Establishment of quantitative RNAi-based forward genetics in *Entamoeba histolytica* and identification of genes required for growth

Akhila Bettadapur[1], Samuel S. Hunter[2*], Rene L. Suleiman[1*], Maura C. Ruyechan[1], Wesley Huang[1], Charles G. Barbieri[3], Hannah W. Miller[1], Tammie S.Y. Tam[1], Matthew L. Settles[2], and Katherine S. Ralston[1#]

[1]Department of Microbiology and Molecular Genetics, University of California, Davis, USA

[2]Genome Center, University of California, Davis, USA

[3]SeqMatic, LLC, Fremont, California, USA

[*]These authors contributed equally to this work

Running Head: Forward genetics and identification of genes required for *Entamoeba histolytica* growth

[#]Address correspondence to Katherine S. Ralston, ksralston@ucdavis.edu

17 **Author Contributions**

18 A.B. designed and performed the experiments, except fragmentation of gDNA,
19 immunofluorescence, and growth curves with expression of full-length or truncated
20 proteins. R.L.S. performed SG1 and SG2 immunofluorescence experiments. M.C.R.
21 performed growth curves with expression of full-length or truncated proteins. W.H.
22 generated transfectants expressing full-length SG2. C.G.B. fragmented gDNA and processed
23 the fragments to enable subsequent plasmid library construction. T.S.Y.T. and H.W.M.
24 performed Gal/GalNAc lectin immunofluorescence experiments. S.S.H. designed the
25 analysis strategy for Illumina sequencing data, and analyzed all Illumina sequencing data.
26 M.L.S. oversaw the Illumina analysis approach. K.S.R. performed growth curves, conceived
27 of the overall approach and oversaw the design and analysis of the experiments. A.B.,
28 S.S.H., and K.S.R. wrote the manuscript.

Abstract

While *Entamoeba histolytica* remains a globally important pathogen, it is dramatically understudied. The tractability of *E. histolytica* has historically been limited, which is largely due to challenging features of its genome. To enable forward genetics, we constructed and validated the first genome-wide *E. histolytica* RNAi knockdown mutant library. This library allows for Illumina deep sequencing analysis for quantitative identification of mutants that are enriched or depleted after selection. We developed a novel analysis pipeline to precisely define and quantify gene fragments. We used the library to perform the first RNAi screen in *E. histolytica* and identified slow growth (SG) mutants. Among genes targeted in SG mutants, many had annotated functions consistent with roles in cellular growth or metabolic pathways. Some targeted genes were annotated as hypothetical or lacked annotated domains, supporting the power of forward genetics in uncovering functional information that cannot be gleaned from databases. While the localization of neither of the proteins targeted in SG1 nor SG2 mutants could be predicted by sequence analysis, we showed experimentally that SG1 localized to the cytoplasm and cell surface, while SG2 localized to the cytoplasm. Overexpression of SG1 led to increased growth, while expression of a truncation mutant did not lead to increased growth, and thus aided in defining functional domains in this protein. Finally, in addition to establishing forward genetics, we uncovered new details of the unusual *E. histolytica* RNAi pathway. These studies dramatically improve the tractability of *E. histolytica* and open up the possibility of applying genetics to improve understanding of this important pathogen.

Author Summary

Entamoeba histolytica is a globally important pathogen that is dramatically understudied. One of the major limitations of this organism is its challenging genome. RNAi is the state-of-the-art tool for genetic manipulation in *E. histolytica,* though the RNAi pathway has several noncanonical features. Here, we harnessed the RNAi pathway to enable RNAi-based forward genetics for the first time in this organism. We validated the RNAi library by performing the first *E. histolytica* RNAi screen and identified slow growth mutants. We showed that independently-generated mutants also exhibited slow growth phenotypes, and we characterized protein localization and domains for some of the identified slow growth genes. The RNAi library that we constructed enables modern, quantitative Illumina deep sequencing analysis to identify mutants that are enriched or depleted after selection. We developed a novel analysis pipeline to precisely define and quantify full-length gene fragments inferred from read mapping. Our approach differs from previous approaches for analysis of RNAi screens, and it better represents the actual DNA fragments and their quantities. This study dramatically improves the tractability of this important pathogen. Moreover, the strategies behind this RNAi library, and its analysis, are novel, and can be applied to other organisms.

67 **Introduction**

68 *Entamoeba histolytica* is the causative agent of amoebiasis in humans. *E. histolytica*
69 is common in developing nations and is responsible for ~50 million infections/year [1]. In an
70 endemic area, approximately 80% of infants are infected with *E. histolytica* [2]. Malnutrition
71 and stunting are associated with childhood *E. histolytica* infections [3]. In children enrolled
72 in the Global Enteric Multicenter Study, *E. histolytica* infection was associated with the
73 highest risk of death between enrollment and follow up [4]. Amoebiasis results in an
74 estimated 15,500 deaths/year in children and 67,900 deaths/year in people of all ages [5].
75 There is no vaccine and therapy is limited [6]. The continued morbidity and mortality
76 indicate that current treatment and prevention approaches are insufficient.

77 Despite its impact on human health, *E. histolytica* is dramatically understudied
78 relative to other parasites. The genome is relatively difficult to manipulate. *E. histolytica* is
79 tetraploid [7]. The genome contains many expanded gene families, which comprise 56% of
80 the proteome [8, 9]. The genome is ~75% A + T, which creates challenges for genomic
81 approaches and has contributed to the relatively poor draft genome quality. The genome is
82 ~20 Mb with ~8,000 genes [8, 10]. For one third of genes, no homologues have been
83 identified [10] and less than half of genes have gene ontology (GO) terms [8]. Thus, with
84 limited sequence homology information and a difficult genome, functional approaches such
85 as forward genetics are needed to identify genes that are relevant to pathogenesis.

86 Historically, the genetic tractability of *E. histolytica* has been limited. Forward
87 genetics have not yet been established in *E. histolytica*, outside of two studies that used a
88 plasmid-based overexpression library to screen for phenotypes [11, 12]. More recently, the

89 endogenous RNAi pathway in *E. histolytica* has been characterized, and exploited as a useful
90 tool for targeted gene knockdown. Very recently, CRISPR/Cas9 has been established in *E.*
91 *histolytica* [13], providing an important new tool for future genetic manipulation of this
92 organism.

93 The existence of an RNAi pathway in *E. histolytica* that could be exploited for gene
94 knockdown was implied by an approach that knocked down expression of amoebapore A
95 [14]. However, the mechanism was unclear, and the expression of off-target genes was
96 affected [15, 16]. A clonal line of stable transfectants carrying the amoebapore A plasmid
97 was named the "G3" strain, and was subsequently used to knockdown additional genes in
98 the context of an amoebapore A knockdown mutant background [15]. Other strategies for
99 gene knockdown have included expression of genes in the antisense orientation [17],
100 expression of long double stranded RNA [18], phagocytosis of bacteria expressing dsRNA
101 [19], and expression of short hairpin RNA [20]. These techniques have been useful for
102 reverse genetic studies, but not robust enough to be effective for every gene of interest.

103 In later studies, the *E. histolytica* RNAi pathway was characterized in detail. Small
104 RNAs were found to be 27nt in length and 5' polyphosphorylated [21]. 5'
105 polyphosphorylated siRNAs have only been seen in *E. histolytica* and *Caenorhabditis elegans*
106 [22]. Supporting that these small RNAs are relevant to previous approaches that exploited
107 gene silencing in *E. histolytica,* in the G3 strain, 27nt small RNAs mapped to amoebapore A,
108 and when additional genes were silenced in this background, new small RNAs mapped to
109 the newly silenced gene [23]. There is no obvious Dicer homologue, but a potential
110 noncanonical Dicer and Argonaute homologues have been characterized [24, 25].

111 Because endogenous small RNAs are polyphosphorylated, it was initially a challenge
112 to exploit this pathway for gene knockdown. A major breakthrough was the discovery that
113 RNAi knockdown could be induced by spreading of silencing [26]. A new "trigger" plasmid
114 approach was established, which relies on endogenous small RNAs directed to short
115 fragment of an endogenously-silenced gene (called the trigger) to drive spreading of
116 silencing to a fragment of gene of interest adjacent to the trigger [26]. This leads to long-
117 term silencing of the gene of interest and the deposition of a repressive histone mark [27].
118 This approach has been successfully applied in many studies, and in *E. invadens* [28-30].
119 In other parasites, RNAi knockdown has been successfully adapted for forward
120 genetic approaches, and this has been very effective in uncovering genes relevant to
121 parasite metabolism [31, 32], DNA repair [33], virulence [34], transmission [35, 36], and
122 drug therapies [37-40]. Initially, these approaches relied on the generation of clonal lines
123 after selection, and Sanger sequencing analysis of individual clones to identify the knocked
124 down gene [31, 32]. More recent studies used next generation sequencing on non-clonal
125 mutant populations in order to identify mutants and quantify their relative abundance [33-
126 40]. The latter approaches are useful because they can quantitatively identify mutants that
127 are enriched by selection, and mutants that are less abundant or are lost after selection.
128 To enable forward genetics, we constructed and validated the first genome-wide *E.*
129 *histolytica* RNAi knockdown mutant plasmid library. Its design allows for Illumina deep
130 sequencing, and our novel analysis pipeline allows for quantitative identification of plasmids
131 that are enriched or depleted after selection. Using a pilot version of the library, we
132 performed the first RNAi screen in *E. histolytica* and identified slow growth mutants. These

133 studies establish the first RNAi screen in *E. histolytica*, thereby dramatically improving the

134 tractability of this pathogen.

135 **Results**

136 **Approach for quantitative RNAi-based forward genetics**

137 We devised an overall strategy for RNAi-based forward genetics in *E. histolytica* (Fig.
138 1). For this approach, a library of gene fragments would be introduced into the *E. histolytica*
139 "trigger" RNAi plasmid (pTrigger) [26], to create an RNAi library (Fig. 1a). Amoebae would
140 then be stably-transfected with the library, and exposed to a selective pressure or screening
141 strategy. Illumina deep sequencing analysis would then be used to sequence the inserts and
142 to compare the selected amoebae to control, non-selected amoebae, in order to identify
143 gene fragments that are quantitatively enriched or depleted. We empirically determined
144 that the stable transfection efficiency with the backbone of pTrigger is ~1/1000 amoebae.
145 This relatively high stable transfection efficiency suggested that transfecting enough
146 amoebae to achieve genome-wide RNAi knockdown would be feasible.

147 The Illumina priming sequences that are necessary for sequencing are typically
148 added to samples by using a nested strategy with two sequential PCR reactions. We
149 modified the trigger plasmid in order to flank the insert with Illumina adaptor sequences,
150 creating pTriggerAdaptor, so that after extracting amoebic genomic DNA (gDNA), only a
151 single PCR step would be needed to generate DNA ready for sequencing (Fig. 1b). Using
152 fewer PCR reactions is expected to reduce the potential for PCR bias that can change the
153 representation of sequences. An additional benefit of this modification is that the original
154 sequences flanking the insert were A/T rich and were not optimal for primer design. The
155 insertion of these Illumina sequences into the knockdown plasmid did not change
156 knockdown efficacy (Fig. 1c).

157 **Construction of a plasmid RNAi library**

158 Building on these findings, we devised a strategy to construct a genome-wide
159 plasmid RNAi library (Fig. 2a). We reasoned that since ~50% of the genome is coding,
160 constructing a gDNA library would be a practical approach. To generate the plasmid library,
161 random fragments of gDNA were inserted into the pTriggerAdaptor plasmid. Library
162 construction was initially done on a small scale, to create a "pilot" library, where plasmids
163 were harvested from a limited number of bacterial colonies. Building on the results of the
164 pilot, a second library was constructed on a larger scale, to make a "final" library that was
165 expected to have genome-wide coverage.

166 Previous studies defined that an insert size ≥ 500 bp is sufficient for knockdown [26,
167 41]. Larger insert sizes are likely to lead to more efficient knockdown, but fragments > 1 kb
168 are problematic for Illumina sequencing. The pilot library was constructed using gDNA
169 fragments between ~400 – 600 bp. Building on the robust Illumina sequencing results from
170 the pilot library (described in detail below), we constructed the final library, where gDNA
171 fragments were larger (between ~500 – 900 bp) in order to maximize the efficacy of RNAi
172 knockdown.

173 To make the RNAi library, high quality, RNA-free, gDNA (Fig. S1a – b) was sonicated
174 and size-selected (Fig. S1c). Y-shaped adaptors were ligated to the fragmented DNA to
175 ensure that each fragment was flanked by sequences matching the trigger plasmid
176 upstream and downstream of the cloning site, in order to enable subsequent Gibson
177 cloning. Adaptor ligation was successful, with an appropriate size shift seen using
178 bioanalyzer analysis (Fig. 2b – c). Gibson cloning and plasmid preps were done *en masse*

179 with pooled plasmids. To check cloning efficiency, individual colonies were obtained and
180 analyzed. Restriction analysis demonstrated that the ligation efficiency was very high, since
181 63 out of 67 plasmids examined contained an insert (Fig. 2d).

182 **The RNAi library has genome-wide coverage**

183 Illumina sequencing was used to analyze the pilot and final libraries, and the
184 fragmented gDNA that was used to construct each library. Notably, our approach for
185 analysis of the sequencing data is novel and differs from previous approaches for analysis of
186 RNAi screens [33-40] since all results presented here are based on full-length DNA-
187 fragments inferred from read mapping results (Fig. S2), rather than on more traditional read
188 mapping coverage or CDS overlap counting strategies. Our approach was designed to better
189 represent the actual DNA fragments present in the plasmid library, instead of just the short
190 ends of these fragments that were sequenced on the Illumina platform. This strategy
191 allowed us to precisely define whether a fragment might target one or multiple genes, and
192 enabled a clearer view of fragment based genome coverage. This approach will also allow
193 us to readily quantify the representation of gene fragments in future RNAi screen
194 experiments, in order to identify specific fragments that are particularly impactful in a
195 particular screen.

196 There did not appear to be bias in gDNA fragmentation or construction of the
197 library, since the nucleotide and dinucleotide frequencies of the fragmented gDNA and the
198 plasmids were the same as the genome (Fig. 3a and S3a). Rarefaction curves and read
199 fragment analysis showed that sequencing was performed to a sufficient depth to identify
200 all unique fragments within the plasmid samples, but the same depth of sequencing was not

201 deep enough to identify all unique fragments within the fragmented gDNA samples (Fig. S3b
202 – c).
203 Illumina sequencing analysis was performed on separate batches of the plasmid
204 library that had been cloned on different days, and on technical sequencing replicates.
205 Similar read counts per fragment were obtained from each batch and each sequencing
206 replicate, showing that cloning and sequencing were reproducible and there was no
207 evidence for bias within these steps (Fig. S4a – b, Table S1). Analysis of the sequencing
208 results showed that the gDNA fragments that were used to create the pilot library had a
209 mean size of 421 bp, and the fragments in the pilot plasmid library had a mean size of 427
210 bp (Fig. 3b). The gDNA fragments that were used to create the final library had a mean size
211 of 540 bp, and the fragments in the pilot plasmid library had a mean size of 555 bp (Fig. 3b).
212 The nearly identical size ranges of the gDNA fragments and cloned fragments further
213 support that there was no bias during library construction.
214 Genome coverage plots of the final plasmid library over the 5 largest contigs in the
215 genome visually showed that coverage of these contigs was complete, and the depth of
216 coverage was consistent (Fig. 3c). The breadth and depth of coverage over a zoomed in
217 region of the largest contig is represented in Fig. 3d. On a genome-wide level, the genome
218 was well covered by fragments that were cloned into the library in both the forward and
219 reverse orientation relative to the orientation of coding sequences (Fig. 4, Table 1).
220 Supporting that the final library contained genome-wide coverage, 97.24% of the genome
221 was represented (Table 1). 8307 of 8333 annotated genes were represented, and the
222 majority (8301 genes) were represented by more than one unique fragment. Of the 26

223 genes that are missing, 22 are likely to be misannotated in the genome, since they were not
224 seen in any sample, including fragmented gDNA (Fig. S5). Additionally, for the majority of
225 these genes, there was no empirical evidence for gene expression in prior RNAseq or
226 proteomic studies (Tables S2 and S3). From these findings, we concluded that the final
227 plasmid library has genome-wide coverage. We named the final library pEhKDL, since it
228 represents an *E. histolytica* knockdown library.

