A Tail of Two Structures: Proteomic and Transcriptomic Analysis of Cercarial Heads and Tails of Schistosome Worms

James Hagerty
Case Western Reserve University https://orcid.org/0000-0001-5507-243X
Hyung Chul Kim
Case Western Reserve University
Emmitt Jolly (emmitt.jolly@case.edu)
Case Western Reserve University https://orcid.org/0000-0002-7141-2307

Article

Keywords: cercarial heads, Transcriptomic, macrostructure

Posted Date: October 7th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-83189/v1

Version of Record: A version of this preprint was published at Communications Biology on July 12th, 2021. See the published version at https://doi.org/10.1038/s42003-021-02366-w.

A Tail of Two Structures: Proteomic and Transcriptomic analysis of cercarial heads and tails of schistosome worms

James R Hagerty[1], Hyung Chul Kim[1], Emmitt R Jolly[1,2*]

[1]*Case Western Reserve University, Department of Biology, Cleveland, OH, USA 44106*

[2]*Case Western Reserve University, Center for Global Health and Disease, Cleveland, OH, USA 44106*

*To whom correspondences should be addressed: emmitt.jolly@case.edu

Abstract

Schistosomes require both molluscan and mammalian hosts for development. The larval stage, cercaria, exits the intermediate snail host and must swim to identify and invade the definitive mammalian host. The cercaria has two macrostructures, the head and the tail. The tail is lost after host invasion, whereas the head develops into a sexually dimorphic adult worm. Translation in the cercaria differs in each macrostructure with higher levels of translation in cercarial tails and little to no translational activity in cercarial heads. We compared the transcriptome and proteome of the cercarial head and tail and observed stark differences between the two macrostructures. We identified unique and differentially expressed transcripts and proteins, including multiple ribosomal components expressed in higher levels in tails than in heads, which may explain the differences in translation levels between heads and tails and reflect the weak correlation between transcription and translation in infectious cercarial heads and tails.

Introduction

Schistosomes have a complex life-cycle that must quickly adapt both physiologically and morphologically to distinct environments. The schistosome infectious larval stage, cercaria, exits the intermediate molluscan host in search of a definitive mammalian host. After identification and penetration of the mammalian host, the cercaria loses its tail, transforms into a schistosomulum, and enters the host circulatory system. Over several weeks, the schistosomulum develops into an adult worm in the mesentery of the liver, pairs with a mate, and produces hundreds of eggs daily. The eggs are excreted out of the host and hatch into transient miracidia. The miracidia then invade and develop into sporocysts in a freshwater snail. The sporocysts produce and release the infectious cercariae, completing the life cycle. The cercaria is transiently free-living and represents the first interaction point of the parasite life-cycle with the human host.

However, while many studies have been performed on cercaria using several -omic approaches, including microarrays, RNA-seq, and proteomics, our understanding of the mechanisms of transcription and translation control of this stage is still limited [1-8]. Collectively, -omic approaches to date have consistently pointed to a substantial upregulation of genes related to glycolysis and metabolism [9], consistent with the functional activity of actively swimming cercaria in search of a mammalian host. Proteomic analysis of whole cercariae and cercarial secretions released from the cercarial head revealed an abundance of proteases, including elastase, and other proteins that are likely involved in host invasion [4,9-11]. The cercarial tail, however, is primarily involved in motility. Given the abundance of proteases in the translation-limited cercarial head and the relatively translation-enhanced cercarial tail, the combinational analysis of the two structures is potentially clouding our understanding of each macrostructure [12,13].

Previously, we proposed treating the free-swimming cercarial stage as two separate -omic entities due to the apparent differences in translational regulation and biological role of the cercarial macrostructures: head and tail [13]. The cercarial head had significantly less global translation levels than the cercarial tail. Here, we add to these initial observations and have analyzed and compared the transcriptome and proteome of the cercarial head and tail. Analysis of the head and tail as distinct structures has allowed us to elucidate potential mechanisms for regulating the observed translational differences and functional roles of heads and tails. We found that cercarial heads and tails: 1) store distinct populations of proteins and transcripts which correlate to their functional roles as macrostructures, 2) differentially regulate translation using ribosomal component composition, 3) have a weak correlation between transcript and protein abundance, and 4) utilize distinct translational repression regimes. Together these findings demonstrate the necessity for treating cercariae as two distinct macrostructures instead of single units for study.

Methods and Materials

Parasite collection

Biomphalaria glabrata snails infected with S*chistosoma mansoni* (NMRI strain) were obtained from Biomedical Research Institute (BRI; Rockville, MD). Cercariae were shed as previously described [14].

Cercarial head and tail separation

The separation and collection of heads and tails were performed as previously described with minor modifications [13]. Samples for mass spectrometry and RNA-seq analysis were washed with PBS and flash-frozen as pellets without the addition of protease inhibitors.

RNA extraction

RNA was extracted from cercarial heads and cercarial tails using the Direct-zol RNA Miniprep Kit (Zymo Research, Irvine, CA) following their standard protocol, including on-column DNAseI digestion. RNA concentration was assessed using a Nanodrop 8000 spectrophotometer (Thermo Scientific, Waltham, MA). Library preparation and barcoding were performed using the Pico stranded library prep kit (Clontech, Mountainview, CA).

Genomic and transcriptomic data analysis

The *S. mansoni* genome sequence and annotation were downloaded from WormBase ParaSite [15,16]. The most recent version (release 14) was used for the analysis presented here.

Six RNA-Seq datasets were used for this study: three sets of in-house cercarial head dataset composed of ~60 million raw paired-end reads from cercarial heads and three sets of in-house cercarial tail dataset composed of ~80 million raw paired-end reads from cercarial tails. Data collection was performed using Illumina HiSEQ 2500 for both head and tail samples. The datasets were checked for quality using FastQC [17], and the

adapters were trimmed using Trimmomatic [18]. The transcripts were quantified using the mapping-mode of Salmon [19] with the *S. mansoni* genome sequence and the annotation used as the reference transcripts.

To further filter the poorly characterized putative genes and pseudogenes in the *S. mansoni* reference annotation, the standalone version of riboPicker [20] was used on the reference transcript sequence to identify and remove 29 transcripts that match rRNAs. Multiple databases were used for the riboPicker process. SILVA rRNA database was used for large subunit (version 132) and small subunit (version 138) rRNA sequences [21]. Rfam was used for 5S and 5.8S subunit rRNA sequences (release 14.2) [22]. As the SILVA small subunit sequence data file was too large to index using riboPicker, it was broken into 6 similar sized files for indexing, and the large, 5S, 5.8S, and the 6 small subunit indices were used together for the analysis.

Statistical analyses were performed with the R environment 4.02 [v3.4.4: 23] with DESeq2 [24], readr [25], and tximport [26] libraries loaded. The log2 fold change shrinkage was performed with the apeglm package [27].

Protein sample preparation

Cercarial heads and tails were thawed on ice for 30 minutes in 300 µL 2% SDS and protease inhibitor cocktail (Sigma, St. Louis, MO). Sonication was performed on all samples at 50% amplitude with a probe sonicator then vortexed; this procedure was repeated for four cycles with an intervening ice incubation between cycles. SDS detergent removal and alkalylation were performed following the FASP protocol [28].

After homogenization was complete, samples were processed using the FASP protocol and Amicon Ultra MWCO 3K filter (Millipore, Billerica, MA). Samples were reduced and akalylated on filter with 10 mM dithiothreitol (Acros, Fair Lawn, NJ) and 25 mM iodoacetamide (Acros, Fair Lawn, NJ). Samples were then concentrated to a final volume of 40 µL in 8 M urea. The sample concentration was determined using the Bradford assay kit (Bio-Rad, Hecules, CA).

Following sample cleanup using FASP, 10 µg of total protein was aliquoted for digestion. The concentration of urea was reduced to 4 M using 50 mM Tris pH 8. Protein digestion was performed using mass spectrometry grade lysyl endopeptidase (Wako Chemicals, Richmond, VA) using a digestion enzyme to substrate ratio of 1:40 and 2-hour incubation at 37^{o}C. Following lysyl digestion, the urea concentration was further reduced to 2M using 50 mM Tris pH 8, and samples were digested using sequencing grade trypsin (Promega, Madison, WI) at a digestion enzyme to substrate ratio of 1:40 overnight at 37 ^{o}C. Samples were then diluted using 0.1% formic acid (Thermo Scientific, Rockford, IL) for LC-MS/MS analysis.

Reverse phase LC-MS/MS

Sample injections of 11 µL containing 600 ng digested peptide was loaded with blank runs intervening between each sample. The cercarial heads and tails were run in triplicate. The Orbitrap Celos Elite mass spectrometer (Thermo Electron, San Jose, CA) equipped with the Waters nanoACQUITY LC system (Waters, Taunton, MA) was used for acquisition. Peptides were desalted in a trap column (180 µm × 20 mm, packed with C18 Symmetry, 5µm, 100Å, Waters, Taunton, MA) and subsequently resolved in a reversed-phase column (75µm x 250 mm nano column, packed with C18 BEH130, 1.7µm, 130Å (Waters, Taunton, MA). Liquid chromatography was carried out at ambient temperature at a flow rate of 300 nL/min using a gradient mixture of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The gradient employed ranged from 4 to 44% solvent B over 210 min. Peptides eluting from the capillary tip were introduced into the nanospray mode with a capillary voltage of 2.4 kV. A full scan was obtained for eluted peptides in the range of 380–1800 atomic mass units followed by twenty-five data-dependent MS/MS scans. MS/MS spectra were generated by collision-induced dissociation of the peptide ions at a normalized collision energy of 35% to create a series of b- and y-ions as major fragments. A one-hour wash was included between each sample.

