *putative ferric reductase* cleaving iron from host transferrin (glycoprotein iron carrier) before transporting into the parasite *via* a *DMT* (44,45).

Micro-exon genes (MEGs), whose structures mainly comprise short exons with lengths being multiples of three bases, are an abundant feature in parasitic helminths (46,47). Despite expression across all developmental stages, and at least 41 MEGs being annotated and assigned into sequence-similarity families (46,48), little is known about their function. *MEG-14* has been shown to interact with an inflammatory-related human protein (49), and *MEG-3*, *MEG-14, MEG-15, MEG-17 and MEG-32*, were previously identified in the oesophagus of schistosomula or adults (50). Multiple MEGs were up-regulated in the lung stage, e.g. 6 of the 72 genes in the major lung stage expression clusters (cluster 8, 24 and 32) were MEGs (S8 Table). Most strikingly, cluster 8 contains eight genes, of which four are micro-exon genes (MEGs); two from the *MEG-2* family and two from *MEG-3* family (S8 Table). Three of these MEGs were previously shown to be up-regulated in schistosomula 3 days after *in vitro* transformation (Smp_159810, Smp_138070, Smp_138080) and detected in schistosomula or egg secretions (48). Furthermore, a pairwise comparison revealed a total of 17 MEGs, from seven families, that were up-regulated in lung stage compared to day-13 schistosomula (S2 Table), with seven MEGs amongst the top 20 lung-stage up-regulated genes (Table 1). In contrast, only one MEG, a member of *MEG-32*, was up-regulated in day-13 schistosomula compared to the lung stage (S2 Table).

Lung stage expressed genes with anti-inflammatory roles

Genes involved in defence against oxidative stress appeared to be up-regulated in lung schistosomula, presumably neutralising reactive oxygen species (ROS) produced during inflammation. For instance, *extracellular superoxide dismutase* was up-regulated by 17-fold compared to day-13 parasites (Smp_174810) (S2 Table and S4 Fig). Superoxide dismutase catalyses the detoxification of superoxide by converting it into hydrogen peroxide and molecular oxygen (45,51). The antioxidant *thioredoxin peroxidase* (Smp_059480) that neutralises peroxide (52) was similarly up-regulated (more than 16-fold) in the lung stage (S2 Table) and thioredoxin glutathione reductase (TGR; Smp_048430), another important enzyme in the redox pathway (53), showed marginally higher expression in the lung stage (S5 Fig). The

388 gene encoding *single Kunitz protease inhibitor* (Smp_147730, S6 Fig), putatively involved in host
389 immune defence (54), was also highly up-regulated in the lung stage (16-fold , adjusted p-value
390 < 1E-100) (S2 Table).

391 These three genes with particularly striking up-regulation in the lung stage belong to
392 the same cluster (cluster 72). Given the possible roles of cluster 72 genes in counteracting
393 oxidative stress and in host immune system evasion, other genes from this cluster were
394 explored in more detail. Cluster 72 contained seven additional genes (Table 2). Three encoded
395 hypothetical proteins of unknown function, while the other four were predicted, based on
396 sequence similarity, to encode proteins with recognisable products. The first, *arginase*
397 (Smp_059980), is hypothesised to counteract the host immune response by depleting l-arginine
398 from blood, thereby preventing it from being used by macrophage in the production of nitric
399 oxide (55,56). The second, a schistosome homologue of *early growth response protein* (EGR,
400 Smp_134870) displayed zinc finger and DNA-binding domains suggesting a role as transcription
401 factor although its target(s), to our knowledge, remain(s) unidentified. The third and fourth
402 proteins contained *PDZ* and *LIM* domains (Smp_156510, Smp_166920) but these also have no
403 obvious link with host immune evasion.

404 **Table 2. Genes in cluster 72.**

Gene identifier	Product name
Smp_059480	thioredoxin peroxidase
Smp_059980	arginase
Smp_074570	hypothetical protein
Smp_114660	hypothetical protein
Smp_134870	early growth response protein
Smp_147730	single Kunitz protease inhibitor; serine type protease inhibitor
Smp_156510	PDZ and LIM domain protein 7
Smp_166920	PDZ and LIM domain protein Zasp

Smp_174810	Extracellular superoxide dismutase (Cu Zn)
Smp_182770	hypothetical protein

Cluster 64 contained genes with similar expression profiles and, together with cluster 72, was annotated with GO terms enriched for redox processes (S12 Table). Clusters 64 and 72 also contained genes involved in cellular metabolism and transport. However, enrichment for these specific GO terms became more pronounced when two additional clusters (71 and 80), with reduced lung stage expression (Fig 4 green boxes), were included in the analysis (S13 and S14 Tables). Amongst the top enriched GO terms for all four clusters was *carbohydrate transport* with two glucose transporters (Smp_012440 and Smp_105410) that have been previously characterised in *S. mansoni* (35).

Hypothetical protein with predicted structure matching a complement cascade regulator

With four genes out of ten in the cluster 72 potentially involved host immune interactions, the three that encoded hypothetical proteins (Smp_074570, Smp_114660, Smp_182770) were investigated further. Peptides encoded by Smp_074570 are abundant (top 10) in secreted extracellular vesicles produced by *S. mansoni* schistosomula (57), suggesting a role in host-parasite interactions. No further published information was found for products of the other two genes, and both contained no known signature domains. I-TASSER protein-structure prediction software (26) was therefore used to align predicted structures against known structures in the Protein Data Bank (29). No match was found for Smp_074570 and Smp_114660, but remarkably, Smp_182770 produced a single high confidence match (TM-score > 0.9) to human complement factor H (CFH; Protein Data Bank ID: 3GAW) (58) (Fig 5).

Fig 5. Predicted structure of Smp_182770 aligned with structure of human CFH. A) Predicted 3D structure of Smp_182770 by I-TASSER server from the input amino acid sequence.

B) Alignment between the predicted structure (blue) and 3D structure of human CFH, in 250 mM NaCl buffer, obtained from PDB (PDB identifier: 3GAW) (red). C) Domain components of human CHF identified by multiple databases through InterProScan web server (31). SCR, short consensus repeat; CCP, complement control protein. Both SCR and CCP are alternative names of the sushi domain.

CFH is a well-characterised regulator of the complement cascade that is normally found on the surface of human cells and prevents complement attack on self (59,60). The human CFH (amino acid sequence from 3GAW entry, PDB) contains *sushi/ccp* domain repeats (Fig 5C) involved in regulating complement cascade (ccp = component control protein). Smp_182770 does not contain the *sushi/ccp* domain repeats but it does encode tandem repeats similar to those expected from mammalian CFH genes (61).

Homologues of the Smp_182770 are present in other *Schistosoma* species as well as other helminths, such as the liver flukes *Fasciola*, *Opisthorchis*, and *Clonorchis*, and the intestinal fluke *Echinostoma* (Fig 6A), suggesting an evolutionarily-conserved role (62,63). The Smp_182770 locus on *S. mansoni* genome is immediately adjacent to Smp_038730, another hypothetical protein. RNA-seq mapping suggested that the two were parts of the same gene (Fig 6B). From gene clustering, Smp_038730 is in cluster 24 which has similar expression pattern to cluster 72, but with noticeably lower expression in the adult stage (Fig 4 yellow box and S1 Fig). In a subsequent version of the *S. mansoni* genome (unpublished), the two genes have been merged to become Smp_334090 (data available on WormbaseParasite.org). The predicted structure of Smp_334090 also resembles that of human CFH (TM-score = 0.830) (S7 Fig).

