Application of barcode sequencing to increase the throughput and complexity of *Plasmodium falciparum* genetic screening

Allan Muhwezi[1], Mehdi Ghorbal[2], Theo Sanderson[2], Maria Ivanova[2], Rizwan Ansari[2], Sarah Harper[2], Wesley Wong[3], Reiner Schulte[1], Gareth Girling[2], Frank Schwach[2], Ellen Bushell[4], Charlotte Beaver[2], Oliver Billker[4] & Julian C. Rayner[1*]

[1]Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

[2]Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK

[3]Harvard T.H. Chan School of Public Health, Harvard University, Cambridge, USA

[4]Umeå University, Umeå, Sweden

Summary

All the pathology and symptoms associated with malaria are caused by the growth of *Plasmodium* parasites inside human red blood cells. This process, which in the case of the major human malaria pathogen *Plasmodium falciparum* takes place over a 48-hour period, involves multiple tightly regulated developmental transitions. Understanding the *P. falciparum* genes that regulate these key processes could lead to the identification of targets for new drugs. However, while large-scale sequencing efforts have led to a good understanding of the *P. falciparum* genome and how it evolves over time and space, a disconnect remains between the amount of genome sequence data available and the amount of data describing what exactly the genes contained within it do – the phenotype. We have generated a panel of 66 *P. falciparum* lines carrying individual gene knockouts tagged with unique DNA barcodes. We then used these lines in a series of assays that combine flow cytometry, cell sorting and DNA barcode quantification using next generation sequencing (Barcode Sequencing or BarSeq) to phenotype key aspects of the parasite life cycle such as growth, replication capacity and cell cycle progression. This approach both yields new

data about individual gene function, and outlines a new approach where barcoded *P. falciparum* lines are used in pooled BarSeq-based assays to generate more precise phenotype data at scale.

Keywords: *Plasmodium falciparum,* malaria, CRISPR-Cas9, experimental genetics, pooled phenotyping, parasite growth, DNA replication, cell cycle

33 **Introduction**

34 *Plasmodium falciparum* parasites cause the most virulent form of human malaria which accounts
35 for hundreds of millions of cases and more than 500,000 deaths annually (WHO 2022). Significant
36 progress has been recently made in controlling this disease, but the repeated emergence and
37 spread of antimalarial drug and insecticide resistance, along with currently limited vaccine
38 coverage, all suggest that further interventions will be required to deliver malaria elimination
39 (Donnelly, Isaacs and Weetman 2016, WHO 2022). The search for new antimalarial targets can
40 be directly informed by the *P. falciparum* genome, which encodes over 5400 genes (Gardner, Hall
41 et al. 2002). Homology and functional prediction reveals many potential intervention targets,
42 including multiple enzymes and transporters, but more than 50% of *P. falciparum* genes lack
43 conservation with model organisms and therefore their function could not be directly predicted by
44 homology alone (Gardner, Hall et al. 2002). Despite extensive annotation efforts, the function of
45 ~35% of the *P. falciparum* genome remain hypothetical (Oberstaller, Otto et al. 2021). More
46 targeted and, critically, scalable efforts are needed to understand *P. falciparum* gene function in
47 order to identify new drug and vaccine targets.

48

49 Experimental genetic manipulation of the *Plasmodium* genome has been central to efforts to
50 characterize gene function since it was first established in the early 1990's (Van Dijk, Waters and
51 Janse 1995, Wu, Sifri et al. 1995). In the subsequent decades, genetics has been used to identify
52 genes involved in many parasite processes such as erythrocyte invasion, drug resistance,
53 metabolism, and virulence. This has been achieved by targeting DNA loci with a range of
54 manipulation strategies such as knock-out, replacement or mutagenesis using either conventional
55 recombination approaches or targeted recombination using CRISPR/Cas9 and zinc finger
56 nucleases (Straimer, Lee et al. 2012, Ghorbal, Gorman et al. 2014). A range of complex
57 conditional tools have also been developed to study the function of essential genes/proteins,
58 including mRNA degradation using inducible promoters and *glm S* ribozymes, protein degradation

using destabilizing domains such as DD and DDD (de Koning-Ward, Gilson and Crabb 2015) and conditional gene deletion using small-molecule induced diCre and FRT recombinases (Combe, Giovannini et al. 2009, Collins, Das et al. 2013). These detailed gene-by-gene approaches have been invaluable and have produced a wealth of insight and data. However many of these approaches are relatively low throughput due to low transfection efficiencies, limited numbers of selectable markers and the need to generate complex vectors and constructs (de Koning-Ward, Gilson and Crabb 2015, Lee, Lindner et al. 2019). More systematic high throughput approaches are needed to both the generation and phenotyping of *P. falciparum* mutant lines to truly expand understanding of the many genes of unknown function.

Despite these challenges, there have been some large genetic screens in *Plasmodium* parasites. Before the development of CRISPR/Cas9, two medium-scale screens of >50 genes encoding kinases or proteins predicted to be involved in cytoadherence were carried out with standard recombination technology (Maier, Rug et al. 2008, Solyakov, Halbert et al. 2011). For the rodent model *P. berghei,* the *Plasmo*GEM (*Plasmodium* Genetic Modification) Project used a lambda recombination and Gateway cloning system to generate targeted barcoded vectors for >2500 genes, which were transfected in pools of 100 vectors at a time. Nextgen sequencing was then used to quantify the relative abundance of each barcode within the mixed pool (Barcode Sequencing or BarSeq) in order to quantify individual parasite growth rates (Bushell, Gomes et al. 2017). This approach has now been used successfully to assign growth phenotypes at blood (Bushell, Gomes et al. 2017), pre-erythrocytic and mosquito (Stanway, Bushell et al. 2019) and gametocyte (Russell, Sanderson et al. 2023) stages in *P. berghei*. The lower transfection efficiency in *P. falciparum* has to date precluded a similar approach in the human pathogen, but here the *piggyBac* transposon system, which uses random insertion of *the piggyBac* transposon at TTAA sites, which occur at high frequencies in the AT-rich *P. falciparum* genome, has been used to generate >30,000 *P. falciparum* mutants and provide a first genome-wide estimate of

85 gene essentiality in this species, based on the pattern of transposon insertion (Zhang, Wang et
86 al. 2018). Both these large-scale screens show that a significant proportion of the *Plasmodium*
87 genome is required for blood-stage growth.
88
89 A major challenge however that still remains is the development of more complex screens that
90 extend beyond just parasite growth measurements. Parasite growth could be impaired for multiple
91 reasons, from limiting one particular step in intra-erythrocytic development (e.g., metabolism or
92 invasion) or by extending the length of the intra-erythrocytic cycle as whole. In addition, the
93 environment and growth conditions are also likely to influence gene function and phenotypes, as
94 has been shown extensively in other organisms such as yeast (Hillenmeyer, Fung et al. 2008).
95 Growth screens alone are therefore only the starting point for understanding gene function. In this
96 work we explored the use of BarSeq to develop more systematic screening approaches
97 specifically for *P. falciparum.* We did so by generating a library of 66 barcoded *P. falciparum*
98 knockout lines using a custom CRISPR/Cas9 construct generation and transfection pipeline, and
99 then screening them both as individual lines and as pools for parasite growth, DNA replication,
100 intra-erythrocytic cycle length and progression, leveraging BarSeq as a method of quantification
101 wherever possible. This generated both information about individual gene function, and also the
102 potential for future larger-scale targeted genetic screens in *P. falciparum* which will accelerate the
103 discovery of new potent drugs and combinatorial therapies with novel cellular targets through
104 screening diverse libraries of small molecules as well as pinpointing gene function across multiple
105 parasite lifecycle stages and environmental perturbations.

106 **Results**

107

108 **Generation of a panel of *Plasmodium falciparum* deletion lines using a CRISPR/Cas9**
109 **pipeline**

110 We developed a CRISPR/Cas9 pipeline to generate a panel of barcoded *P. falciparum* deletion
111 lines for use in mixed pool phenotyping (**Supplementary Figure 1**). Guide RNA (gRNA) and
112 homology repair designs were generated targeting 522 genes using an automated *in silico*
113 construct design pipeline (see Methods). The homologous repair plasmid consists of a positive
114 expression cassette (hDHFR, which can be selected using WR99210) flanked by 5' and 3' regions
115 designed to target repair of the gRNA cleaved locus. An 11-12 nucleotide unique barcode flanked
116 by conserved primer annealing sites was also incorporated into each homologous repair construct
117 to facilitate subsequent high throughput phenotyping using barcode sequencing, BarSeq (Bushell,
118 Gomes et al. 2017). In order to increase the chances of creating deletion lines with growth
119 phenotypes, 392 (75.1%) genes were selected on the basis that deletion of their orthologues in
120 *P. berghei* resulted in a slow growth phenotype (Bushell, Gomes et al. 2017). 31 genes predicted
121 essential and 34 genes predicted dispensable in the same *P. berghei Plasmo*GEM screen were
122 included as controls. An additional 65 genes predicted to be involved in invasion in *P. falciparum*
123 and for which no clear orthologue could be identified in *P. berghei* were also included, again to
124 increase the chance of revealing growth phenotypes.

125

126 Of the 522 construct designs, 234 (44.8%) were successfully cloned using a Gibson assembly
127 pipeline and transfected by electroporation into *P. falciparum* parasite cultures enriched for ring
128 stages. 152 (65.0%) transfectants were recovered after positive selection using the drug
129 WR99210 and thereafter frozen down. Frozen isolates were subsequently revived, and a second
130 round of positive selection performed, of which 93 (61.2%) lines regrew in the presence of
131 WR99210. Positive/negative selection was then carried out to select against maintenance of the

132 repair plasmid and hence for lines in which the homology region had integrated into the genome
133 via double cross-over recombination; this resulted in the recovery of 66 (71.0%) unique lines.
134 Once the combination of both positive and negative selection had been completed for all the lines,
135 the nucleotide sequence identity of the barcode module was then determined using capillary
136 sequencing, followed by PCR-based genotyping using primers flanking the integrated region in
137 order to confirm cassette integration (see Methods). A summary of the proportion of genes, along
138 with the estimated growth phenotypes based on the *Plasmo*GEM *P. berghei* screen
139 (essential/slow/redundant) that progressed through each stage of the CRISPR/Cas9 pipeline is
140 shown in **Table 1** (where the denominator used to calculate the % is the number of
141 constructs/lines that passed the previous step). Details of each construct and its status along the
142 transfection pipeline are included in **Supplementary Table 1**. In certain instances, >1 round of
143 transfections was performed using the same constructs as shown in Supplementary Table 1. A
144 summary of the genotyping results is shown in **Supplementary Table 2**. In general, there was
145 some degree of concordance between predictions based on the *P. berghei Plasmo*GEM data and
146 transfection outcome – of constructs that were transfected, only 2/14 genes (14.2%) identified as
147 essential in *P. berghei* passed combined positive/negative selection, whereas 51/173 (29.5%)
148 and 5/15 (33.3%) of genes identified as slow or redundant in *P. berghei* reached the same stage
149 respectively.

150 **Table 1**. **Outcome of CRISPR/Cas9 pipeline to generate barcoded gene deletion lines in**
151 ***Plasmodium falciparum*.**

*Plasmo*GEM growth phenotype	Construct designs	Transfected constructs	Recovered transfectants	Positive selection (post-thaw)	Negative selection
Essential	31	14 (45.2%)	6 (42.9%)	2 (33.3%)	2 (100%)
Slow	392	173 (44.1%)	108 (62.4%)	74 (68.5%)	51 (68.9%)
Dispensable	34	15 (44.1%)	12 (80.0%)	8 (66.7%)	5 (62.5%)
No orthologue	65	32 (49.2%)	26 (81.3%)	9 (34.6%)	8 (88.9%)
Total	**522**	**234 (44.8%)**	**152 (65.0%)**	**93 (61.2%)**	**66 (71.0%)**

152

153 Several additional lines were generated separate from the automated transfection pipeline.
154 Constructs targeting PfRH3 (PF3D7_1252400) and PfEBA165 (PF3D7_0424300) were included
155 as controls as both have previously been reported as pseudogenes in *P. falciparum*, and previous
156 disruptions of these genes resulted in non-significant growth phenotypes during the blood stages
157 of the parasite life cycle (Taylor, Triglia et al. 2001, Triglia, Thompson and Cowman 2001, Proto,
158 Siegel et al. 2019). Another locus, PfPARE (PF3D7_0709700) was also included as a control
159 since its disruption is also thought to be dispensable in blood stages based on the growth
160 phenotypes of parasite lines with a missense mutation in the protein catalytic site (Istvan, Mallari
161 et al. 2017), although minor growth defects resulting from disruption of the PfPARE locus and
162 tagged with reporter genes have also been reported (Hoshizaki, Jagoe and Lee 2022). Multiple
163 deletion lines with different barcode modules were generated for both PfRH3 and PfPARE to
164 provide multiple controls as well as approaches to account for residual variation and experimental
165 robustness. While the majority of deletions were made in the NF54 background, an additional
166 three deletion lines targeting the PfPARE locus were also made in Dd2, 3D7 and GB4 strains as
167 controls for the effect of parasite genetic background.
168

169 **Assessment of parasite growth using individual lines in arrayed screens**

170 The effect of deleting individual genes on parasite growth was measured using both an arrayed
171 approach, where deletion lines were cultured individually and growth monitored using flow
172 cytometry every 48h, and also using mixed pools containing all deletion lines, where relative
173 growth could be quantitated using BarSeq (Bushell, Gomes et al. 2017). In both assays parasite
174 cultures were sampled every 48h, grown in the same incubators and cultured for five cycles (for
175 more details, see Methods).

