Latent transcriptional variations of individual *Plasmodium falciparum* uncovered by single-cell RNA-seq and fluorescence imaging

Katelyn A. Walzer[1,2], Hélène Fradin[2], Liane Y. Emerson[1,2], David L. Corcoran[2], Jen-Tsan Chi[1,2]*

1 Department of Molecular Genetics and Microbiology, Duke University, Durham, North Carolina, United States of America, **2** Center for Genomic and Computational Biology, Duke University, Durham, North Carolina, United States of America

* jentsan.chi@duke.edu

Abstract

Malaria parasites follow a complex life cycle that consists of multiple stages that span from the human host to the mosquito vector. Among the species causing malaria, *Plasmodium falciparum* is the most lethal, with clinical symptoms manifesting during the intraerythrocytic developmental cycle (IDC). During the IDC, *P. falciparum* progresses through a synchronous and continuous cascade of transcriptional programming previously established using population analyses. While individual parasites are known to exhibit transcriptional variations to evade the host immune system or commit to a sexual fate, such rare expression heterogeneity is largely undetectable on a population level. Therefore, we combined single-cell RNA-sequencing (scRNA-seq) on a microfluidic platform and fluorescence imaging to delineate the transcriptional variations among individual parasites during late asexual and sexual stages. The comparison between asexual and sexual parasites uncovered a set of previously undefined sex-specific genes. Asexual parasites were segregated into three distinct clusters based on the differential expression of genes encoding SERAs, rhoptry proteins, and EXP2 plus transporters. Multiple pseudotime analyses revealed that these stage-specific transitions are distinct. RNA fluorescent *in situ* hybridization of cluster-specific genes validated distinct stage-specific expression and transitions during the IDC and defined the highly variable transcriptional pattern of EXP2. Additionally, these analyses indicated huge variations in the stage-specific transcript levels among parasites. Overall, scRNA-seq and RNA-FISH of *P. falciparum* revealed distinct stage transitions and unexpected degrees of heterogeneity with potential impact on transcriptional regulation during the IDC and adaptive responses to the host.

Author summary

Malaria poses a significant global health problem and is responsible for nearly 445,000 deaths. It is caused by a single-celled parasite that infects a human host and propagates asexually through red blood cells. The most deadly malaria parasite, *P. falciparum*, follows

a tightly synchronous life cycle characterized by a continuous cascade of gene expression programming. Beyond these changes in the population, individual *P. falciparum* parasites are known to exhibit unique and rare expression, particularly to evade the host immune system and also to commit to a sexual fate. But such rare expression is often masked in bulk-cell studies. To understand expression variations among individual parasites, we combined single-cell RNA sequencing and fluorescence imaging to study the gene expression among individual asexual and sexual parasites. We discovered previously undefined sex-specific genes as well as three distinct clusters of late stage asexual parasites largely defined by stage-specific genes. We also showed that EXP2, a marker gene of one cluster, was expressed at multiple times throughout the life cycle. Furthermore, stage-specific gene expression was greatly variable among individual parasites. Our results have great implications for individual parasite survival and adaptation, particularly in response to environmental stresses like the host immune system.

Introduction

Malaria persists as a global health problem, killing nearly 445,000 per year worldwide and resulting in 216 million new cases in 2016 [1]. *Plasmodium falciparum* is responsible for the majority of these deaths, making it the most virulent of the *Plasmodium* species. This unicellular parasite has a complex life cycle, including multiple stages in the human host and *Anopheles* mosquito vector. In particular, the parasite undergoes a 48-hour intraerythrocytic developmental cycle (IDC) in humans, with gene expression commonly thought to occur in a continuous cascade and only when it is needed in the asexual ring, trophozoite, and schizont stages [2]. To examine these distinct transcriptional changes, the parasite population was studied on a bulk-cell level after chemical synchronization [2–6]. In spite of synchronization, the *P. falciparum* population is known to be heterogeneous. For example, one to thirty percent of parasites commits to a sexual fate [7–9] and develops into male and female gametocytes, the only stage of the parasite transmissible to the mosquito. Parasites within the same population may also exhibit known and unknown clonal variations, as shown in the parasites' antigenic variations to evade the host immune system [10]. This suggests that transcript expression is variable from cell-to-cell, an occurrence that is missed when averaged in bulk-cell analyses, even with chemical synchronization.

To overcome the issue of complexity in cellular populations, many groups have applied single-cell RNA-sequencing (scRNA-seq) to uncover cellular heterogeneity and rare gene expression. These findings include the discovery of rare intestinal cell types [11], bimodal gene expression and splicing in immune cells [12], and regulatory variation before and after stress in budding yeast [13]. Recently, scRNA-seq of *P. falciparum* uncovered genes upregulated in sexually committed cells [14, 15], genes upregulated in early stage gametocytes [16], discrete transcriptional changes during the asexual cycle [17], and genes expressed during the entire life cycle, including mosquito stages [18]. These approaches varied in their capture of the single-parasite transcriptome, from thousands of cells with low-coverage (Drop-seq and 10x Genomics) [14, 18–20] to hundreds of cells with either low-coverage (SCRB-seq) [15, 21] or full-length transcript coverage (Smart-Seq2) [16–18, 22]. By capturing thousands of synchronized asexual cells at three time points, scRNA-seq with Drop-seq resolved cell cycle progression of individual parasites into 11 clusters [14]. Higher coverage Smart-Seq2 also organized single-parasite transcriptomes by pseudotime but relied on previously published bulk-cell

analyses to verify cell cycle progression without independent validation, potentially missing rare and uncharacterized gene expression [17].

Here, we describe our development of a single-cell workflow to uncover the heterogeneity of *P. falciparum* populations using the Fluidigm C1 for scRNA-seq, followed by validation with RNA fluorescent *in situ* hybridization (RNA-FISH). To establish our single-cell methods, we first compared expression of single asexual parasites and late-stage gametocytes, uncovering a large number of previously undefined gametocyte-specific genes. We were then able to separate individual trophozoites and schizonts into three distinct clusters based on their specific gene expression. These three clusters were characterized by the expression of exported proteins and transporters, SERA genes, and rhoptries, respectively. This was further verified using RNA-FISH with markers for each cluster: EXP2, SERA4, and RAP1. While two clusters (SERA and rhoptry genes) were verified to be cell-cycle dependent, a gene in the third cluster, EXP2, showed expression throughout multiple times of the life cycle, particularly at the early trophozoite and late schizont stages. Additionally, we have found that expression of the stage-specific marker genes is highly variable among individual parasites and that transition between stages is shorter than and not as smooth as previously thought. Overall, these data allow us to overcome the limitation of bulk-cell analysis and decipher which genes are expressed in each individual cell. Our study highlights the power that single-cell analyses have with regards to dissecting complex populations of cells to uncover hidden gene expression patterns in *P. falciparum*. We expect that a similar approach can be applied to other eukaryotic pathogens to enhance our understanding of cell-to-cell heterogeneity and the functional consequences of such variability.

Results

Development of single-cell RNA-seq for *P. falciparum*

Studies on *P. falciparum*, including large-scale transcriptomic and proteomic profiling of multiple stages of the life cycle [2–6, 23], have mostly focused on bulk-population analyses of synchronized parasites. Even with chemical synchronization, it is possible that there is still significant heterogeneity or rare populations that are masked by bulk-cell analyses that only measure the average gene expression across all cells. Thus, only single-cell studies can reveal previously unrecognized cellular heterogeneity and potential rare/novel cell types. To this end, we used the microfluidic technology of the Fluidigm C1 to capture single parasites and perform scRNA-seq with full transcript coverage. We captured and analyzed 46 asexual parasites and 5 late-stage gametocytes to compare between asexual- and sexual-stage parasites as well as identified the subtypes among asexual parasites. The results from scRNA-seq were further validated using RT-PCR, RNA-FISH, and previously published scRNA-seq data [17].

Parasite cultures were sorbitol synchronized twice, each 48 hours apart, before collection of late stage asexual parasites that were approximately 34 hours post-invasion. Gametocytes were enriched in synchronized cultures by treatment with 50 mM N-acetyl glucosamine for 72 hours and were collected 9 days (stage IV-V) after drug treatment began. Parasite-infected erythrocytes were purified from uninfected erythrocytes by a 40/70% Percoll density gradient and MACS before being loaded onto the Fluidigm C1 Integrated Fluidic Circuit (IFC) for single-cell capture (Fig 1). Each microfluidic chamber was visually inspected under an inverted microscope to validate the presence of a single parasite before further analysis. While late-stage gametocytes had high capture rates (77 percent and above) on the IFC, trophozoite and schizont capture proved more challenging. We tested various ratios of cell number to suspension buffer using a concentration of 400 cells/μL and found that 40 to 50 percent of cells making up the final suspension buffer was optimal, although capture rates were still low (between 24 to 27

Single-parasite capture and sequencing

Percoll purify and magnetically sort to enrich for parasite-infected RBCs
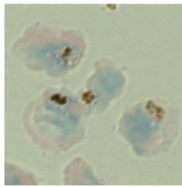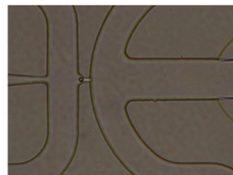

Late stage asexual
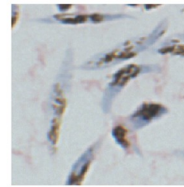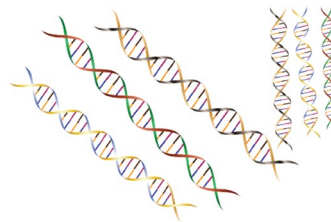

Gametocytes

Load and capture single parasites on C1

Wash and image parasites

Lyse, RT, and Harvest cDNA

Library prep and Illumina sequencing

Fig 1. The workflow of parasite isolation for scRNA-seq. Late stage asexual parasites were collected 34 hours post-invasion while stage IV-V gametocytes were collected on day 9. Either trophozoites and schizonts or late stage gametocytes were first separated from rings, early gametocytes, and uninfected RBCs by a 40/70% Percoll density gradient and were further purified from uninfected RBCs by MACS magnetic sorting. These infected RBCs were loaded onto a C1 IFC for single-cell capture. Each well was visually inspected to ensure the presence of a single parasite. Then the single parasites were lysed and cDNA was synthesized using SMARTer-Seq chemistry. The harvested cDNA was then used to generate a sequencing library for Illumina sequencing.

percent upon visual inspection). Therefore, we used two IFCs to capture a total of 46 single asexual parasites for scRNA-seq library construction using a Smart-Seq protocol. Sequencing was performed on an Illumina HiSeq with single-end 50 bp reads. Generally, gametocytes expressed a greater number of transcripts than asexual cells (Fig 2). On a bulk-cell level, 4805 transcripts were detected in late-stage gametocytes while 4136 and 4313 transcripts were detected in two different asexual samples. At a single-cell level, this number was much lower in both cell types with an average of 3127 transcripts detected in gametocytes. In individual asexual cells, 1321 to 3071 transcripts, with an average of 2002 transcripts, were detected. This number of detected transcripts in our study is consistent with the average of 1712 detected transcripts found in previous scRNA-seq data [17]. Additionally, we analyzed the length distribution of all genes and expressed genes and noted a slight bias towards longer transcripts (S1 Fig), indicating that both long and short transcripts were well represented in the dataset.

Single-cell RNA-seq identifies gametocyte-specific genes and separates asexual parasites into three distinct clusters

As a first step to analyze the heterogeneity of transcriptomes of the 51 single parasites, we performed unsupervised clustering using the clustering method implemented in SC3 [24]. A total of four stable clusters emerged: three clusters of asexual cells and one cluster of gametocytes, visualized using principal component analysis (PCA) (Fig 3A). Among the asexual clusters, cluster 1 was made up of 23 cells, cluster 2 had 12 cells, and cluster 3 had 11 cells. To examine the cell-to-cell similarities and differences within and between clusters, consensus clustering was performed and showed that cells in each cluster exhibited stronger similarity with each other than with cells in other clusters (Fig 3B). In particular, the 5 gametocytes showed very strong correlation with one another but little-to-no correlation with any asexual cells, indicating their distinct transcriptional profiles. Additionally, gametocytes are larger and generally express more transcripts, which may have contributed to their distinct clustering pattern away from asexual parasites.