Protein identification and label-free quantitation

All identification and quantitation were performed using Peaks X+ (Bioinformatics Solutions Inc., Waterloo, ON, CA) [29-31]. De Novo sequencing from spectra was performed with a parent mass error tolerance of 15.0 ppm and a fragment mass error tolerance of 0.5 Da. Fixed and variable modifications accounted for in sequencing: Fixed- carbamidomethylation 57.02, and variable- deamidation (NQ) 0.98, oxidation (M) 15.99. The max variable PTM per peptide was set at 3. Following de novo sequencing PEAKS database search using Wormbase parasite protein annotation release 14 with cRAP contaminants database. Following PEAKS database search PEAKS PTM search was performed with a de novo score threshold of 15 and a peptide hit threshold of 30.0 (-10logP). Following PEAKS PTM search, the SPIDER Homology match search was performed. Following SPIDER homology search ID- directed label-free quantification was performed using the aforementioned PEAKS, PEAKS PTM, and SPIDER searches. The mass error tolerance was set to 20.0 ppm, and retention time shift tolerance was set to 20.0 mins. The false discovery rate (FDR) threshold was set to 0.05. The threshold utilized for significant differential expression of proteins was a -10logP significance of ≥ 13.0 and a Head/Tail (H/T) ratio ≥ 1.6 and ≤ 0.62 for total reporting. All proteins specifically discussed require a more stringent H/T ratio of ≥ 2.0 and ≤ 0.50.

mRNA and protein abundance

DESEQ2 normalized counts were used for correlation mapping and translational repression analysis. PEAKS identifications were analyzed using Scaffold PerSPECtive (version 3.1.0, Proteome Software Inc., Portland, OR) to determine normalized weighted spectral counts. Normalized weighted spectral counts were utilized for correlation mapping.

Gene ontology analysis

Gene ontology (GO) analysis was performed using gProfiler with Wormbase parasite *S. mansoni* GO annotations [32]. The significance threshold for over- and underrepresentation was set at 0.05 and used the g:SCS algorithm for analysis [32]. All significance is reported as –log10 P-value. The gProfiler tool was utilized for both protein and RNA GO analysis.

Correlation and Euler Mapping

Global mRNA and protein abundance were plotted as log10 (normalized count +1) and log10 (normalized weighted spectral count + 1). The comparison of the proteome to the transcriptome utilized linear r^2 correlation, as well as Spearman ranked correlation. The log10 counts were plotted against each other as a direct correlation scatter plot.

Euler diagrams were generated using the Biovenn web tool to give proportional overlapping groups for the two group diagrams[33]. The identified transcripts with an average normalized count from DESEQ2 of ≥ 1.0 and proteins with an FDR ≤ 0.01 are compared. All four group diagrams do not show the proportional sizing of datasets.

Translational repression analysis

In silico translational repression analysis was performed as reported by Lindner et al. [34] with modification. Only undetected protein products were considered for repression analysis to assure higher confidence in putative repressed genes. Consequence tryptic peptide analysis with zero missed cleavages determined the potential for the detection of digested peptides [35].

Statistical Analysis

Statistical tests used in this study were carried out using DEseq2 (RNA-seq), PEAKSX, Scaffold PerSPECtive, and CONSeQuence algorithms (proteomics) as described above. All omic analysis was performed in triplicate with three biological replicates. All p-values and adjusted p-values are included in supplementary data 1. GO analysis was performed using gProfiler and the g:SCS algorithm. The comparison of normalized transcript (DEseq2 normalized counts) and protein abundance (Scaffold PerSPECtive normalized weighted spectral counts) was performed using linear regression and spearman ranked correlation analysis.

Results

Cercarial heads and tails are transcriptomically and proteomically distinct

The functional and developmental roles for cercarial heads and tails are distinct. The cercarial tail gives motility and assists in host penetration; however, it is discarded and does not progress through development in the definitive host. In contrast, the cercarial head develops into the adult worm and stores all the necessary transcripts and proteins necessary for initial entry and adaptation to the definitive host. Since cercariae do not undergo transcription [36]**,** and translation in cercarial heads and tails is differentially regulated [13], we explored conservation of steady-state transcript populations between cercarial heads and tails.

We found a total of 11,663 transcripts in cercariae (Figure 1A). Cercarial heads and tails share 5,839 transcripts. Among these shared genes are four of the most abundant transcripts in cercariae: calmodulin-4 (*Smp*_032990), calcium-binding protein (*Smp*_033000), and two uncharacterized genes *Smp*_195070 and *Smp*_003770 (Supplementary Data 1) [37,38]. Cercarial heads have 211 unique transcripts, and cercarial tails have a large store of 5,613 unique transcripts (Figure 1A).

We next investigated whether the macrostructure specific transcripts identified in the cercaria correlate with the cercarial proteins in these substructures. Combining these RNA-seq data (Figure 1A) with mass spectrometry, we identified 3,108 cercarial proteins (Figure 1B). Cercarial heads and tails share 2,507 of

192 the 3,108 identified proteins. Four hundred (400) proteins are unique to cercarial heads, and 201 are unique
193 to tails. However, while tails have half the number of unique proteins, they have a higher number of overall
194 transcripts. Correspondingly, cercarial tails store a broader diversity of transcripts while cercarial heads store
195 a broader diversity of proteins.

196 Next, we used differential expression analysis to look for variation in the transcriptomes and
197 proteomes of heads and tails. In total, 4,510 transcripts were differentially regulated (log2FoldChange ≥ 2.0
198 and p-value ≤ 0.01) out of 11,663 transcripts identified (Figure 2A). Of the 4,510 differentially regulated
199 transcripts, 4,345 were upregulated in cercarial tails, and 165 were upregulated in cercarial heads (Figure 2A,
200 Supplementary Data 1). To quantify the differential expression of proteins in cercarial heads and tails, we
201 utilized label-free mass spectrometry, identifying a total of 530 differentially expressed proteins (Figure 2B).
202 The distribution of differential protein expression between cercarial heads and tails is relatively even. Cercarial
203 heads have 279 upregulated proteins, and cercarial tails contain 251 upregulated proteins (Figure 2B).

204 **Cercarial heads store ribosomal transcripts while cercarial tails store ribosomal proteins**

205 Given the different functional roles of cercarial heads (host invasion), and cercarial tails (motility, and
206 metabolism), we predicted a varied representation in both unique and differentially expressed transcripts and
207 proteins in the macrostructures. We performed a gene ontology (GO) enrichment analysis to identify over-
208 and under-representation (over/und-rep) of transcripts and proteins in 3 categories: biological process (BP),
209 molecular function (MF), and cellular compartment (CC). The categorized over/under-rep products are
210 separated into two classes: those that are unique to heads or tails (Supplementary Figure 1 and Data 2) and
211 those that are differentially expressed in heads and tails (Figures 3 and 4, Supplementary Data 3). Unique
212 transcripts and proteins were analyzed separately from differentially expressed genes because these
213 transcripts are removed during the filtering of low abundance transcripts and retention matching of protein

spectra. Cercarial tails had a broad array of unique over/under-rep transcript and protein types (Supplementary Figure 1 and Data 2), including an abundance of transcripts related to methyltransferase activity (DIMT1 *Smp*_077020, N6AMT1 *Smp*_024800) and Ras guanyl-nucleotide exchange factor activity (ARHGEF17 *Smp*_167650, ECT2 *Smp*_153520) (Supplementary Figure 1 and Data 2). Unique under-rep transcripts included those the encoded for translation and mRNA interaction (Supplementary Figure 1 and Data 2). For cercarial heads, we did not identify any unique over/under-rep transcript classes or any unique under-rep protein groups as defined by gene ontology. However, unique over-rep cercarial head protein groups identified were associated with mRNA regulation such as POLR2J *Smp*_040710, ENY2 *Smp*_039640, and SF3A3 *Smp*_003630) (Supplementary Figure 1 and Data 2). We found multiple overrepresented GO categories of unique proteins in cercarial tails, including translational and mitochondrial processes (RPL36 *Smp*_035790, SARS *Smp*_057230, and MRPL47 *Smp*_102280) (Supplementary Figure 1 and Data 2). The overrepresentation of translation and ribosome related proteins is of interest, given the underrepresentation of these genes at the transcript level (Supplementary Figure 1 and Data 2). We identified no underrepresented protein GO enrichment classes in unique tail proteins. Notably, heads and tails show little overlap in transcript or protein GO function, with the exception of transcripts related to energy metabolism supporting the observed biology that has the tail functioning as transient motility while the head progresses through development.

Using GO over-rep analysis, we compared transcripts and proteins within each macrostructure to determine if transcripts identified in each macrostructure corresponded with the protein populations in the same structure (Figure 3, Supplementary Data 3). Head transcripts were enriched for several processes, including protein production, ribosomal biogenesis, and chromatin (RPL27 *Smp*_063350, RPS23 *Smp*_074470, MRPS9 *Smp*_333040, and H3F3C *Smp*_082240) (Fig. 3A, Supplementary Data 3) and tail transcripts were enriched for non-coding RNA processing and chromatin marking genes (METTL1 *Smp*_130610, MARS *Smp*_040770, NAT10 *Smp*_144490, and DNMT2 *Smp*_334230) (Figure 3B, Supplementary Data 3). At the

protein level, cercarial heads were enriched for structural proteins and proteases (SAT1 *Smp*_090120, paramyosin *Smp*_021920, cercarial elastases *Smp*_330280 and *Smp*_330290, and PRCP *Smp*_002600) (Figure 3C, Supplementary Data 3), whereas cercarial tails were significantly enriched for translational, ribosomal, stress response, and mitochondrial proteins (Figure 3D, Supplementary Data 3). These include a large population of ribosomal components (RPL11/12 *Smp*_012750, RPL24 *Smp*_001830, RPL26 *Smp*_098960, RPL27a *Smp*_325920, RPS9 *Smp*_180000, RPS13 *Smp*_096750, RPS15 *Smp*_307550, and RPS19 *Smp*_174950) [39-50]. All of the previously described ribosomal component genes are involved ribosomal biogenesis or translational upregulation [39-50].

