

## Multiomic analysis of *Schistosoma mansoni* reveals unique expression profiles in cercarial heads and tails

James R. Hagerty <sup>1</sup>, Hyung Chul Kim<sup>1</sup> & Emmitt R. Jolly <sup>1,2</sup> 

Schistosomes require both molluscan and mammalian hosts for development. The larval cercaria exits the snail host and swims to identify and invade the mammalian host. The cercaria has two macrostructures, the head and the tail. The head invades the host, where it matures into an adult worm. The tail is lost after host invasion. Translation in the cercaria differs in each macrostructure, with higher levels of translation in the cercarial tail and little to no translational activity in the cercarial head. 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 ribosomal components expressed in higher levels in tails than in heads, which may explain the differences in translation levels between heads and tails. We also characterized the weak correlation between transcription and translation in infectious cercarial heads and tails.

<sup>1</sup>Case Western Reserve University, Department of Biology, Cleveland, OH, USA. <sup>2</sup>Case Western Reserve University, Center for Global Health and Disease, Cleveland, OH, USA. email: [emmitt.jolly@case.edu](mailto:emmitt.jolly@case.edu)

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 identifying and penetrating 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 a freshwater snail and develops into mother and daughter sporocysts. The sporocyst produces and releases the infectious cercariae, completing the life cycle. The cercaria is transiently free-living and represents the first interaction point in the parasite life cycle with the human host. This cercarial stage has been studied using several -omic approaches, including microarrays, RNA-seq, and proteomics; however, our understanding of this stage's transcriptional and translational control mechanisms is still limited<sup>1–8</sup>. Collectively, -omic approaches to date have consistently pointed to a substantial upregulation of genes related to glycolysis and metabolism<sup>9</sup>, consistent with the functional activity of actively swimming cercaria in search of a mammalian host. Proteomic analysis of the whole cercariae and the cercarial secretions released from the cercarial head revealed an abundance of proteases, including elastase and other proteins that are likely involved in host invasion<sup>4,9–11</sup>. 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 potentially clouds our understanding of each macrostructure<sup>12,13</sup>.

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<sup>13</sup>. The cercarial head had significantly lower 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 separate 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, and (3) have a weak correlation between transcript and protein abundance. These findings demonstrate the necessity for treating cercariae as two distinctive macrostructures instead of single units for study.

## 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<sup>14,15</sup>. Since cercariae do not undergo transcription<sup>14</sup>, and translation in cercarial heads and tails is differentially regulated<sup>13</sup>, we explored the conservation of steady-state transcript populations between cercarial heads and tails.

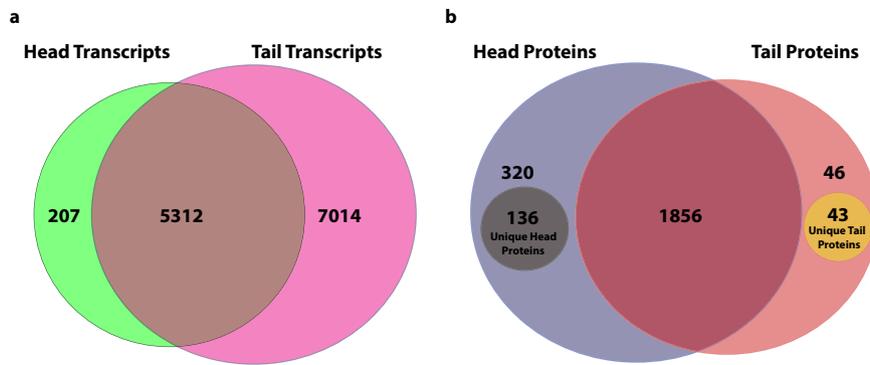
We found a total of 12,533 transcripts in cercariae (Fig. 1a). Cercarial heads and tails share 5,312 transcripts. Among these

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\_318920* (Supplementary Data 1)<sup>16,17</sup>. Cercarial heads have 207 unique transcripts, and cercarial tails have a large store of 7,014 unique transcripts (Fig. 1a). DESeq2 utilizes a Log fold change (LFC) Shrinkage process that reduces the noise created by low abundance transcripts or transcripts only detected in one macrostructure allowing for effective relative comparison of low abundance transcripts<sup>18</sup>.

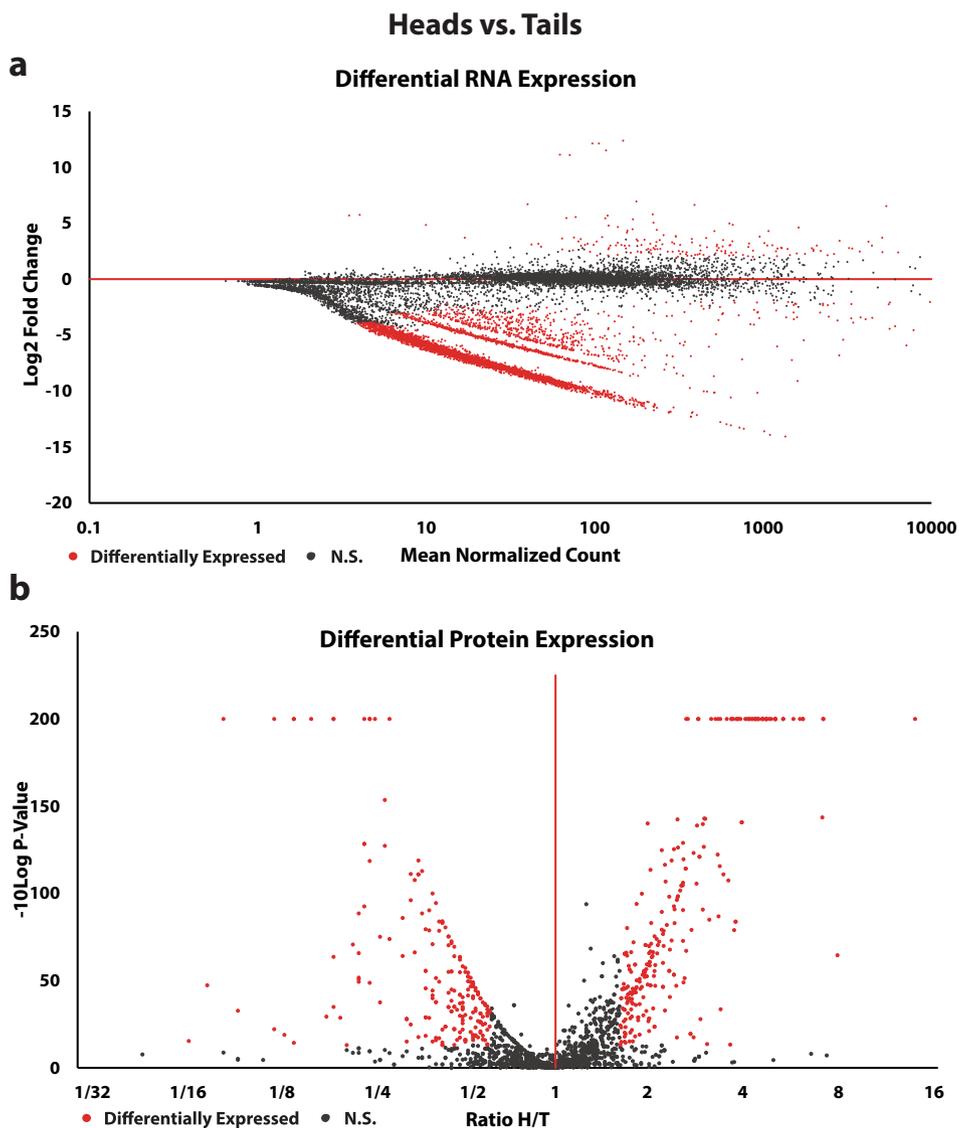
We next investigated protein storage in cercarial heads and tails using label-free mass spectrometry. We identified 2401 proteins and were able to verify 684 of those against the proteins identified by Sotillo et al. (Supplementary Data 4)<sup>19</sup>. Cercarial heads and tails share 1,856 of the 2,401 identified proteins. Four hundred and fifty-six (456) proteins were only identified in  $\geq 2$  replicates of cercarial heads, and 89 were only identified in two or more replicates in cercarial tails. Within the population of proteins fully identified only in heads or tails, 136 unique head proteins were identified in  $\geq 2$  replicates in cercarial heads and were undetected in cercarial tails (Fig. 1b). We also identified 43 unique tail proteins that were identified in  $\geq 2$  replicates in cercarial tails and were undetected in cercarial heads (Fig. 1b). However, while tails have fewer unique proteins compared to heads, they have a higher number of overall transcripts. Unique proteins have been identified in one macrostructure but were not detected in the other macrostructure. The uniquely identified proteins were not analyzed for relative differential expression and are reported separately from our expression analysis and are removed during the filtering in the spectral matching process of label-free quantitation. These proteins are still of interest because they are only stored at detectable levels in a single macrostructure, either the head or the tail. Correspondingly, cercarial tails store a broader diversity of transcripts, while cercarial heads store a broader diversity of proteins.