Cluster-specific genes for gametocytes were identified using the R package SC3 [24]: 340 genes were upregulated in the gametocytes relative to the asexual cells, with an area under the receiver operating characteristic (AUROC) cutoff value of 0.7 (S5 Table). These markers were ordered according to their decreasing AUROC value, with the top 100 differentially expressed in gametocytes shown in Fig 3C. Some of the top expressed genes include known gametocyte markers P25 and P230 [25–27]. Among the 340 gametocyte genes, 185 are annotated as differentially transcribed in females and 73 are annotated as differentially transcribed in males [28]. From the top 100 genes, 8 gametocyte-expressed genes not previously described as gametocyte-specific along with the known marker Pfs16 were chosen for further qPCR validation using population samples from asexual parasites and gametocytes (Fig 3D). Four of these genes have unknown functions while the others putatively encode a peptidase, kinesin, RAP protein, and AAA family ATPase. These 8 genes were all shown to be upregulated in late-stage gametocytes when compared to the asexual stage, with six genes being upregulated more than 10-fold (Fig 3D). Over half of the 340 gametocyte markers have unknown functions, and
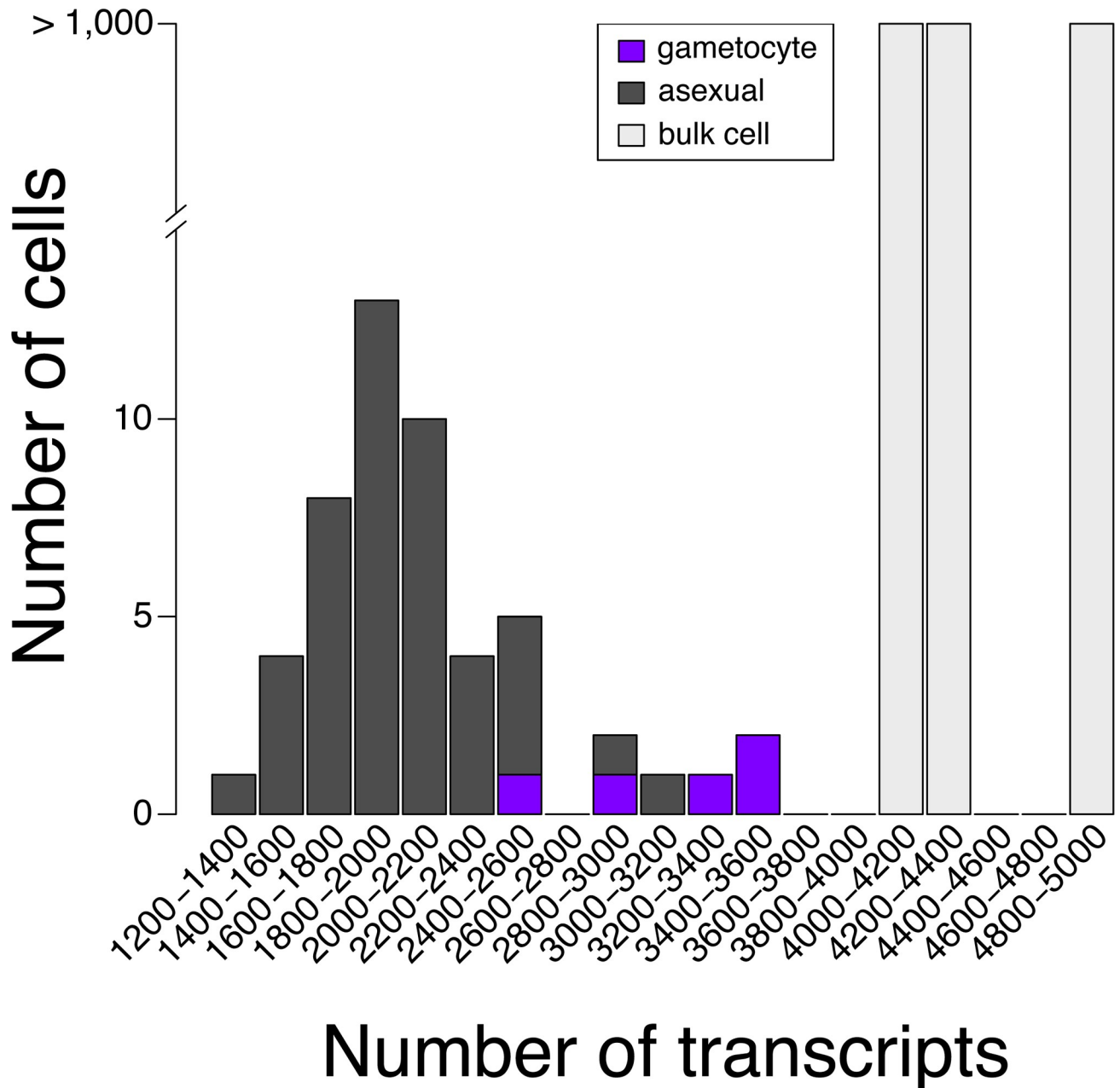
Fig 2. The distribution of expressed transcripts in individual asexual cells, gametocytes, or the total parasite population. Generally, single gametocytes expressed more genes than single asexual cells.

future studies will prioritize these genes to understand their functional roles in gametocyto-genesis and gametogenesis.

Having established that scRNA-seq clearly identifies distinct transcriptional differences between asexual parasites and gametocytes, we further evaluated the heterogeneity among the asexual population. After removing gametocytes, we performed unsupervised clustering using the clustering method implemented in SC3 [24], which separated the remaining 46 single asex-ual parasites into three distinct groups (Fig 4A). Cluster 1 was made up of 26 cells, cluster 2

had 10 cells, and cluster 3 had 10 cells. Since these individual asexual parasites were collected at the same time after synchronization, we performed consensus clustering to group the parasites by gene expression and define their similarities and differences across different clusters.

Fig 3. Single gametocytes cluster separately from three distinct asexual populations and express gametocyte-specific genes revealed by scRNA-seq. A) Principal component analysis (PCA) was used to visualize the cluste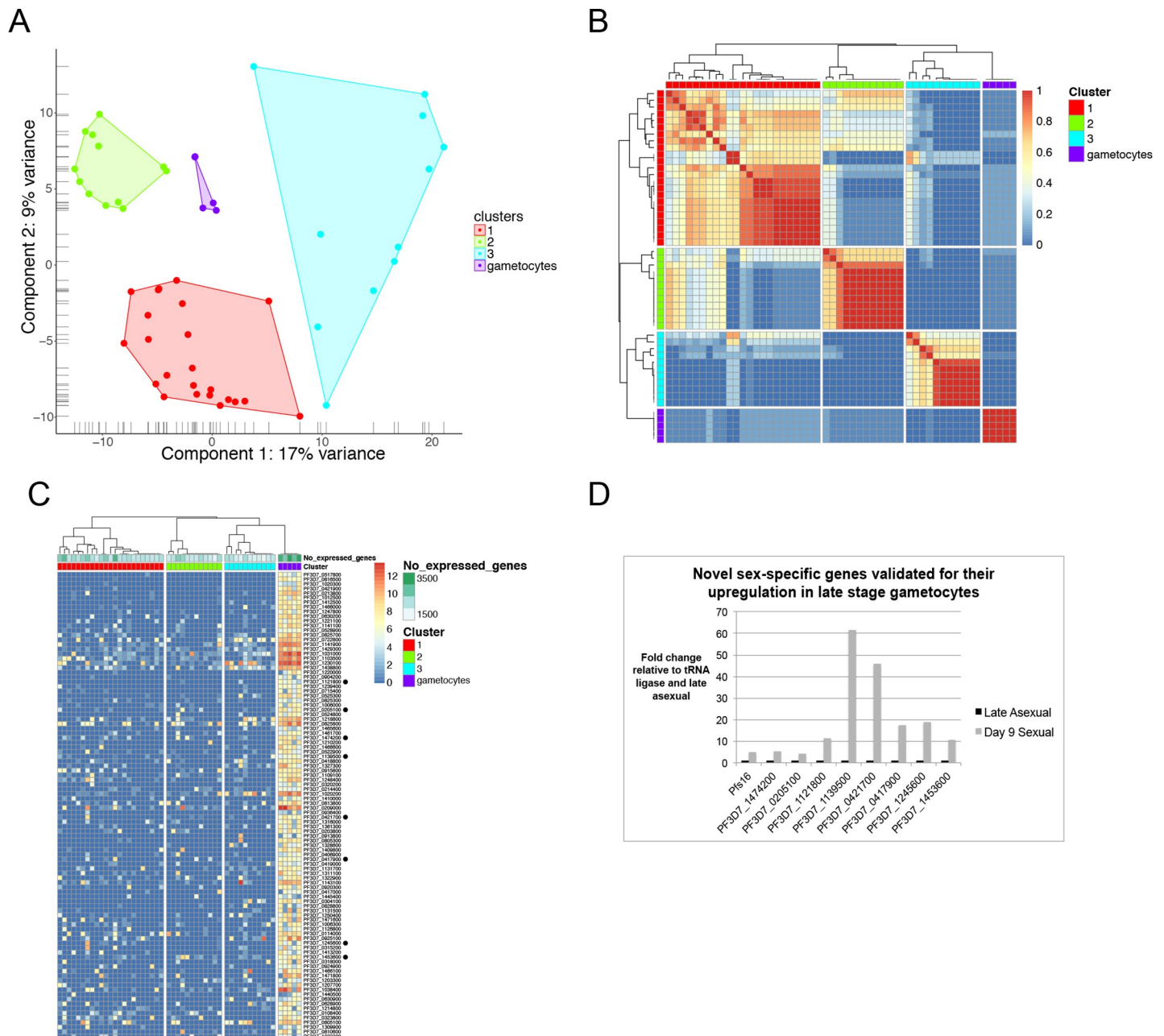ring of 46 late stage asexual cells and 5 late stage gametocytes by gene expression highlighting clear separation of the gametocyte single cells from three distinct asexual populations. Two gametocytes had very similar first and second component coordinates, overlapping on the PCA plot. B) Consensus clustering matrix of all cell pairs displays very strong cell-to-cell similarity amongst gametocytes when compared to asexual cells, which are grouped together based on their similar gene expression. The color scale indicates the likelihood that two cells are arranged in the same cluster (blue: low (0), red: high (1)). C) 340 genes were specifically expressed in gametocytes relative to late stage asexual parasites, with the top 100 genes shown in this heatmap. The SC3 markers function was used to determine the potential that each gene could be a marker for the cluster in which its mean expression value was the highest. Genes are ordered by decreasing area under the receiver operating characteristic (AUROC) with a cutoff of 0.7. D) Eight novel gametocyte genes, denoted by black dots in (C), were chosen for qPCR validation on a bulk-cell level, n = 2. Gametocytes collected on day 9 were compared to late stage asexual parasites. Fold changes are relative to the tRNA ligase control and asexual samples. Pfs16 is a positive control for gametocyte-specific expression.
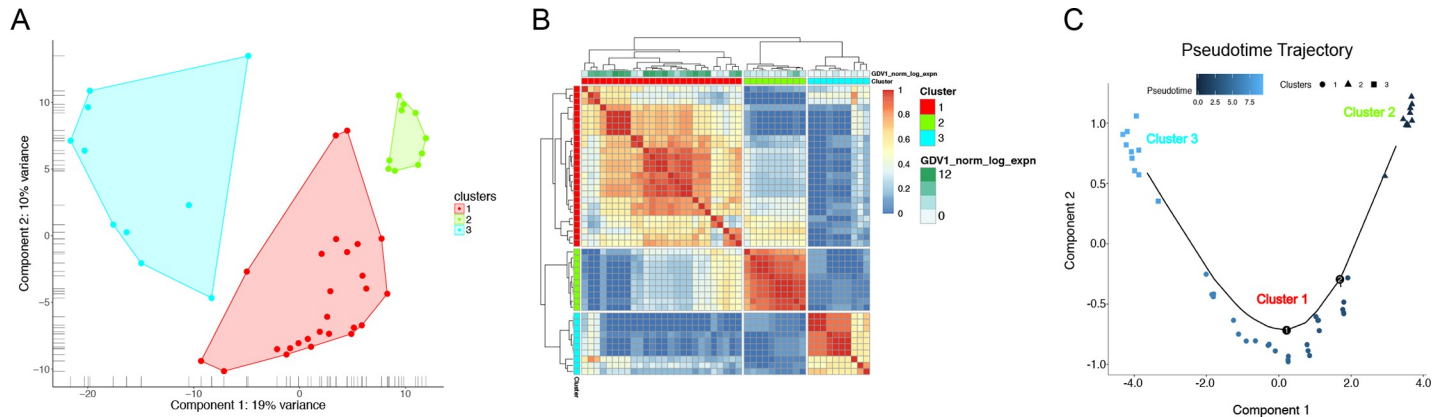
Fig 4. Single-cell RNA-seq expression of late stage asexual parasites identifies three distinct populations. A) Principal component analysis (PCA) was used to visualize the clustering of 46 single late stage asexual transcriptomes. B) Consensus clustering indicates that parasites within the same cluster show the strongest similarity when compared to parasites from other clusters. However, there is significant heterogeneity even between parasites in the same cluster. Blue indicates low similarity between two cells, whereas red indicates high similarity, i.e. that the two cells are always assigned to the same cluster. Single parasites are also annotated for their expression of GDV1 by its log normalized expression. C) Pseudotemporal ordering of 46 single asexual parasites by *monocle* 2. Genes expressed differentially between the 3 clusters detected by *SC3* were used for ordering cells and inferring a trajectory. Cells are colored by pseudotime and shapes indicate their assignment to one of the 3 clusters.

