SUPPLEMENTARY INFORMATION

Revisiting the initial steps of sexual development in the malaria parasite *Plasmodium falciparum*

Cristina Bancells, Oriol Llorà-Batlle, Asaf Poran, Christopher Nötzel, Núria Rovira-Graells, Olivier Elemento, Björn F.C. Kafsack and Alfred Cortés

Supplementary Table 1. Observed frequency of mixed plaques and expected frequency if they originated only from multiply-infected erythrocytes in the overlaid culture. Plaque assays with the E5 and E5-HA-DD lines.

	Exp. 2			Exp. 3		
	E5	E5-HA-DD "SCC"	E5-HA-DD "NCC"	E5	E5-HA-DD "SCC"	E5-HA-DD "NCC"
Multiply-infected erythrocytes in the overlaid culture (%)	1.6	2.9	2.9	3.5	3.4	3.4
Observed mixed plaques (%)	**4.0**	**36.6**	**14.7**	**7.0**	**27.0**	**6.0**
Number of plaques analyzed (*n*)	101	101	102	100	100	100
Expected mixed plaques (%)	0.3	1.3	1.4	0.6	1.6	1.7
p value	0.000	0.000	0.000	0.000	0.000	0.000
Observed mixed plaques among Pfs16-positive plaques (%)	**46.7**	**72.3**	**31.1**	**44.0**	**64.0**	**11.0**
Number of plaques analyzed (*n*)	92	101	103	100	100	100
Expected mixed plaques among Pfs16-positive plaques (%)	1.5	2.5	2.0	3.4	2.9	1.7
p value	0.000	0.000	0.000	0.000	0.000	0.000

The proportion of mixed plaques observed is much higher than expected if they originated only from multiply-infected erythrocytes in the overlaid culture. Observed proportions of mixed plaques were obtained by counting ~100 plaques (top) or ~100 gametocyte-containing plaques (bottom) in each sample of each experiment (*n* is the number of plaques included in the analysis after excluding plaques that could not be unambiguously classified during the image analysis in a computer screen). The proportion of multiply-infected erythrocytes in the overlaid schizont cultures was determined by light microscopy analysis of Giemsa-stained smears when parasites were at the ring stage. Flow cytometry analysis gave similar results, but only microscopy values are shown and were used for all calculations. For two independent experiments, the expected proportion of mixed plaques arising from multiply-infected erythrocytes was calculated according to the method described by Bruce *et al.*[1], and the expected proportion of mixed plaques among plaques containing ≥1 Pfs16-positive parasite was determined as described in the Methods section (an accurate determination of the proportion of multiply-infected erythrocytes was not available for another independent biological replicate). *p* values were obtained using a two-tailed Fisher's exact test. In all cases, the proportion of mixed plaques observed was significantly higher than expected from multiply-infected erythrocytes only. Additionally, most plaques arising from multiply-infected

erythrocytes in the overlaid cultures were excluded during the image analysis step because they had more than one hemozoin-containing residual body (see Methods).

Supplementary Table 2. Observed frequency of mixed plaques and expected frequency if they originated only from multiply-infected erythrocytes in the overlaid culture. Plaque assays with the 3D7-Imp. line.

	Exp. 1	**Exp. 2**
Multiply-infected erythrocytes in the overlaid culture (%)	7.4	2.9
Observed mixed plaques (%)	**12.0**	**17.9**
Number of plaques analyzed (*n*)	100	106
Expected mixed plaques (%)	2.9	1.2
p value	0.000	0.000
Observed mixed plaques among Pfs16-positive plaques (%)	**33.0**	**40.6**
Number of plaques analyzed (*n*)	100	101
Expected mixed plaques among Pfs16-positive plaques (%)	6.0	2.5
p value	0.000	0.000

Comparison between the proportion of mixed plaques observed and the proportion expected if they originated only from multiply-infected erythrocytes in the overlaid culture, for experiments with the parasite line 3D7-Imp. The analysis was performed as in Supplementary Table 1. *p* values were obtained using a two-tailed Fisher's exact test. The number of mixed plaques observed was significantly higher than expected from multiply-infected erythrocytes only.

Supplementary Table 3. Single-cell RNA-seq analysis of genes differentially expressed in cluster 13.

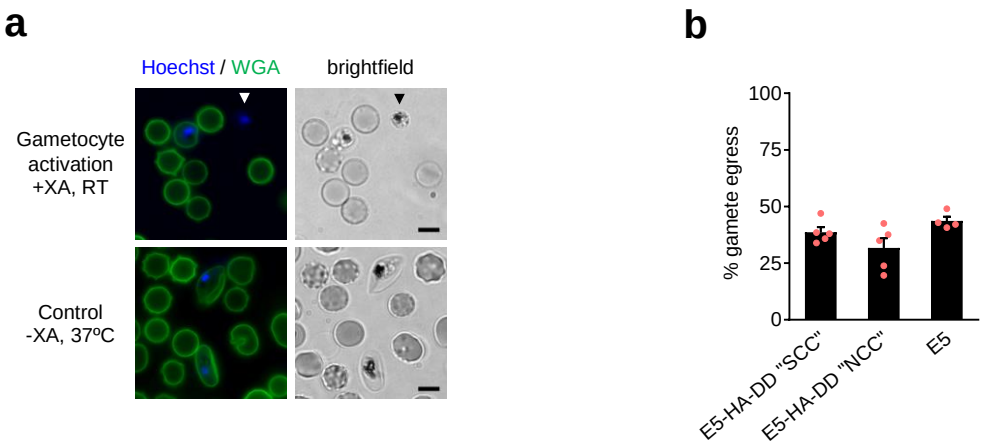
GeneID	Gene Name	Description	% expression in Cycle 2 Cluster 13	% expression outside of cluster 13	% difference in expression	expression fold-change	p.value	Max. Expression blood stage	Peptides detected in T / GC I-II / GC V)
PF3D7_1302100	pfg27	gamete antigen 27/25 (G27/25)	79.5%	10.7%	68.8%	16.6	0	GC II	0/31/29
PF3D7_0423700	etramp4	early transcribed membrane protein 4 (ETRAMP4)	92.3%	30.0%	62.3%	10.7	0	GC V	0/5/2
PF3D7_1467600		conserved Plasmodium protein unknown function	57.8%	7.6%	50.2%	9.4	0	GC II	0/19/0
PF3D7_1102500	gexp02	Plasmodium exported protein (PHISTb) unknown function (GEXP02)	64.5%	9.2%	55.3%	8.5	0	late T, GC V	7/35/24
PF3D7_1476600		Plasmodium exported protein unknown function	38.3%	2.3%	36.0%	8.5	2.07E-259	GC V	no data
PF3D7_1236200		conserved Plasmodium protein unknown function	79.6%	17.7%	61.9%	8.5	0	late T, GC V	no data
PF3D7_1473700	nup116	nucleoporin NUP116/NSP116 putative (NUP116)	55.6%	9.5%	46.1%	6.8	5.14E-268	late T	0/30/7
PF3D7_0214300		conserved Plasmodium protein unknown function	50.3%	10.8%	39.5%	5.4	1.45E-200	GC V	0/1/24
PF3D7_1466200		conserved Plasmodium protein unknown function	68.2%	22.1%	46.1%	4.8	9.84E-253	GC V	0/5/1
PF3D7_1362700		conserved Plasmodium protein unknown function	68.7%	27.0%	41.7%	4.7	3.78E-244	GC V	0/12/0

Differentially expressed genes between cycle 2 cluster 13 cells (1,003 cells) and cycle 1 and cycle 2 cells outside of cluster 13 (6,285 cells). The criteria for different expression was a minimum 4.5-fold change in average expression and 35% difference in the fraction of cells with detectable transcript. *p* values were calculated using a likelihood-ratio test for single-cell gene expression[2], not adjusted for multiple comparisons. The last two columns show expression of the genes in previous transcriptomic[3] or proteomic[4] analysis of asexual stages and gametocytes. “T” refers to trophozoites, “GC” refers to gametocytes.