We used differential expression analysis to look for variation in the transcriptomes and proteomes of heads and tails. In total, 4,511 transcripts were differentially regulated ( $\log_2\text{FoldChange} \geq 2.0$  and  $p\text{-value} \leq 0.01$ ) out of 12,533 transcripts identified in both heads and tails (Fig. 2a). Of the 4,511 differentially regulated transcripts, 164 were upregulated in cercarial heads, and 4,347 were upregulated in cercarial tails (Fig. 2a, Supplementary Data 1). To quantify the differential expression of proteins in cercarial heads and tails, we utilized label-free mass spectrometry, identifying a total of 463 differentially expressed proteins (Fig. 2b). The distribution of differential protein expression between cercarial heads and tails is relatively even. Cercarial heads have 255 upregulated proteins, and cercarial tails contain 208 upregulated proteins (Fig. 2b).

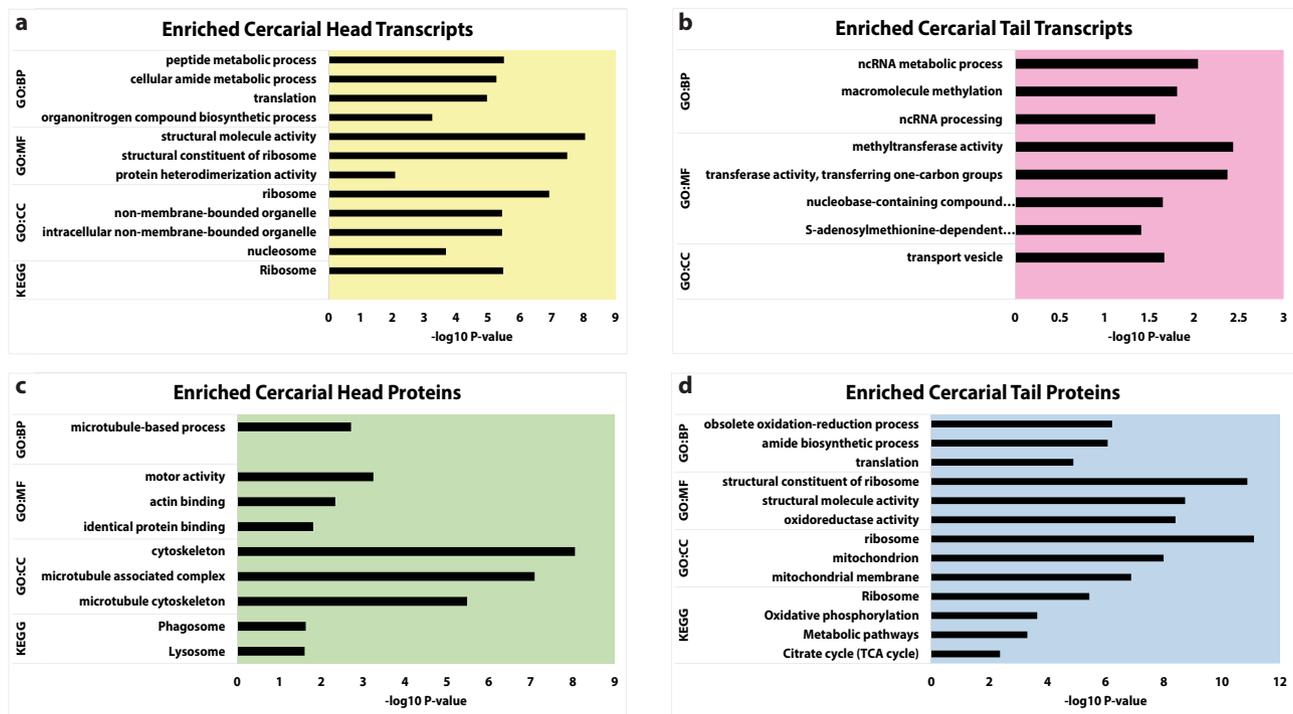
**Cercarial heads store ribosomal transcripts, while cercarial tails store ribosomal proteins.** Given the different functional roles of cercarial heads (host invasion), and cercarial tails (motility and metabolism), we predicted a varied representation in both unique (identified only in head or only in tail samples) and differentially expressed transcripts and proteins in the macrostructures. We performed a gene ontology (GO) enrichment analysis to identify enrichment and reduction of transcript and protein types in 3 categories: biological process (BP), molecular function (MF), and cellular compartment (CC). The enriched status is reported for gene ontology categories if the population of transcripts or proteins that match that category is higher than expected by chance. The reduced status for a GO category is reported if the number of transcripts or proteins that match that category is less than what is expected by chance. To account for false positives and false



**Fig. 1 Total identified proteins and transcripts in cercarial heads and tails.** **a** Venn Diagram showing all transcripts identified above the  $\geq 1.0$  normalized count threshold in one or both macrostructures. **b** Venn Diagram showing proteomic identification of proteins with  $\geq 2$  unique peptides, identified in  $\geq 2$  replicates. **b** includes sub-populations of unique head and unique tail proteins identified in  $\geq 2$  replicates in one macrostructure and 0 replicates in the other macrostructure. One hundred and thirty-six (136) unique head proteins were identified only in cercarial heads, and 43 unique tail proteins were identified only in cercarial tails. All supporting data is contained in Supplementary Data 1. Replication  $n = 3$  independent biological replicates.



**Fig. 2 Both RNA and protein are significantly differentially expressed in cercarial heads and tails.** **a** shows the MA plot of heads compared to tails, with tails being the reference. All red dots (Differentially Expressed) are significantly upregulated transcripts with  $\geq 2.0$  log<sub>2</sub> fold change and have an adjusted  $p$ -value of  $\leq 0.01$ . All black dots (N.S.) are outside these thresholds. **b** shows a volcano plot of differentially expressed proteins heads and tails by the ratio (Head/Tail). All red dots (Differentially Expressed) are proteins with a ratio of  $\geq 1.6$  and a significance  $\geq 13.0$   $-10 \log p$ -value. All black dots (N.S.) fall outside of these thresholds. All supporting data is contained in Supplementary Data 1. Replication  $n = 3$  independent biological replicates.



