Plasmodium falciparum schizont stage transcriptome variation among clinical isolates and laboratory-adapted clones

Sarah J Tarr[1,*], Ofelia Díaz-Ingelmo[1], Lindsay B Stewart[1], Suzanne E Hocking[1], Lee Murray[1], Craig W Duffy[1], Thomas D Otto[2], Lia Chappell[2], Julian C Rayner[2], Gordon A Awandare[3] and David J Conway[1,*]

[1] Pathogen Molecular Biology Department, London School of Hygiene and Tropical Medicine, London, UK

[2] Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK

[3] West African Centre for Cell Biology of Infectious Pathogens, Department of Biochemistry, Cell and Molecular Biology, University of Ghana, Legon, Ghana

[*] Corresponding authors
sarah.tarr@lshtm.ac.uk
david.conway@lshtm.ac.uk

Abstract

Malaria parasite genes exhibit variation in both sequence and expression level. There is much information on sequence polymorphism, but less resolution on natural variation in transcriptomes of parasites at specific developmental stages. This is largely because it is challenging to obtain highly replicated sampling of transcriptomes to overcome potentially confounding technical and biological variation. We address the issue in the major human parasite *Plasmodium falciparum* by obtaining RNA-seq profiles of multiple independent replicate preparations of mature schizont-stage parasites from a panel of clinical isolates recently established in culture and from long-term laboratory-adapted clones. With a goal of robustly identifying variably expressed genes, we show that increasing the numbers of biological sample replicates greatly improves the discovery rate. Generally, six independent replicates of each parasite culture is recommendable as being significantly to lower numbers, although for highly expressed genes variable expression can be detected when fewer replicates are available. A broad comparison identifies genes differing in relative expression between cultured clinical isolates and laboratory-adapted clones. Genes more highly expressed in the laboratory-adapted clones include an AP2 transcription factor gene Pf3D7_0420300 and putative methyl transferase genes. The variable expression of several known merozoite invasion ligands is confirmed, and previously uncharacterised genes are shown to be differentially expressed among clinical isolates. New RT-qPCR assays validate the variation in transcript levels of these genes, and allow quantitation of expression to be extended to a wider panel of clinical isolate samples. These variably expressed genes are new candidates for investigation as potential determinants of alternative parasite developmental pathways or targets of immunity.

Author summary

Understanding parasite diversity and adaptation may require characterisation of gene expression variation, and is vital if chemotherapeutic or vaccine development is to consider new candidate targets, but it is technically challenging to generate precise data on clinical isolates. Here, we analyse the transcriptomes of mature *Plasmodium falciparum* schizonts using RNA-sequencing, using large numbers of biological replicate samples to minimise the impact of inter-replicate variation on observed patterns of differential expression. This identifies genes that are differentially expressed in long term laboratory-adapted parasites

56 and recently cultured clinical isolates, as well as among different clinical isolates. In
57 additional samples of schizonts grown in the first cycle *ex vivo* prior to any erythrocyte
58 invasion, expression levels of a selected panel of these genes vary among isolates, but mean
59 levels are similar to those in the continuously cultured clinical isolates, indicating that the
60 latter are useful for experimental studies requiring biological replication.
61
62 **Introduction**
63 Variation in gene expression is key to survival and reproduction of microbes, affecting
64 diverse phenotypes such as sexual differentiation [1], adaptation to different growth
65 conditions [2] and evasion of host immunity [3]. Malaria occurs as parasites undergo cycles
66 of invasion and asexual replication inside erythrocytes. Towards the end of each cycle
67 schizonts develop which contain multiple merozoites that need to burst from the host cell
68 and invade a new erythrocyte. The transcriptional program is highly synchronised
69 throughout the asexual parasite replication cycle in erythrocytes [4], but some variation
70 exists between parasite clones [5,6]. Analysis of naturally occurring sequence polymorphism
71 in *Plasmodium falciparum* has shown that selection maintains multiple alleles of many
72 merozoite-stage genes [7,8], some of which encode targets of naturally acquired immunity
73 [9-11]. Genes expressed at this stage can also exhibit pronounced variation in transcription
74 among parasites [7,12-14], although the extent of this variation among laboratory-adapted
75 parasite lines and clinical isolates has not been deeply surveyed.
76
77 Accurate quantitation of differential gene expression between organisms or groups of
78 organisms in whole-transcriptome analyses requires biological replicate sampling [15,16].
79 The importance of minimising sampling error in order to detect real differential gene
80 expression is well recognised [17], and some bioinformatic tools can guide the
81 determination of replicate numbers appropriate to experimental designs [18,19]. In
82 transcriptomic studies of malaria parasites, achieving large numbers of replicate sample
83 preparations is difficult, particularly from clinical isolates [5], and most current
84 understanding is derived from a small number of parasite lines that have been culture-
85 adapted for many years. Examination of genome sequences has identified premature stop
86 codon mutants affecting some transcriptional factor genes and cell signalling protein genes
87 [20,21], as well as particular gene deletions and amplifications only observed in parasites

88 adapted to *in vitro* culture [22,23]. Although culture-adapted parasites are phenotypically
89 and transcriptionally diverse [6,12,24], it is not clear to what extent they reflect the diversity
90 of parasites causing clinical malaria. Parasites may be studied in the first cycle of *ex vivo*
91 growth following isolation from patients, and maturation to schizont stages shows variable
92 expression of invasion-related genes [7,13,14], but a lack of biological replicate
93 measurements limits the precision of whole transcriptome analysis of such samples.
94
95 Here we present gene expression profiles of schizont-stage malaria parasites from multiple
96 cultured clinical isolates and laboratory-adapted lines. We conduct RNA-seq analysis with
97 large numbers of replicates per sample, and show that high numbers of biological replicates
98 improves the true-positive discovery rate for identifying differentially expressed genes. This
99 identifies schizont-stage genes that are differentially expressed between laboratory-adapted
100 and cultured clinical isolates, as well as those that are variably expressed among cultured
101 clinical isolates. The results confirm variable transcription in particular genes encoding
102 ligands involved in erythrocyte invasion, and variation in expression of genes that have been
103 less characterised was quantitatively validated by targeted RT-qPCR assays, allowing analysis
104 to be extended to additional isolates from which material is more limited. The expression
105 levels of these genes in a panel of clinical isolates sampled at schizont stage in the first *ex*
106 *vivo* cycle were similar on average to those observed in cultured clinical isolates. This
107 indicates that cultured parasite lines are useful for studying expression of most genes,
108 although a minority of genes are affected by culture adaptation.
109
110 **Results**
111 **Replicate samples of schizont-stage transcriptomes from *P. falciparum* laboratory-adapted**
112 **clones and clinical isolates**
113 Nineteen biological replicate preparations of schizont-stage parasites were made from
114 cultured *P. falciparum* clone 3D7, and between five and ten biological replicate preparations
115 were made from each of three other laboratory-adapted clones of diverse origin and six
116 clinical isolates from Ghana. Transcriptomes were obtained by sequencing of oligo-dT-
117 primed cDNA (RNA-seq), and mapping reads to the 3D7 reference genome sequence. To
118 minimize potential effects of sequence differences between the different parasites, regions
119 of genes with high levels of allelic polymorphism were masked, so that for polymorphic

120 genes only reads mapping to conserved portions were analysed (S1 file). Genes that have
121 been described as deleted in some parasites were not excluded, as most individual deletions
122 are very rare and exclusion would result in unnecessary loss of data for genes unaffected in
123 the isolates sampled here, but we chose to rather filter out cases where deletion was
124 apparent in analysis. The subtelomeric large gene families *var*, *rifin* and *stevor* were
125 excluded as short read sequence mapping is not generally locus-specific for these, so 5,134
126 genes were analysed in total. Mapped RNA-seq reads for each gene were converted to
127 fragments per kilobase of transcript per million mapped reads (FPKM). Analysis of four
128 paired replicates of schizont preparations from E64-treated and untreated cultures of clone
129 3D7 showed that E64 treatment had no major effect, with only a single gene showing $\log_2$
130 fold difference > 2 (S2 File).
131
132 As multiple biological replicates could not have exactly the same distribution of parasite
133 developmental stages, despite being schizont-enriched, to assess the extent of earlier stage
134 transcripts the distributions of FPKMs for each replicate were correlated with FPKM values
135 in reference RNA-seq data for seven cultured time points over a ~48 hour *P. falciparum*
136 cycle [25]. This identified only a small number of replicates that did not have a maximum
137 Spearman's correlation with parasites at either 40 or 48 hours post-invasion, and these
138 were excluded from further analyses (Fig 1). Within each isolate, pairwise correlations of
139 FPKMs among replicates correlated highly, with average Spearman's $\rho > 0.95$ (Fig 2).
140
141 **Numbers of biological replicates improve detection of differentially expressed genes**
142 We assessed the impact of increasing numbers of biological replicates on detection of genes
143 with differing transcript levels between parasite lines. Data were first compared for the 3D7
144 and D10 laboratory clones, which each had material from schizonts purified by density
145 gradient centrifugation from ten independent cultures. Comparing all ten replicates of 3D7
146 and D10 identified 123 genes with $\log_2$ differences of > 2 in relative transcript levels
147 between the two clones (equivalent to at least 4-fold differences). Achieving this high
148 number of biological replicate samples is difficult, so we assessed what proportion of these
149 123 genes were captured as being differentially expressed in 100 random samples of two,
150 four, six and eight replicates within each group. As expected, the true-positive proportion of
151 differentially expressed genes detected (using the same statistical criteria of absolute $\log_2$