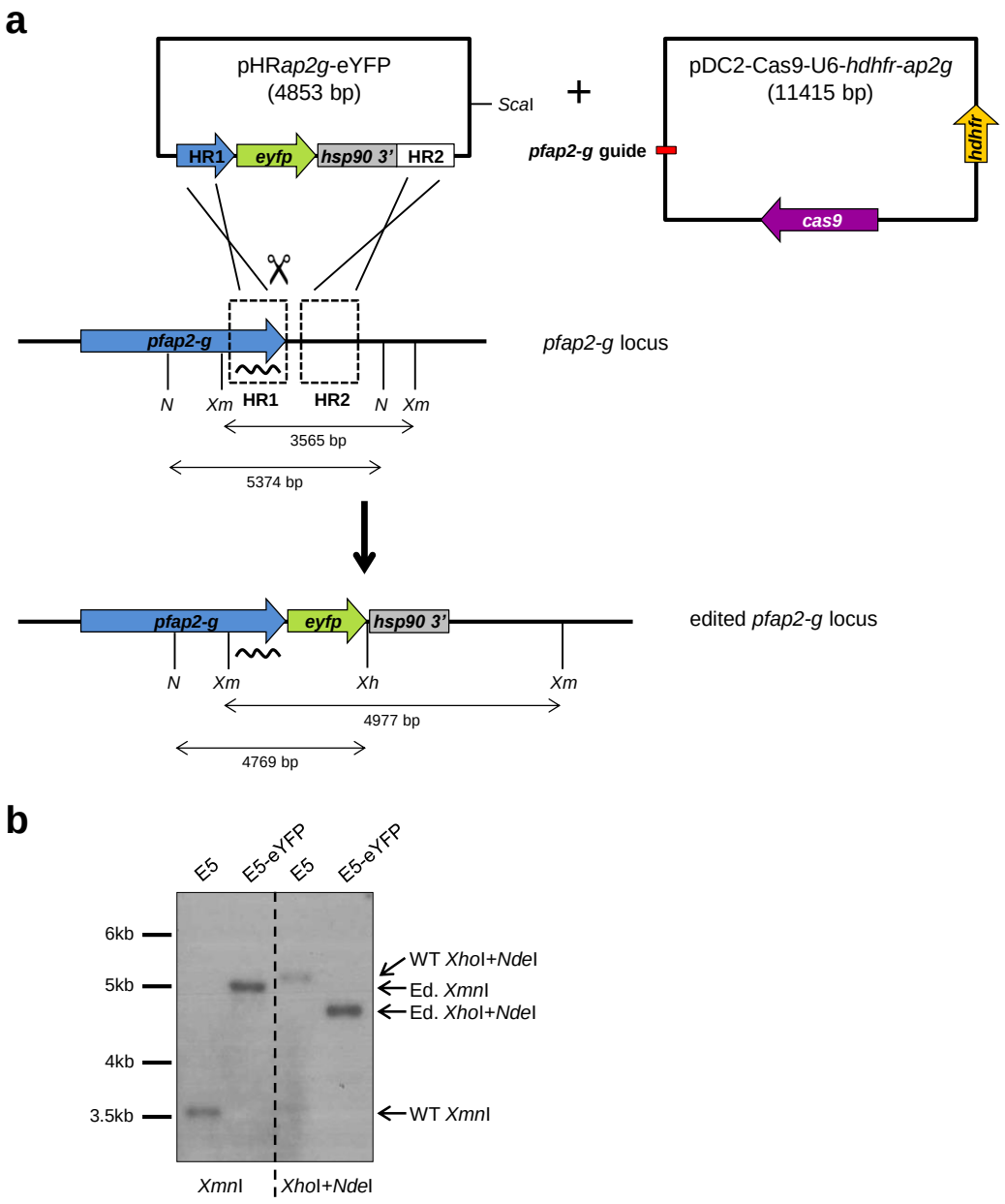
Supplementary Table 4. Primers used in this study.

Primer Name	Sequence 5' – 3'
Primers used for real time PCR	
PF3D7_1222600_F1 (*pfap2-g*) *	AACAACGTTCATTCAATAAATAAGG
PF3D7_1222600_R1 (*pfap2-g*) *	ATGTTAATGTTCCCAAACAACCG
PF3D7_0717700_F (*serrs*) [5]	AAGTAGCAGGTCATCGTGGTT
PF3D7_0717700_R (*serrs*) [5]	TTCGGCACATTCTTCCATAA
PF3D7_0812600_F (*uce*) [# 6]	GGTGTTAGTGGCTCACCAATAGGA
PF3D7_0812600_R (*uce*) [# 6]	GTACCACCTTCCCATGGAGTA
PF3D7_0406200_F (*pfs16*) [7]	TCAGGTGCCTCTCTTCATGCT
PF3D7_0406200_R (*pfs16*) [7]	GCTGAGTTTCTAAAGGCATTTTGTC
PF3D7_1477300_F (*pfg14.744*) [7]	GATGTACCGAAGTATGAGAATGATT
PF3D7_1477300_R (*pfg14.744*) [7]	TGGATAACGGCAAGGATATTTCTT
PF3D7_1302100_F (*pfg27*) [7]	GAAGCGTATCATGAACGACAAGA
PF3D7_1302100_R (*pfg27*) [7]	CTTATTCTTGCTGCTGCGTC
PF3D7_0936600_F (*pfgexp5*) *	GGATGCAAGTCCGAGGTAG
PF3D7_0936600_R (*pfgexp5*) *	TATGTGTACATGATTCCATTGG
Primers used to generate the E5-eYFP transgenic line	
pfap2g_HR1_F	ttgtta**actagt**GAACTTCAAAATGTGAATAATG
pfap2g_HR1_R	ccacaa**cttaag**gATgTTtCTGTTGTTTCCCCCTTTGTG
pfap2g_HR2_F	ttgtca**gaattcatcgat**CATCTTGTGACATTTTTAAATA
pfap2g_HR2_R	gaacaa**ccatgg**CACATATACATGTATGATTACA
pfap2g_gRNA_IF_F	taagtatataatattACCACAAAGGGGGAAACAACgttttagagctagaa
pfap2g_gRNA_IF_R	ttctagctctaaaacGTTGTTTCCCCCTTTGTGGTaatattatatactta
eyfp_IF_F	aacaacagaaacatc**agatct**gcagcagcagcagcagcaGTGAGCAAGGGCGAGGAGC
hsp90 3'_IF_R	aaatgtcacaagatg**gaattc**TATTTGATGAATTAACTACACTTA

All the primers for quantitative (real-time) PCR were used at 200 nM except those indicated with the symbols * (500 nM) and # (250 nM). A reference is provided for primers that had been described before. Restriction sites are indicated in bold and are underlined. Upper case indicates the part of the primer hybridizing with its target sequence.

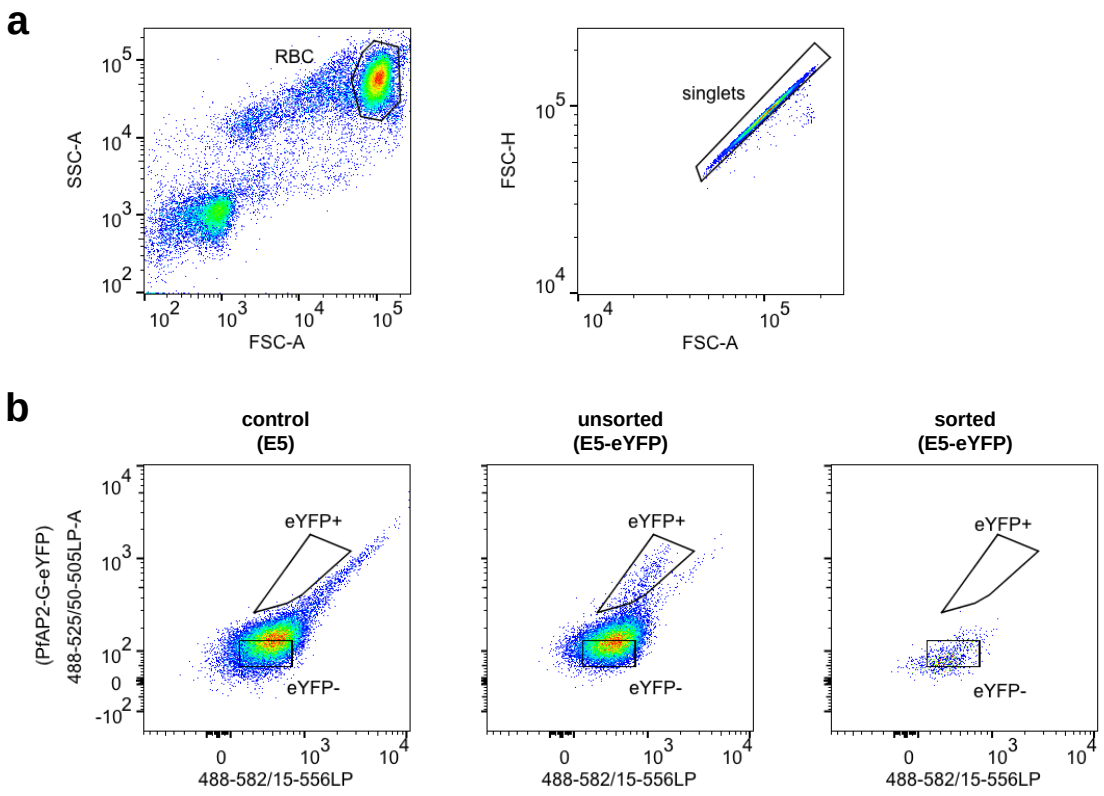


Supplementary Figure 1. Gametocytes formed by either the SCC or the NCC routes have similar gamete egress efficiency. a, Representative images of gametocyte activation and egress assays. Hoechst (blue) stains parasite DNA, whereas WGA-Oregon Green 488 (green) marks erythrocyte membranes. +XA, RT refers to xanthurenic acid addition and low temperature (room temperature) used to activate the gametocytes, whereas -XA, 37ºC refers to control conditions. Activation of gametocytes was evidenced by rounding up (observed in all samples), whereas egress was detected by absence of WGA surface staining (arrowhead). Under control conditions, gametocytes were elongated and did not egress. Images are representative of five independent experiments, with at least two independent samples each. Scale bar, 5 µm. **b,** Quantification of gamete egress efficiency for E5-HA-DD cultures under conditions that result in sexual conversion either via the SCC route or mainly via the NCC route (see Methods), and for the wild type parasite line E5. Parasites were scored as egressed gametes if they were round-shaped and free of WGA signal. 50 to 150 gametocytes were counted for each experiment. Values are the average of five (E5-HA-DD) or four (E5) independent biological replicates, with S.E.M. Red dots are individual data points. Differences in egress efficiency were not significant between any of the parasite lines or conditions (using a two-sided t-test with equal variance).

