Analysis of Long Non-Coding RNA in *Cryptosporidium parvum* Reveals Significant Stage-Specific Antisense Transcription

Yiran Li[1], Rodrigo P. Baptista[1,2], Adam Sateriale[3†], Boris Striepen[3] and Jessica C. Kissinger[1,2,4*]

[1] Institute of Bioinformatics, University of Georgia, Athens, GA, United States, [2] Center for Tropical and Emerging Global Diseases, University of Georgia, Athens, GA, United States, [3] Department of Pathobiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, United States, [4] Department of Genetics, University of Georgia, Athens, GA, United States

Cryptosporidium is a protist parasite that has been identified as the second leading cause of moderate to severe diarrhea in children younger than two and a significant cause of mortality worldwide. *Cryptosporidium* has a complex, obligate, intracellular but extra cytoplasmic lifecycle in a single host. How genes are regulated in this parasite remains largely unknown. Long non-coding RNAs (lncRNAs) play critical regulatory roles, including gene expression across a broad range of organisms. *Cryptosporidium* lncRNAs have been reported to enter the host cell nucleus and affect the host response. However, no systematic study of lncRNAs in *Cryptosporidium* has been conducted to identify additional lncRNAs. In this study, we analyzed a *C. parvum in vitro* strand-specific RNA-seq developmental time series covering both asexual and sexual stages to identify lncRNAs associated with parasite development. In total, we identified 396 novel lncRNAs, mostly antisense, with 86% being differentially expressed. Surprisingly, nearly 10% of annotated mRNAs have an antisense transcript. lncRNAs occur most often at the 3′ end of their corresponding sense mRNA. Putative lncRNA regulatory regions were identified and many appear to encode bidirectional promoters. A positive correlation between lncRNA and upstream mRNA expression was observed. Evolutionary conservation and expression of lncRNA candidates was observed between *C. parvum*, *C. hominis* and *C. baileyi*. Ten *C. parvum* protein-encoding genes with antisense transcripts have *P. falciparum* orthologs that also have antisense transcripts. Three *C. parvum* lncRNAs with exceptional properties (e.g., intron splicing) were experimentally validated using RT-PCR and RT-qPCR. This initial characterization of the *C. parvum* non-coding transcriptome facilitates further investigations into the roles of lncRNAs in parasite development and host-pathogen interactions.

Keywords: lncRNA, RNA regulation, Apicomplexa, parasite development, stranded RNA-Seq, transcriptome

INTRODUCTION

Cryptosporidium is an obligate protist parasite that causes a diarrheal disease called cryptosporidiosis which spreads via an oral-fecal route. Human cryptosporidiosis, mainly caused by *Cryptosporidium parvum* and *Cryptosporidium hominis*, is typically self-limited and causes 1~2 weeks of intense watery diarrhea in people with healthy immune systems. However, the illness may be lethal among the immunocompromised including individuals with AIDS, cancer, and those receiving transplant anti-rejection medications. In recent years, several *Cryptosporidium* species, predominantly *C. hominis*, have been identified as the second most prevalent diarrheal pathogen of infants globally after rotavirus (Bouzid et al., 2013; Kotloff et al., 2013; Painter et al., 2015; Platts-Mills et al., 2015; Sow et al., 2016) and a leading cause of waterborne disease among humans in the United States (Centers for Disease Control and Prevention)[1]. Thus far, Nitazoxanide, the only FDA-approved drug is not effective for use in infants or those with HIV-related immunosuppression (Amadi et al., 2009) i.e. the most susceptible populations, and no vaccine is available (Amadi et al., 2002).

Cryptosporidium has a complex lifecycle in a single host. The *Cryptosporidium* oocyst which is shed in feces is a major extracellular lifecycle stage. It can stay dormant and survive in the environment for months (Drummond et al., 2018). After ingestion of oocysts through contaminated water or food, sporozoites are released which are capable of invading intestinal epithelial cells where both asexual and sexual replication occur. Following invasion, sporozoites develop into trophozoites and undergo asexual replication to generate type I meronts and type II meronts. Type I meronts are thought to be capable of reinvading adjacent cells generating an asexual cycle (Fayer, 2008), while Type II meronts contribute to the formation of microgamonts (male) or macrogamonts (female) to complete the sexual stages (Bouzid et al., 2013). Conventional monolayer cell culture does not permit completion of the life cycle much beyond 48 h post-infection (hpi) (**Figure 1A**), for as of yet poorly understood reasons but gametogenesis does occur (Tandel et al., 2019). The lack of *in vitro* culture has historically impeded the development of new drugs and vaccines for this medically important parasite. Recently, there have been several breakthroughs including genetic manipulation (Vinayak et al., 2015; Sateriale et al., 2020) and lifecycle completion. Several promising approaches have been developed including using a cancer cell line as host (Miller et al., 2018), biphasic and three-dimensional (organoid) culture systems, (Morada et al., 2016; DeCicco RePass et al., 2017; Heo et al., 2018; Cardenas et al., 2020), hollow fiber technology (Yarlett et al., 2020), and air-liquid interface (ALI) cultivation system (Wilke et al., 2019; Wilke et al., 2020). These breakthroughs are enabling better, much needed, studies of the parasite's full life cycle. A better understanding of the conditions and regulatory processes necessary for *Cryptosporidium* development are essential and will prove beneficial for the identification of drug and vaccine targets.

---

[1]https://www.cdc.gov/parasites/crypto/index.html.

The first genome sequence of *C. parvum* was published in 2004 with a genome size of ~9 Mb and ~3800 protein-encoding genes annotated (Abrahamsen et al., 2004). Since this milestone, our understanding of the parasite and its biology have progressed remarkably. Early *in vitro* transcriptome analyses using semi-quantitative RT-PCR over a 72 h post-infection (pi) time course during *in vitro* development revealed complex and dynamic gene expression profiles. Adjacent genes are not generally co-regulated, despite the highly compact genome (Mauzy et al., 2012). Under UV irradiation, *C. parvum* oocysts have shown a vital stress-induced gene expression response according to microarray data (Zhang et al., 2012). mRNA expression related to gametocyte and oocyst formation were studied using RNA sequencing of sorted cells (Tandel et al., 2019). Yet, little is known about the regulation of key developmental transitions. How the parasite regulates gene expression in order to invade hosts, amplify, evade the immune system and interact with their host awaits further discovery.

Most canonical eukaryotic enhancer proteins are not detected in Apicomplexa, the phylum that *Cryptosporidium* belongs to (Iyer et al., 2008), except for the transcriptional activators Myb and zinc finger proteins C2H2 (Cys2His2) and two additional transcription factor families. Instead, an expanded family of apatela-related transcription factors, the AP2 family of proteins (ApiAP2), appear to be the predominant transcription factors in this phylum, including *Cryptosporidium* (Oberstaller et al., 2014; Jeninga et al., 2019). AP2 domains in *C. parvum* are reported to have reduced binding diversity relative to the malaria parasite *Plasmodium falciparum* and proposed to possess less dominancy in transcriptional regulation in *C. parvum* (Campbell et al., 2010; Yarlett et al., 2020). It has been proposed that *C. parvum* is less reliant on ApiAP2 regulators in part because it utilizes E2F/DP1 transcription factors, which are present in *Cryptosporidium* while absent in other studied apicomplexans (Templeton et al., 2004; Yarlett et al., 2020). Based on the similarity of gene expression profiles, it has been suggested that the number of co-expressed gene clusters in *C. parvum* is somewhere between 150 and 200, and putative ApiAP2 and E2F/DP1*cis*-regulatory elements were successfully detected in the upstream region of many co-expressed gene clusters (Oberstaller et al., 2013). Additionally, low levels of DNA methylation in oocysts has been reported in several *Cryptosporidium* spp. (Gong et al., 2017), suggesting the requirement of additional regulatory mechanisms. At the level of post-transcriptional regulation, the RNA interference (RNAi) pathway, which plays a crucial role in gene silencing in most eukaryotes, is considered to be missing in *Cryptosporidium* due to the lack of identifiable genes encoding the microRNA processing machinery or RNA-induced silencing complex (RISC) components (Keeling, 2004). There is much that remains to be discovered with respect to regulation of gene expression in *Cryptosporidium*.

Long non-coding RNAs (lncRNAs) are transcripts without significant protein-encoding capacity that are longer than 200 nt. In eukaryotes, lncRNAs play critical regulatory roles in gene regulation at multiple levels, including transcriptional, post-transcriptional, chromatin modification and nuclear architecture alterations (Marchese et al., 2017). In humans,
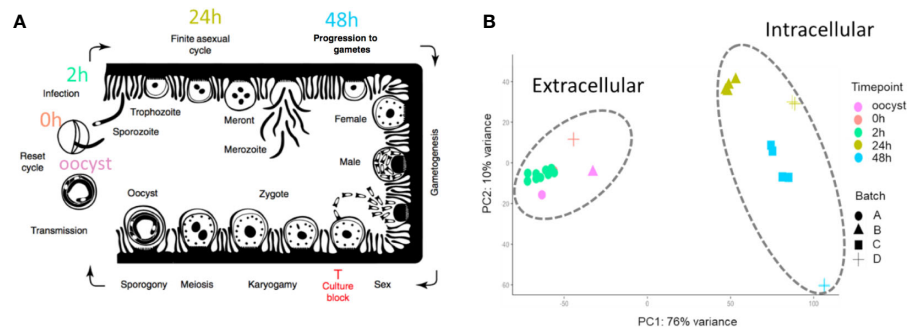
FIGURE 1 | The 33 RNA-Seq data sets used for expression analysis. **(A)** The time points indicate when RNA-Seq samples were collected and the associated *C. parvum* life cycle stage. The schematic model of the *C. parvum* life cycle is modified from (Tandel et al., 2019). "Culture block" refers to the developmental stage where parasites in traditional *in vitro* cultures fail to progress. **(B)** Principal component analysis of 33 *C. parvum* transcriptomes. The analysis is based on the normalized expression level (VST) of the *C. parvum* mRNA and predicted lncRNA genes. Samples collected from different time points are indicated by colors. Various projects/batches are represented by shapes. Batch A includes sample IDs 3–19 without host cells. Batch B includes sample IDs 1, 23–26 from IPEC host cells. Batch C includes sample IDs 27–30 from MDBK host cells. Batch D contains sample IDs 2, 20–22, 31–33 from HCT-8 host cells (**Table 1**).

3300 long intergenic ncRNAs (lincRNAs) were analyzed using chromatin state maps, and ~20% of these RNAs are bound to the polycomb repressive complex (PCR2, a complex with histone methyltransferase activity) (Khalil et al., 2009). Most lncRNAs share many characteristics of mRNAs, such as RNA polymerase II-mediated transcription, a 5′ 7-methylguanosine cap and a 3′ poly(A) tail [6]. The expression of lncRNAs is usually more tissue- or time-specific than mRNA expression (Necsulea et al., 2014; Tsoi et al., 2015). lncRNA sequences are not well conserved across species, but their structure could be conserved due to functional constraints (Ulitsky et al., 2011; Diederichs, 2014). By forming hybrid structural complexes such as RNA-DNA hybrid duplexes or RNA-DNA triplexes, lncRNAs can recruit or scaffold protein complexes to facilitate localization of protein machinery to specific genome target sites (Li et al., 2016). lncRNAs play important roles in regulating occurrence and progression of many diseases. After infected by *C. baileyi*, significant expression changes have been observed in the host (Ren et al., 2018). The mis-regulation of lncRNAs in multi-cellular eukaryotes has been shown to cause tumorigenesis (Chakravarty et al., 2014), cardiovascular diseases (Tang et al., 2019), and neurodegenerative dysfunction (Zhang et al., 2018) and thus can be used as diagnostic biomarkers.

