

Research Article

Identification of 170 New Long Noncoding RNAs in *Schistosoma mansoni*

**Victor F. Oliveira,¹ Lauro A. G. Moares,² Ester A. Mota,¹
Liana K. Jannotti-Passos,³ Paulo M. Z. Coelho,³ Ana C. A. Mattos,³ Flávia F. B. Couto,³
Brian E. Caffrey,⁴ Annalisa Marsico,^{4,5} and Renata Guerra-Sá¹**

¹Departamento de Ciências Biológicas, Núcleo de Pesquisas em Ciências Biológicas, Universidade Federal de Ouro Preto, Morro do Cruzeiro, Ouro Preto, MG, Brazil

²Departamento de Computação, Laboratório de Computação de Sistemas Inteligentes, Universidade Federal de Ouro Preto, Morro do Cruzeiro, Ouro Preto, MG, Brazil

³Centro de Pesquisas René Rachou, Fiocruz, Belo Horizonte, MG, Brazil

⁴Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany

⁵Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany

Correspondence should be addressed to Renata Guerra-Sá; rguerrasa@gmail.com

Received 24 January 2018; Revised 14 June 2018; Accepted 21 June 2018; Published 11 July 2018

Academic Editor: Yudong Cai

Copyright © 2018 Victor F. Oliveira et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Long noncoding RNAs (lncRNAs) are transcripts generally longer than 200 nucleotides with no or poor protein coding potential, and most of their functions are also poorly characterized. Recently, an increasing number of studies have shown that lncRNAs can be involved in various critical biological processes such as organism development or cancer progression. Little, however, is known about their effects in helminth parasites, such as *Schistosoma mansoni*. Here, we present a computational pipeline to identify and characterize lncRNAs from RNA-seq data with high confidence from *S. mansoni* adult worms. Through the utilization of different criteria such as genome localization, exon number, gene length, and stability, we identified 170 new putative lncRNAs. All novel *S. mansoni* lncRNAs have no conserved syntenies including human and mouse. These closest protein coding genes were enriched in 10 significant Gene Ontology terms related to metabolism, transport, and biosynthesis. Fifteen putative lncRNAs showed differential expression, and three displayed sex-specific differential expressions in praziquantel sensitive and resistant adult worm couples. Together, our method can predict a set of novel lncRNAs from the RNA-seq data. Some lncRNAs are shown to be differentially expressed suggesting that those novel lncRNAs can be given high priority in further functional studies focused on praziquantel resistance.

1. Introduction

The trematode *Schistosoma mansoni* is the primary parasite species responsible for schistosomiasis, a chronic debilitating disease. It is considered one of the most devastating tropical diseases in the world with at least 258 million people infected. Furthermore, 800 million people were living in endemic areas at risk of infection with more than 200,000 deaths each year [1, 2]. Its transmission has been reported in more than 78 countries, especially in tropical and subtropical areas such as Central and South America, Africa, and Southeast Asia [3, 4].

Although, in the last few decades, several drugs have been used for the treatment of schistosomiasis, with praziquantel (PZQ) representing the only widely effective agent used [5, 6]. It is effective against all species of schistosomes that infect humans and is relatively cheap and easy to use, but PZQ does not provide a cure, since young schistosomula are mostly resistant to its anthelmintic effects [7]. This drug provides some relief to treated patients. However, young parasites, due to their intrinsic resistance to PZQ, escape elimination during treatment, grow to maturation, and begin to release eggs [5]. This mechanism of resistance is worrying, because, under this

ineffective pressure, drug resistance may also arise in humans, as in murine models [8, 9].

The *S. mansoni* genome is structured in 7 pairs of autosomes and one pair of sex chromosomes (female = ZW, male = ZZ). Chromosomes range in size from 18 to 73 MB and can be distinguished by size, shape, and the C-banding technique [10]. According to the latest annotation, the *S. mansoni* genome is still considered as a draft with 380 Mb and 885 scaffolds. Despite this, about 81% of the bases are organized in these chromosomes. More than 45% of the predicted genes were modified, and the total number was reduced from 11,807 to 10,852 [11].

This parasite has a complex life cycle that involves many larval stages, an intermediate snail, and a final mammalian host. It is believed that the difference and developmental complexity observed between the different evolutionary stages and environments depend on the regulation of gene expression [12]. Several molecules are responsible for gene expression regulation, especially long noncoding RNAs (lncRNAs). They are defined as transcripts longer than 200 nucleotides and do not encode proteins presenting several regulatory functions. They can interact with DNA, RNA molecules, and transcription factors, participating in various biological processes, mainly gene regulation [13].

While our knowledge of the mechanisms and scope of lncRNA-mediated regulation is growing, our understanding of how lncRNAs themselves are regulated is still quite limited [14]. Regulating lncRNA expression would be expected to be an important cellular consideration given that lncRNAs have been implicated in regulating a variety of processes in eukaryotes including imprinting, dosage compensation, cell cycle regulation, pluripotency, retrotransposon silencing, meiotic entry, and telomere length [15–18]. They can also play important roles in numerous disease and physiological metabolism processes, such as X-chromosome inactivation, embryonic development, and pluripotency maintenance [16, 19].

These findings deeply changed disease pathobiology comprehension and led to the emergence of new biological concepts about human diseases, including the parasitic disease. Several methodologies were created to characterize and identify this RNA subtype. Noteworthy, *S. mansoni* researchers used these techniques and described a great picture of these lncRNAs and their participation in the disease processes. To date two studies have been published predicting lncRNAs molecules in schistosomes [20, 21] and only one [21] describes possible lncRNAs' functions of 181 sequences from 7431 total predicted lncRNAs in 5 life cycle stages in this parasite, including canonically spliced putative lincRNA and spliced lncRNAs that are antisense to protein coding genes. These functions were predicted considering that lncRNAs may act by regulating their flanking protein coding gene neighbors in many processes to the rapid adaptation of the parasite to several environments.

