RESEARCH ARTICLE

Discovery of long non-coding RNAs in the liver fluke, *Fasciola hepatica*

Paul McVeigh[1]*, **Erin McCammick**[1], **Emily Robb**[1], **Peter Brophy**[2], **Russell M. Morphew**[2], **Nikki J. Marks**[1], **Aaron G. Maule**[1]

1 School of Biological Sciences, Queen's University Belfast, Northern Ireland, United Kingdom,
2 Department of Life Sciences, Aberystwyth University, Wales, United Kingdom

* paul.mcveigh@qub.ac.uk
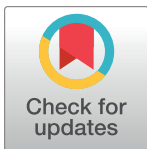
Abstract

Long non-coding (lnc)RNAs are a class of eukaryotic RNA that do not code for protein and are linked with transcriptional regulation, amongst a myriad of other functions. Using a custom *in silico* pipeline we have identified 6,436 putative lncRNA transcripts in the liver fluke parasite, *Fasciola hepatica*, none of which are conserved with those previously described from *Schistosoma mansoni*. *F. hepatica* lncRNAs were distinct from *F. hepatica* mRNAs in transcript length, coding probability, exon/intron composition, expression patterns, and genome distribution. RNA-Seq and digital droplet PCR measurements demonstrated developmentally regulated expression of lncRNAs between intra-mammalian life stages; a similar proportion of lncRNAs (14.2%) and mRNAs (12.8%) were differentially expressed (p<0.001), supporting a functional role for lncRNAs in *F. hepatica* life stages. While most lncRNAs (81%) were intergenic, we identified some that overlapped protein coding loci in antisense (13%) or intronic (6%) configurations. We found no unequivocal evidence for correlated developmental expression within positionally correlated lncRNA:mRNA pairs, but global co-expression analysis identified five lncRNA that were inversely co-regulated with 89 mRNAs, including a large number of functionally essential proteases. The presence of micro (mi)RNA binding sites in 3135 lncRNAs indicates the potential for miRNA-based post-transcriptional regulation of lncRNA, and/or their function as competing endogenous (ce) RNAs. The same annotation pipeline identified 24,141 putative lncRNAs in *F. gigantica*. This first description of lncRNAs in *F. hepatica* provides an avenue to future functional and comparative genomics studies that will provide a new perspective on a poorly understood aspect of parasite biology.

Author summary

With increased understanding of genomes and transcriptomes it has become clear that not all RNAs code for protein, in contrast to the Central Dogma of molecular biology. Nevertheless, many such "non-coding" RNAs have functions in their own right. For example, long non-coding (lnc)RNAs comprise a large proportion of the human genome and control expression of some human protein coding genes. In parasites, lncRNAs

remain understudied and poorly understood. This paper provides the first description of the lncRNA complement of the liver fluke, *Fasciola hepatica*. We show that lncRNA expression changes as parasites develop, in a similar manner to protein-coding mRNA, suggesting important roles in parasite growth and development. Some lncRNAs show correlated developmental expression with mRNAs coding for important parasite proteins, suggesting that they may interact, suggesting regulatory functions for fluke lncRNAs. A larger subset of lncRNA may interact with micro (mi)RNAs, another type of non-coding RNA that are important regulators of gene expression in eukaryotes. These interactions are, again, important additions to our understanding of fluke lncRNA biology. Our ultimate goal is to discover ways to inhibit these interactions that could be used for new medicines for fluke infections which are a global scourge on the farming industry.

Introduction

Fasciola spp. liver fluke are helminth parasites that impact the health and productivity of farm animals, leading to considerable costs for international agricultural economies [1–3]. *Fasciola* also infects humans and is recognized as a zoonotic Neglected Tropical Disease (NTD) pathogen [4]. Fluke pathogenicity is compounded by the prevalence of fluke field populations resistant to key anthelmintics including albendazole, clorsulon, oxyclozanide, rafoxanide and triclabendazole [5,6] and increased prevalence driven by predictions of warmer, wetter, weather patterns associated with the climate crisis [7].

Efforts to develop new treatments for liver fluke infections have been supported by the development of systems-level experimental resources over the past decade, including genome, transcriptome, proteome and functional genomics methodologies for *F. hepatica* [8]. This increased volume and resolution of 'omics data has yielded insights into the non-coding (nc) RNAs of *Fasciola* genomes, where (mi)RNA complements are beginning to be clarified [9–20]., including a subset that appears to be secreted within extracellular vesicles (EVs). Transfer RNA fragments are the only other ncRNA to have been identified from *F. hepatica* transcriptomes [17], but eukaryotic transcriptomes also contain additional ncRNA families that remain poorly characterised. One of the most widely studied over recent years has been the long noncoding (lnc)RNAs, arbitrarily defined as transcripts longer than 200nt that do not code for protein. Interest in this family has largely been driven by perceived importance in human tumour biology [21–23], however increasing numbers of studies are beginning to describe these molecules in non-human animal systems, as well as plants, yeast and prokaryotes [24–29]. The effects of lncRNAs on expression of protein coding transcripts is well recognised, where lncRNAs can regulate expression of overlapping or adjacent protein coding transcripts [30,31].

Amongst lncRNA functions are roles in transcription, translation, protein localization, cellular integrity, the cell cycle, apoptosis, and stem cell pluripotency (for review see [32]). These functions were identified by RNA interference (RNAi) experiments, focusing on cytosolic lncRNAs in human systems [33,34], and increasingly, CRISPR-based perturbation of lncRNA expression in human cell lines [35], and mice [36]. This volume of functional data means that mammalian experimental systems remain our touchstone for understanding lncRNA biology. Functional genomics methods have been less widely applied to invertebrates, but RNAi has yielded functional data in arthropods [37,38] while CRISPR has yielded lncRNA functional insights in *C. elegans* [39,40]. The first lncRNA sequence datasets have now been published from parasitic flatworms, including *Schistosoma mansoni* and *S. japonicum* [41–45],

Echinococcus granulosus [46] and *Macrostomum lignano* [47], and now include the first reports of lncRNA functions in these species [48]. LncRNAs are now being recognised as potential therapeutic targets for new anti-parasitic drugs [49], and rapid progress is being made in development of lncRNA-targeting, oligonucleotide-based drugs for several human diseases [50,51].

LncRNAs are appealing drug targets because they tend to have tightly regulated expression in space and time, and appear to show low conservation between parasite and host [49,52]. These characteristics could imply a lower probability of off-target effects than protein-targeting drugs. LncRNAs could also be targeted by oligonucleotide-based drugs, which have a much lower development cost than small-molecule based therapies [53]. However, this appeal is mitigated somewhat by difficulties in formulating and administering oligonucleotide drugs to specific locations in the target organism–most are not efficiently absorbed by the intestine so must be delivered parenterally [52].

This paper aimed to provide the first description of long non-coding (lnc)RNAs in *F. hepatica*, and generate insights into fluke lncRNA biology. We achieved this aim, showing that putative lncRNAs are: (i) Developmentally regulated during intra-mammalian development of *F. hepatica*; (ii) Expressed in correlation with mRNAs, including proteases important for fluke survival, and (iii) Contain binding sites for miRNAs, suggesting that these two non-coding RNA families may have the capacity to interact in liver fluke. These data provide important insights into potential functions for lncRNAs and will form the basis for future functional genomics studies. Our hope is that these data will catalyse new avenues towards studying, and new perspectives on, liver fluke biology.

Methods

Transcriptome assemblies

Workflow was as described in Fig 1. We began by assembling a non-redundant transcriptome from available *F. hepatica* RNA-Seq datasets (18 biological replicate libraries). These included transcriptome data first published by Cwiklinski et al. [54], consisting of non-stranded, paired-end Illumina sequencing read sets from *F. hepatica* metacercariae (met; $n = 3$), newly excysted juveniles (NEJ) (1h (nej1h; $n = 2$), 3h (nej3h; $n = 2$), 24h (nej24h; $n = 2$) *in vitro*), *in vivo* liver stage parasites recovered from rats at 21 days post oral exposure (juv21d; $n = 1$), adult parasites (Ad; $n = 1$), and eggs (egg; $n = 1$). These fastq files were obtained from the European Nucleotide Archive (ENA) via project PRJEB6904. We supplemented these with stranded paired-end Illumina sequencing datasets for rat-derived juv21d worms (juv21d; $n = 3$), and 21 day *in vitro* maintained worms (invitro21d; $n = 3$). These six libraries are available from ENA project PRJEB49655, samples ERS9656160-ERS9656165, generated as described [19]. The ftp addresses for all libraries used are provided in S1 Dataset. *F. gigantica* lncRNA identification used sequencing datasets from NCBI BioProject PRJNA350370 (S1 Dataset) [55] and an 18h NEJ RNA-Seq dataset [56]. The latter is available from DDBJ/EMBL/GenBank accession GJHP01000000 and can be interrogated *via* https://sequenceserver.ibers.aber.ac.uk/.