Taking advantage of sequencing technologies, numerous lncRNA candidates have been detected in apicomplexans, some with proven regulatory potential during parasite invasion and proliferation processes. These discoveries have ushered in a new era in parasite transcriptomics research (Liao et al., 2014; Siegel et al., 2014; Broadbent et al., 2015; Ramaprasad et al., 2015; Filarsky et al., 2018). In *P. falciparum*, lncRNAs are critical regulators of virulence gene expression and associated with chromatin modifications (Vembar et al., 2014). Likewise, an antisense RNA of the gene *gdv1* was shown to be involved in regulating sexual conversion in *P. falciparum* (Filarsky et al., 2018). In *C. parvum*, putative parasite lncRNAs were found to be delivered into the host nucleus, some of which were experimentally proven to regulate host

genes by hijacking the host histone modification system (Ming et al., 2017; Wang et al., 2017a; Wang et al., 2017b). The importance of lncRNA in *C. parvum* has been demonstrated, but no systematic annotation of lncRNA has been conducted. The systematic identification of lncRNAs will increase the pool of candidate regulatory molecules thus ultimately leading to increased knowledge of the developmental gene regulation in *C. parvum* and control of parasite interactions with its hosts.

In this study, we developed and applied a computational pipeline to systematically identify new lncRNAs in the *C. parvum* genome. We used a set of stranded RNA-seq data collected from multiple lifecycle stages that cover both asexual and sexual developmental stages. We conducted a systematic analysis of lncRNA that includes sequence characteristics, conservation, expression profiles and expression relative to neighboring genes. The results provide new insights into *C. parvum*'s non-coding potential and suggest several areas for further research.

METHODS AND MATERIALS

RNA-Seq Data Pre-Processing/Cleaning

RNA-Seq data sets were downloaded from the NCBI Sequence Read Archive (SRA) and European Nucleotide Archive database (ENA). Detailed information on SRA accession numbers and Bioprojects are listed in **Table 1** and **Supplementary Table S1**. FastQC-v0.11.8 was used to perform quality control of the RNA-Seq reads. Remaining adapters and low-quality bases were trimmed by Trimmomatic-v-0.36 (Bolger et al., 2014) with parameters: Adapters:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:25 MINLEN:25. All reads were scanned with a four-base sliding window and cut when the average Phred quality dropped below 25. Bases from the start and end were removed when the quality score was below 20. The minimum read length was set at 25 bases. The processed reads are referred to as cleaned reads in this work.

TABLE 1 | RNA-Seq data sets used in this study.

ID*	Time point	Host cells	Condition
C. parvum			
1B	Oocyst	NA	Extracellular
2D	Oocyst	NA	Extracellular
3A	0 h	NA	Extracellular
4A	0 h	NA	Extracellular
5A	0 h	NA	Extracellular
6A	0 h	NA	Extracellular
7A	0 h	NA	Extracellular
8A	0 h	NA	Extracellular
9A	0 h	NA	Extracellular
10A	2 h	NA	Extracellular
11A	2 h	NA	Extracellular
12A	2 h	NA	Extracellular
13A	2 h	NA	Extracellular
14A	2 h	NA	Extracellular
15A	2 h	NA	Extracellular
16A	2 h	NA	Extracellular
17A	2 h	NA	Extracellular
18A	2 h	NA	Extracellular
19A	2 h	NA	Extracellular
20D	24 h	HCT-8	Intracellular
21D	24 h	HCT-8	Intracellular
22D	24 h	HCT-8	Intracellular
23B	24 h	IPEC	Intracellular
24B	24 h	IPEC	Intracellular
25B	24 h	IPEC	Intracellular
26B	24 h	IPEC	Intracellular
27C	48 h	MDBK	Intracellular
28C	48 h	MDBK	Intracellular
29C	48 h	MDBK	Intracellular
30C	48 h	MDBK	Intracellular
31D	48 h	HCT-8	Intracellular
32D	48 h	HCT-8	Intracellular
33D	48 h	HCT-8	Intracellular
C. hominis			
34E	Oocyst	NA	Extracellular
35E	Oocyst	NA	Extracellular
36F	Oocyst	NA	Extracellular
37F	Oocyst	NA	Extracellular
C. baileyi			
38G	Oocyst/sporozoite	NA	Extracellular

*Batch with various host cell types and parasites from different projects are indicated with the sample IDs. IDs designated with A, C and E, (Matos et al., 2019); B, (Mirhashemi et al., 2018); D, (Tandel et al., 2019); 36F: SRR1183950; 37F: SRR1183934; 38G: SRR1183952.
Cell line synonyms/origin: HCT-8: Human intestine; IPEC: Intestinal Porcine Epithelial Cell line; MDBK: Bovine kidney.

Read Mapping and Transcript Assembly

Cleaned reads from each sample were mapped to the reference genome sequence for *Cryptosporidium parvum* IOWA-ATCC (BioProject ID:PRJNA573722, Baptista et al. *in prep*) downloaded from CryptoDB v46 (https://cryptodb.org/cryptodb/) using the mapping tool HISAT2-v2.1.0 (Kim et al., 2015) with maximum intron length (–max-intronlen) set at 3000, and remaining parameters as default. Uniquely mapped reads were selected for further study using SAMtools-v1.10 (view -q 10) (Li et al., 2009). StringTie-v2.0.6 (Pertea et al., 2015) was used to reconstruct transcripts using the reference annotation guided method (–rf -j 5 -c 10 -g 1). At least five reads were

required to define a splice junction. A minimum read coverage of 10 was used for transcript prediction. Only overlapping transcript clusters were merged together. The stranded library types were all "fr-firststrand." Transcripts with FPKM lower than three were removed. The transcriptome assemblies from multiple samples were merged into one master transcript file using TACO-v0.7.3 with default settings (Niknafs et al., 2017).

lncRNA Prediction

Transcripts that overlapped with currently annotated mRNAs in the *C. parvum* IOWA-ATCC annotation in CryptoDB v46 with coverage >70% on the same strand were removed using BEDTools-v2.29.2 (Quinlan and Hall, 2010). The remaining transcripts were examined for coding potential using the online tool Coding Potential Calculator (CPC) v0.9 (Kong et al., 2007). Transcripts considered as "coding" by CPC were removed. Potential read-through transcripts resulting from transcription of neighboring mRNAs were removed using two criteria: 1) The transcript was <50 bp from the upstream coding region of another gene on the same strand. 2) The transcript was always transcribed together with the upstream mRNA on the same strand. Finally, the remaining transcripts which occurred in >2 RNA-Seq samples were kept as putative lncRNA candidates for further studies.

Transcriptome Data Normalization and Identification of Differentially Expressed Genes

The raw read counts for both mRNA genes and predicted lncRNAs were calculated using HTSeq-v0.12.4 (Anders et al., 2015). All genes were filtered to require > 50 reads in at least three samples. Variance stabilizing transformation (VST) from DESeq2-v1.28.1 (Love et al., 2014) was used to normalize the expression between samples. Principal components analysis (PCA) of the RNA-Seq samples was performed using the resulting VST values. VST values for each of the mRNA and lncRNA candidates were visualized using the R package pheatmap-v1.0.12 (https://github.com/raivokolde/pheatmap). The K-mean approach in pheatmap was applied to cluster genes based on the expression data.

Differentially expressed genes including mRNAs and lncRNAs were analyzed by EdgeR-v3.30.3 (Robinson et al., 2010) using the raw read count from HTSeq-v0.12.4. The expression time-points compared were between oocyst/sporozoites (oocysts, time point 0 h and sporozoites, 2 h), asexual stage (time point 24 h) and mixed asexual/sexual stage (only select samples from 48 h time point in which gametocyte marker genes are clearly expressed, i.e. batch D). The sex marker gene: cgd6_2090 encodes *Cryptosporidium* oocyst wall protein-1 (COWP1) produced in female gametes and cgd8_2220 encodes the homolog of hapless2 (HAP2) a marker of male gamonts (Tandel et al., 2019). A generalized linear model (GLM) approach was used for differential expression hypothesis testing. P-values were adjusted by the false discovery rate (FDR). Significant differentially expressed genes were declared at a log2-fold change ≥ 1.5 and an FDR < 0.05.

Expression Correlation

The expression correlation analysis between predicted lncRNAs and mRNAs was conducted with normalized expression data from all 33 samples (**Table 1**) of *C. parvum* using the Pearson test. P-values were adjusted by FDR.

Upstream Motif Analysis

MEME v5.0.0 (Bailey and Elkan, 1994) was used to discover motifs that may be present upstream of putative lncRNAs. For the promoter motif search, we extracted the 100 bp of (+) strand sequence upstream of the predicted lncRNAs and searched both strands using MEME for motifs with length six to 50 bp. The parameters were: -dna -mod anr -nmotifs 15 -minw 6 -maxw 50 -objfun classic -markov_order 0 -revcomp.

Evolutionary Conservation

RNA-Seq data sets from *C. hominis* and *C. baileyi* oocysts were mapped to reference genome sequences for *C. hominis* 30976 and *C. baileyi* TAMU-09Q1, respectively, downloaded from CryptoDB v46 (https://cryptodb.org/cryptodb/). Mapped reads were assembled into transcripts using the same methods mentioned above. tblastx from NCBI BLAST v 2.10.0 was used to search for conserved antisense candidates among *C. parvum*, *C. hominis* and *C. baileyi*, with parameter of E-value:1e-5 and best hit retained. To assess conservation among other apicomplexans, *P. falciparum*, antisense and associated sense mRNA data were retrieved from (Broadbent et al., 2015). *P. falciparum* orthologs of the sense mRNAs in *C. parvum* were retrieved from OrthoMCL DB v6.1 (Li et al., 2003). If antisense RNAs were detected between *P. falciparum* and *C. parvum* orthologs, the lncRNAs between *P. falciparum* and *C. parvum* were considered orthologous.

RT-qPCR Validation

We designed PCR primers using the PrimerQuest Tool from IDT (https://www.idtdna.com/pages/tools/primerquest) (**Supplementary Table S2**) to use for expression validation and exon structure confirmation of select lncRNA candidates using RT-PCR and qPCR. RNA was extracted from oocysts and provided by Boris Striepen. The cDNA for each sample was reverse transcribed using the iScript™ cDNA Synthesis Kit (Bio-Rad, Hercules, CA) from 1μg of input RNA. The resulting cDNAs were used as templates for PCR amplification and qPCR detection. Strand-specific primers were designed to amplify antisense RNAs. The 18S rRNA gene of *C. parvum* was used as positive control and samples without RNA or primer were used as a negative control. Each RT-PCR reaction contained 1ul cDNA, 2 μl primer mix (10 μM), 2 μl water and 5 μl MyTaq™ HS Mix (Bioline). RT-PCR was performed in the following conditions: 35 cycles of 15 s at 95°C, 30 s at 64°C. Then the RT-PCR products were run on a 2% agarose gel. The cDNA was also subjected to qPCR with All-in-One qPCR Mix (QP001; GeneCopoeia, Rockville, MD, USA) using the Mx3005P qPCR system (Agilent Technologies, Santa Clara, CA, USA). All reactions, including no-template controls, were run in triplicate. Following amplification, the CT values were determined using fixed threshold settings. lncRNA expression was normalized to 18S rRNA expression.

Data Availability

GenBank accession records CP044415–CP044422 have been updated to include annotation of the lncRNAs identified in this study. These data have also been submitted to CryptoDB.org.