**Fig. 3** GO analysis of overexpressed transcripts and proteins from cercarial heads and tails. Cercarial Heads have significant enrichment for ribosomal transcripts and structural proteins. Cercarial tails have significant enrichment of ribosomal and mitochondrial proteins. All data shown represent enriched categories of differentially expressed genes via GO analysis. Panels **a** and **b** show enriched transcript categories in cercarial heads and tails. **c**, **d** show enriched protein categories from cercarial heads and tails. All supporting data is contained in Supplementary Data 3.

negatives, a weighted correction method, gSCS, based on the overlap of GO categories, is employed<sup>20</sup>. The input proteins for GO enrichment and reduction analysis were separated into two classes: those that are unique only to heads or only to tails (Supplementary Fig. 1 and Supplementary Data 2) and those that are differentially expressed in heads and tails (Figs. 3 and 4, Supplementary Data 3). Unique proteins were analyzed separately from differentially expressed proteins because the unique proteins are removed from the label-free quantitation during retention matching of protein spectra.

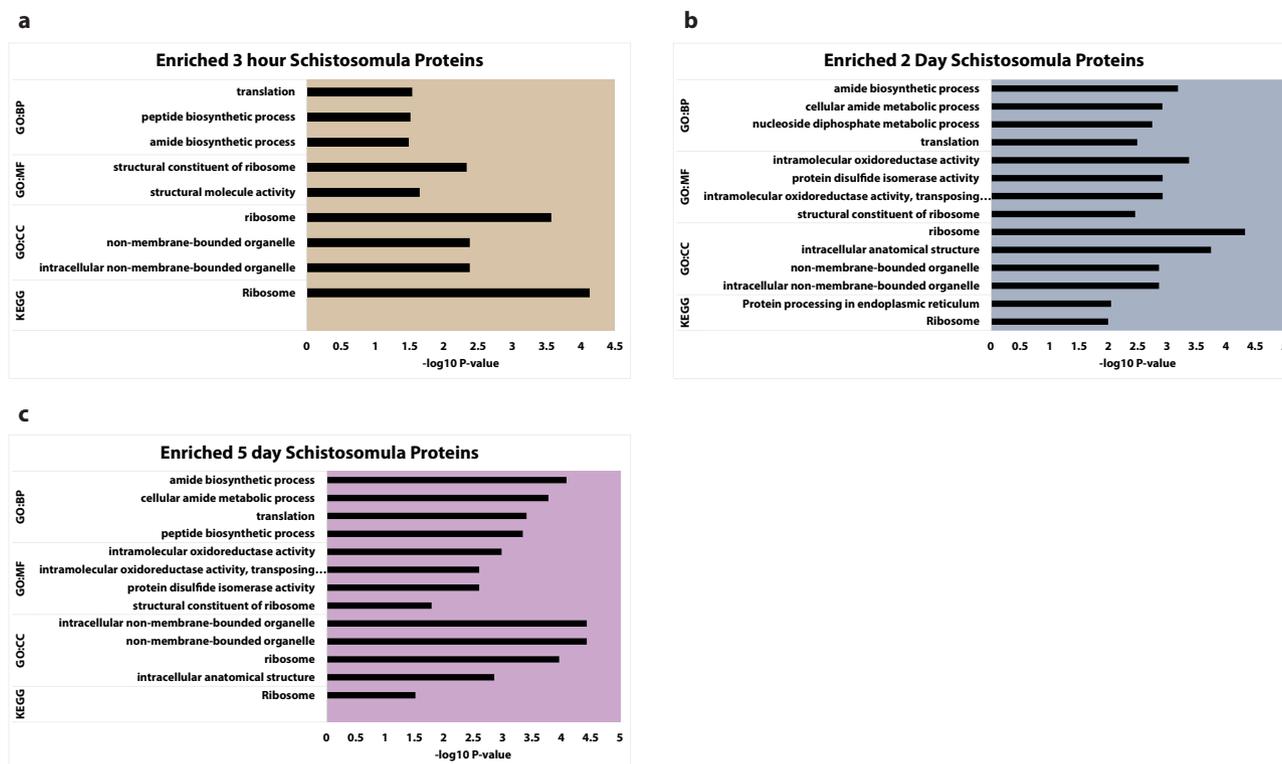
We did not identify any unique enriched or reduced transcript classes or any unique reduced protein classes defined by gene ontology for cercarial heads. We identified unique enriched cercarial head protein groups associated with the spliceosome, such as LSm3 *Smp\_078640*, U1a/U2b *Smp\_069870.1*, and SF3A3 *Smp\_003630* (Supplementary Fig. 1 and Supplementary Data 2). We also identified unique reduced protein groups in cercarial heads related to metabolic pathways. We found multiple enriched GO categories of unique proteins in cercarial tails, including translational and mitochondrial processes (MRPS25 *Smp\_066620*, MRPS5 *Smp\_332830*, and MRPL47 *Smp\_102280*) (Supplementary Fig. 1 and Supplementary Data 2). The enrichment of translation and mitochondrial proteins supports previous findings showing upregulation of metabolic activity (Supplementary Fig. 1 and Data 2). Notably, heads and tails show little overlap in transcript or protein GO function. Cercarial heads lack proteins related to metabolic pathways, and cercarial tails are enriched for proteins related to metabolism, supporting the observed biology of the tail functioning as transient motility. In contrast, the more quiescent head progresses through development.

Using GO enrichment and reduction analysis, we compared differentially expressed transcripts and proteins within each macrostructure (Fig. 3, Supplementary Data 3). Head transcripts

were enriched for several processes, including protein production, ribosomal biogenesis, and chromatin (*RPL27e Smp\_063350*, *RPS23 Smp\_074470*, *MRPS9 Smp\_333040*, and *H3F3C Smp\_082240*) (Fig. 3a, Supplementary Data 3), and tail transcripts were enriched for noncoding RNA processing and chromatin marking genes (*METTL1 Smp\_130610*, *MARS Smp\_040770*, *NAT10 Smp\_144490*, and *DNMT2 Smp\_334230*) (Fig. 3b, Supplementary Data 3).

Cercarial heads were enriched for structural, lysosomal, and phagosomal proteins ( $\alpha$ -Tubulin *Smp\_090120*, Paramyosin *Smp\_021920*, *Smp\_210500* Cathepsin L3, HEXB *Smp\_053900*) (Fig. 3c, Supplementary Data 3). In contrast, cercarial tails were significantly enriched for translational, ribosomal, stress response, and mitochondrial proteins (Fig. 3d, Supplementary Data 3). These include a large population of ribosomal components (*RPL11/12 Smp\_012750*, *RPL24 Smp\_001830*, *RPL27a Smp\_325920*, *RPS9 Smp\_180000*, *RPS13 Smp\_096750*, *RPS15 Smp\_307550*, and *RPS19 Smp\_174950*)<sup>21–32</sup>. The ribosomal component genes described above are involved in ribosomal biogenesis or translational upregulation<sup>21–32</sup>.

We next analyzed upregulated proteins from three-hour, two-day, and five-day schistosomes using data published by Sotillo et al. 2015<sup>19</sup>. Developing schistosomula increase translation rates over multiple days after transformation<sup>13,33</sup>. We could not process the data using our analysis pipeline because the raw data is not publicly available. Over-expression was compared to zero-hour schistosomula. Zero-hour schistosomula are almost equivalent to cercarial heads. The GO enrichment analysis of proteins upregulated in the three time points shows enrichment of protein classes related to translation, ribosomes, and ribosomal components (Fig. 4, Supplementary Data 4). Ribosomal protein RPS13 was overexpressed in all three time points (Supplementary Data 4<sup>19</sup>).



