Published in final edited form as:

Nature. 2017 November 02; 551(7678): 95–99. doi:10.1038/nature24280.

Single-cell RNA-seq reveals a signature of sexual commitment in malaria parasites

Asaf Poran[1,2,3,*], **Christopher Nötzel**[4,5,*], **Omar Aly**[1,2], **Nuria Mencia-Trinchant**[1], **Chantal T. Harris**[4,6], **Monica L. Guzman**[7], **Duane C. Hassane**[1], **Olivier Elemento**[1,2], and **Björn F.C. Kafsack**[5]

[1]Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA

[2]Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA

[3]Physiology, Biophysics and Systems Biology Graduate Program, Weill Cornell Medicine, New York, NY, USA

[4]Biochemistry, Cell & Molecular Biology Graduate Program, Weill Cornell Medicine, New York, NY, USA

[5]Department of Microbiology & Immunology, Weill Cornell Medicine, New York, NY, USA

[6]Immunology and Microbial Pathogenesis Graduate Program, Weill Cornell Medicine, New York, NY, USA

[7]Division of Hematology & Medical Oncology, Department of Medicine, Weill Cornell Medicine, New York, NY, USA

Pathogens have to balance transmission with persistence. For *Plasmodium falciparum*, the most widespread and virulent malaria parasite, persistence within its human host requires continuous asexual replication within red blood cells (RBCs), while its mosquito-borne transmission depends on intra-erythrocytic differentiation into non-replicating sexual stages called gametocytes[1]. Commitment to either fate is determined during the preceding cell cycle that begins with invasion by a single, asexually-committed merozoite and ends, 48 hours later, with a schizont releasing newly formed merozoites, all committed to either continued asexual replication or differentiation into gametocytes[2,3]. Sexual commitment requires activation of *ap2-g* (PF3D7_1222600)[4,5], the transcriptional master regulator of sexual development, from an epigenetically silenced state during asexual replication[6,7].

Correspondence and requests for materials should be addressed to BK (bjk2007@med.cornell.edu) or OE (ole2001@med.cornell.edu).
*These authors contributed equally to this work.

Supplementary information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions:

BK conceived the study. AP, CN, and BK carried out the experiments unless otherwise noted. OA assisted in library preparation. NM, MG, and DH aided in implementing RNA FISH. CH acquired RNA FISH micrographs. AP, CN, and BK analyzed the data. BK and AP wrote the manuscript. CN and OE contributed to the manuscript.

The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper.

AP2-G expression during this “commitment cycle” poises gene expression in nascent merozoites to initiate sexual development through a hitherto unknown mechanism[2,4].

In order to maintain a persistent infection, *ap2-g* expression is limited to a sub-population of parasites (1–30%, depending on genetic background and growth conditions). As sexually-committed schizonts comprise only a sub-population and are morphologically indistinguishable from their asexually-committed counterparts, defining their characteristic gene expression has been difficult using traditional, bulk transcriptome profiling[8]. To determine the transcriptional changes induced by AP2-G within this sub-population, we applied highly-parallel, single-cell RNA sequencing (scRNA-seq)[9] to malaria cultures undergoing sexual commitment. In this first application of scRNA-seq to eukaryotic pathogens, we surveyed over 18,000 single parasite transcriptomes from a conditional AP2-G knockdown (AP2-G-DD) line and NF54 wild-type parasite at multiple stages of development (Fig. 1A, Extended Data Fig. 1, Supplementary Table S1) and found that sexually committed, AP2-G^{+} mature schizonts specifically up-regulate additional regulators of gene expression, including other AP2 transcription factors, histone modifying enzymes, and regulators of nucleosome positioning. These epigenetic regulators likely act to poise the expression/repression of genes necessary for the initiation of gametocyte development in the subsequent cell cycle.

To test if we could reproducibly distinguish multiple parasite stages based on scRNA-seq, we first analyzed a subset of 3,500 single-cell transcriptomes (SCTs) of individual parasites at multiple stages of intra-erythrocytic development. The resulting SCTs readily segregated into distinct clusters by lifecycle stage as determined by time of parasite collection and with little apparent effects from technical or biological replicates (Fig. 1B–C), with the number of individually captured transcripts varying by stage in accordance with their relative RNA content[10].

To identify differences resulting from *ap2-g* expression during the commitment cycle, we used the AP2-G-DD parasite line, in which the endogenous *ap2-g* coding sequence is fused to a destabilization domain that targets the resulting fusion protein for degradation in the absence of the stabilizing Shield1 ligand[4].

During the commitment cycle, synchronized parasites were grown at high parasitemia to favor sexual commitment[11] in the presence of either the ligand (“treated”) or solvent control (“untreated”). We sequenced a total of 12,800 infected erythrocytes isolated during the commitment cycle at 30, 36, and 42 hours post-invasion (hpi), as well as nascent gametocytes (stage I) after 42 hours of development in the subsequent cycle (see methods, Extended Data Fig. 2). The average sexual commitment was found to be 17.5 ± 0.1% in treated cells while no gametocytes were observed in untreated cells (<0.05%) in two independent experiments.

After unsupervised clustering based on similarity in overall gene expression[12], the 10,509 quality-filtered SCTs self-organized in a continuous arc comprised of eleven clusters (1–11) surrounding five central clusters (12–16)(Fig. 2A, Supplementary Video 1). Virtually all cells in the cluster 1–11 continuum were isolated during the commitment cycle and self-

organized primarily based on time of collection (Fig. 2A, Extended Data Fig. 3). The center clusters were composed of either stage I gametocytes (13–14) or cells from a mixture of time points (12, 15–16). Surprisingly, apart from the gametocyte clusters, no others showed notable enrichment for either “treated” or “untreated” cells (Extended Data Fig. 3C), suggesting that cell cycle progression was driving cluster assignment of SCTs irrespective of sexual commitment. This idea was further supported by a comparison of SCTs in clusters 1–11 to published bulk RNA-seq time course data[13] demonstrating that SCTs from progressively later clusters best matched an ordered series of time points (Fig. 2B, Extended Data Fig. 4A).

To further test this hypothesis, we applied the Monocle2 algorithm[14] to test if SCTs could be organized in a continuous gene expression program. Indeed, monocle-ordered SCTs also form a single continuum that recapitulates the cluster 1–11 order (Fig. 2C, Extended Data Fig. 4B). Finally, we ordered cells based on progression through this continuum (pseudo-time) and evaluated expression of gene sets associated with cell cycle progression and merozoite formation, beginning with DNA replication and ending with expression of PfCDPK5, a calcium-dependent protein kinase that triggers merozoites egress (Fig. 2D)[15]. Interestingly, post-mitotic schizonts occupied about two thirds of this continuum, highlighting the tightly orchestrated transcriptional program necessary to assemble merozoite progeny over the final eight hours[16], following the earlier relatively stable activity of genome replication.

The processes of genome replication and merozoite formation are shared by both asexually- and sexually-committing parasites. Accordingly, cell cycle progression was the primary driver of cluster assignment, regardless of the underlying program of sexual commitment. This is consistent with the presence of significant sub-populations of treated cells expressing *ap2-g* within each of these eleven clusters (Fig. 3A). Over the 24–48 hour window, mean *ap2-g* expression and the fraction of AP2-G expressing (AP2-G^{+}) cells, in both treated and untreated cells rises to a moderate peak in cluster 3 before falling again in cluster 4 (Fig. 3B). However, only when AP2-G activity was stabilized by ligand treatment, did the abundance of *ap2-g* transcripts rise rapidly, reaching a maximal 22-fold increase (over cluster 1) in mature schizonts. *ap2-g* expression is significantly higher in treated than in untreated cells across clusters 1–11. Analysis of 3,643 wild-type (NF54) schizonts (36–48 hpi) SCTs recapitulated these findings whether clustered independently or in conjunction with AP2-G-DD SCTs (Extended Data Fig. 5).

Rather than averaging gene expression across the entire population, which is heterogeneous with respect to cell cycle progression and commitment, our single-cell resolution allowed us to directly compare gene expression in cells expressing AP2-G (AP2-G^{+}) and in cells in which the locus remains silenced (AP2-G^{-}). This was made possible only by the large number of SCTs collected, yielding many AP2-G^{+} cells within each cluster, thereby providing sufficient statistical power to determine AP2-G-dependent differential gene expression.

Using this approach, we identified 19 genes that were significantly up-regulated in AP2-G^{+} late stages within at least one cluster of both treated AP2-G-DD (27 genes) and NF54 late

stages (48 genes) (Fig. 3C–D, Extended Data Fig. 6, Supplementary Table S3). In line with our previous observations that AP2-G acts as a transcriptional activator[4], virtually all of the differentially expressed genes were up-regulated in AP2-G$^+$ cells containing an average of 2.6 AP2-G binding sites (3.8x fold enrichment, p<2e-6) within their upstream regions compared with 0.7 binding sites genome-wide. The number of upstream binding sites correlated well with the observed increase in AP2-G$^+$ cells. (Spearman's R=0.52). Six of these shared hits have been linked to interaction with the host cell either via protein export or during the reinvasion process, while seven putatively function in gene regulation based on their predicted interaction with either RNA, DNA, or chromatin. These included two additional ApiAP2-family members, PF3D7_1139300 and PF3D7_1222400, a primate specific ApiAP2 recently found to acquire loss of function mutations in laboratory-adapted strains[17]. Four additional hits are implicated in the regulation of chromatin structure: two SNF2 family helicases (ISWI, SNF2L) that function in chromatin remodeling[18] and two putative histone modifying enzymes: the histone deacetylase HDA1[7], and a previously un-annotated putative histone lysine-specific demethylase PfLSD2 (PF3D7_0801900, see Extended Data Fig. 7)[19].

