Single cell RNA-sequencing identifies the gene expression profile of the asexual and sexual invasive merozoites of the human malaria parasite

Running title: Single cell expression of the malaria free 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. Here we uncovered the gene expression profile of the free merozoites, by doing single cell RNA-sequencing (scRNA-seq). We found that free merozoites express invasion and motility genes that are actively transcribed in schizonts and after merozoite egress. *P. falciparum* parasites undergo at least one IDC before sexually-committing and transforming into the sexual stages, the gametocytes, which are then transmitted to the mosquito vector. Our scRNA-seq dataset identified two early states of sexual commitment, preceding expression of the gametocytogenesis master regulator *ap2-g*. We hypothesize that early committed parasites are primed for sexual differentiation and definitive engagement into gametocytogenesis occurs when *ap2-g* is expressed. Our results contribute to a better understanding of how the parasite prepares for invasion and how it exits asexual growth and enters sexual differentiation. These transitions ensure survival of the parasite in the human host and transmission to the mosquito vector so the gene markers identified here can be therapeutically explored in the future to block malaria invasion or transmission.

39 **Importance**

40 The malaria parasite *Plasmodium falciparum*, which causes the death of over half a
41 million people each year, has a complex life cycle but symptoms only arise when the parasite
42 invades and develops inside the host red blood cells. The merozoites are the parasite stages
43 responsible for invading the human red blood cells and they are one of the few extracellular
44 stages of the lifecycle, even if briefly. However, their gene expression profile had yet to be
45 determined. Here we used single cell RNA-sequencing to establish the gene expression
46 program of free merozoites. We also identified for the first time the gene signature of sexually-
47 committed and sexual merozoites that will later give raise to gametocytes and ensure parasite
48 transmission.

49

50

Introduction

Malaria, responsible for over half a million deaths each year (1), is an infectious disease caused by the unicellular eukaryotic parasite *Plasmodium*. Out of the five species of parasite that cause disease in humans, *P. falciparum*, is the most virulent. The parasite life cycle is complex, alternating between the human host and the mosquito, that acts as both a host and a vector. The intraerythrocytic developmental cycle (IDC), in which the parasite multiplies within the human host red blood cell (RBC) for 48-hours, starts when parasite stages called merozoites, invade the host RBC (Figure 1A). Once inside the RBC, the parasite transforms into rings, then trophozoites and finally schizonts. At the end of schizogony, new merozoites are released into the bloodstream and the IDC restarts. In *P. falciparum*, the IDC ensures parasite multiplication but also transmission because parasites must undergo at least one IDC in order to sexually commit and form gametocytes that are later transmitted to the mosquito vector (2) (Figure 1A). Sexual commitment only occurs in a small subset of cells (<10%).

The merozoite is the less studied stage of the IDC, despite its attractiveness as a drug target given that: (i) it is extracellular, even if briefly; (ii) it is responsible for RBC invasion, which is indispensable for parasite survival; and, (iii) it initiates the IDC, which is the symptomatic life cycle stage of malaria. Studies describing the gene expression profiling of the different stages of the *P. falciparum* life cycle, lack information for the merozoite stage (3) and, it is only recently that it was shown that merozoites are transcriptionally active (4). RBC invasion is a fast process, lasting ~1 minute (5), and merozoites only last 5-15 minutes extracellular (6), so extracellular merozoites have never been isolated and characterized.

Here we used single cell RNA-sequencing to establish for the first time the gene expression signature of individual free merozoites, upon schizont rupture. We find that free merozoites express invasion and motility genes that are actively transcribed in schizonts and merozoites. We also identify two early states of sexual commitment that prepare the parasite to enter gametocytogenesis and be transmitted.

78 Our comprehensive study further sheds light into gene expression in the malaria
79 parasite and unfolds for the first time the single cell expression atlas of the human malaria
80 invasive merozoites, allowing the parasite to invade the RBC.
81
82 **Results**
83
84 **Identification of free merozoites**
85 *Plasmodium falciparum* merozoites have so far been isolated for transcriptional studies
86 by forced rupture of purified schizonts, after parasite treatment with an egress inhibitor that
87 inhibits merozoite egress (4). The recovered merozoites are viable but short-lived (6). In order
88 to record the gene expression profile of *de facto* free merozoites, we decided to not block
89 egress and this way allow natural merozoite release from schizonts. To separate the free
90 merozoites from the schizonts, we used magnetic columns that retain the trophozoites and
91 schizonts (6), which are paramagnetic due to the accumulation of hemozoin, the end product
92 of hemoglobin digestion (Figure 1B). The free merozoites, as well as the rings and non-
93 infected red blood cells, are not retained by the magnetic column and are then recovered in
94 the flow-through of the magnetic purification (Figure 1B). Merozoites cannot be separated from
95 the other cells of the flow-through so, to identify the gene expression profile of the free
96 merozoites, we used single cell RNA-sequencing (scRNA-seq), which is powerful in identifying
97 discrete cell types from mixed cultures.
98 To enrich for the presence of free merozoites in the parasite culture, we single
99 synchronized parasites for two cycles, and when schizonts were visible in Giemsa-stained
100 smears, we collected the culture. We used two different parasite cultures. One containing
101 mainly schizonts (parasite culture 1) and one containing a mix of schizonts and rings (parasite
102 culture 2) (Figure 1B). This strategy was fruitful because free merozoites were identified in
103 both cultures in Giemsa-stained blood smears (Figure 1B). We passed each parasite culture,
104 independently, through magnetic columns, and the purity of the elution and flow-through

105 fractions was then verified by blood smear (Figures S1A-B). The two flow-through fractions
106 were then prepared individually for scRNA-seq using the 10X Genomics technology.

107 After sequencing, we obtained reads from 1464 flow-through cells, 405 from the first
108 parasite culture, and 1059 from the second parasite culture (Table S1). We acquired 200
109 million reads, with a high percentage of sequencing saturation (over 98%) (Table S1). On
110 average, each cell expressed on mean 600 genes (Table S1, Figures S1C-D). After filtering
111 out cells with fewer than 200 genes and over 2500 genes per cell, a common threshold to
112 analyze *P. falciparum* single cell data (3,7,8), we retained 1421 cells in the combined flow-
113 through fractions so 97% of the initial sequenced cells (Table S1).

114 In agreement with what we had observed in blood smears (Figure 1B), the majority of
115 flow-through cells are ring-stage parasites because they express the ring marker *ring-exported*
116 *protein 2* (*rex2*, PF3D7_0936000) (Figure S1C). 2-4% of cells, however, express the
117 merozoite marker *merozoite surface protein 1* (*msp1*, PF3D7_0930300) (Figures S1E-G) and
118 are putative merozoites. We used *msp1* as a merozoite marker because MSP1 is the most
119 abundant merozoite surface protein (9) and it is over-expressed in merozoite cells versus ring
120 cells, as determined by Reers *et al*. (4), that defined a list of genes that are over-expressed in
121 merozoites versus rings or schizonts.

122 Merozoites and ring-staged cells are expected to express different genes so we used
123 Seurat, a toolbox for quality check, analysis and visualization of single cell data (10–12), to
124 non-biasedly identify a cell cluster, in the flow-through fraction, corresponding to free
125 merozoites (Supplementary files 1 and 2). Seurat identified eight expression clusters (Figures
126 1C and S1H-I, Table S2). Differential expression analysis was done to determine which genes
127 are over-expressed in one particular cluster versus the others. These genes are herein called
128 marker genes. Gene ontology (GO) analysis of the marker genes of each flow-through cluster
129 shows that cluster 0 cells express genes specialized in host cell entry and motility (Figure 1D,
130 Table S2), functions associated with invasion, and so with merozoites. Furthermore, 95% of
131 these genes were previously shown to be over-expressed in merozoites (4) and 44% of them
132 encode proteins previously identified in merozoite proteomes (13,14) (Figure S1I, Table S2).

133 Additionally, as expected for merozoites, the cluster 0 cells contain less RNA than other flow-
134 through cluster cells (Figure S1J). This indicates that the cluster 0 cells are free merozoites.
135 We identified 113 free merozoite cells in the two flow-throughs (Table S1), corresponding to
136 less than 10% of the population, as anticipated from an elusive stage that had never before
137 been captured by RNA-sequencing studies.
138

139 **Free merozoites express genes encoding proteins implicated in invasion and**
140 **motility**

141 We identified 102 genes over-expressed in free merozoites versus the other cells of
142 the flow-through (Figures 2A-B, Table S2). The most differentially expressed genes are those
143 encoding the erythrocyte binding antigen 175 (EBA-175), the gamete release protein
144 (GAMER), the reticulocyte binding protein homologue 1 (Rh1), the calcium dependent protein
145 kinase 5 (CDPK5) and the myosin A-tail interacting protein (MTIP) proteins (Figure 2A).
146 CDPK5, Rh1 and EBA-175 have all been shown to play a role in discharge of the parasite
147 apical organelles, micronemes and rhoptries. These Apicomplexa-specific organelles store
148 proteins that are released into the parasite surface prior to invasion (15). CDPK5 and Rh1 are
149 critical for micronemes discharge (16,17). The micronemes store, amongst others, EBA-175.
150 EBA-175 is a parasite ligand that establishes a specific interaction with the host cell receptor
151 glycophorin-A (18). Establishment of this interaction triggers rhoptry discharge (19). Release
152 of the rhoptry proteins requires proper positioning of the organelle, which is under the
153 responsibility of the armadillo-domain containing rhoptry protein (ARO) (20), and *aro* is a
154 marker gene of free merozoites (Figures 2A-B, Table S2).

155 After strong attachment of the parasite to the RBC surface, via the establishment of
156 the ligand-receptor interactions, the parasite reorientates (21). This requires photosensitized
157 INA-labeled protein (PHIL1) (22) and its partner protein PIC5 (23), and both coding genes are
158 over-expressed in free merozoites (Figures 2A-B, Table S2). Afterwards, the parasite ligand
159 Rh5 interacts with the host basigin, and there is formation of the tight junction. Free merozoites
160 over-express the genes coding for Rh5 and its interacting partners RIPR and Cysteine-Rich

161 Protective Antigen (CyRPA) (24), as well as apical membrane antigen 1 (AMA1) (Figures 2A-
162 B, Table S2), member of the tight junction (21). Rh5 only binds basagin, when proteolytically
163 processed by Plasmepsin X (PMX) (25), and *pmx* is also a signature gene of free merozoites
164 (Figures 2A-B, Table S2).

165 Malaria merozoites glide prior to invasion (26). This is supported by the parasite
166 actomyosin machinery, the glideosome (Figure S2B), which also propels the parasite inside
167 the host (27). Free merozoites over-express several of the genes encoding glideosome
168 constituents (27): Myosin A (MyoA), MyoA-tail Interacting Protein (MTIP), Myosin Essential
169 Light Chain (ELC) and Glideosome-Associated Connector (GAC) (Figures 2A-B and S2B,
170 Table S2). We also detect expression of the genes coding MTRAP and TRAMP, which might
171 interact with GAC (26), and several genes coding for inner membrane complex 1 (IMC1)
172 proteins, responsible for anchoring and stabilizing the glideosome (27) (Figures 2A-B and
173 S2B, Table S2). Gliding requires actin polymerization by profilin (27) and we detect over-
174 expression of the *profilin*, *actin I* and *actin depolymerizing factor 1* genes in extracellular
175 merozoites (Figures 2A-B and S2B, Table S2).

176 The gene signature of free merozoites reflects then their invasive and motile nature
177 (Figures 2B and S2A, Table S2). Intriguingly, these cells also over-express three non-coding
178 RNAs (Table S2). Among them, *pf3d7_0108900*, an antisense RNA covering the 5'
179 untranslated region of the *phil1* gene (28), which, as mentioned, is a marker gene of free
180 merozoites. There are also seventeen genes coding for proteins of unknown function that may
181 also be implicated in invasion or motility (Table S2).

182 Tripathi *et al.* showed variation in gene expression between merozoites in schizonts of
183 the same age and established a metric called variability index score (VIS) to reflect this
184 variation (8). We determined the VIS of each of the free merozoites signature genes, by
185 comparing our data to theirs. Most free merozoite genes have a VIS of 1 or more (Figure S2C,
186 Table S2). That is not the case of the marker genes of other clusters (Figure S2C), suggesting
187 that the free merozoite genes in particular are variably expressed among schizonts.

188
189 **Free merozoites genes are transcribed during schizogony and after merozoite**
190 **egress**

191 Merozoites have been shown to be transcriptionally active (4) so we wondered if the
192 free merozoites marker genes are actively transcribed in merozoites. To access this, we first
193 determined the time of peak transcription of each marker gene, as established by a published
194 real-time expression data (29). 54% of genes are peak transcribed at 1 or 48 hpi, when the
195 merozoites prepare to invade/egress and so when they are extracellular or are about to (Figure
196 3A, S3 Table). 37% of genes reach peak transcription beforehand, during schizogony,
197 between 30 and 47hpi, and the remaining 9% are transcribed following invasion, in rings (2-
198 20hpi) or trophozoites (20-30hpi) (Figure 3A, Table S2). In *P. falciparum*, active gene
199 promoters are labelled by the histone marks H3K9ac, H3K27ac and H3K4me3 (30), so we
200 examined if these marks are present in the promoters of the free merozoites signature genes.
201 According to previous Chromatin ImmunoPrecipitation (ChIP-Seq) data (4), the three histone
202 marks are present in the gene promoters, as well as in the transcription start sites, in
203 merozoites and schizonts (Figure 3B), but not in ring-staged parasites (Figure S3A). Binding
204 then drops within gene bodies (Figure 3B). These two comparative analyses indicate that the
205 free merozoite genes are transcribed during schizogony and when the merozoites are
206 extracellular. Bulk RNA-seq (4) indicated that 89% of the free merozoites marker genes are
207 more expressed in merozoites than schizonts so peak expression of the marker genes may
208 be reached after merozoite egress. In ring-staged parasites, H3K27ac is removed from all
209 genes, except from those to be expressed in rings (31), so these analyses also show that the
210 free merozoite genes are no longer transcribed after invasion. This is further supported by the
211 fact that only one of the free merozoites marker genes is, according to (4), over-expressed in
212 rings versus merozoite parasites (Figure S3B, Table S2).