**Fig. 4 GO analysis of enriched proteins in developing schistosomula.** Schistosomula are significantly enriched for ribosomal component proteins at all three developmental time points analyzed. Over-expressed proteins from 3 h, 2 day, and 5 day schistosomula were re-analyzed from publicly available data published by Sotillo et al. 2015<sup>19</sup>. **a** represents enriched GO categories for overexpressed proteins in 3 h schistosomula, **b** represents the same for 2 day schistosomula, and **c** represents the same for 5 day schistosomula. All supporting GO analysis data is contained in Supplementary Data 4.

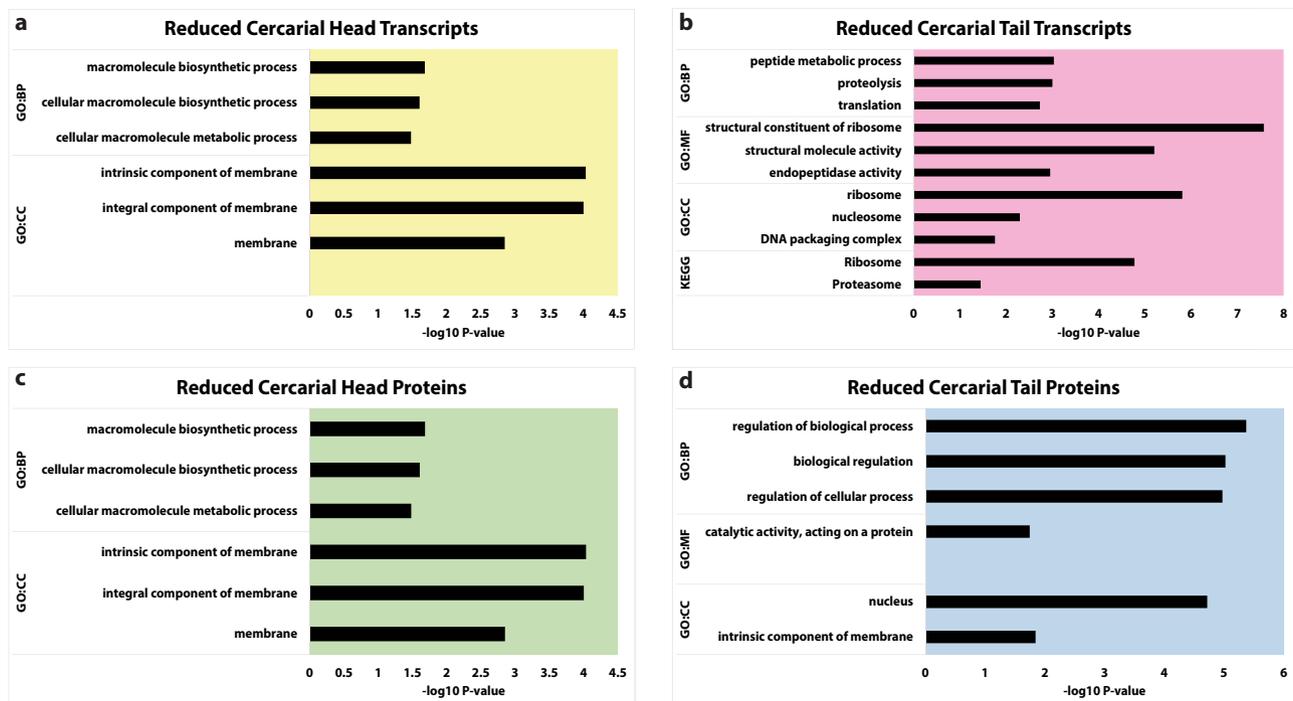
Cercarial heads have few reduced transcript or protein GO classes. The reduced GO status indicates a lack in the expected number of transcripts that fit into a given category based on the size of that category and the size of the input data set. Cercarial heads lacked membrane component transcripts and proteins (Fig. 5a, c, Supplementary Data 3). In cercarial tails, GO analysis highlighted a significant reduction of transcripts involved in translation and peptide biosynthesis. (Fig. 5b, Supplementary Data 3). At the protein level, tails have reduced proteins related to the nucleus and membrane components (Fig. 5d, Supplementary Data 3). Still, they are enriched for ribosomal and translational proteins, although transcripts are significantly lacking in these same categories. Overall, the GO analysis data show that transcript classes in heads or tails unexpectedly do not predict the protein populations.

### Transcripts in cercarial heads and tails are not predictive of translation products.

We further explore the observation that transcripts in heads or tails 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 a Venn diagram. We found that 205 overexpressed head proteins did not have corresponding transcripts, and 152 over-expressed head transcripts did not have corresponding proteins (Fig. 6a). We found that tails have 4235 over-expressed transcripts, for which we were surprisingly unable to identify any corresponding proteins (Fig. 6a). Most surprising was the intersection of over-expressed head proteins and over-expressed tail transcripts. We found 39 proteins that are over-expressed in heads that intersect with transcripts over-expressed in tails. In contrast, we observed only 11 head proteins that intersected with head transcripts. Cercarial tails have the most

robust relationship between differentially expressed transcripts and proteins (Fig. 6a). We identified 71 transcripts that were both overexpressed at the transcript and protein levels in tails. 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 2,254 proteins with respective transcripts from our RNA-seq datasets. We then compared 2,254 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  $R^2 = 0.0605$  and a Spearman correlation of 0.2463 (Fig. 6b). Cercarial tail transcripts and proteins are more highly correlated with an  $R^2 = 0.1564$  and a Spearman correlation of 0.3756 (Fig. 6c). We have previously shown global differences in translational control and expect more specific regulation that represses genes in heads and tails<sup>13</sup>. Given our previous findings, the relationship between over-expressed transcripts and proteins (Fig. 6a), and the lack of correlation between transcript and protein levels (Fig. 6b, c), 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. We verified the RNA-seq abundance of 5 representative genes in cercarial tails using digital droplet PCR. The spearman correlation of the genes was 0.8, the analyzed genes represented high, moderate, and low abundance transcripts by normalized gene counts (Supplementary Fig. 2). We did not observe amplification in our cercarial head samples and could not repeat the tests given material limitations.



**Fig. 5 GO analysis of reduced transcripts and proteins from cercarial heads and tails.** Cercarial Tails have a significant reduction of transcripts related to ribosomal genes and translational maintenance, while cercarial heads show a reduction of both transcript and protein for membrane component genes. All data shown represent reduced categories of differentially expressed genes via GO analysis. **a, b** shows a reduction of transcript categories for cercarial heads and tails, respectively. **c, d** shows a reduction of protein categories from cercarial heads and tails, respectively. All supporting data is contained in Supplementary Data 3.