229 **Forward genetic isolation of slow growing mutants**

230 While the final library was under construction, *E. histolytica* was stably transfected
231 with the pilot library. To determine if the RNAi library could be used to identify genes that
232 contribute to a phenotype, we screened for mutants with abnormal growth. To first ask
233 how many different plasmids were present per stably transfected amoeba, clonal lines were
234 obtained using limiting dilution. Sanger sequencing analysis of clonal lines was consistent
235 with only one unique plasmid (*i.e.,* one unique insert) per transfected amoeba (Fig. S6).

236 To screen for growth phenotypes, a selected number of clonal lines were analyzed.
237 Slow Growth mutants 1 (SG1) and 2 (SG2) were isolated first and chosen for further
238 characterization. These mutants had inserts corresponding to EHI_194310 and EHI_007010,
239 which are annotated as a putative amino acid transporter and a hypothetical protein,
240 respectively (Table 2). These mutants had a statistically significant growth defect (Fig. 5a).
241 We have previously knocked down superoxide dismutase (EHI_159160) and determined
242 that this results in a growth defect; thus, superoxide dismutase RNAi was used as a control.
243 RT-PCR analysis showed that EHI_194310 and EHI_007010 were knocked down in the
244 corresponding mutants (Fig. 5b).

245 To determine if knockdown mutant phenotypes were reproducible, Gibson cloning
246 was used to generate independent knockdown plasmids that contained a fragment of either
247 EHI_194310 or EHI_007010, which differed from the fragment in the library mutants. *E.*
248 *histolytica* was stably transfected with each of these new plasmids to create the
249 independent knockdown mutants, SG1i and SG2i. Growth of heterogeneous, stable
250 transfectants was measured, and these mutants had statistically significant growth defects
251 (Fig. 5c). RT-PCR analysis demonstrated that the corresponding target genes were knocked
252 down (Fig. 5d). The library mutants and independently generated mutants each exhibited
253 statistically significant growth defects (Fig. 5e). Thus, the phenotypes seen in library
254 mutants were reproduced in independently generated mutants.

255 **Localization and domain analysis of proteins targeted in slow growth mutants**

256 To further characterize SG1 and SG2, protein localization was examined. SG1 is
257 annotated as a putative amino acid transporter (Table 2), with predicted transmembrane
258 domains, but no predicted signal peptide sequence. SG2 is annotated as a hypothetical
259 protein (Table 2), with no annotated domains. SG2 also lacks predicted transmembrane
260 domains, predicted lipid modification sites, and a predicted signal peptide sequence. Thus,
261 the localization of neither protein could be definitively inferred from sequence analysis.

262 We found that SG1 was localized to the amoeba surface and was also detected
263 intracellularly in the cytoplasm, and intracellularly with a punctate distribution (Fig. 6a – 6b,
264 S7, S8). Since these transfectants were heterogeneous, there was heterogeneity in the
265 intensity of SG1 staining. The expression level of SG1 was very low (Fig. S8), and was
266 dramatically lower than the expression level of a control, endogenous surface protein (Fig.

267 S9). Interestingly, amoebae overexpressing full-length SG1 consistently exhibited brighter
268 DAPI staining (Fig. 6c) and/or more dividing cells with two nuclei, compared to wild-type
269 amoebae (Fig. 6a – 6b), suggesting that proliferation was higher in these amoebae. Growth
270 analysis showed that amoebae overexpressing full-length SG1 grew faster than vector
271 control transfectants (Fig. 6d). To gain a deeper understanding of domains required for SG1
272 function, a C-terminal truncation mutant was overexpressed, and since it did not lead to
273 increased growth, we concluded that this truncation interfered with protein function (Fig.
274 6d).

275 Unlike SG1, SG2 was not detectable on the amoeba surface (Fig. 7a – 7b, S7, S10).
276 Instead, SG2 exhibited staining throughout the amoeba cytoplasm (Fig. 7a, S10). Since these
277 transfectants were heterogeneous, there was heterogeneity in the intensity of SG2 staining.
278 Like SG1, the expression level of SG2 was very low (Fig. S10), and was dramatically lower
279 than the expression level of a control, endogenous surface protein (Fig. S9). Overexpression
280 of full-length SG2 was not associated with increased growth, and no difference was
281 detected in the intensity of DAPI staining between SG2 overexpressors and wild-type cells
282 (Fig. 7c). Unlike the rabbit antibody used to localize SG1 (Fig. 6a – 6b), the mouse antibody
283 used to localize SG2 was associated with some background nuclear staining in wild-type
284 cells (Fig. 7a – 7b, S7, S11). By contrast, nuclear staining was not pronounced in SG2
285 overexpressors (Fig. S11).

286 **Bioinformatic analysis of slow growing mutants**

287 From the pilot RNAi library transfection, ~64 additional clonal lines were obtained
288 and 10 additional clonal lines with statistically significant growth defects were identified

289 (Fig. S12; Table 2). The insert present in each clonal line was identified by Sanger
290 sequencing, and the results were again consistent with one unique plasmid per transfected
291 amoeba. Inserts ranged in size from 322 to 555 bp. In previous studies that used pTrigger
292 for gene silencing, the gene of interest for knockdown was in the forward orientation and in
293 frame with the trigger [26, 28-30, 41, 42]. In the slow growth mutants, inserts were in the
294 forward or reverse orientation, and in and out of frame (Table 2). Inserts that were in the
295 reverse orientation and out of frame, like SG1, nonetheless led to measurable knockdown
296 (Fig. 5b). Both the directionality of the gene of interest, and the frame, are therefore
297 dispensable for knockdown.

298 Nine of the twelve slow growth mutants contained an insert that mapped to a gene
299 (Table 2). The three mutants with inserts that did not clearly map to an annotated gene
300 (SG3, SG9 and SG12) did map to potentially unannotated ORFs, and also partially mapped
301 (*i.e.,* contained stretches of homology) to annotated genes. BLAST was used to search for
302 homologs of the ORFs that mapped to SG3, SG9, and SG12, and homologs were found in *E.*
303 *dispar* for each ORF (Expect values of 4e -31, 1e -10, 1e -9, respectively), suggesting that
304 these ORFs are genes that were missed in the annotation of the *E. histolytica* genome. The
305 *E. histolytica* genome contains a very large number of gene families [8, 9], evidencing a high
306 level of functional redundancy. One mutant, SG5, contained an insert that mapped to more
307 than one gene (Table 2). Due to the level of identity between the multiple hits, multiple
308 related genes may be knocked down in this mutant.

309 Several mutants contained inserts that mapped to genes with annotated functions
310 consistent with roles necessary for cellular growth, such as an amino acid transporter, a

311 major facilitator transporter, and a splicing factor (Table 2). Three mutants contained
312 inserts with predicted roles in metabolic pathways, based on data from KEGG Metabolic
313 Pathways (https://www.kegg.jp), the MetaCyc Metabolic Pathway Database [43], and
314 Chemical Entities of Biological Interest (ChEBI) database [44] (Table 2). Several mutants
315 contained inserts that mapped to genes annotated as hypothetical, or that lacked
316 annotated domains (Table 2). Thus, in addition to validating the RNAi library approach, the
317 12 slow growth mutants that were uncovered shed light on genes that are required for *E.*
318 *histolytica* growth and emphasize how forward genetics can be used to assign functional
319 information to hypothetical proteins.

320 **Discussion**

321 We constructed and validated the first genome-wide *E. histolytica* RNAi knockdown
322 library. We performed a screen for slow growth mutants and identified 12 new mutants,
323 providing a proof-of-concept that demonstrates the library can be used to identify mutant
324 phenotypes. Independent knockdown mutants phenocopied the growth defects seen in the
325 library mutants, showing that library mutant phenotypes are reproducible. Thus, our studies
326 both identify genes that are critical for growth, and establish RNAi-based forward genetics
327 in *E. histolytica* for the first time. The library represents a valuable new tool for *E.*
328 *histolytica*, that can now be applied towards the study of many processes.

329 Slow growth mutants were identified among clonal knockdown mutant lines. 12
330 mutants were identified among ~80 total clonal lines. Among the identified genes targeted
331 in these mutants, many genes had annotated functions consistent with roles in cellular
332 growth, and/or had predicted roles in metabolic pathways, based on available metabolic
333 pathway databases (https://www.kegg.jp/) [43, 44]. No genes were identified for which a
334 role in cellular growth would be implausible. Several of the identified genes were annotated
335 as hypothetical, or lacked annotated domains. All of the gene hits in our screen had
336 previous evidence of gene expression in RNAseq analyses [45, 46]. Thus, the slow growth
337 phenotype begins to provide functional information to genes for which existing annotation
338 does not inform understanding of protein function. All of the annotated gene hits in our
339 screen had homologues in the related human-infectious *Entamoeba* species *E. dispar, E.*
340 *nuttali* and *E. moshkovskii.* Thus, this type of screen may be informative in identifying
341 potential drug targets that are conserved among human-infectious *Entamoeba* species.

342 In our analysis of slow growth mutants, we seeded cells at a set starting
343 concentration, and then measured the number of cells at various time points thereafter.
344 We defined slow growth phenotypes as those that resulted in significantly lower cell
345 numbers over time, relative to control cells. In the case of cells overexpressing full-length
346 SG1, an increased growth phenotype was also associated with a higher intensity of DAPI
347 staining compared to control amoebae. The assays that we used do not formally distinguish
348 between an altered rate of proliferation and altered longevity of individual cells in culture.
349 The *E. histolytica* genome contains numerous gene families [8, 9]. An advantage of
350 RNAi *vs.* alternative approaches (*e.g.,* gene knockouts) is the potential to simultaneously
351 interfere with multiple gene family members. RNAi is similarly advantageous for genomes
352 that are polyploid, which is also relevant to the tetraploid [7] *E. histolytica* genome. We
353 identified one mutant that contained an insert that mapped to more than one gene. With
354 the high levels of sequence identity, it is likely that multiple genes are knocked down in this
355 mutant. We were able to identify a slow growth phenotype in this mutant, which suggests
356 that the RNAi library may be effective in identifying mutant phenotypes even in the context
357 of high levels of functional redundancy.
358 We have used pTrigger to knockdown many genes in our laboratory, and, in general,
359 have been able to recover knockdown mutants for essentially every targeted gene.
360 Compared to clonal lines, heterogeneous transfectants have typically had more variable
361 levels of knockdown, and sometimes appear to have lower levels of knockdown [47]. We
362 obtain clonal lines by limiting dilution, but another effective available method is cloning by
363 isolating *E. histolytica* colonies on soft agar plates [48]. Among clonal lines obtained from

364 heterogeneous transfectants, we have seen wide variation, including clones with high levels
365 of knockdown and clones that lack knockdown. The SG1 and SG2 clonal lines initially
366 obtained from the library appeared to have an equal or higher level of knockdown than the
367 heterogeneous independent SG1i and SG2i mutants that targeted the same gene. The
368 heterogeneous independent knockdown mutant SG1i had a less severe growth phenotype
369 than the clonal library mutant SG1, which may be due to variation in knockdown in the
370 heterogeneous population. Loss of gene silencing (and loss of a growth phenotype) over
371 time was noted in the heterogeneous SG2i mutants, consistent with the selective
372 disadvantage of the growth phenotype. Related to this point, when a selection is applied to
373 a library mutant population, it will be important to compare selected and non-selected
374 controls, in order to differentiate mutants that are lost due to a growth disadvantage *vs.*
375 mutants that are lost due to selection.

376 The library was designed to facilitate Illumina deep sequencing analysis in order to
377 quantitatively identify gene fragments that are enriched or lost after selection. We
378 constructed pilot and final versions of the library. Multiple lines of evidence supported that
379 there was no bias in library construction, including nucleotide and dinucleotide frequencies,
380 and size comparisons between the gDNA fragments and fragments in the final plasmids.
381 Libraries were prepared in batches that were cloned on different days. Similar read counts
382 per fragment were obtained from batches, again supporting that there was no bias in
383 cloning or sequencing analysis. The majority of the genome (97.24%) was covered in the
384 final plasmid library, and the majority of genes were represented (8307 out of 8333
385 annotated genes). 26 annotated genes are not represented. Of these, 22 genes were not

386 seen in any sample, including gDNA, suggesting that they represent errors in annotation.
387 Consistent with this, there was no published evidence for gene expression for the majority
388 of these genes. The 4 remaining genes that were not seen in the final library had some of
389 the lowest sequencing coverage in the fragmented gDNA sample, suggesting that they were
390 missed due to lack of abundance in the fragmented gDNA that was used for library
391 construction. Thus, with 8307 out of 8311 correctly annotated genes represented, and
392 97.24% of the genome, we concluded that the library represented full genome coverage.
393 A future variation on this approach would be to make a library using fragmented
394 cDNA instead of gDNA. We reasoned that since ~50% of the genome is coding, gDNA is a
395 reasonable approach, since ~50% of the plasmids in the library would lead to gene silencing.
396 A limitation of a gDNA library is that some sequences such as rRNA and tRNA, which are
397 abundant in the genome, are represented in the library, and are not useful for identification
398 of mutant phenotypes. A cDNA library would remove these sequences, but would carry
399 other limitations. The most highly expressed genes would be overrepresented in a cDNA
400 library, and genes that are not expressed in the conditions in which cDNA was obtained
401 would not be represented in the library. The lower abundance of material for fragmentation
402 and cloning is also a consideration. mRNA is a very small proportion of total RNA, and in our
403 initial attempts to prepare cDNA for library construction, insufficient material was
404 recovered. With more initial total RNA, a cDNA library could be constructed in the future,
405 and would be complementary to the gDNA library that was constructed here.
406 We asked whether the frame or the directionality of the insert was critical for
407 knockdown using the trigger plasmid. In previous studies, the gene of interest for

408 knockdown was in the forward orientation and in frame with the trigger [26, 28-30, 41, 42];
409 thus, the frame and orientation requirements had not been tested experimentally. Among
410 the slow growth mutants that we identified, inserts were in the forward or reverse
411 orientation, and in and out of frame. Note that the adaptor sequence that was inserted
412 between the trigger and the insert does not change the frame. Inserts that were in the
413 reverse orientation and out of frame still led to knockdown. Both the directionality of the
414 gene of interest, and the frame, are therefore dispensable for knockdown. Since the RNAi
415 pathway in *E. histolytica* is unusual and mechanistic details are still coming into focus, our
416 results add to understanding of this pathway. The lack of orientation and frame
417 requirements for knockdown is useful, since gDNA fragments are random and thus are
418 ligated into the library plasmid in either orientation and in any frame. Previous studies had
419 defined that a gene fragment ≥ 500 bp is sufficient for knockdown [26, 41]. We found that
420 fragments as small as 322 bp were sufficient to result in slow growth phenotypes,
421 suggesting that efficient knockdown is possible with smaller fragments.
422 *E. histolytica* is dramatically understudied relative to other parasites, even though it
423 heavily impacts human health, causing significant morbidity and mortality. The RNAi library
424 that we developed dramatically improves the tractability of *E. histolytica.* In other parasites,
425 RNAi screens have uncovered genes relevant to molecular and cellular biology, virulence
426 and therapeutics. We look forward to the new information about *E. histolytica* that can be
427 gleaned by applying RNAi screens in the future. Since the design of this RNAi library allows
428 for quantitative identification of gene fragments that are enriched or depleted after
429 selection, we expect that this library can be used to identify genes involved in a wide variety

430 of processes, including essential genes that are relevant to the development of new

431 therapies.