Next, we explored the temporal dynamics of up-regulation in AP2-G$^+$ cells for our 19 hits (Fig. 4A–B, Extended Data Fig. 8). Both of the putative SNF2 helicases are more highly expressed in AP2-G$^+$ cells when DNA replication remains ongoing (pseudo-time 0–20) preceding the initial rise in overall AP2-G expression. Following its peak expression in trophozoites, AP2 PF3D7_1222400 transcript abundance initially falls but rises again in AP2-G$^+$ cells (pseudo-time 20–35) immediately following the initial wave of AP2-G expression. AP2-G transcript abundance increases steadily during schizont segmentation (pseudo-time 30–50) leading to a sharp rise in LSD2, HDA1, and AP2 Pf3D7_1139300 just prior to egress as indicated by CDPK5 expression (pseudo-time >50). Expression of shared hits in untreated AP2-G-DD cells mirrors that in treated AP2-G$^-$ cells. This demonstrates that the altered expression in AP2-G$^+$ cells depends on AP2-G activity. We confirmed AP2-G contingent up-regulation of the hits in late schizonts by qRT-PCR (Extended Data Fig. 9A) and RNA FISH (Fig. 4C–D, Extended Data Fig. 9B).

Since AP2-G is expected, directly or indirectly, to bring about the coordinated expression of a limited set of genes, we wanted to define a network of co-expressed genes in committed schizonts that is centered on AP2-G. To this end, we took advantage of our ability to determine true co-expression (within the same cell) to identify genes that are significantly co-expressed with AP2-G or any of our 19 hits across clusters 1–11 (Extended Data Fig. 10, Supplementary Table S4). We relied on the recurrence of co-expressed gene pairs across all clusters to define a highly interconnected, AP2-G-centered network of significantly co-expressed genes (Fig. 4E).

Due to its stochastic nature, transcriptional regulation of sexual commitment is uniquely suited for exploration for single cell transcriptomics. By comparing thousands of single cell transcriptomes from wild-type NF54 and AP2-G-DD parasites, we identified a shared transcriptional sexual commitment program that is executed as they progress through schizogeny (Fig. 4F). This program is initiated by de-repression of the *ap2-g* locus during DNA replication (pseudo-time 0–15) concomitant with increased expression of the

SWI/SNF helicases ISWI and SNF2L in AP2-G$^+$ cells and may implicate these chromatin-remodeling factors in the replacement or repositioning of nucleosomes that maintain *ap2-g* silencing. As DNA replication concludes, we observed an initial increase in AP2-G expression regardless of AP2-G stabilization, suggesting that this initial rise is not dependent on AP2-G function and consistent with epigenetic de-repression. After reaching an initial peak (around pseudo-time 20), AP2-G expression drops temporarily (around pseudo-time 28) before increasing steadily for the remainder of the cell cycle. Unlike the initial increase, this second wave of expression requires AP2-G stabilization, confirming the existence of a previously proposed positive transcriptional feedback loop[4], that is likely activated by direct binding of AP2-G to the cognate motifs present within its own promoter. Such bi-stable transcriptional switches frequently regulate cell differentiation and prevent the mixed expression of incompatible developmental programs[20–23]. As AP2-G reaches an initial peak, a second, genomically adjacent AP2 (PF3D7_1222400) is up-regulated in AP2-G expressing cells. The period of increased PF3D7_1222400 expression coincides with a temporary drop in AP2-G expression. Combined with recent evidence that culture adaptation selects for loss of function mutations in both AP2-G and PF3D7_1222400, its genomic location and expression raises the possibility that PF3D7_1222400 may act as a sexual commitment checkpoint prior to activation of the feedback loop. Finally, as AP2-G reaches its peak expression immediately prior to egress, a third AP2 factor (PF3D7_1139300) is sharply up-regulated along with the histone modifying enzymes LSD2 and HDA1. We propose that in sexually-committed schizonts these regulators act in concert to poise gene expression necessary for gametocyte development in the subsequent cell cycle and may play a possible role in mating type determination.

The ability to measure gene expression in thousands of individual parasites opens the door for answering key questions that involve small sub-populations of parasites, such as the mechanisms underlying antigenic switching[24] and artimisinin-induced dormancy in resistant parasites[25], with important implications for malaria treatment and vaccines development.

Methods

Parasites and strains

The parasite strains used in this study were AP2-G-DD[4], NF54, and DCJ[26]. They were maintained using established culturing techniques[27]. Cultures were tested for mycoplasma on a semi-annual basis and found negative. Engineered strains were obtained from their lab of origin and tested for the engineered phenotype and expected drug-resistance. DCJ parasites were resistant to Blasticidin-S (Spring 2016). AP2-G-DD parasites were resistant to WR99210 and produced wild-type (NF54) level of gametocytes when cultured in the presence of 0.5 μM Shield1 ligand but not when cultured with solvent control (June 2017).

Gametocyte induction

For synchronous gametocyte induction (adapted from[11]), parasites were double-synchronized with 5% sorbitol solution[27] to achieve a synchrony of ± 6 hours and cultured while shaking (3% hematocrit). Synchronized parasites were set up at 2% late trophozoite stage. Following re-invasion, parasites were maintained at 7–8% parasitemia during the

commitment cycle to induce sexual commitment. Once parasites reached the schizont stage, they were expanded 1:4 in order to relieve stress. After re-invasion, on the first day of gametocyte development (D+1), ring-stage parasites were counted and treated with 50 mM *N*-acetyl-D-glucosamine[27] for three consecutive days in order to kill asexuals and gametocytemia was counted on the fourth day of gametocyte development (D+4). The gametocyte commitment rate was then determined by dividing the D+4 gametocytemia by the D+1 parasitemia, counted before addition of *N*-acetyl-D-glucosamine.

Parasite isolation for scRNA-seq

AP2-G-DD or NF54 parasites were set up for gametocyte induction as described above. For AP2-G-DD, the cultures were split in half after re-invasion on D-2 was completed and treated with 0.5 μM Shld1 or an equal volume of ethanol solvent for the remainder of the experiment[28]. Infected erythrocytes were purified at the desired developmental time-point (on D-1: 30, 36 and 42 hours post invasion; on D+2: 42 hours into gametocyte development) using magnetic columns (LS columns, Milteny Biotec)[27]. On the magnet, columns were washed with media without AlbuMAX, after which the parasite culture was loaded. After washing with media without AlbuMAX and with PBS + 0.01% BSA, hemozoin-containing parasites were eluted off the magnet with PBS + 0.01% BSA. The sample was adjusted to 1.3e5 cells/ml, kept on ice and immediately subjected to Drop-seq.

DCJ parasites were cultured in media supplemented with 5 μg/ml BSD. Synchronized parasites at 18 ± 4 hpi were then stained at 1% hematocrit with 2 μM Hoechst 33342 (HO) for 10 min at room temperature, pelleted and re-suspended in colorless media without AlbuMAX to 0.1% hematocrit. HO positive cells were sorted into PBS + 0.01% BSA using a FACS Aria – II sorter, adjusted to a concentration of 1.3e5 cells/ml, kept on ice and immediately subjected to Drop-seq. All samples were inspected by microscopy and only used for scRNA-seq if singly-infected RBCs > 95%.

See Supplementary Table S1 for the number for samples collected.

Parasite isolation for RNA FISH and qRT-PCR

Schizonts of semi-synchronous AP2-G-DD parasites were isolated using a percoll-sorbitol gradient and reintroduced into culture with fresh blood at 4% parasitemia. After 6 hours shaking in culture, the remaining schizonts were removed using 5% sorbitol lysis, leaving a tightly synchronous population of early ring stages at high parasitemia as desired for induction of sexual commitment (0 – 6 hpi, 4 – 7% parasitemia). The culture was immediately split in two and supplemented with either 0.5 μM Shield1 ligand or solvent control for the remainder of the experiment[28]. 45 hours later, when the parasites were mid-way through re-invasion, highly synchronous mature schizonts were magnetically isolated as described above and used for RNA FISH or bulk RNA extraction for qRT-PCR.

RNA Fluorescent in Situ Hybridization and detection

RNA FISH was carried out using PrimeFlow (ThermoFisher) according to the manufacturer's instructions and stained for DNA with Hoechst33342. Custom probes were designed and synthesized against unique regions of the AP2-G (AlexaFluor 488), CDPK5

(AlexaFluor 647), LSD2 (AlexaFluor 568) and AP2 PF3D7_1139300 (AlexaFluor 568) transcripts by the manufacturer (Supplementary Table S5). Detection was carried out using a BD LSR II flow-cytometer using 350nm, 488nm, 561nm, and 650nm lasers, respectively. To ensure analysis of individual parasites, a single uniform population of parasites (RBC membranes are removed during fixation/permeabilization) was triple gated on FSC-A vs SSC-A, FSC-H vs FSC –A and SSC-H vs SSC-A. Fluorescence gates were set based on fluorescence-minus-one controls (see Extended Data Fig. 9). Mature schizonts, as identified by maximal DNA content and high CDPK5 expression, were analyzed for expression of AP2-G and either LSD2 or PF3D7_1139300.

