

1 **Stage-Specific Long Non-coding RNAs in *Cryptosporidium parvum* as**
2 **Revealed by Stranded RNA-Seq**

3 **Yiran Li¹, Rodrigo P. Baptista^{1,2}, Adam Sateriale^{3#}, Boris Striepen³ and Jessica C.**
4 **Kissinger^{1,2,4,*}**

5 ¹ Institute of Bioinformatics, University of Georgia, Athens, GA, USA

6 ² Center for Tropical and Emerging Global Diseases, University of Georgia, Athens, GA, USA

7 ³ Department of Pathobiology, School of Veterinary Medicine, University of Pennsylvania,
8 Philadelphia, PA, USA

9 ⁴ Department of Genetics, University of Georgia, Athens, GA, USA

10 ***Correspondence:**

11 Jessica C. Kissinger: jkissing@uga.edu

12 [#]Current address: Host-Pathogen Interactions in Cryptosporidiosis Laboratory, The Francis Crick
13 Institute, London, UK

14 **Keywords:** lncRNA₁, RNA regulation₂, Apicomplexa₃, parasite development₄, stranded RNA-
15 Seq₅, transcriptome₆

16 **Abstract**

17 *Cryptosporidium* is a protist parasite that has been identified as the second leading cause of
18 moderate to severe diarrhea in children younger than two and a significant cause of mortality
19 worldwide. *Cryptosporidium* has a complex, obligate, intracellular but extra cytoplasmic lifecycle
20 in a single host. How genes are regulated in this parasite remains largely unknown. Long non-
21 coding RNAs (lncRNAs) play critical regulatory roles, including gene expression across a broad
22 range of organisms. *Cryptosporidium* lncRNAs have been reported to enter the host cell nucleus
23 and affect the host response. However, no systematic study of lncRNAs in *Cryptosporidium* has
24 been conducted to identify additional lncRNAs. In this study, we analyzed a *C. parvum* *in vitro*
25 strand-specific RNA-seq developmental time series covering both asexual and sexual stages to
26 identify lncRNAs associated with parasite development. In total, we identified 396 novel lncRNAs
27 86% of which are differentially expressed. Nearly 10% of annotated mRNAs have an antisense
28 lncRNA. lncRNAs also appear to occur most often at the 3' end of their corresponding sense
29 mRNA. Putative lncRNA regulatory regions were identified and many appear to encode
30 bidirectional promoters. A positive correlation trend between lncRNA and the upstream mRNA
31 expression was observed. Evolutionary conservation and expression of lncRNA candidates was
32 observed between *C. parvum*, *C. hominis* and *C. baileyi*. Ten *C. parvum* protein-encoding genes
33 with antisense transcripts have *P. falciparum* orthologs that also have antisense transcripts.
34 Three *C. parvum* lncRNAs with exceptional properties (e.g., intron splicing) were experimentally
35 validated using RT-PCR and RT-qPCR. We provide an initial characterization of the *C. parvum*
36 37 38

39 non-coding transcriptome to facilitate further investigations into the roles of lncRNAs in parasite
40 development and host-pathogen interactions.
41

42 **Introduction**

43 *Cryptosporidium* is an obligate protist parasite that causes a diarrheal disease called
44 cryptosporidiosis which spreads via an oral-fecal route. Human cryptosporidiosis, mainly caused
45 by *Cryptosporidium parvum* and *Cryptosporidium hominis*, is typically self-limited and causes
46 1~2 weeks of intense watery diarrhea in people with healthy immune systems. However, the illness
47 may be lethal among the immunocompromised including individuals with AIDS, cancer, and those
48 receiving transplant anti-rejection medications. In recent years, several *Cryptosporidium* species,
49 predominantly *C. hominis*, have been identified as the second most prevalent diarrheal pathogen
50 of infants globally after rotavirus (Bouzid, Hunter et al. 2013, Kotloff, Nataro et al. 2013, Painter,
51 Hlavsa et al. 2015, Platts-Mills, Babji et al. 2015, Sow, Muhsen et al. 2016) and a leading cause
52 of waterborne disease among humans in the United States (Prevention). Thus far, Nitazoxanide,
53 the only FDA-approved drug is not effective for use in infants or those with HIV-related
54 immunosuppression (Amadi, Mwiya et al. 2009) i.e. the most susceptible populations, and no
55 vaccine is available (Amadi, Mwiya et al. 2002).

56 *Cryptosporidium* has a complex lifecycle in a single host. The *Cryptosporidium* oocyst
57 which is shed in feces is a major extracellular lifecycle stage. It can stay dormant and survive in
58 the environment for months (Drummond, Boano et al. 2018). After ingestion of oocysts through
59 contaminated water or food, sporozoites are released which are capable of invading intestinal
60 epithelial cells where both asexual and sexual replication occur. Following invasion, sporozoites
61 develop into trophozoites and undergo asexual replication to generate type I meronts and type II
62 meronts. Type I meronts are thought to be capable of reinvading adjacent cells generating an
63 asexual cycle (Fayer 2008), while Type II meronts contribute to the formation of microgamonts
64 (male form) or macrogamonts (female form) to complete the sexual stages (Bouzid, Hunter et al.
65 2013). Conventional monolayer cell culture does not permit completion of the life cycle much
66 beyond 48 hours post-infection (hpi) (**Figure 1A**), for as of yet poorly understood reasons but
67 gametogenesis does occur (Tandel, English et al. 2019). The lack of *in vitro* culture has historically
68 impeded the development of new drugs and vaccines for this medically important parasite.
69 Recently, there have been several breakthroughs including genetic manipulation (Vinayak,
70 Pawlowic et al. 2015, Sateriale, Pawlowic et al. 2020) and lifecycle completion. Several promising
71 approaches have been developed including using a cancer cell line as host (Miller, Josse et al.
72 2018), biphasic and three-dimensional (organoid) culture systems, (Morada, Lee et al. 2016,
73 DeCicco RePass, Chen et al. 2017, Heo, Dutta et al. 2018, Cardenas, Bhalchandra et al. 2020),
74 hollow fiber technology (Yarlett, Morada et al. 2020), and air-liquid interface (ALI) cultivation
75 system (Wilke, Funkhouser-Jones et al. 2019, Wilke, Wang et al. 2020). These breakthroughs are
76 enabling better, much needed, studies of the parasite's full life cycle. A better understanding the
77 conditions and regulatory processes necessary for *Cryptosporidium* development are essential and
78 will prove beneficial for the identification of drug and vaccine targets.

79 The first genome sequence of *C. parvum* was published in 2004 with a genome size of ~9
80 Mb and ~3800 protein-encoding genes annotated (Abrahamsen, Templeton et al. 2004). Since this
81 milestone, our understanding of the parasite and its biology have progressed remarkably. Early *in*

82 *vitro* transcriptome analyses using semi-quantitative RT-PCR over a 72 h post-infection (pi) time
83 course during *in vitro* development revealed complex and dynamic gene expression profiles.
84 Adjacent genes are not generally co-regulated, despite the highly compact genome (Mauzy,
85 Enomoto et al. 2012). Under UV irradiation, *C. parvum* oocysts have shown a vital stress-induced
86 gene expression response according to microarray data (Zhang, Guo et al. 2012). mRNA
87 expression related to gametocyte and oocyst formation were studied using RNA sequencing of
88 sorted cells (Tandel, English et al. 2019). Yet, little is known about the regulation of key
89 developmental transitions. How the parasite regulates gene expression in order to invade hosts,
90 amplify, evade the immune system and interact with their host awaits further discovery.

91 Most canonical eukaryotic enhancer proteins are not detected in Apicomplexa, the phylum
92 that *Cryptosporidium* belongs to (Iyer, Anantharaman et al. 2008), except for the transcriptional
93 activators Myb and zinc finger proteins C2H2 (Cys2His2) and two additional transcription factor
94 families. Instead, an expanded family of apatela-related transcription factors, the AP2 family of
95 proteins (ApiAP2), appear to be the predominant transcription factors in this phylum, including
96 *Cryptosporidium* (Oberstaller, Pumpalova et al. 2014, Jeninga, Quinn et al. 2019). AP2 domains
97 in *C. parvum* are reported to have reduced binding diversity relative to the malaria parasite
98 *Plasmodium falciparum* and proposed to possess less dominancy in transcriptional regulation in
99 *C. parvum* (Campbell, De Silva et al. 2010, Yarlett, Morada et al. 2020). It has been proposed that
100 *C. parvum* is less reliant on ApiAP2 regulators in part because it utilizes E2F/DP1 transcription
101 factors, which are present in *Cryptosporidium* while absent in other studied apicomplexans
102 (Templeton, Iyer et al. 2004, Yarlett, Morada et al. 2020). Based on the similarity of gene
103 expression profiles, it has been suggested that the number of co-expressed gene clusters in *C.*
104 *parvum* is somewhere between 150 and 200, and putative ApiAP2 and E2F/DP1_{cis}-regulatory
105 elements were successfully detected in the upstream region of many co-expressed gene clusters
106 (Oberstaller, Joseph et al. 2013). Additionally, low levels of DNA methylation in oocysts has been
107 reported in several *Cryptosporidium spp.* (Gong, Yin et al. 2017), suggesting the requirement of
108 additional regulatory mechanisms. At the level of post-transcriptional regulation, the RNA
109 interference (RNAi) pathway, which plays a crucial role in gene silencing in most eukaryotes, is
110 considered to be missing in *Cryptosporidium* due to the lack of identifiable genes encoding the
111 microRNA processing machinery or RNA-induced silencing complex (RISC) components
112 (Keeling 2004). There much that remains to be discovered with respect to regulation of gene
113 expression in *Cryptosporidium*.

114 Long non-coding RNAs (lncRNAs) are transcripts without significant protein-encoding
115 capacity that are longer than 200 nt. In eukaryotes, lncRNAs play critical regulatory roles in gene
116 regulation at multiple levels, including transcriptional, post-transcriptional, chromatin
117 modification and nuclear architecture alterations (Marchese, Raimondi et al. 2017). In humans,
118 3,300 long intergenic ncRNAs (lincRNAs) were analyzed using chromatin state maps, and ~20%
119 of these RNAs are bound to the polycomb repressive complex (PCR2, a complex with histone
120 methyltransferase activity) (Khalil, Guttman et al. 2009). Most lncRNAs share many
121 characteristics of mRNAs, such as RNA polymerase II-mediated transcription, a 5' 7-

122 methylguanosine cap and a 3' poly(A) tail [6]. The expression of lncRNAs is usually more tissue-
123 or time-specific than mRNA expression (Necsulea, Soumillon et al. 2014, Tsoi, Iyer et al. 2015).
124 lncRNA sequences are not well conserved across species, but their structure could be conserved
125 due to functional constraints (Ulitsky, Shkumatava et al. 2011, Diederichs 2014). By forming
126 hybrid structural complexes such as RNA-DNA hybrid duplexes or RNA-DNA triplexes,
127 lncRNAs can recruit or scaffold protein complexes to facilitate localization of protein machinery
128 to specific genome target sites (Li, Mo et al. 2016). lncRNAs play important roles in regulating
129 occurrence and progression of many diseases. After infected by *C. baileyi*, significant expression
130 changes have been observed in the host (Ren, Fan et al. 2018). The mis-regulation of lncRNAs in
131 multi-cellular eukaryotes has been shown to cause tumorigenesis (Chakravarty, Sboner et al. 2014),
132 cardiovascular diseases (Tang, Mei et al. 2019), and neurodegenerative dysfunction (Zhang, Luo
133 et al. 2018) and thus can be used as diagnostic biomarkers.

