Stochastic Expression of Invasion Genes in *Plasmodium falciparum* Schizonts

Jaishree Tripathi (tjaishree@ntu.edu.sg)
Nanyang Technological University https://orcid.org/0000-0003-3461-4767
Lei Zhu
Nanyang Technological University
Sourav Nayak
Nanyang Technological University
Michal Stoklasa
Nanyang Technological University
Zbynek Bozdech
Nanynag Technological University https://orcid.org/0000-0002-9830-8446

Article

Keywords: Plasmodium falciparum, Malaria

Posted Date: February 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-152167/v1

Version of Record: A version of this preprint was published at Nature Communications on May 30th, 2022. See the published version at https://doi.org/10.1038/s41467-022-30605-z.

Abstract

Genetically identical cells are known to exhibit differential phenotypes in the same environmental conditions. These phenotypic variants are linked to transcriptional stochasticity and have been shown to contribute towards adaptive flexibility of a wide range of unicellular organisms. Here, we investigated transcriptional heterogeneity and stochastic gene expression in P. falciparum by performing single cell RNA sequencing on blood stage schizonts. Our data reveals significant transcriptional variations in the schizonts stage with a distinct group of highly variable invasion gene transcripts being identified. Moreover, our data reflected several diversification processes including putative developmental "checkpoint"; transcriptomically distinct parasite sub-populations and transcriptional switches in variable gene families (var, rifin, phist). Most of these features of transcriptional variability were preserved in isogenic parasite cell populations (albeit with a lesser amplitude) suggesting a role of epigenetic factors in cell-to-cell transcriptional variations in human malaria parasites. Lastly, we applied quantitative RT-PCR and RNA-FISH approach and confirmed stochastic expression of merozoites surface proteins encoding msp1, msp3, msp7, rhoptry protein encoding rhopH2 and erythrocyte binding antigen encoding eba181, some of which represent key candidates for invasion blocking vaccines.

Introduction

Malaria, an infectious disease caused by *Plasmodium* parasites, remains a threat to human health with 228 million cases and 405,000 deaths recorded in 2018 [1]. Humans can be infected by six different *Plasmodium* species, *P. vivax, P. falciparum, P. ovale, P. malariae, P. knowlesi* and *P. cynomolgi*, with falciparum malaria being the most virulent. During its complex life cycle, *Plasmodium* parasites differentiate into multiple developmental stages, switching between sexual and asexual forms in their primary (mosquito) and intermediate (human) host respectively. Throughout this journey, the parasite thrives in different host environments, selection pressures (such as, host immune response, exposure to antimalarials) and sudden perturbations. This requires remarkable phenotypic plasticity for a single cellular eukaryotic pathogen to survive. Previously, phenotypic plasticity has been demonstrated in *Plasmodium* parasites in relation to sexual conversion rates and ratio, asexual invasion rates and burst size [2]. Within-host environmental factors such as, anaemia, drug treatment, nutritional status, circadian rhythm and age of red blood cells (RBCs) available for invasion have been shown to modulate parasite replicative and reproductive efforts [3–8]. However, the molecular basis underpinning this phenotypic variability still remains uncharacterised.

Phenotypic diversity has been shown to be driven by epigenetic and transcriptional variability in other systems [9,10]. For example, studies on bacteria and yeast [11–13] has demonstrated that cells with multiple 'physiological states' can exist in a genetically identical population [14]. In the context of *Plasmodium* specifically, clonal variation in the expression of members of surface antigen gene families such as *var*, *rifin* and *stevor* is well-known [15–17]. Similarly, expression of solute transport channels, *clag3.1* and *clag3.2*, on the surface of parasitized RBCs is found to be mutually exclusive in isogenic clones [18,19].

Pfap2-g, the master regulator of sexual differentiation, is another gene known to be associated with heterochromatin protein 1 (HP1) and exhibits clonally variant expression [20,21]. These studies confirm the presence of non-genetic transcriptional heterogeneity in *Plasmodium* parasites and calls for a systematic transcriptome-wide investigation of gene expression variability.

One of the latest cutting edge methodologies to study cell-to-cell transcriptional variability and detect highly variable genes (HVGs) in a homogenous cellular population is single cell RNA sequencing (scRNAseq) [22–25]. In the context of malaria, published scRNAseq studies collectively describe transcriptional profiles associated with sexual commitment, development of various *P. falciparum* and *P. berghei* stages (i.e., liver schizonts, ookinetes, sporozoites, blood stages) as well as gene expression patterns of *P. vivax* isolates from non-human primate blood stream infections [26–32]. Undoubtedly, this has provided valuable insights into *Plasmodium* biology, however, up until now, investigations exploring stochastic gene expression within individual parasite forms (developmental stages) remain to be conducted. A key requirement for confident detection of transcriptional stochasticity is a reliable whole transcriptome amplification (WTA) method. Here we optimised the quasilinear multiple annealing and looping based amplification cycles (MALBAC) to suit the AT-rich transcriptome, lower RNA content in *P. falciparum* asexual stages and minimize cell-to-cell technical variation as compared to PCR-based WTA used in previous studies. Using this technique, we generated single cell transcriptomes of two cohorts of *P. falciparum* schizonts originating from non-isogenic and isogenic parasite populations, respectively. Here, we demonstrated several “layers” of transcriptional variations linked to differential life cycle progression, single cell transcriptional subpopulations (SCTS) and stochastic transcriptional variability of genes (HVG) involved in crucial cellular processes such as invasion (and others). We also characterize differential transcriptional patterns for key hypervariable gene families such as *var*, *rifin* and *phist* involved in immune evasion and host-parasite interactions [15,34–37]. Last but not the least, we provide experimental validation of cell-to-cell transcriptional variability of several previously unidentified HVGs, which supports the validity of the derived technique for confident characterizations of stochastic gene expression in malaria parasites.

Results

Multiple Annealing and Looping Based Amplification Cycles (MALBAC) for Single Cell Transcriptomes from P. falciparum Schizonts

Here, we wished to develop a WTA strategy for scRNAseq analysis of the human malaria parasite *P. falciparum* that is based on quasilinear amplification to achieve broad, uniform and quantitatively reproducible transcriptome coverage. The goal was to significantly improve detections of genuine cell-to-cell (stochastic) gene expression variations within otherwise morphologically uniform parasite populations, such as a single developmental stage. The rationale is to minimize extensive PCR cycling that could reduce the complexity of the final cDNA library by amplification bias [38–40] and is known to be a main limitation of the standard techniques applied in previous studies [26–30, 32]. With this rationale, we

tested several (linear/quasilinear) strategies including rolling circle amplification (RCA), multiple displacement amplification (MDA) and MALBAC [40,41]; and compared these to the PCR-based SMARTseq2 method [42]. Comparing 1-cell and 10-cell transcriptomes generated by the four WTA methods, we identify MALBAC as the most suitable technique for the intended study (**Supplementary Fig. 1 and** Fig. 1a). Out of the four approaches, only MALBAC and SMARTseq2 yielded > 50 % parasite-specific sequencing reads, whereas < 10% of the RCA- and MDA-generated reads mapped to the *P. falciparum* genome (**Supplementary Fig. 1**). The results generated by MALBAC, however, exhibited the highest correlations between, both, 1-cell and 10-cell transcriptome replicates, with R^2 coefficients of 0.369 and 0.646, compared to 0.098 and 0.148 for SMARTseq2, respectively (Fig. 1a). Finally the MALBAC results exhibited lower standard deviations (SD) in gene expression across replicates with lower SD dependency on the overall mean transcript abundance (Fig. 1a, **right panel**). This demonstrates that MALBAC captures 1-cell and 10-cell transcriptomes with high quantitative reproducibility across a wide range of transcript abundance, making it most suitable to analyse cell-to-cell gene expression variability.

Hence, we applied MALBAC to generate single cell transcriptomes of *P. falciparum* schizonts collected from non-isogenic and isogenic parasite populations. Briefly, using fluorescence-assisted cell sorting (FACS) we isolated 300 individual schizonts from a highly synchronized culture of the *P. falciparum* 3D7MR4 strain; representing a non-isogenic population. Subsequently, we generated an isogenic clonal culture by serial dilution of the 3D7MR4 parasites and collected 100 individual schizonts after fourteen generations of clonal expansion (presumably from a single founding cell). Overall, we detected 5488 transcripts with 2695 being detected in more than 10 % of the non-isogenic group (**Supplementary Fig. 2**). Similar transcriptome coverage was achieved with the isogenic group of parasites with 5260 transcripts detectable in total. The mean number of transcripts detected per schizont was ~ 1040 in both isogenic and non-isogenic cohorts as compared to ~ 3800 transcripts detected in 100-cell bulk control (n = 15) (Figs. 1b and 1c). This is consistent with the transcriptome generated from samples in which RNA obtained from 10 schizonts was diluted 10-fold to achieve an equivalent amount of total RNA of a single cell (*Pf*RNAdil). Here, we detected a mean of 1145 transcripts per PfRNAdil (n = 49) (Fig. 1b). This represents a significant improvement to the previous studies in which mean/median of genes detected per schizont ranged between 212 to 938 [26,27,29]. To assess the quality of the MALBAC-generated results, we correlated the average of all 1-cell and 100-cell transcriptomes achieving a R^2 value of 0.72, thus confirming that the gene expression signal amplified by MALBAC represents true biological signal detected in bulk samples (Fig. 1d). The R^2 values correlating individual 1-cell and 100-cell transcriptomes ranged from 0.02 to 0.33 (Fig. 1d, **top left inset**). Lastly, majority of the sequenced 1-cell transcriptomes were estimated to be between and around 40 to 44 hours post invasion (HPI) (see below) by correlation to a high resolution bulk 3D7 reference transcriptome [43] (Fig. 1e), thus confirming successful isolation and amplification of late asexual stages.

