The malaria parasite *Plasmodium falciparum* prepares for life within the host cell by activating a complex gene network

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 malaria parasite life cycle. The IDC lasts 48-hours and it initiates when parasite stages called merozoites are released into the bloodstream, following schizont rupture, and invade the host red blood cells. Within the host, the parasite transforms into the ring stage. The merozoite stage is the less studied stage of the IDC, despite its importance for the establishment of infection. Here we did the first transcriptional profiling of purified merozoites throughout development inside the schizont and after egress. We find that merozoites within the schizont, activate a series of tightly regulated gene expression programs encoding the proteins required for host cell invasion and life within the host. After schizont egress, free merozoites still express some of these programs. Upon invasion, the merozoites transcriptional programs halt and ring cells have a distinct transcriptional profile. Each of the gene expression programs encodes for proteins implicated in a specific process and we provide evidence that they 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.

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, causing the majority of the yearly over half a million deaths (WHO). 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) (Figure 1A), in which the parasite multiplies within the human host red blood cell (RBC) for 48-hours, starts when parasite stages called merozoites, are released into the bloodstream, following schizont rupture, and invade the host RBC. Once inside the host cell, the parasite transforms into the ring stage. Some parasites exit the IDC and differentiate into the sexually committed cells, precursor of the male and female gametocytes, the transmission stages, that are uptaken upon a mosquito bite.

The merozoite is the less studied stage of the IDC, despite its attractiveness as a drug target given that it is extracellular, even if briefly[2], it is responsible for RBC invasion, which is indispensable for parasite survival since the parasite is obligatory intracellular, and it initiates the IDC, which is the symptomatic life cycle stage of malaria.

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

factors and epigenetic factors controlling chromatin accessibility as well as post-translational mechanisms.

Our comprehensive study of parasite gene expression before and after RBC invasion identifies potential new drug targets and further sheds light into regulation of gene expression in the malaria parasite.

Results

Merozoites express different sets of genes throughout development inside the schizont

In order to identify the transcription profile of merozoites during schizogony and after schizont rupture, we adapted a protocol described before for the isolation of viable invasive merozoites[5] (Figures 1B and S1A). We synchronized parasites, and when schizonts were visible in Giemsa-stained smears, we collected the culture. The culture also contained non-infected RBCs (90% of cells), ring-stage parasites and likely free merozoites. We then passed the culture through magnetic columns to retain the schizonts, which are magnetic due to the accumulation of hemozoin, the end product of hemoglobin digestion. In the flow-through, we recovered the non-infected RBCs, the ring stage parasites and the free merozoites. After elution from the column, the schizonts were passed through a 1.2μM filter to release the merozoites inside. The isolated merozoite and flow-through cells were then prepared individually for single cell RNA-seq. We obtained a high sequencing saturation and mapability to the parasite genome and detected a similar number of genes (over 4800, ~90% of the parasite genome) in both samples (Table S1). The purified merozoite cells express the merozoite marker *merozoite surface marker 1* (*msp1*, PF3D7_0930300) and do not express the ring marker *ring-exported protein 2* (*rex2*, PF3D7_0936000) (Figures S1B-C). On the other hand, the majority of flow-through cells expresses *rex2* and only residual amounts of *msp1* (Figures S1B-D) and are ring-stage parasites. Furthermore, we detected less genes and transcripts per cell in merozoite samples versus flow-through samples (Figures S1E-F), in accordance with previous findings[3,6].

To analyze the heterogeneity of the merozoite expression programs we used Seurat, a toolbox for quality check, analysis and visualization of single cell data[7–9]. We identified 10 expression clusters (Figure 1C). The 10 clusters were characterized by expression of distinct differentially expressed (DE) genes (Figure S1G and Table S2). Clustering was independent of the number of genes expressed by the cell or the cells RNA content (Figure S1H). When we did gene ontology (GO) analysis, we found that each cluster expresses a specific category of genes (Figure 1C and Table S2). While clusters 5, 6 and 7 express genes related to host cell invasion, in clusters 1, 8, 9 and 10 we detected genes related to nuclear processes, RNA binding and protein synthesis. Clusters 2, 3 and 4 express genes encoding for proteins stored in the dense granules that are later discharged and participate in events taking place post-invasion - formation of the parasitophorous vacuole (PV), wherein the parasite resides after invasion, formation of the *Plasmodium* translocon of exported proteins (PTEX)[10], in the PV membrane, through which parasite proteins are exported into the host cell, and host cell remodeling allowing the parasite to import media nutrients.

Because some clusters were enriched for the same GO-terms, we compared them but we found none to little overlap, except for clusters 5 and 6, related to invasion (Figures 1D and S1I). Clusters 5 and 6 invasion clusters are, however, different from the other invasion cluster, cluster 7 (Figure 1D), and marker genes of a specific invasion cluster are overexpressed in that cluster versus the other two invasion clusters (Figure 1E). Each merozoite cluster corresponds then to a specific expression cluster.

Merozoites go through a transcriptional cascade during schizogony

Pseudotemporal analysis suggests that the merozoite cells, during schizogony, follow one developmental trajectory marked by the sequential activation of each of the expression clusters (Figures 2A and S2A and Table S1). The marker genes of a specific cluster are still expressed in the next cluster but then few to none of those genes are detected in the afterwards cluster (Figure 2B), showing that expression of the genes of a given cluster peaks at a precise time. To determine if the expression clusters correspond to transcriptional

programs, we determined how many of the DE genes of a cluster are transcribed by comparing our dataset to a previously published list of genes actively transcribed throughout the IDC[6]. Clusters 2, 3, 4, 6 and 7 likely correspond to transcriptional programs because over 50% of their DE genes are actively transcribed during schizogony (Figure 2C).

Transcriptional programs are regulated by the binding of transcription factors to gene promoters. The ApiAP2 DNA-binding proteins are the main family of transcription factors in the parasite[11]. The ApiAP2 proteins AP2-I, AP2-LT and AP2-EXP and the bromodomain protein BDP1 have been linked to expression of invasion genes[12–15] so we determined if the genes of the invasion transcriptional programs, clusters 6 and 7, are regulated by these proteins. AP2-I, and associated proteins, regulate the cluster 7 invasion transcriptional program, because 88% of this cluster gene promoters are bound by AP2-I, BDP1, AP2-LT or a combination of the three proteins (Figures 2D-E). That is not the case for the cluster 6 genes (Figure 2D) so to identify the DNA-binding proteins regulating transcription of this and the other three transcriptional programs, we searched for enrichment of DNA sequence motifs in the promoters of the DE genes. We identified two DNA motifs, one enriched in the promoters of clusters 3 and 6 genes, that we named merozoite DNA motif 1, and one enriched upstream of clusters 2 and 4 genes, here named merozoite DNA motif 2 (Figure 2F). The parasite repressor protein AP2-G2 binds to a DNA motif[16,17] similar to merozoite DNA motif 1 (Figure S2B) so we examined if the gene promoters of clusters 3 and 6, containing the merozoite DNA motif 1, are AP2-G2 ChIP-Seq targets and found that about half are (Figure S2C), suggesting that AP2-G2 represses transcription of these clusters. Comparison of merozoite DNA motif 2 with the DNA motifs bound by the ApiAP2 proteins[11] indicate that the uncharacterized ApiAP2 protein PF3D7_1239200 may bind to it (Figure S2D).