134 Taking advantage of sequencing technologies, numerous lncRNA candidates have been
135 detected in apicomplexans, some with proven regulatory potential during parasite invasion and
136 proliferation processes. These discoveries have ushered in a new era in parasite transcriptomics
137 research (Liao, Shen et al. 2014, Siegel, Hon et al. 2014, Broadbent, Broadbent et al. 2015,
138 Ramaprasad, Mourier et al. 2015, Filarsky, Fraschka et al. 2018). In *P. falciparum*, lncRNAs are
139 critical regulators of virulence gene expression and associated with chromatin modifications
140 (Vembar, Scherf et al. 2014). Likewise, an antisense lncRNA of the gene *gdv1* was shown to be
141 involved in regulating sexual conversion in *P. falciparum* (Filarsky, Fraschka et al. 2018). In *C. parvum*,
142 putative parasite lncRNAs were found to be delivered into the host nucleus, some of
143 which were experimentally proven to regulate host genes by hijacking the host histone
144 modification system (Ming, Gong et al. 2017, Wang, Gong et al. 2017, Wang, Gong et al. 2017).
145 The importance of lncRNA in *C. parvum* has been demonstrated, but no systematic annotation of
146 lncRNA has been conducted. The systematic identification of lncRNAs will increase the pool of
147 candidate regulatory molecules thus ultimately leading to increased knowledge of the
148 developmental gene regulation in *C. parvum* and control of parasite interactions with its hosts.

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

155

156 **Methods and Materials**

157 **RNA-Seq data pre-processing/cleaning**

158 RNA-Seq datasets were downloaded from the NCBI Sequence Read Archive (SRA) and
159 European Nucleotide Archive database (ENA). Detailed information on SRA accession numbers
160 and Bioprojects are listed in **Table 1** and **Supplementary Table S1**. FastQC-v0.11.8 was used to
161 perform quality control of the RNA-Seq reads. Remaining adapters and low-quality bases were

162 trimmed by Trimmomatic-v-0.36 (Bolger, Lohse et al. 2014) with parameters: Adapters:2:30:10
163 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:25 MINLEN:25. All reads were scanned
164 with a four-base sliding window and cut when the average Phred quality dropped below 25. Bases
165 from the start and end were cut off when the quality score was below 20. The minimum read length
166 was set at 25 bases. The processed reads are referred to as cleaned reads in this work.
167

168 **Read mapping and transcript assembly**

169 Cleaned reads from each sample were mapped to the reference genome sequence for
170 *Cryptosporidium parvum* IOWA-ATCC (Baptista et al. in prep) downloaded from CryptoDB v46
171 (<https://cryptodb.org/cryptodb/>) using the mapping tool HISAT2-v2.1.0 (Kim, Langmead et al.
172 2015) with maximum intron length (--max-intronlen) set at 3000, and remaining parameters as
173 default. Uniquely mapped reads were selected for further study using SAMtools-v1.10 (view -q
174 10) (Li, Handsaker et al. 2009). StringTie-v2.0.6 (Pertea, Pertea et al. 2015) was used to reconstruct
175 transcripts using the reference annotation guided method (--rf -j 5 -c 10 -g 1). At least five reads
176 were required to define a splice junction. A minimum read coverage of 10 was used for transcript
177 prediction. Only overlapping transcript clusters were merged together. The stranded library types
178 were all “fr-firststrand”. Transcripts with FPKM lower than three were removed. The
179 transcriptome assemblies from multiple samples were merged into one master transcript file using
180 TACO-v0.7.3 with default settings (Niknafs, Pandian et al. 2017).
181

182 **lncRNA prediction**

183 Transcripts that overlapped with currently annotated mRNAs in the *C. parvum* IOWA-
184 ATCC annotation in CryptoDB v46 with coverage >70% on the same strand were removed using
185 BEDTools-v2.29.2 (Quinlan and Hall 2010). The remaining transcripts were examined for coding
186 potential using the online tool Coding Potential Calculator (CPC) v0.9 (Kong, Zhang et al. 2007).
187 Transcripts considered as “coding” by CPC were removed. Potential read-through transcripts
188 resulting from transcription of neighboring mRNAs were removed using two criteria: 1) The
189 transcript was <50 bp from the upstream coding region of another gene on the same strand. 2) The
190 transcript was always transcribed together with the upstream mRNA on the same strand. Finally,
191 the remaining transcripts which occurred in >2 RNA-Seq samples were kept as putative lncRNA
192 candidates for further studies.
193

194 **Transcriptome data normalization and identification of differentially expressed genes**

195 The raw read counts for both mRNA genes and predicted lncRNAs were calculated using
196 HTSeq-v0.12.4 (Anders, Pyl et al. 2015). All genes were filtered to require > 50 reads in at least
197 three samples. Variance stabilizing transformation (VST) from DESeq2-v1.28.1 (Love, Huber et
198 al. 2014) was used to normalize the expression between samples. Principal components analysis
199 (PCA) of the RNA-Seq samples was performed using the resulting VST values. VST values for
200 each of the mRNA and lncRNA candidates were visualized using the R package pheatmap-v1.0.12

201 (<https://github.com/raivokolde/pheatmap>). The K-mean approach in pheatmap was applied to
202 cluster genes based on the expression data.

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

213

214 **Expression correlation**

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

218

219 **Upstream motif analysis**

220 MEME v5.0.0 (Bailey and Elkan 1994) was used to discover motifs that may be present
221 upstream of putative lncRNAs. For the promoter motif search, we extracted the 100 bp of (+)
222 strand sequence upstream of the predicted lncRNAs and searched both strands using MEME for
223 motifs with length six to 50 bp. The parameters were: -dna -mod anr -nmotifs 6 -minw 6 -maxw
224 50 -objfun classic -markov_order 0.

225

226 **Evolutionary conservation**

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

238

239 **RT-qPCR validation**

240 We designed PCR primers using the PrimerQuest Tool from IDT
241 (<https://www.idtdna.com/pages/tools/primerquest>) (**Supplementary Table S2**) to use for
242 expression validation and exon structure confirmation of select lncRNA candidates using RT-PCR
243 and qPCR. RNA was extracted from oocysts and provided by Boris Striepen. The cDNA for each
244 sample was reverse-transcribed using the iScript™ cDNA Synthesis Kit (Bio-Rad, Hercules, CA)
245 from 1 µg of input RNA. The resulting cDNAs were used as templates for PCR amplification and
246 qPCR detection. Strand-specific primers were designed to amplify antisense RNAs. The 18S
247 rRNA gene of *C. parvum* was used as positive control and samples without RNA or primer were
248 used as a negative control. Each RT-PCR reaction contained 1 µl cDNA, 2 µl primer mix (10 µM),
249 2 µl water and 5 µl MyTaq™ HS Mix (Bioline). RT-PCR was performed in the following
250 conditions: 35 cycles of 15 seconds at 95°C, 30 seconds at 64°C. Then the RT-PCR products were
251 run on a 2% agarose gel. The cDNA was also subjected to qPCR with All-in-One qPCR Mix
252 (QP001; GeneCopoeia, Rockville, MD, USA) using the Mx3005P qPCR system (Agilent
253 Technologies, Santa Clara, CA, USA). All reactions, including no-template controls, were run in
254 triplicate. Following amplification, the CT values were determined using fixed threshold settings.
255 lncRNA expression was normalized to 18S rRNA expression.

256

257 **Data availability**

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

261

262 **Results**

263 **Mapping statistics of stranded RNA-Seq data**

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

278

279 **Identification and characteristics of lncRNAs**

280 We began assembly of the *C. parvum* transcriptome using mapped RNA-Seq reads of
281 samples in NCBI Bioproject PRJNA530692 with a high sample quality. However, *C. parvum* has
282 an extremely compact genome sequence. As calculated from the *C. parvum* IOWA-ATCC

283 annotation, the average intergenic distance between the stop and start codon boundaries of
284 neighboring genes is only 504 bp and this distance must also include promoter and UTR regions.
285 The average length of an annotated *C. parvum* ATCC mRNA coding sequence, CDS, is 1802bp.
286 The high gene density and the short distance between genes make it difficult to set UTR boundaries
287 using short-read sequencing data as transcripts overlap and become merged. Transcriptome
288 assembly using RNA-Seq without genome and reference annotation guidance would lead to a high
289 chimerism rate. Thus, we used reference annotation to guide the assembly process and set
290 parameters to minimize the number of artificially fused transcripts. We then used the program
291 TACO v0.7.3 (Multi-sample transcriptome assembly), which employs change point detection to
292 break apart complex loci to lower the number of fused transcripts to obtain a non-redundant master
293 transcriptome from all samples, resulting in 5818 transcripts in total. Of these, 4846 transcripts
294 overlapped with an mRNA on the same strand and thus, were removed. Transcripts which were
295 <200 bp or only detected in a single sample were removed to improve the lncRNA prediction
296 quality. Transcripts that were considered as “coding” by the Coding potential analysis tool CPC
297 were filtered out. To identify and remove potential read-through transcripts, predicted transcripts
298 that were located closer than 50 bp from the coding region of the upstream gene on the same strand
299 and always transcribed together with the upstream mRNA were removed (**Figure 2A**).

300 In total, 396 transcripts, primarily located antisense to an mRNA (**Figure 2B**), were
301 selected as lncRNA candidates for further analysis. Most of the lncRNAs we detected consist of a
302 single exon however five lncRNAs contain introns. This is consistent with the low-intron rate in
303 *Cryptosporidium*. Additional introns are expected to emerge with deeper RNA-Seq data since
304 many lncRNAs have low expression levels. The average length of the lncRNAs and mRNAs is
305 transcripts 1267 bp and 1866 bp (including UTRs), respectively (**Figure 2C**). When compared to
306 mRNAs, one of the most distinguishing features of lncRNAs is the low Open Reading Frame (ORF)
307 coding potential relative to the transcript length. To understand their biogenesis, we searched for
308 potential promoter motifs within 100 bp of the (+) strand upstream from all 396 lncRNAs. This
309 analysis returned five significant motifs (E-value<0.001), the top two, which were also the most
310 dominant motifs, are related to known *Cryptosporidium* transcriptional factor binding motifs for
311 mRNA genes. It included the E2F/DP1 (5'-[C/G]GCGC[G/C]-3') and ApiAP2_1 (5'-
312 BGCATGCAH-3') motifs (**Figure 2D**). This suggests that lncRNA transcriptions have the
313 potential for being regulated independently during parasite development.

314 We further investigated the relative location of the antisense transcripts relative to the
315 mRNA gene body, and we found that many of lncRNAs’ initiation and termination sites are located
316 close to the gene body boundaries, especially the start site of the antisense lncRNA transcript
317 (assembled transcript including untranslated regions, UTRs) (**Figure 2E**). This trend is most
318 apparent with the lncRNA initiation site. When looking at the coverage of the antisense transcript
319 on mRNA, the *C. parvum* lncRNA antisense expression has a bias towards the 3' end of the mRNA
320 transcript (**Figure 2F**). This property has also been reported in other organisms, including the
321 malaria pathogen *Plasmodium falciparum* (Lopez-Barragan, Lemieux et al. 2011, Siegel, Hon et
322 al. 2014, Broadbent, Broadbent et al. 2015).