176

177 To select the most relevant line to be used as a control, growth rates for the control lines PfRH3
178 (variant 1), PfRH3 (variant 2), PfPARE (NF54) and PfEBA165 were tested initially along with the
179 NF54 wild type line. All these control lines had growth rates that overlapped with the wild-type
180 NF54 line, although with some differences in the precise mean Relative Growth Rate (RGR)
181 growth estimates (**Supplementary Figure 2)**. Similarly, wild-type Dd2 and PARE [Dd2], and wild-
182 type 3D7 and PARE [3D7], also had non-significant growth differences, although both wild-type
183 Dd2 and PARE [Dd2] grew faster than 3D7 and NF54 lines, consistent with previous observations
184 (Reilly, Wang et al. 2007, Murray, Stewart et al. 2017). The lack of significant difference between
185 NF54 PfRH3, PfPARE and PfEBA165 knockouts relative to NF54 meant that in theory any could
186 be used as controls to compare relative growth rates with other deletion lines, however PfRH3
187 was selected because it had the closest average RGR to WT NF54.

188

189 Deletion lines were scored as either slow (if the deletion line had a relative growth rate [RGR]
190 significantly lower than the PfRH3 control lines with no overlapping confidence intervals), non-
191 significant (if the deletion line and the PfRH3 control lines had overlapping confidence intervals),
192 or fast (if the deletion line had a RGR significantly higher than the PfRH3 control lines and with
193 no overlapping confidence intervals). Estimates of growth rates from wells that had been initially
194 seeded with synchronised cultures are summarised in **Fig 1a** and show that 51/65 (78.4%) lines

195 had slow growth phenotypes, 11/65 (16.9%) lines had non-significant growth phenotypes,
196 whereas two lines, PfPARE [Dd2] and PfACS9 (PF3D7_0215000) had fast growth phenotypes
197 with EBA175 [Dd2] (PF3D7_0731500) closely bordering significance for fast growth. By contrast,
198 when the same arrayed assays were conducted using non-synchronised starting cultures, only
199 24/70 (34.3%) lines had a slow growth phenotype, while 41/70 (58.6%) were non-significant and
200 5/70 (7.1%) had a fast growth phenotype (shown in **Fig 1b**). Both PfACS9 and PARE [Dd2] had
201 high growth phenotypes in both synchronous and asynchronous conditions. All the control lines
202 in synchronous culture and asynchronous culture had peak distributions within expected ranges
203 and concordant with genetic background (**Fig 1c – 1d**), suggesting robustness in these
204 experiments. There was a significant positive correlation between RGR estimates in synchronous
205 and asynchronous culture (r = 0.60, P < 0.001), shown in **Fig 1e**, however some lines such as
206 EBA175 [Dd2] had negative residuals that were greater than two standard deviations from the
207 PfRH3 control lines.
208
209 **Assessment of parasite growth using pooled lines by BarSeq**
210 For the pooled approach, the deletion lines were grown in a single mixed pool containing all 65
211 lines mixed in equal proportions. This pool was created once, split into triplicates, then sampled
212 over sequential growth cycles. Growth was assessed by measuring mean fold changes in the
213 proportion of each barcode every 48h over five cycles. Growth rates for each deletion line were
214 significantly correlated across the three replicates using multiple linear regression (Adjusted R-
215 squared = 0.60, P < 0.001), which suggested that there were minimal batch effects (shown in
216 **Supplementary Figure 3)**. Two approaches were used to analyse the growth data of the mixed
217 pools, differing primarily in how the confidence intervals were calculated. The first method BarSeq
218 (RGR) was based on estimations of the per 48h cycle fold changes in growth rates for each
219 deletion line normalized to the average per 48h cycle fold change of PfRH3 (variant1) and PfRH3
220 (variant2) controls (shown in **Fig 2a**); this approach is most comparable to RGR for the arrayed

221 mutants. In the second approach BarSeq (MFC), mean fold changes in growth rates were
222 calculated using a regression of the cumulative fold changes independent of the two PfRH3
223 control variants but with the expectation that parasite lines with similar growth profiles would
224 overlap. Growth phenotypes were scored as slow, non-significant and fast following a similar
225 approach described using the arrayed screen (summarized in **Fig 2b**). The results using the
226 BarSeq (MFC) approach show that 46/65 lines (70.8%) had slow growth phenotypes, 19/65
227 (29.2%) had non-significant growth phenotypes, whereas there were no fast growth phenotypes
228 scored. More non-significant phenotypes were observed with the BarSeq (RGR) method, where
229 34/65 lines (52.3%) had slow growth phenotypes, 31/65 (47.7%) had non-significant growth and
230 similarly no fast growth phenotypes were scored. Controls PfEBA165 and PfPARE both had non-
231 significant growth rates relative to PfRH3 (variant 1) and PfRH3 (variant 2), consistent with the
232 arrayed approach. A summary of MFC scores plotted in ranked order is shown in
233 **(Supplementary Figure 4)**. Growth rates generated using BarSeq (MFC) and BarSeq (RGR)
234 were positively correlated (r = 0.93, P < 0.0001) as shown in **(Supplementary Figure 5)**. There
235 was a significant positive correlation between the arrayed and pooled screen (r = 0.54, P < 0.0001)
236 as shown in **Fig 2c**, with all the differences in phenotype calls at the boundaries of a gradient of
237 either low/non-significant or non-significant/high (shown in **Fig 2d**). This also suggests that there
238 were no significant interactions in proliferation of deletion lines in mixed pools under these growth
239 conditions, which might potentially confound growth rates as shown for wild strains of Dd2 and
240 HB3 (Wacker, Turnbull et al. 2012). Growth measurements using the pooled approach were
241 repeated with a smaller subset of genes over multiples rounds, and these were consistent with
242 previous observations. BarSeq therefore appears to be a robust method for measuring growth of
243 *P. falciparum* lines in a scalable manner and has considerable benefits of throughput over an
244 arrayed approach.
245

246 **Validation of growth phenotypes**

247 Growth rates and phenotypes from both the arrayed and mixed pool approach were first validated
248 by comparing them with previously published data from the large-scale *Plasmo*GEM and
249 *Piggy*Bac screens in *P. berghei* and *P. falciparum* respectively (Bushell, Gomes et al. 2017,
250 Zhang, Wang et al. 2018). Growth rates measured using both the arrayed and mixed pool
251 approaches were significantly positively correlated with the *Piggy*Bac mutagenesis fitness score
252 (MFS) and *Plasmo*GEM relative growth rates (shown in **Table 2**). Correlations in RGR were more
253 similar to *Piggy*Bac mutants (which were made in *P. falciparum*), than *Plasmo*GEM mutants
254 (which were made in *P. berghei*). The proportion of mutants with a particular qualitative growth
255 phenotype were also similar across studies (summarised in **Supplementary Figure 6**). This
256 therefore suggested that there was broad agreement between the studies, although discordant
257 matches also existed, most of which bordered on the margins of whether a line was scored as
258 slow or non-significant vs non-significant or fast. There were no obvious GO terms enriched for
259 either the discordant or concordant genes. These differences might underlie biological differences
260 between species, or technical differences in the environment or methodology of the assays.
261

262

263 **Table 2**. Comparison of methods used to measure growth of *Plasmodium* mutants. Parasite gene
264 deletions that were orthologous between all the methods (total of 53 orthologues in *P. berghei*)
265 were initially determined and a correlation test done to assess significance in association.

266

Method	**BarSeq (RGR)**	**BarSeq (MFC)**	***PiggyBac* (MFS)**	***Plasmo*GEM**	***Plasmo*GEM (RGR > 0.5)**
Arrayed (RGR)	0.64 (P < 0.001)	0.57 (P < 0.001)	0.25 (P = 0.02)	0.08 (P = 0.58)	0.39 (P = 0.01)
BarSeq (RGR)	1	0.93 (P < 0.001)	0.41 (P = 0.001)	0.27 (P = 0.05)	0.40 (P = 0.01)
BarSeq (MFC)	0.93 (P < 0.001)	1	0.41 (P = 0.001)	0.30 (P = 0.03)	0.41 (P = 0.01)
PiggyBac (MFS)	0.41 (P = 0.001)	0.41 (P = 0.001)	1	0.43 (P = 0.001)	0.30 (P = 0.06)
*Plasmo*GEM	0.27 (P = 0.05)	0.30 (P = 0.03)	0.43 (P = 0.001)	1	-

267

268 As an additional validation, completely new constructs were generated targeting three genes
269 (PfHECT1/PF3D7_0628100, PfnPrx/PF3D7_1027300, PF3D7_0318500) which resulted in a
270 slow growth phenotype in the initial RGR assessment. Growth estimates of these repeat deletion
271 lines were then determined using the arrayed approach (as described in Methods). Control
272 deletion lines PfRH3 (variant 1) and PfRH3 (variant 2) were included, as well as lines with
273 deletions in PfEBA181 and PfAP2-O which had a slow growth phenotype as shown in this study.
274 Growth phenotypes of these repeat deletion parasite lines (PfHECT1, PfnPrx, and
275 PF3D7_0318500) were all significantly lower than the two PfRH3 controls, consistent with
276 previous observations (shown in **Supplementary Figure 7**).

277

278 **Comparing parasite growth between static and shaking conditions**

279 *Plasmodium* blood-stage parasites encounter diverse environmental conditions *in vivo* such as
280 changes in blood pressure and volume, variation in circulating metabolites, variations in host RBC

281 types and ages, as well as different tissue such as the spleen, liver, bone marrow, all of which
282 might have an effect on parasite growth (Kho, Qotrunnada et al. 2021). Growth is therefore not
283 necessarily a fixed phenotype for a given strain. The advantage of BarSeq is that it provides a
284 scalable method that can be used to assess growth of multiple lines across many different
285 conditions.
286
287 To explore the utility of BarSeq for such studies, we first tested the effect of mechanical stimuli on
288 parasite growth by maintaining mixed pools in shaking and static conditions in parallel, and
289 generating RGR and MFC estimates from three biological replicates as described above. There
290 was a significant correlation between the biological replicates (Adjusted R-squared = 0.80, $P <$
291 0.001) with less variation in shaking conditions compared to static culture (shown in
292 **Supplementary Figure 8**). In shaking culture, a total of 44/65 (67.7%) lines had slow growth
293 phenotype whereas 21/65 (32.3%) lines were non-significant based on RGR estimates. No fast
294 growers were scored using RGR in shaking conditions (shown in **Fig 3a**). A higher proportion of
295 slow growth phenotypes was observed with the MFC method 58/65 (89.2%) whereas 7/65
296 (10.8%) lines were non-significant (shown in **Fig 3b**). There was a significant correlation between
297 RGR and MFC estimates from shaking culture ($r = 0.927$ $P < 0.0001$) as shown in **Supplementary**
298 **Figure 9**.
299
300 Formal comparisons of mean fold changes in growth rates in either shaking or static culture
301 (shown in Figure **4a**) were performed using cumulative fold changes for each deletion line, fitted
302 with a regression line using three independent replicates per cycle. This regression was then used
303 to generate an index we have referred to as the shaking vs static index (SSI_{GR}); deletion lines
304 that grow better in static conditions will have a negative SSI_{GR} index whereas deletion lines that
305 grow better in shaking conditions will have a positive SSI_{GR} index, and this can be exponentiated.

