SUPPLEMENTARY MATERIAL

Single-Cell RNA-Seq Defines Transcriptional Heterogeneity Among *Plasmodium falciparum* Blood-Stage Parasites

Mtakai Ngara[1,2,*], Mia Palmkvist[3*], Sven Sagasser[1,*], Daisy Hjelmqvist[2], Åsa K Björklund[1], Mats Wahlgren[3] , Johan Ankarklev[3,4,5❖] and Rickard Sandberg[1,2❖]

Supplementary figure legends

Supplementary Figure 1:

(**A**) Barplots showing the total number of reads mapped to both the *P. falciparum* (Pf3D7 version3 from PlasmoDB database) and human genome (UCSC Human Genome Assembly Release hg19) per sample of the individual iRBCs. Only individual iRBCs that had at least 10,000 uniquely mapped reads and 200 genes detected were included. The X-axis shows bar plots corresponding to individual samples and the Y-axis shows the total number of mapped reads in log-scale. For each barplot, the number of reads mapped to Pf3D7 and hg19 are colored red and green respectively. Out of the 165 Pf-iRBCs 114 had more reads mapped to the Pf3D7 while the other 51 had more mapped to hg19.

(**B**) Barplots showing the mapping rate of the reads per sample of individual iRBCs (iRBCs). Only iRBCs that had at least 10,000 uniquely mapped reads and 200 genes detected (n=165) were included. The mapping was performed towards the aggregated reference genome constituting the *P. falciparum* (Pf3D7 version3 from the PlasmoDB database) and the human genome (UCSC Human Genome Assembly Release hg19). The X-axis represents the samples and the Y-axis represents the proportions (in %) of total reads that were uniquely mapped (red), multi-mapped (green) and un-mapped (blue).

(**C**) Heatmap of the mean correlation (Spearman) between all individual iRBCs aggregates and iRBC populations per sampling time point. On the X-axis is the number of population-level Pf-iRBCs samples from each post-invasion timepoint of sampling represented by the 'n' in brackets. The Y-axis shows the total number of aggregated iRBCs for each sampling time point. The aggregates were generated from the expression profile of iRBCs with at least 200 *P. falciparum* genes detected and 10,000 uniquely mapped reads. The color legend displays the mean correlation between the aggregates and iRBCs populations. 10, 16, 22, 32, 38, 44 h

(**D**) Scatter plot of the first (X-axis) and second (Y-axis) principal components of the iRBC populations belonging to the six sampling time points color-coded: cyan (T1; timepoint 10hr), darkgreen (T2; 16hr), darkblue (T3; 22hr), orange (T4; 32hr), purple (T5; 38hr) and deep-red (T6; 44hr). Percentage explained variance for the first and second principal components are presented in brackets. Principal Component Analysis (PCA) was based on log-transformed expression (RPKM + 1).

(**E**) Heatmap of representative stage-specific iRBC populations (bulk-controls with at least 500 genes detected and 100,000 uniquely mapped reads) *P. falciparum* gene expression profiles. Clustering was performed based on the pairwise Correlation coefficients (Spearman) between the samples using a hierachical approach with complete linkage. The samples have been colored according to the sampling time points color-coded: cyan (T1; timepoint 10hr), darkgreen (T2; 16hr), darkblue (T3; 22hr), orange (T4; 32hr), purple (T5; 38hr) and deep-red (T6; 44hr). The gene expression levels were initially log-transformed (i.e. log2 (RPKM + 1) and only genes detected in at least two samples.

(**F**) A t-SNE scatter plot showing the first (X-axis) and second (Y-axis) dimension based on the first ten principal components scores derived from single-cell Pf-iRBCs (with at least *P. falciparum* 500 genes detected and 10,000 uniquely mapped reads). Principal components cluster analysis of the 92 single-cell individual-iRBCs (iRBCs) is based on log-normalized expression values of *P. falciparum* genes. Only genes detected in at least two SC-Pf-iRBCs and had a standard deviation >=0.3 quantile

probability across the samples were used. The points are colored according to post-invasion sampling time points of the samples namely: cyan (T1; timepoint 10hr), darkgreen (T2; 16hr), darkblue (T3; 22hr), orange (T4; 32hr), purple (T5; 38hr) and deep-red (T6; 44hr).

Supplementary Figure 2:

(A) An integrated view of the reference *P. falciparum* genome and the aligned reads for each of the three single-iRBCs in which var gene switching is possibly taking place. On the top row and highlighted in red are the two genes' (Pf3D7_1240400 and Pf3D7_1240600) regions where reads were uniquely mapped, the second row has the annotation and structure of the genes (in blue), rows 3 to 5 display the reads and coverage of the genes in the three single-iRBCs. On the last row, is a depiction of the uniqeness profile for the genomic region in which the mappable and unmappable loci are colored purple and green respectively.

(B) Barplot showing the *var* gene expression level of the two *var* genes in each of the three single-iRBCs where multiple *var* gene expression is indicated. The X-axis shows the single-iRBCs and the Y-axis shows the expression levels of the two genes (in log(RPKM)).

(C) Barplot showing rt-qPCR validation of the stage specific expression level of the five conserved hypothetical genes in early (red) and mid (green) stage developing gametocytes compared to late stage asexually replicating parasites (blue). On the Y-axis is the relative normalized fold change of expression and the gene IDs are represented on the X-axis.

