J Med Biosci Suppl 2025;7: S21 - S27 DOI: 10.58837/CHULA.JMBS.7.S1.4

**Original article**

# Comprehensive reconstruction and characterization of long non-coding RNAs in *Opisthorchis viverrini*

**Waraporn Taweesin[1], Siriyakorn Kulwong[2], Wassana Jamnongkan[3] , Wichit Taron[3], Autchasai Siriprayong[4], Sirinya Sitthirak[2], Anchalee Techasen[4] , Nisana Namwat[2,3], Poramate Klanrit[2,3], Watcharin Loilome[2,3], Arporn Wangwiwatsin[2,3,*]**

*[1]Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand*
*[2]Department of Systems Biosciences and Computational Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand*
*[3]Cholangiocarcinoma Research Institute, Khon Kaen University, Khon Kaen, Thailand*
*[4]Department of Medical Technology, Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, Thailand*

## Abstract

***Background***: *Opisthorchis viverrini* is a parasitic flatworm that causes opisthorchiasis and is a major risk factor for cholangiocarcinoma (CCA) in Southeast Asia. Previous studies explored parasite biology, using the available reference genome annotation, specifically the mRNA expression. The information on genomic features of Long Non-Coding RNA (lncRNA), which can affect parasite biological function, is largely unannotated and remains unknown.

***Objectives***: We aimed to reconstruct and classify lncRNA derived from *O. viverrini in vivo* from a definitive host to understand its features and provide a new alternative option for lncRNA as an intervention target against opisthorchiasis.

***Methods***: Our study used rRNA-depleted data obtained by the Illumina NextSeq 550Dx platform to reassemble and classify lncRNA using available public bioinformatic tools.

***Results***: Our findings revealed a total of 38,282 transcripts assembled from uniquely mapped reads. A total of 20,607 transcripts were predicted as long non-coding transcripts of various lengths. In those transcripts, 11,077 antisense and 9,530 sense lncRNAs were accounted for based on transcription direction. Based on genomic location, we classified lncRNAs into four types: long intronic ncRNAs, long exonic ncRNAs, upstream long-intergenic ncRNAs, and downstream long-intergenic ncRNAs.

***Conclusion***: Using an rRNA-depleted library, we provided a comprehensive profile of *O. viverrini* lncRNA and reconstructed 20,607 novel lncRNAs that may interact with protein-coding genes. This information will contribute to opisthorchiasis intervention by providing insight into the biological mechanisms of lncRNAs, serving as stage-specific biomarkers for therapeutic strategies in opisthorchiasis, laying the foundation for research on lncRNA mechanisms, and further improving the reference genome annotation.

***Keywords***: long non-coding RNAs, *O. viverrini*, rRNA depletion, transcriptome assembly.

***Correspondence to: Arporn Wangwiwatsin,** Department of Systems Biosciences and Computational Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand.
E-mail: arpowa@kku.ac.th

*O. viverrini* is a parasitic flatworm that causes opisthorchiasis and is a major risk factor for cholangiocarcinoma (CCA) development in Southeast Asia.[1] Early and effective intervention is important to prevent the high incidence of opisthorchiasis and related diseases. Apart from candidate biomarkers and drug targets, which are mainly derived from protein-coding genes in various diseases, such as breast cancer, hepatocellular carcinoma, and prostate cancer [2–4], other types of RNA are an interesting option, such as lncRNA and microRNA.[5,6] The current reference genome and annotation of *O. viverrini* are publicly available [7] and have been annotated with 16,379 protein-coding genes.[8] However, the information on lncRNA is lacking in the available genome.

LncRNA is a class of RNA longer than 200 nucleotides in sequence length and lacks protein-coding potential for protein synthesis.[9] Despite not being translated into protein, lncRNAs may play crucial roles in organisms. For example, co-expression between mRNA and lncRNA has been reported together with an update on non-annotated lncRNAs of the blood fluke *Schistosoma mansoni*.[10] In *S. mansoni*, lncRNAs are involved in the regulation of gene expression and are essential for reproductive biology [11], as well as in the process of praziquantel drug response.[12] Moreover, specific expression patterns and biological involvement of lncRNAs have been demonstrated in other parasite infections, including *Plasmodium falciparum* [13] and the ruminant parasitic nematode *Haemonchus contortus*. [14] These studies suggested that the lncRNA may play crucial roles in various biological processes necessary for understanding parasite biology.

