



RESEARCH ARTICLE

REVISÉ Probing *Plasmodium falciparum* sexual commitment at the single-cell level [version 3; referees: 2 approved]

Nicolas M.B. Brancucci ^{1,2*}, Mariana De Niz ^{1*}, Timothy J. Straub ^{2,3*},
 Deepali Ravel ^{2*}, Lauriane Sollelis ¹, Bruce W. Birren ³, Till S. Voss ^{4,5},
 Daniel E. Neafsey ^{2,3}, Matthias Marti ^{1,2}

¹Wellcome Centre for Molecular Parasitology, Institute of Infection, Immunity & Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

²Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

³The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

⁴Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland

⁵University of Basel, Basel, Switzerland

* Equal contributors

v3 First published: 13 Jun 2018, 3:70 (doi: [10.12688/wellcomeopenres.14645.1](https://doi.org/10.12688/wellcomeopenres.14645.1))
 Second version: 26 Jul 2018, 3:70 (doi: [10.12688/wellcomeopenres.14645.2](https://doi.org/10.12688/wellcomeopenres.14645.2))
 Latest published: 31 Aug 2018, 3:70 (doi: [10.12688/wellcomeopenres.14645.3](https://doi.org/10.12688/wellcomeopenres.14645.3))

Abstract

Background: Malaria parasites go through major transitions during their complex life cycle, yet the underlying differentiation pathways remain obscure. Here we apply single cell transcriptomics to unravel the program inducing sexual differentiation in *Plasmodium falciparum*. Parasites have to make this essential life-cycle decision in preparation for human-to-mosquito transmission.

Methods: By combining transcriptional profiling with quantitative imaging and genetics, we defined a transcriptional signature in sexually committed cells.

Results: We found this transcriptional signature to be distinct from general changes in parasite metabolism that can be observed in response to commitment-inducing conditions.

Conclusions: This proof-of-concept study provides a template to capture transcriptional diversity in parasite populations containing complex mixtures of different life-cycle stages and developmental programs, with important implications for our understanding of parasite biology and the ongoing malaria elimination campaign.

Keywords

Plasmodium falciparum, single cell, RNAseq, phenotyping, malaria, transmission, sexual differentiation

Open Peer Review

Referee Status:

Invited Referees	
1	2
REVISÉ version 3 published 31 Aug 2018	 report
REVISÉ version 2 published 26 Jul 2018	 report
version 1 published 13 Jun 2018	 report report

- Lyn-Marie Birkholtz** , University of Pretoria, South Africa
Riette van Biljon , University of Pretoria, South Africa
Roelof van Wyk , University of Pretoria, South Africa
- Björn F.C. Kafsack** , Weill Cornell Medicine, USA

Asaf Poran, Weill Cornell Medical College,
USA

Discuss this article

[Comments](#) (0)

Corresponding authors: Daniel E. Neafsey (neafsey@broadinstitute.org), Matthias Marti (Matthias.Marti@glasgow.ac.uk)

Author roles: **Brancucci NMB:** Conceptualization, Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **De Niz M:** Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Straub TJ:** Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Ravel D:** Formal Analysis, Investigation, Writing – Review & Editing; **Solleis L:** Formal Analysis, Investigation; **Birren BW:** Data Curation, Funding Acquisition, Writing – Review & Editing; **Voss TS:** Methodology, Writing – Review & Editing; **Neafsey DE:** Conceptualization, Formal Analysis, Funding Acquisition, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Marti M:** Conceptualization, Investigation, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The project was supported by Senior Investigator Award 172862 and IRS Award 172805 from the Wellcome Trust and a career development award from the Burroughs Wellcome Fund to M.M., and a Centre Award 104111 to the Wellcome Trust Centre for Molecular Parasitology, University of Glasgow. The project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Grant Number U19AI110818 to the Broad Institute. The project was further supported by Swiss National Science Foundation grants (31003A_163258 and BSCGI0_157729) to T.S.V.. N.M.B.B and M.D.N. received PostDoc.Mobility fellowships from the Swiss National Science Foundation (P300PA_160975 and P2BEP3_165396, respectively), and M.D.N. received a Long Term EMBO postdoctoral fellowship (EMBO ALTF 1048-2016).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2018 Brancucci NMB *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Brancucci NMB, De Niz M, Straub TJ *et al.* **Probing *Plasmodium falciparum* sexual commitment at the single-cell level [version 3; referees: 2 approved]** Wellcome Open Research 2018, 3:70 (doi: [10.12688/wellcomeopenres.14645.3](https://doi.org/10.12688/wellcomeopenres.14645.3))

First published: 13 Jun 2018, 3:70 (doi: [10.12688/wellcomeopenres.14645.1](https://doi.org/10.12688/wellcomeopenres.14645.1))

REVISED Amendments from Version 2

In this second revision we have clarified the use of UMIs vs unique reads. We have also changed the graph in [Figure 4a](#) and the figure legend to clarify the phenotype of the Pf2004/164tdTom PMT KO line.

See referee reports

Introduction

Malaria remains a major global health issue, with roughly 200 million infections and more than 400,000 fatal cases caused by *Plasmodium falciparum* each year¹. Despite decades of disease-control and elimination efforts, *P. falciparum* persists in many geographic regions, highlighting the adaptability of this parasite to changing environments. High levels of antigenic diversity have compromised efforts to develop efficacious vaccines, and resistance has evolved to all licensed antimalarial drugs^{2,3}. Malaria parasites have a complex life cycle, and patient blood samples usually contain a mixture of asexually replicating parasites and a small fraction of terminally differentiated sexual-stage parasites. The latter, so-called mature gametocytes are required for parasite transmission to mosquitos. Blood-stage parasite isolates from malaria patients, or from resistance selection or *in vitro* culture adaption experiments, show significant heterogeneity in the transcriptional profile in population-level expression analyses^{4–6}. To date, the transcriptional diversity of such mixed populations has not been captured appropriately, owing to lack of efficient single-cell mRNA profiling methods in *Plasmodium*. In this study, we present a cost-effective and high-throughput pipeline for single-cell transcriptional and phenotypic analysis of *P. falciparum* parasites. We used a digital gene expression (DGE) protocol⁷ to define the transcriptional signature during initiation of parasite sexual differentiation (i.e. sexual commitment) and correlated mRNA profiles with microscopy-based phenotyping. Our study provides a template for capturing transcriptional diversity in heterogeneous parasite populations, which we hope will springboard future endeavors in single cell transcriptomics of *Plasmodium*.

To ensure transmission, malaria parasites invest in the production of gametocytes. During each intra-erythrocytic developmental cycle, a subset of asexually replicating parasites commits to produce sexual progeny⁸. Rates of sexual commitment generally vary between species and conditions, and range from 1–30% in *P. falciparum*⁹. Previous work suggested that one asexual parasite can produce either strictly sexual or strictly asexual progeny (i.e., be sexually or asexually committed)¹⁰; however, conclusive evidence is lacking and the transcriptional profile of the sexually committed parasite has been poorly defined^{11,12}. Two studies have identified AP2-G as a transcription factor that is activated during sexual commitment and required for gametocyte formation^{13,14}. Initial analysis of gene expression in individual sexually committed cells using conditional AP2-G knock-down provided a first transcriptional signature upon AP2-G activation¹⁵. We have recently demonstrated that availability of the host-derived phospholipid lysophosphatidylcholine (LysoPC) acts as an environmental sensor for the regulation

of sexual commitment in *P. falciparum*¹⁶. We have shown that LysoPC is the major building block for phosphatidylcholine biosynthesis in the parasite, and its depletion results in the induction of a compensatory metabolic response and triggers AP2-G activation and sexual commitment in a large part of the parasite population. This observation was facilitated through development of a highly controlled and quantitative *in vitro* assay to induce gametocyte formation^{16,17}. Here we apply this assay to define the transcriptional signature of individual cells at different stages during sexual commitment and validate key findings experimentally.

Results

Development of a single-cell RNA-sequencing (scRNA-seq) pipeline in *P. falciparum*

To capture the transcriptional profile of sexual commitment, cells were exposed to a pulse of defined medium lacking LysoPC (–SerM), as described previously¹⁶. At three time points after LysoPC depletion, 336 individual cells (and 48 control cells in the presence of LysoPC; –SerM/LysoPC) were collected by flow sorting and snap-frozen prior to being processed for DGE ([Figure 1a](#) and [Supplementary Figure 1](#)). We utilized DGE owing to its capacity to quantitatively profile relative expression levels of genes from large numbers of individual cells without investment in specialized hardware, for a total cost of approximately \$30/cell. Reads representing the 3' end of transcripts from each cell were aligned to the *P. falciparum* reference genome (PlasmoDB version 29) and filtered for unique molecular identifier (UMIs) to avoid repeat sampling of the same original RNA molecules, and ribosomal RNA (rRNA) species were removed ([Table 1](#) and [Supplementary File 1](#)). Across cells we detected 3110 genes of the ~4900 genes transcribed at some level in blood stage parasites¹⁸, and a smaller gene set was represented by multiple reads in the majority of cells. We considered genes to be detected if they exhibited at least 15 UMIs among the 881 cells that met our minimum sample quality criteria (described in the Methods). The 500 most highly transcribed genes account for approximately 65% of UMIs across all cells, while the 100 most highly transcribed genes account for approximately 40% of the UMIs ([Figure 1b](#)). We found the number of UMIs per cell to vary across the three time points analyzed due to variation in library quality ([Figure 1c](#) and [Supplementary Figure 2](#)), but cells in the second and third time points exhibited an average of 841 and 1118 UMIs, respectively. Principal component analysis (PCA) of normalized UMIs from highly expressed genes clustered individual cells by time point ([Figure 1d](#)), demonstrating that stage-specific differences in transcriptional profiles are detectable in single *P. falciparum* parasites. Comparison of single cell expression profiles across the three time points to a previously published conventional bulk transcriptomic time series¹⁹ confirms that the stage-specific differences we observe are indicative of cell-cycle progression rather than batch effects (linear regression, $R^2 = 0.43$; $p < 2.2 \times 10^{-16}$). Comparison of transcriptional profiles across time points revealed significantly reduced UMIs observed in cells grown under –SerM conditions compared to control (Wilcoxon rank sum, $p = 0.007$) ([Figure 1e](#)), likely reflecting the reduced merozoite numbers we previously observed under these conditions¹⁶. Next, we compared the transcription levels per gene across single cells with those from the same time points from a population-level

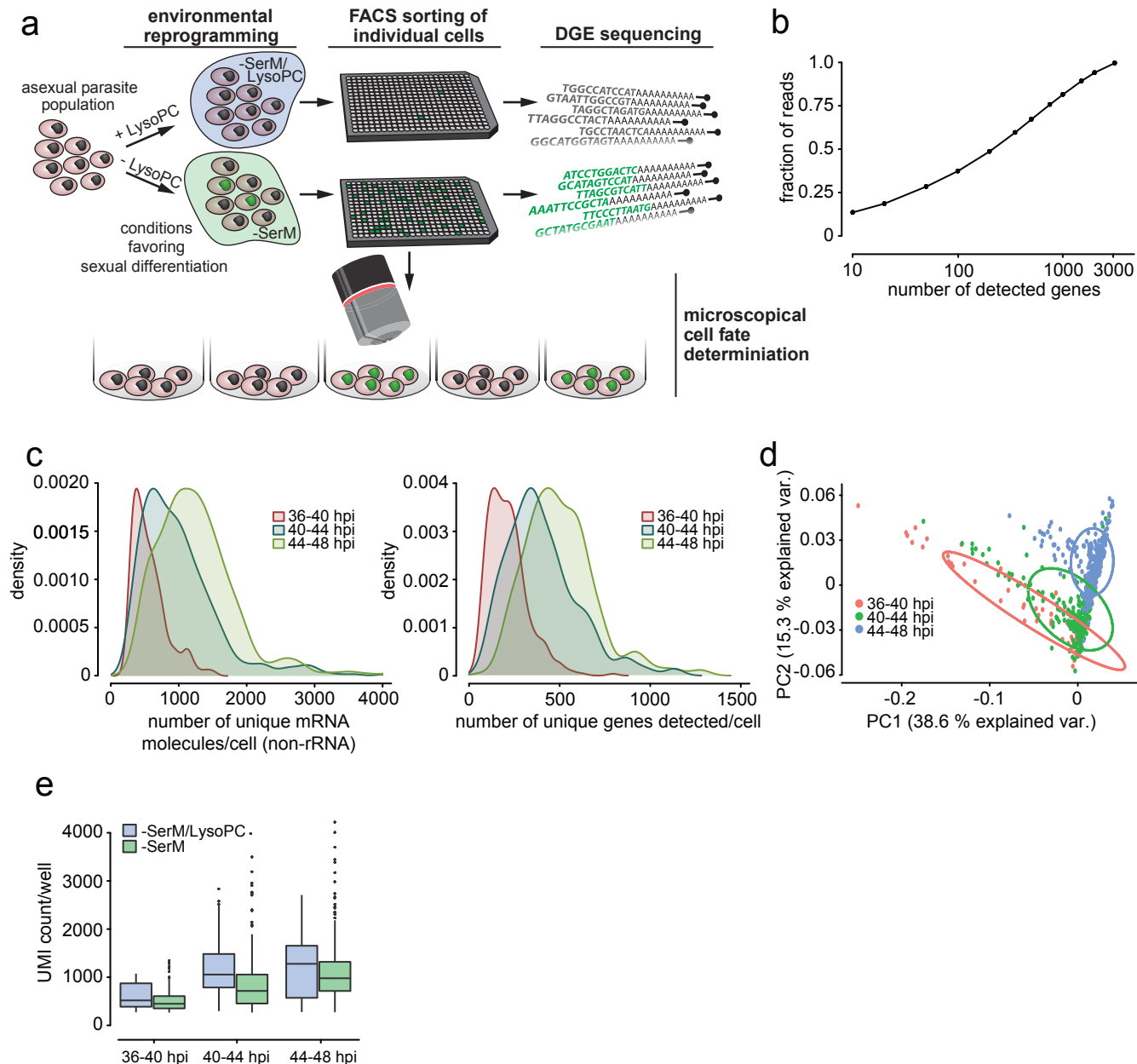


Figure 1. A digital gene expression (DGE) platform for *P. falciparum* single-cell mRNA profiling. (a) Sexual commitment and experimental setup. Synchronized parasites of the Pf2004/164tdTom line are split into 2 populations, one exposed to inducing conditions using lysophosphatidylcholine (LysoPC)-deficient medium (-SerM) and one exposed to -SerM supplemented with LysoPC. Individual cells from each population are sorted by FACS for subsequent scRNAseq analysis. The remainder of cells from both populations is maintained in culture for measurement of parasite multiplication rate and sexual conversion. Sexually committed cells are denoted in green. (b) DGE read counts per gene, sorted by highest to lowest expression in DGE. Approximately 200 genes account for 50% of all DGE reads. (c) Distribution of DGE transcript counts (left panel) and genes detected (right panel) per cell, colored by time point. Density units are arbitrary but reflect the number of cells at a given value in the distribution. As read counts differed among and between time points, they were normalized per time point per cell prior to further analysis. (d) Principal component analysis shows that single cell transcriptomes cluster according to time point. (e) Unique mRNA transcripts per cell across time points are significantly lower in those incubated in SerM compared to those in -SerM supplemented with LysoPC.