Consensus clustering again indicated stronger similarity amongst cells within the same clusters than between them, although cell-to-cell heterogeneity within clusters is more evident after the removal of gametocytes, with some sub-clusters, especially in cluster 1 (Fig 4B), suggesting further heterogeneity within this cluster. In particular, the expression of GDV1, a gene involved in sexual commitment [9, 29], varied in cluster 1 with some high expessers and low expressers (Fig 4B). This indicates that gene expression for sexual commitment may contribute to the heterogeneity within the larger cluster 1. Additionally, multiple pseudotime analyses, with and without these clusters as an indicator variable, revealed the clear ordering of single parasites into three distinct and separate clusters, which started with cluster 2 and progressed to cluster 1, followed by cluster 3 (Fig 4C and S2 Fig).

To identify the genes that best represent each cluster, we used the marker genes function of R package SC3 [24] and ranked the cluster-specific marker genes by AUROC values (Fig 5). Cluster 1 showed expression of many serine repeat antigen (SERA) genes, including SERA4, SERA5, and SERA7, as well as expression of GDV1 (Fig 5A). Cluster 2 was characterized by the expression of a number of exported proteins and transporters (Fig 5B). This includes EXP2, which is a member of the PTEX complex involved in the export of parasite proteins into the host erythrocyte [30, 31]. Recently, EXP2 was also characterized for its role in nutrient transport [32, 33]. MDR1 is another cluster 2 transporter and is found in the food vacuole; mutations in this gene are associated with chloroquine resistance [34]. Meanwhile, MFR5, another cluster 2 gene, was previously identified as a putative transport protein of the major facilitator superfamily (MFS) [35] and may transport similar solutes as other cluster 2 transporters. Known exported proteins among cluster 2 genes included MESA [36, 37], PIESP2 [38], and a PHIST domain-containing gene (PF3D7_0402000) [39], all important for remodelling the human erythrocyte during infection. Cluster 3 contained several rhoptry genes, including RAP1, ROP14, RON6, RAMA, RAP2, RhopH3, RON5, RON2, and RON3 (Fig 5C). Rhoptries are secreted proteins that function in the invasion pathway [40]. An essential AP2 factor (PF3D7_1107800) was also associated with this cluster, similarly to the rhoptry cluster in another scRNA-seq dataset [17]. This co-expression of AP2 and rhoptry genes may suggest a potential regulatory relationship that will need to be further investigated. Furthermore, the
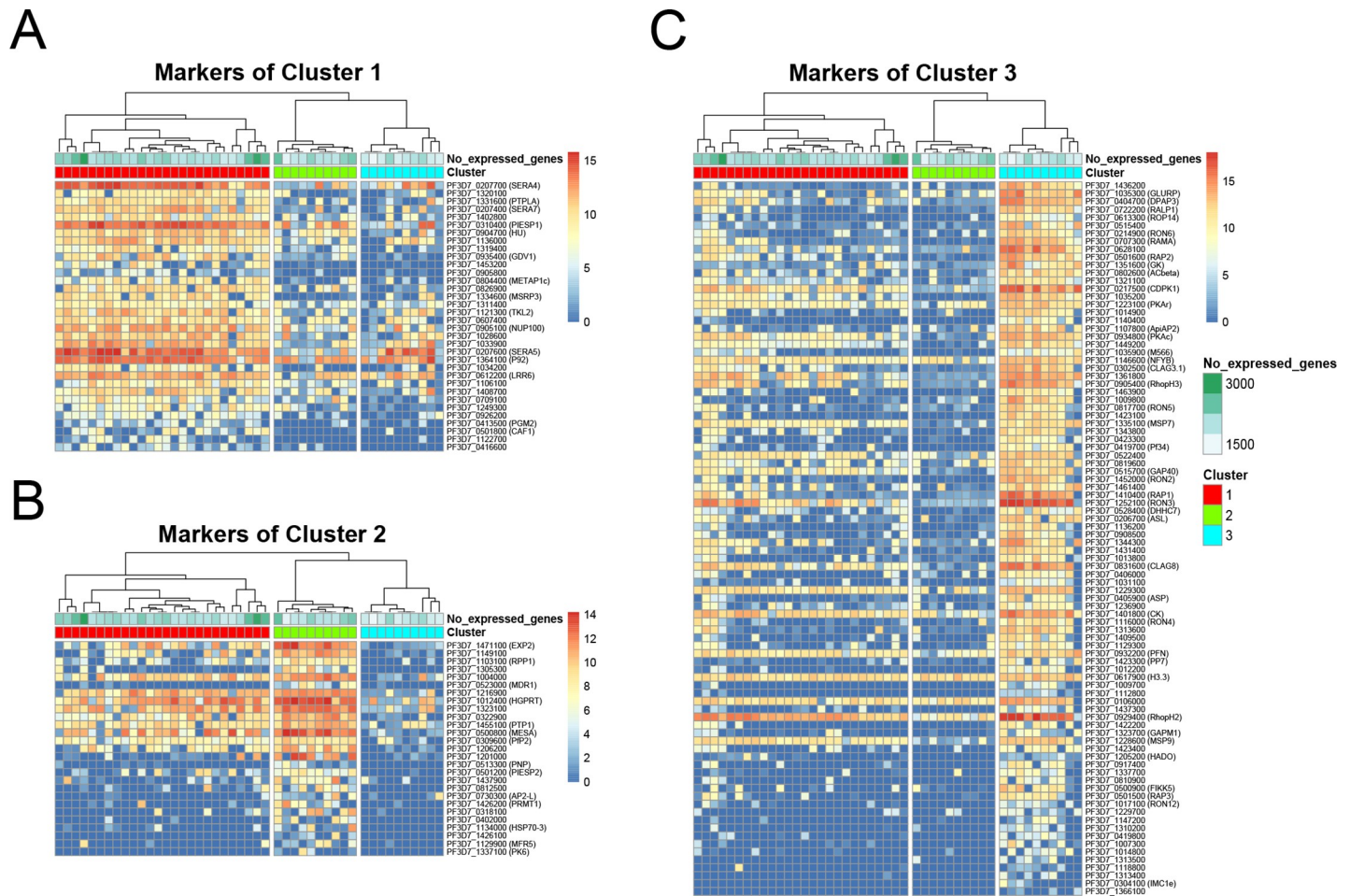
Fig 5. The top marker genes of each cluster were identified by scRNA-seq. The SC3 markers function was used to determine the potential that each gene could be a marker for the cluster in which its mean expression value was the highest. Genes are ordered by decreasing area under the receiver operating characteristic (AUROC) in clusters 1 (A), 2 (B), and 3 (C).

genes expressed in these clusters are consistent with our non-cluster indicator variable pseudo-time analysis, which also exhibits two groups for cluster 1 (S3 Fig).

To validate our clusters and markers, we also analyzed another scRNA-seq dataset [17]. We processed the transcriptomes of 172 late-stage asexual *P. falciparum* the same way as our transcriptomes and clustered the parasites according to our 151 cluster-specific markers. We found that all of our identified markers were also expressed in these parasites (S6 Table) and that the optimal number of clusters was also three (S4 Fig). The 85 markers of the Reid et al. cluster 3 were all markers of our cluster 3. The Reid et al. cluster 2 shared 9 markers with our cluster 1 and 2 with our cluster 3, while the Reid et al. cluster 1 shared 24 markers with our cluster 2 but also 13 with our cluster 1. Of note, EXP2, MDR1, MFR5, MESA, PIESP2, and PF3D7_0402000 are all shared between Reid cluster 1 and our cluster 2 while all three SERA genes are shared between Reid cluster 2 and our cluster 1. GDV1 changed clusters and appears in Reid cluster 1 with the exported proteins, potentially indicating differences between sexual commitment rates or timing, which may be related to experimental conditions or use of a different parasite strain. Therefore, we could cross-validate our results using an independent scRNA-seq study [17].

When comparing the genes expressed in each group of our dataset, we noticed that genes in clusters 1 and 3 seemed to represent known stage-specific markers in the *P. falciparum* IDC. To validate these cluster-specific markers and determine their time-specific progression during the IDC, we employed two validation approaches, 1) comparison to previously published data of IDC progression and 2) RNA fluorescent *in situ* hybridization (RNA-FISH). First, we compared the cluster-specific genes to a previously published microarray dataset, which represents the most meticulous temporal analysis of the *P. falciparum* IDC to date [2]. As expected, genes in clusters 1 and 3 followed a largely time-specific pattern, with cluster 1 characteristic of a trophozoite-to-schizont transition and cluster 3 representative of schizonts (S5 Fig). As cells were synchronized during the ring stage a few hours after re-invasion, the majority of collected asexual single cells were predicted to be 34 hours post invasion (hpi). However, given the limitation of current synchronization methods, the exact timing of any individual parasite may vary. Cluster 1 had the most cells and exhibited strong expression of SERA4, SERA5, and SERA7 (Fig 5A). These genes are highly expressed at 34 hpi, confirming our capture of this stage of the parasite and their identity as cluster 1 markers (S5A Fig). Cluster 3 genes were also characteristic of cell cycle progression and represent schizonts past 34 hpi. In previous data, strong upregulation of cluster 3 genes begins at 32 hpi and continues through 48 hpi, confirming our cluster 3 and capture of these later-stage parasites (S5C Fig). Importantly, our single-cell data indicate that these genes are preferentially expressed in distinct groups of cells (Fig 5). Unlike the synchronized parasite population at 34 hpi which appears to express both SERA and rhoptry genes highly (S5 Fig), our single-cell transcriptomes show distinct transitions in gene expression in single parasites (Figs 4C and 5), a finding further validated using RNA-FISH.

In contrast, the genes denoted as markers of cluster 2 were not expressed at one particular stage in bulk-cell microarray data (S5B Fig). While some genes (e.g., MESA and PIESP2) are highly expressed in early trophozoites, other genes, including MDR1 and AP2-L, are expressed in late schizonts and early rings (S5B Fig). The expression timing of these cluster 2 markers is also distinct from the expected 34 hpi when we captured the parasites. However, because these genes cluster together in single-cell analysis, it appears that cluster 2 specific genes exhibit co-expression not seen previously on a bulk-cell level. In particular, EXP2, the top marker for cluster 2, is a component of the PTEX complex that is responsible for the export of *P. falciparum* proteins into host red blood cells. Because it shows co-expression with MDR1 and other transporter genes, EXP2 was chosen for further validation of its single-cell expression at multiple time points.

RNA fluorescent *in situ* hybridization reveals distinct stage transitions and large variations in gene expression levels

To more clearly define the distinct cell types associated with expression heterogeneity, we used RNA-FISH as a second validation method to determine the expression levels of these cluster-specific transcripts over multiple time points of the IDC. Specifically, we chose SERA4, EXP2, and RAP1 as representative markers for each cluster. Parasites were collected approximately 32 to 38 hours post-synchronization and then staged based on size and morphology [41]. These cells were stained by RNA-FISH for the three markers simultaneously and scored for their stage and expression of each marker. In total, 647 parasites were counted, and 339 parasites exhibited expression of at least one of the three markers and were included in further analyses.