Cercarial heads have few underrepresented transcript or protein GO classes, lacking only membrane component transcripts (Figure 4A and C, Supplementary Data 3). In cercarial tails, GO analysis highlighted significant underrepresentation of transcripts involved in translational control and peptide biosynthesis. (Figure 4B, Supplementary Data 3). At the protein level, tails are underrepresented for proteins related to nucleic acid production (Figure 4D, Supplementary Data 3) but enriched for ribosomal and translational genes, although transcripts are significantly lacking in these same categories. Overall, the GO analysis data show that transcripts classes in heads or tails unexpectedly do not predict the protein populations.

Transcription and translation are not strongly correlated in cercarial heads and tails

We further explore the observation that transcripts in tails or heads do not predict protein classes by testing whether the available transcripts and observed proteins aligned in these populations. We compared transcripts and proteins in heads and tails using Venn diagrams. When we looked for unique transcripts and proteins, we found that 148 head proteins did not have corresponding transcripts, and 2 head transcripts did not have corresponding proteins (Figure 5A). For unique tail proteins and transcripts, we found that tails have 148 unique transcripts for which we were surprisingly able to identify 0 corresponding proteins (Fig 5A). On

the other hand, we found that differentially expressed genes have a somewhat higher correlation between transcript and protein levels than observed for unique genes (Figure 5). Cercarial tails have the most robust relationship between differentially expressed transcripts and proteins (Figure 5B). We identified 91 transcripts that were both overexpressed at the transcript and protein levels in tails. We only identified 12 transcripts in cercarial heads that were overexpressed at the transcript and protein level. Together these findings suggest a weak correlation between protein and transcript levels in cercariae.

To assess the relationship between these two processes more quantitatively, we performed a correlation analysis on the transcriptome and proteome of the cercarial head and tail. We were able to quantify and match 2,884 proteins with respective transcripts from our RNA-seq datasets. We then compared 2,884 matched and quantified transcripts and proteins by their respective normalized abundance measures (Supplementary Data 1). Our analysis shows a limited correlation in cercarial head transcripts and proteins with an R^2= 0.0768 and a Spearman correlation of 0.2813 (Figure 5C). Cercarial tail transcripts and proteins are more highly correlated with an R^2=0.1783 and a Spearman correlation of 0.3912 (Figure 5D). We have previously shown that there are global differences in translational control, and expect more specific regulation that represses genes in heads and tails [13]. Given our previous findings, the relationship between unique proteins and transcripts (Fig. 5A), and the lack of correlation between transcript and protein levels (Fig. 5C-D), it is likely that the mismatch is in part due to the repression of large subsets of transcripts within heads and tails. To further explore these observations, we then analyzed the proteome and transcriptome for putatively repressed genes.

Cercarial heads and tails utilize two different translational repression mechanisms

We next analyzed high abundance transcripts for putative translational repression and found two unique populations of repressed genes in heads and tails as well as a large shared population of transcripts.

Using the top decile of transcripts, we identified 575 translationally repressed transcripts (Figure 6, Supplementary Data 4) between heads and tails. Cercarial heads contain 116 unique repressed transcripts, and cercarial tails contain 291 unique repressed transcripts. Cercarial heads and tails share 84 putatively repressed transcripts (Figure 6, Supplementary Data 4). The shared population of putatively repressed transcripts contained genes related to protein stability and translation, including RANBP9 (*Smp*_161080), ATG2A (*Smp*_148650), NSA1 (*Smp*_054120), SYAP1 (*Smp*_085530), CCNG1 (*Smp*_129360) (Supplementary Data 4). We also identified antiviral and innate immunity-related genes TRIM56 (*Smp*_098550) and AKIRIN2 (*Smp*_076040) (Supplementary Data 4) [51,52]. HSP40 (*Smp*_136540), a known regulator of HSP70 (*Smp*_302170), is putatively repressed in both cercarial heads and tails (Supplementary Data 4) [53]. Interestingly TRPA (*Smp*_342190) is putatively repressed though capsaicin, a TRPA antagonist, is effective in the cercarial stage (Supplementary Data 4) [13,54].

Cercarial heads and tails have large distinct populations of repressed genes as well. Cercarial heads are repressed in multiple ribosomal genes (RPL10 *Smp*_013200, RPL17 *Smp*_024850, RPS24 *Smp*_091640, and NIP7 *Smp*_021760) (Supplementary Data 4). Cercarial tails have multiple genes related to protein degradation repressed (UBE2D1 *Smp*_315750, CUL4B *Smp*_082850, WWP1 *Smp*_123180) (Supplementary Data 4). The findings in heads and tails align with the low levels of translation in heads and the need to maintain proteins during the transient lifespan of the cercarial tail.

Discussion

Previous -omic analysis of whole infectious cercariae showed upregulation of metabolic genes involved in glycolysis, including NADH dehydrogenases, and proteases involved in host invasion, including cercarial elastases and serine proteases. [1-8]. This work supports these previous findings while also elucidating differences in translational control, the correlation of transcription and translation, and identifying specific

repressed targets in heads and tails [13]. Cercarial heads and tails vary significantly in overall translation rates, with heads having limited translation when compared to tails. [13]. This work expands on these previous observations. First, we show using multiple -omics approaches that transcripts and proteins found in the two macrostructures are meaningfully different. Second, we offer insight into the regulation of protein production by comparison of the cercarial head and cercarial tails transcript and protein groups, respectively. Third, we show that transcript levels and protein levels do not have a meaningful correlation in the cercarial life stage. Finally, we identify putatively repressed genes in cercarial heads and tails, thus clearly defining two distinct repressive patterns in heads and tails.

Cercarial heads and tails differ in structure and in function. The cercarial tail is responsible for motility, assists with host invasion mechanically, and is lost upon infection of the host. The cercarial head attaches to the host, produces enzymes to facilitate host invasion, and progresses through development to an adult worm after host infection. We used RNA-seq and label-free mass spectrometry analysis to explore the molecular and regulatory differences in these structures. We identified populations of proteins and transcripts that are unique to the cercarial heads and tails (Figure 1). The cercarial tail contains 5,613 unique transcripts, and the cercarial head has 211 (Figure 1A). The unique proteins show an inverse pattern of storage compared to unique transcripts with the cercarial head containing 400 unique proteins, while the cercarial tail has 201 unique proteins (Figure 1B). We see a clear difference in the storage of unique protein and transcripts, which is somewhat unexpected given the increased rate of translation in cercarial tails compared to cercarial heads. Using differential expression analysis, we identified 165 upregulated transcripts in cercarial heads and 4,345 upregulated transcripts in cercarial tails (Figure 2A, Supplementary Data 1). The role of this large population of stored transcripts in cercarial tails is unclear. Increasing evidence indicates functional roles for mRNA products outside of their coding potential and UTR regulatory regions. For example, P53 mRNA has been shown to have an auto-regulatory function on its own translation, and HIST1C mRNA has been shown to negatively regulate

telomere length [55, 56]. The proteomic differential expression analysis reveals upregulation 279 and 251 proteins for cercarial heads and tails, respectively (Figure 2B, Supplementary Data 1). The abundance of unique and differentially expressed genes across cercarial heads and tails support our previous assertion that the head and tail need to be treated as two distinct functional macrostructures. We were also able to affirm this with GO enrichment that identified previously unreported gene groups at both the transcriptome and proteome levels.

Cercarial heads have an abundance of ribosomal components and translation-related genes (Figure 3A, Supplementary Data 3) while being globally translationally repressed. These transcripts are likely stored for later translation after host invasion. Cercarial head proteins are enriched for mRNA processing, proteolysis, and cytoskeleton related genes (Figure 3C, Supplementary Figure 1, Data 2, and Data 3). The mRNA related genes in cercarial heads include POLR2J, ENY2, and SF3A3. The storage of these proteins mRNA related genes is interesting, given the stalled state of RNA production in cercariae [36,55-57]. The mRNA processing genes may be stored for the early burst of transcription, which begins shortly after invading the definitive host [36]. Moreover, unique cercarial tail transcripts are underrepresented for genes related to DNA synthesis. The two results together support the findings of Roquis *et al.* 2015, showing that cercariae does not produce new nucleic acids but is primed for a massive burst of transcription after host invasion [36]. The detection of proteolytic proteins in cercarial heads (Fig. 3C, Supplementary Data 3) matches with previous analysis showing high levels of these enzymes stored in the cercarial glands [12]. Cercarial tails are enriched for methyltransferase and Ras-GEF genes at the transcript level (Figure 3A, Supplementary Figure 1, Data 2, and Data 3). ARHGEF17 is unique to tails and plays a role in stabilizing actin during oxidative stress [56]. Cercarial tails exhibit higher levels of respiration and likely need stabilization of proteins to support motility. Furthermore, cercarial tails have an overrepresentation of expressed proteins related to translation and mitochondrial maintenance (Figure 3D, Supplementary Figure 1, Data 2, and Data 3). Together we see that the

identification of metabolic genes from tails and proteolytic genes from heads using whole cercarial analysis only gives a shallow look at the functions of either structure [9]. The separation and individual analysis of heads and tails give insight into mechanisms of translational control and development.

Previously, we demonstrated that cercariae have significantly different translational regulation for the heads and tails [13]. Our analysis of the transcriptome and proteome suggests likely mechanisms of translational regulation in heads and tails. Unique cercarial tail transcripts include two notable methyltransferases, DIMT1 and N6AMT1, both of which have been shown to play a role in translational control. DIMT1 is involved in ribosomal biogenesis, and N6AMT1 can produce 5'UTR modifications that act as cap-independent ribosomal binding sites for genes like HSP70 [56, 57]. DIMT1 and N6AMT1 can thus selectively increase cap-independent translation of important genes in cercarial tails. We also observed a large population of proteins related to ribosomal biogenesis, translational initiation, and ribosomal maintenance in cercarial tails (Figure 3D, Supplementary Data 3). These include RPL11/12, RPL24, RPL26, RPL27a, RPS9, RPS13, RPS15, and RPS19 [39-50]. The enrichment of ribosomal proteins is striking when compared to underrepresented transcripts in cercarial tails, and they show underrepresentation of ribosomal transcripts and transcripts related to translation (Figure 4B, Supplementary Data 3). Cercarial heads, on the other hand, show enrichment for ribosomal transcripts (Figure 3A, Supplementary Data 3). Together these results suggest that cercariae have been primed with ribosomal proteins in the tails and have stored dormant transcripts for ribosomal components in the heads. This lack of multiple important ribosomal components could lead to global levels of translational repression [42,45,49]. The analysis of stored transcripts is of note, but given the lack of correlation between our enriched transcripts and proteins, it was becoming clear we needed to understand how predictive RNA abundance is in determining protein abundance.

