Single cell RNA-sequencing of malaria merozoites reveals that the parasite prepares for life within the red blood cell by activating a gene network

Short title: Single cell RNA-seq of malaria merozoites

Laura Piel[1#], Celine Hernandez[1], Yan Jaszczyszyn[1], Delphine Naquin[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 within the human host red blood cell is the symptomatic stage of the life cycle. The IDC 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 did the first transcriptional profiling of merozoites throughout development inside the schizont and after egress. Merozoites within the schizont, activate a series of tightly regulated gene expression programs, each encoding for proteins implicated in a specific process required for invasion and life within the host. After schizont egress, free merozoites still express some of these programs. Upon invasion, ring cells have a distinct transcriptional profile. The gene expression programs are regulated by specific transcription factors and chromatin accessibility as well as post-transcriptional mechanisms. Our work gains unprecedent insight into gene expression prior to and during host cell invasion and identifies potential new antimalarial targets.

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, prior to host cell invasion. We found that, in order to prepare for invasion and life within the host cell, merozoites activate a series of gene expression programs, each encoding for proteins implicated in a specific process. While some of the genes expressed by merozoites during maturation are still expressed in free merozoites, these transcripts are no longer detected after host cell invasion. This implicates tight regulation of gene expression. We suggest that regulation implicates binding of specific transcription factors, epigenetics as well as post-transcriptional mechanisms.

47 **Introduction**

48

49 Malaria, responsible for over half a million deaths each year (1), is an infectious
50 disease caused by the unicellular eukaryotic parasite *Plasmodium*. Out of the five species of
51 parasite that cause disease in humans, *P. falciparum*, is the most virulent, causing the majority
52 of the yearly over half a million deaths (WHO). The parasite life cycle is complex, alternating
53 between the human host and the mosquito, that acts as both a host and a vector. The
54 intraerythrocytic developmental cycle (IDC) (Fig 1A), in which the parasite multiplies within the
55 human host red blood cell (RBC) for 48-hours, starts when parasite stages called merozoites,
56 are released into the bloodstream, following schizont rupture, and invade the host RBC. Once
57 inside the host cell, the parasite transforms into the ring stage. Some parasites exit the IDC
58 and differentiate into the sexually committed cells, precursor of the male and female
59 gametocytes, the transmission stages, that are uptake upon a mosquito bite.

60 The merozoite is the less studied stage of the IDC, despite its attractiveness as a drug
61 target given that it is extracellular, even if briefly (2), it is responsible for RBC invasion, which
62 is indispensable for parasite survival, and it initiates the IDC, which is the symptomatic life
63 cycle stage of malaria.

64 Studies describing the transcriptional profiling of the different stages of the *P.*
65 *falciparum* life cycle, lack information for the merozoite stage (3) and, until recently, it was
66 thought that merozoites were transcriptionally silent (4). Here we establish for the first time the
67 transcriptional signature of merozoites. We find that, during maturation inside the schizont,
68 merozoites activate a succession of tightly regulated expression programs, allowing the
69 parasite to later on express the proteins it needs to invade and survive within the RBC. After
70 schizont exit, free merozoites still express a few of the genes activated during schizogony but,
71 once the parasite invades the RBC and transforms into rings, the merozoites programs shut
72 down and new programs turn on. Each of the merozoite expression programs encodes for
73 proteins implicated in a specific process and we provide evidence that they are regulated by

74 specific transcription factors and epigenetic factors controlling chromatin accessibility as well
75 as post-transcriptional mechanisms.
76 Our comprehensive study of parasite gene expression before and after RBC invasion
77 further sheds light into gene expression in the malaria parasite and unfolds for the first time
78 the expression program allowing the parasite to invade the RBC and start the symptomatic
79 stage of the lifecycle.
80
81 **Results**
82
83 **Merozoites express different sets of genes throughout development inside the schizont**
84 In order to identify the transcription profile of merozoites during schizogony and after
85 schizont rupture, we adapted a protocol described before for the isolation of viable invasive
86 merozoites (5) (Fig 1B; S1A Fig). We synchronized parasites, and when schizonts were visible
87 in Giemsa-stained smears, we collected the culture. The culture also contained non-infected
88 RBCs, ring-stage parasites and free merozoites. We then passed the culture through a
89 magnetic column to retain the schizonts, which are magnetic due to the accumulation of
90 hemozoin, the end product of hemoglobin digestion. In the flow-through, we recovered the
91 non-infected RBCs, the ring stage parasites and the free merozoites. After elution from the
92 column, the schizonts were passed through a 1.2μM filter to release the merozoites inside.
93 The isolated merozoite and the flow-through cells were then prepared individually for single
94 cell RNA-sequencing. We obtained a high sequencing saturation and mapability to the
95 parasite genome and detected a similar number of genes in both samples (over 4800, ~90%
96 of the parasite genome) (S1 Table). Confirming that the purified merozoite cells are merozoite-
97 stage parasites, we detected expression of the merozoite marker *merozoite surface marker 1*
98 (*msp1*, PF3D7_0930300) but not of the ring marker *ring-exported protein 2* (*rex2*,
99 PF3D7_0936000) (S1B-C Figs). On the other hand, the majority of flow-through cells
100 expresses *rex2* (S1B-D Figs) and are therefore ring-stage parasites. Furthermore, we

101 detected less transcripts per cell in merozoite samples versus flow-through samples (S1E-F
102 Figs), in accordance with previous findings (3,6).
103 To analyze the heterogeneity of the merozoite expression programs we used Seurat,
104 a toolbox for quality check, analysis and visualization of single cell data (7–9). We identified
105 ten expression clusters (Fig 1C). Each cluster was characterized by expression of, on average,
106 49 distinct differentially expressed (DE) genes (S1G Fig; S2 Table). Clustering was
107 independent of the number of genes detected per cell or the cells RNA content (S1H Fig).
108 When we did gene ontology (GO) analysis, we found that each cluster expresses a specific
109 category of genes (Fig 1C; S2 Table). While clusters 5, 6 and 7 express genes related to host
110 cell invasion, in clusters 1, 8, 9 and 10 we detected genes related to nuclear processes, RNA
111 binding and protein synthesis. Clusters 2, 3 and 4 express genes encoding for proteins stored
112 in the dense granules that are later discharged and participate in events taking place post-
113 invasion - formation of the parasitophorous vacuole (PV), wherein the parasite resides after
114 invasion, formation of the *Plasmodium* translocon of exported proteins (PTEX) (10), in the PV
115 membrane, through which parasite proteins are exported into the host cell, and host cell
116 remodeling, allowing the parasite to import media nutrients.
117 Because some clusters are enriched for the same GO-terms, we compared them but
118 we found none to little overlap, except for clusters 5 and 6, related to invasion (Fig 1D; S1I
119 Fig). However, the invasion clusters 5 and 6 are different from the other invasion cluster,
120 cluster 7 (Fig 1D), and DE genes of a specific invasion cluster are overexpressed in that cluster
121 versus the other two invasion clusters (Fig 1E). Each merozoite cluster corresponds then to a
122 specific expression program.
123
124 **Merozoites sequentially activate distinct gene expression programs during schizogony**
125 Pseudotemporal analysis indicates that the merozoite cells, during schizogony, follow
126 one developmental trajectory marked by the sequential activation of each of the expression
127 clusters (Fig 2A; S2A Fig; S1 Table). The DE genes of a specific cluster are still expressed in
128 the next cluster but then few to none of those genes are detected in the afterwards cluster (Fig

129 2B), showing that expression of the genes of a given cluster peaks at a precise time. To
130 determine if the expression clusters correspond to transcriptional programs, we determined
131 how many of the DE genes of a cluster are transcribed by comparing our dataset to a
132 previously published list of genes actively transcribed throughout the IDC (6). Clusters 2, 3, 4,
133 6 and 7 likely correspond to transcriptional programs because over 50% of their DE genes are
134 actively transcribed during schizogony (Fig 2C).