In this study, we aimed to predict novel lncRNAs in *S. mansoni* via a computational pipeline and investigate features including possible functions of these sequences. Here we describe a complete new set of 170 lncRNAs in adult worms, from which we selected 15 for expression analysis in male, female, and also, for the first time, praziquantel-resistant

worms which are known to have a differential expression profile in several important genes in relation to susceptible worms [8, 22, 23]. Our results show a differential expression profile of lncRNAs and reinforce the importance of this RNA subtype in schistosome biology.

2. Material and Methods

2.1. Ethics Statement. All experiments that involve animals were authorized by the Ethical Committee for Animal Care of the Federal University of Ouro Preto (CEUA-UFOP protocol 2011/55). These procedures were conducted in accordance with the accepted national and international regulations for laboratory animal use and care.

2.2. Parasites. The *S. mansoni* LE strain was maintained by routine passage through *Biomphalaria glabrata* snails and BALB/c mice. The infected snails were induced to shed cercariae under light exposure for 2 hours. Adult worm parasites were obtained by liver perfusion of mice after 50 days of infection. The *S. mansoni* LE praziquantel-resistant (LE-PZQ) strains were obtained following a described method for inducing resistance to PZQ using infected *B. glabrata* snails [24]. Infected snails were treated 3 times with 100 mg/kg PZQ for five consecutive days with a one-week interval between them. Then, after this treatment the cercariae both from treated snails (LE-PZQ strains) and from nontreated snails (LE strains susceptible) were used to infect two groups of mice. These mice were treated 45 days after infection with 200, 400, or 800 mg/kg PZQ with three PZQ treatments, each treatment administered on 5 consecutive days, with 1-week interval, for selection of less susceptible parasites to PZQ following the method developed by Couto et al. [24]. Then, the LE-PZQ adult worms were obtained by mice liver perfusion, washed in RPMI 1640 (Sigma Chemical Co.), quick-frozen in liquid nitrogen, and stored at -80°C until use.

2.3. Data Set Download. The latest annotation data for *S. mansoni* (Genome Assembly release v5.2) were downloaded from GeneDB database (<http://www.genedb.org>) [25]. In order to perform prediction analysis in this pipeline, a RNA-seq library from 7-week-old mixed sex adult worms was selected and downloaded from ArrayExpress database [11] in the FASTQ format (<http://www.ebi.ac.uk/arrayexpress/>) under accession number E-MTAB-451.

2.4. Initial Transcriptome Assembly for *S. mansoni*. *S. mansoni* genome FASTA file and the annotation data in the GTF format were downloaded from GeneDB. One paired RNA-seq library was downloaded in the FASTQ format provided by sequencing using the Illumina Genome Analyzer IIx platform [11]. The quality was then analyzed using FastQC 0.11.4 [26]. The main parameters analyzed were the quality of the scores on the bases (quality of the Phred set as greater than 20 where it considers 1 error per 100 bases), per sequence quality scores, per sequence GC content, and removal of possible adapters used in the sequencing. After that, these reads were trimmed using Trimmomatic [27], and the following parameters were used: HEADCROP:15, LEADING:20, TRAILING:20,

SLIDINGWINDOW:5:20, and MINLEN:50. Next filtering the quality of the files in the FastQC and removing the adapters, all the reads were mapped using the *S. mansoni* reference genome with the STAR aligner [28]. In order to use this program, it was necessary to index the genome in the format compatible with STAR. Once the indexing was performed, the next step was to perform the mapping of the reads with the *S. mansoni* genome as reference using STAR. The output SAM files of STAR were then converted to its compressed BAM format, and then indexed and sorted with SAMtools for further analysis [29]. Subsequently, the BAM files were assembled by Cufflinks 2.2.1 [30] using *de novo* mode to assemble transcripts for 3 *S. mansoni* stage samples. Finally, assembly of transcripts was performed via Cufflinks, and the final file was submitted to the following pipeline filters for prediction of lncRNAs in *S. mansoni*. These steps were written in and performed with the use of custom Perl and Python scripts.

2.5. Identification of Novel lncRNAs Candidates. Cuffcompare was the first step used to compare the results generated by the *ab initio* assembly with the known annotations present in GeneDB in GTF format. The consensus of novel transcripts were then used for further analysis. As a result of Cuffcompare, all the assemblies that were detected as new transcripts were categorized into 12 different categories according to their location compared with the *S. mansoni* reference genes [30]. We kept the three following classes: unknown intergenic transcript (u), a transfrag falling entirely within a reference intron (i), and exonic overlap with reference on the opposite strand (x). This final file was submitted to the following filters written in the programming languages Python and Perl.

In the second step a filter was used to extract the sequences equal or greater than 200 nucleotides. The third step was done using CPAT (Coding Potential Assessment Tool), an alignment-free method to predict RNA coding potential using four sequence features [31]. Only the transcripts classified by CPAT as noncoding transcripts were added to a new FASTA file. For our proposed filter, we used a prebuilt logit fly model as the classifier with optimum cutoff (CP) 0.39 (CP \geq 0.39 indicates coding sequence; CP $<$ 0.39 indicates noncoding sequence) [31]. The fourth step consisted of the extraction of transcripts that presented putative ORF (Open Reading Frame) smaller than 300 nt using OrfPredictor server [32]. Transcripts with protein coding potential generally have ORF greater than 300 nt in size, generating proteins greater than or equal to 100 aa (amino acids) [33, 34]. In the fifth step we used an FPKM cutoff based on the distribution of the lncRNA expression level. Only FPKM values \geq 2 were included in the final analysis. The sixth step was performed manually using the NCBI database as a reference and lncRNAs sequences from other works [21]. At this step, we manually removed other RNA types, transposons, and predicted protein coding genes that have homology in other *Schistosoma* species. The seventh and final step was performed to remove all the known lncRNAs in *S. mansoni* using BLAST version 2.7.1 [35, 36].

2.6. lncRNAs Features. The novel adult lncRNAs predicted were characterized in terms of genomic localization, exon number, transcript length, and log₂ FPKM using R and R Studio with ggplot2 packaged [37].