We then explored the predictive quality of transcript on protein levels using a matched and normalized correlation plot, as well as Venn diagrams showing overlapping areas of unique and overexpressed genes

(Figure 5). We found an apparent lack of overlap between proteins and transcripts via the Venn diagram mapping of both unique and differentially expressed transcripts in heads (2.2%) total and tails (41.5%) total (Figure 5 A-B). We next analyzed the overall correlation of transcript and protein abundance using normalized counts and normalized weighted spectral counts, respectively. We found a weak positive correlation R^2=0.0768 in cercarial heads as well as cercarial tails R^2= 0.1783 (Fig. 5C-D) between transcript and protein abundance. These values represent a weak correlation compared to multiple studies performed on human tissues [58-60]. An analysis of 29 tissues in humans showed most proteome and transcriptomes have a ~0.50 R^2 correlation [60]. The Spearman ranked correlation, 0.2813 in cercarial heads and 0.3912 in cercarial tails, reveals some predictive power though the relationships are well below even the direct correlation values in human tissues [60]. Therefore, the predictive power of transcripts is low in cercariae. Given the pattern of low correlation in other systems and tissues, it likely that this correlation is low in other developmental stages as well. The unique relationship between undetected transcripts and detectable proteins directed us to analyze the transcriptome and proteome for putatively repressed genes that are defined as high abundance mRNAs with no detected protein.

Using a similar approach to that performed by Lindner et al. 2019 in *Plasmodium* [34], we were able to detect a large population of repressed transcripts in both cercarial heads and tails. A total of 575 likely repressed transcripts with detectable tryptic peptide products were found cercarial heads have 116, cercarial tails have 291, and 84 are shared (Figure 6, Supplementary Data 4). The unique populations our analysis found in cercarial heads and tails suggest two independent repression mechanisms in heads and tails. Repression shared across both heads and tails includes multiple genes known to facilitate protein degradation and also translation, including RANBP9, ATG2A, NSA1, SYAP1, and CCNG1 (Supplementary Data 4). This mixture of repressed production and enhanced stabilization provides an environment that allows protein storage while limiting energy expenditure. Genes related to stress response and innate immunity were also repressed.

TRIM56 and AKIRIN2 are both putatively repressed. In other systems, TRIM56 and AKIRIN2 [51,52] are involved in innate immunity and stress response. Both TRIM56 and AKIRIN2 have been shown to accelerate cell migration and wound healing [61,62]. Repression of these two factors in cercarial heads could be stabilizing the gland structures while maintaining a high abundance of the transcript to allow healing after tail loss and gland secretion. HSP40, a co-factor and modulator of HSP70, is also putatively repressed in both heads and tails. HSP40 can function as a chaperone and also interact with HSP70 to activate ATPase activity and transition HSP70 from the ATP to ADP state [63]. The complex of HSP70-HSP40 can also interact with the ribosomal-associated complex and assist translation and link protein folding to translation directly. We have shown a role for HSP70 in signaling homing behavior in cercariae, but the role of HSP40 is less clear [53]. Two potential roles for HSP40 repression would be the reduction of the HSP70-HSP40 complexing to allow HSP70 to perform its role in host detection and homing or HSP70 ATPase activity needs to remain inhibited to maintain its pre-homing state. TRPA1 is also putatively repressed throughout the cercariae, which is surprising given the effects on viability and motility to cercariae of both capsicum and non-capsicum pepper extracts [54,64].

We have shown that cercarial heads and tails are unique macrostructures. They serve different functions and thus prepare in different ways for those roles. Our finding suggests that the level of ribosomal proteins and ribosomal component mRNAs may be the mechanism for how translation is regulated differently between heads and tails. The storage of essential proteins allows cercarial heads to preserve energy prior to infection of the host while remaining primed to initiate the developmental changes needed after host invasion [65-67]. The transcriptome and proteome are not strongly correlated in cercariae. The lack of correlation could stem from posttranscriptional regulation that leads to massively different translational rates that vary by five orders of magnitude [68]. Finally, we identify a large population of putatively translationally repressed

genes that allow us to distinguish two separate and overlapping mechanisms of translational control being utilized in cercarial heads and tails.

Our findings have led to new emerging questions. What is the role of the large populations of stored transcripts in tails that show no evidence of being translated in this transient structure that doesn't progress through development? What unique patterns and regulatory elements can be utilized for tools in schistosomes? Are these patterns of unique expression different across cell populations not only the macrostructures? We hope to further explore the regulatory mechanisms of translation in cercariae and follow that control throughout the life-cycle next exploring the translational rates and nascent proteins in early schistosomula. The targets of nascent translation in schistosomula can give insight into early founder proteins and minimally required translation for maintenance of the organism. We also intend to explore the repressed targets for meaningful patterns that underpin the specific repression across heads and tails and potentially . Finally we could potentially explore structural differences at the cell type and tissue levels.

References

1 DeMarco, R. *et al.* Protein variation in blood-dwelling schistosome worms generated by differential splicing of micro-exon gene transcripts. *Genome Res* **20**, 1112-1121, doi:10.1101/gr.100099.109 (2010).

2 Jolly, E. R. *et al.* Gene expression patterns during adaptation of a helminth parasite to different environmental niches. *Genome Biol* **8**, R65, doi:gb-2007-8-4-r65 [pii]

10.1186/gb-2007-8-4-r65 (2007).

3 Jones, M. K., Gobert, G. N., Zhang, L., Sunderland, P. & McManus, D. P. The cytoskeleton and motor proteins of human schistosomes and their roles in surface maintenance and host-parasite interactions. *BioEssays : news and reviews in molecular, cellular and developmental biology* **26**, 752-765, doi:10.1002/bies.20058 (2004).

4 Liu, M. *et al.* Proteomic Analysis on Cercariae and Schistosomula in Reference to Potential Proteases Involved in Host Invasion of Schistosoma japonicum Larvae. *Journal of proteome research* **14**, 4623-4634, doi:10.1021/acs.jproteome.5b00465 (2015).

5 Parker-Manuel, S. J., Ivens, A. C., Dillon, G. P. & Wilson, R. A. Gene expression patterns in larval Schistosoma mansoni associated with infection of the mammalian host. *PLoS Negl Trop Dis* **5**, e1274, doi:10.1371/journal.pntd.0001274 (2011).

6 Picard, M. A. *et al.* Sex-Biased Transcriptome of Schistosoma mansoni: Host-Parasite Interaction, Genetic Determinants and Epigenetic Regulators Are Associated with Sexual Differentiation. *PLoS Negl Trop Dis* **10**, e0004930, doi:10.1371/journal.pntd.0004930 (2016).

7 Protasio, A. V. *et al.* A systematically improved high quality genome and transcriptome of the human blood fluke Schistosoma mansoni. *PLoS Negl Trop Dis* **6**, e1455, doi:10.1371/journal.pntd.0001455 (2012).

8 Verjovski-Almeida, S. *et al.* Transcriptome analysis of the acoelomate human parasite Schistosoma mansoni. *Nat Genet* **35**, 148-157, doi:10.1038/ng1237

ng1237 [pii] (2003).

9 Curwen, R. S., Ashton, P. D., Johnston, D. A. & Wilson, R. A. The Schistosoma mansoni soluble proteome: a comparison across four life-cycle stages. *Mol Biochem Parasitol* **138**, 57-66, doi:10.1016/j.molbiopara.2004.06.016 (2004).

10 Cass, C. L. *et al.* Proteomic analysis of Schistosoma mansoni egg secretions. *Mol Biochem Parasitol* **155**, 84-93, doi:10.1016/j.molbiopara.2007.06.002 (2007).

11 Jang-Lee, J. *et al.* Glycomics analysis of Schistosoma mansoni egg and cercarial secretions. *Molecular & cellular proteomics : MCP* **6**, 1485-1499, doi:10.1074/mcp.M700004-MCP200 (2007).

12 Knudsen, G. M., Medzihradszky, K. F., Lim, K. C., Hansell, E. & McKerrow, J. H. Proteomic analysis of Schistosoma mansoni cercarial secretions. *Molecular & cellular proteomics : MCP* **4**, 1862-1875, doi:10.1074/mcp.M500097-MCP200 (2005).

13 Hagerty, J. R. & Jolly, E. R. Heads or tails? Differential translational regulation in cercarial heads and tails of schistosome worms. *PloS one* **14**, e0224358, doi:10.1371/journal.pone.0224358 (2019).

14 Milligan, J. N. & Jolly, E. R. Cercarial transformation and in vitro cultivation of Schistosoma mansoni schistosomules. *J Vis Exp*, doi:10.3791/3191

3191 [pii] (2011).

15 Howe, K. L. *et al.* WormBase 2016: expanding to enable helminth genomic research. *Nucleic acids research* **44**, D774-780, doi:10.1093/nar/gkv1217 (2016).

16 Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P. & Berriman, M. WormBase ParaSite - a comprehensive resource for helminth genomics. *Mol Biochem Parasitol* **215**, 2-10, doi:10.1016/j.molbiopara.2016.11.005 (2017).

17 Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).

18 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

19 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419, doi:10.1038/nmeth.4197 (2017).

20 Schmieder, R., Lim, Y. W. & Edwards, R. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28**, 433-435, doi:10.1093/bioinformatics/btr669 (2012).