RESULTS

Analysis of the *Cryptosporidium* Stranded RNA-Seq Data Used in This Study

To identify and investigate the expression profile of lncRNAs in *C. parvum* during parasite development, we searched for stranded RNA-Seq data sets from *Cryptosporidium* available in public databases. Due to the small volume of *Cryptosporidium* relative to the host cells, RNA-Seq data usually suffer from high host contamination. In this study, we selected samples that had more than 100k *Cryptosporidium* reads generated from the Illumina platform mapped to the reference genome sequence to reduce bias mostly arising from sequencing platform and sequencing depth. In total, 38 stranded-RNA-seq data sets which originated from 33 *C. parvum* samples, four *C. hominis* samples and one *C. baileyi* sample were selected for further analysis. The details and mapping statistics of each sample are shown in **Table 1** and (**Supplementary Table S1**). The *C. parvum* samples represented five lifecycle time points: oocyst, 0 h (sporozoites immediately after oocyst excystation), 2 h (2-h incubation in the medium (Matos et al., 2019)), 24 h (24-h post host cell infection) and 48 h (48-h post host cell infection) (**Figure 1A**). The 24- and 48-h samples were derived from infections in different types of host cells (see details in **Table 1** and **Supplementary Table S1**). The *C. hominis* RNA-Seq and one *C. baileyi* RNA-seq sample were obtained from oocysts.

The transcriptomes of the 33 *C. parvum* RNA-Seq samples were compared by principal component analysis (PCA) based on the normalized mRNA gene expression level in each sample (**Figure 1B**). Extracellular stages, including oocyst and sporozoites from 0 h and 2 h, are differentiated from intracellular stages, including 24 h and 48 h. The transcriptomes of intracellular stages were demonstrated to be more heterogeneous, while extracellular samples formed a relatively compact cluster. This observation was consistent with a previous transcriptome study of *C. parvum* oocysts and intracellular stages (Matos et al., 2019). At time points 24 h and 48 h, different host cells and laboratory procedures were used, which could contribute to the distance observed between samples from the same time point.

To further explore whether sexual commitment was initiated in all 48 h samples, we profiled the transcriptome of marker genes cgd6_2090 and cgd8_2220 (**Supplementary Figure 1**). cgd6_2090 encodes the *Cryptosporidium* oocyst wall protein-1 (COWP1) which is produced in female gametes (Tandel et al., 2019); cgd8_2220 encodes the homolog of hapless2 (HAP2) which is a class of membrane fusion protein required for gamete fusion in a range of organisms including *Plasmodium falciparum* (Liu et al., 2008). HAP2 labeled protein was exclusively found in male gamonts in *C. parvum* (Tandel et al., 2019). In another study, the *C. parvum* transcriptome was elucidated over a 72 h *in vitro* time-course infection with

HCT8 cells using semi-quantitative RT-PCR (Mauzy et al., 2012). The 48 h-specific genes from that study were also examined in the 33 RNA-Seq data sets analyzed here (**Supplementary Figure 2**). Both the sex marker genes and 48 h-specific genes show an expression peak at 48 hr. At 48 h, expression levels from batch D samples are much higher than batch C samples. cgd8_2220 is very low/not expressed at 48 h in batch C samples. These results indicate that both batch C and D samples showed the commitment of sexual development, but commitment was more pronounced in batch D. It is possible that sequencing depth could be influencing this difference as the normalization process (VST) between samples tends to reduce the variation of genes with low read support. Interestingly, batch C and D samples used different host cell types (**Table 1**). Batch D *C. parvum* parasites were cultured in HCT-8 (Human intestine cells) while batch C parasites were cultured in MDBK (Bovine kidney cells). Although sexual stages have been observed in MDBK cells (Villacorta et al., 1996), it is possible that the adaptation of *C. parvum* to MDBK cells is lower than HCT-8 cells as hosts. Thus, a slower or lower conversion rate was observed.

Identification and Characteristics of lncRNAs

We began assembly of the *C. parvum* transcriptome using mapped RNA-Seq reads of samples in NCBI BioProject PRJNA530692 with a high sample quality. However, *C. parvum* has an extremely compact genome sequence. As calculated from the *C. parvum* IOWA-ATCC annotation, the average intergenic distance between the stop and start codon boundaries of neighboring genes is only 504 bp and this distance must also include promoter and UTR regions. The average length of an annotated *C. parvum* ATCC mRNA coding sequence, CDS, is 1853 bp. The high gene density and the short distance between genes make it difficult to accurately define UTR boundaries using short-read sequencing data as transcripts overlap and become merged. Transcriptome assembly using RNA-Seq without genome and reference annotation guidance leads to a high chimerism rate. Thus, we used reference annotation to guide the assembly process and set parameters to minimize the number of artificially fused transcripts. We then used the program TACO v0.7.3 (Multi-sample transcriptome assembly), which employs change point detection to break apart complex loci to lower the number of fused transcripts to obtain a non-redundant master transcriptome from all samples, resulting in 5818 transcripts in total. Of these, 4846 transcripts overlapped with an mRNA on the same strand and thus, were removed. Transcripts which were <200 bp or only detected in a single sample were removed to improve the lncRNA prediction quality. Transcripts that were considered as "coding" by the Coding potential analysis tool CPC were filtered out. To identify and remove potential read-through transcripts, predicted transcripts that were located closer than 50 bp from the coding region of the upstream gene on the same strand and always transcribed together with the upstream mRNA were removed (**Figure 2A**).

In total, 396 transcripts were selected as lncRNA candidates for further analysis, 363 (91.7%) are antisense to mRNAs and only 33 are lincRNAs (**Figure 2B**). This high percentage of antisense RNA transcripts is a noteworthy feature of the compact *C. parvum* genome given that CDSs cover >70% of the genome sequence. Interestingly, the nucleotide composition varies between antisense RNAs and lincRNAs with many antisense RNAs being CT-rich and most of the lincRNAs being AT-rich (**Supplementary Figure 3**).
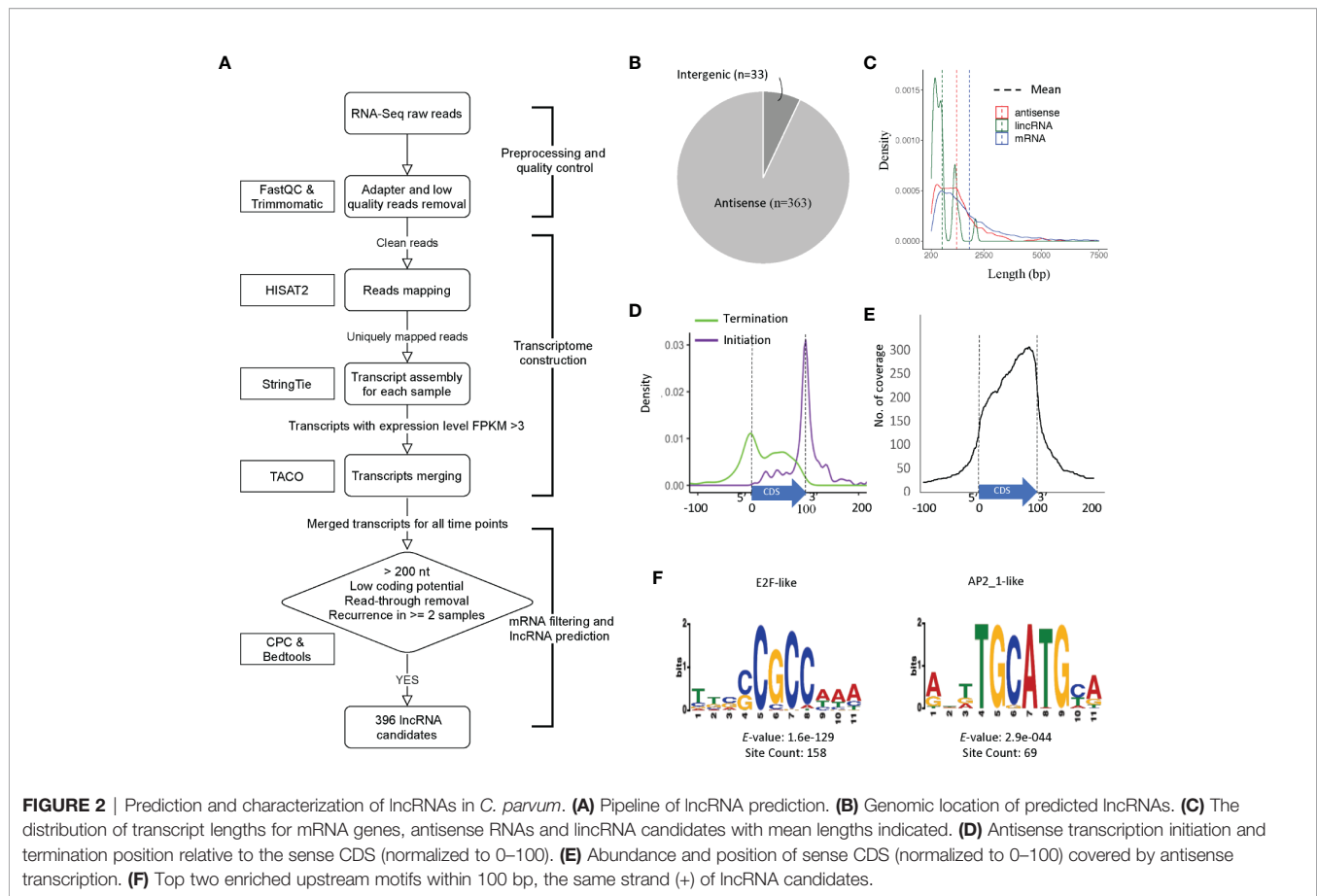
Most of the lncRNAs we detected consist of a single exon however five lncRNAs contain introns. This is consistent with the low-intron rate in *Cryptosporidium*. Additional introns are expected to emerge with deeper RNA-Seq data. The average length of lncRNAs and annotated mRNAs is 1267 bp and 1853 bp, respectively (**Figure 2C**). When compared to mRNAs, one of the most distinguishing features of lncRNAs is the low Open Reading Frame (ORF) coding potential relative to the transcript length.

We further investigated the relative location of the antisense transcripts relative to the annotated sense mRNA coding sequence (CDS), and we found that many of lncRNAs' initiation and termination sites are located close to the sense CDS boundaries. The start site of the antisense lncRNA transcript is highly correlated with the end of the sense strand CDS (**Figure 2D**). When looking at the coverage of the antisense transcript along the mRNA, the *C. parvum* lncRNA antisense expression has a bias towards the 3′ end of the mRNA transcript (**Figure 2E**). This property has also been reported in other organisms, including the malaria pathogen *Plasmodium falciparum* (Lopez-Barragan et al., 2011; Siegel et al., 2014; Broadbent et al., 2015).