306 PfRH3 (variant 1) and PfRH3 (variant 2) were used as controls to estimate 95% confidence
307 interval boundaries for assignment of phenotypes, and again the other controls, PfPARE and
308 PfEBA165, fell within these boundaries **(Fig 4b).** A total of 1/65 (1.5%) deletion lines had a high
309 SSI_{GR}, 20/65 (30.8%) had non-significant SSI_{GR} and 44/65 (67.7%) had low SSI_{GR} estimates,
310 meaning they grew faster in static conditions. The most extreme lines were all deletions in the
311 Erythrocyte Binding Ligands (EBLs), notably PfEBA181 with high SSI_{GR} and both PfEBA140 and
312 EBA175 [Dd2] with low SSI_{GR}, suggesting that shaking significantly affects growth rates in these
313 invasion-related deletion lines and hence potentially modulates gene function of these ligands.
314

315 **Using arrayed and BarSeq approaches to quantitate DNA replication**

316 To further explore the potential for using BarSeq in more complex phenotyping screens, we
317 applied it to estimate DNA replication capacity of each strain. Variations in merozoite number
318 within a single schizont exist between *Plasmodium* species, strains and even within clonal
319 populations (Simon, Stürmer et al. 2021). Merozoite number can be estimated by using flow
320 cytometry to quantitate DNA content in late stage schizonts that have been arrested just before
321 egress using the Protein Kinase G inhibitor, Compound 2 [(Donald, Zhong et al. 2006); see
322 Methods]. As with growth assays, we assessed DNA content in the deletion lines using both an
323 arrayed and mixed pool/BarSeq approach, with the arrayed approach being carried out in four
324 batches.
325

326 Based on the arrayed approach (shown in **Fig 5a**), C2-arrested schizonts from a total of 8/70
327 (11.4%) lines had significantly higher DNA content than the PfRH3 controls, 18/70 (25.7) had
328 lower levels, and 44/70 (62.9%) were non-significantly different. The deletion lines with the lowest
329 DNA content were EBA175 [Dd2], PfACS9 and PARE [Dd2]. It has been previously reported that
330 wild type Dd2 makes an average of 18 daughter merozoites per schizont (Reilly, Wang et al.

331 2007); this was consistent with our observations of both PARE [Dd2] and wild type Dd2 DNA
332 content, both of which were consistently significantly lower than wild type NF54. High levels of
333 kurtosis, a parameter which describes the distribution of the data around the mean estimate, were
334 also observed for PARE [Dd2] and EBA175 [Dd2]. Deletion of BCKDH-E2 resulted in a much
335 broader spread of DNA content compared to other strains, which could also indicate a
336 dysregulation of replication. Several controls were included, and these overlapped with the non-
337 significant distribution of the RH3 controls. The distribution of the DNA content for the individual
338 deletion lines is summarised in **Supplementary Figure 10**.
339
340 BarSeq experiments were performed using two mixed pools of deletion lines, each containing the
341 same six controls - PfPARE, PfACS9, PfRH3 variant 1, PfRH3 variant 2, PfExportin-T
342 (PF3D7_1315900), PfEBA165, along with 17-28 other randomly selected deletion lines. The
343 distribution of the relative DNA fluorescence units of the arrested schizont mixed parasite pools
344 followed a normal distribution (**Supplementary Figure 11**). The mixed pool distribution was then
345 subdivided into quintiles, which were used as gates for flow sorting (shown in **Fig 5b**). DNA was
346 then extracted from each sorted population, and the number of barcodes for each line within each
347 quintile quantitated using BarSeq. Reads were standardized using relative fluorescence units to
348 determine the number of infected red cells per deletion line within each quintile (shown in
349 **Supplementary Figure 12**). The barcode distribution of each deletion lines was scored as either
350 low (if the deletion line had barcode content significantly lower than the PfRH3 control lines with
351 no overlapping confidence intervals), non-significant, or high (if the deletion line had barcode
352 content significantly higher than the PfRH3 control lines and with no overlapping confidence
353 intervals). In group 1, a total of 6/23 (26%) deletion lines had significantly lower barcode content
354 than the PfRH3 controls, 5/23 (21.7%) had higher barcode content and 12/23 (52.2%) were non-
355 significant (**Fig 5c**). In group 2, there were 11/34 (32.3%) low, 7/34 (20.6%) high and 16/34
356 (47.1%) non-significant deletion lines (**Fig 5d**).

357

358 A total of 53 lines were assayed using both the arrayed and BarSeq approach; Overall, there was
359 ~40% concordance between the phenotypes of the two approaches (shown in **Supplementary**
360 **Figure 13**), however the correlation between these two methods was higher for mutants with
361 estimated DNA content < 24 (r = 0.91 P < 0.0001) as shown in **Supplementary Figure 14**.
362 Several deletion lines such as PARE [Dd2], EBA175 [Dd2] and PfACS9 had concordant
363 phenotypes in both arrayed and pooled methods, suggesting a genetic basis for the regulation of
364 DNA content. There was also a consistent negative correlation between DNA content and growth
365 (shown in **Supplementary Figure 15**), in keeping with the known phenotypes of the Dd2 control
366 strain, which grows faster than NF54, but produces fewer merozoites.

367

368 **Using BarSeq to track progression of mutants through the intraerythrocytic developmental**
369 **cell cycle (IDC)**

370 Once inside RBCs, *P. falciparum* parasites undergo an intricate programme of growth and
371 replication through ring, trophozoite and schizont stages, leading to the formation of daughter
372 merozoites which then burst out and infect new RBCs. This intraerythrocytic developmental cycle
373 (IDC) is tightly regulated and its length varies both within and between *Plasmodium* species
374 (Reilly, Wang et al. 2007, Reilly Ayala, Wacker et al. 2010), but the genes that control it are only
375 imperfectly understood (Paul, Miliu et al. 2020). To apply BarSeq to this problem, we used flow
376 sorting based on RNA and DNA fluorescence to separate parasite pools into specific
377 developmental stages - RNA transcription peaks in trophozoite stages, whereas DNA replication
378 only occurs in late trophozoite and schizont stages (shown in **Fig 6a**). Pools of parasites were
379 tightly synchronised, sampled at multiple points over a 48h time course, and at each time point,
380 the parasite population was sorted into the same gates. The relative distribution of each deletion
381 line between gates was then quantitated using BarSeq.

382

383 Three independent runs were performed as described in more detail in the Methods section. The
384 first experiment had a pool of five deletion lines (PfACS9, PfExportin-T, PfChaperone
385 (PF3D7_0920100), PfASP and PfATEL1 (PF3D7_0212900) and was set up as the proof of
386 principle. The changes in barcode proportions between multiple gates sampled every 6h over a
387 48h time course are summarized in **Fig 6b**, and from these trends can be inferred. At 6h, all
388 deletion lines are within the earliest gate, however a bimodal distribution emerges at 12h with
389 PfACS9 and PfChaperone being more advanced in the cell cycle. By 18h, a trimodal distribution
390 is present with PfACS9 the most advanced strain, PfChaperone is intermediate, whereas
391 PfExportin-T, PfASP and PfATEL1 are slowest. By 36h, PfACS9 is further advanced in the cell
392 cycle whereas PfExportin-T is the slowest and by 48h, parasites have started re-invading new
393 RBCs with PfACS9 further advanced in the cell cycle followed by, PfChaperone, PfATEL1/ PfASP
394 and PfExportin-T in that order.
395
396 We explored two methods of quantitating the rate of transition of these deletion lines though the
397 IDC. The first approach following a 'before and after' approach is primarily designed to identify
398 differences in progression after the full time-course. At 48h, PfACS9 had the highest proportion
399 of reads in the early gates, indicating completion of one cycle and beginning of another, whereas
400 PfExportin-T had a higher frequency of reads in the late stages (shown in **Fig 6c**). A weighted
401 mean of the barcode counts was then calculated using ring stage parasites and showed that
402 PfACS9 was the quickest to advance whereas PfExportin-T was the slowest (shown in **Fig 6d**).
403 There was also a significant positive correlation between the weighted transition rate and growth
404 (as measured using the arrayed approach/RGR) (r = 0.98 P < 0.019), although PfASP was an
405 outlier. The second approach used a compartmental model to simulate the transition of reads
406 between gates per hour (shown in **Supplementary Figure 16**). Each compartment records the
407 proportion of sequencing reads observed at time t in that gate on the assumption that the rates of
408 transitions between the gates are constant for a particular deletion line. Results are shown in **Fig**

409 **6e.** As seen using the 'before and after' approach, PfACS9 transitions fastest between gate 3-4,
410 and gate 5-6 and with a fast transition in the late stages. PfASP transitions slowest in the earliest
411 stage (gate 1-3), whereas PfExportin-T generally transitions slowest. All the lines appeared to
412 transition slowest in gates 4-5, and gates 6-7. A mean transition rate from the modelled estimates
413 was also calculated (shown in **Fig 6f**) and again shows a positive correlation with arrayed growth
414 (RGR) ($r = 0.97$ $P < 0.032$), with ASP again an outlier. Both the weighted transition rate and the
415 mean transition modelled rate were positively correlated ($r = 0.96$ $P < 0.010$) as shown in **Fig 6g**.
416
417 The weighted approach was then used to determine the transition rates in two larger pools of
418 deletion lines (shown in **Fig 6h-i**). There was a positive correlation between the weighted
419 transition rate and arrayed growth, as well all RGR from BarSeq data (shown in **Supplementary**
420 **Figure 17-18**), and multiple deletion lines were consistent across all the IDC experiments such
421 as PfACS9 which transitions fastest through the IDC in all pools of deletion lines tested.
422

423 **Discussion**

424 We generated a medium-scale library of targeted *P. falciparum* knockout lines using a custom
425 CRISPR/Cas9. This is only the fourth targeted genetic study of this scale in the major human
426 malaria pathogen, comparable to previous studies of kinases (Solyakov, Halbert et al. 2011),
427 exported proteins (Maier, Rug et al. 2008) and unknown proteins on chromosome 3 (Kimmel,
428 Schmitt et al. 2023). Rather than targeting a specific group of genes, or genes on a specific
429 chromosome, as in previous studies, the gene list was selected to enrich for slow growth
430 phenotypes, based on a previous targeted screen in the rodent model *P. berghei* (Bushell, Gomes
431 et al. 2017, Zhang, Wang et al. 2018). Unlike previous studies however, we focussed on scaling
432 up phenotyping by using BarSeq to assess multiple parasites growing simultaneously in mixed
433 pools. BarSeq was applied to quantitate growth rates *in vitro,* analogous to the way it was applied
434 *in vivo* in *P. berghei,* but also in two completely novel ways – to quantitate DNA multiplication,
435 and to follow cell cycle transitions.

436

437 There was overall a positive correlation between parasite growth rates in our study and those in
438 the *Plasmo*GEM and *Piggy*Bac screens, even though the screens were carried out using very
439 different methodologies. 73.6% of the deletion lines in this study using BarSeq had concordant
440 growth phenotypes with orthologues in the *Plasmo*GEM screen, despite the screens being carried
441 out in different *Plasmodium* species. Discordant phenotypes could be the result of genuine
442 differences in biology between species but could also be driven by methodological differences. In
443 the *Plasmo*GEM screen the full ORF was replaced by a selectable marker, whereas in our *P.*
444 *falciparum* pipeline, the marker deleted only a section of the ORF, which could result in expression
445 of partial gene fragments that might provide partial function. We mitigated this by targeting the 5'-
446 end of the gene to introduce premature stop codons upstream of the sequence and hence
447 increase the probability of generating non-functional truncated genes, but it remains a possibility
448 that in some cases our disruptions were not complete. Similarly, potential differences could exist

449 between culture *in vivo* system as used in the *Plasmo*GEM screen and the *in vitro* system used
450 in this study. Several other large genome-scale screens have reported a higher abundance of
451 dispensable genes in *in vitro* culture and free-living organisms under laboratory conditions
452 (Winzeler, Shoemaker et al. 1999, Giaever, Chu et al. 2002, Sidik, Huet et al. 2016). There was
453 higher correlation in growth estimates with the *Piggy*Bac transposon data, which was carried out
454 *in vitro* in *P. falciparum,* so concordance would be expected to be higher. Discrepancies here
455 could be driven by inaccessibility of certain regions of the genome to TTAA insertion, incomplete
456 saturation or in other cases insertions at multiple sites within the genome, leading to complex
457 phenotypes.
458
459 Some of these differences may reveal new biology. For example, the *Plasmo*GEM screen only
460 identified a single knockout that resulted in faster growth – deletion of the AP2G locus which
461 inactivates gametocyte switching and therefore results in fewer gametocytes being produced,
462 which are a dead-end stage during intraerythrocytic development (Sinha, Hughes et al. 2014,
463 Bushell, Gomes et al. 2017). We observed consistently higher growth phenotypes of an ACS9
464 knockout line which has not been reported in other screens. Confirmation of an intact AP2G locus
465 in these ACS9 deletion line was further confirmed using whole genome sequence analyses
466 (shown in **Supplementary Figure 19**). ACS9 is part of a family of fatty acyl-CoA synthetases that
467 catalyse the formation of a thioester bond between fatty acids and coenzyme-A (CoA) forming
468 acyl-CoA, a precursor for phosphatidic acid which is key to phospholipid metabolism (Athenstaedt
469 and Daum 1999, Watkins 2008). Notably. Although *P. berghei* contains a single direct homologue
470 of ACS9, this ACS9 gene has further expanded through duplication in *P. falciparum* into a family
471 of 10 paralogs (Bethke, Zilversmit et al. 2006). This could possibly explain why it was not identified
472 as a fast growing mutant in the *Plasmo*GEM screen, and points to differences in phospholipid
473 metabolism between *Plasmodium* species.
474