21 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* **41**, D590-596, doi:10.1093/nar/gks1219 (2013).

22 Kalvari, I. *et al.* Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic acids research* **46**, D335-D342, doi:10.1093/nar/gkx1038 (2018).

23 R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2019).

24 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

25 Wickham, H., Hester, J. & Francois, R. readr: Read Rectangular Text Data. (2017).

26 Soneson, C., Love, M. & Robinson, M. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; peer review: 2 approved]. *F1000Research* **4**, doi:10.12688/f1000research.7563.2 (2016).

27 Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084-2092, doi:10.1093/bioinformatics/bty895 (2018).

28 Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362, doi:10.1038/nmeth.1322 (2009).

29 Tran, N. H. *et al.* Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nat Methods* **16**, 63-66, doi:10.1038/s41592-018-0260-3 (2019).

30 Tran, N. H. *et al.* Complete De Novo Assembly of Monoclonal Antibody Sequences. *Scientific reports* **6**, 31730, doi:10.1038/srep31730 (2016).

31 Tran, N. H., Zhang, X., Xin, L., Shan, B. & Li, M. De novo peptide sequencing by deep learning. *Proc Natl Acad Sci U S A* **114**, 8247-8252, doi:10.1073/pnas.1705691114 (2017).

32 Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids research* **47**, W191-W198, doi:10.1093/nar/gkz369 (2019).

33 Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**, 488, doi:10.1186/1471-2164-9-488 (2008).

34 Lindner, S. E. *et al.* Transcriptomics and proteomics reveal two waves of translational repression during the maturation of malaria parasite sporozoites. *Nature communications* **10**, 4964, doi:10.1038/s41467-019-12936-6 (2019).

35 Eyers, C. E. *et al.* CONSeQuence: prediction of reference peptides for absolute quantitative proteomics using consensus machine learning approaches. *Molecular & cellular proteomics : MCP* **10**, M110 003384, doi:10.1074/mcp.M110.003384 (2011).

36 Roquis, D. *et al.* The Epigenome of Schistosoma mansoni Provides Insight about How Cercariae Poise Transcription until Infection. *PLoS Negl Trop Dis* **9**, e0003853, doi:10.1371/journal.pntd.0003853 (2015).

37 Buro, C. *et al.* Transcriptome analyses of inhibitor-treated schistosome females provide evidence for cooperating Src-kinase and TGFbeta receptor pathways controlling mitosis and eggshell formation. *PLoS pathogens* **9**, e1003448, doi:10.1371/journal.ppat.1003448 (2013).

38 Sanchez, M. C., Cupit, P. M., Bu, L. & Cunningham, C. Transcriptomic analysis of reduced sensitivity to praziquantel in Schistosoma mansoni. *Mol Biochem Parasitol* **228**, 6-15, doi:10.1016/j.molbiopara.2018.12.005 (2019).

39 Chen, J., Guo, K. & Kastan, M. B. Interactions of nucleolin and ribosomal protein L26 (RPL26) in translational control of human p53 mRNA. *J Biol Chem* **287**, 16467-16476, doi:10.1074/jbc.M112.349274 (2012).

40 Leger-Silvestre, I. *et al.* Specific Role for Yeast Homologs of the Diamond Blackfan Anemia-associated Rps19 Protein in Ribosome Synthesis. *J Biol Chem* **280**, 38177-38185, doi:10.1074/jbc.M506916200 (2005).

41 Leger-Silvestre, I. *et al.* The ribosomal protein Rps15p is required for nuclear exit of the 40S subunit precursors in yeast. *The EMBO journal* **23**, 2336-2347, doi:10.1038/sj.emboj.7600252 (2004).

42 Lindstrom, M. S. & Nister, M. Silencing of ribosomal protein S9 elicits a multitude of cellular responses inhibiting the growth of cancer cells subsequent to p53 activation. *PloS one* **5**, e9578, doi:10.1371/journal.pone.0009578 (2010).

43 Qi, Y. *et al.* Ribosomal protein L23 negatively regulates cellular apoptosis via the RPL23/Miz-1/c-Myc circuit in higher-risk myelodysplastic syndrome. *Scientific reports* **7**, 2323, doi:10.1038/s41598-017-02403-x (2017).

44 Sloan, K. E., Bohnsack, M. T. & Watkins, N. J. The 5S RNP couples p53 homeostasis to ribosome biogenesis and nucleolar stress. *Cell Rep* **5**, 237-247, doi:10.1016/j.celrep.2013.08.049 (2013).

45 Teng, T., Mercer, C. A., Hexley, P., Thomas, G. & Fumagalli, S. Loss of tumor suppressor RPL5/RPL11 does not induce cell cycle arrest but impedes proliferation due to reduced ribosome content and translation capacity. *Mol Cell Biol* **33**, 4660-4671, doi:10.1128/MCB.01174-13 (2013).

46 Walczak, C. P. *et al.* Ribosomal protein RPL26 is the principal target of UFMylation. *Proc Natl Acad Sci U S A* **116**, 1299-1308, doi:10.1073/pnas.1816202116 (2019).

47 Wen, L. M. *et al.* Molecular Cloning and Characterization of Ribosomal Protein RPS9 in Echinococcus granulosus. *J Parasitol* **103**, 699-707, doi:10.1645/16-164 (2017).

48 Wilson-Edell, K. A. *et al.* RPL24: a potential therapeutic target whose depletion or acetylation inhibits polysome assembly and cancer cell growth. *Oncotarget* **5**, 5165-5176, doi:10.18632/oncotarget.2099 (2014).

49 Zhou, F., Roy, B. & von Arnim, A. G. Translation reinitiation and development are compromised in similar ways by mutations in translation initiation factor eIF3h and the ribosomal protein RPL24. *BMC Plant Biol* **10**, 193, doi:10.1186/1471-2229-10-193 (2010).

50 Cervantes-Salazar, M. *et al.* Dengue virus NS1 protein interacts with the ribosomal protein RPL18: this interaction is required for viral translation and replication in Huh-7 cells. *Virology* **484**, 113-126, doi:10.1016/j.virol.2015.05.017 (2015).

51 Bowman, R. *et al.* Akirin Is Required for Muscle Function and Acts Through the TGF-beta Sma/Mab Signaling Pathway in Caenorhabditis elegans Development. *G3 (Bethesda, Md.)* **10**, 387-400, doi:10.1534/g3.119.400377 (2020).

52 Tsuchida, T. *et al.* The ubiquitin ligase TRIM56 regulates innate immune responses to intracellular double-stranded DNA. *Immunity* **33**, 765-776, doi:10.1016/j.immuni.2010.10.013 (2010).

53 Ishida, K. & Jolly, E. R. Hsp70 May Be a Molecular Regulator of Schistosome Host Invasion. *PLoS Negl Trop Dis* **10**, e0004986, doi:10.1371/journal.pntd.0004986 (2016).

54 Bais, S., Churgin, M. A., Fang-Yen, C. & Greenberg, R. M. Evidence for Novel Pharmacological Sensitivities of Transient Receptor Potential (TRP) Channels in Schistosoma mansoni. *PLoS Negl Trop Dis* **9**, e0004295, doi:10.1371/journal.pntd.0004295 (2015).

55 Sims, R. J., 3rd *et al.* Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Molecular cell* **28**, 665-676, doi:10.1016/j.molcel.2007.11.010 (2007).

56 Kopytova, D. V. *et al.* Multifunctional factor ENY2 is associated with the THO complex and promotes its recruitment onto nascent mRNA. *Genes Dev* **24**, 86-96, doi:10.1101/gad.550010 (2010).

57 Proshkin, S. A., Shematorova, E. K. & Shpakovski, G. V. The Human Isoform of RNA Polymerase II Subunit hRPB11balpha Specifically Interacts with Transcription Factor ATF4. *Int J Mol Sci* **21**, doi:10.3390/ijms21010135 (2019).

58 Bauernfeind, A. L. & Babbitt, C. C. The predictive nature of transcript expression levels on protein expression in adult human brain. *BMC Genomics* **18**, 322, doi:10.1186/s12864-017-3674-x (2017).

59 Moritz, C. P., Muhlhaus, T., Tenzer, S., Schulenborg, T. & Friauf, E. Poor transcript-protein correlation in the brain: negatively correlating gene products reveal neuronal polarity as a potential cause. *Journal of neurochemistry* **149**, 582-604, doi:10.1111/jnc.14664 (2019).

60 Wang, D. *et al.* A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol Syst Biol* **15**, e8503, doi:10.15252/msb.20188503 (2019).

61 Leng, K. *et al.* Akirin2 is modulated by miR-490-3p and facilitates angiogenesis in cholangiocarcinoma through the IL-6/STAT3/VEGFA signaling pathway. *Cell death & disease* **10**, 262, doi:10.1038/s41419-019-1506-4 (2019).

62 Xue, M. *et al.* Regulation of estrogen signaling and breast cancer proliferation by an ubiquitin ligase TRIM56. *Oncogenesis* **8**, 30, doi:10.1038/s41389-019-0139-x (2019).

63 Weyer, F. A., Gumiero, A., Gese, G. V., Lapouge, K. & Sinning, I. Structural insights into a unique Hsp70-Hsp40 interaction in the eukaryotic ribosome-associated complex. *Nat Struct Mol Biol* **24**, 144-151, doi:10.1038/nsmb.3349 (2017).

64 Frischkorn, C. G., Frischkorn, H. E. & Carrazzoni, E. Cercaricidal activity of some essential oils of plants from Brazil. *Naturwissenschaften* **65**, 480-483, doi:10.1007/BF00702834 (1978).

65 Nagai, Y., Gazzinelli, G., de Moraes, G. W. & Pellegrino, J. Protein synthesis during cercaria-schistosomulum transformation and early development of the Schistosoma mansoni larvae. *Comp Biochem Physiol B* **57**, 27-30 (1977).