Figure S1

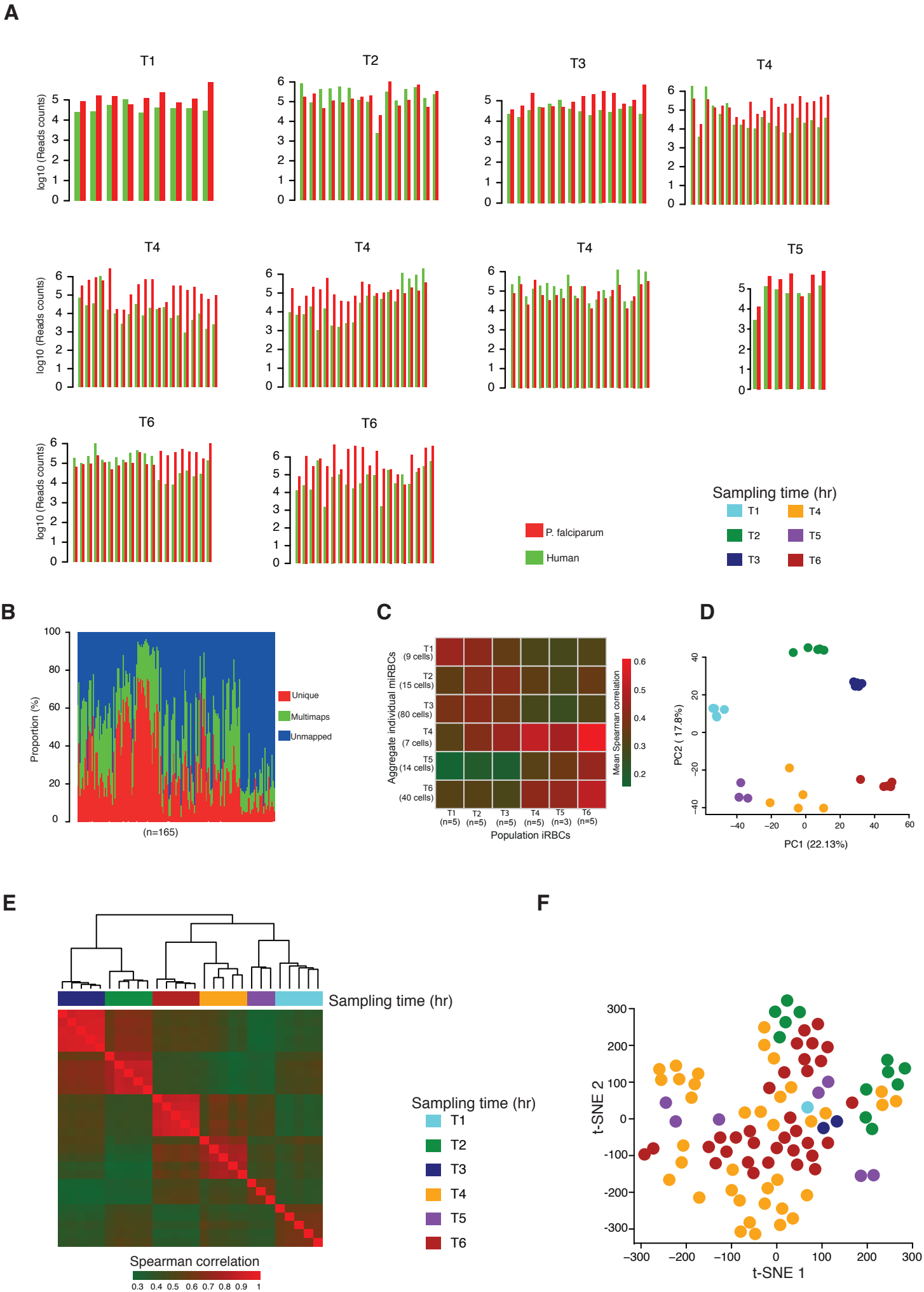
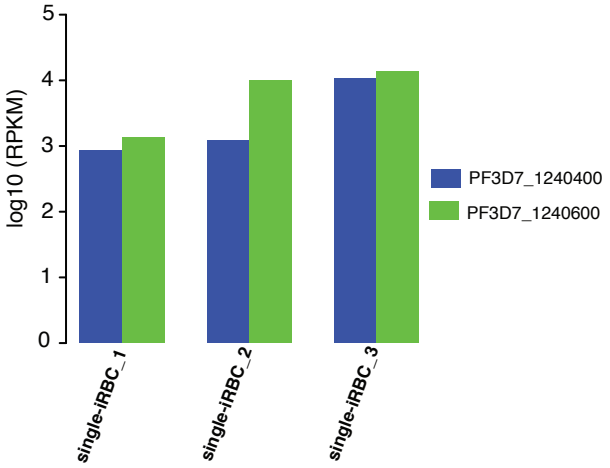

A
T1
T2
T3
T4
T4
T4
T4
T5
T6
T6
log10 (Reads counts)
P. falciparum
Human
Sampling time (hr)
T1
T2
T3
T4
T5
T6
B
Proportion (%)
(n=165)
Unique
Multimaps
Unmapped
C
Aggregate individual miRBCs
T1 (9 cells)
T2 (15 cells)
T3 (80 cells)
T4 (7 cells)
T5 (14 cells)
T6 (40 cells)
Population iRBCs
Mean Spearman correlation
D
PC1 (22.13%)
PC2 (17.8%)
E
Sampling time (hr)
Spearman correlation
F
t-SNE 1
t-SNE 2

Figure S2

A

B



C