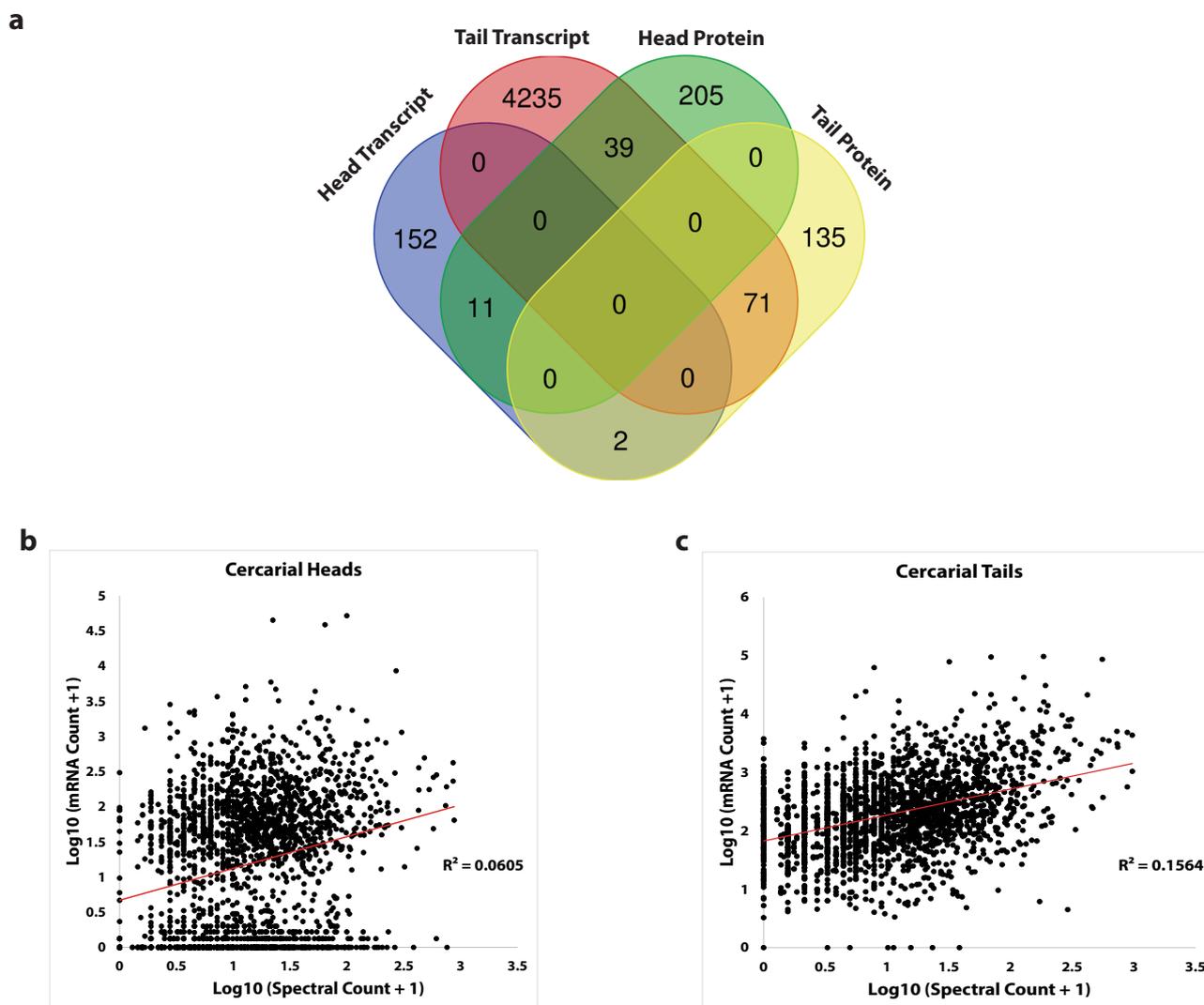
## Discussion

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<sup>1–8</sup>. This work supports these previous findings while also elucidating differences in translational control mechanisms and the lack of correlation between transcription and translation<sup>13</sup>. Cercarial heads and tails vary significantly in overall translation rates, with heads having limited translation compared to tails<sup>13</sup>. This work expands on these previous observations. First, using multiple -omics approaches, we show that transcripts and proteins found in the two cercarial macrostructures are meaningfully different and correspond to the macrostructures' biological functions, quiescent storage, and structural components in heads and metabolic genes to maintain motility in tails. Second, we offer insight into protein production regulation by comparing the cercarial head and cercarial tails transcript and protein groups. Finally, we show that transcript levels and protein levels do not have a meaningful correlation in the cercarial life stage.

Cercarial heads and tails differ in structure and function. The mostly quiescent cercarial head attaches to the host, produces enzymes to facilitate host invasion, and progresses through development to an adult worm after host infection. The cercarial tail is responsible for motility, mechanically assists with host invasion, and is lost upon invasion of the host. 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 unique to the cercarial heads and tails (Fig. 1). The cercarial tail contains 7,014 unique transcripts, and the cercarial head has 207 unique transcripts (Fig. 1a, Supplementary Data 1). The unique proteins show an inverse storage pattern compared to unique transcripts, with the cercarial head containing 136 unique proteins, while the cercarial

tail has 43 unique proteins (Supplementary Data 1). We see a clear difference in the storage of unique protein and transcripts, which is somewhat unexpected given the increased translation rate in cercarial tails compared to cercarial heads. We identified 164 upregulated transcripts in cercarial heads and 4,347 upregulated transcripts in cercarial tails (Fig. 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 translation, and *HISTIC* mRNA can negatively regulate telomere length [55,56]. Developmental regulator *OSK* controls embryonic patterning when translated into protein, while the mRNA independently controls karyosome formation<sup>34</sup>. Long noncoding RNAs are differentially regulated throughout schistosome development and could be playing functional roles in the cercarial head and tail<sup>35,36</sup>. The proteomic differential expression analysis reveals upregulation of 255 and 208 proteins for cercarial heads and tails, respectively (Fig. 2b, Supplementary Data 1). The abundance of unique and differentially expressed transcripts and proteins across cercarial heads and tails supports 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 tails are enriched for ribosomal proteins that increase translational rates, while cercarial heads are enriched for lysosomal and phagosomal proteins related to autophagy (Fig. 3c, d, Supplementary Data 3).

Cercarial heads are enriched for ribosomal components and translation-related transcripts (Fig. 3a, Supplementary Data 3) while generally repressed for translation. These transcripts are likely stored for later translation after host invasion. Cercarial head proteins are enriched for mRNA processing, proteolysis, and



**Fig. 6 Translation and transcription are not highly correlated in cercarial heads and tails.** **a** shows the intersection of differentially expressed transcripts and proteins in cercarial heads and tails. **b, c** 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  $R^2 = 0.0605$  and  $R^2 = 0.1564$  with red linear fit lines show little positive correlation between the transcript and protein abundance in cercarial heads (**b**) and cercarial tails (**c**), respectively. All supporting data is contained in Supplementary Data 1. Replication  $n = 3$  independent biological replicates.

cytoskeleton-related genes (Fig. 3c, Supplementary Fig. 1, Supplementary Data 2, and Supplementary Data 3). The mRNA-related proteins in cercarial heads include LSM3, U1a/U2b, and SF3A3. The storage of these mRNA-related proteins is interesting, given the stalled state of RNA production in cercariae<sup>14,37–39</sup>. The mRNA processing proteins may be stored for the early burst of transcription, which begins shortly after invading the definitive host<sup>14</sup>. These results support Roquis et al. 2015, showing that cercariae do not produce new nucleic acids but are primed for a massive burst of transcription after host invasion<sup>14</sup>. Furthermore, cercarial tails are enriched for differentially expressed proteins related to translation and mitochondrial maintenance MRPS25, MRPS5, and MRPL47 (Fig. 3d, Supplementary Fig. 1, Supplementary Data 2, and Supplementary 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<sup>9</sup>. The separation and individual analysis of heads and tails give insight into mechanisms of translational control and development. We find that the cercarial head is enriched with proteins that can assist in remodeling after host invasion. We can now report that the

cercarial tail is specifically storing ribosomal components and a large population of unused transcripts that had not been previously identified.