2.7. Gene Ontology Enrichment Analyses. The possible lncRNAs' functions were hypothesized with Gene Ontology (GO) enrichment analysis approach (<http://geneontology.org/>). In this analysis, all neighboring genes up to 100,000 bases upstream and downstream of the lncRNAs were selected and then their functions were evaluated [38]. In this way, lncRNAs may act by binding to these protein coding genes by regulating their translation. The list obtained with all GO terms was created and plotted with R version 3.4.2, and R Studio version 1.0.136. The analysis method was based on Fisher's exact test and the -log₁₀ (*P* value) was used to denote the significance of the GO term enrichment.

2.8. lncRNAs Expression Analysis. Approximately 100 mg adult worms were used for total RNA extraction performed according to the manufacturer's specifications (SV Total RNA Isolation System). Expression levels of 15 lncRNAs were quantified by RT-qPCR with Applied Biosystems ABI 7300 by using SYBR-Green PCR Master Mix (Roche®). We designed specific primers for each lncRNA, endogenous control, and positive control using Gene Runner version 6.5.46 (Supplementary Table S1). For the investigated transcripts, three biological replicates were performed and normalized to the endogenous control with specific primers for *S. mansoni EIF4E* [39]. A long terminal repeat (LTR) retrotransposon with 58% cover and 95% identity with *S. mansoni* Saci-4 LTR retrotransposon was used as a comparison parameter because it is a transcript found expressed in all chromosomes. Expression levels were calculated according to the $2^{-\Delta Ct}$ method [40] using the Applied Biosystems 7300 software. All these experiments were performed following MIQE guidelines [41].

2.9. Statistical Analysis. Statistical analysis for RT-qPCR was performed using GraphPad Prism, version 7.0 (San Diego, CA, USA). One-way and two-way analysis of variance (ANOVA), following by Tukey multiple comparisons, were performed to investigate significant differential expression of transcripts throughout the sexes and treatments. In all cases, the differences were considered significant when *P* values were < 0.05 .

3. Results

3.1. Initially Assembled Transcripts. A stranded RNA-seq data set of the whole transcriptome was used to assemble transcripts (Figure 1(a)). We first trimmed more than 10 million raw reads (FASTQ reads) obtaining a total of 6 million clean reads. More than 5 million (82.18%) of them were mapped with STAR to the *S. mansoni* genome (v5.2) and calculation of summary statistics (Supplementary Table S2). *De novo* assembly from the aligned fragments was performed using Cufflinks with all the *ab initio* default parameters to generate

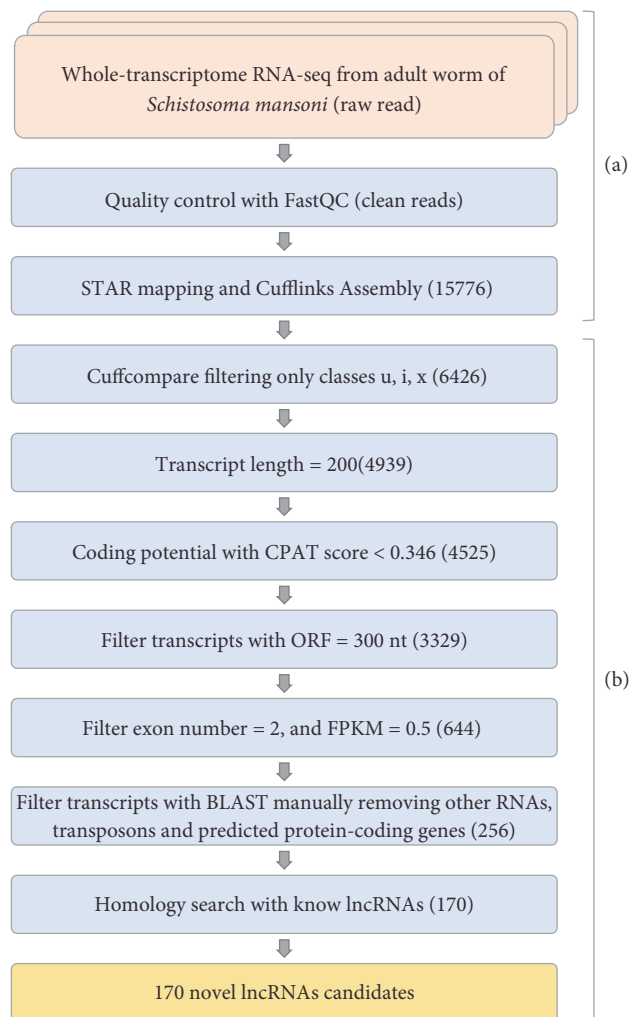


FIGURE 1: Integrative computational pipeline for the identification of lncRNAs in *S. mansoni*. (a) The raw RNA-seq data was preprocessed, aligned with STAR, and assembled using Cufflinks in ab initio mode. (b) The output was analyzed in several steps and algorithms were written in the Python and Perl programming languages. The numbers in parentheses represent the number of transcripts after each filtering step.

15,776 transcripts, which were then processed through the described pipeline.

3.2. Computational Pipeline to Predict Novel lncRNAs. Our computational pipeline applied multiple filters on these transcripts (Figure 1(b)) to predict novel lncRNAs in *S. mansoni* adult worm. First, Cuffcompare removed the assemblies overlapping with transcripts annotated in the reference genome. This includes reconstructed protein coding transcripts or annotated known noncoding transcripts. The selection of classes u, i, and x led to the removal of almost 60% of assembled transcripts. The remaining 6426 transcripts were submitted to a length filter (≥ 200) removing 1487. After passing this length filter, 4525 transcripts were classified by CPAT as noncoding transcripts following the ORF filter (≤ 300 nt) with 3329 transcripts remaining. For both categories,

transcripts with exon number ≥ 2 and FPKM ≥ 0.5 , only 644 transcripts were obtained. Since this study focuses on noncoding transcripts, a manual step was performed removing other RNAs, transposons, and predicted protein coding genes in other *Schistosoma* species. This filter led to 256 transcripts that were submitted to the last step removing all the known lncRNAs in *S. mansoni* by BLAST. From a total of 256 predicted lncRNAs, 170 (66.4%) were considered potentially novel lncRNAs candidates against 86 (33.6 %) that presented a homology (at least $>70\%$) with other lncRNAs in *S. mansoni* (Supplementary Table S3).