Fastq files were mapped against the *F. hepatica* genome (WormBase ParaSite PRJEB25283, database version WBPS14) using HISAT2, with a merged transcript assembly generated through StringTie ([57]; using default parameters throughout). This produced 121,277 transcripts of ≧200 nt length (the default length cutoff within StringTie), which were passed into our lncRNA annotation pipeline (Fig 1). A separate stranded transcriptome was also produced by mapping the three [19] juv21d datasets against the *F. hepatica* genome as described above, and employing the—*fr* option in StringTie to denote a stranded library in "fr-firststrand" format, matching the dUTP-based synthesis method of the stranded library synthesis kit that we used [19].
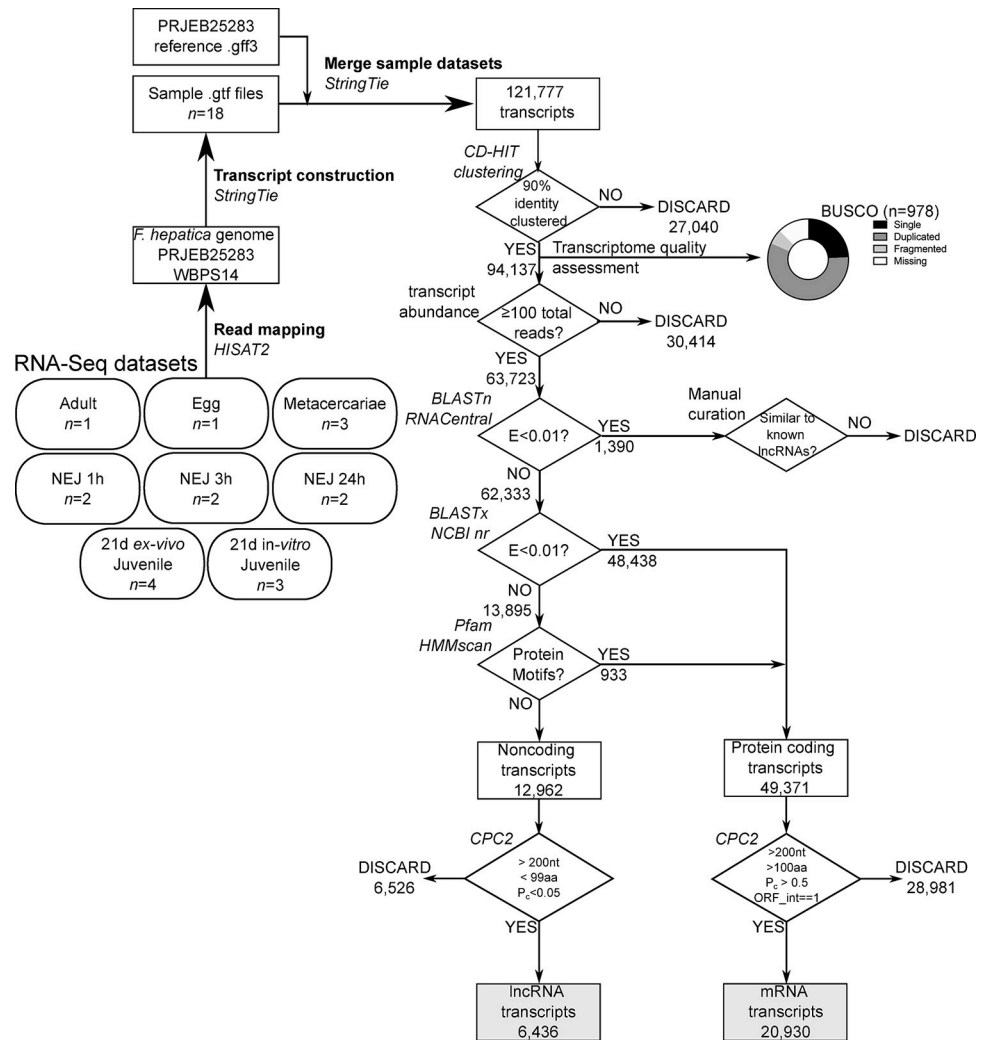
Fig 1. Long non-coding RNA discovery pipeline. Flowchart indicates processing and analysis of RNA-Seq Illumina short read datasets into lncRNA data. Numbers of sequences passing or failing each filter stage of the pipeline are indicated, software tools are named in italics.

Identification of *F. hepatica* lncRNAs

After removal of duplicate transcripts using CD-HIT (using the *cd-hit-est* program, with sequence identity threshold set to 0.9; [58]) (Fig 1), and application of an expression cutoff of ≥100 reads per transcript, contaminating non-coding sequences (ribosomal and transfer RNAs) were removed by comparison with the RNAcentral dataset [59]. Forty sequences possessing similarity (e value < 0.05) with known lncRNAs were replaced back into the dataset. The remaining 63,805 sequences were filtered for similarity with proteins using BLASTx, performed locally using DIAMOND BLAST [60] against the NCBI nr protein sequence database. Transcripts scoring E<0.01 (44,713 sequences) were filtered into a 'coding sequence' bin. Remaining sequences (19,092) were screened for protein motifs using Pfam hmmscan, performed locally, resulting in removal of an additional 455 transcripts with motif hits above the default inclusion threshold. The remaining 18,637 presumed non-coding sequences were all analysed with Coding Potential Calculator (CPC2 [61]) in a process of manual cross-checking. This final step ensured that our lncRNAs all had 'lncRNA-like' characteristics (>200nt length,

encoding ≤100 aa peptide, defined by CPC2 as "noncoding", with a coding probability of <0.05). Our final dataset contained 7,497 lncRNA transcripts from 6,321 genomic loci. Fig 1 also describes a parallel pipeline for identification of coding RNAs (presumed "mRNAs"), all of which were finally assessed by CPC2 as 'coding', with coding probability ≥0.95, with transcript length ≥200nt, and encoded peptides ≥100 aa), yielding 21,697 mRNA transcripts from 9,046 loci. GO analysis was performed for coding RNAs by searching gene IDs within the gene ontology resource (http://geneontology.org/).

Differential expression analyses

To identify transcripts with statistically significant differential expression (DE) between sequential life cycle stages, we used pairwise exact tests within the EdgeR package [62,63]. Only life stages supported by at least 2 biological replicates were used for these tests (pairwise tests performed: met vs nej1h; nej1h vs nej3h; nej3h vs nej24h; nej24h vs juv21d; juv21d vs met). Raw count data, recovered from StringTie using the prepDE.py script described in the StringTie manual, were piped into EdgeR for analysis. EdgeR output files were parsed using custom Python scripts for list comprehension, allowing extraction of fold change data supported by p≤0.001.

Digital droplet PCR

To confirm RNA-Seq transcript expression data, we performed digital droplet (dd)PCR [64] on six randomly selected lncRNAs across three life stages. Total RNA was extracted using Trizol (ThermoFisher) from one adult *F. hepatica* (*n = 3*), and from the same juv21d (*n = 3*) and invitro21d (*n = 3*) RNA samples used for transcriptome sequencing as described in section 2.1. Each cDNA was generated from 500ng total RNA using the High Capacity RNA-to-cDNA kit (Thermo Fisher). Amplification for ddPCR used the EvaGreen Supermix (BioRad), including 200 nM of each primer (Supporting file 11). Data analysis used the Quantasoft package from BioRad.

Genome localisation and expression correlation

For identification of antisense and intronic lncRNAs we used the bedtools package (v2.28; [65]. After generating bed files for transcripts, exons, introns and intergenic regions from our gtf file, we used bedtools *intersect* function to identify overlapping lncRNA and mRNA transcript loci, and sorted these by strand orientation (based on our stranded juv21d datasets) to identify antisense lncRNA:mRNA pairs. Bedtools *intersect* again allowed us to identify intronic lncRNAs by comparing lncRNA exons with mRNA introns, and again, these were sorted by strand orientation. Intergenic lncRNAs were identified as those that were not amongst our lists of antisense or intronic lncRNAs. We used the bedtools *closest* algorithm to identify the single closest upstream and downstream mRNA locus to each intergenic lncRNA locus.

Before examining expression correlation, we parsed all antisense, intronic and intergenic lncRNA:mRNA pairs to identify those pairs in which both members were DE, using a Python script for list comprehension. We then used the 'CORREL' function in MS Excel to calculate the correlation coefficient for lncRNA vs mRNA across TPM expression data from met, nej1h, nej3h, nej24h and juv21d libraries.

Weighted Gene Correlation Network Analysis (WGCNA) was performed using the CEMiTool R package [66].

miRNA response element identification

To examine whether any of our lncRNAs contained binding sites for miRNAs (miRNA response elements; MRE), we employed a consensus prediction method as described by [67].

Working with the 150 miRNAs previously described in *F. hepatica* [9–20], we employed three miRNA target prediction tools to identify binding matches between all *F. hepatica* miRNAs and our lncRNAs; we retained only those lncRNA:miRNA pairs that were identified by all three algorithms. We used local instances of miRANDA [68], RNAhybrid [69](and PITA [70], accepting hits fulfilling the thresholds as used by [67]: miRanda, total score >145, energy < -10; RNAhybrid, p<0.1, energy < -22; PITA, $\Delta\Delta G$ < -10. LncRNA:miRNA pairs identified by all three algorithms were extracted using a custom Python script, and only these consensus pairs were accepted.

Results

Dataset summary

Mapping of the >1.7 bn reads (81.7% overall alignment rate; S1 Dataset) associated with 18 RNA-Seq libraries yielded 121,777 transcripts, which clustered at 90% identity into 94,137 non-redundant transcripts (Fig 1). Using BUSCO [71], we compared the quality of our assembly with the *F. hepatica* transcript assemblies available through WormBase Parasite WBPS14. Our dataset contained more complete BUSCOs (81.7%; PRJNA179522 = 40.4%, PRJEB25283 = 74.1%), and had fewer missing BUSCOs (12.5%; PRJNA179522 = 31.3%, PRJEB25283 = 17.9%), than previous assemblies (Fig 1).

At this point, we filtered the transcriptome to include only sequences represented by at least 100 reads, leaving a total of 63,723 supported transcripts (S2 Dataset). From this dataset, we removed 1,390 sequences with similarity to other classes of ncRNA (ribosomal RNA, transfer RNA etc) listed in the RNACentral dataset. We then extracted protein coding transcripts (containing BLAST or Pfam identities, and a complete ORF) into a counterpart dataset containing 20,930 "mRNA" transcripts (S3 Dataset), which we used to compare and examine the transcriptional and syntenic relationships between lncRNA and mRNA loci. Having removed mRNAs, we extracted our putative lncRNAs (6,436 transcripts measuring ≥200 nt and lacking significant protein coding capacity; S4, S5, S6 Dataset files).

lncRNA and mRNA datasets are quantitatively and qualitatively distinct

The pipeline described in Fig 1 identified 6,436 transcripts fitting the definition of lncRNAs, and 20,930 transcripts with protein coding capacity, which we label here as mRNAs. These datasets were distinct on all measures examined (Fig 2)–lncRNA transcripts were shorter overall (median: lncRNA = 875 nt, mRNA = 3133 nt; Fig 2A), encoded shorter open reading frames (median: lncRNA = 49 aa, mRNA = 397 aa; Fig 2B), displayed lower protein coding probability (median: lncRNA P = 0.025, mRNA P = 1.000; Fig 2C), and were expressed more scarcely (median: lncRNAs 0.122 TPM, mRNA = 0.628 TPM; Fig 2D) than mRNA transcripts. Loci coding for lncRNAs were smaller than mRNA loci (median: lncRNA = 2010 bp, mRNA = 29,261 bp; Fig 2E), with longer exons (median: lncRNA = 319 bp, mRNA = 169 bp; Fig 2F) and shorter introns (median: lncRNA = 322 bp, mRNA = 1412 bp; Fig 2G). LncRNA loci incorporated fewer exons (median = 2 exons; Fig 2H), compared to mRNA loci (median = 7 exons; Fig 2I).