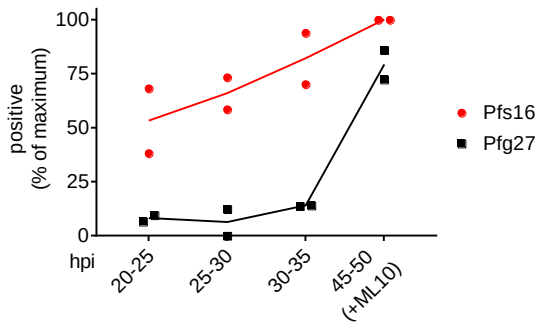


Supplementary Figure 2. Generation of parasites expressing endogenous PfAP2-G fused to eYFP using CRISPR-Cas9 technology. a. Schematic (not to scale) of *pfap2-g* tagging with eYFP using the CRISPR-Cas9 system. The plasmid pDC2-Cas9-U6-*hdhfr-ap2g* was generated by inserting a 20 bp *pfap2-g* guide obtained by annealing complementary oligonucleotides pfap2g_gRNA_IF_F and pfap2g_gRNA_IF_R (all primers are described in Supplementary Table 4) into the *Bbs*I digested plasmid pDC2-Cas9-U6-h*dhfr* [8] using the In-Fusion system (Takara). The scissors indicate the Cas9 cleavage site. The *pfap2-g* homology regions in plasmid pHR*ap2g*-eYFP were PCR-amplified from 3D7 genomic DNA using primer sets pfap2g_HR1_F and pfap2g_HR1_R (HR1) and pfap2g_HR2_F and pfap2g_HR2_R (HR2). Primer pfap2g_HR1_R included a recodonized sequence encoding

the amino acids between the cleavage site and the *pfap2-g* stop codon. The GFP homology regions from plasmid pL6-eGFP-yFCU[9] were sequentially replaced by *pfap2-g* HR1 (using *Spe*I-*Afl*II *sites*) and *pfap2-g* HR2 (*Nco*I-*Eco*RI sites). The resulting plasmid was digested with *Afl*II and *Eco*RI to replace the *hdhfr* gene by a fragment containing a *eyfp-hsp90 3'* sequence, obtained by PCR-amplification from plasmid pDC2-*ap2g*-eYFP (a gift from Manuel Llinás, The Pennsylvania State University, USA) using primers eyfp_IF_F and hsp90 3'_IF_R. The fragment was cloned in frame using the In-Fusion system. Finally, the U6 promoter-terminator guide cassette was excised by sequentially digesting with *Nco*I and *Aat*II, end-blunting using T4 DNA polymerase, and religation. Plasmid pHR*ap2g*-eYFP was linearized with *Sca*I before transfection. The positions of *Xmn*I (*Xm*), *Xho*I (*Xh*) and *Nde*I (*N*) restriction sites used for Southern blot analysis are indicated. The undulating line indicates the position of the probe used for Southern blot, which was PCR amplified using primers pfap2g_HR1_F and pfap2g_HR1_R. **b.** Southern blot analysis of E5 parasites transfected with plasmids pHR*ap2g*-eYFP and pDC2-Cas9-U6-*hdhfr-ap2g* and selected with WR99210 (E5-eYFP line). Parental E5 parasites were also analyzed. "Ed. *Xmn*I*"* and "Ed. *Xho*I+*Nde*I" indicate the position of the bands expected for the correctly edited *pfap2-g* locus in genomic DNA digested with the respective enzymes (4,977 and 4,769 bp, respectively). "WT *Xmn*I*"* and "WT *Xho*I+*Nde*I" indicate the position of the bands expected for the intact *pfap2-g* locus (3,565 and 5,374 bp, respectively). This analysis revealed that the *pfap2-g* locus was edited in essentially all E5-eYFP parasites, as bands corresponding to the wild type locus were not observed. A single Southern blot experiment was performed. Edition of the *pfap2-g* locus was also confirmed by PCR (data not shown).

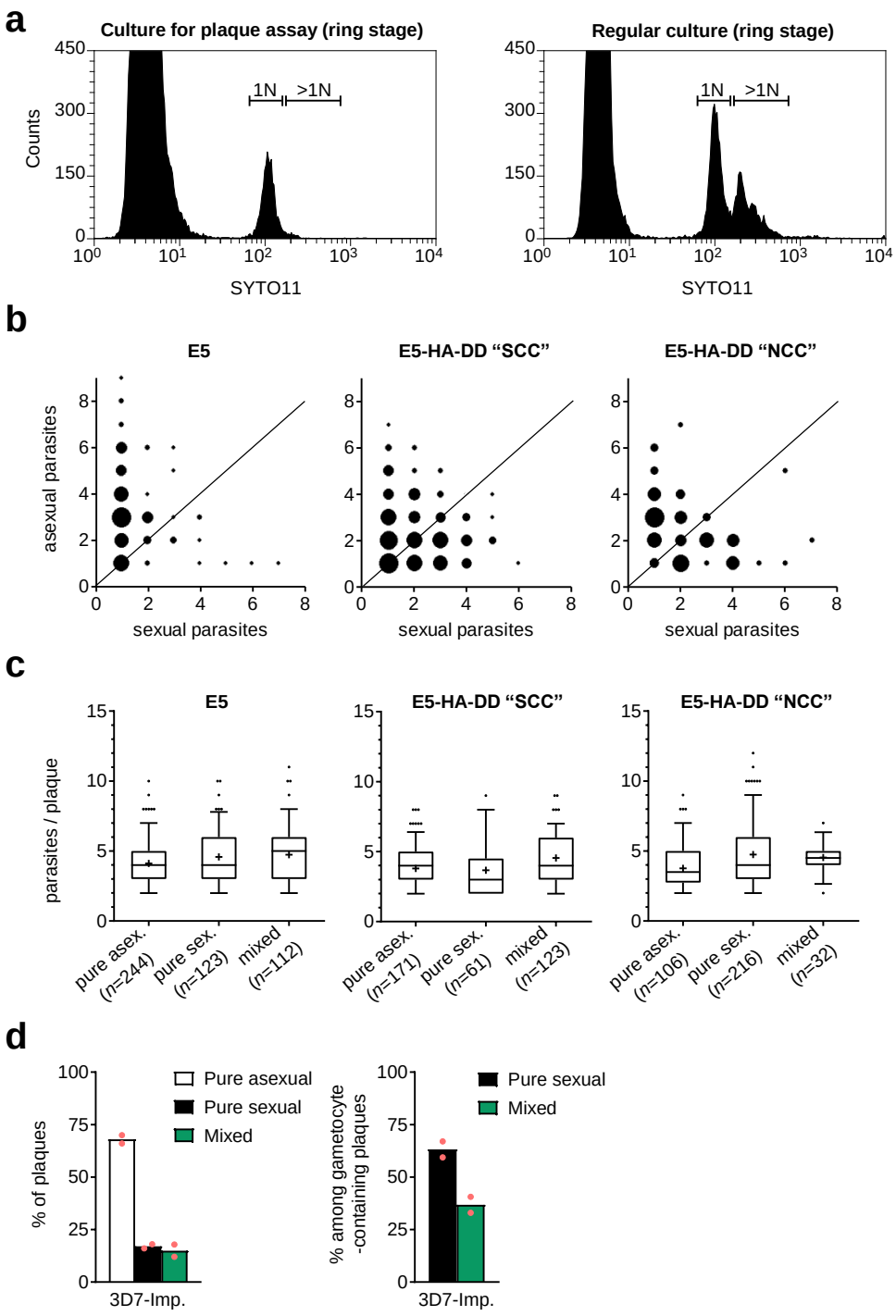


Supplementary Figure 3. FACS-sorting of PfAP2-G-negative schizonts. a, Representative images of the initial gates used to sort PfAP2-G-negative parasites. First, the erythrocytes (RBC) population was identified and gated on SSC-A vs FSC-A plots. Next, singlets were selected on FSC-A vs FSC-H plots. **b,** Representative images of 488-525/50-505LP-A (PfAP2-G-eYFP) vs 488-582/15-556LP (autofluorescence) plots. The plot on the left (control) corresponds to an E5 culture maintained and processed in parallel with E5-eYFP cultures, whereas the plot in the middle (unsorted) is an E5-eYFP culture before sorting and the plot on the right (sorted) is the same E5-eYFP culture after sorting eYFP-negative parasites. The gates used to identify eYFP-positive parasites, which are absent from E5 cultures, and the gate used to sort eYFP-negative parasites, are shown. The latter was defined conservatively to collect only parasites with very low eYFP signal. In the culture shown in this representative example, 1.8% of parasites were scored as eYFP-positive and 51.5% fell within the eYFP-negative gate. In all experiments, none of the sorted eYFP-negative parasites were eYFP-positive (0.0%). The plots shown in panels a and b are representative of eight independent biological replicates (independent sorting experiments). We attempted sorting of eYFP-positive parasites but the yield was very low and the purity never approached 100% (typically 60-70%); sorted eYFP-positive parasites were not used for any of the experiments presented in this article.