3.3. lncRNAs Features. To determine *S. mansoni* lncRNAs features, the genomic localization, exon number, transcript length, and \log_2 FPKM (Figure 2) were analyzed. In this data set we found that the majority of lncRNAs were found on chromosome 1, sexual ZW, and the scaffolds. The majority of the intronic lncRNAs were found on the sexual ZW and the unfinished scaffolds (Figure 2(a)). Moreover, lncRNAs had few exons per transcript (2-3) (Figure 2(b)) with most of them having transcript lengths between 200bp and 2000bp (Figure 2(c)). Finally, many of these lncRNAs could be confidently detected, with FPKM expression level between 2 and 9 (Figure 2(d)).

3.4. Overview of lncRNA RT-qPCR Validation. Based on nearby encoding genes, we selected 15 lncRNAs from a set of 170 putative molecules and the LTR retrotransposon to verify RT-qPCR expressions (Figure 3(a)). This selection was performed based on the enriched GO terms of the analyzed neighboring which were relevant genes for *S. mansoni* biology, for example, Smp_174670 (ubiquitin conjugating enzyme E2 J1), Smp_030780 (putative ubiquitination factor E4a), and Smp_199890 (venom allergen-like 7 protein).

Our relative expression results showed that *Sm-lncRNA 1*, *Sm-lncRNA 2*, *Sm-lncRNA 3*, *Sm-lncRNA 4*, *Sm-lncRNA 7*, *Sm-lncRNA 8*, *Sm-lncRNA 11*, *Sm-lncRNA 13*, *Sm-lncRNA 14*, and *Sm-lncRNA 15* had a low expression without statistical significance. Besides, 4 lncRNAs, *Sm-lncRNA 6*, *Sm-lncRNA 9*, *Sm-lncRNA 10*, and *Sm-lncRNA 12* had a higher expression than this lower group. The most highly expressed lncRNA was *Sm-lncRNA 5*. It was even more highly expressed than the LTR retrotransposon used as a comparison parameter.

We also selected the first three lncRNAs and the LTR retrotransposon to verify the expression between sexes into four groups: Control-Male, Control-Female, PZQ-Male, and PZQ-Female (PZQ-Male and PZQ-Female are related to *S. mansoni* LE praziquantel-resistant strains) (Figure 3(b)). All the 4 Control-Female expressions were higher than Control-Male and PZQ-Female.

3.5. Target Gene Prediction. GO enrichment analyses for the neighborhood target genes were made on three different aspects, namely, biological process (BP), molecular function (MF), and cellular component (CC). The 10 significantly overrepresented GO terms included 389 genes involved in BP, 555 genes involved in MF, and 768 genes involved in CC (Figure 4).

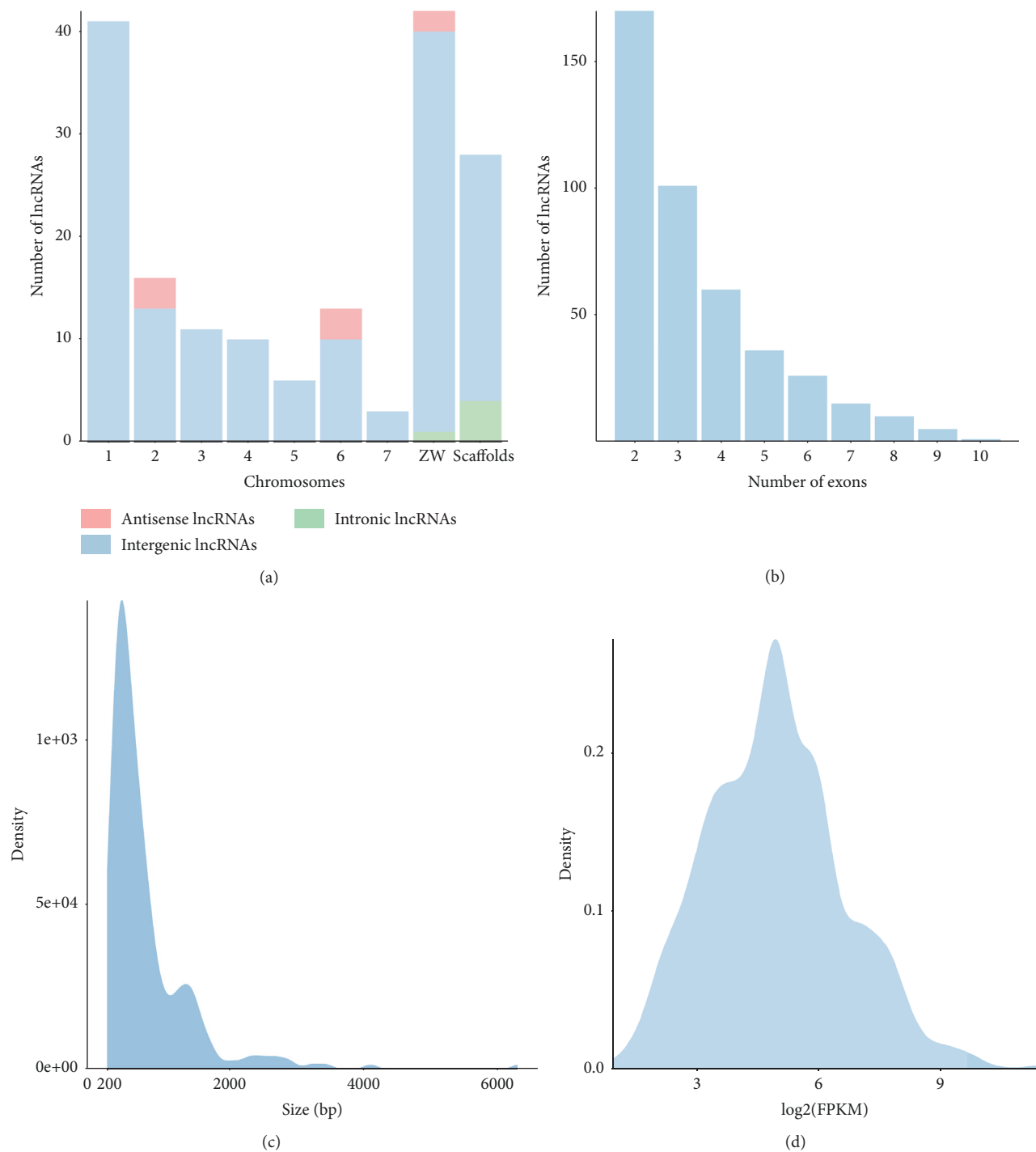


FIGURE 2: Features of *S. mansoni* adult lncRNAs. (a) Genomic localization of lncRNAs, (b) number of exons per transcripts, (c) transcripts length, (d) log₂ FPKM expression.