135 Transcriptional programs are regulated by the binding of transcription factors to gene
136 promoters. The ApiAP2 DNA-binding proteins are the main family of transcription factors in
137 the parasite (11). The ApiAP2 proteins AP2-I, AP2-LT and AP2-EXP and the bromodomain
138 protein BDP1 have been linked to expression of invasion genes (12–15) so we determined if
139 the genes of the invasion transcriptional programs, clusters 6 and 7, are regulated by these
140 proteins. AP2-I and associated proteins, regulate the cluster 7 program because 88% of the
141 gene promoters in this cluster are bound by AP2-I, BDP1, AP2-LT or a combination of the
142 three proteins (Figs 2D-E). That is not the case for the cluster 6 genes (Fig 2D) so to identify
143 the DNA-binding proteins regulating transcription of this and the other three transcriptional
144 programs, we searched for enrichment of DNA sequence motifs in the promoters of the DE
145 genes. We identified two DNA motifs, one enriched in the promoters of clusters 3 and 6 genes,
146 that we named merozoite DNA motif 1, and one enriched upstream of clusters 2 and 4 genes,
147 here named merozoite DNA motif 2 (Fig 2F). A *cis*-regulatory element very similar to
148 merozoite DNA motif 1 was previously identified *in silico* (16) indicating that DNA motif 1 is
149 frequently found in gene promoters. The parasite repressor ApiAP2 protein AP2-G2 (17,18)
150 binds to a DNA motif similar to merozoite DNA motif 1 (S2B Fig) so we examined if the gene
151 promoters of clusters 3 and 6, containing the merozoite DNA motif 1, are AP2-G2 ChIP-Seq
152 targets and found that about half are (S2C Fig), suggesting that AP2-G2 may repress
153 transcription of the DE genes in these clusters. Comparison of merozoite DNA motif 2 with the
154 DNA motifs bound by the ApiAP2 proteins (11) indicate that the uncharacterized ApiAP2
155 protein PF3D7_1239200 may bind to it (S2D Fig).

156 Clusters containing genes enriched for the same merozoite DNA motif are expressed
157 at different pseudotimes. We wondered if this reflects different times of chromatin opening at
158 the promoter regions allowing protein binding earlier or later post-invasion. We thus
159 determined when the promoters of each of the clusters was euchromatic by comparing our
160 gene list to a published ATAC-Seq data (19). The order of activation of the transcription
161 clusters 2, 3, 4, 6 and 7 is linked to the timing of chromatin opening because a larger proportion
162 of gene promoters of later expressed clusters is euchromatic at 40hpi (Fig 2G).

163 Our analysis indicates that clusters 1, 5, 8, 9 and 10 are not actively transcribed (Fig
164 2C) so we wondered if the transcripts detected in these clusters are stabilized. To do that, we
165 determined how many of these clusters genes have been previously labelled as stabilized (6).
166 The majority of the transcripts detected in cluster 5, but not clusters 1, 8, 9 and 10, are
167 stabilized (Fig S2E). However, there is no correlation between the order in which clusters 1,
168 8, 9 and 10 are expressed and the timing of chromatin opening (Fig 2G). Given that the genes
169 expressed in cluster 5 are stabilized and the genes expressed in clusters 1, 8, 9 and 10 are
170 not expressed when most promoters are euchromatic, their regulation is likely post-
171 transcriptional and it is mediated by RNA-binding proteins. Sequence enrichment analysis
172 identified distinct motifs in the 3' untranslated regions (3'utr) of the clusters 1, 5, 8, 9 and 10
173 DE genes (S4 Table). The RNA-binding protein PF3D7_0823200 binds similar motifs to those
174 identified in the 3'utr of clusters 5, 8 and 9 genes (20,21) (S2F Fig) but few genes in these
175 clusters are bound by the protein (S2G Fig). The motifs found in the 3'utr of clusters 1 and 10
176 genes resemble those bound by two putative RNA-binding proteins that remain unstudied
177 (S2F Fig).

178 The invasion clusters 5 and 6 have virtually identical chromatin opening patterns (Fig
179 2G) and, as for cluster 6, the merozoite DNA motif 1 is present in the promoters of cluster 5
180 genes (S2H Fig) and AP2-G2 binds to the gene promoters (S2C Fig), suggesting that
181 regulation of the invasion clusters 5 and 6 involves the same proteins.

182 To establish if there is variation in expression of the merozoite programs in merozoites
183 within a same schizont or all merozoites inside a schizont express the same genes, we

184 quantified which of the clusters have genes with a high variability index score (VIS), a metric
185 established by Tripathi *et al*. to identify genes with high merozoite-to-merozoite variation in
186 mature schizonts (22). Over half of the merozoite clusters 3 and 4 genes, encoding for proteins
187 implicated in host cell remodeling, and 80% and 69% of the genes in the invasion clusters 5
188 and 6, respectively, have a VIS>1 (S2I Fig; S2 Table), suggesting that expression of these
189 particular expression programs is stochastic.
190
191 **Parasites activate different transcriptional programs post-invasion**
192 We performed single cell RNA-Seq on flow-through cells to identify the transcriptional
193 programs expressed by the parasite after schizont rupture. We identified 8 expression
194 clusters, each characterized by the expression of a specific set of genes (Fig 3A; S3A-B Figs;
195 S3 Table). GO-term analysis of the marker genes of each cluster (S3 Table) reveals that three
196 clusters (3, 6 and 7) express genes coding proteins implicated in the formation of the Maurer's
197 cleft, parasite structures from which the exported parasite proteins are sorted and transported
198 to the host cell (Fig 3A) but we did not find any overlapping genes (S3 Table), suggesting that
199 each cluster is a separate expression program. Cluster 1 expresses the *var* genes encoding
200 for Erythrocyte Surface Protein 1, implicated in antigenic variation (Fig 3A), and cluster 2
201 genes code for proteins implicated in sexual commitment (Fig 3A). Apart from cluster 0, which
202 is specialized in host interaction and motility (Fig 3A) and will be discussed later, we believe
203 that these clusters correspond to expression programs expressed by ring-stage parasites
204 because these are the predominant stages in the flow-through, as previously discussed, and
205 they match previous findings (3).
206 Most parasites in the parasite culture collected for magnetic purification were schizont
207 stages so there are few cells in each flow-through cluster (S1 Table). To ensure that this did
208 not bias our results, we sequenced and analyzed flow-through cells coming from a second
209 magnetic purification, containing more ring-stage parasites, and identified the same
210 expression programs (S3C-D Figs; S3 Table).