475 Screens in other organisms have emphasised that gene function can depend heavily on the
476 growth environment, for example in yeast cells where ~80% of the genome is dispensable in rich-
477 medium but nearly all of these genes, totalling 97%, revealed a growth phenotype when the cells
478 were exposed to different and specific environments (Hillenmeyer, Fung et al. 2008). We took
479 advantage of the high-throughput nature of BarSeq phenotyping to quantitate the effect of
480 mechanical stimuli on parasite growth by culturing parasites in parallel in shaking and static
481 culture and comparing growth in these two conditions by developing an SSI index. Parasites are
482 normally grown in static culture; however, several publications have demonstrated improved
483 growth when parasites are cultured in shaking conditions, potentially due to increased random
484 distribution of merozoites leading to fewer erythrocytes being invaded by multiple merozoites, and
485 more equal distribution of nutrients and metabolites through the culture medium (Butcher 1979,
486 Butcher 1981, Zolg, MacLeod et al. 1982, Allen and Kirk 2010). Comparing the SSI index of wild
487 type parasite strains provided evidence for strain-specific effects, with NF54 parasite lines having
488 a positive index and hence growing better in shaking culture whereas Dd2 had a negative index
489 and grows better in static culture. All the control lines (RH3 variant 1, RH3 variant 2, EBA165,
490 PARE) had a similar phenotype to wild type NF54, whereas many of the gene deletions had a
491 slight negative SSI index meaning they had a growth advantage in static culture, suggesting that
492 the effect of specific genotypes on parasite survival could be modulated by specific environmental
493 factors, and that shaking culture can be a way to tease out more subtle growth phenotypes not
494 detected under static conditions. Deletion lines at the extreme end of this distribution were all
495 members of the EBL family, specifically EBA181, EBA140 and EBA175 [Dd2] (Wright and Rayner
496 2014, Cowman, Tonkin et al. 2017). All these ligands bind specific receptors on the RBC surface
497 and are thought to be involved in RBC recognition, deformation and re-orientation of the merozoite
498 during the process of invasion. Knockout of these genes in other studies has also been associated
499 with a reduction in parasite invasion consistent with our results (Weiss, Gilson et al. 2015, Koch,
500 Wright et al. 2017), but the more marked differences revealed in shaking conditions again suggest

501 that varying growth conditions can exacerbate phenotypes, and that in the case of these genes,
502 shaking conditions may influence gene function and better reproduce the situation *in vivo* where
503 merozoites and RBCs interact while exposed to blood flow.
504
505 We also used BarSeq for the first time to compare DNA content, and hence multiplication,
506 between strains. Differences in proliferation rates between strains Dd2 and HB3 have previously
507 in part been associated with differences in the number of merozoites that each produce (Reilly,
508 Wang et al. 2007), but this phenotype has never been explored at high scale. We measured the
509 DNA content of arrested late-stage schizonts from wild type and deletion lines in both an arrayed
510 layout using DNA staining and a pooled format which combined DNA staining with cell sorting
511 and BarSeq. We observed lower DNA content estimates for both PARE [Dd2] and EBA175 [Dd2]
512 than for wildtype NF54, but in line with estimates from (Reilly, Wang et al. 2007) who used
513 microscopy to count the distribution of merozoites within 50 schizonts of wild type Dd2. The high
514 throughput BarSeq approach by contrast allowed us to generate estimates based on counts of
515 >10,000 schizonts allowing reductions in standard errors of estimates and identification of more
516 subtle differences which could be significant in a large parasite biomass such as occurs *in vivo*.
517 Phenotype scores of DNA content varied from either low, non-significant or high and there was
518 overall a negative correlation between DNA content and growth rates, consistent with known
519 growth and merozoite number phenotypes for Dd2 relative to NF54/3D7. Certain deletions
520 resulted in low DNA content estimates for example ACS9, whereas deletions in other genes such
521 as GDH2 resulted in higher estimates, suggesting multiple genes may regulate DNA content. We
522 also assessed variability in DNA content using kurtosis of the distribution of the DNA content and
523 observed that PARE [Dd2] and EBA175 [Dd2] generally had high kurtosis whereas other deletion
524 lines such as BCKDH-E2 had broader variation in the final DNA content. All these might reflect
525 variations in the replication dynamics and regulation associated with parasite genotype. The
526 general consistency between replicates and between the two different methods, testing lines both

527 individually and in pools, emphasize the final DNA content in mature schizonts is relatively stable
528 and that under standard culture conditions is intrinsic to the parasite genetic background, as has
529 been previously suggested (Reilly, Wang et al. 2007). While there was generally consistency
530 between the arrayed and pooled approach, discrepancies were also apparent with certain genes
531 such as AP2-O and nPrx which had significantly higher estimates in the arrayed approach. The
532 arrayed approach uses Hoechst 33342 to directly measure DNA quantity, whereas the pooled
533 approach relies on BarSeq quantitation, which measured the relative distribution of barcodes
534 between sectors of the DNA content distribution, which could account for these differences or
535 biological factors such as cellular events of merogony.
536
537 We also used BarSeq to assess differences in cell-cycle dynamics by using Hoechst and Thiazole
538 Orange to measure DNA and RNA content and hence progression through the cell cycle. While
539 this dual labelling approach to follow intra-erythrocytic progression has been used before
540 (Grimberg, Erickson et al. 2008), in this case we combined it with cell sorting and BarSeq to
541 quantitate parasite progression between cell cycle phases. Several deletion lines had different
542 cycling rates, as measured by two different approaches, a simple start-to-end approach and a
543 more complex compartmental transition model. Deletion of some genes such as the ACS9 locus
544 consistently accelerated progression through the full-length cell-cycle whereas deletion of others
545 such as Exportin-T, putative increased the IDC duration. The resolution in cell-cycle tracking was
546 then increased by shortening the sampling window for a subset of deletion lines. ACS9 had a
547 shorter G1 and S/M phase whereas Exportin-T, putative had longer G1 and S/M phases. Overall,
548 there was a positive correlation between growth and the rate of completion of the cell-cycle.
549 Previous studies using forward genetics approaches to map QTLs associated with IDC duration
550 differences between Dd2 and HB3 strains identified potential candidates on multiple
551 chromosomes associated with regulation of cell cycle length (Reilly Ayala, Wacker et al. 2010),
552 which could further suggest that this trait is polygenic and that it is tightly regulated given the

553 conservation of transcriptomic profiles between geographically diverse *P. falciparum* lines (Llinas,
554 Bozdech et al. 2006). The BarSeq method developed here to track cell cycle dynamics of deletion
555 lines provides a novel approach to pinpoint gene function in specific cellular processes and could
556 be a tool in targeted genetic screens, design and identification of new drug molecules and
557 combinations with novel cellular targets.
558
559 We generally observed stable proliferation rates both in wild type and mutant parasite lines,
560 consistent with the fact that growth is a heritable quantitative trait. The underlying developmental
561 and cellular processes and how these are linked in regulating these growth differences have
562 however been poorly understood, although in some cases these have been positively correlated
563 with sub-phenotypes such as invasion efficiencies, IDC length and merozoite numbers. We
564 developed methods to quantify within-cycle processes of deletion lines using BarSeq. Using
565 simple linear regression, we observed a negative correlation between growth rates and DNA
566 content, and a positive correlation between growth rates and IDC duration. We then modelled the
567 combinatorial effect of DNA content and IDC progression on growth rates using either the support
568 vector regression (SVR) model or the Dense Neural Network (DNN) and observed a significant
569 correlation with growth rates. This could potentially suggest that for some strains, the decrease in
570 merozoite numbers is associated with a shorter cell cycle, which could therefore lead to faster
571 growth. Other sub-phenotypes such as invasion efficiencies might also contribute to differences
572 in overall growth phenotypes and are linked tightly with other sub-phenotypes.
573
574 In summary, we have developed new methods to generate and phenotype a significant set of *P.*
575 *falciparum* deletion lines in a scalable manner using both arrayed and pooled assays. These
576 methods are complementary, with concordant phenotypes from several comparisons, and
577 therefore constitute a framework to pinpoint with greater accuracy the biological processes that
578 specific genes are involved in. To our knowledge BarSeq has never been applied to phenotype

579 either DNA quantity pr cell cycle transition in *Plasmodium* parasites. The ability to use the same
580 barcoded deletion lines in multiple phenotypic approaches argues strongly for the inclusion of
581 barcodes in all future genetic modification experiments, ideally using a common barcoding module
582 to allow for the accumulation of pools across multiple different labs and research groups. Given
583 the difficulty and slow pace of generating modified *P. falciparum* lines, even using CRISPR/Cas9
584 approaches, designing all generated lines to be as useable and flexible as possible, rather than
585 generating them only for a single study, is clearly of the greatest benefit to the malaria research
586 community. The ability to create mutations for 1% of the *P. falciparum* genome in a single lab in
587 this manner also suggests that community-scale whole genome targeted studies should be
588 considered, with appropriate funding support.
589

590 **Acknowledgements**

591 Funding: This work was supported by Wellcome, grant numbers 220266/Z/20/Z (JR) and
592 206194/Z/17/Z (Sanger). For the purpose of open access, the authors have applied a CC BY
593 public copyright license to any Author Accepted Manuscript version arising from this submission.
594 The funders had no role in study design, data collection and analysis, decision to publish or
595 preparation of the manuscript.
596

597 We thank Emily Mallet, Emma Goffin, David Adam Jackson (Sanger CGaP) , Christopher Hall
598 (Sanger core), and Sean Wright and Tristram Bellerby (Sanger DNA Pipelines) for technical
599 support, and Theresa Feltwell, Nadia Cross and the CIMR core support teams for their invaluable
600 logistical support.
601

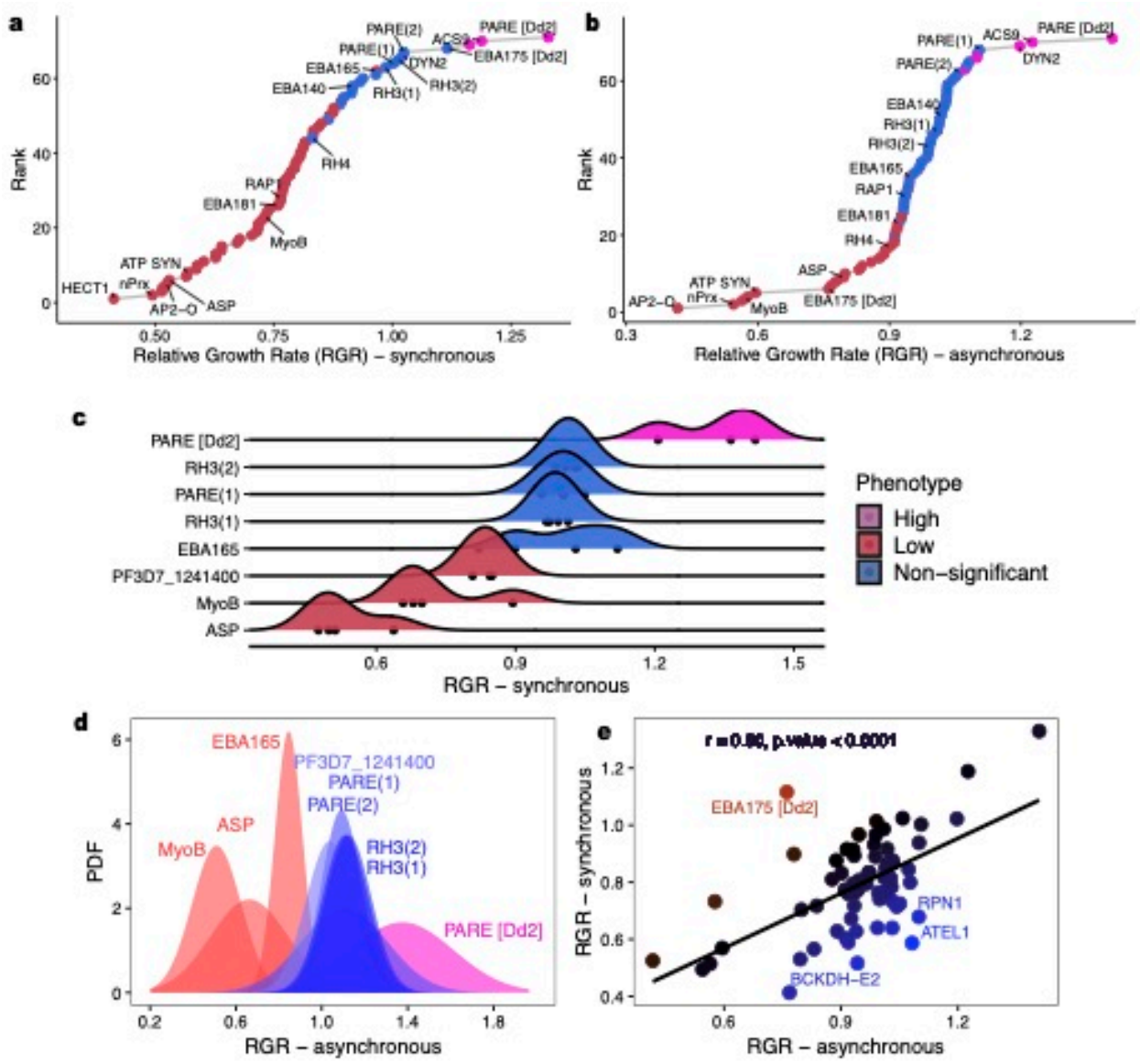
602 **Author contribution**

603 Conceptualization, A.M., O.B. and J.C.R.; methodology, A.M, M.G., T.S., W.W.; data and
604 resource generation, A.M., M.G., M.I., R.A., S.H., R.S., G.G., F.S; validation, A.M.; formal

605 analysis, A.M., and J.C.R.; writing – original draft, A.M., and J.C.R.; writing – review & editing,
606 A.M., W.W., O.B., and J.C.R.; visualization, A.M., and T.S.; supervision, E.B., C.B., O.B. and
607 J.C.R.; funding acquisition, O.B. and J.C.R. All authors read and approved the manuscript.