The largely enriched and meaningful BP terms were related to metabolism, transport, and biosynthesis. The enriched MF terms were predominantly related to binding including catalytic activity and nucleotide binding. As for CC, the most enriched terms were cell, intracellular, and cytoplasm. A detailed table of all 15 expressed lncRNAs was made including the neighboring coding genes and their respective GO entries (Supplementary Table S4).

4. Discussion

Several molecular and biochemical experiments revealed that lncRNAs may play diverse roles and functions in cellular biology. However, given the high abundance of lncRNAs and the poor-genetic conservation between species, the study of these molecules is extremely intriguing because they can be key regulators of species-specific biological processes.

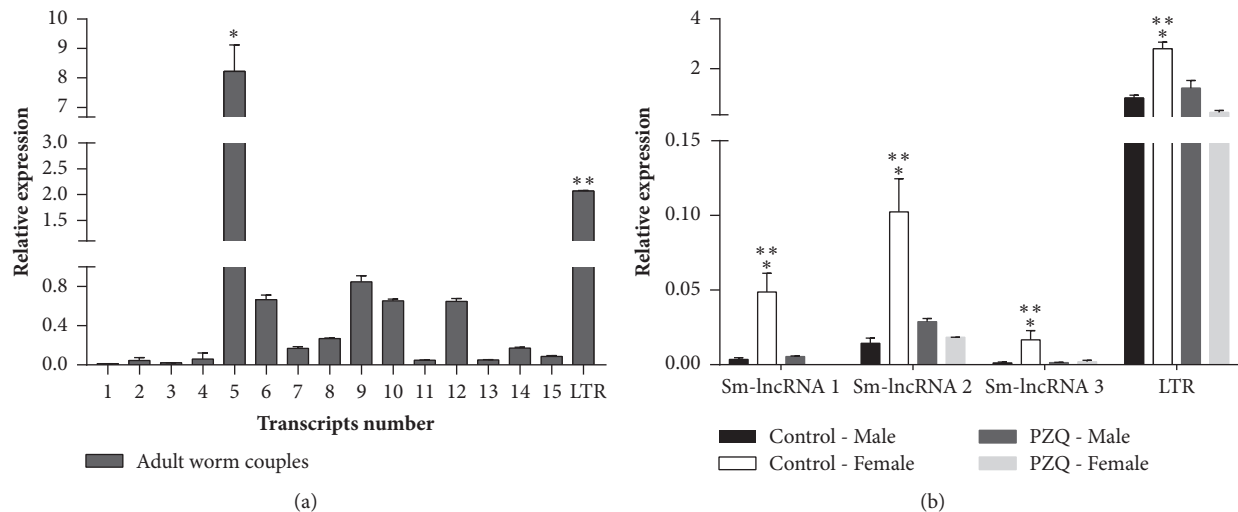


FIGURE 3: Relative expression of lncRNAs by RT-qPCR. (a) Fifteen Sm-lncRNAs and the LTR retrotransposon were selected for validation by RT-qPCR at the parasite adult stage (*significantly different from all other lncRNAs and LTR, **significantly different from all lncRNAs). (b) The first three lncRNAs and the LTR retrotransposon were selected to verify the expression between sexes into four groups Control-Male, Control-Female, PZQ-Male, and PZQ-Female (PZQ-Male and PZQ-Female are related to *S. mansoni* LE praziquantel-resistant strains) (*significantly different from Control-Male, ** different from PZQ-Female). Two-way ANOVA and Tukey's posttest were used for calculating the statistical significance (P value ≤ 0.05).

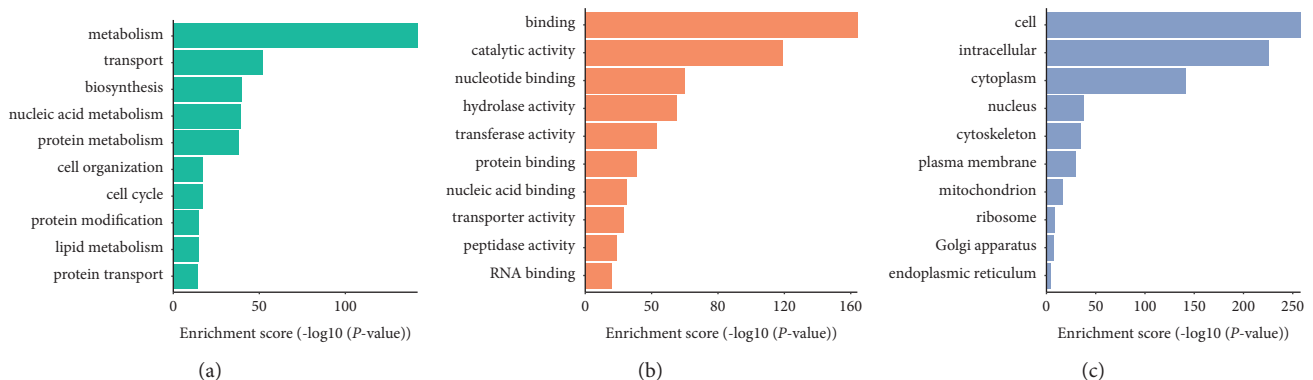


FIGURE 4: GO enrichment analysis of the lncRNA-target genes. The top 10 most enriched GO categories were calculated and plotted: (a) biological process; (b) molecular function; (c) cellular component.

