Single cell expression profile of the human malaria intracellular, extracellular and sexual merozoites

Short title: Single cell expression atlas of the human malaria merozoites

Laura Piel[1#], Celine Hernandez[1], Yan Jaszczyszyn[1], Delphine Naquin[1], Lucille Mellottée[1], Joana M. Santos[1*]

[1]Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France

#current affiliation: Institut Pasteur, Molecular Parasitology and Signaling Unit, Université de Paris, INSERM U1201, France

*Corresponding author: joana.santos@i2bc.paris-saclay.fr

Abstract

The intraerythrocytic developmental cycle (IDC) during which the malaria parasite *Plasmodium falciparum* multiplies asexually initiates when parasite stages called merozoites are released into the bloodstream, following schizont rupture, and invade the host red blood cells. The merozoite is the less studied stage of the IDC, despite its importance for the establishment of infection. We performed the first gene expression profiling of individual merozoites throughout development inside the schizont and after egress. This allowed us to identify for the first time the specific gene signatures of free merozoites, early and late stage intracellular merozoites. We also identified an earlier state of sexual commitment, preceding expression of the gametocytogenesis master regulator *ap2-g*.

Author Summary

The malaria parasite *Plasmodium falciparum*, which causes the death of over half a million people each year, has a complex life cycle but symptoms only arise when the parasite invades and develops inside the host red blood cells. The merozoites are responsible for invading the human red blood cells but because they are only briefly extracellular, their gene expression profile had yet to be determined. Here we used single cell RNA-sequencing to establish the gene expression program of merozoites throughout development inside schizonts and once free. We also identified for the first time the gene signature of sexually-committed and sexual merozoites that will later give raise to gametocytes and ensure parasite transmission.

39 **Introduction**

40 Malaria, responsible for over half a million deaths each year (1), is an infectious
41 disease caused by the unicellular eukaryotic parasite *Plasmodium*. Out of the five species of
42 parasite that cause disease in humans, *P. falciparum*, is the most virulent, causing the majority
43 of the yearly over half a million deaths (WHO). The parasite life cycle is complex, alternating
44 between the human host and the mosquito, that acts as both a host and a vector. The
45 intraerythrocytic developmental cycle (IDC) (Fig 1A), in which the parasite multiplies within the
46 human host red blood cell (RBC) for 48-hours, starts when parasite stages called merozoites,
47 invade the host RBC. After invasion, the parasite transforms into the ring stage and ~20 hours
48 post-invasion (hpi) into trophozoites. At ~30 hpi, new invasive forms called merozoites are
49 formed inside schizonts by schizogony. At the end of the IDC, 30-32 merozoites are released
50 into the bloodstream (2). In *P. falciparum*, the IDC ensures parasite multiplication but also
51 transmission because parasites must undergo at least one IDC in order to sexually commit
52 and form gametocytes that are later transmitted to the mosquito vector (3) (Fig 1A). Sexual
53 commitment only occurs in a small subset of cells and it is triggered by the expression of the
54 gametocyte master regulator AP2-G, a transcription factor (4,5).

55 The merozoite is the less studied stage of the IDC, despite its attractiveness as a drug
56 target given that (i) it is extracellular, even if briefly (6); (ii) it is responsible for RBC invasion,
57 which is indispensable for parasite survival; and, (iii) it initiates the IDC, which is the
58 symptomatic life cycle stage of malaria. Studies describing the gene expression profiling of
59 the different stages of the *P. falciparum* life cycle, lack information for the merozoite stage (7)
60 and, it is only recently that it was shown that merozoites are transcriptionally active (8). RBC
61 invasion is a fast process, lasting ~1 minute (9), and merozoites only last 5-15 minutes
62 extracellular (10), so when the newly formed merozoites egress from schizonts, they must be
63 already equipped with all the proteins they need to find and invade new RBCs. The genes
64 encoding those proteins should then be transcribed before, during merozoite maturation inside
65 schizonts. However, this hypothesis remains to be confirmed.

66 Here we used single cell RNA-sequencing to establish for the first time the gene
67 expression signature of *P. falciparum* merozoites during schizogony and upon schizont
68 rupture. We find that, during maturation inside the schizont, merozoites express nuclear, RNA-
69 binding and translation-related encoding genes, as well as invasion and host cell remodeling
70 genes. After schizont exit, we still detect these transcripts in free merozoites. We also identify
71 for the first time sexually committed merozoites that will later give raise to gametocytes and
72 ensure parasite transmission.
73 Our comprehensive study of parasite gene expression of the malaria merozoites
74 further sheds light into gene expression in the malaria parasite and unfolds for the first time
75 the single cell expression atlas of the human malaria merozoites, allowing the parasite to
76 invade the RBC.
77
78 **Results**
79
80 **Single cell RNA-sequencing of merozoite cells**
81 In order to identify the gene expression profile of the malaria merozoites, we adapted
82 a protocol described before for the isolation of viable invasive merozoites (10) (Fig 1B). The
83 parasite culture was passed through a magnetic column to retain the schizonts, which are
84 magnetic due to the accumulation of hemozoin, the end product of hemoglobin digestion. The
85 purified schizonts were then passed through a 1.2μM filter to release the merozoites inside.
86 Using this method, we could purify merozoites from schizonts as early as 32 hpi (S1A-B Figs).
87 To recover intracellular merozoites, as well as free merozoites, we single synchronized
88 parasites for two cycles, and when schizonts were visible in Giemsa-stained smears (S1C
89 Fig), we collected the culture. Because the culture was not tightly synchronized, it also
90 contained non-infected RBCs, ring-stage parasites and free merozoites (S1C Fig). These cells
91 are not retained by the magnetic column and were then recovered in the flow-through of the
92 magnetic purification. The purity of each fraction was verified by Giemsa-stained blood smears

93 (S1D-E Figs). The merozoite and the flow-through fractions were prepared separately for
94 single cell RNA-sequencing using the 10X Genomics technology.
95 We obtained sequencing data for 1580 isolated merozoite cells and 405 flow-through
96 cells (S1 Table). To obtain a similar number of cells in each fraction, we sequenced flow-
97 through cells coming from a second magnetic purification of a second parasite culture
98 containing more rings and free merozoites and less schizonts (Fig 1B). We obtained a total of
99 1464 cells in the combined flow-through fractions (S1 Table). We acquired a similar number
100 of reads (over 200 million), obtained a high percentage of sequencing saturation (over 96%)
101 and detected a similar number of genes (over 4800, ~90% of the parasite genes) for the cells
102 in the merozoite and flow-through fractions (S1 Table). Smaller cells have less total
103 messenger RNA (11) and we detected about ten times less transcripts per cell of the merozoite
104 fraction than per cell of the flow-through fraction (S1F-H Figs), in accordance with previous
105 studies (7,12). For this reason, whereas for the flow-through samples, we filtered out cells with
106 fewer than 200 genes and over 2500 genes per cell, a common threshold to analyze *P*.
107 *falciparum* single cell data, we used a different filter for the merozoite fraction. We excluded
108 from the merozoite fraction, cells with fewer than 30 genes and more than 2500 genes per cell
109 and less than 80 reads per cell. Similar parameters were used to analyze *P*. *berghei*
110 merozoites (7). After filtering, we retained 968 cells in the merozoite fraction and 1421 cells in
111 the combined flow-through fractions (S1 Table).
112
113 **Identification of free merozoites**
114 To ensure that the cells collected in the merozoite fraction are in fact merozoite-stage
115 parasites, we determined if these cells express known merozoite genes. Merozoite surface
116 protein 1 (MSP1) is the most abundant merozoite surface protein (13) and the encoding gene
117 (PF3D7_0930300) is over-expressed in merozoite cells versus ring cells, as determined by
118 Reers *et al*. (8) that defined a list of genes that are over-expressed in merozoite cells versus
119 rings and schizont cells. We thus used *msp1* as a merozoite marker. *msp1* is expressed by
120 the merozoite fraction cells and these cells do not express the ring marker *ring-exported*

121 *protein 2* (*rex2*, PF3D7_0936000) (Fig 2A and S2A Fig). On the other hand, the majority of
122 flow-through cells expresses *rex2* (Fig 2A and S2A-B Figs) and are therefore ring-stage
123 parasites. Moreover, 59% of the genes expressed by the merozoite fraction cells encode for
124 proteins identified as part of a merozoite proteome (14) (Fig 2B) and, with one exception, none
125 of the genes expressed in merozoite cells were previously found by (8) to be over-expressed
126 in schizonts versus merozoite cells (Fig 2B, S2 Table). 38% and 47% of the genes expressed
127 by the merozoite cells were also previously shown by (8) to be over-expressed in merozoites
128 versus ring and schizont stage cells, respectively (Fig 2B, S2 Table).

129 Having established that the merozoite fraction corresponds to merozoites produced
130 during schizogony, we wanted to identify free, extracellular merozoites. These egressed from
131 schizonts before we did the magnetic purification and were collected in the flow-through. Free
132 merozoites and ring-stage cells, corresponding to 89% of the cells in the flow-through fraction
133 (Fig 2A), are expected to express different genes so we used Seurat, a toolbox for quality
134 check, analysis and visualization of single cell data (15–17), to identify a cell cluster, in the
135 flow-through fraction, corresponding to the free merozoites. We identified eight expression
136 clusters for the flow-through cells (S2C-D Fig). Differential expression analysis was done to
137 determine which genes are differentially expressed in one particular cluster versus the others.
138 These genes are herein called marker genes. Gene ontology (GO) analysis of the marker
139 genes of each flow-through cluster shows that cluster 0 cells express genes specialized in
140 host entry and motility (Fig 2C, S2 and S3 Tables), functions associated with invasion and so
141 with merozoites. As expected from merozoite cells, the cluster 0 cells also contain less RNA
142 than other flow-through cluster cells (S2E Fig). This indicates that the cluster 0 cells represent
143 the gene expression profile of free merozoites. We identified 66 free merozoite cells in the two
144 flow-throughs, corresponding to less than 5% of the population, as expected from an elusive
145 stage that had never before been captured by RNA-sequencing studies.

146

147 **Gene expression signature of intracellular merozoites throughout schizogony**

148 To identify the gene expression signature of intracellular merozoites, we applied
149 Seurat to the cells collected in the merozoite fraction of the magnetic purification. We identified
150 ten merozoite clusters (Fig 3A, S3A Fig). Clustering was independent of the number of genes
151 detected per cell or the cells RNA content (S3B Fig) and each merozoite cluster is
152 characterized by expression of, on average, 49 distinct marker genes (S2 Table).
153 Pseudotemporal analysis indicates that the merozoite cells, during schizogony, follow one
154 developmental trajectory (S3C Fig), as anticipated, since all merozoites egress and invade
155 RBCs. About half of the marker genes of each merozoite cluster encode for unknown or
156 putative proteins (S3D Fig), which is not unexpected since about 30% of the parasite proteins
157 remain unannotated (18). Despite this, in all clusters, the most differentially expressed genes
158 encode known proteins (S2 Table) so we did GO-term analysis of the annotated genes to
159 assign a label to each merozoite cluster. This analysis show that each cluster is characterized
160 by the expression of a specific category of marker genes (Fig 3B; S2 Table). While clusters 5,
161 6 and 7 express marker genes related to host cell invasion, clusters 1, 8, 9 and 10 marker
162 genes are related to nuclear processes, RNA binding and protein synthesis. Clusters 2, 3 and
163 4 express marker genes encoding for proteins stored in the dense granules that are later
164 discharged and participate in events taking place post-invasion - formation of the
165 parasitophorous vacuole (PV), wherein the parasite resides after invasion, formation of the
166 *Plasmodium* translocon of exported proteins (PTEX), in the PV membrane, through which
167 parasite proteins are exported into the host cell, and host cell remodeling, allowing the parasite
168 to import media nutrients. Because some of the merozoite clusters are enriched for the same
169 GO-terms, we compared them but we found none to little overlap, except for clusters 5 and 6,
170 related to invasion (S3E Fig). However, the invasion clusters 5 and 6 are different from the
171 other invasion cluster, cluster 7 (S3E Fig), and marker genes of a specific invasion cluster are
172 over-expressed in that cluster versus the other two invasion clusters (S3F Fig).
173 Since the culture used for the single cell RNA-sequencing was not tightly synchronized
174 and magnetic purification followed by syringe filter allows extraction of merozoites from as
175 early as 32hpi, it is likely that we captured, in the merozoite fraction, merozoites going through

176 schizogony (30-48 hpi). To examine this, we calculated the pseudotime of the different
177 merozoite clusters and found that a different pseudotime is attributed to each cluster (Fig 3C;
178 S1 Table). We also compared our results to a published real-time expression data (19) to
179 determine the time of peak transcription and stability of the marker genes of each cluster and
180 find that not all genes reach peak transcription or stability at the same time (S3G Fig; S2
181 Table). Interestingly, the clusters 8, 9 and 10 marker genes, encoding RNA-binding and
182 translation proteins, reach peak transcription between 12hpi and 30 hpi, so during the ring and
183 trophozoite stages, and reach peak stability between 18 and 30 hpi, at the trophozoite stage
184 (S3G Fig; S2 Table). Moreover, they are more expressed by ring cells than by merozoites
185 (S3H Fig, S2 Table) according to Reers *et al*. (8). That is not the case of the marker genes of
186 the other clusters (S3H Fig; S2 Table). This indicates that these marker genes, while
187 expressed by merozoites, continue to be expressed afterwards during ring development. All
188 of these observations suggest that we collected, in the merozoite fraction, intracellular
189 merozoites of different ages. To test this experimentally, we did single cell RT-qPCR on
190 individual merozoites collected at three different time points (32 hpi, 42 hpi and 47hpi) (S3I
191 Fig). We selected 10 marker genes that reach peak expression in a specific cell cluster (Fig
192 3D) and designed primers for 7 of the genes (S4 Table): *hsp70* (PF3D7_0818900), cluster 1
193 marker gene; *bdp2* (PF3D7_1212900), cluster 2 marker gene; *eba-175* (PF3D7_0731500),
194 cluster 5 marker gene; *gap45* (PF3D7_1222700), cluster 6 marker gene; *msp1*
195 (PF3D7_0930300), cluster 7 marker gene; and *60S L4* (PF3D7_0507100), cluster 9 marker
196 gene. The other 3 genes were determined to be too small or difficult to amplify. For each
197 merozoite (Fig 3E), we quantified expression of two marker genes and scored whether
198 merozoites expressed only one of the marker genes, none of the marker genes or over-
199 expressed one of the transcripts versus the other. We also quantified expression of a transcript
200 previously used as reference for RT-qPCR assays (20) and we only considered cells in which
201 we detected transcription of the reference gene. The percentage of cells only expressing one
202 gene or overexpressing that gene versus another varies, with time, for all the genes pairs we
203 compared (Fig 3E). For instance, cells over-expressing the cluster 1 marker gene versus the

204 cluster 7 marker gene are already detected at 32 hpi, whereas cells over-expressing the
205 cluster 7 marker gene versus the cluster 1 marker gene only start being detected at 42 hpi
206 (Fig 3E, S4 Table). These results show that the different marker genes reach peak expression
207 at different time points and so that we collected merozoite cells throughout development.