RNA-seq experiment¹⁶. The comparison demonstrated that overall expression levels per gene are significantly, though weakly, correlated between single-cell DGE and population-level RNA-seq (F-test, $p < 2.2 \times 10^{-16}$). Notably, the weak correlation we observe

between the two transcriptional datasets highlights the known gene “drop-out” effect of single-cell sequencing^{20,21}. Altogether, these experiments demonstrate that the DGE platform presented here is able to capture mRNA profiles of single *P. falciparum*

Table 1. Mean values per cell. Mean read counts, UMIs and detectable genes per cell at each step in the processing pipeline shown for each time point.

Time point, hpi	Assigned reads	Aligned reads	UMIs	UMIs after excluding rRNA	Genes detectable
36-40	962,253	142,730	7,648	377	212
40-44	1,395,384	292,130	19,069	841	404
44-48	1,128,064	416,588	27,002	1,118	503

UMI, unique molecular identifier.

parasites at sufficient depth to i) detect transcriptional differences between 4-hour time points in the cell cycle and ii) recapitulate overall transcriptional profiles from population-level RNA-seq experiments.

Defining the transcriptional signature of sexual commitment
LysoPC depletion induces sexual commitment in 25–35% of parasites across strains, while the remaining cells continue asexual replication¹⁶. In the population-level experiment, we previously identified a set of 342 genes that were up-regulated upon LysoPC depletion, including the known commitment markers *ap2-g*^{13,14} and *gdl1*²². Using the single cell mRNA profiles, we observed overall concordance in the induction signature when compared to the bulk RNA-seq data (Figure 2a and Supplementary File 2). The increase of cells with detectable *ap2-g* transcripts upon LysoPC depletion (Figure 2b) with the proportion of sexual progeny in the same experiment, confirming that *ap2-g* activation in a cell is predictive of its differentiation state. Considering the mixed commitment of parasites to either the sexual or the asexual pathway under inducing conditions *in vitro*¹⁷, the transcriptional profile from our population-level RNA-seq experiment likely represented a combination of different transcriptional signatures. Specifically, we hypothesized that the changes induced by LysoPC depletion represent both a general response in all cells as well as a sexual commitment-specific activation of *ap2-g* and other factors required to initiate sexual differentiation in a subset of cells. Single-cell consensus clustering (SC3)²³ based on gene expression profiles from 40–48 hpi revealed robust and reproducible separation of individual cells into 12 clusters (Figure 2c, Supplementary Figure 3, Supplementary Figure 4, and Supplementary File 3). Detectable or elevated expression of 10 genes (Figure 2d), including *ap2-g* and *set9*, defined a subset of cells (cluster 11), presumably representing the minimal transcriptional signature of sexually committed cells. This signature was significantly enriched in 36% of all cells (205/560; Wilcoxon rank sum test, corrected $p < 0.05$) comprising five cell clusters (4, 5, 10, 11, and 12; Figure 2d). A total of 125 genes were significantly co-expressed with this transcriptional signature, including several kinases, epigenetic regulators, transcription factors and multiple resident rhoptry proteins (Figure 2e and Supplementary Figure 4). By contrast, ethanolamine kinase (*ek*) and phosphoethanolamine methyl transferase (*pmt*), two key enzymes that are induced upon LysoPC depletion¹⁶ and needed for phosphatidylcholine biosynthesis under these sexual commitment-inducing conditions, were equally expressed across all cells (Figure 2d), rather than co-expressed with this transcriptional signature (Figure 2e).

Comparative analysis of our single-cell data with those from Poran *et al.*¹⁵ and the bulk RNA-seq dataset validated the results by revealing a significant overlap in genes with commitment-associated expression profiles, despite differences in experimental setup and sequencing pipelines (Table 2 and Supplementary File 5). Poran *et al.*¹⁵ also observed an enrichment of AP2-G binding motifs upstream of commitment-associated genes. We observe a similar enrichment of genes with upstream AP2-G binding sites within 2 kb of the translational start site (as defined in Campbell *et al.*²⁴) within our commitment signature gene set (117 out of 125; $p = 0.002$, Fishers exact test, indicating that a significant proportion of the loci we detected may be downstream of AP2-G activation. Altogether, single-cell transcriptional profiling demonstrated that LysoPC depletion induces both a population-wide transcriptional response and a sexual commitment-specific response in a subset of cells.

Schizonts produce either asexual or sexual progeny

Identification of a distinct transcriptional signature in AP2-G-expressing (AP2-G+) cells supports the hypothesis that sexual and asexual progeny are non-uniformly distributed among schizonts. This hypothesis is also reinforced by the recent scRNA-seq study¹⁵, and by previous work using plaque assays and progeny analysis with gametocyte-specific antibodies¹⁰. To validate these observations directly, we investigated sexual commitment in reporter parasites. Parasites were grown in –SerM and –SerM/LysoPC conditions and separated into individual wells for further development. Analysis of progeny from single schizonts confirmed the absence of mixed differentiation states, with progeny numbers corresponding to the number of merozoites per parental schizont (Figure 3a). Moreover, the prevalence of reporter-positive schizonts corresponded to the number of positive progeny (i.e., sexual rings) in the same culture upon LysoPC depletion (Figure 3b). Notably, merozoite numbers were similarly reduced in both sexually committed and asexually committed schizonts in –SerM medium compared to control (Figure 3b, c). These data support the observation of a population-wide transcriptional response, in agreement with the transcriptional activation of *ek* and *pmt* in all cells under –SerM conditions (Figure 2e).

Experimental validation of induction and sexual commitment signatures

To test whether PMT activity is essential in absence of LysoPC (in all cells, irrespective of their differentiation state) and required for sexual commitment, we generated *pmt* knock-out parasites using the CRSIPR/Cas9 technology. Genetic disruption of *pmt*

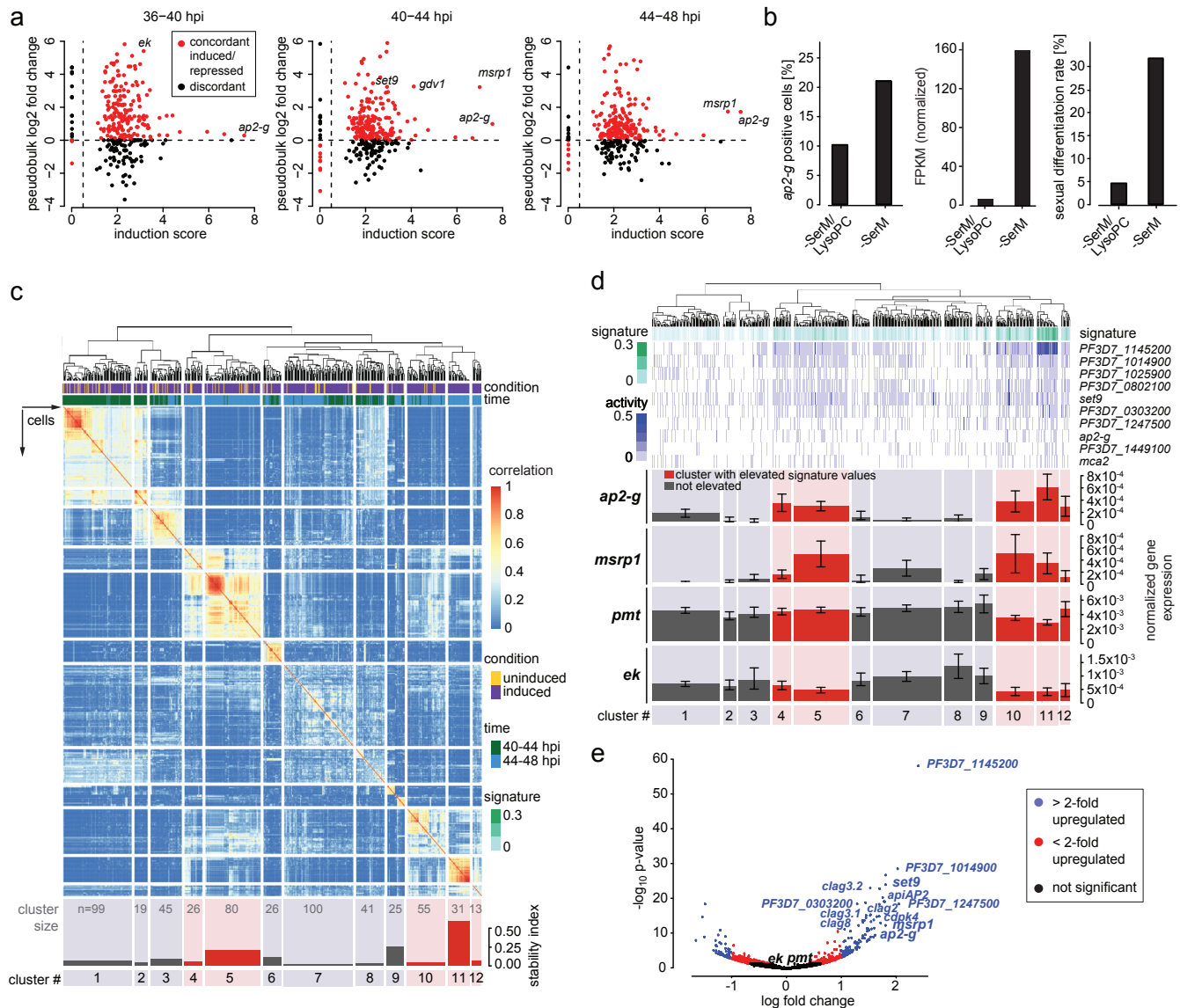


Figure 2. The transcriptional signature of sexually committed cells. (a) Concordance of differential expression between pooled digital gene expression (DGE) reads (pseudobulk, shown as \log_2 fold change) and population-level RNA-seq (defined as \log_2 sum of the ratios of fragments per kilobase of transcript per million mapped reads in inducing over non-inducing conditions, or induction score¹⁶). (b) Comparison of *ap2-g* expression in single cells by DGE (left panel) and in population-level RNA-seq at 44-48 hpi (center). Sexual differentiation rates of cells used for DGE sequencing are shown in the right panel. (c) Single-cell consensus clustering (SC3) defines a signature of sexual commitment. The subset of genes that were significantly differentially expressed from population-level RNA-seq was used to perform consensus clustering on the single cell RNA-seq gene expression. On top, a dendrogram showing k-means clustering results of single cells from 40-44 and 44-48 hpi. Just below, two parameters are shown: i) condition - induced cells are demarcated in purple and uninduced cells are demarcated in yellow; ii) time - cells from 40-44 hpi are in green and those from 44-48 hpi in blue. The rows and columns of the heat map are single cells that have been grouped into 12 clusters on the basis of their expression profiles (coloration indicating strength of clustering). The stability index (bottom) indicates how often those cells clustered together using different k values. Time point and condition from which the cells were derived are indicated, as well as positive expression of several genes of interest. Because *pmt* was highly expressed in most cells, only those with 80th percentile or higher expression are marked. Cluster 11 revealed cells with high *ap2-g* expression and cluster stability. Cluster size is indicated. (d) Transcriptional signature of commitment is enriched in five cellular clusters. Marker genes for SC3 cluster 11 were used to define a transcriptional signature of sexual commitment (see methods). This signature is a metric built on averaging normalized marker gene expression associated with cluster 11 (higher signature values are indicated with darker shades of green). The signature is shown in green above the heat map containing each gene's individual expression values. Below, mean expression values and the standard error for *ap2-g*, *msrp1*, *pmt*, and *ek* are shown for each cluster. Clusters highlighted in red had significantly elevated signature values. Expression of *ap2-g* and *msrp1* is significantly enriched in the clusters highlighted in red, while expression of *pmt* and *ek* are not. (e) Sexually committed cells show increased expression of 125 genes. Single cells from 40-48 hpi separate into subsets with high and low marker signature (above and below the 80th percentile, respectively), with 125 differentially expressed genes measured by scRNA-seq, including *ap2-g*, *msrp1* and *set9*. Plotted *p* values reflect Benjamini-Hochberg correction.

Table 2. Comparison of hits between the three studies. These are results from the R package SuperExactTest, which statistically tests if there is an enrichment of overlapping genes in multiple sets²⁵. The expected overlap is the null hypothesis that there is only random sampling of genes. The p-value is the one-tail probability of observing equal or larger overlapping genes at random. The background size was set to 5,800, which is how many transcripts there are in v29 of the 3D7 transcriptome. For comparison, we used hits from Brancucci *et al.* (bulk)¹⁶, this study (scRNA-seq data in [Supplementary File 4](#), all hits with positive fold change in last two time points; sc) and Poran *et al.*¹⁵ (scRNA-seq data in table S3, all hits; Poran).

Intersections	Degree	Observed overlap	Expected overlap	FE	p-value
Brancucci <i>et al.</i> ¹⁶ (bulk; Supplementary File 2 , TP3/4)	1	268			
Poran <i>et al.</i> ¹⁵ (Table S3)	1	28			
Brancucci <i>et al.</i> (sc; Supplementary File 4 + FC)	1	125			
Poran & bulk	2	9	1.29	6.96	2.67E-06
Poran & sc	2	3	0.60	4.97	0.022
sc & bulk	2	14	5.78	2.42	0.0017
Poran & sc & bulk	3	2	0.03	71.73	3.64E-04

Fe, fold enrichment over null; FC, fold change; sc, single cell.