Furthermore, a deeper profiling of their comprehensive genomic features serves as the initial foundation for research advances. According to current knowledge, the annotated lncRNAs of *O. viverrini* have not been reported. Here, we aimed to reconstruct a new transcriptome assembly from rRNA-depleted libraries across 3 *in vivo* stages obtained from the hamster definitive host and to classify lncRNAs of *O. viverrini*, which is necessary for improving genome annotation and their future use in investigating biomolecular mechanisms, lncRNA-mRNA interaction, and biomarker and drug target candidates for *O. viverrini*. We used publicly available bioinformatics tools to identify and classify the lncRNA category. This work provides an essential step for future investigation of lncRNA functions and helps the research community access an improved genome annotation to advance opisthorchiasis intervention.

## Materials and methods

### *O. viverrini RNA sample preparation*

Nine samples of leftover total RNA from Taweesin et al., 2025 (manuscript under revision), were used in this study. The samples were divided into three groups, with three replicates per group, according to the time points of parasite collection post-infection: 14 days (D14), 42 days (D42), and 180 days (D180).

### *cDNA library preparation* and next-generation sequencing

To enrich lncRNAs, the RNA-seq libraries were prepared using a probe-free Zymo-Seq RiboFree Total RNA Library Kit (R3000, Zymo Research) with a starting RNA amount of 100 ng/sample. In brief, the principle of the probe-free rRNA depletion method removes cDNAs derived from ribosomal RNA (rRNA) by digesting high-abundance rRNA-cDNA hybrids. The remaining cDNAs from mRNAs and other ncRNAs were ligated to adapters for library amplification.[15] The concentrations and average library sizes of cDNA libraries were investigated using the Agilent TapeStation with Agilent D5000 ScreenTape system. The sequencing was performed on Illumina NextSeq 550Dx platform, with paired-end read sequencing, 150 bp in sequence length.

### *Bioinformatics analysis*

All sequence reads of the rRNA-depleted libraries were quality checked using fastQC (v. 0.12.1) [16] and multiQC (v. 1.25.1).[17] All universal Illumina adapters and Zymo adapter sequences were specified before being removed by default parameters of TrimGalore (v. 0.6.7-1).[18] Reads with good quality (Phred quality scores cut-off = 30) were aligned to the reference genome of the *O. viverrini* [7] using STAR (v. 2.7.11b) [19] with *—outSAMstrandField intronMotif* option to identify the strandedness of reads. Mapping outputs were converted from SAM to BAM format and sorted using samtools (v. 1.21).[20] Uniquely mapped reads in sorted BAM files were extracted using samtools with the specific symbol of unique alignment number (NH:i:1). Uniquely mapped reads were reassembled into full-length transcripts and exported as General Transfer Format (GTF) for each of the nine samples using scallop with default parameters (v. 0.10.5).[21] Next, a consensus list of transcriptomes was created using TACO (v. 0.6.2) [22] with parameters *(—filter-min-length 200 –isoform-frac 0.5*). The consensus transcriptome was classified into lncRNA classes using FEELnc (v. 0.01) [23] via Apptainer shell running (v. 1.3.2).[24] All transcripts that passed TACO filtering step were assessed for

their coding potential with CPC2 (version 0.1).[25] We merged the two datasets to select the reliable classified lncRNAs based on two criteria: 1) the filtered assembled transcripts were identified as non-coding potential by the CPC2, and 2) long assembled transcripts were defined as having a high-confidence level by the FEELnc classifier. All data visualization was done using R and RStudio (v. 4.2.3).[26]

## Results

### *Library output*

The Agilent TapeStation analysis showed that our rRNA-depleted libraries of *O. viverrini* were successfully produced with concentrations between 0.64 and 17.6 ng/µl. The average library size ranged from 321 to 587 bp, as the details shown in **Table 1**. This result emphasized a viable approach for studying organisms where a reliable rRNA sequence is not available, rendering probe-based methods unsuitable, by using a universal probe-free rRNA-depletion strategy.