208 **Identification of specific merozoite types**

209 The *P. falciparum* parasite strain we used, W2mef, invades RBCs in a sialic-acid
210 dependent manner but W2mef parasites can switch to the sialic-acid independent invasion
211 pathway, if selected to invade RBCs depleted of sialylated surface proteins upon
212 neuraminidase treatment (21). This switch capacity is correlated with the expression of the
213 parasite ligand Rh4 (21). *rh4* is only expressed by a subset of the W2mef parasites (~6% of
214 the population) (22) and we wondered if we could identify these cells in our dataset. To do
215 that, we quantified the number of cells expressing genes encoding ligands either binding to
216 sialylated host cell receptors (*eba-175*, PF3D7_0731500; *eba-140*, PF3D7_1301600; *eba-*
217 *181*, PF3D7_0102500; and *rh1*, PF3D7_0402300) or to non-sialylated host cell receptors
218 (*rh2a*, PF3D7_1335300; *rh2b*, PF3D7_1335400; and *rh4*, PF3D7_0424200). Whereas 86%
219 of intracellular merozoites express at least one gene encoding for a sialic-acid-dependent
220 ligand, we could not detect *rh2b* expression, and only 8,5% of the population expresses *rh2a*
221 or *rh4* implicated in the sialic-acid-independent invasion pathway (Fig 4A). As expected, *rh4*
222 is only expressed by 6,5% of intracellular merozoites.

223 We also wanted to identify, in the merozoite fraction, the intracellular merozoites that
224 are about to egress from the infected host cell. They are expected to have a similar gene
225 repertoire to free merozoites, which we identified in the flow-through fraction, so to identify
226 these cells, we combined the data from the merozoite and two flow-through cells and filtered
227 all cells using the flow-through parameters (Fig 4B). There are 191 merozoite cells in the
228 integrated dataset (S1 table). These cells are here called late stage merozoites. We identified

229 the genes that were over-expressed in the late stage merozoites versus the flow-through cells
230 (Fig 4C, S5 Table), in order to establish a set of genes to be used as signature, and found 133
231 genes, coding for proteins implicated in host cell invasion, motility, host cell interaction and
232 exit (Fig 4D, S5 Table). All of the over-expressed genes encode proteins identified in the
233 merozoite proteome (S5 Table). Moreover, many of the late stage merozoites signature genes
234 reach peak transcription and stability at 1 or 48 hpi, according to (19), as the merozoites are
235 preparing to egress/invade (S4AFig; S5 Table). This tendency is also seen for the marker
236 genes of free merozoites (S4AFig; S2 Table) and about half of the free merozoites marker
237 genes are signature genes of the late stage merozoites (S4B Fig). As a final confirmation of
238 these cells identity as late merozoites cells, we find that they tend to have a later pseudotime
239 than other intracellular merozoites (Fig 4E), as determined by the merozoite trajectory analysis
240 (Fig 3C; S1 Table).
241
242 **Identification of sexual and sexually committed merozoites**

243 Sexually committed merozoites, which will later give raise to male and female
244 gametocytes have never been identified. Because *P. falciparum parasites* must invade red
245 blood cells before sexually committing (3), some of the free merozoite cells we identified
246 should be sexually committed. To determine which free merozoites are sexual or sexual, we
247 integrated the two flow-through datasets and applied Seurat to it (Fig 5A, S5A Fig, S6 Table).
248 Within the integrated dataset, cells over-expressing the free merozoite marker genes are
249 found within clusters 4 and 13 (Fig 5B, S5B Fig, S6 Table). These cells also contain less RNA
250 and encode less transcripts than other cells (S5C Fig), express the merozoite marker *msp1*
251 (S5D Fig) and encode invasion-related genes (Fig 5C, S6 Table), which are all hallmarks of
252 extracellular merozoites. On the other hand, the other cells of the integrated dataset express
253 the *rex2* ring marker gene and do not express free merozoite marker genes and are therefore
254 rings (S5E Fig). There are few marker genes in common between the clusters 4 and 13 and
255 all of these are signature genes of free merozoites (Fig 5E, S6 Table), indicating that the cells

256 in the two clusters are distinct. Because the cluster 5 cells encode sexual genes (Fig 5C, S6
257 Table) and several of the cluster 12 marker genes encode proteins that have previously been
258 implicated in gametocytogenesis (S6 Table), namely, the ApiAP2 proteins AP2-P, AP2-G5
259 and AP2-Z (12,23,24), we conclude that the free merozoites in cluster 13 are sexually-
260 committed cells (Fig 5F). On the other hand, the free merozoites in cluster 4 are asexual (Fig
261 5F). As expected, there are more asexual free merozoites than sexual free merozoites (S1
262 Table). The branching out of cells into sexual commitment is supported by the trajectory
263 analysis (S5F Fig). The cluster 5 cells over-express bona fide sexual markers (of the 35
264 marker genes, 25 were previously identified as sexual genes (12,25) (S6 Table)) and are thus
265 here called sexual cells (Fig 5F). The cluster 13 cells are herein named committed cells (Fig
266 5F).

267 We then searched for sexual and sexually committed intracellular merozoites in the
268 integrated dataset of the merozoite and flow-through cells used before to identify the late stage
269 merozoites. To do that, we applied Seurat to the integrated dataset (S5G Fig, S1 Table). We
270 identified cells expressing commitment genes in cluster 10 and cells expressing sexual marker
271 genes in cluster 7 (Fig 5G, S5 Table). 24 of the 69 committed cells and 7 of the 134 sexual
272 cells are merozoites (Fig 5G, S1 Table). 6 genes are differentially expressed by sexually
273 committed merozoites and flow-through cells (S5H Fig, S5 Table), suggesting that committed
274 merozoites have a distinct gene expression signature. There are no differentially expressed
275 genes between sexual merozoites and sexual flow-through cells, likely due to the reduced
276 number of sexual merozoites identified (S5 Table). Few of the marker genes of committed
277 cells were previously reported as signature genes of sexually committed parasites or early
278 gametocytes (25,26) (Fig 5H), and *ap2-g* is not expressed in these cells (Fig 5I), indicating
279 that we identified here an earlier state of sexual commitment characterized by the expression
280 of cell cycle and chromatin-related genes (Fig 5J, S5 Table).

281

282 **Regulation of gene expression during merozoite development**

283 There is a good correlation between the time of peak transcription and of peak stability
284 for most merozoite genes (S3G Fig; S2 Table), as determined by (19), suggesting that gene
285 expression during merozoite development is regulated by the binding of transcription factors
286 to specific DNA motifs in the promoter sequences. The ApiAP2 DNA-binding proteins are the
287 main family of transcription factors in the parasite (27) so we searched for ApiAP2 proteins
288 implicated in regulation of merozoite genes expression. The ApiAP2 proteins AP2-I, AP2-LT,
289 AP2-P and AP2-EXP and the bromodomain protein BDP1 have been linked to expression of
290 invasion genes (20,24,28–30). To determine if these proteins regulate any of the merozoite
291 gene clusters related to invasion (clusters 5, 6 and 7), we quantified the percentage of AP2-I,
292 AP2-LT, AP2-P, AP2-EXP and BDP1 target genes, as previously determined by ChIP-Seq
293 (20,24,28–31), in each of the invasion clusters. 82% of the promoters of the cluster 7 marker
294 genes are bound by AP2-I, BDP1 or a combination of the two proteins (S5A Fig) and several
295 of these regions are also bound by AP2-LT, AP2-EXP and AP2-P (Fig 5A; S2 Table). On the
296 other hand, few of the cluster 5 or 6 gene promoters are bound by these proteins (Fig 5A, S2
297 Table). The fact that the cluster 7 marker genes may be regulated differently from the cluster
298 5 and 6 marker genes is supported by the fact that there are differences regarding the time of
299 peak transcription and peak stability of the different invasion cluster marker genes, as
300 established by Painter *et al*. (19). Over 60% of the cluster 7 marker genes are peak transcribed
301 or stabilized between 36 and 48 hpi, i.e., during merozoite development (S5B-C Figs; S2
302 Table). On the other hand, transcription and stability of the clusters 5 and 6 marker genes
303 peaks instead at 1hpi or 48hpi (S5B-C Figs; S2 Table), so at the transition between merozoite
304 egress and host cell invasion.

305 Analysis of published ChIP-Seq data did not identify any other proteins associated with
306 regulation of expression of the merozoite genes so we searched for enrichment of DNA
307 sequence motifs in the promoters of the genes. We identified one DNA motif, enriched in the
308 promoters of the marker genes of clusters 3, 5, 6 and 9, with an AC core, that we named
309 merozoite DNA motif 1 (Fig 5B, S7 Table). A *cis*-regulatory element very similar to merozoite

310 DNA motif 1 was previously identified *in silico* (32), indicating that merozoite DNA motif 1 is
311 frequently found in gene promoters. Comparison of merozoite DNA motif 1 with the database
312 of DNA motifs bound by purified GST-tagged AP2 domains *in vitro* (27), identified AP2-G5
313 (S5D Fig, S7 Table). However, there is no correlation between the presence of merozoite DNA
314 motif 1 and the binding of AP2-G5. 35% of the promoters of clusters 4 and 5 are bound by
315 AP2-G5, as determined by ChIP-Seq (33) (S5E Fig, S2 Table), but we only identified
316 merozoite DNA motif 1 upstream of the genes of cluster 5 (Fig 5B). Upstream of the marker
317 genes of the other clusters, we identified AT-rich DNA motifs, reflecting the composition of the
318 genome of the parasite (S5F Fig).

319 We also determined which *apiap2* genes are expressed by merozoites (Fig 5C, S2
320 Table). AP2-P and AP2-LT, binding upstream of the marker genes of cluster 7, are expressed
321 in merozoites. We also detect expression of AP2-O and AP2-L. The first has been linked to
322 expression of the variant *var* genes in *P. falciparum* (34). SIP2 and AP2-Tel, which in *P.*
323 *falciparum* bind to telomere ends (35,36), are also expressed in merozoite cells. Free
324 merozoites (cluster 0) express almost all *apiap2* genes expressed by intracellular parasites,
325 as well as 10 other genes that we did not detect in intracellular merozoites (Fig 5C, S2 Table).
326

327 **Discussion**

328 Here we used single cell RNA-sequencing to identify the gene expression profile of the
329 RBC invasive stages of the malaria parasite, the merozoites. The merozoites were collected
330 using magnetic purification and a specific filter (10). We validated this approach by showing
331 that merozoites can be extracted from within schizonts as early as 32 hpi, that merozoite cells
332 express the merozoite marker *msp1*, and that several of the merozoite genes identified here
333 encode previously identified merozoite proteins (14). This is only the second time that
334 merozoites are shown to be transcriptionally active. The first study to show this was Reers *et*
335 *al*. (8). 93 or 95% of the genes we define here as expressed by late stage merozoites or
336 extracellular merozoites, respectively, were identified in that study as being over-expressed in
337 merozoites. On the other hand, only 61% of the genes expressed by all intracellular

338 merozoites were previously defined as being over-expressed in merozoites (S6G Fig). This
339 indicates that the early merozoite genes remained understudied and that our work revealed
340 the gene repertoire of both early and late stage intracellular merozoites. This is also supported
341 by the results of the single cell RT-qPCR assays and the fact that all the marker genes of late
342 stage merozoite genes encode proteins identified as part of the merozoite proteome (14),
343 whereas only 59% of the intracellular merozoite marker genes do (S2 and S3 Tables).

344 We show that many of the merozoite invasion cluster 7 genes are bound by AP2-I,
345 BDP1, AP2-LT and AP2-P, suggesting that expression of these genes is regulated by those
346 proteins. We identified in the promoter region of some of the merozoite marker genes, a DNA
347 motif here called merozoite DNA motif 1. It remains unknown what protein binds to merozoite
348 DNA motif 1 and if this motif is important for expression of the genes. Further studies aimed
349 at identifying the regulatory mechanisms implicated in expression of the merozoite genes
350 should clarify this.

351 Tripathi *et al.* showed variation in gene expression between merozoites in schizonts of
352 the same age (37). By comparing our data to theirs we find that, on average, merozoite genes,
353 expressed by both intracellular and extracellular merozoites, have a variability index score of
354 1, suggesting that their expression is stochastic (S6H Fig, S2 Table). This may explain the
355 diversity of chromatin opening patterns we found, even for genes expressed by the cells of
356 the same cluster, with the exception of the marker genes of invasion cluster 7 (S2 Table),
357 when we compared our dataset to a published ATAC-Seq study (38). Cells in which the
358 chromatin is accessible activate gene expression while cells with closed chromatin do not.
359 That is the case of the malaria clonally variant gene *eba-141*, over-expressed by the cluster 5
360 cells, for which increased promoter accessibility leads to increased gene expression (39).
361 Random fluctuation of gene expression in individual cells is a remarkable source of variability
362 (40) and establishes cell identity (41). In the case of malaria, stochasticity of gene expression
363 provides flexibility to the parasite population and ensures productive invasion (42). As
364 calculated here, 6% of the merozoite population of the W2mef parasites expresses the *rh4*
365 ligand, essential for the sialic-acid invasion dependent pathway, and that is why this strain is

366 able to switch invasion pathways. Other genes that are only expressed in a subset of
367 merozoites may dictate other behaviors.