213 H3K4me3 is a marker of open, accessible, chromatin (30), so we examined if the
214 promoters of the free merozoites marker genes are accessible when the genes are being
215 transcribed, i.e., in schizonts and merozoites. We did this by comparing our dataset to a

216 published time-course of chromatin accessibility in *P. falciparum* (32). In that study, 62% of
217 the free merozoites gene promoters are euchromatic between 30 and 40hpi, so during
218 schizogony (Figure 3C, Table S2). This analysis suggests that chromatin accessibility
219 precedes gene transcription in free merozoites. This points towards activation of gene
220 transcription by a transcription factor.

221 The ApiAP2 proteins are the main family of DNA-binding proteins in the Apicomplexa
222 and several have been shown to act as transcription factors (33). We examined expression of
223 *apiap2* genes in free merozoites to identify proteins potentially implicated in regulating
224 expression of the free merozoite genes. We detected expression of 16 *apiap2* genes, including
225 *ap2-lt* (Figure S3C). AP2-LT has been previously found to bind the promoters of several
226 invasion genes (34,35) and to interact with the histone acetylase GCN5, which binds to
227 H3K9ac-labeled promoters (36,37). Since we showed that the free merozoites express
228 invasion genes and these genes promoters are marked by H3K9ac, we determined how many
229 of the free merozoites promoters are bound by AP2-LT or GCN5, according to published ChIP-
230 Seq datasets (34–37). GCN5 binds 40% of the promoters and 60% of these are bound by
231 AP2-LT, suggesting that AP2-LT and GCN5 may regulate expression of the free merozoites
232 genes (Figure 3D, Table S2). On the contrary, AP2-I, which has also been shown to regulate
233 expression of some invasion genes (38) and it is expressed by free merozoites (Figure S3C),
234 only binds upstream of 7% of the free merozoites genes according to (38,39) (Figure 3D, Table
235 S2).

236

237 **Identification of an early sexual commitment state by unsupervised clustering**

238 To increase the power of our scRNA-seq dataset, we integrated the two flow-through
239 datasets and applied Seurat to it (Figures 4A and S4A, Supplementary file 3). This led to the
240 identification of 13 clusters (Table S3). Then we used the list of signature genes of free
241 merozoites, that we previously established (Table S2), to find free merozoite cells in the
242 integrated dataset. Cells expressing the free merozoite marker genes are found in clusters 4

243 and 13 (Figures 4B and S4B). These cells contain less RNA and encode less transcripts than
244 other cells (S4C Figure), express the merozoite marker *msp1* (Figure S4D) and encode
245 invasion-related genes (Figure 4C, Table S3), which are all hallmarks of extracellular
246 merozoites. On the other hand, the other cells of the integrated dataset express the *rex2* ring
247 marker gene and are therefore rings (Figure S4E).

248 The few marker genes in common between the clusters 4 and 13 cells are signature
249 genes of free merozoites (Figure S4F), indicating that the cells in the two clusters correspond
250 to two distinct sub-populations of merozoites. To allocate an identity to these sub-populations,
251 we did trajectory analysis. Cluster 4 cells have the earliest pseudotime (Figures 4D and S4G,
252 Table S1) and express invasion and motility genes (Figure 4C, Table S3) so we conclude that
253 these 141 cells are asexual parasites. 17% of these cells are free merozoites (Table S1). We
254 propose that the other cells are very early rings that still express the gene expression program
255 of free merozoites.

256 The cells trajectory terminates in cluster 5 (Figures 4D and S4G). Cluster 5 parasites
257 express gametocyte-related genes (Figure 4C, Table S3), namely *ap2-g* (Figure 4E), the
258 master regulator of gametocytogenesis (40,41), indicating that they have entered sexual
259 commitment. This is further supported by the fact that, of the 35 marker genes of cluster 5
260 parasites (Figure 4F, Table S3), 25 were previously identified as markers of sexually-
261 committed cells (7,42) (Figure 4G, Table S3) and 57% are ChIP-Seq targets of AP2-G (39)
262 (Figure 4F, Table S3). The cells of clusters 12 and 13, preceding cluster 5, do not express
263 *ap2-g* (Figure 4E). They do, however, over-express genes previously shown to be expressed
264 in gametocyte precursor cells by another scRNA-seq study (42). 29% of the cluster 12 marker
265 genes and 43% of the cluster 13 marker genes (Figure 4H, Table S3) were shown by that
266 study to be expressed by sexually-committed parasites, before bifurcation of the cells into the
267 male and female gametocyte trajectories (42). This indicates that clusters 12 and 13 precede
268 sexual differentiation. We used the definition of Dogga *et al*. (7) that defined the beginning of
269 sexual commitment as the moment when parasites enter the sexual trajectory and here name
270 the cluster 12 parasites, early sexually committed, and the cluster 13 parasites, sexually

271 committed. To distinguish these parasites from the ones in cluster 5, we herein call the cluster
272 5 parasites, sexual. We have then identified two early states of sexual commitment, preceding
273 expression of the *ap2-g* master regulator. The early commitment state (cluster 12) is defined
274 by the expression of 34 marker genes, the commitment state (cluster 13) by the expression of
275 242 marker genes, and there are 22 signature genes of sexual parasites (cluster 5) (Table
276 S3). We identified 60 early sexually-committed parasites, 46 committed parasites and 130
277 sexual parasites (Table S1), corresponding altogether to 17% of the flow-through population,
278 as expected since few parasites exit asexual growth and sexually-commit.
279

280 **Early sexually committed and committed cells have distinct gene expression**
281 **signatures**

282 We compared the gene expression profile of the early committed (cluster 12) and
283 committed parasites (cluster 13) with that of the asexual and the sexual parasites (cluster 4
284 and cluster 5, respectively) to identify differences among them. Apart from the free merozoites
285 marker genes, that as noted before, are also expressed in committed parasites (Figure S4F),
286 marker genes of asexual parasites are under-expressed in early committed and committed
287 parasites (Figure 5A, Table S3). Similarly, marker genes of early committed and committed
288 parasites are down-expressed in asexual parasites (Figures 5B-C, Table S3). Further
289 supporting that asexual and sexually-committed parasites have distinct gene expression
290 profiles, several of the sexual marker genes of cluster 5 are already expressed in sexually-
291 committed parasites, but not in asexual cells (Figure 5D, Table S3). That is the case of
292 *gametocyte exported protein 2* (*gexp02*), *male development protein 4* (*md4*), *parasitophorous*
293 *vacuole membrane protein 16* (*pfs16*), *surface-associated interspersed protein 8.1* (*surfin8.1*)
294 and *early-transcribed membrane protein 4* (*etramp4)* (Figure 5D, Table S3). This concords
295 with previous studies that showed that some early sexual markers are transcribed in the
296 absence of AP2-G (43–46).

297 Many of the marker genes of early-committed and committed parasites encode ApiAP2
298 DNA-binding proteins (Table S3) and committed parasites also over-express genes coding for

299 zinc finger and RNA-binding proteins (Figure S5B, Table S3). This is compatible with the
300 initiation of a distinct gene expression program that prepares the parasite to exit asexual
301 growth and transform into the sexual stages.

302 Early committed genes are still expressed in sexually-committed parasites, and vice-
303 versa, but each cell type has a distinct gene expression profile (Figures 5B-C). GO-term
304 analysis of the marker genes of clusters 12 and 13 show that early committed cells encode
305 proteins to be exported to the host cell cytosol and membrane, and early committed cells code
306 for cytoskeleton and DNA replication proteins (Figure 4C, Table S3).

307 We detect transcripts for female and male markers in early committed and committed
308 parasites, suggesting, like previous studies (42), that some sex-specific transcripts are
309 expressed early on during sexual commitment. Namely, *ap2-g5*, linked to the male lineage
310 (42), is over-expressed by early committed parasites, and *ap2-fg*, *ap2-o* and *ap2-o2*,
311 associated with the female lineage (42,47), are marker genes of committed parasites (Figures
312 S4A-B, Table S3).

313 To identify potential regulators of sexual commitment, we determined which *apiap2*
314 genes are expressed in asexual (cluster 4), early committed (cluster 12), committed (cluster
315 13) and sexual (cluster 5) parasites (Figure 5E). *ap2-lt* is highly expressed in asexual cells
316 and it continues to be expressed during sexual commitment (Figure 5E). As mentioned, AP2-
317 LT may regulate expression of the free merozoites genes but its expression in early
318 gametocytes was also reported (42), so it is likely that, like other ApiAP2 proteins, AP2-LT
319 regulates multiple steps of the parasite lifecycle. *pf3d7_0420300* has the same expression
320 pattern as *ap2-lt* (Figure 5E) and its expression during gametocytogenesis was also recently
321 put forward (42). Expression of *ap2-exp* and *pf3d7_1115500* is higher in asexual and sexual
322 parasites than during sexual commitment (Figure 5E). AP2-EXP regulates expression of
323 genes encoding exported proteins (35,48) so this might reflect a specific need of the asexual
324 and sexual parasites to export proteins to the host cell cytosol and surface. AP2-G5 represses
325 gametocytogenesis by suppressing expression of *ap2-g* (49) and its transcription decreases
326 in sexual cells (Figure 5E).

327 *ap2-p*, which has been linked to female gametocyte development (42), is a marker
328 gene of early committed parasites (Figures 5B and S5A, Table S3), so we used two published
329 ChIP-Seq datasets (50,51), to determine how many of the early commitment gene promoters
330 are bound by AP2-P. Strikingly, 79% of the cluster 12 marker genes are AP2-P ChIP-Seq
331 targets (Figure 5F, Table S3). AP2-P is implicated in chromatin boundary insulation, together
332 with MORC (51), and, according to previous ChIP-Seq assays (51–53), some of the early
333 committed gene promoters are also bound by MORC (Figure 5F, Table S3). Our results then
334 suggest a role for AP2-P in regulating early sexual commitment, perhaps in complex with
335 MORC. *ap2-z*, which is transcribed in gametocytes but only translated in the mosquito, is also
336 highly expressed by early committed parasites (Figure 5E).
337 Six *apiap2* genes are markers of committed cells: *ap2-i*, *pf3d7_1239200*,
338 *pf3d7_0613800*, *ap2-fg*, *ap2-o* and *ap2-o2* (Figure S5B, Table S3). Three others, *ap2-o4*, *ap2-*
339 *g4* and *pf3d7_0934400*, are highly expressed in these cells (Figure 5E). *ap2-o*, *ap2-i*, *ap2-g4*
340 and *pf3d7_1239200* have previously been shown to be expressed in early gametocytes (42)
341 so they may play a role in the early steps of sexual commitment. This is further supported by
342 the fact that AP2-I interacts with AP2-G (39) and that AP2-G4 plays a positive regulatory role
343 in gametocytogenesis (49). However, comparison of published ChIP-Seq datasets of AP2-I
344 and AP2-G4 (38,39,49), with the list of marker genes of committed cells, only identified a few
345 target genes (data not shown).
346 *ap2-g* is only expressed in sexual parasites (Figure 5E) and it is a marker gene of these
347 cells (Figures 4F and 5D, Table S3). Since *ap2-g* is transcribed mainly at the ring and late
348 schizont stages (45), it is not surprising that we detect its expression in our dataset of rings
349 and merozoite parasites. Sexual parasites also express *ap2-fg* and *ap2-o2* (Figure 5E) that
350 are expressed in mature gametocytes (42,47).
351
352 **Early sexually-committed rings can be produced by next cycle conversion**

353 We identified one sexual free merozoite and three sexually-committed merozoites but
354 we were unable to identify early sexually-committed free merozoites (Table S1). Because *P.*
355 *falciparum* parasites can sexually-commit at the ring-stage by same cycle conversion (SCC),
356 without activation of sexual commitment, prior to invasion (45), we wondered if the ring-stage
357 parasites we labelled as early committed are produced by SCC. This would explain why there
358 are no early committed merozoites. Alternatively, the early sexually-committed rings were
359 generated, upon RBC invasion of early committed merozoites (next cycle conversion, NCC) ,
360 and we did not identify early committed merozoites due to reduced cell numbers. There are
361 minor transcriptional differences between sexually-committed parasites produced by NCC or
362 SCC (45) so to explore this, we searched for intracellular parasites expressing the early
363 commitment genes. To do that, we passed the schizonts that had been retained in the
364 magnetic columns through a filter (6) and prepared them for single cell RNA-sequencing. We
365 obtained a similar number of reads, sequencing saturation and reads mapped to the genome
366 in the intracellular cells, as compared to the flow-through cells (Table S1). After filtering with
367 the same parameters used for the flow-through cells, we retained 191 intracellular cells (Table
368 S1).

369 To discover intracellular cells expressing early-committed genes, we integrated the
370 data for the 191 intracellular cells with the combined flow-throughs, and applied Seurat to this
371 new integrated dataset (Supplementary file 4). We identified again 13 clusters (Figures 6A
372 and S6A, Table S1). Cluster 11 cells express early commitment genes, parasites expressing
373 commitment genes are in cluster 10, and cluster 7 cells, succeeding cluster 10, express
374 sexual marker genes (Figure S6B, Table S4). 8 of the 66 early committed parasites and 24 of
375 the 69 committed parasites, are intracellular (Figure 6B, Table S1). Since we identified
376 intracellular parasites expressing early commitment and commitment genes, we conclude that
377 early committed ring-staged parasites can be generated by NCC. This does not exclude that
378 they can also be generated by SCC.

379

380 **Discussion**

381 Here we used single cell RNA-sequencing to identify the gene expression profile of the
382 RBC invasive stages of the malaria parasite, the merozoites. We validated these cells identity
383 by showing that they express the merozoite marker *msp1*, that several of the free merozoite
384 genes identified here encode previously identified merozoite proteins (13,14) and that almost
385 all of the marker genes of these cells were previously shown to be over-expressed in
386 merozoites (4).

387 Extracellular merozoites express invasion and motility genes, as expected from cells
388 that need to glide and invade the host. The promoters of the free merozoites signature genes
389 are already marked with active histone marks during schizogony and continue to be so at
390 merozoites (4), suggesting that the genes are actively transcribed in the two developmental
391 stages. This is further supported by the genes peak transcription time, according to a previous
392 study (29). After invasion, the genes are no longer transcribed. Bulk RNA-seq showed that
393 invasion genes are transcribed during schizogony (54–56) but here we show, by doing
394 comparative analysis, that these genes are also transcribed in free merozoites.