First, we determined the percentage of parasites expressing each marker at each stage. As expected for genes in clusters 1 and 3, SERA4 and RAP1 showed a stage-specific expression
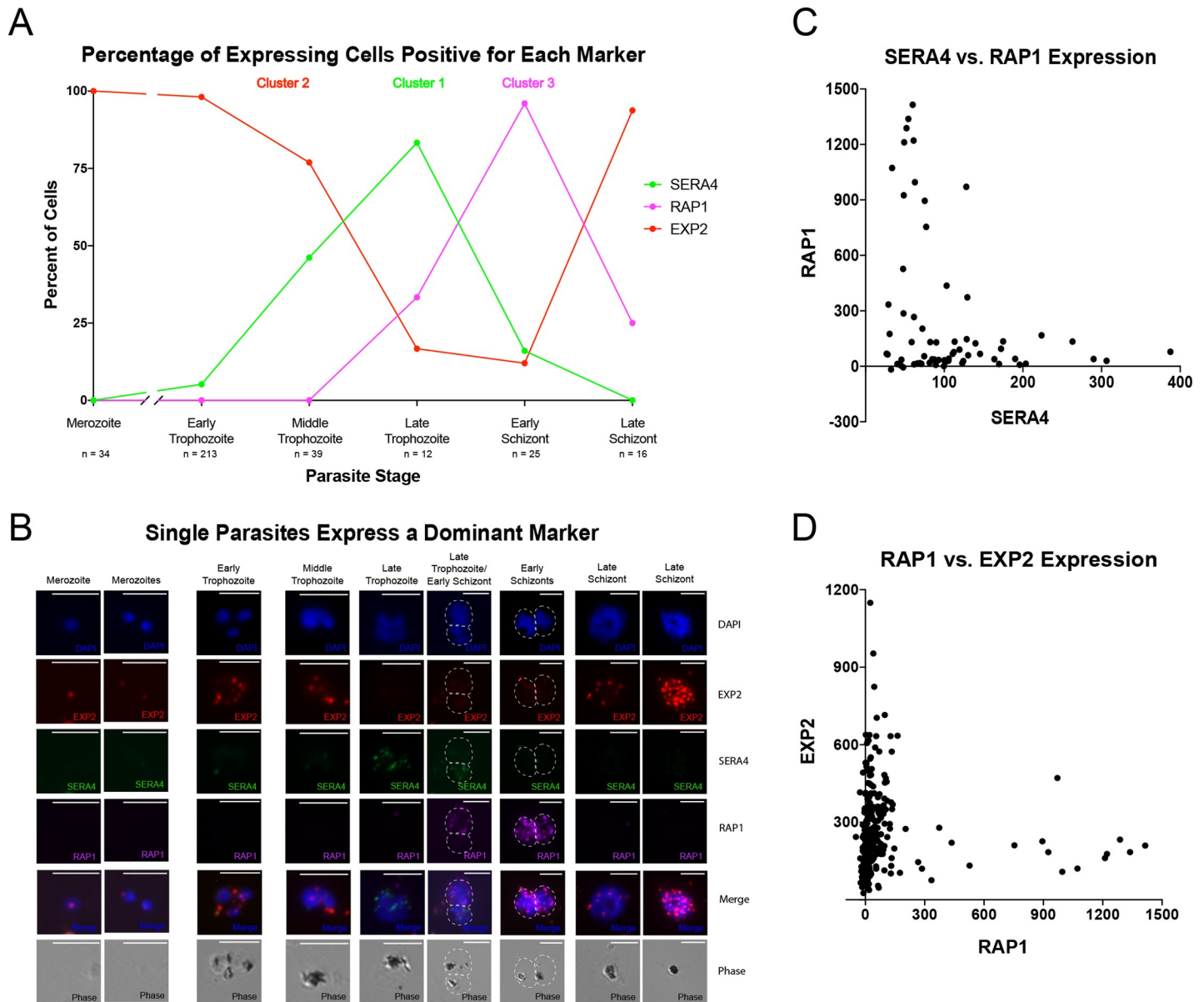
Fig 6. RNA-FISH reveals cell cycle progression of gene expression and distinct pattern among expressing cells for markers SERA4 (cluster 1), EXP2 (cluster 2), and RAP1 (cluster 3). A) The percentage of cells expressing each marker gene was determined for the morphological stage of the parasite and plotted over time, indicating distinct cell cycle progression of gene expression with some overlap between markers during stage transition. The number of expressing cells per stage is indicated by n. B) Representative images for expression of markers during cell cycle progression, with parasites expressing EXP2 (red), SERA4 (green), and RAP1 (magenta). Scale bar: 5 μm. C) The pair-wise correlation between expression values measured for RNA-FISH markers SERA4 and RAP1 reveal wide variation in expression levels and separate expression. Values are relative fluorescence units. D) The pair-wise correlation between expression values measured for RNA-FISH markers RAP1 and EXP2 reveal wide variation in expression levels and separate expression. Values are relative fluorescence units.

(Fig 6A). The percentage of parasites expressing SERA4 began to increase dramatically in middle trophozoites and peaked in late trophozoites, with a majority of late trophozoites expressing this transcript. As the percentage of SERA4-expressing cells decreased markedly in early schizonts (Fig 6A), RAP1 expression peaked and was expressed in nearly all parasites at this stage. Next, we looked at EXP2 expression. Consistent with the previous expression data, EXP2 exhibited a variable and more complex gene expression pattern through the IDC (Fig 6A). It was the only marker expressed in merozoites (Fig 6A and 6B). In addition, EXP2 was

also expressed highly in early trophozoites as well as late dividing schizonts (Fig 6A and 6B). Therefore, the expression of cluster-specific marker genes followed a cluster 2, cluster 1, and cluster 3 progression.
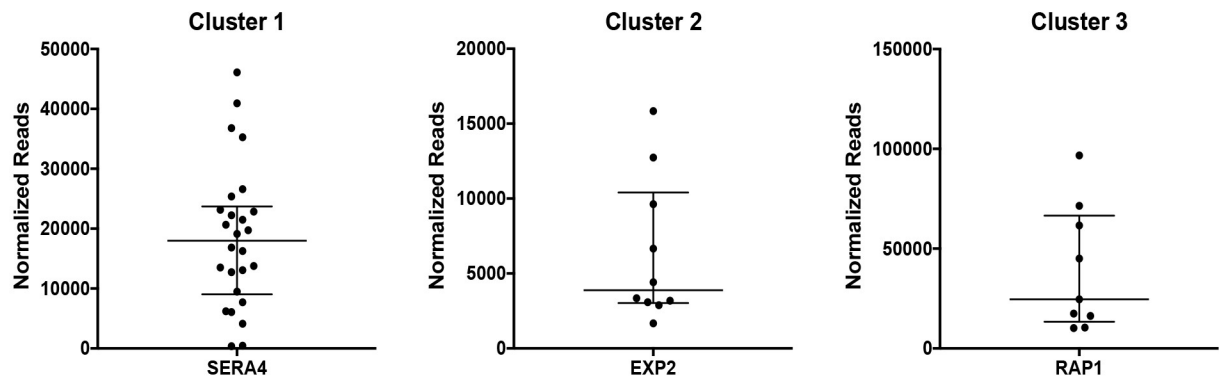
Among individual parasites, the expression of SERA4 and RAP1 generally did not overlap, with dominant expression of either SERA4 or RAP1 in each cell (Fig 6B and 6C). This indicates stage-specific expression of SERA4 and RAP1 in each cell with a distinct transition, a pattern also observed in a previous scRNA-seq study [17]. Such a distinct transition between stages differs from the previously understood smooth transition during progression of the IDC [2] as seen in bulk-cell datasets (S5 Fig). For instance, RAP1 is highly expressed starting at 32 hours post-invasion through 4 hours after reinvasion, spanning around 20 hours, as shown in previous microarray data (S5 Fig). However, our RNA-FISH analysis suggests that its expression primarily occurs only in schizonts 38–44 hours post-invasion (Fig 6A and 6B). Additionally, in late schizonts, the expression of EXP2 was separate from RAP1, which was expressed in schizonts more than four hours from bursting (Fig 6A, 6B and 6D).

When visualizing the scRNA-seq data, marker genes from all three clusters exhibited unexpected large variations in expression levels (Fig 7A). For stage-specific markers SERA4 and RAP1, the normalized reads ranged from 377.8 to 46,103.5 for SERA4 in cluster 1 parasites and 10,223.9 to 96,706.5 for RAP1 in cluster 3 parasites (Fig 7A). When we performed a variance analysis across all parasites, we found a similar result, as SERA4 and RAP1 had large biological variations (S3 Table). These wide variations in transcript levels were validated via RNA-FISH for each stage (Figs 7B, S6 and S7). For instance, the relative fluorescence units of SERA4 in late trophozoites with only SERA4 expression (of the three markers tested) ranged from 87.4 to 223.8 (Fig 7B), indicating a wide range of transcription. To normalize this RNA-FISH data, we chose a gene called actin I with low variance across all parasites from scRNA-seq (S3 Table). As we could only test three markers at once with RNA-FISH, we tested SERA4 and RAP1 along with the actin I control. Late trophozoites only expressing SERA4 and actin I showed a distribution of 100.9 to 258.6 relative fluorescence units for SERA4 (S6A Fig) and the pattern and distribution is similar when normalized to actin I (S6C Fig). Similarly, the expression levels of RAP1 assessed by RNA-FISH in individual early schizonts with only RAP1 expression ranged from 54.6 to 1338.0 relative fluorescence units (Fig 7B). Such individual heterogeneity was further validated with the comparison to actin I, showing that the distribution of RAP1 ranged from 117.2 to 2048.2 relative fluorescence units with normalized values keeping a broad distribution (S6B and S6D Fig). Across hundreds of parasites in multiple stages, this stage-specific and heterogeneous pattern holds true (S7A and S7B Fig). Compared to other stages, late trophozoites exhibit greater SERA4 expression and early schizonts exhibit greater RAP1 expression, with variability in their marker expression (S7A and S7B Fig). Such wide variations of stage-specific SERA4 and RAP1 expression among individual parasites may represent an additional level of regulation not seen in previous analyses. However, since the parasite staging by visualization inspection of morphology is intrinsically imprecise, it is also possible that the apparent varying expression levels may result from minor differences in IDC progression of individual parasites.