432 **Materials and Methods**

433 *Cell culture*

434 *E. histolytica* HM1:IMSS (ATCC 30459) trophozoites were cultured in TYI-S-33 media,
435 supplemented with 15% heat inactivated Adult Bovine Serum (Gemini Bio Products), 80
436 units/ml penicillin and streptomycin (Gibco), and 2.3% Diamond Vitamin solution 80 Tween 40x
437 (Sigma Aldrich), at 35°C. Amoebae were harvested when flasks were approximately 80%
438 confluent.

439 *gDNA Extraction*

440 *E. histolytica* gDNA was harvested using the Quick-DNA miniprep kit (Zymo), cetyl
441 trimethylammonium bromide (CTAB) extraction, or the Wizard gDNA extraction kit (Promega).
442 The Quick-DNA miniprep kit (Zymo) was used according to manufacturer's instructions for cell
443 suspension samples. For CTAB extraction, cells were washed and then lysed at 55°C for 1 hour
444 in a solution of 0.1 M EDTA with 0.25% SDS and 0.1 mg/ml proteinase K. NaCl was added to a
445 final concentration of 0.7 M, pre-warmed CTAB was added to a final concentration of 1%, and
446 samples were incubated for 20 minutes at 65°C. An equal volume of chloroform was added,
447 and samples were centrifuged at 20,600 x *g* for 10 minutes. Supernatants were transferred into
448 fresh tubes, an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added, and
449 samples were centrifuged at 20,600 x *g* for 10 minutes. Supernatants were transferred into
450 fresh tubes, DNA was precipitated with 2.5 volumes of 100% ethanol at room temperature for 5
451 minutes, and samples were centrifuged at 20,600 x *g* for 10 minutes. DNA pellets were washed
452 in 70% ethanol. DNA was left to dry at room temperature and dissolved in water. For DNA
453 harvested using the Quick-DNA miniprep kit or CTAB extraction, RNAseIF (NEB) was used to

454 remove RNA contamination, according to manufacturer's instructions. The Wizard Genomic
455 DNA Purification kit (Promega) was used according to the manufacturer's instructions for the
456 tissue culture cells protocol. Because DNA harvested using the Wizard Genomic DNA
457 Purification kit was of higher quality, gDNA for plasmid library construction was prepared using
458 this kit.

459 *Plasmids and cloning*

460 The trigger plasmid [26] contains 132 base pairs of the trigger gene (EHI_048600) in the
461 pKT3M expression plasmid backbone [49]. Q5 Site Directed Mutagenesis (NEB) was used to
462 insert a SmaI restriction enzyme site after the trigger for cloning purposes, to create "pTrigger".
463 To create knockdown plasmids to target individual genes, pTrigger was linearized using SmaI
464 (NEB) and dephosphorylated using rSAP (NEB). A fragment of the gene of interest was amplified
465 from gDNA using PCR primers that contained homology to pTrigger, and the PCR product was
466 then cloned into the backbone using the Gibson Assembly Ultra kit (Synthetic Genomics Inc.),
467 following the manufacturer's instructions. See Table S4 for the sequences of primers that were
468 used during plasmid construction.

469 To test if the Illumina adaptor sequences interfere with knockdown, pTrigger plasmids
470 were created with and without adaptors. First, a pTrigger plasmid targeting a BAR/SH3 domain-
471 containing protein was created by inserting 812 bp of EHI_091500 (147 – 958). This plasmid was
472 further modified using Q5 Site Directed Mutagenesis (NEB) to add the custom adaptors
473 ACACTCTTTCCCTACACGACGCTCTTCCGATCT and GATCGGAAGAGCACACGTCTGAACTCCAGTCAC,
474 which are compatible with Illumina TruSeq sequencing primers, upstream and downstream of
475 the insert, respectively. After determining that the presence of the Illumina adaptors did not

476 interfere with knockdown, this construct was modified to remove the BAR/SH3 domain-
477 containing protein and replace it with a SmaI restriction site for further cloning purposes, to
478 create pTriggerAdaptor. pTriggerAdaptor was used as the backbone for plasmid library
479 generation.

480 Superoxide dismutase knockdown mutants were generated by inserting 597 bp of
481 EHI_159160 (1 – 597) into pTrigger. Independent knockdown mutants (SG1i) of SG1 were
482 generated by inserting 525 bp of EHI_194310 (32 – 556) into pTrigger. Independent knockdown
483 mutants (SG2i) of SG2 were generated by inserting 487 bp of EHI_007010 (59 – 545) into
484 pTrigger.

485 To create the N terminal FLAG-tag constructs, Q5 Site Directed Mutagenesis (NEB) was
486 used to insert a 3X FLAG epitope tag sequence preceding an XhoI restriction site in the pEhEx
487 expression plasmid backbone [49], to create pEhEx-FLAG. For localization of SG1 (EHI_194310)
488 and SG2 (EHI_007010), each individual full-length gene was cloned into pEhEx-FLAG, to create
489 pEhEx-FLAG-SG1 and pEhEx-FLAG-SG2, respectively. For expression of truncated SG1, nt 1010-
490 1527 were omitted, and the resulting truncated gene was cloned into pEhEx-FLAG, with a stop
491 codon added at the 3' end of the truncated sequence.

492 *Plasmid library construction*

493 *E. histolytica* gDNA was harvested using the Wizard Genomic DNA Purification kit
494 (Promega), following the manufacturer's instructions for the tissue culture cells protocol.
495 RNAse treatment is included in this protocol. Quality and size of gDNA was verified by Qubit
496 and gel electrophoresis. gDNA was further purified and concentrated using TailorMag
497 Purification beads (SeqMatic), and then sheared using a Covaris M220 sonicator. For size

498 selection, sheared gDNA was loaded into three lanes on a 6% TBE gel. The gel was stained using
499 SYBR Safe DNA Gel Stain (Invitrogen), and three bands, covering from ~400 – ~600 bp (for the
500 pilot library) and ~500 – ~900 bp (for the final library) were extracted from the gel using the
501 TailorCut Gel Cutter and Gel Breaker system (SeqMatic). gDNA was purified using TailorMag
502 Purification beads. The sizing and concentration of the purified material was determined by
503 TapeStation (Agilent). Purified gDNA corresponding to the second and third gel slices from each
504 lane was combined. gDNA was further purified and concentrated using TailorMag Purification
505 beads, and then processed using a TruSeq XT library prep kit (Illumina). All steps followed the
506 manufacturer's protocol except: 1. Ligation time was increased to 180 minutes, 2. Custom
507 adaptors were used instead of Illumina TruSeq adaptors. The adaptors that were ligated had
508 the following sequences: Adaptor 1: ACACTCTTTCCCTACACGACGCTCTTCCGATCCCC*T and
509 Adaptor 2: /5'Phos/GGGGATCGGGAGAGCACACGTCTGAACTCTAGTCAC. After adaptor ligation,
510 the gDNA fragments were amplified using 5 rounds of PCR with KAPA HiFi 2X master mix (KAPA
511 Biosystems). See Table S4 for primer sequences. After PCR, fragmented gDNA was purified and
512 concentrated using TailorMag Purification beads.
513 pTriggerAdaptor was linearized using SmaI (NEB) and dephosphorylated using rSAP
514 (NEB). Linearization was verified using gel electrophoresis. 25 ng of fragmented gDNA with
515 ligated adaptors was combined with 25 ng of linearized pTriggerAdaptor. The Gibson Assembly
516 Ultra kit (Synthetic Genomics Inc.) was used according to the manufacturer's instructions. High
517 Efficiency 5-alpha Competent *E. coli* (NEB) were transformed with 3 µl of the Gibson mix and
518 plated on four LB plates containing ampicillin (Sigma). We determined that the Gibson cloning
519 efficiency with pTriggerAdaptor was ~1/250, which was used to calculate the amount of insert

520 needed to generate a library at a given target level of coverage. With an average insert size of
521 ~550 bp, a minimum of ~36,354 plasmids would be needed to fully cover the genome at 1X
522 coverage. For the pilot library, one *E. coli* transformation reaction was performed and 4 plates
523 of bacteria were harvested into 200 ml of LB liquid broth containing ampicillin, and incubated at
524 37°C for 12-18 hours, shaking. For the final library, Gibson cloning, *E. coli* transformation, and
525 plasmid preparations were performed over four separate days to create four batches of final
526 library plasmids. The final plasmid library represents the combined total of the four batches.
527 For the final library, 25 *E. coli* transformation reactions were performed and 100 plates of
528 bacteria were harvested into LB liquid broth containing ampicillin, at a concentration of ~1,000
529 colonies per 200 ml. Plasmids were harvested using the GenElute HP plasmid maxiprep kit
530 (Sigma), using 200 ml of culture per plasmid purification column, and ethanol precipitated
531 according to the manufacturer's instructions.

532 *Illumina sequencing of gDNA fragments and plasmid samples*

533 Fragmented gDNA (with ligated custom adaptors; see 'plasmid library construction'), or
534 plasmid inserts were amplified with KAPA PCR (Kapa Biosystems) using full length TruSeq
535 Illumina primers, as follows: Primer 1:
536 GATCGGAAGAGCACACGTCTGAACTCCAGTCAC[i7]ATCTCGTATGCCGTCTTCTGCTTG and Primer 2:
537 AATGATACGGCGACCACCGAGATCTACAC[i5]ACACTCTTTCCCTACACGACGCTCTTCCGATCT. Three
538 technical replicates were sequenced for each sample. Samples were sequenced on an Illumina
539 MiSeq at either the DNA Technologies and Expression Analysis Core (University of California,
540 Davis), or on an Illumina MiSeq at SeqMatic, LLC. Illumina sequencing data analysis was
541 performed at the Bioinformatics Core (University of California, Davis).

542 *Illumina sequencing analysis*

543 Illumina sequencing data was demultiplexed using barcode sequences added during
544 library preparation with Illumina demultiplexing software. All reads were sequenced as PE
545 151bp runs. The resulting paired reads were then cleaned using HTStream
546 (https://github.com/ibest/HTStream) to 1) remove reads containing PhiX Control Library
547 sequences, 2) trim Illumina sequencing adaptors, and 3) discard primer dimers and other reads
548 where adaptor trimming was unsuccessful.

549 Cleaned reads were mapped against the *E. histolytica* release 46 genome assembly
550 (downloaded from https://amoebadb.org/common/downloads/release-
551 46/EhistolyticaHM1IMSS/fasta/data/AmoebaDB-46_EhistolyticaHM1IMSS_Genome.fasta) using
552 bwa mem v2.0pre1 [50]. Output from bwa was passed to samtools v1.10 [51] for sorting and
553 indexing, and also to extract reads from for visualization of specific loci. Results from mapping
554 and subsequent analysis were visualized and reviewed in Geneious Prime 2020.0.3
555 (https://www.geneious.com). Identifying unique fragments and counting their frequency was
556 done using a custom Python script. Briefly, the script iterates over read-sorted BAM records.
557 Read pairs that were mapped to the same contig with proper orientation, a minimum of 146
558 matched bases, and no more than four bases soft clipped from the 5' end of either read pair
559 were used to identify fragments. Read mapping orientation was used to infer fragment
560 orientation. A counter for each unique fragment was incremented for additional reads
561 supporting the fragment. Reads that were mapped in proper orientation but that did not meet
562 other criteria were compared against the fragment intervals with an incremented count

563 assigned to the highest scoring overlap. As a last step, the script generates summary statistics,
564 an interval table with counts for all samples, and SAF and gff annotation files for each sample.
565 Output from the Python script was then loaded into R v3.6.1 [52] for further analysis
566 and to generate figures. Briefly, data.table [53] was used to load the full set of intervals, order
567 them by genomic position, and to compute summaries by sample group. Tidyr [54] was used for
568 reformatting data, and most plots were generated with ggplot2 [55]. The vegan package [56]
569 was used for calculating saturation curve values. The GenomicRanges package [57] was used in
570 combination with gene annotation (downloaded from
571 https://amoebadb.org/common/downloads/release-
572 46/EhistolyticaHM1IMSS/gff/data/AmoebaDB-46_EhistolyticaHM1IMSS.gff) to determine and
573 count fragments targeting genes, all reported counts used a minimum overlap of 27bp.
574 Genome coverage plots were generated with karyoploteR [58] and heatmaps were plotted
575 using pheatmap [59]. Scripts and more details about the analysis are available at the GitHub
576 repository associated with this publication: https://github.com/samhunter/Bettadapur-et-al-
577 Entamoeba-histolytica-RNAi-library.
578 All sequencing data from this project have been deposited in the Sequence Read Archive
579 (SRA) database (https://www.ncbi.nlm.nih.gov/sra), under project number SUB7459667.
580 *Transfection and clonal lines*
581 Library plasmids were transfected into amoebae using Attractene transfection reagent
582 (Qiagen) [60, 61]. 50 ml M199 (Gibco medium M199 with Earle's salts, L-glutamine, and 2.2
583 g/liter sodium bicarbonate, without phenol red), was supplemented with 0.045 g L-Cysteine,
584 0.3 g HEPES, 0.005 g ascorbic acid (all from Sigma) and the pH was adjusted to 6.8. To 42.5 ml of

585 this medium, 7.5 ml of heat inactivated Adult Bovine Serum (Gemini bio-products) was added
586 to make transfection medium. Amoebae were resuspended in transfection medium at a
587 concentration of $2.5x10^5$ cells/ml, and then 1.8 ml of amoebae, 10 µg plasmid DNA (in 200 µl of
588 transfection medium), and 37.5 µl Attractene reagent (Qiagen) were combined into a 2.0 ml
589 screw-cap cryotube. Cryotubes were incubated horizontally at 35°C for 3 hours. Two cryotubes
590 per transfection were then added to a T25 flask containing 66 ml of fresh TYI media. 24 hours
591 later, 6 µg/ml geneticin (Invitrogen) was added. Stable transfectants were selected and
592 maintained with geneticin at 6 µg/ml.

593 The efficiency of stable transfection was determined by performing limiting dilution.
594 Three independent experiments were performed. Amoebae were transfected using the process
595 described above, except that 24 hours after transfection, amoebae were harvested and
596 resuspended in TYI at a concentration of $2x10^6$ cells/ml. Amoebae were then serially diluted
597 into 96-well plates, in the presence or absence of 6 µg/ml geneticin. Plates were incubated in
598 GasPak EZ pouch (BD Biosciences) and growth was monitored daily. The plates without
599 Geneticin were used to calculate the plating efficiency. In the plates without geneticin, growth
600 was seen in wells corresponding to approximately 2.5 cells on average. In the plates containing
601 geneticin, growth was seen in wells containing 781 to 3125 amoebae. Taken together, the
602 corrected stable transfection efficiency was ~1/1000 amoebae.

603 For generation of clonal mutant lines for phenotypic characterization, limiting dilution
604 was performed in 96-well plates in a GasPak EZ pouch (BD Biosciences). For identification of
605 mutants with growth defects, clonal lines that visually appeared to have slow growth when
606 examined using microscopy were selected for further analysis. RNA was extracted using the

607 Direct-zol RNA kit (Zymo). gDNA (which contains exogeneous plasmids) was extracted using the
608 Quick-DNA miniprep kit (Zymo). Sanger sequencing was used to identify the library insert
609 associated with each clonal line. Sequencing was carried out by Genewiz. To identify the gene
610 targeted for knockdown in each clonal line, the insert sequence was used to search the
611 reference genome. Genes with at least 50nt of continuous sequence identity with the insert
612 were considered hits. See Table S4 for sequencing primer information.

613 *Growth analysis*

614 For growth analysis of RNAi knockdown mutants, clonal stable transfectants (SG1 –
615 SG12) or heterogeneous stable transfectants (SG1i and SG2i) were resuspended to $8x10^3$
616 cells/ml and 1 ml of cells was added to three tubes containing 13 ml fresh TYI with geneticin.
617 Cells were incubated at 35°C. At days 2, 4, and 6, one tube of cells was harvested per day and
618 counted using a TC20 Automatic cell counter (BioRad). Vector control transfected amoebae
619 were used as a negative control, and superoxide dismutase heterogeneous knockdown mutants
620 were used as a positive control for slow growth. Two technical replicates were performed per
621 experiment and three independent growth curve experiments were performed for each
622 mutant.