368 When examining gene expression of sexually-committed cells, we uncovered a primal
369 state of parasite sexual commitment, preceding expression of *ap2-g* and other known
370 gametocyte markers. Future exploration of these cells will allow discovery of proteins
371 implicated in triggering of sexual commitment.

372 In summary, we identified here early and late stage intracellular merozoites, sialic-acid
373 dependent and sialic-independent cells, extracellular and intracellular merozoites, sexually-
374 committed and sexual cells, despite the fact that some of these cells are rare and represent
375 as few as 2% of all sequenced cells (Fig 6D). For each of these cells, we establish their gene
376 expression profile. Our work completes the single cell gene expression atlas of *P. falciparum*,
377 which did not yet contain information for the merozoite stage.

378

379 **Materials and Methods**

380 **Parasite culturing and merozoites purification**

381 *P. falciparum* W2mef parasites were cultured at 37°C in the presence of 5% oxygen and 5%
382 carbon dioxide, in RPMI 1640 media (Fisher) supplemented with hypoxanthine and 0.5%
383 Albumax II (Invitrogen) in 2% hematocrit. Blood was obtained from the Etablissement Français
384 du Sang (EFS).

385 For scRT-qPCR or to determine when merozoites could be released from schizonts, parasites
386 were double synchronized to the ring stage, 3 hours apart, for several cycles by incubating
387 pelleted cultures 1:1 with 0.3M L-Alanine, 10mM HEPES pH 7.5 solution for 10 min at 37°C
388 (43), after which cultures were centrifuged, resuspended in fresh media and placed back in
389 culture. Then, starting at 30 hours post-invasion, schizonts were passed through LD columns
390 (Miltenyi Biotec). The schizonts were eluted from the column with RPMI 1640 and then passed
391 through the 1.2µm filter (Pall) attached to a 10ml syringe to release the merozoites inside as
392 described in (10).

393 For scRNA-seq, a mixed culture of non-infected RBCs, schizonts, rings and free merozoites
394 were magnetically purified. The flow-through, corresponding to ring stage parasites, free
395 merozoites and non-infected red blood cells, was recovered. The elution corresponds to the
396 merozoite fraction. Parasites were counted with a Neubauer chamber and diluted with
397 1XPBS/0.04%BSA to 1 million cells/ml.
398 The purity of each fraction was verified by Giemsa-stained blood smear or by DAPI staining.
399 Images of blood smears were acquired with the widefield microscope Eclipse (Nikon) at 100x
400 amplification. To stain with DAPI, purified schizonts or merozoites were fixed in
401 1XPBS/4%formaldehyde, washed and deposited into glass slides mounted with VectaShield
402 Antifade medium containing DAPI (Vector laboratories) and imaged with a SP8 confocal
403 microscope (Leica) at 63x amplification. Images were treated with ImageJ.
404
405 **Single cell RNA-Sequencing**
406 Approximately 10.000 cells (merozoites or flow-through) were loaded on a Chromium Chip G
407 (10x Genomics) as per the manufacturer's recommendations. Chips containing cell
408 suspensions, gel beads in emulsion (GEMs) and reverse transcription reagents were loaded
409 into the chromium controller (10x Genomics) for single cell partitioning and cell barcoding.
410 Barcoded cDNA and sequencing libraries were generated according to the Chromium Next
411 GEM Single Cell 3' gene expression v3.1 protocol. Quantification and the quality of
412 sequencing libraries was verified on an Agilent Bioanalyzer. Pooled libraries were sequenced
413 on a NextSeq 2000 system (Illumina).
414
415 **Data analysis**
416 Quality control checks on the raw sequencing data were performed using FASTQC (Andrews,
417 2010). FastQ files were analyzed using the CellRanger software (10X Genomics), including
418 alignment, filtering and quantitation of reads on the *Plasmodium falciparum* genome (3D7)
419 and generation of feature-barcode count matrices. During mapping, the reads corresponding
420 to the human host cell were discarded.

421 Genes detected in fewer than three cells were removed. Cells with more than 0.5% of
422 mitochondrial genes were removed. Cells with fewer than 30 genes and more than 1500 genes
423 per cell and less than 100 reads per cell were removed from merozoites. This threshold was
424 chosen because a previous single cell data (37) had shown that schizonts express an average
425 of 1040 genes, which for a schizont producing 30 merozoites, represents 34 genes per
426 merozoite assuming all merozoites express the same number of genes. Cells with fewer than
427 200 genes and more than 2500 genes per cells were removed from rings and merozoites
428 (when compared to rings).
429 The Seurat toolbox version 4.3.0.1 was used to normalize each dataset using SCTransform
430 (44) and cluster the cells. Visualizations and identification of differentially expressed genes
431 were done using the Loupe browser (10X Genomics, version 7).
432 Pseudotime was calculated using Monocle3 R package (45) to organize the merozoite and
433 ring cells.
434 Motif enrichment analysis was done using STREME (46) to identify enriched motifs in the
435 promoter regions (downloaded from PlasmoDB) as compared to scrambled sequences of
436 similar length. TOMTOM (47) was used to compare the *de novo* discovered motifs to
437 previously *in silico* discovered motifs (27).
438 The ATAC-Seq analysis were done by comparing the genes expressed in each cluster to
439 those identified in (38). Whenever more than one peak was assigned to a gene promoter, we
440 only considered the earlier time point. For most genes, there was only one peak of
441 accessibility.
442 The transcription and RNA stability peak time of the marker genes expressed in each cluster
443 was calculated in (19).
444 To determine the percentage of genes encoding proteins detected in mature merozoites by
445 mass spectrometry we compared our dataset to the high-throughput LC-MS/MS of *P.*
446 *falciparum* 3D7 merozoites data generated by Shahid Khan from Leiden University that
447 identified over 1000 proteins (14). We only considered proteins detected with at least 2
448 peptides.

449 The ChIP-Seq analysis were done by comparing the genes expressed in each cluster to the
450 target genes identified in (20,24,28–31,33). Whenever ChIP-Seq data was available from
451 multiple sources, we only considered target genes if found in two datasets.
452 To determine if the cells recovered in the second magnetic purification flow-through express
453 the same expression clusters as the ones recovered in the first purification flow-through, we
454 looked for cells expressing DE genes of each of the flow-through clusters and then manually
455 identified them as belonging to that certain cluster.
456 Venn diagrams were generated using the online tool BioVenn (48).
457 Gene Ontology (GO) enrichment analysis was done using the GO feature at PlasmoDB (49).
458 Orthologue searches were done using OrthoMCL (50).
459 Graphs and statistics were done using GraphPad Prism version 10.0.0 for Mac, GraphPad
460 Software, Boston, Massachusetts USA, www.graphpad.com.
461
462 **Single cell RT-qPCR**
463 The assay was done as described in (51) with few modifications. Briefly, single parasites
464 diluted in water were deposited into 96-well PCR plates containing saponin, water and dNTPs.
465 RT buffer, Ribolock, Maxima RT (Thermo Fischer Scientific) and random hexamers were
466 added to a final volume of 15µl and the plate was incubated at 25°C. The RT reaction was
467 split into 3 qPCR plates and subjected to qPCR using Sso fast Eva Green (BioRad). The
468 primer sequences as well as the qPCR results are available in S4 Table.
469
470 **Data availability**
471 All RNA sequencing data generated in this study has been deposited in NCBI's Gene
472 Expression Omnibus and is accessible through GEO series accession number GSE279639.
473
474 **Acknowledgements**

W2mef parasites were obtained through BEI Resources, NIAID, NIH: *Plasmodium falciparum*, Strain W2mef, MRA-165, contributed by Alan F. Cowman. The PlasmoDB.org and OrthoMCL.org resources were invaluable to this work. We thank A. R. Gomes for helpful discussions, S. Ralph for providing us the merozoite purification protocol and D. Noordemeer and H. Painter for critical reading of the manuscript. We acknowledge the sequencing and bioinformatics expertise of the I2BC High-throughput sequencing facility and the expertise of the Imagerie-Gif Light Microscopy core facility.

Authors contribution

Conceptualization: JMS

Parasite culturing: LP, JMS, LM

Single cell experiment: YJ

Data analysis: DN, CH, JMS

Writing: JMS, CH

Competing interests

The authors declare no competing interests

References

1. Global Malaria Programme (GMP). World malaria report 2024 [Internet]. World Health Organization; 2024. Available from: https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2024

2. Rudlaff RM, Kraemer S, Marshman J, Dvorin JD. Three-dimensional ultrastructure of Plasmodium falciparum throughout cytokinesis. PLOS Pathog. 2020 Jun 8;16(6):e1008587.

3. Zanghí G, Patel H, Smith JL, Camargo N, Bae Y, Hesping E, et al. Genome-wide gene expression profiles throughout human malaria parasite liver stage development in humanized mice. Nat Microbiol. 2025 Feb;10(2):569–84.

4. Kafsack BFC, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, et al. A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature. 2014 Mar 13;507(7491):248–52.

506 5. Josling GA, Russell TJ, Venezia J, Orchard L, van Biljon R, Painter HJ, et al. Dissecting
507 the role of PfAP2-G in malaria gametocytogenesis. Nat Commun. 2020 Mar 20;11:1503.

508 6. Yahata K, Treeck M, Culleton R, Gilberger TW, Kaneko O. Time-Lapse Imaging of Red
509 Blood Cell Invasion by the Rodent Malaria Parasite Plasmodium yoelii. PLOS ONE.
510 2012 Dec 5;7(12):e50780.

511 7. Howick VM, Russell AJC, Andrews T, Heaton H, Reid AJ, Natarajan K, et al. The
512 Malaria Cell Atlas: Single parasite transcriptomes across the complete Plasmodium life
513 cycle. Science. 2019 Aug 23;365(6455):eaaw2619.

514 8. Reers AB, Bautista R, McLellan J, Morales B, Garza R, Bol S, et al. Histone
515 modification analysis reveals common regulators of gene expression in liver and blood
516 stage merozoites of Plasmodium parasites. Epigenetics Chromatin. 2023 Jun 15;16(1):25.

517 9. Geoghegan ND, Evelyn C, Whitehead LW, Pasternak M, McDonald P, Triglia T, et al.
518 4D analysis of malaria parasite invasion offers insights into erythrocyte membrane
519 remodeling and parasitophorous vacuole formation. Nat Commun. 2021 Jun 15;12:3620.

520 10. Boyle MJ, Wilson DW, Richards JS, Riglar DT, Tetteh KKA, Conway DJ, et al. Isolation
521 of viable Plasmodium falciparum merozoites to define erythrocyte invasion events and
522 advance vaccine and drug development. Proc Natl Acad Sci U S A. 2010 Aug
523 10;107(32):14378–83.

524 11. Machado M, Steinke S, Ganter M. Plasmodium Reproduction, Cell Size, and
525 Transcription: How to Cope With Increasing DNA Content? Front Cell Infect Microbiol.
526 2021;11:660679.

527 12. Mohammed M, Dziedziech A, Macedo D, Huppertz F, Veith Y, Postel Z, et al. Single-
528 cell transcriptomics reveal transcriptional programs underlying male and female cell fate
529 during Plasmodium falciparum gametocytogenesis. Nat Commun. 2024 Aug
530 26;15(1):7177.

531 13. Cowman AF, Berry D, Baum J. The cellular and molecular basis for malaria parasite
532 invasion of the human red blood cell. J Cell Biol. 2012 Sep 17;198(6):961–71.

533 14. Kahn S. Merozoite Proteome (3D7) [Internet]. PlasmoDB; 2008. Available from:
534 https://plasmodb.org/plasmo/app/record/dataset/DS_388d40eae3

535 15. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary
536 learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol.
537 2024 Feb;42(2):293–304.

538 16. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated
539 analysis of multimodal single-cell data. Cell. 2021 Jun 24;184(13):3573-3587.e29.

540 17. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell
541 gene expression data. Nat Biotechnol. 2015 May;33(5):495–502.

542 18. Kimmel J, Schmitt M, Sinner A, Jansen PWTC, Mainye S, Ramón-Zamorano G, et al.
543 Gene-by-gene screen of the unknown proteins encoded on Plasmodium falciparum
544 chromosome 3. Cell Syst. 2023 Jan 18;14(1):9-23.e7.

545 19. Painter HJ, Chung NC, Sebastian A, Albert I, Storey JD, Llinás M. Genome-wide real-
546 time in vivo transcriptional dynamics during Plasmodium falciparum blood-stage
547 development. Nat Commun. 2018 Jul 9;9(1):2656.

548 20. Santos JM, Josling G, Ross P, Joshi P, Orchard L, Campbell T, et al. Red Blood Cell
549 Invasion by the Malaria Parasite Is Coordinated by the PfAP2-I Transcription Factor.
550 Cell Host Microbe. 2017 Jun 14;21(6):731-741.e10.

551 21. Stubbs J, Simpson KM, Triglia T, Plouffe D, Tonkin CJ, Duraisingh MT, et al. Molecular
552 mechanism for switching of P. falciparum invasion pathways into human erythrocytes.
553 Science. 2005 Aug 26;309(5739):1384–7.

554 22. Coleman BI, Ribacke U, Manary M, Bei AK, Winzeler EA, Wirth DF, et al. Nuclear
555 repositioning precedes promoter accessibility and is linked to the switching frequency of
556 a Plasmodium falciparum invasion gene. Cell Host Microbe. 2012 Dec 13;12(6):739–50.

557 23. Nishi T, Kaneko I, Iwanaga S, Yuda M. Identification of a novel AP2 transcription factor
558 in zygotes with an essential role in Plasmodium ookinete development. PLoS Pathog.
559 2022 Aug 10;18(8):e1010510.

560 24. Subudhi AK, Green JL, Satyam R, Salunke RP, Lenz T, Shuaib M, et al. DNA-binding
561 protein PfAP2-P regulates parasite pathogenesis during malaria parasite blood stages. Nat
562 Microbiol. 2023 Nov;8(11):2154–69.

563 25. Dogga SK, Rop JC, Cudini J, Farr E, Dara A, Ouologuem D, et al. A single cell atlas of
564 sexual development in Plasmodium falciparum. Science. 2024 May
565 3;384(6695):eadj4088.