Schizont Subpopulations with Unique Transcriptional Profiles Identified in *P. falciparum* Parasites

First, we interrogated the generated dataset for the presence of parasite sub-populations that reflect transcriptional variability at a broader (supracellular) level. Applying a custom designed data quality filter (refer to *Methods*), we select 271 non-isogenic and 79 isogenic 1-cell transcriptomes and subjected these to data normalization and dimensionality reduction by ZINB-WaVE modelling as described [44]. As shown in Figs. 2a and 2b, the first two components of the multidimensional scaling (MDS) account for the maximum differences between 1-cell transcriptomes of both non-isogenic and isogenic parasites. This MDS-based stratification aligned well with the parasite age differentiation (described in Fig. 1e) with a fraction of cells falling into the 36–40 and > 44 HPI groups, respectively (Figs. 2a and 2b). Crucially, similar HPI categorization could be deciphered for the non-isogenic as well as the isogenic population with comparable distribution patterns. Subsequently, we applied resampling-based sequential ensemble clustering (RSEC) to evaluate a possible finer substructure within each dataset that would indicate the existence of distinct subpopulations. Interestingly, eight subpopulations were detected in the non-isogenic parasite group each defined by a specific transcriptional profile (termed as Single-Cell Transcriptional Subpopulations, SCTS) (Fig. 2a). The pseudotemporal ordering followed the estimated age progression of schizonts with SCTS 2,3 and 8 falling onto ~ 36 to 40 HPI and 1,4,5,6 and 7 onto ≥ 44 HPI groups. Interestingly, for the isogenic 1-cell population, identical RSEC clustering parameters did not detect any subpopulations (Fig. 2b). This suggests that while the HPI (age) separation reflects intrinsic properties of schizont maturation of any given schizont population, the SCTSs represent a deeper differentiation of genetically (or otherwise) more diverse parasite cohorts.

Next, we applied the loess regression method and identified 441 genes exhibiting differential expression along the pseudotemporal progression defining the eight identified SCTS (Fig. 2c **and Supplementary Table 1**). Pathway enrichment analyses using the Malaria Parasite Metabolic Pathways (MPM) database showed that each SCTS carries a unique set of biological functions/pathways represented by the differentially expressed genes (**Supplementary Table 2**) [45]. As shown in Fig. 2d, protein folding, COPI/COPII-mediated vesicular trafficking and protein degradation pathways generally showed higher average expression in the SCTS 1, 2 and 3. This included genes such as, coatomer subunits (α,β,γ), endoplasmic reticulum chaperone GRP170 and E3-ubiquitin ligase, etc. Interestingly, ubiquitin activation was recently shown to be essential for schizonts maturation [46]. Expression of pH regulation, nucleosome assembly and phosphatidyl inositol biosynthesis related pathways was also higher in SCTS 1, 2 and 3. This may be indicative of organisation of newly replicated DNA and active membrane biogenesis concomitant with the segregation of daughter parasites (merozoites) during mid-late stage schizogony [47]. On the other hand, expression of parasitophorous vacuole membrane associated proteins and PTEX translocon pathway involved in protein export, was found to be highest in SCTS 4 and 5 mainly consisting of late stage schizonts (i.e. > 44 HPI). Interestingly, pathways linked to invasion and exported proteins showed heterogeneous expression across all SCTS (Fig. 2d). For example, SCTS 1, 4 and 6 showed higher average expression of invasion ligands for erythrocyte receptors, merozoite surface proteins and motility, represented by genes such as, erythrocyte binding antigen (*eba*) 175, merozoite surface protein (*msp*) 5 and 6, and reticulocyte binding protein homologues. On the other hand, cGMP signalling involved in invasion and egress was uniformly highly expressed across all SCTS with only

subtle variations. Taken together, these results suggest the existence of multiple cellular 'physiological states' established within a single parasite population over extended periods of cell diversifications (here prolonged *in vitro* culturing). The SCTS may represent subtle but significant diversion of the schizont maturation process with a putative checkpoint at approximately 44 HPI, all of which may contribute to phenotypic plasticity of malaria parasite prior and during the invasion process.

To assess the relationship between the SCTS-driving transcriptional variation and genetic diversity, we identified single nucleotide polymorphisms (SNPs) from the RNA-seq data for each schizont. A total of 173 and 265 SNPs were detected with an allele frequency > 0.05 in the non-isogenic and isogenic group of parasites, respectively. Using these SNPs, two genetic clusters were observed for the non-isogenic parasites, while no clusters were found in the isogenic group (Figs. 2e and 2f). Interestingly, none of the genetic clusters for non-isogenic parasites showed enrichment of any SCTS, thus suggesting that the observed transcriptional heterogeneity is not driven by the (exonic) SNP profile of the parasites (Fig. 2e). However, we cannot exclude the possibility that other more subtle genetic variations in other regions (such as promoter/intergenic regions) that were beyond detection of the scRNAseq analysis may interline these transcriptional subpopulations.

Single Cell Transcription of Host-Parasite Interaction Factors

Variant expression of surface exported proteins encoded by gene families such as *var*, *rifin* and *phist* has long been of interest due to their immunogenic, host cell remodeling and cytoadhesive properties, essential for parasite survival during asexual blood stages [15,37,48–50]. As such, understanding cell-to-cell variation patterns of their individual members may shed more light into their biological relevance. Indeed, our results showed that these three families have fundamentally different transcriptional profiles suggesting their distinct roles in host-parasite interactions (Figs. 3a-3f). For the *var* gene family, which is believed to be expressed in a mutually exclusive manner, we identify 72 and 62 members out of 105 *var* transcripts (including pseudogenes) expressed at more than 10 FPKM (normalized sequencing reads expressed as fragments per kilobases per million), in the non-isogenic and isogenic parasite populations, respectively. In the non-isogenic schizonts, we found two members, PF3D7_937800 and PF3D7_0400200 detected at the highest frequency in 28.5 % and 27.8% of single cells respectively (Fig. 3a). In contrast, the isogenic parasites expressed one dominant *var* gene transcript, PF3D7_937800, at a frequency of 56.7 % followed by PF3D7_0400200 and PF3D7_0400100, at 27.8 % and 22.7 % respectively. This is consistent with the presumed mutual exclusivity suggesting that PF3D7_937800 was the dominant transcript in the "founding cell" of the isogenic population and is still dominant in more than half of the cells. The two other high frequency *var* genes (PF3D7_0400200 and PF3D7_0400100) likely represent early switches that expanded in the isogenic population within the 14 generations. Even though a single *var* transcript was detected in 17% and 28% of non-isogenic and isogenic cells respectively, a considerable proportion of cells expressed more than one (2–10) *var* transcripts (Fig. 3b). Nonetheless, most of these multiple *var* transcripts are expressed at low levels with a relatively tighter distribution for higher FPKM levels. This suggests that while many cells indeed express one dominant *var* transcript, a

significant proportion of parasite cells express additional members potentially as precursors for *var* gene switching that was estimated to occur in ~ 2% of each generation [51,52].

In contrast, the *rifin* gene family exhibited a much more restricted expression pattern both in the non-isogenic and isogenic populations. Most *rifin* expressing schizonts had only 1–2 *rifin* transcripts expressed at high levels. This indicates that in schizonts, *rifin*s are more tightly regulated compared to the *var* genes (Figs. 3c and 3d). First, 102 and 87 out of 185 *rifin* transcripts (including pseudogenes) were detected at > 10 FPKM in at least one of the non-isogenic and isogenic cells, respectively. Out of these PF3D7_1300600 was expressed at the highest frequency of ~ 12% in non-isogenic parasites. After the isogenic clone expansion, a new *rifin* gene, PF3D7_0223400, gained a slight dominance being expressed in ~ 14% cells while the original PF3D7_1300600 dropped in its frequency to ~ 9%. In both groups, the remaining *rifin* transcripts were expressed in less than 11 % of the population with only slight variations between the non-isogenic and isogenic cell populations. On the contrary, *phist* genes showed expression in relatively larger proportion of schizonts. Expression of PF3D7_1149200, PF3D7_0102200 and PF3D7_1401600 was most frequent in non-isogenic schizonts at ~ 45 %, ~ 38% and ~ 37% as compared to PF3D7_1401600 (52 %), PF3D7_0219800 (34%) and PF3D7_1252800 (28%) in isogenic parasites. Taken together, these data indicate that transcriptional patterns of the *rifin* and *phist* genes undergo only slight shifts upon clonal expansion thus distinguishing these from *var* genes whose mutually exclusive expression is believed to be a key aspect of immune evasion. Interestingly, only a handful of *rifin* genes appear to be expressed in an individual cell, which suggests their redundant biological roles. Conversely, simultaneous expression of a higher proportion of the *phist* gene family may indicate a functional diversification of the individual members most (if not all) of which are required for survival of an individual cell. The ability to change their expression profile after a clonal expansion of the parasite population implicate the *rifin* and *phist* gene family in phenotypic plasticity of the malaria parasites as well.

The functional analyses of the SCTS (previous section) indicated that genes involved in the merozoite invasion are particularly active in terms of their cell-to-cell variation at the schizont stage. To study this further, we quantified transcriptional variability of individual genes by assigning a variability index score (VIS) calculated as the SD of z-score per gene with respect to PfRNAdil. A striking difference was observed in the overall VIS distribution between non-isogenic and isogenic schizonts, with considerably lower indices observed for invasion genes within the isogenic group (Fig. 4a, **left)**. In particular, *msp1, msp3, msp7*, high molecular weight rhoptry protein 2 (*rhopH2*), rhoptry associated protein *rap1* and *eba*181 ranked high on the index (VIS = 9 to 16) in the non-isogenic group whereas much lower VIS were found for the same in isogenic schizonts. As shown in Fig. 4a, the invasion genes can be grouped into two distinct clusters based on their transcription that also outlines their functional significance/relatedness. Gene cluster 1 predominantly consists of merozoite surface proteins (e.g. *msp1*, *msp3*, *msp7*), rhoptry proteins (e.g. *ron2*, *ron4*, *ron5*, *ron12*, *rhopH2*) and glideosome-associated protein (e.g. *gap50*). These genes represent the core invasion factors facilitating the initial merozoite-erythocyte interactions and the process of invasion. In contrast, cluster 2 showed co-expression of reticulocyte

binding protein homologues (*rh1*, *rh2a*, *rh2b*, *rh4*, *rh5*), erythrocyte binding antigens (*eba140, eba175, eba181*) and apical membrane antigen *ama1*. These genes are believed to be crucial for selective surface antigens interactions facilitating merozoite recognition of a suitable erythrocyte for binding [53,54]. There is a strong association of the two clusters with the HPI projection (Fig. 4a **bottom inset**) suggesting their relevance to the merozoite maturation process. However, we also observed presence of HVGs within both clusters suggesting the role of the invasion genes in malaria parasite adaptability; presumably driving distinct invasion phenotypes (see below).