Our variance analysis also indicated EXP2 as a biologically variable gene, though not to the same extent as RAP1 and SERA4 (S3 Table). Among the sequenced cluster 2 parasites, there was dramatic and large individual variation in the levels of EXP2 expression, which ranged from 1669.8 to 15,843.8 normalized reads in scRNA-seq (Fig 7A). Consistent with this data, EXP2-expressing late schizonts showed a distribution of 113.8 to 1149.1 relative fluorescence units (RNA-FISH) (Fig 7B). Such variability in gene expression among individual cells was undetected at a population level, and could indicate the existence of persistently low or high EXP2 expressers within a population or quick bursts of transcription in small numbers of cells
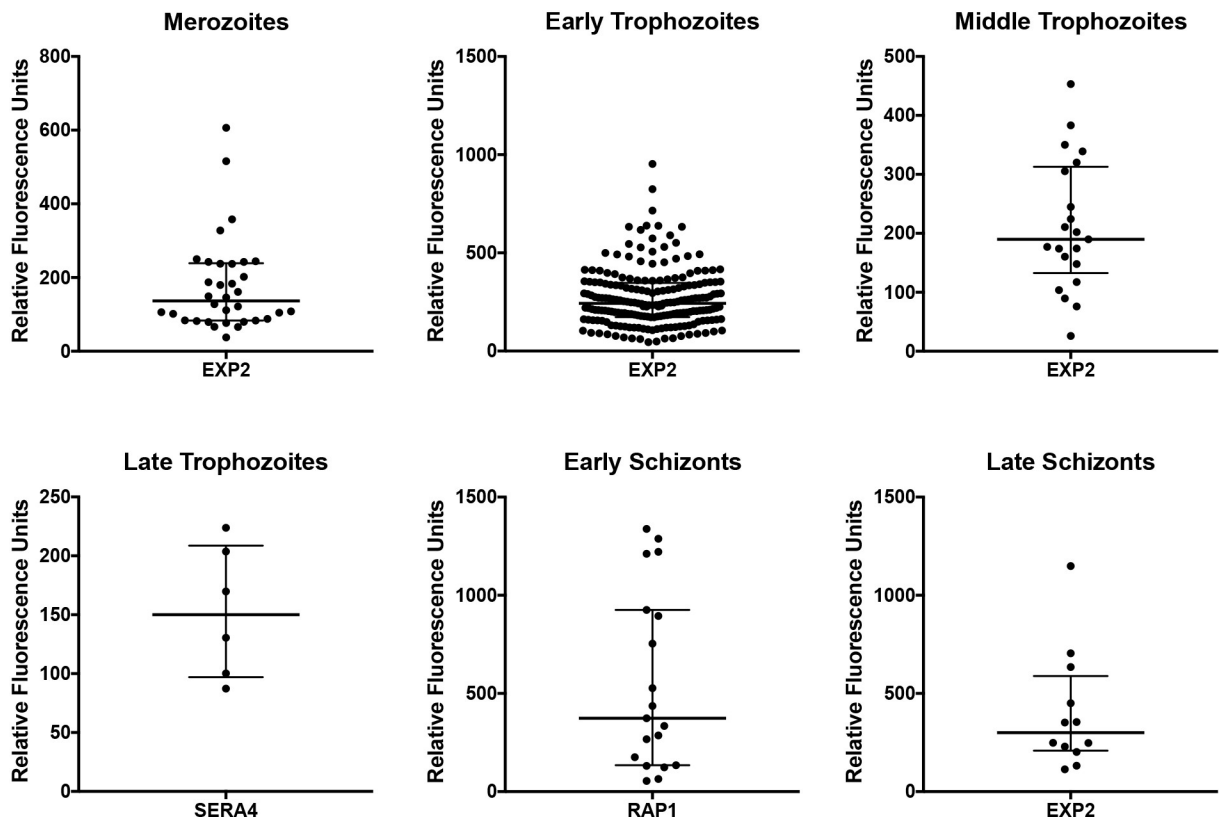
Fig 7. Single-cell analyses indicate wide variation among expressing cells for markers SERA4 (cluster 1), EXP2 (cluster 2), and RAP1 (cluster 3). A) Individual parasite single-cell transcriptomes were assessed for their normalized reads of marker genes representative of each cluster, which showed wide variation in the number of counts per cell. Only expressed values are shown. Error bars represent the median with the interquartile range. B) Individual parasites were measured for their mean fluorescence intensity representative of RNA expression for each marker, which showed wide variation in the expression levels among the exclusive marker-expressing parasites during each stage. Error bars represent the median with the interquartile range.

relative to the larger population. This trend is in line with wide variability in SERA4 and RAP1 expression (Figs 6C, 7A, 7B, S6A–S6D, S7A and S7B). Collectively, these data indicate that there are discrete and distinct transitional changes in the transcriptional program during late stage asexual intraerythrocytic development, a phenomenon previously masked in population analyses due to noise from other cells. An understanding of this unusual and hidden transcriptional timing is fundamental to our knowledge of *P. falciparum* transcriptional regulation and expression, especially as it relates to malaria drug targets.

Discussion

Here we applied the Fluidigm C1 microfluidic system and Smart-Seq to perform scRNA-seq of *P. falciparum* parasites during their asexual and sexual stages. Through scRNA-seq, we have identified a large number of gametocyte-specific genes as well as three unexpected sub-populations of asexual parasites with distinct gene expression. Furthermore, we have validated our scRNA-seq data using RT-PCR (gametocyte-specific genes) and RNA-FISH (cluster-specific markers EXP2, SERA4, and RAP1) across multiple time points. Our cluster-specific markers were further validated with transcriptomic data from a previous scRNA-seq study [17]. While the transcriptional programs of *P. falciparum* during the asexual IDC are highly choreographed with smooth transitions between stages, our scRNA-seq and RNA-FISH data uncover additional layers of heterogeneity not previously observed in population analyses. First, the percentage of SERA4- and RAP1-expressing parasites corresponds to the expected IDC stages and probably contributes to stage-specific expression seen in bulk-cell analyses. However, even among individual parasites scored as the same stage, the expression levels of SERA4 and RAP1 still vary significantly, with high expressers and low expressers. This diversity in expression levels may be due to parasite-to-parasite heterogeneity or subtle differences in cell cycle progression during the same morphological stage. Additionally, the transition between SERA4 to RAP1 expression of individual parasites does not contain as much overlap as predicted by bulk-cell analyses, consistent with another scRNA-seq paper [17]. As rhoptries are secreted during invasion and SERA proteins are involved in parasite egress [40, 42], it appears that their transcriptional pathways might interfere with each other during the late stages of the asexual cycle. Overall, these findings demonstrate the existence of additional regulatory layers and latent heterogeneity in *P. falciparum* populations that have been masked in population analyses. The regulatory mechanisms governing this heterogeneity may provide important novel insights into the transcriptional programs and phenotypic fitness of the parasite population during different stages of the IDC.

Most malaria studies focus on analyzing large populations of cells that can only capture the average gene expression of many parasites. These studies are limited in their ability to detect rare populations and heterogeneous gene expression. With the recent availability of various single-cell based methods, scRNA-seq has been applied in different biological contexts to reveal cellular heterogeneity, identify rare cell types, and unravel developmental pathways. Recently, single-cell approaches have been applied to *P. falciparum*. We previously used the Fluidigm C1 microfluidic system for single-cell qPCR of annotated male- and female-specific genes, where we uncovered a large number of sex-specific transcripts involved in gametocyte development [27]. Other studies have done scRNA-seq on other various platforms. In one study, analysis of thousands of single-cell transcriptomes positive for AP2-G expression identified gene signatures and additional regulators of sexual commitment [14]. A similar study using LysoPC depletion involved analysis of hundreds of AP2-G positive parasites and also identified a gene signature of sexual commitment, although only three genes overlapped with the previous study [15]. Another paper optimized Smart-Seq2 for use in *P. berghei* and *P.*

falciparum populations, using pseudotime and comparison with bulk-cell datasets to order the single-parasite transcriptomes into distinct clusters [17]. Another described a capillary-based platform for single-cell isolation and uncovered novel genes upregulated in early stage gametocytes with a modified Smart-Seq2 protocol [16]. The latest paper used 10x Genomics' droplet system on thousands of parasites, including field samples, to generate single *Plasmodium* transcriptomes across the complete life cycle [18].

Our scRNA-seq analysis generally agrees with previous studies but also reveals some novel insights. First, our comparison of asexual and sexual parasites identified multiple gametocyte-specific genes that were not described in previous proteomic, microarray, or sequencing studies [5, 25, 26, 43], including eight that were validated by qPCR. Of these eight, four have unknown functions and the other four putatively include a peptidase, kinesin, RAP protein, and AAA family ATPase, the most highly induced. Further investigation is needed to elucidate their roles in gametocytogenesis and sexual reproduction. Second, our data suggests a more complicated expression pattern during schizogony, with distinct transitions from SERA to rhoptry expression and then rhoptry to EXP2 expression, previously masked in population analyses. Other scRNA-seq studies also show these distinct transitions by using pseudotime analysis [14, 17]. In one study, expression of microneme and inner membrane complex (IMC) genes followed rhoptry expression, but this was only captured in *P. berghei* and not *P. falciparum* [17]. Drop-seq of thousands of *P. falciparum* parasites was able to capture the subsequent expression of microneme and IMC genes as well as CDPK5 after rhoptry expression in *P. falciparum* [14]. In contrast to these studies, our RNA-FISH data, together with morphological staging, have independently verified and ordered cells according to their gene expression. In particular, this approach showed that EXP2 is expressed at multiple times throughout the life cycle, and similar validation methods may be used to analyze transcripts with non-conventional expression patterns uncovered by scRNA-seq analysis.

One unexpected finding in our study is the existence of a wide distribution in the number of transcripts expressed among different marker-expressing parasites of similar IDC stages determined by visual inspection. Such cell-to-cell variations in the expression level are most strikingly seen in schizonts expressing either RAP1 or EXP2 in our RNA-FISH dataset. Cell-to-cell expression variability in these genes/proteins may result from intrinsic differences between individual parasites or slight differences in cell cycle and differentiation progression. It is important to note that parasite staging by visualisation by light microscopy is inherently imprecise. Therefore, one possible explanation is that scRNA-seq and RNA-FISH may reveal the latent heterogeneity not observable using visual staging methods. Regardless of the underlying mechanisms, such variations may be relevant for the interpretation of many malaria experiments based on stages. These variations may lead to significant functional heterogeneity among different parasites and enhance the overall population fitness, a possibility that may be studied with fluorescent reporter strains of parasites and single-cell imaging. For instance, this heterogeneity may contribute to *P. falciparum*'s ability to adapt to various stresses, different host environments, and drug treatments, with some parasites better equipped to handle fluctuations and/or stresses than others. Recently, it was shown that higher expression levels of EXP2 are correlated with higher numbers of merozoites produced per dividing schizont, indicating that EXP2 is important for parasite fitness [32]. Perhaps such high expressing cells exhibit other phenotypes such as an enhanced invasion ability or a proclivity towards drug resistance. Likewise, the varying expression of MDR1 may allow a sub-population of parasites to survive drug treatment. Such clonal fitness under stress may help to explain the association of MDR1 expression with the emergence of drug resistance. Similarly, rhoptry proteins of *Plasmodium* are implicated in the invasion of new host RBCs. The cell-to-cell variations in the levels of rhoptry expression may affect the parasite's invasion ability and the selection of RBCs to

invade, including new versus old cells. Therefore, such cell-to-cell variations of gene expression may enhance the adaptability of the parasite population to maintain homeostasis as it goes through different host environments and a wide variety of stresses. Thus, these huge cell-to-cell variations in the expression levels could have implications for drug development and resistance, immune pressures, and malaria disease severity and clinical outcomes.

Importantly, as sexual reproduction only occurs in the mosquito, these asexually replicating haploid cells putatively form clonal populations with little genetic diversity. Furthermore, all of the analyzed single parasites come from a population that has been cultured in the same media with identical conditions. Even with these identical genetic and environmental factors, our data clearly indicate that there are incredible levels of cell-to-cell variation which have not been recognized in previous population analyses. Therefore, elucidating the genetic basis and sources of cell-to-cell variations will be of high interest.

First, it is possible that there are epigenetically controlled mechanisms that enhance or repress stage-specific gene expression programs in each individual parasite. Such epigenetic regulation of expression levels in *P. falciparum* has been described for *var* genes [44, 45] and sexual commitment [9]. Additionally, genetically homogenous parasite populations can be epigenetically highly heterogeneous, with transcriptionally diverse populations having the ability to quickly adapt to stress [46]. A previous study has also shown that transcript level variation is strongly influenced by parasite genotype and can be controlled by several regulatory hotspots in the genome, including a region associated with multiple drug resistance [47]. Transcriptional heterogeneity also results from sporadic bursts of gene expression in a small number of parasites that is transient and not passed along to the next generation. Finally, it is possible that cell-to-cell expression variations may also result from the intrinsic and extrinsic genetic noise noted in other systems [48]. RAP1 and other highly variably expressed genes can be used to investigate the underlying mechanisms and potential contributions of transcriptional bursts or genetic noise.

We envision that scRNA-seq combined with RNA-FISH will have broader applications in the parasitology field. While scRNA-seq can provide insight into the many genes expressed in a single parasite, including rare expression, RNA-FISH can be used to identify the spatial distribution of specific transcripts and quantify their levels in individual cells. When combined with morphological analysis, they can be used to uncover distinct but rare parasite phenotypes that may be major contributors of disease. We anticipate that such single-cell approaches can be used to determine gene expression in malaria-infected RBCs in patients during the natural progression of the disease or during drug treatment. We can identify the rare drug-resistant parasites that emerge during drug treatment due to their fitness advantage and devise strategies to prevent drug resistance. Given the small number of parasites committed to sexual differentiation, another obvious application of scRNA-seq is to define the differentiation hierarchy and underlying molecular mechanisms of sexual differentiation into male and female gametocytes to identify mating-type determination [14, 15, 27]. In addition, since RBCs also have RNA transcripts [49–51] with the potential for regulating malaria infection [52, 53] and intercellular communication [54], it will be interesting to perform parallel transcriptome profiling of the host RBCs and infecting parasites to identify novel genetic determinants of invasion. Similar approaches can also be used to dissect host-pathogen interactions in hepatocyte and mosquito host cells.