211 Pseudotemporal analysis suggest that, after RBC invasion, early ring stage parasites
212 are already committed to pursue the IDC by expressing cluster 1 or express instead cluster 2
213 and commit to sexual differentiation (Fig 3B; S3E Fig).
214 Unlike the merozoite clusters, all the ring expression clusters seem to correspond to
215 transcriptional programs because the majority of all clusters DE genes, except for cluster 7, is
216 actively transcribed during the IDC (Fig 3C). Regulation of the ring expression clusters is thus
217 likely mediated by the binding of specific transcription factors to the gene promoters. The time
218 at which the chromatin at the promoter regions of the genes is accessible correlates somehow
219 with the progressive activation of the clusters (S3F Fig). As expected, the sexual commitment
220 transcriptional program corresponding to cluster 2 is regulated by AP2-G, the parasite
221 gametocytogenesis master regulator (23) (Fig 3D). Motif enrichment analysis shows that there
222 are different DNA motifs in the promoters of the genes expressed in the other clusters and
223 indicate, for instance, that the ApiAP2 protein PF3D7_0613800 may regulate the cluster 6
224 genes (S3G Fig; S4 Table). PF3D7_0613800 remains uncharacterized but it is essential for
225 parasite growth (24).
226 We examined if there was any correlation between the transcriptional programs
227 expressed by merozoites and ring stage cells. Merozoite and ring stage cells both express
228 genes encoding for nuclear proteins or proteins implicated in host cell remodeling but the
229 merozoites and ring expression programs are distinct because there is no overlap between
230 the genes expressed by the two types of cells (S2 and S3 Tables). This implies that the
231 merozoite expression programs cease after RBC invasion.
232
233 **Free merozoites express host cell invasion and remodeling genes**
234 We noticed that the cluster 0 of flow-through cells is excluded from the pseudotemporal
235 trajectory (Fig 3B). Furthermore, cluster 0 cells contain a lower amount of RNA (S3B Fig) and
236 the genes expressed in cluster 0 are transcribed at the end of schizogony (1/48hpi) (Fig 4A).
237 This led us to ponder if these cells are free merozoites that egressed from schizonts before
238 we collected the parasite culture and did the magnetic purification. To test this hypothesis, we

239 combined the data from the merozoite and ring cells and filtered the cells, using the same
240 parameters we used for the ring cells, to only select for late stage merozoites. Then we
241 identified the genes that were over-expressed or down-expressed in the merozoites versus
242 the ring cells (Fig 4B). Genes over-expressed in merozoites represent the late stage
243 merozoites transcriptional signature. We found 25 genes over-expressed in merozoites (S5
244 Table), coding for proteins implicated in host cell invasion and host cell remodeling (S4A Fig).
245 Confirming that cluster 0 cells are indeed free merozoites, over 60% of the genes expressed
246 in cluster 0 are part of the late stage merozoites transcriptional signature (Fig 4C). Cluster 0
247 then represents the transcriptional profile of free merozoites, an elusive stage that had never
248 before been captured by transcriptional studies.

249 We determined which of the genes expressed by merozoites during schizogony are
250 still detected in free merozoites. Marker genes of merozoite clusters 2, 3, 4, 5, 6 and 7
251 encoding for genes implicated in host cell remodeling, vacuole formation, motility and invasion
252 are detected in both late stage merozoites and cluster 0 (Fig 4D), indicating that these genes
253 are still expressed in free merozoites. Supporting this, the merozoite DNA motif 1, which we
254 previously showed is present in the promoters of merozoite clusters 3, 5 and 6, is also present
255 in the promoters of genes expressed in free merozoites (S4B Fig). Whereas more than 60%
256 of the invasion merozoite clusters 5 and 6 are still expressed in free merozoites, only 5% of
257 the invasion merozoite cluster 7 genes are (Fig 4D). This, together with the fact that the genes
258 expressed in free merozoites are not regulated by AP2-I and associated proteins (S4C Fig),
259 indicate that only the non-AP2-I merozoite invasion transcriptional program remains active in
260 free merozoites.

261

262 **Discussion**

263 Merozoites are responsible for RBC invasion and initiate the symptomatic phase of the
264 malaria life cycle but they are the less transcriptionally understood stage of the IDC. Our data
265 shows that there are rapid changes in gene expression during formation and segmentation of

266 merozoites (Fig 4E). Merozoites express genes encoding nuclear proteins and proteins
267 destined to the dense granules as well as gene products implicated in host cell interaction, PV
268 formation, invasion and motility allowing infection of new RBCs. There is also a merozoite
269 program dedicated to expression of genes encoding for proteins implicated in RNA-binding
270 and translation. Free merozoites still express the PV and host cell remodeling programs as
271 well as genes implicated in motility and host interaction. All of the merozoite expression
272 programs shut down once the parasite invades the host cell and transforms into the ring-stage.

273 Around 40% of the parasite genome remains unannotated (25) because there are
274 either no recognizable functional domains or it is unknown if the annotated domains are
275 functional. The merozoite program with the higher percentage of unknown genes is cluster 10,
276 specialized in ribosome assembly (S4D Fig). Our study can be used to discover new proteins
277 implicated in this process as well as in vacuole formation, host cell remodeling and others.
278 Moreover, the genes expressed by merozoites do not have, for the most part, human
279 orthologues (S4E Fig), making them interesting from a therapeutic point of view.

280 The gene expression cascade controlling merozoite formation and maturation we
281 identified here implies that gene expression at this stage of the parasite life cycle is tightly
282 regulated. This resembles the developmental gene networks of other organisms (26) where a
283 cascade of transcription factors activates expression of proteins that repress translation of
284 other proteins. We propose that regulation of gene expression during merozoite maturation
285 inside schizonts is transcriptional, involving binding of proteins to specific DNA motifs in the
286 gene promoters, once the chromatin is accessible, and post-transcriptional. On the other hand,
287 regulation of gene expression in ring-stage parasites is mainly transcriptional. This is in
288 accordance with previous studies that found that, at the final stages of schizogony, there is
289 increased expression of RNA-binding proteins and gene expression is mainly the
290 consequence of RNA stabilization, while expression during ring-stage parasites is linked to
291 active transcription (6,27,28).

292 We could show that the genes expressed in the merozoite invasion cluster 7 are
293 regulated by AP2-I, BDP1 and AP2-LT while the genes expressed in the merozoite invasion
294 cluster 6 are not bound by these proteins. The genes of cluster 6 contain, instead, in their
295 promoter region, a DNA motif here called merozoite DNA motif 1. The merozoite DNA motif 1
296 is also present in the promoters of the invasion cluster 5, suggesting that there are two invasion
297 transcriptional programs, one corresponding to cluster 7, regulated by AP2-I and associated
298 proteins, and another one, corresponding to clusters 5 and 6, regulated by other, yet to be
299 determined regulators. Previous to our work it was known that AP2-I only bound a subset of
300 invasion-related genes (12) but it was believed that all invasion-related genes were activated
301 at the same time. Here we show that the AP2-I invasion transcriptional program shuts down
302 as the merozoites egress schizonts while the non-AP2-I transcriptional program remains
303 active. Thus, unlike what was previously believed, expression of some invasion genes
304 continues after schizont rupture (6,29–31). These are the genes coding for proteins implicated
305 in the later steps of invasion, such as host receptor interaction, motility and penetration (12),
306 rather than RBC recognition and attachment. Another difference we found between the two
307 invasion transcriptional programs is that the only the non AP2-I program seems to be
308 stochastic (VIS>1), which is supported by the fact that the chromatin opening pattern of the
309 genes in this program is remarkably uniform compared to that of clusters 5 and 6.
310 Other genes contain in their promoter the merozoite DNA motif 2, which may be bound
311 by the ApiAP2 protein PF3D7_1239200. This protein has recently been detected as part of
312 the parasite MORC complex (32) but it remains uncharacterized. We detected expression of
313 other ApiAP2 proteins as well as RNA-binding proteins throughout merozoite development
314 (S1 Table) but further studies are needed to determine which of these are implicated in
315 regulation of the merozoite expression programs.