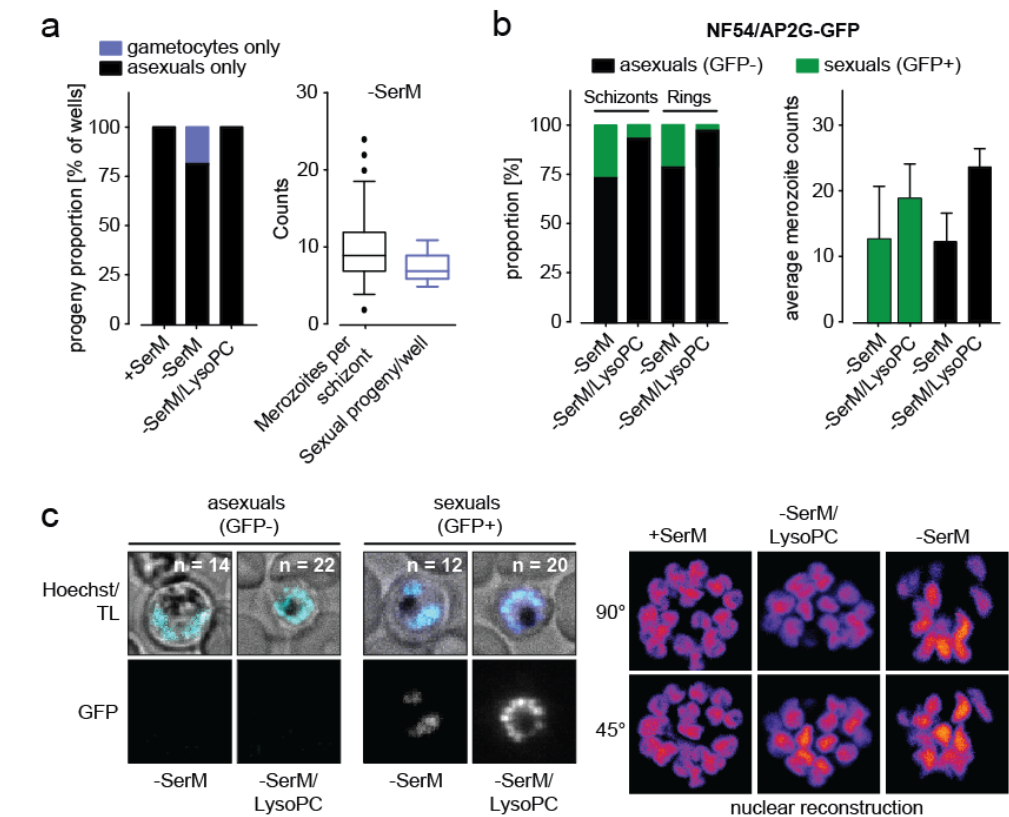


Figure 3. Progeny analysis in single cells. (a) Analysis of progeny from 96 single cells in the Pf2004/164tdTom reporter line (expressing tdTomato reporter under the gametocyte-specific *PF10_0164* promoter¹⁷). Each parasite produces only asexual or gametocyte progeny (left panel). The number of sexual progeny corresponds to the number of daughter merozoites per schizont under lysophosphatidylcholine-deficient medium (–SerM) conditions (right panel). A total of 50 independent measurements were made per condition. (b) Analysis of sexually committed schizonts in the NF54/AP2-G GFP reporter line. The proportion of GFP positive schizonts and the number of GFP positive progeny (rings) correlate and mark sexual cells (gametocyte rings). Both sexually committed (GFP-positive) and asexually committed schizonts show reduced merozoite number under –SerM conditions (right bar graph). GFP-positivity was quantified in 100'000 cells by ImageStream analysis in biological triplicates. Merozoite counts represent the mean of 100 measurements per condition, performed in biological triplicates. (c) Representative images from cells in (b) and selected planes from 3D reconstructions of nuclei are shown in the left and right panels, respectively. Hoechst, nucleic acid stain; TL, transmitted light. n = number of nuclei.

demonstrated that parasites lacking PMT activity are unable to grow in the absence of LysoPC, while growth is unaffected under control conditions. By contrast, sexual commitment levels were not altered in *pmt* knock-out parasites, demonstrating that activation of sexual commitment is independent of the compensatory activation of this metabolic enzyme under LysoPC-limiting conditions (Figure 4a and Supplementary Figure 5, Supplementary Figure 6)¹⁶. In support of these data, imaging flow cytometry also revealed increased protein expression of the second enzyme involved in producing choline from the alternative substrate ethanolamine (ethanolamine kinase, *ek*) in all cells under –SerM conditions (Figure 4b). Confirming the data obtained from the SC3 approach described above, these results show that while EK and PMT are induced upon LysoPC depletion, activation of these enzymes is not specific to the pathway inducing sexual commitment. By contrast, we observed a titratable increase in AP2-G+ cells by inhibition of the essential Kennedy pathway enzyme choline kinase using the specific inhibitor BR23^{16,26} (Figure 4c and Supplementary Figure 6), indicating that a metabolite or enzymatic activity downstream of this reaction translates the LysoPC signal into a cellular response favoring asexual replication over sexual commitment.

As stated above, we also observed enrichment of multiple rhoptry markers in our sexual commitment signature (Figure 2d, e), including *RhopH1/2/3* and several rhoptry neck proteins. Rhoptry genes were generally transcribed at higher levels in all AP2-G+ cells across all three time points, even when accounting for time point and –SerM conditions ($t=4.7$, $p=3.1\times10^{-6}$) (Figure 4d). In agreement with these transcriptional data, analysis of RhopH3 protein expression over time by immunofluorescence microscopy confirmed significantly increased levels in sexually committed cells (AP2-G+) compared to AP2-G- cells and control conditions (Figure 4d). RhopH3 is present in rhoptries and required for merozoite invasion into red blood cells (RBC), and at the infected RBC surface as part of a nutrient channel²⁷. The protein is initially localized to rhoptries and upon invasion it is translocated into the host cell cytosol and surface membrane. Our image analysis demonstrated significantly increased RhopH3 localization in rhoptries but not in the host cell in AP2-G+ cells compared to AP2-G-, confirming that *de novo* synthesis of this protein is elevated in sexually committed schizonts (Figure 4e). Altogether these data validate the findings from the SC3 analysis and highlight the complex nature of the parasite response to changes in LysoPC availability.

Discussion

We have applied scRNA-seq to investigate the signature of sexual commitment in *P. falciparum* malaria parasites, with particular focus on the events that precede the activation of the transcriptional master switch AP2-G.

We have recently identified the host phospholipid LysoPC as a natural repressor of commitment, and LysoPC depletion as a means to induce AP2-G activation *in vitro*¹⁶. Combining this assay with scRNA-seq has revealed a transcriptional signature in sexually committed cells. As expected, there was only modest (but significant) overlap with the signature defined by Poran

et al., as their experiment was not designed to profile commitment-associated processes upstream of AP2-G activation. Apart from *ap2-g*, the only commonly activated gene detected by both studies was the merozoite antigen *msrp1*^{28,29}. *msrp1* has an AP2-G-binding site in its upstream sequence¹⁵, suggesting that it is activated by (and therefore downstream of) AP2-G. We observed significant overlap between our scRNA-seq and the population RNA-seq data¹⁶, as the same experimental set-up was used in these two studies. Shared hits include two kinases, a histone methyl transferase and two additional ApiAP2 transcription factors, further supporting the hypothesis that intracellular signaling events translate nutrient availability into epigenetic changes that ultimately regulate AP2-G. Using a combination of scRNA-seq, quantitative imaging of single cells and reverse genetics we demonstrate that the compensatory activation of *ek* and *pmt* is a general response to nutrient depletion and independent of parasite sexual commitment. Our work also demonstrates that PMT is essential for asexual growth in absence of LysoPC (or choline) as a substrate for PC synthesis. This finding reiterates the compensatory metabolic role of PMT under LysoPC-limiting conditions. By contrast, inhibition of the first enzyme in the Kennedy pathway, choline kinase, using the specific inhibitor BR23 activates AP2-G expression, even in the presence of LysoPC. These data suggest that a metabolite or enzymatic activity in the Kennedy pathway downstream of choline kinase is necessary to repress AP2-G activation. LysoPC uptake is drastically reduced in gametocytes¹⁶, suggesting that PMT activation is both a general response to nutrient depletion and preparation for increased requirement of this alternative substrate arm during gametocyte development. Indeed, previous work has demonstrated that PMT is essential for gametocyte maturation and mosquito infection^{30,31}.

Malaria parasites are exposed to changing environments during their life cycle, and these changes usually coincide with developmental switches. In addition, blood-stage parasites have to adapt to constantly changing nutrient availability. Recent work has demonstrated that the major energy source for the parasite, glucose, is used as an environmental sensor to modulate progeny number and hence virulence via a signaling cascade³². Surprisingly, no effect on transmission was observed under caloric restriction, and glucose complementation did not block sexual conversion in our study¹⁶. These findings suggest the existence of several environmental sensing pathways in the parasite that regulate parasite growth, transmission and possibly other phenotypes, such as immune evasion via antigenic variation. The pathways must be finely tuned in order to ensure an optimal balance between growth and transmission, rather than a population-level shift towards one state. We hypothesize that the decision for commitment is regulated through a bistable biological switch reinforced through a positive feedback loop, such as the known auto-activation of AP2-G^{13,15} and possibly involves kinase-mediated signaling upstream of this transcription factor. In such a scenario, individual cells may show stochastic variation in the threshold concentration of LysoPC required to induce commitment and activate AP2-G. Such threshold-dependent regulation may explain why sexual commitment rates in *P. falciparum* are generally confined to upper limits (i.e., below 50%). While commitment rates can be elevated beyond these limits

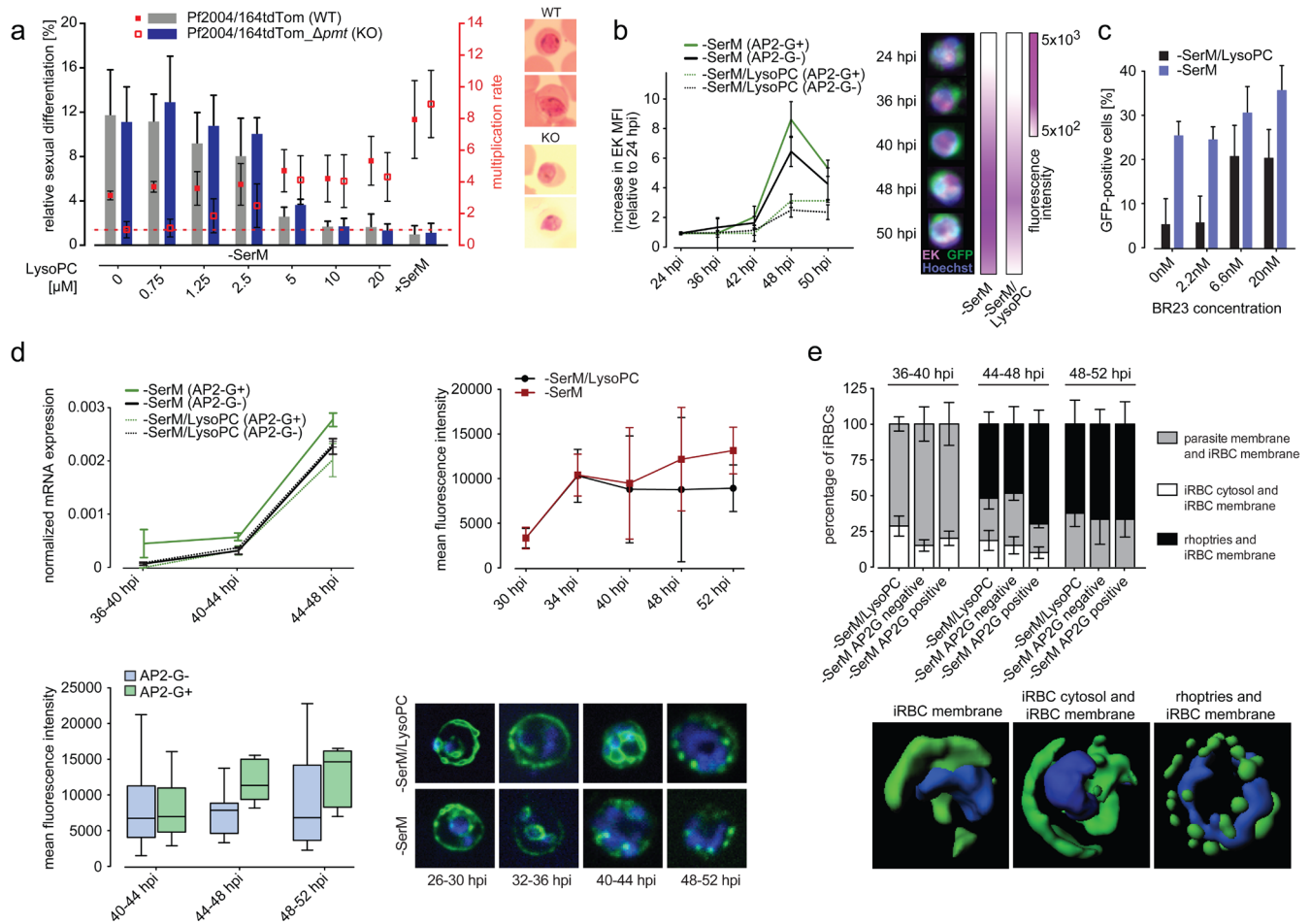


Figure 4. Experimental validation of the single-cell RNA-seq signature. (a) Phosphoethanolamine methyl transferase (PMT) is not required to trigger sexual commitment. *pmt* knock-out parasites (Pf2004/164tdTom_Δ*pmt*) induce sexual commitment in response to lysophosphatidylcholine (LysoPC) depletion at rates that are similar to those of control parasites (left axis). By contrast, asexual growth is significantly impaired in the knock out cell line under LysoPC limiting conditions (right axis). Differences in parasite multiplication increase with lower lipid concentrations, demonstrating that *pmt* activity becomes essential for the compensatory metabolic response induced at low LysoPC levels. The horizontal dashed line indicates complete growth inhibition (multiplication rate of 1). Right panel: Giemsa smears of parasites at the time point of gametocyte readout. No difference in morphology between WT and KO parasite is observed at 0 μM LysoPC, suggesting normal gametocyte development until Stage II. (b) Analysis of EK expression in the NF54/AP2-G GFP reporter line. EK expression increases significantly over time and is induced under LysoPC-deficient medium (-SerM) conditions. Cells were stained with anti-EK antibody (far red), anti-GFP antibody (green) and the nuclear dye Hoechst 33342 (blue). All parasites (AP2-G+ and AP2-G-) show increased EK expression under -SerM conditions. Representative composite images are shown. Data in the left panel are the result of 100,000 measurements performed by ImageStream in biological duplicate experiments with technical triplicates. Error bars represent interquartile ranges. (c) Choline kinase inhibition activates AP2-G expression. Specific inhibition of choline kinase by BR23 induces sexual commitment at growth-limiting concentrations (IC50 of BR23: 3.5 nM). (d) Rhoptry gene expression is increased in AP2-G expressing cells. The average normalized gene expression from DGE was calculated for all rhoptry-related genes. Cells expressing AP2-G have significantly higher levels of overall rhoptry gene transcription ($p=0.001$), even when accounting for time, LysoPC treatment, and raw transcript counts (upper left panel). RhopH3 expression is increased during sexual commitment compared to control. Time course analysis comparing RhopH3 expression between -SerM and -SerM/LysoPC control (upper right panel), and between AP2-G+ and AP2-G- (lower left panel; $p = 0.002$ for time point 2 and $p = 0.01$ when all time points are pooled). Representative images per time point and condition are shown (lower right panel). The number of cells analysed was 100, per time point per condition. (e) RhopH3 localization in -SerM/LysoPC and -SerM treated cells, at 36, 44 and 48 hpi. Localization dynamics were analyzed for rhoptries, infected red blood cell (iRBC) cytosol and iRBC membrane. Time point 2 shows significant differences when comparing AP2-G+ and AP2-G- cells ($p=0.001$ for rhoptries and iRBC membrane; $p=0.004$ for parasite membrane and iRBC membrane). Rendered representative images of the three main localizations are shown. The number of cells analysed was 100, per time point per condition.

in response to genetic modifications introduced under *in vitro* conditions^{11,15}, limiting them in response to environmental triggers will ensure sufficient levels of asexual growth within the human host as a prerequisite for transmission success^{9,33}.

Our data also demonstrate that blood stage schizonts harbor only sexual or asexual merozoites, but not mixed progeny. Moreover, we noted the enrichment of a number of rhoptry-associated proteins in the sexual commitment signature and were able to validate this observation experimentally with RhopH3 on the protein level. Quantitative imaging revealed increased RhopH3 expression in rhoptries, specifically in sexually committed schizonts. Identification of different or differentially expressed sets of merozoite antigens, including *rhoph3* and *msrp1*, in sexual versus asexual merozoites suggests that these two populations may be functionally distinct. Such differences may relate to altered host cell tropism and/or the emerging model of a specific tissue niche for gametocyte formation and development^{34–37}. Sexual commitment rates are generally low and elevated levels of invasion markers may also increase the likelihood of successful invasion and therefore transmission. The concept of distinct merozoite populations is supported by the variegated expression of some invasion ligands³⁸ and the recent identification of differences between hepatic and blood stage merozoites³⁹.