Materials and methods

Parasite strain and culture

P. falciparum strain 3d7a was obtained from the Malaria Research and Reference Reagent Resource Center (MR4). Using standard procedures, parasites were cultured in 30 mL of

RPMI 1640 medium supplemented with 0.5% AlbuMAX II at 2% hematocrit in human B
+ erythrocytes at 3% O_2/5% CO_2 [55, 56].

Culture synchronization and gametocyte induction

To synchronize cultures, parasites were treated with 5% D-sorbitol twice, approximately 48
hours apart during the ring stage [57]. The synchronization was verified by Giemsa stained
blood smears. To enrich for gametocytes, 50 mM N-acetyl glucosamine was added to the cul-
ture on the day following the second 5% D-sorbitol treatment. This addition of N-acetyl glu-
cosamine to cultures with a parasitemia between 6–10% was considered day 0, with drug
treatment continuing for a total of 72 hours with daily medium changes [58]. Gametocyte pro-
duction was monitored with Giemsa stained blood smears. On day 9, before gametocyte col-
lection, the gametocyte culture was treated with 5% D-sorbitol to eliminate any contaminating
trophozoites.

Purification of individual parasites

A 40/70% Percoll density gradient was used to either isolate late stage asexual (approximately
34 hours post-invasion) or day 9 gametocytes from rings, early gametocytes, and uninfected
erythrocytes [58]. To eliminate all uninfected erythrocytes, the sample collected from the 40/
70% interface was run through a MACS LS column, yielding pure trophozoite/schizont and
gametocyte populations. The isolated cells were pelleted by centrifugation and resuspended to
a concentration of at least $2.5x10^5$ cells/mL in complete malaria media.

Single-cell capture and sequencing

To determine the viability and concentration of the isolated cells, 10 μL of cells were mixed
with 10 μL of trypan blue stain (0.4%) and loaded onto a Countess Automated Cell Counter
(Thermo Fisher Scientific). Viable cells were adjusted to concentrations of $3.2x10^5$ cells/mL for
gametocytes and $4.4–4.9x10^5$ cells/mL for late stage asexual parasites. According to Fluidigm
protocol, 60 μL of gametocytes were mixed with 40 μL of Suspension Reagent (Fluidigm)
before being loaded onto a C1 Single-Cell Auto Prep Integrated Fluidic Circuit (IFC) for
mRNA-seq (5–10 μm). We noticed that asexual parasite capture rates at this ratio were low,
and upon further inspection, saw that cells at this concentration fell to the bottom on the cell
inlet, indicating that the asexual parasite concentration was too dense. To optimize the cell
buoyancy and improve capture, we tested various ratios of cell number to suspension buffer
using a concentration of 400 cells/μL. We found that 40 to 50 percent of cells making up the
final suspension buffer was optimal, although capture rates were still low at 24 to 27 percent
(compared to late-stage gametocyte capture rates of 77 percent in this study and 92 percent in
a previous study [27]). Therefore, we used two C1 Single-Cell Auto Prep IFCs to capture 46
synchronized asexual parasites on two different days.

For scRNA-Seq library construction, reagent mixes were prepared according to the Flui-
digm protocol (PN 100–7168 L1). The IFC was primed and loaded with 6 μL of cell mix
according to manufacturer's instructions, using the mRNA Seq: Prime (1771x) and mRNA
Seq: Cell Load (1771x) scripts. Once the cells were loaded, each well of the IFC was checked for
the presence of a single parasite under a microscope. Harvest reagent, lysis mix, RT mix, and
PCR mix were added to the IFC according to Fluidigm protocol, with 1 μL of ERCC spike-in
sequences (Thermo Fisher Scientific, 4456740) added to the lysis mix at a $2.5 x 10^5$ dilution of
the original product. The IFC was then loaded into the C1 system, running the mRNA Seq: RT
& Amp (1771x) script overnight. Amplified products were harvested and diluted, with approx-
imately 3 μL of amplicons diluted in 10 μL of C1 DNA Dilution Reagent (Fluidigm).

Population controls for two asexual and one gametocyte population were prepared in parallel with the C1 single-cell capture according to "Appendix A: Run the Tube Controls". "Protocol A: Tube Controls with Purified RNA" was followed. Samples were vortexed for one minute for the homogenization step, and a gDNA Eliminator spin column was used for the removal of contaminating genomic DNA. When diluting the final products after amplification, 40 μL of C1 DNA Dilution Reagent (Fluidigm) was used to every 1 μL of PCR product.

The Nextera XT kit was then used to generate the sequencing library from the cDNA according to the Fluidigm protocol. Once the libraries were quantified, they were sequenced in one lane on an Illumina HiSeq 2500 with 50 bp single-end reads. 10% PhiX was used as a spike-in control for *P. falciparum*'s highly AT-rich genome. All sequencing data is available at NCBI, SRA accession number SRP151825.

Generating gene counts

Single-cell RNA-seq data was first processed using the TrimGalore toolkit [59] which employs Cutadapt [60] to trim low-quality bases and Illumina sequencing adapters from the 3' end of the reads. Only reads that were 20nt or longer after trimming were kept for further analysis. Using the STAR RNA-seq alignment tool [61], reads were mapped to the *P. falciparum* 3D7 genome and transcriptome sequence (PlasmoDB r13.0), as well as the GRCh37v75 version of the human genome and transcriptome to filter out human RNA originating from the blood sample. Reads were kept for subsequent analysis if they mapped to a single genomic location of the *P. falciparum* genome. Gene counts were compiled using the HTSeq tool [62].

Cell and gene filtering

QC filtering was performed using the R package scater v1.4.0 [63]. As expected, the number of expressed genes for bulk RNA controls was higher (between 4,000 and 5,000) than for sequenced single cells (between 1,200 and 3,600) (Fig 2). Principal component analysis (PCA) based on phenotypic data such as library size, number of expressed genes or spike-in (ERCCs) proportion, did not expose any outliers out of the 51 cells that were sequenced. To filter out low-abundance genes, only genes with average counts at least equal to 1 were kept. 4,694 genes remained after gene filtering when considering all 51 cells, and 4,463 remained when considering the 46 non-gametocyte cells.

Normalization of gene expression

To normalize gene expression, size factors were calculated using the R package scran v1.4.5 [64]. The package scran estimates cell-specific biases by deconvolving size factors using cell pools. The assumption is that most genes are not differentially expressed between cells. Expression values were normalized using the previously generated size factors then log-transformed. Normalization was performed independently on two sets of cells: the whole 51 cells, and the 46 non-gametocyte cells. All downstream applications were performed using normalized gene expression. The normalized gene expression is available in S1 Table (51 cells) and S2 Table (46 cells).

Clustering

To perform clustering, we used the R package SC3 v1.4.2 [24]. After selecting genes that are expressed in at least 10% of cells but also no more than 90% of cells, *SC3* performs k-means clustering on distance matrices between cells that have been transformed with either principal component analysis (PCA) or Laplacian eigenvectors. *SC3* combines the obtained clustering

solutions into a consensus matrix that summarizes how often two cells are part of the same cluster. The final clustering solution is given by the complete-linkage hierarchical clustering of the consensus matrix into k groups. Prior to clustering, *SC3* can estimate an optimal number of clusters using the Tracy-Widom theory on random matrices. When clustering the 51 cells, SC3 estimated the optimal number of clusters to be four, with the five gametocytes forming a stable cluster of their own as measured by its silhouette width of 1. A silhouette width is a quantitative measure between 0 and 1 that describes the cohesion within a cluster, as well as its separation from other clusters, with the best clustering occurring when the average silhouette width is close to 1. The silhouette widths of the three other clusters were 0.51 (cluster 1), 0.81 (cluster 2) and 0.69 (cluster 3) (S8A Fig). To analyze the asexual cells independently, we also performed clustering of the 46 non-gametocyte cells. Consistent with the first clustering analysis, SC3 suggested three clusters, almost identical to the asexual clusters defined in the four-cluster analysis (only three cells clustered differently). Their silhouette widths were 0.49 (cluster 1), 0.88 (cluster 2) and 0.71 (cluster 3) (S8B Fig). Both datasets with gametocytes and without gametocytes were subjected to PCA analysis for visualization of the clusters (Figs 3A and 4A).

Identification of markers

To identify candidate marker genes for each cluster, we used SC3 to evaluate the potential for each gene to be a marker for the cluster in which its mean expression value was the highest. For each cluster, we retained genes with a non-zero probability of being a marker based on a Wilcoxon signed rank test. We then ordered these candidate markers by decreasing area under the receiver operating characteristic (AUROC) for further validation. This was done both for the five gametocytes versus all other cells in the four-cluster analysis, and for each cluster in the three-cluster analysis.

Variance analysis of gene expression

To analyze variance for each gene in the scRNA-seq dataset, the R package *scran* was used to fit a mean-variance trend to the spike-in transcripts (S3 Table). The trend helps quantify the technical component of the variance. The biological component for each gene is then deduced as the difference between the total variance and the technical component.

Inferring a cell trajectory

To generate a cell progression hypothesis, we used the R package *monocle* v2.8.0 [24]. A differential expression test was performed for each gene using the clusters as indicator variables (full model ~cluster and reduced model ~1) to detect genes expressed differentially between clusters. As a result, 84 genes with a q-value from the likelihood ratio tests < 0.01 were chosen as genes to order and infer a trajectory. We then used the DDRTree dimensionality reduction algorithm from *monocle* to align the cells on a branched path, as shown in Fig 4C. The trajectory includes a main path and two very small branches. To confirm this analysis, we also ran *monocle* without using clusters as an indicator variable for selecting genes to represent cell progress. Two additional analyses were run, each using a different set of genes: 1) highly expressed genes with an average gene expression ≥ 10, and 2) genes with high dispersion across cells based on the expected mean-dispersion calculated by *estimateDispersions*. Both trajectories were consistent with the clustering and ordering of cells calculated using *SC3* clusters as indicator variables. However, both additional analyses supported a larger independent branch for a portion of cells from cluster 1 (S2 Fig). Because cluster 2 markers are expressed the earliest [2], cluster 2 cells were used to define a start for all cell trajectories and pseudotime calculated by *monocle*.

To identify genes that co-vary across pseudotime, we used the lineage analysis based on highly expressed genes and focused on the main trajectory. The analysis was independent from our previous *SC3* clustering. Using the *branchTest* function in *monocle*, we identified genes that are differentially expressed between branches originating from both branching points (black circles 1 and 2 in S2 Fig). We restricted the two resulting gene lists to genes with a q-value < 0.01 and obtained 141 unique genes from the union of these two lists. We used the *plot_pseudotime_heatmap* function in *monocle* to visualize genes which expression varies around branch points 1 and 2, and restricted cells to the 32 cells on the main trajectory (S3 Fig). The heatmap displays smooth expression curves and both gene expression and pseudo-time are binned to 100 columns.

Assessing bias towards longer transcripts

Transcript length was deduced from the *P. falciparum* 3D7 PlasmoDB r13.0 transcript sequences. To assess the effect of transcript length, we compared the length distribution of transcripts in the initial dataset (n = 5,452) to that of "expressed genes" remaining after filtering of genes with an average gene count below 1 (n = 4,694).

Validation of asexual clustering

To validate the clustering of the asexual cells into three clusters, we clustered the *P. falciparum* asexual scRNA-seq dataset available from Reid et al. [17] using the 151 markers identified from the clustering of our asexual dataset. We processed the raw count tables the same way as ours. Using the R package *scran*, we detected eight outlier cells that we removed from the dataset, and filtered genes with an average count below 1. We then normalized the resulting dataset that consisted of 4,844 expressed genes and 172 samples. All 151 asexual markers identified earlier were expressed in the Reid et al. asexual dataset and the gene space of the dataset was reduced to these 151 genes for clustering. *SC3* estimated the optimal number of clusters to be three with an average silhouette width of 0.72. After testing the 151 genes as markers for these three clusters with similar thresholds, we compared our three clusters to the obtained three Reid et al. clusters based on shared markers (S6 Table).