152 fold difference > 2) increased with the number of replicates within each group, median true-
153 positive rates being 0.28, 0.52, 0.67 and 0.79, for two, four, six and eight replicates
154 respectively (Fig 3). While the true-positive rate of genes detected increased with the
155 number of replicates, the median false-positive rates were very low in all cases (among the
156 genes that did not show differences with 10 replicates, the proportions showing apparent
157 differences with two, four, six and eight replicates were 0.001, 0.004, 0.005 and 0.004
158 respectively).
159
160 We also examined the ability to detect differential expression when focusing specifically on
161 genes that are highly expressed in schizont-stage samples, as these are the genes most likely
162 to be functionally important for the merozoite invasive stage. The FPKM values for each
163 sample in each of the 100 comparisons were averaged and the genes with non-zero FPKM
164 values were ranked by their maximum expression level in either strain. Then, the
165 differentially expressed genes that fell within the top quartile of most highly expressed
166 genes in the subsampled data were compared to those identified as differentially expressed
167 among the top quartile of genes in the full analysis. This yielded median true-positive rates
168 of 0.45, 0.63, 0.78 and 0.86 for comparisons respectively containing two, four, six and eight
169 replicates, higher proportions than for the analysis of all genes. Based on these data, it is
170 evident that the use of six independent biological replicates for parasite RNA-seq will detect
171 most genes that are differentially expressed, and that slightly fewer replicates are adequate
172 if analyses only need to focus on highly expressed genes.
173
174 **Comparison of gene expression in cultured clinical and laboratory-adapted isolates.**
175 Comparing the transcriptomes of long-term laboratory-adapted and cultured clinical
176 isolates, 134 genes (2.6 % of those analysed) had different transcript levels between the two
177 groups (S1 Table). Among the genes within the top quartile of expression values overall, 20
178 genes (1.6 %) were differentially expressed between the laboratory-adapted and clinical
179 isolate groups, a lower proportion than among the rest of the genes (Odds Ratio 0.52,
180 Fisher's exact P = 0.006). It is possible that there is a higher background of false positive
181 differences for the less expressed genes, due to lower read coverage giving greater noise in
182 the relative ratios. Examining the 20 genes with differing transcript levels among the top
183 quartile of expressed genes, several were seen to contain a deletion in one of the parasite

184 lines or to be members of multigene families. For example, PF3D7_1371600 (*ebl-1*) and
185 PF3D7_0935700 are deleted in the HB3 strain and D10 strain respectively (Fig 4), as
186 previously documented [26,27]. After excluding these genes and others for which one or
187 more of the samples were presumed to contain genetic deletions, and excluding members
188 of sub-telomeric multi-gene families, ten single-locus genes were identified as showing
189 differential expression between laboratory and clinical isolates (Table 1).
190
191 Eight of the ten highly expressed genes showing a difference between the groups had higher
192 transcript levels in the long-term laboratory-adapted clones than in the cultured clinical
193 isolates (Table 1). These include an AP2 transcription factor gene PF3D7_0420300, as well as
194 two predicted methyltransferase genes (PF3D7_0422900 and PF3D7_0522300), and gene
195 PF3D7_0104300 encoding ubiquitin-binding protein 1 which is involved in protein turnover
196 [28] (Fig 4). The only two genes that had higher transcript levels in cultured clinical isolates
197 were PF3D7_1036300 which encodes the merozoite surface protein MSPDBL2, and
198 PF3D7_1302100 which encodes the gamete antigen 27/25 (Fig 4). Previous data has shown
199 MSPDBL2 to be variably expressed among *P. falciparum* isolates [7], and experimental
200 analysis has shown transcript levels to be increased in parasite cultures when the
201 heterochromatin protein 1 is suppressed by gametocyte development protein 1 (GDV1)
202 leading to gametocytogenesis [29,30], whereas gamete antigen 27/25 is a known marker of
203 early gametocytogenesis [30,31]. Gametocyte gene transcripts were not generally more
204 highly expressed in the clinical isolates, and gene PF3D7_1327300 that has been previously
205 described to be transcribed in gametocytes [32] showed lower expression.
206
207 We investigated whether there was an overall association between the genes differentially
208 expressed and targets of HP1 regulation [29,33]. Of the HP1-regulated genes previously
209 identified [29], 118 genes are among those analysed here (*var*, *rifin* and *stevor* gene families
210 were excluded). Of the 134 genes differentially expressed between the two groups, 24 (17.9
211 %) were targets of HP1 regulation, a higher proportion than among the rest of the genes
212 (Odds Ratio 11.3, Fisher's exact $P = 1.3 \times 10^{-5}$). Among 20 differentially expressed genes in
213 the top quartile of expression values, two (10 %) were targets of HP1 regulation, a higher
214 proportion than among the remaining genes in the top quartile (Odds Ratio 7.6, Fisher's
215 exact P = 0.037).

216

217 **Differential expression of schizont-stage genes among cultured clinical isolates**

218 Overall, 214 genes (4.2 %) showed a $\log_2$ fold difference > 2 for at least one of the fifteen
219 pairwise comparisons among cultured clinical isolates. Of those genes in the top quartile of
220 expression levels, 39 genes (3.0 %) showed a difference for at least one pairwise
221 comparison, a lower proportion than among the less expressed genes (Odds Ratio 0.66,
222 Fisher's Exact P = 0.02) (S2 Table). Some of the variably expressed genes encode merozoite
223 proteins previously characterised as transcriptionally variable among *ex vivo* clinical isolates
224 [14], including *dblmsp2* and *msp6*, as well as erythrocyte binding antigen genes *eba-175*,
225 *eba-181*, and *eba-140* (which is also deleted in one of the laboratory adapted clones) (S1
226 Fig).

227

228 There was an over-representation of GDV1-regulated genes [30] among the 214 genes
229 differentially expressed in cultured clinical isolate comparisons (Fisher's exact test odds ratio
230 12.5, Fisher's Exact P = 1.9×10^{-9}). Of the genes differentially expressed among cultured
231 clinical isolates, 64 % were also differentially expressed in comparisons among the long-
232 term laboratory-adapted clones (Odds Ratio 17.4, Fisher's exact P = 2.2×10^{-16}). There was
233 also strong overlap when only considering genes in the top quartile of expression overall, 44
234 % of genes differentially expressed among cultured clinical isolates also being differentially
235 expressed among laboratory-adapted lines (Odds Ratio 7.9, Fisher's exact P = 2.4×10^{-8}).

236

237 To validate the data obtained through RNA-seq, a subset of genes was selected for
238 quantitation by reverse-transcription quantitative PCR (RT-qPCR). We chose differentially
239 expressed genes encoding proteins containing predicted transmembrane domains or signal
240 peptides, thereby likely to enter the parasite secretory pathway, excluding genes previously
241 characterised as variably expressed or members of multi-gene families (S2 Table). Eight
242 genes were selected for assay (S2 Fig), two of which encode known antigens (merozoite-
243 associated tryptophan-rich antigen [34] liver-stage antigen-3 [35]). These gene transcripts
244 were quantified by RT-qPCR in 58 RNA preparations from the laboratory clones and clinical
245 isolates that had been analysed by RNA-seq. Transcript levels were normalised to those of a
246 house-keeping gene (serine-tRNA ligase PF3D7_0717700) and correlated with similarly
247 normalised FPKM expression values for the RNA-seq data. Each gene showed strong positive

248 and highly significant correlation between RNA-seq and RT-qPCR-derived expression
249 measures, all except one having correlation coefficients > 0.8 (Fig 5).
250
251 **Variable expression in first *ex vivo* cycle schizonts**
252 We next determined the expression levels in additional clinical samples from Ghana
253 matured during the first cycle *ex vivo* until a large proportion of parasites were schizonts.
254 For nine of these isolates there was sufficient material for RNA-seq to be performed, and
255 results were analysed for seven of these (two isolates were excluded from analysis as
256 whole-transcriptome staging of FPKM values showed maximal correlation with stages earlier
257 than schizonts in a reference time-course). For six isolates grown *ex vivo* to schizont stage
258 there was sufficient material for quantification of the selected eight gene panel by RT-qPCR.
259 Characterisation of the *ex vivo* samples using either method showed considerable variation
260 in relative expression levels of each gene, but the mean levels of each gene across the
261 isolates were similar to those determined in the cultured parasites that had extensive
262 sample replication (Fig 6). Without sample replication of the first cycle *ex vivo* clinical
263 samples, the contribution of random sampling variation and uncontrolled effects on the
264 expression levels cannot be accounted for, but the observation that for each gene the
265 normalised expression values have a similar mean to that for cultured clinical lines indicates
266 no systematic differences in expression between the initial *ex vivo* cycle and under
267 continuous culture conditions. These results support the use of cultured clinical isolates as a
268 means of studying parasite gene expression to reflect that occurring in original populations,
269 while enabling the necessary experimental replication for many analyses.
270
271 **Discussion**
272
273 Transcriptomic analyses perform optimally with increased levels of sample replication to
274 minimise the impact of stochastic or technical variation in individual samples [17,36]. This is
275 particularly important for studies that focus on a defined developmental stage, in which
276 error can be generated due to subtle differences between samples. However, the nature of
277 clinical samples of malaria parasites is such that replication is technically very difficult. In
278 order to assess the degree of replication useful to robustly identify differentially expressed
279 genes, we first generated RNA-seq profiles from multiple replicates of laboratory-adapted