395 Sexual merozoites had never been identified but since *P. falciparum parasites* must
396 invade RBCs before sexually committing (2), these had to exist. We identified these rare cells
397 as well as sexually-committed parasites. We propose that parasites enter sexual
398 differentiation when they express *ap2-g* and thus called the cells expressing *ap2-g*, sexual.
399 These sexual cells express many of the genes expressed by what Dogga *et al*. called
400 committed cells (7). Here we show that two gene expression commitment programs, that we
401 call early sexual-commitment and sexual-commitment, precede expression of the sexual
402 genes. The clusters corresponding to the early sexually-committed and committed parasites
403 were defined by Seurat, using standard parameters, and were thus found in an unsupervised
404 manner. Each cluster contains several cells and it is characterized by the expression of many
405 marker genes. There are 242 marker genes of sexually-committed cells and 34 markers of
406 early committed cells. In comparison, there are 22 marker genes of sexual cells. Intracellular

407 parasites also express early sexual commitment and commitment genes so expression of
408 these genes is not restricted to free merozoites and rings. Early sexually-committed and
409 committed trophozoites must also exist but we did not have this parasite stage in our cultures.
410 Even though many of the genes expressed by what we here call early committed and
411 committed cells were previously identified as being expressed in early gametocytes (42), the
412 existence of an early sexual commitment state was not reported. We think that in our work we
413 were able to detect more slight changes in gene expression and this way distinguish between
414 early sexually-committed, committed and sexual cells, because we only analyzed two parasite
415 stages, while other studies used mixed stages.

416 We propose that the early sexual commitment steps identified here correspond to an
417 early stage of sexual commitment that primes parasites for sexual differentiation. Definitive
418 commitment into gametocytogenesis would then require *ap2-g* expression. Further studies are
419 required to test this hypothesis.

420 *P. falciparum* parasites can sexually differentiate at the ring-stage, without expressing
421 AP2-G in the previous IDC via the SCC pathway (45). This occurs when AP2-G protein levels
422 reach a threshold sufficient in rings to trigger the sexual differentiation program (45). It is
423 unknown how many of the ring parasites expressing *ap2-g*, of our dataset, also express the
424 AP2-G protein, but since only 7 of the 134 sexual cells are intracellular and, as mentioned,
425 only one of those cells is a free merozoite, it is possible that most of the sexual rings we
426 identified were produced by SCC, as proposed by Bancells *et al.* (45).

427 We propose that AP2-LT regulates expression of the free merozoites genes and that
428 AP2-P may regulate the early commitment expression program because many of the marker
429 genes of free merozoites or early commitment parasites are ChIP-Seq targets of AP2-LT or
430 AP2-P (34,35,50,51), respectively. AP2-LT is part of the SAGA activating complex, together
431 with GCN5 (36,37,57,58), and GCN5 also binds upstream of several free merozoite genes
432 (36,37). *ap2-lt* is also expressed during sexual differentiation, as shown here and in a previous
433 study (42), so this protein may have multiple roles. The precise role of AP2-LT in regulating

434 gene expression requires the establishment of a conditional knockout or knockdown, since
435 gene disruption seems impossible (33).
436 AP2-P, which may or may not regulate expression of the early committed genes,
437 establishes chromatin boundaries, together with MORC, and specific ApiAP2 proteins (51).
438 While many early committed gene promoters are bound by AP2-P (50,51), fewer are bound
439 by MORC (51–53), so it is unclear if AP2-P associates with MORC during sexual commitment.
440 AP2-P has been previously characterized but its implication in sexual commitment was not
441 explored (51–53). The existing AP2-P conditional knockout was done in a parasite strain that
442 is a poor gametocyte-producer (50), so to study the function of the protein in early sexual
443 commitment, a new transgenic line in a more gametocyte prone strain, would have to be
444 established. Alternatively, one could use the glmS conditional knockdown system but this
445 system does not induce complete protein shutdown (51).
446 In summary, we identified here the gene expression profile of extracellular merozoites,
447 as well of the early stages of sexual commitment, despite the rarity of these cells. This study
448 reveals how the parasite prepares for invasion and how it exits asexual growth and enters
449 sexual differentiation. These transitions ensure survival of the parasite in the human host and
450 transmission to the mosquito vector so the gene markers identified here can be therapeutically
451 explored in the future to block malaria invasion or transmission.
452
453 **Materials and Methods**
454 **Parasite culturing and merozoites purification**
455 *P. falciparum* W2mef parasites were cultured at 37°C in the presence of 5% oxygen and 5%
456 carbon dioxide, in RPMI 1640 media (Fisher) supplemented with hypoxanthine and 0.5%
457 Albumax II (Invitrogen) in 2% hematocrit. Blood was obtained from the Etablissement Français
458 du Sang (EFS).

459 Parasite synchronization to the ring stage was done by incubating pelleted cultures 1:1 with
460 0.3M L-Alanine, 10mM HEPES pH 7.5 solution for 10 min at 37°C, after which cultures were
461 centrifuged, resuspended in fresh media and placed back in culture (59).
462 For scRNA-seq, two mixed parasite cultures of non-infected RBCs, schizonts, rings and free
463 merozoites were passed through two LD columns (Miltenyi Biotec), washed with RPMI culture
464 and eluted with RPMI into Falcon tubes. The flow-through, corresponding to the ring stage
465 parasites, free merozoites and non-infected red blood cells, was also recovered. The elution
466 corresponds to the schizont fraction. Schizonts were passed through a 1.2µM filter (Pall) to
467 release the intracellular cells inside (6). After centrifugation, the intracellular or the flow-
468 through cells were counted with a Neubauer chamber and diluted with 1XPBS/0.04%BSA to
469 1 million cells/ml before loading into the Chromium.
470 The purity of each fraction was verified by Giemsa-stained blood smear. Images of blood
471 smears were acquired with the widefield microscope Eclipse (Nikon) at 60x amplification.
472 To stain with DAPI, purified schizonts were fixed in 1XPBS/4%formaldehyde, washed and
473 deposited into glass slides mounted with VectaShield Antifade medium containing DAPI
474 (Vector laboratories) and imaged with a SP8 confocal microscope (Leica) at 63x amplification.
475 Images were treated with ImageJ.
476
477 **Single cell RNA-Sequencing**
478 Approximately 10.000 cells (intracellular or flow-through) were loaded on a Chromium Chip G
479 (10x Genomics) as per the manufacturer's recommendations. Chips containing cell
480 suspensions, gel beads in emulsion (GEMs) and reverse transcription reagents were loaded
481 into the chromium controller (10x Genomics) for single cell partitioning and cell barcoding.
482 Barcoded cDNA and sequencing libraries were generated according to the Chromium Next
483 GEM Single Cell 3' gene expression v3.1 protocol. Quantification and the quality of
484 sequencing libraries was verified on an Agilent Bioanalyzer. Pooled libraries were sequenced
485 on a NextSeq 2000 system (Illumina).
486

487 **Data analysis**

488 Quality control checks on the raw sequencing data were performed using FASTQC (60).
489 FastQ files were analyzed using the CellRanger software (10X Genomics), including
490 alignment, filtering and quantitation of reads on the *Plasmodium falciparum* genome (3D7)
491 and generation of feature-barcode count matrices. During mapping, the reads corresponding
492 to the human host cell were discarded.

493 Genes detected in fewer than three cells were removed. Cells with more than 0.5% of
494 mitochondrial genes were removed. Cells with fewer than 200 genes and more than 2500
495 genes per cells were removed.

496 The Seurat toolbox version 4.3.0.1 was used to normalize each dataset using SCTransform
497 (61) and cluster the cells in an unsupervised manner, using standard conditions (all code used
498 is in the supplementary files). To determine if the cells recovered in the flow-through of the
499 second magnetic purification express the same expression clusters as the ones recovered in
500 the flow-through of the first purification, we looked for cells expressing marker genes of each
501 of the flow-through clusters and then manually identified them as belonging to that cluster.

502 Visualizations and identification of differentially expressed genes were done using the Loupe
503 browser (10X Genomics, version 7).

504 Graphs showing the binding pattern of the active histone marks H3K9ac, H3K4me3 and
505 H3K27ac in the gene promoters and gene bodies of the free merozoites were generated using
506 computeMatrix and plotProfile from the deepTools package (62) of Galaxy (63) using the
507 ChIP-Seq data from.

508 Pseudotime and trajectory analysis were calculated using Monocle3 R package version 1.3.4
509 (64) (thecode used is in the supplementary file 3).

510 The ATAC-Seq analyses were done by comparing the genes expressed in each cluster to
511 those identified in (32). Whenever more than one peak was assigned to a gene promoter, we
512 only considered the earlier time point. For most genes, there was only one peak of
513 accessibility.

514 The transcription peak time of the marker genes expressed in each cluster was calculated in
515 (29).
516 To determine the percentage of genes encoding proteins detected in mature merozoites by
517 mass spectrometry we compared our dataset to the high-throughput LC-MS/MS of *P*.
518 *falciparum* 3D7 merozoites data generated by Shahid Kahn from Leiden University that
519 identified over 1000 proteins (13) or to the data of Kumar *et al*. (14). We only considered
520 proteins detected with at least 2 peptides.
521 The ChIP-Seq comparison analyses were done by determining if the genes expressed in a
522 specific cluster were identified as target genes in (34–39,48,50–53).
523 The percentage of genes over-expressed in merozoites versus rings or schizonts was
524 determined by comparing our dataset to that of (4).
525 Venn diagrams were generated using the tool BioVenn (65).
526 Gene Ontology (GO) enrichment analysis was done using the GO feature at PlasmoDB
527 (release 68) (66).
528 Graph plots were generated with GraphPad Prism version 10.0.0 for Mac, GraphPad
529 Software, Boston, Massachusetts USA, www.graphpad.com.
530
531 **Data availability**
532 All RNA sequencing data generated in this study has been deposited in NCBI's Gene
533 Expression Omnibus and is accessible through GEO series accession number GSE279639.
534
535 **Code availabity**
536 All code used in this study is presented in the Supplementary files.
537
538 **Acknowledgements**
539 This work was funded by a CNRS ATIP-Avenir grant, an IBISA grant and I2BC start-up funds
540 awarded to JMS**.** This work was supported as part of the France 2030 programme ANR-11-
541 IDEX-0003, which includes a research grant awarded by the interdisciplinary program OI

LivingMachines@Work from University Paris-Saclay to JMS. The I2BC High-throughput sequencing facility is supported by *France Génomique* (funded by the French National Program “Investissement d’Avenir” ANR-10-INBS-09). The present work has benefited from Imagerie-Gif core facility supported by l’Agence Nationale de la Recherche (ANR-11-EQPX-0029/Morphoscope, ANR-10-INBS-04/FranceBioImaging; ANR-11-IDEX-0003-02/ Saclay Plant Sciences). W2mef parasites were obtained through BEI Resources, NIAID, NIH: *Plasmodium falciparum*, Strain W2mef, MRA-165, contributed by Alan F. Cowman. The PlasmoDB.org resource was invaluable to this work. We thank D. Noordemeer, V. Jeffers and H. Painter for critical reading of the manuscript, and K. Frenal for providing the glideosome image.

Authors contribution

Conceptualization: JMS

Parasite culturing: LP, JMS, LM

Single cell experiment: LP, JMS, YJ

Data analysis: DN, CH, JMS

Writing: JMS

Competing interests

The authors declare no competing interests

References

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

2. 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.

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

573 4. Reers AB, Bautista R, McLellan J, Morales B, Garza R, Bol S, et al. Histone modification
574 analysis reveals common regulators of gene expression in liver and blood stage
575 merozoites of *Plasmodium* parasites. Epigenetics Chromatin. 2023 June 15;16(1):25.

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

579 6. Boyle MJ, Wilson DW, Richards JS, Riglar DT, Tetteh KKA, Conway DJ, et al. Isolation of
580 viable *Plasmodium falciparum* merozoites to define erythrocyte invasion events and
581 advance vaccine and drug development. Proc Natl Acad Sci U S A. 2010 Aug
582 10;107(32):14378–83.

583 7. Dogga SK, Rop JC, Cudini J, Farr E, Dara A, Ouologuem D, et al. A single cell atlas of
584 sexual development in *Plasmodium falciparum*. Science. 2024 May
585 3;384(6695):eadj4088.

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

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

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

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

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

597 13. Kahn S. Merozoite Proteome (3D7). PlasmoDB; 2008.
598 https://plasmodb.org/plasmo/app/record/dataset/DS_388d40eae3

599 14. Kumar K, Srinivasan P, Nold MJ, Moch JK, Reiter K, Sturdevant D, et al. Profiling invasive
600 *Plasmodium falciparum* merozoites using an integrated omics approach. Sci Rep. 2017
601 Dec 7;7(1):17146.

602 15. Cova MM, Lamarque MH, Lebrun M. How Apicomplexa Parasites Secrete and Build Their
603 Invasion Machinery. Annu Rev Microbiol. 2022 Sept 8;76:619–40.

604 16. Gao X, Gunalan K, Yap SSL, Preiser PR. Triggers of key calcium signals during
605 erythrocyte invasion by *Plasmodium falciparum*. Nat Commun. 2013 Nov 27;4(1):2862.

606 17. Absalon S, Blomqvist K, Rudlaff RM, DeLano TJ, Pollastri MP, Dvorin JD. Calcium-
607 Dependent Protein Kinase 5 Is Required for Release of Egress-Specific Organelles in
608 *Plasmodium falciparum*. mBio. 2018 Feb 27;9(1):10.1128/mbio.00130-18.

609 18. Sim BK, Chitnis CE, Wasniowska K, Hadley TJ, Miller LH. Receptor and ligand domains
610 for invasion of erythrocytes by *Plasmodium falciparum*. Science. 1994 June
611 24;264(5167):1941–4.

612 19. Singh S, Alam MM, Pal-Bhowmick I, Brzostowski JA, Chitnis CE. Distinct External Signals
613 Trigger Sequential Release of Apical Organelles during Erythrocyte Invasion by Malaria
614 Parasites. PLoS Pathog. 2010 Feb 5;6(2):e1000746.

615 20. Ben Chaabene R, Lentini G, Soldati-Favre D. Biogenesis and discharge of the rhoptries:
616 Key organelles for entry and hijack of host cells by the Apicomplexa. Mol Microbiol. 2021
617 Mar;115(3):453–65.

618 21. Cowman AF, Tonkin CJ, Tham WH, Duraisingh MT. The Molecular Basis of Erythrocyte
619 Invasion by Malaria Parasites. Cell Host & Microbe. 2017 Aug 9;22(2):232–45.

620 22. Saini E, Sheokand PK, Sharma V, Agrawal P, Kaur I, Singh S, et al. *Plasmodium*
621 *falciparum* PhIL1-associated complex plays an essential role in merozoite reorientation
622 and invasion of host erythrocytes. PLOS Pathogens. 2021 July 29;17(7):e1009750.