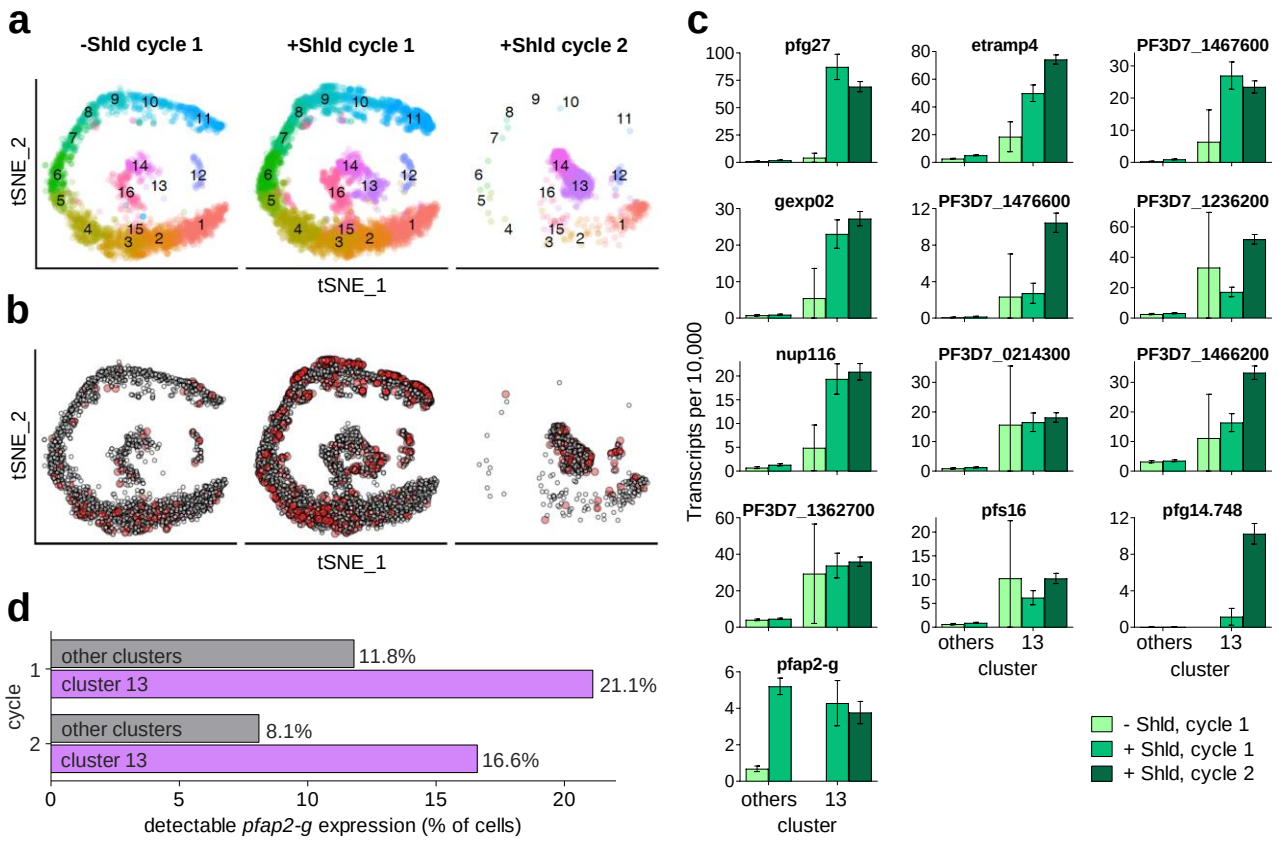
Supplementary Figure 4. Validation of the temporal expression dynamics of the early gametocyte markers Pfs16 and Pfg27. Time-course analysis of the expression of Pfs16 and Pfg27 in E5 cultures synchronized to a 5 h age window. The proportion of positive parasites was determined by IFA with antibodies against Pfs16 and Pfg27 at the times indicated in h post-invasion (hpi). 45-50 hpi samples were treated with ML10 to prevent reinvasion. Values are expressed relative to the highest proportion of positives in each experiment (i.e., Pfs16 at the latest time point). Individual data points and average (line) of two independent biological replicates are shown. For each time point and experiment, >1,000 parasites were scored as positive or negative for each marker. Double staining of the two markers was not possible because both antibodies available were raised in mice. Gametocytes formed at previous cycles are expected to be positive for the two markers from the first time point. They occur at low density because cultures are regularly diluted (asexual parasites multiply every 48 h whereas gametocytes are non-replicative). The temporal pattern observed indicates that the onset of Pfs16 expression in new gametocytes typically occurs between 20-35 hpi or even earlier, whereas Pfg27 expression starts at 35-45 hpi. However, the similar proportion of positive parasites between both markers at the latest time point indicates that sufficiently-mature gametocytes express Pfs16 and Pfg27 and that both are appropriate gametocyte markers, as previously reported[10,11]. While both antibodies clearly mark gametocytes, anti-Pfs16 antibodies show a better signal-to-background ratio, which together with earlier expression makes Pfs16 a superior marker for experiments such as plaque assays. The specificity of anti-Pfs16 antibodies was further demonstrated by experiments with the E5-HA-DD line cultured in the absence of Shld. No single Pfs16-positive parasite was ever observed in these PfAP2-G-deficient cultures (0 Pfs16-positives out of >1,000 parasites scored, and we never observed a Pfs16-positive parasite in such cultures in multiple experiments that were not analyzed quantitatively). In contrast, anti-

Pfg27 antibodies stain a small subset of schizonts (<0.5%) with a signal of similar intensity to gametocytes.



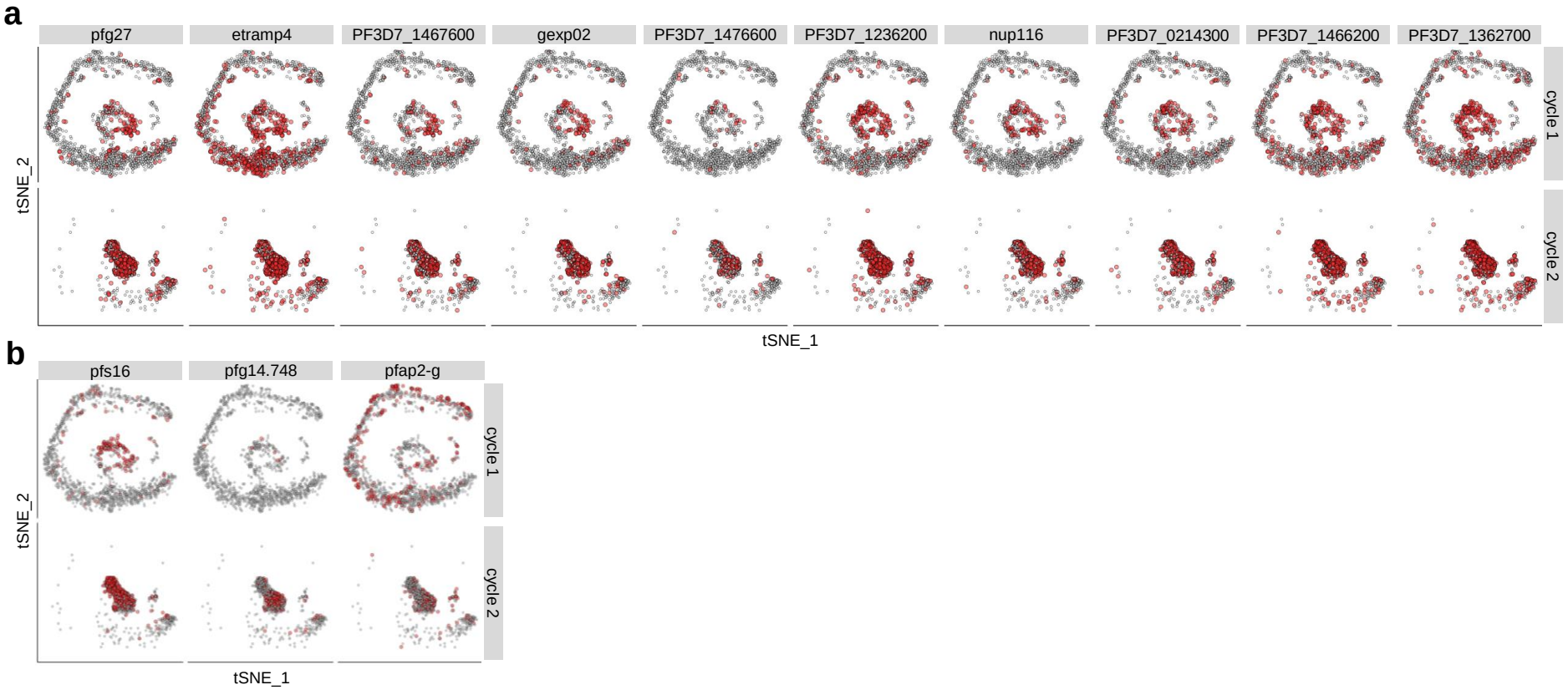
Supplementary Figure 5. The vast majority of mixed plaques do not originate from multiply-infected erythrocytes. a, Representative example of a flow cytometry analysis of synchronized cultures at the ring stage used to generate the overlaid schizont cultures for plaque assays. Cultures at the same stage prepared under regular conditions are shown for comparison. 1N indicates the position of the peak corresponding to single-infected