566 26. Mohammed M, Dziedziech A, Macedo D, Huppertz F, Veith Y, Postel Z, et al. Single-
567 cell transcriptomics reveal transcriptional programs underlying male and female cell fate
568 during Plasmodium falciparum gametocytogenesis. Nat Commun. 2024 Aug
569 26;15(1):7177.

570 27. Campbell TL, Silva EKD, Olszewski KL, Elemento O, Llinás M. Identification and
571 Genome-Wide Prediction of DNA Binding Specificities for the ApiAP2 Family of
572 Regulators from the Malaria Parasite. PLOS Pathog. 2010 Oct 28;6(10):e1001165.

573 28. Josling GA, Petter M, Oehring SC, Gupta AP, Dietz O, Wilson DW, et al. A Plasmodium
574 Falciparum Bromodomain Protein Regulates Invasion Gene Expression. Cell Host
575 Microbe. 2015 Jun 10;17(6):741–51.

576 29. Bonnell VA, Zhang Y, Brown AS, Horton J, Josling GA, Chiu TP, et al. DNA sequence
577 and chromatin differentiate sequence-specific transcription factor binding in the human
578 malaria parasite Plasmodium falciparum. Nucleic Acids Res. 2024 Jul 5;gkae585.

579 30. Russell TJ, De Silva EK, Crowley VM, Shaw-Saliba K, Dube N, Josling G, et al.
580 Inhibitors of ApiAP2 protein DNA binding exhibit multistage activity against
581 Plasmodium parasites. PLoS Pathog. 2022 Oct;18(10):e1010887.

582 31. Shang X, Wang C, Fan Y, Guo G, Wang F, Zhao Y, et al. Genome-wide landscape of
583 ApiAP2 transcription factors reveals a heterochromatin-associated regulatory network

584 during Plasmodium falciparum blood-stage development. Nucleic Acids Res. 2022 Apr
585 8;50(6):3413–31.

586 32. Young JA, Johnson JR, Benner C, Yan SF, Chen K, Le Roch KG, et al. In silico
587 discovery of transcription regulatory elements in Plasmodium falciparum. BMC
588 Genomics. 2008 Feb 7;9:70.

589 33. Shang X, Shen S, Tang J, He X, Zhao Y, Wang C, et al. A cascade of transcriptional
590 repression determines sexual commitment and development in Plasmodium falciparum.
591 Nucleic Acids Res. 2021 Sep 20;49(16):9264–79.

592 34. Cubillos EFG, Prata IO, Fotoran WL, Ranford-Cartwright L, Wunderlich G. The
593 Transcription Factor PfAP2-O Influences Virulence Gene Transcription and Sexual
594 Development in Plasmodium falciparum. Front Cell Infect Microbiol. 2021;11:669088.

595 35. Flueck C, Bartfai R, Niederwieser I, Witmer K, Alako BTF, Moes S, et al. A Major Role
596 for the Plasmodium falciparum ApiAP2 Protein PfSIP2 in Chromosome End Biology.
597 PLoS Pathog. 2010 Feb 26;6(2):e1000784.

598 36. Sierra-Miranda M, Vembar SS, Delgadillo DM, Ávila-López PA, Herrera-Solorio AM,
599 Lozano Amado D, et al. PfAP2Tel, harbouring a non-canonical DNA-binding AP2
600 domain, binds to Plasmodium falciparum telomeres. Cell Microbiol. 2017 Sep;19(9).

601 37. Tripathi J, Zhu L, Nayak S, Stoklasa M, Bozdech Z. Stochastic expression of invasion
602 genes in Plasmodium falciparum schizonts. Nat Commun. 2022 May 30;13(1):3004.

603 38. Toenhake CG, Fraschka SAK, Vijayabaskar MS, Westhead DR, van Heeringen SJ,
604 Bártfai R. Chromatin Accessibility-Based Characterization of the Gene Regulatory
605 Network Underlying Plasmodium falciparum Blood-Stage Development. Cell Host
606 Microbe. 2018 Apr 11;23(4):557-569.e9.

607 39. Ruiz JL, Tena JJ, Bancells C, Cortés A, Gómez-Skarmeta JL, Gómez-Díaz E.
608 Characterization of the accessible genome in the human malaria parasite Plasmodium
609 falciparum. Nucleic Acids Res. 2018 Oct 12;46(18):9414–31.

610 40. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its
611 consequences. Cell. 2008 Oct 17;135(2):216–26.

612 41. van Heyningen V. Stochasticity in genetics and gene regulation. Philos Trans R Soc Lond
613 B Biol Sci. 2024 Apr 22;379(1900):20230476.

614 42. Cortés A, Carret C, Kaneko O, Lim BYSY, Ivens A, Holder AA. Epigenetic Silencing of
615 Plasmodium falciparum Genes Linked to Erythrocyte Invasion. PLOS Pathog. 2007 Aug
616 3;3(8):e107.

617 43. Braun-Breton C, Rosenberry TL, da Silva LP. Induction of the proteolytic activity of a
618 membrane protein in Plasmodium falciparum by phosphatidyl inositol-specific
619 phospholipase C. Nature. 1988 Mar 31;332(6163):457–9.

620 44. Choudhary S, Satija R. Comparison and evaluation of statistical error models for scRNA-
621 seq. Genome Biol. 2022 Jan 18;23(1):27.

622 45. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics
623 and regulators of cell fate decisions are revealed by pseudotemporal ordering of single
624 cells. Nat Biotechnol. 2014 Apr;32(4):381–6.

625 46. Bailey TL. STREME: accurate and versatile sequence motif discovery. Bioinformatics.
626 2021 Sep 29;37(18):2834–40.

627 47. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between
628 motifs. Genome Biol. 2007;8(2):R24.

629 48. Hulsen T, de Vlieg J, Alkema W. BioVenn – a web application for the comparison and
630 visualization of biological lists using area-proportional Venn diagrams. BMC Genomics.
631 2008 Oct 16;9(1):488.

632 49. Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, et al. PlasmoDB:
633 a functional genomic database for malaria parasites. Nucleic Acids Res. 2009
634 Jan;37(Database issue):D539-543.

635 50. Li L, Christian J Stoeckert J, Roos DS. OrthoMCL: Identification of Ortholog Groups for
636 Eukaryotic Genomes. Genome Res. 2003 Sep;13(9):2178.

637 51. Santos JM, Braikia FZ, Oudinet C, Dauba A, Khamlichi AA. Two modes of cis-
638 activation of switch transcription by the IgH superenhancer. Proc Natl Acad Sci U S A.
639 2019 Jul 16;116(29):14708–13.

640

Figure 1

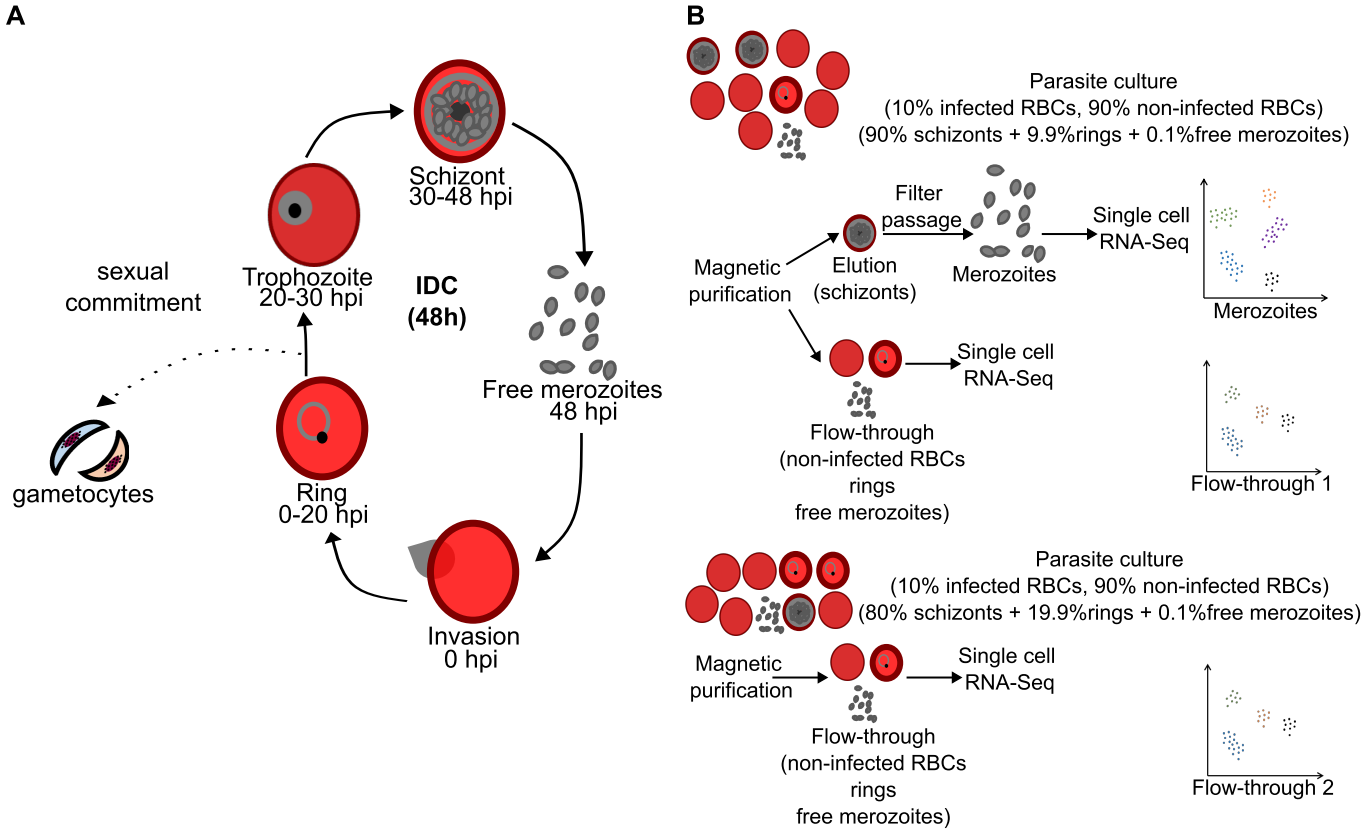
A
Schizont
30-48 hpi
Trophozoite
20-30 hpi
sexual
commitment
IDC
(48h)
Free merozoites
48 hpi
gametocytes
Ring
0-20 hpi
Invasion
0 hpi
B
Parasite culture
(10% infected RBCs, 90% non-infected RBCs)
(90% schizonts + 9.9%rings + 0.1%free merozoites)
Filter
passage
Single cell
RNA-Seq
Magnetic
purification
Elution
(schizonts)
Merozoites
Merozoites
Single cell
RNA-Seq
Flow-through
(non-infected RBCs
rings
free merozoites)
Flow-through 1
Parasite culture
(10% infected RBCs, 90% non-infected RBCs)
(80% schizonts + 19.9%rings + 0.1%free merozoites)
Magnetic
purification
Single cell
RNA-Seq
Flow-through
(non-infected RBCs
rings
free merozoites)
Flow-through 2

Figure 2

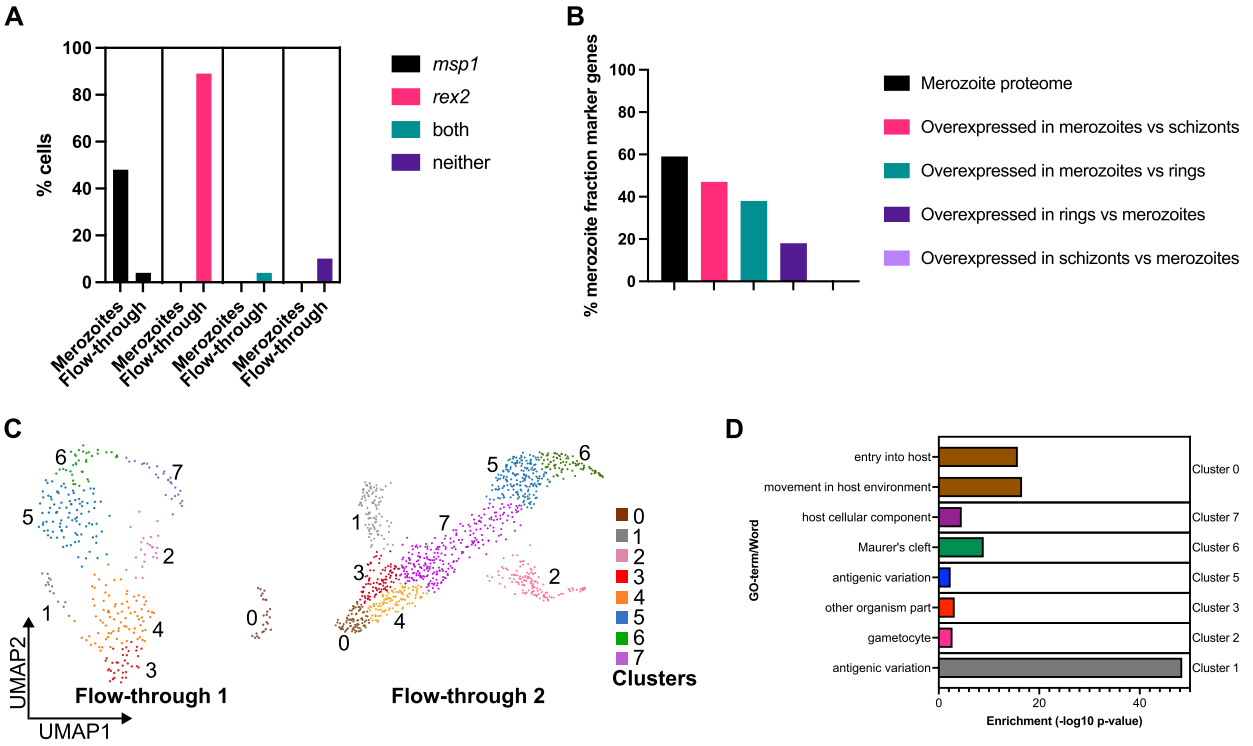
A
% cells
100
80
60
40
20
0
Merozoites
Flow-through
msp1
rex2
both
neither
B
% merozoite fraction marker genes
Merozoite proteome
Overexpressed in merozoites vs schizonts
Overexpressed in merozoites vs rings
Overexpressed in rings vs merozoites
Overexpressed in schizonts vs merozoites
C
UMAP2
UMAP1
Flow-through 1
Flow-through 2
Clusters
D
GO-term/Word
entry into host
movement in host environment
host cellular component
Maurer's cleft
antigenic variation
other organism part
gametocyte
antigenic variation
Cluster 0
Cluster 7
Cluster 6
Cluster 5
Cluster 3
Cluster 2
Cluster 1
Enrichment (-log10 p-value)