316 The gene network we uncovered can serve as a reference to understand gene
317 expression regulation mechanisms implicated in the establishment of malaria infection and to
318 better understand the complex parasite life cycle.

319 **Materials and Methods**

320 **Parasite culturing and merozoites purification**

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

325 Parasite synchronization to the ring stage was done two consecutive cycles by incubating
326 pelleted cultures 1:1 with 0.3M L-Alanine, 10mM HEPES pH 7.5 solution for 10 min at
327 37°C(33), after which cultures were centrifuged, resuspended in fresh media and placed back
328 in culture.

329 A mixed culture of non-infected RBCs, schizonts, rings and free merozoites was passed
330 through LD columns (Miltenyi Biotec). The flow-through, corresponding to ring stage parasites,
331 free merozoites and non-infected red blood cells, was recovered. The schizonts were eluted
332 from the column with RPMI 1640 (Fisher) and then passed through a 1.2µm filter attached to
333 a 10ml syringe to release the merozoites inside as described in(5). This corresponds to the
334 merozoite fraction.

335 The purity of each fraction was verified by Giemsa-stained blood smear.

336 Parasites were counted with a Neubauer chamber and diluted with 1XPBS/0.04%BSA to 1
337 million cells/ml.

338 **Single cell RNA-Sequencing**

339 Approximately 10.000 cells (merozoites or flow-through) were loaded on a Chromium Chip G
340 (10x Genomics) as per the manufacturer's recommendations. Chips containing cell
341 suspensions, gel beads in emulsion (GEMs) and reverse transcription reagents were loaded
342 into the chromium controller (10x Genomics) for single cell partitioning and cell barcoding.
343 Barcoded cDNA and sequencing libraries were generated according to the Chromium Next
344 GEM Single Cell 3' gene expression v3.1 protocol. Quantification and the quality of

345 sequencing libraries was verified on an Agilent Bioanalyzer. Pooled libraries were sequenced
346 on a NextSeq 2000 system (Illumina).

347 **Data analysis**

348 Quality control checks on the raw sequencing data were performed using FASTQC (Andrews,
349 2010). FastQ files were analyzed using the CellRanger software (10X Genomics), including
350 alignment, filtering and quantitation of reads on the *Plasmodium falciparum* genome (3D7)
351 and generation of feature-barcode count matrices. During mapping, the reads corresponding
352 to the human host cell were discarded.

353 Genes detected in fewer than three cells were removed. Cells with more than 0.5% of
354 mitochondrial genes were removed. Cells with fewer than 30 genes and more than 1500 genes
355 per cell and less than 100 reads per cell were removed from merozoites. This threshold was
356 chosen because a previous single cell data (22) had shown that schizonts express an average
357 of 1040 genes, which for a schizont producing 30 merozoites, represents 34 genes per
358 merozoite assuming all merozoites express the same number of genes. Cells with fewer than
359 200 genes and more than 2500 genes per cells were removed from rings and merozoites
360 (when compared to rings).

361 The Seurat toolbox version 4.3.0.1 was used to normalize each dataset using SCTransform
362 (34) and cluster the cells. Visualizations and identification of differentially expressed genes
363 were done using the Loupe browser (10X Genomics, version 7).

364 Pseudotime was calculated using Monocle3 R package (35) to organize the merozoite and
365 ring cells.

366 Motif enrichment analysis was done using STREME(36) to identify enriched motifs in the 5'utr
367 regions (downloaded from PlasmoDB) as compared to scrambled sequences of similar length.
368 TOMTOM(37) was used to compare the *de novo* discovered motifs to previously *in silico*
369 discovered motifs(11). Only motifs found in STREME that were not a series of A and T were
370 considered given the enrichment of these nucleotides in the parasite genome.

371 The ATAC-Seq analysis were done by comparing the genes expressed in each cluster to
372 those identified in (19). We chose this particular ATAC-Seq dataset because peaks are labeled
373 as hpi instead of parasite stage like in (38).
374 The transcription versus RNA stability analysis were done by comparing the genes expressed
375 in each cluster to those labelled as transcribed or stabilized in (6).
376 The ChIP-Seq analysis were done by comparing the genes expressed in each cluster to the
377 target genes identified in (12,15,39) (18). Whenever ChIP-Seq data was available from
378 multiple sources, we filtered out target genes only found in one of the subsets.
379 To determine if the cells recovered in the second magnetic purification flow-through express
380 the same expression clusters as the ones recovered in the first purification flow-through, we
381 looked for cells expressing DE genes of each of the flow-through clusters and then manually
382 identified as belonging to that certain cluster.
383 Venn diagrams were generated using the online tool BioVenn.
384 Gene Ontology (GO) enrichment analysis was done using the GO feature at PlasmoDB (25).
385 Orthologue searches were done using OrthoMCL (40).
386
387 **Data availability**
388 All RNA sequencing data generated in this study will be available in the peer-reviewed
389 versoon.
390
391 **Acknowledgements**
392 W2mef parasites were obtained through BEI Resources, NIAID, NIH: *Plasmodium falciparum*,
393 Strain W2mef, MRA-165, contributed by Alan F. Cowman. The PlasmoDB.org and
394 OrthoMCL.org resources were invaluable to this work. We thank A. R. Gomes for helpful
395 discussions, S. Ralph for providing us the merozoite purification protocol and D. Noordemeer
396 and H. Painter for critical reading of the manuscript. We acknowledge the sequencing and
397 bioinformatics expertise of the I2BC High-throughput sequencing facility.
398

399 **Authors contribution**

400 Conceptualization: JMS

401 Parasite culturing: LP, JMS

402 Single cell experiment: YJ

403 Data analysis: DN, CH, JMS

404 Writing: JMS

405

406 **Funding**

407 This work was funded by a CNRS ATIP-Avenir grant and I2BC start-up funds awarded to JMS.

408 This work was supported as part of the France 2030 programme ANR-11-IDEX-0003, which

409 includes a research grant awarded by the interdisciplinary program OI LivingMachines@Work

410 from University Paris-Saclay to JMS and Laura Eme. *The I2BC High-throughput sequencing*

411 *facility is supported by* France Génomique *(funded by the French National Program*

412 *"Investissement d'Avenir" ANR-10-INBS-09).*

413

414 **Competing interests**

415 The authors declare no competing interests

416

417 **References**

418 1. Global Malaria Programme (GMP). World malaria report 2023. World Health Organization;
419 2023. Available from: https://www.who.int/publications/i/item/9789240086173

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

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

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

429 5. Boyle MJ, Wilson DW, Richards JS, Riglar DT, Tetteh KKA, Conway DJ, et al. Isolation of
430 viable Plasmodium falciparum merozoites to define erythrocyte invasion events and

431 advance vaccine and drug development. Proc Natl Acad Sci U S A. 2010 Aug
432 10;107(32):14378–83.