Bulk-cell qPCR

RNA was extracted from synchronized late stage asexual parasites and day 9 gametocytes using TRIzol Reagent (Invitrogen). The RNeasy Mini Kit (Qiagen) was used to remove DNA by DNase treatment according to Appendix E: DNase Digestion of RNA before RNA Cleanup. cDNA was synthesized from 50 ng of RNA using SuperScript II Reverse Transcriptase (Invitrogen) and 50 ng of random hexamers (Bioline). Primers (Eurofins Genomics) were designed for 8 novel gametocyte-specific genes (S4 Table), which were tested for their gene expression differences using Power SYBR Green (Invitrogen) real-time PCR. Pfs16 was used as a control for gametocyte numbers between asexual and day 9 gametocyte samples. The ΔCt value was calculated relative to the constitutive control, tRNA ligase. The $\Delta\Delta Ct$ value was calculated relative to the asexual sample, with the expression fold change of $2^{-\Delta\Delta Ct}$ representing the increase in expression of genes in the gametocyte sample.

RNA fluorescent *in situ* hybridization

400 μL of a synchronized culture was lysed with 0.15% saponin to isolate single parasites. Briefly, parasites were pelleted and then washed with PBS twice before being resuspended in at least 1 mL of 0.15% saponin. The parasites were then vortexed for one minute before being

placed on ice for 20 minutes. The cells were centrifuged at 2100xg for 12 minutes at 4°C and washed once in PBS. After centrifugation, these cells were fixed in 1 mL of 4% formaldehyde for 30 minutes. The tube was vortexed halfway between the incubation to resuspend the cells. "Appendix A: Sample Preparation Procedure for Suspension Cells" of the QuantiGene View-RNA ISH Cell Assay Kit (Thermo Fisher Scientific) was then followed after the fixation step. 30 μL of poly-D-lysine (Sigma) was used to coat microscope slides, and two hydrophobic barriers were drawn on each slide, one on which probes were added and one in which they were not. Cells were baked on the slides at 50°C for 20 minutes. The assay procedure then continued with the rehydration step, adding the solutions directly to the slide. The probes used were *Plasmodium falciparum* Sera4 (VF4-6000578), Rap1 (VF6-21027), EXP2 (VF1-6000683), and Actin I (VF1-21025), all from Thermo Fisher Scientific. Mounted samples were cured overnight with protection from light and were viewed using a Zeiss Axio Observer 7.1 microscope with a 63x oil objective. DAPI, Cy5, mRFP1.2, and EGFP filters were used to detect DAPI, Cy5, Cy3, and FITC, respectively. The images were processed using Fiji software and individual cells were outlined using the freeform tool and measured for their mean fluorescence signal. The background signal from each image and channel was eliminated from individual cell fluorescence signals. Individual cells were also scored for their stage-specific progression and gene expression, according to previous morphological analysis [41]. Free merozoites were scored after egress from schizonts and were defined by DAPI staining.

Supporting information

S1 Fig. The distribution of transcript lengths from expressed genes (after filtering out genes with an average count below 1) shows slight length bias for longer transcripts when compared to the distribution of transcript lengths from all genes.
(PDF)

S2 Fig. Pseudotemporal ordering of 46 single asexual parasites by *monocle* 2. Highly expressed genes with an average gene expression ≥ 10 were used for ordering cells and inferring a trajectory. Top panel: cells are colored by pseudotime. Bottom panel: cells are colored by cluster assignment.
(PDF)

S3 Fig. Heatmap showing the smoothed, scaled and centered gene expression for 141 genes that show scRNA-seq dynamic expression around branch points 1 and 2 (S2 Fig). Only gene expression in the 32 cells of the main trajectory was used for the heatmap. Cell expression is smoothed and spread across 100 bins distributed uniformly across pseudotime. Genes are grouped by expression pattern.
(PDF)

S4 Fig. Consensus clustering matrix of 172 single asexual *P. falciparum* parasites from Reid et al. represents the optimal three clusters estimated by SC3. Parasites within the same cluster show the strongest similarity when compared to parasites from other clusters, although there is significant heterogeneity even between parasites in the same cluster. The color scale indicates the likelihood that two cells are arranged in the same cluster (blue: low (0), red: high (1)).
(PDF)

S5 Fig. Hierarchical clustering of the top marker genes for each cluster from the microarray 48-hour time course published in Bozdech et al., 2003. The $\log_2$ (Cy5/Cy3) ratios for markers of clusters 1 (A), 2 (B), and 3 (C) were centered by the mean and clustered using complete linkage to examine their expression profile during the IDC. The black arrow represents

the expected stage of the harvested parasite population for scRNA-seq.
(PDF)

S6 Fig. RNA-FISH analyses indicate wide variation among expressing cells for A) SERA4 (late trophozoites) and B) RAP1 (early schizonts). Individual parasites were measured for their mean fluorescence intensity representative of RNA expression for each marker, which showed wide variation in the expression levels among the exclusive marker-expressing parasites. These values were normalized to actin I for C) SERA4 and D) RAP1. Error bars represent the median with the interquartile range.
(PDF)

S7 Fig. RNA-FISH analyses reveal variation of A) SERA4 and B) RAP1 when normalized to actin I control. SERA4 shows highest expression in late trophozoites, while RAP1 shows highest expression in early schizonts. Error bars represent the median with the interquartile range.
(PDF)

S8 Fig. Silhouette widths of single parasites reflect both how well matched individual cells are to their own cluster (cohesion) and how poorly matched they are to neighboring clusters (separation) for A) 51 sexual and asexual cells, B) 46 asexual cells, and C) 172 asexual cells from a previous dataset [17]. A silhouette width value ranges from -1 to +1, with a high value indicating that the cell is similar to its own cluster but not to its neighboring clusters.
(PDF)

S1 Table. Normalized gene expression for 51 cells.
(XLSX)

S2 Table. Normalized gene expression for 46 cells.
(XLSX)

S3 Table. Variance analysis across 51 cells.
(XLSX)

S4 Table. Primers used for bulk-cell qPCR validation of novel gametocyte-specific genes.
(XLSX)

S5 Table. List of 340 gametocyte-specific genes revealed by scRNA-seq. Annotations for significantly differentially expressed male and female genes were obtained from Lasonder et al. [28].
(XLSX)

S6 Table. List of 151 cluster markers shared between our dataset and Reid et al.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Katelyn A. Walzer, Jen-Tsan Chi.

Data curation: Katelyn A. Walzer, Liane Y. Emerson.

Formal analysis: Katelyn A. Walzer, Hélène Fradin, Liane Y. Emerson, David L. Corcoran, Jen-Tsan Chi.

Funding acquisition: Jen-Tsan Chi.

Investigation: Katelyn A. Walzer, Hélène Fradin, Liane Y. Emerson, David L. Corcoran, Jen-Tsan Chi.

Methodology: Katelyn A. Walzer, Hélène Fradin, David L. Corcoran, Jen-Tsan Chi.

Project administration: Jen-Tsan Chi.

Resources: Jen-Tsan Chi.

Supervision: Jen-Tsan Chi.

Writing – original draft: Katelyn A. Walzer, Jen-Tsan Chi.

Writing – review & editing: Katelyn A. Walzer, Hélène Fradin, Liane Y. Emerson, David L. Corcoran, Jen-Tsan Chi.

References

1. World Health Organization. World malaria report. Geneva, Switzerland: World Health Organization. p. volumes.

2. Bozdech Z, Llinas 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. Epub 2003/08/21. https://doi.org/10.1371/journal.pbio.0000005 PMID: 12929205; PubMed Central PMCID: PMC176545.

3. Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Bohme U, et al. New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. Molecular microbiology. 2010; 76(1):12–24. https://doi.org/10.1111/j.1365-2958.2009.07026.x PMID: 20141604; PubMed Central PMCID: PMC2859250.

4. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, et al. Discovery of gene function by expression profiling of the malaria parasite life cycle. Science. 2003; 301(5639):1503–8. https://doi.org/10.1126/science.1087025 PMID: 12893887.

5. Lopez-Barragan MJ, Lemieux J, Quinones M, Williamson KC, Molina-Cruz A, Cui K, et al. Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum. BMC genomics. 2011; 12:587. https://doi.org/10.1186/1471-2164-12-587 PMID: 22129310; PubMed Central PMCID: PMC3266614.

6. Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL. Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains. Nucleic Acids Res. 2006; 34(4):1166–73. https://doi.org/10.1093/nar/gkj517 PMID: 16493140.

7. Kafsack BF, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, et al. A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature. 2014; 507 (7491):248–52. https://doi.org/10.1038/nature12920 PMID: 24572369; PubMed Central PMCID: PMC4040541.

8. Sinha A, Hughes KR, Modrzynska KK, Otto TD, Pfander C, Dickens NJ, et al. A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium. Nature. 2014; 507(7491):253–7. https://doi.org/10.1038/nature12970 PMID: 24572359; PubMed Central PMCID: PMC4105895.

9. Filarsky M, Fraschka SA, Niederwieser I, Brancucci NMB, Carrington E, Carrio E, et al. GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. Science. 2018; 359(6381):1259–63. https://doi.org/10.1126/science.aan6042 PMID: 29590075.

10. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic variation in Plasmodium falciparum. Annu Rev Microbiol. 2008; 62:445–70. Epub 2008/09/13. https://doi.org/10.1146/annurev.micro.61.080706.093134 PMID: 18785843.

11. Grun D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature. 2015; 525(7568):251–5. https://doi.org/10.1038/nature14966 PMID: 26287467.

12. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature. 2013; 498

(7453):236–40. https://doi.org/10.1038/nature12172 PMID: 23685454; PubMed Central PMCID: PMC3683364.

13. Gasch AP, Yu FB, Hose J, Escalante LE, Place M, Bacher R, et al. Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. PLoS biology. 2017; 15 (12):e2004050. https://doi.org/10.1371/journal.pbio.2004050 PMID: 29240790; PubMed Central PMCID: PMC5746276.

14. Poran A, Notzel C, Aly O, Mencia-Trinchant N, Harris CT, Guzman ML, et al. Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. Nature. 2017; 551(7678):95–9. https://doi.org/10.1038/nature24280 PMID: 29094698.

15. Brancucci N, De Niz M, Straub T, Ravel D, Sollelis L, Birren B, et al. Probing Plasmodium falciparum sexual commitment at the single-cell level [version 3; referees: 1 approved, 1 approved with reservations]. Wellcome Open Research. 2018; 3(70). https://doi.org/10.12688/wellcomeopenres.14645.3 PMID: 30320226

16. Ngara M, Palmkvist M, Sagasser S, Hjelmqvist D, Bjorklund AK, Wahlgren M, et al. Exploring parasite heterogeneity using single-cell RNA-seq reveals a gene signature among sexual stage Plasmodium falciparum parasites. Exp Cell Res. 2018. https://doi.org/10.1016/j.yexcr.2018.08.003 PMID: 30096287.

17. Reid AJ, Talman AM, Bennett HM, Gomes AR, Sanders MJ, Illingworth CJR, et al. Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. Elife. 2018; 7. https://doi.org/10.7554/eLife.33105 PMID: 29580379; PubMed Central PMCID: PMC5871331.

18. Howick VM, Russell AJC, Andrews T, Heaton H, Reid AJ, Natarajan K, et al. The Malaria Cell Atlas: Single parasite transcriptomes across the complete Plasmodium life cycle. Science. 2019; 365(6455). Epub 2019/08/24. https://doi.org/10.1126/science.aaw2619 PMID: 31439762.

19. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015; 161(5):1202–14. https://doi.org/10.1016/j.cell.2015.05.002 PMID: 26000488; PubMed Central PMCID: PMC4481139.

20. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nature communications. 2017; 8:14049. https://doi.org/10.1038/ncomms14049 PMID: 28091601; PubMed Central PMCID: PMC5241818 L.M., D.A.M., S.Y.N., M.S.L., P.W.W., C.M.H., R.B., A.W., K.D.N., T.S.M. and B.J.H. are employees of 10x Genomics.

21. Soumillon M, Cacchiarelli D, Semrau S, van Oudenaarden A, Mikkelsen TS. Characterization of directed differentiation by high-throughput single-cell RNA-Seq. bioRxiv. 2014. https://doi.org/10.1101/003236

22. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nature methods. 2013; 10(11):1096–8. https://doi.org/10.1038/nmeth.2639 PMID: 24056875.

23. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, et al. A proteomic view of the Plasmodium falciparum life cycle. Nature. 2002; 419(6906):520–6. Epub 2002/10/09. https://doi.org/10.1038/nature01107 nature01107 [pii]. PMID: 12368866.

24. Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T, et al. SC3: consensus clustering of single-cell RNA-seq data. Nat Methods. 2017; 14(5):483–6. https://doi.org/10.1038/nmeth.4236 PMID: 28346451; PubMed Central PMCID: PMC5410170.

25. Silvestrini F, Bozdech Z, Lanfrancotti A, Di Giulio E, Bultrini E, Picci L, et al. Genome-wide identification of genes upregulated at the onset of gametocytogenesis in Plasmodium falciparum. Mol Biochem Parasitol. 2005; 143(1):100–10. Epub 2005/07/20. https://doi.org/10.1016/j.molbiopara.2005.04.015 PMID: 16026866.

26. Young JA, Fivelman QL, Blair PL, de la Vega P, Le Roch KG, Zhou Y, et al. The Plasmodium falciparum sexual development transcriptome: a microarray analysis using ontology-based pattern identification. Mol Biochem Parasitol. 2005; 143(1):67–79. Epub 2005/07/12. https://doi.org/10.1016/j.molbiopara.2005.05.007 PMID: 16005087.

27. Walzer KA, Kubicki DM, Tang X, Chi JT. Single-Cell Analysis Reveals Distinct Gene Expression and Heterogeneity in Male and Female Plasmodium falciparum Gametocytes. mSphere. 2018; 3(2). https://doi.org/10.1128/mSphere.00130-18 PMID: 29643077; PubMed Central PMCID: PMC5909122.

28. Lasonder E, Rijpma SR, van Schaijk BC, Hoeijmakers WA, Kensche PR, Gresnigt MS, et al. Integrated transcriptomic and proteomic analyses of P. falciparum gametocytes: molecular insight into sex-specific processes and translational repression. Nucleic Acids Res. 2016; 44(13):6087–101. Epub 2016/06/15. https://doi.org/10.1093/nar/gkw536 PMID: 27298255; PubMed Central PMCID: PMC5291273.

29. Eksi S, Morahan BJ, Haile Y, Furuya T, Jiang H, Ali O, et al. Plasmodium falciparum gametocyte development 1 (Pfgdv1) and gametocytogenesis early gene identification and commitment to sexual development. PLoS Pathog. 2012; 8(10):e1002964. Epub 2012/10/25. https://doi.org/10.1371/journal.ppat.1002964 PMID: 23093935; PubMed Central PMCID: PMC3475683.

30. Elsworth B, Matthews K, Nie CQ, Kalanon M, Charnaud SC, Sanders PR, et al. PTEX is an essential nexus for protein export in malaria parasites. Nature. 2014; 511(7511):587–91. https://doi.org/10.1038/nature13555 PMID: 25043043.

31. Beck JR, Muralidharan V, Oksman A, Goldberg DE. PTEX component HSP101 mediates export of diverse malaria effectors into host erythrocytes. Nature. 2014; 511(7511):592–5. https://doi.org/10.1038/nature13574 PMID: 25043010; PubMed Central PMCID: PMC4130291.

32. Garten M, Nasamu AS, Niles JC, Zimmerberg J, Goldberg DE, Beck JR. EXP2 is a nutrient-permeable channel in the vacuolar membrane of Plasmodium and is essential for protein export via PTEX. Nat Microbiol. 2018. https://doi.org/10.1038/s41564-018-0222-7 PMID: 30150733.

33. Gold DA, Kaplan AD, Lis A, Bett GC, Rosowski EE, Cirelli KM, et al. The Toxoplasma Dense Granule Proteins GRA17 and GRA23 Mediate the Movement of Small Molecules between the Host and the Parasitophorous Vacuole. Cell host & microbe. 2015; 17(5):642–52. https://doi.org/10.1016/j.chom.2015.04.003 PMID: 25974303; PubMed Central PMCID: PMC4435723.

34. Koenderink JB, Kavishe RA, Rijpma SR, Russel FG. The ABCs of multidrug resistance in malaria. Trends Parasitol. 2010; 26(9):440–6. https://doi.org/10.1016/j.pt.2010.05.002 PMID: 20541973.

35. Martin RE, Henry RI, Abbey JL, Clements JD, Kirk K. The 'permeome' of the malaria parasite: an overview of the membrane transport proteins of Plasmodium falciparum. Genome Biol. 2005; 6(3):R26. Epub 2005/03/19. gb-2005-6-3-r26 [pii] https://doi.org/10.1186/gb-2005-6-3-r26 PMID: 15774027.

36. Lustigman S, Anders RF, Brown GV, Coppel RL. The mature-parasite-infected erythrocyte surface antigen (MESA) of Plasmodium falciparum associates with the erythrocyte membrane skeletal protein, band 4.1. Molecular and biochemical parasitology. 1990; 38(2):261–70. https://doi.org/10.1016/0166-6851(90)90029-l PMID: 2183050.

37. Magowan C, Coppel RL, Lau AO, Moronne MM, Tchernia G, Mohandas N. Role of the Plasmodium falciparum mature-parasite-infected erythrocyte surface antigen (MESA/PfEMP-2) in malarial infection of erythrocytes. Blood. 1995; 86(8):3196–204. PMID: 7579415.

38. Zhang M, Faou P, Maier AG, Rug M. Plasmodium falciparum exported protein PFE60 influences Maurer's clefts architecture and virulence complex composition. International journal for parasitology. 2018; 48(1):83–95. https://doi.org/10.1016/j.ijpara.2017.09.003 PMID: 29100811.

39. Parish LA, Mai DW, Jones ML, Kitson EL, Rayner JC. A member of the Plasmodium falciparum PHIST family binds to the erythrocyte cytoskeleton component band 4.1. Malaria journal. 2013; 12:160. https://doi.org/10.1186/1475-2875-12-160 PMID: 23663475; PubMed Central PMCID: PMC3658886.

40. Counihan NA, Kalanon M, Coppel RL, de Koning-Ward TF. Plasmodium rhoptry proteins: why order is important. Trends Parasitol. 2013; 29(5):228–36. Epub 2013/04/11. https://doi.org/10.1016/j.pt.2013.03.003 PMID: 23570755.

41. Silamut K, Phu NH, Whitty C, Turner GD, Louwrier K, Mai NT, et al. A quantitative analysis of the microvascular sequestration of malaria parasites in the human brain. Am J Pathol. 1999; 155(2):395–410. https://doi.org/10.1016/S0002-9440(10)65136-X PubMed Central PMCID: PMC1866852. PMID: 10433933

42. Collins CR, Hackett F, Atid J, Tan MSY, Blackman MJ. The Plasmodium falciparum pseudoprotease SERA5 regulates the kinetics and efficiency of malaria parasite egress from host erythrocytes. PLoS Pathog. 2017; 13(7):e1006453. Epub 2017/07/07. https://doi.org/10.1371/journal.ppat.1006453 PMID: 28683142; PubMed Central PMCID: PMC5500368.

43. Lasonder E, Ishihama Y, Andersen JS, Vermunt AM, Pain A, Sauerwein RW, et al. Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. Nature. 2002; 419(6906):537–42. https://doi.org/10.1038/nature01111 PMID: 12368870.

44. Gomez-Diaz E, Yerbanga RS, Lefevre T, Cohuet A, Rowley MJ, Ouedraogo JB, et al. Epigenetic regulation of Plasmodium falciparum clonally variant gene expression during development in Anopheles gambiae. Sci Rep. 2017; 7:40655. https://doi.org/10.1038/srep40655 PMID: 28091569; PubMed Central PMCID: PMC5238449.

45. Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BT, et al. Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. PLoS Pathog. 2009; 5(9):e1000569. https://doi.org/10.1371/journal.ppat.1000569 PMID: 19730695; PubMed Central PMCID: PMC2731224.

46. Rovira-Graells N, Gupta AP, Planet E, Crowley VM, Mok S, Ribas de Pouplana L, et al. Transcriptional variation in the malaria parasite Plasmodium falciparum. Genome Res. 2012; 22(5):925–38. https://doi.org/10.1101/gr.129692.111 PMID: 22415456; PubMed Central PMCID: PMC3337437.

47. Gonzales JM, Patel JJ, Ponmee N, Jiang L, Tan A, Maher SP, et al. Regulatory hotspots in the malaria parasite genome dictate transcriptional variation. PLoS Biol. 2008; 6(9):e238. https://doi.org/10.1371/journal.pbio.0060238 PMID: 18828674; PubMed Central PMCID: PMC2553844.

48. Raser JM, O'Shea EK. Noise in gene expression: origins, consequences, and control. Science. 2005; 309(5743):2010–3. https://doi.org/10.1126/science.1105891 PMID: 16179466; PubMed Central PMCID: PMC1360161.

49. Chen P-H, Hong J, Chi J-T. Discovery, Genomic Analysis, and Functional Role of the Erythrocyte RNAs. Current Pathobiology Reports. 2017:1–6. https://doi.org/10.1007/s40139-017-0124-z

50. Doss J, Corcoran D, Jima D, Telen M, Dave S, Chi J-T. A comprehensive joint analysis of the long and short RNA transcriptomes of human erythrocytes. BMC Genomics. 2015; 16(1):952. https://doi.org/10.1186/s12864-015-2156-2 PMID: 26573221

51. Yang WH, Doss JF, Walzer KA, McNulty SM, Wu J, Roback JD, et al. Angiogenin-mediated tRNA cleavage as a novel feature of stored red blood cells. Br J Haematol. 2018. https://doi.org/10.1111/bjh.15605 PMID: 30368767.

52. Walzer K, Chi J-T. Trans-kingdom small RNA transfer during host-pathogen interactions: the case of P. falciparum and erythrocytes. RNA Biology. https://doi.org/10.1080/15476286.2017.1294307 PMID: 28277932

53. LaMonte G, Philip N, Reardon J, Lacsina JR, Majoros W, Chapman L, et al. Translocation of sickle cell erythrocyte microRNAs into Plasmodium falciparum inhibits parasite translation and contributes to malaria resistance. Cell Host Microbe. 2012; 12(2):187–99. Epub 2012/08/21. https://doi.org/10.1016/j.chom.2012.06.007 PMID: 22901539; PubMed Central PMCID: PMC3442262.

54. Regev-Rudzki N, Wilson Danny W, Carvalho Teresa G, Sisquella X, Coleman Bradley M, Rug M, et al. Cell-Cell Communication between Malaria-Infected Red Blood Cells via Exosome-like Vesicles. Cell. 2013; 153(5):1120–33. https://doi.org/10.1016/j.cell.2013.04.029 PMID: 23683579

55. Trager W, Jensen JB. Human malaria parasites in continuous culture. Science. 1976; 193(4254):673–5. https://doi.org/10.1126/science.781840 PMID: 781840.

56. Cranmer SL, Magowan C, Liang J, Coppel RL, Cooke BM. An alternative to serum for cultivation of Plasmodium falciparum in vitro. Trans R Soc Trop Med Hyg. 1997; 91(3):363–5. Epub 1997/05/01. https://doi.org/10.1016/s0035-9203(97)90110-3 PMID: 9231219.

57. Lambros C, Vanderberg JP. Synchronization of Plasmodium falciparum erythrocytic stages in culture. J Parasitol. 1979; 65(3):418–20. PMID: 383936.

58. Saliba KS, Jacobs-Lorena M. Production of Plasmodium falciparum gametocytes in vitro. Methods in molecular biology. 2013; 923:17–25. https://doi.org/10.1007/978-1-62703-026-7_2 PMID: 22990768.

59. http://www.bioinformatics.babraham.ac.uk/projects/trim_galore.

60. Martin M. CUTADAPT removes adapter sequences from high-throughput sequencing reads2011.

61. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29(1):15–21. Epub 2012/10/30. https://doi.org/10.1093/bioinformatics/bts635 PMID: 23104886; PubMed Central PMCID: PMC3530905.

62. http://www-huber.embl.de/users/anders/HTSeq/.

63. McCarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics. 2017; 33(8):1179–86. https://doi.org/10.1093/bioinformatics/btw777 PMID: 28088763; PubMed Central PMCID: PMC5408845.

64. Lun AT, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res. 2016; 5:2122. https://doi.org/10.12688/f1000research.9501.2 PMID: 27909575; PubMed Central PMCID: PMC5112579.