Stochastic Expression of Highly Variable Genes in *P. falciparum* Schizonts

Further inspection of our data suggested a presence of a small but distinct set of genes with fundamentally broader transcriptional differences indicating stochastic mRNA variations. To identify such HVGs, we applied the F-test analysis (padj < 0.005) to determine transcript variability differences between the 1-cell dataset and PfRNAdils [55]. This revealed two groups of HVGs in non-isogenic (88 genes) and isogenic (45 genes) schizonts with an overlap of 11 genes (Figs. 4b and 4c, **Supplementary Tables 3 and 4**). As expected, invasion genes previously found to have high VIS (such as *msp1, msp3, msp7*, *rap1*, *rhopH2* and *rhopH3*) were enriched amongst the HVGs (Fig. 4b). In addition to these, genes involved in diverse cellular functions, such as, exported protein 2 (*exp2*), endoplasmic reticulum-resident calcium-binding protein (*pferc*), histone 2B, actin I and folate transporter 1 were also highly variable between single non-isogenic schizonts (**Supplementary Table 3**). Interestingly, the HVG group in the isogenic schizonts included somewhat different set of functionalities such as erythrocyte vesicle protein 1, peroxiredoxin, calcium dependent protein kinase 1 (*cdpk1*), *rh2b*, vacuolar protein sorting-associated protein 29, etc. (**Supplementary Table 4**). Crucially, the 11 HVGs that overlapped between the non-isogenic and isogenic schizonts were enriched for invasion antigens including *msp1, msp3*, *rap1*, *rap3, rhopH2* and *rhopH3* (Fig. 4c). These were accompanied by other functions such as *exp2*, AP2 transcription factor, histone H2B, protein phosphatase (putative) and actin I. Interestingly, two of the HVGs, AP2 transcription factor (PF3D7_0420300), and, a putative long chain polyunsaturated fatty acid elongation enzyme (PF3D7_0605900) were also previously reported to have variable expression levels in bulk RNA sequencing of 3D7 schizonts [56]. Taken together, here we identified a small but significant group of genes with fundamentally higher levels of cell-to-cell transcriptional variations that is far beyond differential mRNA levels underlying the HPI projections, SCTS and variable antigenic families (see above). Enrichment of invasion genes in this category (along with other functionalities) suggests that maintaining stochastic expression of these genes is crucial (hence conserved) presumably to ensure successful productive invasion across variable host cell surface receptor repertoire/environments. Maintenance of HVGs even within the isogenic population (presumably devoid of fundamental genetic variations) suggests a role of epigenetic factors in these processes. Indeed, the presence of several *var* genes (10), *rifins* (2) and *surfin1.2* amongst the HVGs in non-isogenic schizonts only supports this concept given their association with a key epigenetic factor HP1, characterizing heterochromatin regions of the *P. falciparum* genome [57].

RNA-FISH and RT-PCR Validations of Highly Variable Gene Expression in Individual Schizonts

As stated in the beginning, here we wished to derive a scRNAseq based protocol that allows us to assess cell-to-cell variations of *P. falciparum* parasites of gene expression with high confidence. Using the derived MALBAC approach seemed to generate such data for schizont stages. However, to gain a stronger confidence in this derived method, we experimentally validated the detected stochastic expression for a subset of the HVGs using RNA fluorescence *in situ* hybridization (RNA-FISH) and quantitative reverse transcription PCR (qRT-PCR) experiments. Specifically, RNA-FISH was performed on individual schizonts by labeling these with *eba181* and *rh5* specific ViewRNA™ probe sets simultaneously (refer to *Methods*), and transcripts levels were quantified through confocal microscopy imaging. As shown in Fig. 4e, *eba181* transcript intensity is variable between individual schizonts in contrast to *rh5*, which is known to be essential for 3D7 invasion [58], with low variability hence not classified as HVG. Clearly, parasites with similar number of daughter nuclei (a proxy of developmental maturation) also exhibit highly variable expression of *eba181* transcript, while, *rh5* RNA-FISH signal remained stable across multiple schizont cells (Fig. 4e). Using a triple labeling protocol, RNA-FISH detected variable degree of transcriptional variability of *msp1*, followed by *msp3* and *msp7* measured simultaneously (Fig. 4f). In addition, we also confirmed HVG status of two other invasion genes, i.e., *gap45* and *rhop2*, along with other genes such as *exp2, Pferc* and *cept* (choline/ethanolaminephosphotransferase, putative) by qRT-PCR and/or RNA-FISH (Fig. 4d, **Supplementary Fig. 3**). Overall, these results not only provide experimental validation for the HVG status of these genes but also substantiate the MALBAC-based analyses of stochastic gene expression in *P. falciparum* schizonts presented in this study.

Discussion

Single cell transcriptomics is a fast evolving research tool bringing broad and deep understanding of essentially all studied biological systems [13,14,22, 59–63]. Without exception, the human malaria parasite has also been a subject of scRNAseq studies in the recent years [26, 28–31,33,64]. One of the main achievements was the assembly of the *Plasmodium* cell atlas delineating single cell transcriptomic pattern for essentially all developmental stages of the entire *P. berghei* life cycle, in general [29]. Here, the scRNASeq methodologies were well suited to compare developmental stages to each other and by that to refine differential transcriptional patterns, previously analysed by “bulk” approaches [65–68]. In this study we wished to complement the previous approaches by focussing on a single developmental stage, asexual blood stage schizont, to characterize stochastic gene expression as a potential mechanism for parasite phenotypic adaptation. This was done by developing MALBAC-based WTA that provides an alternative scRNAseq methodology that is more suitable for such studies. Although somewhat more expensive and labour intensive, MALBAC-scRNAseq allows capturing more transcripts with better quantitative measurements in single cell transcriptomes compared to the standard techniques such as SMARTseq2, Drop-seq or the 10X Chromium platform. This is achieved by minimizing potential amplification bias associated with PCR-based exponential WTA [39,69]. MALBAC provides improved transcriptome sensitivity, reproducibility and a reduction of amplification noise that is mediated by self-annealed amplicon loops in the preamplification step thus facilitating amplification only from the original cDNA strand and reducing amplification bias [40,70]. Moreover, the usage of random priming in MALBAC

reduces preferential amplification of shorter transcripts [41]. Our results demonstrate that MALBAC is suitable for *P. falciparum* single cell transcriptomics that could be used for stochastic gene expression studies of other developmental stages in the future.

An in-depth analysis of the generated single cell transcriptomes revealed multiple "layers" of transcriptional variability in the blood stage schizonts of *P. falciparum*. First, we found cells segregating into 36–40 HPI and > 44 HPI age groups along the pseudotemporal ordering in both non-isogenic and isogenic parasites. This is reminiscent of an intrinsic developmental checkpoint during the schizont development (~ 40 HPI) that is consistent with predicted transcriptional shifts during developmental progression, indicating rapid turning on or shutting down of broad transcriptional modules at once [28]. The existence of an intrinsic developmental checkpoint is also consistent with previous studies demonstrating that nuclear division during schizogony is asynchronous, where individual parasites follow autonomous nuclear division patterns [71,72]. In addition to temporal development, our dataset revealed 8 distinct parasite sub-populations or SCTS identified in non-isogenic schizonts, representing a finer level of transcriptional variation within the population. The detection of 8 STCS may be paralleled by 12 parasite transcriptional clusters detected upon exposure to lysophosphotidyl choline depletion which induced a general metabolic response in all 40 to 48 HPI cells and a sexual commitment-specific signature in a subset of cells [26]. Similarly, Poran *et al* also reported segregation of 30, 36 and 42 HPI asexual NF54 *P. falciparum* cells into 11 clusters [64]. The authors attributed this segregation to developmental progression of parasites by correlation to bulk-RNAseq time course. However, in our study the isogenic schizonts did not show any sub-structure in the population suggesting that cell-based transcriptional substructuring is mediated by cellular differentiation involving various genetic and epigenetic factors that are known to contribute towards transcriptional (and phenotypic) variation [13,73]. Better understanding of the mechanisms underlying cellular divergence will undoubtedly improve our knowledge of malaria parasite phenotypic plasticity, which could aid considerably in the development of future malaria intervention strategies involving all drugs, vaccines and diagnostics.

Identification of *msp1, msp3, msp7, rhopH2*, *rhopH3, rap1, rap3* and *eba181* amongst the HVGs in this study is consistent with previous studies of transcriptional variation in RBC invasion and merozoite ligands genes in clinical isolates [74,75]. For example, a study on Kenyan patients has shown that different clusters of invasion genes are expressed in different isolates [76] suggesting usage of diverse host erythrocyte receptors by epidemiologically relevant *P. falciparum* strains. Additionally, clonally variant expression of merozoite invasion ligands has been demonstrated before for *eba140, rh4* and *rhopH1* and is thought to be epigenetically regulated [18,77–79]. All these findings indicate that maintenance of variable levels/multiple combinations of invasion transcript repertoire in the population, could afford phenotypic plasticity in scenarios of RBC receptor heterogeneity and ligand-specific immune responses [75,80]. Results in this study suggest that the overall transcriptional variability of the invasion antigens occurs at the cellular level such that not all the invasion antigen are utilized by all cells within a given parasite population. Instead, different (sub)combinations of invasion factors may suffice for a single merozoite to invade a new host erythrocyte.