608

609



610

611

612 **Figure 1. Growth phenotypes of *Plasmodium falciparum* deletion lines in static culture**
613 **using arrayed approaches.** *P. falciparum* lines cultured in static culture starting with either
614 synchronous culture (a) or asynchronous culture (b). Parasite proliferation was assessed using
615 flow cytometry. Significance of the individual gene deletions on parasite growth were then
616 estimated by calculating 95% confidence intervals for the mean relative growth rates. PfRH3

617 (variant 1) and PfRH3 (variant 2) were used to determine the boundaries of significance. Several
618 controls were used for both the synchronous approach (c) and asynchronous method (d). There
619 was a significant positive correlation between both synchronous and asynchronous approaches
620 (e).
621

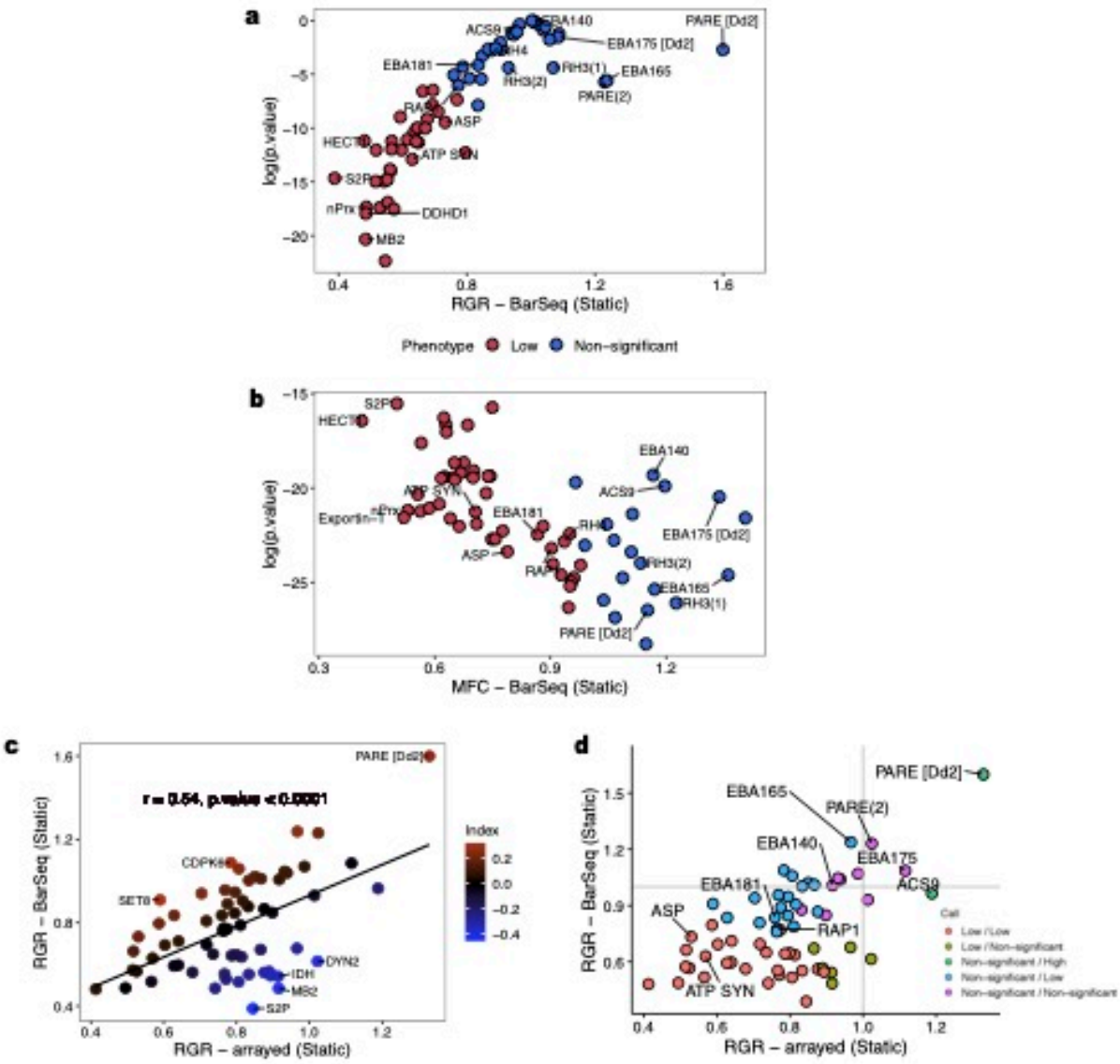
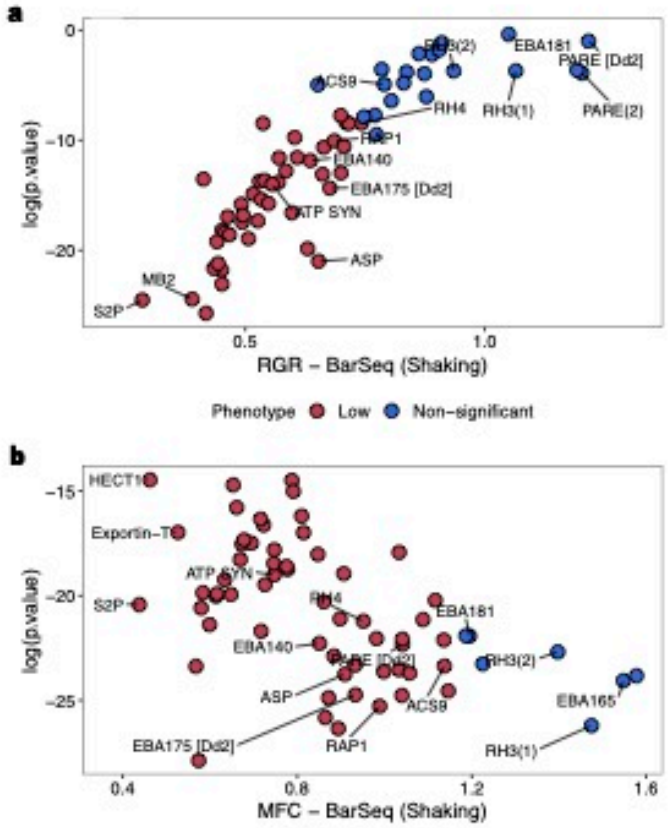


Figure 2. Growth phenotypes of *Plasmodium falciparum* deletion lines in static culture using pooled approaches and BarSeq. *P. falciparum* lines were synchronized, pooled and sampled for growth in static conditions. Parasite proliferation was assessed using barcode quantitation and next generation sequencing. Significance of the individual gene deletions on parasite growth were then estimated by calculating 95% confidence intervals for either the mean relative growth rate (a) or mean fold change estimate (b). PfRH3 (variant 1) and PfRH3 (variant

631 2) were used to determine the boundaries of significance. There was a significant positive
632 correlation between arrayed and pooled growth approaches (c-d).
633

634



635

636 **Figure 3. Effect of shaking on parasite growth.** *P. falciparum* lines were synchronized, pooled
637 and sampled for growth in shaking conditions. Parasite proliferation was assessed using barcode
638 quantitation and next generation sequencing. Significance of the individual gene deletions on
639 parasite growth were then estimated by calculating 95% confidence intervals for either a mean
640 relative growth rate (a) or mean fold change estimate (b).

641

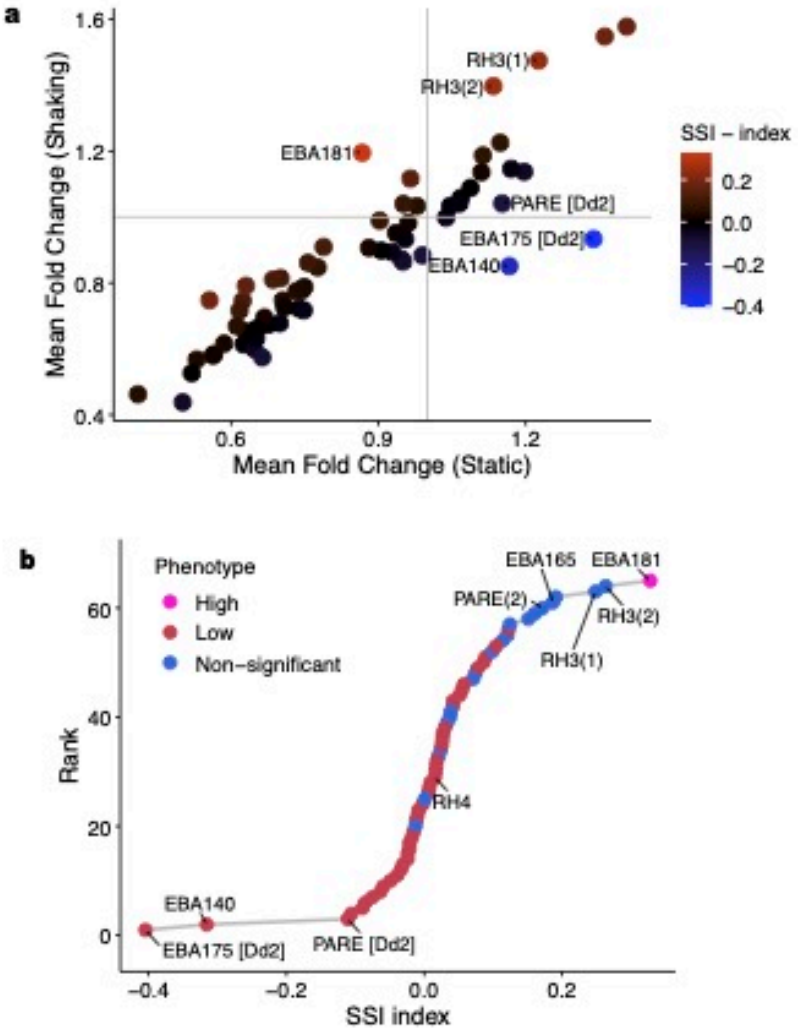


Figure 4. Comparison of shaking and static culture on parasite growth. Growth of *P. falciparum* deletion lines in shaking and static cultures were also compared in parallel (a). Differences in parasite growth were estimated using the shaking vs static index (SSI_{GR}). Control lines PfRH3 (variant 1) and PfRH3 (variant 2) were used to determine boundaries for significance. Phenotypes were scored as either low, non-significant or high (b).