Microscopy

Cells were imaged as z-stacks on a Leica DMI6000 microscope at 1000X magnification using DIC optics and DAPI, GFP, Cy3 filter cubes to detect Hoechst33342, Alexa488 and Alexa536, respectively. Fluorescent channel z-stacks were deconvolved using the ImageJ DeconvolutionLab2 plugin (NLLS algorithm) followed by maximum intensity z-projection and background adjustment.

Quantitative PCR with reverse transcription (qRT-PCR)

Trizol-extracted RNA from saponin-lysed parasites (0.1% in PBS) was used for cDNA synthesis with random hexamers using SuperScript II (ThermoFisher) according to the manufacturer's instructions. Quantitative PCR was performed using gene-specific primer sets (Supplementary Table S6) and iTaq Universal SYBR Green Supermix (Bio-Rad), and relative transcript abundance was determined by normalizing to expression levels of fructose-biphosphate aldolase (PF3D7_1444800) using the ddCt-method[29].

Drop-seq and sequencing analysis pipeline

Single-cell transcriptomic profiles were generated using Drop-seq, a technology designed for highly parallel genome-wide expression profiling of individual cells using nanoliter droplets, as detailed extensively by Macosko et al.[9]. Briefly, single-cell suspension and uniquely barcoded beads were co-localized in droplets using a microfluidics device (see CAD file from http://mccarrolllab.com/dropseq/, manufactured by FlowJEM). The droplets are composed of cell-lysis buffer and serve as compartmentalizing chambers for RNA capture. Flow rates were adjusted to maintain stable droplet formation and increase droplet homogeneity. We then adjusted cell and bead concentrations to accommodate variation in droplet size compared to with the original publication (113um in our system). Doublet rate was estimated with the species-mixing experiment according to Macosko et al.[9]. Examination of cells showed complete lysis within the time required for examination by microscopy (less than 1 min), notably shorter than the time cells spend in droplets during lysis and mRNA capture.

Droplet breakage and single-cell library preparations followed the procedure as described[9]. In short, collected droplets were disrupted and RNA-hybridized beads were extracted. Reverse transcription was performed with template switching to allow for cDNA amplification by PCR. An additional pre-PCR step was added to determine the appropriate number of cycles (17–19 cycles) to achieve a cDNA library at a concentration of 400–

1000ug/μL, as suggested by the protocol. cDNA Samples were purified using Agencourt AMpure XP (Beckman Coulter), and were run on a 2100 BioAnalazyer instrument with a High Sensitivity DNA kit (Agilent Technologies). Samples were prepared for sequencing using the Illumina Nextera XT kit, and sequenced on a NextSeq500 at an average of 70,000 reads per cell. Libraries with high numbers of cells were divided into technical replicates, which were processed independently. Raw reads were processed and aligned (STAR aligner) using the standard Drop-seq pipeline, and according to the "Drop-seq Alignment Cookbook", both found at http://mccarrolllab.com/dropseq/. Reads were aligned to the *Plasmodium falciparum* 3D7 transcriptome (PlasmoDB version 32). For each read, a single optimal mapping position was retained. Unique transcripts mapping to alternative splice variants were combined for subsequent analysis. Single-cell expression matrices were generated using cellular barcode and unique molecular identifiers (UMIs). As Drop-seq RNA capture is designed for poly-adenylated transcripts, we removed known RNA pol I/III transcripts (rRNA, tRNA) prior to analysis. UMIs mapping to ribosomal RNA transcripts comprised less than 5% of reads for 99.5% of our cells. Even though individual malaria parasites contain significantly less RNA than mammalian cells[30], the average number of transcripts captured per asexual late stage parasite or gametocyte was similar to that reported for mammalian cells[9], indicating higher capture efficiency. Transcript capture from ring-stage cells yielded 20–30% of the those in late-stages, reflecting the significantly lower RNA content of ring-stages compared to late stages [30].

Single-cell transcriptome analysis

Data normalization, clustering and differential expression were performed using the "Seurat" R package[12]. Cells in which less than 100/300 detected UMIs (ring/late stages) and genes detected in less than 3 cells were excluded from analysis. SCTs were internally normalized to 10,000 transcripts, log transformed and regressed on the number of UMIs per cells before dimensionality reduction and clustering. We selected ~600–700 highly variable genes using the genes' expression and dispersion (variance/mean), and performed principle component (PC) analysis. The most significant PCs (20–30) were used for clustering and tSNE representations[31]. Clustering resolution was chosen such that visually distinct groups of 100+ cells were assigned to individual clusters.

After filtering lowly expressed genes, pseudo-temporal assignment was performed by applying "Monocle 2" R package to ~1100 most variable genes to order the cells along pseudo-time.

SCT correlation with bulk RNA-seq

Mapping the cells from clusters 1–11 to the time series bulk RNA-seq dataset was performed by calculating Pearson correlation coefficient of each SCT with each of the 8 bulk RNA-seq time points. The cells were ordered by their cluster number, and by the time point of maximal correlation.

Gene-module expression analysis

We defined modules of genes related to specific transcriptional programs activated during different stages of the cell cycle (Supplementary Table 2). We then sum the fraction of the

transcriptome of each cell for all the genes in each module. After normalization by the mean expression of each module across all cells, we plot the "loess" smoothed gene-module expression over pseudo-time.

Differential gene expression in AP2-G^{+} vs AP2-G^{-} cells

We used the Seurat package to calculate differential gene expression between treated AP2-G positive cells and treated AP2-G negative cells in each of clusters 1–11, and repeated this analysis for clusters 1–9 of the NF54 cells. We required that (a) for each cluster treatment groups were larger than 11 cells, (b) the number of cells expressing a gene to increase by at least 17.5%, (c) genes have to be expressed in >10% of the cells of at least one treatment group, and (d) gene expression differed by >2.45-fold between treatment groups. Infinite ratios were capped at the maximal finite value observed. Fig. 3D shows the fold-change of the 19 genes found differentially expressed in both strains, across all 11 clusters of co-clustered AP2-G-DD and NF54 SCTs.

Co-Expression analysis and network construction

For each of clusters 1–11 of the AP2-G-DD SCTs, we used the Fisher's exact test to detect significant co-expression between A) *ap2-g a*nd the 19 shared differentially expressed genes (Fig. 3D), and B) all other genes. We then used the Spearman correlation coefficient and the mean square contingency coefficient of the same gene-pairs for filtering, requiring FDR corrected p-values from the Fisher's test < 0.05, and agreement between the signs of the correlation coefficient and the mean square contingency coefficient. In addition, we required a mean square contingency coefficient > 0.3. The result is a table of significantly co-expressed genes in each cluster (Supplementary Table S4). For co-expression pairs repeating in 5 or more clusters, an undirected weighted network was constructed, where edge weight is the number of occurrences of each gene pair. Networks were generated using the "igraph" R package[32] and visualized using Cytoscape[33].

AP2-G motif calling and enrichment testing

Upstream regions were defined as extending up to 3kb from the ATG start codon. For intergenic regions of less than 3kb containing promoters of two genes, exactly half of the intergenic sequence was assigned to each gene. Within these regions putative AP2-G binding sites were identified using SCANACE[34] and the 6bp-core position-specific weight matrix for AP2-G[35]. The AP2-G motif enrichment analyses were performed using Fisher's exact test on the contingency table of motifs in the compared genes sets.

PfLSD2 Sequence Analysis

The amino acid sequence of PfLSD2 was aligned with 4 other phylogenetically well-distributed syntenic Plasmodium orthologs (*P. vivax, P. ovale curtisi, P. gallinaceum, and P. berghei*) using the phylogeny-aware multiple sequence aligner webPRANK (www.ebi.ac.uk/goldman-srv/webprank/, Supplementary Data 1). Homologous blocks were joined and analyzed for functional conservation using the NCBI Conserved Domain Database (www.ncbi.nlm.nih.gov/cdd/) and PANTHER (Protein ANalysis THrough Evolutionary Relationships, www.pantherdb.org/tools/) (Extended Data Fig. 7).

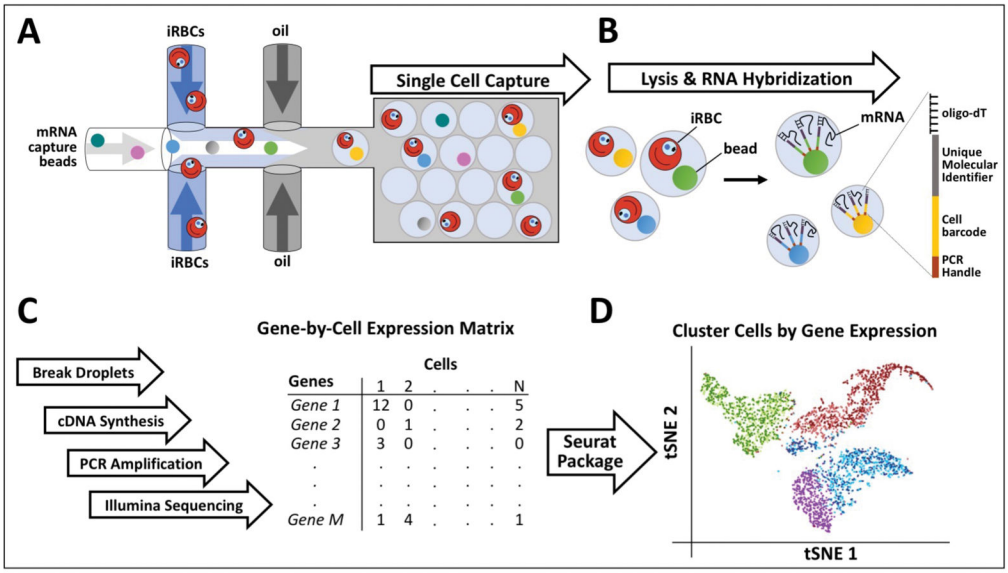
Other statistical analyses

Significance of mean differential gene expression across conditions (Fig. 3B) was assessed using two-sided Welch Two Sample t-test, p-values in multiple hypothesis testing are FDR corrected. Fisher's Test was used to test significance of enrichment in AP2-G binding sites.