### *Pre-processing rRNA-depleted data*

The sequencing output was 3 to 5 million (M) reads/sample. The reads were checked by fastQC and multiQC showing an average Q score of 33.79. GC content was 44%, similar to a previously published study in *O. viverrine*.[7] Read length after adapter trimming is distributed between 132 and 145. For sequence alignment, total reads after the adapter and quality trimming were mapped to the reference genome of the *O. viverrini*[7], resulting in various numbers of uniquely mapped reads (UMR), which were selected for transcriptome assembly to classify lncRNA transcripts. The lowest and highest numbers of UMRs were observed at 24.75% and 76.09%, which were presented in the conditions of D42 and D14, respectively, as the details shown in **Table 2**.

**Table 1.** The concentration and average sequence length of non-coding libraries.

| Sample | Concentration (ng/µl) | Average library size (bp) |
|---|---|---|
| D14_rep1 | 9.87 | 587 |
| D14_rep2 | 17.60 | 426 |
| D14_rep3 | 12.70 | 416 |
| D42_rep1 | 5.82 | 335 |
| D42_rep2 | 6.87 | 328 |
| D42_rep3 | 3.24 | 326 |
| D180_rep1 | 5.44 | 401 |
| D180_rep2 | 0.64 | 334 |
| D180_rep3 | 1.74 | 321 |

**Table 2.** Pre-processing data of rRNA-depleted libraries for lncRNA transcriptome assembly.

| Sample | Forward/ Reverse reads | Total reads (M) | %GC | Average sequence length | Uniquely mapped reads (M) by STAR | Uniquely mapped reads (%) by STAR |
|---|---|---|---|---|---|---|
| D14_rep1 | F | 5,347,334 | 44 | 145.55 | 3,553,279 | 66.45% |
| D14_rep1 | R | 5,347,334 | 43 | 144.55 | | |
| D14_rep2 | F | 5,010,798 | 44 | 145.63 | 3,812,722 | 76.09% |
| D14_rep2 | R | 5,010,798 | 43 | 145.19 | | |
| D14_rep3 | F | 4,294,164 | 46 | 141.48 | 2,792,746 | 65.04% |
| D14_rep3 | R | 4,294,164 | 46 | 142.46 | | |
| D42_rep1 | F | 3,633,967 | 43 | 143.15 | 1,781,447 | 49.02% |
| D42_rep1 | R | 3,633,967 | 42 | 143.52 | | |
| D42_rep2 | F | 4,951,534 | 42 | 143.92 | 1,225,536 | 24.75% |
| D42_rep2 | R | 4,951,534 | 43 | 144.03 | | |
| D42_rep3 | F | 4,955,961 | 44 | 139.91 | 3,115,320 | 62.86% |
| D42_rep3 | R | 4,955,961 | 45 | 141.59 | | |
| D180_rep1 | F | 4,026,285 | 45 | 140.88 | 2,966,768 | 73.68% |
| D180_rep1 | R | 4,026,285 | 43 | 142.07 | | |
| D180_rep2 | F | 4,423,244 | 47 | 132.88 | 3,010,802 | 68.07% |
| D180_rep2 | R | 4,423,244 | 48 | 135.59 | | |
| D180_rep3 | F | 4,597,434 | 45 | 136.17 | 3,434,759 | 74.71% |
| D180_rep3 | R | 4,597,434 | 45 | 139.63 | | |

***Reconstruction of full-length lncRNA transcripts, filtering, and classifying***

Full-length transcripts were reconstructed using a reference-guided approach. UMRs were used for Scallop assembly with reference GTF information, which identified accurate splicing sites from extracted UMRs in each of the sorted BAM files and assembled full-length transcripts. The resulting GTF files representing each sample, now including potential lncRNA locations, were combined to create a consensus GTF using TACO. This process provided a total of 38,282 full-length transcripts in the consensus transcriptome. After filtering out transcripts less than 200 bp and monoexonic transcripts using FEELnc, the analysis yielded 21,985 filtered full-length transcripts. Within these full-length transcripts, 50,960 interactions were predicted, and 379 transcripts without predicted interaction were identified. The probability of coding potential in the remaining 21,985 full-length transcripts was assessed using CPC2, revealing 20,607 transcripts predicted as long non-coding transcripts with a high confidence level in predicted lncRNA classes.