erythrocytes, whereas >1N are multiply-infected erythrocytes. Data analysis was performed using Summit v4.3 software. The plots are representative of two independent experiments (with two independent samples each). The comparison of the proportion of mixed plaques observed with the proportion expected if they originated only from multiply-infected erythrocytes in the overlaid cultures is presented in Supplementary Tables 1 and 2. The flow cytometry quantification of 1N and >1N cells was not used for any of the calculations shown in these Supplementary Tables. **b**, Distribution of sexual and asexual parasites in mixed plaques. Each dot represents a number of mixed plaques proportional to the size of the dot. The distribution observed roughly reflects the sexual conversion rate of each parasite line: the E5 line shows a clear predominance of asexual parasites in its mixed plaques, whereas the E5-HA-DD line shows a similar abundance of sexual and asexual parasites. This is consistent with the lower sexual conversion rate of E5 compared to E5-HA-DD (Fig. 2c in the main article). If mixed plaques originated mainly from multiply-infected erythrocytes, the distribution of sexual and asexual parasites within mixed plaques would be similar for all parasite lines. Plots include data from three independent biological replicates. **c**, Number of parasites in pure and mixed plaques. The number of parasites is similar in the different types of plaques: including all experiments with the different parasite lines there were an average of 3.94 parasites/plaque in pure asexual plaques, 4.53 in pure sexual plaques, and 4.63 in mixed plaques. Although the difference between pure asexual and other types of plaques was statistically significant (p=0.000 using a two-tailed unpaired t-test with unequal variance), this difference is of low magnitude and unlikely to have any biological significance. *n* indicates the number of plaques included in the analysis (plaques are from three independent experiments). Boxes show median and interquartile range, whereas whiskers are the 5-95 percentiles, "+" indicates the mean, and dots are outliers. **d**, Plaque assays performed with the parasite line 3D7-Imp. (the 3D7 stock at Imperial College)[12,13]. Experiments were performed and quantified as the experiments in Fig. 2c-d in the main article. Additional analysis of these experiments is presented in Supplementary Table 2. At least 100 plaques (left bar chart) or 100 plaques containing ≥1 Pfs16-positive sexual parasite (right bar chart) were counted in each experiment. Values are the average of two independent experiments, with individual data points (red dots).

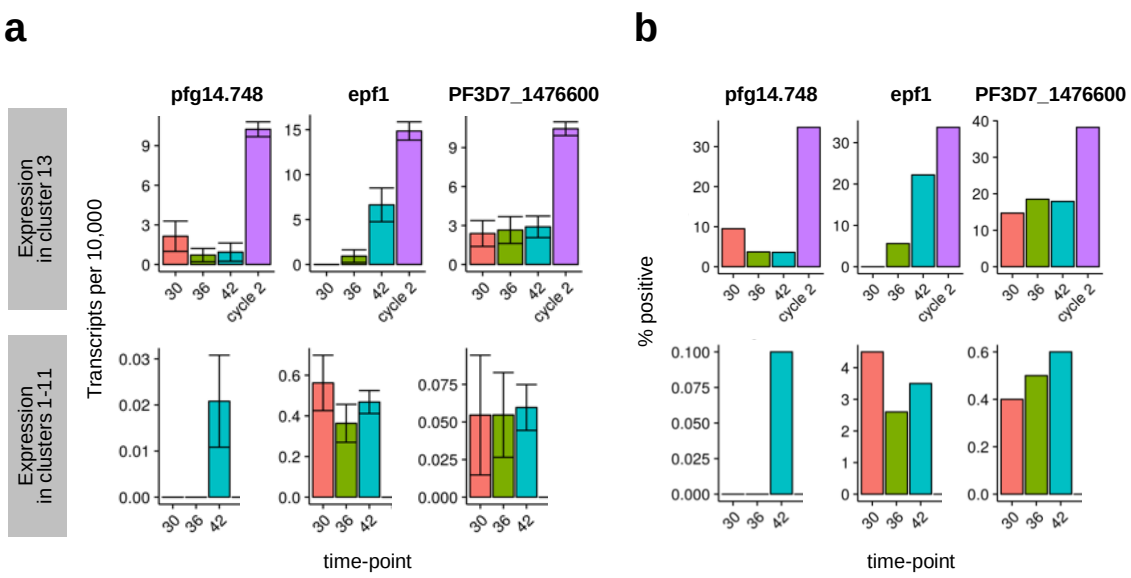


Supplementary Figure 6. Single-cell RNA-seq reveals a cluster of cells that corresponds to stage I gametocytes. a, Cluster analysis of single-cell RNA-seq data. Panels a and b show 2D representations of the single-cell transcriptomes using the tSNE algorithm. This non-linear dimensionality reduction algorithm aims to preserve similarity in the high-dimensional space and represent it as proximity in this 2D plot. The axes represent the two dimensions; however, they do not directly represent transcriptional features. Single-cell RNA-seq was performed on E5-HA-DD samples collected at ~30, ~36 or ~42 hpi of the same cycle of treatment with either solvent control (left panel, -Shld cycle 1, 3,037 cells) or Shld (middle panel, +Shld cycle 1, 5,736 cells), or samples collected at ~42 hpi of the next cycle (right panel, +Shld cycle 2, 1,736 cells). All samples were magnet-purified before the analysis to remove ring-stage parasites and retain only gametocytes and pigmented trophozoites/schizonts. Cycle 2 samples were treated with GlcNAc after reinvasion[14]. The majority of cells from cycle 2 (containing mainly stage I gametocytes after GlcNAc treatment and magnet purification) fall within clusters 13 or 14. Cluster 13 also contains cells from cycle 1 cultures treated with Shld, but virtually no cells from untreated cycle 1 cultures. **b,** Expression of *pfap2-g* (shown in red) in the different samples. The *pfap2-g* transcripts are detected in a significantly lower proportion of cells and only at earlier trophozoite time-points

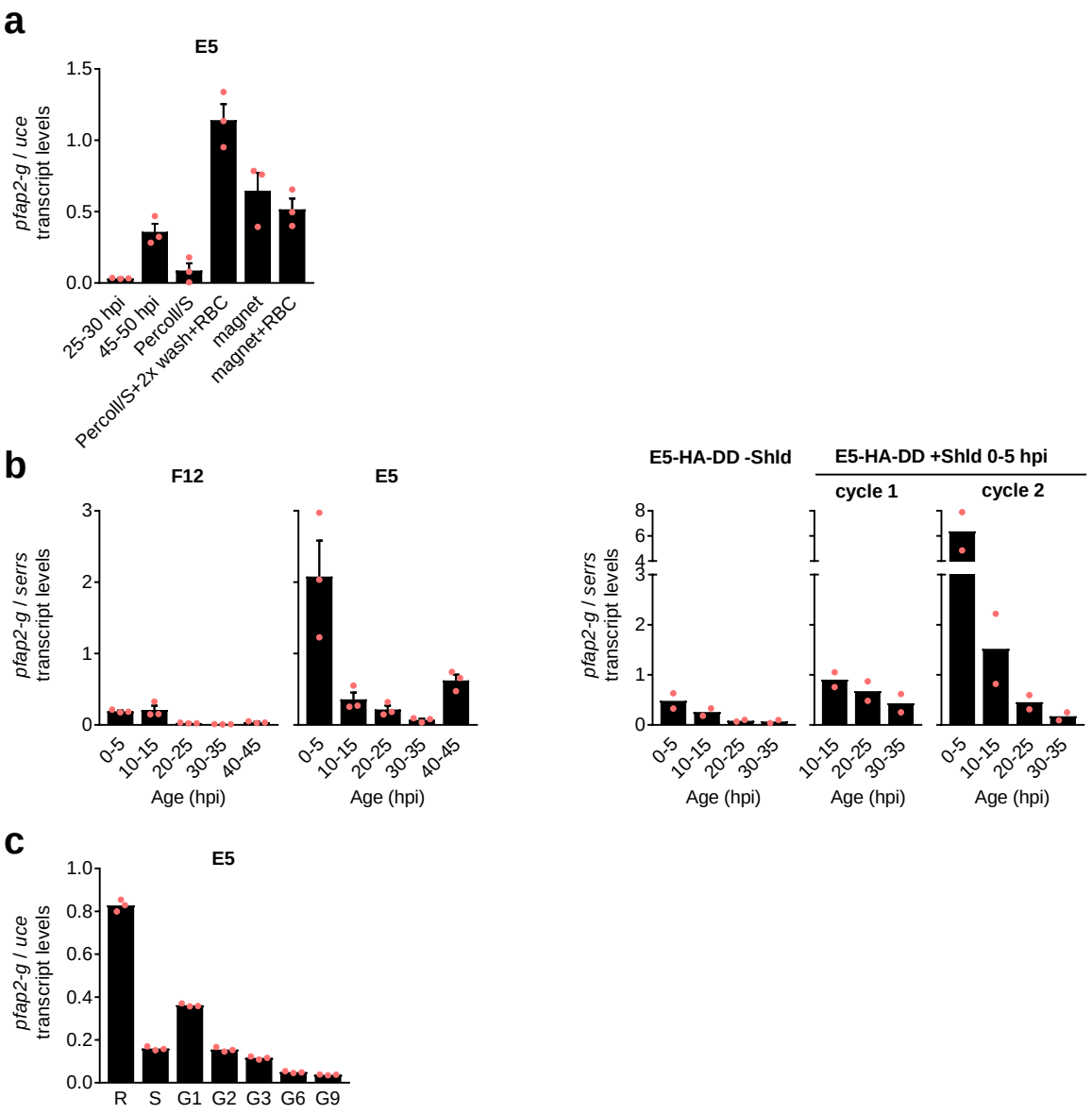
in untreated cultures. The same cells as in panel a are shown. **c,** Expression of genes most highly expressed in cluster 13 cycle 2 cells compared to non-cluster 13 cells (also see Supplementary Table 3 and Supplementary Fig. 7) and additional selected markers. Average expression in cluster 13 or other clusters is shown for untreated cycle 1 cells (others: 3,030 cells, cluster 13: 7 cells), Shld-treated cycle 1 cells (others: 5,552 cells, cluster 13: 184 cells) or Shld-treated cycle 2 cells (cluster 13: 1,003 cells). The very few untreated cells that fall within cluster 13 (7 cells) express some cluster 13 markers but have very low levels of transcripts for some established gametocytes genes such as *pfg27*, *pfg14.748* or *pfap2-g*. This can explain the absence of gametocytes in untreated cultures in spite of containing a few cells that fall within cluster 13. Expression is normalized to 10,000 transcripts per cell. Mean and 95% confidence intervals are shown. **d,** Percentage of Shld-treated cells with detectable *pfap2-g* transcripts from cycle 1 and from cycle 2 is shown for cluster 13 (stage I gametocytes; 184 cells for cycle 1, 1,003 cells for cycle 2) and for other clusters (5,552 cells for cycle 1, 733 cells for cycle 2).



