Figures and figure supplements

Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites

Adam J Reid et al
Figure 1. Establishment of a robust protocol for single-cell transcriptomic analysis of Plasmodium parasites. (A) Overview of the single-cell RNAseq protocol. Steps in the original Smart-seq2 protocol (Picelli et al., 2013) that resulted in significant gains are highlighted in orange. (B) Relative numbers of reads mapping to coding RNA and rDNA for our initial sequencing trial, averaged over all cells in that trial (n = 5). (C) The protocol was evaluated using qPCR of the msp-1 transcript (PF3D7_09303000) on sorted pools of 10 asexual parasites (n = 8) (significance from Mann-Whitney test, p<0.05 *p<0.01 **p<0.001 ***). The following reagents were tested: Oligo(dT)s containing a terminal anchoring base (A,G,C, V) or not (T) and of varying lengths (20 Ts vs. 30 Ts); four reverse transcriptase enzymes; 25 or 30 cycles of preamplification. (D) Relative numbers of reads mapping to coding RNA and rDNA for optimisation trials (6, 5, 6, 6 cells, respectively) and the main P. falciparum gametocyte (n = 237), P. berghei mixed blood (n = 182) and P. falciparum asexual (n = 189) datasets (final three bars). Asterisks indicate selected significant differences between proportions of reads mapping to coding genes, calculated using Mann-Whitney U (p<0.05 *p<0.01 **p<0.001 ***).

DOI: https://doi.org/10.7554/eLife.33105.003
Figure 2. Assessment of single-cell transcriptome sequence purity, diversity and accuracy. (A) Individually sorted *P. falciparum* and *P. berghei* cells from a mixed pool revealed no doublets and little contamination. (B) Distributions of numbers of genes identified as expressed in our three main datasets. (C) Expressed genes (those with at least 10 reads in at least five cells) were representative of average gene length, suggesting that although the reverse transcriptase might not copy the whole of long transcripts, fragments of long genes are still detected. (D) Sequencing library preparation often introduces end bias, where either the 5’ or 3’ end of transcripts tend to be better covered. Our protocol introduced a small 5’-bias, which could be attributable to the reverse transcription sometimes initiating within transcripts in internal polyA regions, rather than in the 3’ poly-A tail.

DOI: https://doi.org/10.7554/eLife.33105.005
Figure 2—figure supplement 1. Dual sorting of P. berghei and P. falciparum cells shows that contamination from ambient RNA is low. (A) Purified asexual late blood stage of GFP P. falciparum and mCherry P. berghei were mixed at a 1:1 ratio, inactivated in RNAlater, and sorted individually by flow cytometry, gated on respective fluorescent channels. (B) P. falciparum transcripts contaminating P. berghei cells and (C) P. berghei transcripts contaminating P. falciparum cells suggesting that contaminants reflect leakage from lysed cells (D) Histogram of the number of reads (log-transformed) from P. berghei cells that mapped to genomes of P. berghei and P. falciparum. A small proportion of detected genes were from the incorrect species and these tended to be expressed at low levels. (E) Histogram of the number of reads (log-transformed) from P. falciparum cells that mapped to P. falciparum and P. berghei.

DOI: https://doi.org/10.7554/eLife.33105.006
Figure 2—figure supplement 2. The GC content of transcript fragments agreed well with the GC content of genes. There was no apparent over- or under-representation of GC-rich regions.
DOI: https://doi.org/10.7554/eLife.33105.007
Different cell types were successfully resolved using single-cell transcriptome. (A) A combination of Principal Components Analysis (PCA), k-means clustering and comparison to bulk RNA-seq datasets was used to classify 144 high-quality *P. berghei* single cells, and revealed three distinct subpopulations. Outliers may represent erythrocytes infected with both sexual and asexual stages or early stages in gametocyte development. (B) Three well-established markers of the male, female and asexual lineages (*Mair et al.*, 2006; *Liu et al.*, 2008; *Moss et al.*, 2012) are concordant with our classification.