Clusters associated with the same merozoite DNA motif are expressed at different pseudotimes. We wondered if this reflects different times of chromatin opening at the promoter regions allowing protein binding earlier or later post-invasion. We determined when the promoters of each of the clusters was euchromatic by comparing our gene list to a published ATAC-Seq data[18]. The order of activation of the transcription clusters 2, 3, 4, 6 and 7 is linked

to the timing of chromatin opening because a larger proportion of gene promoters of later expressed clusters is euchromatic at 40hpi (Figure 2G).

Our analysis indicates that clusters 1, 5, 8, 9 and 10 are not actively transcribed (Figure 2C) so we wondered if the transcripts detected in these clusters are stabilized. To do that, we determined how many of these clusters genes were labelled as stabilized by Painter *et al*. [6]. While the majority of the transcripts detected in cluster 5 are stabilized, that is not true for clusters 1, 8, 9 and 10 (Figure S2E). However, there is no correlation between the order in which clusters 1, 8, 9 and 10 are expressed and the timing of chromatin opening (Figure 2G). Given that the genes expressed in cluster 5 are stabilized and the genes expressed in clusters 1, 8, 9 and 10 are not expressed when most promoters are euchromatic, their regulation is likely post-transcriptional and mediated by RNA-binding proteins. Sequence enrichment analysis identified distinct motifs in the 3' untranslated regions (3'utr) of the 1, 5, 8, 9 and 10 DE genes (Table S4). The RNA-binding protein PF3D7_0823200 binds similar motifs to those identified in the 3'utr of clusters 5,8 and 9 genes[19,20] (Figure S2F) but few genes in these clusters are bound by the protein (Figure S2G). The motifs found in the 3'utr of clusters 1 and 10 genes resemble those bound by two other putative RNA-binding proteins that remain unstudied (Figure S2F).

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

To establish if there is variation in expression of the merozoite programs in merozoites within a same schizont or all merozoites inside a schizont express the same genes, we quantified which of the clusters have genes with a high variability index score (VIS), a metric established by Tripathi *et al*. to identify genes with high merozoite-to-merozoite variation in mature schizonts[21]. Over half of the merozoite clusters 3 and 4 genes, encoding for proteins implicated in host cell remodeling, and 80% and 69% of the genes in the invasion clusters 5

and 6, respectively, have a VIS>1 (Figure S2I and Table S2), suggesting that expression of these particular expression programs is stochastic.

Parasites activate different transcriptional programs post-invasion

We performed single cell RNA-Seq on flow-through cells to identify the transcriptional programs expressed by the parasite after schizont rupture. We identified 8 expression clusters, each characterized by the expression of a specific set of genes (Figures 3A and S3A-B and Table S3). GO-term analysis of the marker genes of each cluster (Table S3) reveal that clusters 3, 6 and 7 express genes coding proteins implicated in the formation of the Maurer's cleft, parasite structures from which the exported parasite proteins are sorted and transported to the host cell (Figure 3A) but we did not find any overlapping genes (Table S3), suggesting that they are separate expression programs. Cluster 1 expresses the *var* genes encoding for EMP1 (erythrocyte surface protein 1), implicated in antigenic variation (Figure 3A), and cluster 2 genes code for proteins implicated in sexual commitment (Figure 3A). Apart from cluster 0, which is specialized in host interaction and motility (Figure 3A) and will be discussed later, we believe that these clusters correspond to expression programs expressed by ring-stage parasites because these are the predominant stages in the flow-through, as previously discussed. Since most parasites in the parasite culture collected for magnetic purification are schizont staged, there are few cells in each cluster (Table S1), but sequencing and analysis of cells coming from a second magnetic purification, containing more flow-through cells, identified the same expression programs (Figure S3C).

Pseudotemporal analysis suggest that, after RBC invasion, early ring stage parasites are already committed to pursue the IDC by expressing cluster 1 or express instead cluster 2 and commit to sexual differentiation. Then ring cells express clusters 3 to 7 encoding for proteins implicated in host cell remodeling and nuclear functions (Figures 3B and S3D).

Unlike the merozoite clusters, all the ring expression clusters seem to correspond to transcriptional programs because the majority of all clusters DE genes, except for cluster 7, is actively transcribed (Figure 3C). Regulation of the ring transcriptional programs is thus likely

mediated by the binding of specific transcription factors to the gene promoters. The time at which the chromatin at the promoter regions of the genes is accessible correlates somehow with the progressive activation of the clusters (Figure S3E). As expected, the sexual commitment transcriptional program corresponding to cluster 2 is regulated by AP2-G, the parasite gametocytogenesis master regulator[22] (Figure 3D). Motif enrichment analysis shows that there are different DNA motifs in the promoters of the genes expressed in the other clusters and indicate, for instance, that the ApiAP2 protein PF3D7_0613800 may regulate the cluster 6 genes (Figure S3F and Table S4). PF3D7_0613800 remains uncharacterized but it is essential for parasite growth[23].

We examined if there was any correlation between the transcriptional programs expressed by merozoites and ring stage cells. Merozoite and ring stage cells both express genes encoding for nuclear proteins or proteins implicated in host cell remodeling but the merozoites and ring transcriptional programs are distinct because there is no overlap between the genes expressed by the two types of cells (Tables S2 and S3). This implies that the merozoite transcriptional programs cease after RBC invasion and new programs are activated.