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

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

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

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

443 10. de Koning-Ward TF, Gilson PR, Boddey JA, Rug M, Smith BJ, Papenfuss AT, et al. A
444 newly discovered protein export machine in malaria parasites. Nature. 2009 Jun
445 18;459(7249):945–9.

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

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

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

455 14. Cui L, Miao J, Furuya T, Li X, Su X zhuan, Cui L. PfGCN5-Mediated Histone H3 Acetylation
456 Plays a Key Role in Gene Expression in Plasmodium falciparum. Eukaryot Cell. 2007
457 Jul;6(7):1219–27.

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

461 16. Young JA, Johnson JR, Benner C, Yan SF, Chen K, Le Roch KG, et al. In silico discovery
462 of transcription regulatory elements in Plasmodium falciparum. BMC Genomics. 2008 Feb
463 7;9:70.

464 17. Singh S, Santos JM, Orchard LM, Yamada N, van Biljon R, Painter HJ, et al. The PfAP2-
465 G2 transcription factor is a critical regulator of gametocyte maturation. Mol Microbiol. 2021
466 May;115(5):1005–24.

467 18. Shang X, Wang C, Fan Y, Guo G, Wang F, Zhao Y, et al. Genome-wide landscape of
468 ApiAP2 transcription factors reveals a heterochromatin-associated regulatory network
469 during Plasmodium falciparum blood-stage development. Nucleic Acids Res. 2022 Apr
470 8;50(6):3413–31.

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

475 20. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, et al. A compendium of
476 RNA-binding motifs for decoding gene regulation. Nature. 2013 Jul;499(7457):172–7.

477 21. Hollin T, Le Roch KG. From Genes to Transcripts, a Tightly Regulated Journey in
478 Plasmodium. Front Cell Infect Microbiol. 2020;10:801.

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

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

484 24. Modrzynska K, Pfander C, Chappell L, Yu L, Suarez C, Dundas K, et al. A Knockout
485 Screen of ApiAP2 Genes Reveals Networks of Interacting Transcriptional Regulators
486 Controlling the Plasmodium Life Cycle. Cell Host Microbe. 2017 Jan 11;21(1):11–22.

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

490 26. Davidson EH, McClay DR, Hood L. Regulatory gene networks and the properties of the
491 developmental process. Proc Natl Acad Sci. 2003 Feb 18;100(4):1475–80.

492 27. Whole-genome analysis of mRNA decay in Plasmodium falciparum reveals a global
493 lengthening of mRNA half-life during the intra-erythrocytic development cycle - PubMed
494 [Internet]. [cited 2024 Sep 16]. Available from: https://pubmed-ncbi-nlm-nih-
495 gov.insb.bib.cnrs.fr/17612404/

496 28. Bunnik EM, Batugedara G, Saraf A, Prudhomme J, Florens L, Le Roch KG. The mRNA-
497 bound proteome of the human malaria parasite Plasmodium falciparum. Genome Biol.
498 2016 Jul 5;17(1):147.

499 29. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, et al. Discovery of gene
500 function by expression profiling of the malaria parasite life cycle. Science. 2003 Sep
501 12;301(5639):1503–8.

502 30. Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. The Transcriptome of the
503 Intraerythrocytic Developmental Cycle of Plasmodium falciparum. PLOS Biol. 2003 Aug
504 18;1(1):e5.

505 31. Lu XM, Batugedara G, Lee M, Prudhomme J, Bunnik EM, Le Roch KG. Nascent RNA
506 sequencing reveals mechanisms of gene regulation in the human malaria parasite
507 Plasmodium falciparum. Nucleic Acids Res. 2017 Jul 27;45(13):7825–40.

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

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

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

35. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells | Nature Biotechnology [Internet]. [cited 2024 Sep 16]. Available from: https://www-nature-com.insb.bib.cnrs.fr/articles/nbt.2859

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

37. Quantifying similarity between motifs - PubMed [Internet]. [cited 2024 Aug 2]. Available from: https://pubmed.ncbi.nlm.nih.gov/17324271/

38. Jl R, Jj T, C B, A C, Jl GS, E GD. Characterization of the accessible genome in the human malaria parasite Plasmodium falciparum. Nucleic Acids Res [Internet]. 2018 Oct 12 [cited 2024 Sep 30];46(18). Available from: https://pubmed.ncbi.nlm.nih.gov/30016465/