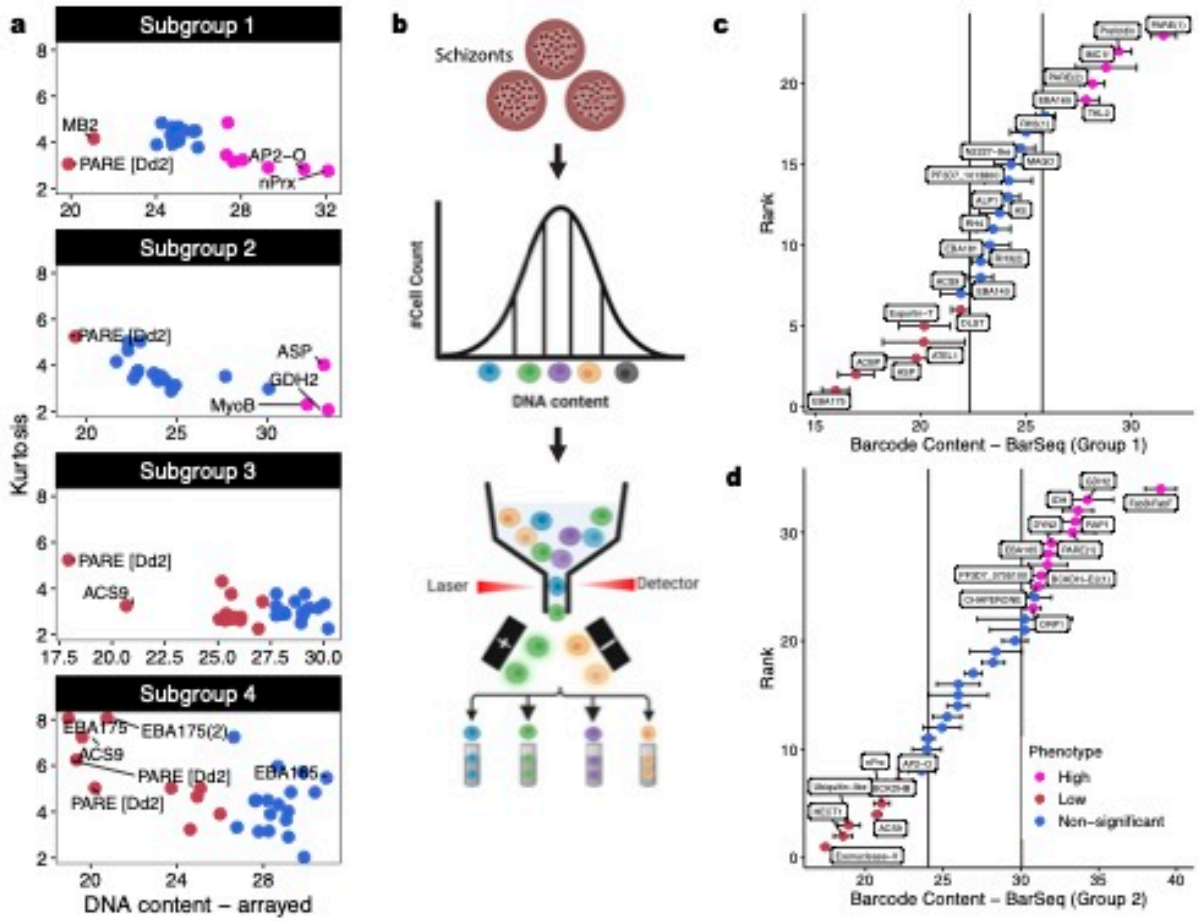


Figure 5. Estimates of DNA content within mature schizonts of *Plasmodium falciparum* deletion lines using arrayed, flow sorting and BarSeq approaches. Parasite DNA and barcode content was quantitated using either an arrayed or mixed pool BarSeq approach. For the arrayed approach, mature schizonts were stained with Hoechst 33342 and quantitated using flow cytometry (a). The PfRH3 lines were used as controls to determine boundaries of significance with phenotypes scores either as low, non-significant or high. For the mixed pool method, mature stage parasites were acquired on a cell sorter and their distribution subdivided into quintiles (b). Schizont populations from each window were then sorted and quantitated using next generation

659 sequencing. A frequency distribution was used to calculate the average barcode content for each
660 deletion line (c-d). The PfRH3 lines were used as controls to determine boundaries of significance
661 with phenotypes scores either as low, non-significant or high.
662

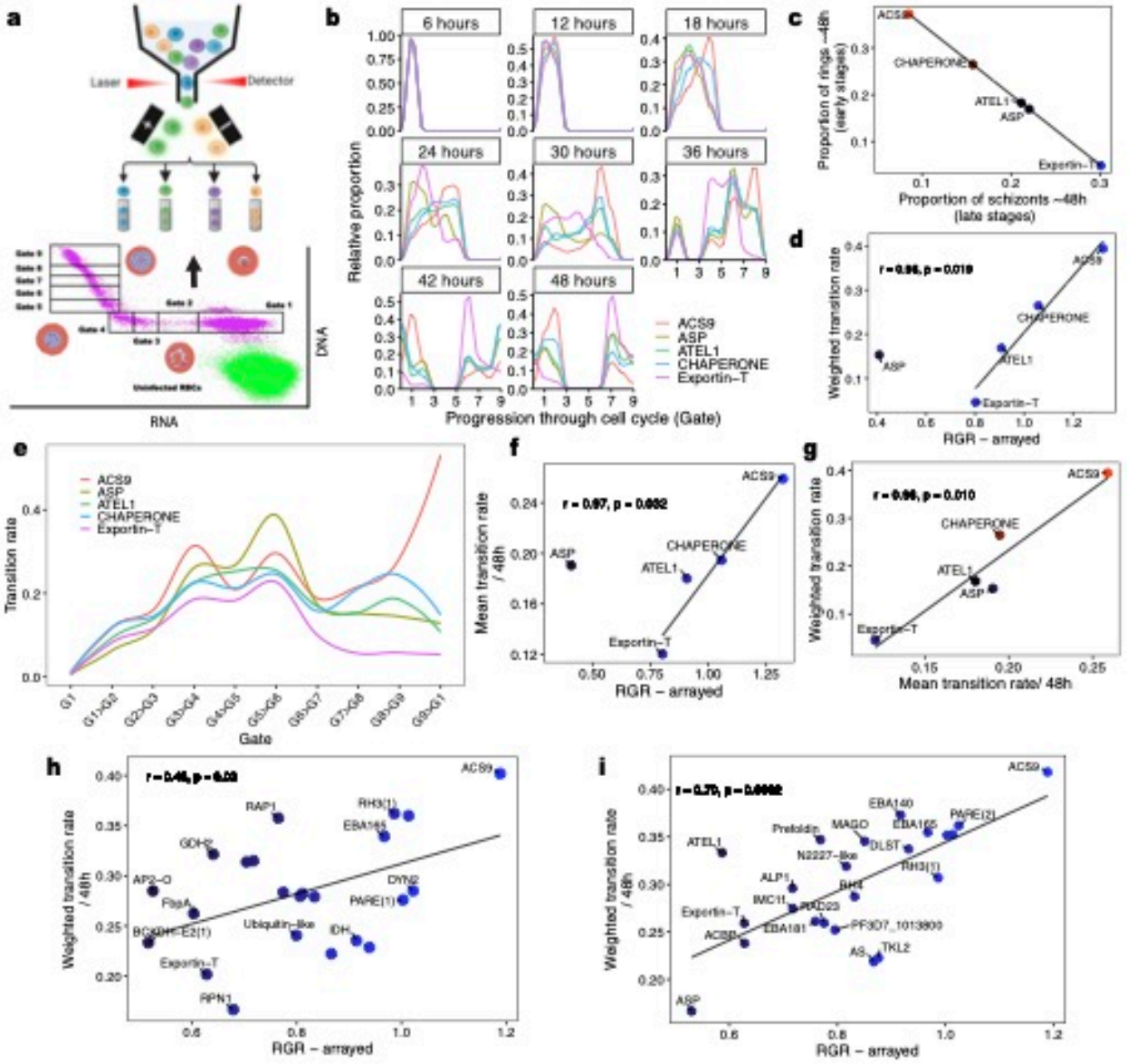


Figure 6. Intraerythrocytic developmental cell cycle dynamics of *Plasmodium falciparum* deletion lines determined using flow sorting and BarSeq. Synchronous parasite cultures were pooled and sorted based on their RNA/DNA content using a BD influx cell sorter (a). The relative proportion of reads of each deletion line was then assessed over a 48h time course (b). The rate of transition was then determined using a weighted approach (c-d). A second approach using a

compartmental model was also used to determine the transition rate per hour between each of the gates (e). A mean transition across the cycle was also determined from this model and is positively correlated with growth (f). Both the weighted approach as well as the modelling approach had a positive correlation (g). Similar weighted transition rates were also observed for the larger pools, and these were all significantly correlated with growth and associated with genetic background (h-i).

679 **References**

680

681 Allen, R. J. and K. Kirk (2010). "Plasmodium falciparum culture: the benefits of shaking."
682 Molecular and biochemical parasitology **169**(1): 63-65.

683

684 Athenstaedt, K. and G. Daum (1999). "Phosphatidic acid, a key intermediate in lipid
685 metabolism." European journal of biochemistry **266**(1): 1-16.

686

687 Bethke, L. L., et al. (2006). "Duplication, gene conversion, and genetic diversity in the species-
688 specific acyl-CoA synthetase gene family of Plasmodium falciparum." Molecular and
689 biochemical parasitology **150**(1): 10-24.

690

691 Bushell, E., et al. (2017). "Functional profiling of a Plasmodium genome reveals an abundance
692 of essential genes." Cell **170**(2): 260-272. e268.

693

694 Butcher, G. (1979). "Factors affecting the in vitro culture of Plasmodium falciparum and
695 Plasmodium knowlesi." Bulletin of the World Health Organization **57**: 17-26.

696

697 Butcher, G. (1981). "A comparison of static thin layer and suspension cultures for the
698 maintenance in vitro of Plasmodium falciparum." Annals of Tropical Medicine & Parasitology
699 **75**(1): 7-17.

700

701 Collins, C. R., et al. (2013). "Robust inducible cre recombinase activity in the human malaria
702 parasite p lasmodium falciparum enables efficient gene deletion within a single asexual
703 erythrocytic growth cycle." Molecular microbiology **88**(4): 687-701.

704

705 Combe, A., et al. (2009). "Clonal conditional mutagenesis in malaria parasites." Cell host &
706 microbe **5**(4): 386-396.
707
708 Cowman, A. F., et al. (2017). "The molecular basis of erythrocyte invasion by malaria parasites."
709 Cell host & microbe **22**(2): 232-245.
710
711 de Koning-Ward, T. F., et al. (2015). "Advances in molecular genetic systems in malaria."
712 Nature Reviews Microbiology **13**(6): 373-387.
713
714 Donald, R. G., et al. (2006). "Anticoccidial kinase inhibitors: identification of protein kinase
715 targets secondary to cGMP-dependent protein kinase." Molecular and biochemical parasitology
716 **149**(1): 86-98.
717
718 Donnelly, M. J., et al. (2016). "Identification, validation, and application of molecular diagnostics
719 for insecticide resistance in malaria vectors." Trends in parasitology **32**(3): 197-206.
720
721 Gardner, M. J., et al. (2002). "Genome sequence of the human malaria parasite Plasmodium
722 falciparum." Nature **419**(6906): 498-511.
723
724 Ghorbal, M., et al. (2014). "Genome editing in the human malaria parasite Plasmodium
725 falciparum using the CRISPR-Cas9 system." Nature biotechnology **32**(8): 819-821.
726
727 Giaever, G., et al. (2002). "Functional profiling of the Saccharomyces cerevisiae genome."
728 Nature **418**(6896): 387-391.
729

730 Gibson, D. G., et al. (2008). "Complete chemical synthesis, assembly, and cloning of a
731 Mycoplasma genitalium genome." Science **319**(5867): 1215-1220.
732
733 Grimberg, B. T., et al. (2008). "Monitoring Plasmodium falciparum growth and development by
734 UV flow cytometry using an optimized Hoechst-thiazole orange staining strategy." Cytometry
735 Part A: The Journal of the International Society for Analytical Cytology **73**(6): 546-554.
736
737 Hillenmeyer, M. E., et al. (2008). "The chemical genomic portrait of yeast: uncovering a
738 phenotype for all genes." Science **320**(5874): 362-365.
739
740 Hoshizaki, J., et al. (2022). "Efficient generation of mNeonGreen Plasmodium falciparum
741 reporter lines enables quantitative fitness analysis." Frontiers in cellular and infection
742 microbiology: 1408.
743
744 Istvan, E. S., et al. (2017). "Esterase mutation is a mechanism of resistance to antimalarial
745 compounds." Nature communications **8**(1): 1-8.
746
747 Kho, S., et al. (2021). "Hidden biomass of intact malaria parasites in the human spleen." New
748 England Journal of Medicine **384**(21): 2067-2069.
749
750 Kimmel, J., et al. (2023). "Gene-by-gene screen of the unknown proteins encoded on
751 Plasmodium falciparum chromosome 3." Cell Systems **14**(1): 9-23. e27.
752
753 Koch, M., et al. (2017). "Plasmodium falciparum erythrocyte-binding antigen 175 triggers a
754 biophysical change in the red blood cell that facilitates invasion." Proceedings of the National
755 Academy of Sciences **114**(16): 4225-4230.

Lee, M., et al. (2019). Cutting back Malaria: CRISPR/Cas9 genome editing of Plasmodium. Brief Funct Genomics 18: 281–289.

Llinas, M., et al. (2006). "Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains." Nucleic Acids Research **34**(4): 1166-1173.

Maier, A. G., et al. (2008). "Exported proteins required for virulence and rigidity of Plasmodium falciparum-infected human erythrocytes." Cell **134**(1): 48-61.

Murray, L., et al. (2017). "Multiplication rate variation in the human malaria parasite Plasmodium falciparum." Scientific reports **7**(1): 6436.