Figure 3

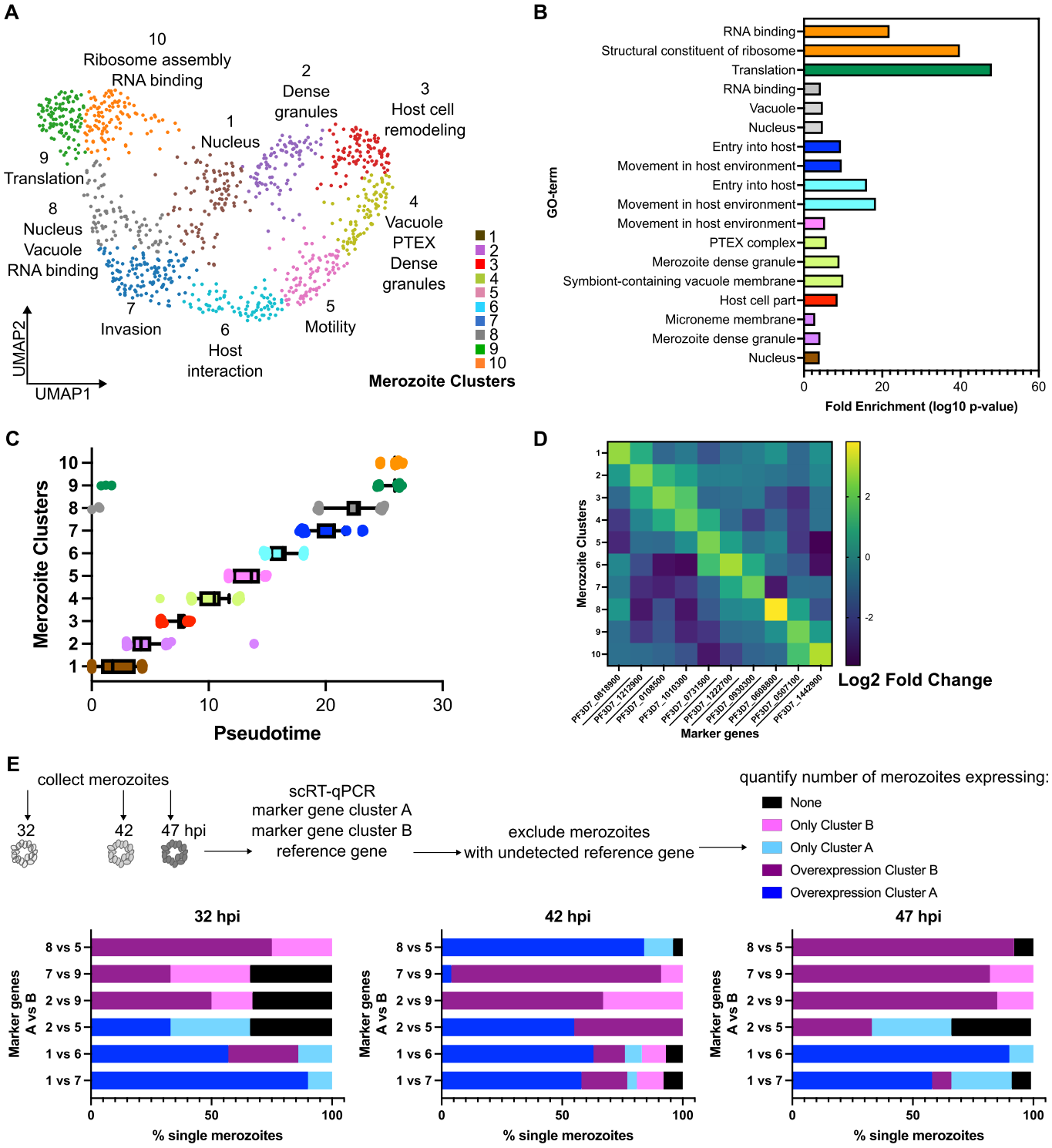
A
10
Ribosome assembly
RNA binding
2
Dense
granules
3
Host cell
remodeling
1
Nucleus
9
Translation
8
Nucleus
Vacuole
RNA binding
4
Vacuole
PTEX
Dense
granules
7
Invasion
6
Host
interaction
5
Motility
UMAP2
UMAP1
1
2
3
4
5
6
7
8
9
10
Merozoite Clusters
B
RNA binding
Structural constituent of ribosome
Translation
RNA binding
Vacuole
Nucleus
Entry into host
Movement in host environment
Entry into host
Movement in host environment
Movement in host environment
PTEX complex
Merozoite dense granule
Symbiont-containing vacuole membrane
Host cell part
Microneme membrane
Merozoite dense granule
Nucleus
GO-term
0
20
40
60
Fold Enrichment (log10 p-value)
C
Merozoite Clusters
Pseudotime
0
10
20
30
D
Merozoite Clusters
Marker genes
Log2 Fold Change
E
collect merozoites
32
42
47 hpi
scRT-qPCR
marker gene cluster A
marker gene cluster B
reference gene
exclude merozoites
with undetected reference gene
quantify number of merozoites expressing:
None
Only Cluster B
Only Cluster A
Overexpression Cluster B
Overexpression Cluster A
32 hpi
42 hpi
47 hpi
Marker genes
A vs B
8 vs 5
7 vs 9
2 vs 9
2 vs 5
1 vs 6
1 vs 7
% single merozoites

Figure 4

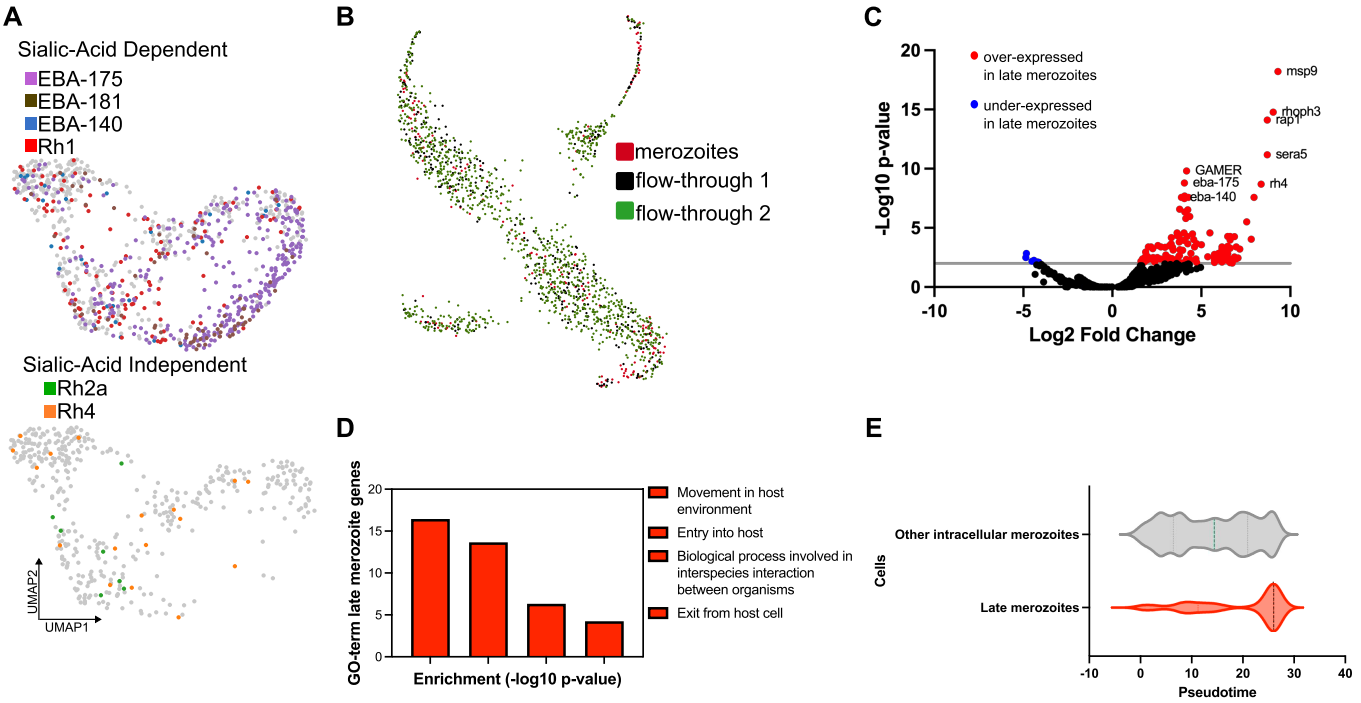
A
Sialic-Acid Dependent
EBA-175
EBA-181
EBA-140
Rh1
Sialic-Acid Independent
Rh2a
Rh4
UMAP2
UMAP1
B
merozoites
flow-through 1
flow-through 2
C
over-expressed in late merozoites
under-expressed in late merozoites
msp9
rhoph3
rap1
sera5
GAMER
eba-175
eba-140
rh4
-Log10 p-value
Log2 Fold Change
D
GO-term late merozoite genes
Enrichment (-log10 p-value)
Movement in host environment
Entry into host
Biological process involved in interspecies interaction between organisms
Exit from host cell
E
Other intracellular merozoites
Late merozoites
Cells
Pseudotime

Figure 5

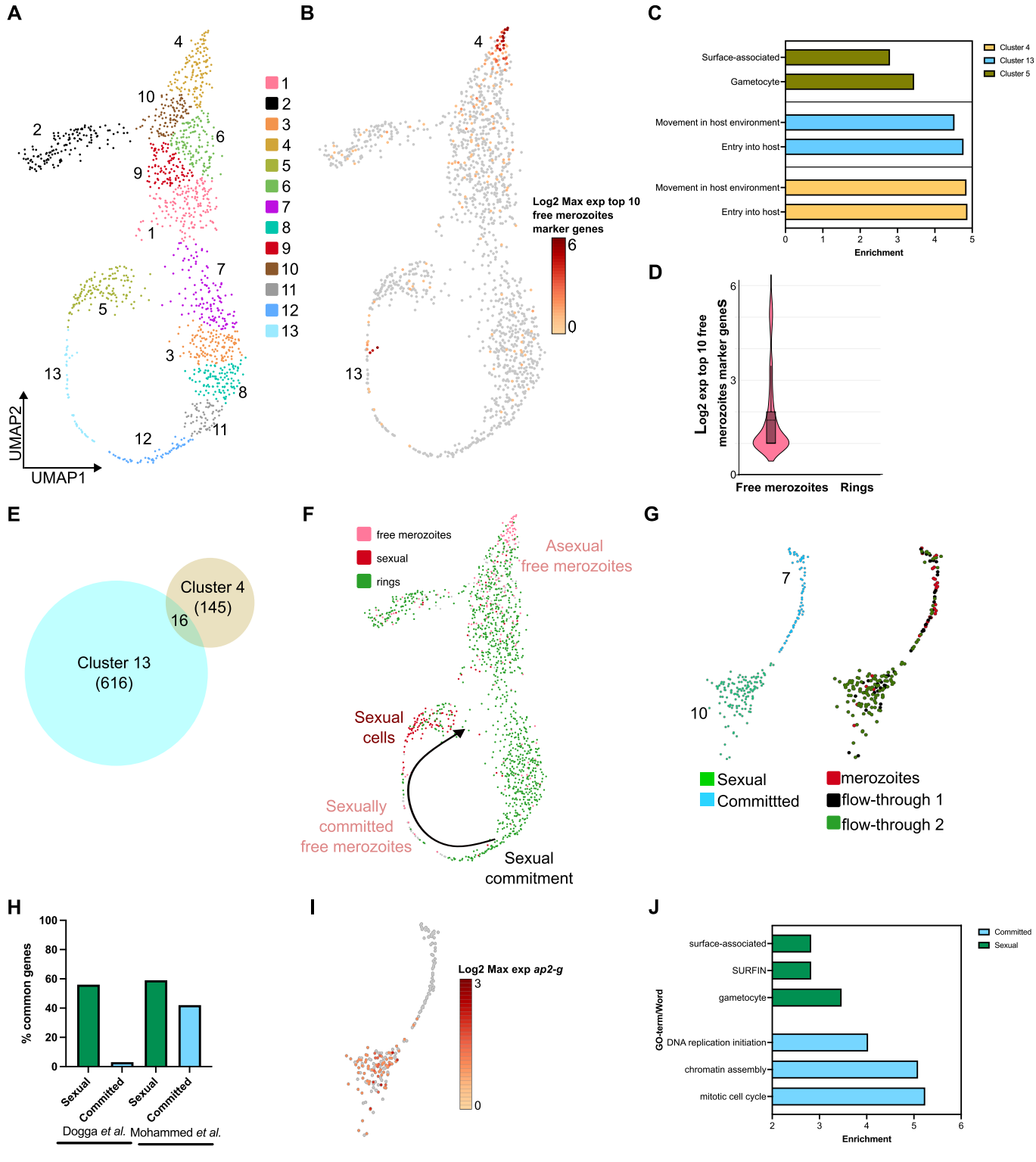
A
UMAP2
UMAP1
1
2
3
4
5
6
7
8
9
10
11
12
13
B
Log2 Max exp top 10 free merozoites marker genes
6
0
C
Cluster 4
Cluster 13
Cluster 5
Surface-associated
Gametocyte
Movement in host environment
Entry into host
Movement in host environment
Entry into host
Enrichment
D
Log2 exp top 10 free merozoites marker genes
Free merozoites
Rings
E
Cluster 4 (145)
16
Cluster 13 (616)
F
free merozoites
sexual
rings
Asexual free merozoites
Sexual cells
Sexually committed free merozoites
Sexual commitment
G
Sexual Committtted
merozoites
flow-through 1
flow-through 2
H
% common genes
Sexual
Committed
Sexual
Committed
Dogga et al.
Mohammed et al.
I
Log2 Max exp ap2-g
3
0
J
Committed
Sexual
surface-associated
SURFIN
gametocyte
DNA replication initiation
chromatin assembly
mitotic cell cycle
GO-term/Word
Enrichment