Here we wish to suggest that better understanding of stochasticity of merozoite invasion antigens will be crucial for future decision making in the development of an efficacious invasion-blocking vaccine, that up until now focused mostly on single candidate blood-stage vaccines such as MSP1, AMA1, Rh5, EBA-175, MSP2 or MSP3 [81–86]. Our results suggest that a successful strategy for an invasion vaccine might require targeting several invasion factors simultaneously, such as previously considered multi blood-stage antigen vaccines including the GMZ2 (consisting of conserved domains of GLURP and MSP3) and the MSP1/MSP2/RESA triple antigen vaccine [87,88]. As such, multi-target approaches can capture large proportions of invading merozoites helping to overcome additional challenges of invasion blocking vaccines including extensive polymorphism of the surface antigens and the short timespan for which merozoites are exposed to the immune system that requires a very high antibody titre for efficient inhibition [81,82,89]. To overcome these hurdles, recently other alternative strategies utilizing parasite genomic information and reverse vaccinology to discover new blood-stage vaccine candidates have been suggested [90–93]. Our data on HVGs and stochastic expression of several merozoite surface proteins transcripts can contribute to this by applying a 'transcriptomics-to-vaccines' approach. One of such crucial information generated in this study is represented by our discovery of stable cell-to-cell transcription of one of the key invasion antigen PfRH5; that contrast many other antigens (Fig. 4e). Interestingly PfRH5 is currently considered to have one of the highest potential as invasion vaccine target candidate, whose interaction with erythrocyte surface protein Basigin appears to be invariantly essential for invasion [58,94,95]. Our results strengthen this potential and provide future precedence for cross-referencing of single cell transcriptomics data for new malaria intervention strategies.

In addition to invasion genes, we also discovered variation in a wide range of other (non-invasion) genes, such as, *Pferc*, *cept*, aquaglyceroporin, AP2 transcription factor and thioredoxin reductase involved in a variety of functions such as, parasite egress [96], nutrient/protein channel, lipid metabolism, glycerol transport, transcription regulation and redox homeostasis respectively. This suggests the presence of unique 'transient' transcriptional states of individual parasites in a population, some of which may be suitable in defence to antagonists or coping to fluctuating environments; a strategy known as bet hedging. Similar to our finding, previous study on budding yeast has reported HVGs representing diverse cellular functions such as mitochondria and heat shock response, vitamin B1 metabolism, amino acid starvation response and membrane biology [24]. We propose that stochastic expression of diverse gene groups could be a potential mechanism underlying spontaneous adaptation of malaria parasites. It has been suggested previously that few *Plasmodium* specific phenomenon where bet-hedging could possibly play a role maybe during asexual replication, drug induced dormancy and recrudescence, sexual conversion rate, and successful invasion of red blood cells [2,6,97–99]. Future studies investigating transcriptional variability and stochastic gene expression in *P. falciparum* parasites under perturbed growth conditions, such as, exposure to drugs, heat shock, nutrient starvation etc. are required to reveal novel mechanisms of phenotypic adaptation in malaria parasites.

Methods

Experimental Model

P. falciparum 3D7 strain was maintained in RPMI 1640 medium (Gibco) supplemented with Albumax I (Gibco) (0.25%), hypoxanthine (Sigma) (0.1 mM), Sodium bicarbonate (Sigma) (2 g/L), and gentamicin (Gibco) (50 µg/L) in 2 % haematocrit at 37 ^{o}C on shaker. Cultures were gassed with malaria gas (5% CO_2, 3% O_2, and 92% N_2) after daily medium change. Freshly washed human RBCs were added to the culture every alternate day. Parasitemia and parasite morphology were assessed daily by microscopic examination of blood smears stained with Giemsa (1:10 dilution, Sigma). For synchronisation, ring stage culture were treated with 5% D-Sorbitol for 15 minutes with intermittent shaking. Treated culture were washed twice in culture media (2000 rpm, acc 9, brake 1) and resuspended in fresh medium.

Single Cell Sorting

A sorbitol synchronised *P. falciparum* 3D7 schizont culture (40 to 44 HPI) was used for performing single cell sorting. 20 µl of infected RBCs (iRBC) were washed in 1X PBS twice and stained with SYBR green dye (final dilution 0.2 x) for 30 minutes at 37 ^{o}C in the dark. Next, the iRBC pellet was washed > 5 times with 1X PBS before final resuspension in 1X PBS solution ready for FACS. Uninfected RBCs were also stained in parallel as a control for gating strategy. iRBCs were sorted on BD FACSAria™ into 4 µl of lysis buffer (5x FS buffer = 100 µl, 0.1M DTT = 25 µl, 10mM dNTP = 25 µl, 40U/ul RNaseOUT = 5 µl, GAT-dT Primer (100 µM) = 12.5 µl, H2O = 331.5 µl, Triton X114 = 1 µl) in 96 well plates, spinned briefly and processed immediately or stored at -80 ^{o}C for later.

MALBAC Whole Transcriptome Amplification

Single cell lysate obtained after FACS was subjected to reverse transcription and WTA using MALBAC protocol optimised for *P. falciparum* transcriptome. Briefly, cell lysate was exposed to 70 degrees for 90 seconds in a thermocycler. For reverse transcription SuperScript II, T4 gene32 and RNaseOut inhibitor were added to the lysate and thermocycled at 4°C, 2 min, 10°C, 2 min, 20°C, 2 min, 30°C, 2 min, 40°C, 2 min, 42°C, 50 min, 70°C, 15 min and finally maintained at 4°C, hold. Next, first strand cDNA was amplified by 35 cycles of MALBAC using *Pyrococcus* derived Deep Vent® DNA polymerase and 1 µM GAT-7N primers. This was followed by 19 cycles of PCR using 0.5 µM GAT-COM primers. Primer sequences for each step are: GAT-12dT. 5- GTG AGT GAT GGT TGA GGT AGT GTG GAG TTT TTT TTT TTT – 3 (used during reverse transcription); GAT-7N. 5- GTG AGT GAT GGT TGA GGT AGT GTG GAG NNN NNN N -3 (used during MALBAC pre-amplification); GAT-COM. 5- GTG AGT GAT GGT TGA GGT AGT GTG GAG – 3 (used during PCR). Amplified cDNA was purified using Agentcourt Ampure XP magnetic beads according to the manufacturer's instructions.

RNA sequencing

Purified MALBAC amplicon was used directly for preparing sequencing libraries using Illumina Nextera XT kit following manufacturer's instructions. Purified cDNA libraries were QC'ed on Agilent Bioanalyser 2100 using High-Sensitivity DNA chips. Libraries with optimal size distribution of 300 to 900 bp were pooled (20 samples per lane) and sequenced on Illumina HiSeq4000 platform generating 150 b x 150 b paired end reads with 110 Gb data output generated per lane. A minimum of 25 million reads were generated per library.

qRT-PCR

qRT-PCR was performed on purified MALBAC amplicon from individual schizonts for targeted genes using the Applied Biosystems™ SYBR™ Select Master Mix. Forward and reverse primers for each gene were designed using the NCBI Primer BLAST tool with annealing temperature set to 60°C and product size limited to 150 to 250 bp. All primers were ordered from Integrated DNA Technology at 100 µM stock concentration. A cDNA input of 5 ng was used per sample during qRT-PCR for normalised comparison of individual transcript level (represented by Ct values) variation between single cells.

RNA-FISH and Imaging

Customised RNA-FISH probes for individual genes were ordered from Thermo Fisher Scientific. Prior to performing RNA-FISH, thin smears of *P. falciparum* 3D7 schizont culture (38 to 42 hpi) was made on polylysine coated glass slides. Upon air drying, smears were fixed using 4% PFA and 0.008% Glutaraldehyde solution for 45 minutes. The rest of the steps of RNA-FISH labelling was performed using the Invitrogen™ ViewRNA™ ISH Cell assay kit following manufacturer's instructions. Stained blood smears were imaged using confocal microscope LSM710 at 500X (optical zooming) and image analysis was performed on ZENlite software.

Subcloning

P. falciparum 3D7 parasites were subcloned using serial dilution cloning. Culture was diluted in uninfected RBCs suspended at 2% haematocrit in culture media until 0.1 to 0.5 parasites were obtained per well theoretically. Media was changed on day 4, day 7, day 10 and every other day thereafter. 1 to 2 µl of fresh RBC were added to each well from day 7 onwards on every third day. Positive wells were screened from day 14 onwards by Hoechst staining (8mM) on BD FACSAria™.

Data Analysis

Raw reads obtained from the sequencer were checked for overall quality and trimmed to remove low quality bases from 3' ends, adapters, and amplification primers using Trim Galore. HISAT2 aligner was used to perform alignment to the *P. falciparum* genome. Only paired reads with proper orientation mapped to unique location of the genome were considered for counting. Gene specific read counts were calculated using BEDTools. Fragments per kilo base per million mapped reads (FPKM) were then calculated and used for further analysis.

Single-cells clustering

To investigate the transcriptomic heterogeneity of the parent parasite population, we applied clustering analysis to the 295 single cells. Before clustering, we filtered out those samples having a significantly lower (3 time mad lower than median) mapping rate or total read counts; also samples were discarded if they presented < 75% of the 269 highly appeared genes which were detected in 95% of dilution samples. Finally, there were 271 samples used for further study. And here, we considered a total of 4934 genes which were detectable in at least 5 cells excluding mitochondrial genes. The 271 single-cells were clustered using the method of resampling-based sequential ensemble clustering (RSEC) implemented in R with the package ClusterExperiment. The low-dimensional matrix used for clustering was specifically computed using zinbwave function of the R package zinbwave which K = 50 was asked for the dimensionality reduction. In practice, we applied k-means to generate a collection of candidate clustering using k0s = 4:5; obtained the consensus clustering based on the setting consensusProportion = 0.7, consensusMinSize = 5 and minSizes = 1; sequentially merged similar clusters based on the setting alphas = c(0.2), betas = 0.75, clusterFunction = "hierarchical01" and seqArgs = list(remain.n = 10, top.can = 5). And the stability of clusters were controlled by 500 times subsampling with 70% samples used each time as the parameters were set as subsampleArgs = list(resamp.num = 500, samp.p = 0.7, clusterFunction="kmeans", clusterArgs = list(nstart = 5)). Then, we asked the RSEC to merge the resulted clusters which had less than 15% differential expression using mergeMethod = "adjP" and mergeCutoff = 0.15 (**Supplementary Fig. 4a**). Lower mergeCutoff at 0.05 only generated smaller clusters within two of the clusters from mergeCutoff = 0.15 setting (**Supplementary Fig. 4b**). At last, we obtained 8 major clusters with 196 of 271 samples. We assigned the left 75 samples, which form clusters with less than 5 cells under the required conditions, to their nearest cluster using the assignUnassigned function of ClusterExperiment.