280 clones, and used the data to calculate the true- and false-discovery rates obtained using
281 fewer replicates. Consistent with studies on other cultured eukaryotes [17], we show that
282 for malaria parasites increasing the sample replication improves the true-positive discovery
283 rate in identifying variably expressed genes, and indicate that six independent replicates are
284 useful to balance accuracy and experimental feasibility. Where these numbers of replicates
285 are not obtainable, it is recommended to focus on genes that are highly expressed in order
286 to achieve as much accuracy as possible per gene.
287
288 It has been proposed that spontaneous transcriptional variation within parasite populations
289 is a strategy that ensures fitness of parasites facing a range of potential and changing
290 selective pressures [6], and it is likely that the rates and mechanisms of regulation differ
291 among most genes. In comparison between long-term laboratory-adapted clones and
292 cultured clinical isolates, we found that among the genes most differentially expressed were
293 an AP2-transcription factor [37] and two methyltransferases, the functional roles of which
294 remain to be investigated. Most of the differentially expressed genes had lower
295 transcription in the clinical isolates, suggesting that these may be normally repressed. HP1 is
296 responsible for extensive gene silencing in malaria parasites [30], and we showed an over
297 representation of HP1-regulated genes among those differentially expressed between
298 laboratory-adapted clones and cultured clinical isolates. It would be interesting to
299 undertake further studies directly on HP1 and other chromatin modifications in recent
300 clinical isolates, as experimental studies have so far focused on long-term laboratory-
301 adapted clones. Interestingly, among the top quartile of genes expressed, the only two
302 genes that were more highly expressed in clinical isolates were an early marker of
303 gametocytogenesis (*Pfs25/27)* and a merozoite surface protein gene (*mspdbl2*) encoding a
304 product expressed all merozoites within only a minority of schizonts [7] which may be a
305 marker of parasites committed to sexual development [29]. Studies of gametocytogenesis in
306 clinical isolates will be important to understand the process of commitment to sexual
307 differentiation and transmission, as also indicated by transcriptome comparisons that
308 sampled different stages without having multiple replicates for each stage [5].
309
310 Through pairwise comparisons of individual cultured clinical isolates, we identified many
311 genes that were differentially expressed among clinical isolates. Some of the genes

312 identified have previously been shown to be differentially expressed through targeted qRT-
313 PCR assays of schizont-stage *ex vivo* cultured clinical isolates [7,13,14]. Focusing on genes
314 for which variable expression has not been previously studied, we selected a panel of eight
315 highly differentially expressed genes that encode secreted proteins. Of these, the
316 merozoite-associated tryptophan-rich antigen and liver stage antigen 3 have been
317 previously identified as merozoite proteins [34,35], and another gene (PF3D7_0423900) is
318 adjacent to loci encoding the cysteine-rich protective antigen (CyRPA) and reticulocyte
319 binding homologue 5 (Rh5) which are targets of antibodies that inhibit merozoite invasion.
320 We extensively validated the RNA-seq data by RT-qPCR for these genes, and extended
321 quantitation to additional *ex vivo* clinical isolate samples. Despite the lack of sample
322 replicates for the latter, results showed that average expression levels for each gene were
323 similar to those for the cultured samples that had biological replicates.
324
325 Together, these data highlight new candidates for investigation as potential determinants of
326 alternative developmental pathways or targets of immunity. Characterisation of gene
327 transcript levels is only one level of analysis, as proteomic characterisation [38] and analysis
328 of phosphorylation and other protein modifications [39,40] may also identify functional
329 determinants, and large-scale analysis of parasite protein variation has not yet been
330 attempted on clinical isolates. A more complete description of transcriptome variation may
331 also be derived in future by analysis of individual infected cells, which has not been
332 undertaken on parasites from clinical isolates, as initial single-cell sequencing studies of
333 laboratory model malaria parasites have only recently begun to probe the expression
334 pathways of different life-cycle stages [41-43]. Scaling up transcriptomic analyses to clinical
335 studies is challenging, and biological replicate preparations of parasite samples may only be
336 feasible in a limited number of cases. In an epidemiological context, reduction of error also
337 requires large sample sizes and tight definition of clinical phenotypes, and more statistical
338 analysis and modelling of covariates may be appropriate, ideally including data on host
339 transcriptome variation within the same samples [44].
340
341 **Methods**
342 **Ethical approval.** The study of clinical parasite transcriptomes was approved by the Ethics
343 committees of the Ghana Health Service, the Noguchi Memorial Institute for Medical

344 Research, University of Ghana, the Kintampo Health Research Centre, the Navrongo Health
345 Research Centre and the London School of Hygiene and Tropical Medicine. Written
346 informed consent was obtained from parents or legal guardians of all participating children,
347 and additional assent was received from children over 10 years old.
348
349 ***P. falciparum* isolates sampled from clinical malaria cases.** Blood samples were collected
350 from clinical malaria cases attending Ghana government health facilities between 2012 and
351 2013, in Kintampo (Brong-Ahafo Region of central Ghana), and Navrongo (Kassena-Nankana
352 East Municipality, in the Upper East Region of northern Ghana). Patients were eligible to
353 participate in the study if they had uncomplicated clinical malaria, were aged 2–14 years,
354 tested positive for *P. falciparum* malaria by lateral flow rapid diagnostic test and slide
355 microscopy, and had not taken antimalarial drugs during the 72 hours preceding sample
356 collection. Venous blood samples (up to 5 ml) were collected into heparinized Vacutainer
357 tubes (BD Biosciences). Blood samples were centrifuged, plasma separated and leukocyte
358 buffy coat cell layer removed, and erythrocytes were cryopreserved in glycerolyte and
359 stored frozen at −80°C or in liquid nitrogen until shipment on dry ice to the London School
360 of Hygiene and Tropical Medicine.
361
362 **Parasite culture and enrichment for schizont stages.** Parasites from *ex vivo* clinical samples
363 were matured within the original patient erythrocytes at 2–5 % haematoctit for up to 48 h in
364 RPMI 1640 medium containing 2 % human AB serum (GE Healthcare) and 0.3 % Albumax II
365 (Thermo Fisher Scientific) under an atmosphere of 5 % O_2, 5 % CO_2, and 90 % N_2 at 37 °C,
366 until most parasites were at the schizont stage of parasite development, at which point RNA
367 was extracted directly from the bulk cultures. Laboratory-adapted clones and cultured
368 clinical isolates were grown in erythrocytes at 2–5 % haematocrit in RPMI 1640 medium
369 containing 0.5 % Albumax II, at 37 °C. Laboratory-adapted clones were maintained under
370 atmospheric air with 5 % CO_2, and cultured clinical isolates were maintained in 5 % O_2, 5 %
371 CO_2, 90 % N_2. Schizonts were isolated by either MACS or discontinuous density
372 centrifugation methods.
373 For MACS purification of cultured clinical isolates, parasites were prepared from 100 ml
374 cultures with at least 0.7 % of erythrocytes containing schizonts. Parasites were isolated by
375 magnetic purification using magnetic LD Separation columns (Miltenyi Biotech). One column

376 was used per 25 ml culture, and washed twice in 3 ml of culture medium at room
377 temperature. Parasite culture was pelleted at 500 g for 5 minutes, then re-suspended in 3
378 ml culture medium per 1 ml of packed cell volume. The re-suspended material was bound to
379 the MACS column, which was then washed three times with 3 ml culture medium, and
380 schizonts were eluted twice by removing the magnet from the column and forcing 2 ml
381 culture medium through the column into a 15 ml collection tube. Finally, the schizonts were
382 pelleted by centrifugation at 500 g for 5 minutes and the pellet volume was estimated, with
383 0.5 µl used for Giemsa-stained microscopical examination to assess staging, 1 µl added back
384 to 250 µl culture at 0.8 % hematocrit to follow the progression. Remaining parasites were
385 re-suspended in 1.5 ml of culture medium with 10 µM E64 in a 12 well plate, and parasites
386 were incubated for 5.5 hours in 5 % CO_2 at 37 °C, before centrifugation in a 1.5 ml tube and
387 processing for RNA extraction.
388 For discontinuous density centrifugation purification, parasites were maintained as 25 ml
389 cultures at 2.5 % hematocrit. Cultures were used when > 1 % erythrocytes contained
390 parasites with multiple nuclei. Schizonts were purified using 70 % Percoll (GE
391 Healthcare)/2.93 % sorbitol/PBS overlayed with 35 % Percoll/1.47 % sorbitol/PBS, which was
392 overlayed with parasitized erythrocytes, allowing the schizonts to be separated by
393 centrifugation at 2500 g for 10 minutes at 24 °C, with a light brake at the end of
394 centrifugation. Purified schizonts were washed once in complete medium and the pellet
395 volume was estimated. Six pellet volumes of 50 % haematocrit erythrocytes were added to
396 the pellet and resuspended for slide examination. The sample was smeared, and
397 resuspended in 6 ml of complete culture medium. Of this, 1 ml was used as a control
398 untreated sample to track parasite egress. E64 was added to the remaining 5 ml at a final
399 conentraction of 10 µM. Control and E64-treated samples were placed at 37 °C, 5 % CO_2
400 static incubator for 5.5 hours. Schizonts from the E64-treated culture were overlaid on 70 %
401 Percoll/2.93 % sorbitol/PBS and separated by centrifugation at 2500 g for 10 minutes at 24
402 °C, with a light brake at the end of centrifugation. The schizont layer was washed once in
403 complete culture medium and final cell pellets of approximately 10 - 20 µl were used for
404 RNA extraction.
405
406 **RNA extraction.** Pellets were resuspended 500 µl in TRIzol® reagent (Thermo Fisher
407 Scientific) and stored at -80 °C until RNA extraction following manufacturer's instructions.