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

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

Figure 1

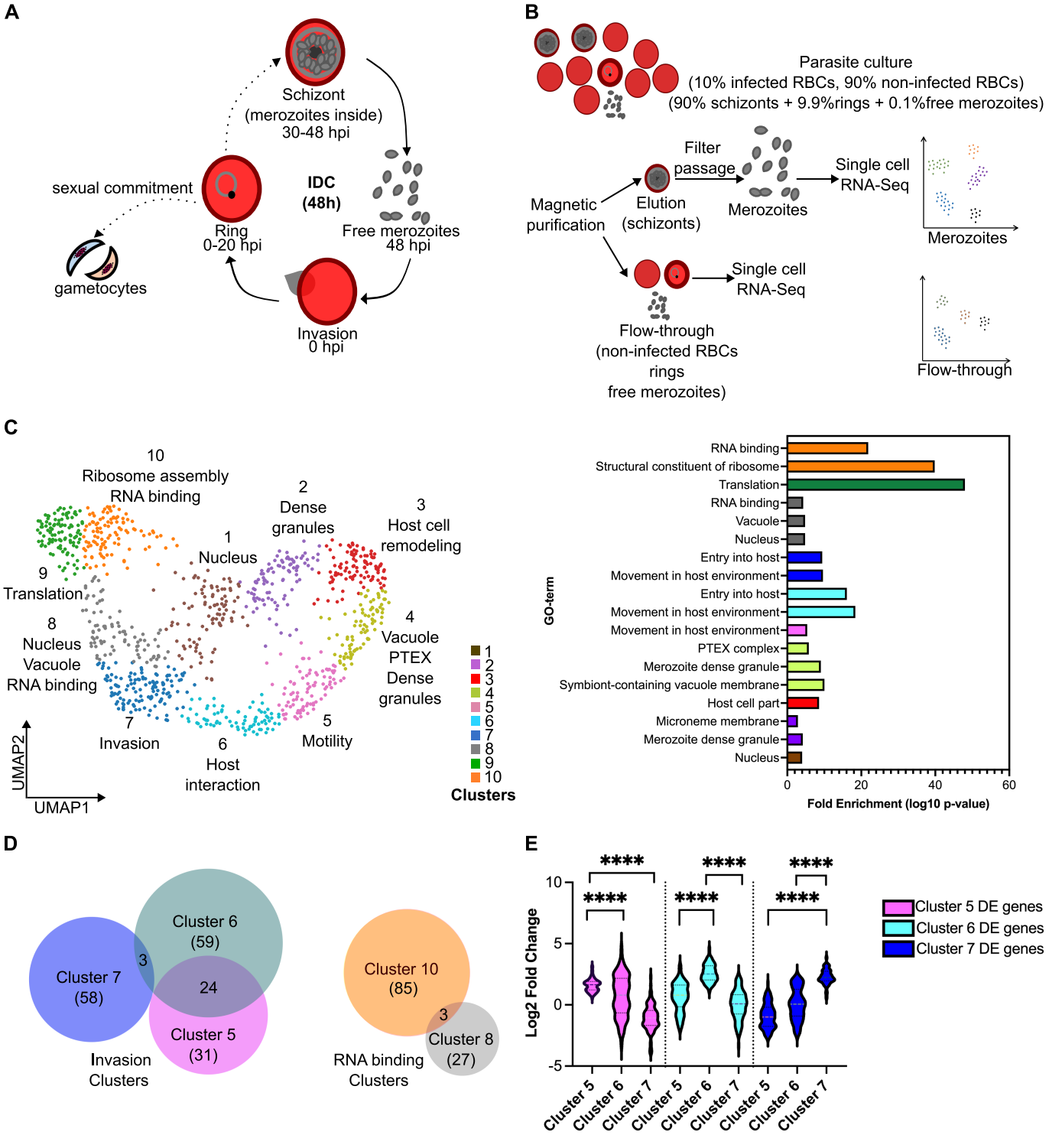
A
Schizont
(merozoites inside)
30-48 hpi
sexual commitment
IDC
(48h)
Ring
0-20 hpi
Free merozoites
48 hpi
gametocytes
Invasion
0 hpi
B
Parasite culture
(10% infected RBCs, 90% non-infected RBCs)
(90% schizonts + 9.9%rings + 0.1%free merozoites)
Filter
passage
Single cell
RNA-Seq
Magnetic
purification
Elution
(schizonts)
Merozoites
Merozoites
Single cell
RNA-Seq
Flow-through
(non-infected RBCs
rings
free merozoites)
Flow-through
C
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
Clusters
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)
D
Cluster 6
(59)
Cluster 7
(58)
3
24
Cluster 5
(31)
Invasion
Clusters
Cluster 10
(85)
3
Cluster 8
(27)
RNA binding
Clusters
E
****
****
****
****
****
****
10
5
0
-5
Log2 Fold Change
Cluster 5
Cluster 6
Cluster 7
Cluster 5
Cluster 6
Cluster 7
Cluster 5
Cluster 6
Cluster 7
Cluster 5 DE genes
Cluster 6 DE genes
Cluster 7 DE genes

Figure 1 – **During schizogony merozoite cells express different sets of genes. 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. 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** The parasite culture used contained 10% infected RBCs, corresponding to 90% schizont-stage parasites, 9.9% ring-stage parasites and 0.1% free merozoites. To obtain single cell RNA-seq data from merozoites and flow-through cells, independently, schizonts were purified from the culture by magnetic purification and then the merozoites inside were released by filter passage. In the flow-through, we collected the non-infected RBCs, the ring-stage parasites and the free merozoites. **C** UMAP representation showing the merozoite cells grouped by expression cluster (the number of cells in each cluster is in Supplementary Data 1). Gene ontology (GO) analysis of the genes differentially expressed (DE) in each merozoite cluster shows that each cluster expresses genes encoding proteins implicated in different processes. Cell and bar graphs are colored according to the corresponding cluster. Only the most significant GO-terms are shown. For the complete GO-term analysis, check Supplementary Data 2. A heatmap showing expression of 100 randomly selected genes in the 10 clusters is in Supplementary Fig. S1G. The distribution of the number of genes/UMIs per cell in each cluster is in Supplementary Fig. 1H. **D** Comparison of clusters enriched for the same GO-term. There is little overlap between invasion clusters 6 and 7 or 5 and 7 but invasion clusters 5 and 6 are similar. Clusters 8 and 10 (enriched for GO-term RNA binding) are not similar. Circles are colored according to the corresponding cluster. Comparison of other clusters is in Supplementary Fig. 1I. **E** 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. Differences between values were evaluated by an unpaired t-test. **** $P < 0.0001$.

Figure 2

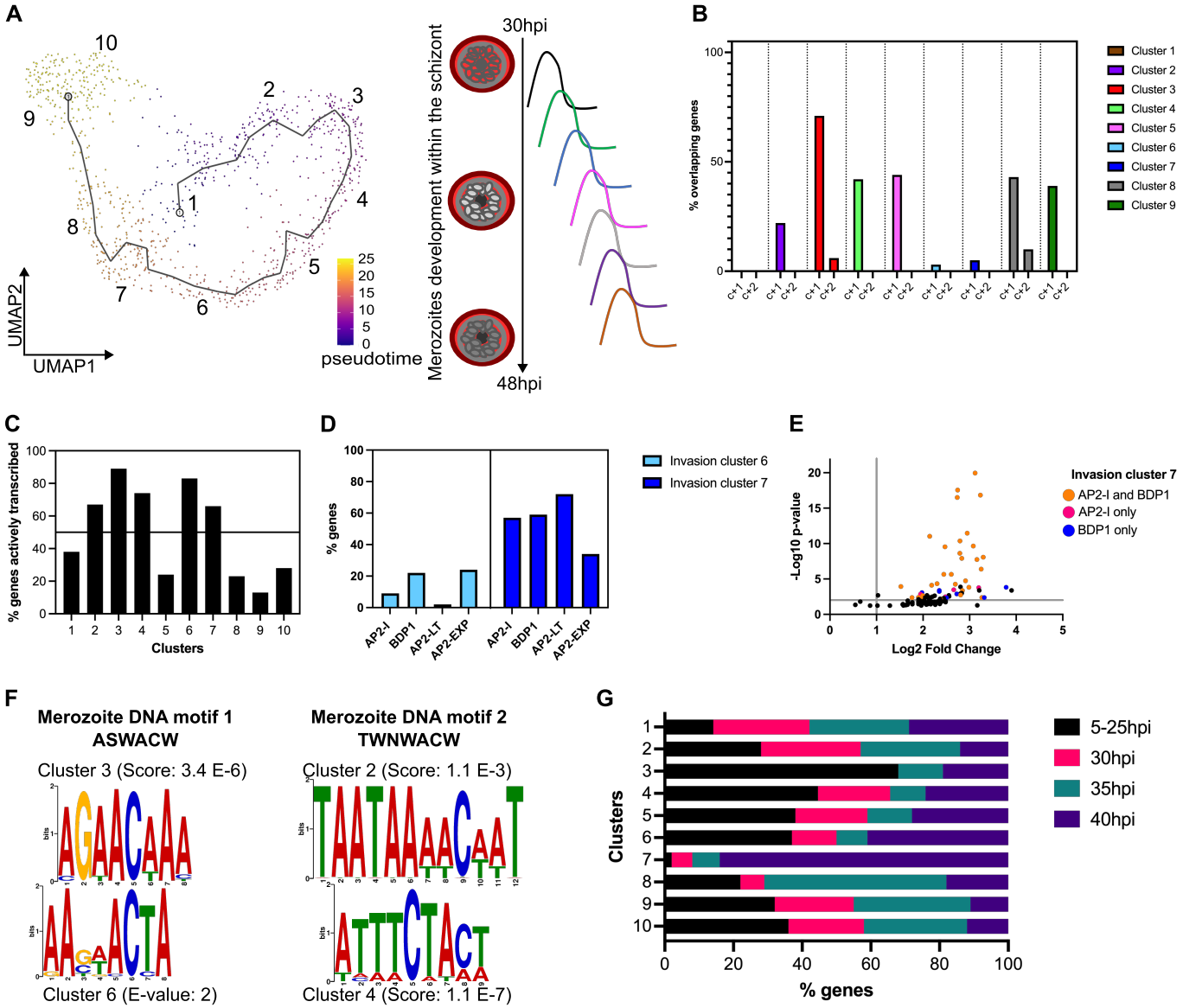
A
UMAP2
UMAP1
pseudotime
Merozoites development within the schizont
30hpi
48hpi
B
% overlapping genes
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
C
% genes actively transcribed
Clusters
D
% genes
Invasion cluster 6
Invasion cluster 7
AP2-I
BDP1
AP2-LT
AP2-EXP
E
-Log10 p-value
Log2 Fold Change
Invasion cluster 7
AP2-I and BDP1
AP2-I only
BDP1 only
F
Merozoite DNA motif 1
ASWACW
Cluster 3 (Score: 3.4 E-6)
Cluster 6 (E-value: 2)
Merozoite DNA motif 2
TWNWACW
Cluster 2 (Score: 1.1 E-3)
Cluster 4 (Score: 1.1 E-7)
G
Clusters
% genes
5-25hpi
30hpi
35hpi
40hpi