We repeated the same procedure of clustering analysis for clone cells. Out of 98 single-cells from a clone, 79 passed the filtering criteria and their transcriptome were used for the cluster detection by RSEC algorithm. Given the smaller number of single-cells studied here, we reset the parameters of consensusMinSize = 5 and seqArgs = list(remain.n = 10, top.can = 3). At last, a single cluster was obtained at the same criteria above that > 10% differential expression were required to create two distinct clusters.

We also lowered down the mergeCutoff to 0.05 which reported 5 clusters as shown in **Supplementary Fig. 5** with maximum 8% differential expression between clusters.

Declarations

DATA AND CODE AVAILABILITY

The RNA sequencing dataset generated during this study are pending approval for submission on GEO database (due to >1 TB data volume).

The code generated during this study have not been deposited in a public repository because most tools used for analysis are publicly available. The quality trimming code is available from the corresponding author upon request.

ACKNOWLEDGEMENTS

We thank Dr. Abdul Rashid Bin Mohammad Muzaki and Dr. Lai Soak Kuan for technical support during FACS and confocal microscopy at School of Biological Sciences, Nanyang Technological University, Singapore.

AUTHOR CONTRIBUTIONS

J.T. and Z.B. conceptualized the study. J.T. designed and performed all experiments, analysed and interpreted the data and wrote the manuscript. L.Z. performed pseudotime and RSEC data analysis. S.N. performed sequencing data processing. M.S. performed qRT-PCR experiments. Z.B. acquired funding, supervised the study and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

References

1. World Health Organization. *World malaria report 2019*. *WHO Regional Office for Africa* https://www.who.int/news-room/fact-sheets/detail/malaria (2019).
2. Mideo, N. & Reece, S. E. Plasticity in parasite phenotypes: evolutionary and ecological implications for disease. *Future Microbiol.* **7**, 17–24 (2012).
3. Carter, L. M. *et al.* Stress and sex in malaria parasites. *Evol. Med. Public Heal.* 135–147 (2013) doi:10.1093/emph/eot011.
4. Mancio-Silva, L. *et al.* Nutrient sensing modulates malaria parasite virulence. *Nature* **547**, 213–216 (2017).

5. Birget, P. L. G., Repton, C., O'Donnell, A. J., Schneider, P. & Reece, S. E. Phenotypic plasticity in reproductive effort: Malaria parasites respond to resource availability. *Proc. R. Soc. B Biol. Sci.* **284**, (2017).
6. Reece, S. E., Ramiro, R. S. & Nussey, D. H. Plastic parasites : sophisticated strategies for survival and reproduction ? *Evol. Appl.* **2**, 11–23 (2009).
7. Subudhi, A. K. *et al.* Malaria parasites regulate intra-erythrocytic development duration via serpentine receptor 10 to coordinate with host rhythms. *Nat. Commun.* **11**, 1–15 (2020).
8. Schneider, P. *et al.* Adaptive plasticity in the gametocyte conversion rate of malaria parasites. *PLoS Pathog.* **14**, 1–21 (2018).
9. Shaffer, S. M. *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).
10. Ecker, S., Pancaldi, V., Valencia, A., Beck, S. & Paul, D. S. Epigenetic and Transcriptional Variability Shape Phenotypic Plasticity. *Biossays* **40**, (2018).
11. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial Persistence as a Phenotypic Switch. **305**, 1622–1626 (2004).
12. Rotem, E., Loinger, A., Ronin, I., Levin-reisman, I. & Gabay, C. Regulation of phenotypic variability by a threshold- based mechanism underlies bacterial persistence. **107**, (2010).
13. Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: From theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).
14. Gasch, A. P. *et al.* Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. *PLoS Biol.* **15**, 1–28 (2017).
15. Scherf, A., Lopez-Rubio, J. J. & Riviere, L. Antigenic variation in Plasmodium falciparum. *Annu. Rev. Microbiol.* **62**, 445–470 (2008).
16. Kyes, S. A., Rowe, J. A., Kriek, N. & Newbold, C. I. Rifins: A second family of clonally variant proteins expressed on the surface of red cells infected with Plasmodium falciparum. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9333–9338 (1999).
17. Lavazec, C., Sanyal, S. & Templeton, T. J. Expression switching in the stevor and Pfmc-2TM superfamilies in Plasmodium falciparum. *Mol. Microbiol.* **64**, 1621–1634 (2007).
18. Voss, T. S., Bozdech, Z. & Bartfai, R. Epigenetic memory takes center stage in the survival strategy of malaria parasites. *Curr. Opin. Microbiol.* **20**, 88–95 (2014).
19. Cortés, A. *et al.* Epigenetic silencing of Plasmodium falciparum genes linked to erythrocyte invasion. *PLoS Pathog.* **3**, 1023–1035 (2007).
20. Brancucci, N. M. B. *et al.* Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host Microbe* **16**, 165–176 (2014).
21. Kafsack, B. F. C. *et al.* A transcriptional switch underlies commitment to sexual development in human malaria parasites. *Nature* **507**, 248–252 (2014).

22. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The Technology and Biology of Single-Cell RNA Sequencing. *Mol. Cell* **58**, 610–620 (2015).
23. Osorio, D. *et al.* Single-Cell Expression Variability Implies Cell Function. *Cells* **9**, (2020).
24. Saint, M. *et al.* Single-cell imaging and RNA sequencing reveal patterns of gene expression heterogeneity during fission yeast growth and adaptation. *Nat. Microbiol.* **4**, 480–491 (2019).
25. Mantsoki, A., Devailly, G. & Joshi, A. Gene expression variability in mammalian embryonic stem cells using single cell RNA-seq data. *Comput. Biol. Chem.* **63**, 52–61 (2016).
26. Brancucci, N. M. B. *et al.* Probing Plasmodium falciparum sexual commitment at the single-cell level. *Wellcome Open Res.* **3**, (2018).
27. Poran, A. *et al.* Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature* **551**, 95–99 (2017).
28. Reid, A. J. *et al.* Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. *Elife* **7**, 1–29 (2018).
29. Howick, V. M. *et al.* The malaria cell atlas: Single parasite transcriptomes across the complete Plasmodium life cycle. *Science (80-.).* **365**, (2019).
30. Sà, J. M., Cannon, M. V., Caleon, R. L., Wellems, T. E. & Serre, D. Single-cell transcription analysis of Plasmodium vivax blood-stage parasites identifies stage- And species-specific profiles of expression. *PLoS Biol.* **18**, 1–27 (2020).
31. Ngara, M., Palmkvist, M., Sagasser, S., Hjelmqvist, D. & Björklund, Å. K. Exploring parasite heterogeneity using single-cell RNA-seq reveals a gene signature among sexual stage Plasmodium falciparum parasites. *Exp. Cell Res.* **371**, 130–138 (2018).
32. Walzer, K. A., Fradin, H., Emerson, L. Y., Corcoran, D. L. & Chi, J. T. Latent transcriptional variations of individual Plasmodium falciparum uncovered by single-cell RNA-seq and fluorescence imaging. *PLoS Genet.* **15**, 1–26 (2019).
33. Walzer, K. A., Fradin, H., Emerson, L. Y., Corcoran, D. L. & Chi, J.-T. Latent transcriptional variations of individual Plasmodium falciparum uncovered by single- cell RNA-seq and fluorescence imaging. *Plos Genet.* **15**, 1–26 (2019).
34. Turner, L. *et al.* Severe malaria is associated with parasite binding to endothelial protein C receptor. *Nature* **498**, 502–505 (2013).
35. Mwakalinga, S. B. *et al.* Expression of a type B RIFIN in Plasmodium falciparum merozoites and gametes. 1–12 (2012).
36. Niang, M. *et al.* STEVOR is a Plasmodium falciparum erythrocyte binding protein that mediates merozoite invasion and rosetting. *Cell Host Microbe* **16**, 81–93 (2014).
37. Warncke, J. D., Vakonakis, I. & Beck, H. Plasmodium Helical Interspersed Subtelomeric (PHIST) Proteins , at the Center of Host Cell Remodeling. **80**, 905–927 (2016).
38. Polz, M. F. & Cavanaugh, C. M. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **64**, 3724–3730 (1998).

39. Ahmed, I., Tucci, F. A., Aflalo, A., Smith, K. G. C. & Bashford-Rogers, R. J. M. Ultrasensitive amplicon barcoding for next-generation sequencing facilitating sequence error and amplification-bias correction. *Sci. Rep.* **10**, 1–10 (2020).
40. Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-Wide Detection of Single Nucleotide and Copy Number Variations of a Single Human Cell. *Science (80-.).* **338**, 1622–1626 (2012).
41. Chapman, A. R. *et al.* Single cell transcriptome amplification with MALBAC. *PLoS One* **10**, 1–12 (2015).
42. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
43. Kucharski, M. *et al.* A comprehensive RNA handling and transcriptomics guide for high-throughput processing of Plasmodium blood-stage samples. *Malar. J.* **19**, 1–16 (2020).
44. Risso, D., Perraudeau, F., Gribkova, S., Dudoit, S. & Vert, J. ZINB-WaVE : A general and flexible method for signal extraction from single-cell RNA-seq data. (2017).
45. Ginsburg, H. Progress in in silico functional genomics: the malaria Metabolic Pathways database. *Trends Parasitol.* **22**, 238–240 (2006).
46. Green, J. L. *et al.* Ubiquitin activation is essential for schizont maturation in Plasmodium falciparum blood-stage development. *PLoS Pathog.* **16**, 1–29 (2020).
47. Matthews, H., Duffy, C. W. & Merrick, C. J. Checks and balances? DNA replication and the cell cycle in Plasmodium. *Parasites and Vectors* **11**, 1–13 (2018).
48. Borst, P., Bitter, W., McCulloch, R., Leeuwen, F. Van & Rudenko, G. Antigenic Variation in Malaria. *Cell* **82**, 1–4 (1995).
49. Su, X. *et al.* The Large Diverse Gene Family var Encodes Proteins Involved in Cytoadherence and Antigenic Variation of Plasmodium falciparum-Infected Erythrocytes. *Cell* **82**, 89–100 (1995).
50. Goel, S. *et al.* RIFINs are adhesins implicated in severe Plasmodium falciparum malaria. *Nat. Med.* **21**, 314–317 (2015).
51. Roberts, D. J. *et al.* Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature* **357**, 689–692 (1992).
52. Horrocks, P., Pinches, R., Christodoulou, Z., Kyes, S. A. & Newbold, C. I. Variable var transition rates underlie antigenic variation in malaria. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11129–11134 (2004).
53. Beeson, J. G. *et al.* Merozoite surface proteins in red blood cell invasion, immunity and vaccines against malaria. *FEMS Microbiol. Rev.* **40**, 343–372 (2016).
54. Baum, J. *et al.* A conserved molecular motor drives cell invasion and gliding motility across malaria life cycle stages and other apicomplexan parasites. *J. Biol. Chem.* **281**, 5197–5208 (2006).
55. Ouyang, W., An, Q., Zhao, J. & Qin, H. Integrating mean and variance heterogeneities to identify differentially expressed genes. *BMC Bioinformatics* **17**, 1–14 (2016).
56. Tarr, S. J. *et al.* Schizont transcriptome variation among clinical isolates and laboratory-adapted clones of the malaria parasite Plasmodium falciparum. *BMC Genomics* **19**, 1–13 (2018).