In the predicted non-coding group, of 20,607 transcripts, the average length of lncRNAs was 724 bp, with the shortest and longest lncRNAs being 236 and 18,872 bp respectively, suggesting *O. viverrini* lncRNAs vary widely in size, as shown in **Figure 1A**. We also classified lncRNA into two major categories based on transcript direction and genomic location using FEELnc classifier mode. Based on transcription directions, from all 20,607 transcripts, 11,077 transcripts were classified as antisense lncRNAs, and 9,530 transcripts as sense lncRNAs, as shown in **Figure 1B**. By genomic locations, four types of lncRNAs were identified, including long intronic non-coding RNAs (10,660 transcripts), upstream lincRNAs (4,583 transcripts), downstream lincRNAs (3,810 transcripts), and long exonic ncRNAs (1,554 transcripts), as shown in **Figure 1C**.

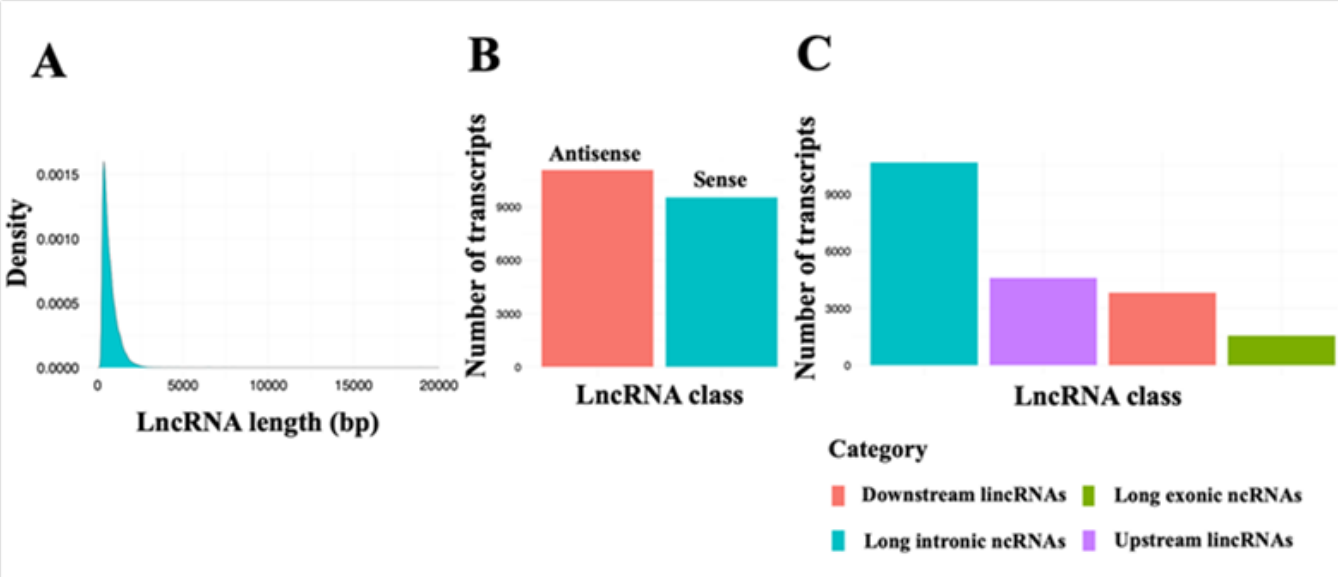


**Figure 1.** Classification of non-coding potential transcripts and lncRNA types in *O. viverrini*. A total of 20,607 full-length transcripts were predicted as non-coding potential transcripts by the CPC2 and high-confidence classified groups of lncRNA with the FEELnc classifier mode. **(A)** Average lncRNA length distribution. **(B)** LncRNA types categorized by the transcription direction: 1) antisense lncRNA, transcribed from the opposite strand of a gene, complementary to the sense strand, 2) sense lncRNAs, transcribed from the same strand of a gene, and sometimes contain exons from the gene. **(C)** LncRNA types categorized by the genomic location: 1) long intronic ncRNAs, transcribed entirely from the introns without an exonic region, 2) upstream long intergenic non-coding RNA (lincRNA), transcribed from upstream of a gene, 3) downstream lincRNAs, transcribed from downstream of a gene, and 4) long exonic ncRNA, transcribed from exonic regions.