In summary, our study is one of the first single-cell transcriptomic studies of *Plasmodium* parasites, and applies this approach to define the signature of parasite commitment to the sexual pathway. Complementary chemical and genetic experiments validate the transcriptional data and define the different roles of the Kennedy pathway and its alternative substrate arm in growth and sexual commitment. Together with the recent studies by Poran *et al.*¹⁵ and Reid *et al.*⁴⁰, our scRNA-seq approach introduces methodology that provides a blueprint to investigate key knowledge gaps in the complex parasite cycle that are driven by transcriptional variation across cells. As such, these studies also represent powerful tools in addition to the available genetic and phenotypic assays in order to track parasite behavior during the ongoing malaria elimination campaign.

Methods

Generation of CRISPR/Cas9 *pmt* knock-out cell line

To disrupt the *Pfpm*t gene (PF3D7_1343000) in the Pf2004/164tdTom cell line, we generated the transfection vector pB_{gC}-Pfpmt. The pB_{gC} Cas9/sgRNA mother plasmid contains expression cassettes for (i) the *Streptococcus pyogenes* Cas9 enzyme, (ii) the single guide RNA (sgRNA), and (iii) the BSD resistance marker flanked by homology boxes¹⁶. To generate pB_{gC}-Pfpmt, two complementary oligonucleotides (sgRNA_{pmt_F}, sgRNA_{pmt_R}) encoding the sgRNA target sequence for *pmt* and appropriate single-stranded overhangs were annealed and inserted into the sgRNA expression cassette using *BsaI*-digested pB_{gC} and In-Fusion cloning. The *pmt* sgRNA target sequence (ATATTCTTCAACAGTTATTA) is positioned in the coding sequence 254 bp upstream of the stop codon. The donor cassette was generated using In-Fusion® HD cloning kit (Takara) to join PCR fragments *pmt* homology region 1 amplified from Pf2004_164Tdtom gDNA using primers *pmt_HR1_F* and *pmt_HR1_R*, and *pmt* homology region 2 generated using

primers *pmt_HR2_F* and *pmt_HR2_R*. PCR was performed with Phusion HF DNA polymerase (NEB) with the following thermocycling conditions: initial denaturation at 95°C for 2 min, thermocycling with denaturation at 95°C for 30 sec, annealing gradient ramp from at 50°C to 58°C for 30 sec, elongation at 61°C for 1 min for 8 cycles, 22 additional cycles [denaturation at 95°C for 30 sec, annealing at 57°C for 30 sec, elongation at 61°C for 1 min] and final elongation step at 61°C for 8 min. In the edited strain (Pf2004/164tdTom-Δ*pmt*), integration of *bsd* resistance marker was verified by PCR using primers *pmt_5'UTR_F* and *BSD_5'UTR_R* primer (testing 5' integration) and *BSD_3'UTR_F* and *pmt_3'UTR_R* (testing 3' integration). PCR conditions were as follows: initial denaturation at 95°C for 2 min, thermocycling with denaturation at 95°C for 30 sec, annealing at 51°C for 30 sec, and elongation at 61°C for 1 min for 25 cycles, and final elongation step at 61°C for 8 min. All oligonucleotide sequences are provided in [Supplementary File 6](#).

Parasite culture

P. falciparum cell culture was performed as described previously^{17,41}. Culture media consisted of RPMI-1640 supplemented with 25 mM HEPES, 100 μM hypoxanthine, 24 mM sodium bicarbonate, and gentamicin (all from Sigma-Aldrich). Serum media was generated by additional supplementation of 10% O+ human serum (The Interstate Companies), while serum-free medium was generated by additional supplementation of 0.39% fatty acid-free BSA, 30 μM oleic acid, and 30 μM palmitic acid (all from Sigma-Aldrich). To generate serum-free media supplemented with 20 μM LysoPC, LysoPC (Avanti Polar Lipids) was dissolved in ethanol and dried on culture dishes prior to culture addition. All cell cultures were maintained in a 5% CO₂, 1% O₂ and 94 % N₂ gas mixture (Med-Tech Gases). All experiments used a clone of the Pf2004/164Tdtomato *P. falciparum* line¹⁷, which requires addition of 4 nM WR99210 (Jacobus Pharmaceuticals) to select for stable maintenance of episomes.

Parasite transfection and selection of transgenic parasites

Pf2004/164tdTom ring stage parasites were transfected with 100 μg pB_{gC}-Pfpmt plasmid using electroporation conditions as described⁴². To select for parasites carrying the a disrupted *pmt* locus (Pf2004/164tdTom-Δ*pmt*), transfected parasites were grown in presence of 2.5 μg/ml blasticidin-S deaminase for the first 10 days and then in absence of drug pressure until a stably propagating parasite population was established (approximately 4 weeks post-transfection). Successful disruption of the *pmt* gene in the 164tdTom-Δ*pmt* population was confirmed by PCR on gDNA DNeasy blood and tissue kit (Qiagen), with thermocycling and primers, as described under “Generation of CRISPR/Cas9 *pmt* knock-out cell line”. All oligonucleotide sequences are provided in [Supplementary File 6](#).

Gametocyte induction and sample collection

Parasite cultures of the Pf2004/164tdTom line were sorbitol-synchronized (Sigma-Aldrich) as described previously⁴³ to a 4-hour cycle window using four consecutive treatments across two growth cycles. At 28–32 hours post-invasion, parasites were washed in –SerM media and plated with or without LysoPC. At 4, 8, and 12 hours post induction, cells were stained with Vybrant DyeCycle Violet (Life Technologies) in

HBSS (Thermo Fisher Scientific). Stained cells were pelleted and resuspended in phenol red-free PBS supplemented with 0.3% RNaseOUT (Life Technologies). Collection Twin.tec PCR 384-well plates (Eppendorf) were prepared by addition of 5 µl of cell lysis buffer to each well; lysis buffer consisted of nuclease-free water (Thermo Fisher Scientific) supplemented with 0.2% HF buffer (New England Biolabs). Single Vybrant Dye Cycle Violet-positive cells were sorted using a Beckman Coulter MoFlo Astrios EQ. At each time point, 48 cells from the non-induced culture (+LysoPC) and 336 cells from the induced culture (-LysoPC) were sorted into the collection plate and snap frozen on dry ice. Cultures were maintained for an additional two cycles for parasitemia and gametocytemia measurements. The media of both non-induced and induced cultures was replaced with RPMI/serum 12 hours post-induction and daily thereafter.

Conversion rate measurement

To measure parasitemia, samples were stained with SYBR Green (Life Technologies) and measured by flow cytometry using a MACSQuant Analyzer 10 (Miltenyi Biotec). To measure gametocytemia, samples were again stained with SYBR Green and measured by flow cytometry using the BioRad S3. Gametocyte conversion rate was calculated as described previously¹⁷.

Sequencing

We used an optimized version of single-cell RNA barcoding and sequencing⁶ (SCRB-seq⁷), with some modifications to enable high throughput automated liquid handling. Briefly, cells were thawed on ice and each sample was lysed using 1 µl of 1 mg/ml Proteinase K (ThermoFisher Scientific) followed by incubation at 50°C for 15 min, then 95°C for 10 min uncovered to evaporate all liquid. Reverse transcription was performed in 2 µl reactions containing 0.4 µl of 5X Maxima H Minus Reverse Transcriptase Buffer (ThermoScientific), 0.2 µl dNTPs (10 mM each; NEB), 0.02 µl of 100 µM E5V6NEXT primer (/5Me-isodC/iisodG/iMe-isodC/ACACTCTTTCCCTACACGACGCrGrGrG; IDT), 1 µl of 2 µM E3V6NEXT barcoded primer (see Soumilion *et al.*⁷ for barcoded primer sequences) and 0.125 µl Maxima H Minus Reverse Transcriptase (200 U/µl) (Thermo Fisher Scientific). Reactions were incubated at 42°C for 90 minutes followed by incubation at 80°C for 10 minutes, then pooled together. Pooled cDNA was purified using a DNA Clean & Concentrator kit-5 (Zymo Research) according to the manufacturer's recommended protocol for single-stranded DNA. Purified cDNA was then exonuclease-treated in a 20 µl reaction using 2 µl of 10X Exonuclease I Reaction Buffer (NEB) and 1 µl of Exonuclease I enzyme (NEB). Reactions were incubated at 37°C for 30 minutes, then at 80°C for 20 minutes. The cDNA was then amplified in a 50 µl reaction using the Advantage 2 PCR Kit (Clontech) as follows: 5 µl of 10X Advantage 2 PCR Buffer, 1 µl dNTPs, 1 µl of 10 µM SINGV6 primer (/5Biosg/ACACTCTTTCCCTACA CGACGC) (IDT) and 1 µl Advantage 2 Polymerase Mix (Clontech) using the following cycling program: 95°C, 1 min; 18 cycles of 95°C, 15 sec, 65°C, 30 sec, 68°C 6 min; 72°C 10 min. Reactions were purified using AMPure XP beads (Beckman Coulter) according to the manufacturer's recommended protocol. Illumina sequencing libraries were prepared using 1 ng of purified cDNA as input material for the Nextera XT kit (Illumina) according to the manufacturer's recommended protocol, with the exception of using the P5NEXTPT5 primer (AATGATACGGCGACCACCG

AGATCTACACTCTTTCCCTACACGACGCTCTTCC*G*A*T*C*T) (IDT) instead of the standard i5 primers contained in the kit. Final libraries were purified using AMPure XP beads (Beckman Coulter) according to the manufacturer's recommended protocol and prepared for sequencing on an Illumina NextSeq500 using paired reads of 17bp (library insert read1) + 46bp (library insert read2).

DGE read processing

DGE reads have the following organization. Read1 of the pair contains a well barcode, which signifies which well in a 384-well plate a read came from, and a unique molecular index (UMI). The UMI is used to distinguish reads representing different transcripts from reads originating from PCR duplicates. Read2 contains 45 nucleotides originating from the 3' end of an mRNA transcript.

Reads from each time point were processed as follows. Reads were assigned to a well using the well barcode, allowing for one mismatch between sequence and barcode (all barcodes are at least two nucleotides different from each other). Reads without valid UMIs were also discarded. Read sequences were aligned to the [PlasmoDB *Plasmodium falciparum* 3D7 transcript](#) (v29) downloaded on Oct 19, 2016 using [bwa aln](#) (v 0.7.10), allowing for up to four mismatches between the read and the reference. This mismatch-tolerance approach yielded more usable data than trimming ends of reads exhibiting a drop-off in base quality. Reads were also discarded if there were 20 or more As in a row (i.e., polyA tail) or if the sequence is mapped to more than 20 different transcripts. If there were duplicate UMIs, only one was counted. We aligned reads to the [hg19 human genome](#) using the same alignment parameters as above to evaluate whether reads that did not map to the *P. falciparum* reference transcriptome instead aligned to the human transcriptome, but observed an overall low alignment rate, suggesting that reads not assignable to *P. falciparum* most likely represent low quality and/or low complexity sequence rather than human contamination. To confirm our suspicion, we used [FastQC](#) (v 0.11.4) to assess sequence quality on aligned and unaligned reads to *P. falciparum*.

Further downstream processing of read counts

Post-processing and statistical analyses were performed using [R](#) version 3.3.1. UMI counts for all ribosomal RNA transcripts were removed. Genes that were expressed in fewer than 20 wells across all three time points were also excluded, which removed 2,576 genes from the 5,686 non-rRNA genes, to give a total of 3,110 genes measured. Wells containing fewer than 300 filtered UMIs were then removed, leaving 191/384 wells from 36-40 hpi, 326/384 wells from 40-44 hpi, and 364/384 wells from 44-48 hpi. Following filtering and removal, each time point was normalized using [edgeR](#) (v 3.14.0), such that each well's total normalized expression values summed to 1. The rRNA genes that were used for filtering are: *PF3D7_0112300*, *PF3D7_0112500*, *PF3D7_0112700*, *PF3D7_0531600*, *PF3D7_0531800*, *PF3D7_0532000*, *PF3D7_0725600*, *PF3D7_0725800*, *PF3D7_0726000*, *PF3D7_0801100*, *PF3D7_0801200*, *PF3D7_0830000*, *PF3D7_0830200*, *PF3D7_1148600*, *PF3D7_1148620*, *PF3D7_1148640*, *PF3D7_1371000*, *PF3D7_1371200*, *PF3D7_1371300*, *PF3D7_1418500*, *PF3D7_1418600*, *PF3D7_1418700*, *PF3D7_API04900*, *PF3D7_API05700*, *PF3D7_API05900*, *PF3D7_API06700*.

Downstream analysis

Principal component analysis (PCA). PCA was performed in R using the function `prcomp`, with center set to true and scale set to false (as data were already normalized). PCA was performed on edgeR-normalized read counts across all three time points (combined). PCA plots were visualized using the R package `ggbiplot` (v 0.55).

Comparison of DGE expression profiles to Bozdech *et al.*¹⁹ expression profiles. Normalized expression profiles from 46 time points from 1–48 hpi at 1-hour intervals were obtained from supplementary data from Bozdech *et al.*¹⁹. Only genes that were observed in both datasets were retained, and further filtered to the highest 100 expressed genes by DGE (though these results appeared to be robust to varying numbers of included genes). The Pearson correlation coefficient was obtained for normalized gene expression from each individual cell to each individual time point described by Bozdech *et al.*¹⁹. Then, each cell was assigned to the maximally correlated time point from the above study. A linear regression model was generated in R comparing the time (in hpi) the cell came from (*i.e.* 38–42, 46 hpi) to the maximally correlated time (in hpi) in ¹⁹. Adjusted R-squared and p-values obtained from the F test are reported.

Detection of differentially expressed genes. Since DGE read counts were sparse, we first pooled UMI counts from wells at random into three distinct replicates for induced and uninduced separately, for each time point, creating “pseudobulk” gene expression profiles. Then, to detect differentially expressed genes, edgeR was employed, comparing pseudobulk counts of induced vs uninduced, for each time point. Differential expression was determined using a general linear model likelihood ratio test (`glmLRT`). Log₂ fold change and p-value are reported for each gene. Multiple hypothesis correction was done using the Benjamini & Hochberg method.

SC3. The R package SC3 was used to perform single cell consensus clustering on normalized gene expression of those genes that were detected to be differentially expressed in the bulk RNA-seq experiment. The built in function using the Tracy-Widom theory on random matrices to estimate the optimal number of clusters k was used to pick the number of clusters. Cluster strength was determined by iterating over $k-1$ to $k+1$ clusters in SC3. The marker signature was calculated by averaging all marker genes' expression from cluster 11 from SC3 with k equal to 12, normalizing genes across cells first, so as to give equal weight to all genes, regardless of how highly expressed they are. Enrichment of the marker signature was determined using one-vs-rest Wilcoxon rank sum test for each cluster with Benjamini & Hochberg correction.