Fig 6. Homologous relationship of Smp_182770 and alignment of RNA-seq reads to genomic locations. A) Homologues of Smp_182770. The information was obtained from WormBase ParaSite release 8 (16). Gene identifiers from the WormBase ParaSite are shown in the parentheses after the species names. B) Artemis screenshot showing the genomic region that contains Smp_182770 and RNA-seq reads mapped from multiple RNA-seq libraries (two

454 top rows). Information on homologues of genes was obtained from WormBase ParaSite release
455 8 and 9 (16) based on genetree generated by Compara pipeline (64).

456 Discussion

457 We have described a complete transcriptome time-course of *S. mansoni* covering key
458 developmental stages during the intra-mammalian infection, including the first RNA-seq
459 transcriptome of *in vivo* lung schistosomula. The data recapitulated known major biological
460 changes for well-characterised parts of the life cycle in the mammalian host, as well as provided
461 novel insights on molecular processes underlying parasite development and interaction with
462 the host. In particular, analysis of the lung developmental stage highlighted striking signalling
463 pathways, including those related to developmental control, cell differentiation and
464 neuropeptide signalling. In addition, tentative strategies for immune evasion such as up-
465 regulation of iron homeostasis, oxidative stress-related genes, and anti-inflammatory genes
466 were revealed.

467 In early stages of *S. mansoni* infection, schistosomula elongate and migrate within lung
468 capillaries (65). Although the cues involved in these changes are unknown (66), it is believed
469 that schistosomula differentiate or remodel their existing cells, rather than trigger cell division
470 (3,34). The up-regulation of signalling-related genes in this developmental stage emphasises
471 processes required for a successful migration through lung capillaries. Furthermore, changes in
472 development-related genes may play roles in the remodelling or trigger developmental onset
473 observed in later stages. Neuronal signalling could involve both the sensing of stimuli and motor
474 responses, and it is one of the biological processes disturbed in irradiated schistosomes (67)
475 that exit lung capillaries into the alveoli and eventually die during the lung migration (7,65,68).
476 The migration through the lung capillaries represents another challenge for the schistosomula.
477 Moving within narrow lung capillaries, the schistosomula are in direct contact with the capillary
478 endothelium (65) and they may interrupt blood flow, increasing the risk of microthrombosis or

479 leading to an accumulation of circulating immune cells (69). However, it has been shown that
480 lung schistosomula can resist direct cytotoxic immune killing (6,68,70,71). Inflammation can be
481 observed in the lung during the infection, but it appears not to be associated with the location of
482 migrating parasites, but instead triggered by damaged host tissue (72). The accumulation of
483 immune cells, instead of killing the parasites, is thought to either disrupt blood vessels, inducing
484 microhemorrhage and causing the parasites to migrate into alveoli, or it may act as a plug that
485 blocks migration (5,7).

486 Our data revealed that multiple genes with potential immune evasion or protective roles
487 are up-regulated. One example was the Kunitz protease inhibitor that has an anti-coagulation
488 and anti-inflammation role (54) and has been proposed as a vaccine candidate against *S.*
489 *mansoni* (73). Neutralisation of oxidative stress by up-regulating dismutase and peroxidase
490 systems may mitigate the effects of inflammation caused by migrating schistosomula. Both ROS
491 and iron metabolism are involved in the host immune response; therefore, iron homeostasis
492 and ROS neutralisation have likely evolved as an immune evasion strategy in these parasites
493 (45). Schistosomula also clearly face the risk of oxidative damage from the uniquely high oxygen
494 tension in the lung (74,75) and therefore have a further requirement for maintaining redox
495 balance. In rodents, expression of thioredoxin is induced by oxidative stress (75). The up-
496 regulation of schistosome ROS metabolism likely also depends on changes in the environment
497 rather than a hard-wired gene expression programme. In contrast to our *in vivo* results from
498 lung schistosomula, genes related to stress and immune evasion are expressed at a lower level
499 when the parasites are cultivated *in vitro* (8), highlighting the need to study parasites obtained
500 from their most natural environment to understand relevant biological processes.

501 Multiple genes with tentative roles in immune evasion displayed particularly high
502 expression in the lung stage, moderately high expression in adult stages and lower expression in
503 the liver stages. In particular, a previously uncharacterised gene annotated as a hypothetical
504 protein, with this expression pattern was predicted to be structurally similar to the human

505 complement factor H (CFH), a regulator of the complement cascade. In the mammals, CFH
506 cleaves C3b, a central protein in the complement cascade (76). The regulatory function of CFH
507 involves two other molecules - its co-factor complement factor I (CFI) (60) and decay-
508 accelerating factor (DAF) (59). To date, a *S. mansoni* gene with sequence or structural similarity
509 to CFI has not been described, but the parasites do possess a serine protease, m28, that cleaves
510 C3bi (a function normally carried out by CFI (59,77)). For DAF, a native *S. mansoni* form has not
511 been reported but schistosomes can acquire DAF from host blood cells (78,79). Given the nature
512 of other genes sharing its expression pattern, the structural match to human CFH, and its
513 phylogenetic relationship, it is tempting to speculate that the CFH-like gene in schistosome may
514 have a role in immunomodulation. Nevertheless, host immune system proteins have been
515 shown to interact with schistosomes in their developmental signalling pathways (80). CFH in
516 mammals not only cleaves C3b, but also has been implicated in lymphocyte extracellular
517 secretion and DNA synthesis (80). In addition, schistosomes have a 130 kDa surface protein that
518 can bind to C3, inactivating the C3 complement-activation, and the binding stimulates renewal
519 of the parasite surface membrane (76,81). To have another, CFH-like protein identified here,
520 inhibiting the downstream product of C3 activation, may seem redundant. However, its
521 potential roles in immunomodulation should not be ruled out. It is common for pathways to
522 have regulators at various steps, particularly for C3 which is focal for complement cascade
523 activation and is involved in clearing *S. mansoni* after praziquantel administration (76,82).

524 The data produced from this study will serve as a unique resource for the research
525 community to explore changes across intra-mammalian stages of schistosome development.
526 Our particular focus on the lung stage demonstrated consistency with previous observations
527 and introduced potential new players in host-parasite interactions and parasite development.
528 Further investigation and functional validation of genes identified here will help decipher
529 mechanisms for parasite long-term survival within the mammalian host, exposing
530 vulnerabilities that can be exploited to develop new control strategies for this neglected tropical
531 pathogen.

Acknowledgement

We thank Prof Karl Hoffmann and Dr Cinzia Cantacessi for their comments on the study and the first version of this manuscript. We thank multiple members of the Parasite Genomics team at the Wellcome Sanger Institute for their comments and input for the experimental design and analysis; in particular, we thank Drs Hayley Bennett, Lia Chappell, James Cotton, Stephen Doyle, Magda Lotkowska, Thomas Otto, Kate Rawlinson, Adam Reid, and Gavin Rutledge. The infrastructure used for the analysis is maintained by the core IT Service and the Pathogen Informatics teams at the Wellcome Sanger Institute.

References

1. Gryseels B, Polman K, Clerinx J, Kestens L. Human schistosomiasis. Lancet. 2006 Sep;368(9541):1106–18.
2. Crellen T, Walker M, Lamberton PHL, Kabatereine NB, Tukahebwa EM, Cotton JA, et al. Reduced efficacy of praziquantel against *Schistosoma mansoni* is associated with multiple rounds of mass drug administration. Clinical Infectious Diseases. 2016 Jul 28;63(9):1151–1159.
3. Clegg JA. *In vitro* cultivation of *Schistosoma mansoni*. Experimental Parasitology. 1965;16(2):133–147.
4. Georgi JR, Wade SE, Dean DA. Attrition and temporal distribution of *Schistosoma mansoni* and *S. haematobium* schistosomula in laboratory mice. Parasitology. 1986 Apr;93(1):55.
5. Wilson RA. The saga of schistosome migration and attrition. Parasitology. 2009;136(12):1581–1592.
6. Crabtree JE, Wilson RA. The role of pulmonary cellular reactions in the resistance of vaccinated mice to *Schistosoma mansoni*. Parasite Immunology. 1986 May;8(3):265–285.