Figure 6

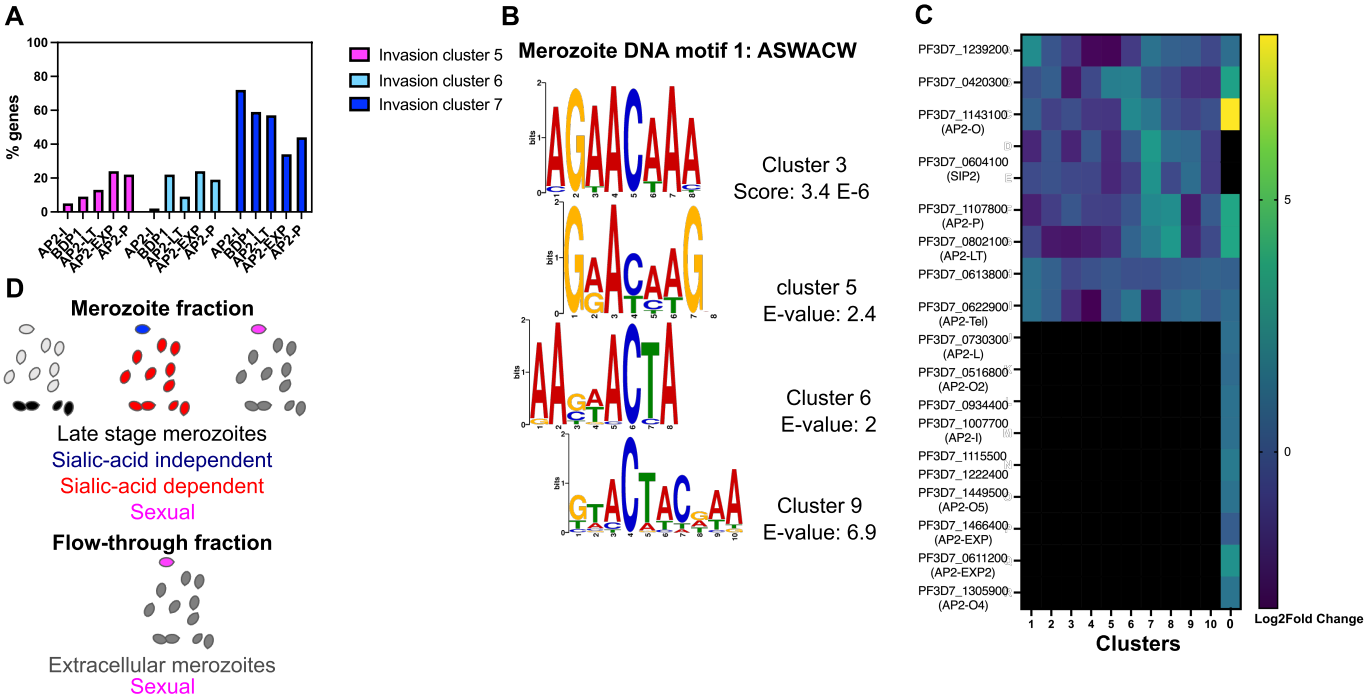
A
% genes
Invasion cluster 5
Invasion cluster 6
Invasion cluster 7
B
Merozoite DNA motif 1: ASWACW
Cluster 3
Score: 3.4 E-6
cluster 5
E-value: 2.4
Cluster 6
E-value: 2
Cluster 9
E-value: 6.9
C
PF3D7_1239200
PF3D7_0420300
PF3D7_1143100 (AP2-O)
PF3D7_0604100 (SIP2)
PF3D7_1107800 (AP2-P)
PF3D7_0802100 (AP2-LT)
PF3D7_0613800
PF3D7_0622900 (AP2-Tel)
PF3D7_0730300 (AP2-L)
PF3D7_0516800 (AP2-O2)
PF3D7_0934400
PF3D7_1007700 (AP2-I)
PF3D7_1115500
PF3D7_1222400
PF3D7_1449500 (AP2-O5)
PF3D7_1466400 (AP2-EXP)
PF3D7_0611200 (AP2-EXP2)
PF3D7_1305900 (AP2-O4)
Clusters
Log2Fold Change
D
Merozoite fraction
Late stage merozoites
Sialic-acid independent
Sialic-acid dependent
Sexual
Flow-through fraction
Extracellular merozoites
Sexual

Figure S1

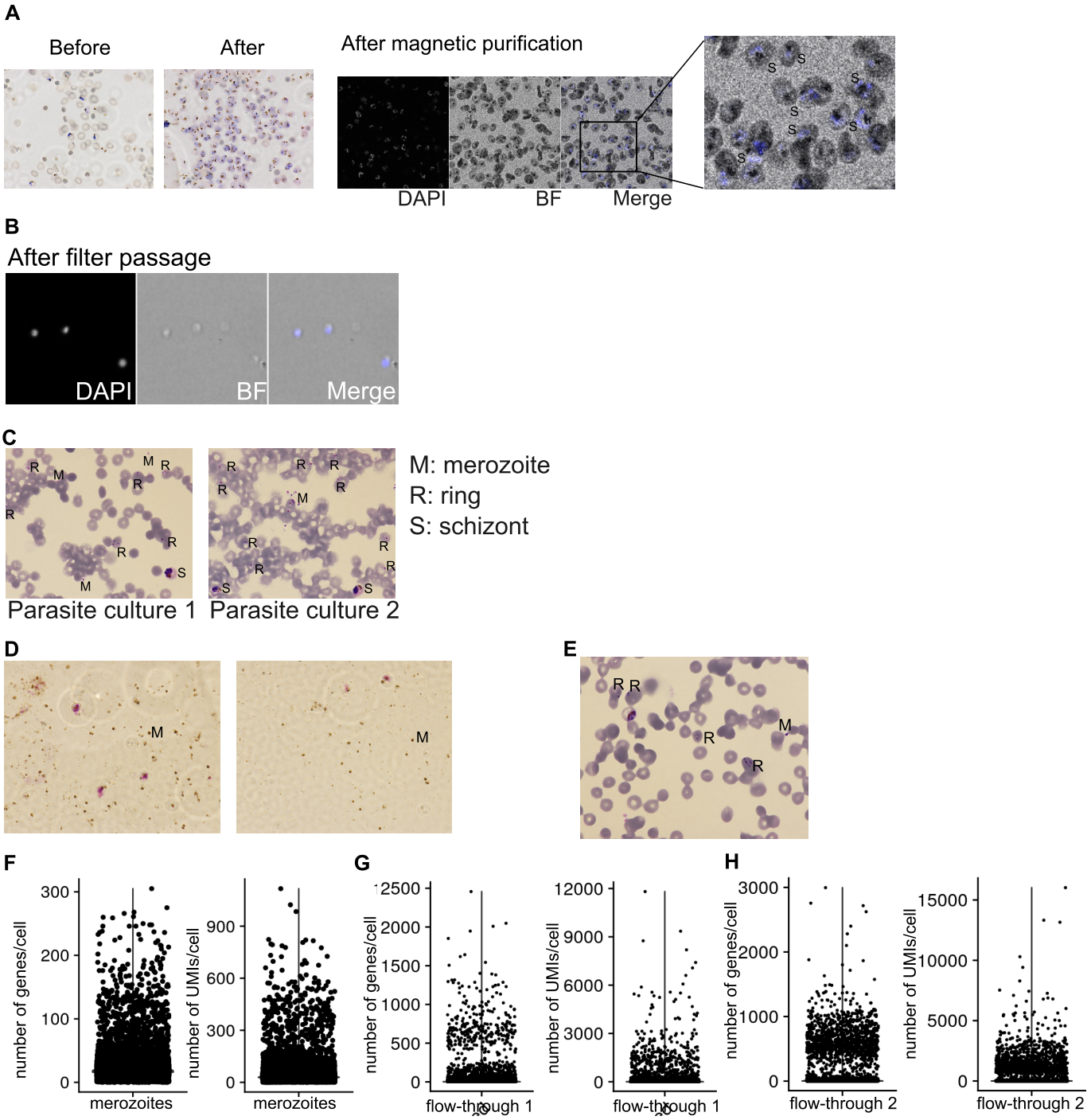
A
Before
After
After magnetic purification
DAPI
BF
Merge
B
After filter passage
DAPI
BF
Merge
C
M: merozoite
R: ring
S: schizont
Parasite culture 1
Parasite culture 2
D
E
F
number of genes/cell
number of UMIs/cell
merozoites
G
flow-through 1
H
flow-through 2

Figure S2

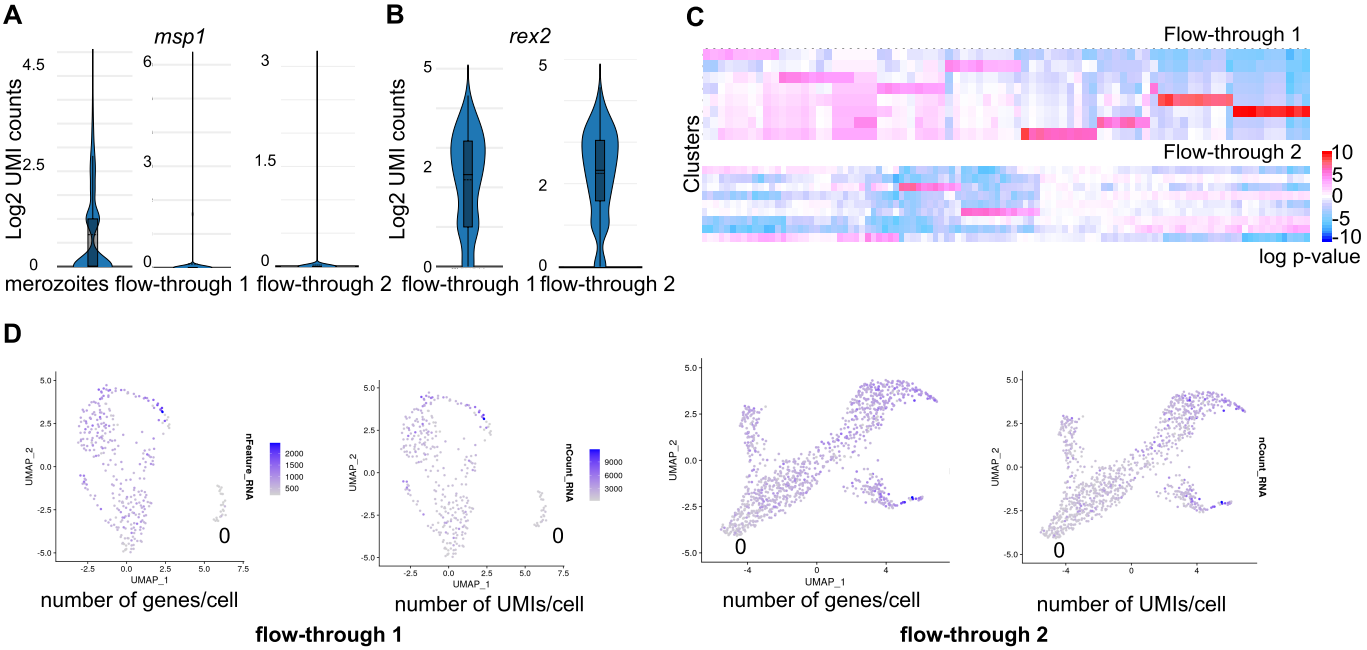
A
msp1
Log2 UMI counts
merozoites
flow-through 1
flow-through 2
B
rex2
Log2 UMI counts
flow-through 1
flow-through 2
C
Flow-through 1
Flow-through 2
Clusters
log p-value
D
number of genes/cell
number of UMIs/cell
flow-through 1
number of genes/cell
number of UMIs/cell
flow-through 2