408 Final RNA pellets were resuspended in 100 µl RNase-free H_2O. A second RNA clean-up and
409 on-column DNase treatment was carried out using RNeasy mini columns (Qiagen), with RNA
410 eluted in 30-50 µl RNase-free H_2O and concentration quantified by Qubit High Sensitivity
411 RNA Assay (Thermo Fisher Scientific). Samples containing at least 500 ng RNA were
412 considered for analysis after the RNA integrity was checked on an Agilent Bioanalyzer using
413 RNA 6000 Nano reagents and chips (Agilent Genomics).
414
415 **RNA-seq library preparation and sequencing**. Laboratory isolate and cultured clinical
416 isolate RNA-seq libraries were prepared using TruSeq Stranded mRNA Library Prep Kit
417 (Illumina) using 500 ng – 1 µg RNA as per the Illumina TruSeq Stranded mRNA protocol for
418 MiSeq sequencers. Libraries were validated on an Agilent Bioanalyzer using DNA 1000
419 reagents and chips (Agilent Genomics) to quantify library sizes and confirm the absence of
420 primer dimers. Libraries were quantified using a KAPA Universal Library Quantification kit
421 (Roche Diagnostics Limited) on a 7500 Fast Real-Time PCR System (Thermo Fisher Scientific)
422 and library concentrations were adjusted for library size. 12 – 15 pM pooled libraries were
423 sequenced on a MiSeq System (Illumina) using a MiSeq Reagent Kit v3 (Illumina) with 2 x 75
424 cycles.
425 *Ex vivo* schizont-enriched *P. falciparum* isolate RNA-seq libraries were prepared using a
426 modified protocol. PolyA+ RNA (mRNA) was selected using magnetic oligo-d(T) beads.
427 mRNA was reverse transcribed using Superscript III® (Thermo Fisher Scientific), primed using
428 oligo-d(T) primers. dUTP was included during second-strand cDNA synthesis. The resulting
429 double stranded cDNA was fragmented using a Covaris AFA sonicator. Sheared double
430 stranded cDNA was dA-tailed, end repaired, and "PCR-free" barcoded sequencing adaptors
431 (Bioo Scientific) [45] were ligated (NEB). Libraries were cleaned up twice, using solid phase
432 reversible immobilisation beads, and eluted in EB buffer (Qiagen). Second strand cDNA was
433 removed using uracil-specific excision reagent enzyme mix (NEB). Libraries were quantified
434 by quantitative PCR prior to sequencing on an Illumina MiSeq sequencer.
435
436 **Data curation and analysis.** Raw Illumina sequence reads were aligned to the *P. falciparum*
437 3D7 v3 genome and converted to 'bam' format using samtools [46]. Reads with MAPQ
438 scores < 60 were removed. Reads were counted using the "summarizeOverlaps" feature of
439 the GenomicAlignments package [47] in R, against a previously published *P. falciparum*

440 genome annotation file that had been masked for regions of polymorphism (detailed in S2
441 File) to remove known regions of genomic polymorphism, highly polymorphic gene families
442 and duplicated genes.
443 Fragments Per Kilobase of transcript per Million mapped reads (FPKMs) for the data here
444 and those of previously published data on a developmental time course of the clone 3D7
445 [25] were calculated using the 'fpkm' function of DESeq2 [48] in R. FPKMs for all genes in
446 each of our parasite preparations were correlated using a Spearman's Rank correlation with
447 FPKMs for each of the seven time points (0, 8, 16, 24, 32, 40 and 48 hours post-invasion; S1
448 Table). Replicates with peak Spearman correlation values of $\rho > 0.7$ at the latest time points
449 (40 or 48 hours) were included for further analysis. Differential expression analysis was
450 conducted using DESeq2 in R.
451
452 **RT-PCR assays.** 150 – 500 ng total RNA from each preparation of parasite schizonts was
453 reverse transcribed using Superscript II® (Thermo Fisher Scientific) with 250 ng random
454 hexamer oligonucleotide primers per 20 µl reaction. Quantitative PCR (qPCR) was carried
455 out using SYBR® Select Master Mix (Thermo Fisher Scientific) with 500 nM forward and
456 reverse primers, in a Prism 7500 Fast qPCR machine. For each gene, threshold-cycle values
457 were quantified against a serially diluted genomic DNA (Dd2 strain) standard curve, run on
458 the same plate. Cycling parameters were: 50°C for 2 minutes, 95°C for 2 minutes followed
459 by 40 cycles of 95 °C for 15 seconds and 60 °C for 1 minute. All wells were run as 10 µl
460 volumes in technical duplicate. The qPCR copy numbers were normalised against copies of a
461 house-keeping gene, PF3D7_0717700 [49]. PCR primer pair sequences are as follows:
462 PF3D7_0102700 5'-CAACCAGACAAACAACAAGAAA-3' and 5'-TCCATTCTGATGCTTTCCAC-3',
463 PF3D7_0220000 5'-GTAAATGGTGAATTAGCTAGTGAAG-3' and 5'-
464 CCTTTATATCTTCGGCTTCTTCT-3', PF3D7_0423900 5'-GAGAAGAAGCCAAAGTAAATGAAC-3'
465 and 5'-GAATGTGTCAAGTGCATCATAA-3', PF3D7_0605900 5'-CGCAATAACAAGAAGTCACAAA-
466 3' and 5'-AAGATTGTAGGAGGGTGTTAAT-3', PF3D7_0935300 5'-GGGCTGCTGTTATACCTTG-3'
467 and 5'-AGAATGAGGAGGTTCAGTTTG-3', PF3D7_1252900 5'-
468 CCTTAGTAGAACTTCAAAGTGACA-3' and 5'-TGTAACCAACTACGAAATTCCC-3',
469 PF3D7_1461700 5'-TGCTGACTTTCAAGAGATAAGG-3' and 5'-
470 TTTCATTTGTTCATTTTGTACAACC-3', PF3D7_1476500 5'-CTTCGATTCACAAATGCAGAAA-3' and

471 5'-CGTATTCCACACCTTGTTCTAT-3', PF3D7_0717700 5'-AAGTAGCAGGTCATCGTGGTT-3' and
472 5'-GTTCGGCACATTCTTCCATAA-3'.
473
474 **Data open access**
475 All *P. falciparum* RNA-seq data from the laboratory-adapted and cultured clinical isolates
476 have been submitted for open access at the Gene Expression Omnibus
477 (https://www.ncbi.nlm.nih.gov/geo/), entry: GSE113718. The *P. falciparum* RNA-seq data
478 from the first cycle e*x vivo* isolates have been submitted to the European Nucleotide
479 Accession Archive (https://www.ebi.ac.uk/ena), study: ERP103955.
480
481 **Acknowledgements**
482 We are grateful to malaria patients who contributed blood samples for analysis of parasite
483 transcription, and to colleagues who supported the clinical sample collection and initial
484 laboratory processing. We thank Alistair Miles and Antoine Claessens for help with
485 identification of genomic coordinates containing highly polymorphic sequences within
486 genes.
487
488 **References**
489 1. Mata J, Wilbrey A, Bahler J (2007) Transcriptional regulatory network for sexual differentiation in
490 fission yeast. Genome Biol 8: R217.
491 2. Cases I, de Lorenzo V, Ouzounis CA (2003) Transcription regulation and environmental adaptation
492 in bacteria. Trends Microbiol 11: 248-253.
493 3. Cestari I, Stuart K (2018) Transcriptional Regulation of Telomeric Expression Sites and Antigenic
494 Variation in Trypanosomes. Curr Genomics 19: 119-132.
495 4. Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, et al. (2003) The transcriptome of the
496 intraerythrocytic developmental cycle of *P. falciparum*. PLoS Biology 1: e5.
497 5. Mackinnon MJ, Li J, Mok S, Kortok MM, Marsh K, et al. (2009) Comparative transcriptional and
498 genomic analysis of *Plasmodium falciparum* field isolates. PLoS Pathog 5: e1000644.
499 6. Rovira-Graells N, Gupta AP, Planet E, Crowley VM, Mok S, et al. (2012) Transcriptional variation in
500 the malaria parasite *Plasmodium falciparum.* Genome Res 22: 925-938.
501 7. Amambua-Ngwa A, Tetteh KK, Manske M, Gomez-Escobar N, Stewart LB, et al. (2012) Population
502 genomic scan for candidate signatures of balancing selection to guide antigen
503 characterization in malaria parasites. PLoS Genet 8: e1002992.
504 8. Ochola LI, Tetteh KK, Stewart LB, Riitho V, Marsh K, et al. (2010) Allele frequency-based and
505 polymorphism-versus-divergence indices of balancing selection in a new filtered set of
506 polymorphic genes in *Plasmodium falciparum*. Mol Biol Evol 27: 2344-2351.
507 9. Osier FH, Mackinnon MJ, Crosnier C, Fegan G, Kamuyu G, et al. (2014) New antigens for a
508 multicomponent blood-stage malaria vaccine. Science Translational Medicine 6: 247ra102-
509 247ra102.