F. hepatica lncRNAs are dissimilar to lncRNAs from non-*Fasciola* species

To identify lncRNA orthologues in the closely related species, *F. gigantica*, we assembled an *F. gigantica* lncRNA dataset from available RNA-Seq datasets [23,56] using the pipeline described in Fig 1. This comprised 24,141 sequences (S7 Dataset). As described by Maciel *et al.* [45], we compared the *F. hepatica* and *F. gigantica* datasets using blastn, finding 2596 one-to-one blastn
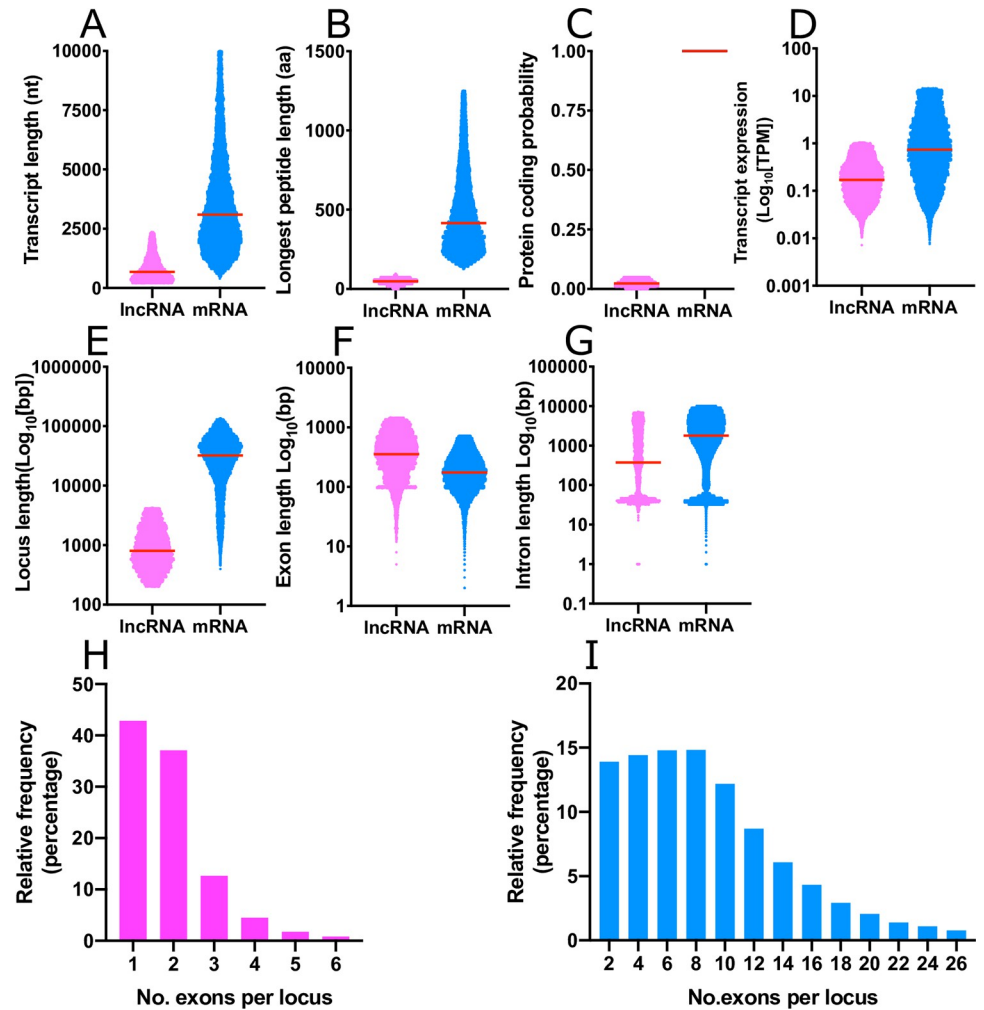
Fig 2. Statistical summary of mRNA and long noncoding (lnc)RNA annotations from *Fasciola hepatica*. LncRNAs and mRNAs are qualitatively and quantitatively distinct across all measures examined: (A) Transcript length; (B) Open reading frame length; (C) Probability of protein coding capacity; (D) Transcript expression/abundance; (E) Genomic locus length; (F) Exon length; (G) Intron length; and (H, I) Number of exons per locus. Each scatter graph (A-G) is composed of individual datapoints, with dataset median illustrated by a horizontal red bar. In all graphs, lncRNAs are magenta, mRNA are blue.

https://doi.org/10.1371/journal.pntd.0011663.g002

hits with an e-value cutoff of 1e-3 (S8 Dataset). We parsed these to a set of 522 *F. gigantica* lncRNAs matching an *F. hepatica* lncRNA at ≥90% sequence identity over ≥50% of the sequence length.

We also compared *F. hepatica* lncRNAs with published lncRNAs from other flatworms for where those sequences were openly available, using both CD-HIT and BLASTn. *F. hepatica* lncRNAs were compared with the >10,000 lncRNA transcripts reported from *S. mansoni* [41,72]. CD-HIT identified no evidence for lncRNA sequence similarity between *F. hepatica* and *S. mansoni*, while BLASTn identified a single, low-scoring orthologous pair, matching the *F. hepatica* lncRNA STRG.25709.1 with SmLINC02629-Ibu from *S. mansoni* (bitscore 56.5, E-value 4e[-07]). Likewise, BLASTn comparison with *M. lignano* lncRNAs [47] found no similarity, while comparisons with the RNAcentral non-coding RNA database found no similarity with the >200,000 mammalian lncRNAs within that dataset.

lncRNA transcripts are differentially expressed during development

Differential expression (DE) analyses employed only life stages represented by at least two biological replicate libraries (met, nej1h, nej3h, nej24h, juv21d, invitro21d). Fig 3 and S9 Dataset show that all of these libraries bore stage-specific lncRNAs (met = 4; nej1h = 4; nej3h = 3; nej24h = 1; juv21d = 63; invitro21d = 13), and mRNAs (met = 42; nej1h = 35; nej3h = 45; nej24h = 30; juv21d = 225; invitro21d = 97). Using edgeR's exact test algorithm, pairwise comparisons were performed between developmentally sequential life stages, defining DE transcripts as those with a statistically significant difference (p≤0.001) in at least one of these comparisons (Fig 3A and S10 Dataset). A total of 911 DE lncRNA transcripts (Fig 3B; 14.2% of
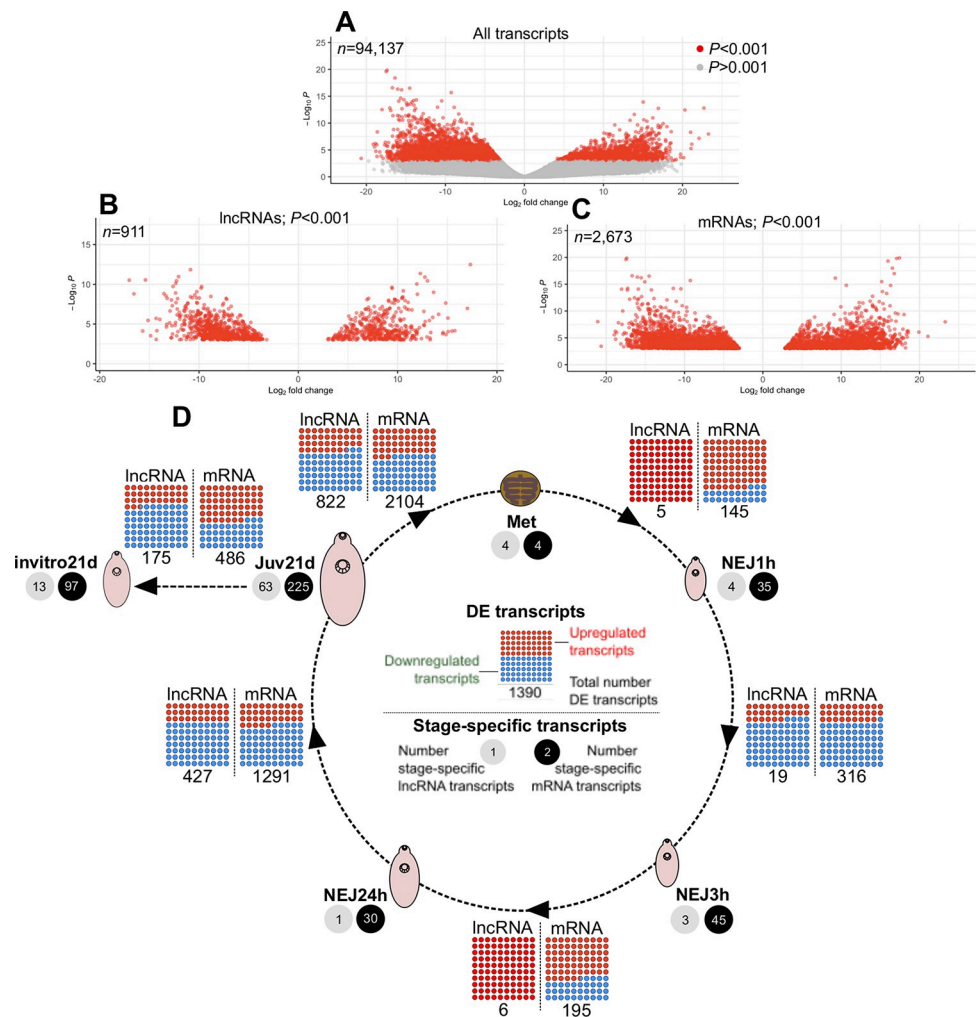

Fig 3. Developmental regulation of long non-coding (lnc)RNAs in intra-mammalian stages of *Fasciola hepatica*. Statistically significant differential expression (DE) was determined for lncRNA and mRNA transcripts during developmental transitions, showing that F. hepatica lncRNAs are dynamically-regulated during intra-mammalian development. A-C. Distribution of transcript fold change vs P value across all comparisons. A, all transcripts, with P-value cutoff as indicated on y-axis; B, Fold change distribution of DE lncRNAs; C, Fold change distribution of DE mRNAs. D, Summary data from DE analysis of transcripts in metacercariae (Met), newly excysted juvenile (NEJ) maintained in vitro for 1h post excystment (NEJ1h), 3h post excystment (NEJ3h) or 24h post excystment (NEJ24h), and ex vivo parasites recovered from rat livers at 21 days post infection (Juv21d). Numbers of stage-specific lncRNAs and mRNAs at each life stage are indicated, as are the numbers of DE lncRNA and mRNA transcripts, and the proportion of each upregulated or downregulated, between each developmental transition.

https://doi.org/10.1371/journal.pntd.0011663.g003

all lncRNAs), and 2,673 DE mRNA transcripts (Fig 3C; 12.8% of all mRNAs) were identified. DE transcripts were found in every pairwise comparison (Fig 3D). To confirm these data, six randomly selected lncRNAs were tested for expression across adult, juv21d and invitro21d life stages using ddPCR; in most cases the ddPCR expression pattern correlated well with the NGS expression pattern (S11 Dataset).