57. Filarsky, M. *et al.* GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. *Science (80-.).* **359**, 1259–1263 (2018).
58. Baum, J. *et al.* Reticulocyte-binding protein homologue 5 - An essential adhesin involved in invasion of human erythrocytes by Plasmodium falciparum. *Int. J. Parasitol.* **39**, 371–380 (2009).
59. Raj, A. & Oudenaarden, A. Van. Nature , Nurture , or Chance : Stochastic Gene Expression and Its Consequences. *Cell* **135**, (2008).
60. Weinberger, L. S., Burnett, J. C., Toettcher, J. E., Arkin, A. P. & Schaffer, D. V. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* **122**, 169–182 (2005).
61. Freddolino, P. L., Yang, J., Momen-Roknabadi, A. & Tavazoie, S. Stochastic tuning of gene expression enables cellular adaptation in the absence of pre-existing regulatory circuitry. *Elife* **7**, 1–34 (2018).
62. Song, Q., Ando, A., Jiang, N., Ikeda, Y. & Chen, Z. J. Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in Arabidopsis female gametophytes. *Genome Biol.* **21**, 1–18 (2020).
63. Bost, P. *et al.* Host-Viral Infection Maps Reveal Signatures of Severe COVID-19 Patients. *Cell* **181**, 1475–1488 (2020).
64. Poran, A. *et al.* Single-cell RNA-seq reveals a signature of sexual commitment in malaria parasites. *Nature* **551**, 95–99 (2017).
65. Caldelari, R. *et al.* Transcriptome analysis of Plasmodium berghei during exo-erythrocytic development. *Malar. J.* 1–20 (2019) doi:10.1186/s12936-019-2968-7.
66. Bozdech, Z. *et al.* Expression profiling of the schizont and trophozoite stages of Plasmodium falciparum with a long-oligonucleotide microarray. *Genome Biol.* **4**, (2003).
67. Roch, K. G. Le *et al.* Discovery of Gene Function by Expression Profiling of the Malaria Parasite Life Cycle. *Science (80-.).* **301**, 1503–1509 (2003).
68. Otto, T. D. *et al.* New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. *Mol. Microbiol.* **76**, 12–24 (2010).
69. Krehenwinkel, H., Wolf, M., Lim, J. Y., Simison, W. B. & Gillespie, R. G. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Sci. Rep.* **7**, 1–12 (2017).
70. Chen, M. *et al.* Comparison of Multiple Displacement Amplification (MDA) and Multiple Annealing and Looping-Based Amplification Cycles (MALBAC) in Single-Cell Sequencing. *PLoS One* **9**, 1–12 (2014).
71. Reininger, L., Wilkes, J. M., Bourgade, H., Miranda-saavedra, D. & Doerig, C. An essential Aurora-related kinase transiently associates with spindle pole bodies during Plasmodium falciparum erythrocytic schizogony. *Mol. Microbiol.* **79**, 205–221 (2011).
72. Rudlaff, R. M., Kraemer, S., Marshman, J. & Dvorin, J. D. Three-dimensional ultrastructure of Plasmodium falciparum throughout cytokinesis. *PLoS Pathog.* **16**, 1–21 (2020).

73. Huang, S. Non-genetic heterogeneity of cells in development: More than just noise. *Development* **136**, 3853–3862 (2009).
74. Gomez-Escobar, N. *et al.* Erythrocyte Invasion and Merozoite Ligand Gene Expression in Severe and Mild Plasmodium falciparum Malaria Erythrocyte Invasion and Merozoite Ligand Gene Expression in Severe and Mild Plasmodium falciparum Malaria. *J. Infect. Dis.* **201**, 444–452 (2010).
75. Bowyer, P. W. *et al.* Variation in Plasmodium falciparum erythrocyte invasion phenotypes and merozoite ligand gene expression across different populations in areas of malaria endemicity. *Infect. Immun.* **83**, 2575–2582 (2015).
76. Nery, S. *et al.* Expression of Plasmodium falciparum genes involved in erythrocyte invasion varies among isolates cultured directly from patients. *Mol. Biochem. Parasitol.* **149**, 208–215 (2006).
77. Jiang, L. *et al.* Epigenetic control of the variable expression of a Plasmodium falciparum receptor protein for erythrocyte invasion. *Proc. Natl. Acad. Sci.* **107**, 2224–2229 (2009).
78. Coleman, B. I. *et al.* Article Nuclear Repositioning Precedes Promoter Accessibility and Is Linked to the Switching Frequency of a Plasmodium falciparum Invasion Gene. *Cell Host Microbe* **12**, 739–750 (2012).
79. Crowley, V. M., Rovira-graells, N., Pouplana, L. R. De & Cortés, A. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant Plasmodium falciparum genes linked to erythrocyte invasion. *Mol. Microbiol.* **80**, 391–406 (2011).
80. Nyarko, P. B. *et al.* Investigating a Plasmodium falciparum erythrocyte invasion phenotype switch at the whole transcriptome level. *Sci. Rep.* **10**, 1–10 (2020).
81. Duffy, P. E. & Gorres, J. P. Malaria vaccines since 2000 : progress , priorities , products. *npj Vaccines* **5**, 1–9 (2020).
82. Hill, A. V. S. Vaccines against malaria. 2806–2814 (2011) doi:10.1098/rstb.2011.0091.
83. Goodman, A. L. & Draper, S. J. Blood-stage malaria vaccines — recent progress and future challenges. *Ann. Trop. Med. Parasitol.* **104**, 189–211 (2013).
84. Coelho, C. H., Yai, J., Doritchamou, A., Zaidi, I. & Duffy, P. E. Advances in malaria vaccine development : report from the 2017 malaria vaccine symposium. *npj Vaccines* **34**, 1–5 (2017).
85. Douglas, A. D. *et al.* The Blood-Stage Malaria Antigen PfRH5 is Susceptible to Vaccine-Inducible Cross-Strain Neutralizing Antibody. *Nat. Commun.* **2**, (2012).
86. Spring, M. D. *et al.* Phase 1 / 2a Study of the Malaria Vaccine Candidate Apical Membrane Antigen-1 (AMA-1) Administered in Adjuvant System AS01B or AS02A. *PLoS One* **4**, e5254 (2009).
87. Sirima, S. B. *et al.* A phase 2b randomized , controlled trial of the efficacy of the GMZ2 malaria vaccine in African children. *Vaccine* **34**, 4536–4542 (2016).
88. Genton, B. *et al.* A Recombinant Blood-Stage Malaria Vaccine Reduces Plasmodium falciparum Density and Exerts Selective Pressure on Parasite Populations in a Phase 1 – 2b Trial in Papua New Guinea. *J. Infect. Dis.* **185**, 820–827 (2002).

89. Wang, R., Smith, J. D. & Kappe, S. H. I. Advances and challenges in malaria vaccine development. *Expert Rev. Mol. Med.* **11**, 1–26 (2009).
90. Rappuoli, R. Reverse vaccinology. *Curr. Opin. Microbiol.* **3**, 445–450 (2000).
91. Frimpong, A., Kusi, K. A., Ofori, M. F. & Ndifon, W. Novel Strategies for Malaria Vaccine Design. *Front. Immunol.* **9**, 1–14 (2018).
92. Doolan, D. L. *et al.* Identification of Plasmodium falciparum antigens by antigenic analysis of genomic and proteomic data. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9952–9957 (2003).
93. Doolan, D. L. *et al.* Utilization of genomic sequence information to develop malaria vaccines. *J. Exp. Biol.* **206**, 3789–3802 (2003).
94. Crosnier, C. *et al.* Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum. *Nature* **480**, 534–537 (2011).
95. Bustamante, L. Y. *et al.* A full-length recombinant Plasmodium falciparum PfRH5 protein induces inhibitory antibodies that are effective across common PfRH5 genetic variants. *Vaccine* **31**, 373–379 (2013).
96. Fierro, M. A. *et al.* An endoplasmic reticulum CREC family protein regulates the egress proteolytic cascade in malaria parasites. *MBio* **11**, 1–21 (2020).
97. Birget, P. L. G., Prior, K. F., Savill, N. J., Steer, L. & Reece, S. E. Plasticity and genetic variation in traits underpinning asexual replication of the rodent malaria parasite , Plasmodium chabaudi. *Malar. J.* **18**, 1–11 (2019).
98. Schneider, P., Chan, B. H. K., Reece, S. E. & Read, A. F. Does the drug sensitivity of malaria parasites depend on their virulence? *Malar. J.* **7**, 1–11 (2008).
99. Cohen, N. R., Lobritz, M. A. & Collins, J. J. Microbial persistence and the road to drug resistance. *Cell Host Microbe* **13**, 632–642 (2013).