Free merozoites express host cell invasion and remodeling genes

We noticed that the cluster 0 of flow-through cells is further away from the other clusters in the UMAP projection and it is excluded from the pseudotemporal trajectory (Figures 3A-B). Furthermore, cluster 0 cells contain a lower amount of RNA (Figure S3B) and the genes expressed in cluster 0 are not transcribed at the ring-stage (2-20hpi) but at the end of schizogony (1/48hpi) (Figure 4A). This led us to ponder if these cells are free merozoites that egressed from schizonts before we collected the parasite culture and did the magnetic purification. To test this hypothesis, we combined the data from the merozoite and ring cells and filtered the cells, using the same parameters we used for the ring cells, to only select for late stage merozoites. Then we identified the genes that were over-expressed or down-expressed in the merozoites versus the ring cells (Figure 4B). Genes over-expressed in

merozoites represent the late stage merozoites transcriptional signature. We found 25 genes over-expressed in merozoites (Table S5), coding for proteins implicated in host cell invasion and host cell remodeling (Figure S4A). Confirming that cluster 0 cells are indeed free merozoites, over 60% of the genes expressed in cluster 0 are part of the late stage merozoites transcriptional signature (Figure 4C). Cluster 0 then represents the transcriptional profile of free merozoites, an elusive stage that had never before been captured by transcriptional studies.

We determined which of the genes expressed by merozoites during schizogony are still detected in free merozoites. Marker genes of merozoite clusters 2, 3, 4, 5, 6 and 7 encoding for genes implicated in host cell remodeling, vacuole formation, motility and invasion are detected in both late stage merozoites and cluster 0 (Figure 4D), indicating that these genes are still expressed in free merozoites. Supporting this, the merozoite DNA motif 1, which we previously showed is present in the promoters of merozoite clusters 3, 5 and 6, is also present in the promoters of genes expressed in free merozoites (Figure S4B). Whereas more than 60% of the invasion merozoite clusters 5 and 6 are still expressed in free merozoites, only 5% of the invasion merozoite cluster 7 genes are (Figure 4D). This indicates that only the non-AP2-I merozoite invasion transcriptional program remains active in free merozoites. Indeed, the genes expressed in free merozoites are not regulated by AP2-I and associated proteins (Figure S4C). Thus, unlike what was previously believed, expression of some invasion genes continues after schizont rupture[6,24–26]. These are the genes coding for proteins implicated in the later steps of invasion, such as host receptor interaction, motility and penetration[12], rather than RBC recognition and attachment.

Discussion

Our data indicates that there are rapid changes in gene expression during formation and segmentation of merozoites (Figure 4E). Merozoites express genes encoding nuclear proteins and proteins destined to the dense granules as well as gene products implicated in

host cell interaction, PV formation, invasion and motility allowing infection of new RBCs. There is also a merozoite program dedicated to expression of genes encoding for proteins implicated in RNA binding and translation. Free merozoites still express the PV and host cell remodeling programs as well as genes implicated in motility and host interaction. All of the merozoite expression programs shut down once the parasite invades the host cell and transforms into ring-stage parasites. The newly activated ring transcriptional programs allow parasites to decide between committing to gametocytogenesis or continue asexual development inside the RBC.

Around 40% of the parasite genome remains unannotated[27] because there are either no recognizable functional domains or it is unknown if the annotated domains are functional. The ring clusters are particularly enriched in unannotated genes (Figure S4D) and the merozoite program with the higher percentage of unknown genes is cluster 10 specialized in ribosome assembly (Figure S4D), reflecting how little we still know about translation in the parasite. Our study can be used to discover new proteins implicated in this process as well as vacuole formation, host cell remodeling and others.

Merozoites are responsible for RBC invasion and initiate the symptomatic phase of the malaria life cycle but they are the less transcriptionally understood stage of the IDC. Here we identified a transcriptional cascade controlling merozoite formation and maturation during schizogony, implying that gene expression at this stage of the parasite life cycle is tightly regulated. This resembles the developmental gene networks of other organisms[28] where regulation of gene expression implicates a cascade of transcription factors and expression of proteins that repress translation of other proteins. We propose that regulation of gene expression during merozoite maturation inside schizonts is transcriptional, involving binding of proteins to specific DNA motifs in the gene promoters, once the chromatin is accessible, and post-transcriptional. On the other hand, regulation of gene expression in ring-stage parasites is mainly transcriptional. This is in accordance with previous studies that found that at the final stages of schizogony, there is increased expression of RNA-binding proteins and gene expression is mainly the consequence of RNA stabilization, while expression during ring-

stage parasites is linked to active transcription[6,29,30]. We did detect expression of different ApiAP2 proteins and RNA-binding proteins throughout merozoite development (Table S1) but further studies are needed to determine which of these are implicated in regulation of the merozoite expression programs.

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

Material and methods

Parasite culturing and merozoites purification

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

Parasite synchronization to the ring stage was done two consecutive cycles by incubating pelleted cultures 1:1 with 0.3M L-Alanine, 10mM HEPES pH 7.5 solution for 10 min at 37°C[31], after which cultures were centrifuged, resuspended in fresh media and placed back in culture.

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

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

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

Single cell RNA-Seq

Approximately 10 000 cells were loaded on a Chromium Chip B (10X Genomics) as per the manufacturer's recommendations. Chips containing cell suspensions, gel beads in emulsion (GEMs) and reverse transcription reagents were loaded into the Chromium Controller (10X Genomics) for single cell partitioning and cell barcoding. Barcoded cDNA libraries were generated according to the Chromium Single Cell 3' gene expression protocol. Quantification and the quality of sequencing libraries was verified on an Agilent Bioanalyzer. Pooled libraries were sequenced on a NextSeq system (Illumina).

Data analysis

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

Genes detected in fewer than three cells were removed. Cells with more than 0.5% of mitochondrial genes were removed. Cells with fewer than 30 genes and more than 1500 genes per cell and less than 100 reads per cell were removed from merozoites. This threshold was chosen because a previous single cell data[21] had shown that schizonts express an average of 1040 genes, which for a schizont producing 30 merozoites, represents 34 genes per merozoite assuming all merozoites express the same number of genes. Cells with fewer than 200 genes and more than 2500 genes per cells were removed from rings and merozoites (when compared to rings).

The Seurat toolbox version 4.3.0.1 was used to normalize each dataset using SCTransform[32] and cluster the cells. Visualizations and identification of differentially expressed genes were done using the Loupe browser (10X Genomics, version 7).

Pseudotime was calculated using Monocle3 R package[33] to organize the merozoite and ring cells.

Motif enrichment analysis was done using STREME[34] to identify enriched motifs in the 5'utr regions (downloaded from PlasmoDB) as compared to scrambled sequences of similar length. TOMTOM[35] was used to compare the *de novo* discovered motifs to previously *in silico* discovered motifs[11]. Only motifs found in STREME that were not a series of A and T were considered given the enrichment of these nucleotides in the parasite genome.

The ATAC-Seq analysis were done by comparing the genes expressed in each cluster to those identified in [18].