623 For growth analysis of cells overexpressing full-length or truncated SG proteins,
624 geneticin concentration was doubled daily until it reached 48 µg/ml, in order to increase
625 plasmid copy number and lead to protein overexpression. Geneticin concentration was also
626 increased in the same manner in vector control transfectants. Heterogeneous stable
627 transfectants were resuspended to $5x10^4$ cells/ml and 1 ml of cells was added to five tubes
628 containing 13 ml fresh TYI with 48 µg/ml geneticin. Cells were incubated at 35°C. At days 1, 2, 3,

629 4 and 5, one tube of cells was harvested per day and counted using a TC20 Automatic cell
630 counter (BioRad). Vector control transfected amoebae were used as a negative control. Four
631 technical replicates were performed per experiment and three independent growth curve
632 experiments were performed for each cell line. Growth curves were also performed at 24 µg/ml
633 geneticin, which led to the same growth phenotypes seen at higher concentrations of geneticin.
634 *RT-PCR*
635 RT-PCR was used to assess knockdown in transfectants. RNA was extracted using the
636 Direct-zol RNA kit (Zymo). DNAseI (Turbo) treatment and SuperScript Reverse Transcription II
637 (Invitrogen) was performed as per manufacturer instructions to obtain cDNA. RT-PCR was
638 performed on cDNA harvested from transfectants containing the following constructs: BAR/SH3
639 domain-containing protein with and without Illumina adaptors, Superoxide Dismutase;
640 SG1/SG1 independent knock down; and SG2/SG2 independent knockdown. GAPDH
641 (EHI_187020) RT-PCR was used as a positive control. See Table S4 for RT-PCR primer sequences.
642 *Immunofluorescence assays*
643 For immunofluorescence studies, geneticin concentration was doubled daily until it
644 reached 24 µg/ml, in order to increase plasmid copy number and lead to protein
645 overexpression. Amoebae were washed in M199s and fixed with 4% paraformaldehyde in PBS
646 for 30 minutes at room temperature. Samples were permeabilized with 0.2% Triton X-100, or
647 were not permeabilized. For permeabilized samples, for SG1, all subsequent blocking, antibody
648 incubation and wash steps included 0.1% Triton X-100. For permeabilized samples, for SG2, all
649 subsequent blocking, antibody incubation and wash steps included 0.1% Tween 20. For non-
650 permeabilized samples, all subsequent blocking, antibody incubation and wash steps included

651 0.1% Tween 20. Samples were blocked with goat serum (Jackson ImmunoResearch
652 Laboratories, Inc.) and bovine serum albumin (Gemini). Antibody labeling was carried out in
653 solution. For SG1, a rabbit anti-FLAG primary antibody (Abcam Rabbit Anti-DDDDK tag, binds to
654 FLAG tag sequence, antibody ab1162) was used, followed by an alpaca anti-rabbit Alexa 647
655 secondary antibody (Jackson ImmunoResearch Laboratories, Inc.). For SG2, a mouse anti-FLAG
656 primary (Sigma Mouse Anti-FLAG M2 monoclonal antibody) was used, followed by an alpaca
657 anti-mouse Cy5 antibody (Jackson ImmunoResearch Laboratories). Samples were stained with
658 DAPI at 40 µg/ml. Data were collected using an ImageStreamX Mark II flow cytometer,
659 equipped with a 40X objective. 10,000 events were collected for each sample and data were
660 analyzed using Amnis IDEAS software (Fig. S7).

661 For detection of the endogenous Gal/GalNAc lectin, amoebae transfected with vector
662 control (pEhEx) were washed and resuspended in M199s medium and labeled with CellTracker
663 green 5-chloromethylfluorescein diacetate (CMFDA; Invitrogen) at 186 ng/ml for 10 minutes at
664 35°C. Samples were fixed and processed as above, without permeabilization. The *E. histolytica*
665 Gal/GalNAc lectin antibody (Clone 7F4; a gift from William A. Petri, Jr., University of Virginia)
666 was used, followed by a Cy5 AffiniPure Goat Anti-Mouse secondary antibody (Jackson
667 ImmunoResearch Laboratories, Inc.). Amoebae were imaged and analyzed as above.

668 *Statistical analysis*

669 GraphPad Prism was used for statistical analyses. Raw growth analysis data were
670 analyzed using a two-way ANOVA with a post-hoc Dunnett's multiple comparisons test;
671 multiplicity-adjusted *p*-values are shown. Normalized growth analysis data were analyzed using
672 student's unpaired *t* test statistical analysis and normalization (baseline removal). Mean values

673 and standard deviations are shown in the figures, with *p*-values reported as follows: ns = $p >$

674 0.05, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

675 **Acknowledgements**

676 We thank Scott Dawson and Su-Ju Lin for helpful discussions. We thank Lutz
677 Froenicke for helpful advice on designing the library to enable Illumina sequencing. DNA
678 sequencing was carried out by the DNA Technologies and Expression Analysis Cores at the
679 UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-
680 01. A.B. was supported by a fellowship from the UC Davis Training Program in Biomolecular
681 Technology. M.C.R. was supported by the NIH T32 Comparative Medical Science Training
682 Program, T32OD011147. T.S.Y.T. was supported by the Khaira Family Experiential Learning
683 Award through the UC Davis College of Biological Sciences. This work was funded by NIH
684 grants 1R01AI146914, 5R21AI154163 and a Pew Scholarship awarded to K.S.R.

Competing Interests

The authors declare that they have no competing interests.

687 **References**

688 1. WHO/PAHO/UNESCO report. A consultation with experts on amoebiasis. Mexico City,
689 Mexico 28-29 January, 1997. Epidemiol Bull. 1997;18(1):13-4. Epub 1997/03/01. PubMed PMID:
690 9197085.
691 2. Gilchrist CA, Petri SE, Schneider BN, Reichman DJ, Jiang N, Begum S, et al. Role of the
692 Gut Microbiota of Children in Diarrhea Due to the Protozoan Parasite Entamoeba histolytica. J
693 Infect Dis. 2016;213(10):1579-85. Epub 2015/12/30. doi: 10.1093/infdis/jiv772. PubMed PMID:
694 26712950; PubMed Central PMCID: PMCPMC4837909.
695 3. Mondal D, Petri WA, Jr., Sack RB, Kirkpatrick BD, Haque R. Entamoeba histolytica-
696 associated diarrheal illness is negatively associated with the growth of preschool children:
697 evidence from a prospective study. Trans R Soc Trop Med Hyg. 2006;100(11):1032-8. Epub
698 2006/05/30. doi: S0035-9203(06)00066-6 [pii]
699 10.1016/j.trstmh.2005.12.012 [doi]. PubMed PMID: 16730764.
700 4. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al. Burden
701 and aetiology of diarrhoeal disease in infants and young children in developing countries (the
702 Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet.
703 2013;382(9888):209-22. doi: 10.1016/S0140-6736(13)60844-2. PubMed PMID: 23680352.
704 5. Mortality GBD, Causes of Death C. Global, regional, and national age-sex specific all-
705 cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for
706 the Global Burden of Disease Study 2013. Lancet. 2015;385(9963):117-71. Epub 2014/12/23.
707 doi: 10.1016/S0140-6736(14)61682-2. PubMed PMID: 25530442; PubMed Central PMCID:
708 PMCPMC4340604.
709 6. Gonzales MLM, Dans LF, Sio-Aguilar J. Antiamoebic drugs for treating amoebic colitis.
710 Cochrane Database Syst Rev. 2019;1:CD006085. Epub 2019/01/10. doi:
711 10.1002/14651858.CD006085.pub3. PubMed PMID: 30624763; PubMed Central PMCID:
712 PMCPMC6326239.
713 7. Willhoeft U, Tannich E. The electrophoretic karyotype of Entamoeba histolytica. Mol
714 Biochem Parasitol. 1999;99(1):41-53. Epub 1999/04/24. doi: 10.1016/s0166-6851(98)00178-9.
715 PubMed PMID: 10215023.
716 8. Lorenzi HA, Puiu D, Miller JR, Brinkac LM, Amedeo P, Hall N, et al. New assembly,
717 reannotation and analysis of the Entamoeba histolytica genome reveal new genomic features
718 and protein content information. PLoS Negl Trop Dis. 2010;4(6):e716. Epub 2010/06/19. doi:
719 10.1371/journal.pntd.0000716. PubMed PMID: 20559563; PubMed Central PMCID:
720 PMCPMC2886108.
721 9. Wilson IW, Weedall GD, Lorenzi H, Howcroft T, Hon CC, Deloger M, et al. Genetic
722 Diversity and Gene Family Expansions in Members of the Genus Entamoeba. Genome Biol Evol.
723 2019;11(3):688-705. Epub 2019/01/23. doi: 10.1093/gbe/evz009. PubMed PMID: 30668670;
724 PubMed Central PMCID: PMCPMC6414313.
725 10. Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, et al. The genome
726 of the protist parasite Entamoeba histolytica. Nature. 2005;433(7028):865-8. Epub 2005/02/25.
727 doi: nature03291 [pii]
728 10.1038/nature03291 [doi]. PubMed PMID: 15729342.

729 11. King AV, Welter BH, Koushik AB, Gordon LN, Temesvari LA. A genome-wide over-
730 expression screen identifies genes involved in phagocytosis in the human protozoan parasite,
731 Entamoeba histolytica. PLoS One. 2012;7(8):e43025. doi: 10.1371/journal.pone.0043025.
732 PubMed PMID: 22905196; PubMed Central PMCID: PMC3419234.
733 12. Koushik AB, Welter BH, Rock ML, Temesvari LA. A genomewide overexpression screen
734 identifies genes involved in the phosphatidylinositol 3-kinase pathway in the human protozoan
735 parasite Entamoeba histolytica. Eukaryot Cell. 2014;13(3):401-11. doi: 10.1128/EC.00329-13.
736 PubMed PMID: 24442890; PubMed Central PMCID: PMC3957588.
737 13. Kangussu-Marcolino MM, Morgado P, Manna D, Yee H, Singh U. Development of a
738 CRISPR/Cas9 system in Entamoeba histolytica: proof of concept. Int J Parasitol. 2021;51(2-
739 3):193-200. Epub 2020/12/03. doi: 10.1016/j.ijpara.2020.09.005. PubMed PMID: 33264648;
740 PubMed Central PMCID: PMCPMC7880892.
741 14. Bracha R, Nuchamowitz Y, Leippe M, Mirelman D. Antisense inhibition of amoebapore
742 expression in Entamoeba histolytica causes a decrease in amoebic virulence. Mol Microbiol.
743 1999;34(3):463-72. Epub 1999/11/17. doi: mmi1607 [pii]. PubMed PMID: 10564488.
744 15. Bracha R, Nuchamowitz Y, Anbar M, Mirelman D. Transcriptional silencing of multiple
745 genes in trophozoites of Entamoeba histolytica. PLoS Pathog. 2006;2(5):e48. Epub 2006/05/31.
746 doi: 10.1371/journal.ppat.0020048 [doi]. PubMed PMID: 16733544.
747 16. Bracha R, Nuchamowitz Y, Mirelman D. Transcriptional silencing of an amoebapore gene
748 in Entamoeba histolytica: molecular analysis and effect on pathogenicity. Eukaryot Cell.
749 2003;2(2):295-305. Epub 2003/04/10. PubMed PMID: 12684379; PubMed Central PMCID:
750 PMC154849.
751 17. Stock RP, Olvera A, Sanchez R, Saralegui A, Scarfi S, Sanchez-Lopez R, et al. Inhibition of
752 gene expression in Entamoeba histolytica with antisense peptide nucleic acid oligomers. Nat
753 Biotechnol. 2001;19(3):231-4. Epub 2001/03/07. doi: 10.1038/85671. PubMed PMID:
754 11231555.
755 18. Kaur G, Lohia A. Inhibition of gene expression with double strand RNA interference in
756 Entamoeba histolytica. Biochemical and biophysical research communications.
757 2004;320(4):1118-22. Epub 2004/07/14. doi: 10.1016/j.bbrc.2004.06.064. PubMed PMID:
758 15249205.
759 19. Solis CF, Santi-Rocca J, Perdomo D, Weber C, Guillen N. Use of bacterially expressed
760 dsRNA to downregulate Entamoeba histolytica gene expression. PLoS One. 2009;4(12):e8424.
761 Epub 2009/12/29. doi: 10.1371/journal.pone.0008424. PubMed PMID: 20037645; PubMed
762 Central PMCID: PMC2793006.
763 20. Linford AS, Moreno H, Good KR, Zhang H, Singh U, Petri WA, Jr. Short hairpin RNA-
764 mediated knockdown of protein expression in Entamoeba histolytica. BMC Microbiol.
765 2009;9(1):38. Epub 2009/02/19. doi: 1471-2180-9-38 [pii]
766 10.1186/1471-2180-9-38 [doi]. PubMed PMID: 19222852.
767 21. Zhang H, Ehrenkaufer GM, Pompey JM, Hackney JA, Singh U. Small RNAs with 5'-
768 polyphosphate termini associate with a Piwi-related protein and regulate gene expression in
769 the single-celled eukaryote Entamoeba histolytica. PLoS Pathog. 2008;4(11):e1000219. Epub
770 2008/12/02. doi: 10.1371/journal.ppat.1000219. PubMed PMID: 19043551; PubMed Central
771 PMCID: PMC2582682.

772 22. Sijen T, Steiner FA, Thijssen KL, Plasterk RH. Secondary siRNAs result from unprimed
773 RNA synthesis and form a distinct class. Science. 2007;315(5809):244-7. Epub 2006/12/13. doi:
774 10.1126/science.1136699. PubMed PMID: 17158288.
775 23. Zhang H, Alramini H, Tran V, Singh U. Nucleus-localized antisense small RNAs with 5'-
776 polyphosphate termini regulate long term transcriptional gene silencing in Entamoeba
777 histolytica G3 strain. J Biol Chem. 2011;286(52):44467-79. Epub 2011/11/04. doi: M111.278184
778 [pii]
779 10.1074/jbc.M111.278184. PubMed PMID: 22049083; PubMed Central PMCID: PMC3247957.
780 24. Pompey JM, Foda B, Singh U. A Single RNaseIII Domain Protein from Entamoeba
781 histolytica Has dsRNA Cleavage Activity and Can Help Mediate RNAi Gene Silencing in a
782 Heterologous System. PLoS One. 2015;10(7):e0133740. doi: 10.1371/journal.pone.0133740.
783 PubMed PMID: 26230096; PubMed Central PMCID: PMC4521922.
784 25. Zhang H, Tran V, Manna D, Ehrenkaufer G, Singh U. Functional Characterization of
785 Entamoeba histolytica Argonaute Proteins Reveals a Repetitive DR-Rich Motif Region That
786 Controls Nuclear Localization. mSphere. 2019;4(5). Epub 2019/10/18. doi:
787 10.1128/mSphere.00580-19. PubMed PMID: 31619501; PubMed Central PMCID:
788 PMCPMC6796981.
789 26. Morf L, Pearson RJ, Wang AS, Singh U. Robust gene silencing mediated by antisense
790 small RNAs in the pathogenic protist Entamoeba histolytica. Nucleic Acids Res. 2013. doi:
791 10.1093/nar/gkt717. PubMed PMID: 23935116.
792 27. Foda BM, Singh U. Dimethylated H3K27 Is a Repressive Epigenetic Histone Mark in the
793 Protist Entamoeba histolytica and Is Significantly Enriched in Genes Silenced via the RNAi
794 Pathway. J Biol Chem. 2015;290(34):21114-30. doi: 10.1074/jbc.M115.647263. PubMed PMID:
795 26149683; PubMed Central PMCID: PMC4543668.
796 28. Pompey JM, Morf L, Singh U. RNAi pathway genes are resistant to small RNA mediated
797 gene silencing in the protozoan parasite Entamoeba histolytica. PLoS One. 2014;9(9):e106477.
798 Epub 2014/09/10. doi: 10.1371/journal.pone.0106477. PubMed PMID: 25198343; PubMed
799 Central PMCID: PMCPMC4157801.
800 29. Manna D, Lentz CS, Ehrenkaufer GM, Suresh S, Bhat A, Singh U. An NAD(+)-dependent
801 novel transcription factor controls stage conversion in Entamoeba. eLife. 2018;7. Epub
802 2018/10/31. doi: 10.7554/eLife.37912. PubMed PMID: 30375973; PubMed Central PMCID:
803 PMCPMC6207428.
804 30. Suresh S, Ehrenkaufer G, Zhang H, Singh U. Development of RNA Interference Trigger-
805 Mediated Gene Silencing in Entamoeba invadens. Infect Immun. 2016;84(4):964-75. Epub
806 2016/01/21. doi: 10.1128/IAI.01161-15. PubMed PMID: 26787723; PubMed Central PMCID:
807 PMCPMC4807475.
808 31. Morris JC, Wang Z, Drew ME, Englund PT. Glycolysis modulates trypanosome
809 glycoprotein expression as revealed by an RNAi library. EMBO J. 2002;21(17):4429-38. Epub
810 2002/08/29. PubMed PMID: 12198145; PubMed Central PMCID: PMC125414.
811 32. Drew ME, Morris JC, Wang Z, Wells L, Sanchez M, Landfear SM, et al. The adenosine
812 analog tubercidin inhibits glycolysis in Trypanosoma brucei as revealed by an RNA interference
813 library. J Biol Chem. 2003;278(47):46596-600. Epub 2003/09/16. doi: 10.1074/jbc.M309320200.
814 PubMed PMID: 12972414.