Although most lncRNAs originated from the antisense strand, the number of lncRNAs from the sense strand was not substantially different. When categorized by genomic location, long intronic RNAs were the most abundant, indicating that, despite being derived from the sense strand, the regions within these lncRNAs are primarily intronic, consistent with the nature of lncRNAs. Some regulatory mechanisms clearly play roles in controlling the expression and splicing of these lncRNAs, which appear to overlap with the locations of protein-coding genes, warranting further investigation.

## Discussion

Here, we revealed the successful ncRNAs library preparation, specifically to reassemble and identify novel lncRNAs of the *O. viverrini* species. Using the universal probe-free rRNA depletion strategy to generate non-coding libraries, the obtained library sizes (321-587 bp) were shorter than the human RNA library size prepared by the same method (~560 bp) as indicated in the kit instruction. This is likely due to differences between species or variation in the amount and quality of input RNA. As expected, approximately 3-5 M reads/sample were sequenced with good quality in base calling of over 30 Phred score (>99.9% accuracy). The average GC content of 44.33 bp in the rRNA-depleted libraries was lower than mRNA GC content (47.8%) but higher than genomic DNA GC contents (43.7%) for *O. viverrini* [7], suggesting the inclusion of other types of RNA other than mRNA in our dataset. The wide range of UMRs (24%-76.09%) may reflect the nature of lncRNAs, which often contain repetitive elements and intronic sequences that complicate mapping, especially with short-read sequencing data. Additionally, the incomplete *O. viverrini* reference genome, generated from short-read sequencing, often has gaps near repeat regions, which may further reduce alignment efficiency. Additionally, although the FASTQ files showed good quality with an overall score over 33, variability in depletion efficiency during library preparation and differences in input RNA quality may also contribute to the observed range.

Reconstruction of transcriptome assembly revealed a total of 38,282 transcripts, and 20,607 transcripts were predicted as lncRNAs. The average lncRNA transcript length (727 bp) was close to the median length of 738 bp in lncRNA of another worm species, *H. concortus*.[14] Our results revealed that the intronic ncRNAs are the most common class found in lncRNAs based on genome coordinates. The majority, as classified by potential mechanism of action, were antisense lncRNAs. These results emphasized that lncRNA of *O. viverrini* can originate from various genomic locations and transcriptional directions, representing all major types of lncRNA known.[27]

Intriguingly, our data identified only 1,019 transcripts, excluding mono-exonic genes, with coding potential, whereas the reference genome annotates 16,379 protein-coding genes. Notably, the conducted rRNA depletion method could potentially enrich lncRNAs and lessen the more abundant mRNA, reflected in the lower number of coding transcripts compared to 16,379 coding genes, derived from mRNA poly-A enrichment in previous work.[8] This suggests a potential bias from a probe-free rRNA-depletion method, which might not be compatible for studying mRNA-coding genes, and other highly-expressed genes. While using this dataset to profile the lncRNA sequences would be sensible, quantitative measurement of the lncRNA expression needs to be done with caution and should be verified with another method.

However, the high number of transcripts originating from the antisense strand, with the majority derived from intronic regions, suggests they are unlikely to be coding transcripts misannotated as lncRNA, supporting the validity of our lncRNA annotation. Nevertheless, further validation of the identified lncRNAs is warranted. Validation using RT-qPCR is planned to help confirm the robustness of this dataset. Moreover, additional methods for coding potential prediction should be explored, or a reference-free approach using *de novo* assembly of RNA-seq reads could be tested.