66 Atkinson, K. H. & Atkinson, B. G. Protein synthesis in vivo by Schistosoma mansoni cercariae. *Mol Biochem Parasitol* **4**, 205-216 (1981).

67 Wiest, P. M., Tartakoff, A. M., Aikawa, M. & Mahmoud, A. A. Inhibition of surface membrane maturation in schistosomula of Schistosoma mansoni. *Proc Natl Acad Sci U S A* **85**, 3825-3829 (1988).

68 Edfors, F. *et al.* Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol Syst Biol* **12**, 883, doi:10.15252/msb.20167144 (2016).

Figure 1. Total identified proteins and transcripts in cercarial heads and tails. Euler diagram shows the proportional overlap of shared proteins and transcripts as well as uniquely identified genes. (A) all transcripts identified above the ≥ 1.0 normalized count threshold in one or both macrostructures. (B) the same for proteomic identification of proteins with a PEAKS significance threshold ≥ 20. (unless it's the journal format)

Figure 2. Both RNA and protein are significantly differentially expressed in cercarial heads and tails. Panel A shows the MA plot of heads compared to tails with tails being the reference. All (●) are significantly upregulated transcripts with ≥ 2.0 log2 fold change and have an adjusted p-value of ≤ 0.01. All (●) are outside these thresholds. Panel B shows a volcano plot of differentially expressed proteins heads and tails by the ratio (Head/Tail). All (●) are proteins that have a ratio of ≥ 1.6 and a significance ≥ than 13.0 -10 log p-value. All (●) fall outside of these thresholds.

Figure 3. GO analysis of over-expressed transcripts and proteins from cercarial heads and tails. Cercarial tails have a significant overrepresentation of ribosomal and mitochondrial proteins. All data shown represent overrepresented categories of differentially expressed genes via GO analysis. Panels A and B show overrepresented transcripts in cercarial heads and tails. Panels C and D show overrepresented proteins from cercarial heads and tails.

Figure 4. GO analysis of under-represented transcripts and proteins from cercarial heads and tails. Cercarial Tails have a significant underrepresentation of transcripts related to ribosomal genes and translational maintenance, while cercarial heads show underrepresentation of both transcript and protein for membrane component genes. All data shown represent underrepresented categories of differentially expressed genes via GO analysis. Panels A and B show underrepresentation of transcripts for cercarial heads and tails, respectively. Panels C and D show underrepresentation of proteins from cercarial heads and tails, respectively.

Figure 5. Translation and transcription are not highly correlated in cercarial heads and tails. Panels A and B show the intersection of unique transcripts and proteins (A), and differentially expressed transcripts and proteins in (B) from cercarial heads and tails. Panel C and D show a scatter plot of normalized abundance from transcript along the X-axis and normalized abundance from protein along the Y-axis for heads and tails individually. The R^2= 0.0768 and R^2= 0.1783

show little positive correlation between the transcript and protein abundance in cercarial heads (C) and cercarial tails (D), respectively.

Figure 6. Translational repression applies in an asymmetrical and overlapping manner in cercarial heads and tails. An Euler diagram proportionally represents the distribution of translationally repressed transcripts in cercarial heads alone in olive (116), cercarial tails alone in blue (291), and shared transcripts in the intersection (84).

Acknowledgments

I would like to thank the entire Jolly lab for review and edits to the final manuscript. Research reported in this publication was supported by the following grants: National Institutes of Allergy and Infectious Disease of the NIH, grant R21AI137577; B. glabrata snails were provided by the NIAID Schistosomiasis Resource Center of the Biomedical Research Institute (Rockville, MD) through NIH-NIAID Contract HHSN272201700014I for distribution through BEI Resources.

Competing interests

The authors declare no competing interests.

Author Contributions

EJ and JH were responsible for the project conceptualization. JH and HCK were responsible for experimentation and methodology. EJ was responsible for Project administration and Resources. JH, EJ, and HCK were responsible for manuscript writing. EJ, JH, and HCK were responsible for review & editing of the manuscript.

Supplementary Items

1. Full list of identification, differential expression, and abundance of transcripts and proteins -Supplementary Data 1.xlsx
2. Full GO enrichment term lists for unique transcripts and proteins -Supplementary Data 2.xlsx
3. Full GO enrichment term lists for Differentially expression transcripts and proteins –Supplementary Data 3.xlsx

4. Full list of Putatively translationally repressed genes in heads and tails-Supplementary Data 4.xlsx

Supplementary Figure 1. GO analysis of unique transcripts and proteins from cercarial heads and tails. Panel A shows overrepresented transcripts in cercarial Tails. Panel B highlights underrepresented transcripts for cercarial tails across the categories above. Panel C and D are overrepresented unique proteins from cercarial heads and tails, respectively. No underrepresented protein groups were found for cercarial heads or tails. No over or underrepresented transcripts were found via GO analysis for cercarial heads. GO terms are shown across biological process (BP), molecular function (MF), and cell compartment (CC).

Figures

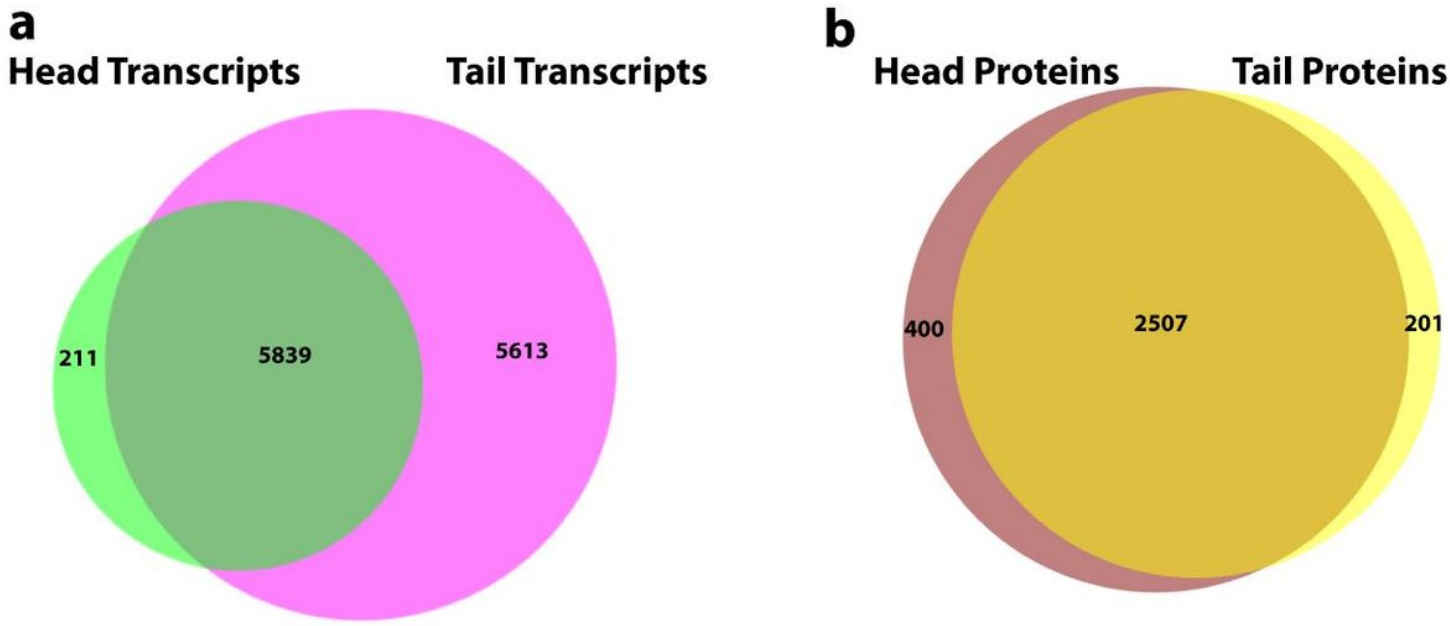

Figure 1

Total identified proteins and transcripts in cercarial heads and tails. Euler diagram shows the proportional overlap of shared proteins and transcripts as well as uniquely identified genes. (A) all transcripts identified above the ≥ 1.0 normalized count threshold in one or both macrostructures. (B) the same for proteomic identification of proteins with a PEAKS significance threshold ≥ 20. (unless it's the journal format)