Data and code availability

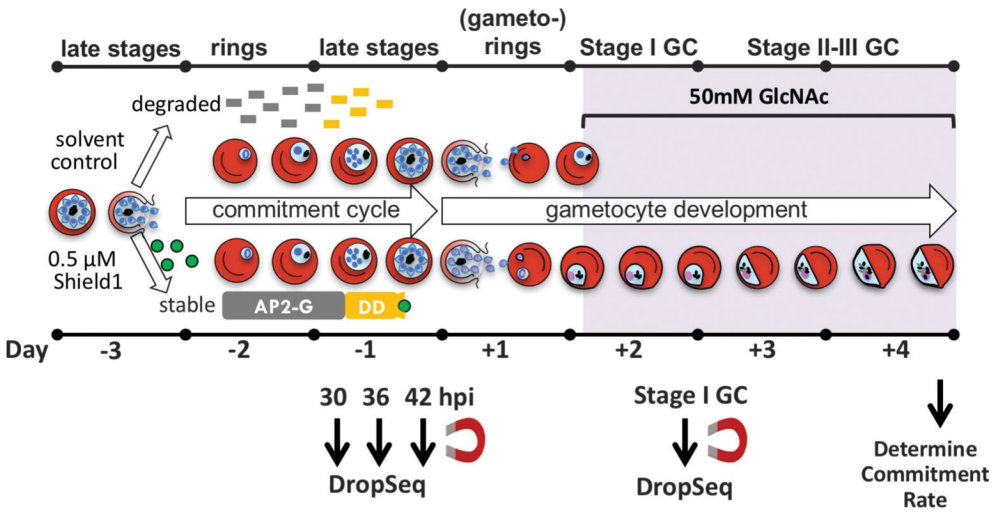
The single cell RNA sequencing data of this study has been deposited to the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) with the study accession code SRP116718. The scripts used for analysis and figure generation are available on www.KafsackLab.org/code.

Extended Data



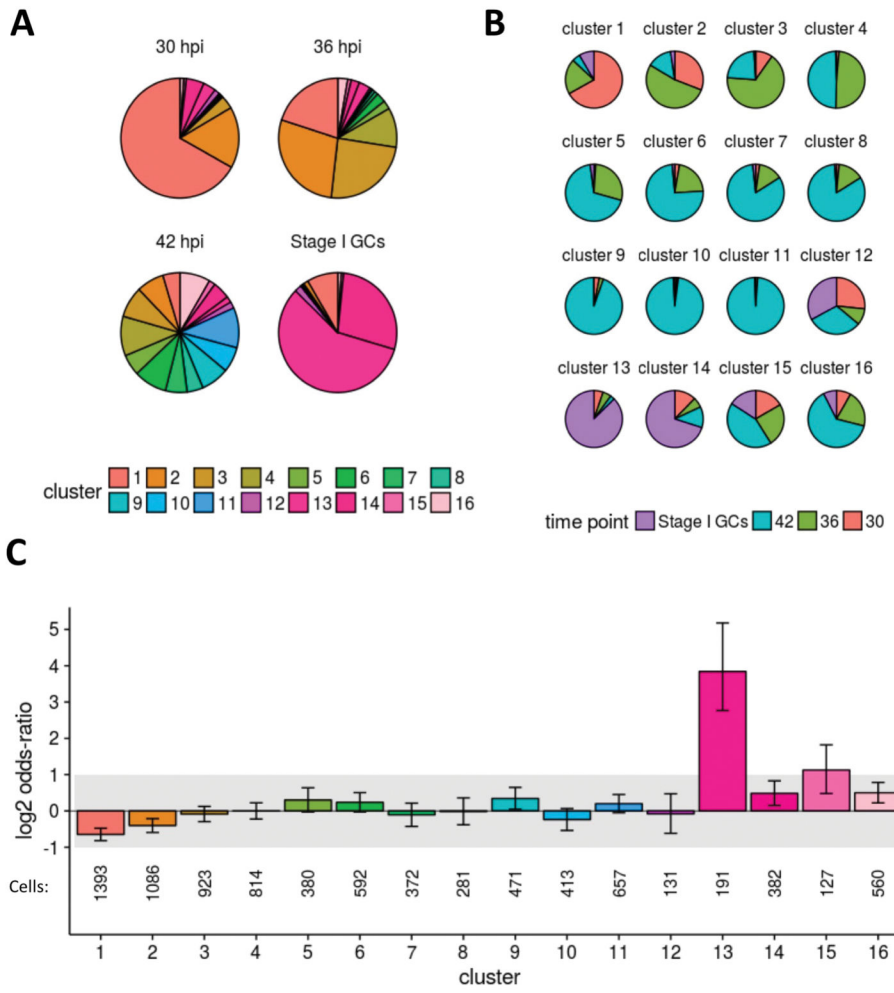
Extended Data Figure 1. Drop-seq single-cell sequencing and analysis workflow

(A) Single infected RBCs and uniquely barcoded beads are captured in droplets of cell lysis buffer using a microfluidics device. **(B)** Released mRNAs are capture on individually barcoded poly-dT oligos. **(C)** Template-switch cDNA synthesis labels each captured transcripts with cell-specific barcode and a unique molecular identifier (UMI). Following library preparation and Illumina sequencing, individual transcripts are mapped and counted within each cell **D.** The resulting expression matrix of single cell transcriptomes is used for clustering and analysis using the Seurat package of scRNA-seq analysis tools.



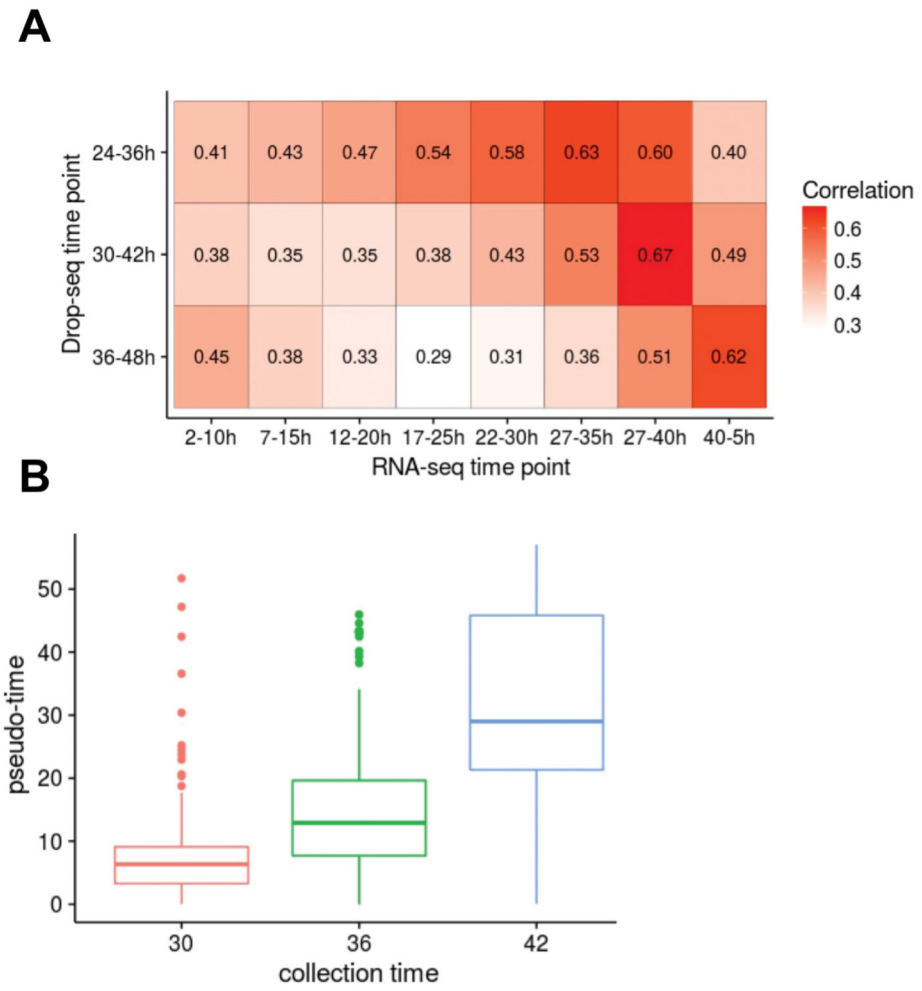
Extended Data Figure 2. Experimental overview

Fusion of the endogenous *ap2-g* coding sequence with the ddFKBP destabilization domain makes sexual commitment conditional on treatment with 0.5 µM Shield1 ligand. Unless ligand is added, AP2-G is targeted for proteolytic degradation. At the start of the commitment cycle, AP2-G-DD parasites were split into two cultures and treated with ligand or solvent control. Cultures were maintained under conditions inducing sexual commitment and infected RBCs were purified for Drop-seq at 30h, 36h and 42h during the commitment cycle, as well as, after 42 hours of gametocyte development in the subsequent cycle.