623 23. Wichers JS, Wunderlich J, Heincke D, Pazicky S, Strauss J, Schmitt M, et al. Identification
624 of novel inner membrane complex and apical annuli proteins of the malaria parasite
625 *Plasmodium falciparum*. Cellular Microbiology. 2021;23(9):e13341.

626 24. Volz JC, Yap A, Sisquella X, Thompson JK, Lim NTY, Whitehead LW, et al. Essential Role
627 of the PfRh5/PfRipr/CyRPA Complex during *Plasmodium falciparum* Invasion of
628 Erythrocytes. Cell Host Microbe. 2016 July 13;20(1):60–71.

629 25. Triglia T, Scally SW, Seager BA, Pasternak M, Dagley LF, Cowman AF. Plasmepsin X
630 activates the PCRCR complex of *Plasmodium falciparum* by processing PfRh5 for
631 erythrocyte invasion. Nat Commun. 2023 Apr 19;14(1):2219.

632 26. Yahata K, Hart MN, Davies H, Asada M, Wassmer SC, Templeton TJ, et al. Gliding motility
633 of *Plasmodium* merozoites. Proc Natl Acad Sci U S A. 2021 Nov
634 30;118(48):e2114442118.

635 27. Frénal K, Dubremetz JF, Lebrun M, Soldati-Favre D. Gliding motility powers invasion and
636 egress in Apicomplexa. Nat Rev Microbiol. 2017 Nov;15(11):645–60.

637 28. Hoshizaki J, Adjalley S, Thathy V, Judge K, Berriman M, Reid AJ, et al. A manually curated
638 annotation characterises genomic features of P. falciparum lncRNA. 2022.
639 https://europepmc.org/article/PPR/PPR516393

640 29. Painter HJ, Chung NC, Sebastian A, Albert I, Storey JD, Llinás M. Genome-wide real-time
641 in vivo transcriptional dynamics during *Plasmodium falciparum* blood-stage development.
642 Nat Commun. 2018 July 9;9(1):2656.

643 30. Connacher J, von Grüning H, Birkholtz L. Histone Modification Landscapes as a Roadmap
644 for Malaria Parasite Development. Front Cell Dev Biol. 2022 Apr 1;10:848797.

645 31. Tang J, Chisholm SA, Yeoh LM, Gilson PR, Papenfuss AT, Day KP, et al. Histone
646 modifications associated with gene expression and genome accessibility are dynamically
647 enriched at *Plasmodium falciparum* regulatory sequences. Epigenetics Chromatin. 2020
648 Nov 23;13(1):50.

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

653 33. Singhal R, Prata IO, Bonnell VA, Llinás M. Unraveling the complexities of ApiAP2
654 regulation in *Plasmodium falciparum*. Trends Parasitol. 2024 Nov;40(11):987–99.

655 34. Bonnell VA, Zhang Y, Brown AS, Horton J, Josling GA, Chiu TP, et al. DNA sequence and
656 chromatin differentiate sequence-specific transcription factor binding in the human malaria
657 parasite *Plasmodium falciparum*. Nucleic Acids Res. 2024 July 5;gkae585.

658 35. Shang X, Wang C, Fan Y, Guo G, Wang F, Zhao Y, et al. Genome-wide landscape of
659 ApiAP2 transcription factors reveals a heterochromatin-associated regulatory network
660 during *Plasmodium falciparum* blood-stage development. Nucleic Acids Res. 2022 Apr
661 8;50(6):3413–31.

662 36. Lucky AB, Shakri AR, Liang X, Min H, Li XL, Adapa SR, et al. GCN5 Is a Master Regulator
663 of Gene Expression in the Malaria Parasite *Plasmodium falciparum*. Cells. 2025 June
664 10;14(12):876.

665 37. Tang J, Yeoh LM, Grotz MD, Goodman CD, Chisholm SA, Nguyen HHT, et al. PfGCN5 is
666 essential for *Plasmodium falciparum* survival and transmission and regulates Pf H2B.Z
667 acetylation and chromatin structure. Nucleic Acids Res. 2025 Apr 11;53(6):gkaf218.

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

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

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

676 41. Sinha A, Hughes KR, Modrzynska KK, Otto TD, Pfander C, Dickens NJ, et al. A cascade
677 of DNA-binding proteins for sexual commitment and development in *Plasmodium*. Nature.
678 2014 Mar 13;507(7491):253–7.

679 42. Mohammed M, Dziedziech A, Macedo D, Huppertz F, Veith Y, Postel Z, et al. Single-cell
680 transcriptomics reveal transcriptional programs underlying male and female cell fate
681 during *Plasmodium falciparum* gametocytogenesis. Nat Commun. 2024 Aug
682 26;15(1):7177.

683 43. Tibúrcio M, Dixon MWA, Looker O, Younis SY, Tilley L, Alano P. Specific expression and
684 export of the *Plasmodium falciparum* Gametocyte EXported Protein-5 marks the
685 gametocyte ring stage. Malaria Journal. 2015 Aug 28;14(1):334.

686 44. Painter HJ, Carrasquilla M, Llinás M. Capturing in vivo RNA transcriptional dynamics from
687 the malaria parasite *Plasmodium falciparum*. Genome Res. 2017 June;27(6):1074–86.

688 45. Bancells C, Llorà-Batlle O, Poran A, Nötzel C, Rovira-Graells N, Elemento O, et al.
689 Revisiting the initial steps of sexual development in the malaria parasite *Plasmodium*
690 *falciparum*. Nat Microbiol. 2019 Jan;4(1):144–54.

691 46. Rovira-Graells N, Gupta AP, Planet E, Crowley VM, Mok S, Ribas de Pouplana L, et al.
692 Transcriptional variation in the malaria parasite *Plasmodium falciparum*. Genome Res.
693 2012 May;22(5):925–38.

694 47. Russell AJC, Sanderson T, Bushell E, Talman AM, Anar B, Girling G, et al. Regulators of
695 male and female sexual development are critical for the transmission of a malaria parasite.
696 Cell Host & Microbe. 2023 Feb 8;31(2):305-319.e10.

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

700 49. Shang X, Shen S, Tang J, He X, Zhao Y, Wang C, et al. A cascade of transcriptional
701 repression determines sexual commitment and development in *Plasmodium falciparum*.
702 Nucleic Acids Research. 2021 Sept 20;49(16):9264–79.

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

706 51. Singh P, Serizay J, Couble J, Cabahug MD, Rosa C, Chen P, et al. High-resolution map
707 of the *Plasmodium falciparum* genome reveals MORC/ApiAP2-mediated links between
708 distant, functionally related genes. Nat Microbiol. 2025 July;10(7):1665–83.

709 52. Chahine ZM, Gupta M, Lenz T, Hollin T, Abel S, Banks CA, et al. PfMORC protein
710 regulates chromatin accessibility and transcriptional repression in the human malaria
711 parasite, *Plasmodium falciparum*. eLife. 2024 Oct 17;12.

712 53. Singh MK, Bonnell VA, Silva ITD, Santiago VF, Moraes MS, Adderley J, et al. A
713 *Plasmodium falciparum* MORC protein complex modulates epigenetic control of gene
714 expression through interaction with heterochromatin. eLife. 2024 Oct 16;12:RP92201.

715 54. Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. The Transcriptome of the
716 Intraerythrocytic Developmental Cycle of *Plasmodium falciparum*. PLOS Biology. 2003
717 Aug 18;1(1):e5.

718 55. Chappell L, Ross P, Orchard L, Russell TJ, Otto TD, Berriman M, et al. Refining the
719 transcriptome of the human malaria parasite *Plasmodium falciparum* using amplification-
720 free RNA-seq. BMC Genomics. 2020 June 8;21(1):395.

721 56. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, et al. Discovery of gene
722 function by expression profiling of the malaria parasite life cycle. Science. 2003 Sept
723 12;301(5639):1503–8.

724 57. Cui L, Miao J, Furuya T, Li X, Su X zhuan, Cui L. PfGCN5-Mediated Histone H3 Acetylation
725 Plays a Key Role in Gene Expression in *Plasmodium falciparum*. Eukaryot Cell. 2007
726 July;6(7):1219–27.

727 58. Miao J, Wang C, Lucky AB, Liang X, Min H, Adapa SR, et al. A unique GCN5 histone
728 acetyltransferase complex controls erythrocyte invasion and virulence in the malaria
729 parasite *Plasmodium falciparum*. PLOS Pathogens. 2021 Aug 17;17(8):e1009351.

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

733 60. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data . 2010.
734 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

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

737 62. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a
738 next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016
739 July 8;44(W1):W160–5.

740 63. The Galaxy Community. The Galaxy platform for accessible, reproducible, and
741 collaborative data analyses: 2024 update. Nucleic Acids Res. 2024 July 5;52(W1):W83–
742 94.

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

746 65. Hulsen T, de Vlieg J, Alkema W. BioVenn – a web application for the comparison and
747 visualization of biological lists using area-proportional Venn diagrams. BMC Genomics.
748 2008 Oct 16;9(1):488.

749 66. Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, et al. PlasmoDB:
750 a functional genomic database for malaria parasites. Nucleic Acids Res. 2009
751 Jan;37:D539-543.

752

753 **Figure legends**
754

755 **Figure 1 – Single cell RNA-sequencing of merozoites.**

756 (**A**) During the intraerythrocytic developmental cycle (IDC), lasting 48 hours, parasite forms
757 called merozoites egress from schizonts and invade the red blood cell (RBC) to become first
758 rings and later on trophozoites. These transform into schizonts, within which new merozoites
759 are formed. Parasites can also sexually commit, leaving the IDC, eventually differentiating into
760 gametocytes. The merozoites are only briefly extracellular, after schizont egress and before
761 RBC invasion. Numbers below each stage indicate hours post-invasion, hpi. (**B**) Parasites
762 were magnetically purified from two different cultures containing 10% infected RBCs. As
763 determined by Giemsa-stained blood smears, the first parasite culture contained 90%
764 schizont-stage parasites, 9.9% ring-stage parasites and 0.1% free merozoites and the second
765 one 80% schizont-stage parasites, 19.9% ring-stage parasites and 0.1% free merozoites (R:
766 rings, S: schizonts, M: merozoites). The schizonts were retained in the magnetic columns and
767 in the flow-through, we collected the non-infected RBCs, the ring-stage parasites and the free
768 merozoites. The flow-through fractions from the two parasite cultures were processed
769 independently for single cell RNA-seq using 10x Genomics (the number of sequenced cells is
770 in Table S1). (**C**) Uniform Manifold Approximation and Projection (UMAP) of the flow-through
771 cells, from the two parasite cultures, colored according to the Seurat cluster. The number of
772 cells on each cluster is in Table S1 and the code used for the Seurat analysis is in the
773 supplementary files 1 and 2. (**D**) GO-term analysis of the marker genes of each flow-through
774 cluster. Because there is no GO-term associated with gametocytes, word search analysis was
775 done instead for cluster 2 genes. Only some GO-terms/words are shown. The list of marker
776 genes as well as of all GO-terms and words is in Table S2.
777

778 **Figure 2 – Gene signature of the free merozoites.**

779 (**A**) Volcano plot of the genes expressed in free merozoites. Genes overexpressed in these
780 cells, with a p-value equal or less than 0,01, versus other cells of the flow-through, are in red.
781 Some of the gene names are shown. The complete list of genes is in Table S2. (**B**) Heatmap
782 showing the log2 fold change expression value of the marker genes of the free merozoites. If
783 the genes code proteins implicated in entry into the host or in motility, according to GO-term
784 analysis, is indicated by a green or blue dot, respectively. The log2 fold change for each gene
785 and the full GO-term analysis is in Table S2 and Figure S2A.

786

787 **Figure 3 – Free merozoite genes are transcribed in schizonts and merozoites.**

788 (**A**) As determined by Painter *et al.* (29), most marker genes of free merozoites are peak
789 transcribed in schizonts (30-47hpi, purple) or merozoites (1/48hpi, pink), whereas a minority
790 are peak transcribed in rings (2-20hpi, black) or trophozoites (20-30hpi, green). The timing of
791 peak transcription for each gene is in Table S2. (**B**) Graph-plots showing binding of the histone
792 marks H3K9ac (dark blue), H3K27ac (light blue) and H3K4me3 (green and orange) to the
793 gene promoters, transcription start sites (TSS) and gene bodies of the free merozoite genes,
794 in schizont and merozoite parasites, as determined by the log2 ChIP/Input calculated by Reers
795 *et al*. (4). Binding in rings is in Figure S3A. (**C**) According to the ATAC-Seq dataset of
796 Toenhaken *et al.* (32), most free merozoites gene promoters are euchromatic in schizonts
797 (t30, t35 and t40 in blue, orange and gray, respectively), while less than 40% are euchromatic
798 in rings (t05, t10 and t15 in black, pink and sage, respectively) or trophozoites (t20 and t25, in
799 in green and purple, respectively). The ATAC-Seq peak for each gene is in Table S2. (**D**) Bar
800 graph showing how many free merozoite genes are ChIP-Seq targets of AP2-LT, GCN5 or
801 AP2-I, according to previous datasets (35–39). Which genes are ChIP-Seq targets of these
802 proteins is in Table S2.