Figure S3

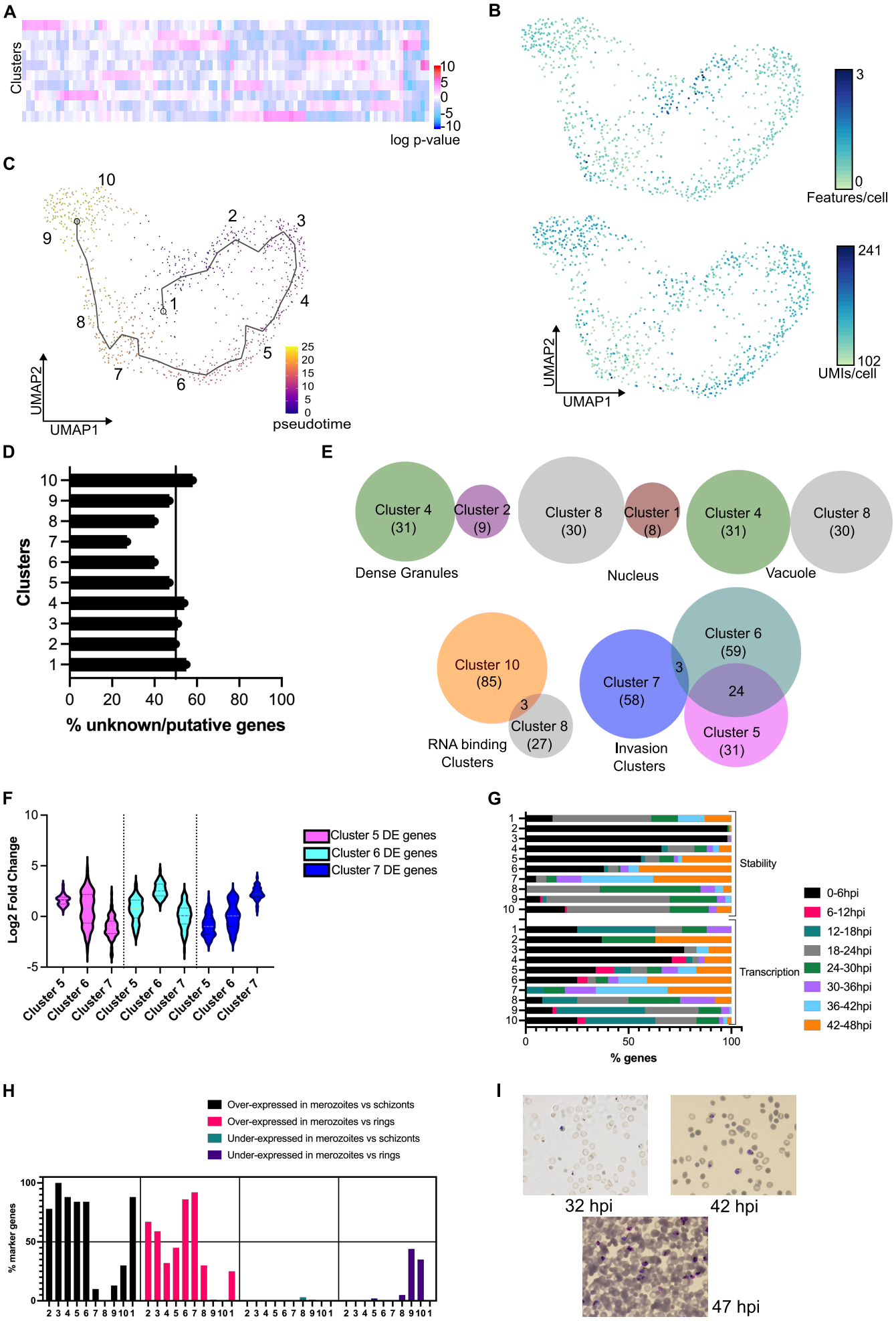
A
Clusters
10
5
0
-5
-10
log p-value
B
3
0
Features/cell
241
102
UMIs/cell
UMAP2
UMAP1
C
10
9
2
3
1
4
8
5
7
6
25
20
15
10
5
0
pseudotime
UMAP2
UMAP1
D
Clusters
10
9
8
7
6
5
4
3
2
1
0
20
40
60
80
100
% unknown/putative genes
E
Cluster 4 (31)
Cluster 2 (9)
Dense Granules
Cluster 8 (30)
Cluster 1 (8)
Nucleus
Cluster 4 (31)
Cluster 8 (30)
Vacuole
Cluster 10 (85)
3
Cluster 8 (27)
RNA binding Clusters
Cluster 6 (59)
3
Cluster 7 (58)
24
Cluster 5 (31)
Invasion Clusters
F
Log2 Fold Change
10
5
0
-5
Cluster 5 DE genes
Cluster 6 DE genes
Cluster 7 DE genes
Cluster 5
Cluster 6
Cluster 7
Cluster 5
Cluster 6
Cluster 7
Cluster 5
Cluster 6
Cluster 7
G
Stability
Transcription
0
50
100
% genes
0-6hpi
6-12hpi
12-18hpi
18-24hpi
24-30hpi
30-36hpi
36-42hpi
42-48hpi
H
Over-expressed in merozoites vs schizonts
Over-expressed in merozoites vs rings
Under-expressed in merozoites vs schizonts
Under-expressed in merozoites vs rings
% marker genes
100
50
0
I
32 hpi
42 hpi
47 hpi

Figure S4

A
100
80
60
40
20
0
% genes
Free merozoites
Late merozoites
Free merozoites
Late merozoites
Transcription peak time 1/48hpi
Stability peak time 1/48hpi
B
Cluster 0
marker genes
(52)
50
Overexpressed in
late merozoites
(83)

Figure S5

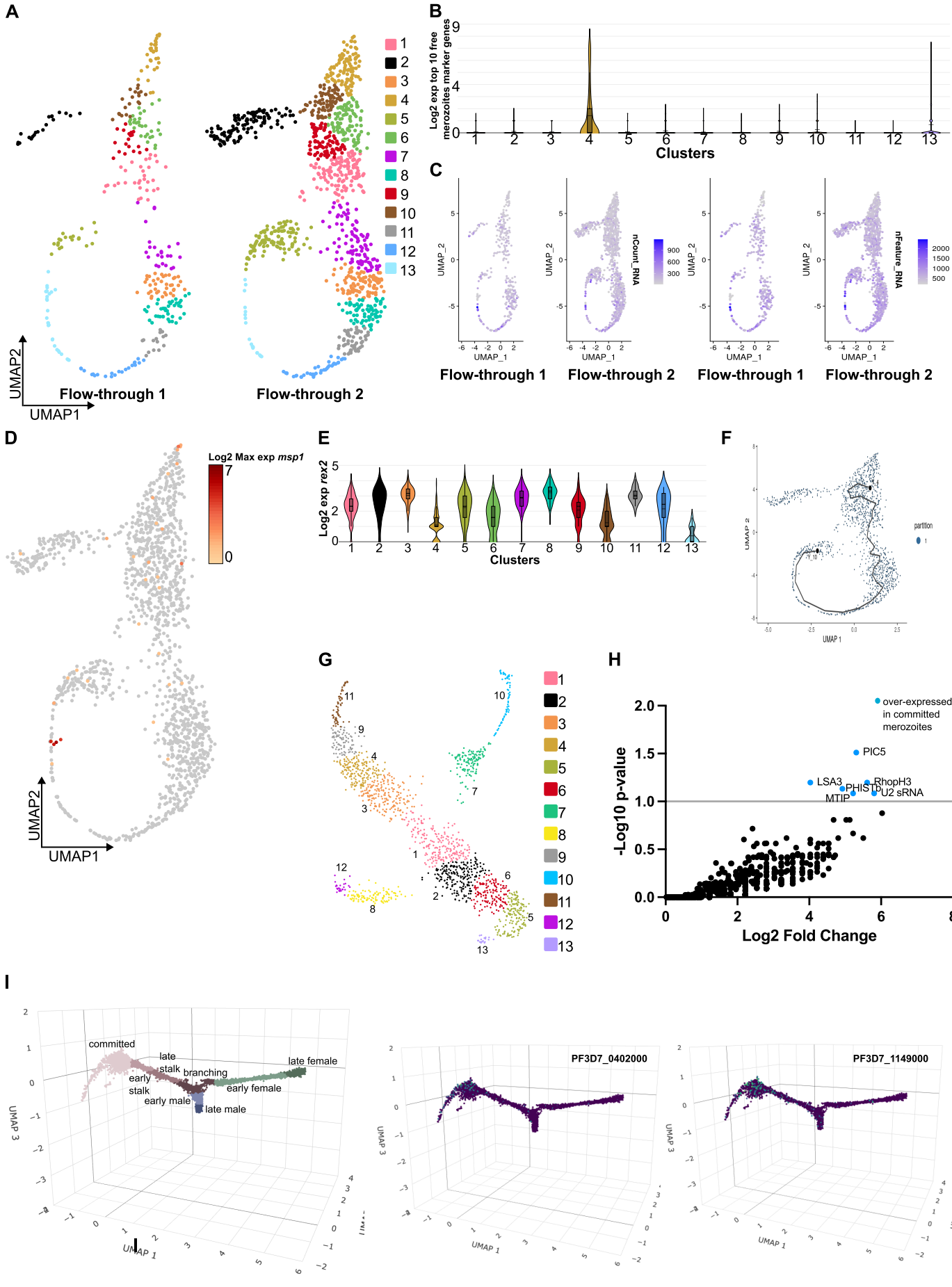
A
UMAP2
UMAP1
Flow-through 1
Flow-through 2
1
2
3
4
5
6
7
8
9
10
11
12
13
B
Log2 exp top 10 free merozoites marker genes
Clusters
C
UMAP_2
UMAP_1
nCount_RNA
nFeature_RNA
Flow-through 1
Flow-through 2
Flow-through 1
Flow-through 2
D
Log2 Max exp msp1
UMAP2
UMAP1
E
Log2 exp rex2
Clusters
F
partition
UMAP 1
UMAP 2
G
H
over-expressed in committed merozoites
PIC5
LSA3
RhopH3
PHIST
U2 sRNA
MTIP
-Log10 p-value
Log2 Fold Change
I
committed
late stalk
early stalk
branching
late female
early female
early male
late male
UMAP 1
UMAP 2
UMAP 3
PF3D7_0402000
PF3D7_1149000

Figure S6

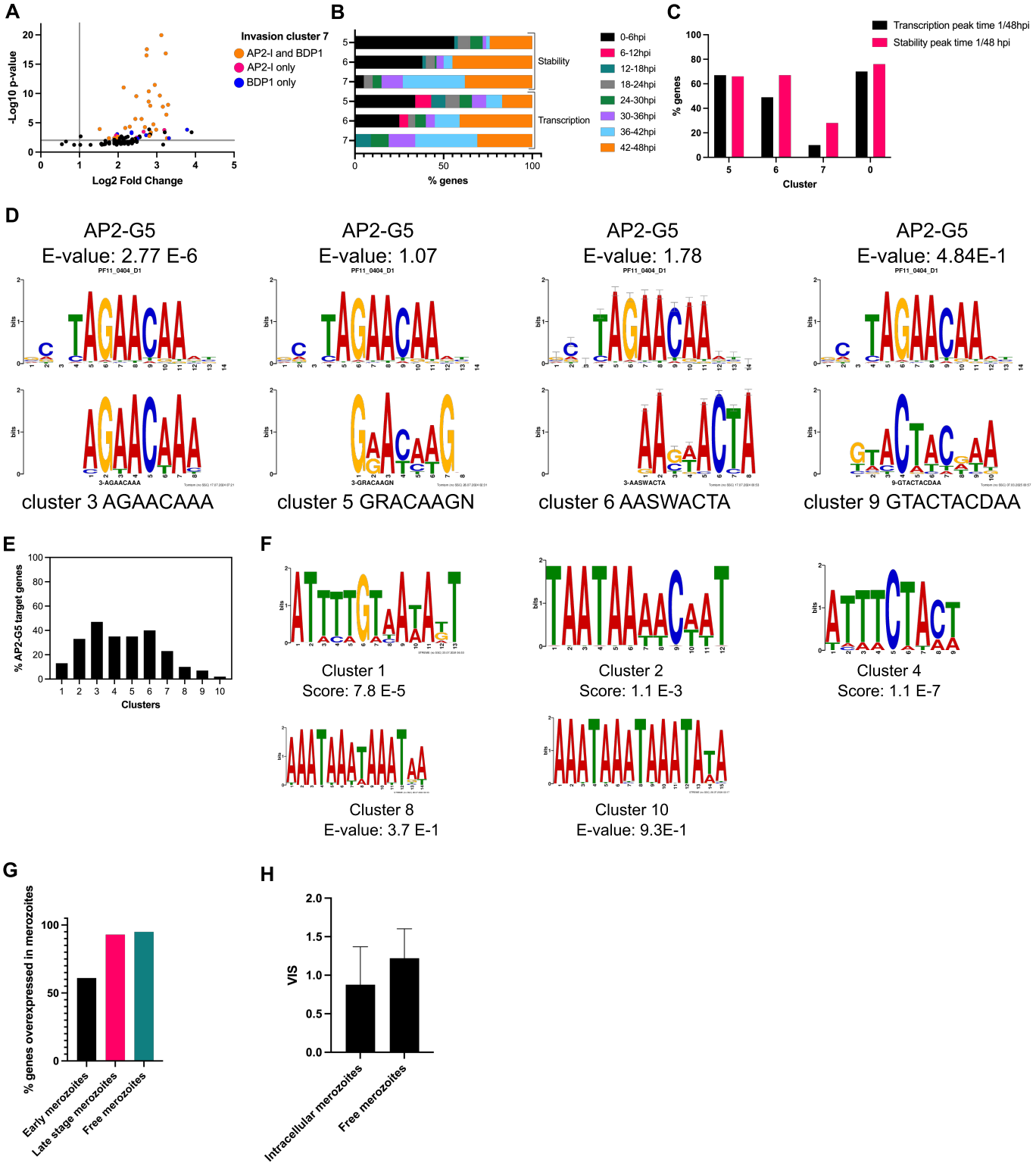
A
-Log10 p-value
Log2 Fold Change
Invasion cluster 7
AP2-I and BDP1
AP2-I only
BDP1 only
B
Stability
Transcription
% genes
0-6hpi
6-12hpi
12-18hpi
18-24hpi
24-30hpi
30-36hpi
36-42hpi
42-48hpi
C
Transcription peak time 1/48hpi
Stability peak time 1/48 hpi
% genes
Cluster
D
AP2-G5
E-value: 2.77 E-6
AP2-G5
E-value: 1.07
AP2-G5
E-value: 1.78
AP2-G5
E-value: 4.84E-1
cluster 3 AGAACAAA
cluster 5 GRACAAGN
cluster 6 AASWACTA
cluster 9 GTACTACDAA
E
% AP2-G5 target genes
Clusters
F
Cluster 1
Score: 7.8 E-5
Cluster 2
Score: 1.1 E-3
Cluster 4
Score: 1.1 E-7
Cluster 8
E-value: 3.7 E-1
Cluster 10
E-value: 9.3E-1
G
% genes overexpressed in merozoites
Early merozoites
Late stage merozoites
Free merozoites
H
VIS
Intracellular merozoites
Free merozoites

Fig 1 – Single cell RNA-sequencing of merozoites.

(**A**) During the intraerythrocytic developmental cycle (IDC), lasting 48 hours, parasite forms called merozoites egress from schizonts and invade the red blood cell (RBC) to become ring cells and later on trophozoites. Parasites either sexual commit, eventually differentiating into gametocytes, or continue the IDC, transforming into schizonts, within which new merozoites are formed (numbers below each stage indicate hour post-invasion, hpi). (**B**) Schizonts were magnetically purified from either a parasite culture containing 10% infected RBCs, corresponding to 90% schizont-stage parasites, 9.9% ring-stage parasites and 0.1% free merozoites (top) or corresponding to 80% schizont-stage parasites, 19.9% ring-stage parasites and 0.1% free merozoites (bottom). The merozoites inside the schizonts collected from the first parasite culture were released by filter passage and processed for single cell RNA-seq (merozoite fraction). In the flow-through of the magnetic purification of the two parasite cultures, we collected the non-infected RBCs, the ring-stage parasites and the free merozoites. The flow-through fractions from the two parasite cultures were processed independently for single cell RNA-seq (flow-through fractions 1 and 2).