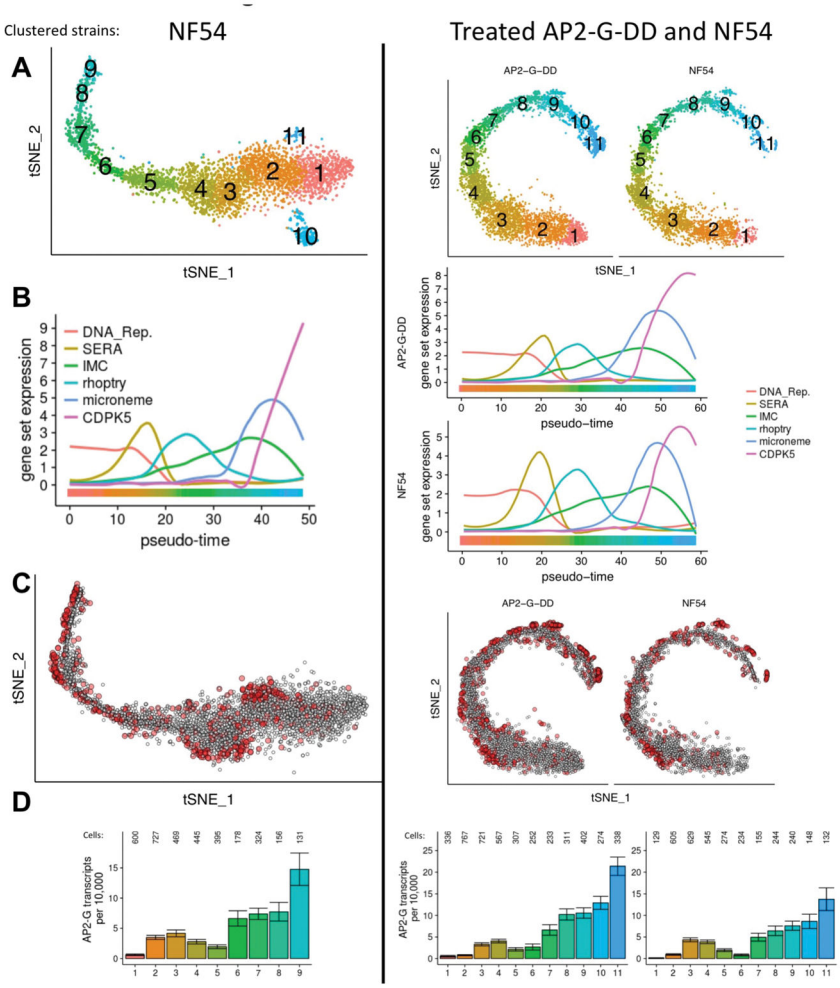
Extended Data Figure 3. Analysis of AP2-G-DD cluster composition

(A) Cluster assignment of SCTs collected at 30, 36, and 42 h.p.i and stage I gametocytes. **(B)** Cluster composition by sample time point. **(C)** Cluster-wise enrichment for treated vs. untreated cells by cluster for cells collected during the commitment cycle. Positive and negative log-odds-ratio values indicate enrichment for treated and untreated cells, respectively. Error bars indicate the 95% confidence interval. Number of cells is shown.



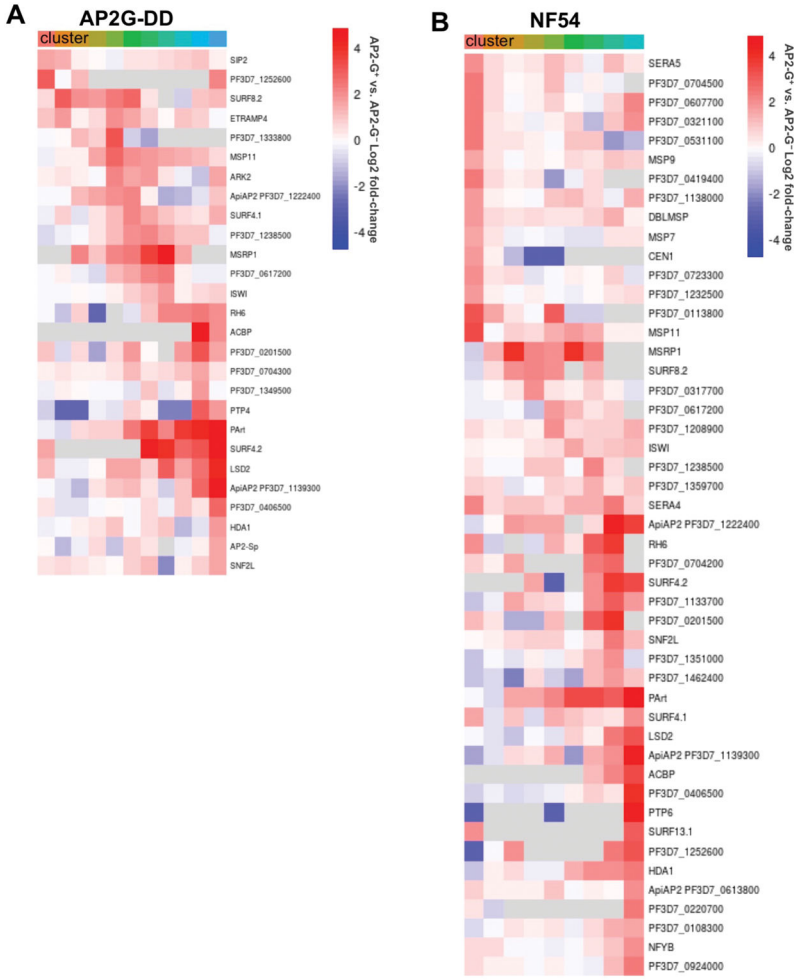
Extended Data Figure 4. Single-cell transcriptomes averaged by collection time match corresponding bulk RNA-seq time points (A) and pseudo-time assignment (B)

(A) The averaged single-cell expression profiles for each of the collected samples ($n_{24\text{-}36h}$=1202, $n_{30\text{-}42h}$=1536, $n_{36\text{-}48h}$=6035) were correlated with published bulk RNA-seq time points. Pearson's correlation coefficents are shown in heatmap boxes. Maximal correlation values indicate agreement between collection time point and transcriptome mapping. **(B)** The distributions of pseudo-time assignment of treated cluster 1–11 SCTs grouped by collection time (n_{30h}=508, n_{36h}=821, n_{42h}=3356). Boxes indicate the interquartile range; whiskers extend 1.5x interquartile range from the box.



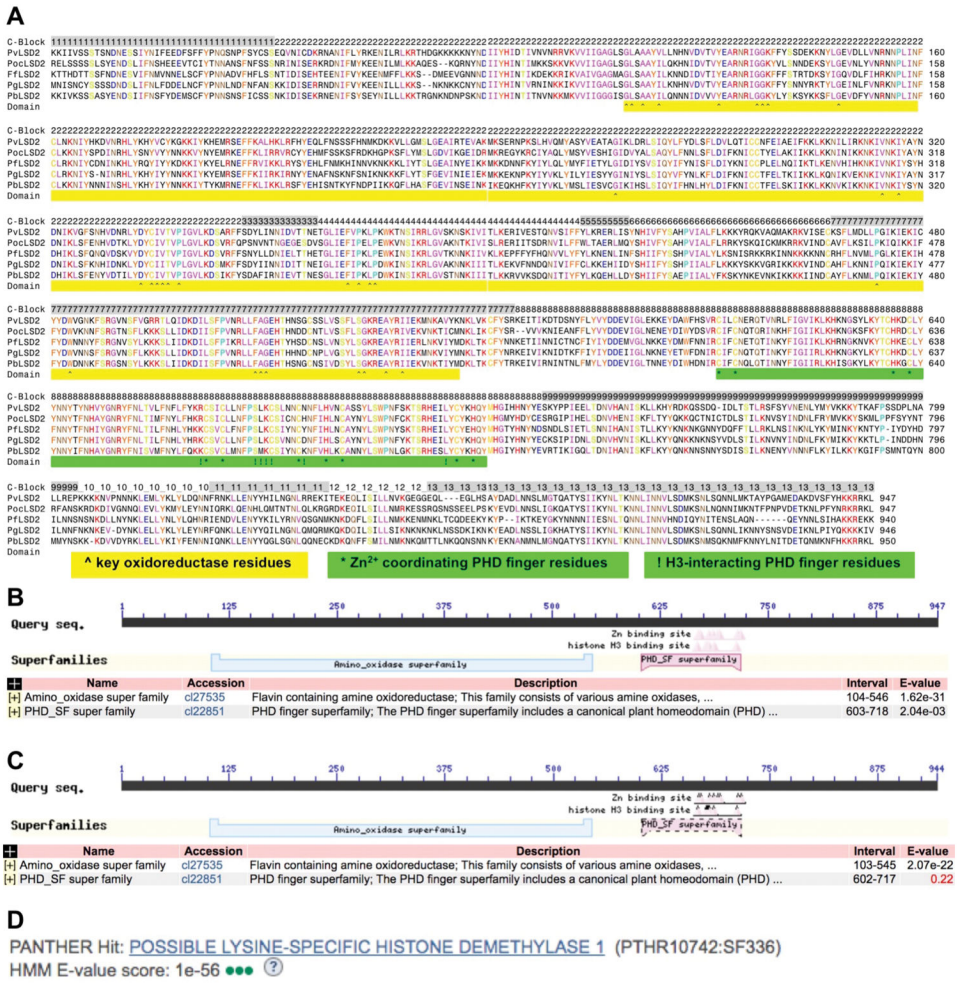
Extended Data Figure 5. Analysis of NF54 late-stage SCTs clustered independently (left) or co-clustered with treated AP2-G-DD SCTs (right)

(A) tSNE plot of cluster 1–11 SCTs**. (B)** Gene set expression as a function of pseudo-time. Color bar indicates cluster assignment along pseudo-time. **(C)** tSNE plot showing AP2-G^{+} cells in red. Number of cells are indicated. **(D)** Mean *ap2-g* expression per 10,000 transcripts by cluster. Error bars are SEM. Number of cells is indicated.