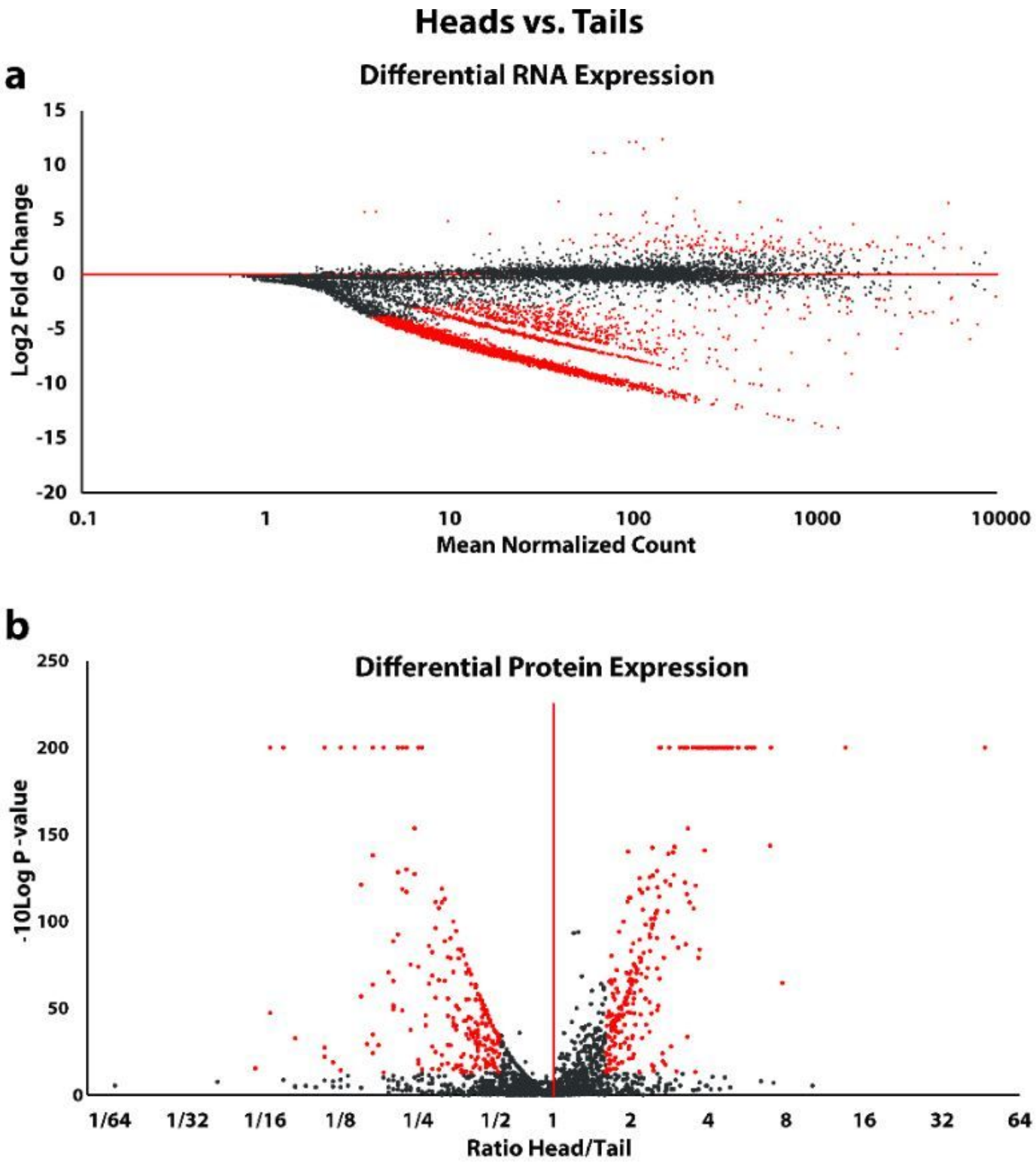

Figure 2

Both RNA and protein are significantly differentially expressed in cercarial heads and tails. Panel A shows the MA plot of heads compared to tails with tails being the reference. All (▯) are significantly upregulated transcripts with ≥ 2.0 log2 fold change and have an adjusted p-value of ≤ 0.01. All (▯) are outside these thresholds. Panel B shows a volcano plot of differentially expressed proteins heads and tails by the ratio

(Head/Tail). All (▯) are proteins that have a ratio of ≥ 1.6 and a significance ≥ than 13.0 -10 log p-value. All (▯) fall outside of these thresholds.

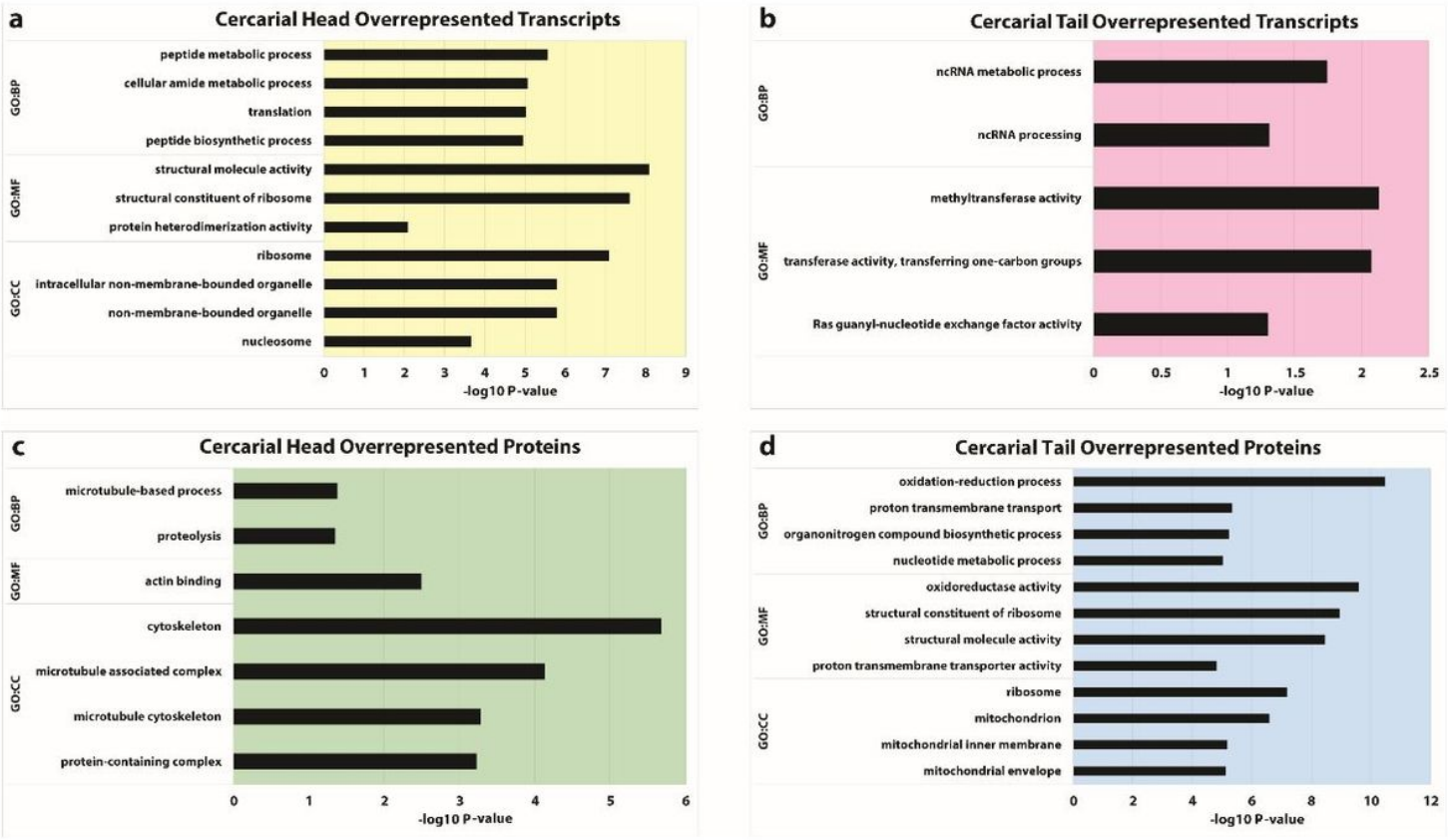

Figure 3

GO analysis of over-expressed transcripts and proteins from cercarial heads and tails. Cercarial tails have a significant overrepresentation of ribosomal and mitochondrial proteins. All data shown represent overrepresented categories of differentially expressed genes via GO analysis. Panels A and B show overrepresented transcripts in cercarial heads and tails. Panels C and D show overrepresented proteins from cercarial heads and tails.

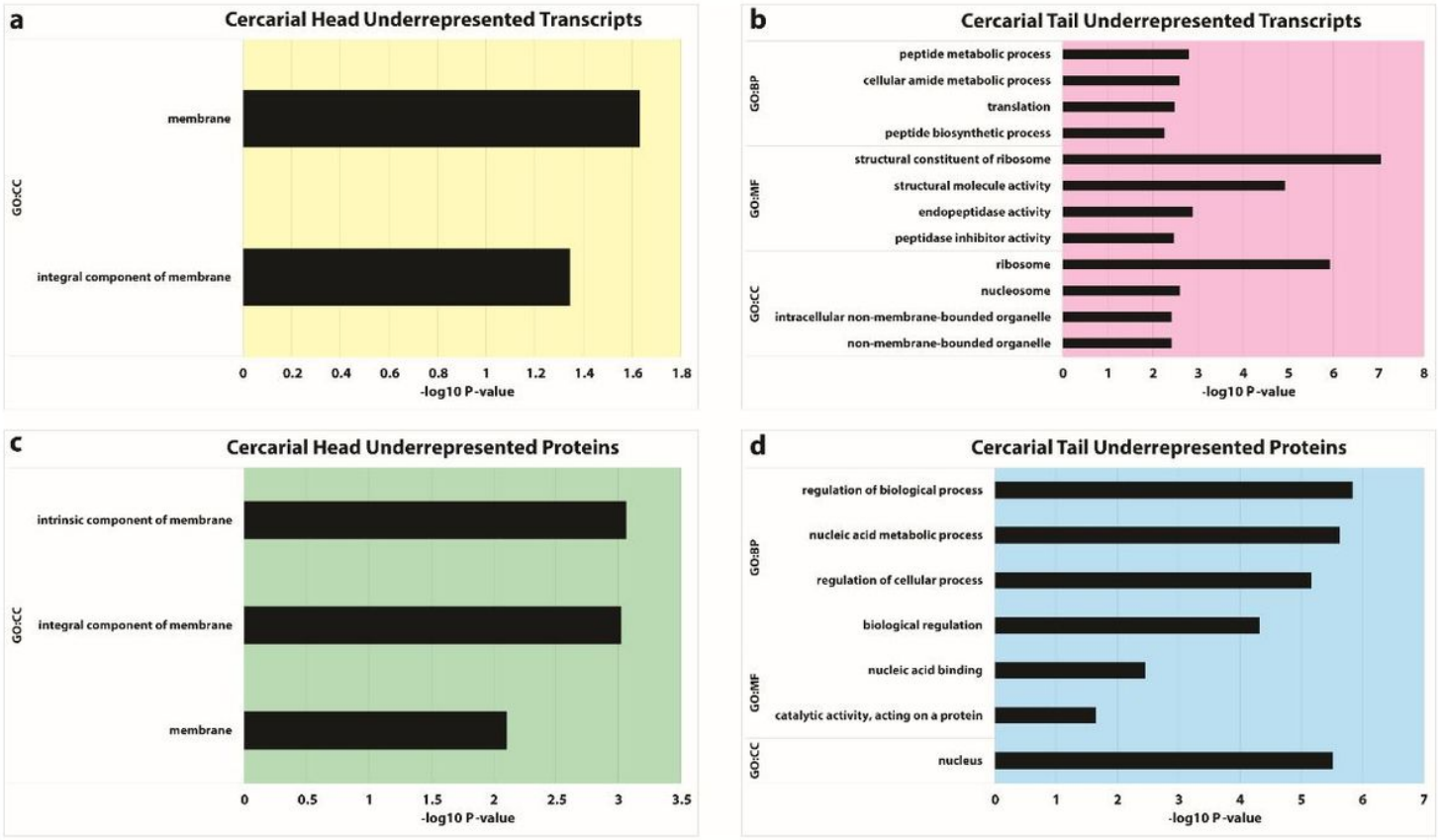

Figure 4

GO analysis of under-represented transcripts and proteins from cercarial heads and tails. Cercarial Tails have a significant underrepresentation of transcripts related to ribosomal genes and translational maintenance, while cercarial heads show underrepresentation of both transcript and protein for membrane component genes. All data shown represent underrepresented categories of differentially expressed genes via GO analysis. Panels A and B show underrepresentation of transcripts for cercarial heads and tails, respectively. Panels C and D show underrepresentation of proteins from cercarial heads and tails, respectively.