323
324 **The *Cryptosporidium* transcriptome varies developmentally and by host**
325 Before profiling the expression of lncRNA candidates, we first compared the
326 transcriptomes of the 33 *C. parvum* RNA-Seq samples by principal component analysis (PCA)
327 based on the normalized mRNA and lncRNA gene expression level of each sample (**Figure 1B**).
328 Extracellular stages, including oocyst and sporozoites from 0 h and 2 h, are differentiated from

329 intracellular stages, including 24 h and 48 h. The transcriptomes of intracellular stages were
330 demonstrated to be more heterogeneous, while extracellular samples formed a relatively compact
331 cluster. This observation was consistent with a previous transcriptome study of *C. parvum* oocysts
332 and intracellular stages (Matos, McEvoy et al. 2019). At time points 24 h and 48 h, different host
333 cells and laboratory procedures were used, which could contribute to the distance observed
334 between samples from the same time point.

335 To further explore whether sexual commitment was initiated in all 48 h samples, we
336 profiled the transcriptome of marker genes cgd6_2090 and cgd8_2220 (**Supplementary Figure**
337 **1**). cgd6_2090 encodes the *Cryptosporidium* oocyst wall protein-1 (COWP1) which is produced
338 in female gametes (Tandel, English et al. 2019); cgd8_2220 encodes the homolog of hapless2
339 (HAP2) which is a class of membrane fusion protein required for gamete fusion in a range of
340 organisms including *Plasmodium falciparum* (Liu, Tewari et al. 2008). HAP2 labeled protein was
341 exclusively found in male gamonts in *C. parvum* (Tandel, English et al. 2019). In another study,
342 the *C. parvum* transcriptome was elucidated over a 72 h *in vitro* time-course infection with HCT8
343 cells using semi-quantitative RT-PCR (Mauzy, Enomoto et al. 2012). The 48 h-specific genes from
344 that study were also examined in the 33 RNA-Seq data sets analyzed here (**Supplementary Figure**
345 **2**). Both the sex marker genes and 48 h-specific genes show an expression peak at 48 hr. At 48 h,
346 expression levels from batch D samples are much higher than batch C samples. cgd8_2220 is very
347 low/not expressed at 48 h in batch C samples. These results indicate that both batch C and D
348 samples showed the commitment of sexual development, but commitment was more pronounced
349 in batch D. It is possible that sequencing depth could be influencing this difference as the
350 normalization process (VST) between samples tends to reduce the variation of genes with low read
351 support. Interestingly, batch C and D samples used different host cell types (**Table 1**). Batch D *C.*
352 *parvum* parasites were cultured in HCT-8 (Human intestine cells) while batch C parasites were
353 cultured in MDBK (Bovine kidney cells). Although sexual stages have been observed in MDBK
354 cells (Villacorta, de Graaf et al. 1996), it is possible that the adaptation of *C. parvum* to MDBK
355 cells is lower than HCT-8 cells as hosts. Thus, a slower or lower conversion rate was observed.
356

357 **lncRNAs are developmentally regulated**

358 We visualized gene expression profiles across the 33 RNA-Seq samples used in this study
359 (**Table 1**) for both mRNA (**Figure 3A**) and lncRNAs (**Figure 3B**). To identify genes with a similar
360 expression profile, we used the k-mean algorithm to cluster mRNAs and lncRNAs separately. The
361 k value was selected as the smallest value that allowed the separation of genes from different time
362 points while keeping genes from different samples of the same timepoint together despite the batch
363 effects present within each time point. As a result, mRNAs and lncRNAs were clustered into seven
364 and nine broad co-expression groups, respectively.

365 The expression of mRNAs in the extracellular stages (oocysts and sporozoites from 0h and
366 2h) showed a similar expression trend that is quite distinct from the latter two stages. For mRNAs,
367 genes from cluster 1 and cluster 2 were more highly expressed in the extracellular stages but still
368 show expression in the intracellular stages when the vast majority of mRNAs are active. This result
369 is consistent with another transcriptome study of *C. parvum*, which used semi-quantitative RT-
370 PCR over a 72 h time course during *in vitro* development (Mauzy, Enomoto et al. 2012). On the
371 contrary, many lncRNAs showed enriched expression in extracellular stages (oocyst, 0 h, 2 h) and
372 some had expression later at the intracellular sexual development stage (48 h) while the asexual
373 stage (24 h) showed the least lncRNA expression. It is noteworthy that both mRNA and lncRNA
374 have gene sets that are specifically turned on at 48 h (mRNA cluster 5 and lncRNA cluster 5). The

375 average expression level of lncRNAs suggests they were more abundant or were actively expressed
376 at the oocyst, 0 h, 2 h and 48 h stages (**Figure 3C**) As was observed in the PCA to assess batch
377 effect (**Figure 1B**) there is increased variation at the 48hr time point. Compared to mRNAs which
378 show more upregulation when transitioning from oocyst/sporozoite to the asexual stage (1715
379 genes vs 1270 genes), lncRNAs have many more genes downregulated (218 vs 80) (**Figure 3D**).
380 Comparing the asexual stage at 24 h to the sexually activated stage at 48 h, fewer mRNAs showed
381 differential expression, with both having ~400 genes upregulated and downregulated. Very few
382 lncRNAs were downregulated in the transition from the asexual to sexually activated stage but 85
383 lncRNAs were upregulated. The 85 upregulated lncRNAs did not significantly overlap with the
384 lncRNAs that were downregulated between the extracellular to the asexual stage. Here we only
385 used 48 h samples from batch D to analyze differential expression since this batch has clear sexual
386 stage marker gene expression, as discussed above. The developmentally regulated expression
387 pattern of lncRNAs is indicative of their importance in extracellular and sexual stages. It is
388 important to note that overall, the levels of lncRNA expression are lower than the expression levels
389 observed for mRNAs (Figure 3).
390

391 **Correlation of LncRNA expression with neighboring mRNA expression**

392 LncRNA mediated gene regulation can be achieved by various mechanisms (Li, Baptista
393 et al. 2020). One mechanism is transcriptional interference that usually results in repression of the
394 target gene. LncRNAs can also regulate target gene expression through epigenetic mechanisms.
395 Additionally, translational regulation by lncRNA has also been reported. Therefore, to understand
396 the potential roles of lncRNA transcription or transcripts, we studied the correlation between
397 lncRNA and neighboring gene mRNA expression in *C. parvum* (**Supplementary Table 3**). We
398 found that compared to random gene pairs, lncRNAs and their upstream mRNAs have a higher
399 positive correlation of expression level (**Figure 4**) despite the fact that potential read-through
400 transcripts have already been removed. Bidirectional promoters have been reported in many
401 organisms especially species with compact genome sequences. In *Giardia lamblia*, bidirectional
402 transcription is considered to be an inherent feature of promoters and contributes to an abundance
403 of antisense transcripts throughout the genome (Teodorovic, Walls et al. 2007). Thus, some
404 transcriptionally positive correlated lncRNA and upstream mRNAs pairs in *C. parvum* are
405 expected to share bidirectional promoters. However, a large proportion of lncRNA and neighbor
406 mRNAs do not show an apparent expression correlation.
407

408 **Many lncRNAs are conserved between *C. parvum*, *C. hominis* and *C. baileyi***

409 Evolutionary conservation of a lncRNA can imply functional importance. lncRNAs can be
410 conserved in different dimensions: the sequence, structure, function, and expression from syntenic
411 loci (Diederichs 2014). lncRNAs are considered to be poorly conserved at the primary sequence
412 level between genera as reported in many higher eukaryotes. Here, we looked for expression
413 conservation of *C. parvum* lncRNA in two other *Cryptosporidium* species from syntenic loci with
414 available stranded RNA-Seq data that include *C. hominis*, a very close relative of *C. parvum* and
415 *C. baileyi*, a distant relative that infects birds (Slapeta 2013). First, we assembled the
416 oocyst/sporozoite transcriptome (the only stranded samples that exist) of *C. hominis* 30976 and *C.*
417 *baileyi* TAMU-09Q1 by the same methods as used previously except we did not use reference
418 genome annotation guidance (-G). A total of 167 *C. parvum* antisense lncRNAs were detected in
419 both *C. hominis* and *C. baileyi* (**Supplementary Table 4**). Of these, 10 are putatively conserved
420 in *P. falciparum* (**Supplementary Table 5**) based on the presence of antisense lncRNA expression

421 of the orthologous sense gene in *P. falciparum* (Broadbent, Broadbent et al. 2015). No significant
422 sequence similarities were detected among the antisense lncRNAs of these orthologs in *C. parvum*
423 and *P. falciparum*, this is not surprising as little similarity would be found at the level of the sense
424 mRNA's either given the evolutionary distance and AT bias of *P. falciparum*.
425
426

427 Since the samples of *C. hominis* 30976 and *C. baileyi* TAMU-09Q1 were from
428 oocysts/sporozoites, we focused on 48 of the 167 conserved *C. parvum* lncRNAs that showed a
429 higher expression level in the extracellular stages (**Figure 5**). The corresponding sense mRNAs
430 were involved in various biological processes. Translation related functions, including
431 translational initiation (cgd7_2430, translational initiation factor eIF-5) and protein folding
432 (cgd2_1800, DnaJ domain-containing protein), were seen in the sense mRNAs. A positive
433 correlation of 0.78 and a negative correlation of -0.78 was calculated for cgd7_2430 sense-
434 antisense pair and cgd2_1800 sense-antisense pair, respectively. In addition, a few mRNAs that
435 encode putative secreted proteins (cgd5_10 and cgd4_3550) also showed a high positive
436 correlation of expression with the corresponding antisense.
437

438 **lncRNA prediction validation**

439 In the RNA-Seq data, Cp_lnc_51 was expressed in oocyst/sporozoites while the associated
440 sense mRNA cgd1_380 (Ubiquinone biosynthesis protein COQ4) was seen to be silenced. The
441 expression levels for each were validated by stranded RT-qPCR (**Figure 6A**). To validate lncRNA
442 by strand-specific RT-PCR (Ho, Donaldson et al. 2010), a specific RT primer was designed for
443 each gene to generate the cDNA with strand information retained (**Supplementary Table 2**).
444 Cp_lnc_51 contains an intron. The splicing of Cp_lnc_51 was confirmed by RT-PCR and agarose
445 gel electrophoresis to assess the transcript size (**Figure 6B**). We randomly selected additional five
446 lncRNAs for validation. Two out of five, Cp_lnc_82 (**Figure 6C**) and Cp_lnc_93 (**Figure 6D**)
447 were validated by qPCR. The relative expression of sense-antisense is consistent with the RNA-
448 Seq data.
449
450

451 **Discussion**

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

461 To investigate the expression relationship of lncRNAs and their neighboring mRNA
462 encoding genes, we calculated the expression correlation of different type of gene pairs by Pearson
463 coefficient and noticed a higher positive correlation of expression between lncRNA and its
464 upstream mRNAs compared to random gene pairs. Many sense and antisense pairs also showed a
465 positive correlation of expression. Notably, spurious correlations of gene expression can happen
466 if the biological variation among samples is too large. Due to the challenge of *in vitro* culture for

467 *C. parvum* and very low volume, hence number of the parasite transcripts compared to their host
468 cells, samples from the early intracellular stages are rare and usually contain very low levels of
469 parasite transcripts. In this study, the transcriptome data from early intracellular stages was absent.
470 We detected a bimodal shape for the distribution of expression correlation with random gene pairs
471 showing trends at both high positive and negative values, probably due to spurious correlation.
472 However, lncRNAs showed much less negative correlation of expression with both upstream and
473 corresponding sense mRNA than random gene pairs. Thus, the higher positive correlation between
474 lncRNA and the neighboring mRNAs may suggest pervasive bidirectional promoters in *C. parvum*.
475 Another possibility is that lncRNAs function as positive regulators of the neighboring mRNA
476 expression. In *P. falciparum*, ncRNAs derived from GC-rich elements that are interspersed among
477 the internal chromosomal *var* gene clusters are hypothesized to play a role in *var* gene activation
478 while the mechanism is unclear (Guizetti, Barcons-Simon et al. 2016, Barcons-Simon, Cordon-
479 Obras et al. 2020). lncRNAs have been associated with chromatin remodeling to achieve
480 transcriptional regulation in many studies (Li, Baptista et al. 2020). One example is that a lncRNA
481 HOTTIP transcribed from the 5' tip of the HOXA locus that coordinates the activation of HOXA
482 genes by maintaining active chromatin (Wang, Yang et al. 2011).