803 **Figure 4 – Identification of early sexually-committed and sexually-committed cells.**

804 (**A**) Seurat clusters of the integrated flow-through dataset (the full list of genes expressed in
805 each cluster is in Table S3, the number of cells on each cluster is in Table S1 and the Seurat

806 analysis are in supplementary file 3). (**B**) There are two sub-groups of free merozoites, one in
807 cluster 4 and one in cluster 13, as determined by the expression of the top 10 free merozoite
808 markers genes. (**C**) GO-term analysis of the marker genes of clusters 4, 5, 12 and 13 showing
809 that cluster 4 cells express invasion-related genes and are thus asexual, cluster 12 cells code
810 exported proteins, cluster 13 cells express DNA replication, chromatin and cell-cycle genes,
811 while cluster 5 cells express gametocyte-related genes (the full GO-term analysis are in Table
812 S3). Bars are colored according to the cluster. (**D**) Pseudotemporal analysis of the merozoite
813 clusters showing that there is only one trajectory. The cells pseudotime is in Table S1 and the
814 Monocle analysis is in supplementary file 3. (**E**) UMAP representation showing that cells in
815 cluster 5 express *ap2-g*, indicating that they are destined to become gametocytes and are
816 therefore sexual. (**F**) Volcano plot of the genes expressed in cluster 5 cells versus the other
817 cells of the integrated dataset. Genes overexpressed with a p-value equal or less than 0,01
818 are in red. AP2-G ChIP-Seq targets, as determined by Josling *et al*. (39), are in blue. The
819 names of selected genes are shown. The full list of genes is in Table S3. (**G**) Graph showing
820 that many of the signature genes of cluster 5 cells (green) were previously reported by Dogga
821 *et al*. (7) or Mohammed *et al*. (42) as being up-regulated in committed cells or early
822 gametocytes, respectively. Several of the marker genes of cluster 12 (dark blue) or cluster 13
823 cells (light blue) were also detected in the Mohammed *et al*. dataset (42) (the list of common
824 genes between datasets is in Table S3). Bars are colored according to the Seurat cluster.
825
826 **Figure 5 – Comparison of the early sexually-committed and sexually-committed cells.**
827 (**A**) Heatmap showing the log2 expression levels of the signature genes of asexual cells
828 (cluster 4) on those cells as well as in the early committed, committed and sexual cells (cluster
829 12, cluster 13 and cluster 5, respectively), showing that, apart from the free merozoite gene
830 markers, which are also expressed in committed cells, asexual cells have a distinct gene
831 expression profile from other cells. Annotated rows correspond to genes expressed in free
832 merozoites. (**B**) Heatmap showing the log2 expression levels of the signature genes of early
833 committed cells (cluster 12) on those cells as well as in the asexual, committed and sexual

834 cells (cluster 4, cluster 13 and cluster 5, respectively), showing that some of those genes are
835 expressed in committed and sexual cells but not in asexual cells. (**C**) Heatmap showing the
836 log2 expression levels of the signature genes of committed cells (cluster 13) on those cells as
837 well as in the asexual, early committed and sexual cells (cluster 4, cluster 12 and cluster 5,
838 respectively), showing that several of those genes are already expressed in early committed
839 cells. (**D**) Heatmap showing the log2 expression levels of the signature genes of sexual cells
840 (cluster 5) on those cells as well as in the asexual, early committed and committed cells
841 (cluster 4, cluster 12 and cluster 13, respectively), showing that many of these genes are
842 already expressed in committed cells. The log2 expression data for **A**, **B**, **C** and **D** is on Table
843 S3. (**E**) Heatmap showing the log2 expression levels of *apiap2* genes in asexual, early
844 committed, committed and sexual cells (cluster 4, cluster 12, cluster 13 and cluster 5,
845 respectively). (**F**) Bar graph showing how many early committed genes are ChIP-Seq targets
846 of AP2-P or MORC, according to previous datasets (50–53). Which genes are targets is in
847 Table S3.

848 **Figure 6 – Early sexually-committed and sexually-committed genes are expressed by**
849 **intracellular parasites.**

850 (**A**) UMAP plot showing which Seurat clusters of the integrated dataset of the flow-through
851 and intracellular cells correspond to early committed, committed and sexual cells. The number
852 of cells in each cluster is in Table S1 and the Seurat analysis are in supplementary file 4. The
853 list of genes expressed by each cluster is in Table S4. The other cluster numbers are in Figure
854 S6A. (**B**) UMAP plots showing the number of intracellular cells (red) or flow-through cells
855 (black and green) on each cluster. The number and the type of cells (intracellular, flow-
856 through) in each cluster is in Table S1.

857

858 **Supplementary Information**
859

860 **Figure S1 – Preparation of parasites for single cell RNA-sequencing**.

861 (**A**) Giemsa-stained blood smears of schizonts before (left) and after (right) magnetic
862 purification. DAPI-staining of the purified schizonts (S) is also shown (BF: brightfield). Higher
863 magnification of a DAPI staining region is shown. (**B**) Giemsa-stained blood smear of the flow-
864 through fraction of the parasite culture 1 used for single cell RNA-sequencing showing the
865 presence of rings (R) and free merozoites (M). (**C**) Plots showing the number of genes and
866 UMIs detected per flow-through cell of the parasite culture 1. (**D**) Plots showing the number of
867 genes and UMIs detected per flow-through cell of the parasite culture 2. In **C** and **D**, each dot
868 represents a cell. (**E**) Percentage of flow-through cells expressing the merozoite marker *msp1*
869 or the ring marker *rex2*, showing that some cells express the merozoite marker. (**F**) Violin plot
870 showing the expression level of the merozoite marker *msp1* in the flow-through cells. (**G**) Violin
871 plot showing the expression level of the ring marker *rex2* in the flow-through cells. (**H**)
872 Heatmap showing differential expression of 100 randomly selected genes in the different
873 clusters of the two flow-throughs. Each row represents a cluster. (**I**) Bar graph showing how
874 many of the marker genes of the cells of the cluster 0 are over-expressed in merozoites versus
875 schizonts (black) or are over-expressed in merozoites versus rings (pink) according to Reers
876 *et al*. (4), or encode proteins identified in the two merozoite proteomes (green) (13,14) (The
877 log2 expression values determined by Reers *et al*. for each gene and which genes encode
878 proteins detected in the proteomes is in Table S2). (**J**) UMAP representation of the flow-
879 through cells of the two parasite cultures colored by number of genes or UMIs (reads) detected
880 per cell, showing that cluster 0 cells express less genes and have less UMIs per cell.
881
882
883 **Figure S2** – **Analysis of the marker genes of free merozoites**.
884 (**A**) GO-term analysis of the marker genes of the free merozoites. (**B**) Scheme of the parasite
885 glideosome, modified from (27), with the authorization of the authors, showing the different
886 molecular components. The annotated proteins are encoded by genes over-expressed in free
887 merozoites. PM: plasma membrane, IMC: inner membrane complex. (**C**) Graph showing the
888 average variability index score (VIS) of each free merozoite gene (pink) as compared to the

889 marker genes of other clusters (blue), as calculated by Tripathi *et al*. (8). The VIS score for
890 each gene is in Table S2.
891
892 **Figure S3** – **Gene expression analysis of the free merozoites**.
893 (**A**) Graph-plots showing binding of the histone marks H3K9ac (light blue), H3K27ac (yellow)
894 and H3K4me3 (dark blue), to the promoters, transcription start sites (TSS) and gene bodies
895 of the free merozoite genes in ring-stage parasites, as determined by Reers *et al*. (4). (**B**)
896 Percentage of marker genes of free merozoites over-expressed in merozoites versus
897 schizonts or rings or under-expressed in merozoites versus schizonts or rings, as determined
898 by Reers *et al*. (4). (**C**) Heatmap showing expression of *apiap2* genes in free merozoites.
899
900 **Figure S4** – **Identification of two early sexual commitment states**.
901 (**A**) UMAP representation of the Seurat clusters of the combined flow-through dataset for each
902 flow-through sample (the number of cells in each cluster is in Table S1, the marker genes of
903 each cluster is in Table S3, the code used is in supplementary file 3). (**B**) Graph showing
904 higher expression of the top 10 free merozoites marker genes in clusters 4 and 13. (**C**) UMAP
905 representation showing that some of the cells in clusters 4 and 13 contain less RNA (nCount)
906 and encode less transcripts (nFeature). UMAP representation showing that cells in cluster 4
907 and cluster 13 cells express the merozoite marker *msp1* (**D**) and under-express the ring
908 marker *rex2* (**E**), indicating that they are free merozoites. Violin plots are colored according to
909 the cluster. (**F**) There are only 16 common marker genes between the clusters 4 and 13 and
910 all of these are markers of free merozoites (the identity of these genes is in Table S3). (**G**)
911 Pseudotime of the cells in each cluster as determined by Monocle. Boxes are colored
912 according to the corresponding cluster. The pseudotime of each cell is Table S1 and the
913 Monocle analysis is in supplementary file 3.
914
915 **Figure S5** – **Marker genes of early committed and committed cells**.

916 Volcano plots showing in red the genes over-expressed by the early committed cells – cluster
917 12 (**A**), or the committed cells – cluster 13 (**B**) (the full list of genes is in Table S3). In (**B**), in
918 blue, are free signature genes. The names of some of the genes are shown.
919
920 **Figure S6** – **Identification of intracellular early committed and committed cells**.
921 (**A**) UMAP plot showing the Seurat clusters of the integrated dataset of the flow-throughs and
922 intracellular cells. Cells are colored according to the cluster (the number of cells in each cluster
923 is in Table S1 and the Seurat analysis are in supplementary file 4). (**B**) GO-term analysis of
924 the marker genes of early committed (cluster 11), committed (cluster 10) and sexual cells
925 (cluster 7). Bar graphs are colored according to the corresponding cluster. Only the most
926 significant GO-terms are shown. For the complete GO-term analysis, refer to Table S4.
927
928 **Table S1 – Number of cells and reads obtained after single cell RNA-sequencing,**
929 **number of cells in each Seurat cluster for the intracellular and flow-through cells.**
930
931 **Table S2 – Gene expression and GO-term analysis of the genes over-expressed on the**
932 **flow-through SEURAT clusters (it includes both flow-throughs). It includes the list of**
933 **marker genes of free merozoites. It includes the peak transcription, the ATAC-seq peak**
934 **and the VIS analysis, the comparative ChIP-Seq analysis and the comparison with the**
935 **Reers *et al*. data, as well as with the merozoite proteomes. The AP2-LT ChIP-Seq data is**
936 **from Bonnell *et al*. and Shang *et al*., the AP2-I ChIP-Seq data is from Santos *et al*. and Josling**
937 ***et al*., the GCN5 ChIP-Seq data is from Lucky *et al*. and Tang *et al*., the transcription and**
938 **stability peak time is from Painter *et al*., the ATAC-Seq data is from Toenhake *et al*., the VIS**
939 **data is from Tripathi *et al*., the merozoite vs ring/schizonts log2 expression data is from**
940 **Reers *et al*., the merozoite proteomes data is from Khan *et al*. and Kumar *et al*.**
941

942 **Table S3 – Gene expression and GO-term analysis of the genes over-expressed on the**
943 **integrated flow-through SEURAT clusters. It includes the list of marker genes of**
944 **asexual, early committed, committed and sexual cells, the pseudotime analysis, the**
945 **comparison with other sexual commitment scRNA-seq datasets and the comparative**
946 **ChIP-Seq analysis. The AP2-LT ChIP-Seq data is from Bonnell *et al*. and Shang *et al*., the**
947 **AP2-I ChIP-Seq data is from Santos *et al*. and Josling *et al*., the AP2-G ChIP-Seq data is from**
948 **Josling *et al*., the MORC ChIP-Seq data is from Singh *et al*. 2024, Singh et al. 2025 and**
949 **Chahine *et al*., the AP2-P ChIP-Seq data is from Singh *et al*. 2025 and Subudhi *et al*., the**
950 **sexual commitment data is from Dogga *et al*. and Mohammed *et al*..**
951
952 **Table S4 – Gene expression and GO-term analysis of the genes over-expressed on the**
953 **integrated dataset of the intracellular cells and two flow-throughs SEURAT clusters.**
954
955 **Supplementary file 1 – Code used to analyze the flow-through cells of parasite culture**
956 **1. It includes the Seurat analysis.**
957 **Supplementary file 2 – Code used to analyze the flow-through cells of parasite culture**
958 **2. It includes the Seurat analysis.**
959 **Supplementary file 3 – Code used to integrate the two flow-through datasets. It includes**
960 **the Seurat and Monocle analysis.**
961 **Supplementary file 4 – Code used to integrate the two flow-through datasets with the**
962 **intracellular cells. It includes the Seurat analysis.**
963
964
965
966
967
968

Figure 1

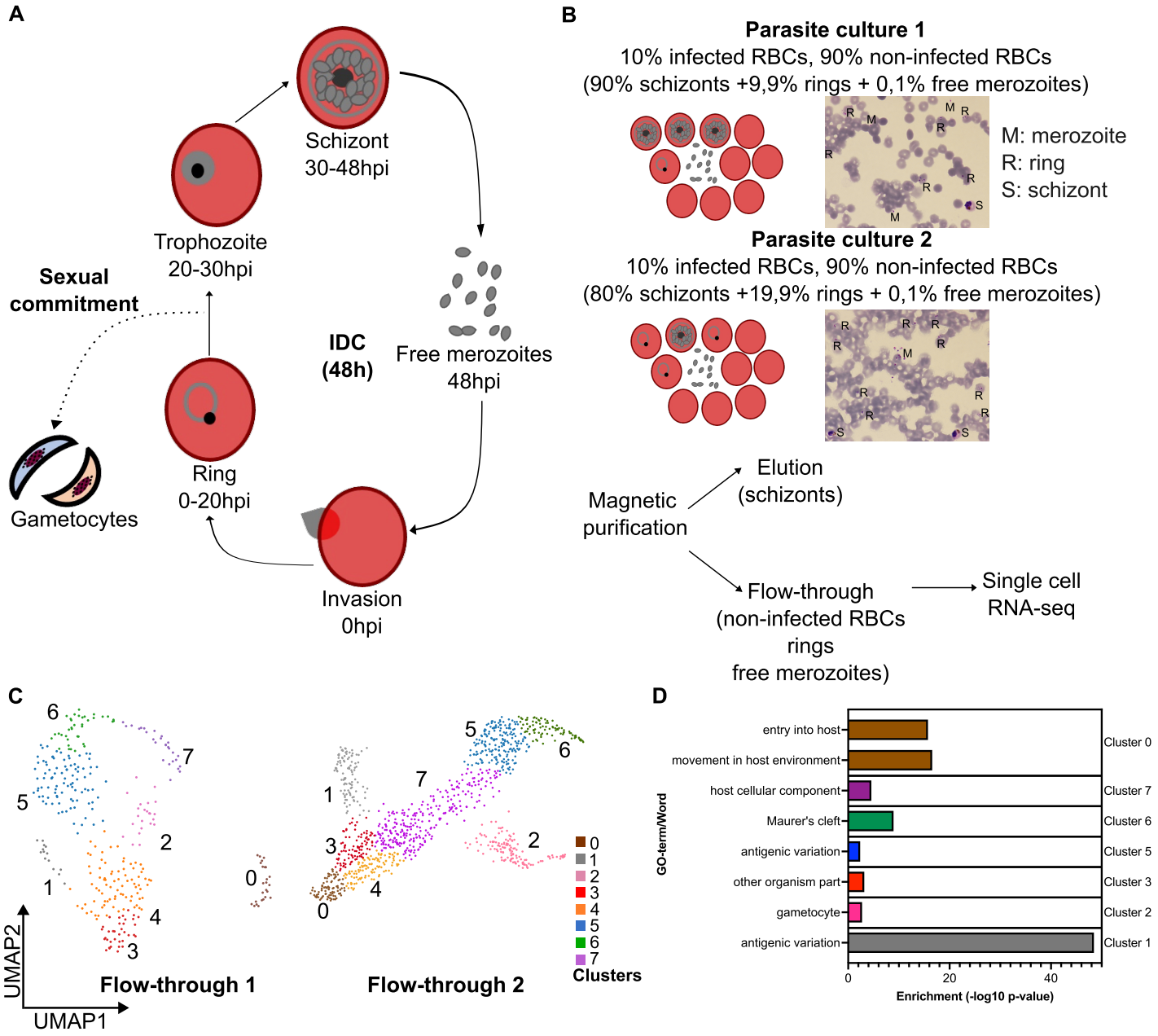

A
Schizont
30-48hpi
Trophozoite
20-30hpi
Sexual
commitment
IDC
(48h)
Free merozoites
48hpi
Ring
0-20hpi
Gametocytes
Invasion
0hpi
B
Parasite culture 1
10% infected RBCs, 90% non-infected RBCs
(90% schizonts +9,9% rings + 0,1% free merozoites)
M: merozoite
R: ring
S: schizont
Parasite culture 2
10% infected RBCs, 90% non-infected RBCs
(80% schizonts +19,9% rings + 0,1% free merozoites)
Magnetic
purification
Elution
(schizonts)
Flow-through
(non-infected RBCs
rings
free merozoites)
Single cell
RNA-seq
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 2