Other statistics. All statistics were performed in R. The p-values comparing *ap2-g* expression between inducing conditions as well as time points were calculated using a Fisher's exact test on the contingency tables. Linear regression for population-level RNA-seq versus DGE expression was performed using a linear model (`lm` in R), with adjusted R² and p-values reported. Correlation coefficients and p-values for *ap2-g* versus other genes were calculated using a paired samples Pearson's product

moment correlation test (`cor.test` in R). The overlap between single-cell RNA-seq experiments was evaluated with the R package `SuperExactTest` (v 1.0.0), which calculates the size of intersections of multiple sets and evaluates their statistical significance, as described previously²⁵.

Parasite live imaging and quantitative analysis of progeny numbers

Prior to imaging, *P. falciparum* trophozoite-infected RBCs were incubated in either of 3 different conditions, namely minimal fatty acid medium (–SerM), –SerM medium supplemented with 20 μM LysoPC (–SerM/LysoPC), or serum-complemented media (+SerM). Following 4–6 hour incubation in these inducing or non-inducing conditions, all parasites were incubated in RPMI/serum, and allowed to develop until early schizont (34–38 hpi), late schizont (40–48 hpi), or ring stages (48–54 hpi).

For merozoite quantification, late trophozoites/early schizonts were processed as described by Grüning *et al.*⁴⁴ and Brancucci *et al.*¹⁶. Briefly, cells from either –SerM, –SerM/LysoPC, or +SerM culture were arrested on the glass bottom of a sterile, concanavalin A-coated dish. Concanavalin A (ConA; Sigma-Aldrich) was dissolved at a concentration of 0.5 mg per ml in dH₂O, and 200 μl distributed uniformly on the glass surface of a 27 mm dish (Scientific Laboratory Supplies). ConA was added to the dish for 30 minutes at 37°C. It was then washed off twice using 1x PBS, after which the culture was re-suspended in PBS, added to the glass bottom of the dish, and allowed to settle for 10–20 minutes at 37°C. Thereafter, non-bound cells were washed off using PBS leaving a monolayer in the glass bottom, and 2 ml of pre-warmed PBS were added to the dish for imaging. Cells were viewed using a Zeiss Observer Z1 spinning disc confocal microscope equipped with an incubation chamber, a Yokogawa CSU-X1 filter wheel and spinning disc unit, a Photometrics Evolve 512 delta EM-CCD camera, two laser lines: 488nm, 405nm, and an α-Plan-Apochromat 100x 1.46 NA DIC VIS immersion oil lens. Parasites were immobilized in a glass-bottom dish as described above, stained with Hoechst 33342, and 50–72 z-stacks (0.2-μm step size) obtained using bright field and the 488 nm and 405 nm laser line, to determine the number of nuclei in RBCs infected with AP2G-GFP-positive or AP2G-GFP-negative parasites (when using the NF54 AP2G-GFP parasite line) or alternatively in all infected red blood cells of the Pf2004164tdTom parasite line. Quantifications were performed at either 2- or 4-hour intervals between 24 and 54 hpi.

Collection of individual *P. falciparum* parasites and progeny analysis

Between 36 and 40 hpi parasites at a parasitemia of 2% were allowed to settle on a glass-bottomed dish at a hematocrit of 0.05%. This enabled capturing of single parasites by pipette aspiration. Using a 100x oil immersion objective, and DIC or BF illumination, parasites were identified by hemozoin crystal presence or schizont structure when possible. Parasites were transferred to an intermediate glass-bottomed dish, where the presence of a single parasite was confirmed, and the parasite was ultimately transferred to a single well of a 96-well plate (glass bottom) at 0.5 % hematocrit in complete RPMI. This was repeated for 96 parasites per condition within a 4-hour time window. In the

96-well plates, each schizont was able to successfully rupture, after which merozoites successfully invaded neighboring uninfected RBCs. Four days post-infection, Hoechst 33342 at a concentration of 1:1000 was added to each well, and the wells viewed using a Zeiss Observer Z1 (specifications described above). Two proxies were analyzed using a 40x water objective, namely number of nuclei detected per well (using a 405 nm laser line), and number of TdTomato-positive cells per well (using a 594 nm laser line), the latter being exclusive to gametocyte progeny. The total number of TdTomato-positive cells and Hoechst-positive cells per well was quantified.

Parasite fixed imaging and quantitative analysis of AP2-G, ethanolamine kinase and RhopH3

For imaging of AP2-G and ethanolamine kinase levels in fixed cells at various times post-infection under –SerM and –SerM/LysoPC conditions, parasites were collected from culture at 24, 36, 40, 42, 44, 48, and 50 hpi, washed 3 times with 1x PBS, and fixed in 4% paraformaldehyde/0.01% glutaraldehyde in PBS for 1 hour, followed by permeabilization using 0.1% Triton X for 10 minutes. Parasites were then washed and blocked in 3% BSA overnight. All samples were incubated for 1.5 hours with a mouse anti-ethanolamine kinase antibody (dilution 1:100)⁴⁵, and a polyclonal rabbit anti-GFP antibody (dilution 1:200, Abcam #ab6556) diluted in 1 % FBS in PBS. After 3 washes in PBS with 1 % FBS, AlexaFluor488 monoclonal anti-rabbit IgG (1:500, Thermo Fisher #A27034), and AlexaFluor647 monoclonal anti-mouse IgG (1:500, BD Biosciences #557681) was added to the cells and incubated for 30 min. After 3 additional washes, the nuclear dye Hoechst 33342 diluted 1:1000 in Hank's Balanced Salt Solution (Sigma-Aldrich) was added for 30 minutes at RT.

RhopH3 was used as previously published²⁷, namely using a 1:1 ratio of acetone:methanol, followed by blocking in 3% milk powder. The RhopH3 rabbit antibody was diluted 1:500 and incubated for 40 min at room temperature. For detection, either AlexaFluor647 or AlexaFluor488 anti-rabbit antibodies were used at a dilution of 1:1000 for 30 minutes at room temperature.

Samples were examined using a Zeiss observer Z1 spinning disc confocal microscope, equipped with a Yokogawa CSU-X1 filter wheel and spinning disc unit, a Photometrics Evolve 512 delta EM-CCD camera, three laser lines (405, 488, and/or 642 nm) and an α -Plan-Apochromat 100x 1.46 NA DIC VIS immersion oil lens. Images were collected in line sequential mode with a dwell time of 1.1 μ s and z-increments of 0.19 μ m. Additionally, images were acquired using an ImageStream X Mark II imaging flow cytometer, using a 40x objective. Images were acquired using Zen 2012 software, and processed using ImageJ (version 2.0.0-rc-43/1.51r), Ideas (version 4.0), or Imaris

software (version 8.2.0). For quantitation of rhoptry localization, surface rendering and segmentation was performed using Imaris, and further analysis was done using ImageJ.

Data availability

All raw data from experimental procedures (imaging and flow cytometry experiments) are available on OSF under project title “Probing *Plasmodium falciparum* sexual commitment at the single cell level”: <https://doi.org/10.17605/OSF.IO/QMXVZ>⁴⁶. Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication). The .czi (Carl Zeiss imaging files) and .lif (Leica imaging files) files can be accessed using the [Bio-Formats](#) plugin for [ImageJ](#); the .fcs (flow cytometry input files) and .wps files can be accessed using [FlowJo](#).

Data from expression profiling by high-throughput sequencing of the *Plasmodium falciparum* NF54 line, GSE96066. <http://identifiers.org/geo:GSE96066>.

Grant information

The project was supported by Senior Investigator Award 172862 and IRS Award 172805 from the Wellcome Trust and a career development award from the Burroughs Wellcome Fund to M.M., and a Centre Award 104111 to the Wellcome Trust Centre for Molecular Parasitology, University of Glasgow. The project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Grant Number U19AI110818 to the Broad Institute. The project was further supported by Swiss National Science Foundation grants (31003A_163258 and BSCGIO_157729) to T.S.V., N.M.B.B and M.D.N. received PostDoc.Mobility fellowships from the Swiss National Science Foundation (P300PA_160975 and P2BEP3_165396, respectively), and M.D.N. received a Long Term EMBO postdoctoral fellowship (EMBO ALTF 1048-2016).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We thank Dr. Rachel Cerdan (University of Montpellier, France) and Dr. Sanjai Desai (Laboratory of Malaria and Vector Research, NIAID, National Institutes of Health, Rockville, United States) for kindly providing the ethanolamine kinase antibody and the RhopH3 antibody, respectively. We thank Patricia Rogers at the Broad Institute Flow Cytometry core for performing the single-cell sorting and Scott Steelman in the Broad Technology Labs for DGE library construction. We thank Caroline Cusick, James Bochicchio and Sinéad Chapman at the Broad Institute for project management.

Supplementary material

Supplementary Figure 1. DNA content of collected samples. At each time point, cells were VybrantViolet-stained to label (parasite) DNA. Plots show VybrantViolet intensity for infected cells from each sample. Lysophosphatidylcholine (LysoPC)-deficient medium (–SerM) is shown on the left; –SerM supplemented with LysoPC is shown on the right.

[Click here to access the data.](#)

Supplementary Figure 2. Quality of reads across time points and aligned vs unaligned. FastQC analysis reveals that aligned reads from 36-40 hpi are of lower quality, thus accounting for decreased number of UMI counts in that time point. Unaligned reads across all three time points have substantially lower-quality scores, likely explaining why these reads could not be aligned. In addition, unaligned reads are enriched in polyA tails, as exemplified by the increase in %A at the 3' end. (a) Per position base quality scores for each time point, separated by aligned vs unaligned reads. (b) Per read mean base quality for each time point, separated by aligned vs unaligned reads. (c) Per base sequence content of unaligned reads.

[Click here to access the data.](#)

Supplementary Figure 3. Single-cell consensus clustering (SC3) based on all filtered cells from 36-48 hpi. (a) Expression correlation across all cells by SC3. The dendrogram shows separation of cells into 12 clusters. Positive expression of *set9* (dark blue), *pmt* (80th percentile, orange), *msrp1* (cyan), and *ap2-g* (magenta) is indicated. The stability index of each cluster is given (bottom). (b) Marker genes of cluster #6 from SC3 of 36-48 hpi cells. The heat map shows transcriptional activity, with higher (red), moderate (yellow), and low (blue) levels. The top seven marker genes in cluster #6 are identical to those detected in cluster #11 from SC3 of 40-48 hpi.

[Click here to access the data.](#)

Supplementary Figure 4. Expression dynamics of commitment markers. (a) Life cycle expression dynamics of single-cell consensus clustering (SC3) marker genes. Normalized gene expression of marker genes from SC3 cluster #11 per time point from 1-48 hpi based on Bozdech *et al.*¹⁹. Only marker genes observed in 19 are shown. Expression of these marker genes peaks around 38 hpi. (b) Life cycle expression dynamics of rhostry marker genes. All rhostry-related genes from Bozdech *et al.*¹⁹ are shown. The majority of these genes are also highly expressed after 36 hpi, with a peak around 39 hpi.

[Click here to access the data.](#)

Supplementary Figure 5. Cloning strategy for Pfpmt knock out. To disrupt the Pfpmt gene (PF3D7_1343000) in parasites, we generated the CRISPR/Cas9 transfection vector pB_gC-Pfpmt. This vector was obtained by inserting a *bsd* resistance cassette flanked by the 5' and 3' homology regions (HR) into the p_gC mother plasmid¹⁶. (a), Schematic maps of the wild-type *pmt* locus in Pf2004/164tdTom parasites (top), the pB_gC-Pfpmt transfection vector (center), and the disrupted *pmt* locus after CRISPR/Cas9-based integration of the *bsd* resistance marker (bottom). The position of the *pmt* sgRNA target sequence 254 bp upstream of the *stop* codon is indicated. The pB_gC-pmt Cas9/sgRNA plasmid contains expression cassettes for Cas9, the sgRNA, and two donor homology regions (5'HR and 3'HR) flanking the *bsd* resistance marker facilitate repair of the Cas9-induced double-strand break and deletion of *pmt* gene by double crossover homologous recombination. Blue and red arrows indicate positions of PCR primer binding sites used to confirm successful gene editing. (b) PCR on gDNA isolated from Pf2004/164tdTom_Δpmt parasites and Pf2004 wild-type control parasites. Primers *pmt_5'UTR_F* and *pmt_3'UTR_F* bind to wild-type sequences outside the *pmt* 5'HR and *ap2-g* 3'HR homology regions and amplify a 2148 bp fragment from the wild type *pmt* locus (note the edited locus did not yield a product due to the large fragment size). The *pmt_5'UTR_F*/BSD_5'UTR_R and BSD_3'UTR_F/*pmt_3'UTR_R* primer combinations are specific for the edited locus in Pf2004/164tdTom_Δpmt parasites and amplify 957 and 861 bp fragments, respectively.

[Click here to access the data.](#)

Supplementary Figure 6. Growth phenotypes of parasites with impaired Kennedy pathway metabolism. (a) *pmt* knock-out parasites (Pf2004/164tdTom_Δpmt) show reduced growth compared to control parasites (Pf2004/164tdTom) under lysophosphatidylcholine (LysoPC)-limiting conditions only. The horizontal line indicates complete inhibition of growth (multiplication rate of 1). (b) LysoPC depletion shifts inhibitory concentration of BR23 (IC₅₀ -SerM: 1.42 nM, 95% CI 1.23-1.64; IC₅₀ -SerM/LysoPC: 3.54 nM, 95% CI 2.9-4.66). BR23 efficiently inhibits growth in presence of LysoPC, demonstrating the essential role of its target choline kinase. In absence of LysoPC, parasites become even more sensitive to BR23 suggesting that this drug also blocks the compensatory response through inhibition of the reported EK activity of choline kinase²⁶.

[Click here to access the data.](#)

Supplementary File 1. Normalized read counts and unique molecular identifiers (UMIs) across single cells.

[Click here to access the data.](#)

Supplementary File 2. Significantly differentially expressed genes in single cells under -SerM conditions.

[Click here to access the data.](#)

Supplementary File 3. Cluster membership and marker gene results from single cell consensus clustering (SC3) analysis.

[Click here to access the data.](#)

Supplementary File 4. Genes differentially expressed in cells with a marker signature above the 80th percentile across 40-44 and 44-48 hpi.

[Click here to access the data.](#)

Supplementary File 5. Overlapping hits between the three studies. Shown are all 22 hits that overlap between at least two out of the three studies. AP2-G and MSRP1 are the only hits overlapping in all three studies, confirming that AP2-G expression represents a major commitment signal. The small set of overlapping hits likely defines a minimal machinery driving the sexual commitment pathway. This core set includes several transcriptional and cell cycle regulators, as we well as signaling molecules but no metabolic enzymes, confirming our experimental data.

[Click here to access the data.](#)

Supplementary File 6. Oligonucleotides used in this study.