Previously, we demonstrated that cercariae have significantly different translational regulation profiles for the heads and tails<sup>13</sup>. Our analysis of the transcriptome and proteome suggests likely mechanisms of translational regulation in heads and tails. We observed a large population of proteins related to ribosomal biogenesis, translational initiation, and ribosomal maintenance in cercarial tails (Fig. 3d, Supplementary Data 3). These include RPL11/12, RPL24, RPL27a, RPS9, RPS13, RPS15, and RPS19<sup>21–32</sup>. More specifically, the ribosomal proteins, RPS19, RPL23, RPL27a, RPS13, RPS15, and RPL24, are enriched in translationally efficient polysomes compared to inefficient monosomes<sup>40</sup>. The opposing GO enrichment of ribosomal component proteins and GO reduction of ribosomal component transcripts is striking in the cercarial tail. The cercarial tail stores significant populations of ribosomal proteins but lacks transcripts for ribosomal proteins. (Figs. 3b and 5b, Supplementary Data 3). On the other hand, cercarial heads are enriched for ribosomal transcripts (Fig. 3a, Supplementary Data 3). Together these

**Table 1 Over-expressed (O.E.) ribosomal component proteins in cercarial tails.**

Accession	Ribosomal Component	Cercarial Tail	3 h Schistosomula	2 Day Schistosomula	5 Day Schistosomula
Smp_180000	RPS9	O.E.	—	O.E.	—
Smp_096750	RPS13	O.E.	O.E.	O.E.	O.E.
Smp_074780	RPS18	O.E.	O.E.	—	—
Smp_174950	RPS19	O.E.	O.E.	O.E.	—
Smp_012750	RPL12	O.E.	—	O.E.	O.E.
Smp_022640	RPL13	O.E.	O.E.	—	O.E.
Smp_007900	RPL 23	O.E.	—	—	—
Smp_325920	RPL27a	O.E.	—	—	—
Smp_307550	RPS15	O.E.	—	—	—
Smp_001830	RPL24	O.E.	—	—	—
Smp_018990	RPL9	O.E.	ID	ID	ID
Smp_032260	RPL15	O.E.	—	—	—
Smp_038510	RPL6	O.E.	—	—	—
Smp_101450	RPL7	O.E.	—	—	—
Smp_024850	RPL17	O.E.	—	—	—
Smp_090230	RPL13a	O.E.	ID	ID	ID
Smp_013470	RPS2	O.E.	ID	ID	ID
Smp_210310	RPL7a	O.E.	—	—	—
Smp_175740	RPL14	O.E.	—	—	—

Overexpressed (O.E.) ribosomal component proteins in the cercarial tail were compared to the dataset of Sotillo et al. 2015<sup>19</sup>. (O.E.) denotes over-expression, (—) denotes no additional information, (ID) denotes verification of protein without over-expression. All supporting data is contained in Supplementary Data 1 and Sotillo et al. 2015<sup>19</sup>.

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<sup>24,27,31</sup>.

We next verified the correlation of the ribosomal component proteins with increases in translational rates during development in schistosomula using publicly available proteomics data from Sotillo et al. 2015<sup>19</sup>. Ribosomal component proteins, (RPS9 *Smp\_180000*, RPS13 *Smp\_096750*, RPS18 *Smp\_074780*, RPS19 *Smp\_174950*, RPL12 *Smp\_012750*, and RPL13 *Smp\_022640*) are over-expressed in both cercarial tails and developing schistosomula (Table 1, Supplementary Data 1, and Supplementary Data 4). All six components are over-expressed in the translationally active cercarial tail and at least two other time points in schistosomula between three hours and five days of development (Table 1). RPS2 *Smp\_013470*, RPL9 *Smp\_018990*, and RPL13a *Smp\_090230* are over-expressed in cercarial tails but were not over-expressed in developing schistosomula, but their identification was validated (Table 1, Supplementary Data 1, and Supplementary Data 4). The cercarial tail and developing schistosomula are all more translationally active than the cercarial head<sup>13,33,41</sup>. Interestingly elongation factors were not overexpressed in schistosomula until day two and maintained overexpression at day five compared to zero-hour. We did not identify differential protein expression of translational initiation or elongation factors in our comparison of cercarial heads and cercarial tails (Supplementary Data 1).

We then explored the predictive quality of transcript on protein levels in cercariae using a matched and normalized correlation plot, as well as a Venn diagram showing overlapping areas of overexpressed genes (Fig. 6). 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 (Fig. 6a). Next, we analyzed the overall correlation of transcript and protein abundance using normalized counts and normalized weighted spectral counts. We found a weak positive correlation  $R^2 = 0.0605$  in cercarial heads and cercarial tails  $R^2 = 0.1564$  (Fig. 6b, c) between transcript and protein abundance. These values represent a weak

correlation compared to multiple studies performed on human tissues<sup>42–44</sup>. An analysis of 29 tissues in humans showed that most proteome and transcriptomes have a ~0.50  $R^2$  correlation<sup>44</sup>. The Spearman ranked correlation, 0.2463 in cercarial heads and 0.3756 in cercarial tails, reveals some predictive power though the relationships are well below the direct correlation values in human tissues<sup>44</sup>. Therefore, the predictive power of transcripts is low in cercariae. Given the pattern of low correlation in other systems and tissues, it is likely that this correlation is low in other developmental stages.

We have demonstrated that cercarial heads and tails are unique macrostructures that serve different functions and prepare in different ways for those roles. The cercarial head is primed with mRNA processing transcripts, motor proteins, and autophagy-related proteins to allow for the large burst of transcription and remodeling after transformation (Fig. 3a, c, Supplementary Data 1). The cercarial tail is enriched with ribosomal proteins and metabolic proteins to maintain motility and assist in honing into a potential host (Fig. 3b, d, Supplementary Data 1). Our finding suggests that the number of ribosomal proteins and ribosomal component mRNAs may be the mechanism for how translation is regulated differently between cercarial 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<sup>15,45,46</sup>. The transcriptome and proteome are not strongly correlated in cercariae. The lack of correlation could stem from posttranscriptional regulation, leading to massively different translational rates that vary by five orders of magnitude<sup>47</sup>.

Our findings have opened new questions. What role do the large populations of stored transcripts in tails that do not appear to be translated play in this transient structure that does not progress through development? Interestingly the cercarial tail can develop after separating the head and proliferate new cells and growths in culture<sup>48</sup>. The signals for transformation after the loss of the tail that leads to growth remain unclear. Although the tail is not directly involved in development after transformation, it may store transcripts and proteins related to development that receive a similar signal to begin growth when separated from the head. Furthermore, we know mRNAs escape translational repression

and are translated in the head and the early schistosomula. 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 as they develop toward adulthood, particularly after host invasion in early schistosomula. We propose two likely mechanisms for translational regulation in cercariae: ribosomal quantity and ribosomal heterogeneity. Ribosomal quantity means an increased overall abundance of actively translating ribosomes are available; thus, translational rates are increased. Ribosomal heterogeneity proposes a different model that does not require a difference in the number of ribosomes but differences in the type of available ribosomes. These differences can include different ratios of ribosomal component proteins, differences in rRNA, and differences in post-translational modifications. We identified seven overexpressed ribosomal proteins found in higher ratios in actively translating polysomes<sup>40</sup>. These targets of nascent translation in schistosomula can give insight into early founder proteins and minimally required translation for maintenance of the organism.

## Methods

**Parasite collection.** Biomphalaria glabrata snails infected with *Schistosoma mansoni* (NMRI strain) were obtained from the Biomedical Research Institute (BRI; Rockville, MD). Cercariae were shed from infected *B. glabrata* snails were kept in total darkness for 48 h before exposure to intense white light for 1.5 h to release cercariae. Cercariae were collected then concentrated at  $1000 \times g$  for 15 min at 4 °C<sup>49</sup>.