483 Functional enrichment analysis is challenging for this parasite due to the large number of
484 uncharacterized proteins and incomplete pathways (Rider and Zhu 2010). Functional analysis for
485 antisense associated sense mRNA by GO enrichment didn't significantly define key biological
486 processes. However, the enriched expression of lncRNA in extracellular stages and late
487 intracellular stages, the time point when the parasite starts to have sexual commitment and produce
488 gametes, suggests potential critical roles that lncRNAs may play during these life cycle stages.
489 One possibility is that lncRNAs are involved in transcriptome pre-loading in macrogamont that
490 will eventually become an oocyst. It is also possible that lncRNA play a role in transcriptional
491 regulation or that antisense lncRNAs may play roles in the post-transcriptional process. It has been
492 reported that lncRNAs can regulate translation by stabilizing mRNAs, triggering mRNA
493 degradation or triggering translation process by interactions with associated machineries (Li,
494 Baptista et al. 2020). As shown in this study, antisense transcripts have a strong bias towards
495 covering the 3' end of the sense mRNA. This property has also been reported in other organisms,
496 including the malaria pathogen *P. falciparum* (Siegel, Hon et al. 2014, Broadbent, Broadbent et al.
497 2015). As these authors suggest, one possibility is that antisense RNAs arise from promiscuous
498 transcription initiation from nucleosome depleted regions (Siegel, Hon et al. 2014). It is also
499 known that the 3' UTR of mRNAs can contain elements that are important for transcript cleavage,
500 stability, translation and mRNA localization. The 3' UTR serves as a binding site for numerous
501 regulatory elements including proteins and microRNAs (Jia, Yao et al. 2013, Tushev, Glock et al.
502 2018). In humans, the antisense *KAT5* gene has been reported to promoted the usage of distal
503 polyA (pA) site in the sense gene *RNASEH2C*, which generated a longer 3' untranslated region (3'
504 UTR) and produced less protein, accompanied by slowed cell growth (Shen, Li et al. 2018).
505 Whether the 3' end bias of antisense expression related to its function and translation repressor
506 need further investigation. One future direction is to take advantage of single-cell sequencing
507 approaches and look at the transcriptomic details of male and female gametocytes. It will be
508 interesting to see if lncRNAs are specifically expressed in male and female gametocytes and
509 whether or not some lncRNAs are restricted specifically to these gamonts or if they are also
510 detected elsewhere, e.g., in oocysts where they could, perhaps, have a role in transcriptional or
511 post-transcriptional gene regulation, or mRNA stability. The amount of active transcription as
512 opposed to RNA pre-loading in the female gamont (future oocyst) is not known.

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

518 A significant roadblock in lncRNA research is the determination of their function. Genetics
519 studies are particularly tricky with anti-sense transcripts of the sequences overlap coding
520 sequences closely, as they do in *C. parvum* because genetic alterations of the sequence affect the
521 sense and anti-sense transcripts. lncRNAs with similar functions often lack sequence similarity
522 (Kirk, Kim et al. 2018). Many known lncRNAs function by interacting with proteins. Proteins
523 often bind RNA through short motifs (three to eight bp in length) (Ray, Kazan et al. 2013). It was
524 hypothesized that lncRNAs with shared functions should harbor motif composition similarities
525 (Kirk, Kim et al. 2018). In this study, nucleotide composition of lncRNAs varies between those
526 that are antisense and those that are encoded in intergenic regions. We see many lncRNAs have
527 higher CT-rich content compared to mRNAs (**Supplementary Figure 3**). Interestingly, lncRNAs
528 can be grouped into CT-rich and AT-rich, with most of the intergenic lncRNAs belonging to the
529 AT-rich group. The difference of nucleotide composition gives rise to the speculation that the
530 machineries interacting with antisense and intergenic lncRNAs may be different in *C. parvum*.

531 lncRNA prediction using short reads in organisms with compact genome sequences,
532 including *C. parvum* is limited due to the difficulty of separating independent lncRNA
533 transcription from neighboring transcriptional read-through noise. In this study, we used a
534 customized pipeline with strict criteria designed to minimize false positives from background noise
535 such as transcriptional read-through. Many antisense transcriptions cover all or most of the sense
536 mRNA transcript. To further improve the discovery of full-length lncRNA and any isoforms, long-
537 read approaches such as Iso-Seq (Pacific Biosystems) and single molecule pore-sequencing
538 approaches [Oxford Nanopore Technologies (ONT)] are needed. Although obtaining sufficient
539 high-quality RNA from intracellular stages is still challenging, hybrid capture approaches (Gnirke,
540 Melnikov et al. 2009, Amorim-Vaz, Tran Vdu et al. 2015) can be utilized to obtain *Cryptosporidium*
541 RNA to be used for, direct RNA sequencing on the ONT platform providing additional insights
542 into the RNA biology of *Cryptosporidium*. Besides, long-read sequencing would also enable
543 better annotation of mRNA UTR boundaries (Chappell, Ross et al. 2020), which can be used to
544 further investigate the 3' UTR bias of antisense transcription as observed in this study.

545 It is important to understand how species evolve and adapt to their environment. Due to
546 the poor conservation of lncRNA reported in higher eukaryotes (Johnsson, Lipovich et al. 2014)
547 and the large phylogenetic distance among *Cryptosporidium* species (Slapeta 2013), it is
548 noteworthy that many lncRNAs detected in *C. parvum* were also seen expressed in *C. baileyi*. It
549 indicates that RNA regulation could be a common and critical strategy for *Cryptosporidium* gene
550 regulation or interactions with their hosts. The discovery of conserved antisense lncRNA
551 expression between *C. parvum* and *P. falciparum* orthologs revealed that many important mRNAs
552 have antisense expression including a methyltransferase protein, a palmitoyltransferase and a
553 copper transporter. lncRNAs function by interaction with DNA, RNA or proteins. Thus, the
554 structure of lncRNAs could be under stronger selection than their sequence. Since most lncRNAs
555 are antisense in *C. parvum*, to separate conservation of lncRNA from the conservation of mRNA
556 sequence could provide further insights into lncRNA evolution. Selection pressures that
557 independently act to maintain sequence and secondary structure features can lead to incongruent
558 conservation of sequence and structure. As a consequence, it is possible that analogous base pairs

559 no longer correspond to homologous sequence positions. Thus, possible selection pressures
560 independently acting on sequence and structure should be taken into account (Nowick, Walter
561 Costa et al. 2019). Despite the increasing acknowledgment that ncRNAs are functional, tests for
562 ncRNAs under either positive or negative selective did not exist until recently (Walter Costa,
563 Honer Zu Siederdissen et al. 2019). This type of analysis will assist in identifying candidates to
564 prioritize for further functional lncRNA investigations.
565

566

Table 1. Mapping statistics of RNA-Seq datasets

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

567

568

569 *Batch with various host cell types and parasites from different projects are indicated with the sample IDs. IDs
570 designated with A, C and E: (Matos, McEvoy et al. 2019); B: (Mirhashemi, Noubary et al. 2018); D: (Tandel,
571 English et al. 2019); 36F: SRR1183950; 37F: SRR1183934; 38G: SRR1183952.

572 Cell line synonyms/origin: HCT-8: Human intestine; IPEC: Intestinal Porcine Epithelial Cell line; MDBK: Bovine
573 kidney

574 **Figure legends**

575
576 **Figure 1. The 33 RNA-Seq datasets used for expression analysis.** A) The time points indicate
577 when RNA-Seq samples were collected and the associated *C. parvum* life cycle stage. The
578 schematic model of the *C. parvum* life cycle is reproduced from (Tandel, English et al. 2019). B)
579 Principal component analysis of 33 *C. parvum* transcriptomes. The analysis is based on the
580 normalized expression level (VST) of the *C. parvum* mRNA and predicted lncRNA genes.
581 Samples collected from different time points are indicated by colors. Various projects/batches
582 are represented by shapes. Batch A includes sample IDs 3-9 without host cells, Batch B includes
583 sample IDs 1, 23-26 with host cell type of IPEC, Batch C includes sample IDs 27-30 26 with
584 host cell type of MDBK, Batch D contains sample IDs 2, 20-22, 31-33 with host cell type of
585 HCT-8 (Table 1).

586
587 **Figure 2. Prediction and characterization of lncRNAs in *C. parvum*.** A) Pipeline of lncRNA
588 prediction. B) Genomic location of predicted lncRNAs. C) The distribution of transcript length
589 of mRNA genes and lncRNA candidates. D) Enriched upstream motifs within 100 bp, the same
590 strand (+) of lncRNA candidates E) Antisense transcription initiation and termination position
591 relative to the sense gene body (normalized to 0-100). F) Abundance and position of sense gene
592 body (normalized to 0-100) covered by antisense transcription

593
594 **Figure 3. Developmentally regulated lncRNAs.** A) Heatmap of mRNA expression across 33
595 RNA-Seq samples (Table 1). B) Heatmap of lncRNA expression across 33 RNA-Seq samples
596 (Table 1). Expression clusters generated by K-means are indicated by colored bars on the left-
597 most edge. C) The average expression level and standard deviation of lncRNAs at each time
598 point. D) Differentially expressed (DE) genes are compared between developmental transitions
599 (oocyst/sporozoite stage, 24 h asexual stage and and 48 h sexual stages. The arrows indicate the
600 direction of change in gene expression. Normalized gene expression values are colored as
601 indicated in the scale located between panels A and B with yellow indicating the highest levels.