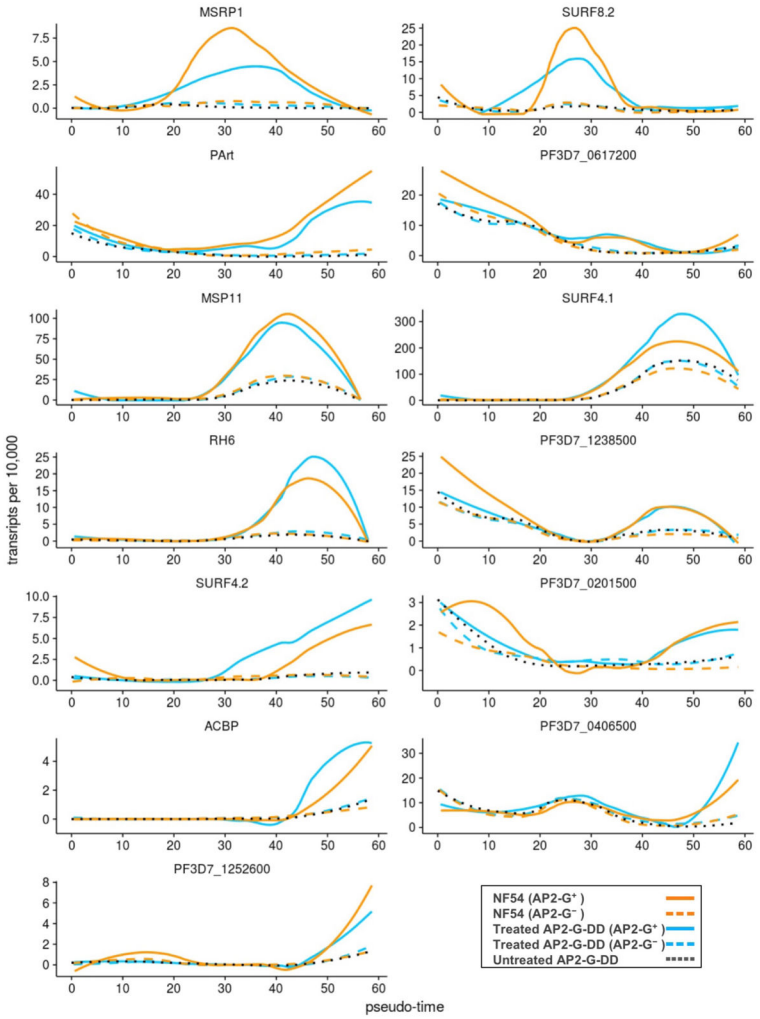
Extended Data Figure 6. Differential expression analysis in treated AP2-G-DD and NF54 cells

The fold change of differentially expressed genes in treated AP2-G+ cells compared to AP2-G- cells in AP2-G-DD SCTs **(A)** or independently clustered NF54 SCTs **(B)**. grey = not detected.



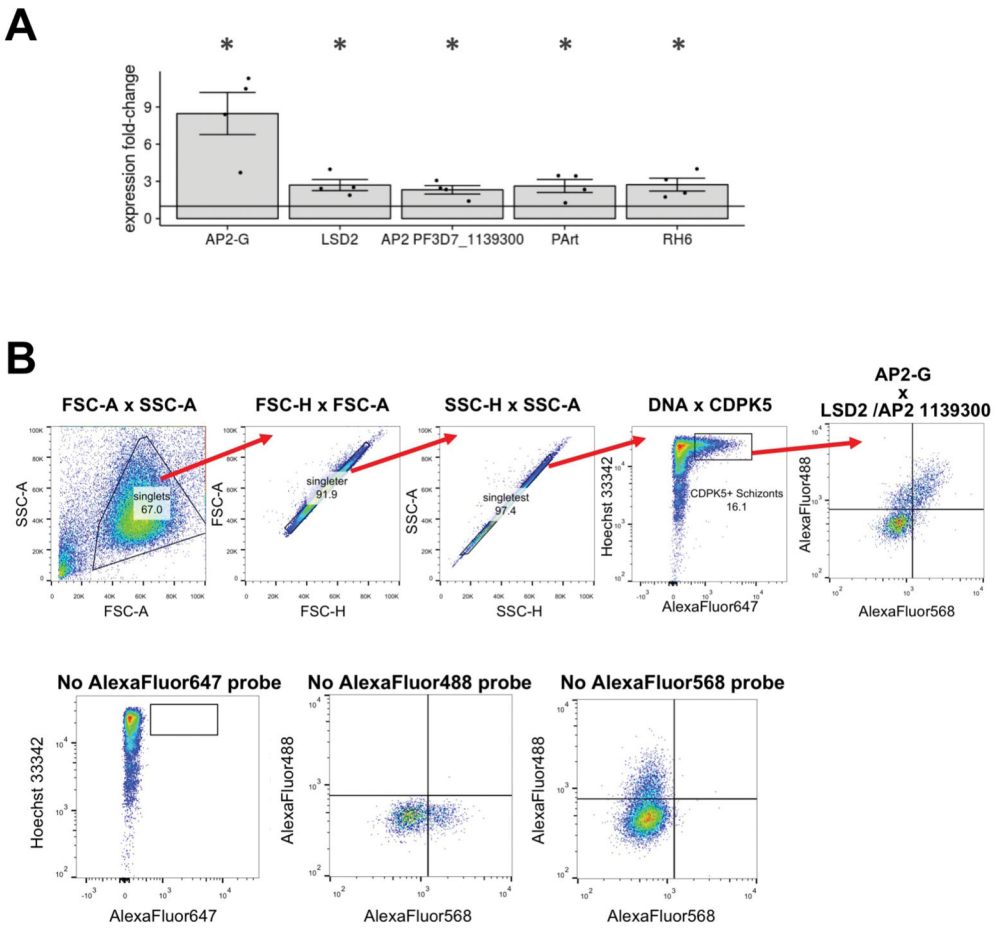
Extended Data Figure 7. Homology-based functional annotation of PF3D7_0801900 as a putative histone lysine specific demethylase

(A) Alignment of conserved blocks (C-blocks) for P. falciparum LSD2 (PF3D7_0801900) and syntenic orthologs in *P. vivax* (PVP01_0118300), *P. ovale curtisi* (PocGH01_01025900), *P. gallinaceum* (PGAL8A_00078100), and *P. berghei* (PBANKA_1228300). Colored residues are conserved. Key functional residues for the flavin-containing amine oxidoreductase (PF01593, yellow) and extended plant homeodomain (PHD) finger (cd15571, green) domains are indicated. **(B)** NCBI Conserved Domain Database hits for PvLSD2 conserved blocks. **(C)** NCBI Conserved Domain Database hits for PfLSD2 conserved blocks. **(D)** Single PANTHER DB functional annotation hit for PfLSD2 conserved blocks.



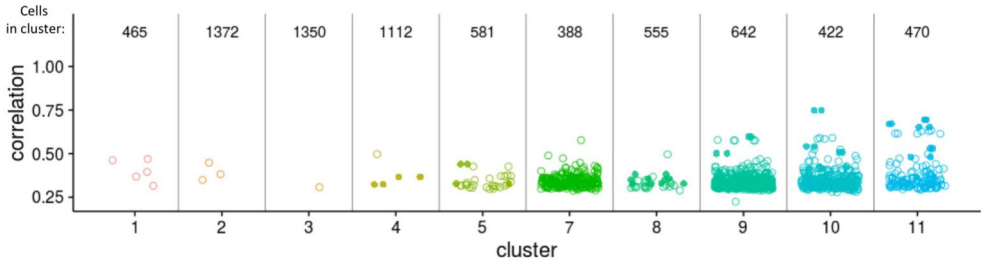
Extended Data Figure 8.

Differential expression of putative regulators in AP2-G^{+} cells (solid) and AP2-G^{-} (dashed) cells of NF54 (orange) and treated AP2-G-DD (blue), as well as in untreated AP2-G-DD cells (dotted black) of Figure 3D hits not shown in Figure 4B.



Extended Data Figure 9. Validation of single cell findings

(A) Expression fold-change in treated vs. untreated AP2-G-DD schizonts as determined by quantitative RT-PCR. Four independent biological replicates, * indicates significantly higher than 1.0 (one-sided p-value < 0.05). Bar height indicates the mean fold-change across replicates, error bars are SEM. **(B)** RNA FISH quantification gating schema (top row) and fluorescence minus-one controls (bottom row) for data shown in Figure 4D. Results are representative of three independent experiments.