The transcription versus RNA stability analysis were done by comparing the genes expressed in each cluster to those labelled as transcribed or stabilized in [6].

The ChIP-Seq analysis were done by comparing the genes expressed in each cluster to the target genes identified in [12,15,36,17]. Whenever ChIP-Seq data was available from multiple sources, we filtered out target genes only found in one of the subsets.

Venn diagrams were generated using the online tool BioVenn.

Gene Ontology (GO) enrichment analysis was done using the GO feature at PlasmoDB[27].

Data availability

All RNA sequencing data generated in this study will be available upon peer-reviewed publication.

Acknowledgements

W2mef parasites were obtained through BEI Resources, NIAID, NIH: *Plasmodium falciparum*, Strain W2mef, MRA-165, contributed by Alan F. Cowman. The parasite genome was downloaded from PlasmoDB.org, which resources were invaluable to this work. We thank A. R. Gomes for helpful discussions and D. Noordemeer and H. Painter for critical reading of the manuscript. We thank S. Ralph for providing us the merozoite purification protocol. We acknowledge the sequencing and bioinformatics expertise of the I2BC High-throughput sequencing facility, supported by *France Génomique* (funded by the French National Program "Investissement d'Avenir" ANR-10-INBS-09). This work was funded by a CNRS ATIP-Avenir

grant. This work was supported as part of the France 2030 programme ANR-11-IDEX-0003, which includes a research grant awarded by the interdisciplinary program OI LivingMachines@Work from University Paris-Saclay.

Author contribution

Conceptualization: JMS

Parasite culturing: LP, JMS

Single cell experiment: YJ

Data analysis: DN, CH, JMS

Writing: JMS

Competing interests

The authors declare no competing interests

References

1. Global Malaria Programme (GMP) (2023). World malaria report 2023.

2. Yahata, K., Treeck, M., Culleton, R., Gilberger, T.-W., and Kaneko, O. (2012). Time-Lapse Imaging of Red Blood Cell Invasion by the Rodent Malaria Parasite Plasmodium yoelii. PLOS ONE *7*, e50780. https://doi.org/10.1371/journal.pone.0050780.

3. Howick, V.M., Russell, A.J.C., Andrews, T., Heaton, H., Reid, A.J., Natarajan, K., Butungi, H., Metcalf, T., Verzier, L.H., Rayner, J.C., et al. (2019). The Malaria Cell Atlas: Single parasite transcriptomes across the complete Plasmodium life cycle. Science *365*, eaaw2619. https://doi.org/10.1126/science.aaw2619.

4. Reers, A.B., Bautista, R., McLellan, J., Morales, B., Garza, R., Bol, S., Hanson, K.K., and Bunnik, E.M. (2023). Histone modification analysis reveals common regulators of gene expression in liver and blood stage merozoites of Plasmodium parasites. Epigenetics Chromatin *16*, 25. https://doi.org/10.1186/s13072-023-00500-y.

5. Boyle, M.J., Wilson, D.W., Richards, J.S., Riglar, D.T., Tetteh, K.K.A., Conway, D.J., Ralph, S.A., Baum, J., and Beeson, J.G. (2010). Isolation of viable Plasmodium falciparum merozoites to define erythrocyte invasion events and advance vaccine and drug development. Proc. Natl. Acad. Sci. U. S. A. *107*, 14378–14383. https://doi.org/10.1073/pnas.1009198107.

6. Painter, H.J., Chung, N.C., Sebastian, A., Albert, I., Storey, J.D., and Llinás, M. (2018). Genome-wide real-time in vivo transcriptional dynamics during Plasmodium falciparum

blood-stage development. Nat. Commun. *9*, 2656. https://doi.org/10.1038/s41467-018-04966-3.

7. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., et al. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat. Biotechnol. *42*, 293–304. https://doi.org/10.1038/s41587-023-01767-y.

8. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

9. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502. https://doi.org/10.1038/nbt.3192.

10. de Koning-Ward, T.F., Gilson, P.R., Boddey, J.A., Rug, M., Smith, B.J., Papenfuss, A.T., Sanders, P.R., Lundie, R.J., Maier, A.G., Cowman, A.F., et al. (2009). A newly discovered protein export machine in malaria parasites. Nature *459*, 945–949. https://doi.org/10.1038/nature08104.

11. Campbell, T.L., Silva, E.K.D., Olszewski, K.L., Elemento, O., and Llinás, M. (2010). Identification and Genome-Wide Prediction of DNA Binding Specificities for the ApiAP2 Family of Regulators from the Malaria Parasite. PLOS Pathog. *6*, e1001165. https://doi.org/10.1371/journal.ppat.1001165.

12. Santos, J.M., Josling, G., Ross, P., Joshi, P., Orchard, L., Campbell, T., Schieler, A., Cristea, I.M., and Llinás, M. (2017). Red Blood Cell Invasion by the Malaria Parasite Is Coordinated by the PfAP2-I Transcription Factor. Cell Host Microbe *21*, 731-741.e10. https://doi.org/10.1016/j.chom.2017.05.006.

13. Josling, G.A., Petter, M., Oehring, S.C., Gupta, A.P., Dietz, O., Wilson, D.W., Schubert, T., Längst, G., Gilson, P.R., Crabb, B.S., et al. (2015). A Plasmodium Falciparum Bromodomain Protein Regulates Invasion Gene Expression. Cell Host Microbe *17*, 741–751. https://doi.org/10.1016/j.chom.2015.05.009.

14. Cui, L., Miao, J., Furuya, T., Li, X., Su, X., and Cui, L. (2007). PfGCN5-Mediated Histone H3 Acetylation Plays a Key Role in Gene Expression in Plasmodium falciparum. Eukaryot. Cell *6*, 1219–1227. https://doi.org/10.1128/EC.00062-07.

15. Bonnell, V.A., Zhang, Y., Brown, A.S., Horton, J., Josling, G.A., Chiu, T.-P., Rohs, R., Mahony, S., Gordân, R., and Llinás, M. (2024). DNA sequence and chromatin differentiate sequence-specific transcription factor binding in the human malaria parasite Plasmodium falciparum. Nucleic Acids Res., gkae585. https://doi.org/10.1093/nar/gkae585.

16. Singh, S., Santos, J.M., Orchard, L.M., Yamada, N., van Biljon, R., Painter, H.J., Mahony, S., and Llinás, M. (2021). The PfAP2-G2 transcription factor is a critical regulator of gametocyte maturation. Mol. Microbiol. *115*, 1005–1024. https://doi.org/10.1111/mmi.14676.