Fig 2 – Identification of free merozoites.

(**A**) Percentage of merozoites and flow-through cells expressing only *msp1*, only *rap2* or the two genes, showing that the merozoite cells express the merozoite marker *msp1* but not the ring marker *rap2*. (**B**) Bar graph showing how many genes expressed by the cells of the merozoite fraction encode proteins identified in the Khan merozoite proteome (black) or are overexpressed or downexpressed in merozoites versus schizonts (pink and light purple, respectively) or are overexpressed or downexpressed in merozoites versus rings cells (green and dark purple, respectively), according to Reers *et al*.. (**C**) UMAP representation of the flow-through cells colored according to the SEURAT cluster. (**D**) GO-term analysis of the marker genes of each flow-through cluster. Because there is no GO-term associated with gametocytes, word search analysis was done instead for cluster 2 genes. The full list of GO-terms and words is in the S2 Table (cluster 0) and S3 Table (other clusters).

Fig 3 – Gene expression of the intracellular merozoites.

(**A**) UMAP representation showing the merozoite cells grouped by SEURAT cluster (the number of cells in each cluster is in S1 Table). Cells are colored according to the corresponding cluster. A heatmap showing expression of 100 randomly selected genes in the 10 clusters is in S3A Fig. The distribution of the number of genes/UMIs per cell in each cluster is in S3B Fig. (**B**) GO-term analysis of the genes differentially expressed in each merozoite cluster shows that each cluster expresses genes encoding proteins implicated in different processes. Bar graphs are colored according to the corresponding cluster. Only the most significant GO-terms are shown. For the complete GO-term analysis, check S2 Table. (**C**) Merozoite clusters pseudotime as determined by Monocle. Boxes are colored according to the corresponding cluster. The pseudotime of each merozoite cell is S1 Table. (**D**) Heatmap showing expression of selected marker genes of each merozoite cluster. Underlined are transcripts for which we designed qPCR primers. (**E**) Results of the single cell RT-qPCR on single merozoites collected at 32hpi, 42hpi and 47hpi, showing the percentage of merozoites expressing none of the marker genes, only one of the genes or overexpressing one of the marker genes versus the other. Cells overexpressing one or another marker gene are detected at different time points, suggesting that we collected merozoites of different ages in the merozoite fraction of the magnetic purification. The raw data of the RT-qPCR assays is in S4 Table.

Fig 4 – Identification of sialic-acid independent and late stage merozoites.

(**A**) UMAP representations of the merozoite cells colored according to expression of genes encoding for parasite ligands of the sialic-acid dependent invasion pathway (top) (EBA-175, purple; EBA-140, blue; EBA-181, brown; and Rh1, red) or of the sialic-acid independent invasion pathway (bottom) (Rh2a, green; and Rh4, orange) showing that few cells express sialic-acid independent ligands. (**B**) UMAP representation of the combined dataset of late stage merozoites and flow-through cells. Cells are colored according to the fraction they originated from (merozoite or flow-through). (**C**) Volcano plot showing the genes over-expressed (red) or under-expressed (blue) in late stage merozoite cells versus ring cells. The name of the encoding proteins for some of the overexpressed genes is shown (the full list of genes is in S5 Table). (**D**) GO-term analysis of the genes overexpressed in late stage merozoites. The full list of genes overexpressed in late stage merozoites and the GO-term analysis are in S5 Table. (**E**) Late merozoite cells (red) tend to have a later pseudotime than other intracellular merozoites (gray) (the pseudotime of each cell is in S1 Table). This was determined by identifying in the merozoite fraction, the barcodes assigned as late merozoites in the combined dataset of late stage merozoites and flow-through cells. The pseudotime was attributed by Monocle when it calculated a trajectory for the cells of the merozoite clusters.

Fig 5 – Identification of sexual merozoites.

(**A**) Seurat clusters of the integrated flow-through dataset (the full list of genes expressed in each cluster is in S6 Table and the number of cells on each cluster is in S1 Table). (**B**) There are two subsets of free merozoites, one in cluster 4 and one in cluster 13, as determined by over-expression of the top 10 free merozoite markers. (**C**) GO-term analysis of the marker genes of clusters 4 and 13 show that these cells expressed invasion-related genes, while cluster 5 cells express gametocyte-related genes (the full GO-term analysis are in S6 Table). Bars are colored according to the cluster. (**D**) There are only 16 common marker genes between the clusters 4 and 13 and all of these are markers of free merozoites (the identity of these genes is in S6 Table). (**E**) UMAP showing which cells express the top 10 free merozoite markers (free merozoites, in pink), *ap2-g* (sexual cells, in red) or *rex2* (rings, in green). One subset of free merozoites is asexual while the other one is sexually-committed. The sexual commitment pathway is shown. (**G**) Cluster 7 cells of the integrated merozoite and flow-throughs dataset are sexually-committed cells (blue) whereas cluster 10 cells are sexual cells (green), as determined by differentially gene expression analysis (S5 Table). Which cells are merozoites or flow-through is show in red or black/green, respectively. (**H**) Graph showing that more signature genes of sexual cells (green) were previously reported by Dogga *et al*. or Mohammed *et al*. as being over-expressed in committed cells or early gametocytes, as compared to the marker genes of committed cells (blue) (the list of genes is in S5 Table). (I) Sexually-committed cells do not express *ap2-g*. (**J**) GO-term analysis showing that committed cells express DNA replication, chromatin and cell-cycle genes (blue), while sexual cells (green) express gametocyte-related genes (the full GO-term analysis is in S5 Table).

Fig 6 – **Regulation of expression of the merozoite genes.**

(**A**) Graph showing the percentage of cluster 5 (pink), cluster 6 (light blue) or cluster 7 marker genes (dark blue) that are ChIP-Seq targets of AP2-I, BDP1, AP2-P, AP2-EXP or AP2-LT, as determined by Santos *et al*., Josling *et al*., Subudi *et al*., Russel *et al.*, Bonnell *et al*. and Shang *et al.*, respectively. The list of target genes for each cluster is in S2 Table. (**B**) STREME analysis identified an enriched DNA motif, here called merozoite DNA motif 1, in the promoter region of the marker genes of clusters 3, 5, 6 and 9. E-value and Scores indicate the significance of the found DNA motif. The DREME and TOMTOM analysis are in S6 Table. (**C**) Heatmap showing expression of *apiap2* genes in intracellular (cluster 1 to 10) and free merozoites (cluster 0). Black boxes indicate absence of expression. The expression data is in S2 Table. (**D**) Scheme of the type of merozoites identified in the merozoite and flow-through fractions.

Supplementary Information

S1 Fig – **Parasites purification**.

(**A**) Giemsa-stained blood smears of 32 hpi schizonts before (left) and after (right) magnetic purification. DAPI staining of the purified schizonts (S) is also shown (BF: brightfield). Higher magnification of a DAPI staining region is shown. (**B**) DAPI staining of 32 hpi merozoites after filter passage of the schizonts in A. (**C**) Giemsa-stained blood smear of the parasite cultures used for single cell RNA-sequencing showing the presence of rings (R), schizonts (S) and free merozoites (M). (**D**) Giemsa-stained blood smears of the purified merozoite fraction used for single cell RNA-sequencing showing the presence of merozoites (M). (**E**) Giemsa-stained blood smear of the flow-through fraction 1 used for single cell RNA-sequencing showing the presence of rings (R) and free merozoites (M). (**F**) Plots showing the number of genes and UMIs detected per merozoite cell. (**G**) Plots showing the number of genes and UMIs detected per flow-through 1 cell. (**H**) Plots showing the number of genes and UMIs detected per flow-through 2 cell. In (**F**), (**G**) and (**H**), each dot represents a cell.

S2 Fig – **Merozoite cells analysis**.

(**A**) Violin plot showing the expression level of *msp1* in merozoite and flow-through cells. (**B**) Violin plot showing the expression level of *rex2* in flow-through cells (the transcript was not detected in merozoites). (**C**) Heatmap showing differential expression of 100 randomly selected genes in the different clusters of the two flow-throughs. Each row represents a cluster. (**D**) Left: UMAP representation of the flow-through cells colored by number of genes detected per cell. Right: UMAP representation of the flow-through cells colored by number of UMIs (reads) detected per cell.

S3 Fig – **Merozoite clusters analysis**.

(**A**) Heatmap showing differential expression of 100 randomly selected genes in the different merozoite clusters. Each row represents a cluster. (**B**) Top: UMAP representation of merozoite cells colored by the number of genes detected per cell. Bottom: UMAP representation of merozoite cells colored by number of UMIs (reads) detected per cell. (**C**) Pseudotemporal analysis of the merozoite clusters showing that there is only one trajectory. The cells pseudotime is in S1 Table. (**D**) Percentage of genes annotated as putative or unknown in the different merozoite clusters. (**E**) Venn diagram showing no overlap between marker genes expressed in clusters 2 and 4 (GO-term dense granules), clusters 1 and 8 (GO-term nucleus) and 4 and 8 (GO-term vacuole) and very little between clusters 8 and 10 (GO-term RNA binding). There is little overlap between invasion clusters 6 and 7 or 5 and 7 but invasion clusters 5 and 6 are similar. Circles are colored according to the corresponding cluster and in brackets are the number of marker genes. (**F**) Differentially expressed (DE) genes of the invasion clusters 5, 6 and 7 are significantly overexpressed in that cluster compared to the others. Violin plots are colored according to the set of DE genes analyzed. (**G**) Timing of peak transcription or peak stability for the marker genes of each merozoite cluster, as determined by Painter *et al*. The timing of peak transcription and stability for each gene is in S2 Table. (**H**) Percentage of marker genes of each merozoite cluster over-expressed in merozoites versus schizonts (black) or rings (pink) or under-expressed in merozoites versus schizonts (green) or rings (purple), as determined by Reers *et al*.. (**I**) Giemsa-stain of the schizont cultures used for scRT-qPCR (before magnetic purification and filter passage).

S4 Fig – **Comparison of late stage merozoites with free merozoites**.

(**A**) Percentage of genes expressed by free merozoites or late stage merozoites that are peak transcribed (black) or peak stabilized (pink) at 1 or 48 hpi as determined by Painter *et al*.. (**B**) Many of the genes overexpressed in late merozoite cells are marker genes of free merozoite cells.

S5 Fig – **Analysis of sexual merozoites**.

(**A**) UMAP representation of the Seurat clusters of the combined flow-through dataset showing the number of cells in each flow-through sample (the number of cells in each cluster is in S1 Table, the marker genes of each cluster is in S5 Table). (**B**) Graph showing higher expression of the top 10 free merozoites marker genes in clusters 4 and 13. (**C**) UMAP representation showing that some of the cells in clusters 4 and 13 contain less RNA (nCount) and encode less transcripts (nFeature). (**D**) Some of the cluster 4 and cluster 13 cells express *msp1*, indicating that they are free merozoites. (**E**) rex2 is less expressed in the cluster 4 and cluster 13 cells. Violin plots are colored according to the cluster. (**F**) UMAP representation of the trajectory of the cells in the combined flow-through dataset. (**G**) UMAP representation of the Seurat clusters of the combined merozoite and flow-through dataset (the number of cells in each cluster is in S1 Table, the marker genes of each cluster is in S5 Table). (**H**) Volcano plot showing in blue the genes overexpressed by the sexually-committed merozoite cells of cluster 10 (the full list of genes is in S5 Table).

S6 Fig – Expression timing and stochasticity analysis of the merozoite genes.

(**A**) The majority of marker genes expressed in merozoite invasion cluster 7 are ChIP-Seq targets of AP2-I (pink), BDP1 (blue) or both (orange), as determined by Santos *et al*. and Josling *et al*., respectively. (**B**) Timing of peak transcription or peak stability for the marker genes of each of the merozoite invasion clusters, as determined by Painter *et al*. The timing of peak transcription and stability for each gene is in S2 Table. (**C**) Percentage of marker genes of the invasion clusters 5, 6 or 7 or of free merozoites with peak transcription (black) or stability (pink) at 1 or 48 hpi, as determined by Painter *et al*., showing that only the marker genes of clusters 5 and 6 reach peak expression at this time points. (**D**) Merozoite DNA motif 1, found by STREME, resembles the AP2-G5 DNA-binding motif as determined by TOMTOM. The STREME and TOMTOM results are in S7 Table. (**E**) Genes with merozoite DNA motif 1 are not AP2-G5 ChIP-Seq targets, as determined by Shang *et al*.. The list of AP2-G5 target genes in each cluster is in S2 Table. (**F**) In the gene promoters of clusters 1, 2, 4, 8 and 10, we only identified AT-rich DNA motifs. (**G**) Graph showing the percentage of marker genes of early merozoites (black), late stage merozoites (pink) and free merozoites (green) previously found to be overexpressed in merozoites by Reers *et al*.. (**H**) Graph showing the average variability index score (VIS score), as calculated by Tripathi *et al*., for the marker genes expressed by the merozoite clusters or by free merozoites. The VIS score for each gene is on S2 Table.

S1 Table – Number of cells and reads obtained after single cell RNA-sequencing, number of cells in each Seurat cluster for the merozoites and flow-through samples, pseudotime of the merozoite cells.

S2 Table – Gene expression and GO-term analysis of the merozoite cells, including all clusters of the merozoite fraction (1-10) and the cluster 0 of the flow-through fractions (free merozoites).

S3 Table – Gene expression and GO-term analysis of the flow-through clusters (except cluster 0).

S4 Table – List of primers used for single cell RT-qPCR and raw data from the single cell RT-qPCR assays.

S5 Table – Gene expression and GO-term analysis of the genes over-expressed and down-expressed in late merozoites versus ring cells as well as of each SEURAT cluster. It includes the list of marker genes of committed and sexual cells.

S6 Table – Gene expression and GO-term analysis of the genes over-expressed and down-expressed in each cluster of the combined flow-through clusters. It includes the list of marker genes of committed and sexual cells.

S7 Table – STREME and TOMTOM results.