**Cercarial head and tail separation.** After collection, cercariae were incubated in incomplete DMEM (Gibco) with 10% ethanol supplementation to prevent further development. Cercarial heads and tails were then separated by vortexing and passing through a 22-gauge needle. The mixed head and tail samples were washed with incomplete DMEM and segregated using an ice-cold 70% percoll gradient and centrifugation at  $1000 \times g$  for 25 min at 4 °C. Head and tail fractions were washed with incomplete DMEM 3 times and visualized for contamination. Samples were then washed in 1x PBS 3 times, pelleted, and flash-frozen<sup>13</sup>.

**RNA extraction.** RNA was extracted from ~60,000 cercarial heads and cercarial tails per replicate 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<sup>50,51</sup>. 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<sup>52</sup>, and the adapters were trimmed using Trimmomatic<sup>53</sup>. The datasets were then aligned to the *S. mansoni* genome using HISAT2<sup>54</sup>; then, the transcripts were assembled with the *S. mansoni* annotation as the reference using Stringtie<sup>55</sup>. The transcripts were quantified using the mapping mode of Salmon<sup>56</sup> with the *S. mansoni* genome sequence and the annotation used as the reference transcripts.

Additional filtering to remove any transcripts that match any snail host transcript was performed. The three sets of uninfected *B. glabrata* RNA-seq datasets (accession number PRJNA602007) were quality-checked, trimmed, aligned, and assembled as described above. Three transcripts that match snail transcripts were further removed prior to the differential gene expression analysis.

To further filter the poorly characterized putative genes and pseudogenes in the *S. mansoni* reference annotation, the standalone version of riboPicker<sup>57</sup> 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<sup>58</sup>. Rfam was used for 5 S and 5.8 S subunit rRNA sequences (release 14.2)<sup>59,60</sup>. 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, 5 S, 5.8 S, and the 6 small subunit indices were used together for the analysis.

Statistical analyses were performed with the R environment 4.0.2 v3.4.4: with DESeq2<sup>18</sup>, readr<sup>61</sup>, and tximport<sup>62</sup> libraries loaded. The log<sub>2</sub> fold change shrinkage was performed with the apeglm package<sup>63</sup>.

**Digital droplet PCR.** Absolute quantitation of representative transcripts was performed via digital droplet PCR using the Biorad QX200 system. Biorad Eva-Green Supermix was added at a 1x concentration to 20  $\mu$ L reactions. Each reaction used a primer concentration of 150  $\mu$ M, and reactions were performed in duplicate. Each reaction contained 10 ng input cDNA or RNA only for negative controls. No template controls were performed in single replicates. cDNA generation using Superscript kit following the standard protocol with oligodT primer. Quantitation was performed using ddpcRquant webtool<sup>64</sup>. No template control samples were merged for threshold analysis, and default settings were used for processing<sup>64</sup>. Primer sequences are in Supplementary Data 1. Thermocycler conditions: 95 °C for 5 min, 95 °C for 30 sec and 60 °C for 1 min 40 cycles, 4 °C for 5 min, 90 °C for 5 min.

**Protein sample preparation.** Cercarial heads and tails, ~60,000 per replicate, were thawed on ice for 30 min in a 300  $\mu$ L 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 alkylation were performed following the FASP protocol<sup>65</sup>.

After homogenization was complete, samples were processed using the FASP protocol and Amicon Ultra MWCO 3 K filter (Millipore, Billerica, MA). Samples were reduced and alkylated 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  $\mu$ L in 8 M urea. The sample concentration was determined using the Bradford assay kit (Bio-Rad, Hercules, CA).

Following sample cleanup using FASP, 10  $\mu$ g 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 °C. Following lysyl digestion, the urea concentration was further reduced to 2 M 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 °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  $\mu$ L 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 Velos 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  $\mu$ m  $\times$  20 mm, packed with C18 Symmetry, 5  $\mu$ m, 100 Å, Waters, Taunton, MA) and subsequently resolved in a reversed-phase column (75  $\mu$ m  $\times$  250 mm nano column, packed with C18 BEH130, 1.7  $\mu$ m, 130 Å (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)<sup>66–68</sup>. 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 min. Total ion current (TIC) normalization was applied to all samples. 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  $\geq 13.0$  and a Head/Tail (H/T) ratio  $\geq 1.6$  and  $\leq 0.62$  for total reporting. All proteins specifically discussed require a more stringent H/T ratio of  $\geq 2.0$  and  $\leq 0.50$ . All identified proteins were detected with  $\geq 2$  unique peptides and detected in  $\geq 2$  of the analyzed replicates. Unique

proteins are identified in  $\geq 2$  replicates in one macrostructure and 0 replicates in the other macrostructure.

**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<sup>20</sup>. The significance thresholds for enrichment or reduction were set at 0.05 and used the gSCS algorithm for analysis<sup>20</sup>. All significance is reported as  $-\log_{10}$  P-value. The gProfiler tool was utilized for both protein and mRNA GO analysis. Gene Ontology and KEGG pathway enrichment and reduction are reported.

**Correlation and Venn diagram mapping.** Global mRNA and protein abundance were plotted as  $\log_{10}$  (normalized count + 1) and  $\log_{10}$  (normalized weighted spectral count + 1). The comparison of the proteome to the transcriptome utilized linear  $R^2$  correlation and Spearman ranked correlation. The  $\log_{10}$  counts were plotted against each other as a direct correlation scatter plot.

mRNA abundance measured by RNA-seq was plotted against mRNA abundance as measured by ddPCR using  $\log_{10}$  of normalized count from RNA-seq and  $\log_{10}$  of copies/ $\mu$ L from ddPCR quantitation. Linear  $R^2$  and Spearman ranked correlation are reported. The plot was a direct correlation scatter plot.

**Statistics and reproducibility.** Statistical tests used in this study were carried out using DESeq2 (RNA-seq), PEAKSX, and Scaffold PerSPECTive as described above. All omic analysis was performed in triplicate with three biological replicates. All p-values and adjusted p-values for -omic identification and differential expression analysis are included in Supplementary Data 1 and Supplementary Data 4. GO analysis was performed using gProfiler and the gSCS algorithm. All GO analysis p-values are reported in Supplementary Data 2-4. 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.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All data supporting the findings of this paper are contained within the paper or with the provided Supplementary Information and Supplementary Data files. A summary of source data for each figure is listed in the figure legends and the Supplementary Information file. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD026435 and <https://doi.org/10.6019/PXD026435><sup>69</sup>. The RNA-Seq data analyzed in this study are available on the National Center for Biotechnology Information (NCBI) database under the BioProject PRJNA734345<sup>70</sup>. All other data are available from the corresponding author upon reasonable request.

Received: 24 September 2020; Accepted: 15 June 2021;