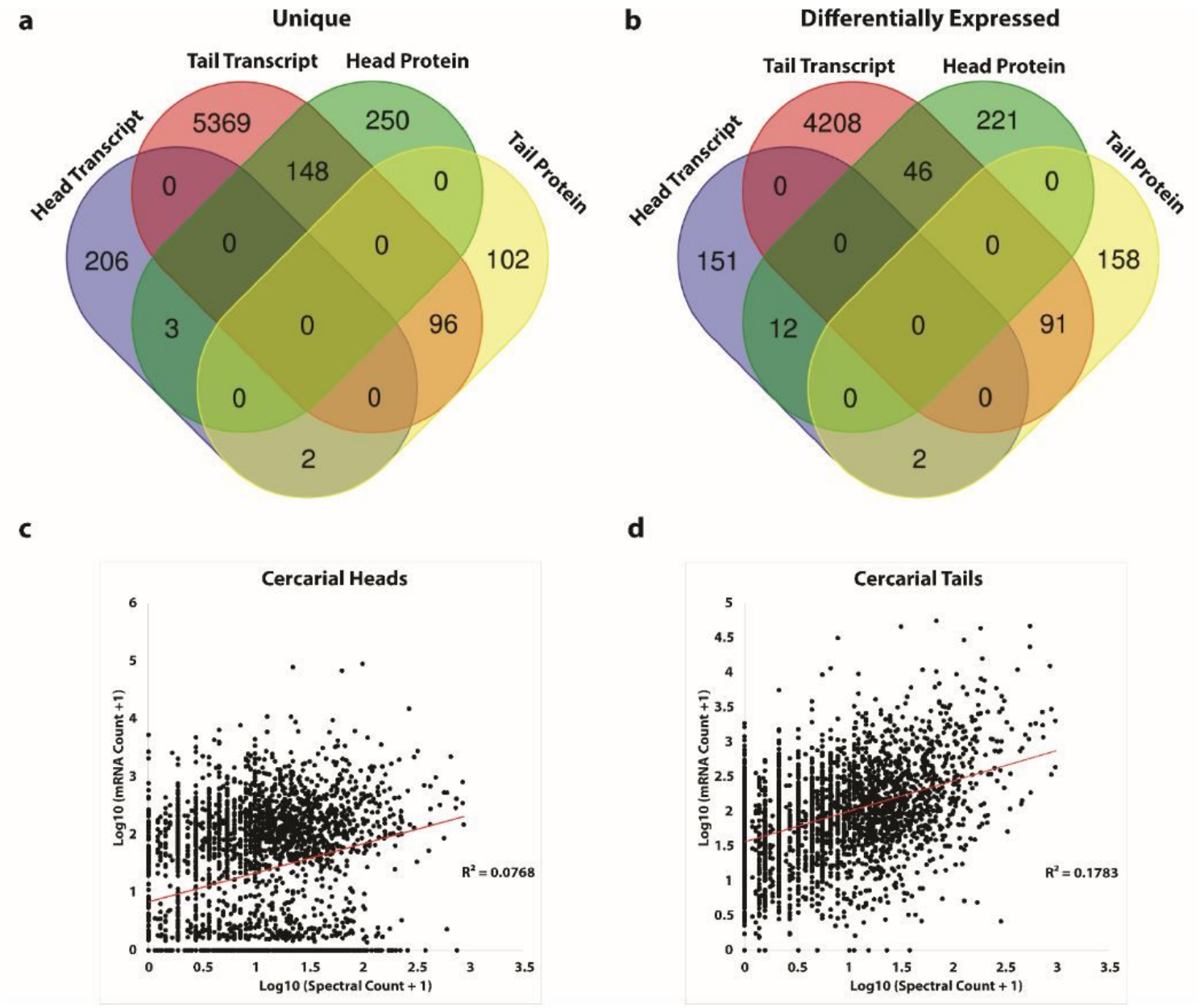

Figure 5

Translation and transcription are not highly correlated in cercarial heads and tails. Panels A and B show the intersection of unique transcripts and proteins (A), and differentially expressed transcripts and proteins in (B) from cercarial heads and tails. Panel C and D show a scatter plot of normalized abundance from transcript along the X-axis and normalized abundance from protein along the Y-axis for heads and tails individually. The R2= 0.0768 and R2= 0.1783 show little positive correlation between the transcript and protein abundance in cercarial heads (C) and cercarial tails (D), respectively.

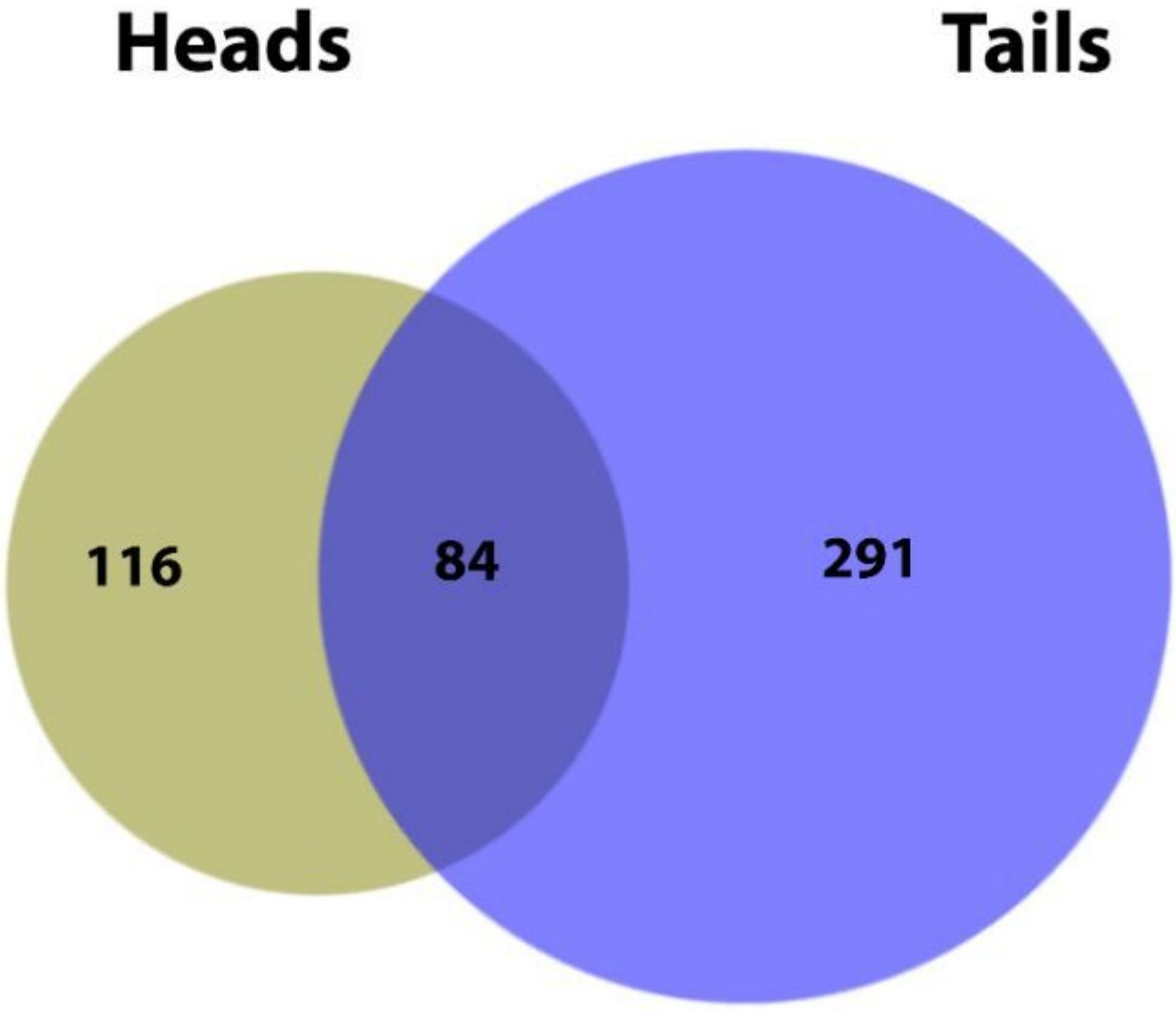

Figure 6

Translational repression applies in an asymmetrical and overlapping manner in cercarial heads and tails. An Euler diagram proportionally represents the distribution of translationally repressed transcripts in cercarial heads alone in olive (116), cercarial tails alone in blue (291), and shared transcripts in the intersection (84).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf
- SupplementaryFigure1.tif

- SupplementaryData1.xlsx
- SupplementaryData2.xlsx
- SupplementaryData3.xlsx
- SupplementaryData4.xlsx