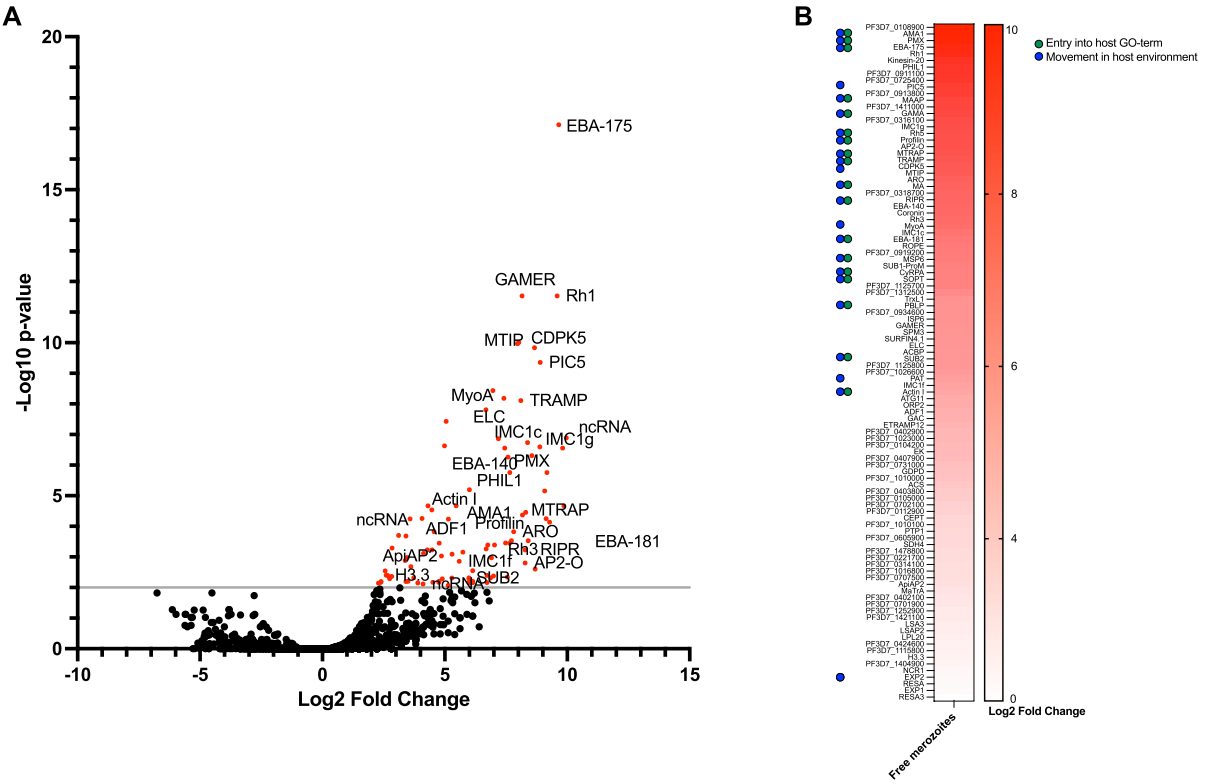

A
-Log10 p-value
Log2 Fold Change
EBA-175
GAMER
Rh1
MTIP
CDPK5
PIC5
MyoA
TRAMP
ELC
IMC1c
ncRNA
IMC1g
EBA-140
PMX
PHIL1
Actin I
AMA1
MTRAP
ncRNA
ADF1
Profilin
ARO
EBA-181
ApiAP2
Rh3
RIPR
IMC1f
AP2-O
H3.3
SUB2
B
Entry into host GO-term
Movement in host environment
Log2 Fold Change
Free merozoites

Figure 3

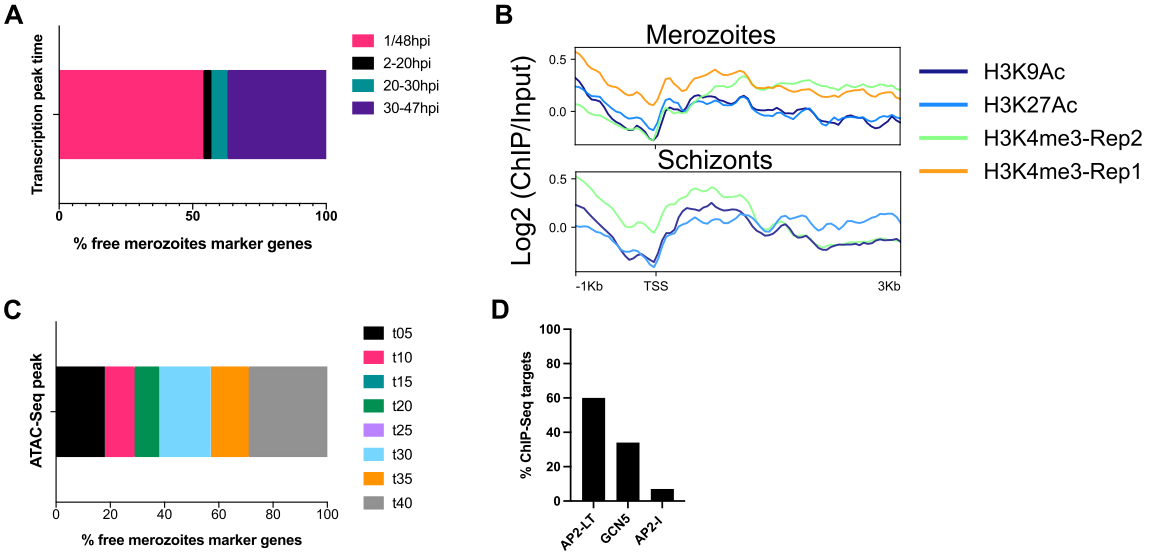

A
Transcription peak time
1/48hpi
2-20hpi
20-30hpi
30-47hpi
0
50
100
% free merozoites marker genes
B
Merozoites
Schizonts
Log2 (ChIP/Input)
0.5
0.0
-1Kb
TSS
3Kb
H3K9Ac
H3K27Ac
H3K4me3-Rep2
H3K4me3-Rep1
C
ATAC-Seq peak
t05
t10
t15
t20
t25
t30
t35
t40
0
20
40
60
80
100
% free merozoites marker genes
D
% ChIP-Seq targets
100
80
60
40
20
0
AP2-LT
GCN5
AP2-I

Figure 4

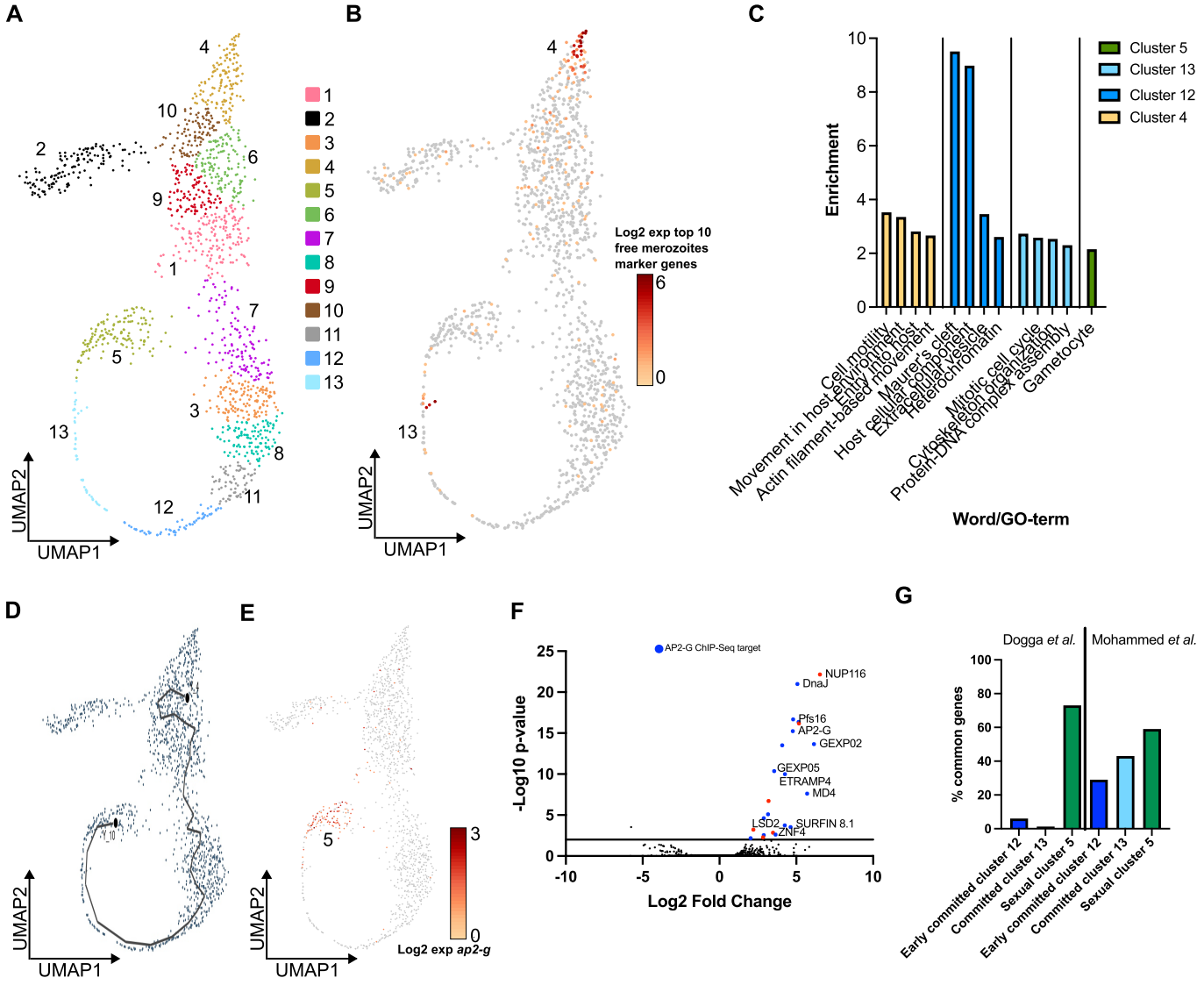

A
B
C
D
E
F
G
UMAP1
UMAP2
Log2 exp top 10 free merozoites marker genes
Log2 exp ap2-g
Enrichment
Word/GO-term
Cluster 5
Cluster 13
Cluster 12
Cluster 4
Cell motility
Movement in host environment
Entry into host
Actin filament-based movement
Host cellular component
Maurer's cleft
Extracellular vesicle
Heterochromatin
Mitotic cell cycle
Cytoskeleton organization
Protein-DNA complex assembly
Gametocyte
AP2-G ChIP-Seq target
NUP116
DnaJ
Pfs16
AP2-G
GEXP02
GEXP05
ETRAMP4
MD4
LSD2
SURFIN 8.1
ZNF4
-Log10 p-value
Log2 Fold Change
Dogga et al.
Mohammed et al.
% common genes
Early committed cluster 12
Committed cluster 13
Sexual cluster 5
Early committed cluster 12
Committed cluster 13
Sexual cluster 5