Published online: 12 July 2021

## References

- 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 (2010).
- Jolly, E. R. et al. Gene expression patterns during adaptation of a helminth parasite to different environmental niches. *Genome Biol.* **8**, R65 (2007).
- 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: N. Rev. Mol. Cell. Dev. Biol.* **26**, 752–765 (2004).
- Liu, M. et al. Proteomic analysis on Cercariae and Schistosomula in reference to potential proteases involved in host invasion of schistosoma japonicum Larvae. *J. proteome Res.* **14**, 4623–4634 (2015).
- 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 (2011).
- 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 (2016).
- 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 (2012).
- Verjovski-Almeida, S. et al. Transcriptome analysis of the acelomate human parasite *Schistosoma mansoni*. *Nat. Genet.* **35**, 148–157 (2003).
- 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 (2004).
- Cass, C. L. et al. Proteomic analysis of *Schistosoma mansoni* egg secretions. *Mol. Biochem. Parasitol.* **155**, 84–93 (2007).
- Jang-Lee, J. et al. Glycomics analysis of *Schistosoma mansoni* egg and cercarial secretions. *Mol. Cell. Proteom.* **6**, 1485–1499 (2007).
- Knudsen, G. M., Medzihradsky, K. F., Lim, K. C., Hansell, E. & McKerrow, J. H. Proteomic analysis of *Schistosoma mansoni* cercarial secretions. *Mol. Cell. Proteom.* **4**, 1862–1875 (2005).
- Hagerty, J. R. & Jolly, E. R. Heads or tails? Differential translational regulation in cercarial heads and tails of schistosome worms. *PLoS ONE* **14**, e0224358 (2019).
- Roquis, D. et al. The epigenome of *Schistosoma mansoni* provides insight about how cercariae poise transcription until infection. *PLoS Negl. Trop. Dis.* **9**, e0003853 (2015).
- 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. USA* **85**, 3825–3829 (1988).
- 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 Pathog.* **9**, e1003448 (2013).
- 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 (2019).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Sotillo, J., Pearson, M., Becker, L., Mulvenna, J. & Loukas, A. A quantitative proteomic analysis of the tegumental proteins from *Schistosoma mansoni* schistosomula reveals novel potential therapeutic targets. *Int. J. Parasitol.* **45**, 505–516 (2015).
- Raudvere et al. gProfiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids Res.* **47**, W191–W198 (2019).
- 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 (2012).
- 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 (2005).
- Leger-Silvestre, I. et al. The ribosomal protein Rps15p is required for nuclear exit of the 40S subunit precursors in yeast. *EMBO J.* **23**, 2336–2347 (2004).
- 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 (2010).
- Qi, Y. et al. Ribosomal protein L23 negatively regulates cellular apoptosis via the RPL23/Miz-1/c-Myc circuit in higher-risk myelodysplastic syndrome. *Sci. Rep.* **7**, 2323 (2017).
- 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 (2013).
- 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 (2013).
- Walczak, C. P. et al. Ribosomal protein RPL26 is the principal target of UFMylation. *Proc. Natl Acad. Sci. USA* **116**, 1299–1308 (2019).
- Wen, L. M. et al. Molecular cloning and characterization of ribosomal protein RPS9 in *Echinococcus granulosus*. *J. Parasitol.* **103**, 699–707 (2017).
- 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 (2014).
- 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 (2010).
- 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 (2015).

33. Harrop, R. & Wilson, R. A. Protein synthesis and release by cultured schistosomula of *Schistosoma mansoni*. *Parasitology* **107**, 265–274 (1993).
34. Kanke, M. et al. oskar RNA plays multiple noncoding roles to support oogenesis and maintain integrity of the germline/soma distinction. *RNA* **21**, 1096–1109 (2015).
35. Kim, H. C., Khalil, A. M. & Jolly, E. R. LncRNAs in molluscan and mammalian stages of parasitic schistosomes are developmentally-regulated and coordinately expressed with protein-coding genes. *RNA Biol.* **17**, 805–815 (2020).
36. Vasconcelos, E. J. R. et al. The *Schistosoma mansoni* genome encodes thousands of long non-coding RNAs predicted to be functional at different parasite life-cycle stages. *Sci. Rep.* **7**, 10508 (2017).
37. Sims, R. J. 3rd et al. Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol. cell* **28**, 665–676 (2007).
38. Wu, D. et al. Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation. *Cell Res.* **24**, 233–246 (2014).
39. Saldi, T., Wilusz, C., MacMorris, M. & Blumenthal, T. Functional redundancy of worm spliceosomal proteins U1A and U2B<sup>+</sup>. *Proc. Natl Acad. Sci. USA* **104**, 9753–9757 (2007).
40. Slavov, N., Semrau, S., Airoidi, E., Budnik, B. & van Oudenaarden, A. Differential Stoichiometry among core ribosomal proteins. *Cell Rep.* **13**, 865–873 (2015).
41. Blanton, R. E. & Licate, L. S. Developmental regulation of protein synthesis in schistosomes. *Mol. Biochem. Parasitol.* **51**, 201–208 (1992).
42. Bauernfeind, A. L. & Babbitt, C. C. The predictive nature of transcript expression levels on protein expression in adult human brain. *BMC Genomics* **18**, 322 (2017).
43. Moritz, C. P., Muhlhaus, T., Tenzer, S., Schulenburg, T. & Friauf, E. Poor transcript-protein correlation in the brain: negatively correlating gene products reveal neuronal polarity as a potential cause. *J. Neurochem.* **149**, 582–604 (2019).
44. Wang, D. et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503 (2019).
45. 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).
46. Atkinson, K. H. & Atkinson, B. G. Protein synthesis in vivo by *Schistosoma mansoni* cercariae. *Mol. Biochem. Parasitol.* **4**, 205–216 (1981).
47. Edfors, F. et al. Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol. Syst. Biol.* **12**, 883 (2016).
48. Coultas, K. A. & Zhang, S.-M. In vitro cercariae transformation: comparison of mechanical and nonmechanical methods and observation of morphological changes of detached cercariae tails. *J. Parasitol.* **98**, 1257–1261 (2012).
49. Milligan, J. N. & Jolly, E. R. Cercarial transformation and in vitro cultivation of *Schistosoma mansoni* schistosomules. *J. Vis. Exp.* <https://doi.org/10.3791/31913191> [pii] (2011).
50. Howe, K. L. et al. WormBase 2016: expanding to enable helminth genomic research. *Nucleic acids Res.* **44**, D774–D780 (2016).
51. 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 (2017).
52. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).
53. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
54. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
55. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
56. 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 (2017).
57. Schmieder, R., Lim, Y. W. & Edwards, R. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28**, 433–435 (2012).
58. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids Res.* **41**, D590–D596 (2013).
59. Kalvari, I. et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic acids Res.* **46**, D335–D342 (2018).
60. R. Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2019).
61. Wickham, H., Hester, J. & Francois, R. readr: Read Rectangular Text Data. (2017).
62. 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**, <https://doi.org/10.12688/f1000research.7563.2> (2016).
63. 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 (2018).
64. Trypsteen, W. et al. ddpcRquant: threshold determination for single channel droplet digital PCR experiments. *Anal. Bioanal. Chem.* **407**, 5827–5834 (2015).
65. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
66. Tran, N. H. et al. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nat. Methods* **16**, 63–66 (2019).
67. Tran, N. H. et al. Complete de novo assembly of monoclonal antibody sequences. *Sci. Rep.* **6**, 31730 (2016).
68. Tran, N. H., Zhang, X., Xin, L., Shan, B. & Li, M. De novo peptide sequencing by deep learning. *Proc. Natl Acad. Sci. USA* **114**, 8247–8252 (2017).
69. Hagerty, J. R., Kim, H. C. & Jolly, E. R. Multiomic analysis of *Schistosoma mansoni* reveals unique expression profiles in cercarial heads and tails RNA-Seq datasets. *NCBI SRA PRJNA734345* (2021).
70. Hagerty, J. R., Kim, H. C. & Jolly, E. R. Multiomic analysis of *Schistosoma mansoni* reveals unique expression profiles in cercarial heads and tails proteomics datasets. *PRIDE* <https://doi.org/10.6019/PXD026435> (2021).

## Acknowledgements

We 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.

## Author contributions

E.J. and J.H. were responsible for the project conceptualization. J.H. and H.C.K. were responsible for experimentation and methodology. E.J. was responsible for Project administration and Resources. J.H., E.J., and H.C.K. were responsible for manuscript writing. E.J., J.H., and H.C.K. were responsible for the review & editing of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02366-w>.

**Correspondence** and requests for materials should be addressed to E.R.J.

**Peer review information** *Communications Biology* thanks Javier Sotillo and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: Carolina Perdigoto and George Inglis. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021