DOI: https://doi.org/10.7554/eLife.33105.008
Figure 3—figure supplement 1. We detect stage-specific transcripts at a variety of expression levels. Stage-specific genes at different expression levels, were identified from RNA-seq data from (Otto et al., 2014) for (A) asexual stages, (B) male gametocytes and (C) female gametocytes. Mean FPKM between replicate samples is displayed here.
DOI: https://doi.org/10.7554/eLife.33105.009
Figure 3—figure supplement 2. Principal Components Analysis and classification of *P. falciparum* gametocyte cells. (A) A combination of Principal Components Analysis (PCA), k-means clustering and comparison to bulk RNA-seq datasets was used to classify 191 high-quality *P. falciparum* gametocytes. A consensus of clustering and comparison to bulk RNA-seq allowed us to distinguish male gametocytes and female gametocytes. (B) A PCA plot showing single-cell transcriptomes of *P. falciparum* asexual cells annotated using bulk transcriptome datasets from Lopez et al. and Otto et al.

DOI: https://doi.org/10.7554/eLife.33105.010
Figure 4. Single-cell RNA-seq reveals hidden transcriptional variation in the asexual cell cycle. (A) Pseudotime ordering (using [Trapnell et al., 2014]) of the asexual cells in was in close agreement with bulk RNA-seq datasets (predicted stage = consensus; see Materials and methods). (B) Pseudotime ordering (using [Trapnell et al., 2014]) of the 125 P. falciparum late asexual cells was in close agreement with bulk RNA-seq datasets (predicted timepoint from [Otto et al., 2010], predicted stage = consensus; see Materials and methods). (C) Differentially expressed genes (identified using M3Drop [Andrews and Hemberg, 2016]) were clustered along pseudotime revealing groups of genes with abrupt expression profile changes during late asexual cycle. Functional enrichment in the clusters was in agreement with the expected shift from the growing trophozoite to the budding schizont (IMC = Inner Membrane Complex; micronemes and rhoptries are secretory organelles). ‘Hoo’ is the most similar timepoint in development in the Hoo et al. (2016) dataset.

DOI: https://doi.org/10.7554/eLife.33105.011
Figure 4—figure supplement 1. Pseudotime reconstruction of the late asexual trajectory of *P. falciparum*. PCA of 155 *P. falciparum* cells colored by pseudotime (A) or Monocle state (B); identified trajectory branches are displayed as circled numbers 1 and 2. (C) Differentially expressed genes were plotted along pseudotime for cells in the main trajectory (States 1, 3 and 4). The number of genes per cell is displayed on top of the heat map, whilst the pseudotime and the stage prediction (Otto et al., 2010) displayed on the side of the heatmap. The transition between trophozoites and schizonts is associated with a hard transcriptional shift, as seen for *P. berghei*.

DOI: https://doi.org/10.7554/eLife.33105.012
The same subsets of transcripts show different patterns of expression around the end of the asexual cell cycle in conventional bulk RNA-seq data and pseudotime reconstructions of single cell RNAseq data. A shared set of 651 genes identified as following a
sigmoidal expression pattern through the intraerythrocytic developmental cycle (see Materials and methods) are shown in both bulk transcriptome data (Hoo et al., 2016) (A) and single-cell data ordered by pseudotime (B) for P. berghei. A much more dramatic shift in gene expression is observed in the single-cell transcriptome data. A similar pattern is observed between P. falciparum bulk (Otto et al., 2010) (C) and single-cell (D) RNA-seq. In panels b and d, gene expression patterns are mean-normalised l-scran values. Only late stage parasites (grey arrows in bulk reference datasets) are expected to be present in the single-cell datasets.

DOI: https://doi.org/10.7554/eLife.33105.013
Figure 4—figure supplement 3. Recently published low-coverage, high-throughput single-cell RNA-seq data supports our finding of step changes in gene expression in the *P. falciparum* asexual cycle. A heatmap showing logged, mean-normalised expression values for late asexual parasites from *Poran et al., 2017* ordered by pseudotime. Genes were ordered as for *Figure 4—figure supplement 2A* showing that this data also exhibits a discontinuous expression pattern, unlike bulk RNA-seq data of the same part of the cycle.