Figure 5

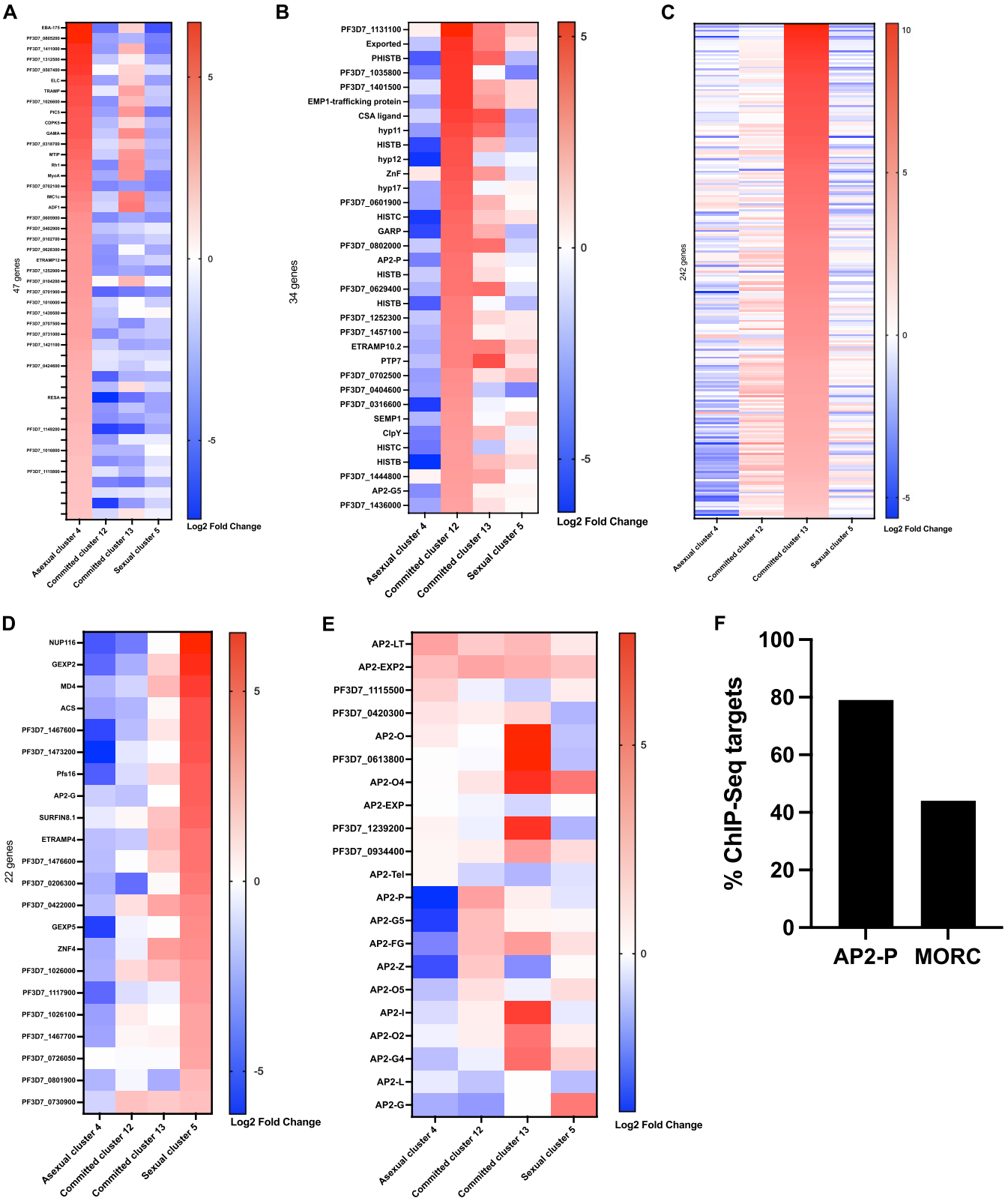

A
47 genes
Asexual cluster 4
Committed cluster 12
Committed cluster 13
Sexual cluster 5
Log2 Fold Change
B
34 genes
PF3D7_1131100
Exported
PHISTB
PF3D7_1035800
PF3D7_1401500
EMP1-trafficking protein
CSA ligand
hyp11
HISTB
hyp12
ZnF
hyp17
PF3D7_0601900
HISTC
GARP
PF3D7_0802000
AP2-P
HISTB
PF3D7_0629400
HISTB
PF3D7_1252300
PF3D7_1457100
ETRAMP10.2
PTP7
PF3D7_0702500
PF3D7_0404600
PF3D7_0316600
SEMP1
ClpY
HISTC
HISTB
PF3D7_1444800
AP2-G5
PF3D7_1436000
Asexual cluster 4
Committed cluster 12
Committed cluster 13
Sexual cluster 5
Log2 Fold Change
C
242 genes
Asexual cluster 4
Committed cluster 12
Committed cluster 13
Sexual cluster 5
Log2 Fold Change
D
22 genes
NUP116
GEXP2
MD4
ACS
PF3D7_1467600
PF3D7_1473200
Pfs16
AP2-G
SURFIN8.1
ETRAMP4
PF3D7_1476600
PF3D7_0206300
PF3D7_0422000
GEXP5
ZNF4
PF3D7_1026000
PF3D7_1117900
PF3D7_1026100
PF3D7_1467700
PF3D7_0726050
PF3D7_0801900
PF3D7_0730900
Asexual cluster 4
Committed cluster 12
Committed cluster 13
Sexual cluster 5
Log2 Fold Change
E
AP2-LT
AP2-EXP2
PF3D7_1115500
PF3D7_0420300
AP2-O
PF3D7_0613800
AP2-O4
AP2-EXP
PF3D7_1239200
PF3D7_0934400
AP2-Tel
AP2-P
AP2-G5
AP2-FG
AP2-Z
AP2-O5
AP2-I
AP2-O2
AP2-G4
AP2-L
AP2-G
Asexual cluster 4
Committed cluster 12
Committed cluster 13
Sexual cluster 5
Log2 Fold Change
F
% ChIP-Seq targets
100
80
60
40
20
0
AP2-P
MORC

Figure 6

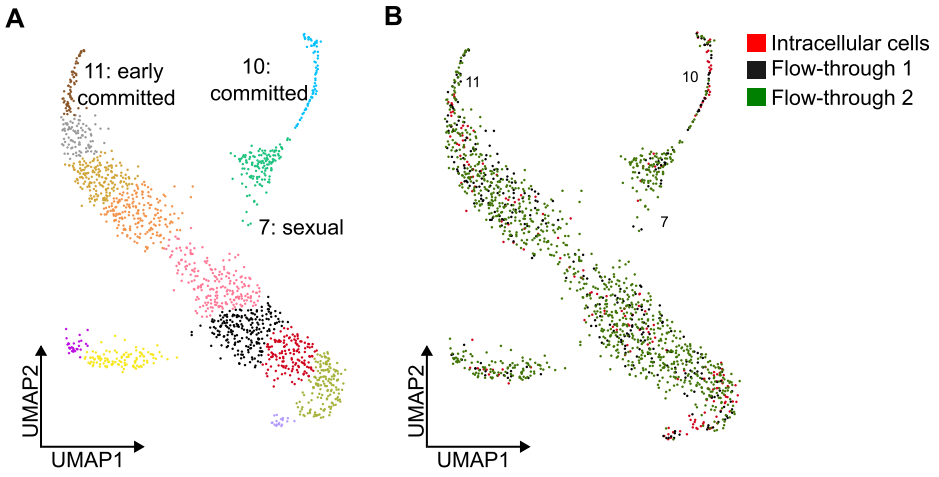

A
11: early
committed
10:
committed
7: sexual
UMAP2
UMAP1
B
11
10
7
UMAP2
UMAP1
Intracellular cells
Flow-through 1
Flow-through 2

Figure S1

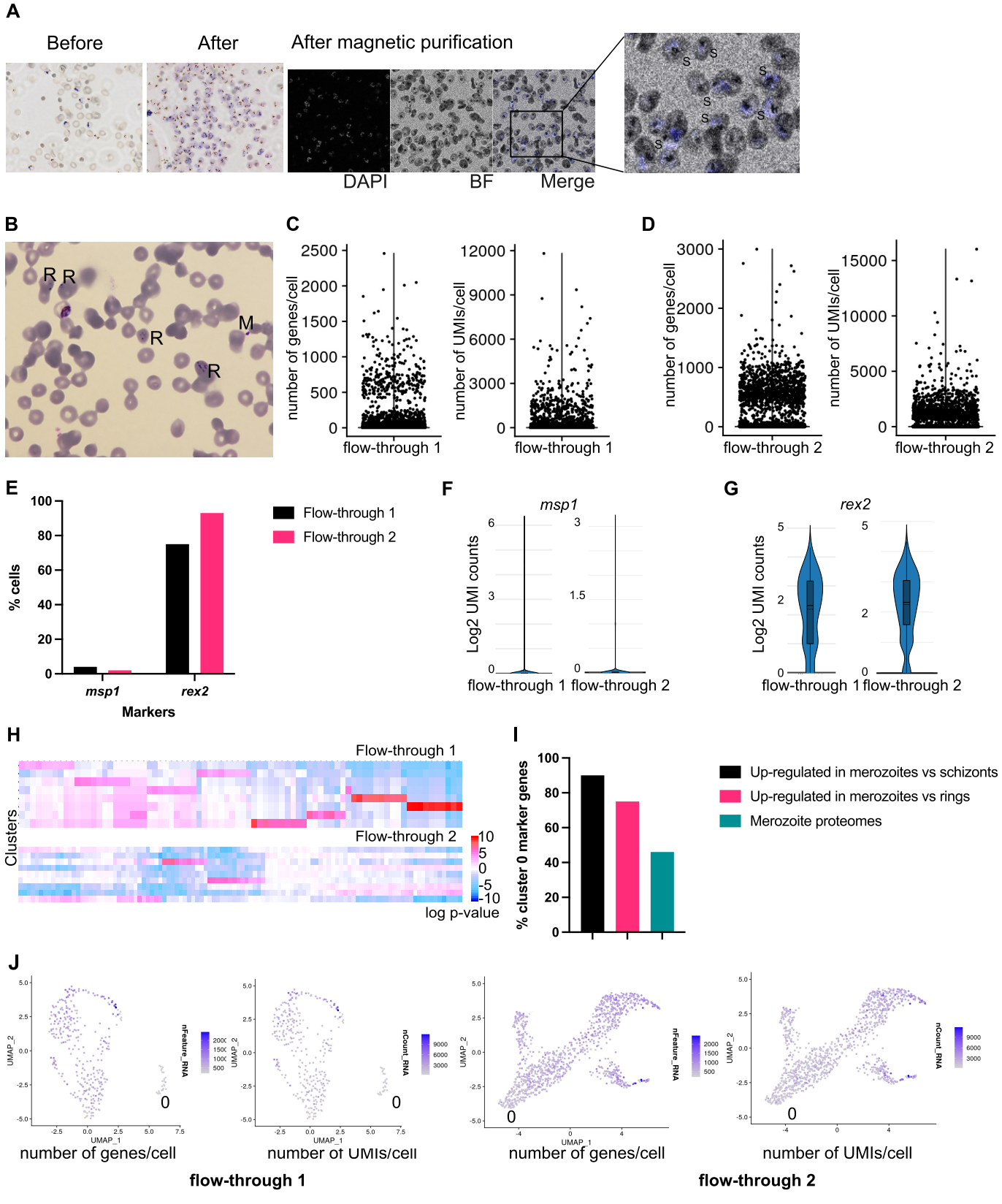

A
Before
After
After magnetic purification
DAPI
BF
Merge
B
R
M
C
number of genes/cell
number of UMIs/cell
flow-through 1
D
flow-through 2
E
% cells
Flow-through 1
Flow-through 2
msp1
rex2
Markers
F
msp1
Log2 UMI counts
G
rex2
H
Clusters
log p-value
I
% cluster 0 marker genes
Up-regulated in merozoites vs schizonts
Up-regulated in merozoites vs rings
Merozoite proteomes
J
UMAP_1
UMAP_2
nFeature_RNA
nCount_RNA
number of genes/cell
number of UMIs/cell
flow-through 1
flow-through 2

Figure S2

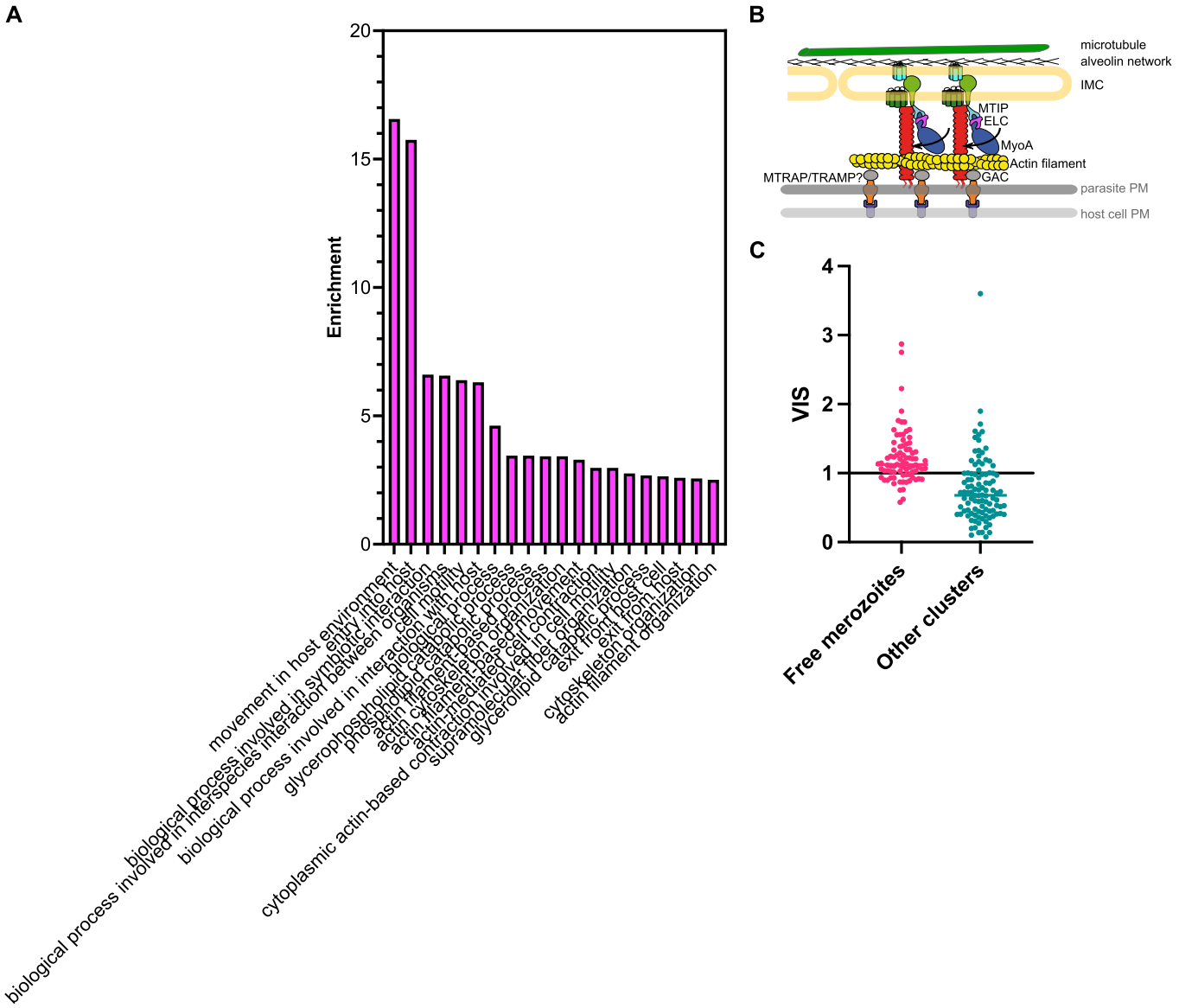
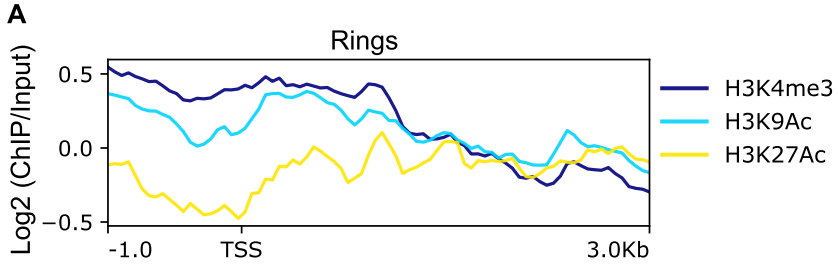
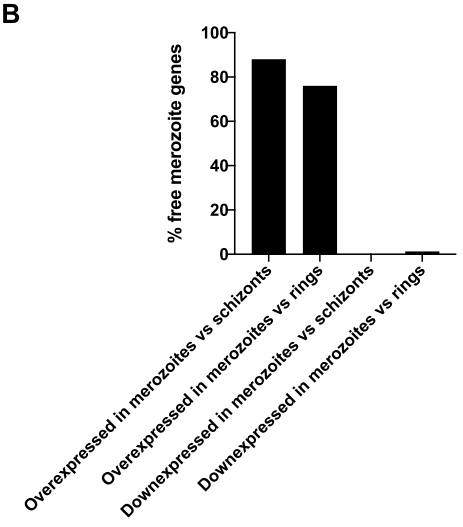