17. Shang, X., Wang, C., Fan, Y., Guo, G., Wang, F., Zhao, Y., Sheng, F., Tang, J., He, X., Yu, X., et al. (2022). Genome-wide landscape of ApiAP2 transcription factors reveals a heterochromatin-associated regulatory network during Plasmodium falciparum blood-

stage development. Nucleic Acids Res. *50*, 3413–3431. https://doi.org/10.1093/nar/gkac176.

18. Toenhake, C.G., Fraschka, S.A.-K., Vijayabaskar, M.S., Westhead, D.R., van Heeringen, S.J., and Bártfai, R. (2018). Chromatin Accessibility-Based Characterization of the Gene Regulatory Network Underlying Plasmodium falciparum Blood-Stage Development. Cell Host Microbe *23*, 557-569.e9. https://doi.org/10.1016/j.chom.2018.03.007.

19. Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172–177. https://doi.org/10.1038/nature12311.

20. Hollin, T., and Le Roch, K.G. (2020). From Genes to Transcripts, a Tightly Regulated Journey in Plasmodium. Front. Cell. Infect. Microbiol. *10*, 801. https://doi.org/10.3389/fcimb.2020.618454.

21. Tripathi, J., Zhu, L., Nayak, S., Stoklasa, M., and Bozdech, Z. (2022). Stochastic expression of invasion genes in Plasmodium falciparum schizonts. Nat. Commun. *13*, 3004. https://doi.org/10.1038/s41467-022-30605-z.

22. Kafsack, B.F.C., Rovira-Graells, N., Clark, T.G., Bancells, C., Crowley, V.M., Campino, S.G., Williams, A.E., Drought, L.G., Kwiatkowski, D.P., Baker, D.A., et al. (2014). A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature *507*, 248–252. https://doi.org/10.1038/nature12920.

23. Modrzynska, K., Pfander, C., Chappell, L., Yu, L., Suarez, C., Dundas, K., Gomes, A.R., Goulding, D., Rayner, J.C., Choudhary, J., et al. (2017). A Knockout Screen of ApiAP2 Genes Reveals Networks of Interacting Transcriptional Regulators Controlling the Plasmodium Life Cycle. Cell Host Microbe *21*, 11–22. https://doi.org/10.1016/j.chom.2016.12.003.

24. Le Roch, K.G., Zhou, Y., Blair, P.L., Grainger, M., Moch, J.K., Haynes, J.D., De La Vega, P., Holder, A.A., Batalov, S., Carucci, D.J., et al. (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. Science *301*, 1503–1508. https://doi.org/10.1126/science.1087025.

25. Bozdech, Z., Llinás, M., Pulliam, B.L., Wong, E.D., Zhu, J., and DeRisi, J.L. (2003). The Transcriptome of the Intraerythrocytic Developmental Cycle of Plasmodium falciparum. PLOS Biol. *1*, e5. https://doi.org/10.1371/journal.pbio.0000005.

26. Lu, X.M., Batugedara, G., Lee, M., Prudhomme, J., Bunnik, E.M., and Le Roch, K.G. (2017). Nascent RNA sequencing reveals mechanisms of gene regulation in the human malaria parasite Plasmodium falciparum. Nucleic Acids Res. *45*, 7825–7840. https://doi.org/10.1093/nar/gkx464.

27. Aurrecoechea, C., Brestelli, J., Brunk, B.P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S., et al. (2009). PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res. *37*, D539-543. https://doi.org/10.1093/nar/gkn814.

28. Davidson, E.H., McClay, D.R., and Hood, L. (2003). Regulatory gene networks and the properties of the developmental process. Proc. Natl. Acad. Sci. *100*, 1475–1480. https://doi.org/10.1073/pnas.0437746100.

29. Whole-genome analysis of mRNA decay in Plasmodium falciparum reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle - PubMed https://pubmed-ncbi-nlm-nih-gov.insb.bib.cnrs.fr/17612404/.

30. Bunnik, E.M., Batugedara, G., Saraf, A., Prudhomme, J., Florens, L., and Le Roch, K.G. (2016). The mRNA-bound proteome of the human malaria parasite Plasmodium falciparum. Genome Biol. *17*, 147. https://doi.org/10.1186/s13059-016-1014-0.

31. Braun-Breton, C., Rosenberry, T.L., and da Silva, L.P. (1988). Induction of the proteolytic activity of a membrane protein in Plasmodium falciparum by phosphatidyl inositol-specific phospholipase C. Nature *332*, 457–459. https://doi.org/10.1038/332457a0.

32. Choudhary, S., and Satija, R. (2022). Comparison and evaluation of statistical error models for scRNA-seq. Genome Biol. *23*, 27. https://doi.org/10.1186/s13059-021-02584-9.

33. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells | Nature Biotechnology https://www-nature-com.insb.bib.cnrs.fr/articles/nbt.2859.

34. Bailey, T.L. (2021). STREME: accurate and versatile sequence motif discovery. Bioinformatics *37*, 2834–2840. https://doi.org/10.1093/bioinformatics/btab203.

35. Quantifying similarity between motifs - PubMed https://pubmed.ncbi.nlm.nih.gov/17324271/.

36. Russell, T.J., De Silva, E.K., Crowley, V.M., Shaw-Saliba, K., Dube, N., Josling, G., Pasaje, C.F.A., Kouskoumvekaki, I., Panagiotou, G., Niles, J.C., et al. (2022). Inhibitors of ApiAP2 protein DNA binding exhibit multistage activity against Plasmodium parasites. PLoS Pathog. *18*, e1010887. https://doi.org/10.1371/journal.ppat.1010887.