DOI: https://doi.org/10.7554/eLife.33105.014
Figure 4—figure supplement 4. Analysis of the co-expression pattern of the ApiAP2 family of transcription factors (TFs) in asexual parasites. (A) Expression of Plasmodium ApiAP2 genes in asexual parasites. Orthologous genes are presented on the same rows. (B) A co-expression network for P. berghei was built using significant positive and negative correlations (p<0.05 by Pearson’s correlation) between TFs and weighted according to their correlation coefficient. Numbers on the network refers to the TFs as presented in A.

DOI: https://doi.org/10.7554/eLife.33105.015
Figure 5. After removing the signal of cell cycle progression, we identify a new class of cell-cycle independent variable genes. (A) *P. falciparum* genes with $\geq 50\%$ of their variance attributed to cell-cycle associated latent variable one vary in pseudotime. After removing variation associated with the cell cycle, 56 genes with $\geq 50\%$ of their variance remained. Highly enriched functional terms associated with the two sets of genes are shown. (B) Here, we show that cell-cycle independent variable transcripts have similar half lives to genes in general during the ring and trophozoite stages. However, during the schizont stage and later, they are significantly longer. The data was derived from (Shock et al., 2007). (C) A conservation score, calculated based on mean amino acid substitution between *P. berghei* and *P. falciparum* proteins, was plotted against expression level (scran-l) for each cell-cycle-dependent and each cell-cycle-independent gene in *P. falciparum*. Density plots show the distributions of each of these parameters, highlighting that cell-cycle-independent genes tend to have higher conservation scores, but similar expression levels.

DOI: https://doi.org/10.7554/eLife.33105.016
Figure 5—figure supplement 1. Latent factor analysis of expression variation in cell cycle genes. We found that only the first two latent variables explained at least 5% of variation in cell cycle genes (red line).

DOI: https://doi.org/10.7554/eLife.33105.017
Figure 6. Multigene families show variable expression within and between sexual stages of both *P. berghei* and *P. falciparum*. (a) The heatmap shows gene expression levels for multigene family members differentially expressed between male and female *P. berghei* gametocytes. * gene variably expressed within male (orange) or female (green), *Lpl* = lysophospholipase, *ema1* = erythrocyte membrane antigen 1, (b) Read counts for var mRNAs in *P. falciparum* female gametocyte single cells and female and male gametocyte populations from bulk RNA-seq data. Only reads which spanned the var introns and only genes with at least two such reads were included. There were insufficient male single cells for analysis.

DOI: https://doi.org/10.7554/eLife.33105.018
Figure 6—figure supplement 1. Multigene families show variable expression in sexual stages of both *P. berghei* and *P. falciparum*, respectively. (A) Pir gene expression was highly variable across male gametocytes. In addition, more pir genes were expressed in males than females. These are distinct subfamilies of pir genes from those variably expressed in asexual stages. (B) Read counts for var mRNAs in *P. falciparum* female gametocyte single cells and female and male gametocyte populations from bulk RNA-seq data. Only reads which spanned the var introns and only genes with at least two such reads were included. There were insufficient male single-cells for analysis. For var genes, we highlight the type (e.g. upsB, upsC or var2csa) and whether we found evidence for expression of the mRNA, rather than reads that might represent ncRNAs known to be expressed from these genes.

Epf1 = exported protein family 1, gbp = glycophorin binding protein, emp3 = erythrocyte membrane protein 3.

DOI: https://doi.org/10.7554/eLife.33105.019
Figure 6—figure supplement 2. Analysis of the co-expression pattern of the ApiAP2 family of transcription factors (TFs) in sexual parasites. (A) Expression of *Plasmodium* ApiAP2 genes in sexual parasites. Orthologous genes are presented on the same rows. (B) A co-expression network for *P. berghei* was built using significant positive and negative correlations (p<0.05 by Pearson’s correlation) between TFs and weighted according to their correlation coefficient. Numbers on the network refers to the TFs as presented in A.

DOI: https://doi.org/10.7554/eLife.33105.020