Positional relationships and expression correlations between lncRNAs and mRNAs

Based on genomic location and directionality relative to protein-coding loci, we identified three lncRNA types in *F. hepatica*: (i) antisense lncRNAs: expressed from a lncRNA exon on the opposite DNA strand to a protein-coding locus, and overlapping a protein-coding exon by at least 1bp; (ii) intronic lncRNAs: expressed from lncRNA exons in either orientation that reside within an intron of a protein coding locus and do not overlap an exon; (iii) intergenic lncRNAs: expressed from distinct loci that do not intersect with any protein coding locus. To delineate our lncRNAs into these categories, we used the bedtools software package, and stranded juv21d libraries (as described in Methods). Our first approach was to search for sense:antisense overlaps between lncRNA and mRNA transcripts using the bedtools *intersect* algorithm, which yielded 536 non-redundant lncRNA:mRNA transcript pairs, overlapping in opposite transcriptional orientations. The same approach identified 534 similarly oriented mRNA:mRNA pairs (i.e. transcripts in antisense orientation, where both strands encode protein), and 17 lncRNA:lncRNA pairs (i.e. transcripts in antisense orientation, where neither strand encodes protein) (S12 Dataset). In every case but one, antisense lncRNAs overlapped protein coding exons by a minimum of 94 nt, with a median overlap of 3343 nt (S12 Dataset).

Parsing the data to identify antisense pairs in which both members showed statistically significant differential expression (DE) identified 14 DE lncRNA:mRNA pairs and five DE mRNA:mRNA pairs. There were no lncRNA:lncRNA pairs in which both members were DE (S12 Dataset). For lncRNA:mRNA pairs we calculated the Pearson correlation coefficient (CC) of lncRNA vs mRNA expression (TPM) across all life stages (S12 Dataset). Accepting >0.99 and <-0.99 as cut-offs identified just seven pairs of "expression correlated" lncRNA:mRNA transcript pairs, where seven of these showed positive correlation, while two showed negative correlation.

After removing antisense transcripts from the dataset, we identified 30 intronic lncRNA:mRNA pairs (median CC = 0.407; S13 Dataset). No pairs consisted of both members as DE, and none passed the +/-0.9 CC cut-off. Finally, we identified 3795 intergenic lncRNAs where the closest transcript was an mRNA (S14 Dataset). In 29 pairs both members were DE, and only three pairs passed the +/-0.9 CC cut-off.

To explore *trans* interactions in more detail, we employed a weighted correlation network analysis approach. Here, we sought evidence for co-expression between all lncRNAs and all mRNAs across all libraries with multiple biological replicates. CEMiTool identified a single co-expression module, containing 89 mRNAs and five lncRNAs (Fig 4 and S15 Dataset). The mean TPM of the lncRNAs and mRNAs within this module were inversely correlated across longitudinally sampled libraries (Fig 4), suggesting that they may be co-regulated. Interestingly, 47% of mRNA transcripts within the module were associated with peptidase activity, including Cathepsin L and B, and legumain, potentially linking five lncRNAs with regulation of these important virulence genes.

F. hepatica lncRNAs contain miRNA binding sites

LncRNAs contain binding sites for miRNAs, implying either that they serve as "sponges" for "soaking up" individual miRNAs, or that their expression is regulated by miRNA binding and
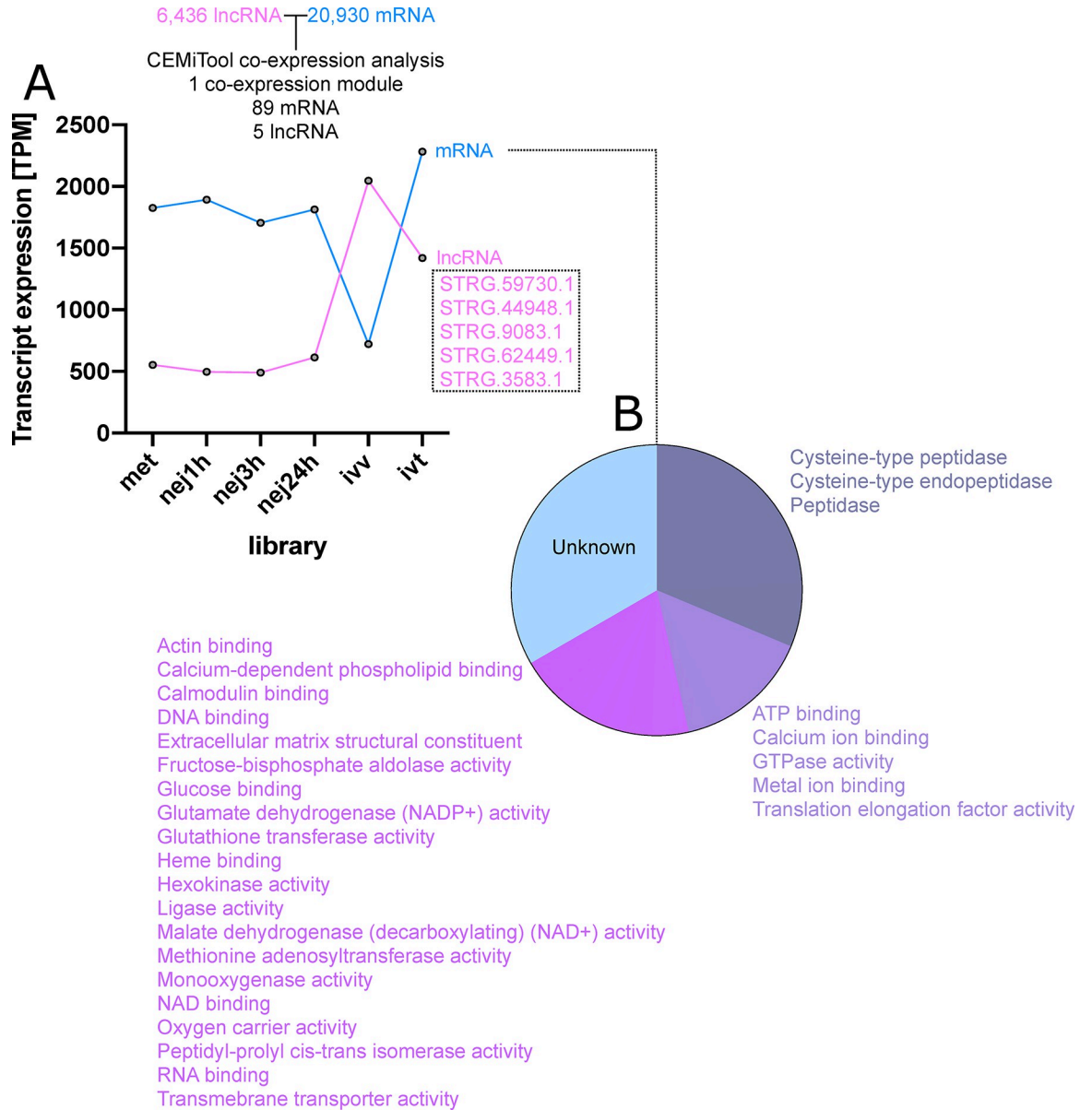
Fig 4. Co-expression of long noncoding (lnc)RNAs and mRNAs in *Fasciola hepatica*. (A) Global co-expression analysis of all lncRNAs and mRNAs in our dataset. CEMiTool identified one co-expression module in which five lncRNAs and 89 mRNAs show inversely correlated expression across life stage RNA-Seq libraries. (B) The 89 mRNAs comprise mostly cysteine proteases in addition to a range of other sequences as indicated. The remaining one third of matching mRNAs code for unknown sequence types that lacked BLAST homology and identifiable sequence motifs.

https://doi.org/10.1371/journal.pntd.0011663.g004

transcriptional destruction. Previous reports have identified at least 150 miRNAs in *F. hepatica* [9–20]. Using three miRNA binding prediction tools (miRANDA, PITA, RNAhybrid) we analysed these miRNAs in the context of binding to our lncRNAs, producing a consensus set of 4104 lncRNA:miRNA pairs that were present in outputs from all three tools (we rejected any matches that were not predicted by all three tools). S16 Dataset shows that these matches incorporated all 150 miRNAs, and 2618 lncRNAs. Individual lncRNAs contained binding sites for up to eight distinct miRNAs, with each miRNA predicted to bind up to 234 lncRNAs.
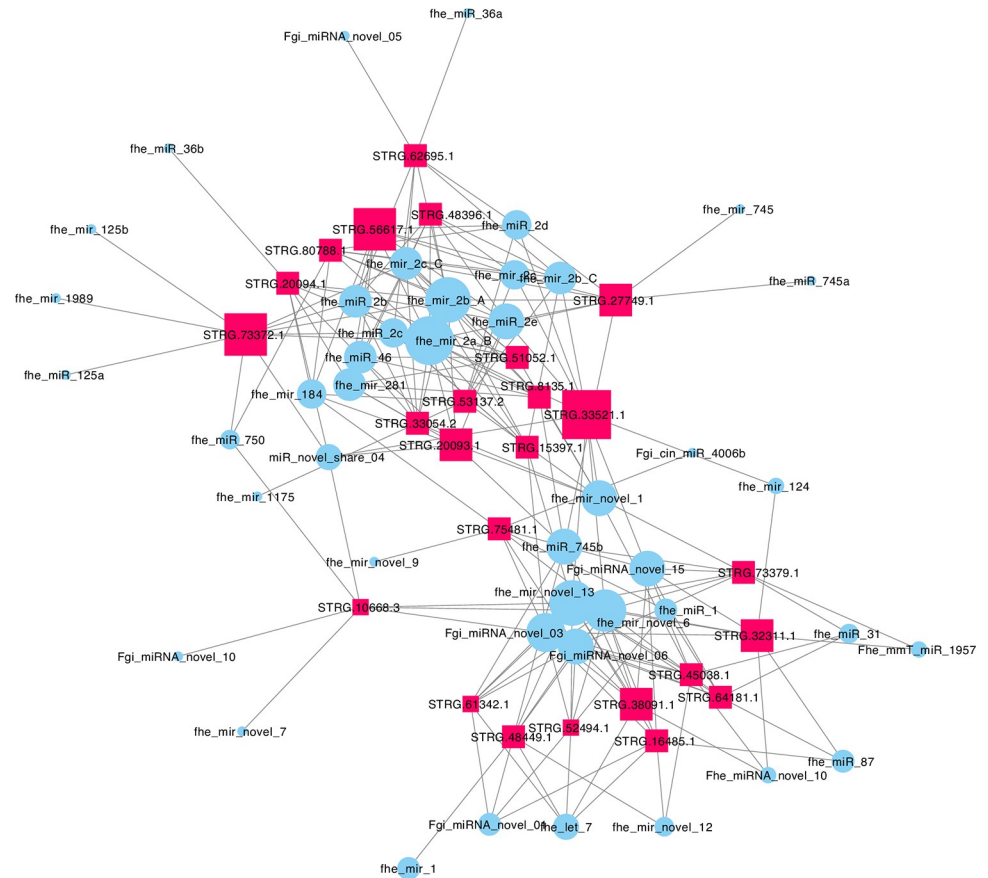
Fig 5. Network representation of computationally predicted binding interactions between *Fasciola hepatica* long noncoding (lnc)RNAs and micro (mi)RNAs. This network represents an excerpt of the lncRNA:miRNA network detailed in S16 Dataset, showing only the lncRNAs that matched eight or more miRNAs. LncRNA nodes are represented by red squares, miRNA nodes by blue circles. Node size is proportional to the number of connections with other nodes. Connections between nodes are represented by grey edges.

https://doi.org/10.1371/journal.pntd.0011663.g005

Output data from PITA illustrated the number of individual binding sites per individual lncRNA:miRNA pair: while 95% of lncRNAs contained ten or fewer binding sites, 5% displayed up to 246. Fig 5 shows an excerpt of this lncRNA:miRNA network, focusing on the lncRNAs matching eight or more miRNAs. Within this network, STRG.61441.1 was the major hub lncRNA, binding to eight miRNAs. Fhe-pubnovelmiR-7 was the major hub miRNA, binding to 234 lncRNAs.