Notably, there are great prospects for a better understanding of *S. mansoni* biology, but much work will be needed to elucidate the specific role of lncRNAs in this parasite lifestyle. Thus, pipeline development for the *S. mansoni* genome is of extreme importance due to the lack of adapted and specific methodologies for the genus *Schistosoma*.

Current methodologies for predicting lncRNA genes are species-specific following the genome intrinsic characteristics of each species which increase data generation sensitivity and specificity. Each methodology used follows a flow chart that best adapts the steps and programs used [42–45]. This is a very important point because *S. mansoni* has particular characteristics since some of the repetitive fractions of DNA consist of tandemly repeated ribosomal genes of which there are 500-1000 copies per genome that represent 1.8-3.6% of the total DNA and nonribosomal repetitive sequences that comprise at least a further 2.0% of the total DNA [46].

In this pipeline, we introduced a manual step to remove other RNAs, transposons, and predicted protein coding genes in other *Schistosoma* species. In Figure 1, 39% of the initially putative lncRNAs showed significant homology with retrotransposons. This finding indicates an important difference of our pipeline. In addition, 33.6% of the putative lncRNAs identified were identical to those recently described for *S. mansoni* [21], reinforcing the hypothesis that the steps which were described in both studies are ideal for the identification of lncRNA in the genus *Schistosoma*.

In this work, we selected some lncRNAs to initiate expression studies, using lncRNA assumptions as criteria that are located in the proximity of genes involved with important pathways to *S. mansoni* survival, like the ubiquitin proteasome system, ribosomal protein (Supplementary Table S4), and others. In total, 100% of novel lncRNAs analyzed were expressed in adult worm couples (Figure 3). This is consistent

with data from Vasconcelos et al. [21] who found similar percentages of expressed lncRNAs in *S. mansoni* adult worms.

The act of transcribing lncRNAs can have profound consequences on the ability of nearby genes to be expressed. For example, transcription of a lncRNA across the promoter region of a downstream protein coding gene directly interferes with transcription factor binding and thus prevents the protein coding gene from being expressed [47, 48]. We detected high levels of *Sm-lncRNA 5* expression related to proteasome B1 subunit, responsible for peptidylglutamyl activity. Curiously, it is the lowest activity of the proteasome of *S. mansoni* [49]. The proteasomes form a pivotal component for the ubiquitin proteasome system (UPS). It is implicated in protein ubiquitination, proteolysis, and degradation and was set as essential to *S. mansoni* biology [49]. Moreover, the *Sm-lncRNA 12* is differentially expressed in *S. mansoni* and occupies the neighborhood of E2 gene [50], leading us to hypothesize this participation on ubiquitin proteasome regulation. These findings reinforce themselves because differentially expressed lncRNAs, identified from human placentas, may regulate their associated mRNAs through several mechanisms and connect the UPS with infection-inflammation pathways [51].

Another set of putative lncRNAs, neighbor genes related to protein synthesis described in this work (Supplementary Table S4), are related to protein synthesis. Recently, lncRNAs have emerged as key players in the stress responses in plants [52] and eukaryotic cellular senescence [53]. The stress response is a feature of the *S. mansoni* lifestyle in both vertebrate and invertebrates hosts [54].

Furthermore, lncRNA levels dynamically change in response to various drugs. These alterations affect gene expression involved in several cells function such as cycle arrest, inhibition of apoptosis, and DNA damage repair [55]. *S. mansoni* exposure to a drug, particularly PZQ, which is used in a single dose or repeatedly in reinfection, may induce drug resistance or reduced susceptibility over time [56]. The difference in gene expression between male and female in PZQ resistance worms suggest that lncRNAs can also be involved in drug resistance mechanisms in *S. mansoni*.

Several studies have provided insights into how genomic neighborhoods could influence gene expression levels, with important consequences for evolution, development, and disease [57].

5. Conclusions

In conclusion, we have identified a novel set of 170 lncRNAs in *S. mansoni* expressed in male, female, and PZQ resistant adult worms. These observations suggest that lncRNAs may be significant in parasite biology and be useful therapeutic targets. Further studies are required to dissect the function and mechanism of action of these RNA subtypes in normal biology and life cycle progression.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the following Brazilian research agencies: FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais, CBB APQ-01880-13), NuBio UFOP (Núcleo de Bioinformática da Universidade Federal de Ouro Preto), and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). The authors thank the following transcriptome initiatives: the São Paulo Transcriptome Consortium, the Minas Gerais Genome Network, and the Wellcome Trust Genome Initiative (UK).

Supplementary Materials

Supplementary Table S1: primers pairs sequences selected for RT-qPCR experiments. *Supplementary Table S2*: detailed data obtained and mapping conditions for adult worm sample in RNA-seq. *Supplementary Table S3*: set of 170 putative lncRNAs identified by our computational pipeline. *Supplementary Table S4*: set of 15 expressed lncRNAs with the neighboring coding genes and their respective GO entries. (*Supplementary Materials*)