555 7. Dean DA, Mangold BL. Evidence that both normal and immune elimination of *Schistosoma*
556 *mansoni* take place at the lung stage of migration prior to parasite death. Am J Trop Med
557 Hyg. 1992;47(2):238–248.
558 8. Chai M, McManus DP, McInnes R, Moertel L, Tran M, Loukas A, et al. Transcriptome
559 profiling of lung schistosomula, *in vitro* cultured schistosomula and adult *Schistosoma*
560 *japonicum*. Cellular and Molecular Life Sciences. 2006 Mar;63(7):919–929.
561 9. Jolly ER, Chin C-S, Miller S, Bahgat MM, Lim KC, DeRisi J, et al. Gene expression patterns
562 during adaptation of a helminth parasite to different environmental niches. Genome
563 biology. 2007 Jan;8(4):R65.
564 10. Fitzpatrick JM, Peak E, Perally S, Chalmers IW, Barrett J, Yoshino TP, et al. Anti-
565 schistosomal intervention targets identified by lifecycle transcriptomic analyses. PLoS
566 Neglected Tropical Diseases. 2009 Jan;3(11):e543.
567 11. Gobert GN, Moertel L, Brindley PJ, McManus DP. Developmental gene expression profiles of
568 the human pathogen *Schistosoma japonicum*. BMC Genomics. 2009;10:128.
569 12. Parker-Manuel SJ, Ivens AC, Dillon GP, Wilson RA. Gene expression patterns in larval
570 *Schistosoma mansoni* associated with infection of the mammalian host. PLoS Neglected
571 Tropical Diseases. 2011 Aug;5(8):e1274.
572 13. Tucker MS, Karunaratne LB, Lewis FA, Freitas TC, Liang Y. Schistosomiasis. In: Coligan JE,
573 Bierer BE, Margulies DH, Shevach EM, Strober W, editors. Current Protocols in
574 Immunology. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2013. p. 19.1.1-19.1.58. Available
575 from: http://doi.wiley.com/10.1002/0471142735.im1901s103 Cited [cited 8 July 2019]
576 14. Basch PF. Cultivation of *Schistosoma mansoni in vitro*. I. Establishment of cultures from
577 cercariae and development until pairing. The Journal of Parasitology. 1981 Apr;67(2):179–
578 85.
579 15. Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M, Aslett MA, et al. A systematically
580 improved high quality genome and transcriptome of the human blood fluke *Schistosoma*
581 *mansoni*. PLoS Neglected Tropical Diseases. 2012;6(1).

582 16. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite – a comprehensive
583 resource for helminth genomics. Molecular and Biochemical Parasitology. 2017 Jul;215:2–
584 10.
585 17. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate
586 alignment of transcriptomes in the presence of insertions, deletions and gene fusions.
587 Genome Biology. 2013;14(4):R36.
588 18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
589 Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009
590 Aug;25(16):2078–9.

591 19. Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput
592 sequencing data. Bioinformatics. 2015 Jan 15;31(2):166–169.
593 20. RStudio Team. RStudio: Integrated Development for R [Internet]. Boston, MA: RStudio,
594 Inc.; 2016. Available from: http://www.rstudio.com/.
595 21. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna,
596 Austria: R Foundation for Statistical Computing; 2016. Available from: https://www.R-
597 project.org/

598 22. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
599 seq data with DESeq2. Genome Biology. 2014 Dec 5;15(12):550.
600 23. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in
601 behavior genetics research. Behav Brain Res. 2001 Nov 1;125(1):279–284.
602 24. Wehrens R, Buydens LMC. Self- and Super-organising maps in R: the kohonen package. J
603 Stat Softw. 2007;21(5).
604 25. Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. 2016. DOI:
605 10.18129/B9.bioc.topGO
606 26. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008
607 Jan 23;9:40.

608 27. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure
609 and function prediction. Nat Protoc. 2010 Apr;5(4):725–738.
610 28. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and
611 function prediction. Nat Methods. 2015 Jan;12(1):7–8.
612 29. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data
613 Bank. Nucleic Acids Res. 2000 Jan 1;28(1):235–242.
614 30. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF
615 Chimera—a visualization system for exploratory research and analysis. J Comput Chem.
616 2004 Oct;25(13):1605–12.
617 31. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017—
618 beyond protein family and domain annotations. Nucleic Acids Research. 2017 Jan
619 4;45(D1):D190–D199.
620 32. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, et al. CATH: comprehensive
621 structural and functional annotations for genome sequences. Nucleic Acids Research. 2015
622 Jan 28;43(D1):D376–D381.
623 33. Miller P, Wilson RA. Migration of the schistosomula of *Schistosoma mansoni* from the lungs
624 to the hepatic portal system. Parasitology. 1980;80:267–288.
625 34. Lawson JR, Wilson RA. Metabolic changes associated with the migration of the
626 schistosomulum of *Schistosoma mansoni* in the mammal host. Parasitology. 1980
627 Oct;81(2):325–36.
628 35. Krautz-Peterson G, Simoes M, Faghiri Z, Ndegwa D, Oliveira G, Shoemaker CB, et al.
629 Suppressing glucose transporter gene expression in schistosomes impairs parasite feeding
630 and decreases survival in the mammalian host. PLoS Pathogens. 2010;6(6).
631 36. Delcroix M, Sajid M, Caffrey CR, Lim K-C, Dvorák J, Hsieh I, et al. A multienzyme network
632 functions in intestinal protein digestion by a platyhelminth parasite. J Biol Chem. 2006 Dec
633 22;281(51):39316–39329.

634 37. Fitzpatrick JM, Hirai Y, Hirai H, Hoffmann KF. Schistosome egg production is dependent
635 upon the activities of two developmentally regulated tyrosinases. The FASEB Journal. 2007
636 Mar;21(3):823–35.
637 38. Buro C, Oliveira KC, Lu Z, Leutner S, Beckmann S, Dissous C, et al. Transcriptome analyses
638 of inhibitor-treated schistosome females provide evidence for cooperating Src-kinase and
639 TGFβ receptor pathways controlling mitosis and eggshell formation. PLoS Pathogens.
640 2013;9(6):e1003448.

641 39. Kamachi Y, Kondoh H. Sox proteins: regulators of cell fate specification and differentiation.
642 Development. 2013 Oct;140(20):4129–44.
643 40. Paulson AF, Prasad MS, Thuringer AH, Manzerra P. Regulation of cadherin expression in
644 nervous system development. Cell Adh Migr. 2014;8(1):19–28.
645 41. Park WJ, Liu J, Adler PN. Frizzled gene expression and development of tissue polarity in the
646 *Drosophila* wing. Dev Genet. 1994;15(4):383–389.
647 42. Logan CY, Nusse R. The Wnt signaling pathway in development and disease. Annu Rev Cell
648 Dev Biol. 2004;20:781–810.
649 43. Tsuda M, Sasaoka Y, Kiso M, Abe K, Haraguchi S, Kobayashi S, et al. Conserved role of nanos
650 proteins in germ cell development. Science. 2003 Aug 29;301(5637):1239–41.