Figure 1

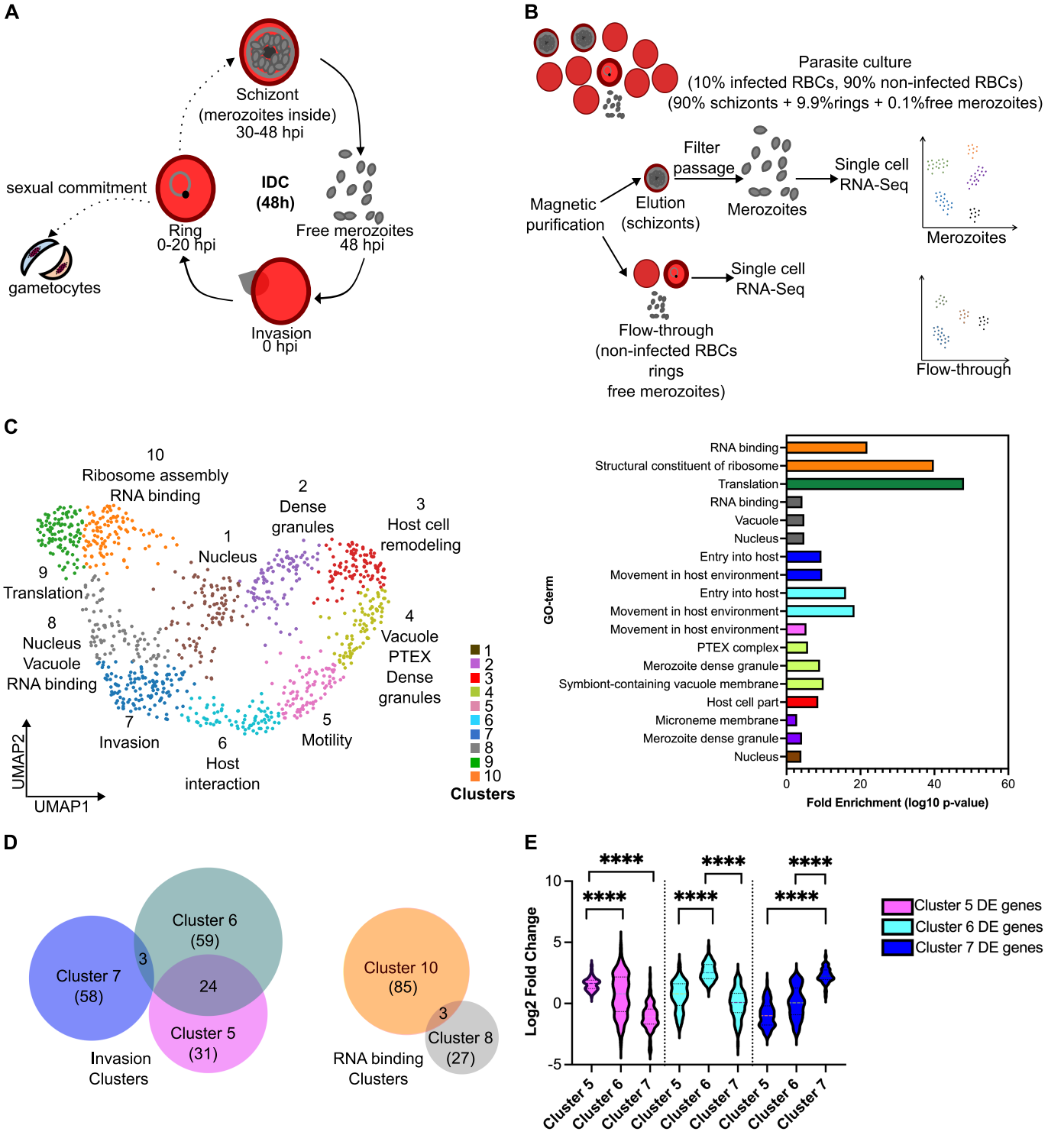
A
Schizont
(merozoites inside)
30-48 hpi
sexual commitment
IDC
(48h)
Ring
0-20 hpi
Free merozoites
48 hpi
gametocytes
Invasion
0 hpi
B
Parasite culture
(10% infected RBCs, 90% non-infected RBCs)
(90% schizonts + 9.9%rings + 0.1%free merozoites)
Filter
passage
Single cell
RNA-Seq
Magnetic
purification
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 and eventually differentiate into gametocytes or continue the IDC, transforming into schizonts, within which new merozoites are formed (numbers below each stage indicate when these events occur post-invasion, hpi). **B**. Parasite cultures are mainly constituted of non-infected RBCs and the one used for the experiment 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 (rings and free merozoites) 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. A scheme of the experimental pipeline is in Figure S1A. The number of genes/UMIs/mitochondrial genes per merozoite or flow through cell is in Figures S1E-F. **C**. UMAP representation showing the merozoite cells grouped by expression cluster (the number of cells in each cluster is in Table S1). 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. The results of the GO-term analysis are in Table S2. A heatmap showing expression of 100 randomly selected genes in the 10 clusters is in Figure S1G. The distribution of the number of genes/UMIs per cell in each cluster is in Figure S1H. **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 Figure S1I. **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