Supplementary Figure 7. Single-cell RNA-seq analysis of genes differentially expressed in cluster 13 and other gametocyte markers. a, Expression of genes differentially expressed between cycle 2 cluster 13 cells and cycle 1 and cycle 2 cells outside of cluster 13. Differentially expressed genes are described in Supplementary Table 3. Expression is shown for Shld-treated cells in cycle 1 (top row) and cycle 2 (bottom row, composed mainly of stage I gametocytes). Equal numbers of randomly sampled cells (1,736 cells) are shown for cycle 1 and cycle 2. Cells with detectable transcripts are shown in red. See Supplementary Figure 6a legend for a detailed definition of tSNE axes. **b,** Same analysis for additional selected markers.

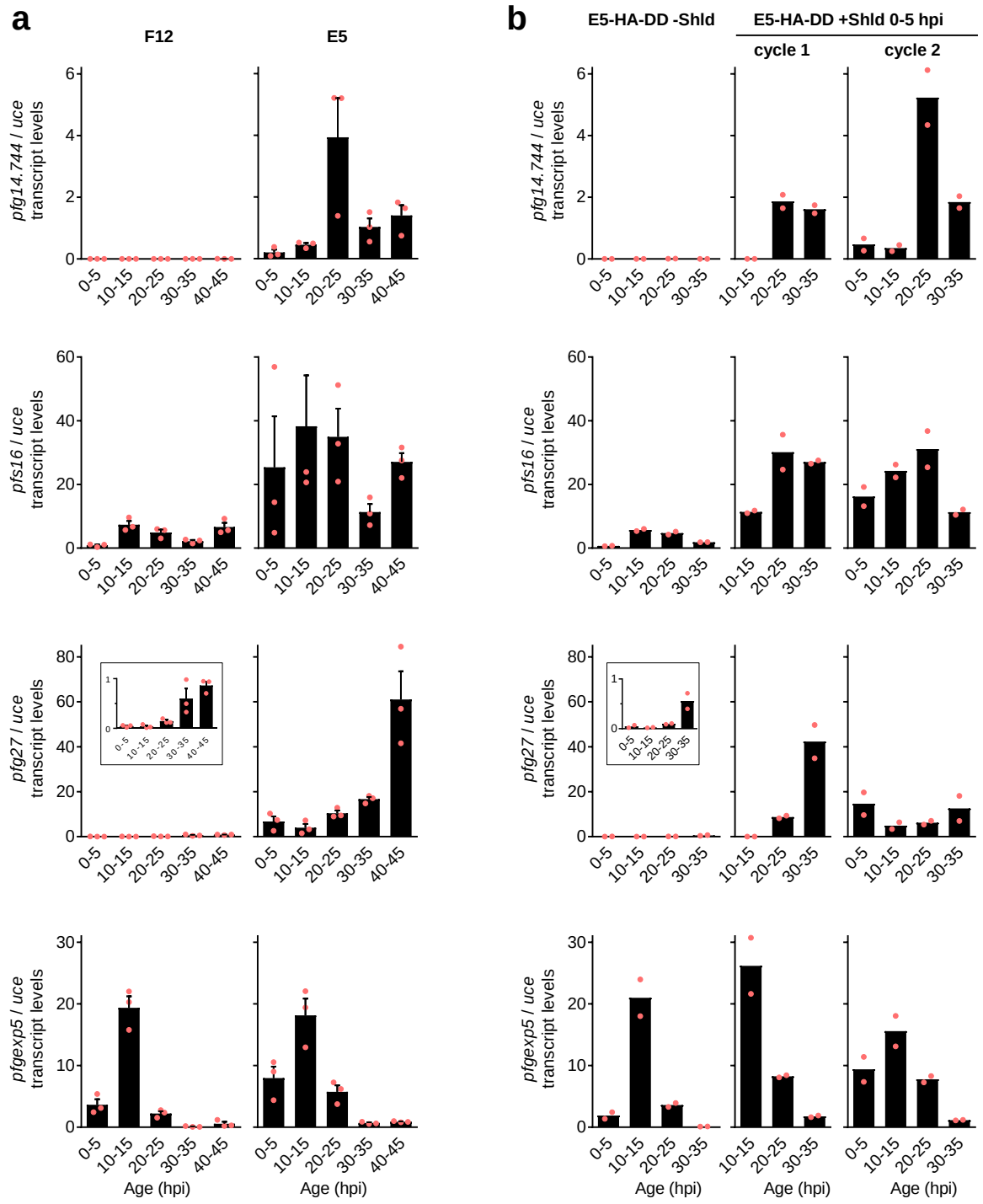


Supplementary Figure 8. Genes differentially expressed between cluster 13 cells from cycle 1 and cycle 2. The criteria to identify differentially expressed genes were a minimum 3-fold change in average expression and 20% difference in the fraction of cells with detectable transcripts. IDs: PF3D7_1477700 (*pfg14.748*); PF3D7_0114000 (*epf1*). Only genes showing higher expression in cycle 2 gametocytes are presented because the majority of genes showing higher expression in cycle 1 gametocytes are among the most abundantly transcribed genes in asexual parasites and their higher number of transcripts in cycle 1 gametocytes may be attributable to technical issues of single-cell RNA-seq. **a,** Average expression of differentially expressed genes in Shld-treated cells within cluster 13 from cycle 1 (time-points ~30 hpi: 41 cells; ~36 hpi: 47 cells; and ~42 hpi: 96 cells) and cycle 2 (1,003 cells) (top), or within clusters 1-11 cycle 1 (~30 hpi: 508 cells; ~36 hpi: 821 cells; and ~42 hpi: 3,356 cells) (bottom). Values are average number of transcripts per 10,000 transcripts, with S.E.M. For all genes, expression in cluster 13 was significantly higher in cycle 2 cells than in cycle 1 cells collected at either time point, using a two-sided t-test (*pfg14.748*: p=3x10^{-8}, 8x10^{-28} and 2x10^{-21} for 30, 36 and 42 h, respectively; *epf1*: p=5x10^{-44}, 2x10^{-25} and 2x10^{-4}; PF3D7_1476600: p=9x10^{-10}, 4x10^{-9}, 8x10^{-13}). **b,** Proportion of cells with detectable transcripts of the differentially-expressed genes.



Supplementary Figure 9. Analysis of *pfap2-g* relative transcript levels at different stages of development. a, Reverse transcription-quantitative PCR (RT-qPCR) analysis of *pfap2-g* transcript levels in mature schizonts. RNA was prepared from tightly synchronized E5 cultures at 25-30 and 40-45 hpi. At the latter time point the culture contains abundant rings in addition to schizonts, which potentially may account for the *pfap2-g* transcripts. Schizonts were purified from 40-45 hpi cultures using Percoll-sorbitol 80-60-40% gradients (Percoll/S) or using magnetic columns (magnet) before Trizol extraction of RNA. Percoll-sorbitol gradients were used instead or regular Percoll purification (see Methods) to obtain increased purity (~99%), but similar results were obtained when using regular Percoll purification (not shown). We also analyzed samples in which we extracted RNA from schizonts purified in Percoll-sorbitol gradients, performed an additional wash (the regular protocol includes a single washing step after Percoll