651 44. Smyth DJ, Glanfield A, McManus DP, Hacker E, Blair D, Anderson GJ, et al. Two isoforms of a
652 divalent metal transporter (dmt1) in *Schistosoma mansoni* suggest a surface-associated
653 pathway for iron absorption in schistosomes. Journal of Biological Chemistry. 2006
654 Jan;281(4):2242–2248.
655 45. Glanfield A, McManus DP, Anderson GJ, Jones MK. Pumping iron: a potential target for
656 novel therapeutics against schistosomes. Trends in Parasitology. 2007 Dec;23(12):583–8.
657 46. Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, et al. The genome of
658 the blood fluke *Schistosoma mansoni*. Nature. 2009 Jul 16;460(7253):352–8.

659 47. Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, Brooks KL, et al. The
660 genomes of four tapeworm species reveal adaptations to parasitism. Nature. 2013 Mar
661 13;496(7443):57–63.
662 48. DeMarco R, Mathieson W, Manuel SJ, Dillon GP, Curwen RS, Ashton PD, et al. Protein
663 variation in blood-dwelling schistosome worms generated by differential splicing of
664 micro-exon gene transcripts. Genome research. 2010 Aug;20(8):1112–21.
665 49. Orcia D, Zeraik AE, Lopes JLS, Macedo JNA, Santos CR dos, Oliveira KC, et al. Interaction of
666 an esophageal MEG protein from schistosomes with a human S100 protein involved in
667 inflammatory response. Biochimica et Biophysica Acta (BBA) - General Subjects. 2017
668 Jan;1861(1):3490–7.
669 50. Wilson RA, Li XH, MacDonald S, Neves LX, Vitoriano-Souza J, Leite LCC, et al. The
670 schistosome esophagus is a 'hotspot' for microexon and lysosomal hydrolase gene
671 expression: Implications for blood processing. Ribeiro JMC, editor. PLOS Neglected
672 Tropical Diseases. 2015 Dec 7;9(12):e0004272.
673 51. Afonso V, Champy R, Mitrovic D, Collin P, Lomri A. Reactive oxygen species and superoxide
674 dismutases: Role in joint diseases. Joint Bone Spine. 2007 Jul;74(4):324–329.
675 52. Kwatia MA, Botkin DJ, Williams DL. Molecular and enzymatic characterization of
676 *Schistosoma mansoni* thioredoxin peroxidase. The Journal of parasitology.
677 2000;86(5):908–915.
678 53. Prast-Nielsen S, Huang H-H, Williams DL. Thioredoxin glutathione reductase: Its role in
679 redox biology and potential as a target for drugs against neglected diseases. Biochimica et
680 Biophysica Acta (BBA) - General Subjects. 2011 Dec;1810(12):1262–71.
681 54. Ranasinghe SL, Fischer K, Gobert GN, McManus DP. Functional expression of a novel Kunitz
682 type protease inhibitor from the human blood fluke *Schistosoma mansoni*. Parasites &
683 Vectors. 2015 Aug 4;8(1):408.

684 55. Fitzpatrick JM, Fuentes JM, Chalmers IW, Wynn TA, Modolell M, Hoffmann KF, et al.
685 *Schistosoma mansoni* arginase shares functional similarities with human orthologs but

depends upon disulphide bridges for enzymatic activity. Int J Parasitol. 2009 Feb;39(3):267–279.

56. Hai Y, Edwards JE, Van Zandt MC, Hoffmann KF, Christianson DW. Crystal structure of *Schistosoma mansoni* arginase, a potential drug target for the treatment of schistosomiasis. Biochemistry. 2014;53(28):4671–4684.

57. Nowacki FC, Swain MT, Klychnikov OI, Niazi U, Ivens A, Quintana JF, et al. Protein and small non-coding RNA-enriched extracellular vesicles are released by the pathogenic blood fluke *Schistosoma mansoni*. Journal of Extracellular Vesicles. 2015;4:1–16.

58. Okemefuna AI, Nan R, Gor J, Perkins SJ. Electrostatic interactions contribute to the folded-back conformation of wild type human factor H. Journal of Molecular Biology. 2009 Aug;391(1):98–118.

59. Morgan HP, Schmidt CQ, Guariento M, Blaum BS, Gillespie D, Herbert AP, et al. Structural basis for engagement by complement factor H of C3b on a self surface. Nature structural & molecular biology. 2011;18(4):463–70.

60. Medjeral-Thomas N, Pickering MC. The complement factor H-related proteins. Immunological Reviews. 2016;274(1):191–201.

61. Kalenda YDJ, Kato K, Goto Y, Fujii Y, Hamano S. Tandem repeat recombinant proteins as potential antigens for the sero-diagnosis of *Schistosoma mansoni* infection. Parasitology International. 2015 Dec;64(6):503–12.

62. Sirisinha S, Rattanasiriwilai W, Puengtomwatanakul S, Sobhon P, Saitongdee P, Koonchornboon T. Complement-mediated killing of *Opisthorchis viverrini* via activation of the alternative pathway. Int J Parasitol. 1986 Aug;16(4):341–346.

63. Schroeder H, Skelly PJ, Zipfel PF, Losson B, Vanderplasschen A. Subversion of complement by hematophagous parasites. Developmental & Comparative Immunology. 2009 Jan;33(1):5–13.

64. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 2009 Feb;19(2):327–335.

65. Crabtree JE, Wilson RA. *Schistosoma mansoni*: an ultrastructural examination of pulmonary migration. Parasitology. 1986 Apr;92 (Pt 2):343–54.

66. Kusel JR, Al-Adhami BH, Doenhoff MJ. The schistosome in the mammalian host: understanding the mechanisms of adaptation. Parasitology. 2007 Oct;134(Pt 11):1477–526.

67. Dillon GP, Feltwell T, Skelton J, Coulson PS, Wilson RA, Ivens AC. Altered patterns of gene expression underlying the enhanced immunogenicity of radiation-attenuated schistosomes. PLoS neglected tropical diseases. 2008 Jan;2(5):e240.

68. Mastin A, Bickle QD, Wilson RA. An ultrastructural examination of irradiated, immunizing schistosomula of *Schistosoma mansoni* during their extended stay in the lungs. Parasitology. 1985 Aug;91 (Pt 1):101–110.

69. Mebius MM, van Genderen PJJ, Urbanus RT, Tielens AGM, de Groot PG, van Hellemond JJ. Interference with the host haemostatic system by schistosomes. PLoS Pathogens. 2013 Jan;9(12):e1003781.

70. Bentley AG, Carlisle AS, Phillips SM. Ultrastructural analysis of the cellular response to *Schistosoma mansoni*. II. Inflammatory response in rodent skin. The American Journal of Tropical Medicine and Hygiene. 1981;30(4):815–824.

71. McLaren DJ, Terry RJ. The protective role of acquired host antigens during schistosome maturation. Parasite Immunology. 1982;4:129–148.

72. Burke ML, McGarvey L, McSorley HJ, Bielefeldt-Ohmann H, McManus DP, Gobert GN. Migrating *Schistosoma japonicum* schistosomula induce an innate immune response and wound healing in the murine lung. Molecular immunology. 2011 Oct;49(1):191–200.

73. Hernández-Goenaga J, López-Abán J, Protasio AV, Vicente Santiago B, del Olmo E, Vanegas M, et al. Peptides derived of kunitz-type serine protease inhibitor as potential vaccine

against experimental schistosomiasis. bioRxiv [Internet]. 2019 May 10 [cited 2019 Aug 14]; Available from: http://biorxiv.org/lookup/doi/10.1101/634360

74. Dirks RC, Faiman MD. Free radical formation and lipid peroxidation in rat and mouse cerebral cortex slices exposed to high oxygen pressure. Brain Research. 1982 Sep;248(2):355–360.

75. Kobayashi-Miura M, Shioji K, Hoshino Y, Masutani H, Nakamura H, Yodoi J. Oxygen sensing and redox signaling: the role of thioredoxin in embryonic development and cardiac diseases. Am J Physiol Heart Circ Physiol. 2007 May;292(5):H2040–2050.