Figure 2 – Merozoites go through a transcriptional cascade during schizogony. A. UMAP projection and pseudotemporal analysis of the merozoite clusters showing that clusters are expressed sequentially. The cells pseudotime is in Supplementary Data 1 and the clusters pseudotime is in Supplementary Fig. S2A. **B**. Genes are still detected in the next cluster (c+1) but in the cluster afterwards (c+2) detection diminishes or is absent. Bars are colored according to cluster. **C**. Comparison of the genes expressed in each cluster with published data establishing which genes are actively transcribed or stabilized during the IDC[5] find that over 50% of genes of clusters 2, 3, 4, 6 and 7 are actively transcribed. **D**. Only the invasion genes expressed in cluster 7 are ChIP-Seq targets of AP2-I, BDP1 or AP2-LT[11-14]. **E**. The majority of genes expressed in invasion cluster 7 are ChIP-Seq targets of AP2-I (pink), BDP1 (blue) or both (orange). **F**. STREME analysis showing that a DNA motif, here called merozoite DNA motif 1, is present in the promoter region of the DE genes of clusters 3 and 6, while in clusters 2 and 4 gene promoters, there is merozoite DNA motif 2. E-value and Scores indicate the significance of the found DNA motif. The DREME and TOMTOM analysis are in Supplementary Data 4. **G**. Comparison of the genes expressed in each cluster with published ATAC-Seq data[17] shows that the pseudotemporal progression of the actively transcribed clusters (2, 3, 4, 6 and 7) concurs with the sequence of chromatin accessibility, with 84% of cluster 7 genes being euchromatic at 40hpi while 86% of cluster 2 genes are euchromatic at 5-35hpi. There is no concordance between pseudotime and chromatin accessibility in the non-actively transcribed clusters (1, 5, 8, 9 and 10), except for cluster 5. The indicated times correspond to the earliest assigned ATAC-Seq peak. The ATAC-Seq analysis is in Supplementary Data 2.

Figure 3

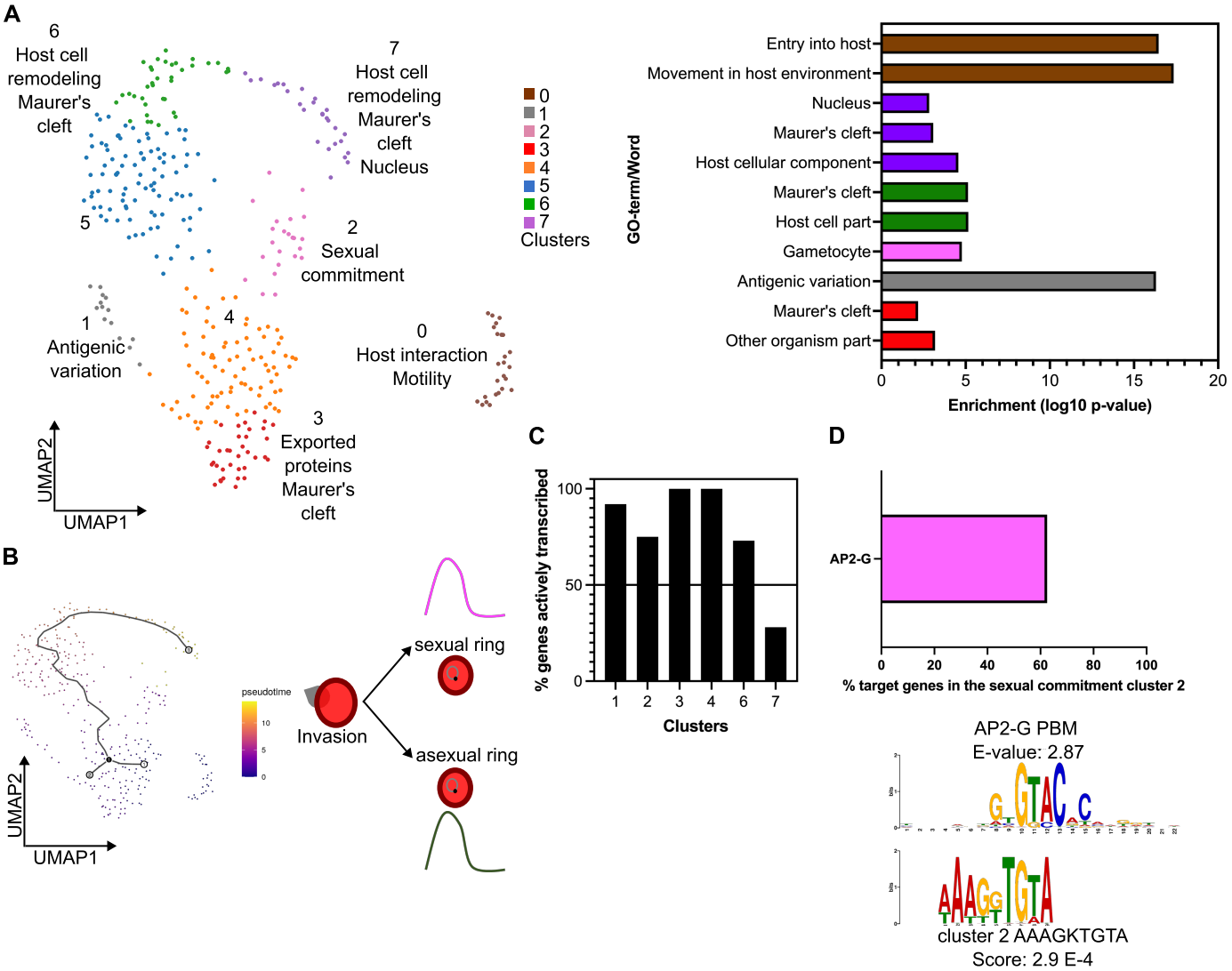
A
6
Host cell
remodeling
Maurer's
cleft
7
Host cell
remodeling
Maurer's
cleft
Nucleus
5
2
Sexual
commitment
1
Antigenic
variation
4
0
Host interaction
Motility
3
Exported
proteins
Maurer's
cleft
UMAP2
UMAP1
0
1
2
3
4
5
6
7
Clusters
GO-term/Word
Entry into host
Movement in host environment
Nucleus
Maurer's cleft
Host cellular component
Maurer's cleft
Host cell part
Gametocyte
Antigenic variation
Maurer's cleft
Other organism part
0
5
10
15
20
Enrichment (log10 p-value)
B
pseudotime
10
5
0
UMAP2
UMAP1
Invasion
sexual ring
asexual ring
C
% genes actively transcribed
100
50
0
1
2
3
4
6
7
Clusters
D
AP2-G
0
20
40
60
80
100
% target genes in the sexual commitment cluster 2
AP2-G PBM
E-value: 2.87
cluster 2 AAAGKTGTA
Score: 2.9 E-4