A
Enrichment
20
15
10
5
0
movement in host environment
entry into host
biological process involved in symbiotic interaction
biological process involved in interspecies interaction between organisms
cell motility
biological process involved in interaction with host
phospholipid catabolic process
glycerophospholipid catabolic process
actin filament-based process
actin cytoskeleton organization
actin filament-based movement
actin-mediated cell contraction
cytoplasmic actin-based contraction involved in cell motility
supramolecular fiber organization
glycerolipid catabolic process
exit from host cell
cytoskeleton organization
exit from host
actin filament organization
B
microtubule
alveolin network
IMC
MTIP
ELC
MyoA
Actin filament
MTRAP/TRAMP?
GAC
parasite PM
host cell PM
C
VIS
4
3
2
1
0
Free merozoites
Other clusters

Figure S3


A
Rings
Log2 (ChIP/Input)
0.5
0.0
−0.5
-1.0
TSS
3.0Kb
H3K4me3
H3K9Ac
H3K27Ac


B
% free merozoite genes
100
80
60
40
20
0
Overexpressed in merozoites vs schizonts
Overexpressed in merozoites vs rings
Downexpressed in merozoites vs schizonts
Downexpressed in merozoites vs rings

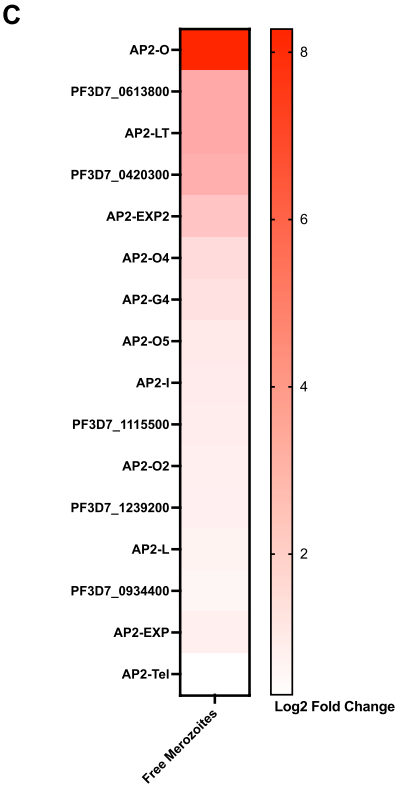

C
AP2-O
PF3D7_0613800
AP2-LT
PF3D7_0420300
AP2-EXP2
AP2-O4
AP2-G4
AP2-O5
AP2-I
PF3D7_1115500
AP2-O2
PF3D7_1239200
AP2-L
PF3D7_0934400
AP2-EXP
AP2-Tel
Free Merozoites
8
6
4
2
Log2 Fold Change

Figure S4

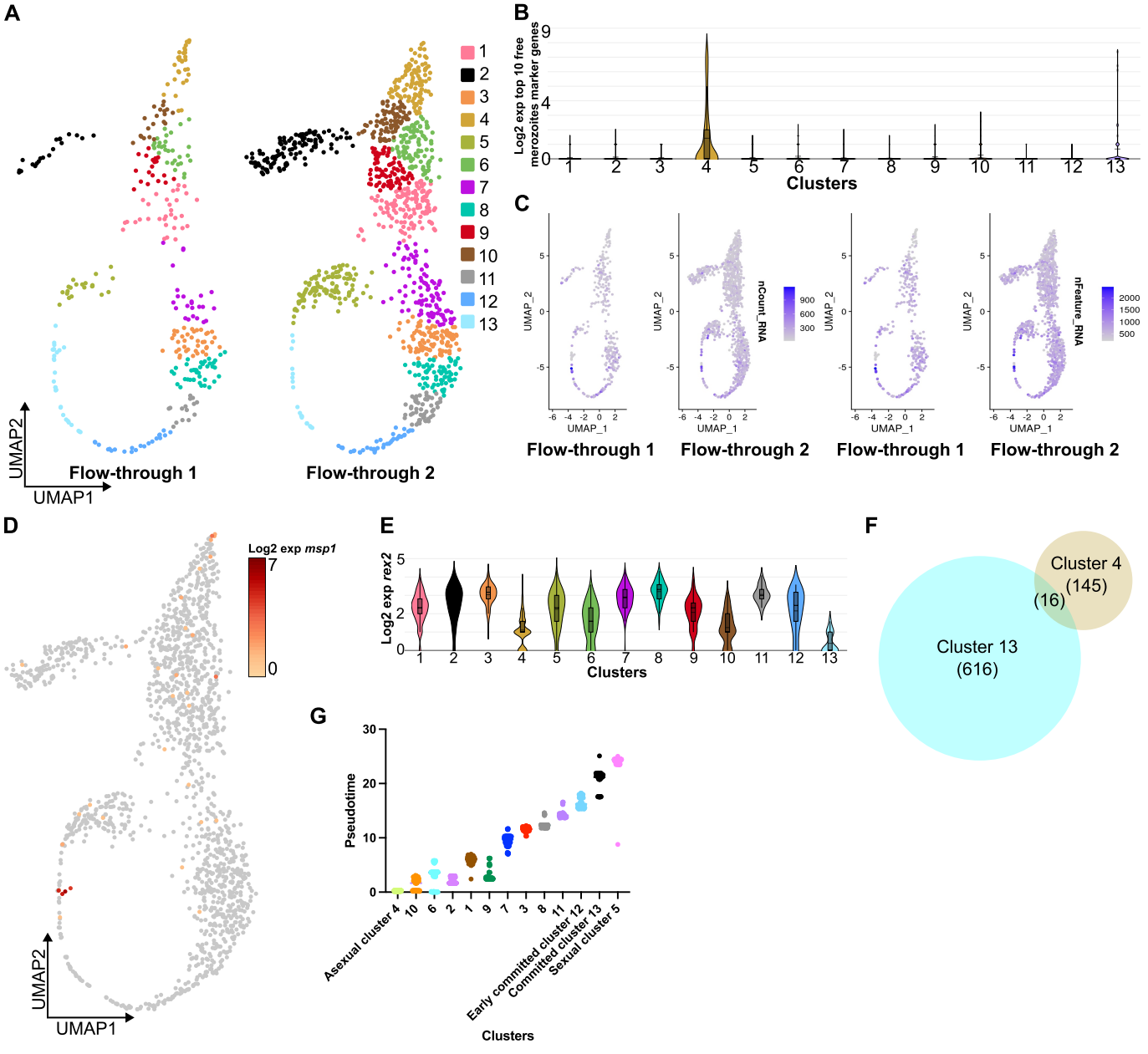

A
1
2
3
4
5
6
7
8
9
10
11
12
13
UMAP2
UMAP1
Flow-through 1
Flow-through 2
B
Log2 exp top 10 free merozoites marker genes
Clusters
C
UMAP_2
UMAP_1
nCount_RNA
nFeature_RNA
Flow-through 1
Flow-through 2
Flow-through 1
Flow-through 2
D
Log2 exp msp1
UMAP2
UMAP1
E
Log2 exp rex2
Clusters
F
Cluster 4
(145)
(16)
Cluster 13
(616)
G
Pseudotime
Asexual cluster 4
Early committed cluster 12
Committed cluster 13
Sexual cluster 5
Clusters

Figure S5

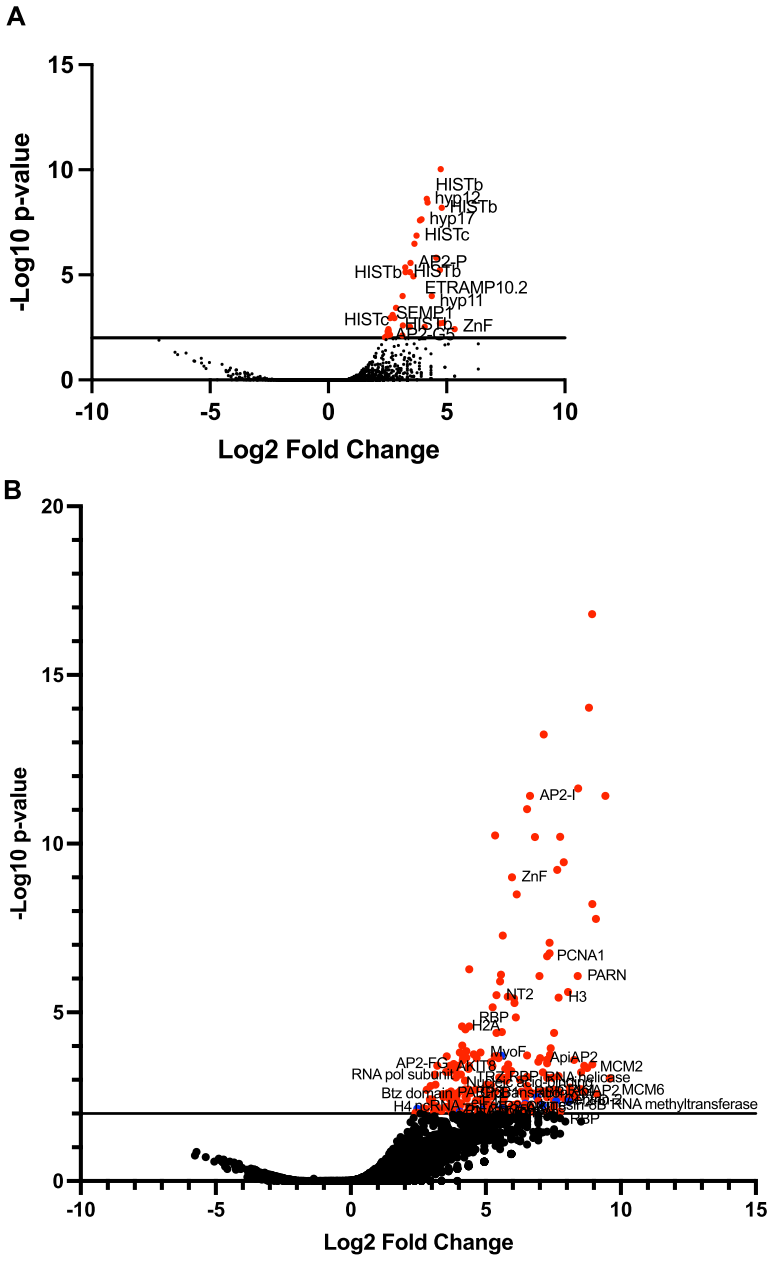
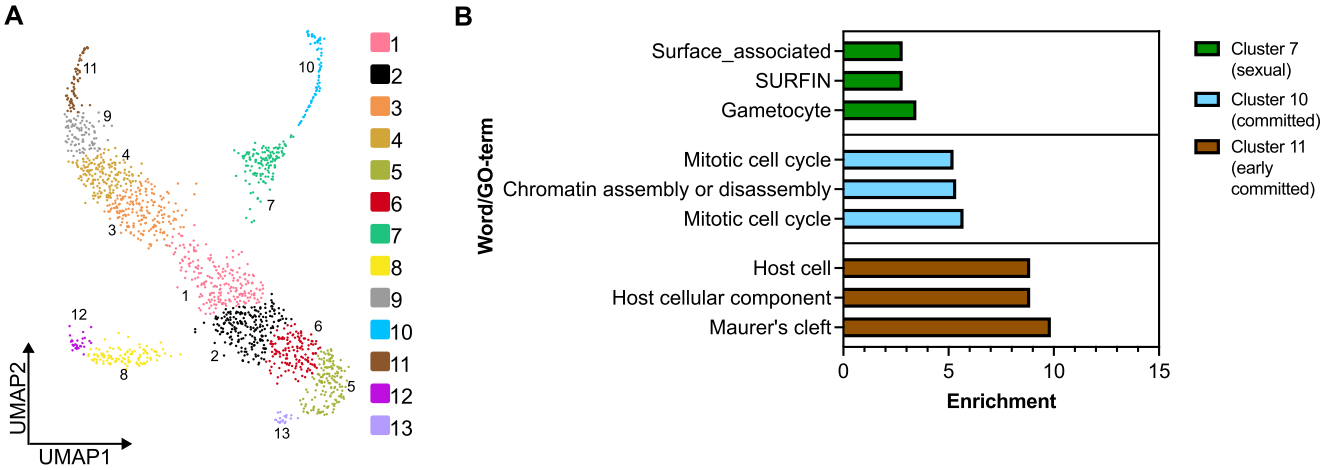

A
-Log10 p-value
Log2 Fold Change
HISTb
hyp12
HISTb
hyp17
HISTc
AP2-P
HISTb
HISTb
ETRAMP10.2
hyp11
SEMP1
HISTc
HISTb
ZnF
AP2-G5
B
-Log10 p-value
Log2 Fold Change
AP2-I
ZnF
PCNA1
PARN
NT2
H3
RBP
H2A
MyoF
ApiAP2
MCM2
AP2-FG
RNA pol subunit
RNA helicase
Btz domain
MCM6
RNA methyltransferase
RBP

Figure S6


A
UMAP2
UMAP1
1
2
3
4
5
6
7
8
9
10
11
12
13
B
Word/GO-term
Surface_associated
SURFIN
Gametocyte
Mitotic cell cycle
Chromatin assembly or disassembly
Mitotic cell cycle
Host cell
Host cellular component
Maurer's cleft
0
5
10
15
Enrichment
Cluster 7 (sexual)
Cluster 10 (committed)
Cluster 11 (early committed)