Extended Data Figure 10. Expression correlation coefficients for pairs of significantly co-expressed genes by cluster

For each cluster, all genes across all the treated ap2-g-dd and NF54 cells in the cluster were evaluated for co-expression (phi >0.3) with each of the 19 shared hits in figure 3D.

Spearman correlation of expression is shown for highly co-expressed gene pairs. Solid symbols indicate gene pairs including AP2-G. Number of cells used to evaluate co-expression is shown.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the WCM Genomics and Flow Cytometry core facilities, and Gabrielle Suppa for technical assistance. This work was supported by WCM internal startup funds (BK) and the NSF CAREER award (DBI-10549646, OE), LLS SCOR (7006-13 & 7012016, OE), Hirschl Trust Award (OE), Starr Cancer Consortium (I6-A618, OE), NIH 1R01CA194547 (OE). AP and CN were supported by WCM graduate fellowships.

References

1Cowman AF, Healer J, Marapana D, Marsh K. Malaria: Biology and Disease. Cell. 2016; 167:610–624. [PubMed: 27768886]

2Silvestrini F, Alano P, Williams JL. Commitment to the production of male and female gametocytes in the human malaria parasite Plasmodium falciparum. Parasitology. 2000; 121(Pt 5):465–471. [PubMed: 11128797]

3Bruce MC, Alano P, Duthie S, Carter R. Commitment of the malaria parasite Plasmodium falciparum to sexual and asexual development. Parasitology. 1990; 100(Pt 2):191–200. [PubMed: 2189114]

4Kafsack BFC, et al. A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature. 2014; 507:248–252. [PubMed: 24572369]

5Sinha A, et al. A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium. Nature. 2014; 507:253–257. [PubMed: 24572359]

6Brancucci NMB, et al. Heterochromatin protein 1 secures survival and transmission of malaria parasites. Cell Host Microbe. 2014; 16:165–176. [PubMed: 25121746]

7Coleman BI, et al. A Plasmodium falciparum Histone Deacetylase Regulates Antigenic Variation and Gametocyte Conversion. Cell Host Microbe. 2014; 16:177–186. [PubMed: 25121747]

8Pelle KG, et al. Shared elements of host-targeting pathways among apicomplexan parasites of differing life styles. Cell Microbiol. 2015; doi: 10.1111/cmi.12460

9Macosko EZ, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015; 161:1202–1214. [PubMed: 26000488]

10Sims JS, et al. Patterns of gene-specific and total transcriptional activity during the Plasmodium falciparum intraerythrocytic developmental cycle. Eukaryotic Cell. 2009; 8:327–338. [PubMed: 19151330]

11Fivelman QL, et al. Improved synchronous production of Plasmodium falciparum gametocytes in vitro. Mol Biochem Parasitol. 2007; 154:119–123. [PubMed: 17521751]

12Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015; 33:495–502. [PubMed: 25867923]

13Kensche PR, et al. Nucleic Acids Res. 2016; 44:2110–2124. [PubMed: 26578577]

14Qiu X, et al. Single-cell mRNA quantification and differential analysis with Census. Nat Meth. 2017; 14:309–315.

15Dvorin JD, et al. A Plant-Like Kinase in Plasmodium falciparum Regulates Parasite Egress from Erythrocytes. Science. 2010; 328:910–912. [PubMed: 20466936]

16Bozdech Z, et al. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. PLoS Biol. 2003; 1:E5. [PubMed: 12929205]

17Claessens A, Affara M, Assefa SA, Kwiatkowski DP, Conway DJ. Culture adaptation of malaria parasites selects for convergent loss-of-function mutants. Sci Rep. 2017; 7:srep41303.

18Narlikar GJ, Sundaramoorthy R, Owen-Hughes T. Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes. Cell. 2013; 154:490–503. [PubMed: 23911317]

19Volz J, et al. Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in Plasmodium falciparum. International Journal for Parasitology. 2010; 40:109–121. [PubMed: 19765590]

20Park BO, Ahrends R, Teruel MN. Consecutive Positive Feedback Loops Create a Bistable Switch that Controls Preadipocyte-to-Adipocyte Conversion. Cell Reports. 2012; 2:976–990. [PubMed: 23063366]

21Laurent M, Kellershohn N. Multistability: a major means of differentiation and evolution in biological systems. Trends in Biochemical Sciences. 1999; 24:418–422. [PubMed: 10542403]

22Chickarmane V, Troein C, Nuber UA, Sauro HM, Peterson C. Transcriptional Dynamics of the Embryonic Stem Cell Switch. PLoS Comput Biol. 2006; 2:e123. [PubMed: 16978048]

23Bhattacharya S, et al. A Bistable Switch Underlying B-Cell Differentiation and Its Disruption by the Environmental Contaminant 2,3,7,8-Tetrachlorodibenzo-p-dioxin. Toxicol Sci. 2010; 115:51–65. [PubMed: 20123757]

24Guizetti J, Scherf A. Silence, activate, poise, and switch! Mechanisms of antigenic variation in Plasmodium falciparum. Cell Microbiol. 2013; doi: 10.1111/cmi.12115

25Hott A, et al. Artemisinin-resistant Plasmodium falciparum parasites exhibit altered patterns of development in infected erythrocytes. Antimicrobial Agents & Chemotherapy. 2015; 59:3156–3167. [PubMed: 25779582]

26Dzikowski R, Frank M, Deitsch K. Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. PLoS Pathog. 2006; 2:e22. [PubMed: 16518466]

27Moll K, Ljungström I, Perlmann H, Scherf A. Methods in malaria research. Manassas. 2008

28Armstrong CM, Goldberg DE. An FKBP destabilization domain modulates protein levels in Plasmodium falciparum. Nat Meth. 2007; 4:1007–1009.

29Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) Method. Methods. 2001; 25:402–408. [PubMed: 11846609]

30Martin RE, Henry RI, Abbey JL, Clements JD, Kirk K. The ‘permeome’ of the malaria parasite: an overview of the membrane transport proteins of Plasmodium falciparum. Genome Biol. 2005; 6:R26. [PubMed: 15774027]

31Maaten LVD, Hinton G. Visualizing Data using t-SNE. Journal of Machine Learning Research. 2008; 9:2579–2605.

32Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal. 2006

33Lopes CT, et al. Cytoscape Web: an interactive web-based network browser. Bioinformatics. 2010; 26:2347–2348. [PubMed: 20656902]

34Roth FP, Hughes JD, Estep PW, Church GM. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nat Biotechnol. 1998; 16:939–945. [PubMed: 9788350]

35Campbell TL, De Silva EK, Olszewski KL, Elemento O, Llinás M. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. PLoS Pathog. 2010; 6:e1001165. [PubMed: 21060817]

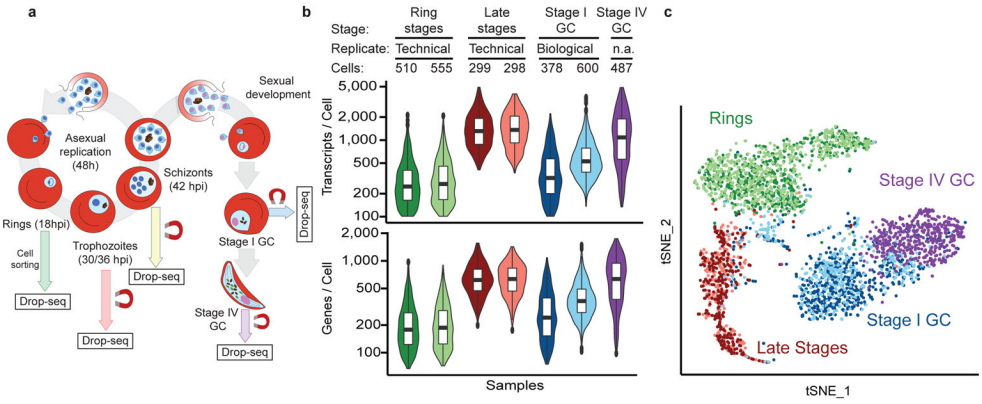


Figure 1. Single cell RNA-seq differentiates between *P. falciparum* life cycle stages in a reproducible manner

(A) Single-cell transcriptomes were collected from early (ring) and late asexual blood stages, as well as stage I and stage IV gametocytes (GC). **(B)** Distributions of unique transcripts per cells and expressed genes per cell for collected stages across technical and biological replicates. n.a. indicates single experiment. The number of cells (n) in each sample is indicated. Boxes indicate the interquartile range; whiskers extend 1.5x interquartile range from the box. **(C)** Unsupervised clustering of 2,829 quality-filtered, single-cell transcriptomes collected from 4 parasite blood stages. Colors match the experiments shown in 1B.