Oberstaller, J., et al. (2021). "Essential genes of the parasitic apicomplexa." Trends in parasitology **37**(4): 304-316.

Paul, A. S., et al. (2020). "Co-option of Plasmodium falciparum PP1 for egress from host erythrocytes." Nature communications **11**(1): 3532.

Proto, W. R., et al. (2019). "Adaptation of Plasmodium falciparum to humans involved the loss of an ape-specific erythrocyte invasion ligand." Nature communications **10**(1): 4512.

Reilly Ayala, H. B., et al. (2010). "Quantitative trait loci mapping reveals candidate pathways regulating cell cycle duration in Plasmodium falciparum." BMC genomics **11**: 1-14.

781 Reilly Ayala, H. B., et al. (2010). "Quantitative trait loci mapping reveals candidate pathways
782 regulating cell cycle duration in Plasmodium falciparum." BMC genomics **11**(1): 1-14.
783
784 Reilly, H. B., et al. (2007). "Quantitative dissection of clone-specific growth rates in cultured
785 malaria parasites." International Journal for Parasitology **37**(14): 1599-1607.
786
787 Russell, A. J., et al. (2023). "Regulators of male and female sexual development are critical for
788 the transmission of a malaria parasite." Cell host & microbe **31**(2): 305-319. e310.
789
790 Sidik, S. M., et al. (2016). "A genome-wide CRISPR screen in Toxoplasma identifies essential
791 apicomplexan genes." Cell **166**(6): 1423-1435. e1412.
792
793 Sinha, A., et al. (2014). "A cascade of DNA-binding proteins for sexual commitment and
794 development in Plasmodium." Nature **507**(7491): 253-257.
795
796 Solyakov, L., et al. (2011). "Global kinomic and phospho-proteomic analyses of the human
797 malaria parasite Plasmodium falciparum." Nature communications **2**(1): 565.
798
799 Stanway, R. R., et al. (2019). "Genome-scale identification of essential metabolic processes for
800 targeting the Plasmodium liver stage." Cell **179**(5): 1112-1128. e1126.
801
802 Straimer, J., et al. (2012). "Site-specific genome editing in Plasmodium falciparum using
803 engineered zinc-finger nucleases." Nature methods **9**(10): 993-998.
804

805 Taylor, H. M., et al. (2001). "Plasmodium falciparum homologue of the genes for Plasmodium
806 vivax and Plasmodium yoelii adhesive proteins, which is transcribed but not translated."
807 Infection and immunity **69**(6): 3635-3645.
808
809 Triglia, T., et al. (2001). "An EBA175 homologue which is transcribed but not translated in
810 erythrocytic stages of Plasmodium falciparum." Molecular and biochemical parasitology **116**(1):
811 55-63.
812
813 Van Dijk, M., et al. (1995). "Stable transfection of malaria parasite blood stages." Science
814 **268**(5215): 1358-1362.
815
816 Wacker, M. A., et al. (2012). "Quantification of multiple infections of Plasmodium falciparum in
817 vitro." Malaria journal **11**: 1-9.
818
819 Watkins, P. A. (2008). "Very-long-chain acyl-CoA synthetases." Journal of Biological Chemistry
820 **283**(4): 1773-1777.
821
822 Weiss, G. E., et al. (2015). "Revealing the sequence and resulting cellular morphology of
823 receptor-ligand interactions during Plasmodium falciparum invasion of erythrocytes." PLoS
824 pathogens **11**(2): e1004670.
825
826 WHO (2022). World malaria report 2022, World Health Organization.
827
828 Winzeler, E. A., et al. (1999). "Functional characterization of the S. cerevisiae genome by gene
829 deletion and parallel analysis." Science **285**(5429): 901-906.
830

831 Wright, G. J. and J. C. Rayner (2014). "Plasmodium falciparum erythrocyte invasion: combining
832 function with immune evasion." PLoS pathogens **10**(3): e1003943.
833
834 Wu, Y., et al. (1995). "Transfection of Plasmodium falciparum within human red blood cells."
835 Proceedings of the National Academy of Sciences **92**(4): 973-977.
836
837 Zhang, M., et al. (2018). "Uncovering the essential genes of the human malaria parasite
838 Plasmodium falciparum by saturation mutagenesis." Science **360**(6388).
839
840 Zolg, J. W., et al. (1982). "Plasmodium falciparum: modifications of the in vitro culture conditions
841 improving parasitic yields." The Journal of Parasitology: 1072-1080.
842
843
844

845 **Methods**

846 **Creation of deletion plasmid constructs**

847 An in-house automated bioinformatics pipeline was developed to generate designs of deletion
848 constructs at scale. Two starter plasmids, pDC2-cam-coCas9-U6-yDHOODH and pCC1-hDHFR-
849 yFCU were used as a base for the gRNA and homology repair templates respectively (Maier, Rug
850 et al. 2008).

851

852 Prior to cloning the target constructs, frozen stocks of both pDC2-cam-coCas9-U6-yDHOODH
853 and pCC1-hDHFR-yFCU wild-type plasmids in XL-10 competent cells (Agilent) were initially
854 expanded by inoculation in 200mL LB broth supplemented with 100ug/mL Ampicillin and grown
855 for 18h at 37°C overnight under shaking conditions at 200rpm. This bacterial culture was then
856 pelleted by centrifugation at 4000rpm for 10min. Plasmid DNA was purified using a midi prep kit
857 (Macherey-Nagel) as per the manufacturer's instructions.

858

859 To generate donor template constructs, primers were used to amplify parasite DNA fragments
860 (approx. 1200bp) flanking the cleavage target site at both the 5' and 3' ends. Short DNA adapters
861 complementary to digested plasmids were also attached to these primers during the design
862 process. This generates two fragments referred to hereafter as H5 and H3 homology regions. The
863 amplicons were run through agarose gels to ensure that the right product sizes were obtained,
864 and PCR amplicons were purified using the Min-elute 96 UF PCR purification kit (Qiagen)
865 following the manufacturer's instructions.

866

867 pCC1-hDHFR-yFCU base plasmids were digested using SacII and EcoRI restriction enzymes in
868 preparation for Gibson Assembly (Gibson, Benders et al. 2008). Confirmation of successful
869 plasmid digestion was done using agarose gel electrophoresis. The digested reaction mix was

870 then purified prior to Gibson Assembly using a Zymo DNA clean and concentrator kit (Zymo
871 Research) following the manufacturer's instructions. Purified samples were quantitated using a
872 nanodrop to determine DNA concentrations. A total of 0.2 to 1pmol of assembly products in a
873 ratio between 1:1 and 1:5 of inserts to backbone was used for the Gibson assembly following the
874 manufacturer's instructions.
875
876 The assembled constructs were then cloned using XL-10 chemically competent cells (Agilent),
877 and plasmid DNA was then purified using a commercial kit (Macherey-Nagel) followed by capillary
878 sequencing. Capillary sequence files were trimmed using Benchling software and sequence
879 alignments to confirm correct DNA fragment insertions were generated using algorithms such as
880 ClustalOmega. Positive colonies were expanded in 200mL LB broth supplemented with 100ug/uL
881 Ampicillin and incubated for 18h at 37°C overnight under shaking conditions at 200rpm. Plasmid
882 DNA was then purified using a midi prep kit (Macherey-Nagel) in preparation for transfections.
883 The cut pDC2-cam-coCas9-U6-yDHOODH was then dephosphorylated to avoid re-ligation in
884 subsequent enzymatic steps. The gRNA sequences were pre-phosphorylated with adaptor
885 sequences complementary to cut pDC2-cam-coCas9-U6-yDHOODH that were then used for the
886 ligation step. Cloning was done using XL-10 chemically competent cells (Agilent) followed by
887 capillary sequencing to identify positive clones from which plasmid DNA was purified in
888 preparation for transfections using a midi prep kit (Macherey-Nagel).
889
890 ***P. falciparum* transfections**
891 The purified targeting constructs were initially precipitated to generate a concentrated mixture that
892 was also free from contaminating salts. To do this, a total of 60ug assembled pCC1 and 30ug
893 assembled pDC2- gRNA plasmids for each target gene were pooled and mixed with 1/10 volume
894 of 3M sodium acetate of the total plasmid volume. Samples were mixed by flicking the tube 4-5

895 times. 1mL of cold 100% Ethanol was then added to the sample and mixed by inverting 10 times
896 followed by centrifugation at 15000rpm for 30min at 5°C. The supernatant was then discarded
897 followed by addition of 1mL of cold 70% Ethanol and centrifugation at 15000rpm for 30min at 5°C.
898 The supernatant was then discarded, and the sample further pelleted at 15000rpm for 30min at
899 5°C. The supernatant was then discarded, removing as much ethanol as possible without
900 disturbing the pellet, and thereafter allowing the pellet to dry until it was translucent. The plasmid
901 DNA was then resuspended in 50uL Cytomix solution ready for transfection.
902
903 Transfections were performed using ring-stage parasites. Cultures of *P. falciparum* NF54 parasite
904 line at 4-5% parasitaemia were initially washed in warm 30 mL Cytomix per 2mL of packed iRBCs
905 at 1500rpm for 5min and the supernatant discarded. A total of 300uL warm Cytomix was then
906 added per 150uL of iRBC pellet. The resuspended DNA was then mixed with 450uL of the
907 iRBC/Cytomix mixture and transferred to a 0.2cm Bio-Rad cuvette and thereafter electroporated
908 using a BioRad electroporator (Gene Pulser Xcell, model number 1652660) using the following
909 settings: Exponential mode, 310V Resistance: infinite, 950μF Cuvette: 2mm. After electrorotation,
910 samples were then taken out of the cuvette and transferred to a 5mL culture 6-well plate, mixed
911 with fresh RBCs at 2% hematocrit and incubated at 37°C ready for the next day.
912
913 **Post transfection parasite culture**
914 The next steps after transfection were to maintain and keep the cultures on drug selection as
915 needed until the desired mutants could be detected by microscopy. Positive selection was
916 performed using 2.5nM WR99210 (Jacobus) and this was initially added to the cultures one day
917 post the transfection. Smears were then made on day-2 post transfection to check for parasite
918 death. Drug media was changed every day for the first six days post transfection followed by a 2-
919 3-day cycle per week. Fresh uninfected RBCs were added every week post drug selection by

920 discarding 1/3 of old culture and replacing this with an equal amount of fresh RBCs. Smears were
921 performed twice a week until parasites appeared. The time taken for parasites to be detected
922 under WR99210 selection varied from 8-21 days. Positive cultures were then frozen down until
923 rounds of transfections had been completed, thus generating an arrayed library of resistant lines.
924 Once the list of transfections had been completed, frozen lines were then thawed out and grown
925 in 2.5nM WR99210 (Jacobus), followed by culture in a combination of both negative and positive
926 selection using a mixture of 7mM 5-flurocytosine [5-FC] (Sigma) and 2.5nM WR99210 (Jacobus)
927 respectively for three cycles. Parasite were grown for at least three cycles in complete medium
928 supplemented with 7mM 5-flurocytosine [5-FC] and 2.5nM WR99210 to select ideally for double-
929 crossover recombinants. Resistant parasites were then expanded out and pellets directly frozen
930 at either -20°C for genotyping or in glycerolyte stocks at -80°C for future recovery and use.
931

932 **Genotyping of deletion parasites**

933 Transgenic parasites were genotyped to confirm the desired gene disruptions. Capillary
934 sequencing using a conserved primer present in the cassette of all the targeting constructs was
935 performed to verify the identity of the barcode module that the parasites were carrying. PCRs
936 were also performed to assess gene disruptions by agarose gel electrophoresis. Primers were
937 designed to exclusively amplify either the wild-type or the transgenic locus and therefore
938 distinguish modified and unmodified loci. A complete list of all primers used both in plasmid
939 construction and genotyping is shown in **Supplementary Table 3**.
940

941 **BarSeq**

942 Barcode Sequencing (BarSeq) was used in multiple assays to test relative growth rates, DNA
943 content and cell cycle dynamics of deletion strains. In these experiments, genomic DNA was
944 extracted from parasite DNA pellets using the Qiagen DNeasy Blood & Tissue Kit (Qiagen),