[Click here to access the data.](#)

References

- WHO: WHO Malaria Report 2016. 2016.
[Reference Source](#)
- Neafsey DE, Juraska M, Bedford T, *et al.*: Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *N Engl J Med.* 2015; 373(21): 2025–2037.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ménard D, Khim N, Beghain J, *et al.*: A Worldwide Map of *Plasmodium falciparum* K13-Propeller Polymorphisms. *N Engl J Med.* 2016; 374(25): 2453–2464.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Daily JP, Scanfield D, Pochet N, *et al.*: Distinct physiological states of *Plasmodium falciparum* in malaria-infected patients. *Nature.* 2007; 450(7172): 1091–1095.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mok S, Ashley EA, Ferreira PE, *et al.*: Drug resistance. Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science.* 2015; 347(6220): 431–435.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Herman JD, Rice DP, Ribacke U, *et al.*: A genomic and evolutionary approach reveals non-genetic drug resistance in malaria. *Genome Biol.* 2014; 15(11): 511.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Soumilion M, Cacchiarelli D, Semrau S, *et al.*: Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv.* 2014.
[Publisher Full Text](#)
- Nilsson SK, Childs LM, Buckee C, *et al.*: Targeting Human Transmission Biology for Malaria Elimination. *PLoS Pathog.* 2015; 11(6): e1004871.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pollitt LC, Mideo N, Drew DR, *et al.*: Competition and the evolution of reproductive restraint in malaria parasites. *Am Nat.* 2011; 177(3): 358–367.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bruce MC, Alano P, Duthie S, *et al.*: Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development. *Parasitology.* 1990; 100 Pt 2: 191–200.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Brancucci NMB, Bertschi NL, Zhu L, *et al.*: Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host Microbe.* 2014; 16(2): 165–176.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Pelle KG, Oh K, Buchholz K, *et al.*: Transcriptional profiling defines dynamics of parasite tissue sequestration during malaria infection. *Genome Med.* 2015; 7(1): 19.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kafsack BF, Rovira-Graells N, Clark TG, *et al.*: A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature.* 2014; 507(7491): 248–252.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sinha A, Hughes KR, Modrzynska KK, *et al.*: A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature.* 2014; 507(7491): 253–257.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Poran A, Nötzel C, Aly O, *et al.*: Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature.* 2017; 551(7678): 95–99.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brancucci NMB, Gerdt JP, Wang C, *et al.*: Lysophosphatidylcholine Regulates Sexual Stage Differentiation in the Human Malaria Parasite *Plasmodium falciparum*. *Cell.* 2017; 171(7): 1532–1544.e15.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brancucci NM, Goldowitz I, Buchholz K, *et al.*: An assay to probe *Plasmodium falciparum* growth, transmission stage formation and early gametocyte development. *Nat Protoc.* 2015; 10(8): 1131–1142.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Otto TD, Wilinski D, Assefa S, *et al.*: New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Mol Microbiol.* 2010; 76(1): 12–24.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bozdech Z, Llinás M, Pulliam BL, *et al.*: The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* 2003; 1(1): E5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bacher R, Kendziorowski C: Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol.* 2016; 17: 63.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Haque A, Engel J, Teichmann SA, *et al.*: A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* 2017; 9(1): 75.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Eksi S, Morahan BJ, Haile Y, *et al.*: *Plasmodium falciparum* gametocyte development 1 (*Pfgdv1*) and gametocytogenesis early gene identification and commitment to sexual development. *PLoS Pathog.* 2012; 8(10): e1002964.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kiselev VY, Kirschner K, Schaub MT, *et al.*: SC3: consensus clustering of single-cell RNA-seq data. *Nat Methods.* 2017; 14(5): 483–486.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Campbell TL, De Silva EK, Olszewski KL, *et al.*: Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog.* 2010; 6(10): e1001165.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wang M, Zhao Y, Zhang B: Efficient Test and Visualization of Multi-Set Intersections. *Sci Rep.* 2015; 5: 16923.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Serran-Aguilera L, Denton H, Rubio-Ruiz B, *et al.*: *Plasmodium falciparum* Choline Kinase Inhibition Leads to a Major Decrease in Phosphatidylethanolamine Causing Parasite Death. *Sci Rep.* 2016; 6: 33189.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ito D, Schureck MA, Desai SA: An essential dual-function complex mediates erythrocyte invasion and channel-mediated nutrient uptake in malaria parasites. *eLife.* 2017; 6: pii: e23485.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Perrin AJ, Bartholdson SJ, Wright GJ: P-selectin is a host receptor for *Plasmodium* MSP7 ligands. *Malar J.* 2015; 14: 238.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kadekoppala M, Ogun SA, Howell S, *et al.*: Systematic genetic analysis of the *Plasmodium falciparum* MSP7-like family reveals differences in protein expression, location, and importance in asexual growth of the blood-stage parasite. *Eukaryot Cell.* 2010; 9(7): 1064–1074.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pessi G, Kociubinski G, Mamoun CB: A pathway for phosphatidylcholine

- biosynthesis in *Plasmodium falciparum* involving phosphoethanolamine methylation. *Proc Natl Acad Sci U S A*. 2004; **101**(16): 6206–6211.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Bobenchik AM, Witola WH, Augagneur Y, *et al.*: ***Plasmodium falciparum* phosphoethanolamine methyltransferase is essential for malaria transmission.** *Proc Natl Acad Sci U S A*. 2013; **110**(45): 18262–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 32. Mancio-Silva L, Slavic K, Grilo Ruivo MT, *et al.*: **Nutrient sensing modulates malaria parasite virulence.** *Nature*. 2017; **547**(7662): 213–216.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 33. Greischar MA, Mideo N, Read AF, *et al.*: **Predicting optimal transmission investment in malaria parasites.** *Evolution*. 2016; **70**(7): 1542–1558.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 34. Joice R, Nilsson SK, Montgomery J, *et al.*: ***Plasmodium falciparum* transmission stages accumulate in the human bone marrow.** *Sci Transl Med*. 2014; **6**(244): 244re245.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 35. Aguilar R, Magallon-Tejada A, Achtman AH, *et al.*: **Molecular evidence for the localization of *Plasmodium falciparum* immature gametocytes in bone marrow.** *Blood*. 2014; **123**(7): 959–966.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 36. De Niz M, Meibalan E, Mejia P, *et al.*: ***Plasmodium* gametocytes display homing and vascular transmigration in the host bone marrow.** *Sci Adv*. 2018; **4**(5): eaat3775.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 37. Obaldia N 3rd, Meibalan E, Sa JM, *et al.*: **Bone Marrow Is a Major Parasite Reservoir in *Plasmodium vivax* Infection.** *mBio*. 2018; **9**(3): pii: e00625-18.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 38. Stubbs J, Simpson KM, Triglia T, *et al.*: **Molecular mechanism for switching of *P. falciparum* invasion pathways into human erythrocytes.** *Science*. 2005; **309**(5739): 1384–1387.
[PubMed Abstract](#) | [Publisher Full Text](#)
 39. Hopp CS, Bennett BL, Mishra S, *et al.*: **Deletion of the rodent malaria ortholog for falcipain-1 highlights differences between hepatic and blood stage merozoites.** *PLoS Pathog*. 2017; **13**(9): e1006586.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 40. Reid AJ, Talman AM, Bennett HM, *et al.*: **Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites.** *eLife*. 2018; **7**: pii: e33105.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 41. Trager W, Jensen JB: **Human malaria parasites in continuous culture.** *Science*. 1976; **193**(4254): 673–675.
[PubMed Abstract](#) | [Publisher Full Text](#)
 42. Voss TS, Healer J, Marty AJ, *et al.*: **A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria.** *Nature*. 2006; **439**(7079): 1004–1008.
[PubMed Abstract](#) | [Publisher Full Text](#)
 43. Lambros C, Vanderberg JP: **Synchronization of *Plasmodium falciparum* erythrocytic stages in culture.** *J Parasitol*. 1979; **65**(3): 418–420.
[PubMed Abstract](#) | [Publisher Full Text](#)
 44. Grüning C, Heiber A, Kruse F, *et al.*: **Development and host cell modifications of *Plasmodium falciparum* blood stages in four dimensions.** *Nat Commun*. 2011; **2**: 165.
[PubMed Abstract](#) | [Publisher Full Text](#)
 45. Alberge B, Gannoun-Zaki L, Bascunana C, *et al.*: **Comparison of the cellular and biochemical properties of *Plasmodium falciparum* choline and ethanolamine kinases.** *Biochem J*. 2009; **425**(1): 149–158.
[PubMed Abstract](#) | [Publisher Full Text](#)
 46. Marti M, De Niz M, Ravel D, *et al.*: **Probing *Plasmodium Falciparum* Sexual Commitment at the Single Cell Level.** *Open Science Framework*. 2018.
<http://www.doi.org/10.17605/OSF.IO/QMXVZ>

Open Peer Review

Current Referee Status:



Version 3

Referee Report 17 September 2018

doi:[10.21956/wellcomeopenres.16113.r33800](https://doi.org/10.21956/wellcomeopenres.16113.r33800)



Björn F.C. Kafsack 

Department of Microbiology and Immunology, Weill Cornell Medicine, New York City, NY, USA

The authors have address my remaining concerns. All I would ask for is the inclusion of higher resolution images with improved color balance for the Giemsa stains associated with Figure 4a.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2

Referee Report 21 August 2018

doi:[10.21956/wellcomeopenres.16028.r33579](https://doi.org/10.21956/wellcomeopenres.16028.r33579)



Björn F.C. Kafsack ¹, **Asaf Poran**²

¹ Department of Microbiology and Immunology, Weill Cornell Medicine, New York City, NY, USA

² Weill Cornell Medical College, New York, NY, USA

The initial version of the manuscript raised two major concerns with respect to (1) quality filtering of the single cell sequencing data and (2) regarding the quantification of sexual commitment in the Pf2004/164dtTom_Δpmt. This revised manuscript fully removes our concerns regarding the former but unfortunately left the latter unaddressed.

Quality Filtering

Our issues regarding QC were based the use of “unique reads” to indicated UMIs since a single transcript with a single UMI can give rise to hundreds of unique (non-identical) reads because the tagmentation step occurs following cDNA PCR amplification. Since the QC threshold was actually 300 UMIs rather than 300 unique reads, concerns regarding the data quality are resolved. To improve clarity, “unique reads” should be replaced by “UMIs” or “individual transcripts”, particularly in the description of quality filtering.

Quantification of sexual commitment in the Pf2004/164dtTom_Δpmt line.

I am unclear how providing the actual data underlying the normalized values being presented in Figure 4A could be considered beyond the scope of this manuscript. The answers to these questions seem integral to the interpretation of the data presented.

Without providing the actual gametocyte frequencies along-side the normalized counts it would be impossible to assess whether loss of *pmt* results in change in commitment at all and how well suppression of this can be determined under various LysoPC concentrations. This is important since the results in Bobenchick et al. show a reduction the *pmt*Δ line at all time points. It is therefore not possible to discern whether the *pmt*Δ line is defective solely in gametocyte maturation or both commitment and maturation.

The results presented Fig. 4A of this manuscript are normalized within each strain to the commitment level at 0uM LysoPC. But since the Pf2004/164dtTom_Δ*pmt* line seems unable to replicate (the replication rate is 1, ie no increase in parasitemia) under those conditions, it begs the questions what that 100% value represents. Are the Pf2004/164dtTom_Δ*pmt* parasite converting into gametocytes without replication under Serum-free conditions or does each schizonts only yield a single invasion event on average? To support the author's claim that PMT is not required for the repression of commitment by LysoPC, it is important to understand what actually being repressed if Pf2004/164dtTom_Δ*pmt* do not replicate and gametocytes do not develop under these conditions.

Additionally, if under serum/LysoPC-free conditions Pf2004/164dtTom_Δ*pmt* gametocytes don't develop as stated by the authors, providing visual examples of Pf2004/164dtTom_Δ*pmt* gametocytes under these conditions is critical to know what is actually being counted particularly since these cells provide the reference value for all LysoPC concentrations in the Pf2004/164dtTom_Δ*pmt* line.

All our other points were been sufficiently addressed.

Competing Interests: No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 27 Aug 2018

Matthias Marti, University of Glasgow, UK

Quality Filtering

Our issues regarding QC were based the use of "unique reads" to indicated UMIs since a single transcript with a single UMI can give rise to hundreds of unique (non-identical) reads because the tagmentation step occurs following cDNA PCR amplification. Since the QC threshold was actually 300 UMIs rather than 300 unique reads, concerns regarding the data quality are resolved. To improve clarity, "unique reads" should be replaced by "UMIs" or "individual transcripts", particularly in the description of quality filtering.

We have clarified throughout the manuscript that analyses are based on UMIs rather than unique reads.

Quantification of sexual commitment in the Pf2004/164dtTom_Δpmt line.

I am unclear how providing the actual data underlying the normalized values being presented in Figure 4A could be considered beyond the scope of this manuscript. The answers to these questions seem integral to the interpretation of the data presented.

Without providing the actual gametocyte frequencies along-side the normalized counts it would be impossible to assess whether loss of *pmt* results in change in commitment at all and how well suppression of this can be determined under various LysoPC concentrations. This is important since the results in Bobenchick et al. show a reduction the *pmt*Δ line at all time points. It is therefore not possible to discern whether the *pmt*Δ line is defective solely in gametocyte maturation or both commitment and maturation.

The results presented Fig. 4A of this manuscript are normalized within each strain to the commitment level at 0μM LysoPC. But since the Pf2004/164dtTom_Δpmt line seems unable to replicate (the replication rate is 1, ie no increase in parasitemia) under those conditions, it begs the questions what that 100% value represents. Are the Pf2004/164dtTom_Δpmt parasite converting into gametocytes without replication under Serum-free conditions or does each schizonts only yield a single invasion event on average? To support the author's claim that PMT is not required for the repression of commitment by LysoPC, it is important to understand what actually being repressed if Pf2004/164dtTom_Δpmt do not replicate and gametocytes do not develop under these conditions.

We have revised Figure 4a and now show absolute conversion rates rather than normalized rates. As mentioned in the figure legend reduced PMR in the PMT KO under limited or no LysoPC conditions demonstrates that PMT is essential under these conditions. But conversion rates are titratable down to a PMR of 1 (hence, at 12% conversion rate every 12th surviving schizont produces gametocyte progeny). Therefore, PMT activity is not required for sexual commitment.

Additionally, if under serum/LysoPC-free conditions Pf2004/164dtTom_Δpmt gametocytes don't develop as stated by the authors, providing visual examples of Pf2004/164dtTom_Δpmt gametocytes under these conditions is critical to know what is actually being counted particularly since these cells provide the reference value for all LysoPC concentrations in the Pf2004/164dtTom_Δpmt line.

We have also included Giemsa smears from WT and KO parasites at the time point of gametocyte readout (Stage II), at 0 μM LysoPC. There is no apparent difference in morphology suggesting gametocyte development is not affected in the KO until Stage II. The figure legend has been modified accordingly.

However, we agree that gametocyte viability is likely reduced at Stage II (i.e., our time point of readout) in the PMT KO, as effects on later development and transmission have been observed (Pessi *et al.*, PNAS 2004). As conversion rates are unchanged compared to WT even in complete absence of LysoPC, we conclude that both asexual and sexual stages are equally affected in the PMT KO.

Competing Interests: No competing interests were disclosed

Version 1

Referee Report 09 July 2018

doi:10.21956/wellcomeopenres.15947.r33322

 Björn F.C. Kafsack ¹, Asaf Poran²

¹ Department of Microbiology and Immunology, Weill Cornell Medicine, New York City, NY, USA

² Weill Cornell Medical College, New York, NY, USA

In their recent study in cell, the authors convincingly showed that sexual conversion in *P. falciparum* is higher responsive to availability of the serum metabolite LysoPC.