815 33. Stortz JA, Serafim TD, Alsford S, Wilkes J, Fernandez-Cortes F, Hamilton G, et al.
816 Genome-wide and protein kinase-focused RNAi screens reveal conserved and novel damage
817 response pathways in Trypanosoma brucei. PLoS Pathog. 2017;13(7):e1006477. Epub
818 2017/07/26. doi: 10.1371/journal.ppat.1006477. PubMed PMID: 28742144; PubMed Central
819 PMCID: PMCPMC5542689.
820 34. Glover L, Hutchinson S, Alsford S, Horn D. VEX1 controls the allelic exclusion required for
821 antigenic variation in trypanosomes. Proc Natl Acad Sci U S A. 2016;113(26):7225-30. Epub
822 2016/05/27. doi: 10.1073/pnas.1600344113. PubMed PMID: 27226299; PubMed Central
823 PMCID: PMCPMC4932947.
824 35. Mony BM, MacGregor P, Ivens A, Rojas F, Cowton A, Young J, et al. Genome-wide
825 dissection of the quorum sensing signalling pathway in Trypanosoma brucei. Nature.
826 2014;505(7485):681-5. Epub 2013/12/18. doi: 10.1038/nature12864. PubMed PMID:
827 24336212; PubMed Central PMCID: PMCPMC3908871.
828 36. Rico E, Ivens A, Glover L, Horn D, Matthews KR. Genome-wide RNAi selection identifies a
829 regulator of transmission stage-enriched gene families and cell-type differentiation in
830 Trypanosoma brucei. PLoS Pathog. 2017;13(3):e1006279. Epub 2017/03/24. doi:
831 10.1371/journal.ppat.1006279. PubMed PMID: 28334017; PubMed Central PMCID:
832 PMCPMC5380359.
833 37. Alsford S, Eckert S, Baker N, Glover L, Sanchez-Flores A, Leung KF, et al. High-throughput
834 decoding of antitrypanosomal drug efficacy and resistance. Nature. 2012;482(7384):232-6.
835 Epub 2012/01/27. doi: 10.1038/nature10771. PubMed PMID: 22278056; PubMed Central
836 PMCID: PMCPMC3303116.
837 38. Collett CF, Kitson C, Baker N, Steele-Stallard HB, Santrot MV, Hutchinson S, et al.
838 Chemogenomic Profiling of Antileishmanial Efficacy and Resistance in the Related Kinetoplastid
839 Parasite Trypanosoma brucei. Antimicrob Agents Chemother. 2019;63(8). Epub 2019/06/05.
840 doi: 10.1128/AAC.00795-19. PubMed PMID: 31160283; PubMed Central PMCID:
841 PMCPMC6658743.
842 39. Trenaman A, Glover L, Hutchinson S, Horn D. A post-transcriptional respiratome regulon
843 in trypanosomes. Nucleic Acids Res. 2019;47(13):7063-77. Epub 2019/05/28. doi:
844 10.1093/nar/gkz455. PubMed PMID: 31127277; PubMed Central PMCID: PMCPMC6648352.
845 40. Zhang N, Zoltner M, Leung KF, Scullion P, Hutchinson S, Del Pino RC, et al. Host-parasite
846 co-metabolic activation of antitrypanosomal aminomethyl-benzoxaboroles. PLoS Pathog.
847 2018;14(2):e1006850. Epub 2018/02/10. doi: 10.1371/journal.ppat.1006850. PubMed PMID:
848 29425238; PubMed Central PMCID: PMCPMC5823473.
849 41. Khalil MI, Foda BM, Suresh S, Singh U. Technical advances in trigger-induced RNA
850 interference gene silencing in the parasite Entamoeba histolytica. Int J Parasitol.
851 2016;46(3):205-12. doi: 10.1016/j.ijpara.2015.11.004. PubMed PMID: 26747561; PubMed
852 Central PMCID: PMCPMC4767557.
853 42. Rastew E, Morf L, Singh U. Entamoeba histolytica rhomboid protease 1 has a role in
854 migration and motility as validated by two independent genetic approaches. Exp Parasitol.
855 2015. doi: 10.1016/j.exppara.2015.04.004. PubMed PMID: 25889553.
856 43. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, et al. The MetaCyc
857 database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome

858 databases. Nucleic Acids Res. 2016;44(D1):D471-80. Epub 2015/11/04. doi:
859 10.1093/nar/gkv1164. PubMed PMID: 26527732; PubMed Central PMCID: PMCPMC4702838.
860 44. Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, Kale N, et al. The ChEBI reference
861 database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic
862 Acids Res. 2013;41(Database issue):D456-63. Epub 2012/11/28. doi: 10.1093/nar/gks1146.
863 PubMed PMID: 23180789; PubMed Central PMCID: PMCPMC3531142.
864 45. Shabardina V, Kischka T, Kmita H, Suzuki Y, Makalowski W. Environmental adaptation of
865 Acanthamoeba castellanii and Entamoeba histolytica at genome level as seen by comparative
866 genomic analysis. Int J Biol Sci. 2018;14(3):306-20. Epub 2018/03/22. doi: 10.7150/ijbs.23869.
867 PubMed PMID: 29559848; PubMed Central PMCID: PMCPMC5859476.
868 46. Hon CC, Weber C, Sismeiro O, Proux C, Koutero M, Deloger M, et al. Quantification of
869 stochastic noise of splicing and polyadenylation in Entamoeba histolytica. Nucleic Acids Res.
870 2013;41(3):1936-52. Epub 2012/12/22. doi: 10.1093/nar/gks1271. PubMed PMID: 23258700;
871 PubMed Central PMCID: PMCPMC3561952.
872 47. Suleiman RL, Ralston KS. Growth and genetic manipulation of Entamoeba histolytica.
873 Curr Protoc. 2021;Submitted.
874 48. Mueller DE, Petri WA, Jr. Clonal growth in Petri dishes of Entamoeba histolytica. Trans R
875 Soc Trop Med Hyg. 1995;89(1):123. Epub 1995/01/01. doi: 10.1016/0035-9203(95)90684-3.
876 PubMed PMID: 7747296.
877 49. Saito-Nakano Y, Yasuda T, Nakada-Tsukui K, Leippe M, Nozaki T. Rab5-associated
878 vacuoles play a unique role in phagocytosis of the enteric protozoan parasite Entamoeba
879 histolytica. J Biol Chem. 2004;279(47):49497-507. Epub 2004/09/07. doi:
880 10.1074/jbc.M403987200 [doi]
881 M403987200 [pii]. PubMed PMID: 15347665.
882 50. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
883 arXiv e-prints. 2013.
884 51. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
885 Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 2009/06/10.
886 doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID:
887 PMCPMC2723002.
888 52. Team RC. R: A language and environment for statistical computing. R Foundation for
889 statistical computing. 2019.
890 53. Dowle M, Srinivasan A. data.table: Extension of `data.frame`. R package version 1.12.8.
891 2019.
892 54. Wickham H, Henry L. tidyr: Tidy Messy Data. R package version 1.0.2. 2020.
893 55. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.
894 2016.
895 56. Okansen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, et al.
896 vegan: Community Ecology Package. R package version 2.5-6. 2019.
897 57. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, et al. Software for
898 computing and annotating genomic ranges. PLoS Comput Biol. 2013;9(8):e1003118. Epub
899 2013/08/21. doi: 10.1371/journal.pcbi.1003118. PubMed PMID: 23950696; PubMed Central
900 PMCID: PMCPMC3738458.

901 58. Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes
902 displaying arbitrary data. Bioinformatics. 2017;33(19):3088-90. Epub 2017/06/03. doi:
903 10.1093/bioinformatics/btx346. PubMed PMID: 28575171; PubMed Central PMCID:
904 PMCPMC5870550.
905 59. Kolde R. pheatmap: Pretty Heatmaps. R package version 1.0.12. 2019.
906 60. Ralston KS, Solga MD, Mackey-Lawrence NM, Somlata, Bhattacharya A, Petri WA, Jr.
907 Trogocytosis by Entamoeba histolytica contributes to cell killing and tissue invasion. Nature.
908 2014;508(7497):526-30. doi: 10.1038/nature13242. PubMed PMID: 24717428; PubMed Central
909 PMCID: PMC4006097.
910 61. Miller HW, Suleiman RL, Ralston KS. Trogocytosis by Entamoeba histolytica Mediates
911 Acquisition and Display of Human Cell Membrane Proteins and Evasion of Lysis by Human
912 Serum. MBio. 2019;10(2). doi: 10.1128/mBio.00068-19. PubMed PMID: 31040235; PubMed
913 Central PMCID: PMCPMC6495370.
914

915 **Figure Legends**

916 **Figure 1. RNAi library strategy.**

917 **a,** The plasmid RNAi library contains random fragments of gDNA cloned into the trigger
918 RNAi plasmid (pTriggerAdaptor). Different plasmid inserts are represented by different
919 colors in this schematic. A heterogeneous population of stably transfected RNAi knockdown
920 mutants can be exposed to selection or screening. To identify gene fragments that are
921 represented in the population after selection/screening, gDNA is extracted for analysis.
922 gDNA extracted from non-selected amoebae is used as a control. **b,** Schematic
923 representation of the plasmid RNAi library. The trigger plasmid is shown, with different
924 inserts represented by different colors. The "trigger" is a 132 bp fragment of EHI_048600.
925 pTrigger was modified to create pTriggerAdaptor, containing custom Illumina TruSeq
926 adaptor sequences (IA) upstream and downstream of the cloning site to facilitate amplicon
927 sequencing. A single round of PCR is needed to add the remaining Illumina sequences
928 (shown in dark gray) and barcodes necessary for Illumina deep sequencing analysis. **c,** An
929 812 bp fragment of a BAR/SH3 domain-containing protein (EHI_091530) was inserted into
930 pTrigger with or without custom Illumina TruSeq adaptors flanking the cloning site. The
931 BAR/SH3 domain-containing protein was in the forward orientation and in frame relative to
932 the trigger sequence. Stably transfected heterogeneous mutants were assessed for
933 knockdown of the targeted gene using RT-PCR, with (+) or without (-) reverse transcriptase.
934 GAPDH (EHI_187020) was used as a loading control.

935 **Fig. 2. RNAi library construction.**

936 **a,** Steps to create the library. *E. histolytica* gDNA was extracted and treated with RNAse.

937 Sonication, followed by gel purification, was used to isolate gDNA fragments. Fragments
938 were end-repaired, A-tailed, and ligated to Y-shaped adaptors. The adaptors contain
939 sequences corresponding to the region upstream and downstream of the cloning site in
940 pTriggerAdaptor. The Y-shaped design ensures that, after PCR to make these sequences
941 double-stranded, each gDNA fragment will be flanked by both the upstream/IA plasmid
942 sequence (represented in black), and the downstream/IA sequence (represented in gray).
943 Finally, the adaptor-ligated gDNA fragments were cloned into pTriggerAdaptor using Gibson
944 cloning. **b – c,** Bioanalyzer traces show appropriate size shift after adaptor ligation,
945 supporting that Y-shaped adaptors were successfully ligated. Example data are shown from
946 optimization experiments that were performed to establish the protocols for library
947 construction. **d,** Gibson cloning and plasmid preps were done *en masse* with pooled
948 plasmids. To check cloning efficiency, individual colonies were obtained and analyzed.
949 Shown are plasmid samples and restriction analysis from the pilot plasmid library
950 preparation. Restriction analysis was performed with restriction enzymes that flank the
951 cloning site (U, uncut; D, digested). Essentially all plasmids that were analyzed contained an
952 insert of the appropriate size range.

953 **Figure 3. The final RNAi library has genome-wide coverage.**

954 **a,** Nucleotide frequency analysis of the HM1:IMSS reference genome, the pilot and final
955 gDNA fragments, and the pilot and final plasmid libraries. Stacked bar plots show the
956 percentage of each base. The nucleotide composition of the gDNA fragments and the final
957 plasmid library is the same as the genome. **b,** Histograms showing the distribution of
958 fragments sizes for the unique fragments in the pilot and final gDNA fragments, and the

959 pilot and final plasmid libraries. Size in nucleotides is plotted on the X axis, and probability
960 distribution density is plotted on the Y axis. **c,** Genome coverage plots of the final library,
961 over the five largest contigs in the reference genome. The positions of unique fragments
962 and the depth in fragments overlapping each genomic position are shown. The positions of
963 genes are shown with black and green bars. Fragments that are in the same orientation as
964 genes that are in the forward orientation in the reference genome are shown in black, and
965 fragments that are in the same orientation as genes that are in the reverse orientation in
966 the reference genome are shown in green. **d,** Visual representation of fragment coverage in
967 the final library, over 100 kb of contig DS571145.

968 **Figure 4. Fragment directionality and coverage analysis.**

969 Heat map showing the number of unique fragments per gene in the pilot and final gDNA
970 fragments, and the pilot and final plasmid libraries. The number of unique fragments (Log_2) for
971 each gene are plotted (color value, see legend). Stranded fragments are in the same orientation
972 as genes. Unstranded fragments are the total number of fragments, both in the same
973 orientation as genes and in the opposite orientation as genes.

974 **Figure 5. Identification of slow growth mutants, with phenotypes that are reproducible in**
975 **independently-generated knockdown mutants.**

976 **a,** Growth analysis of clonal knockdown mutant lines obtained from screening the
977 heterogeneous library transfectants for growth defects. Amoebae transfected with
978 pTriggerAdaptor (vector control) were used as a negative control, and superoxide dismutase
979 (EHI_159160) knockdown mutants were used as a positive control for slow growth. The
980 library mutants SG1 and SG2 each exhibited a significant growth defect when compared to

981 vector control transfectants. **b,** RT-PCR analysis demonstrates that the genes corresponding
982 to the plasmid inserts in SG1 and SG2 (EHI_194310 and EHI_007010, respectively), were
983 knocked down. **c,** Knockdown mutants were independently generated by using fragments of
984 these genes that differed from the fragments in the library plasmids. These independent
985 mutants are referred to as SG1i and SG2i. Growth analysis was performed using
986 heterogeneous stable transfectants, and growth mutants exhibited a significant growth
987 defect when compared to vector control. **d,** RT-PCR analysis demonstrates that EHI_194310
988 and EHI_007010 were knocked down. **e,** Growth phenotypes in SG1, SG2, SG1i and SG2i.
989 Growth values for each mutant were normalized to vector control transfectants. Mutants
990 exhibited statistically significant growth defects relative to vector control transfectants.
991 **Figure 6. SG1 localizes to the cytoplasm and cell surface, and overexpression of SG1**
992 **enhances growth.**
993 Amoebae were stably transfected with a plasmid for expression of full-length FLAG-tagged
994 SG1. Immunofluorescence was used to determine the localization of SG1 in heterogeneous
995 transfectants and imaging flow cytometry was used for analysis. **a,** Localization of SG1 in
996 FLAG-SG1 transfectants (left panel) and control staining of wild-type cells (right panel).
997 Samples were permeabilized prior to antibody staining. Six random cells are shown for each
998 condition. Shown from left to right are bright field images (BF), rabbit anti-FLAG antibody
999 staining (FLAG, red), and DAPI staining (DAPI, blue). The numbers in the BF images indicate
1000 the object/image number. **b,** Localization of SG1 in FLAG-SG1 transfectants (left panel) and
1001 control staining of wild-type cells (right panel), as in panel A, except samples were not
1002 permeabilized. **c,** Histograms showing the intensity of the DAPI staining in images collected

1003 from FLAG-SG1 transfectants (left panel) or wild-type cells (right panel). The mean

1004 fluorescence intensity (MFI) is indicated on each histogram. **d,** Growth analysis of

1005 heterogeneous transfectants expressing full-length SG1 or a C-terminal SG1 truncation

1006 mutant. Amoebae transfected with pEhEx (vector control) were used as control.