Discussion

Fasciolosis, caused by *F. hepatica* and *F. gigantica*, is an important veterinary and zoonotic disease which requires improved diagnostic and control approaches if it is to be sustainably combatted in the long term. A key approach to identification of new control targets and diagnostic biomarkers is through increased understanding of fundamental parasite biology and host-parasite interactions. This study provides a solid foundation of data describing lncRNAs in *F. hepatica* for the first time, through which this need can be addressed by ongoing research. This study has performed the first classification of lncRNAs in *F. hepatica*, linking these to potential roles in parasite development and co-regulation of miRNAs and mRNAs. This work represents an essential precursor to functional understanding of liver fluke lncRNA biology.

Unlike other systems, *F. hepatica* does not yet have a unified description of the protein-coding transcriptome that would have enabled a simple subtraction-based approach to lncRNA discovery. Disparities can be seen in the varying gene complements presented by currently available studies [54,73,74]. This meant that we needed to generate a consensus transcriptome before filtering lncRNAs from mRNAs within those datasets. Our consensus transcriptome combined 18 publicly available and newly generated libraries, which together encompassed 94,137 non-redundant transcripts. This represented by far the largest transcriptome yet described for *F. hepatica*, with previous datasets describing 22,676 [54,73], 16,806 (WormBase ParaSite release WBPS14), or 14,462 [74] transcripts, although it is unclear whether those previously published datasets retained non-coding sequences. BUSCO metrics confirmed that our transcriptome assembly was the most complete yet published for *Fasciola*, giving us confidence in proceeding to lncRNA discovery. This transcriptome was then filtered to include only sequences represented by at least 100 reads (an arbitrary selection which aimed to balance well-supported transcript models while maintaining adequate detection of rare transcripts), followed by removal of irrelevant classes of ncRNA, and protein coding transcripts. Our final pool of putative lncRNAs represented 6,436 transcripts measuring ≥200 nt and lacking significant protein coding capacity. A separate pool of protein-coding transcripts contained 20,930 mRNAs.

To identify lncRNA transcripts that we may have mis-annotated, or those from published datasets that may have been erroneously annotated as protein-coding, we used BLAST to compare all lncRNAs with previously published *F. hepatica* transcript annotations. BLASTx did not identify any identical matches between our lncRNAs and previously predicted proteins. BLASTn analyses identified 37 lncRNAs matching 20 sequences previously annotated as mRNAs. However, we found that these matching sequences had either no identifiable protein sequence characteristics and/or were classed as hypothetical proteins. We therefore retained these in our lncRNA dataset as putative lncRNAs.

Previous literature from other species showed that lncRNAs were distinct from mRNAs across a wide range of measures [37,75–77]. We confirmed that these findings also applied to our dataset, showing that lncRNAs were shorter than mRNAs in gene locus length, overall transcript length and open reading frame length. LncRNA loci also tended to display fewer, longer exons and shorter introns than mRNA loci. LncRNAs were also expressed more scarcely than mRNA transcripts. Inter-species comparisons of nucleic acid sequences can provide insights into evolution, with sequence conservation potentially indicating functional conservation. We used public datasets and our pipeline to identify lncRNAs in *F. gigantica*. Comparison with *F. hepatica* lncRNAs identified at least 522 orthologues (defined as those showing ≥ 90% sequence identity across ≥50% of the aligned sequence). This suggests common evolutionary origin of lncRNAs between these two species although the majority of lncRNAs in both species appeared to be species specific.

There are currently limited opportunities to make this comparison for lncRNAs with other flatworms, with *S. mansoni*, *S. japonicum*, *E. granulosus* and *M. lignano* the only other flatworm lncRNA datasets currently available [41–47]. In comparing *F. hepatica* lncRNAs with these datasets we found an almost complete lack of lncRNA sequence conservation. This absence of primary sequence similarity between genera is not surprising and suggests that, like in other organisms, the rapid evolution of lncRNAs invalidates primary sequence similarity as a means of identifying lncRNA homologs [78–80]. Similarly, BLASTn comparisons with the RNAcentral non-coding RNA database found no similarity between our lncRNAs and the >200,000 mammalian lncRNAs within that dataset. This distinction between host and parasite lncRNA sequences could provide a source of new diagnostic biomarkers, if future work can show that lncRNAs are secreted by fluke or other helminths, as is the case for miRNAs [81].

Rather than primary sequence similarity, lncRNAs are thought to be conserved between species according to spatiotemporal and syntenic locus expression [82]. Indeed, Maciel et al. [45] showed for the first time that synteny conservation can be used for lncRNA discovery in flatworms, where 14% of *S. japonicum* intergenic lncRNAs had conserved synteny with those in *S. mansoni*. Other studies have also shown lncRNA sequence conservation between *S. mansoni*, *S. japonicum* and *S. haematobium*, albeit with lower similarity than that seen in mRNA comparisons, and using small sample sizes and an older version of the *S. mansoni* genome [42]. Syntenic comparisons have not yet been performed between flatworm genera, but undoubtedly represent a necessary future direction for lncRNA discovery in flatworms.

Differential expression (DE) analyses employed only life stages represented by at least two biological replicate libraries (met, nej1h, nej3h, nej24h, juv21d, invitro21d). All these libraries bore stage-specific lncRNAs and mRNAs, with 911 DE lncRNA transcripts and 2,673 DE mRNA transcripts identified in pairwise comparisons between sequential lifestages. DE transcripts were found in all pairwise comparisons, with the number of DE transcripts roughly proportional to the time-course length between stages–the fewest between met:nej1h (5 lncRNA, 145 mRNA), and the most between juv21d:met (822 lncRNA, 2104 mRNA). Comparison of juv21d with invitro21d samples (respectively, flukes recovered from rat livers 21 days after infection, or maintained *in vitro* for 21 days [19,83], demonstrated the presence of 13 lncRNAs found uniquely in invitro21d samples, with 63 found uniquely in juv21d worms. Given the key developmental differences between these groups [19], these lncRNAs have potential importance for fluke development, and represent priority targets for functional genomics experiments. The strikingly similar proportions of DE lncRNAs and mRNAs (respectively, 14.2% and 12.8%), shows that lncRNAs are at least as transcriptionally dynamic as coding RNAs, and supports the hypothesis that they have essential and stage-specific functions during the *F. hepatica* intra-mammalian life cycle.

While our focus was not to reannotate the mRNA transcriptome of *F. hepatica*, our lncRNA annotation method did also generate protein-coding mRNAs and associated DE data. Briefly, the most highly regulated mRNA transcripts ($\geq$16 fold) in our dataset closely reflect previous observations [54,73], and include sequences within GO processes describing cell adhesion, cell division, glycolysis, metal ion homeostasis, muscle function, proteolysis, protein synthesis and modification, RNA transcriptional control and signal transduction.

The purpose of these data was to identify and explore potential interactions between lncRNAs and mRNAs. Various classification schemes exist for lncRNAs, but one of the most commonly used groups sequences based on their genomic location and directionality relative to protein-coding loci [82]. According to this classification scheme, we used a subtractive approach to identify antisense, intronic and intergenic lncRNA types in *F. hepatica*. These classifications are important because they can inform potential interactions with protein coding genes, for example *cis* natural antisense transcripts (NATs) and intronic lncRNAs can affect the expression of corresponding sense transcripts from their "host" gene, while *trans*-antisense and intergenic lncRNAs can impact expression of distant genomic loci. Our searches yielded 536 antisense overlapping lncRNA and mRNA transcript pairs (where the mRNA was considered "sense"). These data also identified 534 complementary mRNA:mRNA pairs, and 17 complementary lncRNA:lncRNA pairs. These data show that lncRNA and/or mRNA loci may overlap on the genome, to produce pairs of antisense/complementary-oriented transcripts. The mRNA:mRNA pairings in particular are relevant to RNA interference (RNAi) experiments, where design of double stranded (ds)RNA triggers to a target gene on one of these strands could lead to unintended off-target effects on the opposite-strand gene.

Given evidence that lncRNAs can interact transcriptionally in *cis* or *trans* fashion with mRNAs, we looked for evidence of transcriptional interaction between antisense transcripts.

For example, data from human cell lines show that antisense lncRNAs can regulate the expression of the mRNAs with which they overlap [30,31]. We reasoned that correlated expression might be most clearly identified in cases where both members of an antisense pair were DE. Parsing the data to identify pairs in which this was the case identified 14 DE lncRNA:mRNA pairs. We also found five DE mRNA:mRNA pairs, but no lncRNA:lncRNA pairs in which both members were DE. In case these criteria were prohibitively restrictive, we also calculated the Pearson correlation coefficient (CC) of lncRNA vs mRNA expression (TPM) for all 536 pairs across all life stages. Accepting >0.99 and <-0.99 as cut-offs identified just 53 pairs of DE lncRNA/mRNA transcript pairs that were highly correlated. These similar data suggested that correlated expression between *cis* oriented antisense transcripts might not be a widespread phenomenon, but we cannot rule out the possibility that some antisense lncRNA regulation of *cis* mRNA transcripts may occur. Functional genomics experiments will be necessary to test hypotheses around linkages between individual transcripts, and separate interacting transcripts from physically overlapping transcripts that do not interact [84–87]. As well as informing studies on evolution of the *F. hepatica* genome, these data have important practical implications in avoiding off-target effects in RNAi or CRISPR/Cas9 experimental design.

After removing these antisense transcripts from the dataset, we used similar methods to identify 30 intronic lncRNA:mRNA pairs (i.e. lncRNAs located within an mRNA intron in either orientation). No pair consisted of both members as DE, and none passed the +/-0.9 CC cut-off. Finally, we identified 3795 intergenic lncRNAs where the closest transcript was an mRNA (Supporting file 12). In 29 pairs both members were DE, and only three pairs passed the +/-0.9 CC cut-off. Further exploration of these data could focus on intronic lncRNAs as a potential source of long interspersed element (LINE) type transposable elements (TEs).