In this they use LysoPC depletion to explore the explore transcriptional heterogeneity related to sexual commitment at the single cell level using a single-cell RNA sequencing approach. Single cells from three time points during schizogony are collected and sequenced, both under inducing and uninducing conditions.

Relatively to previously published single-cell studies, the author detect low numbers of UMI / genes, with respect to the high sequencing depth. This low depth leads to weak association with bulk RNA seq data. The authors claim to define a gene signature specific to sexual commitment (10 genes), first through clustering of their data using SC3, following by a correlation analysis (125 genes).

The authors then use AP2-G GFP and microscopy of single-cells in single wells to track and quantify fate-determination. Using this approach, the authors are able to confirm that the merozoites from a single schizonts are either all asexually- or sexually- committed, without mixing. The authors show that LysoPC depletion (-SerM conditions) both increases sexual commitment and reduces the number of merozoites progeny.

The authors use flow cytometry and microscopy to validate some of the single-cell transcriptional analysis, such as the upregulation of PMT and EK enzymes in all parasites under LysoPC depletion, regardless of their sexual commitment. The authors also show that as predicted by transcriptomics, expression of rhopty genes is altered in sexually committed parasites.

Major Concerns:

1. Quality of single cell transcriptomes

Comparing the results of this study to the two other published scRNA-seq studies of malaria blood stages (Poran et al., Reid et al.) shows that the DGE scRNA-seq platform used yields single cell transcriptomes of substantially lower quality across a variety of measures.

	This study	Reid et al.	Poran et al.
Sequencing depth (read/cell)	~1e6	~1-3e6	8e4
Uniquely mapped reads	15-36%	75%	>80%
UMI/reads from rRNA	95-97%	17-34%	<5%
Unique mRNA genes- detected/cell/50k reads	20-50	184	352
Cells passing QC	88%	80-90%	82%

While the fraction of cells passing quality control is similar in all three studies, despite similar sequencing depths, the QC threshold used in this study (>300 unique reads/cell) is nearly two orders of magnitude lower than Reid et al. (25,000 reads/cell & >1000 genes detected).

It is also much lower than the (>300 UMI/cell, about 20,000 reads per cell) our study.

The technical quality of the single cell data raises significant concerns to the reproducibility of the single cell analyses.

Given the high cost (nearly \$35,000 = \$30/cell * 384 cells/tp * 3 time-points) for the analysis of approximately 1000 cells, the claim that DGE is "cost-effective & high-throughput" should be qualified, in light of the available alternative approaches.

This might suggest a substantial problem with the quality of collected cells (e.g. effect of sorting or Vibrant Violet staining), the effect of cryo-preservation, or library preparation.

The resulting transcriptional noise is reflected by the very low correlation with bulk RNAseq collected under the same conditions ($R^2=0.1-0.15$). In light of such weak correlation, the claim that these single cell profiles "recapitulate overall transcriptional profiles from population-level RNA-seq experiments" should be reevaluated.

The study by Reid et al, eLife 2018 also used a well-based method and sequenced to similar depths, and includes optimization of the library generation process that resolves many of the issue noted above. Given the similarities in approach and problems encountered, the lack of citation or discussion of the Reid et al. study is a critical gap in this manuscript.

2. Lack of commitment phenotype in Pf2004/164tdTom Δ pmt

To test whether the given upregulation of *pmt* under LysoPC-deprivation is required for sexual conversion, Δ pmt was deleted in the Pf2004/164tdTom background. Figure 4a shows that while loss of *pmt* had an impact on the parasite's ability to grow under low LysoPC concentrations, no impact on sexual conversion was observed.

This finding is quite surprising given the study by Bobenchik et al. (PNAS 2013, Fig 2) that gametocyte formation was drastically reduced in PMT knockout parasites and that this phenotype could be fully restored to WT levels by genetic complementation of *pmt*. This study should be referenced and possible explanations of these differences should be discussed.

Are the Δ pmt parasite still forming GCs despite being unable to replicate in the absence of LysoPC as indicated? Please provide actual % gametocytemia in addition to relative gametocytemia would help answer this question. As would images to validate the gametocyte morphology of Δ pmt tdTomato positive cells under these conditions

In the interpretation of figure 4, the text states "In support of these data, imaging flow cytometry also revealed increased protein expression of the second enzyme involved in producing choline from the alternative substrate ethanolamine (ethanolamine kinase, ek) in all cells under –SerM conditions"

Are the authors suggesting that that upregulations of EK can compensate for loss of PMT?

EK converts ethanolamine into phospho-ethanolamine, which is converted into phospho-choline by PMT.

Additional Notes:

Fig 2a:

Indicating the expression of ap2-g in all 3 time points would be useful.

Figure 2b

Bars shown in the left panel represent two completely different readouts (%ap2-g+c for single cell & FPKM for bulk) and should not be placed side by side in the same panel.

Also, the text states that "Both analyses demonstrate that ap2-g expression is significantly increased over time (Figure 2b)" but Fig. 2b has no time component.

Fig 2c:

"40-48 hpi revealed robust and reproducible separation of individual cells into 12 clusters"

9/12 clusters appear to have a stability index of 0.1. Clusters 5 and 11 appear to be the only stable clusters. The heatmap strongly suggests that Cluster 9 is composed of 2 clusters.

The text also states that "Expression of 10 genes, including ap2-g and set9, defined a subset of cells (cluster 11)" but 2c appears to show that set9 was detected in the majority of cells regardless of cluster.

Why are cells with >80%ile pmt expression indicated when the text states that *pmt* and *ek* "were equally expressed across all cells" and the fact the scRNAseq is quite poor at making quantitative comparisons?

Fig 2d:

Moreover, the data in Fig 2d seem to contradict the earlier statement that *pmt* and *ek* aren't equally expressed across the clusters but rather lower in clusters with high "signature" scores.

The text states that "Expression of ap2-g and msrp1 is significantly enriched in the clusters highlighted in red, while expression of pmt and ek are not" While PMT and EK are not higher in the clusters, they do appear lower particularly for clusters 10-11.

For clarity, color coding of significant difference (higher or lower) of mean gene expression in these clusters should be done individually by gene/cluster rather than for all 4 genes in a column.

What fraction of the variation captured by the signature is just the serine/threonine kinase PF3D7_1145200? The heatmap would suggest it's a or, possibly, the major component.

Providing gene names when available in the heatmap portion would help interpretation.

The upstream regions of "commitment signature" gene identified in Poran et al. were substantially enriched in AP2-G binding sites. A similar analysis would be a valuable addition to help evaluate whether differences in signature genes in this study may be upstream or downstream of ap2-g expression.

Fig 2e.

Please clarify the ratio being calculated on the x-axis. The legend states:

"Single cells from 40-48 hpi separate into subsets with high and low marker signature (above and below the 80th percentile, respectively), with 125 differentially expressed genes measured by scRNA-seq, including ap2-g, msrp1 and set9."

Does that mean that the fold-change being calculated is between the top 20% of signature expressing

cells vs the rest? That seems unlikely as the argument would then become circular.

Since AP2-G expression is the current standard for commitment, plotting the ratio of AP2-G+ vs AP2-G- cells may be the most appropriate. Please indicate in the legend if the plotted p-values were Benjamini-Hochberg adjusted.

Fig 4d

The number of cells analyzed should be indicated.

Fig 4e

The legend indicates that all three compartments include the "iRBC membrane". I suspect this is in error as compartment should be mutually exclusive to sum to 100%.

The number of cells analyzed should be indicated.

The methods are unclear as to how these compartments were defined in the absence of additional markers for the RBC PM or PVM.

The text seems to indicate that the microscopy results in fig 4e confirms upregulation of RhopH3. However, the last time-point shows no differences levels of RhopH3, suggests that RhopH3 expression may begin earlier in AP2-G+ cells but that expression in AP2-G- cells catches up yielding similar RhopH3 distribution in mature AP2-G+ and AP2-G- schizonts, as shown.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 23 Jul 2018

Matthias Marti, University of Glasgow, UK

Major Concerns:

1. Quality of single cell transcriptomes

Comparing the results of this study to the two other published scRNA-seq studies of malaria blood stages (Poran et al., Reid et al.) shows that the DGE scRNA-seq platform used yields single cell transcriptomes of substantially lower quality across a variety of measures.

	This study	Reid et al.	Poran et al.
Sequencing depth (read/cell)	~1e6	~1-3e6	8e4
Uniquely mapped reads	15-36%	75%	>80%
UMI/reads from rRNA	95-97%	17-34%	<5%
Unique mRNA genes-detected/cell/50k reads	20-50	184	352
Cells passing QC	88%	80-90%	82%

While the fraction of cells passing quality control is similar in all three studies, despite similar sequencing depths, the QC threshold used in this study (>300 unique reads/cell) is nearly two orders of magnitude lower than Reid et al. (25,000 reads/cell & >1000 genes detected). It is also much lower than the (>300 UMI/cell, about 20,000 reads per cell) our study. The technical quality of the single cell data raises significant concerns to the reproducibility of the single cell analyses.

Our QC threshold was 300 UMI/cell, regardless of the number of total reads coming from that cell. We acknowledge that our data were produced less efficiently due to poorer rDNA depletion, which could be improved if the method were implemented in the future. Nevertheless, we feel that the biological results are reproducible. Please note below our analysis of enriched AP2-G binding sites in genes that are positively differentially expressed in our list of loci associated with sexual commitment signature.

Given the high cost (nearly \$35,000 = \$30/cell * 384 cells/tp * 3 time-points) for the analysis of approximately 1000 cells, the claim that DGE is "cost-effective & high-throughput" should be qualified, in light of the available alternative approaches.

We have moderated language regarding cost-effectiveness. We do note that the protocol has limited hardware requirements relative to 10X or DropSeq.

This might suggest a substantial problem with the quality of collected cells (e.g. effect of sorting or Vibrant Violet staining), the effect of cryo-preservation, or library preparation.

Please note that cells have not been cryopreserved prior to sorting.

The resulting transcriptional noise is reflected by the very low correlation with bulk RNAseq collected under the same conditions ($R^2=0.1-0.15$). In light of such weak correlation, the claim that these single cell profiles "recapitulate overall transcriptional profiles from population-level RNA-seq experiments" should be reevaluated.

There is a significant, albeit weak, positive correlation, indicating that overall, there is quite a bit of noise from the single cell data, but the general trends from the bulk RNA-seq

hold. We have moderated the language discussing this result and have removed Figure 1f.

The study by Reid et al, eLife 2018 also used a well-based method and sequenced to similar depths, and includes optimization of the library generation process that resolves many of the issue noted above. Given the similarities in approach and problems encountered, the lack of citation or discussion of the Reid et al. study is a critical gap in this manuscript.

Thank you for your suggestion. We have added Reid *et al.* to the references.

2. Lack of commitment phenotype in Pf2004/164tdTom_Δpmt

To test whether the given upregulation of *pmt* under LysoPC-deprivation is required for sexual conversion, Δpmt was deleted in the Pf2004/164tdTom background. Figure 4a shows that while loss of *pmt* had an impact on the parasite's ability to grow under low LysoPC concentrations, no impact on sexual conversion was observed.

This finding is quite surprising given the study by Bobenchik et al. (PNAS 2013, Fig 2) that gametocyte formation was drastically reduced in PMT knockout parasites and that this phenotype could be fully restored to WT levels by genetic complementation of *pmt*. This study should be referenced and possible explanations of these differences should be discussed.

Indeed, our data demonstrate that PMT is not required to repress sexual commitment. However, it is required for gametocyte development and maturation as shown by Bobenchik et al. This is also supported by our own preliminary experiments. We have added the Bobenchik *et al.* reference in the discussion.

Are the Δpmt parasite still forming GCs despite being unable to replicate in the absence of LysoPC as indicated? Please provide actual % gametocytemia in addition to relative gametocytemia would help answer this question. As would images to validate the gametocyte morphology of Δpmt tdTomato positive cells under these conditions.

This is an interesting question and subject of ongoing studies in our lab, but outside of the scope of this MS.

In the interpretation of figure 4, the text states "In support of these data, imaging flow cytometry also revealed increased protein expression of the second enzyme involved in producing choline from the alternative substrate ethanolamine (ethanolamine kinase, ek) in all cells under –SerM conditions"

Are the authors suggesting that that upregulations of EK can compensate for loss of PMT? EK converts ethanolamine into phospho-ethanolamine, which is converted into phospho-choline by PMT.

It does not refer to the PMT KO data but the -SerM conditions of WT Pf2004/164tdTom parasites, as shown in Figure 4b and referenced in the text accordingly.

Additional Notes:

Fig 2a:

Indicating the expression of ap2-g in all 3 time points would be useful.

This is a good suggestion - we have now marked *ap2-g* across the three time points.

Figure 2b

Bars shown in the left panel represent two completely different readouts (%*ap2-g+c* for single cell & FPKM for bulk) and should not be placed side by side in the same panel.

Also, the text states that "Both analyses demonstrate that *ap2-g* expression is significantly increased over time (Figure 2b)" but Fig. 2b has no time component.

Thank you for catching this error. We agree that it may be misleading and will adjust the figure accordingly. We have updated the text with a more accurate description of panel b.

Fig 2c:

"40-48 hpi revealed robust and reproducible separation of individual cells into 12 clusters"

9/12 clusters appear to have a stability index of 0.1. Clusters 5 and 11 appear to be the only stable clusters. The heatmap strongly suggests that Cluster 9 is composed of 2 clusters.

Altogether, these issues may derive from the complex and incomplete nature of single-cell transcriptomes. A lower stability index indicates that upon choosing either 11 or 13 clusters, these cells were not entirely clustered together, which indicates how robust a cluster is to the a priori number of clusters. We acknowledge there may be some strong sub-clustering, especially within cluster 9 as you mentioned. We opted for 12 clusters as this was estimated by the Tracy-Widom theory to be the optimal number of clusters. We could obviously choose a larger number of clusters to have "stronger" clusters, but at the cost of potentially splitting up relevant and real clusters of cells.

The text also states that "Expression of 10 genes, including *ap2-g* and *set9*, defined a subset of cells (cluster 11)" but 2c appears to show that *set9* was detected in the majority of cells regardless of cluster.

***Set9* was detected in a majority of cells, but *set9* expression was elevated in cluster 11. We have adjusted the text to be more clear.**

Why are cells with >80%ile *pmt* expression indicated when the text states that *pmt* and *ek* "were equally expressed across all cells" and the fact the scRNAseq is quite poor at making quantitative comparisons?

We agree that the presentation of *pmt* in Figure 2c is potentially confusing, given that figure 2d much more clearly and quantitatively illustrates the expression level of *pmt* and *ek* in the cells comprising each cluster. We have removed *pmt* and the other genes previously displayed on the top rows of Figure 2c, and have amended the text accordingly.

Fig 2d:

Moreover, the data in Fig 2d seem to contradict the earlier statement that *pmt* and *ek* aren't equally expressed across the clusters but rather lower in clusters with high "signature" scores.

The text states that "Expression of *ap2-g* and *msrp1* is significantly enriched in the clusters

highlighted in red, while expression of pmt and ek are not" While PMT and EK are not higher in the clusters, they do appear lower particularly for clusters 10-11.

The difference is not statistically significant.