945 following the manufacturer's instructions. A nested PCR was performed to generate the BarSeq
946 sequencing libraries.
947
948 For the first PCR which attaches the adaptor overhangs, each reaction mix had a total of 25uL,
949 containing, 0.2uM of the forward and reverse primers (sequences), distilled water, 2uL of DNA
950 template, and 1X of 12.5uL CloneAmp containing both the DNA polymerase enzyme and dNTPs.
951 Samples were amplified at an initial denaturation temperature of 95°C for 5min, followed by 30
952 cycles of denaturation at 95°C for 30s, then annealing at 54°C for 30 s, and extension at 72°C
953 for 3s. Samples were finally extended for 5min at 68°C.
954
955 For the second PCR which attaches the index barcodes needed for cluster generation and
956 demultiplexing, samples were amplified using index primers from Illumina that are unique to each
957 experimental condition. Each reaction volume contained a total 0.2uM of both the forward and
958 reverse index primers, 2uL of neat amplicon product from the first PCR, 12.5uL of CloneAmp
959 containing both the DNA polymerase enzyme and dNTPs, and water to make up a total volume
960 of 25uL. Thermocycling conditions were set at an initial temperature denaturation of 95°C for
961 2min, followed by 12 cycles of denaturation at 95°C for 30s, then annealing at 68°C for 15s, and
962 extension at 68°C for 8s. Samples were finally extended for 5min at 68°C.
963
964 Prior to next generation sequencing, libraries were initially cleaned using a Qiagen MiniElute UF
965 PCR purification kit (Qiagen) as per the manufacturer's instructions. The concentration of each of
966 the purified samples was initially measured using the Quant-iT PicoGreen dsDNA reagent
967 (Thermo Fisher Scientific) or Qubit. 125ng of each sample was then pooled and re-measured
968 using the Qubit fluorometric broad sensitivity quantitation kit (Thermo Fisher Scientific). The DNA
969 pool was then diluted to a concentration of 4nM and sequenced at the Wellcome Sanger. Libraries

970 were run using a Miseq Reagent Kit v2 Paired-End Reads 2 x 150bp with the cluster generation
971 and sequencing all performed using a Miseq system, which generated 24-30 million reads
972
973 **Growth measurements using arrayed and pooled approaches**
974 For the arrayed approach, the lines were divided into four subgroups of 22-27 lines each, which
975 were then assayed sequentially to each other. This approach was the only one that was
976 technically feasible, and took a period of three months, but had the risk of increasing variation in
977 growth phenotypes due to batch effects. To allow for comparison between experiments, eight
978 controls - PfRH3 (variant 1), PfRH3 (variant 2), PfPARE, PfRNA binding (PF3D7_1241400),
979 PfASP (PF3D7_0405900), PfPARE [Dd2], PfMyoB (PF3D7_0503600) and PfEBA165, were
980 included as much as possible in each of the four subgroups. To allow growth to be compared
981 between groups and both the arrayed and pooled approaches, growth was expressed relative to
982 an average of the two PfRH3 control lines and these were also used to assign qualitative growth
983 phenotypes using confidence intervals.
984
985 Growth assays for the arrayed approach were performed starting with either asynchronous or
986 synchronous cultures. In the latter, parasites were enriched using LD MACS columns. To do this,
987 the LD MACS columns were placed on a magnetic stand and equilibrated with 3mL RPMI, and
988 thereafter loaded with 4mL of parasites at 50% haematocrit. Once the flow through with iRBCs
989 was complete, the columns were washed 2X using 4mL RPMI and thereafter eluted off the magnet
990 in 4mL RPMI. The eluate at this point was now enriched with late-stage parasites which were then
991 allowed to re-invade fresh uninfected RBCs. Experiments were then seeded the following day,
992 with parasite cultures that were predominantly (>90%) ring stages. The asynchronous stages
993 were seeded with variable proportions of rings, trophozoites and schizont stage parasites.
994

995 Mixed pools of parasite deletion lines were maintained in either static or shaking culture at 60rpm
996 on an INFORS HT bench-top shaker and cultured in the same incubator. A simple index to
997 formally assess growth phenotypes between shaking and static conditions for each deletion line
998 was generated and referred to as the shaking vs static index (SSI). The RH3 controls were used
999 as baseline and phenotypes were scored as low if the mean SSI_{GR} estimate of the deletion lines
1000 was lower than the PfRH3 controls and with non-overlapping confidence intervals, non-significant
1001 if the deletion lines had overlapping confidence intervals with the PfRH3 controls or high if the
1002 mean SSI_{GR} estimate of the deletion lines was higher than the PfRH3 controls and with non-
1003 overlapping confidence intervals.
1004
1005 **DNA content measurements**
1006 Parasites at mid-schizont stages were treated with Compound 2 drug to block egress of mature
1007 segmented schizont stages (Donald, Zhong et al. 2006), stained with 4uM Hoechst 33342 and
1008 staining intensity acquired on a Fortessa (BD Biosciences) flow cytometer. The average total DNA
1009 content per schizont from the arrayed experiments was determined by assessing the difference
1010 in relative fluorescence units of schizont stage parasites normalized to the fluorescence units in
1011 ring stage parasites, which contain only one copy of the *P. falciparum* genome. Experiments were
1012 performed in four subgroups. Additional parameters such as kurtosis and skewness were also
1013 determined. A total of six controls including PARE [Dd2], PfRH3 (variant 1), PfRH3 (variant 2),
1014 PfEBA165, PfPARE and MyoB were included in each of the subgroups. The two PfRH3 lines
1015 PfRH3 (variant 1) and PfRH3 (variant 2) were used to determine the boundaries of the confidence
1016 intervals for phenotype assignment. The DNA content of the deletion lines was scored as either
1017 low (if the mean DNA content was lower that the controls and had non-overlapping confidence
1018 intervals to the PfRH3 controls), or non-significant (if the confidence intervals of the mean DNA
1019 content overlapped with the PfRH3 controls) or high (if the mean DNA content was higher than
1020 the PfRH3 controls and with non-overlapping confidence intervals to the PfRH3 controls).

1021

1022 For BarSeq assessment of schizont DNA content, mixed pools of synchronous parasite cultures
1023 at mid-schizont stages were incubated in 1.5uM Compound 2 for 8 hours, fixed in 25% Methanol
1024 for 15 min, washed with 1X PBS, stained with 4uM Hoechst 33342, and then sorted into five
1025 operational gates (shown in **Supplementary Figure 11**) using the BD influx cell sorter (BD
1026 Biosciences). The total number of cells sorted ranged from 2000 – 150000. These volumes were
1027 then spun down and frozen. A volume of 2uL of this were then used for the BarSeq experiments
1028 as detailed in the methods to generate reads. A frequency distribution table was used to calculate
1029 the mean barcode content and confidence intervals. Significance in the mean barcode content
1030 was determined by comparing each deletion line to PfRH3 (variants 1 and 2). The barcode content
1031 of the deletion lines was scored as either low (if the mean barcode content was lower that the
1032 controls and had non-overlapping confidence intervals to the PfRH3 controls), or non-significant
1033 (if the confidence intervals of the mean barcode content overlapped with the PfRH3 controls) or
1034 high (if the mean barcode content was higher than the PfRH3 controls and with non-overlapping
1035 confidence intervals to the PfRH3 controls).

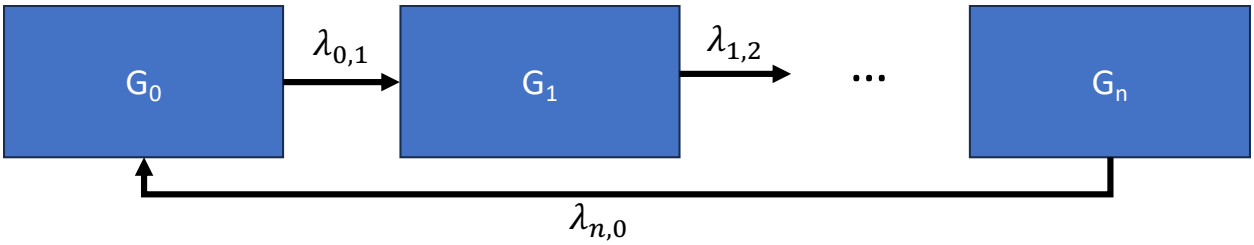
1036

1037 **IDC progression**

1038 Synchronous parasites at schizont stages were pooled and allowed to invade uninfected RBCs
1039 on a shaker at 60rpm. After 3-6 hours, the pooled cultures were treated with 5% Sorbitol to remove
1040 non-egressed schizonts. For the proof of principle experiment, samples were sorted at intervals
1041 of 6h over a 48h time course using the BD influx cell sorter (BD Biosciences). At each time of
1042 sampling, 125uL of a 2% culture was fixed in 25% Methanol for 15 min, washed with 1X PBS and
1043 then stained in 2mL mixture of 4uM Hoechst 33342/ 0.1ng/mL Thiazole Orange until parasites
1044 were ready for acquisition. The total number of cells sorted in each gate ranged from 2000 –
1045 150000 from multiple gates as shown in the results and density was dependent on the stage of

the cycle. These volumes were then spun down and frozen down. A volume of 2uL of this were then used to extract gDNA for the BarSeq experiments.

Two methods were used to measure the transition rates of parasites through the gates – a 'before and after' weighted approach, and also a compartmental model to simulate the transition of reads between gates. Each compartment records the proportion of sequencing reads observed at time *t* in that gate. We modelled independent transition rates $\lambda_{x,x+1}$ for each pair of sequential gates.



Basic structure of the compartmental model.

For each gate, the change in read proportion is modelled as:

$$\frac{dG_x}{dt} = -\lambda_{x,x+1} + \lambda_{x-1,x} \quad \text{if x not equal to n or 0} \tag{1}$$

$$\frac{dG_x}{dt} = -\lambda_{n,0} + \lambda_{n-1,n} \quad \text{if x equal to n}$$

$$\frac{dG_x}{dt} = -\lambda_{x,x+1} + \lambda_{n,0} \quad \text{if x equal to 0}$$

where $\lambda_{x,x+1}$ is the transition rate from the current gate to the next gate, $\lambda_{x-1,x}$ is the current gate to the next gate, and *dt* is measured in hours. Note that the model assumes that the transition rates are constant through time. It furthermore assumes that all the sequencing reads in a given compartment will eventually transition to the next compartment.

1063

1064 The transition rates of the compartmental model were independently calibrated to each of the
1065 genes examined. These rates were calibrated by minimizing the mean square error between the
1066 proportion of reads in each gate collected at hour *t* and the simulated results at time *t*.

1067

1068 For the bigger pools, we used a weighted 'before and after' approach to assess the transition
1069 rates of parasites through the IDC.

1070

1071 **Modelling the effect of changes in DNA content and cell cycle progression on growth rates**

1072 The combination of DNA content and IDC progression were selected as the independent variables
1073 (X), and parasite growth rates chosen as the dependent variable (y). The dataset was then divided
1074 into training and testing sets using an 80/20 split through the train_test_split function from
1075 scikitlearn. To ensure that the features were on a similar scale, particularly important for the
1076 models used, the features were standardized using StandardScaler. For the Support Vector
1077 Regression (SVR) model, the features DNA content and IDC progression were standardized
1078 using StandardScaler to ensure that each feature contributed equally to the model. The SVR
1079 model was implemented using the SVR class from scikit-learn, with a polynomial kernel
1080 (kernel='poly'), a regularization parameter C=100, and an epsilon value of 0.1. The polynomial
1081 kernel was selected to capture the non-linear relationships between the features and the target
1082 variable. The model was trained on the standardized training data, and predictions were made on
1083 the test set. The performance of the model was evaluated using the Mean Squared Error (MSE)
1084 metric, which resulted in an MSE of 0.016. A scatter plot was generated to visualize the
1085 relationship between the true and predicted values. Specific values were also defined and
1086 standardized for prediction, demonstrating the model's prediction capabilities for unseen data
1087 points. A Dense Neural Network (DNN) was constructed using TensorFlow's Keras API for the
1088 task of regression. The features DNA content and IDC progression were standardized prior to

training to ensure consistent input to the model. The network architecture consisted of three fully connected layers: the first layer with 64 neurons and ReLU activation, the second layer with 32 neurons and ReLU activation, and a final output layer with a single neuron to predict the continuous target outcome variable. The model was compiled using the Adam optimizer and the Mean Squared Error (MSE) loss function. Training was conducted over 20 epochs with a batch size of 32, and a validation split of 20% was used to monitor the model's performance on unseen data during training. The model achieved a test MSE of 0.059. Additionally, a plot of the training and validation loss across epochs was generated to assess the model's learning process. The trained DNN was used to make predictions on both the test set and specific predefined values, which were standardized similarly to the training data, showcasing its ability to generalize to new inputs. For the Random Forest Regression model, the RandomForestRegressor from scikit-learn was employed with 100 trees (n_estimators=100). The model was trained on the standardized training data, and predictions were made on the test set. Model performance was assessed using the Mean Squared Error (MSE) metric, yielding an MSE of 0.028. The relationship between the true and predicted values was visualized with a scatter plot. Additionally, specific values were defined for prediction purposes, standardized, and passed through the trained model to obtain predicted outcomes, demonstrating the model's ability to generalize to new data points.

Data analysis

Fastq files were generated from the illumina sequencing pipelines and contained raw barcode counts. An in-house script in Perl was used to generate read counts in excel files that could then be manipulated using different statistical methods and software. Scripts in R software were generated to analyse both the arrayed and BarSeq data. A range of statistical approaches were used nd details of these are discussed in the results sections.