This dataset represents the first description of lncRNAs in *O. viverrini* and provides a foundation for future studies on parasite biology and biomarker potential. Future investigations, integrating lncRNA and mRNA expression data, could explore co-expression and co-localization, laying the groundwork for functional research on lncRNA mechanisms, gene regulation, and potential stage-specific biomarkers and drug targets. For example, in *S. mansoni*, transcription factors shared between lncRNAs and mRNAs regulate promoter activity, suggesting lncRNA-mRNA interactions, with lncRNAs co-expressed with Praziquantel-controlled mRNAs.[28] In *Toxoplasma*

*gondii*-infected mice, gene expression analyses of mRNAs and lncRNAs identified co-localization in key pathways, such as immune responses and metabolic changes, across different infection stages.[29] Regarding biomarkers, lncRNAs like *PVT1* and *uc002mbe.2* have been suggested as potential hepatocellular carcinoma (HCC) markers in serum, distinguishing HCC patients from healthy individuals.[30] Likewise, *LINC01271* has been implicated as a candidate for metastasis in breast cancer. [31]

## Conclusion

In conclusion, we have demonstrated that our library preparation method with the rRNA removal, sequencing, and bioinformatics pipeline was suitable for the reconstruction of a new transcriptome assembly and lncRNA classification, with novel 20,607 lncRNAs identified. Our findings provide the first comprehensive catalog of lncRNAs in *O. viverrini*, offering a valuable resource for future studies on their functional roles for research and intervention strategies.

## Acknowledgments

This project was supported by the National Science Research and Innovation Fund (NSRF) under the Basic Research Fund of Khon Kaen University through Cholangiocarcinoma Research Institute to AW. WaT is sponsored through the Institute for the Promotion of Teaching Science and Technology studentship. We thank Mr. Taweesak Sarnboon for his assistance with parasite collection and infection. We thank members of the Systems Biosciences lab, Department of Systems Biosciences and Computational Medicine, and Department of Biochemistry, Faculty of Medicine, Khon Kaen University, for their valuable support. Computationally-intensive data analyses were performed using high-performance computing resources provided by the Office of Digital Technology, Khon Kaen University. Finally, we thank all available public bioinformatics tools.

## Conflict of interest statement

The authors declare no conflicts of interest related to this work.

## Data sharing statement

Raw data (FASTQ) for this study have been deposited in the European Nucleotide Archive (ENA) under the accession number ERP172029.

## References

1. Khuntikeo N, Titapun A, Loilome W, Yongvanit P, Thinkhamrop B, Chamadol N, et al. Current perspectives on opisthorchiasis control and cholangiocarcinoma detection in Southeast Asia. Front Med 2018;5:117.
2. Han Z, Feng W, Hu R, Ge Q, Ma W, Zhang W, et al. RNA-seq profiling reveals PBMC RNA as a potential biomarker for hepatocellular carcinoma. Sci Rep 2021;11:17797.
3. Nikitina AS, Sharova EI, Danilenko SA, Butusova TB, Vasiliev AO, Govorov AV, et al. Novel RNA biomarkers of prostate cancer revealed by RNA-seq analysis of formalin-fixed samples obtained from Russian patients. Oncotarget 2017;8:32990-3001.
4. Sinicropi D, Qu K, Collin F, Crager M, Liu ML, Pelham RJ, et al. Whole transcriptome RNA-Seq analysis of breast cancer recurrence risk using formalin-fixed paraffin-embedded tumor tissue. PLoS One 2012;7: e40092
5. Yamada A, Yu P, Lin W, Okugawa Y, Boland CR, Goel A. A RNA-Sequencing approach for the identification of novel long non-coding RNA biomarkers in colorectal cancer. Sci Rep 2018;8:575.
6. Zhang W, Zang J, Jing X, Sun Z, Yan W, Yang D, et al. Identification of candidate miRNA biomarkers from miRNA regulatory network with application to prostate cancer. J Transl Med 2014;12:66.
7. Young ND, Nagarajan N, Lin SJ, Korhonen PK, Jex AR, Hall RS, et al. The Opisthorchis viverrini genome provides insights into life in the bile duct. Nat Commun 2014;5:4378.
8. Jex AR, Young ND, Sripa J, Hall RS, Scheerlinck JP, Laha T, et al. Molecular Changes in Opisthorchis viverrini (Southeast Asian Liver Fluke) during the Transition from the Juvenile to the Adult Stage. PLoS Negl Trop Dis 2012;6:e1916.
9. Bridges MC, Daulagala AC, Kourtidis A. LNCcation: lncRNA localization and function. J Cell Biol. 2021; 220:e202009045.
10. 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.
11. 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:e1011369.