76. Skelly PJ. Intravascular schistosomes and complement. Trends in Parasitology. 2004 Aug;20(8):370–4.

77. Fishelson Z. Novel mechanisms of immune evasion by *Schistosoma mansoni*. Mem Inst Oswaldo Cruz. 1995 Apr;90(2):289–292.

78. Horta MF, Ramalho-Pinto FJ. Role of human decay-accelerating factor in the evasion of *Schistosoma mansoni* from the complement-mediated killing *in vitro*. The Journal of experimental medicine. 1991 Dec;174(6):1399–406.

79. Ramalho-Pinto FJ, Carvalho EM, Horta MF. Mechanisms of evasion of *Schistosoma mansoni* schistosomula to the lethal activity of complement. Mem Inst Oswaldo Cruz. 1992;87 Suppl 4:111–116.

80. Inal JM. Parasite interaction with host complement: beyond attack regulation. Trends in Parasitology. 2004 Sep;20(9):407–12.

81. Silva EE, Clarke MW, Podesta RB. Characterization of a C3 receptor on the envelope of *Schistosoma mansoni*. Journal of Immunology. 1993;151:7057–7066.

82. La Flamme AC, MacDonald AS, Huxtable CR, Carroll M, Pearce EJ. Lack of C3 affects Th2 response development and the sequelae of chemotherapy in schistosomiasis. J Immunol. 2003 Jan 1;170(1):470–6.

Supporting information

S1 Fig. Clusters of genes based on timecourse expression pattern, with fixed y-axes. Expression profile of genes differentially expressed in at least one time point clustered into 96 groups. The clustering was done on mean-normalised regularized log transformation (rlog-transformed) of raw read counts. X-axes represent six time points from this dataset; y-axes represent the mean-normalised rlog-transformed. Unlike Fig 4, this supplementary figure show clusters with the fixed range for y-axis across all clusters in order to visualise clusters with the largest changes over time.

S2 Fig. Expression of genes encoding ferritin-heavy chain over the timecourse. Each dot represents one replicate from each of the time points. Y-axis represent normalised counts from DESeq2.

S3 Fig. Expression of genes encoding putative ferric reductase (cytochrome b 561; Smp_136400) and Divalent Metal Transporter (DMT1) (Smp_099850) over the timecourse. Each dot represents one replicate from each of the time points. Y-axis represent normalised counts from DESeq2.

S4 Fig. Expression of genes encoding two extracellular superoxide dismutases (Smp_174810 and Smp_095980) over the timecourse. Expression of Smp_174810 and Smp_095980 with y axes on log scale. Each dot represents one replicate from each of the time points. Y-axis represent normalised counts from DESeq2. Smp_095980 was identified as differentially expressed (S2 Table and S4 Fig), but this figure shows that its expression in the lung stage was high in only one out of seven replicates.

S5 Fig. Expression of genes encoding thioredoxin glutathione reductase over the timecourse. Each dot represents one replicate from each of the time points. Y-axis represent normalised counts from DESeq2. Log_2FC between D13/D06 is -0.52, adjusted p-value for differential expression between D13/D06 is 4.09675e-21.

789 **S6 Fig. Expression of genes encoding single Kunitz serine protease inhibitor over the**
790 **timecourse.** Each dot represents one replicate from each of the time points. Y-axis represent
791 normalised counts from DESeq2.
792 **S7 Fig. Predicted structure of Smp_334090 aligned with structure of human CFH.** A)
793 Predicted 3D structure of Smp_334090 (a resulting merge of Smp_182770 and Smp_038730 in
794 the most recent version of *S. mansoni* genome; unpublished) by I-TASSER server from the input
795 amino acid sequence. B) Alignment between the predicted structure (blue) and 3D structure of
796 human CFH (in 137 mM NaCl buffer) obtained from PDB (PDB identifier: 3GAV) (red).
797 **S1 Table.** Genes that were differentially expressed in at least one time point (likelihood ratio
798 test, adjusted p-value < 0.01)
799 **S2 Table.** Differentially expressed genes between day-13 and lung stage *S. mansoni*
800 **S3 Table.** Enriched GO terms of genes differentially expressed between day-13 and lung stage *S.*
801 *mansoni*
802 **S4 Table.** Differentially expressed genes between day-28 and day-21 *S. mansoni*
803 **S5 Table.** Enriched GO terms of genes differentially expressed between day-28 and day-21 *S.*
804 *mansoni*
805 **S6 Table.** Differentially expressed genes between day-35 and day-28 *S. mansoni*
806 **S7 Table.** Enriched GO terms of genes differentially expressed between day-35 and day-28 *S.*
807 *mansoni*
808 **S8 Table.** Genes differentially expressed in at least one time point identified by their clustered
809 expression profiles
810 **S9 Table.** GO term enrichment of genes with high expression during liver stages (genes in
811 cluster 1, 2, 9, 10, 17, 25, 26, 27, 33, 34)
812 **S10 Table.** GO enrichment of genes up-regulated in lung stage (genes in cluster 8, 24, 32)
813 **S11 Table.** Enriched GO terms of genes with high expression in lung stage followed by a steady
814 decline toward adult (genes in cluster 5, 6, 7, 13, 14, 15, 16, 21, 22, 23, 31)

S12 Table. Enriched GO terms of genes with high expression in lung stage, low expression in liver stages, and increased expression in adult stages (genes in cluster 64, 72)

S13 Table. Enriched GO terms of genes with high expression in lung stage, low expression in liver stages, and increased expression in adult stages (genes in cluster 71, 80)

S14 Table. Enriched GO terms of genes with high expression in lung stage, low expression in liver stages, and increased expression in adult stages (genes in cluster 64, 71, 72, 80)

S1 File. Modified Basch media components.

S2 File. Smansoni_longest_transcript_htseq_friendly.gff. Genome annotation file used for generating read counts.

S3 File. Final_counts. A zipped folder containing the following files:

D06_SM_1_17675_4_1.htseq-count.txt

D06_SM_2_17675_4_2.htseq-count.txt

D06_SM_3_17675_4_3.htseq-count.txt

D06_SM_4_17675_4_4.htseq-count.txt

D06_SM_5_17675_4_5.htseq-count.txt

D06_SM_6_17675_4_6.htseq-count.txt

D06_SM_7_17675_4_7.htseq-count.txt

D13_SM_1_17675_4_8.htseq-count.txt

D13_SM_2_17675_4_9.htseq-count.txt

D13_SM_3_17675_4_10.htseq-count.txt

D17_SM_1_17675_4_11.htseq-count.txt

836 D17_SM_2_17675_4_12.htseq-count.txt

837 D17_SM_3_17675_4_13.htseq-count.txt

838 D21_SM_1_17675_4_14.htseq-count.txt

839 D21_SM_2_17675_4_15.htseq-count.txt

840 D21_SM_3_17675_4_16.htseq-count.txt

841 D28_SM_1_17675_4_17.htseq-count.txt

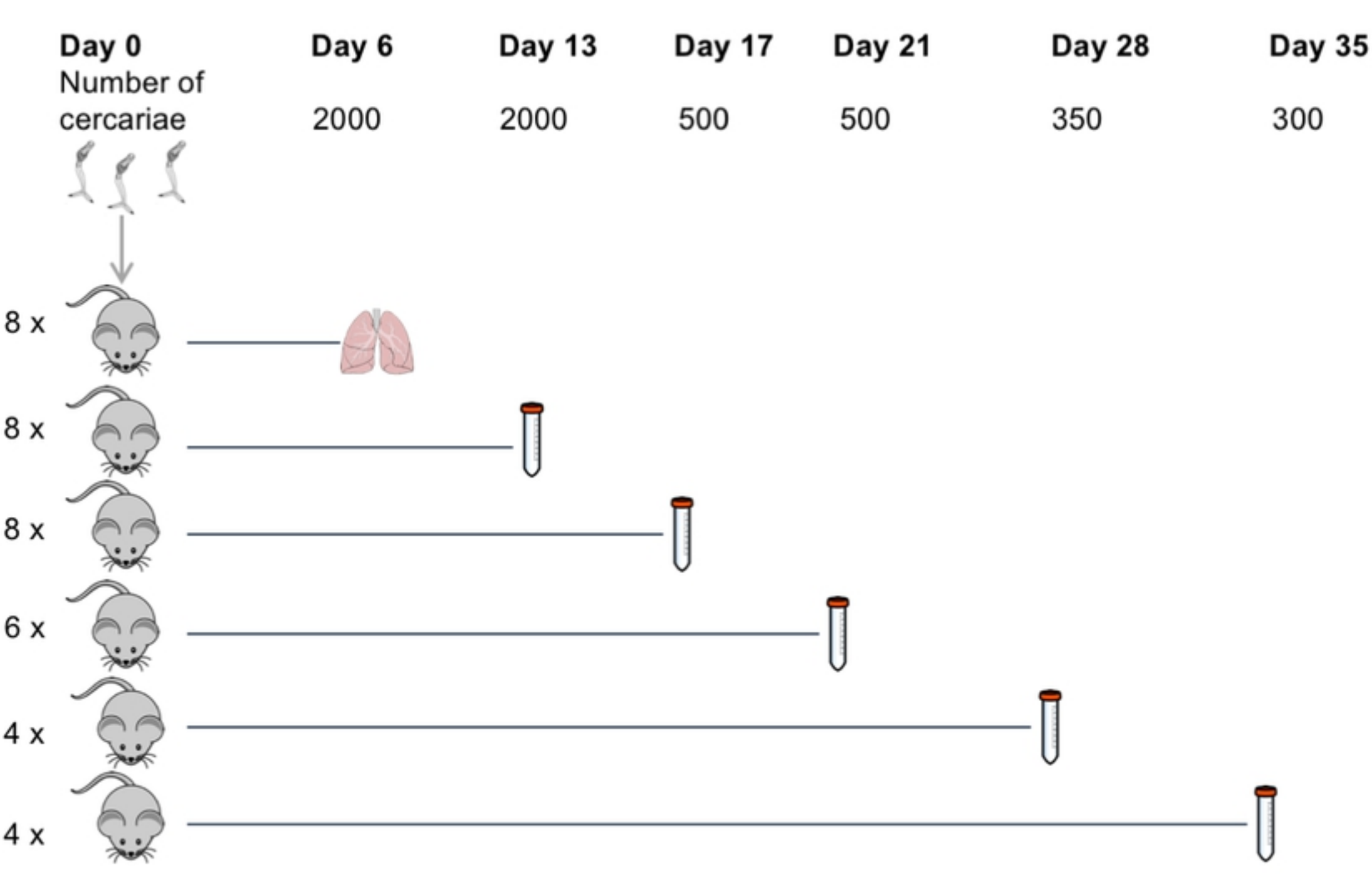
842 D28_SM_2_17675_4_18.htseq-count.txt

843 D28_SM_3_17675_4_19.htseq-count.txt

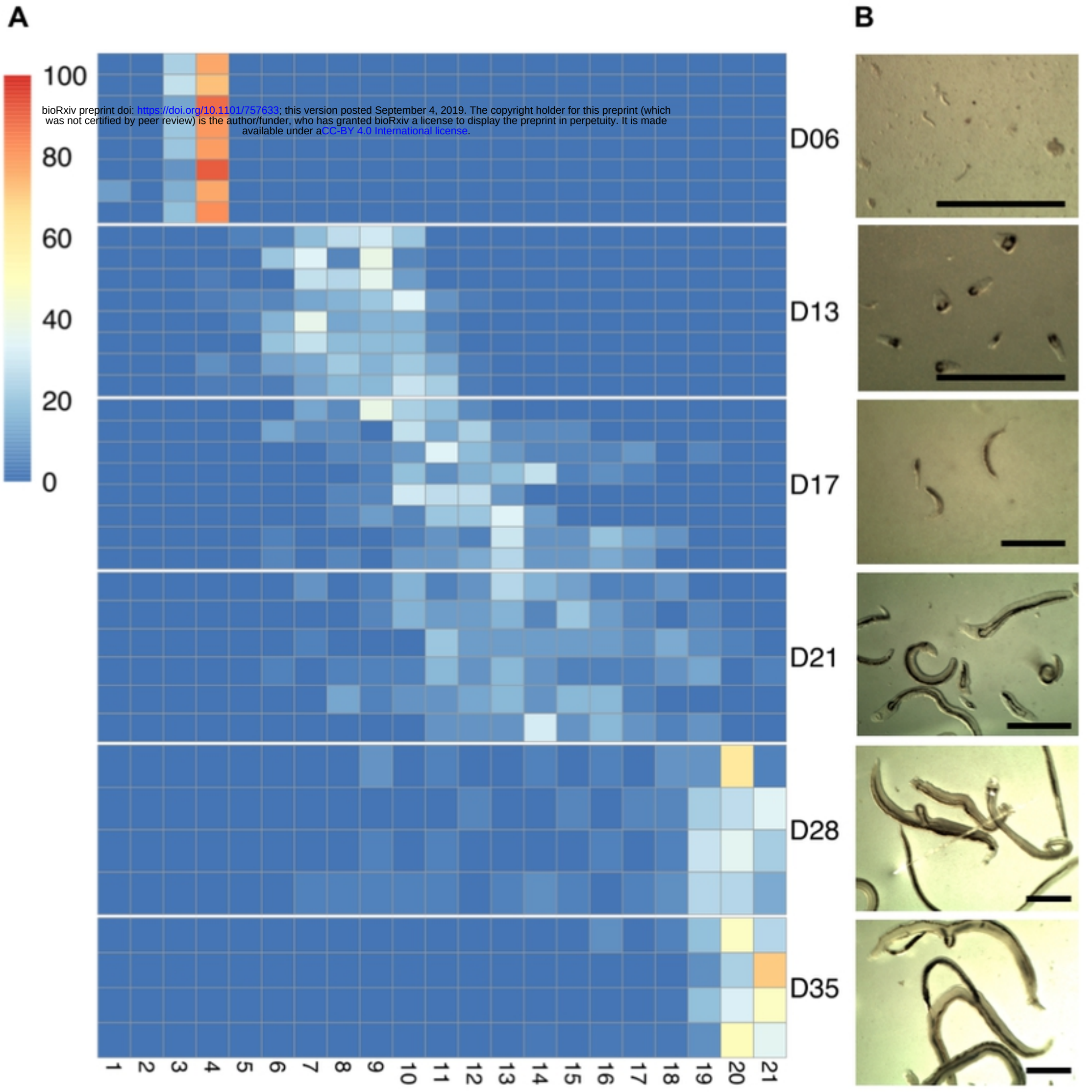
844 D35_SM_1_17675_4_20.htseq-count.txt

845 D35_SM_2_17675_4_21.htseq-count.txt

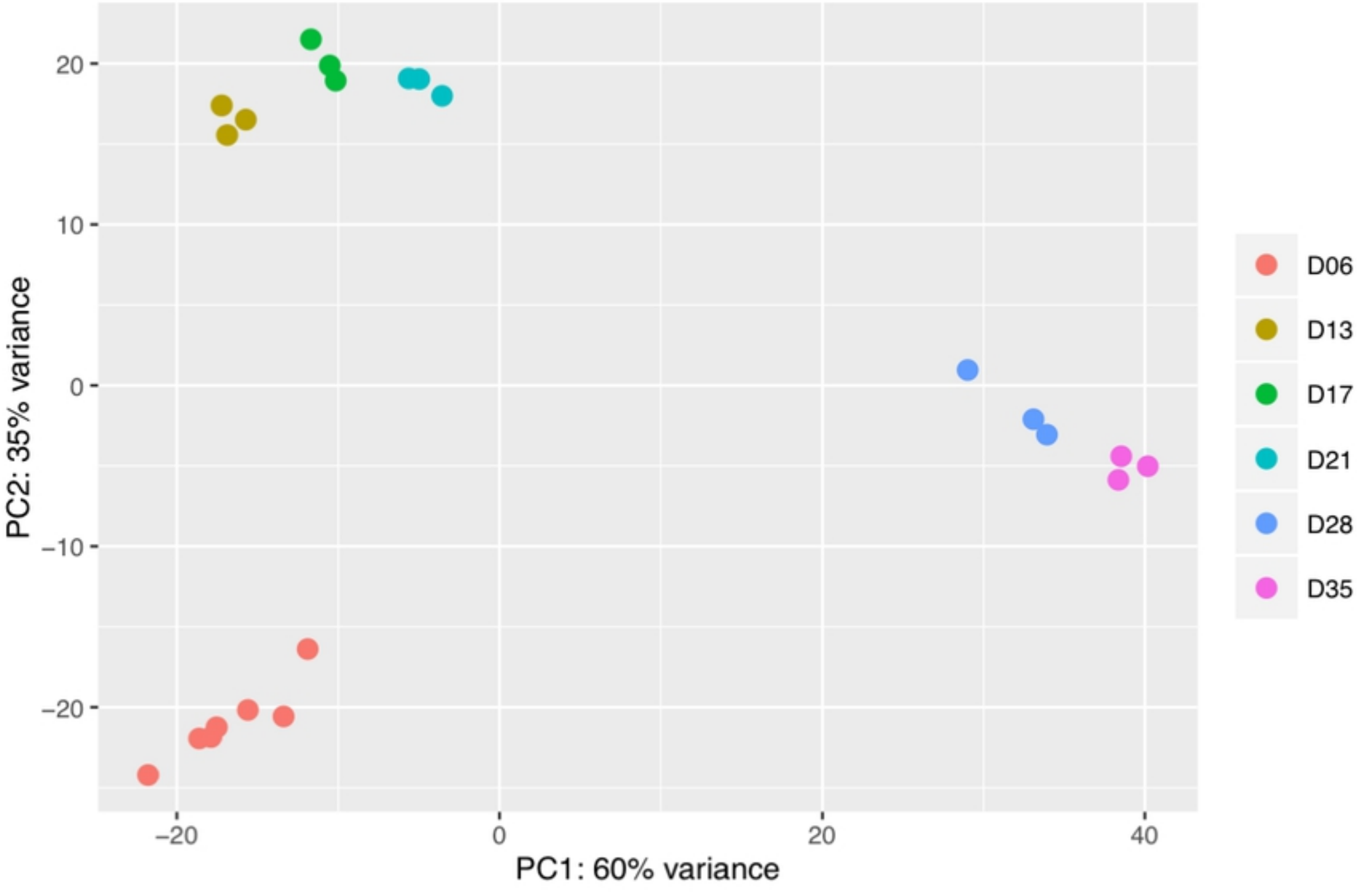
846 D35_SM_3_17675_4_22.htseq-count.txt


Day 0
Day 6
Day 13
Day 17
Day 21
Day 28
Day 35
Number of cercariae
2000
2000
500
500
350
300
8 x
8 x
8 x
6 x
4 x
4 x

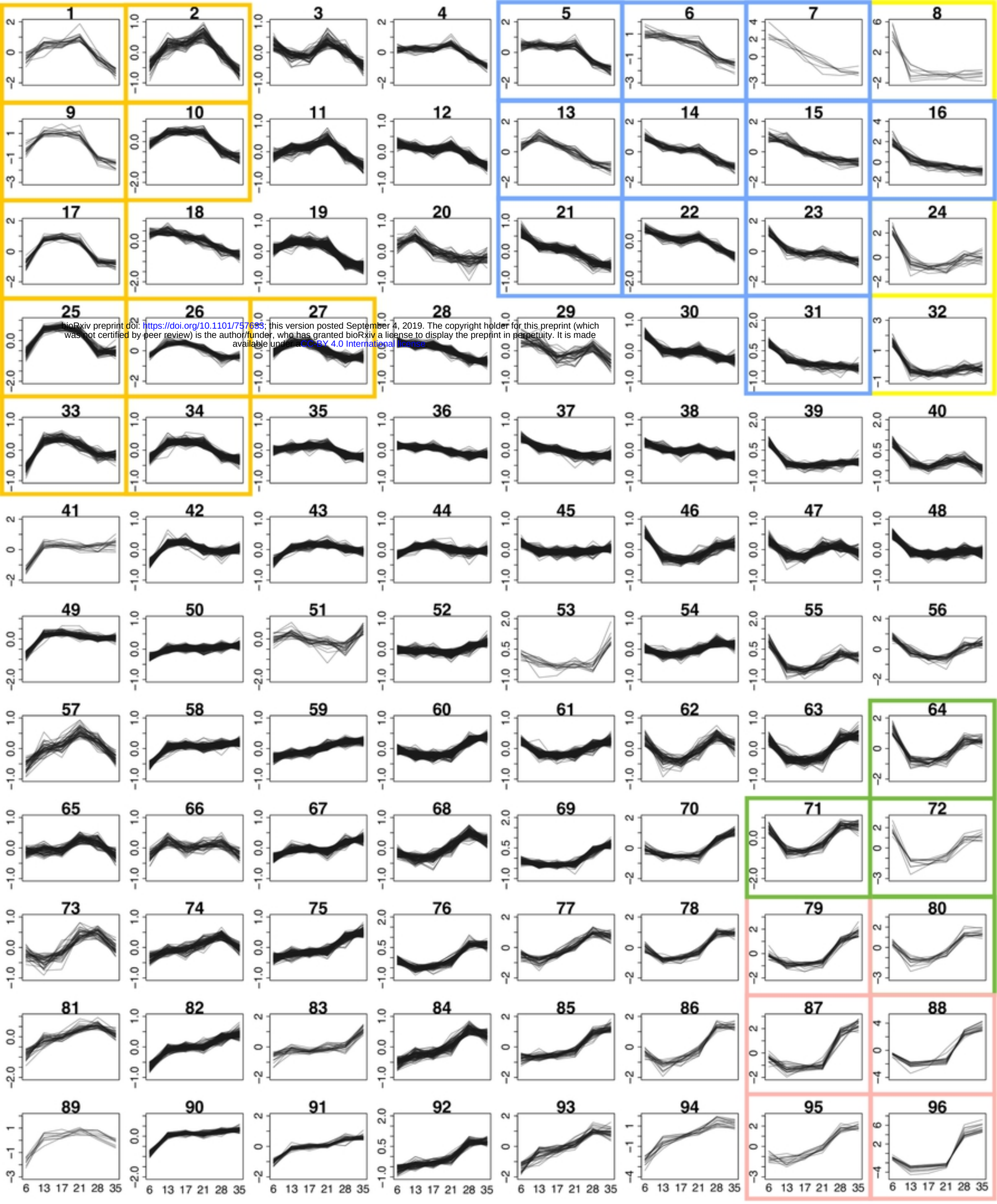
Figure



Figure


20
10
0
−10
−20
PC2: 35% variance
−20
0
20
40
PC1: 60% variance
D06
D13
D17
D21
D28
D35