602
603 **Figure 4. lncRNA expression relative to neighboring mRNAs.** The Pearson correlation
604 coefficient was used to measure the expression correlation of different types of gene pair
605 relationships using VST expression levels from the 33 RNA-Seq samples. Random genes pairs
606 were genes that were randomly selected from any lncRNA candidates or mRNA genes. The
607 median value is indicated by a vertical line in each box plot. A graphical representation of the
608 relative position of the mRNA being evaluated to the lncRNA is indicated on the right side. The
609 antisense lncRNA (red arrow) is shown relative to the upstream and downstream mRNAs on the
610 same strand as the sense mRNA.

611
612 **Figure 5. Highly expressed *C. parvum* lncRNAs with conservation and expression in *C.*
613 *hominis* and *C. baileyi* oocysts.** The heatmap visualizes the lncRNA expression level of 48
614 conserved *C. parvum* genes across 33 RNA-Seq samples, grouped as extracellular
615 (oocyst/sporozoite, 0 h and 2 h) and intracellular (24 h and 48 h) stages. The lncRNA name and
616 the corresponding sense mRNA description are listed. An mRNA description of “NA” indicates
617 the lncRNA is intergenic. The sense-antisense expression correlation coefficient is shown in the
618 bracket . The color scale is shown on the left. Yellow indicates high levels of expression.

619

620 **Figure 6. lncRNA candidates expression and intron structure validation.** A) The expression
621 level of lncRNA Cp_lnc_51 and the corresponding sense mRNA CPATCC_0039020 validated
622 by RT-qPCR, the expression was normalized to 18S. The annotated genome model is shown on
623 the top with RNA-Seq reads mapped to the genomic region. Location of RT primers and PCR
624 primers for each gene are shown with the gene models. Reads are separated by the mapped
625 strand: forward strand (F) and reverse strand (R). The intron structure of Cp_lnc_51 is indicated
626 by the split reads. B) Splicing of the Cp_lnc_51 transcript is supported by the various length of
627 transcripts from intron splicing, shown on agarose gel with the expected size. The expected size
628 of the products with/without intron is indicated next to the gene name in panel A. 18S is used as
629 positive control with expected size of 239bp. Control 1 is negative control without RT primer but
630 only PCR primers of CPATCC_0039020 added. Control 2 is also negative control with both RT
631 and PCR primers of CPATCC_0039020 but no RNA template added. C) The expression level of
632 lncRNA Cp_lnc_82 and the corresponding sense mRNA CPATCC_0030480 validated by RT-
633 qPCR, the expression was normalized to 18S. D) The expression level of lncRNA Cp_lnc_93
634 and the corresponding sense mRNA CPATCC_0030780 validated by RT-qPCR, the expression
635 was normalized to 18S. The RNA-Seq coverage in C and D is with range 0-100 CPM (counts per
636 million reads mapped).

637

638

639

640

641

642

643 **References**

644

645 Abrahamsen, M. S., T. J. Templeton, S. Enomoto, J. E. Abrahante, G. Zhu, C. A. Lancto, M.

646 Deng, C. Liu, G. Widmer, S. Tzipori, G. A. Buck, P. Xu, A. T. Bankier, P. H. Dear, B. A.

647 Konfortov, H. F. Spriggs, L. Iyer, V. Anantharaman, L. Aravind and V. Kapur (2004).

648 "Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*." *Science*

649 **304**(5669): 441-445.

650 Amadi, B., M. Mwiya, J. Musuku, A. Watuka, S. Sianongo, A. Ayoub and P. Kelly (2002).

651 "Effect of nitazoxanide on morbidity and mortality in Zambian children with

652 cryptosporidiosis: a randomised controlled trial." *Lancet* **360**(9343): 1375-1380.

653 Amadi, B., M. Mwiya, S. Sianongo, L. Payne, A. Watuka, M. Katubulushi and P. Kelly (2009).

654 "High dose prolonged treatment with nitazoxanide is not effective for cryptosporidiosis in

655 HIV positive Zambian children: a randomised controlled trial." *BMC Infect Dis* **9**: 195.

656 Amorim-Vaz, S., T. Tran Vdu, S. Pradervand, M. Pagni, A. T. Coste and D. Sanglard (2015).

657 "RNA Enrichment Method for Quantitative Transcriptional Analysis of Pathogens *In*

658 *Vivo* Applied to the *Fungus Candida albicans*." *mBio* **6**(5): e00942-00915.

659 Anders, S., P. T. Pyl and W. Huber (2015). "HTSeq--a Python framework to work with high-

660 throughput sequencing data." *Bioinformatics* **31**(2): 166-169.

661 Bailey, T. L. and C. Elkan (1994). "Fitting a mixture model by expectation maximization to

662 discover motifs in biopolymers." *Proc Int Conf Intell Syst Mol Biol* **2**: 28-36.

663 Barcons-Simon, A., C. Cordon-Obras, J. Guizetti, J. M. Bryant and A. Scherf (2020). "CRISPR

664 Interference of a Clonally Variant GC-Rich Noncoding RNA Family Leads to General

665 Repression of var Genes in *Plasmodium falciparum*." *mBio* **11**(1).

666 Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina

667 sequence data." *Bioinformatics* **30**(15): 2114-2120.

668 Bouzid, M., P. R. Hunter, R. M. Chalmers and K. M. Tyler (2013). "*Cryptosporidium*

669 pathogenicity and virulence." *Clin Microbiol Rev* **26**(1): 115-134.

670 Broadbent, K. M., J. C. Broadbent, U. Ribacke, D. Wirth, J. L. Rinn and P. C. Sabeti (2015).

671 "Strand-specific RNA sequencing in *Plasmodium falciparum* malaria identifies

672 developmentally regulated long non-coding RNA and circular RNA." *BMC Genomics*

673 **16**: 454.

674 Campbell, T. L., E. K. De Silva, K. L. Olszewski, O. Elemento and M. Llinas (2010).

675 "Identification and genome-wide prediction of DNA binding specificities for the ApiAP2

676 family of regulators from the malaria parasite." *PLoS Pathog* **6**(10): e1001165.

677 Cardenas, D., S. Bhalchandra, H. Lamisere, Y. Chen, X. L. Zeng, S. Ramani, U. C. Karandikar,

678 D. L. Kaplan, M. K. Estes and H. D. Ward (2020). "Two- and Three-Dimensional

679 Bioengineered Human Intestinal Tissue Models for *Cryptosporidium*." *Methods Mol Biol*

680 **2052**: 373-402.

681 Chakravarty, D., A. Sboner, S. S. Nair, E. Giannopoulou, R. Li, S. Hennig, J. M. Mosquera, J.

682 Pauwels, K. Park, M. Kossai, T. Y. MacDonald, J. Fontugne, N. Erho, I. A. Vergara, M.

683 Ghadessi, E. Davicioni, R. B. Jenkins, N. Palanisamy, Z. Chen, S. Nakagawa, T. Hirose,

684 N. H. Bander, H. Beltran, A. H. Fox, O. Elemento and M. A. Rubin (2014). "The

685 oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate

686 cancer." *Nat Commun* **5**: 5383.

687 Chappell, L., P. Ross, L. Orchard, T. J. Russell, T. D. Otto, M. Berriman, J. C. Rayner and M.
688 Llinas (2020). "Refining the transcriptome of the human malaria parasite *Plasmodium*
689 *falciparum* using amplification-free RNA-seq." *BMC Genomics* **21**(1): 395.

690 DeCicco RePass, M. A., Y. Chen, Y. Lin, W. Zhou, D. L. Kaplan and H. D. Ward (2017).
691 "Novel Bioengineered Three-Dimensional Human Intestinal Model for Long-Term
692 Infection of *Cryptosporidium parvum*." *Infect Immun* **85**(3).

693 Diederichs, S. (2014). "The four dimensions of noncoding RNA conservation." *Trends Genet*
694 **30**(4): 121-123.

695 Drummond, J. D., F. Boano, E. R. Atwill, X. Li, T. Harter and A. I. Packman (2018).
696 "Cryptosporidium oocyst persistence in agricultural streams -a mobile-immobile model
697 framework assessment." *Sci Rep* **8**(1): 4603.

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

700 Filarsky, M., S. A. Fraschka, I. Niederwieser, N. M. B. Brancucci, E. Carrington, E. Carrio, S.
701 Moes, P. Jenoe, R. Bartfai and T. S. Voss (2018). "GDV1 induces sexual commitment of
702 malaria parasites by antagonizing HP1-dependent gene silencing." *Science* **359**(6381):
703 1259-1263.

704 Gnrke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, G.
705 Giannoukos, S. Fisher, C. Russ, S. Gabriel, D. B. Jaffe, E. S. Lander and C. Nusbaum
706 (2009). "Solution hybrid selection with ultra-long oligonucleotides for massively parallel
707 targeted sequencing." *Nat Biotechnol* **27**(2): 182-189.

708 Gong, Z., H. Yin, X. Ma, B. Liu, Z. Han, L. Gou and J. Cai (2017). "Widespread 5-
709 methylcytosine in the genomes of avian Coccidia and other apicomplexan parasites
710 detected by an ELISA-based method." *Parasitol Res* **116**(5): 1573-1579.

711 Guizetti, J., A. Barcons-Simon and A. Scherf (2016). "Trans-acting GC-rich non-coding RNA at
712 var expression site modulates gene counting in malaria parasite." *Nucleic Acids Res*
713 **44**(20): 9710-9718.

714 Heo, I., D. Dutta, D. A. Schaefer, N. Iakobachvili, B. Artegiani, N. Sachs, K. E. Boonekamp, G.
715 Bowden, A. P. A. Hendrickx, R. J. L. Willems, P. J. Peters, M. W. Riggs, R. O'Connor
716 and H. Clevers (2018). "Modelling *Cryptosporidium* infection in human small intestinal
717 and lung organoids." *Nat Microbiol* **3**(7): 814-823.

718 Ho, E. C., M. E. Donaldson and B. J. Saville (2010). "Detection of antisense RNA transcripts by
719 strand-specific RT-PCR." *Methods Mol Biol* **630**: 125-138.

720 Iyer, L. M., V. Anantharaman, M. Y. Wolf and L. Aravind (2008). "Comparative genomics of
721 transcription factors and chromatin proteins in parasitic protists and other eukaryotes." *Int*
722 *J Parasitol* **38**(1): 1-31.

723 Jeninga, M. D., J. E. Quinn and M. Petter (2019). "ApiAP2 Transcription Factors in
724 Apicomplexan Parasites." *Pathogens* **8**(2).

725 Jia, J., P. Yao, A. Arif and P. L. Fox (2013). "Regulation and dysregulation of 3'UTR-mediated
726 translational control." *Curr Opin Genet Dev* **23**(1): 29-34.

727 Johnsson, P., L. Lipovich, D. Grander and K. V. Morris (2014). "Evolutionary conservation of
728 long non-coding RNAs; sequence, structure, function." *Biochim Biophys Acta* **1840**(3):
729 1063-1071.

730 Keeling, P. J. (2004). "Reduction and compaction in the genome of the apicomplexan parasite
731 *Cryptosporidium parvum*." *Dev Cell* **6**(5): 614-616.

732 Khalil, A. M., M. Guttman, M. Huarte, M. Garber, A. Raj, D. Rivea Morales, K. Thomas, A.
733 Presser, B. E. Bernstein, A. van Oudenaarden, A. Regev, E. S. Lander and J. L. Rinn
734 (2009). "Many human large intergenic noncoding RNAs associate with chromatin-
735 modifying complexes and affect gene expression." *Proc Natl Acad Sci U S A* **106**(28):
736 11667-11672.