These findings differ from those from other systems, where lncRNAs have been shown to affect transcription of antisense overlapping and neighbouring protein-coding genes [30,31,88]. This prompted us to explore *trans* lncRNA interactions in more detail, where lncRNAs may regulate distant mRNA loci through epigenetic interaction [89]. We approached this using weighted correlation network analysis, via the CEMiTool R package, focusing broadly on co-expression between all lncRNAs and all mRNAs across all libraries in our datasets. This identified a single co-expression module, containing 89 mRNAs and five lncRNAs, which showed inversely correlated expression suggesting interaction and/or co-regulation. Interestingly, 47% of mRNA transcripts within the module were associated with peptidase activity, including Cathepsin L and B, and legumain (Fig 4). This is unsurprising given the preponderance of proteases amongst the most highly regulated *F. hepatica* transcripts [54,73]; these data represent the first link between lncRNAs and potential co-regulation of protease expression in *F. hepatica*. Future experiments should cement this link by silencing one or more of the lncRNAs and assaying for fluctuation in linked protease transcripts. Interrupting this regulatory interaction with lncRNA-targeting, oligonucleotide-based drugs [50,51], could represent a route towards new therapeutics for liver fluke infections. This idea is supported by the existence of licensed nucleic-acid-based drugs for several human conditions. Antisense oligonucleotide therapeutics, of the type that could be developed to target lncRNAs, include fomivirsen, which was used to treat cytomegalovirus retinitis [90] before being superceded by newer therapies, tofersen, which inhibits translation of superoxide dismutase in amytrophic lateral sclerosis [91] and eteplirsen, which causes exon skipping in the dystrophin transcript, used in therapy of Duchenne muscular dystrophy [92].

One of the most widespread hypotheses for lncRNA function is the competing endogenous (ce)RNA hypothesis, which holds that lncRNAs contain miRNA binding sites (known as miRNA Response Elements, or MREs), enabling them to act as 'sponges' for miRNAs. This 'sponging' phenomenon is thought to promote competition for miRNA binding with cognate

mRNA targets, enabling fine control of miRNA regulation of mRNA target transcripts [93,94]. Alternatively, the presence of miRNA binding sites may simply indicate that lncRNAs are transcriptionally regulated by miRNA binding in the same manner as mRNAs [95]. Since both possibilities could yield important insights into liver fluke lncRNA biology, we used in silico tools to explore the presence of miRNA binding sites on lncRNAs.

One hundred and fifty miRNAs have been previously reported in *F. hepatica* [9–20]. We generated in silico binding predictions between these miRNAs and our lncRNAs, producing a consensus set of 4104 lncRNA:miRNA pairs. These matches incorporated all 150 miRNAs, and 2618 lncRNAs. Individual lncRNAs contained binding sites for up to eight distinct miRNAs, with each miRNA predicted to bind up to 234 lncRNAs. Output data from PITA illustrated the number of individual binding sites per individual lncRNA:miRNA pair: while 95% of lncRNAs contained ten or fewer binding sites, the remaining 5% displayed up to 246. Network analysis identified STRG.61441.1 as the major hub lncRNA, binding to eight miRNAs. Fhe-pubnovelmiR-7 was the major hub miRNA, binding to 234 lncRNAs. These data support the possibility of miRNA-lncRNA interactions, that might manifest as either traditional miRNA-driven post-transcriptional regulation of lncRNA expression, or a ceRNA function for the described miRNAs. Experimental evidence will be required to test these hypotheses further. If supported, these could provide an additional route to new therapies similar to those described for lncRNAs above; indeed miRNA inhibiting drugs are in clinical trials for many conditions including haemophilia, hepatitis B, hypercholesterolaemia, nonalcoholic fatty liver disease and various cancers [96], suggesting the possibility of translating these approaches into the anthelmintic sphere.

This study has profiled the first set of lncRNAs in the liver fluke, *F. hepatica*. These non-coding RNAs were expressed across multiple intra-mammalian developmental stages, showing dynamic regulation between life-stages that suggests life-stage specific functions. *In silico* analyses supported important roles for lncRNAs in transcriptional regulation including: (i) an inverse correlation of lncRNA expression with mRNAs, suggesting co-regulation of these sequences, and; (ii) the widespread location of miRNA binding sites on lncRNAs, suggesting miRNA regulation of lncRNA, or vice versa. These data represent a steppingstone towards an understanding of non-coding RNA biology in *F. hepatica*, an area which remains poorly understood across eukaryotes but could expose new therapeutic and diagnostic options for parasite infections globally.

Supporting information

S1 Dataset. Hisat2 alignment statistics of Illumina RNA-Seq reads during mapping to *Fasciola hepatica* genome (PRJEB25283, WBPS14).
(DOCX)

S2 Dataset. Mapped assembled transcript sequences each supported by at least 100 RNA-Seq reads. Entire dataset.
(ZIP)

S3 Dataset. Mapped assembled transcript sequences each supported by at least 100 RNA-Seq reads. Protein-coding sequences only.
(TXT)

S4 Dataset. Mapped assembled transcript sequences each supported by at least 100 RNA-Seq reads. Long non-coding RNA sequences only.
(TXT)

S5 Dataset. BLASTx comparison of predicted long non-coding RNA sequences with predicted proteins from the *Fasciola hepatica* genome (WBPS14 PRJEB25283).
(XLSX)

S6 Dataset. BLASTn comparison of predicted long non-coding RNA sequences with mRNA transcripts from the *Fasciola hepatica* genome (WBPS14 PRJEB25283).
(XLSX)

S7 Dataset. BLASTn comparison of *Fasciola hepatica* long non-coding RNA sequences with a *Fasciola gigantica* transcriptome (Davey et al., 2022) [56].
(TXT)

S8 Dataset. Long non-coding RNA sequences identified from a *Fasciola gigantica* transcriptome (Davey et al., 2022) [56].
(XLSX)

S9 Dataset. Long non-coding RNAs and mRNAs expressed in only one *Fasciola hepatica* life stage.
(XLSX)

S10 Dataset. Differentially expressed long non-coding RNAs and mRNAs in comparisons between sequential developmental stages.
(XLSX)

S11 Dataset. Comparison of expression level of six long non-coding (lnc)RNAs measured with next generation sequencing (NGS) (blue), and digital droplet (dd)PCR (red). Each lncRNA was measured across three life-stage libraries: Adult *Fasciola hepatica*, 21 day in vitro juvenile *F. hepatica* (ivt) and 21 day in vivo juvenile *F. hepatica* (ivv). Each datapoint represents the mean±SEM of at least three biological replicates. Under each graph, the primer sets used for ddPCR amplification are indicated.
(PDF)

S12 Dataset. Overlapping antisense oriented long non-coding RNA vs mRNA pairs. Note that there are three tabs comparing: 1. Antisense lncRNA vs mRNA; 2. Antisense mRNA vs mRNA; 3. Antisense lncRNA vs lncRNA.
(XLSX)

S13 Dataset. Intronic long non-coding RNA vs mRNA pairs.
(XLSX)

S14 Dataset. Intergenic long non-coding RNA vs mRNA pairs.
(XLSX)

S15 Dataset. Co-expression module containing long non-coding RNA and mRNA transcripts.
(XLSX)

S16 Dataset. Consensus predictions of interactions between Fasciola hepatica long non-coding RNA and microRNA sequences.
(XLSX)

Author Contributions

Conceptualization: Paul McVeigh, Aaron G. Maule.

References

1. Spithill TW, Smooker PM, Copeman DB. Fasciola gigantica: Epidemiology, control, immunology and molecular biology. In: *Fasciolosis*. CABI; 1999. pp. 465–525.

2. Gray G.D., Copland R.S. and Copeman D.B. (eds) 2008. Overcoming liver fluke as a constraint to ruminant production in South-East Asia. ACIAR Monograph No. 133, pp. 155.

3. NADIS. Liver fluke control in sheep. 2016. Available from: https://www.nadis.org.uk/disease-a-z/sheep/liver-fluke-control-in-sheep/.

4. Hotez PJ, Brindley PJ, Bethony JM, King CH, Pearce EJ, Jacobson J. (2008). Helminth infections: the great neglected tropical diseases. *J Clin Invest* 118(4), 1311–1321. https://doi.org/10.1172/JCI34261 PMID: 18382743

5. Kelley JM, Elliott TP, Beddoe T, Anderson G, Skuce P, Spithill TW. Current Threat of Triclabendazole Resistance in Fasciola hepatica. Trends Parasitol. 2016; 32(6), 458–469. https://doi.org/10.1016/j.pt.2016.03.002 PMID: 27049013

6. Fairweather I, Brennan GP, Hanna REB, Robinson MW, Skuce PJ. Drug resistance in liver flukes. Int J Parasitol Drugs Drug Resist. 2020; 12, 39–59. https://doi.org/10.1016/j.ijpddr.2019.11.003 PMID: 32179499

7. Alba A, Vazquez AA, Hurtrez-Bousses S. Towards the comprehension of fasciolosis (re-)emergence: an integrative overview. Parasitology 2021; 148(4), 385–407. https://doi.org/10.1017/S0031182020002255 PMID: 33261674

8. McVeigh P., McCusker P., Robb E., Wells D., Gardiner E., Mousley A., et al. (2018). Reasons to Be Nervous about Flukicide Discovery. *Trends Parasitol* 34(3), 184–196. https://doi.org/10.1016/j.pt.2017.11.010 PMID: 29269027

9. Xu MJ, Ai L, Fu JH, Nisbet AJ, Liu QY, Chen MX, et al. Comparative characterization of microRNAs from the liver flukes Fasciola gigantica and F. hepatica. Plos One 2012; 7(12). https://doi.org/10.1371/journal.pone.0053387 PMID: 23300925

10. Fontenla S, Dell'Oca N, Smircich P, Tort JF, Siles-Lucas M. The miRnome of Fasciola hepatica juveniles endorses the existence of a reduced set of highly divergent micro RNAs in parasitic flatworms. Int J Parasitol. 2015; 45(14), 901–913. https://doi.org/10.1016/j.ijpara.2015.06.007 PMID: 26432296

11. Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL. A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. Annu Rev Genet. 2015; 49, 213–242. https://doi.org/10.1146/annurev-genet-120213-092023 PMID: 26473382

12. Guo X, Guo A. Profiling circulating microRNAs in serum of Fasciola gigantica-infected buffalo. Mol Biochem Parasitol. 2019; 232, 111201. https://doi.org/10.1016/j.molbiopara.2019.111201 PMID: 31377228

13. Ovchinnikov VY, Kashina EV, Mordvinov VA, Fromm B. EV-transported microRNAs of Schistosoma mansoni and Fasciola hepatica: Potential targets in definitive hosts. Infect Genet Evol. 2020; 85, 104528. https://doi.org/10.1016/j.meegid.2020.104528 PMID: 32891875

14. Ricafrente A, Nguyen H, Tran N, Donnelly S. An Evaluation of the Fasciola hepatica miRnome Predicts a Targeted Regulation of Mammalian Innate Immune Responses. Front Immunol. 2021; 11, 608686. https://doi.org/10.3389/fimmu.2020.608686 PMID: 33584684