To better understand their transcription, we searched for potential promoter motifs within 100 bp of the (+) strand upstream from all 396 lncRNAs. This analysis returned five significant motifs (E-value <0.001) for 190 lncRNAs (**Supplementary Figure 4** and **Supplementary Table S3**). The top two motifs are known *Cryptosporidium* transcriptional factor binding motifs E2F/DP1 (5′-[C/G]GCGC[G/C]-3′) and ApiAP2_1 (5′-BGCATGCAH-3′) motifs (Oberstaller et al., 2013), both of which are palindromic (**Figure 2F**). Of the 190 lncRNAs with significant upstream motifs, 142 lncRNAs had only a single type of motif (93 with E2F/DP1 and 19 with ApiAP2_1). 70 (75.3%) of the 93 lncRNAs with upstream E2F/DP1 motifs had only one motif, 21(22.6%) had two occurrences. Of the 19 lncRNAs with upstream ApiAP2_1 motifs, 12 (63.2%) had only one occurrence of the motif and 6 (31.6%) had two. Co-occurrence of the E2F/DP1 and ApiAP2_1 motif was only observed upstream of 17 lncRNAs. The spacing of the motifs relative to the transcription start site was assessed and no clear pattern was observed (**Supplemental Figure 5**). These combined results suggest that lncRNA transcripts have the potential for being regulated independently during parasite development.

lncRNAs Are Developmentally Regulated

We visualized gene expression profiles across the 33 RNA-Seq samples used in this study (**Table 1**) for both mRNA (**Figure 3A**)

FIGURE 2 | Prediction and characterization of lncRNAs in *C. parvum*. **(A)** Pipeline of lncRNA prediction. **(B)** Genomic location of predicted lncRNAs. **(C)** The distribution of transcript lengths for mRNA genes, antisense RNAs and lincRNA candidates with mean lengths indicated. **(D)** Antisense transcription initiation and termination position relative to the sense CDS (normalized to 0–100). **(E)** Abundance and position of sense CDS (normalized to 0–100) covered by antisense transcription. **(F)** Top two enriched upstream motifs within 100 bp, the same strand (+) of lncRNA candidates.

and lncRNAs (**Figure 3B**). To identify genes with a similar expression profile, we used a k-means algorithm to cluster mRNAs and lncRNAs separately (**Supplementary Table S4** and **Supplementary Table S5**). The k value was selected as the smallest value that allowed the separation of genes from different time points while keeping genes from different samples of the same timepoint together despite the batch effects present within each time point. As a result, mRNAs and lncRNAs were clustered into seven and nine broad co-expression groups, respectively. Select highly expressed lncRNAs from clusters 2, 4 and 5 are visualized as examples (**Supplementary Figure 6**).

The expression of mRNAs in the extracellular stages (oocysts and sporozoites from 0h and 2h) showed a similar expression trend that is quite distinct from the latter two stages. For mRNAs, genes from cluster 1 and cluster 2 were more highly expressed in the extracellular stages but still show expression in the intracellular stages when the vast majority of mRNAs are active. This result is consistent with another transcriptome study of *C. parvum*, which used semi-quantitative RT-PCR over a 72 h time course during *in vitro* development (Mauzy et al., 2012). On the contrary, many lncRNAs showed enriched expression in extracellular stages (oocyst, 0 h, 2 h) and some had expression later at the intracellular sexual development stage (48 h) (**Supplementary Table S6**) while the asexual stage (24 h) showed the least lncRNA expression. When considering the

correlation of antisense transcription to parasite development (according to the peak expression time point), 162 antisense transcripts were classified as extracellular stage enriched (oocyst, 0 h and sporozoite 2 h), 50 as merozoite enriched (24 h) and 151 as sexual stage enriched (48 h). It is noteworthy that both mRNAs and lncRNAs have gene sets that are specifically turned on at 48 h (mRNA cluster 5 and lncRNA cluster 4). The average expression level of lncRNAs suggests they are more abundant or were actively expressed in oocysts, 0 h, 2 h and gamont formation 48 h stages than the 24 h merozoite stage (p value <.0001, t-test) (**Figure 3C**). As was observed in the PCA to assess batch effect (**Figure 1B),** there is increased variation at the 48 h time point. Compared to mRNAs which show more upregulation when transitioning from oocyst/sporozoite to the asexual stage (1715 genes vs 1270 genes), lncRNAs have many more genes downregulated (218 vs 80) (**Figure 3D**). Comparing the asexual stage at 24 h to the sexually activated stage at 48 h, fewer mRNAs showed differential expression, with both having ~400 genes upregulated and downregulated. Very few lncRNAs were downregulated in the transition from the asexual to sexually activated stage but 85 lncRNAs were upregulated. The 85 upregulated lncRNAs did not significantly overlap with the lncRNAs that were downregulated between the extracellular to the asexual stage. Here we only used 48 h samples from batch D to analyze differential expression since this batch has clear sexual
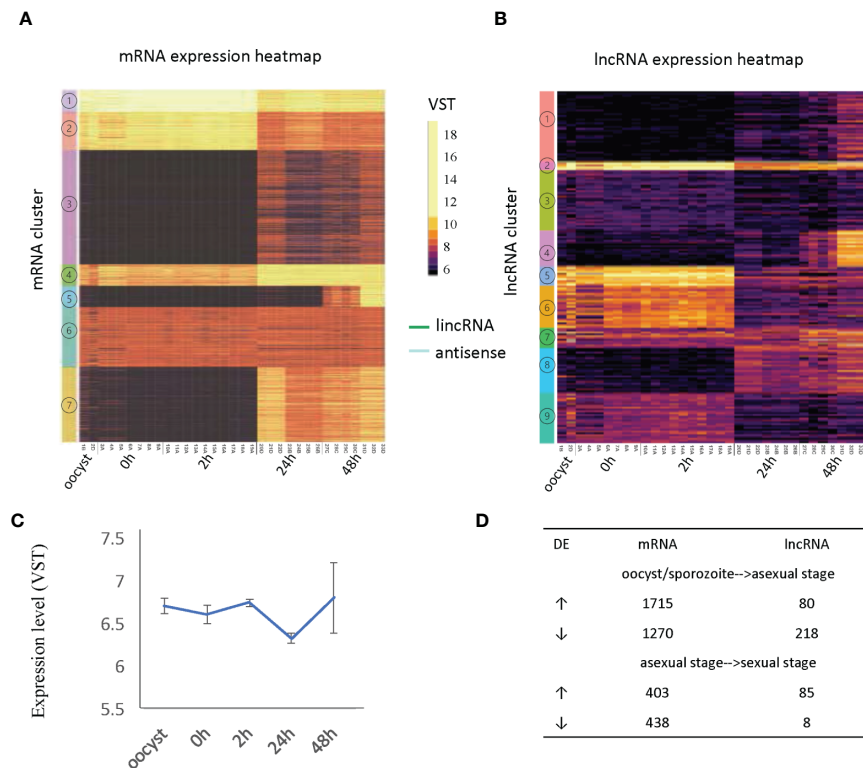
FIGURE 3 | Developmentally regulated lncRNAs. **(A)** Heatmap of mRNA expression across 33 RNA-Seq samples (**Table 1**). **(B)** Heatmap of lncRNA expression (396 transcripts) across 33 RNA-Seq samples (**Table 1**). Expression clusters generated by K-means are indicated by colored bars on the left-most edge. The 33 lincRNAs are identified by dark green lines on the righthand side of the figure. **(C)** The average expression level and standard deviation of lncRNAs at each time point. **(D)** Differentially expressed (DE) genes are compared between developmental transitions (oocyst/sporozoite stage, 24 h asexual stage and 48 h sexual stages). The arrows indicate the direction of change in gene expression. Normalized gene expression values are colored as indicated in the scale located between panels A and B with yellow indicating the highest levels.

stage marker gene expression, as discussed above. The developmentally regulated expression pattern of lncRNAs is indicative of their importance in extracellular and sexual stages. It is important to note that overall, the levels of lncRNA expression are lower than the expression levels observed for mRNAs (**Figure 3**).

Correlation of lncRNA Expression With Sense and Neighboring mRNA Expression

LncRNA mediated gene regulation can be achieved by various mechanisms (Li et al., 2020). One mechanism is transcriptional interference that usually results in repression of the target gene. LncRNAs can also activate target gene expression through epigenetic mechanisms. Additionally, translational regulation by lncRNA has also been reported. Therefore, to understand the potential roles of lncRNA transcription or transcripts, we studied the correlation between lncRNA and neighboring gene mRNA expression in *C. parvum* using the 33 RNA-Seq samples (**Supplementary Table S5**).

First, to estimate the background transcriptional correlation between genes, we calculated the expression correlation between 10,000 random gene pairs. 5748 pairs had a positive Pearson's correlation with an average r value of 0.56 while 4250 pairs had a

negative Pearson's correlation with an average r value of −0.51. Thus, we consider r > 0.56 as an indicator of positive correlation and r > 0.70 as significant correlation and r < −0.51 as an indicator of negative correlation with r < −0.7 as significant negative correlation and r values between −0.51 and 0.56 as indicative of no correlation.

Next, transcriptional correlations between the sense mRNA gene and its antisense as well as between mRNAs located upstream and downstream (either DNA strand) of the antisense transcript were analyzed (**Figure 4A**). Most sense-antisense transcript pairs do not show a strong expression correlation but those that do, tend to have a positive correlation. We note that the levels of positive correlation vary by developmental stage (**Figure 4B**). For the antisense transcripts that have a negative correlation with the sense mRNA, we note that the antisense transcripts tend to have their expression peak in extracellular stages (**Figure 4B**). To explore the possibility of a repressive role for these negatively correlated anti-sense transcripts in oocysts, we examined GO enrichment for the corresponding sense mRNAs. No significant GO functions were enriched because most sense mRNAs are uncharacterized protein-encoding genes. However, one sense-antisense pair is noteworthy. Cp_lnc_7-RA is highly expressed in oocyst/
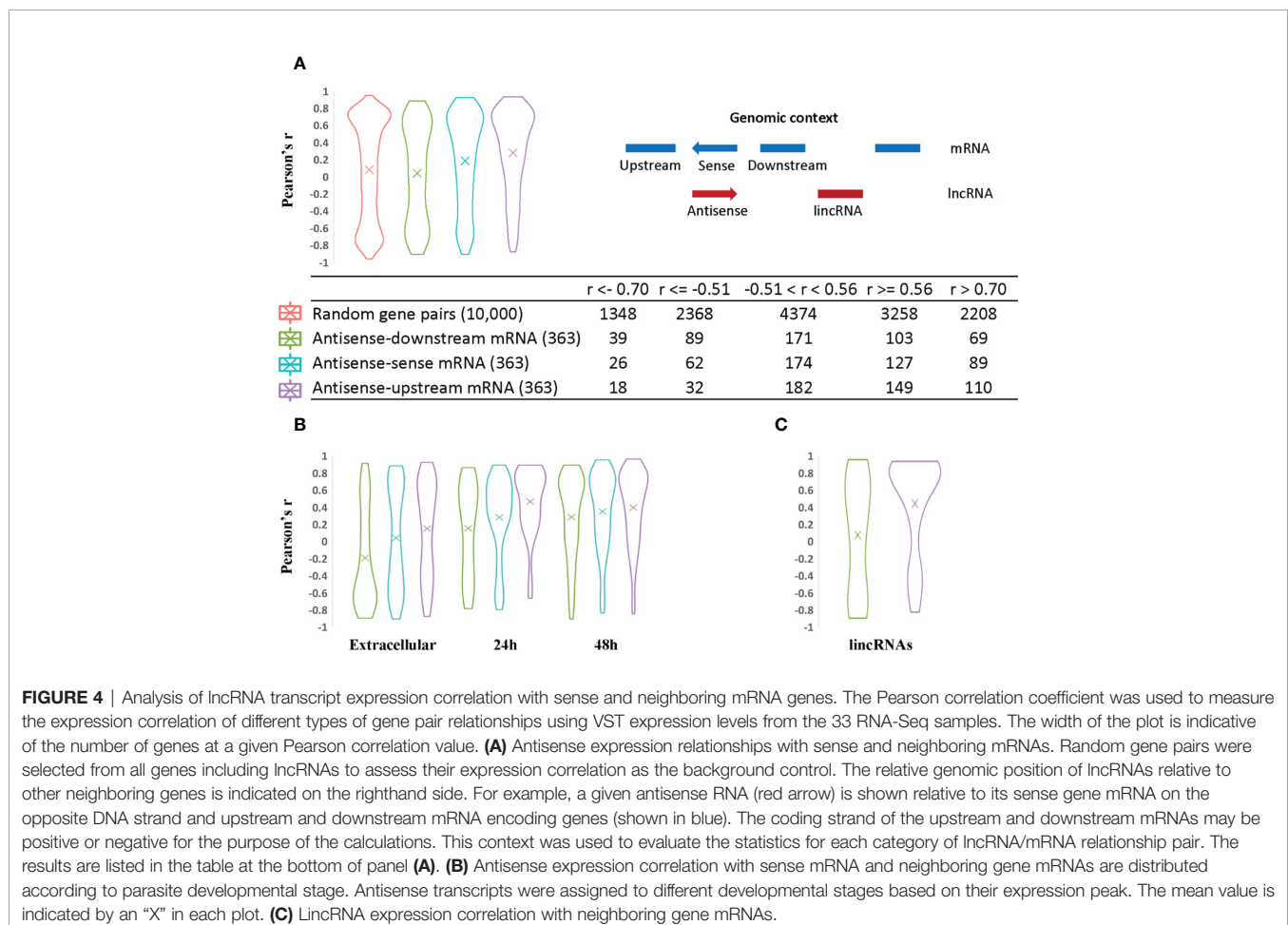
sporozoites and negatively correlated with its sense mRNA CPATCC_0034040 (Guanine nucleotide exchange factor/ Initiation factor 2 subunit family) with Pearson's r of −0.90. This protein facilitates and controls protein synthesis in eukaryotes by mediating guanine nucleotide exchange (Gordiyenko et al., 2014). We note that CPATCC_0034040 is silenced in oocyst/sporozoites and becomes active in later stages. Further experiments are needed to determine if Cp_lnc_7-RA plays any role in this regulation.