737 Kim, D., B. Langmead and S. L. Salzberg (2015). "HISAT: a fast spliced aligner with low
738 memory requirements." *Nat Methods* **12**(4): 357-360.

739 Kirk, J. M., S. O. Kim, K. Inoue, M. J. Smola, D. M. Lee, M. D. Schertzer, J. S. Wooten, A. R.
740 Baker, D. Sprague, D. W. Collins, C. R. Horning, S. Wang, Q. Chen, K. M. Weeks, P. J.
741 Mucha and J. M. Calabrese (2018). "Functional classification of long non-coding RNAs
742 by k-mer content." *Nat Genet* **50**(10): 1474-1482.

743 Kong, L., Y. Zhang, Z. Q. Ye, X. Q. Liu, S. Q. Zhao, L. Wei and G. Gao (2007). "CPC: assess
744 the protein-coding potential of transcripts using sequence features and support vector
745 machine." *Nucleic Acids Res* **35**(Web Server issue): W345-349.

746 Kotloff, K. L., J. P. Nataro, W. C. Blackwelder, D. Nasrin, T. H. Farag, S. Panchalingam, Y. Wu,
747 S. O. Sow, D. Sur, R. F. Breiman, A. S. Faruque, A. K. Zaidi, D. Saha, P. L. Alonso, B.
748 Tamboura, D. Sanogo, U. Onwuchekwa, B. Manna, T. Ramamurthy, S. Kanungo, J. B.
749 Ochieng, R. Omore, J. O. Oundo, A. Hossain, S. K. Das, S. Ahmed, S. Qureshi, F.
750 Quadri, R. A. Adegbola, M. Antonio, M. J. Hossain, A. Akinsola, I. Mandomando, T.
751 Nhampossa, S. Acacio, K. Biswas, C. E. O'Reilly, E. D. Mintz, L. Y. Berkeley, K.
752 Muhsen, H. Sommerfelt, R. M. Robins-Browne and M. M. Levine (2013). "Burden and
753 aetiology of diarrhoeal disease in infants and young children in developing countries (the
754 Global Enteric Multicenter Study, GEMS): a prospective, case-control study." *Lancet*
755 **382**(9888): 209-222.

756 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R.
757 Durbin and S. Genome Project Data Processing (2009). "The Sequence Alignment/Map
758 format and SAMtools." *Bioinformatics* **25**(16): 2078-2079.

759 Li, L., C. J. Stoeckert, Jr. and D. S. Roos (2003). "OrthoMCL: identification of ortholog groups
760 for eukaryotic genomes." *Genome Res* **13**(9): 2178-2189.

761 Li, T., X. Mo, L. Fu, B. Xiao and J. Guo (2016). "Molecular mechanisms of long noncoding
762 RNAs on gastric cancer." *Oncotarget* **7**(8): 8601-8612.

763 Li, Y., R. P. Baptista and J. C. Kissinger (2020). "Noncoding RNAs in Apicomplexan Parasites:
764 An Update." *Trends Parasitol*.

765 Liao, Q., J. Shen, J. Liu, X. Sun, G. Zhao, Y. Chang, L. Xu, X. Li, Y. Zhao, H. Zheng, Y. Zhao
766 and Z. Wu (2014). "Genome-wide identification and functional annotation of
767 *Plasmodium falciparum* long noncoding RNAs from RNA-seq data." *Parasitol Res*
768 **113**(4): 1269-1281.

769 Liu, Y., R. Tewari, J. Ning, A. M. Blagborough, S. Garbom, J. Pei, N. V. Grishin, R. E. Steele,
770 R. E. Sinden, W. J. Snell and O. Billker (2008). "The conserved plant sterility gene
771 HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and
772 *Plasmodium* gametes." *Genes Dev* **22**(8): 1051-1068.

773 Lopez-Barragan, M. J., J. Lemieux, M. Quinones, K. C. Williamson, A. Molina-Cruz, K. Cui, C.
774 Barillas-Mury, K. Zhao and X. Z. Su (2011). "Directional gene expression and antisense
775 transcripts in sexual and asexual stages of *Plasmodium falciparum*." *BMC Genomics* **12**:
776 587.

777 Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and
778 dispersion for RNA-seq data with DESeq2." *Genome Biol* **15**(12): 550.

779 Marchese, F. P., I. Raimondi and M. Huarte (2017). "The multidimensional mechanisms of long
780 noncoding RNA function." *Genome Biol* **18**(1): 206.

781 Matos, L. V. S., J. McEvoy, S. Tzipori, K. D. S. Bresciani and G. Widmer (2019). "The
782 transcriptome of *Cryptosporidium* oocysts and intracellular stages." *Sci Rep* **9**(1): 7856.

783 Mauzy, M. J., S. Enomoto, C. A. Lancto, M. S. Abrahamsen and M. S. Rutherford (2012). "The
784 *Cryptosporidium parvum* transcriptome during in vitro development." *PLoS One* **7**(3):
785 e31715.

786 Miller, C. N., L. Josse, I. Brown, B. Blakeman, J. Povey, L. Yiangou, M. Price, J. Cinatl, Jr., W.
787 F. Xue, M. Michaelis and A. D. Tsaoasis (2018). "A cell culture platform for
788 *Cryptosporidium* that enables long-term cultivation and new tools for the systematic
789 investigation of its biology." *Int J Parasitol* **48**(3-4): 197-201.

790 Ming, Z., A. Y. Gong, Y. Wang, X. T. Zhang, M. Li, N. W. Mathy, J. K. Strauss-Soukup and X.
791 M. Chen (2017). "Involvement of *Cryptosporidium parvum* Cdg7_Flc_1000 RNA in the
792 Attenuation of Intestinal Epithelial Cell Migration via Trans-suppression of Host Cell
793 SMPD3 Gene." *J Infect Dis*.

794 Mirhashemi, M. E., F. Noubary, S. Chapman-Bonofigli, S. Tzipori, G. S. Huggins and G.
795 Widmer (2018). "Transcriptome analysis of pig intestinal cell monolayers infected with
796 *Cryptosporidium parvum* asexual stages." *Parasit Vectors* **11**(1): 176.

797 Morada, M., S. Lee, L. Gunther-Cummins, L. M. Weiss, G. Widmer, S. Tzipori and N. Yarlett
798 (2016). "Continuous culture of *Cryptosporidium parvum* using hollow fiber technology."
799 *Int J Parasitol* **46**(1): 21-29.

800 Necsulea, A., M. Soumillon, M. Warnefors, A. Liechti, T. Daish, U. Zeller, J. C. Baker, F.
801 Grutzner and H. Kaessmann (2014). "The evolution of lncRNA repertoires and
802 expression patterns in tetrapods." *Nature* **505**(7485): 635-640.

803 Niknafs, Y. S., B. Pandian, H. K. Iyer, A. M. Chinnaiyan and M. K. Iyer (2017). "TACO
804 produces robust multisample transcriptome assemblies from RNA-seq." *Nat Methods*
805 **14**(1): 68-70.

806 Nowick, K., M. B. Walter Costa, C. Honer Zu Siederdissen and P. F. Stadler (2019). "Selection
807 Pressures on RNA Sequences and Structures." *Evol Bioinform Online* **15**:
808 1176934319871919.

809 Oberstaller, J., S. J. Joseph and J. C. Kissinger (2013). "Genome-wide upstream motif analysis of
810 *Cryptosporidium parvum* genes clustered by expression profile." *BMC Genomics* **14**:
811 516.

812 Oberstaller, J., Y. Pumpalova, A. Schieler, M. Llinas and J. C. Kissinger (2014). "The
813 *Cryptosporidium parvum* ApiAP2 gene family: insights into the evolution of
814 apicomplexan AP2 regulatory systems." *Nucleic Acids Res* **42**(13): 8271-8284.

815 Painter, J. E., M. C. Hlavsa, S. A. Collier, L. Xiao, J. S. Yoder, C. Centers for Disease and
816 Prevention (2015). "Cryptosporidiosis surveillance -- United States, 2011-2012."
817 *MMWR Suppl* **64**(3): 1-14.

818 Pertea, M., G. M. Pertea, C. M. Antonescu, T. C. Chang, J. T. Mendell and S. L. Salzberg
819 (2015). "StringTie enables improved reconstruction of a transcriptome from RNA-seq
820 reads." *Nat Biotechnol* **33**(3): 290-295.

821 Platts-Mills, J. A., S. Babji, L. Bodhidatta, J. Gratz, R. Haque, A. Havit, B. J. McCormick, M.
822 McGrath, M. P. Olortegui, A. Samie, S. Shakoor, D. Mondal, I. F. Lima, D. Hariraju, B.

823 B. Rayamajhi, S. Qureshi, F. Kabir, P. P. Yori, B. Mufamadi, C. Amour, J. D. Carreon, S.
824 A. Richard, D. Lang, P. Bessong, E. Mduma, T. Ahmed, A. A. Lima, C. J. Mason, A. K.
825 Zaidi, Z. A. Bhutta, M. Kosek, R. L. Guerrant, M. Gottlieb, M. Miller, G. Kang, E. R.
826 Hourt and M.-E. N. Investigators (2015). "Pathogen-specific burdens of community
827 diarrhoea in developing countries: a multisite birth cohort study (MAL-ED)." *Lancet
828 Glob Health* **3**(9): e564-575.

829 Prevention, C. f. D. C. a. "Parasites - *Cryptosporidium*." Retrieved 22Nov 2017, from
830 <https://www.cdc.gov/parasites/crypto/index.html>.

831 Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing
832 genomic features." *Bioinformatics* **26**(6): 841-842.

833 Ramaprasad, A., T. Mourier, R. Naeem, T. B. Malas, E. Moussa, A. Panigrahi, S. J. Vermont, T.
834 D. Otto, J. Wastling and A. Pain (2015). "Comprehensive evaluation of *Toxoplasma
835 gondii* VEG and *Neospora caninum* LIV genomes with tachyzoite stage transcriptome
836 and proteome defines novel transcript features." *PLoS One* **10**(4): e0124473.

837 Ray, D., H. Kazan, K. B. Cook, M. T. Weirauch, H. S. Najafabadi, X. Li, S. Gueroussov, M.
838 Albu, H. Zheng, A. Yang, H. Na, M. Irimia, L. H. Matzat, R. K. Dale, S. A. Smith, C. A.
839 Yarosh, S. M. Kelly, B. Nabet, D. Mecenas, W. Li, R. S. Laishram, M. Qiao, H. D.
840 Lipshitz, F. Piano, A. H. Corbett, R. P. Carstens, B. J. Frey, R. A. Anderson, K. W.
841 Lynch, L. O. Penalva, E. P. Lei, A. G. Fraser, B. J. Blencowe, Q. D. Morris and T. R.
842 Hughes (2013). "A compendium of RNA-binding motifs for decoding gene regulation."
843 *Nature* **499**(7457): 172-177.

844 Ren, G. J., X. C. Fan, T. L. Liu, S. S. Wang and G. H. Zhao (2018). "Genome-wide analysis of
845 differentially expressed profiles of mRNAs, lncRNAs and circRNAs during
846 *Cryptosporidium baileyi* infection." *BMC Genomics* **19**(1): 356.