15. Ricafrente A, Cwiklinski K, Nguyen H, Dalton JP, Donnelly S. Stage-specific miRNAs regulate gene expression associated with growth, development and parasite-host interaction during the intra-mammalian migration of the zoonotic helminth parasite Fasciola hepatica. BMC Genomics 2022; 23(1):419. https://doi.org/10.1186/s12864-022-08644-z PMID: 35659245

16. Herron CM, O'Connor A, Robb E, McCammick E, Hill H, Marks NJ, Robinson MW, Maule AG, McVeigh P. Developmental regulation and functional prediction of microRNAs in an expanded Fasciola hepatica miRNome. Front Cell Infect Microbiol. 2022; 12:811123. https://doi.org/10.3389/fcimb.2022.811123 PMID: 35223544

17. Hu RS, Zhang XX, Ma QN, Elsheikha HM, Ehsan M, Zhao Q. Differential expression of microRNAs and tRNA fragments mediate the adaptation of the liver fluke Fasciola gigantica to its intermediate snail and definitive mammalian hosts. Int J Parasitol. 2021; 51(5), 405–414. https://doi.org/10.1016/j.ijpara.2020.10.009 PMID: 33513403

18. Tran N, Ricafrente A, To J, Lund M, Marques TM, Gama-Carvalho M, Cwiklinski K, Dalton JP, Donnelly S. Fasciola hepatica hijacks host macrophage miRNA machinery to modulate early innate immune responses. Sci Rep. 2021; 11(1), 6712. https://doi.org/10.1038/s41598-021-86125-1 PMID: 33762636

19. Robb E, McCammick E, Wells D, McVeigh P, Gardiner E, Armstrong R, et al. Transcriptomic analysis reveals a role for the nervous system in regulating growth and development of Fasciola hepatica juveniles. PLoS Negl Trop Dis. 2022; 16(11):e0010854. https://doi.org/10.1371/journal.pntd.0010854

20. Fontenla S, Langleib M, de la Torre-Escudero E, Dominguez MF, Robinson MW, Tort J. Role of Fasciola hepatica small RNAs in the interaction with the mammalian host. Front Cell Infect Microbiol. 2022; 11, 812141. https://doi.org/10.3389/fcimb.2021.812141 PMID: 35155272

21. Cossu AM, Mosca L, Zappavigna S, Misso G, Bocchetti M, De Micco F, et al. Long Non-coding RNAs as Important Biomarkers in Laryngeal Cancer and Other Head and Neck Tumours. Int J Mol Sci. 2019; 20(14). https://doi.org/10.3390/ijms20143444 PMID: 31336999

22. Lecerf C, Le Bourhis X, Adriaenssens E. The long non-coding RNA H19: an active player with multiple facets to sustain the hallmarks of cancer. Cell Mol Life Sci. 2019; 76(23), 4673–4687. https://doi.org/10.1007/s00018-019-03240-z PMID: 31338555

23. Zhao J, Li L, Han ZY, Wang ZX, Qin LX. Long noncoding RNAs, emerging and versatile regulators of tumor-induced angiogenesis. Am J Cancer Res. 2019; 9(7), 1367–1381. PMID: 31392075

24. Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, et al. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 1992; 71(3), 527–542. https://doi.org/10.1016/0092-8674(92)90520-m PMID: 1423611

25. Bernstein HD, Zopf D, Freymann DM, Walter P. Functional substitution of the signal recognition particle 54-kDa subunit by its Escherichia coli homolog. Proc Natl Acad Sci USA. 1993; 90(11), 5229–5233. https://doi.org/10.1073/pnas.90.11.5229 PMID: 8389475

26. Clemson CM, McNeil JA, Willard HF, Lawrence JB. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J Cell Biol. 1996; 132(3), 259–275. https://doi.org/10.1083/jcb.132.3.259 PMID: 8636206

27. Reeves MB, Davies AA, McSharry BP, Wilkinson GW, Sinclair JH. Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. Science 2007; 316(5829), 1345–1348. https://doi.org/10.1126/science.1142984 PMID: 17540903

28. Houseley J, Rubbi L, Grunstein M, Tollerve D, and Vogelauer M. A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. Mol Cell. 2008; 32(5), 685–695. https://doi.org/10.1016/j.molcel.2008.09.027 PMID: 19061643

29. Swiezewski S, Liu F, Magusin A, Dean C. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature 2009; 462(7274), 799–802. https://doi.org/10.1038/nature08618 PMID: 20010688

30. Balbin OA, Malik R, Dhanasekaran SM, Prensner JR, Cao X, Wu YM, et al. The landscape of antisense gene expression in human cancers. Genome Res. 2015; 25(7), 1068–1079. https://doi.org/10.1101/gr.180596.114 PMID: 26063736

31. Jadaliha M, Gholamalamdari O, Tang W, Zhang Y, Petracovici A, Hao Q. A natural antisense lncRNA controls breast cancer progression by promoting tumor suppressor gene mRNA stability. PLoS Genet. 2018; 14(11), e1007802. https://doi.org/10.1371/journal.pgen.1007802 PMID: 30496290

32. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. RNA Biol. 2013; 10(6), 925–933. https://doi.org/10.4161/rna.24604 PMID: 23696037

33. Goyal A, Myacheva K, Gross M, Klingenberg M, Duran Arque B, Diederichs S. Challenges of CRISPR/Cas9 applications for long non-coding RNA genes. Nucleic Acids Res. 2017; 45(3), e12. https://doi.org/10.1093/nar/gkw883 PMID: 28180319

34. Salehi S, Taheri MN, Azarpira N, Zare A, Behzad-Behbahani A. State of the art technologies to explore long non-coding RNAs in cancer. J Cell Mol Med. 2017; 21(12), 3120–3140. https://doi.org/10.1111/jcmm.13238 PMID: 28631377

35. Qi F, Tan B, Ma F, Zhu B, Zhang L, Liu X, et al. A synthetic light-switchable system based on CRISPR Cas13a regulates the expression of lncRNA MALAT1 and affects the malignant phenotype of bladder cancer cells. Int J Biol Sci. 2019; 15(8), 1630–1636. https://doi.org/10.7150/ijbs.33772 PMID: 31360106

36. Butler AA, Johnston DR, Kaur S, Lubin FD. Long noncoding RNA NEAT1 mediates neuronal histone methylation and age-related memory impairment. Sci Signal. 2019; 12(588). https://doi.org/10.1126/scisignal.aaw9277 PMID: 31266852

37. Etebari K, Asad S, Zhang G, Asgari S. Identification of Aedes aegypti Long Intergenic Non-coding RNAs and Their Association with Wolbachia and Dengue Virus Infection. PLoS Negl Trop Dis. 2016; 10(10), e0005069. https://doi.org/10.1371/journal.pntd.0005069 PMID: 27760142

38. Yang D, Lian T, Tu J, Gaur U, Mao X, Fan X, et al. (2016). LncRNA mediated regulation of aging pathways in Drosophila melanogaster during dietary restriction. Aging (Albany NY) 2016; 8(9), 2182–2203. https://doi.org/10.18632/aging.101062 PMID: 27687893

39. Akay A, Jordan D, Navarro IC, Wrzesinski T, Ponting CP, Miska EA, et al. Identification of functional long non-coding RNAs in C. elegans. BMC Biol. 2019; 17(1), 14. https://doi.org/10.1186/s12915-019-0635-7 PMID: 30777050

40. Wei S, Chen H, Dzakah EE, Yu B, Wang X, Fu T, et al. Systematic evaluation of C. elegans lincRNAs with CRISPR knockout mutants. Genome Biol. 2019; 20(1), 7. https://doi.org/10.1186/s13059-018-1619-6 PMID: 30621757

41. Vasconcelos EJR, daSilva LF, Pires DS, Lavezzo GM, Pereira ASA, Amaral MS, 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. 2017; 7(1), 10508. https://doi.org/10.1038/s41598-017-10853-6 PMID: 28874839

42. Liao Q, Zhang Y, Zhu Y, Chen J, Dong C, Tao Y, et al. Identification of long noncoding RNAs in Schistosoma mansoni and Schistosoma japonicum. Exp Parasitol. 2018; 191, 82–87. https://doi.org/10.1016/j.exppara.2018.07.001 PMID: 29981293

43. Oliveira VF, Moares LAG, Mota EA, Jannotti-Passos LK, Coelho PMZ, Mattos ACA, et al. Identification of 170 New Long Noncoding RNAs in Schistosoma mansoni. Biomed Res Int. 2018; 1264697. https://doi.org/10.1155/2018/1264697 PMID: 30112357

44. Kim HC, Khalil AM, Jolly ER. LncRNAs in molluscan and mammalian stages of parasitic schistosomes are developmentally-regulated and coordinately expressed with protein-coding genes. RNA Biol. 2020; 17(6), 805–815. https://doi.org/10.1080/15476286.2020.1729594 PMID: 32131676

45. Maciel LF, Morales-Vicente DA, Verjovski-Almeida S. Dynamic Expression of Long Non-Coding RNAs Throughout Parasite Sexual and Neural Maturation in Schistosoma japonicum. Noncoding RNA 2020; 6(2). https://doi.org/10.3390/ncrna6020015 PMID: 32244675

46. Zhang X, Gong W, Cao S, Yin J, Zhang J, Cao, et al. Comprehensive Analysis of Non-coding RNA Profiles of Exosome-Like Vesicles From the Protoscoleces and Hydatid Cyst Fluid of Echinococcus granulosus. Front Cell Infect Microbiol. 2020; 10, 316. https://doi.org/10.3389/fcimb.2020.00316 PMID: 32793506

47. Azlan A, Halim MA, Azzam G. Genome-wide identification and characterization of long intergenic non-coding RNAs in the regenerative flatworm Macrostomum lignano. Genomics 2020; 112(2), 1273–1281. https://doi.org/10.1016/j.ygeno.2019.07.016 PMID: 31381967

48. Silveira GO, Coelho HS, Pereira ASA, Miyasato PA, Santos DW, Maciel LF, et al. Long non-coding RNAs are essential for Schistosoma mansoni pairing-dependent adult worm homeostasis and fertility. PLoS Pathog. 2023; 19(5):e1011369. https://doi.org/10.1371/journal.ppat.1011369 PMID: 37146077

49. Silveira GO, Coelho HS, Amaral MS, Verjovsky-Almeida S. Long non-coding RNAs as possible therapeutic targets in protozoa, and in Schistosoma and other helminths. Parasitol Res. 2022; 121(4), 1091–1115. https://doi.org/10.1007/s00436-021-07384-5 PMID: 34859292

50. Blokhin I, Khorkova O, Hsaio J, Wahlestedt C. Developments in lncRNA drug discovery: where are we heading? Expert Opin Drug Discov. 2018; 13(9), 837–849. https://doi.org/10.1080/17460441.2018.1501024 PMID: 30078338