510 10. Richards JS, Arumugam TU, Reiling L, Healer J, Hodder AN, et al. (2013) Identification and
511 prioritization of merozoite antigens as targets of protective human immunity to *Plasmodium*
512 *falciparum* malaria for vaccine and biomarker development. J Immunol 191: 795-809.
513 11. Tetteh KK, Osier FH, Salanti A, Kamuyu G, Drought L, et al. (2013) Analysis of antibodies to newly
514 described *Plasmodium falciparum* merozoite antigens supports MSPDBL2 as a predicted
515 target of naturally acquired immunity. Infect Immun 81: 3835-3842.
516 12. Stubbs J, Simpson KM, Triglia T, Plouffe D, Tonkin CJ, et al. (2005) Molecular mechanism for
517 switching of *P. falciparum* invasion pathways into human erythrocytes. Science 309: 1384-
518 1387.
519 13. Gomez-Escobar N, Amambua-Ngwa A, Walther M, Okebe J, Ebonyi A, et al. (2010) Erythrocyte
520 invasion and merozoite ligand gene expression in severe and mild *Plasmodium falciparum*
521 malaria. J Infect Dis 201: 444-452.
522 14. Bowyer PW, Stewart LB, Aspeling-Jones H, Mensah-Brown HE, Ahouidi AD, et al. (2015) Variation
523 in *Plasmodium falciparum* erythrocyte invasion phenotypes and merozoite ligand gene
524 expression across different populations in areas of malaria endemicity. Infect Immun 83:
525 2575-2582.
526 15. Hansen KD, Wu Z, Irizarry RA, Leek JT (2011) Sequencing technology does not eliminate biological
527 variability. Nat Biotechnol 29: 572-573.
528 16. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell.
529 Science 297: 1183-1186.
530 17. Schurch NJ, Schofield P, Gierlinski M, Cole C, Sherstnev A, et al. (2016) How many biological
531 replicates are needed in an RNA-seq experiment and which differential expression tool
532 should you use? RNA 22: 839-851.
533 18. Seo J, Gordish-Dressman H, Hoffman EP (2006) An interactive power analysis tool for microarray
534 hypothesis testing and generation. Bioinformatics 22: 808-814.
535 19. Busby MA, Stewart C, Miller CA, Grzeda KR, Marth GT (2013) Scotty: a web tool for designing
536 RNA-Seq experiments to measure differential gene expression. Bioinformatics 29: 656-657.
537 20. Claessens A, Affara M, Assefa SA, Kwiatkowski DP, Conway DJ (2017) Culture adaptation of
538 malaria parasites selects for convergent loss-of-function mutants. Sci Rep 7: 41303.
539 21. Kafsack BF, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, et al. (2014) A transcriptional
540 switch underlies commitment to sexual development in malaria parasites. Nature 507: 248-
541 252.
542 22. Cheeseman IH, Gomez-Escobar N, Carret CK, Ivens A, Stewart LB, et al. (2009) Gene copy number
543 variation throughout the *Plasmodium falciparum* genome. BMC Genomics 10: 353.
544 23. Jeffares DC, Pain A, Berry A, Cox AV, Stalker J, et al. (2007) Genome variation and evolution of
545 the malaria parasite *Plasmodium falciparum*. Nat Genet 39: 120-125.
546 24. Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL (2006) Comparative whole genome
547 transcriptome analysis of three *Plasmodium falciparum* strains. Nucleic Acids Res 34: 1166-
548 1173.
549 25. Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, et al. (2010) New insights into the blood-stage
550 transcriptome of *Plasmodium falciparum* using RNA-Seq. Mol Microbiol 76: 12-24.
551 26. Peterson DS, Wellems TE (2000) EBL-1, a putative erythrocyte binding protein of *Plasmodium*
552 *falciparum*, maps within a favored linkage group in two genetic crosses. Mol Biochem
553 Parasitol 105: 105-113.
554 27. Nacer A, Roux E, Pomel S, Scheidig-Benatar C, Sakamoto H, et al. (2011) Clag9 is not essential for
555 PfEMP1 surface expression in non-cytoadherent *Plasmodium falciparum* parasites with a
556 chromosome 9 deletion. PLoS One 6: e29039.
557 28. Henriques G, Hallett RL, Beshir KB, Gadalla NB, Johnson RE, et al. (2014) Directional selection at
558 the pfmdr1, pfcrt, pfubp1, and pfap2mu loci of *Plasmodium falciparum* in Kenyan children
559 treated with ACT. J Infect Dis 210: 2001-2008.

560 29. Filarsky M, Fraschka SA, Niederwieser I, Brancucci NMB, Carrington E, et al. (2018) GDV1 induces
561 sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing.
562 Science 359: 1259-1263.
563 30. Brancucci NMB, Bertschi NL, Zhu L, Niederwieser I, Chin WH, et al. (2014) Heterochromatin
564 protein 1 secures survival and transmission of malaria parasites. Cell Host Microbe 16: 165-
565 176.
566 31. Carter R, Graves PM, Creasey A, Byrne K, Read D, et al. (1989) *Plasmodium falciparum*: an
567 abundant stage-specific protein expressed during early gametocyte development. Exp
568 Parasitol 69: 140-149.
569 32. Lu XM, Batugedara G, Lee M, Prudhomme J, Bunnik EM, et al. (2017) Nascent RNA sequencing
570 reveals mechanisms of gene regulation in the human malaria parasite *Plasmodium*
571 *falciparum*. Nucleic Acids Res 45: 7825-7840.
572 33. Fraschka SA, Filarsky M, Hoo R, Niederwieser I, Yam XY, et al. (2018) Comparative
573 heterochromatin profiling reveals conserved and unique epigenome signatures linked to
574 adaptation and development of malaria parasites. Cell Host Microbe 23: 407-420 e408.
575 34. Ntumngia FB, Bouyou-Akotet MK, Uhlemann AC, Mordmuller B, Kremsner PG, et al. (2004)
576 Characterisation of a tryptophan-rich *Plasmodium falciparum* antigen associated with
577 merozoites. Mol Biochem Parasitol 137: 349-353.
578 35. Morita M, Takashima E, Ito D, Miura K, Thongkukiatkul A, et al. (2017) Immunoscreening of
579 *Plasmodium falciparum* proteins expressed in a wheat germ cell-free system reveals a novel
580 malaria vaccine candidate. Sci Rep 7: 46086.
581 36. Robinson MD, Smyth GK (2007) Moderated statistical tests for assessing differences in tag
582 abundance. Bioinformatics 23: 2881-2887.
583 37. Balaji S, Babu MM, Iyer LM, Aravind L (2005) Discovery of the principal specific transcription
584 factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA
585 binding domains. Nucleic Acids Research 33: 3994-4006.
586 38. Kumar K, Srinivasan P, Nold MJ, Moch JK, Reiter K, et al. (2017) Profiling invasive *Plasmodium*
587 *falciparum* merozoites using an integrated omics approach. Sci Rep 7: 17146.
588 39. Doerig C, Rayner JC, Scherf A, Tobin AB (2015) Post-translational protein modifications in malaria
589 parasites. Nat Rev Microbiol 13: 160-172.
590 40. Jones ML, Collins MO, Goulding D, Choudhary JS, Rayner JC (2012) Analysis of protein
591 palmitoylation reveals a pervasive role in *Plasmodium* development and pathogenesis. Cell
592 Host Microbe 12: 246-258.
593 41. Poran A, Notzel C, Aly O, Mencia-Trinchant N, Harris CT, et al. (2017) Single-cell RNA sequencing
594 reveals a signature of sexual commitment in malaria parasites. Nature 551: 95-99.
595 42. Reid AJ, Talman AM, Bennett HM, Gomes AR, Sanders MJ, et al. (2018) Single-cell RNA-seq
596 reveals hidden transcriptional variation in malaria parasites. Elife 7.
597 43. Walzer KA, Kubicki DM, Tang X, Chi JT (2018) Single-cell analysis reveals distinct gene expression
598 and heterogeneity in male and female *Plasmodium falciparum* gametocytes. mSphere 3.
599 44. Lee HJ, Georgiadou A, Otto TD, Levin M, Coin LJ, et al. (2018) Transcriptomic studies of malaria: a
600 paradigm for investigation of systemic host-pathogen interactions. Microbiol Mol Biol Rev
601 82.
602 45. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, et al. (2009) Amplification-free Illumina
603 sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased
604 genomes. Nat Methods 6: 291-295.
605 46. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map
606 format and SAMtools. Bioinformatics 25: 2078-2079.
607 47. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, et al. (2013) Software for computing and
608 annotating genomic ranges. PLoS Comput Biol 9: e1003118.
609 48. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-
610 seq data with DESeq2. Genome Biol 15: 550.