1007 **Figure 7. SG2 localizes to the cytoplasm.**

1008 Amoebae were stably transfected with a plasmid for expression of full-length FLAG-tagged

1009 SG2. Immunofluorescence was used to determine the localization of SG2 in heterogeneous

1010 transfectants and imaging flow cytometry was used for analysis. **a,** Localization of SG2 in

1011 FLAG-SG2 transfectants (left panel) and control staining of wild-type cells (right panel).

1012 Samples were permeabilized prior to antibody staining. Six random cells are shown for each

1013 condition. Shown from left to right are bright field images (BF), mouse anti-FLAG antibody

1014 staining (FLAG, red), and DAPI staining (DAPI, blue). The numbers in the BF images indicate

1015 the object/image number. **b,** Localization of SG2 in FLAG-SG2 transfectants (left panel) and

1016 control staining of wild-type cells (right panel), as in panel A, except samples were not

1017 permeabilized. **c,** Histograms showing the intensity of the DAPI staining in images collected

1018 from FLAG-SG2 transfectants (left panel) or wild-type cells (right panel). The mean

1019 fluorescence intensity (MFI) is indicated on each histogram.

1020 **Table Legends**

1021 **Table 1. Library coverage analysis.**

1022 Analysis of coverage in the pilot and final gDNA fragments, and the pilot and final plasmid
1023 libraries. Genomic coverage (reported as bp covered, forward, reverse, and unstranded)
1024 was calculated using fragments of the appropriate orientation, inferred from mapped reads
1025 and the *coverage* function in the GenomicRanges R/Bioconductor package. The total
1026 number of covered bp are indicated, and the corresponding percentage of the HM1:IMSS
1027 reference genome that this represents. Forward fragments are in the same orientation as
1028 genes that are in the forward orientation in the reference genome. Reverse fragments are
1029 in the same orientation as genes that are in the reverse orientation in the reference
1030 genome. Stranded fragments are in the same orientation as genes, and are the combined
1031 total of both forward and reverse fragments. Unstranded fragments are the total number of
1032 fragments, both in the same orientation as genes in the reference genome and in the
1033 opposite orientation as genes. Genes that are targeted are those with at least one unique
1034 fragment with greater than or equal to 27nt that overlaps the gene. Genes that are missed
1035 are those that do not fit the criteria for targeted genes. Genes that are targeted by multiple
1036 fragments are those with two or more unique fragments. Counts were calculated using the
1037 countOverlaps function in the GenomicRanges package and annotation from AmoebaDB.
1038 The number of unique fragments is the total number of unique fragments in each sample.

1039 **Table 2. Identified slow growth mutants.**

1040 Summary of growth phenotypes of SG1 – SG12 slow growth mutants, and analysis of the
1041 sequences present in each corresponding library plasmid. Growth values represent cell

1042 numbers normalized to vector control transfectants on day 4 of growth assays. The
1043 corresponding raw data growth assay data are found in Fig. 5a and Fig. S12a, S12c, S12e).
1044 Mutants exhibited statistically significant growth defects relative to vector control
1045 transfectants; *p*-values from ANOVA analyses are shown (see Fig. S12a, S12c, S12e). The
1046 insert sizes are shown in bp. Insert directionality relative to the corresponding coding
1047 sequence in the annotated genome is shown. Insert frame relative to the trigger sequence
1048 in the pTriggerAdaptor is shown. The accession numbers for coding sequences that inserts
1049 correspond to are indicated, with available annotation information and GO terms for each
1050 gene (NA, not applicable). Predicted roles in metabolic pathways are shown; genes that are
1051 predicted to function in more than one metabolic pathway are indicated as "several." These
1052 predictions are based on data from KEGG Metabolic Pathways (https://www.kegg.jp), the
1053 MetaCyc Metabolic Pathway Database [43], and Chemical Entities of Biological Interest
1054 (ChEBI) database [44].

1055 **Supplementary Figure Legends**

1056 **Figure S1. Preparation of RNA-free gDNA from *E. histolytica.***

1057 Gel electrophoresis analysis of gDNA extracted using the Wizard gDNA extraction kit (Promega),
1058 Quick-DNA miniprep kit (Zymo), and CTAB extraction. **a,** gDNA isolated using the Wizard kit had
1059 no apparent RNA contamination, and the gDNA appeared to be of high quality. **b,** gDNA
1060 isolated using the Quick-DNA kit or CTAB was contaminated with RNA. RNA contamination was
1061 removed by treatment with RNAseIF and subsequent purification. **c,** Bioanalyzer traces of pilot
1062 and final gDNA samples, after fragmentation and size selection.

1063 **Figure S2. Fragment Mapping Analysis.**

1064 Example view of mapped genomic fragments ("intervals") represented in the final plasmid
1065 library. For comparison, the mapped paired end reads that give rise to each fragment are
1066 shown below each fragment. Shown are results from final plasmid batch two, sequencing
1067 replicate one, mapped to contig DS571153.

1068 **Figure S3. Appropriate sequencing depth and lack of bias during library construction.**

1069 **a,** Dinucleotide frequency analysis of the HM1:IMSS reference genome, the pilot and final gDNA
1070 fragments, and the pilot and final plasmid libraries. Stacked bar plots show the percentage of
1071 each dinucleotide. The dinucleotide composition of the gDNA fragments and the final plasmid
1072 library is the same as the genome. **b – c,** Rarefaction curves for pilot (panel b) and final (panel c)
1073 samples. gDNA fragment samples are shown in grey and plasmid samples are shown in orange.
1074 The number of fragments detected with increasing sequence reads shows if sequencing was
1075 performed to saturation. Three sequencing replicates are shown for pilot gDNA samples and
1076 three sequencing replicates are shown for pilot plasmid samples. Three sequencing replicates

1077 are shown for final gDNA samples and three sequencing replicates are shown for each of four
1078 batches of final plasmid samples.

1079 **Figure S4. Reproducible fragment and read counts from biological and technical replicates.**
1080 "Batches" of the final plasmid library were cloned on different days, and represent biological
1081 replicates. "Reps" correspond to independently prepared sequencing preparations, and
1082 represent technical replicates. **a,** Box plot of read counts per fragment in plasmid samples, with
1083 pilot plasmids in orange and final plasmids in grey. Log_2 read counts are shown; outliers
1084 (outside of 2nd and 98th percentile) are plotted as points, and the box represents the 25th, and
1085 75th percentile with median indicated as a bold line. Most fragments were represented by
1086 multiple sequenced reads. **b,** Stacked bar plots of the total number of fragments detected per
1087 sample shows that replicates are similar to each other in the number of fragments identified
1088 and that roughly equal numbers of forward and reverse fragments were produced consistently.
1089 Forward fragments are in the same orientation as genes that are in the forward orientation in
1090 the reference genome. Reverse fragments are in the same orientation as genes that are in the
1091 reverse orientation in the reference genome.

1092 **Figure S5. Annotated genes that were missed or were least covered by sequencing reads.**
1093 Heat map showing the number of unique fragments per gene, for the 50 least covered genes in
1094 the pilot and final gDNA fragments, and the pilot and final plasmid libraries. The color intensity
1095 indicates Log_2 fragment coverage, with dark blue indicating no coverage. Stranded fragments
1096 are in the same orientation as genes. Unstranded fragments are the total number of fragments,
1097 both in the same orientation as genes and in the opposite orientation as genes.

1098 **Figure S6. Clonal transfectants contain only one library plasmid.**

1099 SG1 gDNA was extracted using the Quick-DNA miniprep kit and PCR amplified using the CS5'
1100 and CS3' primers that sit in the pTriggerAdaptor backbone. Sanger sequencing was performed
1101 using the RNAi Sequencing Library F primer. Adaptor 2 starts at the "GGGG" series of
1102 nucleotides, and upstream of this is the SG1 (EHI_194310) sequence. Noncompeting peaks in
1103 the chromatogram suggest that there is one unique genome fragment present in this mutant.

1104 **Figure S7. Gating scheme for imaging flow cytometry analysis.**

1105 Shown is the gating strategy that was used to analyze imaging flow cytometry data, with the
1106 percentage of gated events, and number of gated events, shown below each plot. This
1107 example illustrates the gating of the wild-type sample shown in Figure 7. 10,000 events
1108 were collected. **a,** In-focus events were gated using a gradient of brightfield (BF). **b,** Events
1109 gated in panel a were refined to remove images containing more than one amoeba, by
1110 examining the area and the aspect ratio of the object(s) masked in the brightfield image. **c,**
1111 Events gated in panel b were refined to remove images that contained saturation in the
1112 FLAG channel (channel 11), by examining the raw max pixel intensity (i.e., the intensity of
1113 the brightest pixel in the image). **d,** Events gated in panel c as "unsaturated" were further
1114 examined to identify images that contained both FLAG and DAPI signal, by examining the
1115 intensity of the object masked in the DAPI and FLAG channels. **e,** Events gated in panel d as
1116 "+DAPI/FLAG" were further examined to characterize the degree of colocalization (bright
1117 detail similarity) between the DAPI and FLAG signals.

1118 **Figure S8. Image analysis for FLAG-SG1 transfectants demonstrates that FLAG staining is**
1119 **above background in both permeabilized and non-permeabilized cells.**

1120 Amoebae were stably transfected with a plasmid for expression of full-length FLAG-tagged

1121 SG1. Immunofluorescence was used to determine the localization of SG1 in heterogeneous
1122 transfectants and imaging flow cytometry was used for analysis. The mean fluorescence
1123 intensity (MFI) is indicated on each histogram. **a,** Intensity of FLAG staining in FLAG-SG1
1124 transfectants (left panel) and control staining of wild-type cells (right panel). Samples were
1125 permeabilized prior to antibody staining. **b,** Intensity of FLAG staining in FLAG-SG1
1126 transfectants (left panel) and control staining of wild-type cells (right panel). Samples were
1127 not permeabilized prior to antibody staining.

1128 **Figure S9. Gal/GalNAc lectin immunofluorescence and image analysis.**

1129 Amoebae were stably transfected with a vector control plasmid (pEhEx) and stained with
1130 the cytoplasmic dye CMFDA. Immunofluorescence was used to determine the localization of
1131 the endogenous Gal/GalNAc lectin, and imaging flow cytometry was used for analysis. **a,**
1132 Localization of the Gal/GalNAc lectin (left panel) and control staining of samples in which
1133 primary antibody was omitted (right panel). Samples were not permeabilized prior to
1134 antibody staining. Six random cells are shown for each condition. Shown from left to right
1135 are bright field images (BF), CMFDA staining (CMFDA, green), and mouse anti-Gal/GalNAc
1136 antibody staining (Lectin, red). The numbers in the BF images indicate the object/image
1137 number. **b,** Histograms showing the intensity of the Gal/GalNAc lectin staining (left panel) or
1138 control staining of samples in which primary antibody was omitted (right panel). The mean
1139 fluorescence intensity (MFI) is indicated on each histogram.

1140 **Figure S10. Image analysis for FLAG-SG1 transfectants demonstrates that FLAG staining is**
1141 **above background in permeabilized cells only.**

1142 Amoebae were stably transfected with a plasmid for expression of full-length FLAG-tagged

1143 SG2. Immunofluorescence was used to determine the localization of SG2 in heterogeneous
1144 transfectants and imaging flow cytometry was used for analysis. The mean fluorescence
1145 intensity (MFI) is indicated on each histogram. **a,** Intensity of FLAG staining in FLAG-SG2
1146 transfectants (left panel) and control staining of wild-type cells (right panel). Samples were
1147 permeabilized prior to antibody staining. **b,** Intensity of FLAG staining in FLAG-SG2
1148 transfectants (left panel) and control staining of wild-type cells (right panel). Samples were
1149 not permeabilized prior to antibody staining.

1150 **Figure S11. Nuclear cross-reactivity of the mouse anti-FLAG antibody is only seen in**
1151 **permeabilized cells.**

1152 Nuclear cross-reactivity of the mouse anti-FLAG antibody was not seen in non-
1153 permeabilized cells, supporting that the sample preparation steps did not inadvertently
1154 permeabilize samples, and further supporting the surface staining patterns seen in non-
1155 permeabilized FLAG-SG1 cells. Colocalization analysis of FLAG and DAPI staining is shown.
1156 The percentage of DAPI+/FLAG+ images with colocalization of the two markers are
1157 indicated. Amoebae were stably transfected with a plasmid for expression of full-length
1158 FLAG-tagged SG2, or wild-type cells were analyzed as a control. **a,** Colocalization analysis of
1159 FLAG and DAPI staining in FLAG-SG2 transfectants (left panel) or wild-type cells (right
1160 panel). Samples were permeabilized prior to antibody staining. **b,** Colocalization analysis of
1161 FLAG and DAPI staining in FLAG-SG2 transfectants (left panel) or wild-type cells (right
1162 panel). Samples were not permeabilized prior to antibody staining.

1163 **Figure S12. Identification of slow growing RNAi knockdown mutants.**

1164 Growth analysis of mutants SG1 – SG12. **a, c, e,** Growth analysis demonstrates that selected

1165 clonal knockdown mutant lines have significant growth defects compared to vector control

1166 transfectants. **b, d, f**, Growth was normalized and compared to vector control transfectants

1167 on each day. Mutants exhibited statistically significant growth defects relative to vector

1168 control transfectants.

1169 **Supplementary Table Legends**

1170 **Table S1. Sequencing and fragment count statistics.**

1171 Total number of paired reads mapping to the same contig (mappedPairs) and the number of
1172 fragments inferred from these mapped pairs for each sample. Reads were mapped with bwa
1173 mem, and processed from the resulting SAM formatted file using a custom Python script.
1174 Three sequencing replicates were performed for pilot gDNA samples and three sequencing
1175 replicates were performed for pilot plasmid samples. Three sequencing replicates were
1176 performed for final gDNA samples and three sequencing replicates were performed for each of
1177 four batches of final plasmid samples.

1178 **Table S2. Expression data for annotated genes that were missing in the final plasmid library.**

1179 Available empirical expression data for genes that were missing in the final plasmid library
1180 (genes that are represented in blue in Fig. S5). Accession numbers, annotation information, and
1181 the number of paralogs in the reference genome are indicated. Presence (+) or absence (-) in
1182 gDNA and plasmid datasets is shown. Gene expression from available RNAseq datasets [45, 46]
1183 is indicated, as the FPKM + 1 value. The highest value for FPKM + 1 for each gene is shown. The
1184 total number of spectra that have been identified in proteomics studies is indicated.

1185 **Table S3. Expression data for representative genes present in the final plasmid library.**

1186 To provide a frame of reference for the expression data in Table S2, shown are the available
1187 empirical expression data for 10 random, representative genes that were present in the final
1188 plasmid library. Accession numbers, annotation information, and the number of paralogs in the
1189 reference genome are indicated. Presence (+) or absence (-) in gDNA and plasmid datasets is
1190 shown. Gene expression from available RNAseq datasets [45, 46] is indicated, as the FPKM + 1

1191 value. The highest value for FPKM + 1 for each gene is shown. The total number of spectra that

1192 have been identified in proteomics studies is indicated.