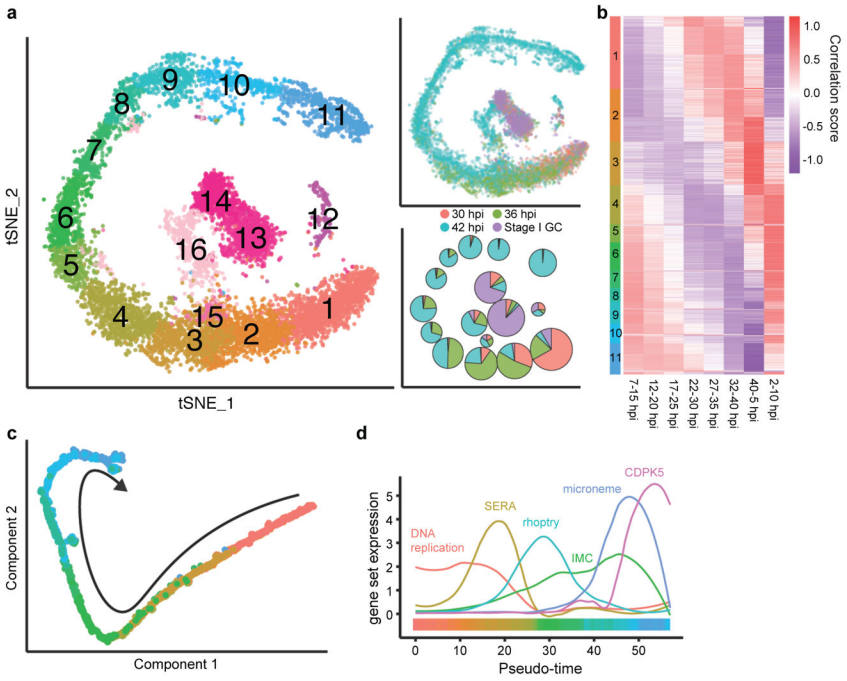


Figure 2. Single cell RNA-seq of malaria parasites successfully captures cell cycle progression and differentiation

(A) tSNE plots of SCTs collected from AP2-G-DD parasites at three asexual time-points and stage I gametocytes colored by cluster assignment (left) or collection time (top right). Bottom right depicts cluster composition by collection time. **(B)** Correlation of SCTs in clusters 1–11 (rows) to a bulk RNA-seq time-course (columns). **(C)** Dimensionality-reduction and pseudo-time ordering of the SCTs assigned to clusters 1–11 (color). **(D)** Scaled expression level of specific parasitic gene modules as a function of pseudo-time. Color bar indicates average cluster assignment along pseudo-time.

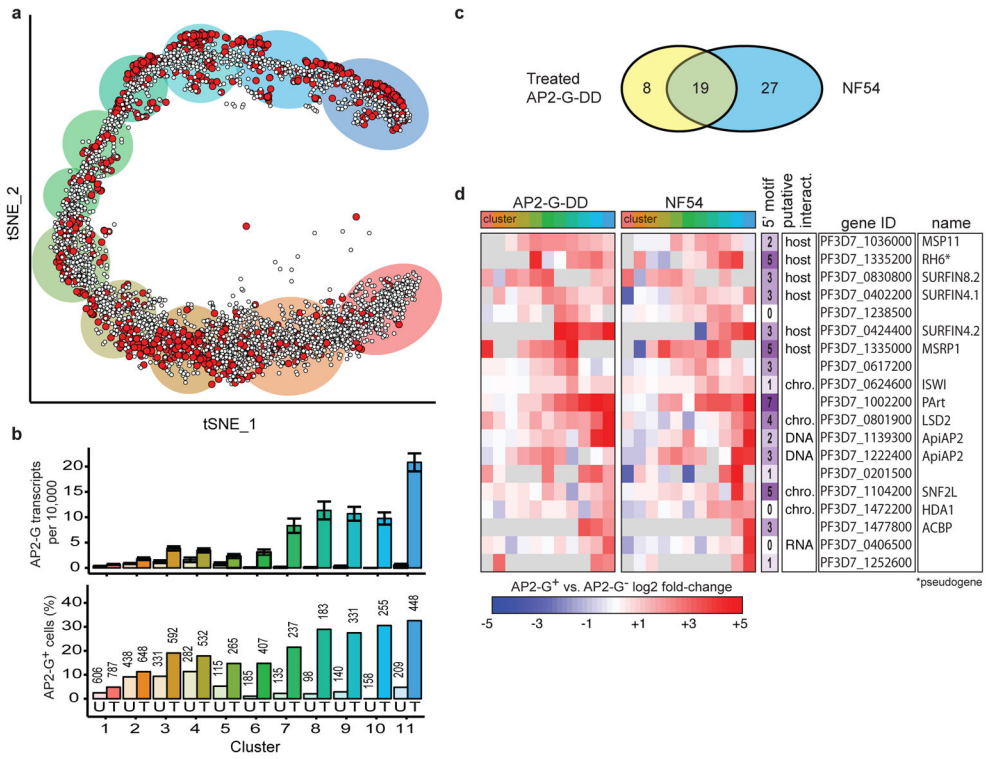


Figure 3. AP2-G specific gene expression in committing parasites

(A) 19.0% of treated AP2-G-DD parasites expressed *ap2-g* (red). Color outlines indicate cluster assignment. **(B)** *ap2-g* in (T)reated and (U)ntreated AP2-G-DD parasites by cluster. Mean expression (top) and percent AP2-G$^+$ (bottom) with number of cells indicated. Error bars are SEM. **(C–D)** Differential gene expression in AP2-G$^+$ vs. AP2-G$^-$ cells of shared hits between AP2-G-DD and NF54 parasites, along with the number of upstream AP2-G binding sites (purple) and putative interaction based on annotation (host= host cell, chro.= chromatin). Grey indicates transcripts were not detected.

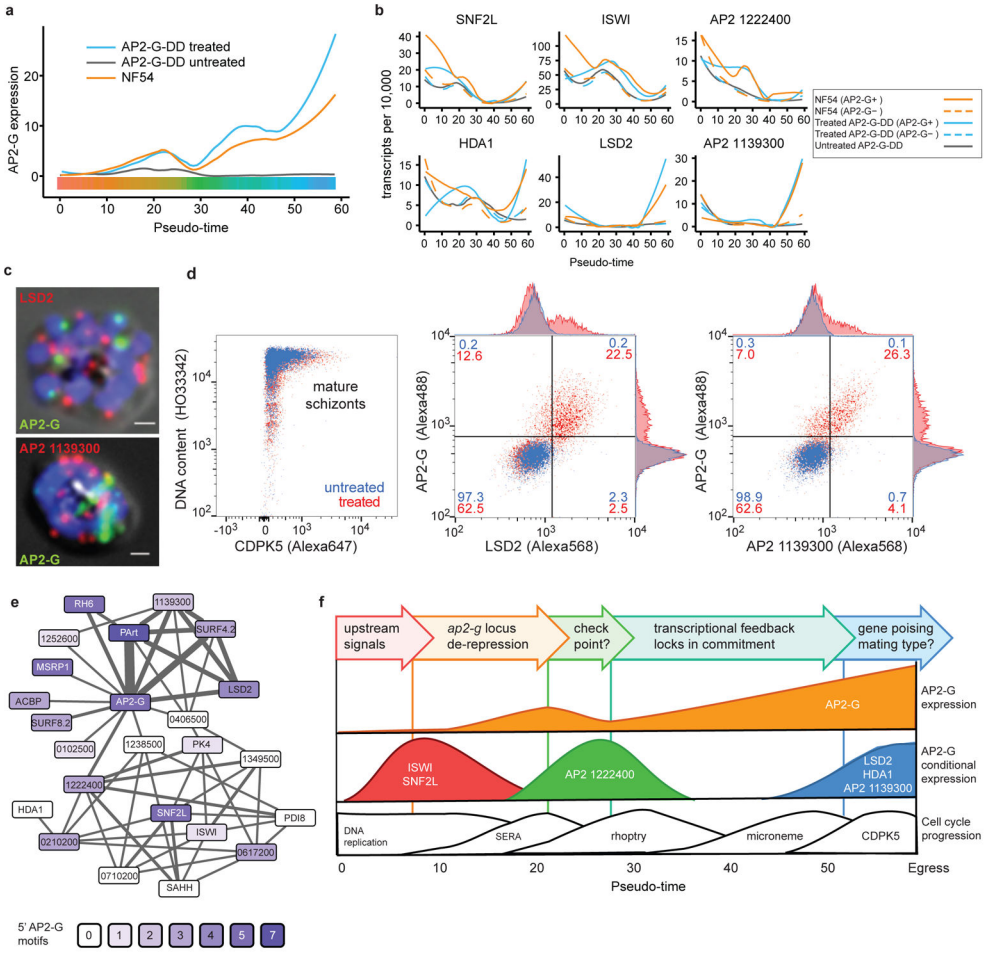


Figure 4. Sexual commitment-specific expression

(A) *ap2-g* expression in NF54 (orange), treated (blue), and untreated (grey) AP2-G-DD parasites versus pseudo-time. **(B)** Expression in AP2-G^{+} (solid) and AP2-G^{-} (dashed) cells of NF54 (orange), treated AP2-G-DD (blue), and untreated AP2-G-DD cells (gray). **(C)** RNA FISH of AP2-G and LSD2 (top) or PF3D7_1139300 (bottom) in schizonts. DNA=blue. **(D)** Co-expression of AP2-G with LSD2 (middle) or PF3D7_1139300 (right) by RNA FISH in treated (red) and untreated (blue) mature (CDPK5^{+}) schizonts (representative of n=3). **(E)** Co-expression network in sexually-committed schizonts. Line thickness indicates number of clusters with co-expression. **(F)** Model of transcriptional regulation during *P. falciparum* sexual commitment.