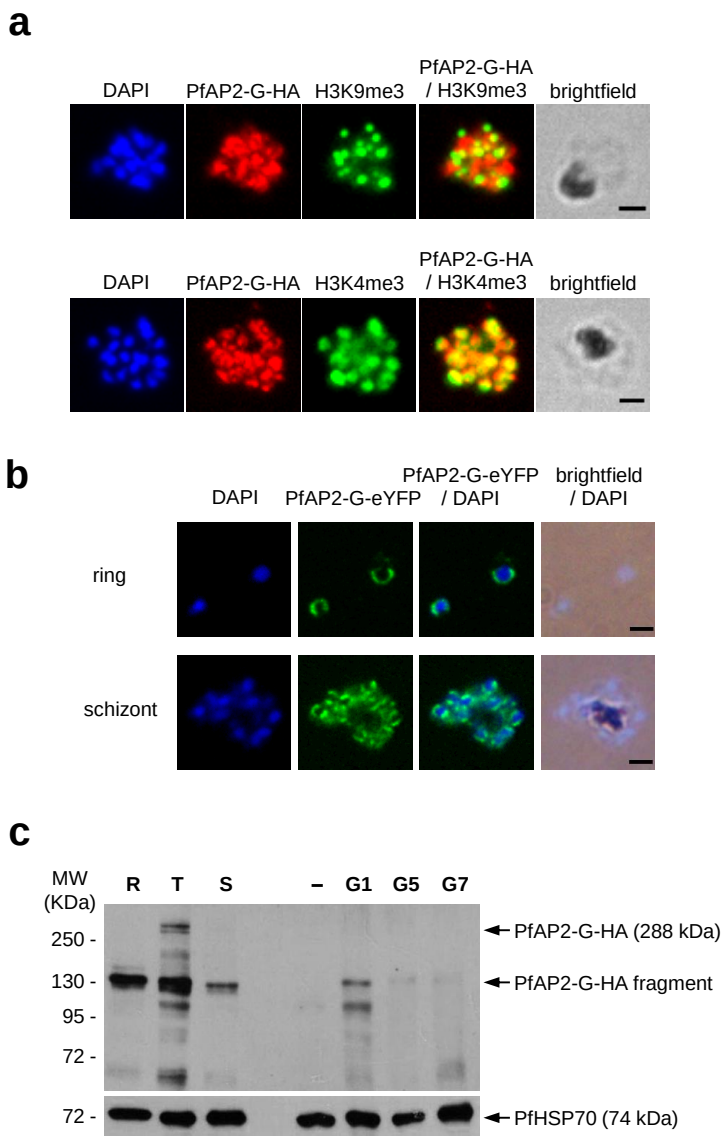
purification), and added uninfected erythrocytes immediately before Trizol extraction (Percoll/S+2x wash+RBC), and samples in which we magnet-purified schizonts and added uninfected erythrocytes immediately before Trizol extraction (magnet+RBC). Together with the experiments in Fig. 4b in the main article, these experiments show that *pfap2-g* can be abundantly expressed in mature schizonts. However, we measured very low transcript levels for this gene in late schizonts when Trizol extraction was performed directly on Percoll-purified schizont pellets, suggesting that left over Percoll selectively interferes with Trizol extraction of some transcripts. Thus, Trizol extraction directly from Percoll-purified schizonts can give artifactual results and should be avoided. Transcript levels were normalized against *ubiquitin-conjugating enzyme (uce).* Values are the average of three independent biological replicates (red dots), with S.E.M. **b,** Transcriptional time-course analysis of tightly synchronized cultures (5 h age window) of the parasite lines F12, E5 and E5-HA-DD (same samples as in Fig. 4a,c of the main article) normalized against *seryl tRNA synthetase* (*serrs*). While the overall temporal patterns of relative expression are similar to those observed when normalizing against *uce* (Fig. 4), apparent higher relative *pfap2-g* transcript levels are observed at 0-5 hpi. This is attributable to very low *serrs* expression at this time point (www.plasmodb.org). Values are the average of three (F12 and E5) or two (E5-HA-DD) independent biological replicates (red dots), with S.E.M. (F12 and E5 panels). **c,** *pfap2-g* transcript levels decrease with gametocyte maturation. RT-qPCR analysis of *pfap2-g* relative transcript levels in synchronized cultures of the E5 line at the late-ring stage (R), schizont stage (S, also containing ring-stages of the next cycle), and different stages of gametocyte development (G1, G2, G3, G6 and G9 correspond to days 1, 2, 3, 6 and 9 after GlcNAc addition to cultures at the ring stage, respectively). Sorbitol lysis was performed in the G1 and G2 samples to eliminate growth-arrested asexual parasites (although dead asexual parasites could not be completely removed). Transcript levels cannot be directly compared between rings/schizonts samples and gametocyte preparations because in the former only a subpopulation of the parasites is sexually committed and expresses *pfap2-g*, whereas in the latter essentially all parasites are sexual forms. Transcript levels were normalized against *ubiquitin-conjugating enzyme* (*uce*). Values are the average of three technical replicates (red dots).



Supplementary Figure 10. Transcriptional analysis of early gametocyte markers. a, Reverse transcription-quantitative PCR (RT-qPCR) time-course analysis of early gametocyte markers in tightly synchronized cultures (5 h age window) of the parasite lines F12 and E5. **b,** Analysis of the same markers in the transgenic parasite line E5-HA-DD maintained in the absence of Shld (-Shld) or with Shld added at 0-5 hpi. Samples are the same as in Fig. 4 of the main article. Insets show the same data with a different y-axis scale to highlight differential expression between time points. Transcript levels were normalized against *ubiquitin-conjugating enzyme (uce).* Values are the average of three

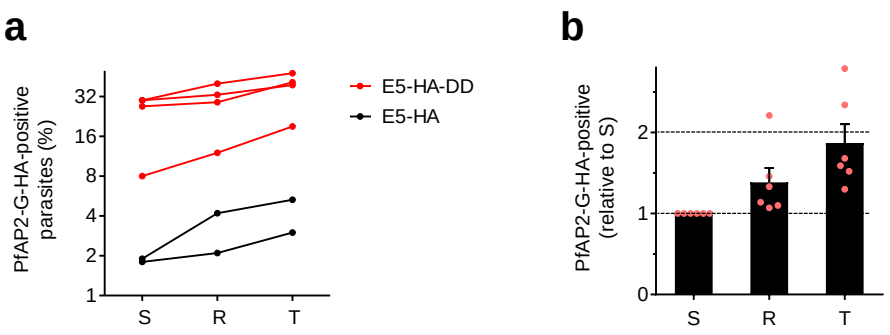
(panel a) or two (panel b) independent biological replicates (red dots), with S.E.M. (panel a).

pfs16[10] was the earliest gene to show an increase in transcript levels relative to untreated cultures when PfAP2-G was stabilized with Shld in the E5-HA-DD line. *pf14.744*[15] transcripts were the most specific of the markers tested, as transcripts for this gene were not detected in F12 and E5-HA-DD without Shld (however, at the protein level other markers such as Pfs16 are also highly specific. By IFA, we never observed Pfs16-positive parasites in the E5-HA-DD line in the absence of Shld). Transcripts of *pfs16* and *pfg27*[11] were detected in F12 and E5-HA-DD without Shld, but at much lower levels than in E5 and E5-HA-DD with Shld. Another recently described early gametocyte marker, *pfgexp5*[16], was expressed at high levels in the absence of functional PfAP2-G and gametocyte production (F12 and E5-HA-DD without Shld), which needs to be taken into account for the interpretation of the results of epidemiological studies or clinical trials using transcripts of this gene as a gametocyte marker[17]. However, this gene showed a marked increase in transcript levels at 0-5 hpi of cycle 2 after Shld addition.



Supplementary Figure 11. Subnuclear location of PfAP2-G and expression during gametocyte development. a, Representative IFA images of committed schizonts of the E5-HA line stained with antibodies against the HA tag and the H3K9me3 or H3K4me3 post-translational histone modifications, which mark heterochromatin and euchromatin, respectively. Experiments with anti-H3K9me3 and anti-H3K4me3 were independently repeated two and eight times, respectively, with similar results. PfAP2-G-HA does not preferentially localize within heterochromatin foci. Scale bar, 2 µm. **b**, In many parasites the PfAP2-G signal appears concentrated in the nuclear periphery. IFA images of ring and schizont stage parasites of the E5-eYFP line clearly showing this distribution of the PfAP2-G-eYFP signal are shown. IFA was performed with an anti-GFP antibody. Images are representative of four independent experiments. Scale bar, 2 µm. **c,** Western blot analysis of PfAP2-G-HA in E5-HA cultures synchronized to the ring (R), trophozoite (T) or schizont (S) stages, and in preparations of gametocytes collected at day 1 (G1), 5 (G5) or

7 (G7) after treatment of cultures at the ring stage with GlcNAc (day 0). Depending on the stage, we used between ~10^6 and ~10^7 parasites (the number of nuclei and total protein content varies between different stages). For the day 1 sample, sorbitol lysis was performed to eliminate asexual trophozoites, although contaminating arrested/ dead trophozoites could not be completely removed. Lane "–" is a negative control consisting of parasites of the wild type line E5 (no tag in PfAP2-G) at the trophozoite/schizont stage. Hybridization of the same membranes with antibodies against PfHSP70 was used as a loading control. The image is representative of two independent biological replicates, although variability was observed in the distribution of incomplete fragments of the protein.