Compared to expression of random gene pairs, antisense transcripts and their upstream mRNAs have a higher positive correlation of expression level on average (**Figure 4A**), despite the fact that potential read-through transcripts have already been removed (309 out of 363 antisense genes are arranged in a head-to-head orientation with the corresponding upstream mRNA). 110 antisense genes have high positive expression correlation with their upstream mRNAs (r > 0.7) while only 69 and 89 correlated with the downstream and sense mRNAs, respectively. LincRNAs also show a higher positive expression correlation with upstream mRNAs (**Figure 4C**). One possible explanation for the higher positive transcription correlation between lncRNAs and upstream mRNAs in *C. parvum* is the potential existence of bidirectional promoters. Bidirectional promoters have been reported in many organisms especially species with compact genome sequences. In *Giardia lamblia*, bidirectional transcription is considered to be an inherent feature of promoters and contributes to an abundance of antisense transcripts (Teodorovic et al., 2007).

When antisense expression is compared to downstream mRNA expression, more negative correlations are observed relative to antisense-sense mRNA and antisense-upstream mRNA pairs. This may suggest the presence of transcriptional interference such as transcriptional machinery collision between antisense and its downstream mRNA (Hobson et al., 2012; Kindgren et al., 2018) as one possible explanation.

Finally, we note that a large proportion of antisense transcripts (102 out of 363) and their corresponding sense, upstream and downstream neighboring mRNAs do not show significant expression correlations, i.e. their r values are between −0.51 and 0.56 (**Figure 4A; Supplementary Table S7**). Seven of these uncorrelated transcripts have been reported to be delivered into the host cell nucleus (Wang et al., 2017b) (**Supplementary Table S7**). Another reason for a lack of correlation may stem from possible roles in post-transcriptional and translational processes rather than transcriptional regulation.

	r <- 0.70	r <= -0.51	-0.51 < r < 0.56	r >= 0.56	r > 0.70
Random gene pairs (10,000)	1348	2368	4374	3258	2208
Antisense-downstream mRNA (363)	39	89	171	103	69
Antisense-sense mRNA (363)	26	62	174	127	89
Antisense-upstream mRNA (363)	18	32	182	149	110

FIGURE 4 | Analysis of lncRNA transcript expression correlation with sense and neighboring mRNA genes. The Pearson correlation coefficient was used to measure the expression correlation of different types of gene pair relationships using VST expression levels from the 33 RNA-Seq samples. The width of the plot is indicative of the number of genes at a given Pearson correlation value. **(A)** Antisense expression relationships with sense and neighboring mRNAs. Random gene pairs were selected from all genes including lncRNAs to assess their expression correlation as the background control. The relative genomic position of lncRNAs relative to other neighboring genes is indicated on the righthand side. For example, a given antisense RNA (red arrow) is shown relative to its sense gene mRNA on the opposite DNA strand and upstream and downstream mRNA encoding genes (shown in blue). The coding strand of the upstream and downstream mRNAs may be positive or negative for the purpose of the calculations. This context was used to evaluate the statistics for each category of lncRNA/mRNA relationship pair. The results are listed in the table at the bottom of panel **(A)**. **(B)** Antisense expression correlation with sense mRNA and neighboring gene mRNAs are distributed according to parasite developmental stage. Antisense transcripts were assigned to different developmental stages based on their expression peak. The mean value is indicated by an "X" in each plot. **(C)** LincRNA expression correlation with neighboring gene mRNAs.

Many lncRNAs Are Conserved Between *Cryptosporidium parvum*, *Cryptosporidium hominis*, and *Cryptosporidium baileyi*

Evolutionary conservation of a lncRNA can imply functional importance. lncRNAs can be conserved in different dimensions: the sequence, structure, function, and expression from syntenic loci (Diederichs, 2014). lncRNAs are considered to be poorly conserved at the primary sequence level between genera as reported in many higher eukaryotes. Here, we looked for expression conservation of *C. parvum* lncRNA in two other *Cryptosporidium* species from syntenic loci (observed conserved antisense expression from orthologous genes) with available stranded RNA-Seq data that include *C. hominis*, a very close relative of *C. parvum* and *C. baileyi*, a distant relative that infects birds (Slapeta, 2013). First, we assembled the oocyst/sporozoite transcriptome (the only stranded samples that exist) of *C. hominis* 30976 and *C. baileyi* TAMU-09Q1 by the same methods as used previously except we did not use reference genome annotation guidance (-G). A total of 167 *C. parvum* antisense lncRNAs were detected in both *C. hominis* and *C. baileyi* (**Supplementary Table S8**). Of these, 10 are putatively conserved in *P. falciparum* (**Supplementary Table S9**) based on the presence of antisense lncRNA expression of the orthologus sense gene in *P. falciparum* (Broadbent et al., 2015). The sense mRNAs were related to functions that included a methyltransferase protein, a palmitoyltransferase and a copper transporter. No significant sequence similarities were detected among the antisense lncRNAs of these orthologs in *C. parvum* and *P. falciparum.* This is not surprising given the evolutionary distance and AT bias of *P. falciparum.*

Since the samples of *C. hominis* 30976 and *C. baileyi* TAMU-09Q1 were from oocysts/sporozoites, we focused on 48 of the 167 conserved *C. parvum* lncRNAs that showed a higher expression level in the extracellular stages (**Figure 5**). The corresponding sense mRNAs were involved in various biological processes. Translation related functions, including translational initiation (cgd7_2430, translational initiation factor eIF-5) and protein folding (cgd2_1800, DnaJ domain-containing protein), were seen in the sense mRNAs. A positive correlation of 0.78 and a negative correlation of −0.78 was calculated for cgd7_2430 sense-antisense pair and cgd2_1800 sense-antisense pair, respectively. In addition, a few mRNAs that encode putative secreted proteins (cgd5_10 and cgd4_3550) also showed a high positive correlation of expression with the corresponding antisense.

lncRNA Prediction Validation

In the RNA-Seq data, Cp_lnc_51 was expressed in oocysts/sporozoites while the associated sense mRNA cgd1_380 (Ubiquinone biosynthesis protein COQ4) was seen to be silenced the RNA-Seq data. The expression levels for the sense and antisense transcripts were validated by stranded RT-qPCR (**Figure 6A**). To validate lncRNA by strand-specific RT-PCR (Ho et al., 2010), a specific RT primer was designed for each gene to generate the appropriate stranded cDNA (**Supplementary Table S2**). Cp_lnc_51 contains an intron as does its sense mRNA. Splicing of Cp_lnc_51 was confirmed by RT-PCR and agarose gel

electrophoresis to assess the transcript size (**Figure 6B**), however, the splicing is observed in all RT-PCR products. The existence of splicing was also observed in the sense transcript (**Figure 6B**). We randomly selected five additional lncRNAs that were expressed in RNA-Seq data for oocyst/sporozoites for validation. Two out of the five, Cp_lnc_82 (**Figure 6C**) and Cp_lnc_93 (**Figure 6D**) were validated by qPCR. The relative expression of sense-antisense is consistent with the RNA-Seq data.

DISCUSSION

In this study, we utilized stranded RNA-Seq data from multiple time points during parasite development to systematically identify and characterize lncRNAs in *C. parvum*. 396 high-confidence lncRNAs were identified, 363 (91.7%) occur as antisense transcripts to mRNAs and 33 are encoded in intergenic locations. Nearly 10% of predicted mRNAs are covered by an antisense lncRNA. This level of antisense transcription suggests an important function for lncRNA in *C. parvum*. The lncRNAs were analyzed to determine expression profiles, promoter motifs for coordinately expressed transcripts, transcriptional relationships with upstream and downstream mRNAs and conservation among three *Cryptosporidium* species with stranded RNA-Seq data available.

To investigate the expression relationship of lncRNAs and their neighboring mRNA encoding genes, we calculated the expression correlation of different types of gene pairs by Pearson coefficient and noticed a higher positive correlation of expression between lncRNAs and their upstream mRNAs (primarily from the opposite strand) compared to random gene pairs. Notably, spurious correlations of gene expression can happen if the biological variation among samples is too large. Due to the challenge of *in vitro* culture for *C. parvum* and very low number of the parasite transcripts compared to their host cells, samples from the early intracellular stages are rare and usually contain very low levels of parasite transcripts. In this study, transcriptome data from early (< 24 h) intracellular timepoints are absent. We detected a bimodal shape for the distribution of expression correlation with random gene pairs showing trends at both high positive and negative values, probably due to spurious correlation. However, antisense transcripts showed much less negative correlation of expression with upstream than random gene pairs. One possible explanation for the higher positive correlation between lncRNA and their neighboring mRNAs would be the presence of bidirectional promoters in *C. parvum*. Another possibility is that lncRNAs function as positive regulators of neighboring mRNA expression. In *P. falciparum*, ncRNAs derived from GC-rich elements that are interspersed among the internal chromosomal *var* gene clusters are hypothesized to play a role in *var* gene activation while the mechanism is unclear (Guizetti et al., 2016; Barcons-Simon et al., 2020). lncRNAs have also been associated with chromatin remodeling to achieve transcriptional regulation in many studies (Li et al., 2020). One example is a lncRNA named HOTTIP which is transcribed from the 5′ tip of the HOXA locus that coordinates the activation of HOXA genes by maintaining active chromatin (Wang et al., 2011).

FIGURE 5 | Highly expressed *C. parvum* lncRNAs with conservation and expression in *C. hominis* and *C. baileyi* oocysts. The heatmap visualizes the lncRNA expression level of 48 conserved *C. parvum* genes across 33 RNA-Seq samples, grouped as extracellular (oocyst/sporozoite, 0 h and 2 h) and intracellular (24 h and 48 h) stages. The lncRNA name and the corresponding sense mRNA description are listed. An mRNA description of "NA" indicates the lncRNA is intergenic. The sense-antisense expression correlation coefficient is shown in the bracket. The color scale is shown on the left. Yellow indicates high levels of expression.