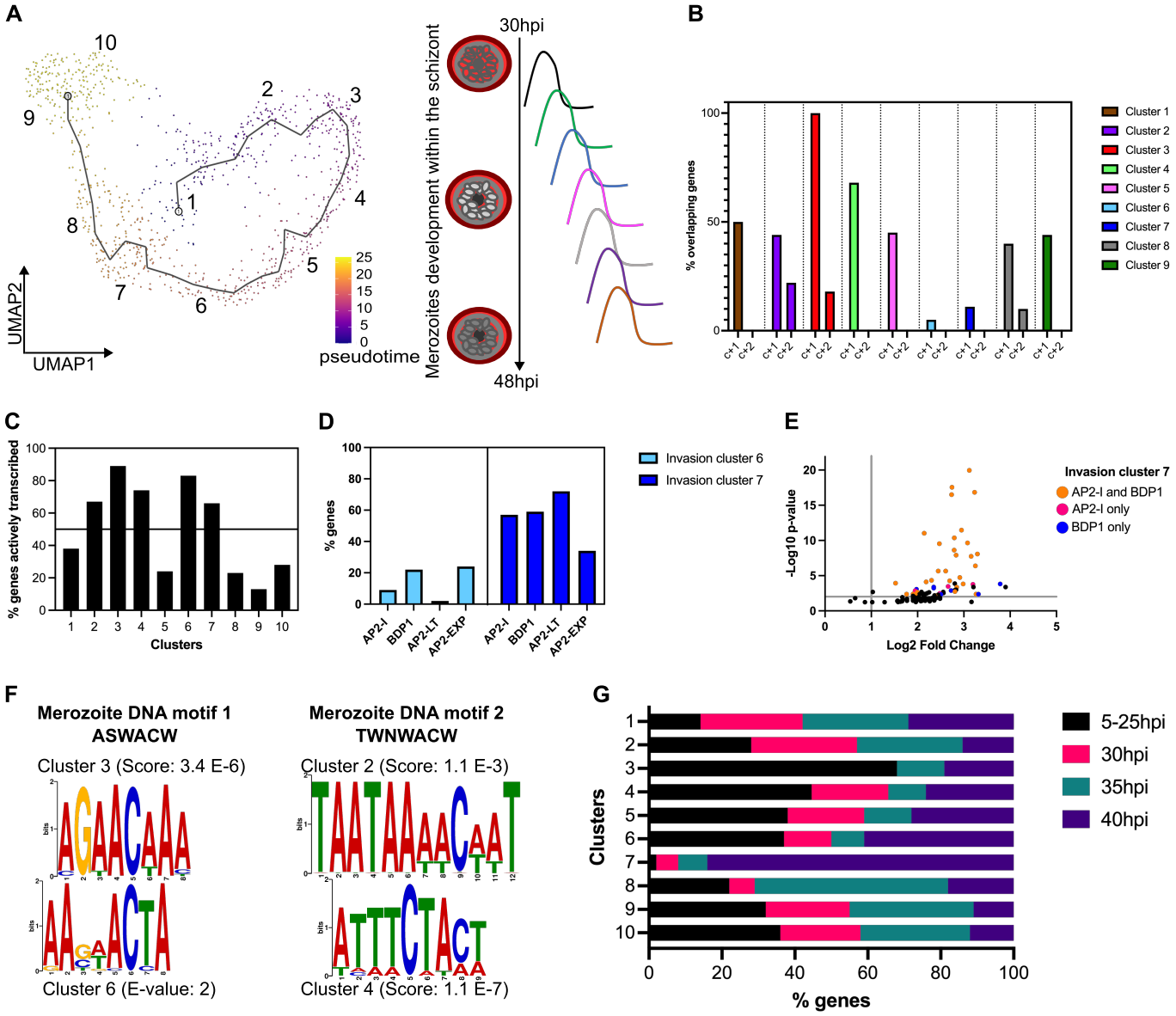
A
10
9
2
3
1
4
8
5
7
6
UMAP2
UMAP1
25
20
15
10
5
0
pseudotime
30hpi
Merozoites development within the schizont
48hpi
B
% overlapping genes
100
50
0
c+1
c+2
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
C
% genes actively transcribed
Clusters
D
% genes
AP2-I
BDP1
AP2-LT
AP2-EXP
Invasion cluster 6
Invasion cluster 7
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 in cascade. The cells pseudotime is in Table S1 and the clusters pseudotime is in Figure 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] found that over 50% of genes of clusters 2, 3, 4, 6 and 7 are actively transcribed between 30h and 48hpi. **D**. Only the invasion genes expressed in cluster 7 are ChIP-Seq targets of AP2-I, BDP1 or AP2-LT[11-14]. **E**. Graph showing that the majority of genes expressed in invasion cluster 7 are ChIP-Seq targets of AP2-I, BDP1 or both. In orange are AP2-I and BDP1 ChIP-Seq targets, in pink AP2-I targets and in blue BDP1 targets. The ChIP-Seq analysis are in Tables S1 and S2. **F**. STREME analysis showing that a DNA motif, here called merozoite DNA motif 1, is present in the promoter region of the genes expressed in 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. Comparison of the merozoite DNA motif 1 with the AP2-G2 DNA-binding motif[15-16] is in Figure S2C and comparison of the merozoite DNA motif 2 with the *in vitro* DNA-binding motif of PF3D7_1239200[10] is in Figure S2E. The percentage of genes bound by AP2-G2 in each invasion cluster is in Figure S2D. The DREME and TOMTOM analysis are in Table S4. **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 Table S2 and the percentage of each category is in Table S1.

Figure 3

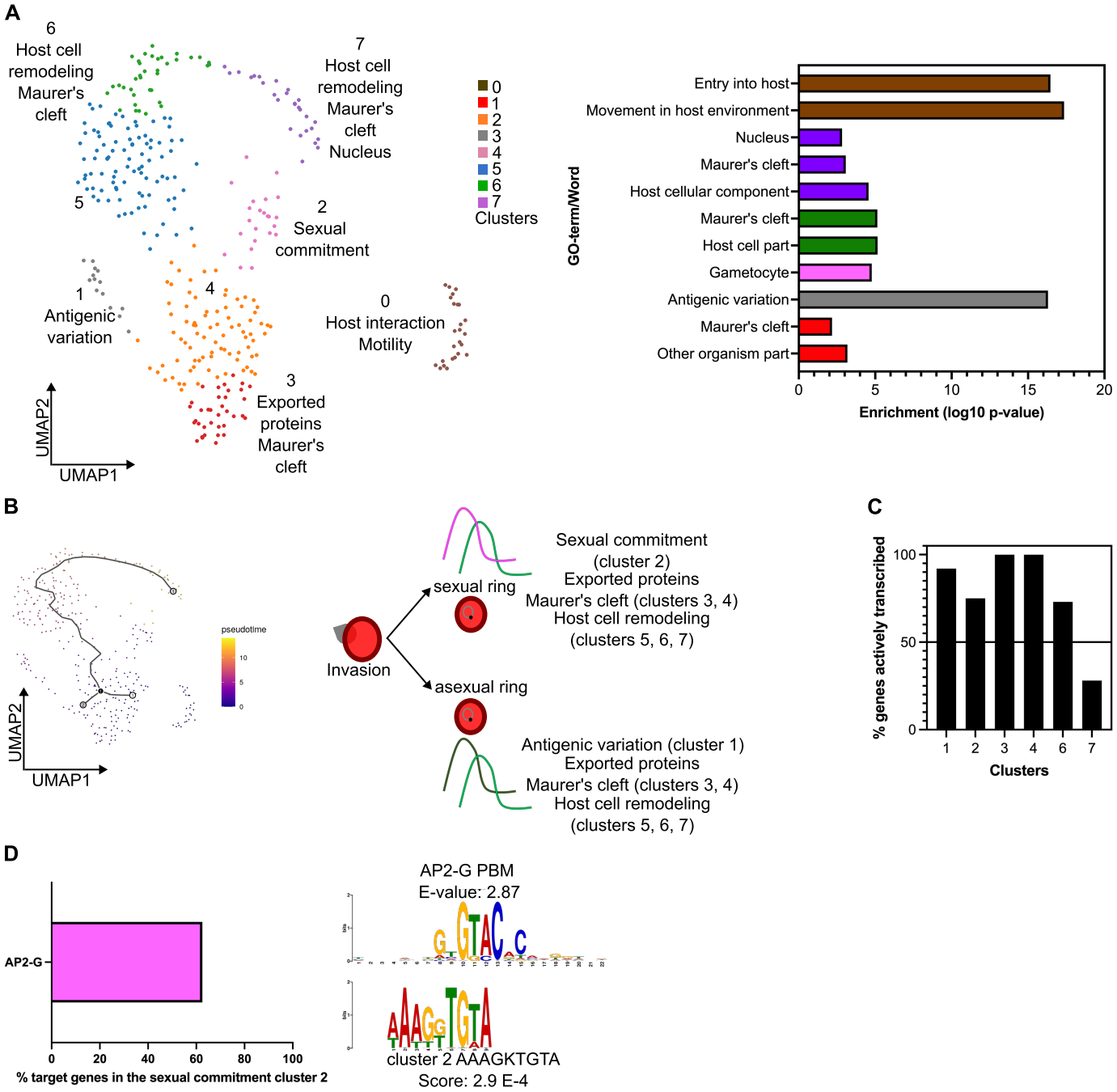
A
6
Host cell
remodeling
Maurer's
cleft
7
Host cell
remodeling
Maurer's
cleft
Nucleus
5
2
Sexual
commitment
1
Antigenic
variation
4
0
Host interaction
Motility
3
Exported
proteins
Maurer's
cleft
UMAP2
UMAP1
0
1
2
3
4
5
6
7
Clusters
GO-term/Word
Entry into host
Movement in host environment
Nucleus
Maurer's 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
Sexual commitment
(cluster 2)
Exported proteins
Maurer's cleft (clusters 3, 4)
Host cell remodeling
(clusters 5, 6, 7)
asexual ring
Antigenic variation (cluster 1)
Exported proteins
Maurer's cleft (clusters 3, 4)
Host cell remodeling
(clusters 5, 6, 7)
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 transcriptional programs 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 Table S1). 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 Table S3. A heatmap showing expression of 100 randomly selected genes in the 8 clusters is in Figure S3A. The distribution of the number of genes/UMIs per cell in each cluster is in Figure S3B. **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 Table S1 and the clusters pseudotime is in Figure S3C. **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 Tables S1 and S3. The DNA motifs enriched in the promoters of the other clusters genes is in Figure S3D. The DREME and TOMTOM analysis of all clusters is in Table S4.