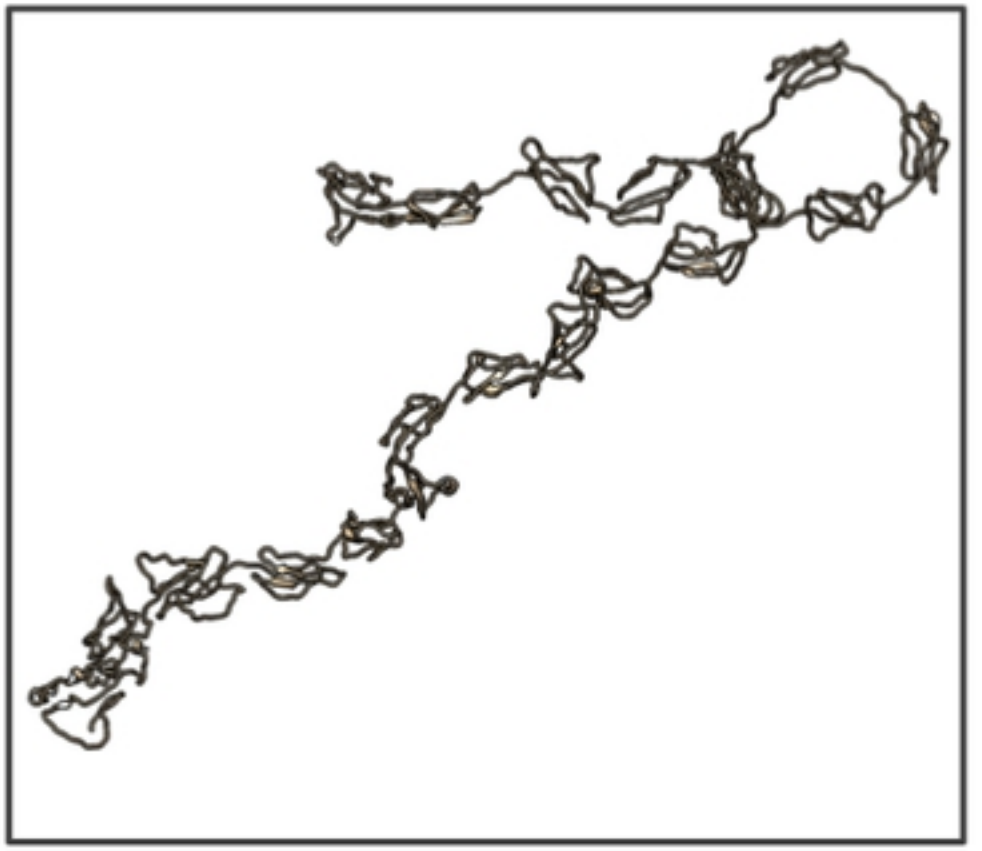
Figure



Figure

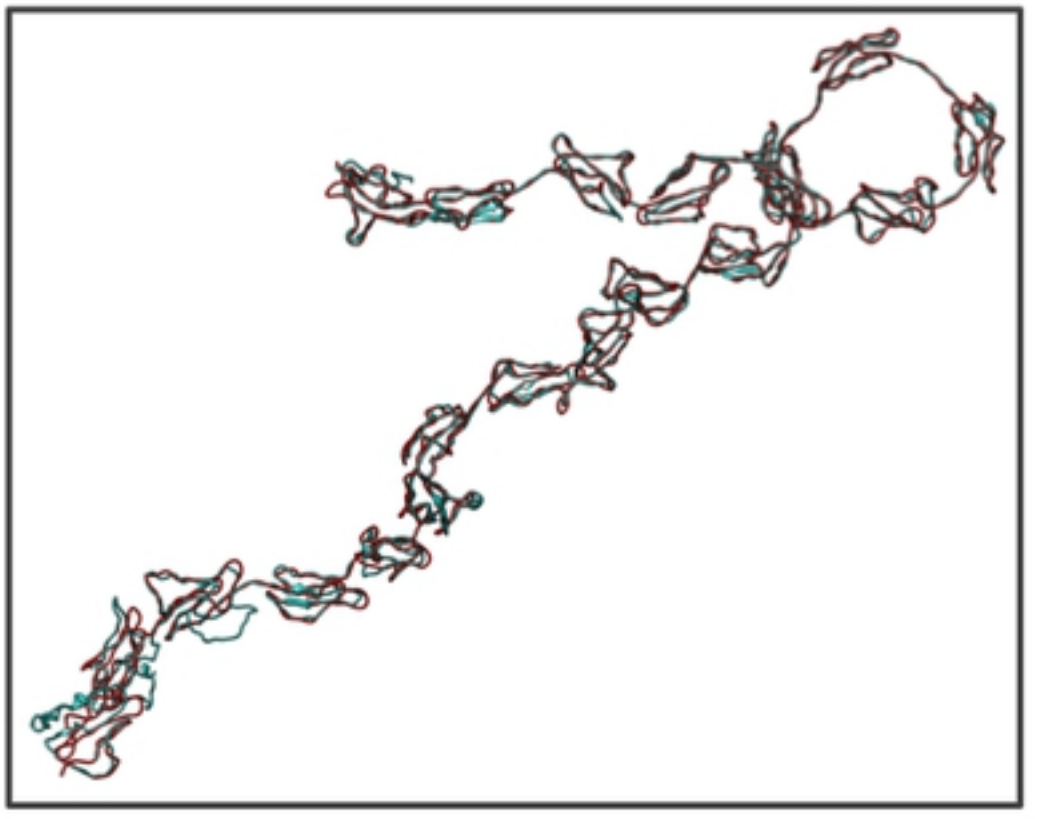
A

Predicted structure of Smp_182770



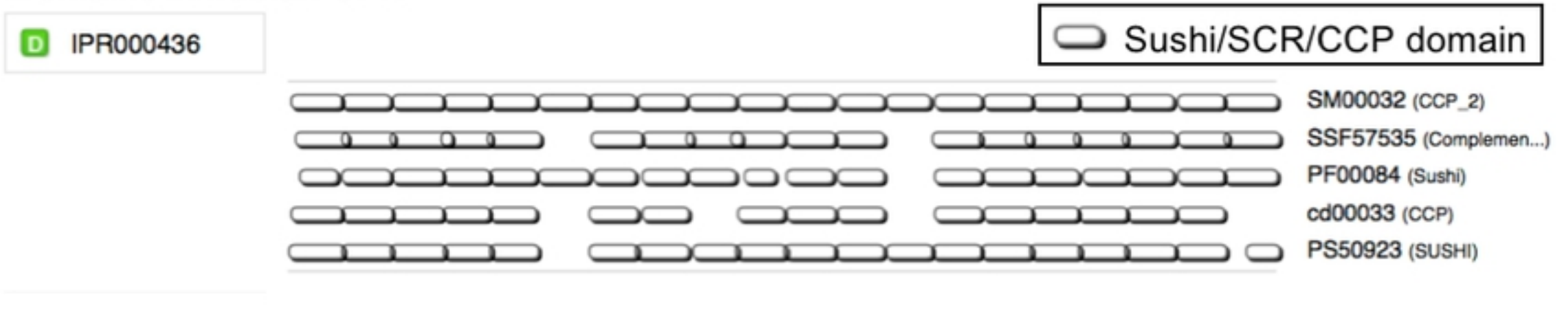
B

Alignment of Smp_182770 predicted structure with PDB structure of human CHF in solution



C

Domains in human CFH



Figure

Figure