Antisense transcripts have a strong bias towards covering the 3′ end of the sense mRNA. This property has also been reported in other organisms, including the malaria pathogen *P. falciparum* (Siegel et al., 2014; Broadbent et al., 2015). As these authors suggest, one possibility is that antisense RNAs arise from promiscuous transcription initiation from nucleosome depleted regions (Siegel et al., 2014). It is also known that the 3′ UTR of mRNAs can contain elements that are important for transcript cleavage, stability, translation and mRNA localization. The 3′ UTR serves as a binding site for numerous regulatory elements including proteins and microRNAs (Jia et al., 2013; Tushev et al., 2018). In humans, the antisense *KAT5* gene has been reported to promoted the usage of distal polyA (pA) site in the sense gene *RNASEH2C*, which generated a longer 3′ untranslated region (3′ UTR) and produced less protein, accompanied by slowed cell growth (Shen et al., 2018). Whether the 3′ end bias of antisense expression related to its function and translation repressor need further investigation.

The enriched expression of lncRNAs in extracellular and late intracellular stages, the time point when the parasite expresses sexual commitment and produces gametes, suggests potential critical roles that lncRNAs may play during these life cycle stages. We specifically analyzed the transcription correlation between sense-antisense pairs in extracellular stages and late intracellular stages. Both significant positive and negative correlations between mRNA sense-antisense transcript pairs are observed in *C. parvum* which indicates that diverse RNA regulation strategies exist. Antisense transcripts/transcription as a negative regulator of sense mRNAs has been reported in many organisms including *P. falciparum* (Filarsky et al., 2018), which could also hold true in *C. parvum* for the negatively correlated sense-antisense pairs. It has been also reported that lncRNAs can regulate translation by stabilizing mRNAs, and in some cases triggering the translation process by interactions with the associated machinery (Li et al., 2020). For transcriptionally
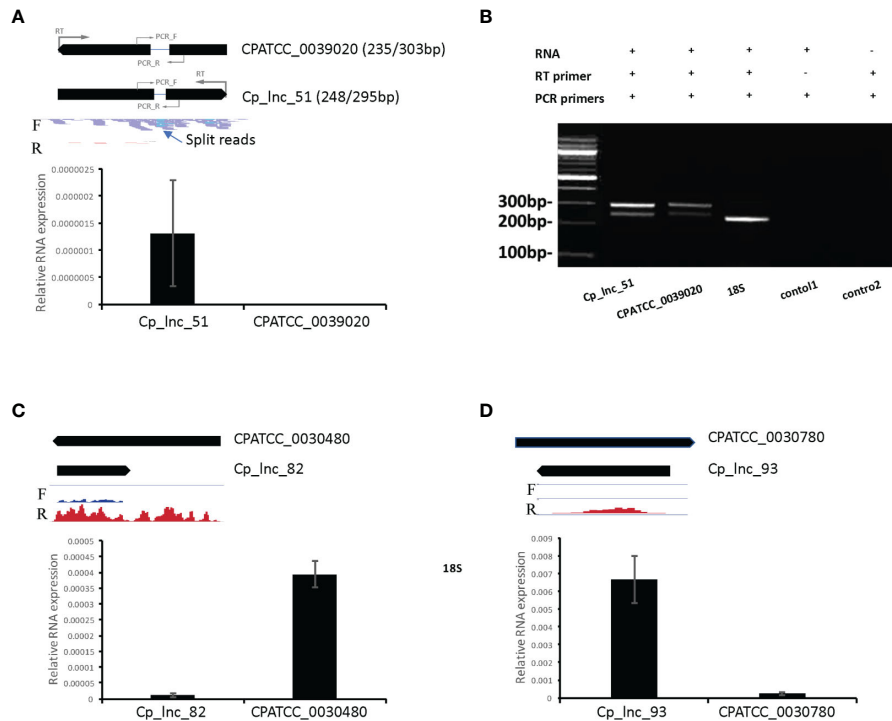
FIGURE 6 | lncRNA candidate expression and intron structure validation. **(A)** The expression levels of lncRNA Cp_lnc_51 and its corresponding sense mRNA CPATCC_0039020 were validated by RT-qPCR of RNA from oocyst/sporozoite the expression was normalized to 18S expression levels. The annotated gene model is shown on the top with RNA-Seq reads mapped to the genomic region. Location of RT primers and PCR primers for each gene are shown with the gene models. Reads are separated by the mapped strand: forward strand (F) and reverse strand (R). The intron structure of Cp_lnc_51 is indicated by the presence of split RNA-Seq reads. **(B)** Agarose gel electrophoresis of RT-PCR products for the sense and antisense transcripts. Splicing of the Cp_lnc_51 transcript and its sense mRNA CPATCC_0039020 are supported by detection of both spliced and unspliced forms for each of the transcripts of the expected sizes. The expected size of the products with/without their intron is indicated next to the gene name in panel A. 18S is used as positive control with expected size of 239bp. Control 1 is a negative control without the RT primer for CPATCC_0039020. Control 2 is a negative control for CPATCC_0039020 with both RT and PCR primers but no RNA template added. **(C)** The expression levels of lncRNA Cp_lnc_82 and its corresponding sense mRNA CPATCC_0030480 were validated by RT-qPCR. The expression was normalized to 18S expression levels. **(D)** The expression levels of lncRNA Cp_lnc_93 and its corresponding sense mRNA CPATCC_0030780 validated by RT-qPCR. The expression was normalized to 18S expression levels. The RNA-Seq coverage in C and D is with range 0–100 CPM (counts per million reads mapped).

positively correlated sense-antisense transcript pairs, we raise the testable hypothesis that lncRNAs may be involved in stabilizing transcripts that may be preloaded in macrogamonts that will eventually become oocysts. To test this hypothesis, one future direction is to take advantage of single-cell sequencing approaches and look at the transcriptomic details of male and female gametocytes. It will be interesting to see if lncRNAs are specifically expressed in male and female gametocytes and whether or not some lncRNAs are restricted specifically to these gamonts or if they are also detected elsewhere, e.g., in oocysts where they could, perhaps, have a role in transcriptional or post-transcriptional gene regulation, or mRNA stability. The amount of active transcription as opposed to RNA pre-loading in oocysts is not known.

Twenty-two *C. parvum* lncRNAs have been detected in the host cell nucleus (Wang et al., 2017a). Of these, 18 were detected in this study. Motif analysis was conducted on the exported lncRNA transcript sequences, but no significant similarity or motif was detected relative to the larger pool of lncRNA candidates identified in this study. This raises the question of what signal is responsible for lncRNA export. Further studies are needed.

A significant roadblock in lncRNA research is the determination of their function. Genetics studies are particularly tricky with antisense transcripts of the sequences overlap coding sequences closely, as they do in *C. parvum* because genetic alterations of the sequence affect the sense and antisense transcripts. lncRNAs with similar functions often lack sequence similarity (Kirk et al., 2018). Many known lncRNAs function by interacting with proteins. Proteins often bind RNA through short motifs (three to eight bp in length) (Ray et al., 2013). It was hypothesized that lncRNAs with shared functions should harbor motif composition similarities (Kirk et al., 2018). In this study, the difference in nucleotide composition between antisense RNA and lincRNA transcripts gives rise to the speculation that the machineries interacting with antisense and intergenic lncRNAs may be different.

lncRNA prediction using short reads in organisms with compact genome sequences, including *C. parvum* is limited due to the difficulty of separating independent lncRNA transcription from neighboring transcriptional read-through noise. In this study, we used a customized pipeline with strict criteria designed to minimize false positives from background noise such as transcriptional read-through. Many antisense transcriptions

overlap all or most of the sense mRNA transcript. To further improve the discovery of full-length lncRNA and any isoforms, long-read approaches such as Iso-Seq (Pacific Biosystems) and single molecule pore-sequencing approaches [Oxford Nanopore Technologies (ONT)] are needed. Although obtaining sufficient high-quality RNA from intracellular stages is still challenging, hybrid capture approaches (Gnirke et al., 2009; Amorim-Vaz et al., 2015) can be utilized to obtain *Cryptosporidium* RNA to be used for direct RNA sequencing on the ONT platform providing additional insights into the RNA biology of *Cryptosporidium*. Long-read sequencing also enables better annotation of mRNA UTR boundaries (Chappell et al., 2020) which can be used to further investigate the 3′ bias (relative to the sense transcript) of antisense transcription observed in this study.

The RNA-Seq samples from intracellular stages used in this study suffered from host contamination though we set thresholds for sample selection. For extracellular samples (sample ID 1-19 for *C. parvum*), mapping rate ranges from 76%–97% with the majority >95%. The low mapping rate dropped below 4% in intracellular. Low numbers of parasite-specific reads can lead to an underestimated number of lncRNAs, with lowly expressed lncRNAs possibly being undetectable. On the flip side, a lower false positive prediction rate is expected since the goal is to primarily characterize high-quality lncRNAs with good expression level by setting thresholds to filter out low-quality candidates.

Another limitation of this study results from the fact that the lncRNAs discussed here are exclusively generated from conventional poly(A)-enriched libraries. Thus, only polyadenylated transcripts will be identified. While many lncRNAs in the literature share characteristics of mRNAs, such as RNA polymerase II-mediated transcription, a 5′ 7-methylguanosine cap and a 3′ polyadenylated tail (Sun et al., 2018), many non-polyadenylated lncRNAs have been reported including enhancer-derived lncRNAs (Li et al., 2016) and RNA Polymerase III transcribed lncRNAs (Sun et al., 2018). Thus, determination of the full repertoire of lncRNAs requires an analysis of total RNA. The contributions of polyA+ and polyA- lncRNAs to the complete repertoire of lncRNA remains unclear including in model organism species. This is in part due to the instability of polyA- RNAs (Li et al., 2016) and the caveats of Illumina short read sequencing coupled with transcriptome assembly algorithms that predict transcripts (small RNAs excluded) which suffer from low 3′ and 5′ coverage (Uszczynska-Ratajczak et al., 2018). In *C. parvum*, the non-poly(A) selected library suffered from high levels of rRNA contamination (data not shown) since no commercial kits specific for apicomplexan or other protist parasite rRNA removal are available. In an rRNA depleted non-poly(A)+ selected total RNA library from *C. parvum* oocysts/sporozoites generated by our lab (data not shown), we were able to detect 134 (82.7%) out of the 162 antisense transcripts with peak expression in extracellular stages found in this study suggesting that we have not missed a huge population of non-poly(A) transcripts by focusing on the available poly(A)+ libraries (Li et al., in prep). Only ~20 new potential non-polyadenylated lncRNAs were detected in the rRNA depleted library. This finding suggests the possibility that lncRNAs in *C. parvum* are primarily polyadenylated.

It is important to understand how species evolve and adapt to their environment. Due to the poor conservation of lncRNA reported in higher eukaryotes (Johnsson et al., 2014) and the large phylogenetic distance among *Cryptosporidium* species (Slapeta, 2013), it is noteworthy that many lncRNAs detected in *C. parvum* were also seen expressed in *C. baileyi*. It indicates that RNA regulation could be a common and critical strategy for *Cryptosporidium* gene regulation or interactions with their hosts. The conserved antisense RNA expression between *C. parvum* and *P. falciparum* orthologs did not share detectable sequence similarity. However, the structure of lncRNAs could be under stronger selection than their sequence. Since most lncRNAs are antisense in *C. parvum*, to separate conservation of lncRNA from the conservation of mRNA sequence could provide further insights into lncRNA evolution. Selection pressures that independently act to maintain sequence and secondary structure features can lead to incongruent conservation of sequence and structure. As a consequence, it is possible that analogous base pairs no longer correspond to homologous sequence positions. Thus, possible selection pressures independently acting on sequence and structure should be taken into account (Nowick et al., 2019). Despite the increasing acknowledgment that ncRNAs are functional, tests for ncRNAs under either positive or negative selection did not exist until recently (Walter Costa et al., 2019). This type of analysis will assist in identifying candidates to prioritize for further functional lncRNA investigations.