611 49. Salanti A, Staalsoe T, Lavstsen T, Jensen AT, Sowa MP, et al. (2003) Selective upregulation of a
612 single distinctly structured var gene in chondroitin sulphate A-adhering *Plasmodium*
613 *falciparum* involved in pregnancy-associated malaria. Mol Microbiol 49: 179-191.

614

615

616 **Table 1.** *P. falciparum* genes in the top quartile of expression showing a difference in
617 schizont stage transcript levels between long-term laboratory-adapted clones and recent
618 clinical isolates in culture

Gene ID	Product annotation	Log_2 fold difference	Wald test
Higher transcription in long-term laboratory-adapted clones			
PF3D7_0422900	methyltransferase, putative	-2.50	-6.48
PF3D7_0501300	skeleton-binding protein 1	-2.50	-4.98
PF3D7_0410000	erythrocyte vesicle protein 1	-2.38	-10.62
PF3D7_0104300	ubiquitin carboxyl-terminal hydrolase 1, putative	-2.18	-6.13
PF3D7_0522300	18S rRNA (guanine-N(7))-methyltransferase, putative	-2.15	-8.67
PF3D7_0420300	AP2 domain transcription factor, putative	-2.02	-7.19
PF3D7_1030200	claudin-like apicomplexan microneme protein, putative	-2.00	-9.21
PF3D7_1327300	conserved *Plasmodium* protein, unknown function	-2.00	-4.97
Higher transcription in clinical isolates			
PF3D7_1036300	duffy binding-like merozoite surface protein 2	3.02	5.32
PF3D7_1302100	gamete antigen 27/25	2.02	3.99

619

620

621 **Figure 1.** Multiple replicate biological preparations of *P. falciparum* transcriptomes from
622 four long-term laboratory adapted clones and six Ghanaian clinical isolates, showing FPKM
623 values across all genes correlated with previous data from seven time-points across the *P.*
624 *falciparum* asexual erythrocytic cycle [25]. Each individual parasite culture was enriched for
625 schizont-stage parasites, with egress blocked using E64 treatment for 5.5 hours, and
626 parasites were purified using MACS purification for all of the lines, except for the D10 and
627 HB3 clones and ten of the 3D7 replicates which were purified using Percoll density gradient
628 centrifugation. Red lines plot data for samples with peak correlation at either 40 or 48 hours
629 post-invasion, and grey lines plot replicates that did not maximally correlate with either of
630 these timepoints and which were subsequently excluded from the analysis.
631
632 **Figure 2.** Distribution of correlations of FPKM expression values among biological sample
633 replicate transcriptomes for each isolate or clone. The final numbers of replicates, and
634 median Spearman's correlation coefficients among replicates were: 3D7 n=19, ρ=0.89; D10
635 n=10, ρ=0.93; Dd2 n=6, ρ=0.95; HB3 n=7, ρ=0.89; 271 n=3, ρ=0.94; 278 n=5, ρ=0.91; 280
636 n=6, ρ=0.93; 286 n=6, ρ=0.97; 293 n=6, ρ=0.93; 296 n=3, ρ=0.90.
637
638 **Figure 3. Increased sample replication improves discovery of differentially expressed**
639 **genes.** Assessment of the proportion of genes captured as being differentially expressed
640 between two different parasite clones (3D7 and D10), by taking 100 random samples of two,
641 four, six and eight replicates of each (out of ten initially analysed replicates that identified
642 123 genes with $\log_2$ differences of > 2 in relative transcript levels between the two clones).
643
644 **Figure 4 Differential gene transcript levels among preparations of schizont-enriched P.**
645 **falciparum cultures of laboratory-adapted clones and clinical isolates**. Plots of normalised
646 RNA-seq read counts for eight of the genes showing differences between four laboratory-
647 adapted clones and six clinical isolates. Each point shows the value for an independent
648 biological replicate sample. The two top left panels show examples of genes for which one
649 of the laboratory-adapted clones had a deletion (HB3 and D10 for the respective genes).
650 These, and other genes with suspected or known deletions in the sampled parasites, are
651 excluded from the list of differentially expressed genes in Table 1.
652

653 **Figure 5. High correlations between RNA-seq and RT-qPCR expression measures.** Eight
654 genes identified as differentially expressed among clinical isolates by RNA-seq were
655 validated by RT-qPCR for 49 of the independent RNA preparations from the four laboratory-
656 adapted clones and six cultured clinical isolates under study. Scatter plots for each gene
657 show the gene of interest (GOI) copies (log_2-transformed) normalised to house-keeping
658 gene copies (HKG; PF3D7_0717700), correlated with GOI FPKM values (log_2-transformed)
659 normalised to HKG FPKM values. Spearman's correlation coefficients are shown for each
660 plot. For PF3D7_1461700, one sample had an FPKM value of 0 which was not plotted but is
661 included in the calculation of correlation.
662
663 **Figure 6. Transcript levels of variably expressed genes for first cycle *ex vivo* clinical**
664 **samples compared with continuously cultured isolates.** For the laboratory and cultured
665 clinical isolates, each point represents the mean log_2 normalised expression value for
666 multiple replicates of each sample under study. For the *ex vivo* clinical isolates, each point
667 represents log_2 normalised expression values for the individual, unreplicated samples under
668 study. Horizontal bars indicate the mean log_2 normalised expression value for all samples
669 within the group.
670
671
672

673 **Supplementary Material**

674

675 **S1 Figure. Differential expression of merozoite invasion-related genes among schizonts**
676 **from different parasite cultures.** Distributions of read counts (normalised to library size) for
677 eight genes, for replicated laboratory-adapted and clinical isolate samples, showing data
678 from each replicate culture preparation of each strain.

679

680 **S2 Figure. Gene expression levels for eight genes differentially expressed among clinical**
681 **isolates.** Distributions of read counts (normalised to library size) for eight genes, showing
682 data from each replicate culture preparation of each strain.

683

684

685 **S1 Table.** Genes differentially expressed between clinical and laboratory lines by absolute
686 log2 fold > 2 (at least 4 fold difference on average). For genes among the top quartile of
687 expression values genome-wide (top 18 genes in the table), members of multigene families
688 and genes in which strain-specific deletions may be responsible for differences are
689 annotated with asterisks (*).

690

691 **S2 Table.** Log_2 fold changes of genes differentially expressed among pairwise comparisons
692 of six cultured clinical isolates with multiple schizont preparations of each, among genes
693 within the top quartile of expression overall.

694

695

696 **S1 File.** Masking of *P. falciparum* genome annotation file for polymorphic regions of genes.
697

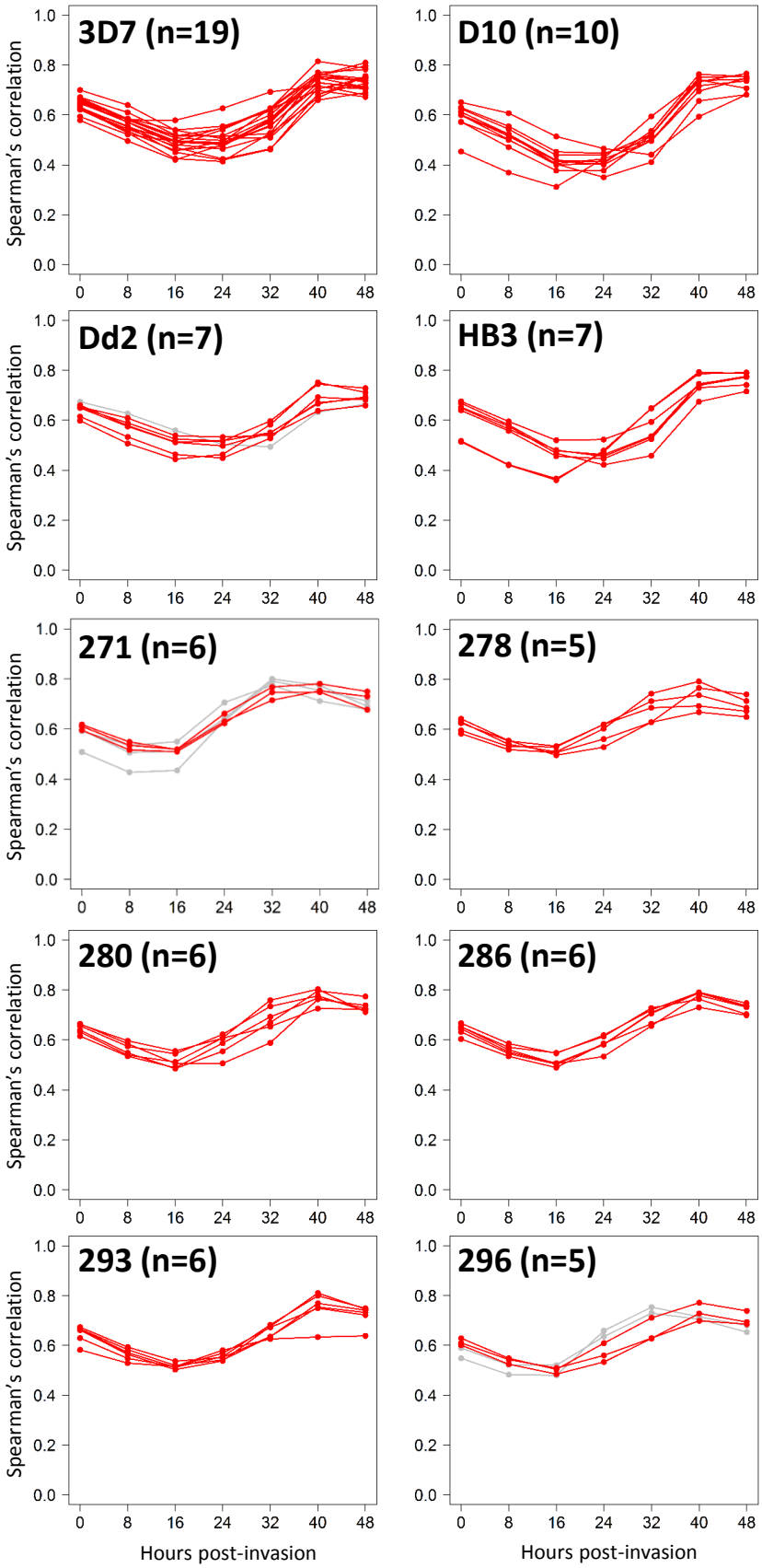
698 **S2 File.** Analysis of paired E64-treated and untreated *P. falciparum* 3D7 schizonts
699