Figures

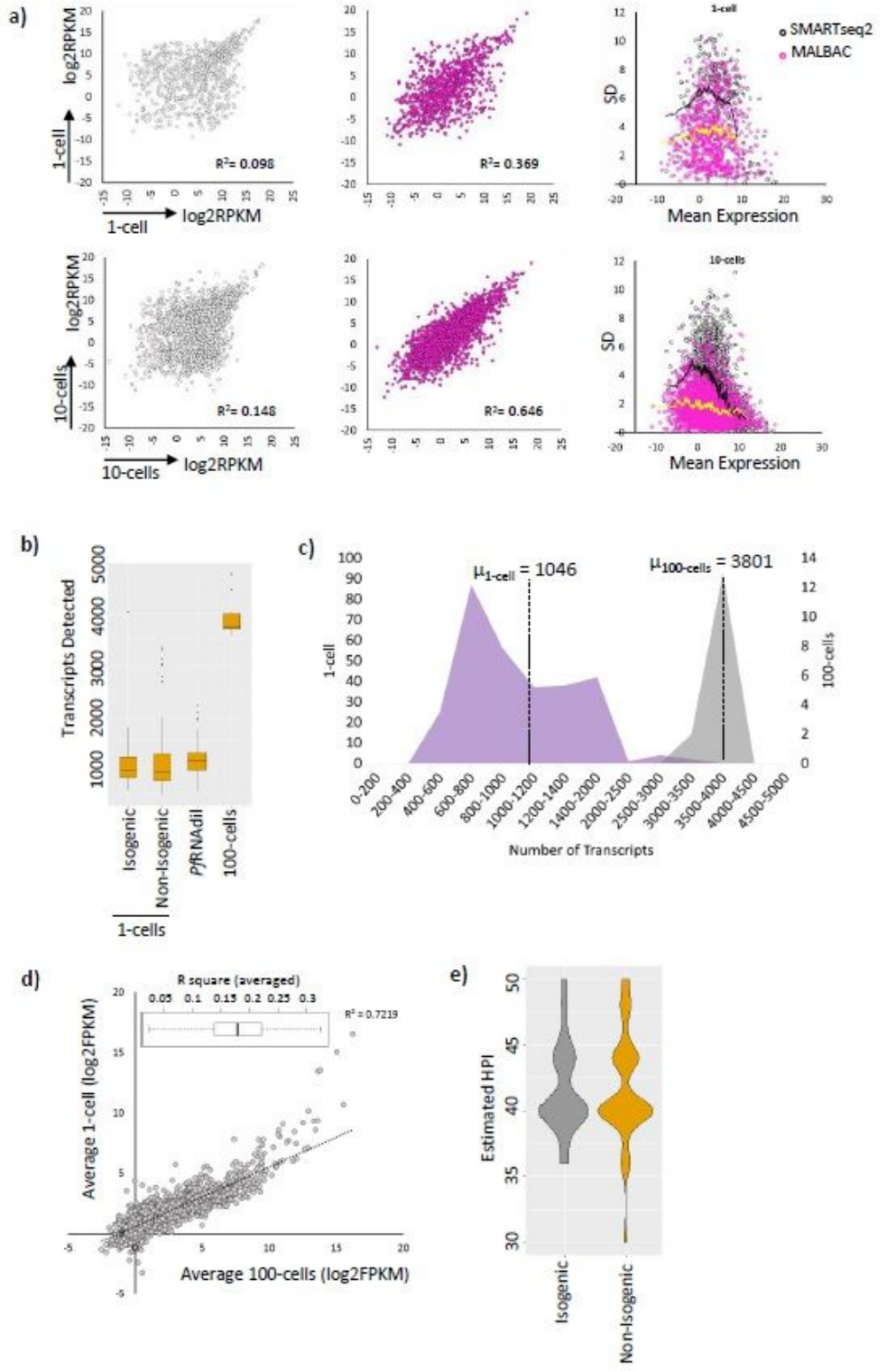


Figure 1

Highly Reproducible MALBAC-based Amplification of Single Cell Transcriptomes from P. falciparum Schizonts. a) Scatter plots showing correlations between 1-cell and 10-cell transcriptomes amplified by SMARTseq2 or MALBAC. Transcript expression is shown as log2RPKM values on the x and y-axis. The right most panel depicts relationship between standard deviation (SD) and mean transcript expression (log2RPKM) across all three 1-cell and 10-cell replicates for the two techniques. b) A boxplot showing

total transcripts detected in 1-schizonts (isogenic and non-isogenic), PfRNAdil and 100-cell bulk samples. c) A frequency distribution plot of number of transcripts detected in 1-schizonts and 100-schizonts samples. d) Correlation between averaged 1-cell and 100-cell transcriptomes. The range of R2 values for individual 1-cell and 100-cell transcriptome correlations is shown as a boxplot (inset). e) A violin plot depicting estimated age distribution of non-isogenic and isogenic schizonts.

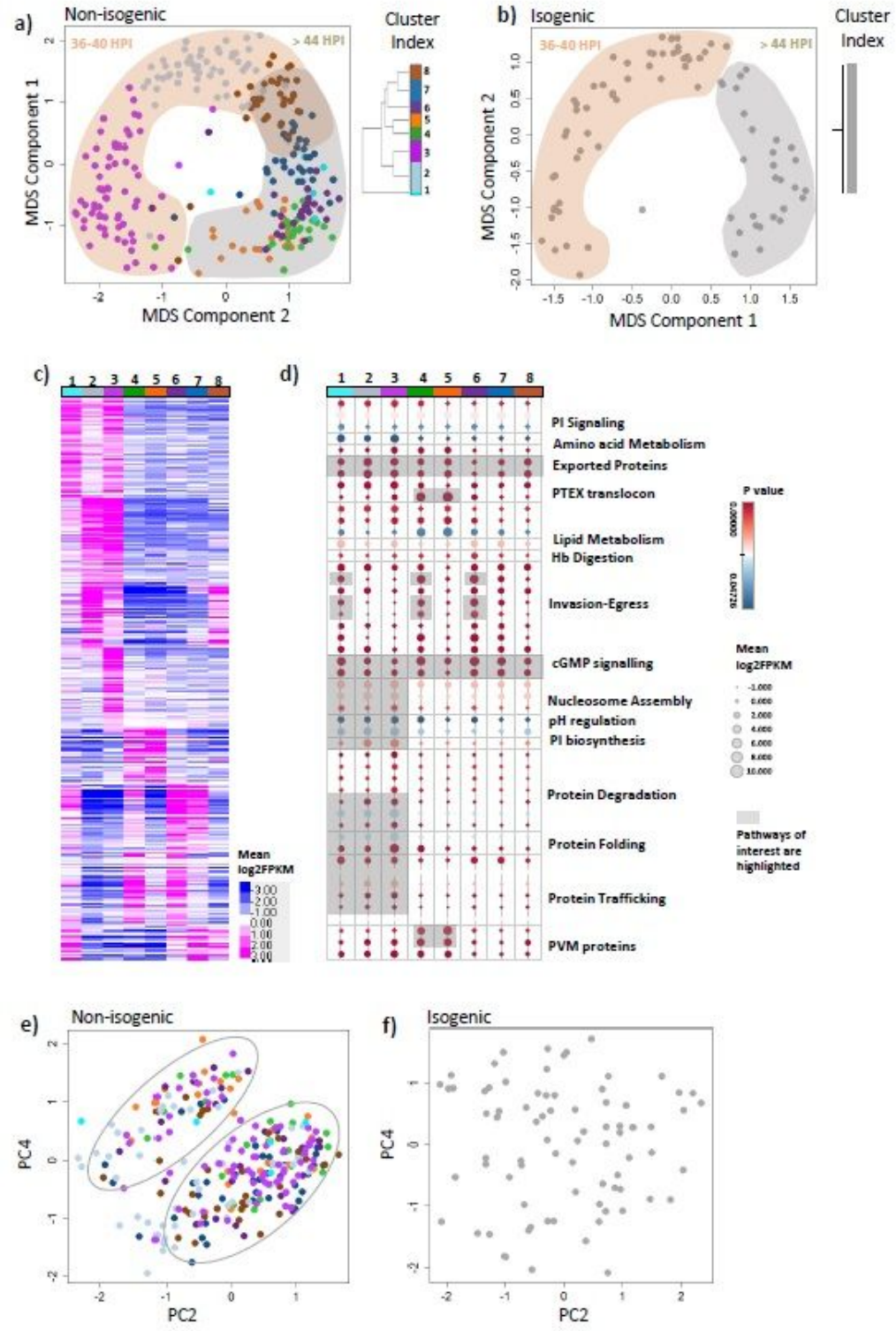


Figure 2

Schizont Subpopulations with Unique Transcriptional Profiles Identified in Non-Isogenic P. falciparum Parasites. A multidimensional scaling (MDS) plot showing first two components (MDS1 and MDS2) depicting maximum differences between individual schizonts in a) non-isogenic and b) isogenic group. RSEC-based SCTS hierarchy is shown in right hand inset (for both (a) and (b)) with each SCTS represented by a different colour for non-isogenic schizonts. The two major age groups (36-40 HPI and >44 HPI) identified has been overlaid on the MDS plot. c) A heatmap showing average expression of differentially expressed (DE) genes (mean normalised log2FPKM) in each SCTS of non-isogenic parasites. d) A dot plot of various functional pathways enriched in non-isogenic parasite SCTS. The colour and size of each circle depicts p-value and mean log2FPKM value respectively with pathways of interest highlighted in grey. e) , f) A PCA plot of RNA SNPs (MAF >0.05) shown for both non-isogenic and isogenic group respectively. Roughly two genetic clusters were found for non-isogenic group (marked as grey dotted ovals in (e)). The colour of each circle corresponds to the parasite SCTS (described in a and b above) that individual cell belongs to.