847 Rider, S. D., Jr. and G. Zhu (2010). "*Cryptosporidium*: genomic and biochemical features." *Exp
848 Parasitol* **124**(1): 2-9.

849 Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). "edgeR: a Bioconductor package for
850 differential expression analysis of digital gene expression data." *Bioinformatics* **26**(1):
851 139-140.

852 Sateriale, A., M. Pawlowic, S. Vinayak, C. Brooks and B. Striepen (2020). "Genetic
853 Manipulation of *Cryptosporidium parvum* with CRISPR/Cas9." *Methods Mol Biol* **2052**:
854 219-228.

855 Shen, T., H. Li, Y. Song, J. Yao, M. Han, M. Yu, G. Wei and T. Ni (2018). "Antisense
856 transcription regulates the expression of sense gene via alternative polyadenylation."
857 *Protein Cell* **9**(6): 540-552.

858 Siegel, T. N., C. C. Hon, Q. Zhang, J. J. Lopez-Rubio, C. Scheidig-Benatar, R. M. Martins, O.
859 Sismeiro, J. Y. Coppee and A. Scherf (2014). "Strand-specific RNA-Seq reveals
860 widespread and developmentally regulated transcription of natural antisense transcripts in
861 *Plasmodium falciparum*." *BMC Genomics* **15**: 150.

862 Slapeta, J. (2013). "Cryptosporidiosis and *Cryptosporidium* species in animals and humans: a
863 thirty colour rainbow?" *Int J Parasitol* **43**(12-13): 957-970.

864 Sow, S. O., K. Muhsen, D. Nasrin, W. C. Blackwelder, Y. Wu, T. H. Farag, S. Panchalingam, D.
865 Sur, A. K. Zaidi, A. S. Faruque, D. Saha, R. Adegbola, P. L. Alonso, R. F. Breiman, Q.
866 Bassat, B. Tamboura, D. Sanogo, U. Onwuchekwa, B. Manna, T. Ramamurthy, S.
867 Kanungo, S. Ahmed, S. Qureshi, F. Quadri, A. Hossain, S. K. Das, M. Antonio, M. J.
868 Hossain, I. Mandomando, T. Nhampossa, S. Acacio, R. Omore, J. O. Oundo, J. B.

869 Ochieng, E. D. Mintz, C. E. O'Reilly, L. Y. Berkeley, S. Livio, S. M. Tenant, H.
870 Sommerfelt, J. P. Nataro, T. Ziv-Baran, R. M. Robins-Browne, V. Mishcherkin, J. Zhang,
871 J. Liu, E. R. Houpt, K. L. Kotloff and M. M. Levine (2016). "The Burden of
872 *Cryptosporidium* Diarrheal Disease among Children < 24 Months of Age in
873 Moderate/High Mortality Regions of Sub-Saharan Africa and South Asia, Utilizing Data
874 from the Global Enteric Multicenter Study (GEMS)." *PLoS Negl Trop Dis* **10**(5):
875 e0004729.

876 Tandel, J., E. D. English, A. Sateriale, J. A. Gullicksrud, D. P. Beiting, M. C. Sullivan, B.
877 Pinkston and B. Striepen (2019). "Life cycle progression and sexual development of the
878 apicomplexan parasite *Cryptosporidium parvum*." *Nat Microbiol* **4**(12): 2226-2236.

879 Tang, R., X. Mei, Y. C. Wang, X. B. Cui, G. Zhang, W. Li and S. Y. Chen (2019). "LncRNA
880 GAS5 regulates vascular smooth muscle cell cycle arrest and apoptosis via p53 pathway."
881 *Biochim Biophys Acta Mol Basis Dis* **1865**(9): 2516-2525.

882 Templeton, T. J., L. M. Iyer, V. Anantharaman, S. Enomoto, J. E. Abrahante, G. M.
883 Subramanian, S. L. Hoffman, M. S. Abrahamsen and L. Aravind (2004). "Comparative
884 analysis of apicomplexa and genomic diversity in eukaryotes." *Genome Res* **14**(9): 1686-
885 1695.

886 Teodorovic, S., C. D. Walls and H. G. Elmendorf (2007). "Bidirectional transcription is an
887 inherent feature of *Giardia lamblia* promoters and contributes to an abundance of sterile
888 antisense transcripts throughout the genome." *Nucleic Acids Res* **35**(8): 2544-2553.

889 Tsoi, L. C., M. K. Iyer, P. E. Stuart, W. R. Swindell, J. E. Gudjonsson, T. Tejasvi, M. K. Sarkar,
890 B. Li, J. Ding, J. J. Voorhees, H. M. Kang, R. P. Nair, A. M. Chinnaiyan, G. R. Abecasis
891 and J. T. Elder (2015). "Analysis of long non-coding RNAs highlights tissue-specific
892 expression patterns and epigenetic profiles in normal and psoriatic skin." *Genome Biol*
893 **16**: 24.

894 Tushev, G., C. Glock, M. Heumuller, A. Biever, M. Jovanovic and E. M. Schuman (2018).
895 "Alternative 3' UTRs Modify the Localization, Regulatory Potential, Stability, and
896 Plasticity of mRNAs in *Neuronal Compartments*." *Neuron* **98**(3): 495-511 e496.

897 Ulitsky, I., A. Shkumatava, C. H. Jan, H. Sive and D. P. Bartel (2011). "Conserved function of
898 lincRNAs in vertebrate embryonic development despite rapid sequence evolution." *Cell*
899 **147**(7): 1537-1550.

900 Vembar, S. S., A. Scherf and T. N. Siegel (2014). "Noncoding RNAs as emerging regulators of
901 *Plasmodium falciparum* virulence gene expression." *Curr Opin Microbiol* **20**: 153-161.

902 Villacorta, I., D. de Graaf, G. Charlier and J. E. Peeters (1996). "Complete development of
903 *Cryptosporidium parvum* in MDBK cells." *FEMS Microbiol Lett* **142**(1): 129-132.

904 Vinayak, S., M. C. Pawlowic, A. Sateriale, C. F. Brooks, C. J. Studstill, Y. Bar-Peled, M. J.
905 Cipriano and B. Striepen (2015). "Genetic modification of the diarrhoeal pathogen
906 *Cryptosporidium parvum*." *Nature* **523**(7561): 477-480.

907 Walter Costa, M. B., C. Honer Zu Siederdissen, M. Dunjic, P. F. Stadler and K. Nowick (2019).
908 "SSS-test: a novel test for detecting positive selection on RNA secondary structure."
909 *BMC Bioinformatics* **20**(1): 151.

910 Wang, K. C., Y. W. Yang, B. Liu, A. Sanyal, R. Corces-Zimmerman, Y. Chen, B. R. Lajoie, A.
911 Protacio, R. A. Flynn, R. A. Gupta, J. Wysocka, M. Lei, J. Dekker, J. A. Helms and H. Y.
912 Chang (2011). "A long noncoding RNA maintains active chromatin to coordinate
913 homeotic gene expression." *Nature* **472**(7341): 120-124.

914 Wang, Y., A. Y. Gong, S. Ma, X. Chen, Y. Li, C. J. Su, D. Norall, J. Chen, J. K. Strauss-Soukup
915 and X. M. Chen (2017). "Delivery of Parasite RNA Transcripts Into Infected Epithelial
916 Cells During *Cryptosporidium* Infection and Its Potential Impact on Host Gene
917 Transcription." *J Infect Dis* **215**(4): 636-643.

918 Wang, Y., A. Y. Gong, S. Ma, X. Chen, J. K. Strauss-Soukup and X. M. Chen (2017). "Delivery
919 of parasite Cdg7_Flc_0990 RNA transcript into intestinal epithelial cells during
920 *Cryptosporidium parvum* infection suppresses host cell gene transcription through
921 epigenetic mechanisms." *Cell Microbiol* **19**(11).

922 Wilke, G., L. J. Funkhouser-Jones, Y. Wang, S. Ravindran, Q. Wang, W. L. Beatty, M. T.
923 Baldridge, K. L. VanDussen, B. Shen, M. S. Kuhlenschmidt, T. B. Kuhlenschmidt, W. H.
924 Witola, T. S. Stappenbeck and L. D. Sibley (2019). "A Stem-Cell-Derived Platform
925 Enables Complete *Cryptosporidium* Development *In Vitro* and Genetic Tractability." *Cell*
926 *Host Microbe* **26**(1): 123-134 e128.

927 Wilke, G., Y. Wang, S. Ravindran, T. Stappenbeck, W. H. Witola and L. D. Sibley (2020). "In
928 Vitro Culture of *Cryptosporidium parvum* Using Stem Cell-Derived Intestinal Epithelial
929 Monolayers." *Methods Mol Biol* **2052**: 351-372.

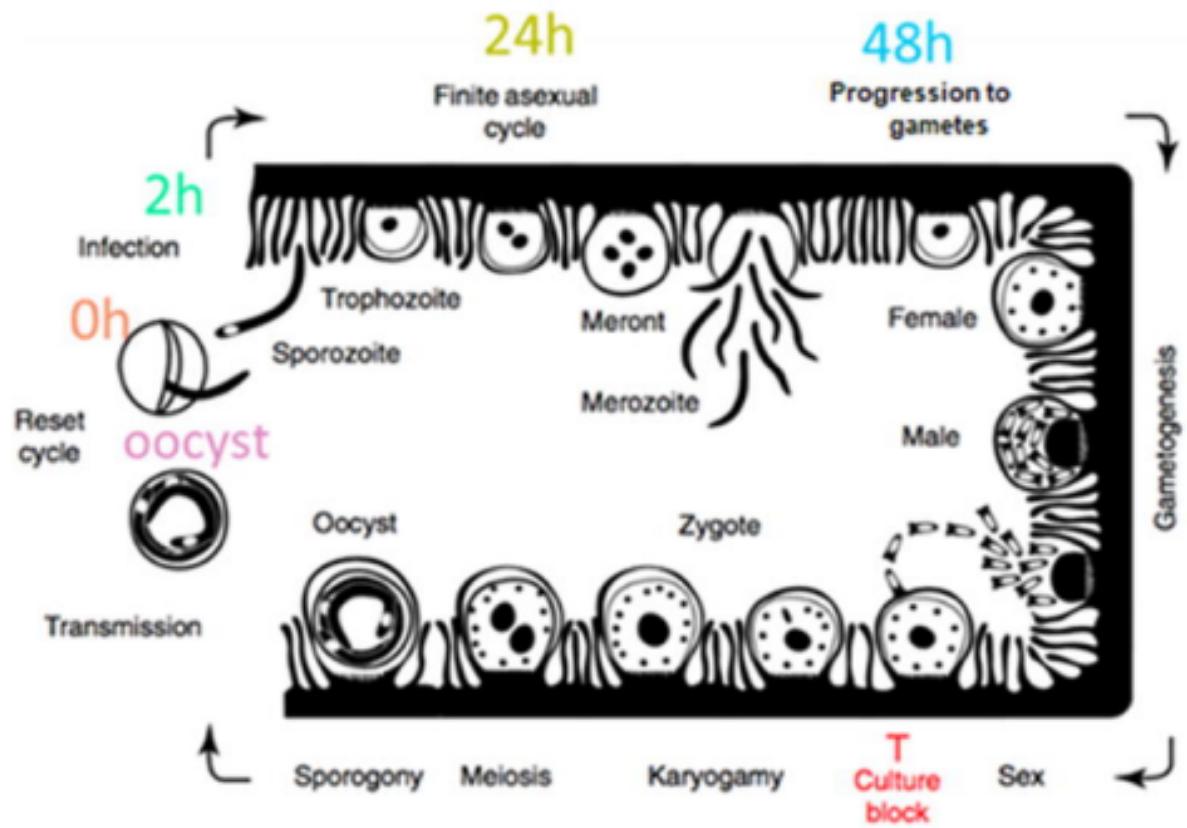
930 Yarlett, N., M. Morada, M. Gobin, W. Van Voorhis and S. Arnold (2020). "In Vitro Culture of
931 *Cryptosporidium parvum* Using Hollow Fiber Bioreactor: Applications for Simultaneous
932 Pharmacokinetic and Pharmacodynamic Evaluation of Test Compounds." *Methods Mol*
933 *Biol* **2052**: 335-350.