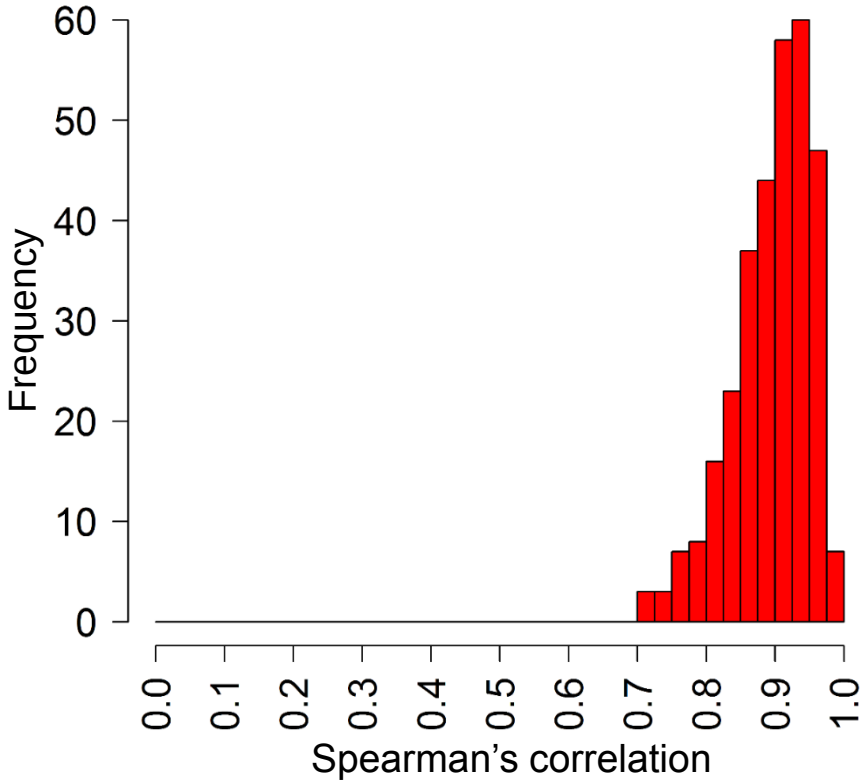
700 **S3 File.** Sequence annotation file Pf3D7.May2015.NoSplice.LSHTM.gtf (see S1 Table for
701 details)