References

- [1] P. Steinmann, J. Keiser, R. Bos, M. Tanner, and J. Utzinger, "Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk," *The Lancet Infectious Diseases*, vol. 6, no. 7, pp. 411–425, 2006.
- [2] S. A.-L. Thétiot-Laurent, J. Boissier, A. Robert, and B. Meunier, "Schistosomiasis chemotherapy," *Angewandte Chemie International Edition*, vol. 52, no. 31, pp. 7936–7956, 2013.
- [3] D. J. Gray, A. G. Ross, Y.-S. Li, and D. P. McManus, "Diagnosis and management of schistosomiasis," *BMJ*, vol. 342, no. 7807, Article ID d2651, 2011.
- [4] WHO, "Schistosomiasis. Fact sheet No. 115, January 2012," <http://www.who.int/mediacentre/factsheets/fs115/en/index.html>, Accessed on Nov. 19, 2013.
- [5] H. You, D. P. McManus, and G. N. Gobert, "Current and prospective chemotherapy options for schistosomiasis," *Expert Opinion on Orphan Drugs*, vol. 3, no. 2, pp. 195–205, 2015.
- [6] M. J. Doenhoff, D. Cioli, and J. Utzinger, "Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis," *Current Opinion in Infectious Diseases*, vol. 21, no. 6, pp. 659–667, 2008.
- [7] P. Hagan, C. C. Appleton, G. C. Coles, J. R. Kusel, and L.-A. Tchuem-Tchuente, "Schistosomiasis control: keep taking the tablets," *Trends in Parasitology*, vol. 20, no. 2, pp. 92–97, 2004.
- [8] A. D. Aragon, R. A. Imani, V. R. Blackburn et al., "Towards an understanding of the mechanism of action of praziquantel," *Molecular and Biochemical Parasitology*, vol. 164, no. 1, pp. 57–65, 2009.
- [9] L. Pica-Mattoccia and D. Cioli, "Sex- and stage-related sensitivity of *Schistosoma mansoni* to in vivo and in vitro praziquantel treatment," *International Journal for Parasitology*, vol. 34, no. 4, pp. 527–533, 2004.

- [10] A. I. Grossman, R. B. Short, and G. D. Cain, "Karyotype evolution and sex chromosome differentiation in schistosomes (Trematoda, Schistosomatidae)," *Chromosoma*, vol. 84, no. 3, pp. 413–430, 1981.
- [11] A. V. Protasio, I. J. Tsai, A. Babbage et al., "A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*," *PLOS Neglected Tropical Diseases*, vol. 6, no. 1, Article ID e1455, 2012.
- [12] Z.-G. Han, P. J. Brindley, S.-Y. Wang, and C. Zhu, "Schistosoma genomics: new perspectives on schistosome biology and host-parasite interaction," *Annual Review of Genomics and Human Genetics*, vol. 10, pp. 211–240, 2009.
- [13] C. P. Ponting, P. L. Oliver, and W. Reik, "Evolution and functions of long noncoding RNAs," *Cell*, vol. 136, no. 4, pp. 629–641, 2009.
- [14] S. Leone and R. Santoro, "Challenges in the analysis of long noncoding RNA functionality," *FEBS Letters*, vol. 590, no. 15, pp. 2342–2353, 2016.
- [15] J. Cao, "The functional role of long non-coding RNAs and epigenetics," *Biological Procedures Online*, vol. 16, article 11, 2014.
- [16] R. Li, H. Zhu, and Y. Luo, "Understanding the functions of long non-coding RNAs through their higher-order structures," *International Journal of Molecular Sciences*, vol. 17, no. 5, article no. 702, 2016.
- [17] Y. Fang and M. J. Fullwood, "Roles, functions, and mechanisms of long non-coding RNAs in cancer," *Genomics, Proteomics & Bioinformatics*, vol. 14, no. 1, pp. 42–54, 2016.
- [18] I. Grammatikakis, A. C. Panda, K. Abdelmohsen, and M. Gorospe, "Long noncoding RNAs (lncRNAs) and the molecular hallmarks of aging," *AGING*, vol. 6, no. 12, pp. 992–1009, 2014.
- [19] L.-L. Chen and G. G. Carmichael, "Decoding the function of nuclear long non-coding RNAs," *Current Opinion in Cell Biology*, vol. 22, no. 3, pp. 357–364, 2010.
- [20] C. C. Copeland, M. Marz, D. Rose et al., "Homology-based annotation of non-coding RNAs in the genomes of *Schistosoma mansoni* and *Schistosoma japonicum*," *BMC Genomics*, vol. 10, article no. 1471, p. 464, 2009.
- [21] E. J. Vasconcelos, L. F. daSilva, D. S. Pires et al., "The *Schistosoma mansoni* genome encodes thousands of long non-coding RNAs predicted to be functional at different parasite life-cycle stages," *Scientific Reports*, vol. 7, no. 1, 2017.
- [22] J. Hines-Kay, P. M. Cupit, M. C. Sanchez, G. H. Rosenberg, B. Hanelt, and C. Cunningham, "Transcriptional analysis of *Schistosoma mansoni* treated with praziquantel in vitro," *Molecular and Biochemical Parasitology*, vol. 186, no. 2, pp. 87–94, 2012.
- [23] M. C. Sanchez, K. V. Krasnec, A. S. Parra et al., "Effect of praziquantel on the differential expression of mouse hepatic genes and parasite ATP binding cassette transporter gene family members during *Schistosoma mansoni* infection," *PLOS Neglected Tropical Diseases*, vol. 11, no. 6, Article ID e0005691, 2017.
- [24] F. F. B. Couto, P. M. Z. Coelho, N. Araújo et al., "Schistosoma mansoni: a method for inducing resistance to praziquantel using infected *Biomphalaria glabrata* snails," *Memórias do Instituto Oswaldo Cruz*, vol. 106, no. 2, pp. 153–157, 2011.
- [25] M. Berriman, B. J. Haas, P. T. Loverde et al., "The genome of the blood fluke *Schistosoma mansoni*," *Nature*, vol. 460, no. 7253, pp. 352–358, 2009.
- [26] S. Andrews, "FastQC: a quality control tool for high throughput sequence data," 2010, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- [27] A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: a flexible trimmer for Illumina sequence data," *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, 2014.
- [28] A. Dobin, C. A. Davis, F. Schlesinger et al., "STAR: ultrafast universal RNA-seq aligner," *Bioinformatics*, vol. 29, no. 1, pp. 15–21, 2013.
- [29] H. Li, B. Handsaker, A. Wysoker et al., "The sequence alignment/map format and SAMtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009.
- [30] C. Trapnell, A. Roberts, L. Goff et al., "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks," *Nature Protocols*, vol. 7, no. 3, pp. 562–578, 2012.
- [31] L. Wang, H. J. Park, and S. Dasari, "Coding-potential assessment tool using an alignment-free logistic regression model," *Nucleic Acids Research*, vol. 41, no. 6, article e74, 2013.
- [32] X. J. Min, G. Butler, R. Storms, and A. Tsang, "OrfPredictor: predicting protein-coding regions in EST-derived sequences," *Nucleic Acids Research*, vol. 33, no. 2, pp. W677–W680, 2005.
- [33] M. Clamp, B. Fry, M. Kamal et al., "Distinguishing protein-coding and noncoding genes in the human genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 49, pp. 19428–19433, 2007.
- [34] M. E. Dinger, K. C. Pang, T. R. Mercer, and J. S. Mattick, "Differentiating protein-coding and noncoding RNA: challenges and ambiguities," *PLoS Computational Biology*, vol. 4, no. 11, Article ID e1000176, 2008.
- [35] C. Camacho, G. Coulouris, V. Avagyan et al., "BLAST+: architecture and applications," *BMC Bioinformatics*, vol. 10, article 421, 2009.
- [36] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.
- [37] H. Wickham, *ggplot2: elegant graphics for data analysis*, Springer Publishing Company, Incorporated, 2009.
- [38] U. A. Ørom, T. Derrien, M. Beringer et al., "Long noncoding RNAs with enhancer-like function in human cells," *Cell*, vol. 143, no. 1, pp. 46–58, 2010.
- [39] S. Liu, P. Cai, N. Hou et al., "Genome-wide identification and characterization of a panel of house-keeping genes in *Schistosoma japonicum*," *Molecular and Biochemical Parasitology*, vol. 182, no. 1–2, pp. 75–82, 2012.
- [40] K. J. Livak and T. D. Schmittgen, "Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method," *Methods*, vol. 25, no. 4, pp. 402–408, 2001.
- [41] S. A. Bustin, V. Benes, J. A. Garson et al., "The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments," *Clinical Chemistry*, vol. 55, no. 4, pp. 611–622, 2009.
- [42] Y. Zhao, H. Luo, X. Chen, Y. Xiao, and R. Chen, "Computational methods to predict long noncoding RNA functions based on co-expression network," *Methods in Molecular Biology*, vol. 1182, pp. 209–218, 2014.
- [43] N. E. Iltott and C. P. Ponting, "Predicting long non-coding RNAs using RNA sequencing," *Methods*, vol. 63, no. 1, pp. 50–59, 2013.
- [44] Y. Zhang I, H. Huang, D. Zhang et al., "A review on recent computational methods for predicting noncoding RNAs," *BioMed Research International*, vol. 2017, pp. 1–14, 2017.
- [45] H. Luo, D. Bu, L. Sun, S. Fang, Z. Liu, and Y. Zhao, "Identification and function annotation of long intervening noncoding