Figure 4

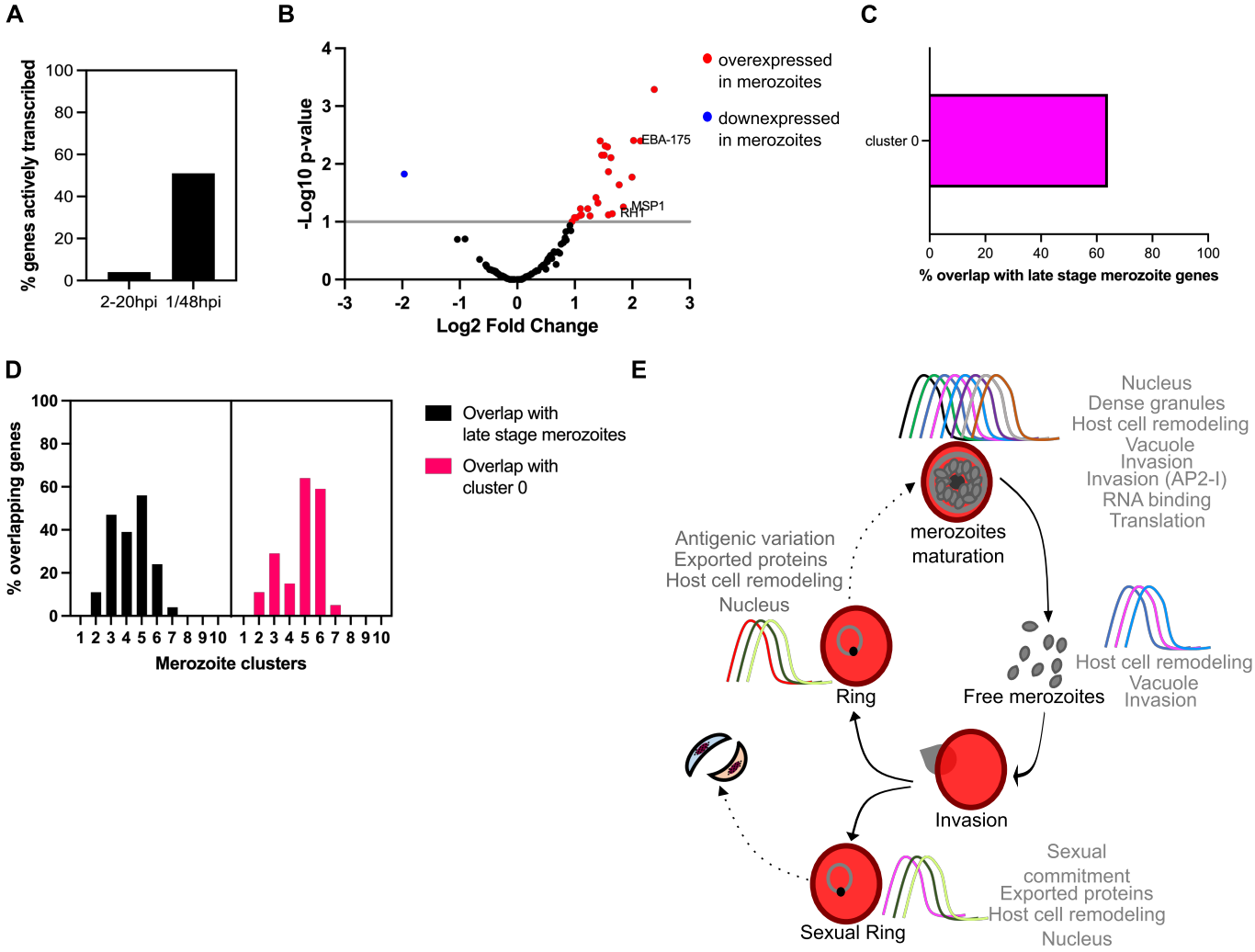
A
% genes actively transcribed
100
80
60
40
20
0
2-20hpi
1/48hpi
B
-Log10 p-value
Log2 Fold Change
EBA-175
MSP1
RH1
overexpressed in merozoites
downexpressed in merozoites
C
cluster 0
% overlap with late stage merozoite genes
D
% overlapping genes
Merozoite clusters
Overlap with late stage merozoites
Overlap with cluster 0
E
Nucleus
Dense granules
Host cell remodeling
Vacuole
Invasion
Invasion (AP2-I)
RNA binding
Translation
merozoites maturation
Antigenic variation
Exported proteins
Host cell remodeling
Nucleus
Ring
Free merozoites
Host cell remodeling
Vacuole
Invasion
Invasion
Sexual commitment
Exported proteins
Host cell remodeling
Nucleus
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. The full list of genes is in Table S5. The GO-term analysis of the genes overexpressed in late merozoite cells is in Figure S4A and Table S5. There was only one gene overexpressed in rings so GO-term analysis were not done. **C**. 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. **D**. 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.