934 Zhang, H., F. Guo, H. Zhou and G. Zhu (2012). "Transcriptome analysis reveals unique
935 metabolic features in the *Cryptosporidium parvum* Oocysts associated with
936 environmental survival and stresses." *BMC Genomics* **13**: 647.

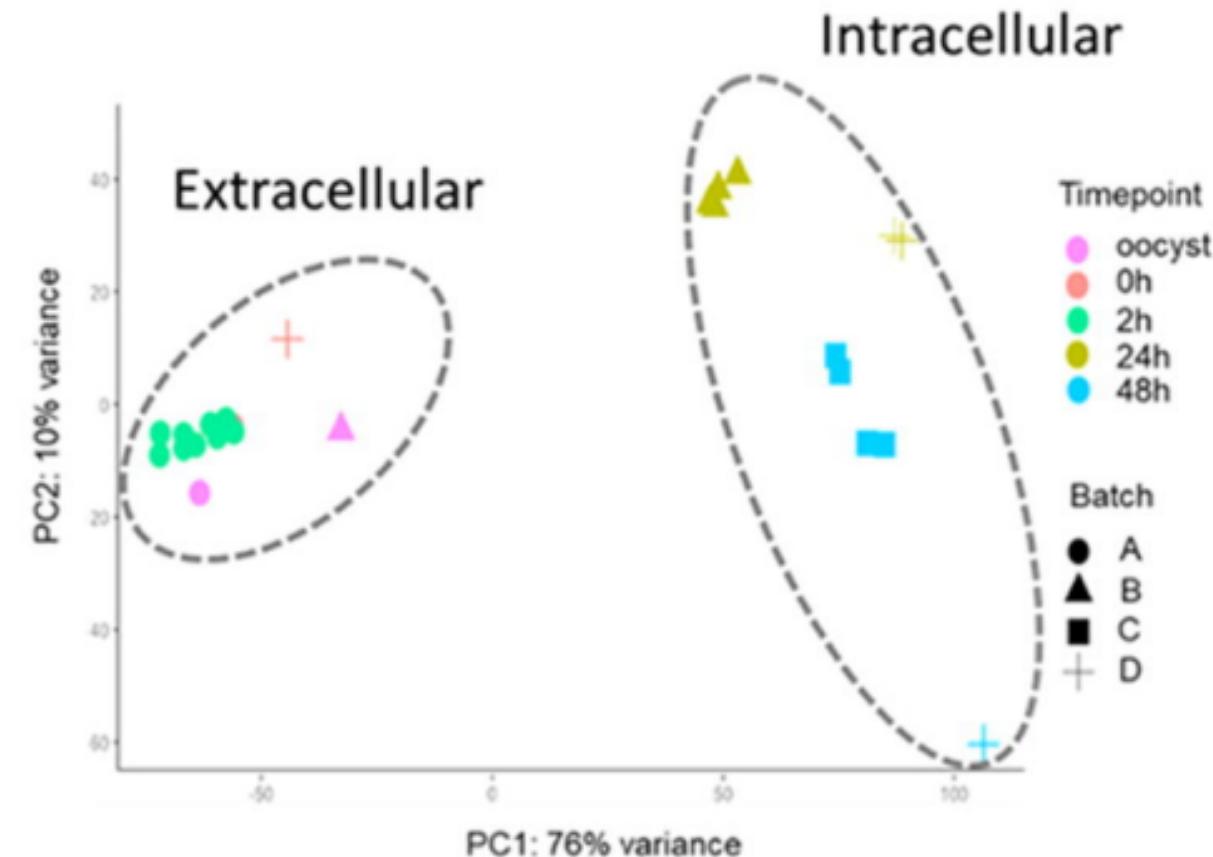
937 Zhang, L., X. Luo, F. Chen, W. Yuan, X. Xiao, X. Zhang, Y. Dong, Y. Zhang and Y. Liu (2018).
938 "LncRNA SNHG1 regulates cerebrovascular pathologies as a competing endogenous
939 RNA through HIF-1alpha/VEGF signaling in ischemic stroke." *J Cell Biochem* **119**(7):
940 5460-5472.

941

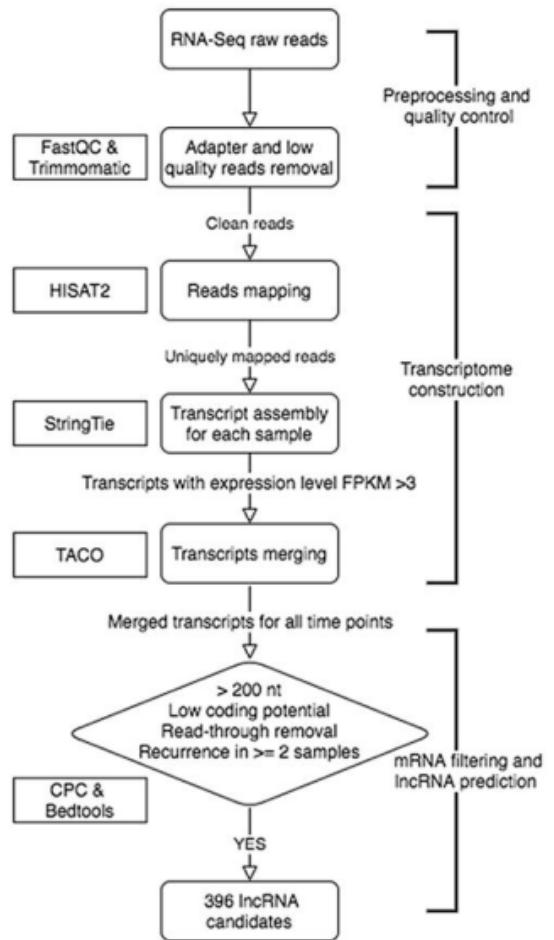
A



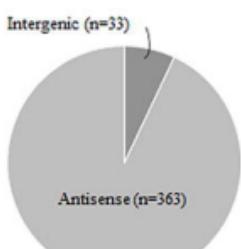
B



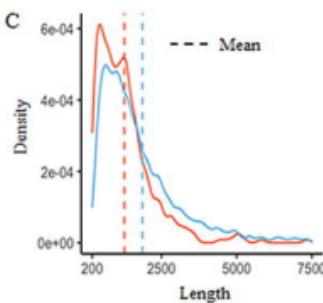
A



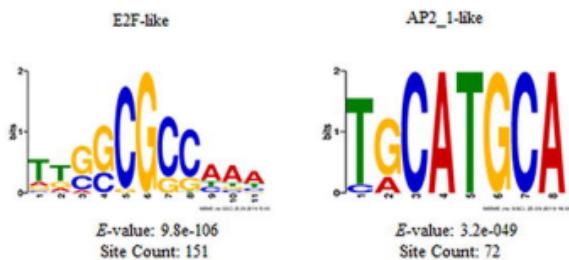
B



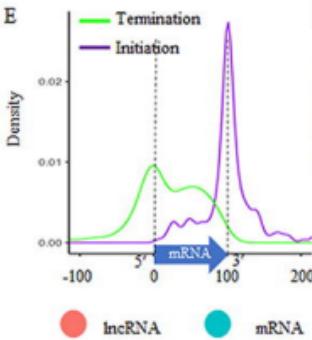
C



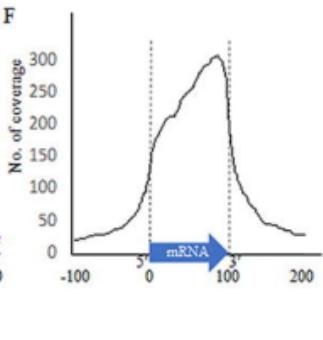
D



E

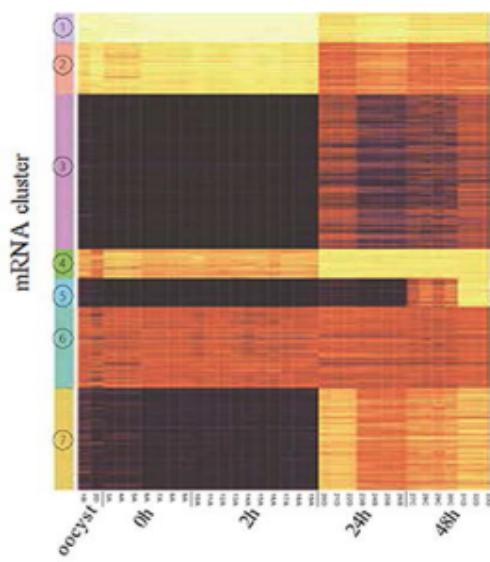


F



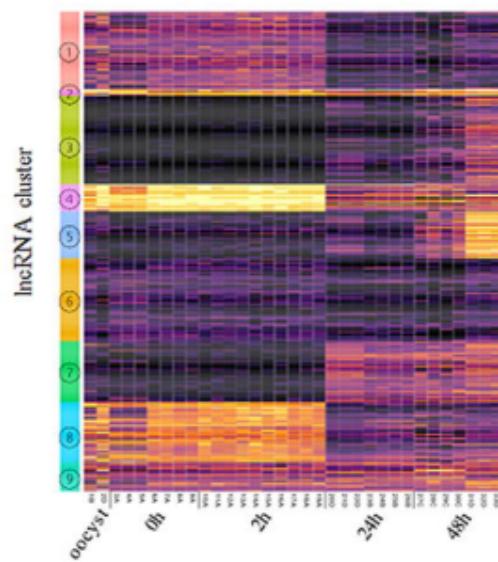
A

mRNA expression heatmap

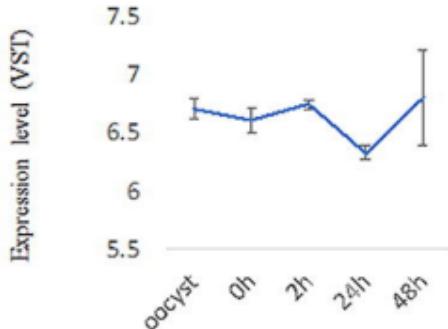


B

lncRNA expression heatmap



C



D

DE	mRNA	lncRNA
<i>oocyst/sporozoite-->asexual stage</i>		
↑	1715	80
↓	1270	218
<i>asexual stage-->sexual stage</i>		
↑	403	85
↓	438	8