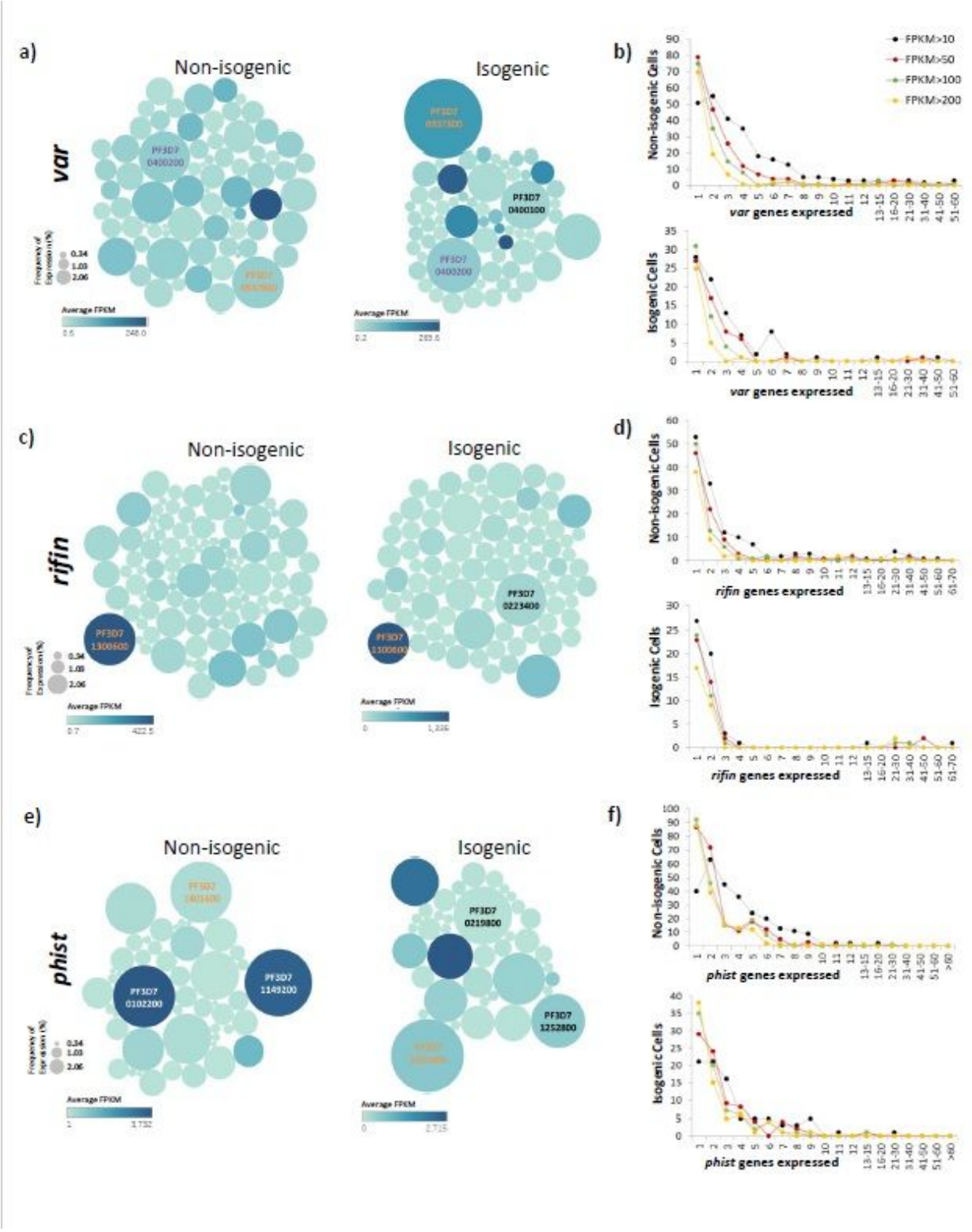
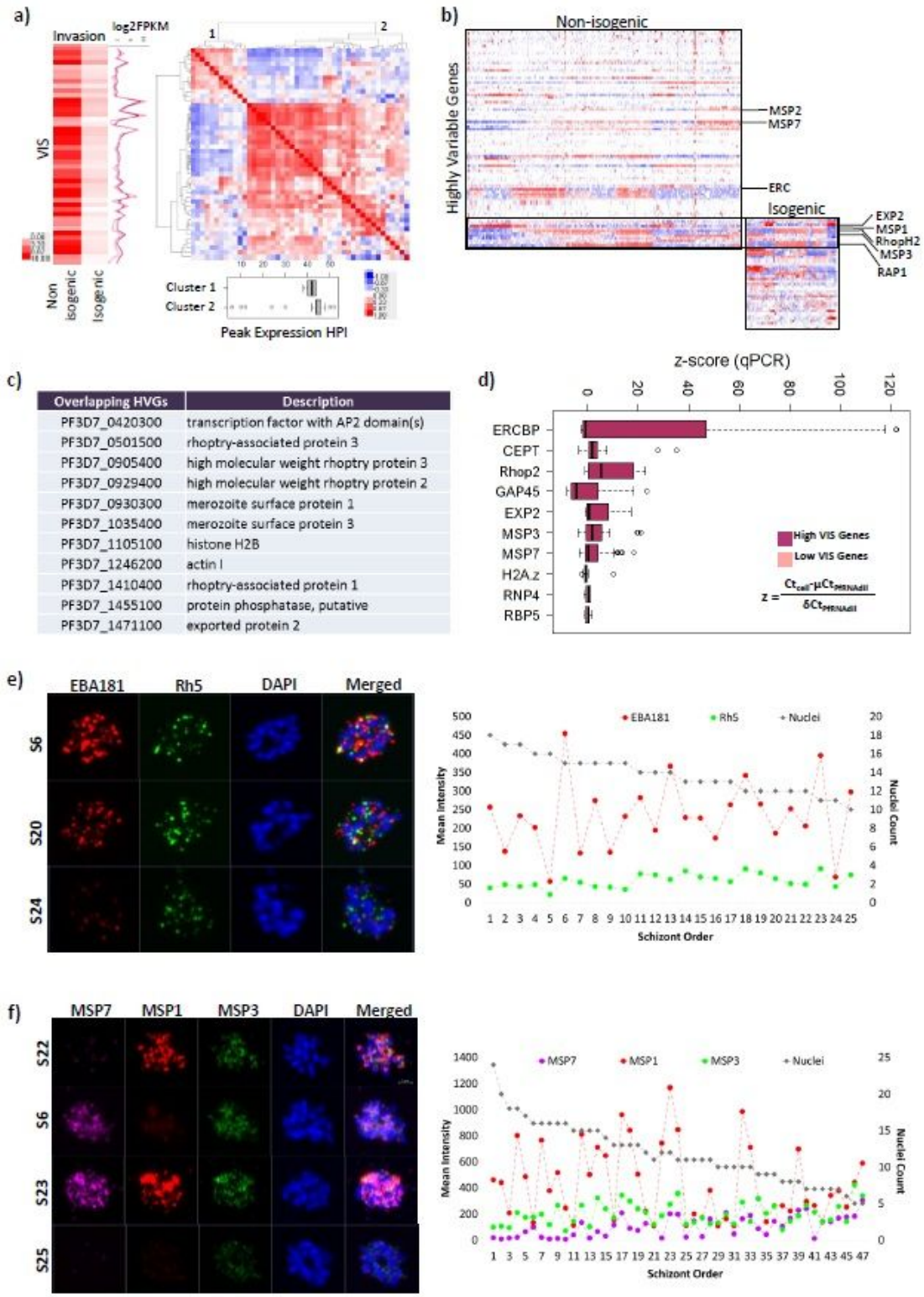


Figure 3

Transcriptional profile of Variant Surface Antigen Encoding Genes. Bubble plots depicting the frequency of expression and average expression (FPKM) for a) var, c) rifin and e) phist genes in both non-isogenic and isogenic schizonts with same gene IDs highlighted in same colour in the two parasite groups. A distribution of number of b) var, d) rifin and f) phist transcripts expressed per cell is shown as a line plot

for non-isogenic and isogenic parasites. Different colours represent the distribution of transcripts expressed above 10 (black), 50 (red) , 100 (green) or 200 (yellow) FPKM in individual schizonts.



Overlapping HVGs	Description
PF3D7_0420300	transcription factor with AP2 domain(s)
PF3D7_0501500	rhoptry-associated protein 3
PF3D7_0905400	high molecular weight rhoptry protein 3
PF3D7_0929400	high molecular weight rhoptry protein 2
PF3D7_0930300	merozoite surface protein 1
PF3D7_1035400	merozoite surface protein 3
PF3D7_1105100	histone H2B
PF3D7_1246200	actin I
PF3D7_1410400	rhoptry-associated protein 1
PF3D7_1455100	protein phosphatase, putative
PF3D7_1471100	exported protein 2

Figure 4

Stochastic Gene Expression in P. falciparum Schizonts. a) A heatmap showing VIS for invasion genes in non-isogenic and isogenic schizonts. Expression of each gene (log2FPKM) is shown as a line graph. Co-expression of invasion genes shown as a heatmap on the right hand side with peak expression of each

gene module (1 and 2) shown as a boxplot (bottom). b) A heatmap showing expression of HVGs (log2FPKM) in non-isogenic and isogenic schizonts. Higher and lower expression (log2FPKM) is shown in red and blue respectively. c) HVGs overlapping between non-isogenic and isogenic schizonts are listed. d) A boxplot showing z-score calculated from C¬t values for each gene transcript measured in 1-schizonts and PfRNAdil by qPCR. Genes with high and low VIS are shown in magenta and pink respectively. e), f) RNA-FISH to measure transcript levels of e) eba181, rh5, and, (f) msp1, msp3, msp7 in individual schizonts. Representative confocal images of single 3D7MR4 schizonts stained for e) eba181 (red) and rh5 (green), and, f) msp1 (red), msp3 (green) and msp7 (purple) transcripts using customised RNA probes are shown on the left. S6, S20, S24 and S6, S22, S23, S25 refers to the schizont order in the line graph for e and f respectively for which the image is shown. Parasite nuclei was stained with DAPI. Scale bar = 5 µm. Line graphs on the right hand side show the mean intensity quantified for each transcript in individual parasites using the ZENlite software.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1DifferentiallyExpressedGenes.xlsx
- SupplementaryTable2MPMPathwaysEnrichment.xlsx
- SupplementaryTable3HighlyVariableGenesNonIsogenicSchizonts.xlsx
- SupplementaryTable4HighlyVariableGenesIsogenicSchizonts.xlsx
- SupplementaryTable5qRTPCRPrimers.xlsx
- SupplementaryFigfinal.pdf