1193 **Table S4. Primer sequences.**

1194 Sequences of the primers used to generate and sequence plasmids, perform RT-PCR, and

1195 create the plasmid RNAi library.

Figure 1

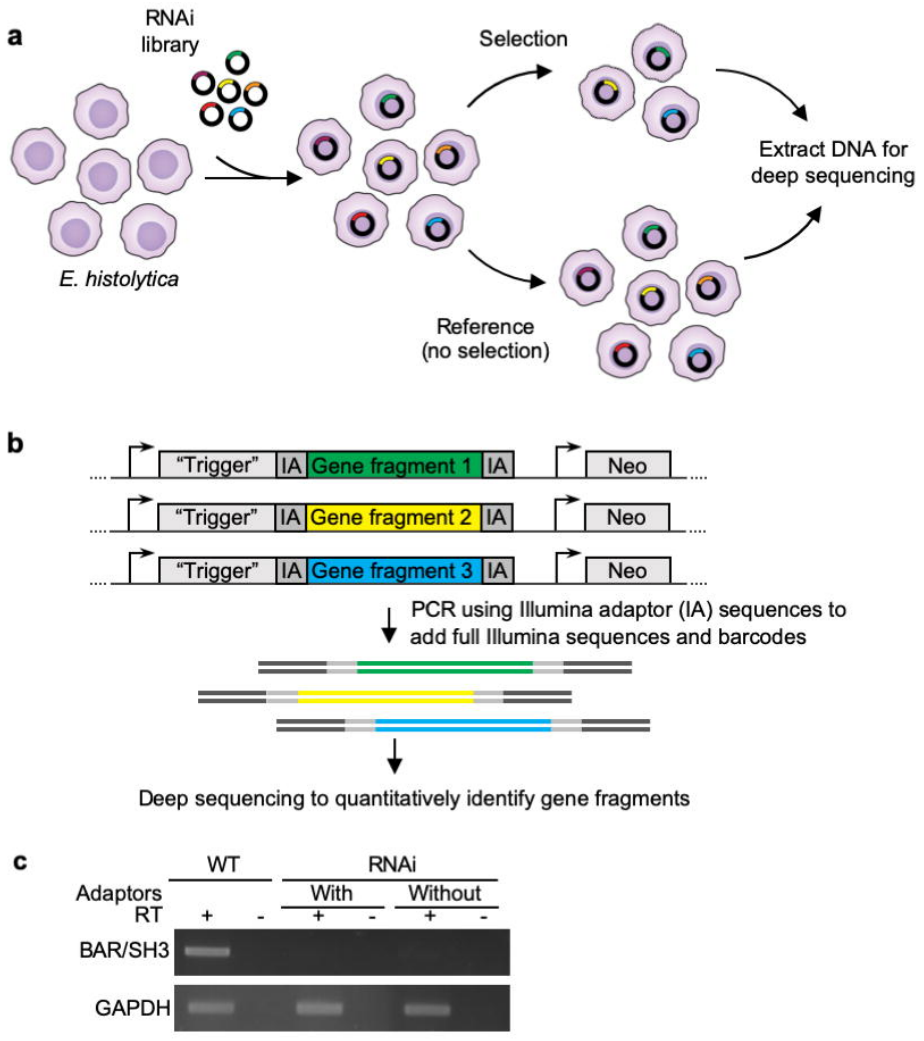
a
RNAi
library
E. histolytica
Selection
Reference
(no selection)
Extract DNA for
deep sequencing
b
"Trigger"
IA
Gene fragment 1
Gene fragment 2
Gene fragment 3
Neo
PCR using Illumina adaptor (IA) sequences to
add full Illumina sequences and barcodes
Deep sequencing to quantitatively identify gene fragments
c
WT
RNAi
Adaptors
With
Without
RT
+
-
BAR/SH3
GAPDH

Figure 2

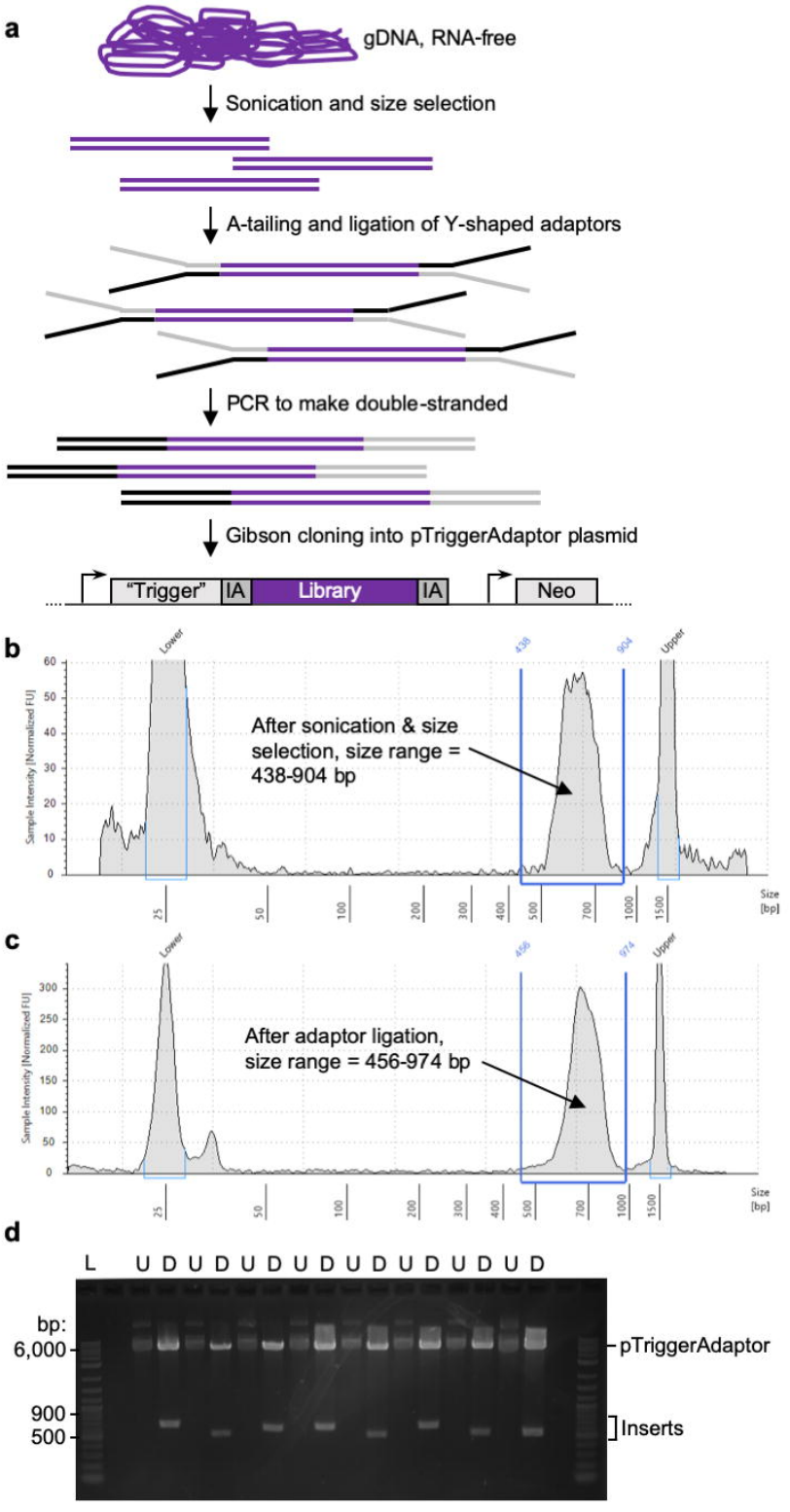
a
gDNA, RNA-free
Sonication and size selection
A-tailing and ligation of Y-shaped adaptors
PCR to make double-stranded
Gibson cloning into pTriggerAdaptor plasmid
"Trigger"
IA
Library
IA
Neo
b
Lower
438
904
Upper
After sonication & size selection, size range = 438-904 bp
Sample Intensity [Normalized FU]
Size [bp]
c
Lower
456
974
Upper
After adaptor ligation, size range = 456-974 bp
Sample Intensity [Normalized FU]
Size [bp]
d
L U D U D U D U D U D U D U D U D
bp:
6,000
900
500
pTriggerAdaptor
Inserts

Figure 3

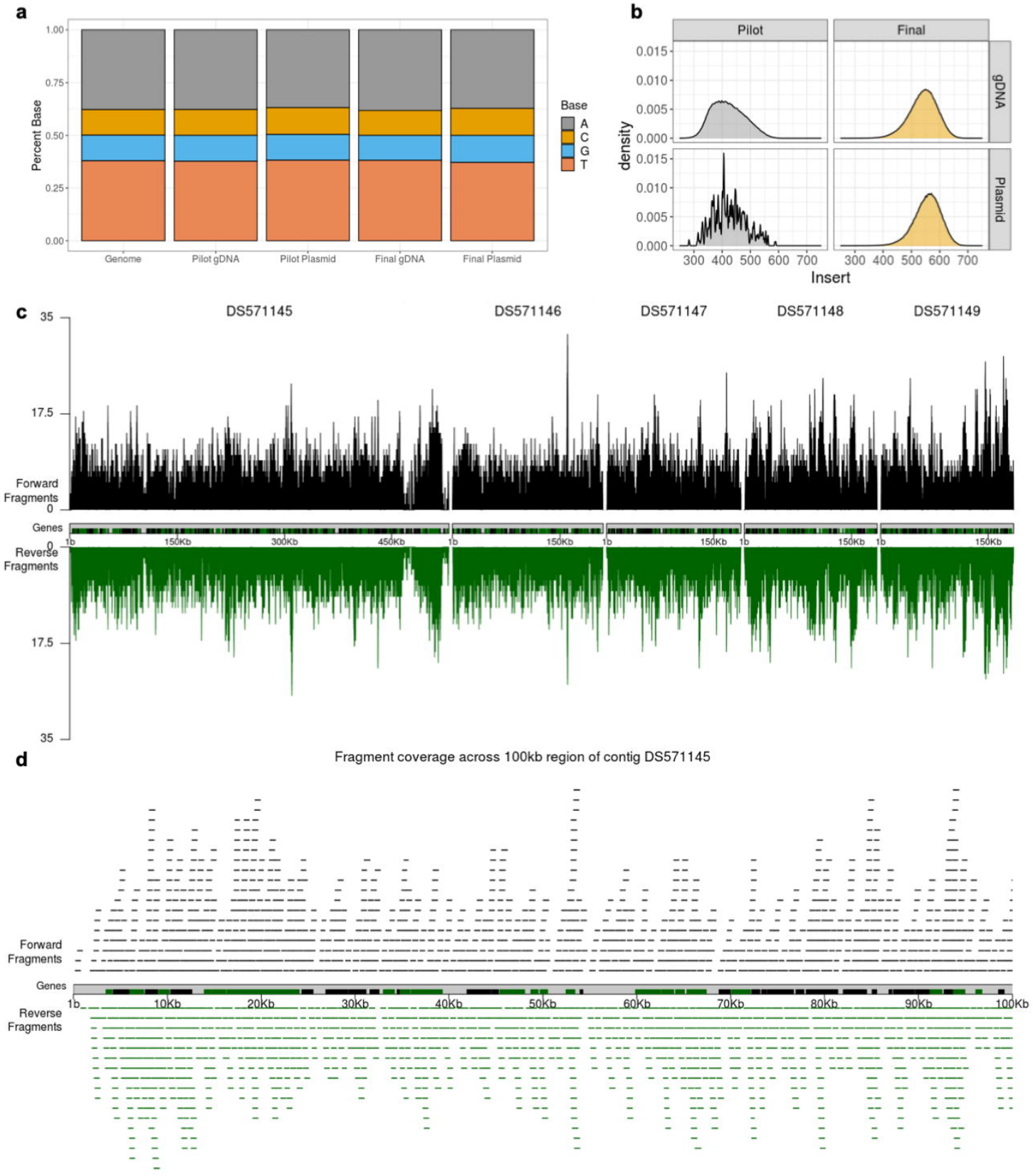
a
Percent Base
1.00
0.75
0.50
0.25
0.00
Genome
Pilot gDNA
Pilot Plasmid
Final gDNA
Final Plasmid
Base
A
C
G
T
b
Pilot
Final
gDNA
Plasmid
density
0.015
0.010
0.005
0.000
300 400 500 600 700
Insert
c
DS571145
DS571146
DS571147
DS571148
DS571149
35
17.5
0
Forward Fragments
Genes
Reverse Fragments
1b
150Kb
300Kb
450Kb
d
Fragment coverage across 100kb region of contig DS571145
Forward Fragments
Genes
Reverse Fragments
1b
10Kb
20Kb
30Kb
40Kb
50Kb
60Kb
70Kb
80Kb
90Kb
100Kb

Figure 4

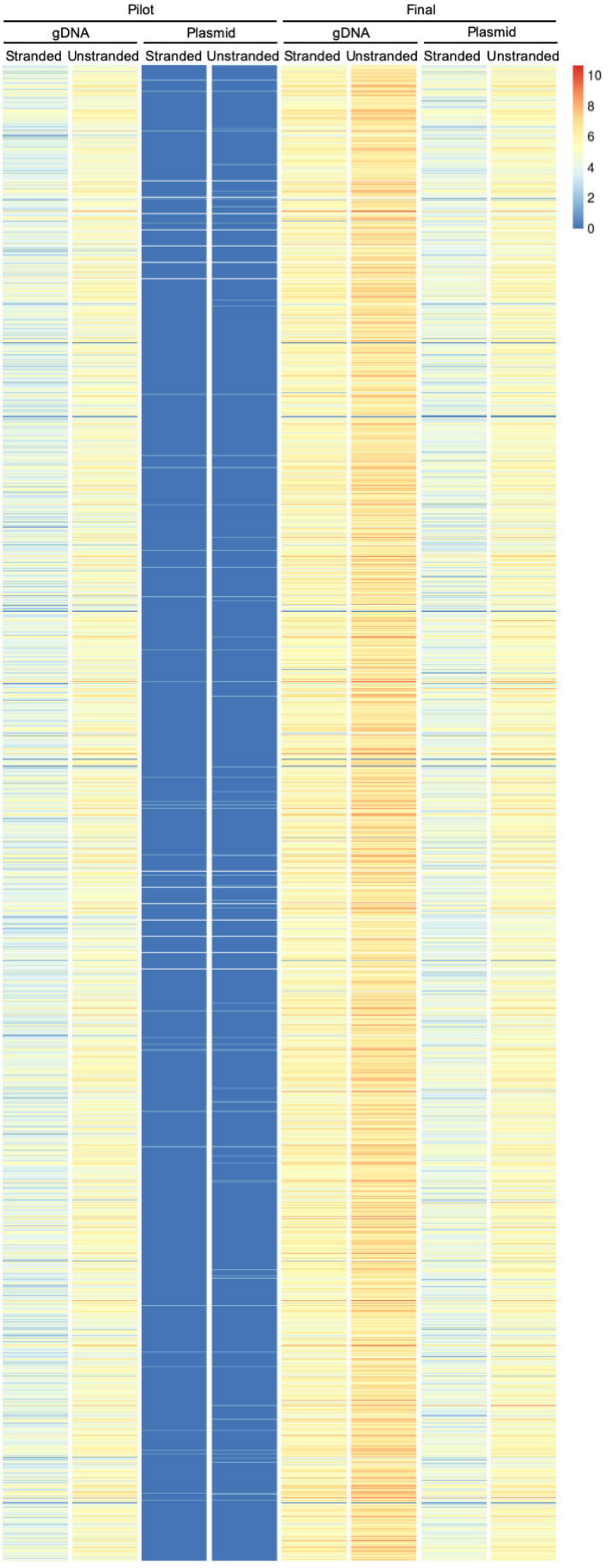
Pilot
Final
gDNA
Plasmid
gDNA
Plasmid
Stranded
Unstranded
Stranded
Unstranded
Stranded
Unstranded
Stranded
Unstranded
10
8
6
4
2
0