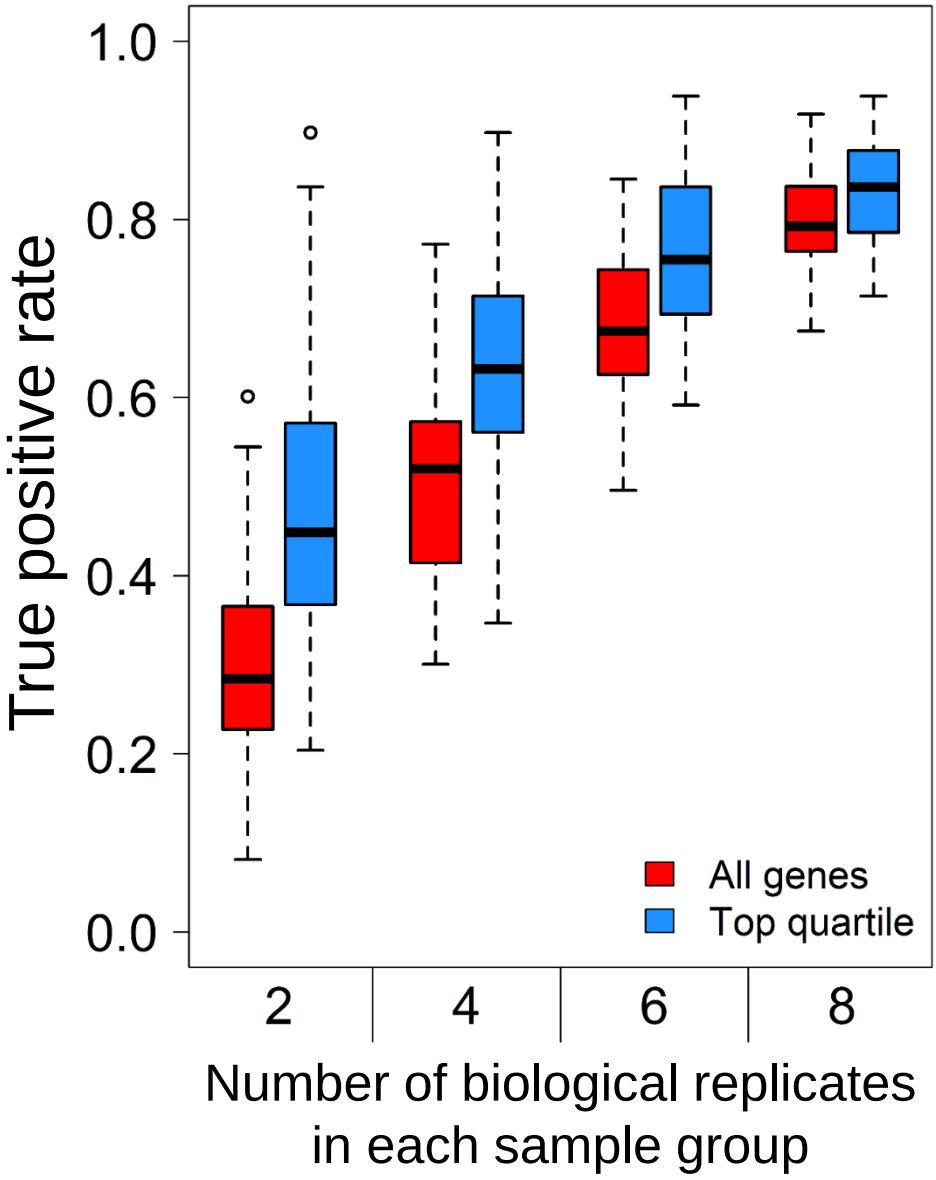
702

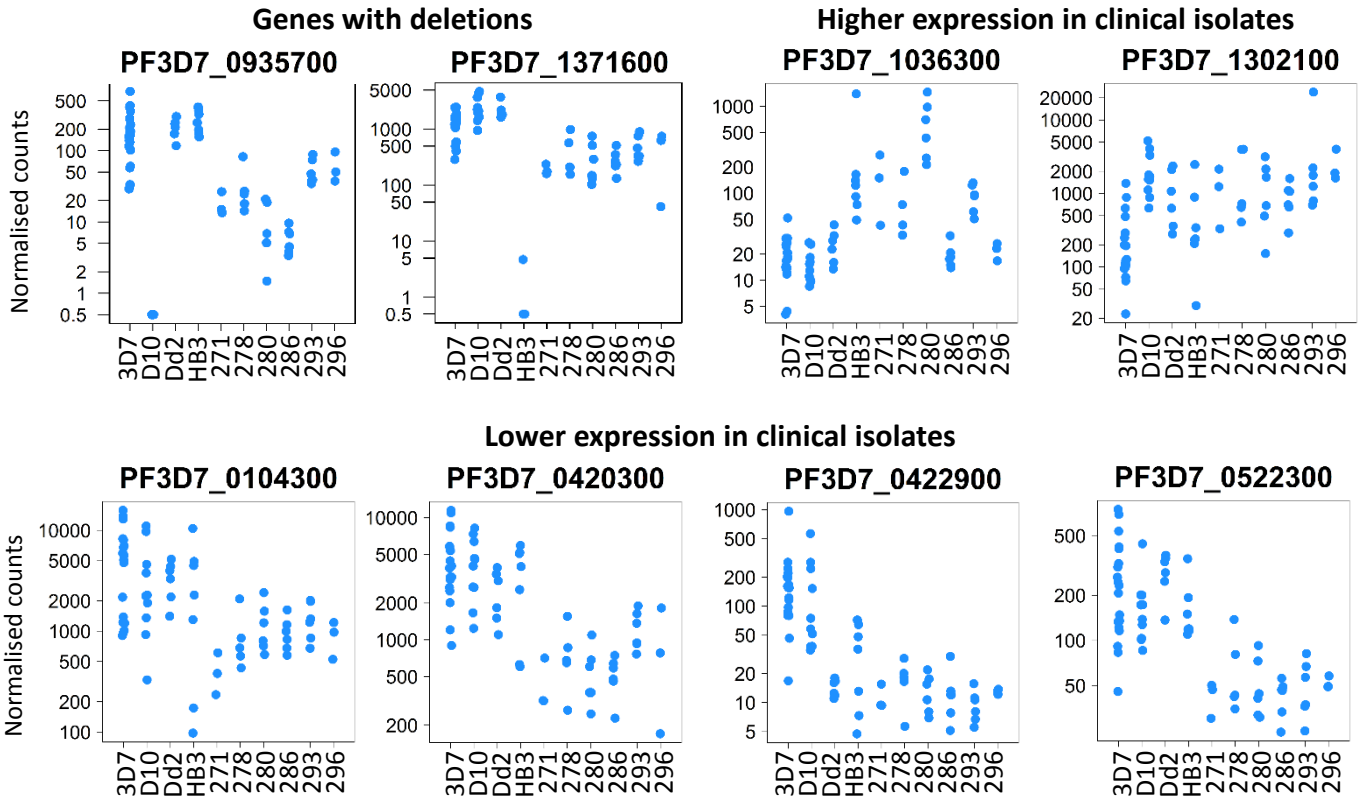
703 **S4 File.** Sequence annotation file GTF_VarRifStev_filtered out.gtf (see S1 Table for details)

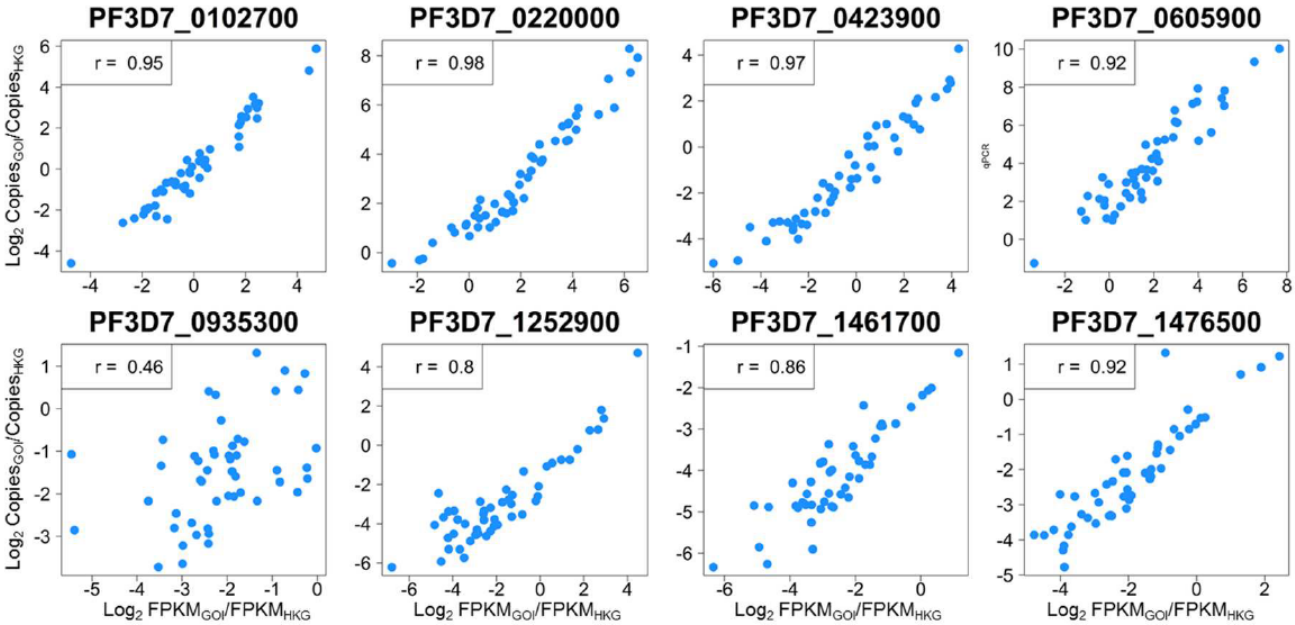
704

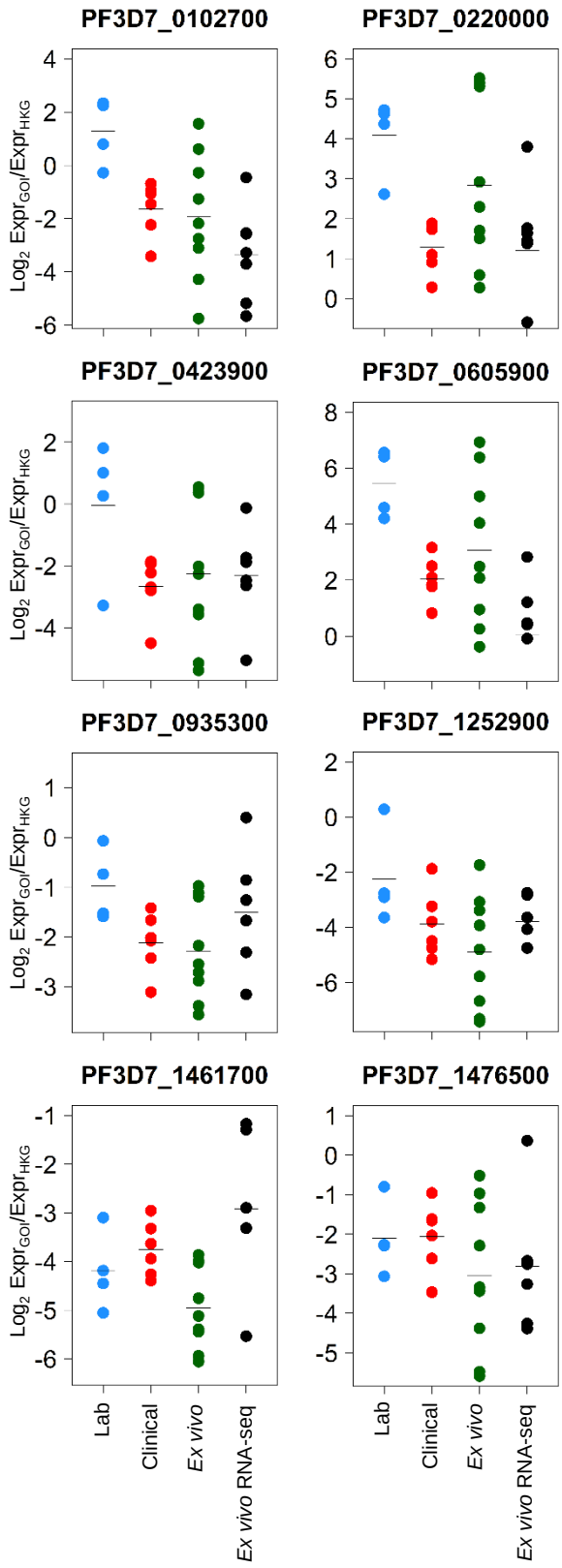

3D7 (n=19)
D10 (n=10)
Dd2 (n=7)
HB3 (n=7)
271 (n=6)
278 (n=5)
280 (n=6)
286 (n=6)
293 (n=6)
296 (n=5)
Spearman's correlation
Hours post-invasion
0.0
0.2
0.4
0.6
0.8
1.0
0
8
16
24
32
40
48


Frequency
60
50
40
30
20
10
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Spearman's correlation


1.0
0.8
0.6
0.4
0.2
0.0
True positive rate
All genes
Top quartile
2
4
6
8
Number of biological replicates
in each sample group


Genes with deletions
Higher expression in clinical isolates
Lower expression in clinical isolates
PF3D7_0935700
PF3D7_1371600
PF3D7_1036300
PF3D7_1302100
PF3D7_0104300
PF3D7_0420300
PF3D7_0422900
PF3D7_0522300
Normalised counts
3D7
D10
Dd2
HB3
271
278
280
286
293
296


PF3D7_0102700
r = 0.95
PF3D7_0220000
r = 0.98
PF3D7_0423900
r = 0.97
PF3D7_0605900
r = 0.92
PF3D7_0935300
r = 0.46
PF3D7_1252900
r = 0.8
PF3D7_1461700
r = 0.86
PF3D7_1476500
r = 0.92
Log2 CopiesGOI/CopiesHKG
Log2 FPKMGOI/FPKMHKG


PF3D7_0102700
PF3D7_0220000
PF3D7_0423900
PF3D7_0605900
PF3D7_0935300
PF3D7_1252900
PF3D7_1461700
PF3D7_1476500
Log2 ExprGOI/ExprHKG
Lab
Clinical
Ex vivo
Ex vivo RNA-seq