51. Chen Y, Li Z, Chen X, Zhang S. Long non-coding RNAs: From disease code to drug role. Acta Pharm Sin B, 2021; 11(2), 340–354. https://doi.org/10.1016/j.apsb.2020.10.001 PMID: 33643816

52. Khorkova O, Stahl J, Joji A, Volmar CH, Zeier Z, Wahlestedt C. Long non-coding RNA-targeting therapeutics: discovery and development update. Expert Opin Drug Discov. 2023; https://doi.org/10.1080/17460441.2023.2236552 PMID: 37466388

53. Crooke ST, Baker BF, Crooke RM. Antisense technology: an overview and prospectus. Nat Rev Drug Discov. 2021; 20(6), 427–453. https://doi.org/10.1038/s41573-021-00162-z PMID: 33762737

54. Cwiklinski K, Jewhurst H, McVeigh P, Barbour T, Maule AG, Tort J, et al. Infection by the Helminth Parasite Fasciola hepatica Requires Rapid Regulation of Metabolic, Virulence, and Invasive Factors to Adjust to Its Mammalian Host. Mol Cell Proteomics. 2018; 17(4), 792–809. https://doi.org/10.1074/mcp.RA117.000445 PMID: 29321187

55. Zhang XX., Cwiklinski K., Hu RS. et al. Complex and dynamic transcriptional changes allow the helminth *Fasciola gigantica* to adjust to its intermediate snail and definitive mammalian hosts. BMC Genomics 2019; 20, 729. https://doi.org/10.1186/s12864-019-6103-5

56. Davey SD, Chalmers IW, Fernandez-Fuentes N, Swain MT, Smith D, Abbas Abidi SM, et al. In silico characterisation of the complete Ly6 protein family in Fasciola gigantica supported through transcriptomics of the newly-excysted juveniles. Mol Omics. 2022; 18(1), 45–56. https://doi.org/10.1039/d1mo00254f PMID: 34781332

57. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016; 11(9), 1650–1667. https://doi.org/10.1038/nprot.2016.095 PMID: 27560171

58. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006; 22, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158 PMID: 16731699

59. The RNAcentral Consortium. RNAcentral: a hub of information for non-coding RNA sequences. Nucleic Acids Res. 2019; 47(D1), D1250–D1251. https://doi.org/10.1093/nar/gky1206 PMID: 30535383

60. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015; 12(1), 59–60. https://doi.org/10.1038/nmeth.3176 PMID: 25402007

61. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res. 2017; 45(W1), W12–W16. https://doi.org/10.1093/nar/gkx428 PMID: 28521017

62. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010; 26(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616 PMID: 19910308

63. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012; 40(10), 4288–4297. https://doi.org/10.1093/nar/gks042 PMID: 22287627

64. Taylor SC, Laperriere G, Germain H. Droplet digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. Sci Rep. 2017; 7, 2409 (2017). https://doi.org/10.1038/s41598-017-02217-x PMID: 28546538

65. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010; 26(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

66. Russo PST, Ferreira GR, Cardoza LE, Bürger MC, Arias-Carrasco R, Maruyama SR, et al. CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. BMC Bioinformatics 2018; 56, https://doi.org/10.1186/s12859-018-2053-1 PMID: 29458351

67. Gillan V, Maitland K, Laing R, Gu H, Marks ND, Winter AD. Increased Expression of a MicroRNA Correlates with Anthelmintic Resistance in Parasitic Nematodes. Front Cell Infect Microbiol. 2017; 7, 452. https://doi.org/10.3389/fcimb.2017.00452 PMID: 29209592

68. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in Drosophila. Genome Biol. 2003; 5(1), R1. https://doi.org/10.1186/gb-2003-5-1-r1 PMID: 14709173

69. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. RNA 2004; 10(10), 1507–1517. https://doi.org/10.1261/rna.5248604 PMID: 15383676

70. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nat Genet. 2007; 39(10), 1278–1284. https://doi.org/10.1038/ng2135 PMID: 17893677

71. Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods Mol Biol. 2019; 1962, 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14 PMID: 31020564

72. Maciel LF, Morales-Vicente DA, Silveira GO, Ribeiro RO, Olberg GGO, Pires DS, et al. Weighted Gene Co-Expression Analyses Point to Long Non-Coding RNA Hub Genes at Different Schistosoma mansoni Life-Cycle Stages. Front Genet. 2019; 10, 823. https://doi.org/10.3389/fgene.2019.00823 PMID: 31572441

73. Cwiklinski K, Dalton JP, Dufresne PJ, La Course J, Williams DJ, Hodgkinson J, et al. The Fasciola hepatica genome: gene duplication and polymorphism reveals adaptation to the host environment and

the capacity for rapid evolution. Genome Biol. 2015; 16, 71. https://doi.org/10.1186/s13059-015-0632-2 PMID: 25887684

74. McNulty SN, Tort JF, Rinaldi G, Fischer K, Rosa BA, Smircich P, et al. Genomes of Fasciola hepatica from the Americas reveal colonization with Neorickettsia endobacteria related to the agents of Potomac horse and Human Sennetsu fevers. PLoS Genet. 2017; 13(1):e1006537. https://doi.org/10.1371/journal.pgen.1006537 PMID: 28060841

75. Feng D, Li Q, Yu H, Kong L, Du S. Transcriptional profiling of long non-coding RNAs in mantle of Crassostrea gigas and their association with shell pigmentation. Sci Rep. 2018; 8(1), 1436. https://doi.org/10.1038/s41598-018-19950-6 PMID: 29362405

76. Wang Y, Xu T, He W, Shen X, Zhao Q, Bai J, et al. Genome-wide identification and characterization of putative lncRNAs in the diamondback moth, Plutella xylostella (L.). Genomics 2018; 110(1), 35–42. https://doi.org/10.1016/j.ygeno.2017.08.003 PMID: 28789862

77. Zhou QZ, Fang SM, Zhang Q, Yu QY, Zhang Z. Identification and comparison of long non-coding RNAs in the silk gland between domestic and wild silkworms. Insect Sci. 2018; 25(4), 604–616. https://doi.org/10.1111/1744-7917.12443 PMID: 28111905

78. Diederichs S. The four dimensions of noncoding RNA conservation. Trends Genet. 2014; 30(4), 121–123. https://doi.org/10.1016/j.tig.2014.01.004 PMID: 24613441

79. Johnsson P, Lipovich L, Grander D, Morris KV. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochim Biophys Acta. 2014; 1840(3), 1063–1071. https://doi.org/10.1016/j.bbagen.2013.10.035 PMID: 24184936

80. Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Rep. 2015; 11(7), 1110–1122. https://doi.org/10.1016/j.celrep.2015.04.023 PMID: 25959816

81. He X, Pan W. Host-parasite interactions mediated by cross-species microRNAs. Trends Parasitol. 2022; 38(6), 478–488. https://doi.org/10.1016/j.pt.2022.02.011 PMID: 35307299

82. Jarroux J, Morillon A, Pinskaya M. History, Discovery, and Classification of lncRNAs. Adv Exp Med Biol. 2017; 1008, 1–46. https://doi.org/10.1007/978-981-10-5203-3_1 PMID: 28815535

83. McCusker P, McVeigh P, Rathinasamy V, Toet H, McCammick E, O'Connor A, et al. Stimulating neoblast-like cell proliferation in juvenile Fasciola hepatica supports growth and progression towards the adult phenotype in vitro. Plos Negl Trop Dis. 2016; 10(9):e0004994 https://doi.org/10.1371/journal.pntd.0004994 PMID: 27622752

84. Karlin S, Chen C, Gentles AJ, Cleary M. Associations between human disease genes and overlapping gene groups and multiple amino acid runs. Proc Natl Acad Sci USA. 2002; 99(26), 17008–17013. https://doi.org/10.1073/pnas.262658799 PMID: 12473749

85. Veeramachaneni V, Makalowski W, Galdzicki M, Sood R, Makalowska I. Mammalian overlapping genes: the comparative perspective. Genome Res. 2004; 14(2), 280–286. https://doi.org/10.1101/gr.1590904 PMID: 14762064

86. Kim DS, Cho CY, Huh JW, Kim HS, Cho HG. EVOG: a database for evolutionary analysis of overlapping genes. Nucleic Acids Res. 2009; 37(Database issue), D698–702. https://doi.org/10.1093/nar/gkn813 PMID: 18986995

87. Behura SK, Severson DW. Overlapping genes of Aedes aegypti: evolutionary implications from comparison with orthologs of Anopheles gambiae and other insects. BMC Evol Biol. 2013; 13, 124. https://doi.org/10.1186/1471-2148-13-124 PMID: 23777277

88. Liu SJ, Horlbeck MA, Cho SW, Birk HS, Malatesta M, He D, et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science 2017; 355(6320). https://doi.org/10.1126/science.aah7111 PMID: 27980086

89. Ransohoff JD, Wei Y, Khavari PA. The functions and unique features of long intergenic non-coding RNA. Nat Rev Mol Cell Biol. 2018; 19(3), 143–157. https://doi.org/10.1038/nrm.2017.104 PMID: 29138516

90. de Smet MD, Meenken CJ, van den Horn GJ. Fomivirsen—a phosphorothioate oligonucleotide for the treatment of CMV retinitis. Ocul Immunol Inflamm. 1999; 7(3–4), 189–198. https://doi.org/10.1076/ocii.7.3.189.4007 PMID: 10611727

91. Smith RA, Miller TM, Yamanaka K, Monia BP, Condon TF, Hung G, et al. Antisense oligonucleotide therapy for neurodegenerative disease. J Clin Invest. 2006; 116(8), 2290–2296. https://doi.org/10.1172/JCI25424 PMID: 16878173

92. Lim KRQ, Maruyama R, Yokota T. Eteplirsen in the treatment of Duchenne muscular dystrophy. Drug Des Devel Ther. 2017; 11, 533–545. https://doi.org/10.2147/DDDT.S97635 PMID: 28280301

93. Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. Nat Rev Genet. 2016; 17(5), 272–283. https://doi.org/10.1038/nrg.2016.20 PMID: 27040487

94. Ulitsky I. Interactions between short and long noncoding RNAs. FEBS Lett. 2018; 592(17), 2874–2883. https://doi.org/10.1002/1873-3468.13085 PMID: 29749606

95. Yamamura S, Imai-Sumida M, Tanaka Y, Dahiya R. Interaction and cross-talk between non-coding RNAs. Cell Mol Life Sci. 2018; 75(3), 467–484. https://doi.org/10.1007/s00018-017-2626-6 PMID: 28840253

96. Zhang S, Cheng Z, Wang Y, Han Y. (2021). The risks of miRNA therapeutics: In a drug target perspective. Drug Design Dev Ther. 2021; 15, 721. https://doi.org/10.2147/DDDT.S288859