- RNAs,” *Briefings in Bioinformatics*, vol. 18, no. 5, pp. 789–797, 2017.
- [46] A. J. G. Simpson, A. Sher, and T. F. McCutchan, “The genome of *Schistosoma mansoni*: Isolation of DNA, its size, bases and repetitive sequences,” *Molecular and Biochemical Parasitology*, vol. 6, no. 2, pp. 125–137, 1982.
- [47] J. A. Martens, L. Laprade, and F. Winston, “Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene,” *Nature*, vol. 429, no. 6991, pp. 571–574, 2004.
- [48] S. Katayama, Y. Tomaru, and T. Kasukawa, “Antisense transcription in the mammalian transcriptome,” *Science*, vol. 309, no. 5740, pp. 1564–1566, 2005.
- [49] R. Guerra-Sá, W. Castro-Borges, E. A. Evangelista, I. C. Kettelhut, and V. Rodrigues, “*Schistosoma mansoni*: functional proteasomes are required for development in the vertebrate host,” *Experimental Parasitology emphasizes*, vol. 109, no. 4, pp. 228–236, 2005.
- [50] M. P. Costa, V. F. Oliveira, R. V. Pereira et al., “In silico analysis and developmental expression of ubiquitin-conjugating enzymes in *Schistosoma mansoni*,” *Parasitology Research*, vol. 114, no. 5, pp. 1769–1777, 2015.
- [51] X. Luo, J. Pan, L. Wang et al., “Epigenetic regulation of lncRNA connects ubiquitin-proteasome system with infection-inflammation in preterm births and preterm premature rupture of membranes,” *BMC Pregnancy and Childbirth*, vol. 15, no. 1, Article ID 35, 2015.
- [52] J. Wang, X. Meng, O. B. Dobrovolskaya, Y. L. Orlov, and M. Chen, “Non-coding RNAs and their roles in stress response in plants,” *Genomics, Proteomics & Bioinformatics*, vol. 15, no. 5, pp. 301–312, 2017.
- [53] C. Kim, D. Kang, E. K. Lee, and J.-S. Lee, “Long noncoding RNAs and RNA-binding proteins in oxidative stress, cellular senescence, and age-related diseases,” *Oxidative Medicine and Cellular Longevity*, vol. 2017, pp. 1–21, 2017.
- [54] M. Knight, W. Ittiprasert, H. D. Arican-Goktas, and J. M. Bridger, “Epigenetic modulation, stress and plasticity in susceptibility of the snail host, *Biomphalaria glabrata*, to *Schistosoma mansoni* infection,” *International Journal for Parasitology*, vol. 46, no. 7, pp. 389–394, 2016.
- [55] L. Lipovich, R. Johnson, and C.-Y. Lin, “MacroRNA underdogs in a microRNA world: evolutionary, regulatory, and biomedical significance of mammalian long non-protein-coding RNA,” *Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms*, vol. 1799, no. 9, pp. 597–615, 2010.
- [56] A. Pinto-Almeida, T. Mendes, R. N. de Oliveira et al., “Morphological Characteristics of *Schistosoma mansoni* PZQ-Resistant and -Susceptible Strains Are Different in Presence of Praziquantel,” *Frontiers in Microbiology*, vol. 7, 2016.
- [57] N. E. Sanjana, J. Wright, K. Zheng et al., “High-resolution interrogation of functional elements in the noncoding genome,” *Science*, vol. 353, no. 6307, pp. 1545–1549, 2016.