12. Jardim Poli P, Fischer-Carvalho A, Tahira AC, Chan JD, Verjovski-Almeida S, Sena Amaral M. Long non-coding RNA levels are modulated in schistosoma mansoni following in vivo praziquantel exposure. Noncoding RNA 2024;10:27.
13. Batugedara G, Lu XM, Hristov B, Abel S, Chahine Z, Hollin T, et al. Novel insights into the role of long non-coding RNA in the human malaria parasite, Plasmodium falciparum. Nat Commun 2023;14:5086.
14. Zhou C, Tuersong W, Liu L, Di W, He L, Li F, et al. Non-coding RNA in the gut of the blood-feeding parasitic worm, Haemonchus contortus. Vet Res 2024;55:1.
15. Zymo Research. Zymo-Seq RiboFree® Total RNA Library Kit one for all. Universal rRNA depletion for any organism. California: Zymo Research; 2023.
16. Andrews S. FastQC: A quality control tool for high throughput sequence data. [Internet]. Cambridge: Babraham Bioinformatics;c2023[cited 2025 Apr 27]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
17. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016;32:3047-8.
18. Krueger F, James F, Ewels P, Afyounian E, Weinstein M, Schuster-Boeckler B, et al. FelixKrueger/TrimGalore: v0.6.10 - add default decompression path [software]. Zenodo; 2023 [cited 2025 Mar 13]. Available from: https://zenodo.org/record/7598955.
19. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2012;29:15-21.
20. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience 2021;10:giab008.
21. Shao M, Kingsford C. Accurate assembly of transcripts through phase-preserving graph decomposition. Nat Biotechnol 2017;35:1167-9.
22. Niknafs YS, Pandian B, Iyer HK, Chinnaiyan AM, Iyer MK. TACO produces robust multisample transcriptome assemblies from RNA-seq. Nat Methods 2017;14:68-70.
23. Wucher V, Legeai F, Hédan B, Rizk G, Lagoutte L, Leeb T, et al. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. Nucleic Acids Res 2017;45:e57.
24. Kurtzer GM, Sochat V, Bauer MW. Singularity: Scientific containers for mobility of compute. PLoS One 2017;12:e0177459.
25. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res 2017;45:W12-6.
26. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2003.
27. Silveira GO, Coelho HS, Amaral MS, Verjovski-Almeida S. Long non-coding RNAs as possible therapeutic targets in protozoa, and in Schistosoma and other helminths. Parasitol Res 2022;121:1091-115.
28. Jardim Poli P, Fischer-Carvalho A, Tahira AC, Chan JD, Verjovski-Almeida S, Sena Amaral M. Long Non-Coding RNA Levels Are Modulated in Schistosoma mansoni following In Vivo Praziquantel Exposure. Noncoding RNA 2024;10:27.
29. Zou Y, Yang X, Chen C, Ma H, Cao HW, Jiang J, et al. Transcriptomic profiling of long non-coding RNAs and messenger RNAs in the liver of mice during Toxoplasma gondii infection. Parasit Vectors 2024;17:20.
30. Yu J, Han J, Zhang J, Li G, Liu H, Cui X, et al. The long noncoding RNAs PVT1 and uc002mbe.2 in sera provide a new supplementary method for hepatocellular carcinoma diagnosis. Medicine (Baltimore) 2016;95:e4436.
31. Chang KC, Diermeier SD, Yu AT, Brine LD, Russo S, Bhatia S, et al. MaTAR25 lncRNA regulates the Tensin1 gene to impact breast cancer progression. Nat Commun 2020;11:6438.