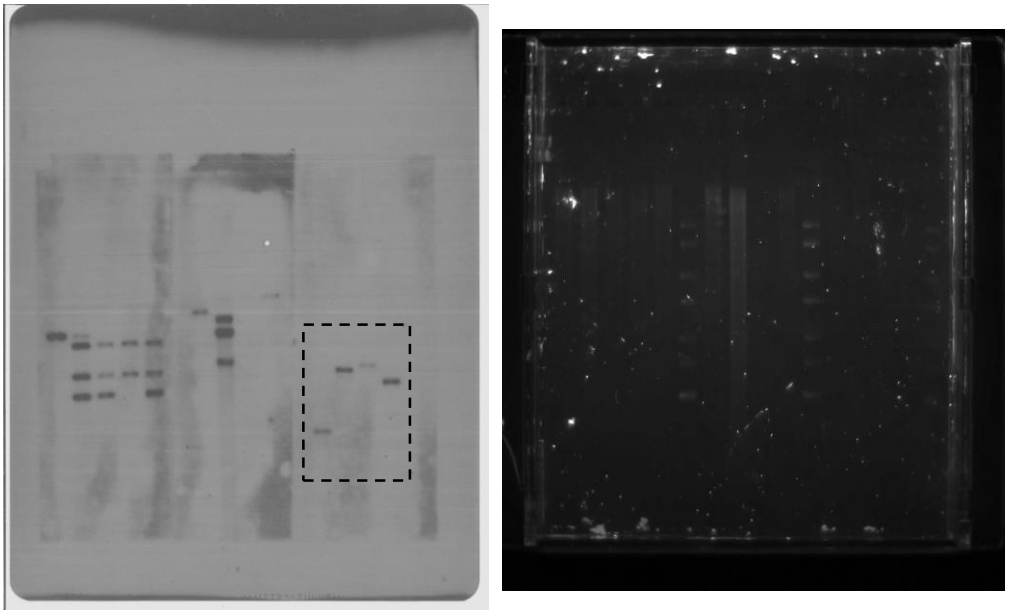


Supplementary Figure 12. Abundance of PfAP2-G-positive parasites in cultures synchronized to different stages of intraerythrocytic development. a, The proportion of PfAP2-G-HA-positive parasites was determined by IFA in synchronized cultures of the E5-HA and E5-HA-DD lines in consecutive schizont (S), ring (R) and trophozoite (T) stages. In the E5-HA-DD line, PfAP2-G was stabilized with Shld at the ring stage of the previous cycle. Each black or red line corresponds to a separate experiment. **b**, Proportion of PfAP2-G-positive parasites in synchronized cultures at the ring or trophozoite stage relative to the proportion of PfAP2-G-positive schizonts in the previous cycle. Values are the average of six independent experiments (red dots) using the E5-HA (*n*=2) and E5-HA-DD (*n*=4) lines, with S.E.M. The increase in the proportion of PfAP2-G-HA-positive parasites was significant from schizonts to trophozoites (p=0.013) but not from schizonts to rings (p=0.080) using a two-sided t-test with unequal variance.

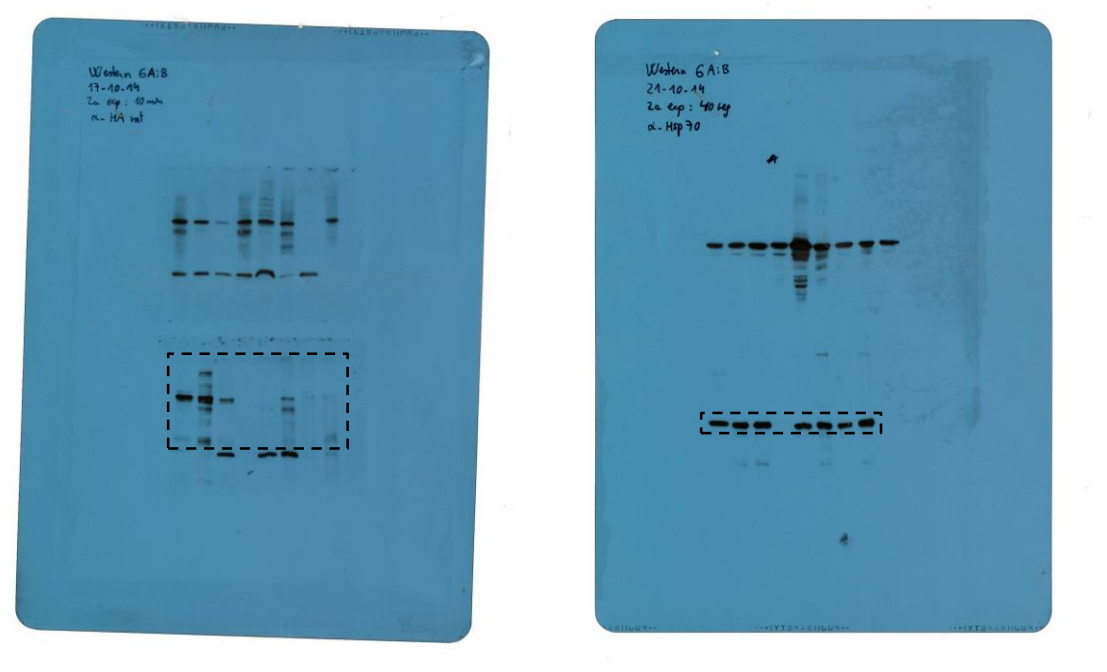
If the onset of PfAP2-G expression was always at the schizont stage, PfAP2-G-positive parasites in synchronized cultures at the schizont, ring or trophozoite stage would only include committed schizonts, sexual rings, and stage I gametocytes, respectively. Then, assuming that the average number of new parasites derived from a committed schizont is the same as for a non-committed schizont, the proportion of PfAP2-G-positive parasites in cultures at the ring or trophozoite stage would be expected to be the same as the proportion of positive parasites at the schizont stage of the previous cycle. Instead, if PfAP2-G activation (and sexual commitment) can start at the ring or trophozoite stage, a higher proportion of PfAP2-G-positive parasites would be expected in the ring and trophozoite cultures, as in addition to sexual rings or stage I gametocytes originated from sexually-committed schizonts of the previous cycle, they would include PfAP2-G-positive newly committed parasites (of which some would convert via the SCC route and some via

the NCC route). Hence, the higher proportion of PfAP2-G-positive parasites in ring and trophozoite cultures supports the idea that PfAP2-G expression and subsequent sexual commitment can start as early as the ring or trophozoite stage.

a



b



Supplementary Figure 13. Full length blots. a, Full length blot for Supplementary Fig. 2b. Left, autoradiography. Right, SYBR green-stained agarose gel used to determine the position of the molecular weight markers. **b,** Full length blots for Supplementary Fig. 11c. In the left image, the band excluded from the cropped image is non-specific, as it appears in the negative control lane (a wild type parasite line not containing the tag against which the antibody is directed, see Supplementary Fig. 11c).

Supplementary References

1 Bruce, M. C., Alano, P., Duthie, S. & Carter, R. Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development. *Parasitology* **100 Pt 2**, 191-200 (1990).

2 McDavid, A. *et al.* Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* **29**, 461-467 (2013).

3 Lopez-Barragan, M. J. *et al.* Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. *BMC Genomics* **12**, 587 (2011).

4 Silvestrini, F. *et al.* Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*. *Mol Cell Proteomics* **9**, 1437-1448 (2010).

5 Salanti, A. *et al.* Selective upregulation of a single distinctly structured *var* gene in chondroitin sulphate A-adhering *Plasmodium falciparum* involved in pregnancy-associated malaria. *Mol Microbiol* **49**, 179-191 (2003).

6 Aguilar, R. *et al.* Molecular evidence for the localization of *Plasmodium falciparum* immature gametocytes in bone marrow. *Blood* **123**, 959-966 (2014).

7 Rovira-Graells, N. *et al.* Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Res* **22**, 925-938 (2012).

8 Lim, M. Y. *et al.* UDP-galactose and acetyl-CoA transporters as *Plasmodium* multidrug resistance genes. *Nat Microbiol* **1**, 16166 (2016).

9 Ghorbal, M. *et al.* Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. *Nat Biotechnol* **32**, 819-821 (2014).

10 Bruce, M. C., Carter, R. N., Nakamura, K., Aikawa, M. & Carter, R. Cellular location and temporal expression of the *Plasmodium falciparum* sexual stage antigen Pfs16. *Mol Biochem Parasitol* **65**, 11-22 (1994).

11 Carter, R. *et al. Plasmodium falciparum*: an abundant stage-specific protein expressed during early gametocyte development. *Exp Parasitol* **69**, 140-149 (1989).

12 Delves, M. J. *et al.* Routine in vitro culture of *P. falciparum* gametocytes to evaluate novel transmission-blocking interventions. *Nat Protoc* **11**, 1668-1680 (2016).

13 Delves, M. J. *et al.* Male and female *Plasmodium falciparum* mature gametocytes show different responses to antimalarial drugs. *Antimicrob Agents Chemother* **57**, 3268-3274 (2013).

14 Poran, A. *et al.* Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature* **551**, 95-99 (2017).

15 Eksi, S. *et al.* Identification of a subtelomeric gene family expressed during the asexual-sexual stage transition in *Plasmodium falciparum*. *Mol Biochem Parasitol* **143**, 90-99 (2005).

16 Tiburcio, M. *et al.* Specific expression and export of the *Plasmodium falciparum* Gametocyte EXported Protein-5 marks the gametocyte ring stage. *Malar J* **14**, 334 (2015).

17 Farid, R., Dixon, M. W., Tilley, L. & McCarthy, J. S. Initiation of gametocytogenesis at very low parasite density in *Plasmodium falciparum* infection. *J Infect Dis* **215**, 1167-1174 (2017).