For clarity, color coding of significant difference (higher or lower) of mean gene expression in these clusters should be done individually by gene/cluster rather than for all 4 genes in a column.

These colors indicated that the signature was elevated, not the individual gene expressions. We have attempted to clarify this in the figure and legend.

What fraction of the variation captured by the signature is just the serine/threonine kinase PF3D7_1145200? The heatmap would suggest it's a or, possibly, the major component.

The signature used normalized expression values such that highly expressed genes are not weighed more than lower genes. As referee 1 pointed out, the scale in panel d does not clearly illustrate expression level differences, and we have attempted to rectify this with a different color map.

Providing gene names when available in the heatmap portion would help interpretation.

We have done so.

The upstream regions of "commitment signature" gene identified in Poran et al. were substantially enrich in AP2-G binding sites. A similar analysis would be a valuable addition to help evaluate whether differences in signature genes in this study may be upstream or downstream of ap2-g expression.

Thank you for this suggestion. We have performed this analysis using predicted AP2-G binding motifs from Campbell *et al.* 2010, which identified such motifs up to 2 kb upstream of translational start sites. Overall, 84% of the genes we detected with our DGE data had AP2-G binding sites, but 94% (117/125) of the genes we identified as being positively differentially expressed in our commitment signature gene set harbored an AP2-G binding site, a significant enrichment (Fishers exact test, $P=0.002$) that could indicate a significant proportion of the loci we detected are expressed downstream of AP2-G activation. We have included this new analysis result in the text.

Fig 2e.

Please clarify the ratio being calculated on the x-axis. The legend states:

"Single cells from 40-48 hpi separate into subsets with high and low marker signature (above and below the 80th percentile, respectively), with 125 differentially expressed genes measured by scRNA-seq, including ap2-g, msrp1 and set9."

Does that mean that the fold-change being calculated is between the top 20% of signature expressing cells vs the rest? That seems unlikely as the argument would then become circular.

Since AP2-G expression is the current standard for commitment, plotting the ratio of AP2-G+ vs AP2-G- cells may be the most appropriate. Please indicate in the legend if the plotted p-values

were Benjamini-Hochberg adjusted.

We have clarified the ratio in the text. We used bulk-verified genes to create the marker signature, and show results from those in c and d. In e, we used all genes detected in single cell. We did expect the genes used to create the signature to appear in this analysis, but we also wanted to look for potentially other genes that were not picked up by the bulk analysis.

The p-values were corrected and we have more clearly denoted this.

Fig 4d

The number of cells analyzed should be indicated.

The number of cells analysed was 100 per time point, per condition. This is now denoted in the figure legend.

Fig 4e

The legend indicates that all three compartments include the “iRBC membrane”. I suspect this is in error as compartment should be mutually exclusive to sum to 100%.

“iRBC membrane” is correct. The iRBC membrane staining remains constant as it reflects RhopH3 localization from the previous cycle. Therefore, the compartments where labeling changes in each condition are the parasite membrane, the iRBC cytosol, or the Rhoptries.

The number of cells analyzed should be indicated.

The number of cells analysed in all our work was 100 per time point, per condition. We have included this in the figure legend as well.

The methods are unclear as to how these compartments were defined in the absence of additional markers for the RBC PM or PVM.

These compartments are a proxy of maturation and/or commitment, and not a detailed co-localization analysis. Using DIC, bright field and fluorescence microscopy, it is possible to define the parasite and the RBC boundaries, as well as the RBC cytosol. By fluorescence intensity it is possible to define the rhoptries and distinguish them from other types of signal arising from more disperse presence. The image quality and resolution we obtained are comparable to those of other publications using the same antibodies.

The text seems to indicate that the microscopy results in fig 4e confirms upregulation of RhopH3. However, the last time-point shows no differences levels of RhopH3, suggests that RhopH3 expression may begin earlier in AP2-G+ cells but that expression in AP2-G- cells catches up yielding similar RhopH3 distribution in mature AP2-G+ and AP2-G- schizonts, as shown.

The data in 4d demonstrate that de novo rhoptry gene transcription (based on normalized mRNA expression of all rhoptry-related genes) and rhoptry protein synthesis (based on RhopH3 mean fluorescence intensity) is increased across time points in AP2-G+ cells compared to AP2-G- cells, suggesting overall increased levels rather than just a slight

shift in cell cycle in AP2-G+ cells. This difference is only partially explained by rhoptry signal intensity of RhopH3, however, as pointed out by this reviewer: rhoptry signal intensity (RhopH3) is only increased in the second time point in AP2-G+ compared to AP2G- cells. These were our observations in multiple repeated experiments, suggesting different dynamics in committed and non-committed parasites. As parasites reach full maturation at the schizont stage, this difference disappears. While there is no mechanistic observation on the reason for this, the observation was consistent and we considered it appropriate to report as such.

Competing Interests: No competing interests were disclosed.

Referee Report 04 July 2018

doi:[10.21956/wellcomeopenres.15947.r33320](https://doi.org/10.21956/wellcomeopenres.15947.r33320)



Lyn-Marie Birkholtz ¹, **Riette van Biljon** ², **Roelof van Wyk** ²

¹ Malaria Parasite Molecular Laboratory, Department of Biochemistry, Institute for Sustainable Malaria Control, Medical Research Council Collaborating Centre for Malaria Research, University of Pretoria, Pretoria, South Africa

² University of Pretoria, Pretoria, South Africa

The ability of malaria parasites to differentiate into sexual forms is the basis for continued survival and propagation of disease. The paper by Brancucci et al., provides a new dataset describing commitment to gametocytogenesis by evaluating the transcriptional signature in sexually committed cells on a single-cell level. The strength of the paper lies in the identification of signatures associated only with differentiation, within a background of general metabolic and life cycle changes, and this was mediated by the development of a digital gene expression (DGE) platform. The authors argued that this allowed the heterogeneity of parasite populations to be captured on a transcriptional level.

The authors combine their previous system, constricting lysoPC as environmental sensor, to produce gametocytes and subsequently capture individual committed cells to evaluate transcriptional signatures, confirmed through imaging. The transcriptional signatures are compared with bulk transcriptome datasets previously produced by these authors¹ and others (Bozdech *et al.*²), but significantly also with previous single-cell sequencing data from committed parasites (Poran *et al.*³). The work presented here by Brancucci correlates with their previous findings on bulk parasite cultures with similar experimental designs. This iterates that the observations made from bulk parasite cultures are indeed associated with commitment as delineated and confirmed by the scRNA-seq data presented here. Additionally, the manuscript provides information on genes involved in the processes preceding activation of the commitment master switch, AP2-G, which was not evaluated in the Poran study.

The paper concludes with functional evaluation of the involvement of *pmt* in sexual commitment, concluding that activation thereof is independent of sexual commitment under lysoPC constriction. This leads to the postulation of a threshold-dependent regulation of lysoPC activation of commitment, and interesting proposal worthy of further evaluation.

The paper therefore presents a timely contribution of a dataset that will aid further explorations and

validations of signatures associated with gametocyte commitment. With the DGE platform, the paper makes a powerful resource to the malaria research community.

The manuscript is well written, the figures nicely conceived and presented and the methods appropriately reported. The quality of the experimental system (DGE) and data produced is very good, given the limitations on starting material within this system, and should be taken up by the community as platform for further investigations. I support indexing of this paper.

Technical aspects that could be clarified include:

Figure 1a: The explanation as to the experimental setup needs clarification, including cell lines used for the single-cell transcriptome study, as this is not clearly defined in the methodology section. To avoid confusion, please clearly indicate if the single-cell transcriptomics was performed on non-modified lines, the Pf2004/164dTom line or the *pmt* knockout line. It is also not clear from the figure or legend what the 'green' population for FACS sorting and microscopic evaluation refers to, besides indicating the populations tagged for DGE?

Figure 1d: The platform is validated through transcriptional differences observed between time points and a statement that 'single-cell transcriptomes generally recapitulate overall transcriptional profiles from population-level RNAseq experiments'. However, care should be taken in as to the extent that correlations can be used to support these conclusions. The data presented in Fig 1d, is used to justify a major conclusion from figure 1- that the scRNA profiling is able to discriminate 4 h of temporal development. However, the single correlation (at R^2 of 0.43) with the complete bulk transcriptome time course from Bozdech 2003 (only the top 100 genes) is not entirely convincing (should perhaps just be included in supplementary data?).

Fig 1f. Poor correlations of 0.1 and 0.15 is observed for the top 100 most transcribed genes between the scRNA-seq data and population-level data (from Brancucci 2017), and the authors conclude this to be 'significantly through weakly' correlated. However, such low correlations are almost indicative of no correlation, even though significant. Could the authors elaborate why such low correlation is seen between the scRNA-seq and their own population-seq data produced under the same experimental conditions, but a correlation of 0.4 is seen when compared with the Bozdech 2003 population-seq data? As the authors rightly state, this could be a reflection of the loss of genes in single-cell sequencing, and should be left at that.

Fig 2a. This figure has some findings that could be really important that is only presented in the scRNA-seq dataset and lacking in the bulk RNA dataset. Genes with an induction score of 0, but with positive pseudobulk \log_2FC values are therefore only seen at the scRNA-seq levels and potentially lost or obscured in bulk datasets. Could the authors elaborate on these genes as potential novel aspects of their dataset?

Fig 2b: The relevance of the additional FPKM axis is not clear? Please clarify in the figure legend.

Fig 2c: The signature value is not properly explained.

Fig 2d: The activity scale used is not sufficient to show genes that express, with white indicating high expression even though it appears mostly inactive. The use of a 5-colour system may be more appropriate or transformations that place the values in a more comparable range.

Table 2. The comparison with the Poran 2017 data clearly highlights the genes that correlate, but also indicates that the data here identifies genes involved in the processes upstream of AP2-G activation. It would be important to highlight the numbers and individual genes involved in these processes, which currently are not clearly stipulated, but do present a novel finding of this study.

References

1. Brancucci NMB, Gerdt JP, Wang C, De Niz M, Philip N, Adapa SR, Zhang M, Hitz E, Niederwieser I, Boltryk SD, Laffitte MC, Clark MA, Grüning C, Ravel D, Blancke Soares A, Demas A, Bopp S, Rubio-Ruiz B, Conejo-Garcia A, Wirth DF, Gendaszewska-Darmach E, Duraisingh MT, Adams JH, Voss TS, Waters AP, Jiang RHY, Clardy J, Marti M: Lysophosphatidylcholine Regulates Sexual Stage Differentiation in the Human Malaria Parasite *Plasmodium falciparum*. *Cell*. 2017; **171** (7): 1532-1544.e15 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, DeRisi JL: The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol*. 2003; **1** (1): E5 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Poran A, Nötzel C, Aly O, Mencia-Trinchant N, Harris C, Guzman M, Hassane D, Elemento O, Kafsack B: Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature*. 2017. [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 23 Jul 2018

Matthias Marti, University of Glasgow, UK

Figure 1a: The explanation as to the experimental setup needs clarification, including cell lines used for the single-cell transcriptome study, as this is not clearly defined in the methodology section. To avoid confusion, please clearly indicate if the single-cell transcriptomics was performed on non-modified lines, the Pf2004/164dTom line or the *pmt* knockout line. It is also not clear from

the figure or legend what the 'green' population for FACS sorting and microscopic evaluation refers to, besides indicating the populations tagged for DGE?

The line used for flow sorting and scRNAseq is Pf2004/164tdTom, hence the same as used in the previous cell paper by Brancucci et al. This has been clarified in the methods section and the figure legend. The green population represents sexually committed cells. This has also been added to figure legend 1.

Figure 1d: The platform is validated through transcriptional differences observed between time points and a statement that 'single-cell transcriptomes generally recapitulate overall transcriptional profiles from population-level RNAseq experiments'. However, care should be taken in as to the extent that correlations can be used to support these conclusions. The data presented in Fig 1d, is used to justify a major conclusion from figure 1- that the scRNA profiling is able to discriminate 4 h of temporal development. However, the single correlation (at R2 of 0.43) with the complete bulk transcriptome time course from Bozdech 2003 (only the top 100 genes) is not entirely convincing (should perhaps just be included in supplementary data?).

The single cell data are much sparser than the bulk transcriptomics data, and gene dropout effects may lower the correlation. Using a standard linear regression will not be robust to gene dropout and the sparseness of these data, but nevertheless show a modest, significant correlation. We acknowledge that the plot illustrating the correlation between scRNA and bulk profiles is not compelling (Fig 1f) and have consequently removed it.

Fig 1f. Poor correlations of 0.1 and 0.15 is observed for the top 100 most transcribed genes between the scRNA-seq data and population-level data (from Brancucci 2017), and the authors conclude this to be 'significantly through weakly' correlated. However, such low correlations are almost indicative of no correlation, even though significant. Could the authors elaborate why such low correlation is seen between the scRNA-seq and their own population-seq data produced under the same experimental conditions, but a correlation of 0.4 is seen when compared with the Bozdech 2003 population-seq data? As the authors rightly state, this could be a reflection of the loss of genes in single-cell sequencing, and should be left at that.

Again, gene dropout is a likely culprit here for why the single cell data only weakly correlates with bulk transcriptomics data. We agree the significance is driven by numbers, rather than a strong correlation. However, in general, especially for the 100 most expressed genes, a higher relative expression in the population data led to a higher relative expression in the single cell data, which was our point with this panel.

Fig 2a. This figure has some findings that could be really important that is only presented in the scRNA-seq dataset and lacking in the bulk RNA dataset. Genes with an induction score of 0, but with positive pseudobulk log2FC values are therefore only seen at the scRNA-seq levels and potentially lost or obscured in bulk datasets. Could the authors elaborate on these genes as potential novel aspects of their dataset?

Some of those genes were not significantly differentially expressed by single cell, but those that were include ETRAMP14, CLAG3.1, and Rhoph3. After multiple hypothesis correction, only CLAG3.1 remains. The single-cell data were unbalanced, with more induced than uninduced cells; this, combined with gene dropout, may lead to some false

discovery, though p-value correction would in theory overcome this issue. We instead chose to focus on genes that were corroborated by the bulk dataset, as we have higher confidence in those results than discordant genes.

Fig 2b: The relevance of the additional FPKM axis is not clear? Please clarify in the figure legend.

FPKM is the axis for the population data. We acknowledge this is a confusing presentation and have separated the first plot in Fig 2b into two separate plots.

Fig 2c: The signature value is not properly explained.

The signature value is explained in the methods, but was unfortunately omitted from the legend. We have now added it to the legend to correct this.

Fig 2d: The activity scale used is not sufficient to show genes that express, with white indicating high expression even though it appears mostly inactive. The use of a 5-colour system may be more appropriate or transformations that place the values in a more comparable range.

Thank you for your suggestions. We adjusted the scale to improve clarity.

Table 2. The comparison with the Poran 2017 data clearly highlights the genes that correlate, but also indicates that the data here identifies genes involved in the processes upstream of AP2-G activation. It would be important to highlight the numbers and individual genes involved in these processes, which currently are not clearly stipulated, but do present a novel finding of this study.

Please see our response below to REVIEWER 1, in which we have further investigated the genes which may be upstream vs. downstream of AP2-G activation through an analysis of AP2-G binding sites in promotor regions. We have included a more detailed description of the genes in the revised draft.

Competing Interests: No competing interests were disclosed.