Figure 3 – The merozoite genes are not expressed after invasion. **A**. UMAP representation showing the flow-through cells grouped by expression cluster (the number of cells in each cluster is in Supplementary Data 1). According to GO-term analysis of the DE genes of each cluster, these cells express genes implicated in host cell remodeling, nuclear processes, antigenic variation, invasion or sexual commitment. There were not enough DE genes in clusters 4 and 5 to perform GO-term analysis. Because there is no GO-term associated with gametocytes, word search analysis was done instead for cluster 2 genes. Bar graphs are colored according to the corresponding cluster. Only the most significant GO-terms are shown. The full list of GO-terms and words is in Supplementary Data 3. A heatmap showing expression of 100 randomly selected genes in the 8 clusters is in Supplementary Fig. 3A. The distribution of the number of genes/UMIs per cell in each cluster is in Supplementary Fig. 3B. **B**. UMAP projection and schematic of the pseudotemporal ordering of the ring clusters showing that ring cells either commit to sexual development and express cluster 2 or go on to express the antigenic variation genes. Numbers indicate expression clusters. The cells pseudotime is in Supplementary Data 1 and the clusters pseudotime is in Supplementary Fig. 3C. **C**. Comparison of genes expressed in each cluster with published transcription data[5] found that over 50% of the genes expressed in all clusters, except cluster 7, are actively transcribed between 1 and 20hpi (rings developmental timing). **D**. The DNA-binding motif of AP2-G is enriched in the promoter of cluster 2 genes and a majority of these cluster genes are AP2-G ChIP-Seq targets[22]. The ChIP-Seq analysis are in Supplementary Data 3. The DNA motifs enriched in the promoters of the other clusters genes is in Supplementary Fig. 3D.

Figure 4

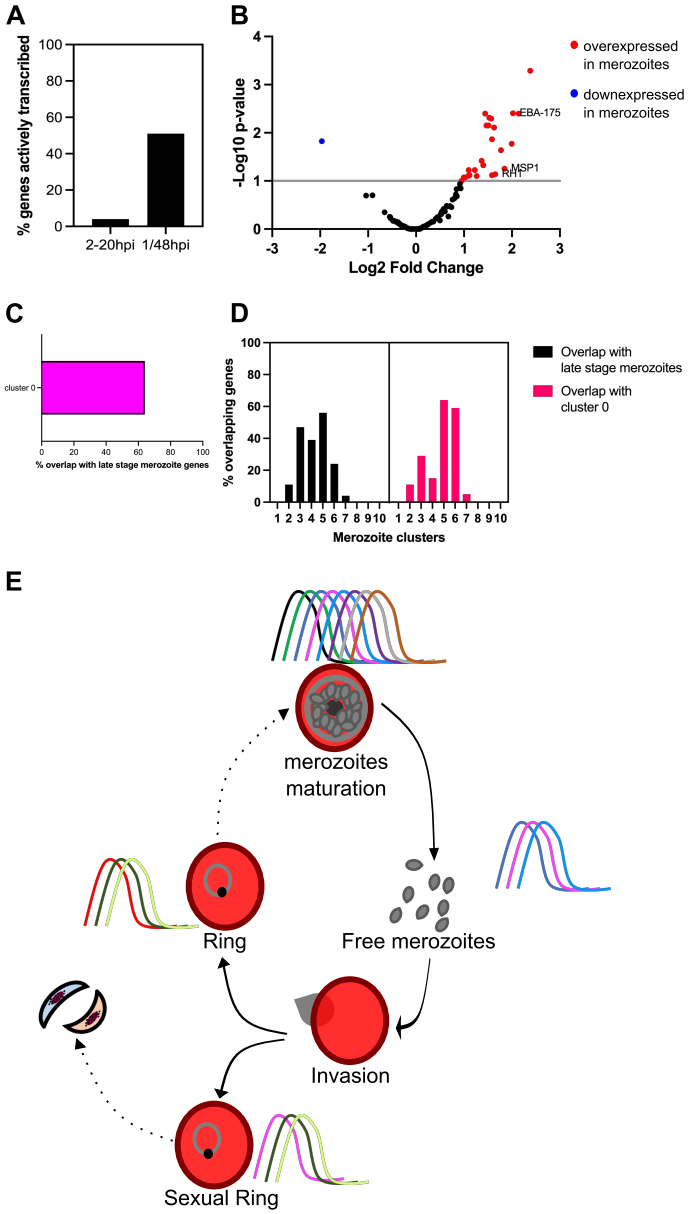
A
% genes actively transcribed
100
80
60
40
20
0
2-20hpi
1/48hpi
B
-Log10 p-value
Log2 Fold Change
overexpressed in merozoites
downexpressed in merozoites
EBA-175
MSP1
RH1
C
cluster 0
% overlap with late stage merozoite genes
D
% overlapping genes
Merozoite clusters
Overlap with late stage merozoites
Overlap with cluster 0
E
merozoites maturation
Ring
Free merozoites
Invasion
Sexual Ring

Figure 4 – **Transcriptional signature of free merozoite cells**. **A**. The majority of genes detected in cluster 0 of flow-through cells are actively transcribed at 1/48hpi. **B**. Volcano plot showing that invasion genes, such as *msp1*, *eba-175* and *rh1* are overexpressed in late merozoite cells versus ring cells. There was only one gene overexpressed in rings. The full list of genes and GO-term analysis are in Supplementary Data 5. **C**. The majority of genes overexpressed in late merozoite cells is detected in cluster 0 cells. **D**. Merozoite clusters 3, 5 and 6 are still expressed in free merozoites since over 50% of those cluster genes are detected in late stage merozoites and in cluster 0 cells. **E**. Scheme of the merozoites transcriptional cascade and progression of gene expression upon schizont rupture, in free merozoites, and after host cell invasion, in ring-stage cells.