DATA AVAILABILITY STATEMENT

The data sets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2020.608298/full#supplementary-material

REFERENCES

Abrahamsen, M. S., Templeton, T. J., Enomoto, S., Abrahante, J. E., Zhu, G., Lancto, C. A., et al. (2004). Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304 (5669), 441–445. doi: 10.1126/science.1094786

Amadi, B., Mwiya, M., Musuku, J., Watuka, A., Sianongo, S., Ayoub, A., et al. (2002). Effect of nitazoxanide on morbidity and mortality in Zambian children with cryptosporidiosis: a randomised controlled trial. *Lancet* 360 (9343), 1375–1380. doi: 10.1016/S0140-6736(02)11401-2

Amadi, B., Mwiya, M., Sianongo, S., Payne, L., Watuka, A., Katubulushi, M., et al. (2009). High dose prolonged treatment with nitazoxanide is not effective for cryptosporidiosis in HIV positive Zambian children: a randomised controlled trial. *BMC Infect. Dis.* 9, 195. doi: 10.1186/1471-2334-9-195

Amorim-Vaz, S., Tran Vdu, T., Pradervand, S., Pagni, M., Coste, A. T., and Sanglard, D. (2015). RNA Enrichment Method for Quantitative Transcriptional Analysis of Pathogens In Vivo Applied to the Fungus Candida albicans. *mBio* 6 (5), e00942–e00915. doi: 10.1128/mBio.00942-15

Anders, S., Pyl, P. T., and Huber, W. (2015). "HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31 (2), 166–169. doi: 10.1093/bioinformatics/btu638

Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.

Barcons-Simon, A., Cordon-Obras, C., Guizetti, J., Bryant, J. M., and Scherf, A. (2020). CRISPR Interference of a Clonally Variant GC-Rich Noncoding RNA Family Leads to General Repression of var Genes in *Plasmodium falciparum*. *mBio* 11 (1), e03054-19. doi: 10.1128/mBio.03054-19

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Bouzid, M., Hunter, P. R., Chalmers, R. M., and Tyler, K. M. (2013). *Cryptosporidium* pathogenicity and virulence. *Clin. Microbiol. Rev.* 26 (1), 115–134. doi: 10.1128/CMR.00076-12

Broadbent, K. M., Broadbent, J. C., Ribacke, U., Wirth, D., Rinn, J. L., and Sabeti, P. C. (2015). Strand-specific RNA sequencing in *Plasmodium falciparum* malaria identifies developmentally regulated long non-coding RNA and circular RNA. *BMC Genomics* 16, 454. doi: 10.1186/s12864-015-1603-4

Campbell, T. L., De Silva, E. K., Olszewski, K. L., Elemento, O., and Llinas, M. (2010). Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PloS Pathog.* 6 (10), e1001165. doi: 10.1371/journal.ppat.1001165

Cardenas, D., Bhalchandra, S., Lamisere, H., Chen, Y., Zeng, X. L., Ramani, S., et al. (2020). Two- and Three-Dimensional Bioengineered Human Intestinal Tissue Models for *Cryptosporidium*. *Methods Mol. Biol.* 2052, 373–402. doi: 10.1007/978-1-4939-9748-0_21

Chakravarty, D., Sboner, A., Nair, S. S., Giannopoulou, E., Li, R., Hennig, S., et al. (2014). The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat. Commun.* 5, 5383. doi: 10.1038/ncomms6383

Chappell, L., Ross, P., Orchard, L., Russell, T. J., Otto, T. D., Berriman, M., et al. (2020). Refining the transcriptome of the human malaria parasite *Plasmodium falciparum* using amplification-free RNA-seq. *BMC Genomics* 21 (1), 395. doi: 10.1186/s12864-020-06787-5

DeCicco RePass, M. A., Chen, Y., Lin, Y., Zhou, W., Kaplan, D. L., and Ward, H. D. (2017). Novel Bioengineered Three-Dimensional Human Intestinal Model for Long-Term Infection of *Cryptosporidium parvum*. *Infect. Immun.* 85 (3), e00731-16. doi: 10.1128/IAI.00731-16

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

Drummond, J. D., Boano, F., Atwill, E. R., Li, X., Harter, T., and Packman, A.II (2018). *Cryptosporidium* oocyst persistence in agricultural streams -a mobile-immobile model framework assessment. *Sci. Rep.* 8 (1), 4603. doi: 10.1038/s41598-018-22784-x

Fayer, R. (2008). *General biology. Cryptosporidium and Cryptosporidiosis. R. F. a. L. Xiao* (Boca Raton, London: CRC Press), 1–42. IWA Pub.

Filarsky, M., Fraschka, S. A., Niederwieser, I., Brancucci, N. M. B., Carrington, E., Carrio, E., et al. (2018). GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. *Science* 359 (6381), 1259–1263. doi: 10.1126/science.aan6042

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., et al. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* 27 (2), 182–189. doi: 10.1038/nbt.1523

Gong, Z., Yin, H., Ma, X., Liu, B., Han, Z., Gou, L., et al. (2017). Widespread 5-methylcytosine in the genomes of avian Coccidia and other apicomplexan parasites detected by an ELISA-based method. *Parasitol. Res.* 116 (5), 1573–1579. doi: 10.1007/s00436-017-5434-x

Gordiyenko, Y., Schmidt, C., Jennings, M. D., Matak-Vinkovic, D., Pavitt, G. D., and Robinson, C. V. (2014). eIF2B is a decameric guanine nucleotide exchange factor with a gamma2epsilon2 tetrameric core. *Nat. Commun.* 5, 3902. doi: 10.1038/ncomms4902

Guizetti, J., Barcons-Simon, A., and Scherf, A. (2016). Trans-acting GC-rich non-coding RNA at var expression site modulates gene counting in malaria parasite. *Nucleic Acids Res.* 44 (20), 9710–9718. doi: 10.1093/nar/gkw664

Heo, I., Dutta, D., Schaefer, D. A., Iakobachvili, N., Artegiani, B., Sachs, N., et al. (2018). Modelling *Cryptosporidium* infection in human small intestinal and lung organoids. *Nat. Microbiol.* 3 (7), 814–823. doi: 10.1038/s41564-018-0177-8

Ho, E. C., Donaldson, M. E., and Saville, B. J. (2010). Detection of antisense RNA transcripts by strand-specific RT-PCR. *Methods Mol. Biol.* 630, 125–138. doi: 10.1007/978-1-60761-629-0_9

Hobson, D. J., Wei, W., Steinmetz, L. M., and Svejstrup, J. Q. (2012). RNA polymerase II collision interrupts convergent transcription. *Mol. Cell* 48 (3), 365–374. doi: 10.1016/j.molcel.2012.08.027

Iyer, L. M., Anantharaman, V., Wolf, M. Y., and Aravind, L. (2008). Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int. J. Parasitol.* 38 (1), 1–31. doi: 10.1016/j.ijpara.2007.07.018

Jeninga, M. D., Quinn, J. E., and Petter, M. (2019). ApiAP2 Transcription Factors in Apicomplexan Parasites. *Pathogens* 8 (2), 47. doi: 10.3390/pathogens8020047

Jia, J., Yao, P., Arif, A., and Fox, P. L. (2013). Regulation and dysregulation of 3'UTR-mediated translational control. *Curr. Opin. Genet. Dev.* 23 (1), 29–34. doi: 10.1016/j.gde.2012.12.004

Johnsson, P., Lipovich, L., Grander, D., and Morris, K. V. (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim. Biophys. Acta* 1840 (3), 1063–1071. doi: 10.1016/j.bbagen.2013.10.035

Keeling, P. J. (2004). Reduction and compaction in the genome of the apicomplexan parasite *Cryptosporidium parvum*. *Dev. Cell* 6 (5), 614–616. doi: 10.1016/S1534-5807(04)00135-2

Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 106 (28), 11667–11672. doi: 10.1073/pnas.0904715106

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi: 10.1038/nmeth.3317

Kindgren, P., Ard, R., Ivanov, M., and Marquardt, S. (2018). Transcriptional read-through of the long non-coding RNA SVALKA governs plant cold acclimation. *Nat. Commun.* 9 (1), 4561. doi: 10.1038/s41467-018-07010-6

Kirk, J. M., Kim, S. O., Inoue, K., Smola, M. J., Lee, D. M., Schertzer, M. D., et al. (2018). Functional classification of long non-coding RNAs by k-mer content. *Nat. Genet.* 50 (10), 1474–1482. doi: 10.1038/s41588-018-0207-8

Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35 (Web Server issue), W345–W349. doi: 10.1093/nar/gkm391

Kotloff, K. L., Nataro, J. P., Blackwelder, W. C., Nasrin, D., Farag, T. H., Panchalingam, S., et al. (2013). Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* 382 (9888), 209–222. doi: 10.1016/S0140-6736(13)60844-2

Li, L., Stoeckert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13 (9), 2178–2189. doi: 10.1101/gr.1224503

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, Y., Baptista, R. P., and Kissinger, J. C. (2020). Noncoding RNAs in Apicomplexan Parasites: An Update. *Trends Parasitol.* 36 (10), 835–849. doi: 10.1016/j.pt.2020.07.006

Li, T., Mo, X., Fu, L., Xiao, B., and Guo, J. (2016). Molecular mechanisms of long noncoding RNAs on gastric cancer. *Oncotarget* 7 (8), 8601–8612. doi: 10.18632/oncotarget.6926

Li, W., Notani, D., and Rosenfeld, M. G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.* 17 (4), 207–223. doi: 10.1038/nrg.2016.4

Liao, Q., Shen, J., Liu, J., Sun, X., Zhao, G., Chang, Y., et al. (2014). Genome-wide identification and functional annotation of *Plasmodium falciparum* long noncoding RNAs from RNA-seq data. *Parasitol. Res.* 113 (4), 1269–1281. doi: 10.1007/s00436-014-3765-4

Liu, Y., Tewari, R., Ning, J., Blagborough, A. M., Garbom, S., Pei, J., et al. (2008). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in Chlamydomonas and *Plasmodium* gametes. *Genes Dev.* 22 (8), 1051–1068. doi: 10.1101/gad.1656508

Lopez-Barragan, M. J., Lemieux, J., Quinones, M., Williamson, K. C., Molina-Cruz, A., Cui, K., et al. (2011). Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. *BMC Genomics* 12, 587. doi: 10.1186/1471-2164-12-587

Love, M.II, Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi: 10.1186/s13059-014-0550-8

Marchese, F. P., Raimondi, I., and Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* 18 (1), 206. doi: 10.1186/s13059-017-1348-2

Matos, L. V. S., McEvoy, J., Tzipori, S., Bresciani, K. D. S., and Widmer, G. (2019). The transcriptome of *Cryptosporidium* oocysts and intracellular stages. *Sci. Rep.* 9 (1), 7856. doi: 10.1038/s41598-019-44289-x

Mauzy, M. J., Enomoto, S., Lancto, C. A., Abrahamsen, M. S., and Rutherford, M. S. (2012). The *Cryptosporidium parvum* transcriptome during in vitro development. *PloS One* 7 (3), e31715. doi: 10.1371/journal.pone.0031715

Miller, C. N., Josse, L., Brown, I., Blakeman, B., Povey, J., Yiangou, L., et al. (2018). A cell culture platform for *Cryptosporidium* that enables long-term cultivation and new tools for the systematic investigation of its biology. *Int. J. Parasitol.* 48 (3-4), 197–201. doi: 10.1016/j.ijpara.2017.10.001