Figure 5

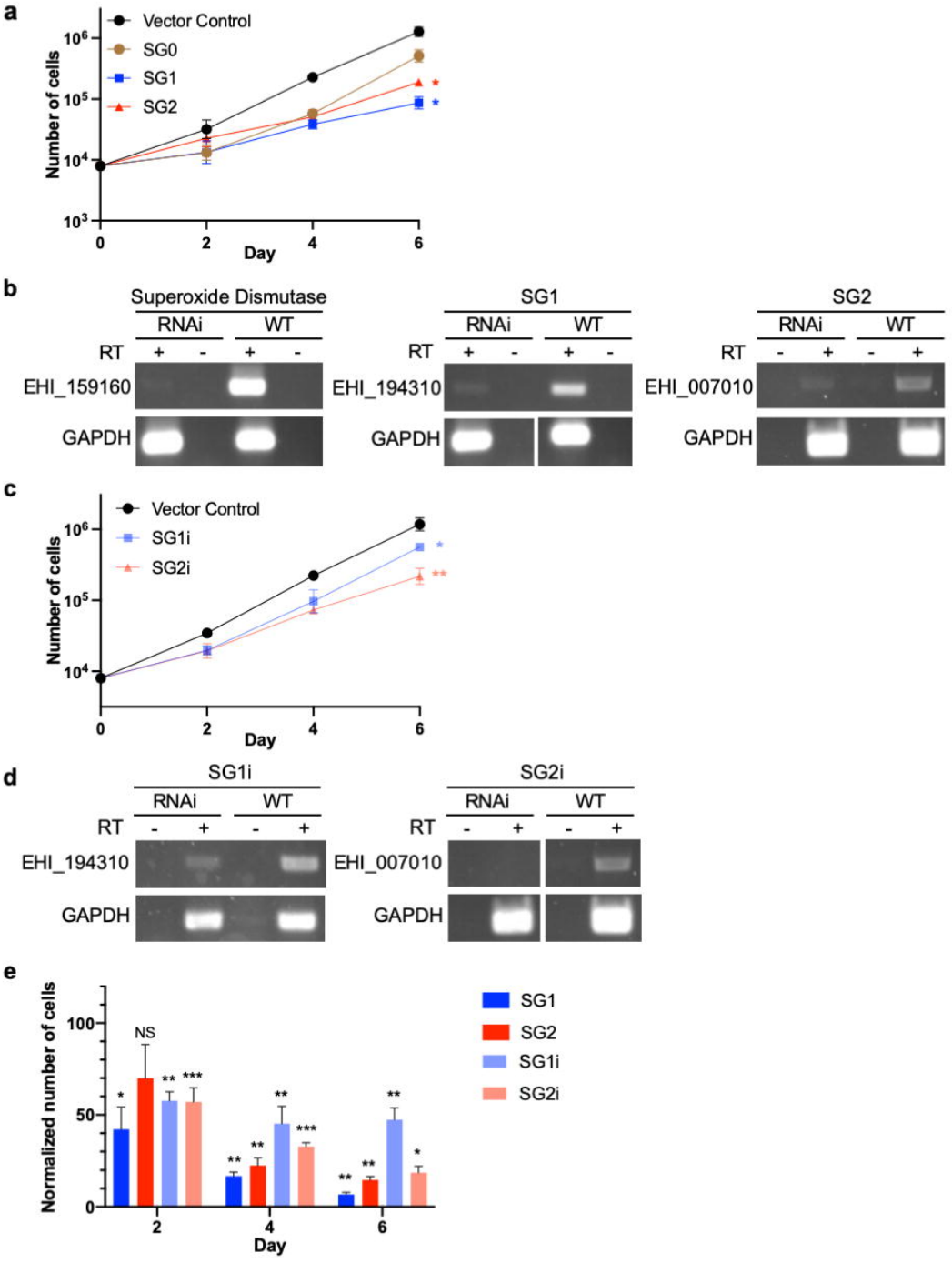
a
Vector Control
SG0
SG1
SG2
Number of cells
Day
b
Superoxide Dismutase
SG1
SG2
RNAi
WT
RT
EHI_159160
EHI_194310
EHI_007010
GAPDH
c
Vector Control
SG1i
SG2i
Number of cells
Day
d
SG1i
SG2i
RNAi
WT
RT
EHI_194310
EHI_007010
GAPDH
e
Normalized number of cells
Day
NS
SG1
SG2
SG1i
SG2i

Figure 6

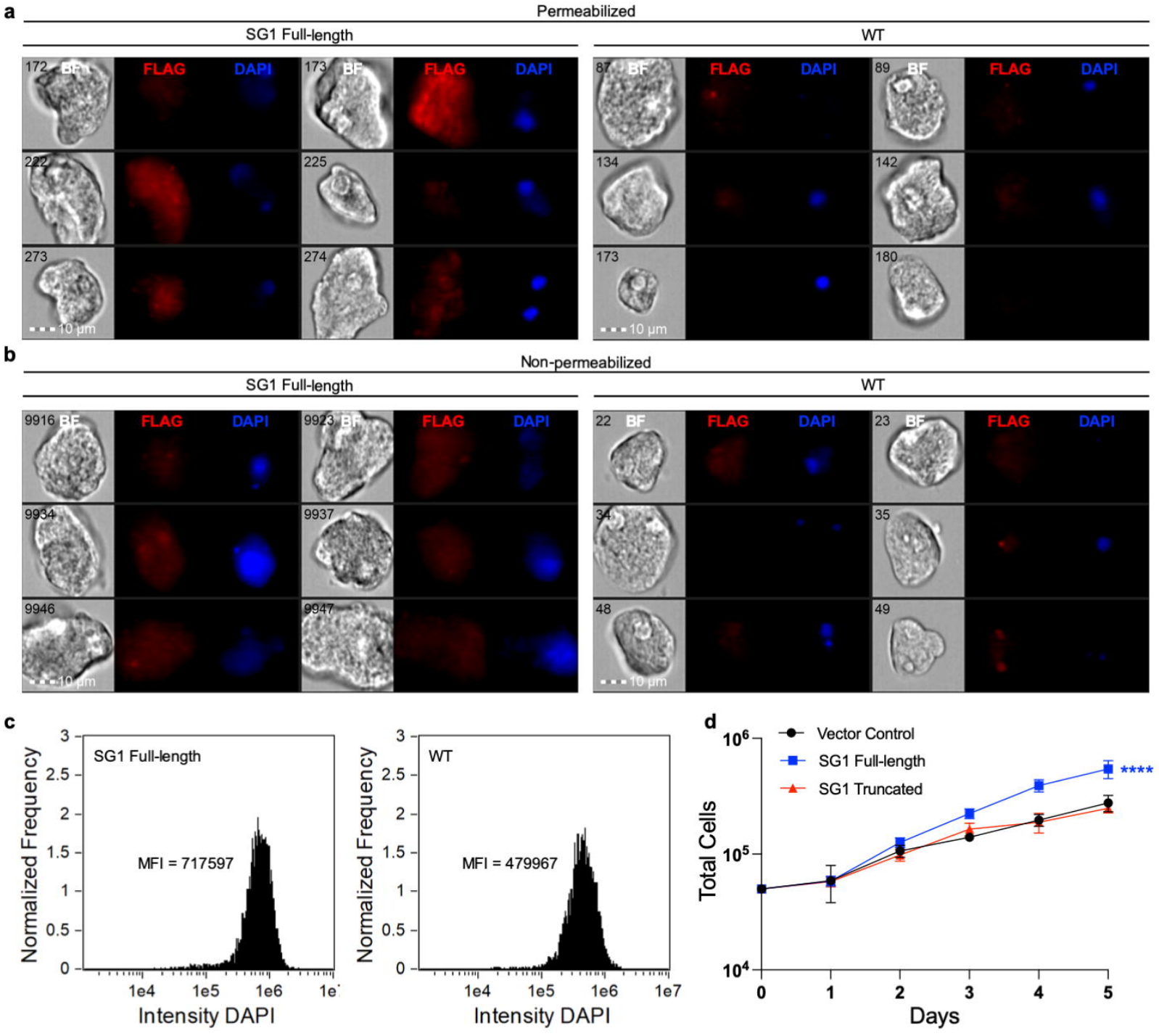
a
Permeabilized
SG1 Full-length
WT
BF
FLAG
DAPI
10 µm
b
Non-permeabilized
SG1 Full-length
WT
c
SG1 Full-length
MFI = 717597
WT
MFI = 479967
Normalized Frequency
Intensity DAPI
d
Vector Control
SG1 Full-length
SG1 Truncated
****
Total Cells
Days

Figure 7

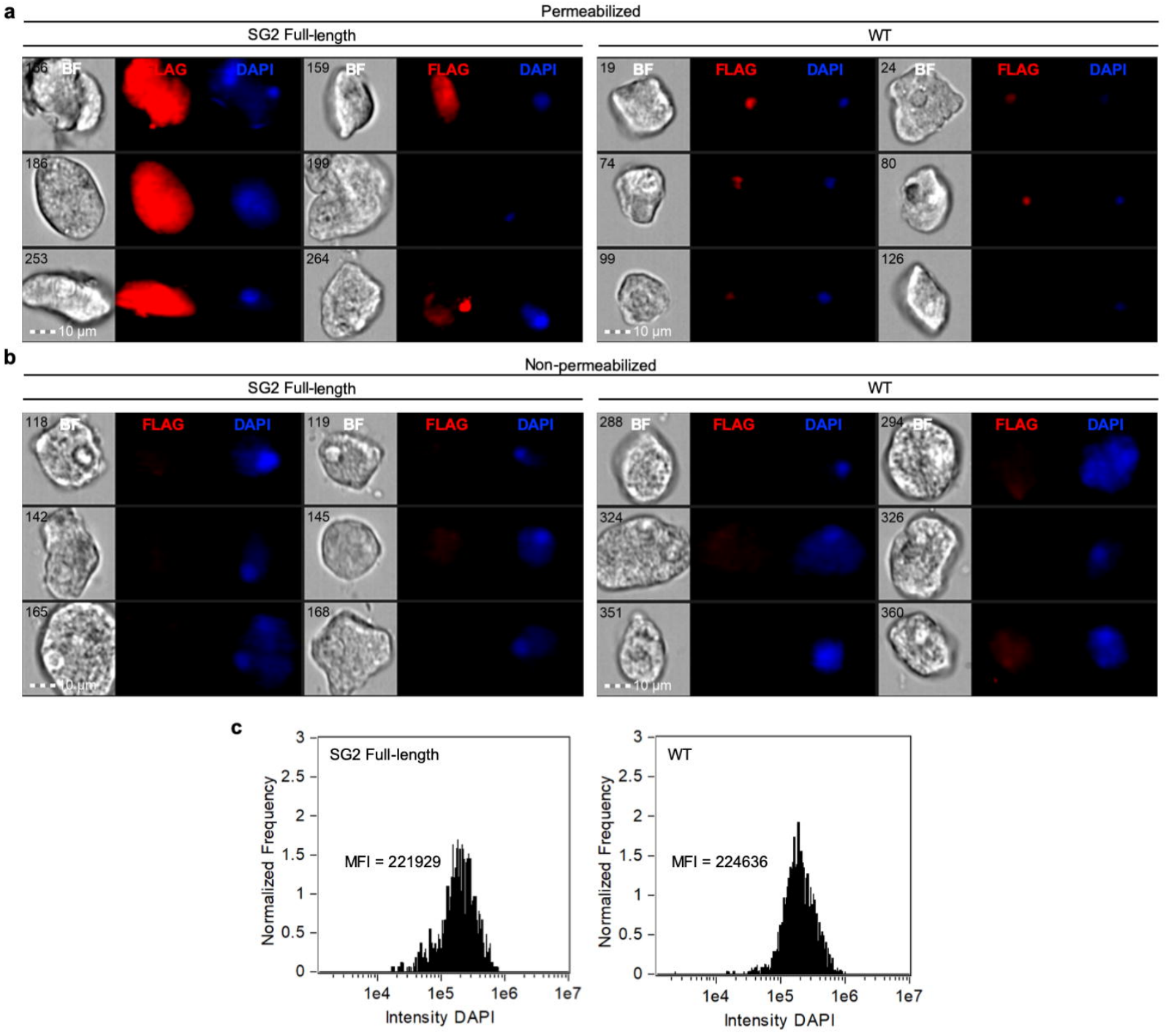
a
Permeabilized
SG2 Full-length
WT
BF
FLAG
DAPI
10 µm
b
Non-permeabilized
SG2 Full-length
WT
c
SG2 Full-length
MFI = 221929
WT
MFI = 224636
Normalized Frequency
Intensity DAPI
1e4
1e5
1e6
1e7

Table 1

	Pilot		Final	
	gDNA	Plasmid	gDNA	Plasmid
Covered bp, forward	19961090	93257	20500538	19730276
Covered %, forward	96.0	0.5	98.6	94.9
Covered bp, reverse	19949512	99274	20499865	19734222
Covered %, reverse	96.0	0.5	98.6	94.9
Covered bp, unstranded	20401392	191674	20571421	20225883
Covered %, unstranded	98.1	0.9	98.9	97.2
Genes targeted, stranded	8299	152	8310	8302
Genes missed, stranded	34	8181	23	31
Mean fragments per gene, stranded	21	0	59.2	23.3
Genes targeted by multiple fragments, stranded	8270	54	8308	8276
Genes targeted, unstranded	8310	305	8310	8307
Genes missed, unstranded	23	8028	23	26
Mean fragments per gene, unstranded	42.9	0.1	119.9	47.2
Genes targeted by multiple fragments, unstranded	8307	111	8310	8301
Number of unique fragments	537066	621	1509696	489981

Table 2

Name	Normalized growth	*p*-value	Insert size (bp)	Insert Direction	Insert Frame	Accession	Annotation	Annotated Domains	GO term	Metabolic Pathways
SG1	16.74 +/- 2.17	***	451	Reverse	Out	EHI_194310	Amino acid transporter, putative	Amino acid transporter, transmembrane domain	NA	NA
SG2	22.58 +/- 4.10	**	544	Forward	In	EHI_007010	Hypothetical protein	NA	NA	NA
SG3	72.87 +/- 6.54	*	403	Forward	Out	DS571580	NA	NA	NA	NA
SG4	9.45 +/- 0.78	****	351	Reverse	Out	EHI_110390	Transporter, major facilitator family	MFS transporter superfamily	NA	NA
SG5	14.50 +/- 0.91	****	463	Forward	In	EHI_024220	Hypothetical protein	Importin beta family	Protein import into nucleus	NA
				Forward	In	EHI_160920	Hypothetical protein	Importin beta family	Protein import into nucleus	NA
				Forward	In	EHI_171760	Hypothetical protein	Importin beta family	Obsolete protein import into nucleus, docking	NA
SG6	15.19 +/- 2.23	****	380	Reverse	In	EHI_093960	Pre-mRNA splicing factor, putative	Pre-mRNA-processing factor 6/Prp1/STA1	mRNA splicing, *via* spliceosome	Several
SG7	19.59 +/- 1.89	**	555	Reverse	In	EHI_009580	Tyrosine kinase, putative	Protein kinase-like domain superfamily	Protein phosphorylation	Several
SG8	65.35 +/- 5.05	***	370	Forward	In	EHI_148960	Hypothetical protein	NA	NA	NA
SG9	11.08 +/- 0.96	ns	352	Forward	In	DS571952	NA	NA	NA	NA
SG10	11.45 +/- 1.29	****	425	Forward	In	EHI_069320	C2 domain containing protein	C2 domain	NA	Several
SG11	40.47 +/- 6.77	**	408	Reverse	Out	EHI_098820	Hypothetical protein	Dbl homology (DH) domain; Armadillo-type fold	Regulation of Rho protein signal transduction	NA
SG12	18.09 +/- 1.83	**	322	Reverse	In	DS571214	NA	NA	NA	NA