Ming, Z., Gong, A. Y., Wang, Y., Zhang, X. T., Li, M., Mathy, N. W., et al. (2017). Involvement of *Cryptosporidium parvum* Cdg7_Flc_1000 RNA in the Attenuation of Intestinal Epithelial Cell Migration via Trans-suppression of Host Cell SMPD3 Gene. *J. Infect. Dis.* 217 (1), 122–133. doi: 10.1093/infdis/jix392

Mirhashemi, M. E., Noubary, F., Chapman-Bonofiglio, S., Tzipori, S., Huggins, G. S., and Widmer, G. (2018). Transcriptome analysis of pig intestinal cell monolayers infected with *Cryptosporidium parvum* asexual stages. *Parasit. Vectors* 11 (1), 176. doi: 10.1186/s13071-018-2754-3

Morada, M., Lee, S., Gunther-Cummins, L., Weiss, L. M., Widmer, G., Tzipori, S., et al. (2016). Continuous culture of *Cryptosporidium parvum* using hollow fiber technology. *Int. J. Parasitol.* 46 (1), 21–29. doi: 10.1016/j.ijpara.2015.07.006

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., et al. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505 (7485), 635–640. doi: 10.1038/nature12943

Niknafs, Y. S., Pandian, B., Iyer, H. K., Chinnaiyan, A. M., and Iyer, M. K. (2017). TACO produces robust multisample transcriptome assemblies from RNA-seq. *Nat. Methods* 14 (1), 68–70. doi: 10.1038/nmeth.4078

Nowick, K., Walter Costa, M. B., Honer Zu Siederdissen, C., and Stadler, P. F. (2019). Selection Pressures on RNA Sequences and Structures. *Evol. Bioinform. Online* 15, 1176934319871919. doi: 10.1177/1176934319871919

Oberstaller, J., Joseph, S. J., and Kissinger, J. C. (2013). Genome-wide upstream motif analysis of *Cryptosporidium parvum* genes clustered by expression profile. *BMC Genomics* 14, 516. doi: 10.1186/1471-2164-14-516

Oberstaller, J., Pumpalova, Y., Schieler, A., Llinas, M., and Kissinger, J. C. (2014). The *Cryptosporidium parvum* ApiAP2 gene family: insights into the evolution of apicomplexan AP2 regulatory systems. *Nucleic Acids Res.* 42 (13), 8271–8284. doi: 10.1093/nar/gku500

Painter, J. E., Hlavsa, M. C., Collier, S. A., Xiao, L., Yoder, J. S.C. Centers for Disease and Prevention (2015). Cryptosporidiosis surveillance – United State-2012. *MMWR Suppl.* 64 (3), 1–14.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33 (3), 290–295. doi: 10.1038/nbt.3122

Platts-Mills, J. A., Babji, S., Bodhidatta, L., Gratz, J., Haque, R., Havt, A., et al. (2015). Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). *Lancet Glob. Health* 3 (9), e564–e575. doi: 10.1016/S2214-109X(15)00151-5

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6), 841–842. doi: 10.1093/bioinformatics/btq033

Ramaprasad, A., Mourier, T., Naeem, R., Malas, T. B., Moussa, E., Panigrahi, A., et al. (2015). Comprehensive evaluation of *Toxoplasma gondii* VEG and Neospora caninum LIV genomes with tachyzoite stage transcriptome and proteome defines novel transcript features. *PloS One* 10 (4), e0124473. doi: 10.1371/journal.pone.0124473

Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499 (7457), 172–177. doi: 10.1038/nature12311

Ren, G. J., Fan, X. C., Liu, T. L., Wang, S. S., and Zhao, G. H. (2018). Genome-wide analysis of differentially expressed profiles of mRNAs, lncRNAs and circRNAs during *Cryptosporidium baileyi* infection. *BMC Genomics* 19 (1), 356. doi: 10.1186/s12864-018-4754-2

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616

Sateriale, A., Pawlowic, M., Vinayak, S., Brooks, C., and Striepen, B. (2020). Genetic Manipulation of *Cryptosporidium parvum* with CRISPR/Cas9. *Methods Mol. Biol.* 2052, 219–228. doi: 10.1007/978-1-4939-9748-0_13

Shen, T., Li, H., Song, Y., Yao, J., Han, M., Yu, M., et al. (2018). Antisense transcription regulates the expression of sense gene via alternative polyadenylation. *Protein Cell* 9 (6), 540–552. doi: 10.1007/s13238-017-0497-0

Siegel, T. N., Hon, C. C., Zhang, Q., Lopez-Rubio, J. J., Scheidig-Benatar, C., Martins, R. M., et al. (2014). Strand-specific RNA-Seq reveals widespread and developmentally regulated transcription of natural antisense transcripts in *Plasmodium falciparum*. *BMC Genomics* 15, 150. doi: 10.1186/1471-2164-15-150

Slapeta, J. (2013). Cryptosporidiosis and *Cryptosporidium* species in animals and humans: a thirty colour rainbow? *Int. J. Parasitol.* 43 (12-13), 957–970. doi: 10.1016/j.ijpara.2013.07.005

Sow, S. O., Muhsen, K., Nasrin, D., Blackwelder, W. C., Wu, Y., Farag, T. H., et al. (2016). The Burden of *Cryptosporidium* Diarrheal Disease among Children < 24 Months of Age in Moderate/High Mortality Regions of Sub-Saharan Africa and South Asia, Utilizing Data from the Global Enteric Multicenter Study (GEMS). *PloS Negl. Trop. Dis.* 10 (5), e0004729. doi: 10.1371/journal.pntd.0004729

Sun, Q., Hao, Q., and Prasanth, K. V. (2018). Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression. *Trends Genet.* 34 (2), 142–157. doi: 10.1016/j.tig.2017.11.005

Tandel, J., English, E. D., Sateriale, A., Gullicksrud, J. A., Beiting, D. P., Sullivan, M. C., et al. (2019). Life cycle progression and sexual development of the apicomplexan parasite *Cryptosporidium parvum*. *Nat. Microbiol.* 4 (12), 2226–2236. doi: 10.1038/s41564-019-0539-x

Tang, R., Mei, X., Wang, Y. C., Cui, X. B., Zhang, G., Li, W., et al. (2019). LncRNA GAS5 regulates vascular smooth muscle cell cycle arrest and apoptosis via p53 pathway. *Biochim. Biophys. Acta Mol. Basis Dis.* 1865 (9), 2516–2525. doi: 10.1016/j.bbadis.2019.05.022

Templeton, T. J., Iyer, L. M., Anantharaman, V., Enomoto, S., Abrahante, J. E., Subramanian, G. M., et al. (2004). Comparative analysis of apicomplexa and genomic diversity in eukaryotes. *Genome Res.* 14 (9), 1686–1695. doi: 10.1101/gr.2615304

Teodorovic, S., Walls, C. D., and Elmendorf, H. G. (2007). Bidirectional transcription is an inherent feature of Giardia lamblia promoters and contributes to an abundance of sterile antisense transcripts throughout the genome. *Nucleic Acids Res.* 35 (8), 2544–2553. doi: 10.1093/nar/gkm105

Tsoi, L. C., Iyer, M. K., Stuart, P. E., Swindell, W. R., Gudjonsson, J. E., Tejasvi, T., et al. (2015). Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin. *Genome Biol.* 16, 24. doi: 10.1186/s13059-014-0570-4

Tushev, G., Glock, C., Heumuller, M., Biever, A., Jovanovic, M., and Schuman, E. M. (2018). Alternative 3' UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments. *Neuron* 98 (3), 495–511 e496. doi: 10.1016/j.neuron.2018.03.030

Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H., and Bartel, D. P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147 (7), 1537–1550. doi: 10.1016/j.cell.2011.11.055

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigo, R., and Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.* 19 (9), 535–548. doi: 10.1038/s41576-018-0017-y

Vembar, S. S., Scherf, A., and Siegel, T. N. (2014). Noncoding RNAs as emerging regulators of *Plasmodium falciparum* virulence gene expression. *Curr. Opin. Microbiol.* 20, 153–161. doi: 10.1016/j.mib.2014.06.013

Villacorta, I., de Graaf, D., Charlier, G., and Peeters, J. E. (1996). Complete development of *Cryptosporidium parvum* in MDBK cells. *FEMS Microbiol. Lett.* 142 (1), 129–132. doi: 10.1111/j.1574-6968.1996.tb08419.x

Vinayak, S., Pawlowic, M. C., Sateriale, A., Brooks, C. F., Studstill, C. J., Bar-Peled, Y., et al. (2015). Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*. *Nature* 523 (7561), 477–480. doi: 10.1038/nature14651

Walter Costa, M. B., Honer Zu Siederdissen, C., Dunjic, M., Stadler, P. F., and Nowick, K. (2019). SSS-test: a novel test for detecting positive selection on RNA secondary structure. *BMC Bioinf.* 20 (1), 151. doi: 10.1186/s12859-019-2711-y

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472 (7341), 120–124. doi: 10.1038/nature09819

Wang, Y., Gong, A. Y., Ma, S., Chen, X., Li, Y., Su, C. J., et al. (2017a). Delivery of Parasite RNA Transcripts Into Infected Epithelial Cells During *Cryptosporidium* Infection and Its Potential Impact on Host Gene Transcription. *J. Infect. Dis.* 215 (4), 636–643. doi: 10.1093/infdis/jiw607

Wang, Y., Gong, A. Y., Ma, S., Chen, X., Strauss-Soukup, J. K., and Chen, X. M. (2017b). Delivery of parasite Cdg7_Flc_0990 RNA transcript into intestinal epithelial cells during *Cryptosporidium parvum* infection suppresses host cell

gene transcription through epigenetic mechanisms. *Cell Microbiol.* 19 (11), 1–21. doi: 10.1111/cmi.12760

Wilke, G., Funkhouser-Jones, L. J., Wang, Y., Ravindran, S., Wang, Q., Beatty, W. L., et al. (2019). A Stem-Cell-Derived Platform Enables Complete *Cryptosporidium* Development In Vitro and Genetic Tractability. *Cell Host Microbe* 26 (1), 123–134.e128. doi: 10.1016/j.chom.2019.05.007

Wilke, G., Wang, Y., Ravindran, S., Stappenbeck, T., Witola, W. H., and Sibley, L. D. (2020). In Vitro Culture of *Cryptosporidium parvum* Using Stem Cell-Derived Intestinal Epithelial Monolayers. *Methods Mol. Biol.* 2052, 351–372. doi: 10.1007/978-1-4939-9748-0_20

Yarlett, N., Morada, M., Gobin, M., Van Voorhis, W., and Arnold, S. (2020). In Vitro Culture of *Cryptosporidium parvum* Using Hollow Fiber Bioreactor: Applications for Simultaneous Pharmacokinetic and Pharmacodynamic Evaluation of Test Compounds. *Methods Mol. Biol.* 2052, 335–350. doi: 10.1007/978-1-4939-9748-0_19

Zhang, H., Guo, F., Zhou, H., and Zhu, G. (2012). Transcriptome analysis reveals unique metabolic features in the *Cryptosporidium parvum* Oocysts associated with environmental survival and stresses. *BMC Genomics* 13, 647. doi: 10.1186/1471-2164-13-647

Zhang, L., Luo, X., Chen, F., Yuan, W., Xiao, X., Zhang, X., et al. (2018). LncRNA SNHG1 regulates cerebrovascular pathologies as a competing endogenous RNA through HIF-1alpha/VEGF signaling in ischemic stroke. *J. Cell Biochem.* 119 (7), 5460–5472. doi: 10.1002/jcb.26705